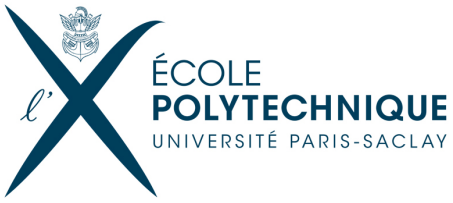
Tau lepton reconstruction and search for Higgs bosons decaying to tau pairs in the CMS experiment at the LHC.

Ivo Nicolas Naranjo Fong



ÉCOLE POLYTECHNIQUE

THÈSE DE DOCTORAT

Présentée par

Ivo Nicolás Naranjo Fong

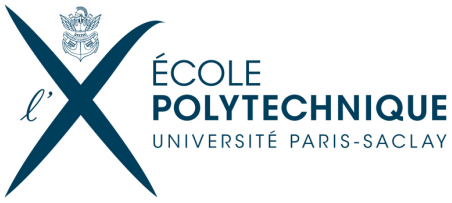
en vue d'obtenir le grade de

DOCTEUR EN SCIENCES

Reconstruction des leptons tau et recherche de bosons de Higgs se désintégrant en paire de leptons tau dans l'expérience CMS auprès du LHC

Soutenue le 1er Octobre 2014 devant le jury composé de :

Dr.	Pierre	Fayet	Président du Jury
Dr.	Suzanne	Gascon	Rapporteur
Dr.	Arnaud	Lucotte	Rapporteur
Dr.	Pascal	Paganini	Directeur de thèse
Dr.	Boris	Tuchming	Examinateur
Dr.	Mónica	Vázquez Acosta	Examinateur



ÉCOLE POLYTECHNIQUE

PHILOSOPHIÆ DOCTOR THESIS

Presented by

Ivo Nicolás Naranjo Fong

Submitted in fulfillment of the requirements for the degree of

DOCTEUR EN SCIENCES

Tau lepton reconstruction and search for Higgs bosons decaying to tau pairs in the CMS experiment at the LHC

Defended on October the 1^{st} 2014 in front of the comitee:

Dr.	Pierre	Fayet	President
Dr.	Suzanne	Gascon	Referee
Dr.	Arnaud	Lucotte	Referee
Dr.	Pascal	Paganini	Supervisor
Dr.	Boris	Tuchming	Examiner
Dr.	Mónica	Vázquez Acosta	Examiner

Abstract

This thesis presents a search for Higgs bosons decaying into tau pairs in the context of the Standard Model (SM) and its minimal supersymmetric extension (MSSM) with the Compact Muon Solenoid (CMS) experiment, using the data collected during the first years of operation at the LHC.

After introducing the theoretical context relevant for the SM and MSSM Higgs bosons searches, the corresponding phenomenological aspects are discussed. The discovery of a scalar boson by the ATLAS and CMS Collaborations in 2012 is then presented. The design and performance of the CMS experiment are described as well as the identification and measurements of physical objects needed for the analysis. The accent is put on the tau lepton reconstruction including in particular a novel tau isolation that uses the lifetime information. We optimized this tool in terms of tau identification efficiency and fake tau rejection, particularly suited for boosted taus. A new technique for rejecting the electrons faking taus is also presented: a discriminator based on a multivariate analysis which considerably reduces the contamination from electrons faking taus while preserving the efficiency of identification is exposed.

The SM and MSSM Higgs boson searches in the di-tau final state are presented. The dataset corresponds to an integrated luminosity of 4.9 fb^{-1} at $\sqrt{s} = 7$ TeV and 19.7 fb^{-1} at $\sqrt{s} = 8$ TeV recorded by CMS respectively in 2011 and 2012. The focus is put on the most sensitive semi-leptonic decay channels: $H \to \tau\tau \to l\tau_h$ where l is either an electron or muon and τ_h stands for hadronic tau decays. After combining all the decay channels in SM Higgs boson to di-tau search and in particular using the new anti-electron discriminator, an excess of observed events with respect to the background-only hypothesis is found. The statistical significance of this excess corresponds to 3.2 standard deviations at a mass of 125 GeV. The measured signal strength is compatible with the SM Higgs signal expectation. This result constitutes an evidence for a coupling between the scalar boson and leptons.

Finally, a new strategy in the search for the MSSM Higgs bosons is presented in this thesis. It combines both the novel tau isolation technique and a new refined categorization based on the τ_h transverse momentum. With respect to the previous CMS public result (2013), the sensitivity of the analysis presented in this document is improved by more than 30%. This is equivalent to an increase of the total integrated luminosity by a factor 3 to 4. The interpretation of the model independent upper limits on the cross-section times branching ratio of the production process shows no significant excess over the Standard Model backgrounds.

Résumé

Cette thèse présente une recherche du boson de Higgs du Modèle Standard (MS) et de son extension supersymétrique minimale (MSSM) dans la voie de désintégration en paires de leptons tau avec l'expérience CMS pendant les premières années d'exploitation du LHC.

Après une introduction du contexte théorique pertinent pour la recherche du boson de Higgs du MS et du MSSM, les aspects phénoménologiques sont exposés. La découverte en 2012 d'un boson scalaire par les collaborations ATLAS et CMS est ensuite présentée. La description du détecteur CMS est abordée et la reconstruction des objets nécessaires a l'analyse est décrite. Une attention particulière est portée à la reconstruction des leptons tau avec notamment un algorithme d'isolation original incluant des informations sur le temps de vie. Nous avons optimisé cet outil en termes d'efficacité d'identification et de réjection de bruit de fond. Une nouvelle technique visant à supprimer des électrons reconstruits a tord comme des taus est également présentée: le développement d'un discriminateur anti-électron, basé sur une analyse multivariée réduit considérablement cette contamination tout en préservant l'efficacité d'identification.

La recherche du boson de Higgs du MS et MSSM dans ses désintégrations en paires de taus est exposée. Les données utilisées correspondent à une luminosité intégrée de 4.9 fb^{-1} avec une énergie dans le centre de masse $\sqrt{s} = 7$ TeV et 19.7 fb^{-1} avec $\sqrt{s} = 8$ TeV collectées par CMS en 2011 et 2012 respectivement. Une attention particulière est portée aux canaux de désintégration semi-leptoniques: $H \to \tau\tau \to l\tau_h$ où l représente un électron ou un muon et τ_h une désintégration hadronique de tau.

La combinaison de tous les canaux de désintégration du Higgs du MS en paire de taus avec en particulier l'utilisation du nouveau discriminateur anti-électron conduit à un excès d'événements par rapport aux bruits de fond attendus. L'interprétation statistique correspond à 3.2 déviations standard pour une masse de 125 GeV et la force du signal est compatible avec le boson de Higgs du Modèle Standard. Ce résultat constitue une indication directe du couplage du boson de Higgs aux leptons.

Finalement, une stratégie originale de recherche de bosons de Higgs du MSSM est présentée. Elle combine d'une part la nouvelle technique d'isolation de tau et d'autre part une définition optimisée des catégories d'analyse fondées sur l'impulsion transverse des taus. La sensibilité de l'analyse est ainsi augmentée de plus de 30% par rapport au dernier résultat public de CMS (2013) ce qui est équivalent à un accroissement de 3 à 4 fois la luminosité totale intégrée. Aucun excès d'événements n'est observé par rapport aux bruits de fonds prédits par le MS suite à l'interprétation modèle indépendante des limites supérieures sur la section efficace fois le rapport d'embranchement pour chacun des modes de production.

Las cosas tienen vida propia, todo es cuestión de despertarle el ánima.

Gabriel García Márquez, *Cien años de soledad*

Remerciements

Je voudrais remercier tous ceux qui m'ont soutenu et aidé pendant ces trois dernières années. Le travail décrit dans cette thèse n'est pas individuel mais le fruit d'innombrables échanges, discussions, apprentissages et encouragements des personnes à mes côtés.

Tout d'abord je voudrais remercier mon directeur de thèse Pascal Paganini. L'autonomie et le soutien que tu m'as accordé ont été indispensables dès le début de mon stage de M2. Je tiens a témoigner de ta disponibilité constante tout le long de ma thèse, tes conseils et tes encouragements, toujours en train de veiller à mettre en valeur mon travail. La mise en place des corrections pour le manuscrit ainsi que la préparation de ma soutenance ont été d'une aide inestimable. Ton ouverture d'esprit lors des moments de doute ont permis de me rassurer à chaque étape de ce long chemin. Merci pour ta confiance.

Je remercie les membres de mon jury de thèse. Suzanne Gascon-Shotkin et Arnaud Lucotte pour avoir accepté de lire ce manuscrit en tant que rapporteurs. Je remercie également Pierre Fayet d'avoir présidé ma soutenance ainsi que Monica Vásquez Acosta et Boris Tuchming d'avoir accepté de juger mon travail.

Je tiens à remercier aussi l'ensemble du Laboratoire Leprince-Ringuet. D'abord un grand merci au service administratif du LLR pour leur bonne humeur. Merci au service informatique pour leur efficacité et disponibilité. Je remercie Yves Sirois, directeur du groupe CMS du LLR pour son soutien et son appui renforcés lors des moments critiques de l'analyse. I want to thank in particular Christian Veelken and Michal Bluj. Thank you for introducing me to the Tau World, your valuable teachings and your support of my work all along my thesis. Je voudrais remercier Florian Beaudette pour les discussions et les conseils très utiles. Je garderai des très bons souvenirs des escapades KFC avec Alex qui m'ont permis de me détendre autour d'un bon sceau d'ailes de poulet :-). Je remercie aussi les autres membres du LLR avec qui j'ai pu partager d'agréables moments, leur convivialité et expérience ont marqué ces trois dernières années: Stephanie, Roberto, Christophe, Philippe, Claude, Thomas.

Olivier, tu m'as offert ton amitié dès ton arrivée. Tes constants coups de main ont permis de mouler le travail qui est présenté dans ce document. Merci pour tes innombrables conseils, relectures, suggestions et corrections lors de l'écriture de ce manuscrit et aussi lors de la préparation de ma soutenance. Ton ouverture et soutien ont su me rassurer dans les (nombreux) moments critiques de cette dernière année. Je t'en remercie et je suis certain que ce travail reste entre de très bonnes mains.

Nadir, avec l'année passée à travailler ensemble nous sommes devenus des compagnons de route et des combattants H2Tau avec Arun :-). Travailler avec toi a été une expérience enrichissante aussi bien du point de vue intellectuel que humain. Ce temps nous a permis

de forger une amitié au delà du simple échange entre thésards; j'espère que le soutien que je t'ai apporté lors de ta rédaction a été à la hauteur de celui que tu m'as accordé. Tu as accepté d'être mon "coach" de thèse, je te remercie pour ta présence lors de mes questionnements, parfois existentiels et pour ton amitié. Tu as été d'une aide morale inestimable.

Yacine , merci pour ta complicité et ton amitié tout le long des trois dernières années et en particulier lors de notre dernière ligne droite. Nous avons pu partager des moments exigeants lors de la rédaction intense à la BNF et la préparation de nos soutenances mais aussi des moments de détente et d'évasion nécessaires dans la vie de thésard: le footing, la muscu, les jam session, les sorties improbables à Paris et à Massy. Je te souhaite que des réussites et je suis certain que nos chemins vont se recroiser dans le futur. J'ai aussi une pensée chaleureuse pour Claudia.

Je remercie aussi notre cher et unique Luca, toujours de bonne humeur ainsi que Hue pour tes blagues incompréhensibles :-). Merci pour les Laser Quest à Massy et des sorties resto à Paris.

Je remercie mes autres camarades docteurs et doctorants. Je pense particulièrement à Lorenzo, merci pour tes conseils et ton aide dans la prise en main des outils d'analyse. Merci aux autres ex-doctorants pour le basket@LLR et les pauses café: Misha, Alice, Benjamin, David, Jonathan, J-P, Helene. Je souhaite une bonne continuation à Nicolas, Simon, Luca C, Thomas, Phillip, Iurii, Dan.

I would like to thank my CERN friends Xavier, Andrés, Somnath, Lucio and Juan for their friendship. Je veux aussi remercier mes amis non parisiens: Raymond pour les sorties à Genève et ailleurs, ainsi que Anthony et Antoine avec qui malgré la distance et le temps écoulé ont su garder une amitié intacte.

También quiero agradecer igualmente a mis amigos hispanohablantes. Gracias Eva por haber sido mi "colloc" durante tres meses y por alegrarme las noches con "Pobre Pablo". Francis, gracias por estar ahí a pesar de la distancia. Juan Pablo, al final no pude ir a verte a Toulouse, me voy a "rattraper" y voy próximamente, gracias por tu amistad incondicional. Y Lombana, gracias por acompañarme en cada paso que doy, por tu amistad y confianza sin limites. Gracias por las navidades y vacaciones juntos hermanos.

Gracias a mi madre, gracias por tu fuerza y consejo que me han guiado desde mi llegada a Europa. Tu apoyo incondicional siempre a estado presente, tanto en las alegrías como en los momentos más difíciles. Tu personalidad y tu corazón han marcado mis pasos y son para mi un ejemplo a seguir. Gracias por tu amor y paciencia durante mis años de estudio.

Agradezco tiernamente a mi familia en Colombia y Ecuador. Particularmente a mi papá, Elo, Luisa, Sarita, mi hermano y Katty. Su apoyo permanente, la educación y el amor que me dan han sido ingredientes determinantes en mi construcción como persona. Gracias por estar siempre presentes demostrando su confianza aún tras mis largos momentos de ausencia durante los últimos años.

Je voudrais remercier aussi ma belle famille qui a su m'accueillir avec leurs bras gros ouverts pendant ces dernières années. Mes passages sporadiques à Camps ont été pour moi des réels souffles de détente.

Finalmente, quiero darle las gracias a Elodie, gracias por soportarme durante estos últimos años, y sobretodo los últimos meses y días ! Tu paciencia sin limites, tu apoyo incondicional, tu personalidad fuerte pero sensible, me han dado fuerza para cumplir todos mis objetivos, a nivel profesional como personal. No me queda mas que agradecerte simplemente por ser quien eres y por brindarme tanto amor.

Contents

Introduction

The Standard Model (SM) of particle physics [1–3] is a theoretical framework that describes successfully the elementary particles and their fundamental interactions. In this gauge theory, the matter fields (quarks and leptons) interact via the exchange of force carriers (vector bosons). Numerous precision measurements in various high energy physics experiments during the last decades confirmed the corresponding SM predictions. One of the Standard Model greatest achievements is the electroweak unification, introduced by Glashow, Salam, Ward and Weinberg [4, 5], where the weak and electromagnetic interactions are described in a common framework. This model predicted the existence of three new vector bosons (the charged $W^{\pm}$ and the neutral Z^0), which were indeed observed at the CERN SPS in 1983. In order to explain massive weak gauge bosons, Brout, Englert, Higgs, Guralnik, Hagen and Kibble in 1964 suggested that their masses are generated through the spontaneous symmetry breaking mechanism [6–10] of the electroweak $SU(2)_L \otimes U(1)_Y$ symmetry group. This mechanism, now called BEH mechanism implies the existence of a scalar boson with $J^{CP} = 0^{++}$, called generally Higgs boson. The fermion masses are then described by Yukawa couplings between the fermions and the Higgs field. Therefore, this last missing piece in the SM is critical in order to confirm the spontaneous symmetry breaking mechanism. On top of that, its properties should be measured in order to test the theory in detail.

The discovery of the first elementary scalar boson was announced on the 4^{th} of July 2012 at CERN by the ATLAS [11] and CMS [12] collaborations, using the proton-proton collisions produced by the Large Hadron Collider (LHC) [13] near Geneva. It led to the 2013 Nobel Prize in Physics rewarding P. Higgs and F. Englert for their theoretical prediction. Since then, several properties related to this new boson have been measured, in order to probe its consistency with the SM Higgs boson. One important property to be established is the existence of Yukawa couplings between the Higgs boson and the elementary fermions (quarks and leptons), in order to generate their mass. The only channel capable to probe the Higgs boson coupling to charged leptons, with the LHC collision data currently available, is the Higgs channel to a pair of τ-leptons.

The SM presents many limitations. It describes only three of the four fundamental interactions as it does not include the gravitational force. The Higgs boson mass is not protected for large radiative corrections in the SM. Besides, the origin of the fermion mass hierarchy, but also the neutrino oscillations and masses, are still unexplained. Another case for new physics is the non baryonic matter (Dark Matter) present in the Universe

which is not predicted by the SM. The SM is believed to be an effective theory at the energy scale of the electroweak symmetry breaking. The presence of an additional symmetry would solve in an elegant way some of these considerations. One of the most attractive theories is Super-Symmetry (SUSY) [14–21]. In particular the minimal Super-Symmetric scenario (MSSM) [22–28], in which a second doublet of scalar fields are added, leading to 5 physical Higgs bosons, two charged $H^{\pm}$, two neutral h, H and one pseudo-scalar A. In this model, the coupling of the Higgs bosons to τ-leptons turns out to be enhanced. This motivates strongly the search for neutral MSSM Higgs bosons decaying to τ-leptons.

This thesis describes my work within the CMS collaboration. After three years of operation, the CMS detector recorded data corresponding to an integrated luminosity of 25 fb^{-1}, at a center-of-mass energy of 7 TeV in 2011 (4.9 fb^{-1}) and 8 TeV in 2012 (19.8 fb^{-1}). This document is divided in three parts. The aim of the first part is to introduce the theoretical framework and the experimental setup. In chapter 1, I introduce the SM formalism and the BEH mechanism as well as a description of the MSSM extension. The second chapter is dedicated to the Higgs boson phenomenology and presents the Higgs bosons searches before the discovery and latest results at LHC. A special emphasis is put in the di-τ final state since due to the high mass of the τ-lepton, it plays a special role in the Higgs boson searches for both SM and MSSM. Chapter 3 describes the LHC complex and the CMS apparatus, placed at the collision point 5 of the LHC ring. The CMS sub-detectors are reviewed; the trigger and data acquisition system is briefly described. Finally, the algorithms used to reconstruct physics objects in CMS, in particular the Particle Flow strategy, are described.

The second part describes the hadronic τ-lepton reconstruction. The τ is the only lepton decaying into hadrons, denoted τ_h hereafter. The τ_h lepton reconstruction, detailed in chapter 4, uses the output of the Particle Flow algorithm to identify the τ_h decay products. An emphasis is put on the development of a novel τ_h isolation using the τ lifetime information, aiming at reducing significantly the overwhelming jet contamination from the LHC proton collisions environment. In addition, as the electron signature in the detector may be misinterpreted as a τ_h, a performant anti-electron discriminator is necessary to reject this background. Chapter 5 focuses on the various discriminators against electrons. The development and validation of a MVA-based anti-electron discriminator is described, where I have provided original contributions.

The third and last part of this thesis is devoted to the search for the SM and MSSM Higgs bosons decaying to τ pairs. In the context of this thesis, I focus on the semi-leptonic channels: $H \to \tau\tau \to \ell\tau_h$, where ℓ is either an electron or a muon and τ_h a hadronic decay of the τ-lepton. First I introduce the tools needed to perform such analysis, in chapter 6. A description of the di-τ mass reconstruction and the statistical tools are presented. In chapter 7, a description of the SM Higgs boson search is detailed. The analysis strategy is presented and the categories aiming to exploit the different production modes are described. My personal contribution to the analysis are highlighted. An important improvement is brought by the new anti-electron discriminator in the $e\tau_h$ channel, due to the large reduction of the $Z/\gamma^* \to ee$ background. The novel τ_h isolation brings a significant improvement in the MSSM Higgs boson search, as the lifetime information is suited for

boosted τ_h topologies. In addition, a strong accent is put on a refined categorization I proposed for the MSSM analysis, leading to an increase in the sensitivity. In addition, a better treatment of the model dependence related to the uncertainty on the Higgs boson transverse momenta in the gluon fusion process is exposed. Finally, all decay channels are combined into a full statistical interpretation of the results. In the MSSM analysis, the interpretation is done by computing the generic model independent upper limits on the cross section times branching ratio for each of the production modes.

Part I

The Electroweak Symmetry Breaking : theoretical framework and experimental searches

Chapter 1

The Standard Model of particle physics

Chapter content

In this chapter we will introduce the theoretical framework of the work presented in this thesis. The Standard Model (SM) of the electroweak and strong interactions of elementary particle physics [1–3] provides so far the best description of the basic constituents of matter and their fundamental interactions. Since its introduction in the late 60s, it has successfully passed numerous stringent tests and has shown an impressive predictive power on the physics at energies below 100 GeV.

In section 1.1 we will do a historical overview of the elementary particles [29], before introducing theß Standard Model lagrangian. The spontaneous symmetry breaking mechanism and the Higgs boson are presented in section 1.2. A brief description on the theory Beyond Standard Model is given in section 1.3, focusing on the Minimal Supersymmetric Model (MSSM).

1.1 The Standard Model of strong and electroweak interactions

1.1.1 Historical overview

During the XIX^{th} century the atom was supposed to be the most fundamental constituent of matter. The electron (e) discovery by Thomson in 1897, the proton (p) by Rutherford and the neutron by Chadwik in 1932 changed this picture. Atoms were indeed composed of these fundamental particles. The photon (γ) was first introduced by Planck and Einstein as the quanta of the electromagnetic field in the early 1900s and its interactions were studied by the Compton scattering experiment in 1923. In the 1920s the basic representation of an interaction existed already for the electromagnetism: matter particles, fermions, interacting through an exchange of a vector boson, here the photon. The relativistic equation of motion for the wave function of the electron was then proposed by Dirac in 1928. This was the origin of quantum electrodynamics (QED) [30, 31].

Then the question of how the same sign protons inside an atom could hold together arised. A new interaction between them, stronger than the electromagnetism, should exist. And which was its vector particle? Yukawa presented the pion (π) as the equivalent of the photon for the nuclear interaction in 1934 and predicted its mass to be between the electron and the proton. So he defined three species of particles depending on their mass: the light particles *leptons* like the electron, the *mesons* like the pion and the *baryons* with higher mass like the neutron and proton. Between 1947 and 1960, a zoology of new mesons and baryons were discovered ($K, \rho, \eta, \Lambda, \Sigma, \Xi$, etc.), until they got classified in the quark model presented by Gell-Mann in 1964. The quarks, which could have three flavors: up (u), down (d) and strange (s) were the fundamental constituents of the hadronic matter: mesons (quark-antiquark) and baryons (three quarks). This model was completed by the experimental discovery of the charm (c) in 1974 as well as the bottom (b) and top (t) in 1977 and 1995 respectively. The observation of the $\Delta^{++}(uuu)$ baryon composed by three up quarks led to introduce new quantum number in order to avoid the violation of the Pauli principle. A new degree of freedom was the introduced, the *color* charge, taking three possible values: red, green and blue. The quarks in the hadrons were interacting

via the strong interaction, associated to the new charge. The vector bosons of the strong interaction are the gluons (g). The strong interaction theory developed in 1973 is then called Quantum ChromoDynamics (QCD) [32–35].

Meanwhile, the lifetime of particles decaying by strong interaction (10^{-23} s) and the lifetime of new hadrons discovered in the 1950s (10^{-10} s) led to think that another interaction was in the game. A weak interaction would be responsible for these decays [36,37]. It was found to be the one explaining the energy spectra of the electrons produced in the β-decay of a neutron to a proton, studied extensively by Fermi in the 1930s. An anti-neutrino ($\bar{\nu_e}$) is additionally produced in this decay carrying part of the energy. Neutrinos, as the other leptons does not interact via the strong interaction, and neither the electromagnetic as being neutrals. In 1983, the mediators of the weak interaction were discovered at CERN, the charged $W^{\pm}$ and neutral Z^0 currents [38,39]. To complete the picture, two other leptons were discovered at very different times in history. The muon (μ) was discovered in 1936 and the tau lepton (τ) in 1975. Each of them is associated with a neutrino ν_μ and ν_τ.

Finally we have the frame of the Standard Model describing three fundamental interactions: electromagnetic, weak (merged into the electroweak interaction) and strong. Each one is associated with a corresponding spin-one vector boson mediating the interaction between the spin-$\frac{1}{2}$ fermion fields. The corresponding physical bosons and associated fields are:

- 8 gluons (g) for the strong interaction;
- the Z^0 and the $W^{\pm}$ for the weak interaction;
- the photon (γ) for the electromagnetism.

The matter is organized in three families of quarks and leptons:

- 3 lepton doublets composed by a charged particle interacting both electromagnetically and weakly (the electron e, the muon μ and the tau τ) and by a neutral particle interacting only weakly (the electron neutrino ν_e, the muon neutrino ν_μ, the tau neutrino ν_τ).
- 3 quark doublets composed by a particle of charge $+2/3$ (up (u), charm (c), top (t)) and a particle of charge $-1/3$ (down (d), strange (s), bottom (b). The quarks are sensitive to the three interactions.

In figure 1.1 the Standard Model particles and their properties are presented. One fundamental particle remains missing in this picture. In 1964 Higgs, Brout, Englert, Guralnik, Hagen and Kibble proposed a mechanism explaining the experimental observation of massive vector bosons [6–10]. Introducing this mechanism in the theory has as a necessary consequence the existence of a new scalar particle called Higgs boson (H). The discovery of a Higgs-like particle was announced by the ATLAS and CMS collaborations in 2012 having a mass around 125 GeV [40,41].

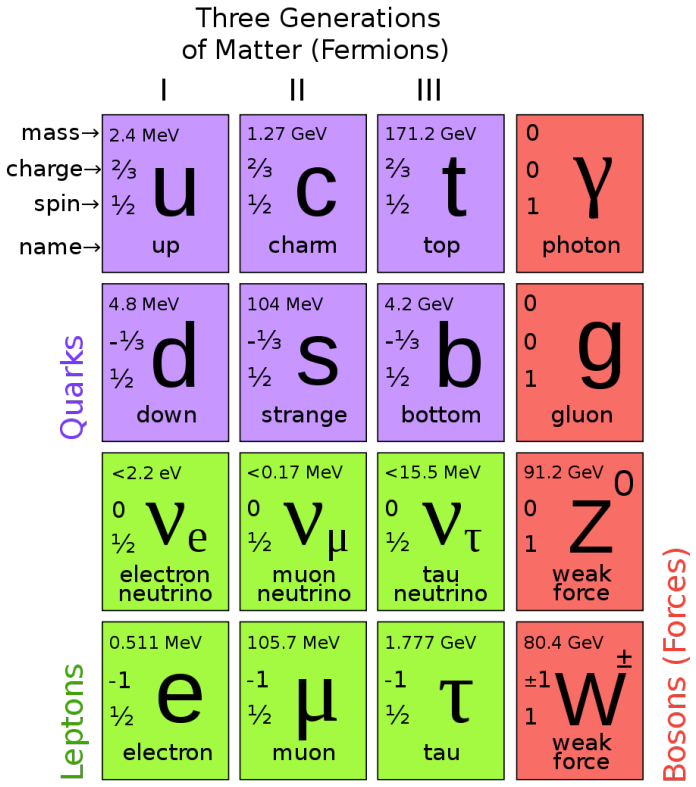


Figure 1.1: *Standard Model particles.*

1.1.2 The Standard Model Lagrangian

The electroweak theory introduced by Glashow, Salam, Ward and Weinberg describes the weak and electromagnetic interactions between quarks and leptons. It is described by the gauge symmetry group $SU(2)_L \otimes U(1)_Y$. The strong interaction, based on the quantum chromodynamics (QCD) theory, is described by the gauge symmetry group $SU(3)_C$. In the Standard Model these three forces are described imposing the invariance of the theory under gauge transformations, and the interaction between two particles of matter are modeled by the exchange of the gauge bosons. In this section, we present briefly the Standard Model of strong and electroweak interactions, a detailed review is presented in Ref. [42, 43].

The Standard Model mathematically describes all the elementary particles using fields which are either scalars for spin-0 particles, bi-spinors for spin-1/2 particles and vectors for spin-1 particles. A gauge theory is a quantum field theory where the Lagrangian is invariant under a group of local transformations of the fields. The gauge transformations, form a Lie group which is the symmetry group or the gauge group of the theory. Gauge fields appear in the Lagrangian imposing the invariance under the local group transformations. This is called (local) gauge invariance. The quanta of the gauge fields are called gauge bosons. In the Standard Model, particles are described by quantum fields, spinors for the fermions and vectors for the bosons. Two basic ingredients to describe the quantum field

theory are:

- The gauge group describing the interactions.
- The representation of the spinors under the gauge group that describe the particles of matter.

The Standard Model is a gauge theory with the symmetry group $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$. The gauge boson associated to the generator of the $U(1)_Y$ group (*hypercharge*) is the field B_μ, the three fields $W_\mu^{1,2,3}$ are associated to the generators of the $SU(2)_L$ group (*weak isospin*) and complete the electroweak sector. For the strong interaction, eight fields $G_F^{1,\dots,8}$ correspond to the generators of the $SU(3)_C$ group (*color*).The strong charge is the *color* and the particle fields interacting strongly are "colored" and carry a quantum number r,b or g (standing for red, blue and green). The field strengths are given by [42]:

$$\begin{aligned} G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2 \epsilon^{abc} W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.1}$$

where g_s and g_2 are the coupling strengths of the strong and the weak force respectively, f^{abc} the tensor of the structure constants of $SU(3)_C$ and ϵ^{abc} the antisymmetric tensor. The non-abelian structure of the $SU(3)_C$ and $SU(2)_L$ groups causes the vector boson self interaction in the equation, the coupling constant g_1 for $U(1)_Y$ appears only in the interactions with fermions.

Electroweak unification

The discovery of neutral currents and the observation of CP violating decays of neutral kaons by Cronin and Fitch suggested that the weak interaction is related to the electromagnetic one. The electroweak unification was proposed by Glashow, Salam, Ward and Weinberg [4,5]. The electroweak theory is based on the symmetry group $SU(2)_L \otimes U(1)_Y$, we can write the covariant derivative:

$$\begin{aligned} D_\mu &= \partial_\mu + i g_1 Y B_\mu + i g_2 \tfrac{\sigma^\alpha}{2} W^\alpha_\mu \\ &= \partial_\mu + \underbrace{i g_1 Y B_\mu + i g_2 \frac{\sigma^3}{2} W^3_\mu}_{\textit{neutral current contribution}} + \underbrace{i g_2 (\frac{\sigma^1}{2} W^1_\mu + \frac{\sigma^2}{2} W^2_\mu)}_{\textit{charged current contribution}} \end{aligned} \tag{1.2}$$

where we can separate the neutral current and the charged current components of the interaction term. The last two terms are represented by the two physical $W^\pm$ bosons and the two first terms represent the electromagnetic field interaction A_μ and a new field Z^0_μ, which was predicted by this formalism. They are linked with the gauge bosons trough the transformation:

$$\begin{pmatrix} A_\mu \\ Z^0_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}$$

where θ_W is the Weinberg angle. We can then write the physical vector bosons in function of the gauge bosons:

$$\begin{aligned} Z_\mu &= -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3 \\ A_\mu &= \cos\theta_W B_\mu + \sin\theta_W W_\mu^3 \\ W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \end{aligned} \tag{1.3}$$

And the relation between the weak and electromagnetic couplings depend on the θ_W angle:

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2+g_2^2}} \quad \sin\theta_W = \frac{g_1}{\sqrt{g_1^2+g_2^2}} \tag{1.4}$$

According to Noether's theorem for every continuous symmetry there is a conserved quantity [44]. So for each symmetry group there is a quantum number which is conserved. The quantum numbers related to the $SU(2)_L \otimes U(1)_Y$ symmetry are the isospin $\overrightarrow{T}$ and the hypercharge Y respectively. Q the electrical charge for the electromagnetism is related to the third component of the isospin and the hypercharge:

$$Q = T_3 + \frac{Y}{2} \tag{1.5}$$

The fermion fields families are arranged in 3 bi-spinor representations per family, the left-handed fermions disposed in weak isodoublets, and the right-handed in weak isosinglets, after applying the left (or right) chiral projection $\psi_{L(R)} = \frac{1}{2}(1 \pm \gamma^5)\psi$. First for the leptonic fields left-handed doublets:

$$T = \frac{1}{2} : \begin{cases} T_3 = +1/2 \\ T_3 = -1/2 \end{cases} \quad L_e = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, L_\mu = \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, L_\tau = \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$

And the right-handed singlets [1]:

$$T = 0 : \qquad e_R \qquad \mu_R \qquad \tau_R$$

And similarly for the quark fields, where we use the weak eigenstates ($d_L^{'}$,$s_L^{'}$,$b_L^{'}$), superpositions of the strong interaction (or mass) eigenstates (d_L,s_L,b_L) following the CKM (Cabbibo, Kobayashi and Mashkawa) matrix [45, 46]. The Standard Model framework describes them by 12 bi-spinor fields per family:

$$T = \frac{1}{2} : \begin{cases} T_3 = +1/2 \\ T_3 = -1/2 \end{cases} \quad Q_1 = \begin{pmatrix} u_L \\ d_L^{'} \end{pmatrix}, Q_2 = \begin{pmatrix} c_L \\ s_L^{'} \end{pmatrix}, Q_3 = \begin{pmatrix} t_L \\ b_L^{'} \end{pmatrix}$$

And the right-handed singlets:

$$T = 0 : \qquad u_R \qquad d_R \qquad s_R \qquad c_R \qquad b_R \qquad t_R$$

[1] The right handed neutrinos do not interact with the gauge fields. They are sterile we will not consider them in the Standard Model Lagrangian.

The lagrangian of the Standard Model

The three generations of fermions can then be referred in a compact way showing their quantum numbers:

$$Q(3,2)_{1/3}; \quad L(1,2)_{-1}; \quad u_R(3,1)_{4/3}; \quad d_R(3,1)_{-2/3}; \quad e_R(1,1)_{-2} \tag{1.6}$$

where the first(second) parameter in the parenthesis is the $SU(3)(SU(2))$ representation and the subindex indicate the hypercharge.

The most general Lagrangian invariant under the symmetry $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$ without mass terms for the fermions and gauge bosons is of the form :

$$\begin{aligned}\mathcal{L}_{SM} = &-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ \bar{L_i}iD_\mu\gamma^\mu L_i + \bar{e_{Ri}}iD_\mu\gamma^\mu e_{Ri} + \bar{Q_i}iD_\mu\gamma^\mu Q_i + \bar{u_{Ri}}iD_\mu\gamma^\mu u_{Ri} + \bar{d_{Ri}}iD_\mu\gamma^\mu d_{Ri}\end{aligned} \tag{1.7}$$

using the covariant derivative:

$$D_\mu = \partial_\mu + ig_1 Y B_\mu + ig_2\frac{\sigma^\alpha}{2}W^\alpha_\mu + ig_s\frac{\lambda_b}{2}G^b_\mu \tag{1.8}$$

σ^α and λ_b are a representation of the generators of $SU(2)_L$ and $SU(3)$ respectively. σ^α are the three Pauli matrices and λ_b the 8 Gell-Mann matrices.

The first line of equation 1.7 represent the dynamic term of the gauge fields (and the non-abelian interaction of gauge fields), and the second line the interaction terms between the gauge fields with the fermions. However it does not include any mass term of the form $m_f\psi\bar{\psi}$, $m_B^2 B_\mu B^\mu$, $m_W^2 W_\mu W^\mu$, $m_g^2 G_F G^\mu$. The gluons are experimentally massless particles ($m_g = 0$), so the $SU(3)_C$ gauge invariance of the strong interaction is unbroken. But if we add mass terms of the form $\frac{1}{2}m_V^2 W_\mu W^\mu$ they will violate the electroweak $SU(2)_L \otimes U(1)_Y$ gauge invariance. We can illustrate this with the example of the photon in QED:

$$\frac{1}{2}m_A^2 A_\mu A^\mu \rightarrow \frac{1}{2}m_A^2(A_\mu - \frac{1}{e}\partial_\mu\alpha)(A\mu - \frac{1}{e}\partial\mu\alpha) \neq \frac{1}{2}m_A^2 A_\mu A^\mu \tag{1.9}$$

Compatible with a massless photon and preserving the $U(1)_Q$ gauge invariance. This is contradictory with the experimental observation of massive $W^\pm$ and Z^0 gauge bosons making the weak a short distance interaction.

In the case of fermions, if we add a mass term of the form $-m_f\psi\bar{\psi}$ we would have, taking for example the electron:

$$-m_e\bar{e}e = -m_e\bar{e}(\frac{1}{2}(1-\gamma_5) + \frac{1}{2}(1+\gamma_5))e = -m_e(\bar{e}_R e_L + \bar{e}_L e_R) \tag{1.10}$$

As e_L belongs to a doublet and e_R belongs to a singlet under $SU(2)_L$ this term is not invariant under isospin symmetry. This shows that we cannot include mass terms in the lagrangian without breaking the electroweak gauge symmetry.

As experimentally we have massive fermions and bosons, we have to find a way to generate the mass terms without violating the $SU(2)_L \otimes U(1)_Y$ gauge invariance. An answer to this question was proposed in the sixties: the spontaneous symmetry breaking mechanism or the Brout-Englert-Higgs mechanism. In the next section we will introduce it in the case of the Standard Model.

1.2 Spontaneous Symmetry Breaking and the Higgs boson

In order to generate the weak vector boson masses, the mechanism of electroweak symmetry breaking, or *Brout-Englert-Higgs mechanism* (BEH) was proposed in the mid sixties by Higgs, Brout, Englert, Guralnik, Hagen and Kibble [6–10]. Similar to the condensed matter phenomena, the symmetry of the theory is spontaneously broken by the vacuum, thus generating in a simple way the gauge boson masses. The simplest way to implement the Higgs mechanism is by introducing a doublet of $SU(2)$ complex scalar fields. The fermion masses are also generated by *ad hoc* Yukawa couplings to the same scalar field and its conjugate. The discovery of this particle is then an essential probe of the electroweak symmetry breaking. In the minimal version of the Standard Model, one single scalar is predicted and the last missing parameter before the Higgs boson discovery is its mass M_H.

The term spontaneous in the symmetry breaking means that the Lagrangian stays invariant under gauge transformations of the symmetry group but the ground state (or vacuum state) is not invariant and breaks the symmetry in the Lagrangian. This phenomena appears for example in ferromagnetism where in a ferromagnetic material the spins of the electrons are oriented in the same direction below a critical temperature T_c. The vacuum state then has to chose spontaneously a random direction.

1.2.1 The Brout-Englert-Higgs mechanism

The spontaneous symmetry breaking trough the BEH mechanism introduces a complex scalar $SU(2)_L$ doublet of the form with hypercharge $Y_\phi = 1$:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 - i\phi_2 \\ \phi_3 - i\phi_4 \end{pmatrix} \tag{1.11}$$

An additional term in the Lagrangian, a potential term $V(\Phi, \Phi^*)$, is responsible for the symmetry breaking:

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger D_\mu \Phi - V(\Phi, \Phi^*), \text{with } V(\Phi, \Phi^*) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.12}$$

When $\mu^2 < 0$ and $\lambda > 0$ we have the non trivial solution where the ground state is not unique but it is located on a continuous ring in a complex plane, as seen in figure 1.2. More appropriately, the figure presents the case of one complex field, whereas in the Standard Model we have two complex scalar fields. The state at $\Phi = 0$ is symmetric, but it is an excited state, once the arbitrary choice of the ground state is done when the $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken.

As the vacuum is electrically neutral the ground state does not have a charged component and the neutral component should have after applying a convenient gauge transformation:

$$\langle \Phi \rangle_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tag{1.13}$$

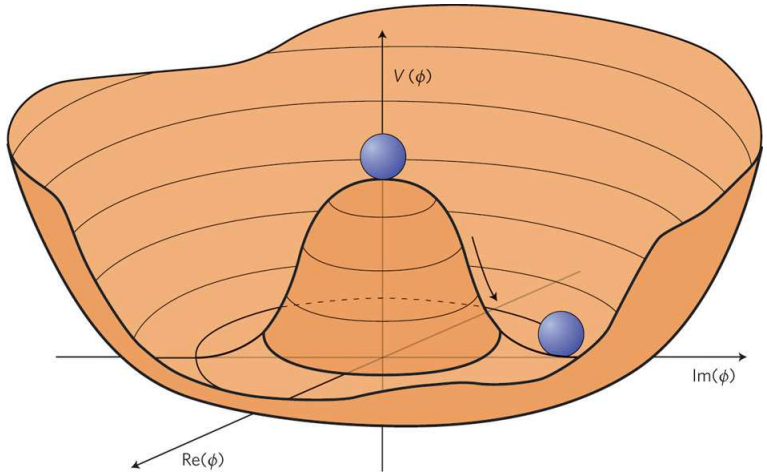


Figure 1.2: *Potential of the complex scalar field $V(\phi)$ for the case $\mu^2 < 0$.*

With the vacuum expectation value (v.e.v.) v obtained by minimizing the potential:

$$v = \sqrt{-\mu^2/\lambda} \tag{1.14}$$

The ground state being neutral it is invariant under a $U(1)_Q$ transformation. Referring to the Nambu-Goldstone theorem [47–50]: for every spontaneously broken symmetry the theory generates massless scalar spin-0 particles (Goldstone bosons), their number is equal to the number of broken generators. The $SU(2)_L \otimes U(1)_Y$ symmetry breaking generates then three degrees of freedom. Expanding around the ground state, with $a = 1...3$:

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{i\theta_a(x)\sigma^a(x)} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.15}$$

where θ_a and h are real scalar fields. We can arbitrary perform a gauge transformation that reabsorb the Goldstone's bosons:

$$\Phi \rightarrow \Phi'(x) = \frac{1}{\sqrt{2}} e^{i(\theta_a(x) - \frac{\alpha_a(x)}{2})\sigma^a(x)} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.16}$$

And chosing $\alpha_a(x) = 2\theta_a(x)$ the field reduces only to a neutral excitation:

$$\Phi'(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.17}$$

After the BEH mechanism, one degree of freedom is left: the quantum of the real field $h(x)$ is associated to the scalar BEH boson, or simply the Higgs boson.

Gauge boson masses and couplings with the Higgs boson

If we inject the field $\Phi'(x)$ in the lagrangian $\mathcal{L}_{Higgs}$ (1.12) and using the covariant derivative for $SU(2)_L \otimes U(1)_Y$ given in (1.2) and we add it to the electroweak gauge

lagrangian $\mathcal{L}^{EW}_{gauge} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$ we have:

$$\begin{aligned}\mathcal{L}^{EW}_{gauge} + \mathcal{L}_{Higgs} =& \frac{1}{2}\left(\partial_\mu h \partial^\mu h + 2\mu^2\right) \\ & - \frac{1}{4}\left(W^\pm_{\mu\nu}\right)^\dagger W^{\pm\mu\nu} + \frac{1}{2}\left(\frac{g_2 v}{2}\right)^2 \left(W^\pm_\mu\right)^\dagger W^{+\mu} \\ & - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{2}\left(\frac{g_2 v}{2\cos\theta_W}\right)^2 Z_\mu Z^\mu \\ & - \frac{1}{4}A_{\mu\nu}A^{\mu\nu} \\ & + \frac{g_2^2 v}{2}hW^-_\mu W^{+\mu} + \frac{g_2^2}{4}h^2 W^-_\mu W^{+\mu} + \frac{g_2^2 v}{4\cos^2\theta_W}hZ_\mu Z^\mu + \frac{g_2^2}{8\cos^2\theta_W}h^2 Z_\mu Z^\mu \\ & + \frac{\mu^2}{v}h^3 + \frac{\mu^2}{4v^2}h^4\end{aligned} \tag{1.18}$$

Applying the Euler-Lagrange equation to the first term we obtain the Klein-Gordon equation for the Higgs boson having a mass:

$$m_H^2 = 2\lambda v^2 = -2\mu^2 \tag{1.19}$$

The second and third lines make appear the mass terms for the $W^\pm$ and Z^0 bosons:

$$\begin{aligned} m_{W^\pm} =& \frac{g_2 v}{2} \\ m_{Z^0} =& \frac{g_2 v}{2\cos\theta_W} = \frac{m_{W^\pm}}{cos\theta_W}\end{aligned} \tag{1.20}$$

There is no mass term for the photon field (A_μ), meaning that we have indeed a massless photon. After the electroweak symmetry breaking the symmetry $U(1)_Q$ emerges. The two last lines in the equation describe the Higgs boson couplings with the gauge bosons ($HWW, HZZ, HHWW$ and $HHZZ$), and the Higgs boson self-interactions (HHH and $HHHH$).

Relating equation 1.20 to the Fermi constant G_F of the effective weak theory, by imposing the equality between transition amplitude according to Fermi theory and the form of Weinberg-Salam model, we have:

$$\frac{G_F}{\sqrt{2}} = \frac{g_2^2}{8m_W^2} \Rightarrow v^2 = \frac{1}{\sqrt{2}G_F} \Rightarrow v \approx 246 \text{ MeV} \tag{1.21}$$

which sets the energy scale of the electroweak symmetry breaking. We have then the gauge vector couplings to the scalar field, proportional to the gauge boson mass:

$$g_{HVV} = \frac{2M_V^2}{v} = 2\left(\sqrt{2}G_F\right)^{1/2} M_V^2 \text{ and : } g_{HHVV} = \frac{2M_V^2}{v^2} = 2\sqrt{2}G_F M_V^2 \tag{1.22}$$

And cubic and quartic self-interaction couplings:

$$g_{H^3} = \frac{3M_H^2}{v} \text{ and : } g_{H^4} = \frac{3M_H^2}{v^2} \tag{1.23}$$

We finally have a theoretical model based on the electroweak symmetry group, including a fundamental scalar field and a mechanism giving rise to the mass of the weak gauge boson and letting the photon massless.

Fermion masses and couplings to the Higgs boson

In order to complete the electroweak lagrangian, we have to introduce the fermion masses through a lagrangian including Yukawa interaction terms. As shown in equation 1.10, the fermion mass term can be decomposed as $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$. A way to couple the left-handed and right-handed states without explicitly breaking the gauge invariance is by introducing the Higgs doublet coupled to the fermions through Yukawa couplings in the lagrangian:

$$\mathcal{L}_{Yukawa} = \sum_{D'}(-g_{D'}\bar{L}_{D'}\phi D'_R - g_{D'}\bar{D'}_R\phi^\dagger L_{D'}) + \sum_{U}(-g_U\bar{L}_U\phi U_R - g_U\bar{U}_R\phi^\dagger L_U) \quad (1.24)$$

Where U and D' are the first and second component of the $SU(2)_L$ doublet. That is to say the charged leptons and u, c, t quarks for one side, and the neutrinos and the d', s', b' quarks on the other side. Here d', s', b' refers to the weak eigenstates.

The introduction of the electroweak symmetry breaking by replacing ϕ by $\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$ 1.17 leads to:

$$\mathcal{L}_{Yukawa} = \sum_{D'}(-m_{D'}\bar{\psi}_{D'}\psi_{D'} - \frac{m_{D'}}{v}\bar{\psi}_{D'}\psi_{D'}h) + \sum_{U}(-m_U\bar{\psi}_U\psi_U - \frac{m_U}{v}\bar{\psi}_U\psi_U h) \quad (1.25)$$

with the relation between the fermion masses and the corresponding Yukawa coupling:

$$m_f = \frac{g_f v}{\sqrt{2}} \quad (1.26)$$

And the coupling between the Higgs boson and fermions is then proportional to the fermion mass:

$$g_{Hff} = \frac{m_f}{v} = (\sqrt{2}G_F)^{1/2}m_f \quad (1.27)$$

The fermion masses depend on the Higgs-fermion Yukawa coupling g_f that can only be obtained from experimental measurements. As the Yukawa couplings do not come from a gauge principle they are a phenomenological model and not a predictive theory.

1.2.2 Theoretical constraints on the Higgs boson mass

The Standard Model theory is valid up to some energy domain, or cut-off Λ_C. Above this energy new physics should appear in order to describe Nature. For example at very large energy scale the theory should take into account gravity. The energy scale limit where the perturbation theory in the Standard Model stay valid can impose constrains on the Higgs boson mass. These constrains can come from unitarity [51–54] in VV scattering amplitudes, perturbativity of the Higgs self-coupling and the vacuum state stability.

If the Higgs boson did not exist the diffusion amplitude of the scattering process $V_L V_L \to V_L V_L$ would diverge, breaking the unitarity condition. If we include the first order radiative corrections, we should include the Higgs boson exchange in the Feynman diagrams. Applying the unitarity condition on the diffusion matrix as showed in [42], leads to the condition on the Higgs boson mass:

$$\frac{M_H^2}{8\pi v^2} < \frac{1}{2} \Rightarrow M_H < 870 \text{ GeV} \tag{1.28}$$

By imposing the theory to stay perturbative up to high values of the energy scale, one can constrain the Higgs boson mass. The $\lambda = M_H^2/v^2$ coupling constant, like the other coupling constants of renormalizable theories [55], also runs with energy scale.

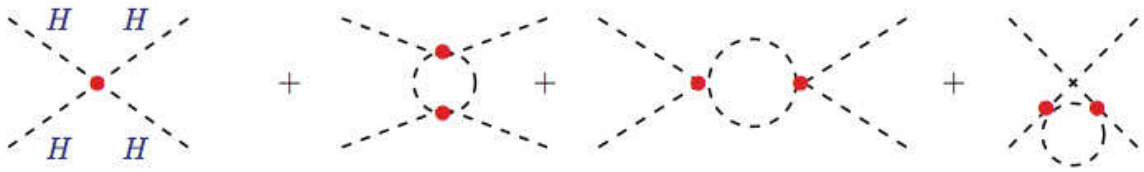


Figure 1.3: *Quartic coupling Feynman diagram at the tree level and for one loop Higgs self coupling. From Ref. [42].*

Considering only the Higgs contribution in the one-loop radiative corrections of the quartic coupling of the Higgs boson showed in figure 1.3, one has :

$$\lambda(Q^2) = \lambda(v^2)\left[1 - \frac{3}{4\pi^2}\lambda(v^2)\log\frac{Q^2}{v^2}\right]^{-1} \tag{1.29}$$

The energy cut-off Λ_C, below which the self-coupling λ remains finite, is the solution at a Landau pole depending on M_H:

$$\Lambda_C = v\exp\left(\frac{2\pi^2 v^2}{3M_H^2}\right) \tag{1.30}$$

This validity limit of the theory, also called triviality condition, defined a region in the plane (Λ, M_H) where the Standard Model works as an incomplete theory. Outside this region, some new physics should restore the perturbativity. This translates in the upper bound limit of the Higgs boson mass.

Otherwise, for small values of λ ($\lambda \ll g_t, g_1, g_2$), the top quark contribution could determine a negative value of $\lambda(Q^2)$. This would violate the vacuum state stability. Hence a lower bound on M_H emerges depending on the cut-off scale value Λ_C.

Figure 1.4 presents the combined lower and upper bounds on the Higgs boson mass M_H as function of the cut-off scale Λ_C. One can see that these constraints imply that if the Standard Model is valid up to the Grand Unification scale, i.e. the energy scale at which a unification of strong, electromagnetic and weak interactions is expected, $\Lambda_{GUT} \sim 10^{16}$ GeV, the Higgs boson mass should stand in the range: 130 GeV $< m_H <$ 180 GeV.

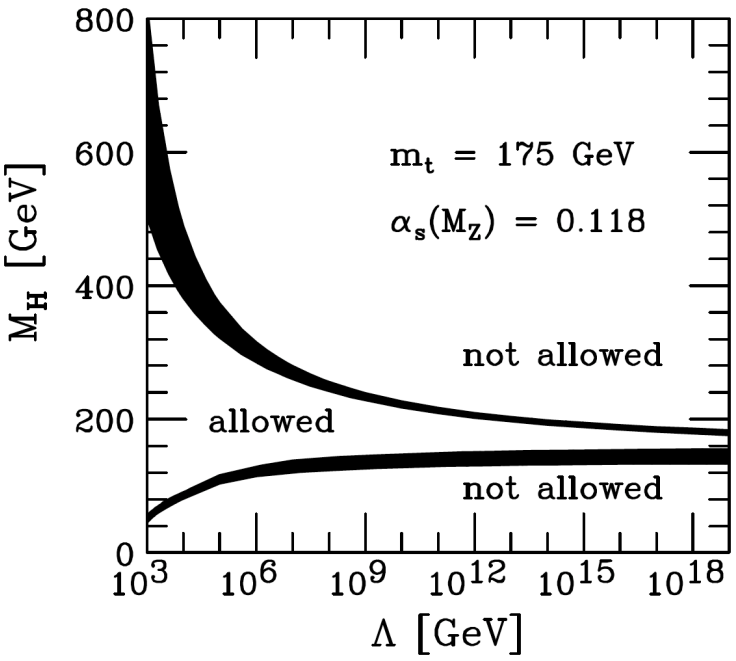


Figure 1.4: *The triviality (upper) bound and the vacuum stability (lower) bound on the Higgs boson mass as a function of the cut-off scale for New Physics. From Ref. [42].*

A Higgs boson mass in the range $124 - 126$ GeV, observed in the recent results by CMS and ATLAS leads to a Higgs potential in the so-called metastable region becoming unstable at an energy scale around 10^{11} GeV. Data favor metastability of the electroweak vacuum but with a lifetime longer than the age of the Universe, resulting on a very small probability of quantum tunneling [56].

1.2.3 Higgs boson decay modes

Assuming the Standard Model, the Higgs boson mass is the unique missing unknown parameter. Once its mass is given its couplings to other particles is determined. Then its decay widths and production modes can be set for a given mass. The SM Higgs boson will couple strongly with the heaviest particles, and decay to the heaviest ones in the phase space allowed by the kinematics. The Standard Model Higgs boson has the quantum numbers $J^{PC} = 0^{++}$. Once the mass is determined, the angular and energy distributions of its decay modes are fixed.

The partial widths for the Higgs boson decays to the following particles can be uniquely predicted [2]:

$$\begin{aligned} M_Z &= 91.187 \text{ GeV}, \; M_W = 80.425 \text{ GeV}, \; M_\tau = 1.777 \text{ GeV}, \; m_\mu = 0.106 \text{ GeV}, \\ m_b &= 4.88 \pm 0.07 \text{ GeV}, \; m_c = 1.64 \pm 0.07 \text{ GeV} \end{aligned} \tag{1.31}$$

[2] We neglect the electron and light quark masses that are too small to be considered.

Fermionic decay modes

As showed in equation 1.27, the Higgs boson coupling to a pair of fermions is at tree level proportional to the fermion mass. Then in the Born approximation the Higgs boson decay width into fermions is given by:

$$\Gamma_{Born}(H \to f\bar{f}) = \frac{G_F N_c}{4\sqrt{2}\pi} M_H m_f^2 \beta_f^3 \tag{1.32}$$

where $\beta = (1 - 4m_f^2/M_H^2)^{1/2}$ is the velocity of the final state fermions and $N_c = 3(1)$ the color factor for quarks (leptons). In the limit where $M_H \simeq 2m_f$, $\Gamma(H \to f\bar{f}) \sim \beta_f^3$ the decay widths presents a strong suppression. In the case of a pseudo scalar A boson ($J^{PC} = 0^{-+}$), the decay width is proportional to β_f instead of β_f^3, the suppression is lighter.

In the case of leptons the Higgs decay to τ leptons is the main decay mode, because of their high mass (1.777 GeV). For the quarks decays, QCD radiative corrections can be large and then should be included. The next-to-leading order (NLO) decay width in the limit $M_H \gg 2m_f$ is given by:

$$\Gamma_{NLO}(H \to q\bar{q}) \simeq \frac{3G_F}{4\sqrt{2}\pi} M_H m_q^2 \left[1 + \frac{4}{3}\frac{\alpha_s}{\pi}\left(\frac{9}{4} + \frac{3}{2}\log\frac{m_q^2}{M_H^2}\right)\right] \tag{1.33}$$

with α_s the strong coupling constant. The logarithmic factor can be large for the light quarks: u, d, s, c, b. For a scale $M_H \simeq 100$, the decay width for the quark $b(c)$ is suppressed by a factor 1.5(2).

Bosonic decay modes

The Higgs boson decay into massive vector bosons $W^{\pm}, Z^0$ is proportional to the HVV coupling:

$$\mathcal{L}(HVV) = \left(\sqrt{2}G_F\right)^{1/2} M_V^2 H V^\mu V_\mu \tag{1.34}$$

The Higgs couplings are therefore proportional to M_V^2. For $M_H < 2M_V$, at least one of the vector bosons is produced off-shell and decays into fermions, giving rise to four body decays. Also, the partial decay width is proportional to G_F^2. Hence it is then suppressed with respect to the dominant low mass decay mode $H \to b\bar{b}$. But there is a compensation due to the fact that the Higgs couplings to W bosons is much larger than the Yukawa couplings to b quarks, becoming the main decay mode when $M_H > 130$ GeV. Here the decay width $\Gamma \to VV^*$ reaches its maximum when the vector bosons are produced at rest in the Higgs frame.

Loop induced bosonic decay modes

The Higgs boson does not couple to photons (γ) and gluons (g) since they are massless. Higgs decays to gluons ($H \to gg$), photons ($H \to \gamma\gamma$), as well as $Z\gamma$ ($H \to Z\gamma$) are done through loops involving massive (charged or colored) particles as showed in figure 1.5. The

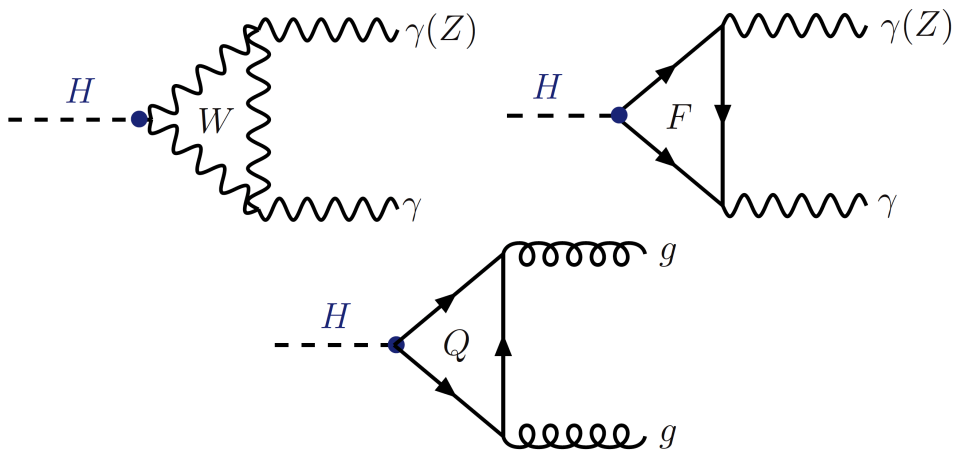


Figure 1.5: *Loop induced decays of the Higgs boson. F denotes a heavy charged fermion while Q is a heavy quark. From Ref. [42].*

presence of new vertices, bringing extra powers in α or α_S suppresses the corresponding decay widths. But this effect is compensated by the Higgs coupling to the heavy particles in the loop and for low M_H this decays become significant.

The decay to two photons is one of the main detection channels of the Higgs boson due to its clear signature. The Feynman diagram contains charged particles loops: quark top and W boson that couples to the photons. The partial decay width in the Born approximation reads:

$$\Gamma_{Born}(H \to \gamma\gamma) = \frac{G_F M_H^3 \alpha^2}{128\sqrt{2}\pi^3} \left| \sum_f N_c Q_f^2 A_{1/2}^H(\tau_f) + A_1^H(\tau_W) \right|^2 \tag{1.35}$$

with $\tau_i = M_H^2/4M_i^2$ and $A_{1/2}^H(\tau_f), A_1^H(\tau_W)$ the form factors for the fermions and vectors respectively.

As for $H \to \gamma\gamma$, the decay $H \to Z\gamma$ is mediated by loops of W and top. The Higgs mass should exceed M_Z to permit this decay. In the limit $M_H \gg M_Z$ the decay width is the same than for $H \to \gamma\gamma$, replacing the photon couplings to the W boson and top quark by the Z ones.

The Higgs boson decay to gluons induces quark loops dominated by top quark and in lesser extent bottom quark contributions. The partial width is:

$$\Gamma_{Born}(H \to gg) = \frac{G_F M_H^3 \alpha_S^2}{36\sqrt{2}\pi^3} \left| \frac{3}{4} \sum_Q A_{1/2}^H(\tau_Q) \right|^2 \tag{1.36}$$

with the quark form factor $A_{1/2}^H(\tau_Q)$ normalized such that it tends to 0 when $\tau_Q \ll 1$.

The total width of the SM Higgs boson as a function of M_H is shown in figure 1.6. The Higgs boson is very narrow in the low mass range ($\Gamma_H < 10$ MeV), growing rapidly for Higgs masses above 130 GeV and reaching ~ 1 GeV at the ZZ threshold. For larger masses (> 500 GeV) the Higgs boson is very large and its width is comparable to its mass.

For $m_H = 1$ TeV it has a total decay width $\Gamma_H \sim 700$ GeV. The width for $m_H = 125$ GeV is ~ 4 MeV.

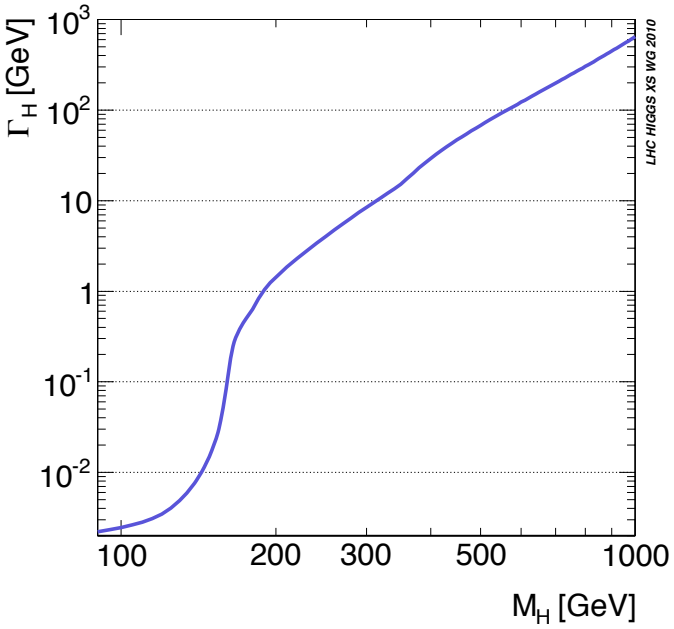


Figure 1.6: *Total width* $\Gamma(H \to anything) = \Gamma_H$ *of the Standard Model Higgs boson in function of its mass* M_H*, From Ref. [57].*

The branching ratio of the channel $H \to X_f$ is given by:

$$BR(H \to X_f) = \frac{\Gamma(H \to X_f)}{\Gamma(H \to anything)} \tag{1.37}$$

In figure 1.7 the branching ratios of the Higgs boson decay are presented in the left plot. Only the channels kinematically permitted and having branching ratios higher than 10^{-4} are plotted. The right plot shows the branching ratio times the production cross section for each mass, for a center-of-mass energy of 8 TeV. These are computed by the *LHC Cross Section Working Group* [57–60].

The "low mass" region, 110 GeV $< M_H <$ 130 GeV, is dominated by the decay $H \to b\bar{b}$ having a branching ratio between 50 and 75%. Followed by the channels $H \to \tau\tau$ (branching ratio between 6 and 9%), $H \to gg$ (branching ratio between 6 and 10%) and $H \to c\bar{c}$ (branching ratio between 2 and 3%). The decays $H \to \gamma\gamma$ and $H \to Z\gamma$ are very rare with branching ratios around the per mille level, and for the decays into muon pair around 10^{-4}. The branching ratio of the channel $H \to WW^*$ increases from 1% at $M_H \sim 100$ GeV to 30% at $M_H \sim 130$ GeV, where the channel $H \to ZZ^*$ reaches the percent level.

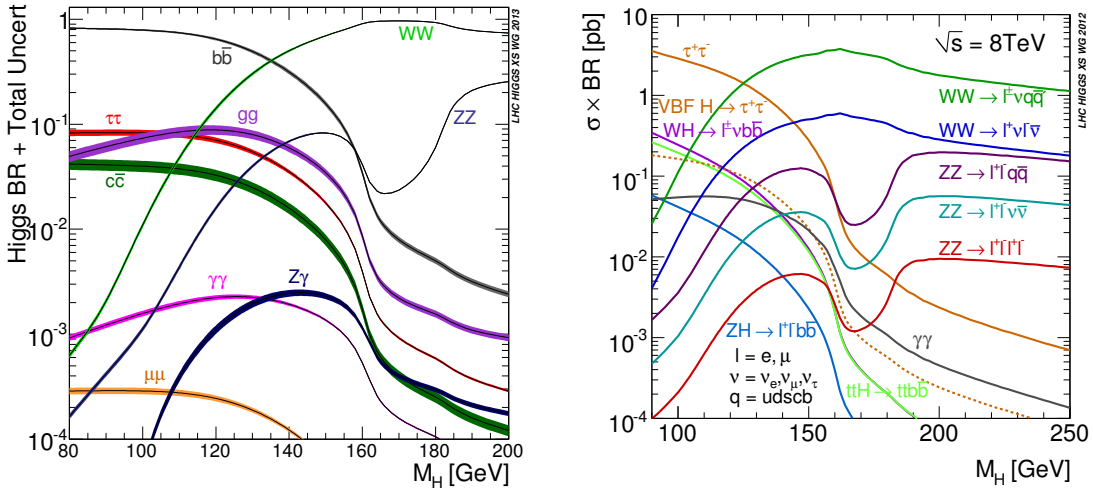


Figure 1.7: *Branching ratios (left) and branching ratios times cross section (right) of the Higgs boson decay in function of the Higgs boson mass M_H with a center of mass energy $\sqrt{s} = 8$ TeV.*

1.3 Beyond the SM

The Standard Model is believed to be an effective theory up to a certain energy scale. It has several theoretical and phenomenological limitations.

As presented before, the Standard Model is based in the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ and describes three over the four fundamental interactions in Nature, it does not include the gravitational interaction. The electroweak unification as presented in section 1.1.2 is achieved, therefore a more general theory is expected to unify also the strong interaction in the so called Grand Unified Theory (GUT), that should be achieved based on a single symmetry group. Experimental data on high precision measurements have shown that at the GUT scale ($\sim 10^{16}$ GeV) the coupling constants does not meet [61–64], raising the unification problem.

As discussed in section 1.2.2 it should exist a cut-off scale Λ from which the perturbative treatment in not valid anymore and new physics should appear. But the radiative corrections to the Higgs boson mass (squared) are quadratic with Λ [65], becoming divergent at high values. For the one loop fermion contribution in figure 1.8:

$$\Delta M_H^2 = N_f \frac{\lambda_f^2}{8\pi^2}\left[-\Lambda^2 + 6m_f^2 \log\frac{\Lambda}{m_f} - 2m_{f^2}\right] + \mathcal{O}(1/\Lambda^2) \tag{1.38}$$

with N_f the number of fermions and $\lambda_f = \sqrt{2}m_f/v$ the Yukawa coupling.

If Λ is chosen to be near the GUT scale or the Planck scale (10^{18} GeV) , then the Higgs boson would tend to be very massive and in order to not violate the unitarity constrain and to be close to the electroweak symmetry breaking scale v, an arbitrary choice of parameters compensating the radiative corrections have to be performed. This is the fine tuning problem where an adjustment of this unnatural choice has to be done with an enormous precision up to $\mathcal{O}(10^{-30})$ [67–69]. The Higgs boson mass has no protection for

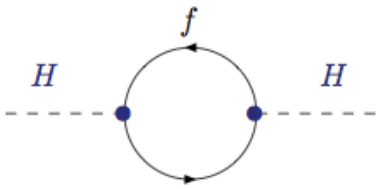


Figure 1.8: *One loop fermion contribution to the Higgs boson mass [66].*

large radiative corrections, unlike fermions and photons that are protected by the chiral and local gauge symmetries. This fine tuning problem is caused by the lack of a new symmetry protecting the Higgs boson mass.

Another reason to call for new physics is the non baryonic matter (Dark Matter) present in the universe (25%) which is not taken into account in the Standard Model. A neutral weak interacting particle candidate is needed and it is not provided in the Standard Model. Also it does not explain the hierarchy of the fermion masses, the electron mass is three orders of magnitude lower than the τ lepton mass. The neutrino ν_e has a mass at least 10^4 lower than the electron one, this is incompatible with the original Standard Model where the neutrinos are massless. The fact that $\Lambda \gg v$ is also not explained in the Standard Model, rising the naturalness or fine tuning problem.

1.3.1 Motivations for supersymmetry

In order to solve the drawbacks enumerated before, one of the most attractive extensions of the Standard Model is Supersymmetry (SUSY) [14–21], which provides a symmetry between bosons and fermions. For each fermion and boson, the corresponding supersymmetric partner , slepton, squark or boson SUSY particle (with suffix *ino*), has the same quantum numbers and mass (if Supersymmetry is not a true symmetry). In figure 1.9 the Standard Model particles and their supersymmetric particles are presented.

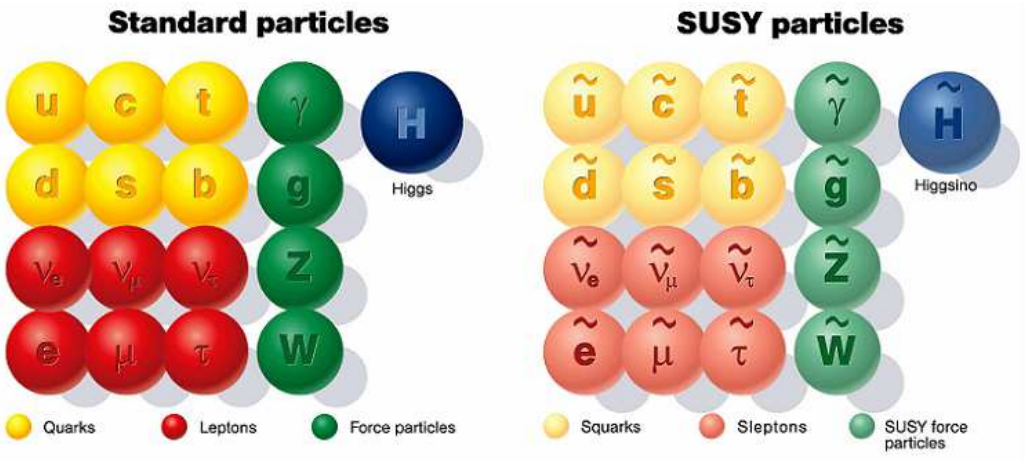


Figure 1.9: *Standard Model particles and their supersymmetric partners.*

Being an extension of the Poincaré group in quantum field theory, it gives a connection to gravity. Hence it is a good candidate to be the fundamental theory of Nature and its interactions. The contribution of the supersymmetric particles correct for the difference in the slopes of the coupling constants and make them intersect at an energy scale about 10^{16} GeV.

If we add new scalar particles, partners of the standard fermions they will contribute to the Higgs boson mass radiative corrections as shown in figure 1.10.

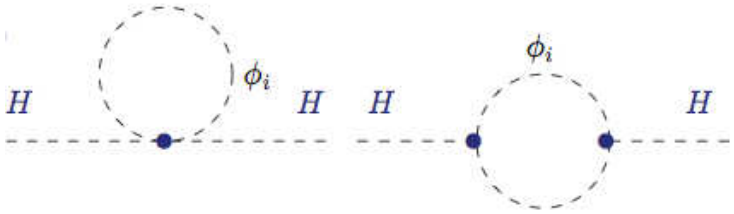


Figure 1.10: *Scalar loops contribution to the Higgs boson mass [66].*

The contribution to the radiative correction from N_S scalar particles with the coupling to the Higgs boson λ_S and a mass m_S is given by [66]:

$$\Delta M_H^2 = \frac{\lambda_S N_S}{16\pi^2}\left[-\Lambda^2 + 2m_S^2 \log\frac{\Lambda}{m_S}\right] - \frac{\lambda_S^2 N_S}{16\pi^2} v^2 \left[-1 + 2\log\frac{\Lambda}{m_S}\right] + \mathcal{O}(1/\Lambda^2) \quad (1.39)$$

By taking the couplings between the Higgs boson and the scalar particles to verify $\lambda_f^2 = -\lambda_S$ and $N_S = 2N_f$ we obtain by adding 1.38 and 1.39:

$$\Delta M_H^2 = \frac{N_f \lambda_f^2}{4\pi^2}\left[(m_f^2 - m_S^2)\log\frac{\Lambda}{m_S} + 3m_f^2 \log\frac{m_S}{m_f}\right] + \mathcal{O}(1/\Lambda^2) \quad (1.40)$$

The quadratic divergences disappear. The additional supersymmetric particles avoid large radiative corrections on the Higgs boson mass by absorbing the quadratic divergences from their standard partners. The Higgs boson mass is protected by this supersymmetry, solving the hierarchy problem. We can note that if the supersymmetric partner has exactly the same mass as the standard particle the Higgs boson mass squared correction vanishes.

Finally SUSY also provides a good candidate for the Dark Matter : the lightest supersymmetric particle, the neutralino is stable in minimal supersymmetric extensions of the Standard Model through the introduction of the R-parity symmetry (ensuring lepton and baryon number conservation).

However, SUSY theory has to be broken as none of the predicted partners of the existing fermions and bosons have been observed. Moreover, by looking at the term in $m_f^2 - m_S^2$ in the radiative corrections in equation 1.40, SUSY particles should not be with a mass higher than ~ 1 TeV in order to preserve the Higgs boson in the electroweak scale and to avoid excessive fine-tuning. This sets a bound on the SUSY scale that can be defined as the geometrical average of the two stop masses:

$$M_S = \sqrt{m_{\tilde{t_1}} m_{\tilde{t_2}}} < 2 \text{ TeV} \quad (1.41)$$

1.3.2 Minimal Supersymmetric Standard Model (MSSM)

The simplest extension of the Standard Model in the Supersymmetry framework is the Minimal Supersymmetric Standard Model (MSSM) [22–28] that we will describe in this section. In this case, the particle masses are obtained by introducing two isodoublets of complex scalar fields of opposite hypercharge. After introducing the electroweak symmetry breaking, five scalar Higgs bosons are generated: two CP-even neutral h and H, a pseudoscalar A boson, and a pair of charged scalars $H^{\pm}$. Only two free parameters are left and are usually taken to be M_A, the mass of the pseudo scalar Higgs boson, and tanβ, the ratio of the two vacuum expectation values. Following, we will describe briefly the Higgs sector in the MSSM context. Details can be found in a review in Ref. [66].

1.3.3 MSSM framework

The Minimal Supersymmetric Standard Model is based on the following hypothesis: minimal gauge group and particle content, Yukawa interactions R-parity conservation and minimal set of SUSY-breaking terms. The MSSM is based on the same geometry group as the Standard Model, $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, ensuring the minimal gauge group and minimal particle content. So each vector boson of the Standard Model is arranged with its corresponding spin-$\frac{1}{2}$ super-partner (the bino $\tilde{B}$, three winos $\tilde{W}_a$ and eight gluinos $\tilde{G}_a$) in vector super-multiplets. The spin-0 quarks and leptons, called squarks and sleptons, are arranged in three generations $(\widehat{Q}, \widehat{U_R}, \widehat{D_R}, \widehat{L}, \widehat{E_R})$. In order to avoid chiral anomalies in the theory two chiral super-fields $\widehat{H_1}$ and $\widehat{H_2}$ are introduced with hypercharge -1 and $+1$ respectively, giving the fermion masses of isospin $-\frac{1}{2}$ and $+\frac{1}{2}$ through their scalar components H_1 and H_2. In tables 1.1 and 1.2 the particle content of the MSSM is summarized.

Superfields	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	Particle content
$\hat{G}_a$	8	1	0	G_a^μ, $\tilde{G}_a$
$\hat{W}_a$	1	3	0	W_a^μ, $\tilde{W}_a$
$\hat{B}$	1	1	0	B^μ, $\tilde{B}$

Table 1.1: *The MSSM super-partners of the gauge and their quantum numbers.*

The Higgs doublet will introduce five scalar Higgs bosons through the electroweak symmetry breaking. Their super-partners, the higgsinos, after mixing with the winos and bino, give raise to the massive eigenstates the charginos $\chi^{\pm}_{1,2}$ and neutralinos $\chi^0_{1,2,3,4}$. In order to preserve the lepton and baryon numbers the conservation of the R-parity is imposed [70]. The R-parity is defined as:

$$R_p = (-1)^{2s+3B+L} \tag{1.42}$$

with L and B the lepton and baryon numbers respectively and s the spin. So that $R_p = 1$ for the Standard model particles and $R_p = -1$ for their supersymmetric partners.

Superfields	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	Particle content
$\hat{Q}$	3	2	$\frac{1}{3}$	(u_L, d_L), $(\tilde{u_L}, \tilde{d_L})$
$\hat{U^c}$	$\bar{3}$	1	$-\frac{4}{3}$	u_R, $\tilde{u}_R^*$
$\hat{D^c}$	$\bar{3}$	1	$\frac{2}{3}$	$\bar{d}_R$, $\tilde{d}_R^*$
$\hat{L}$	1	2	-1	(ν_L, e_L), $(\tilde{\nu_L}, \tilde{e}_L)$
$\hat{E^c}$	1	1	2	$\bar{e}_R$, $\tilde{e}_R^*$
$\hat{H}_1$	1	2	-1	H_1, $\tilde{H}_1$
$\hat{H}_2$	1	2	1	H_2, $\tilde{H}_2$

Table 1.2: *The MSSM fermion super-partners and the Higgs bosons.*

Finally, in order to defined the so-called unconstrained MSSM, a minimal set of SUSY-breaking terms is imposed in the Supersymmetric lagrangian. The terms that break SUSY are the mass terms for the supersymmetric partners (gluinos, winos, binos and sfermions), the Higgs bosons mass terms and the trilinear couplings between the sfermions and Higgs bosons. So the soft SUSY-breaking scalar potential reads:

$$V_{soft} = -\mathcal{L}_{sfermions} - \mathcal{L}_{Higgs} - \mathcal{L}_{trilinear} \tag{1.43}$$

On top of the 19 free parameters from the Standard Model (assuming massless neutrinos), soft-SUSY breaking terms adds $\mathcal{O}(100)$ new parameters, leading to a very complicated treatment and interpretation of the model. Those can be reduced significantly by defining a phenomenological more treatable MSSM by requiring that:

- All the SUSY breaking terms are real, avoiding a new source of CP-violation.
- sfermions and trilinear couplings are diagonal, ensuring the absence of FCNC at the tree level.
- soft-SUSY breaking masses and trilinear couplings are the same for the first and second fermion generations

This defines the "phenomenological" MSSM (*pMSSM*), reducing to 22 the number of input parameters of the theory:

- $tan\beta$: the ratio of the two Higgs doublets vacuum expectation values.
- $m_{H_1}^2, m_{H_2}^2$: the Higgs mass parameters squared.
- M_1, M_2, M_3: the bino, wino and gluino mass parameters.
- $m_{\tilde{q}}, m_{\tilde{u}_R}, m_{\tilde{d}_R}, m_{\tilde{l}}, m_{\tilde{d}_R}$: the first/second generation sfermion mass parameters
- A_u, A_d, A_e: the first/second generation trilinear couplings.

- $m_{\tilde{Q}}, m_{\tilde{t}_R}, m_{\tilde{b}_R}, m_{\tilde{L}}, m_{\tilde{\tau}_R}$: the third generation sfermion mass parameters.
- A_t, A_b, A_τ: the third generation trilinear couplings.

This model is much more easier to investigate than the unconstrained MSSM.

1.3.4 Higgs sector in the MSSM

As pointed out in the previous section, in the MSSM we introduce two doublets of complex scalar fields having opposite hypercharge in order to break the electroweak symmetry:

$$H_1(Y_{H_1} = -1) = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix}, H_2(Y_{H_1} = +1) = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \tag{1.44}$$

The Higgs potential V_H can be decomposed in three parts. First the terms describing the quartic Higgs interactions:

$$V_D = \frac{g_2^2}{8}\left[4|H_1^\dagger \cdot H_2|^2 - 2|H_1|^2|H_2|^2 + (|H_1|^2)^2 + (|H_2|^2)^2\right] + \frac{g_1^2}{8}(|H_2|^2)^2 - (|H_1|^2)^2 \tag{1.45}$$

A term coming from the super-potential:

$$V_F = \mu^2(|H_1|^2 + |H_2|^2) \tag{1.46}$$

And a term coming from the soft SUSY-breaking scalar potential of equation 1.43 scalar Higgs mass and bilinear terms from $\mathcal{L}_{Higgs}$:

$$V_{soft} = m_{H_1}^2 H_1^\dagger H_1 + m_{H_2}^2 H_2^\dagger H_2 + B\mu(H_2 \cdot H_1 + h.c.) \tag{1.47}$$

We have then the Higgs potential, expanding the neutral and charged components using 1.44:

$$\begin{aligned} V_H &= V_D + V_F + V_{soft} \\ &= \bar{m}_1^2(|H_1^0|^2 + |H_1^-|^2) + \bar{m}_2^2(|H_2^0|^2 + |H_2^+|^2) - \bar{m}_3^2(H_1^- H_2^+ - H_1^0 H_2^0 + h.c.) \\ &+ \frac{g_2^2 + g_1^2}{8}(|H_1^0|^2 + |H_1^-|^2 - |H_2^0|^2 - |H_2^+|^2) + \frac{g_2^2}{2}|H_1^{-*} H_1^0 + H_2^{-*} H_2^0| \end{aligned} \tag{1.48}$$

where we have defined the mass squared terms $\bar{m}_1^2 = |\mu|^2 + m_{H_1}^2, \bar{m}_2^2 = |\mu|^2 + m_{H_2}^2$ and $\bar{m}_3^2 = B\mu$.

As in the Standard Model case, we require the minimum of the Higgs potential to break the $SU(2)_L \otimes U(1)_L$ group, while preserving the electromagnetic symmetry $U(1)_Q$.

The neutral components of the Higgs doublets acquire *vacuum expectation values* (vev):

$$\langle H_1^0 \rangle = \frac{v_1}{\sqrt{2}}, \langle H_2^0 \rangle = \frac{v_2}{\sqrt{2}} \tag{1.49}$$

where

$$(v_1^2 + v_2^2) = v^2 = \frac{4M_Z^2}{g_2^2 + g_1^2} = (246 \text{ GeV})^2 \tag{1.50}$$

and the ratio of vevs:

$$\tan\beta = \frac{v_2}{v_1} = \frac{v\sin\beta}{v\cos\beta} \tag{1.51}$$

Applying the minimization conditions leads to the relations:

$$\begin{aligned} B\mu &= \frac{(m_{H_1}^2 - m_{H_2}^2)\tan 2\beta + M_Z^2 \sin 2\beta}{2} \\ \mu^2 &= \frac{(m_{H_2}^2 sin^2\beta - m_{H_1}^2\cos^2\beta)}{\cos^{2\beta}} - \frac{M_Z^2}{2} \end{aligned} \tag{1.52}$$

This shows that if m_{H_1}, m_{H_2} and $\tan\beta$ are known the values of B and μ^2 are fixed, letting the sign of μ undefined, thus lowering the number of free parameters. Then, expanding the two Higgs doublet around the vacuum into real and imaginary parts as for the Standard Model case (1.15):

$$\begin{aligned} H_1 = (H_1^0, H_1^-) &= \frac{1}{\sqrt{2}}(v_1 + H_1^0 + iP_1^0, H_1^-) \\ H_2 = (H_2^+, H_2^0) &= \frac{1}{\sqrt{2}}(H_2^+, v_2 + H_2^0 + iP_2^0) \end{aligned} \tag{1.53}$$

where the real components correspond to the scalar Higgs bosons (CP-even) and the imaginary components correspond to the pseudo scalar Higgs bosons (CP-odd) and the Goldstone bosons. The mass eigenstates are then written:

$$\begin{pmatrix} H \\ h \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} H_1^0 \\ H_2^0 \end{pmatrix} \tag{1.54}$$

$$\begin{pmatrix} G^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} P_1^0 \\ P_2^0 \end{pmatrix} \tag{1.55}$$

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} H_1^\pm \\ H_2^\pm \end{pmatrix} \tag{1.56}$$

where G^0 and $G^\pm$ are the massless Goldstone bosons that are absorbed as in the Standard Model case by a gauge transformation. The corresponding mixing angle β for the CP-odd case and the angle α for the CP-even case are defined as:

$$\cos 2\alpha = -\cos 2\beta \frac{M_A^2 - M_Z^2}{M_H^2 - M_h^2} \text{ , and } \sin 2\alpha = -\sin 2\beta \frac{M_H^2 + M_h^2}{M_H^2 - M_h^2} \tag{1.57}$$

And the CP-even Higgs boson masses at the tree level reads:

$$M_{h,H}^2 = \frac{1}{2}\left[M_A^2 + M_Z^2 \pm \sqrt{(M_A^2 + M_Z^2)^2 - 4M_A^2 M_Z^2 \cos^2 2\beta}\right] \tag{1.58}$$

We see that the supersymmetric structure of the theory imposes strong constraints on the Higgs sector. The Higgs self-interactions are not independent parameters but they can

be expressed in terms of the electroweak gauge coupling constants. As a result, all Higgs sector parameters at tree-level ($M_h, M_H, M_A, M_{H\pm}, \beta$ and α) are determined by two free parameters, which usually are taken to be $\tan\beta$ and M_A. Also the mass spectrum endures hierarchy constrains, $M_H > max(M_A, M_Z)$, $M_{H_\pm} > M_W$ and an important constrain on the lightest h boson at the tree level derived from 1.58:

$$M_h^{tree} \leq min(M_A, M_Z) \cdot |\cos 2\beta| \leq M_Z \tag{1.59}$$

Radiative corrections on equation 1.58 leads to:

$$M_{h,H}^2 = \frac{1}{2}\left[M_A^2 + M_Z^2 + \epsilon \pm \sqrt{1 - \frac{M_A^2 M_Z^2 \cos^2 2\beta + \epsilon(M_A^2 \sin^2\beta + M_Z^2\cos^2\beta)}{((M_A^2+M_Z^2)^2+\epsilon)^2}}\right] \tag{1.60}$$

with

$$\epsilon = \frac{3g_2^2 m_t^4}{8\pi^2 m_W^2}\left(\ln\left(\frac{M_S^2}{m_t^2} + \frac{X_t^2}{M_S^2}\left(1 - \frac{X_t^2}{12M_S^2}\right)\right)\right)$$

where M_S is the arithmetic average of the stop masses $M_S = \frac{1}{2}(m_{\tilde{t}_1} + m_{\tilde{t}_2})$ and X_t the stop mixing parameter defined as $X_t = A_t - \mu\cot\beta$.

We can see that the correction evolves quadratically with the top quark mass and logarithmically with the stop masses, leading to a large radiative correction of several tens of GeV. Then the maximum value of the h mass is $M_h^{max} \sim 140$ GeV, making it kinematically unaccessible during the LEP2 direct searches. Radiative corrections at the one-loop level depend also on the stop mixing parameter maximizing the h mass in the scenario when it reaches the maximum mixing (also called "maximum mixing" scenario). We can then define the 2 scenarios:

$$\begin{aligned} \text{maximal mixing scenario: } & X_t = A_t - \mu\cot\beta \sim \sqrt{6}M_S \\ \text{no mixing scenario: } & X_t = 0 \end{aligned} \tag{1.61}$$

Including the radiative corrections for the other Higgs bosons, in the limit $M_A \gg M_Z$:

$$\begin{aligned} M_H &\to M_A\left[1 + \frac{M_Z^2\sin^2 2\beta + \epsilon\cos^2\beta}{2M_A^2}\right] \\ M_{H^\pm} &\to M_A\left[1 + \frac{M_W^2}{2M_A^2}\right] \end{aligned} \tag{1.62}$$

In this limit the three masses become degenerate and we have $M_H \approx M_{H^\pm} \approx M_A$. In figure 1.11 the MSSM Higgs bosons masses are function of M_A for two different values of $\tan\beta$ (3 and 30) and for the two scenarios described above.

Gauge boson couplings with the MSSM Higgs bosons

The coupling of the Higgs bosons are derived from the kinetic terms of the H_1 and H_2 fields in the lagrangian.

$$\mathcal{L}_{kin} = (D^\mu H_1)^\dagger (D_\mu H_1) + (D^\mu H_2)^\dagger (D_\mu H_2) \tag{1.63}$$

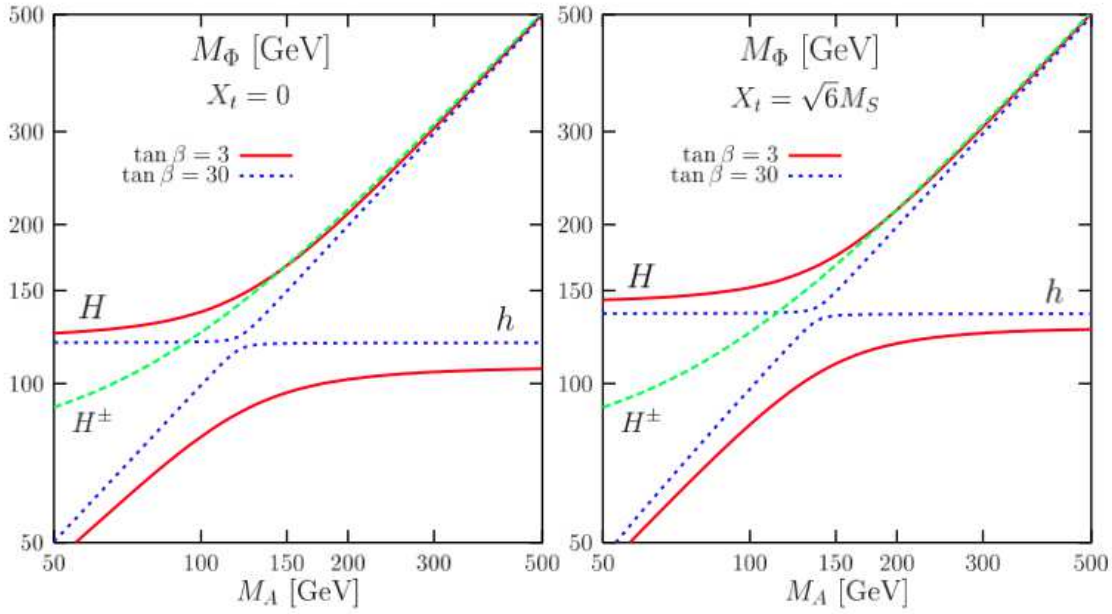


Figure 1.11: *The masses of the MSSM Higgs bosons as function of M_A for two values* $\tan\beta = 3$ *and 30, in the no mixing (left) and maximal mixing (right) scenarios with $M_S = 2$ TeV and all the other SUSY parameters set to 1 TeV. From Ref. [66].*

One can then identify the three types of couplings, the trilinear couplings $V_\mu H_i H_j$ and $V_\mu V_\nu H_i$ and the couplings between two gauge and two Higgs bosons $V_\mu V_\nu H_i H_j$.

In the first case as the photon is massless there is no $H\gamma\gamma$ and $HZ\gamma$ couplings. The couplings of the Higgs bosons h or H to W and Z pairs typically are either proportional to $\cos(\beta-\alpha)$ or $\sin(\beta-\alpha)$. Then the couplings G_{hVV} and G_{HVV} are complementary: $G^2_{hVV} + G^2_{HVV} = g^2_{H_{SM}VV}$, their sum of their squares is equal to the square of the Standard Model Higgs boson.

For the second kind of coupling (between one gauge boson and two Higgs bosons) CP-invariance requirement avoids Zhh, ZHh, ZHH and ZAA as the two Higgs bosons should have opposite parity.

Yukawa couplings with the MSSM Higgs bosons

Supersymmetry imposes that the two Higgs doublets H_1 and H_2 generate the masses of the fermions with respectively weak isospin $-\frac{1}{2}$ and $+\frac{1}{2}$. The Yukawa lagrangian part of the super potential can be written as function of the scalar fields H_1 and H_2 and decomposed as for the Standard Model using the left and right projection operators ($P_{L/R} = \frac{1}{2}(1 \mp \gamma_5)$).

$$\mathcal{L}_{Yukawa} = -\lambda_u[\bar{u}P_L u H_2^0 - \bar{u}P_L d H_2^+] - \lambda_d[\bar{d}P_L d H_1^0 - \bar{d}P_L u H_1^-] + h.c. \tag{1.64}$$

We can then derive the Yukawa couplings achieved when the neutral components of the

Higgs fields reach their corresponding vev v_1 and v_2, for the first fermion family:

$$\begin{aligned} \lambda_u &= \frac{\sqrt{2}m_u}{v_2} = \frac{\sqrt{2}m_u}{v\sin\beta} \\ \lambda_d &= \frac{\sqrt{2}m_d}{v_1} = \frac{\sqrt{2}m_d}{v\cos\beta} \end{aligned} \tag{1.65}$$

Replacing the H_1 and H_2 fields as function of the physical fields we finally get the Yukawa lagrangian in terms of the fermion masses:

$$\begin{aligned} \mathcal{L}_{Yukawa} = &- \frac{g_2 m_u}{2M_W \sin\beta}[\bar{u}u(H\sin\alpha + h\cos\alpha) - iu\gamma_5 uA\cos\beta] \\ &- \frac{g_2 m_d}{2M_W \cos\beta}[\bar{d}d(H\cos\alpha - h\sin\alpha) - id\gamma_5 dA\sin\beta] \\ &+ \frac{g_2}{2\sqrt{2}M_W} V_{ud}\{H^+\bar{u}[m_d\tan\beta(1+\gamma_5) + m_u\cot\beta(1-\gamma_5)]d + h.c.\} \end{aligned} \tag{1.66}$$

We can note the importance of the $\tan\beta$ parameter. For values of $\tan\beta > 1$ the A and $H^\pm$ couplings to down-type fermions is enhanced and the couplings to up-type fermions is suppressed. Then the couplings of the Higgs bosons to b quarks ($\propto m_b\tan\beta$) is strongly enhanced for large $\tan\beta$ values and the coupling to top quarks ($\propto m_t/\tan\beta$) becomes weak. This justifies the search of the MSSM Higgs bosons with the production mode associated to b quarks, and the decay to tau leptons presented in this thesis. The Higgs boson search in the channel $H \to \tau\tau$ is one of the most sensitive channels to explore the Higgs sector in the MSSM.

In table 1.3 we summarize the couplings of the neutral MSSM Higgs bosons normalized to the Standard Model Higgs couplings. The fermionic couplings are normalized to $g_{H_{SM}ff} = [\sqrt{2}G_F]^{1/2}m_f$, the coupling to two bosons is normalized to $g_{H_{SM}VV} = 2[\sqrt{2}G_F]^{1/2}m_V^2]$. The coupling between two Higgs bosons and one gauge boson normalized to $g_W = [\sqrt{2}G_F]^{1/2}m_W$ for $g_{\Phi H^\pm W^\mp}$ and normalized to $g_Z = [\sqrt{2}G_F]^{1/2}m_Z$ for $g_{\Phi AZ}$.

Φ	$g_{\bar{u}u}$	$g_{\bar{d}d}$	$g_{\Phi VV}$	$g_{\Phi AZ}$	$g_{\Phi H^\pm W^\mp}$
H_{SM}	1	1	1	0	0
h	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\sin(\beta-\alpha)$	$\cos(\beta-\alpha)$	$\mp\cos(\beta-\alpha)$
H	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\cos(\beta-\alpha)$	$-\sin(\beta-\alpha)$	$\pm\sin(\beta-\alpha)$
A	$\cot\beta$	$\tan\beta$	1	0	1

Table 1.3: *The MSSM neutral Higgs bosons couplings to fermions and gauge bosons normalized to the Standard Model Higgs boson couplings (see text). From Ref. [66].*

Theoretical bounds on $\tan\beta$

In general $\tan\beta$ is a free parameter of the theory, but its value can be constrained when the perturbativity of the Higgs boson couplings to fermions is required. As shown in table 1.3 and from equation 1.66, the couplings of the pseudo scalar boson A and the charged Higgs bosons to spin $+\frac{1}{2}$ $(-\frac{1}{2})$ are proportional to $\cot\beta$ $(\tan\beta)$.And for large (small) values of M_A the same condition applies to the h (H) boson. Requiring the Yukawa coupling to be smaller enough ($\lesssim \sqrt{4\pi}$) imposes the condition $0.3 \lesssim \tan\beta \lesssim 150$.

The requirement in constrained MSSM models of universal boundary conditions at the GUT scale leads to a stronger condition of $\tan\beta$:

$$1 \lesssim \tan\beta \lesssim m_t/m_b \tag{1.67}$$

For a value of the SUSY scale of $M_S \sim 1$ TeV, requiring the unification of the Yukawa couplings at the GUT scale and taking $m_t \sim 175$ GeV and $m_b \sim 4.25$ GeV one obtains the bounds $1.5 \lesssim \tan\beta \lesssim$ 50-60.

MSSM Higgs bosons decay width and branching ratios

The width and branching ratios of the 4 Higgs particles can be computed for the "maximal mixing" scenario defined in the formula 1.61. In the figure 1.12 the total width of the four Higgs bosons (h, H, A and $H^\pm$) are displayed as function of their masses for two values of $\tan\beta$ (3 and 30).

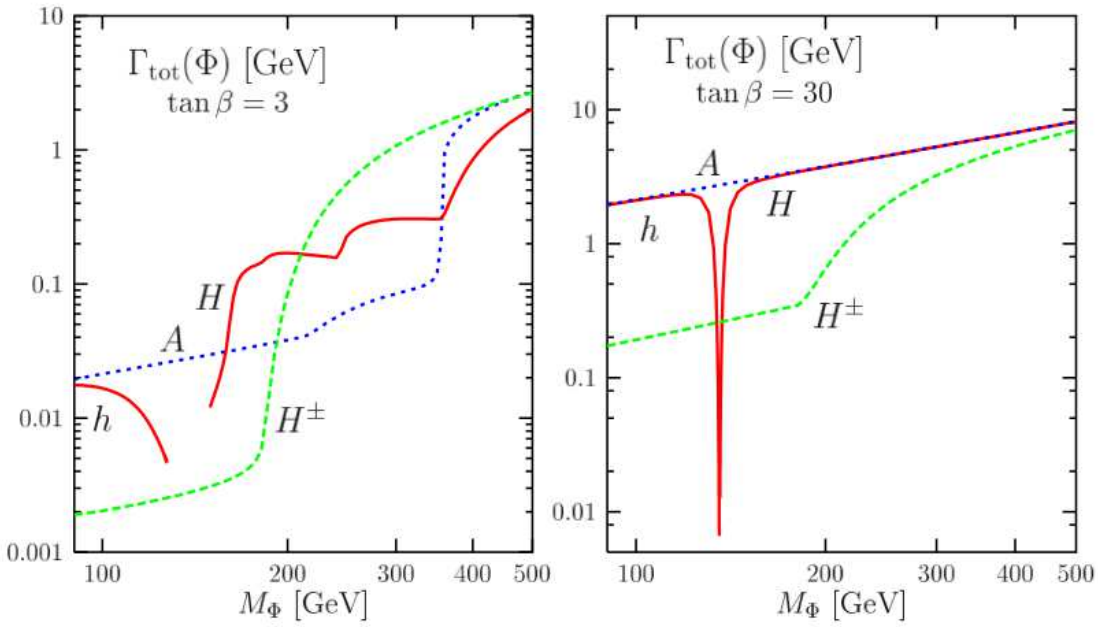


Figure 1.12: *Total width of the four Higgs bosons (h and H in red, A in blue and $H^\pm$ in green) are displayed in function of their masses for two values of* $\tan\beta$ *(3 and 30). From Ref. [66].*

In the figure 1.13 the branching ratios for the lighter CP-even h boson, the heavier

CP-even H boson and the CP-odd A boson are shown as function of their mass for two values of $\tan\beta$ (3 and 30).

1.3.5 MSSM regimes

In this section we will describe the MSSM regimes following the different values of the theory parameters. The most interesting being the decoupling regime, where there is only one Standard Model-like Higgs boson and four heavy Higgses.

The decoupling regime

This regime is reached when there is only one light Higgs boson h and the others Higgs fields are very massive and degenerate: $M_Z \ll M_A \sim M_H \sim M_{H^\pm}$. As described in the previous section it is reached when $M_A \gg M_Z$ and for large $\tan\beta$ values (typically $M_A \gtrsim 300$ GeV and $\tan\beta \gtrsim 10$). This regime is reached at $M_A \gtrsim 150$ GeV for $\tan\beta = 3$ or $M_A \gtrsim$ 400-500 GeV for $\tan\beta = 30$. The couplings of the CP-even h boson to the Standard Model gauge bosons results on the ones expected for the Standard Model Higgs boson. As shown previously, the couplings g_{hVV} and g_{HVV} are complementary and then:

$$\begin{aligned} g_{HVV} &= \cos(\beta-\alpha) \to 0 \\ g_{hVV} &= \sin(\beta-\alpha) \to 1 \end{aligned} \tag{1.68}$$

Also for the fermion couplings, we found for $M_A \gg M_Z$ and $\tan\beta \gg 1$:

$$\begin{aligned} g_{huu} &\to 1 \\ g_{hdd} &\to 1 \\ g_{Huu} &\to -\cot\beta \\ g_{Hdd} &\to \tan\beta \end{aligned} \tag{1.69}$$

The couplings of the lightest Higgs boson approaches the ones of the Standard Model. These limits are reached quicker for bigger values of $\tan\beta$. Hence, the MSSM reduces to the Standard Model phenomenology with only one light Higgs boson h. The heavier neutral bosons are degenerate in mass $M_A \sim M_H$ and their couplings to gauge bosons almost vanish and their couplings to fermions are (inversely) proportional to $\tan\beta$ for the isospin $(+\frac{1}{2})$ $-\frac{1}{2}$. That is why this regime is called decoupling regime: the heavier Higgs bosons decouple and the MSSM sector is reduced to the Standard Model Higgs sector with a Higgs boson having a mass $m_h \lesssim 140$.

The total width is small, $\Gamma_h \lesssim \mathcal{O}(10$ MeV), comparable to the one of the Standard Model Higgs boson. The dominant decay modes into $b\bar{b}$ and WW^* are dominant, followed by the decays into $\tau^+\tau^-, gg, c\bar{c}$ and ZZ^* with branching ratios of the order of the percent level. Finally the decays into $\gamma\gamma$ and $Z\gamma$ have branching ratios at the order of the per mil level.

The decay modes of the heavier Higgs boson depends on the value of $\tan\beta$. For high values, $\tan\beta \gg 1$, the couplings to down type fermions is strongly enhanced, resulting in a decay of the H and A bosons almost exclusively to $b\bar{b}$ ($\sim 90\%$) and $\tau^+\tau^-$ ($\sim 10\%$). The

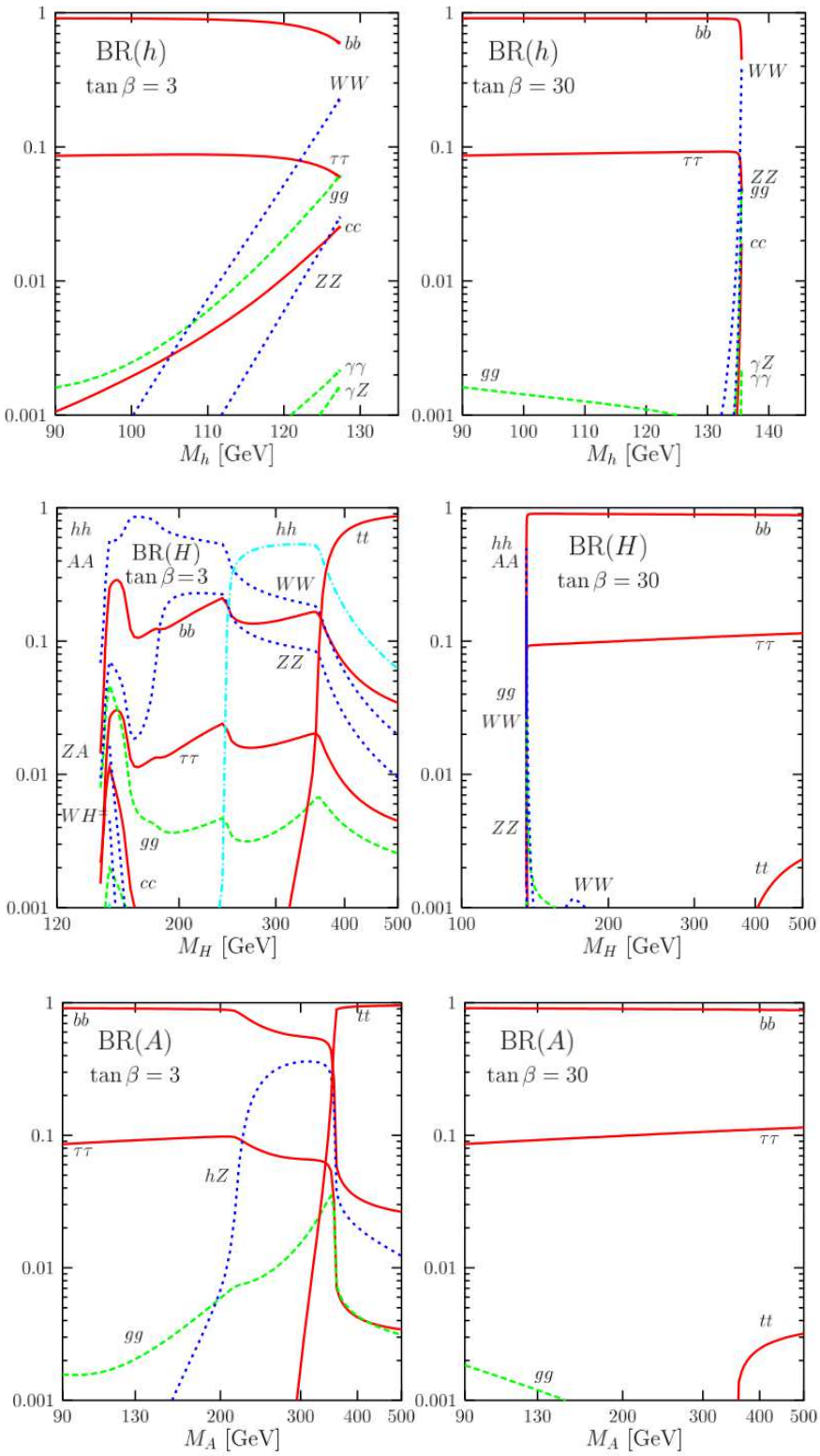


Figure 1.13: *Branching ratios for the lighter CP-even MSSM h boson (top), heavier CP-even MSSM H boson (middle) and CP-odd MSSM A boson(bottom) bosons in function of its mass for* $\tan\beta = 3$ *(left) and* $\tan\beta = 30$ *(right). From Ref. [66].*

decay into $t\bar{t}$ is achieved when allowed kinematically and the other decays are strongly suppressed. For low values of $\tan\beta$ the decay to top pairs $t\bar{t}$ is dominant. The charged Higgs $H^{\pm}$ decays dominantly into tb pairs and for large $\tan\beta$, a fraction of cases to $\tau\nu_{\tau}$ ($\sim 10\%$).

The anti-decoupling regime

This regime is achieved when the pseudo scalar Higgs mass is low, $M_A \ll M_Z$. Here the CP-even light Higgs boson has a mass $M_h \simeq M_A \cos 2\beta$ and the heavier CP-even Higgs has a mass $M_H \simeq M_Z(1 + \sin^2 2\beta/M_Z)$. The phenomenology is the opposite as in the decoupling regime. For high values of $\tan\beta$ the light boson is degenerate in mass with A ($M_h \approx M_A$) and the H boson is degenerate in mass with Z. The roles of H and h are reversed compared to the decoupling regime. In this regime $\cos(\beta-\alpha)$ approaches 1 while $\sin(\beta-\alpha)$ is small. We have then for $M_A \ll M_Z$:

$$\begin{aligned} g_{huu} &\to \cot\beta \\ g_{hdd} &\to -\tan\beta \\ g_{Huu} &\to 1 \\ g_{Hdd} &\to 1 \end{aligned} \tag{1.70}$$

So for large $\tan\beta$ the H boson has couplings similar to the Standard Model case, the light Higgs h is degenerate in mass with the pseudoscalar A and have very suppressed couplings to the gauge bosons and isospin $+\frac{1}{2}$ fermions while its couplings to isospin $-\frac{1}{2}$ fermions are enhanced.

The intense coupling regime

When the pseudo scalar Higgs mass is near the Z mass, the three neutral Higs bosons are degenerate in mass $M_A \sim M_H \sim M_h \sim M_h^{max}$. The degeneracy is enhanced for large $\tan\beta$ values. In this regime we have $\cos^2(\beta-\alpha) \sim \sin^2(\beta-\alpha) \sim \frac{1}{2}$. And thus for the couplings:

$$\begin{aligned} g_{HVV} \simeq g_{hVV} \simeq g_{huu} \simeq g_{Huu} &\simeq \frac{1}{\sqrt{2}} \\ g_{Hdd} \simeq g_{hdd} &\simeq \tan\beta \end{aligned} \tag{1.71}$$

The intermediate coupling regime

This regime happens for low values of $\tan\beta$ ($\lesssim 3-5$) and for moderate masses of the pseudo scalar Higgs boson ($M_A \lesssim 300-500$ GeV). The couplings of the neutral Higgs bosons to isospin $(+\frac{1}{2})$ $-\frac{1}{2}$ fermions are not strongly suppressed (enhanced) because $\tan\beta$ has low values. In this regime H, A and $H^{\pm}$ are relatively heavy but they still couple to gauge bosons and up-type fermions.

Vanishing-coupling regime

The last regime is the one where $\tan\beta$ reaches large values and the pseudo scalar Higgs boson has intermediate to large values. It is possible that combining these parameters with the other parameters in the MSSM that enter the radiative corrections to have a suppression of the couplings to fermions or gauge bosons on one of the CP-even Higgs. This results from the cancellation of the tree-level terms with the radiative corrections.

Summary

We can arbitrarily define the regimes in the MSSM Higgs sector as follows [66]:

$$\begin{aligned}
\text{decoupling regime} &: cos^2(\beta-\alpha) \leq 0.05 \\
\text{anti-decoupling regime} &: cos^2(\beta-\alpha) \geq 0.9 \\
\text{intense-coupling regime} &: g^2_{hbb} \text{ and } g^2_{Hbb} \geq 30 \\
\text{vanishing-coupling regime} &: g^2_{hbb} \leq 0.1 \\
\text{intermediate-coupling regime} &: M_A \gtrsim 2M_Z, g^2_{Htt}/g^2_{Hbb} \geq 10^{-2} \\
&: M_A \lesssim 2M_Z \text{ (complementary region)}
\end{aligned} \tag{1.72}$$

The MSSM Higgs boson search presented in this thesis is done in the decoupling regime.

Chapter 2

Higgs boson phenomenology

Chapter content

In this chapter, the experimental searches of the Higgs boson from the Standard Model and of its Minimal SuperSymmetric extension are presented. First a review of Standard Model Higgs searches before the LHC is given, with a special focus on the results from the LEP and Tevatron accelerators. Then the searches of the Standard Model Higgs boson at the LHC are presented. The discovery of a new SM-like Higgs boson by the ATLAS and CMS collaborations in 2012 [71, 72] is described in details. Finally, the last section in this chapter introduces the direct MSSM Higgs bosons searches at the LHC.

2.1 SM Higgs boson discovery

2.1.1 Pre-LHC searches

In the last decades, bounds on the Higgs boson mass were provided by direct searches and, indirectly, by precision measurements in the electroweak sector. The LEP collaborations (ALEPH, DELPHI, OPAL and L3), from 1989 to 2000 measured the properties of the $W^{\pm}$ and Z^0 bosons with an outstanding precision, while putting also direct bounds on the Higgs boson mass. The Tevatron collaborations ($D\emptyset$ and CDF), discovered the top quark and explored new ranges of energy, putting extra constraints on the presence of the Higgs boson.

Indirect constraints

Electroweak precision measurements are sensitive to radiative corrections which depend logarithmically on m_H. Taking all the precision electroweak data into account (Z^0 precision measurements by LEP collaborations and $W^{\pm}$ and top quark mass measurements by the Tevatron experiments) one can perform a global fit where the Higgs mass m_H is taken as the floating parameter. The update from March 2012 performs the fit with a top quark mass of 173 ± 0.9 GeV and a $W^{\pm}$ boson mass of 80.385 ± 0.015 GeV. By minimizing $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ as function of m_H, as shown in figure 2.1, the value of the most likely SM Higgs boson mass is [73]:

$$m_H = 94^{+29}_{-24}\ \text{GeV}/c^2 \tag{2.1}$$

with an upper limit at the 95% confidence level on the SM of:

$$m_H < 152\ \text{GeV}/c^2 \tag{2.2}$$

The limit increases to 171 GeV when including the LEP2 direct search limit of 114 GeV shown in yellow. The Tevatron experiments CDF and $D\emptyset$ direct search for the Standard Model Higgs boson are also displayed excluding the mass range of 156 GeV to 177 GeV at 95%CL. At that time, the LHC experiments were able to exclude a range of 127 GeV to 600 GeV (December 2011 LHC presentations of ATLAS and CMS) also displayed in the yellow bands.

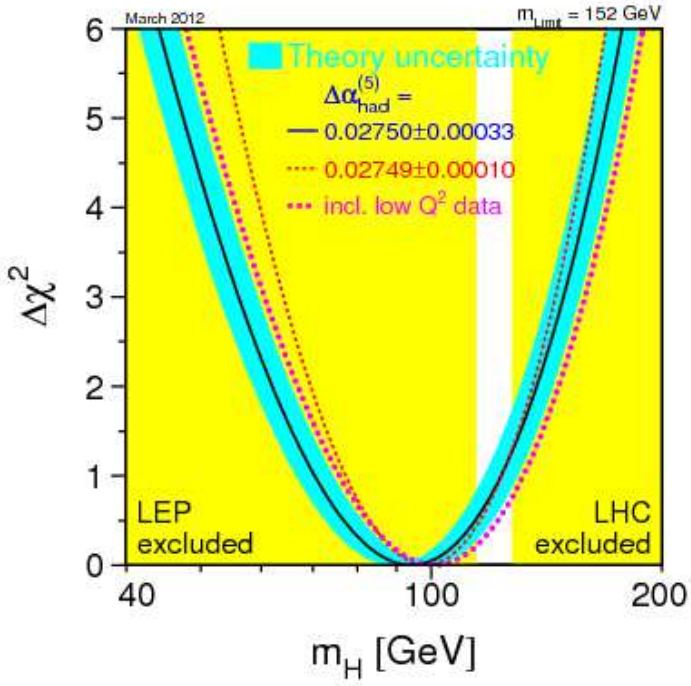


Figure 2.1: *The χ^2 of the fit to the electro-weak precision data as function of m_H. The continuous line results from high energy precision measurement. The blue band takes theoretical uncertainties (caused by neglecting higher order corrections) into account. From Ref. [73].*

Direct searches at LEP

During the first phase of LEP also called LEP1, the Higgs boson was searched at collisions at energies near the Z boson resonance. This searches proved the absence of any Higgs boson signal by the four collaborations at LEP1, and allowed to set the 95% Confidence Level limit of $M_H \geq 65.2$ GeV on the SM Higgs boson mass [74].

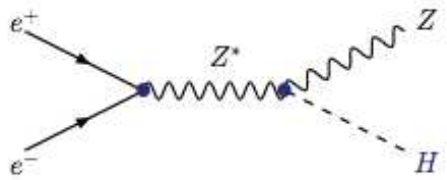


Figure 2.2: *Main production mode of the Higgs boson in e^+e^- collisions at LEP2.*

The search for the Standard Model Higgs boson was extended in LEP2, where the collisions reached a center of mass energy of $\sqrt{s} = 209$ GeV, maximal for the LEP collider. The main production mode was the so called *Higgsstrahlung* process (*c.f.* figure 2.2) where the Higgs boson is produced in association with an on-shell Z boson coming from an off-shell Z^* boson from the electron-positron pair annihilation. The maximal Higgs boson mass reached was then $M_H < \sqrt{s} - M_Z \approx 118$ GeV. The production cross section for the

Higgsstrahlung process is:

$$\sigma(e^+e^- \to ZH) = \frac{G_\mu^2 M_Z^4}{96\pi s}\left[1 + (4\sin^2(\theta_W) - 1)^2\right]\lambda^{1/2}\frac{\lambda + 12M_Z^2/s}{(1 - M_Z^2/s)^2} \tag{2.3}$$

where $\lambda = (1 - M_H^2/s - M_Z^2/s)^2 - 4M_H^2M_Z^2/s^2$ is the phase space function for the two-particle system.

The LEP2 phase data-taking periods gradually increased the center of mass energy recording a total integrated luminosity of 2461 pb^{-1}. With a typical production cross section of the order of 100 fb for a Higgs boson of $M_H = 115$ GeV the expected number of events was around 10 for each experiment. No significant excess was observed. Therefore, limits on the production cross section were set, which exclude the presence of a SM Higgs boson up to a mass of 114.4 GeV [75] as shown in figure 2.3.

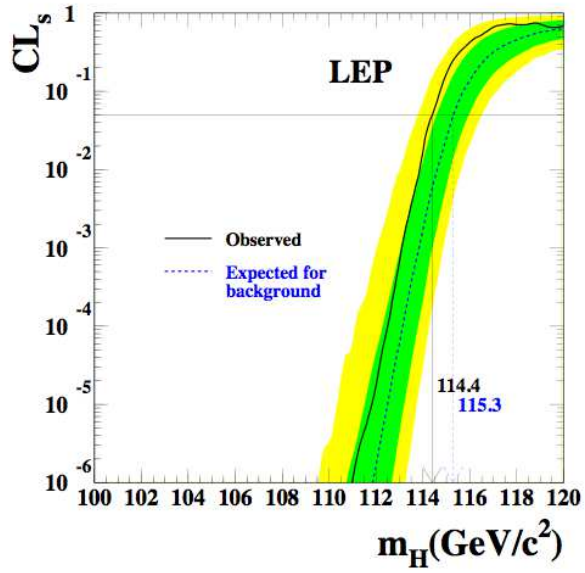


Figure 2.3: *The ratio $CL_s = CL_{s+b}/CL_b$ for the signal plus background hypothesis. Solid line: observation; dashed line: median background expectation. The green and yellow bands around the median expected line correspond to the 68% and 95% probability bands. The intersection of the horizontal line for $CL_s = 0.05$ with the observed curve is used to define the 95% confidence level lower bound on the mass of the Standard Model Higgs boson. From Ref. [75].*

Direct search at Tevatron

$\bar{p}p$ collisions at the Fermilab Tevatron collider were performed during the first run between 1992 and 1996 at a center of mass energy of $\sqrt{s} = 1.8$ TeV. At the end of this run, each experiment recorded an integrated luminosity of 100 pb^{-1}. During the Run 2 (between 2001 and 2011) the collisions at $\sqrt{s} = 1.96$ TeV were done with a much higher instantaneous luminosity allowing to record a total integrated luminosity up to 10 fb^{-1} per experiment.

Using this data, the CDF and D0 collaborations searches for a SM Higgs boson produced via the following production modes: gluon fusion, vector boson fusion and associate production with a W or Z boson. These processes will be detailed in the next section. The Higgs boson searches were performed in the following decay channels: $H \to b\bar{b}, H \to WW, H \to WW, H \to \tau\tau$ and $H \to \gamma\gamma$.

A combination with the data collected at $\sqrt{s} = 1.96$ TeV was presented on July 2^{nd} 2012, two days before the discovery announced by the ATLAS and CMS Collaborations (*c.f.* next section), leading to an evidence for a new particle in the search for the Standard Model Higgs boson [76]. In figure 2.4 the observed and expected limit in the combined search are shown.

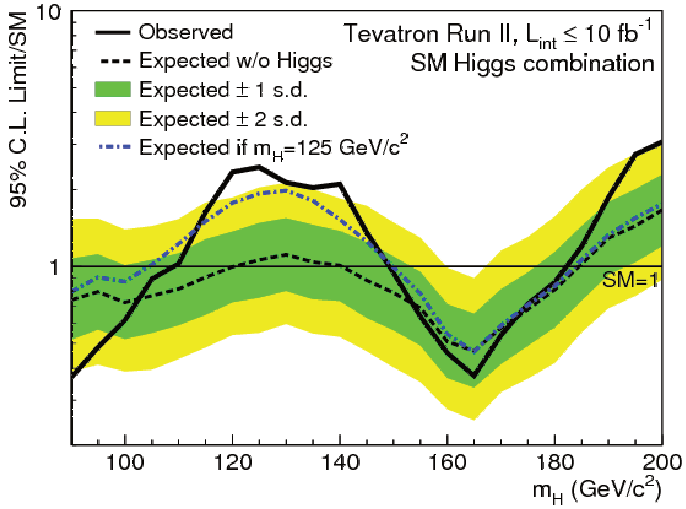


Figure 2.4: *Observed and median expected (for the background-only hypothesis) 95% C.L. Bayesian upper production limits expressed as multiples of the SM cross section as a function of Higgs boson mass for the combined CDF and D0 searches in all decay modes. The dark- and light-shaded bands indicate, respectively, the one and two s.d probability regions in which the limits are expected to fluctuate in the absence of signal. The blue short-dashed line shows median expected limits assuming the SM Higgs boson is present at* $m_H = 125\ GeV/c^2$. *From Ref. [76].*

Hence, before the advent of the LHC, the LEP and Tevatron collaborations had established strong constraints on the Higgs boson mass.

2.1.2 Production modes at the LHC

As we saw before (in section 1.2.1), the couplings of the Standard Model Higgs boson are proportional to the vector boson mass squared for vector boson couplings and to the fermion mass for the fermionic Yukawa couplings. This leads to production modes preferentially involving heavy particles. We will limit ourselves here to the four main production modes of the Standard Model Higgs boson at the LHC: the gluon fusion, the vector boson fusion, the *Higgsstrahlung* process and the associated production with heavy

quarks:

$$\begin{aligned} \text{associated production with W/Z} &: q\bar{q} \to V + H \\ \text{vector boson fusion} &: qq \to VV^* \to qq + H \\ \text{gluon-gluon fusion} &: gg \to H \\ \text{associated production with heavy quarks} &: gg, q\bar{q} \to Q\bar{Q} + H \end{aligned} \tag{2.4}$$

The production cross sections depend on m_H, the Higgs boson mass as well as on the center of mass energy of the proton-proton collision ($\sqrt{s}$). The cross section is lower for higher m_H and increases with $\sqrt{s}$. In figure 2.5 the Standard Model Higgs boson production cross sections are displayed.

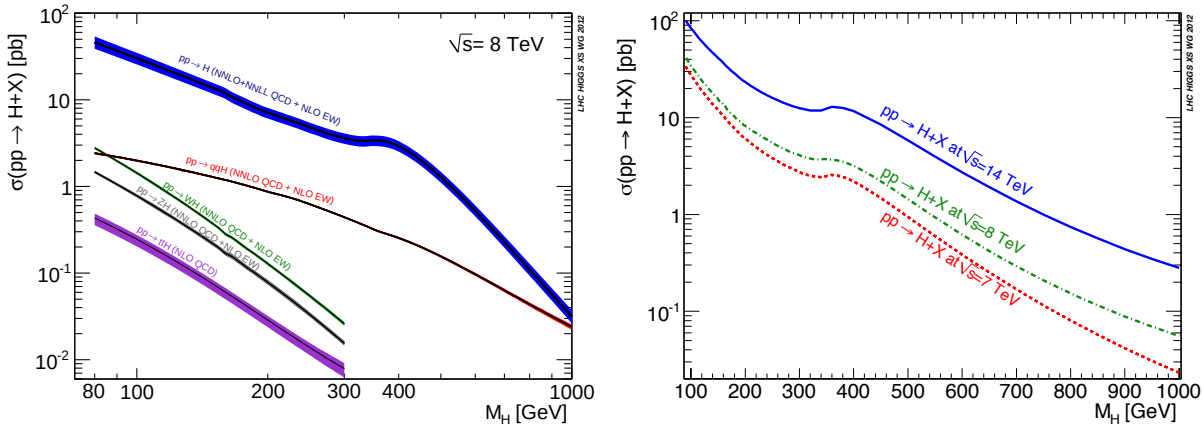


Figure 2.5: *Standard Model Higgs boson production cross section as a function of its mass for two center of mass collisions scenarios: $\sqrt{s} = 8$ TeV (left): the four production mechanisms are displayed, gluon-gluon fusion (blue), vector boson fusion (red), associated production with W/Z (grey/green) and associated production with heavy quarks (violet). Right: the total cross sections for $\sqrt{s} = 7$, 8 and 14 TeV. From Ref. [58].*

The uncertainty bands in the plots of figure 2.5 correspond to the QCD and EW theoretical uncertainties (missing orders, PDFs, etc.).

The associated production with W/Z bosons

In the associated production with massive gauge bosons, also called *Higgsstrahlung* process, the Higgs boson is radiated from a virtual boson as shown in figure 2.6 (V=Z or W). The charged leptons and missing transverse energy originating from the leptonic decays of the gauge bosons can be used to tag these events and the Higgs boson can be searched in its decay channels.

The *Higgsstrahlung* process can be decomposed into two sub-processes: the production of the virtual vector V^* and the subsequent decay into a real vector boson and the Higgs boson ($V^* \to HV$). Leading to the partonic cross section of the form:

$$\hat{\sigma}(\hat{s}) = \int_0^{\hat{s}} dp^2_{V^*} \sigma\left(p^2_{V^*}\right) \frac{d\Gamma(V^* \to VH)}{dp^2_{V^*}} \tag{2.5}$$

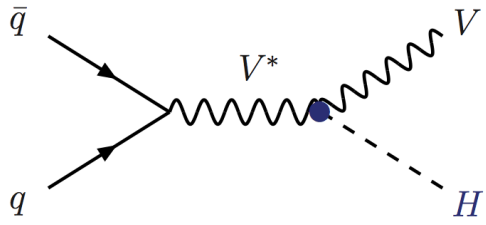


Figure 2.6: *Feynman diagram of the Standard Model Higgs boson associated production with V=W/Z mechanism.*

The total cross section for this process is known to NNLO accuracy in QCD [77, 78] and NLO in the electroweak expansion [79]. The cross sections for the production with $W^{\pm}$ bosons is $\sim$ 2 times larger than the one with Z^0. It is even more favored due to the corresponding branching ratio to charged leptons (electron or muons): $BR(W^{\pm} \to \ell^{\pm}\nu) \sim 20\%$ while $BR(Z^0 \to \ell^+\ell^-) \sim 6\%$.

The hadronic decays of the gauge bosons can be also used to tag this production mode in the channels $VH \to qq' + H$. This channels are favored by the higher branching ratio of the hadronic vector boson decays.

As showed previously, this was the main Higgs boson production mode at LEP. It was the one with the second largest cross section at the Tevatron. At the LHC, it accounts for about 5% of the Higgs boson produced and represents the third main contribution.

The vector boson fusion

In the vector boson fusion (VBF) production mode, the Higgs boson is produced through the fusion of two $W^{\pm}$ or Z^0 bosons which have been radiated from large momentum incoming quarks. Figure 2.7 shows the Feyman diagram for this production mechanism at leading order.

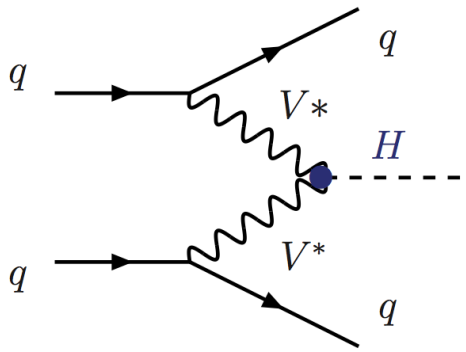


Figure 2.7: *Feynman diagram of the Standard Model Higgs boson vector boson fusion production mechanism.*

At leading order, this process is purely electroweak and due to the lack of color exchange between the two incoming quark currents, the QCD activity is concentrated

around the outgoing quarks. The outgoing quarks have typically a transverse momentum of $\sim m_V/2$, they hadronize in the forward and backward regions of the detector and the Higgs decay products lean in the central region. This event topology constitutes a specific signature of the VBF production process and a targeted selection can be used to greatly suppress the other SM backgrounds.

The partonic cross section for the leading order VBF can be computed in the limit where the transverse momentum of the outgoing quarks is small compared to the center-of-mass energy $\hat{s}$. In this approximation we have [80]:

$$\hat{\sigma}_{Born}(\hat{s}) = \frac{1}{16M_W^2}\left(\frac{\alpha}{\sin^2\theta_W}\right)^3\left[\left(1+\frac{M_H^2}{\hat{s}}\right)\log\frac{\hat{s}}{M_H^2} - 2 + 2\frac{M_H^2}{\hat{s}}\right] \tag{2.6}$$

We can note then the cross section increases with $\hat{s}$, which explain the importance of this process at the LHC with respect to the Tevatron. At the LHC, the fusion of W^+W^- have a rate ~ 3 times larger then the Z^0 fusion due to the higher coupling of the W boson to quarks.

The VBF cross section is known at NLO accuracy in QCD including electroweak corrections [81–83]. Approximations including NNLO corrections are also computed [84]. The QCD corrections increases the cross section by 5 to 10% and the scale dependence is less than 1-2%. About 7% of all the Higgs boson produced come from the VBF production mode and represents the second main production mode at the LHC.

The gluon-gluon fusion

The gluon fusion mechanism is the main production mode of the Higgs boson at the LHC. In this process, the Higgs boson production is mediated by triangular loops of heavy quarks, dominated by the top quark. The leading order Feynman diagram is presented in figure 2.8.

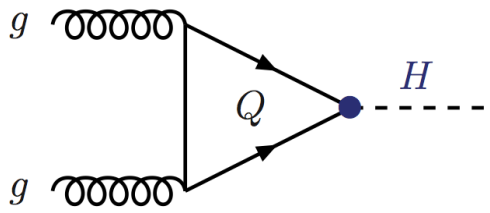


Figure 2.8: *Feynman diagram of the Standard Model Higgs boson gluon-gluon fusion production mechanism.*

At leading order the partonic cross section is proportional to the Born width in equation 1.36, being dominated by the strong interaction and of $\mathcal{O}(\alpha_S^2)$:

$$\begin{aligned}\hat{\sigma}_{LO}(gg\to H) &= \sigma_0^H M_H^2 \delta(\hat{s}-M_H^2)\\ \sigma_0^H &= \frac{G_F\alpha_s^2(\mu_R^2)}{288\sqrt{2}\pi}\left|\frac{3}{4}\sum_q A_{1/2}^H(\tau_Q)\right|^2\end{aligned} \tag{2.7}$$

where $\hat{s}$ is the gg invariant energy squared, $A^H_{1/2}(\tau_Q)$ the form factor with $\tau_Q = M_H^2/4m_Q^2$.

The cross section decreases steeply with M_H. It is at the level of ~ 30 pb for $M_H \sim 100$ GeV. A kink at $M_H \sim 350$ GeV, near the $t\bar{t}$ threshold, is due to the fact that the imaginary part of the Hgg amplitude $A^H_{1/2}(\tau_Q)$ suddenly increases.

The cross section is computed at the next-to-next-to-leading order (NNLO) accuracy in QCD using the $m_t \to \infty$ approximation [85–87]. The next-to-leading order (NLO) corrections increase the total cross section by about $80-100\%$ with respect to the leading order (LO) computation. The third order NNLO corrections additionally increase the NLO cross section by $\sim 25\%$. The re-summation of the soft-gluon contributions at the next-to-next leading logarithms (NNLL) [88] improves the accuracy of the cross section calculation and increases it by an additional $6-9\%$ as well as a reduces the sensitivity to the choice of the factorization (μ_F) and renormalization scales (μ_R). Electroweak corrections are computed at NLO and included in the total cross section [89, 90].

The cross section theoretical uncertainty is attributed by varying the factorization and renormalization scales within $M_H/2 < \mu_R, \mu_F < 2M_H$. The main contribution to this uncertainty comes from the residual scale dependence of the NNLO+NNLL calculation ($\sim 10\%$), the parton density functions and α_S. The gluon fusion process originates $\sim 87\%$ of the Higgs with $M_H = 125$ GeV produced at the LHC at $\sqrt{s} = 8$ TeV.

The associated production with heavy quarks

The associated production with heavy quarks ($Q = t, b$) is the fourth main production mode of the Higgs boson at the LHC. It proceeds at tree level from $q\bar{q}$ annihilation and gluon fusion mechanisms. The corresponding Feyman diagrams are displayed in figure 2.9.

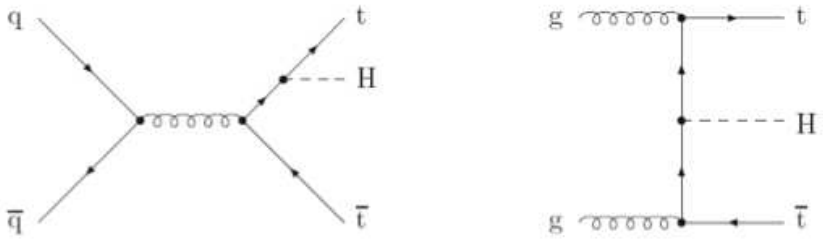


Figure 2.9: *Leading Feynman diagrams of the Standard Model Higgs boson associated production with heavy quarks mechanism.*

The partonic cross section computation for this process is very difficult even at LO due to the large combinatorics of Feynman diagrams and the final state phase space containing three particles. The inclusive cross section is known up to NLO in QCD [91–93], and it increases the LO prediction by $\sim 20\%$. The ttH cross section is about two orders of magnitude smaller than the gluon fusion process and is responsible for near 1% of the produced Higgs bosons.

This production mode has the advantage to directly probe the Higgs couplings to the top quarks and then plays a primary role in the study of the electroweak symmetry breaking.

2.1.3 Searches at the LHC

At the LHC, a variety of Higgs boson search channels are studied. Their relative importance is based on the mass range under study, production cross section, branching ratios, background contamination and resolution. In the case of the CMS experiment, a complete review on the prospectives of analyses is done in Ref. [12]. In the following, I present the main channels analyzed during the Run I.

$H \to \gamma\gamma$

At the LHC, it is the most sensitive decay mode in the low mass range. This decay mode is suited for masses up to $\sim$ 150 GeV. The signature in the detector consists on two energetic photons in the final state, a very clean signature. It is one of the two so-called high resolution channels. The analysis is based on the search of a peak in the reconstructed invariant mass distribution of the two photons on top of the background continuum. The main source of background is the QCD $\gamma\gamma$ production.

This channel is very interesting despite the low branching ratio and the high background contamination. The clean signature in the detector allows to reconstruct precisely the Higgs boson mass, limited only by the experimental resolution ($\sim 1-2$ GeV). An excellent energy resolution and an efficient photon reconstruction are then needed to perform this analysis. Due to the loops in this Higgs boson decay (see section 1.2.3) this channel gives access indirectly to the couplings to the top quark and the $W^{\pm}$ bosons. The VBF, VH and ttH production topologies can be used to exploit categories with higher signal purity and enhance the analysis sensitivity. This channel also allows to perform the measurement of the spin and CP quantum numbers of the Higgs boson.

$H \to ZZ^{(*)} \to 4\ell$

The $H \to ZZ^{(*)}$ channel has a branching fraction of $\mathcal{O}(1\%)$ at $m_H = 120$ GeV raising to $\mathcal{O}(10\%)$ at $m_H = 150$ GeV. It drops quickly at about $m_H = 2m_W$ due to the rising of the $H \to WW$ channel with on-shell W bosons. Then, at higher masses it becomes the second dominating channel. The main source of the moderate background contribution consists in QCD and electroweak di-boson production. The reducible background contributions are kept to moderate levels thanks to the efficient lepton reconstruction and isolation requirements imposed on the leptons. In particular the $H \to ZZ^{(*)} \to 4\ell$ channel is the other high resolution channel at the LHC (with $H \to \gamma\gamma$) and is called the "golden channel" due to its very large signal to background ratio of $\sim 50-100\%$.

This analysis can cover a wide range of Higgs boson masses from 120 GeV to 1000 GeV. The lepton trigger and reconstruction are critical in this channel. At low mass, maintaining high efficiencies is a great challenge due to the presence of an off-shell Z^* boson

decaying to low energy leptons (typically with a few GeV transverse momentum). At high masses this channels is the most sensitive at the LHC.

The clean signature in the final state allows a precise measurement of the Higgs boson mass. It can be used also to measure the Higgs boson properties as the spin and CP quantum numbers through the analysis of the angles between the decay products.

$H \to WW^{(*)} \to \ell\nu\ell\nu$

The $H \to WW$ channel has a branching ratio higher than 15% for $m_H > 125$ GeV, it blows quickly with M_H and becomes the dominant channel for masses higher than $2m_W$. The most interesting final state is the leptonic decays of the two W bosons, because of the lower background contamination. The signature consists of two isolated leptons and missing transverse energy due to the presence of the elusive neutrinos.

This channel has a poor resolution on the Higgs boson mass due to the presence of missing energy. The main background contributions are the WW production, $t\bar{t}$, single top and $Z + jets$ productions. Due to the spin conservation in the $H \to WW^{(*)}$ and $W \to \ell\nu$ decays, the angular correlations between the two leptons and the MET direction can be used to reduce the background. The VBF topology can also help to define a high signal purity sample which improves the sensitivity of the analysis. This channel gives direct access to the Higgs boson couplings to the W boson. The study of the angular distributions are good handles to measure the spin and CP quantum numbers of the Higgs boson.

$H \to b\bar{b}$

The $H \to b\bar{b}$ decay channel is dominant in the low mass region $m_H < 135$ GeV. Due to the extremely large $b\bar{b}$ QCD production at the LHC it is a very challenging channel. In order to improve the sensitivity, the associated production VH is tagged where the Higgs boson is produced in boosted topologies. The presence of missing transverse energy and/or leptons in the final state is used to select events with associated production with a $Z \to \nu\nu$, $W \to \ell\nu$ or $Z \to \ell\ell$.

This channel is the only one that gives direct access to the Higgs boson coupling to down-type quarks. A moderate Higgs boson mass resolution is achieved. An efficient b-tagging is needed to distinguish b-quark jets coming from c-quark and light-quark jets from electroweak, QCD and Z+jets main backgrounds.

$H \to \tau\tau$

With the current LHC data, this analysis is the only one capable to probe the Higgs boson couplings to leptons with the first data collected at the LHC. The $H \to \tau\tau$ has an intermediate branching ratio, favorable at low mass, around 6% at $m_H = 125$ GeV. In the $H \to \tau\tau$ searches, the main backgrounds are the Drell-Yann $Z \to \tau\tau$ production, W+jets and QCD multi-jets production as well as the $Z \to \ell\ell$ production where one (or both) of the leptons are mis-identified as a τ_h. A performant τ lepton reconstruction in its hadronic decay modes (τ_h) is a key point in this analysis due to the large decay branching ratio ($\sim 2/3$).

The most promising final states are the semi-leptonic ones ($H \to \tau\tau \to \ell\tau_h$), because they combine the high efficiency of the lepton trigger and reconstruction and the high $\tau \to \tau_h$ branching ratio. The distinctive topologies, such as the one of the VBF or VH production processes, can be used to enhance the analysis sensitivity. In the gluon fusion process an initial state radiation of a gluon can induce a boost in the Higgs boson system, then requiring the presence of an additional jet can reduce the background contamination and improve the di-τ mass resolution.

The τ_h reconstruction, my contributions to the electron mis-identification as τ_h are detailed in chapters 4 and 5. The CMS analysis and my contributions to it are also presented in chapters 6 and 7.

2.1.4 Higgs boson discovery at the LHC

On the 4^{th} of July 2012 the ATLAS and CMS collaborations announced the discovery of a new particle in the search for the SM Higgs boson. The discovery was achieved by combining the analyses of the 5 decay channels discussed previously [71, 72].

For this publication ATLAS recorded 4.8 fb^{-1} at $\sqrt{s} = 7$ TeV and 5.8 fb^{-1} at $\sqrt{s} = 8$ TeV. The production of a neutral boson was observed in the combination of the $H \to ZZ^{(*)} \to 4\ell$, $H \to WW^{(*)} \to e\mu$ at 8 TeV and $H \to ZZ^{(*)}$, $H \to WW^{(*)}$, $H \to b\bar{b}$ and $H \to \tau\tau$ at 7 TeV. The excess of observed events had a statistical significance of 5.9 standard deviations and the mass was measured to be $126.0 \pm 0.4(stat.) \pm 0.4(syst.)$. In figure 2.10 the p-value as a function of the Higgs boson mass is displayed.

The CMS result was obtained by analyzing the recorded integrated luminosity of 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV. An excess of events was observed with respect to the background-only hypothesis, with a statistical significance of 5.0 standard deviations. The measured mass reads $125.3 \pm 0.4(stat.) \pm 0.5(syst.)$. In figure 2.11, the CMS combined p-value is displayed as a function of the Higgs boson mass.

This result presented at CERN and widely followed by the particle physics community culminates the work made during decades in the Higgs boson searches. The Nobel price of Physics was awarded to Peter Higgs and Francois Englert in 2013 for the theoretical prediction of the Higgs boson.

2.2 MSSM Higgs bosons searches

In this section I present the neutral MSSM Higgs boson production modes at the LHC and its constrains and searches. Usually, the search of the MSSM Higgs bosons is done by using customary benchmark scenarios where all the SUSY parameters are fixed, excepting M_A and $\tan\beta$. In the context of each scenario, the search is therefore performed as a scan of the (M_A, $\tan\beta$) parameter space. The most common is the m_h^{max} scenario. A description of the different MSSM scenarios considered for the search is available in appendix A.

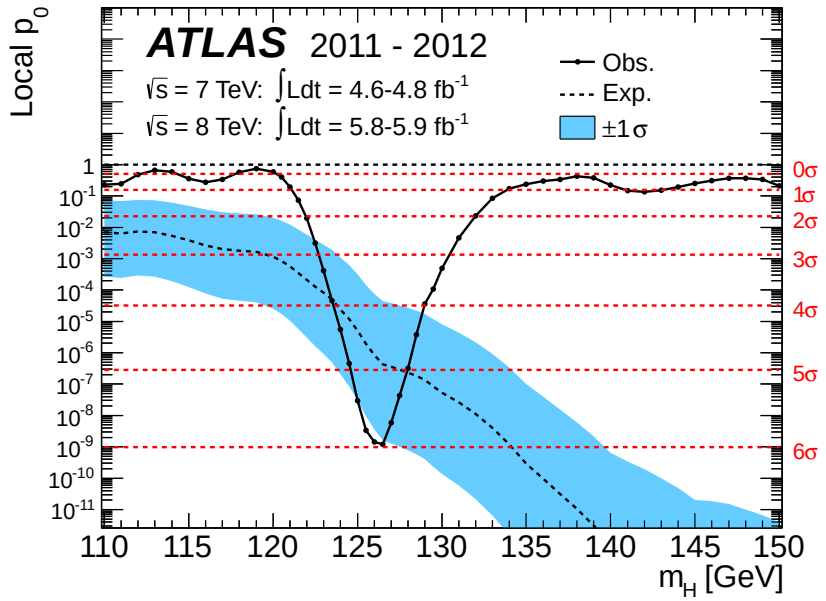


Figure 2.10: *The observed (solid) local p_0 as a function of m_H in the low mass range. The dashed curve shows the expected local p_0 under the hypothesis of a SM Higgs boson signal at that mass with its plus/minus one sigma band. The horizontal dashed lines indicate the p-values corresponding to significances of 1 to 6 sigma. From Ref. [71].*

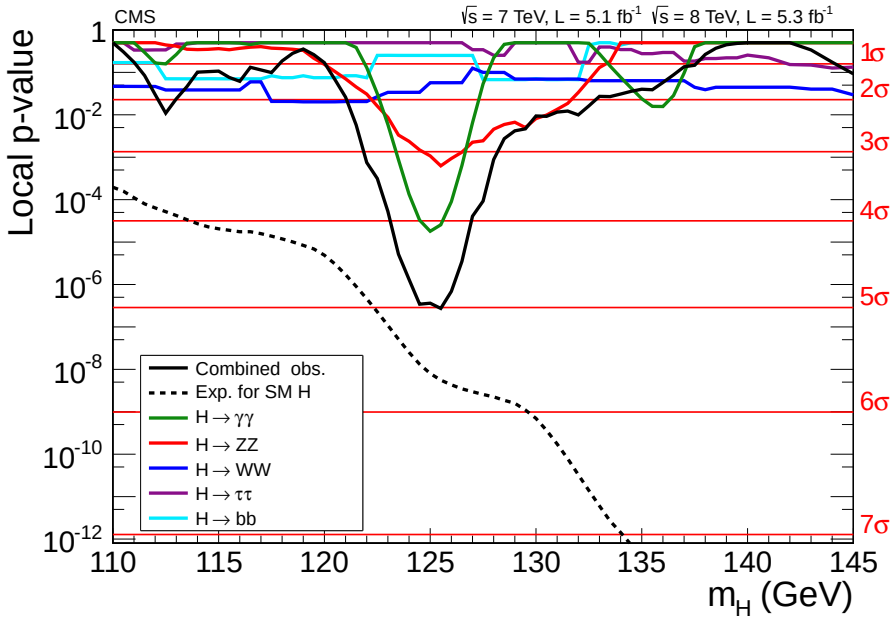


Figure 2.11: *The observed local p-value for the five decay modes and the overall combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H. From Ref. [72].*

2.2.1 Pre-LHC searches

Direct searches for MSSM Higgs bosons have been performed at the LEP and Tevatron experiments. To perform these analysis the m_h^{max} benchmark scenario has usually been used. At high values of $\tan\beta$, the MSSM Higgs bosons decay almost exclusively to $b\bar{b}$ and $\tau\tau$, due to the enhancement of the coupling to down type fermions discussed in section 1.3.4.

Direct search at LEP

The extension of the LEP2 energy up to $\sqrt{s} = 209$ GeV was mainly motivated by the search for the Higgs boson. At these energies, the main production modes for the MSSM Higgs bosons were the *Higgsstrahlung* process and the associated production of CP-even and CP-odd Higgs bosons. The figure 2.12 presents the corresponding Feynman diagrams.

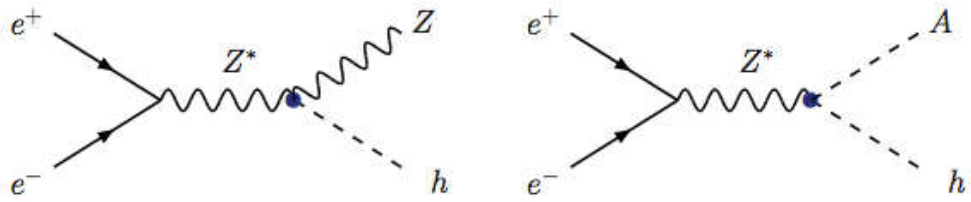


Figure 2.12: *MSSM neutral Higgs boson production diagrams at LEP.*

The search for the neutral MSSM Higgs bosons have been performed using the $\phi \to b\bar{b}$ and $\phi \to \tau\tau$ final states ($\phi = A/h$). The combination of all the final states topologies showed no excess of observed data and a combined 95% CL limits on the h and A boson masses have been set to [94] $M_h > 92.8$ GeV and $M_A > 93.4$ GeV within the m_h^{max} scenario with a top quark mass of $m_t = 174.3$ GeV. In order to set the upper bound limits we define:

$$S_{95} = \sigma_{max}/\sigma_{ref}$$

where σ_{max} is the largest cross section compatible with the data and σ_{ref} is a reference cross section. For the *Higgsstrahlung* process it is taken to be the SM production cross section and for final topologies motivated by the pair production σ_{ref} is taken to be the corresponding MSSM production cross section. In figure 2.13, S_{95} for the *Higgsstrahlung* process and the excluded region in the $M_A - \tan\beta$ plane are shown within the m_h^{max} scenario.

Direct searches at Tevatron

The Tevatron collaborations have published the combination of searches for neutral MSSM Higgs bosons also in the $\phi \to b\bar{b}$ and $\phi \to \tau\tau$ final states.

The $\phi \to b\bar{b}$ channel was analyzed using events with three or more b-tagged jets in the final state, using 2.6 fb^{-1} of recorded luminosity by CDF and 5.2 fb^{-1} by $D\emptyset$ [95]. Limits on the product of the cross section and the branching ratio are extracted. An

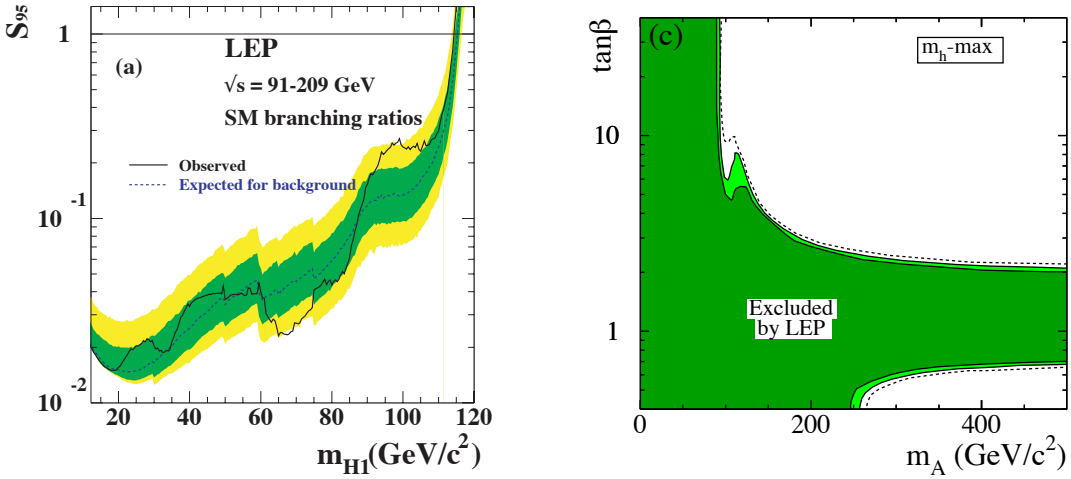


Figure 2.13: *The 95% CL upper bounds, S_{95} for the combined Higgsstrahlung process (left); the solid line corresponds to the observed limits, the green and yellow bands around the median expectation (dashed line) correspond to the 68% and 95% probability bands and the horizontal line correspond to the SM cross section. Excluded region in the $M_A - \tan\beta$ plane (right), at 95% CL (light green) and 99.7% CL (dark green) within the m_h^{max} scenario; the dashed lines correspond to the regions expected to be excluded at 95% CL in the basis of MC simulation with no signal. From Ref. [94].*

excess of observed events over the SM background expectation are observed for $M_\phi = 120$ and 140 GeV with significances of 2.5 and 2.6 standard deviations respectively. The significance of the excesses in the combined analysis is reduced to ~ 2 standard deviations after applying the look elsewhere effect[1] [75]. In figure 2.14 the model independent limits and the interpretation in the $(M_A, \tan\beta)$ plane for the m_h^{max} scenario are presented.

The di-τ final states have been also exploited. The combination of the search for a neutral Higgs boson using 1.8 fb^{-1} of recorded luminosity by CDF and 2.2 fb^{-1} by $D\emptyset$ is reported in Ref. [96]. Three different channels depending on the τ-lepton decay mode were analyzed: $e\tau_h$, $\mu\tau_h$ and $e\mu$, where τ_h refers to a hadronic τ decay. The 1.8 fb^{-1} of recorded luminosity by CDF and 1.0 fb^{-1} by $D\emptyset$ are combined with an additional 1.0 fb^{-1} in the $\mu\tau_h$ final state collected at $D\emptyset$.

The observed limits are in good agreement with expectation with no evidence for significant excess for $90 < M_A < 200$ GeV. In figure 2.15, the model independent limits and the interpretation $(M_A, \tan\beta)$ plane for the m_h^{max} scenario are presented.

2.2.2 Production modes at the LHC

In the decoupling regime presented in section 1.3.5, the couplings of the light CP-even Higgs boson converges to the ones of the SM. A difference with respect to the Standard

[1] A standard convention is used to account for the effect that it is more likely to find a deviation (under the background-only) hypothesis when several mass regions are probed compared with only a single hypothesis.

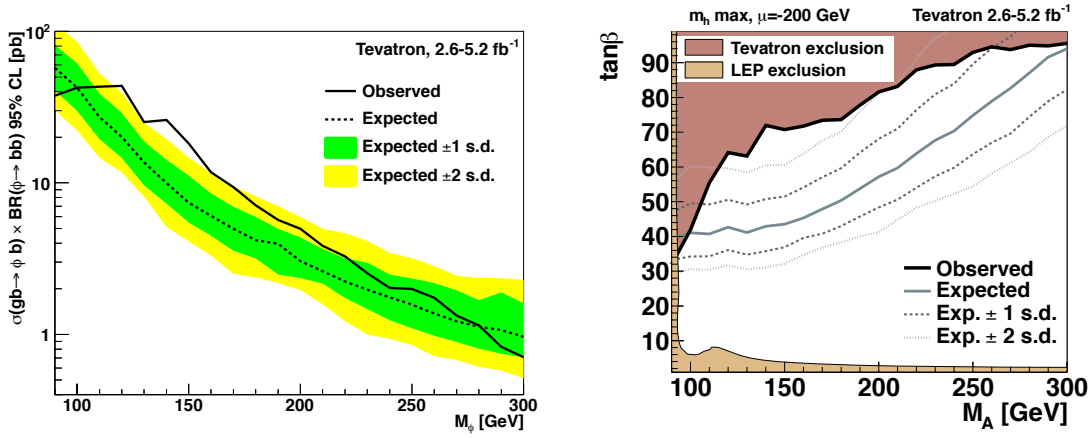


Figure 2.14: *Model independent 95% C.L. upper limits (left) on the product of cross section and branching ratio for the combined analyses. The green and yellow regions correspond to the one and two standard deviation bands around the median expected limit. 95% C.L. lower limit in the* $(M_A, \tan\beta)$ *plane for the* m_h^{max} *scenario (right). The exclusion limit obtained from the LEP experiments is also shown. From Ref. [95]*

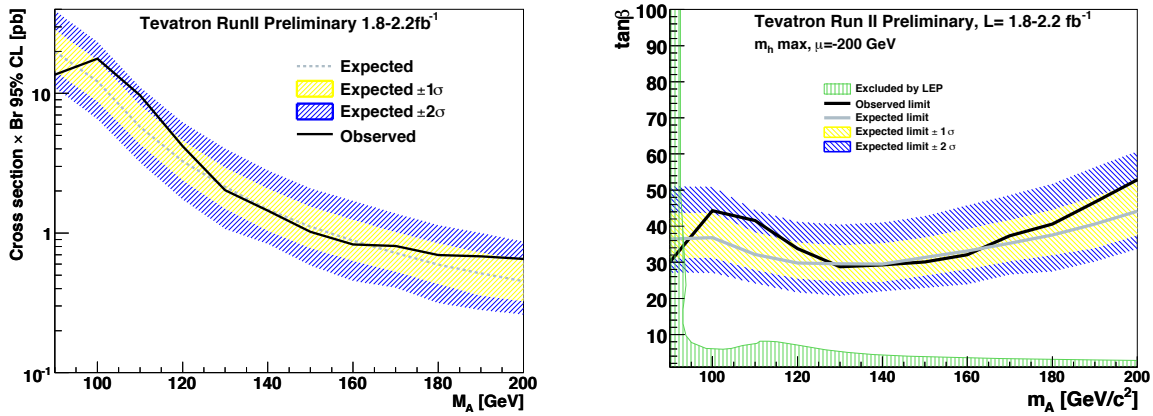


Figure 2.15: *95% Confidence limits on cross section × branching ratio (left). The solid black and dashed grey lines show the observed and expected limits respectively. The yellow and blue hatched bands around the expected limit show the 1 and 2σ deviations from the expectation. 95% Confidence limits in the* $(M_A, \tan\beta)$ *plane for the* m_h^{max} *scenario (right). The black line denotes the observed limit, the grey line the expected limit and the hatched yellow and blue regions denote the* ± 1 *and* 2σ *bands around the expectation. The shaded light-green area shows the limits from LEP. From Ref. [96].*

Model searches is that the production processes can include additional contributions from SUSY particles or even lighter MSSM Higgs bosons if kinematically allowed. Also, contrarily to the Standard Model where the Higgs boson mass is a free parameter, the production processes requires an evaluation of the Higgs boson masses and mixing contribution terms on the radiative corrections.

The associated production with W/Z bosons and vector boson fusion

The *Higgsstrahlung* and VBF processes are relevant for SM like states, as the light MSSM Higgs boson in the decoupling regime. The same computation as for the SM case is done. The additional SUSY QCD corrections are computed and are found to be small [97]. For the vector boson fusion process the SUSY electroweak radiative corrections are at the level of 1% and they are unknown for the associated production process.

The gluon-gluon fusion

The gluon fusion process is dominant for small values of $\tan\beta$. The production introduces a loop of top or bottom quarks as for the SM case, but additional contributions appear from stop and sbottom as presented in the leading order Feynman diagram in figure 2.16.

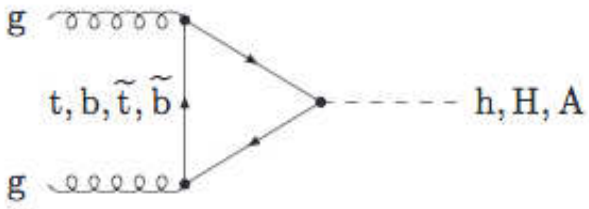


Figure 2.16: *Main Feynman diagram at leading order for the gluon fusion ($gg \to h/H/A$) production mode.*

The SUSY QCD corrections are included using the heavy-quark limit and including the full mass quark dependence [98–100]. As a result the cross section increases by up to 100% (50%) for small (large) $\tan\beta$. The NNLL resummation presented for the SM case can be used in the region where the heavy quark limit is still valid, *i.e.* for $\tan\beta \lesssim 5$.

The associated production with heavy quarks

The Higgs boson radiation of top quarks is only relevant for the light Higgs boson, because it has properties very similar to the SM one. The SUSY QCD corrections are very moderate [97]. In the case of larger $\tan\beta$ values, the Higgs boson couplings to down-type fermions are enhanced. The Higgs boson radiation of bottom quarks is the dominant production mode for large $\tan\beta$ values. In figure 2.17, the main leading order Feynman diagrams for this production mode are presented.

The NLO QCD and electroweak corrections to this production mode have been calculated [101–103] and completed with the NLO SUSY QCD corrections [104].

In figure 2.18, the MSSM Higgs bosons cross section summary is displayed using the m_h^{max} scenario, for two different values of $\tan\beta$. The two production modes exploited in the search for MSSM Higgs bosons are the gluon fusion and the associated production with bottom quarks, as they are the dominant ones and the other production modes are covered in the SM search.

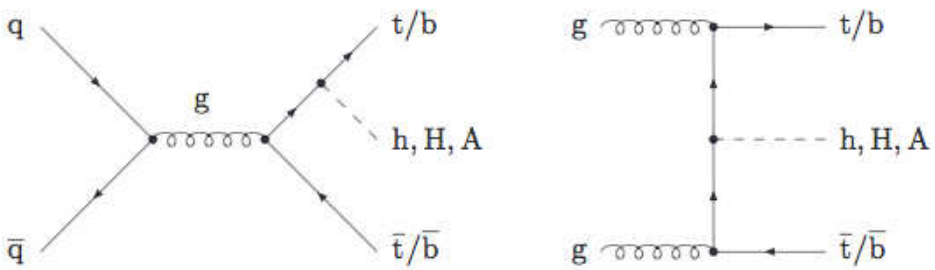


Figure 2.17: *Main Feynman diagrams at leading order for the heavy quark associated production mode: $gg/q\bar{q} \to Q\bar{Q} + h/H/A$ (up), $gb \to b + h/H/A$ (down).*

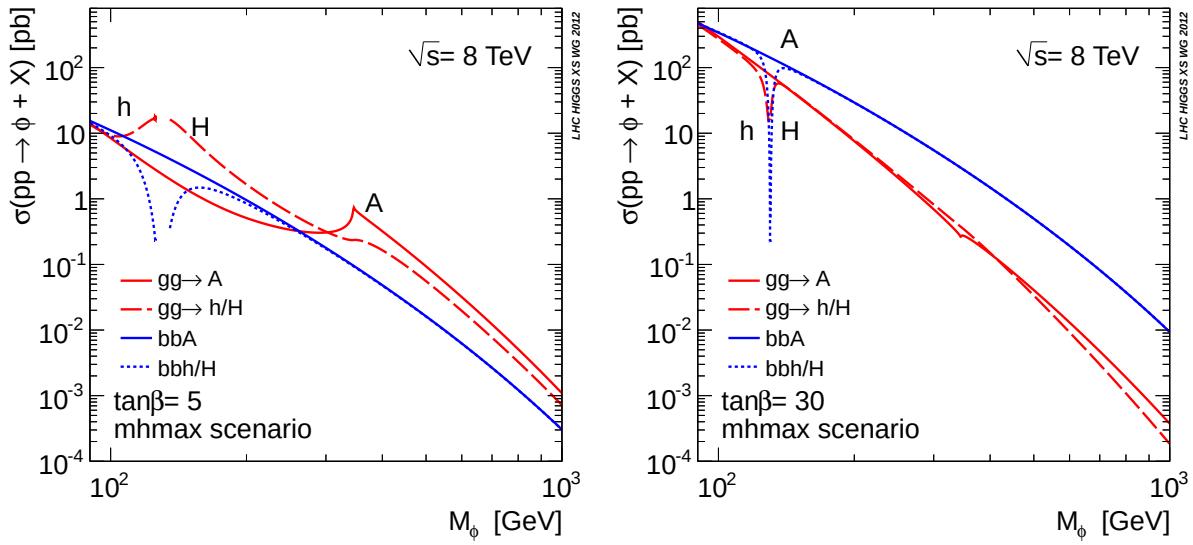


Figure 2.18: *MSSM Higgs boson cross section summary for the m_h^{max} scenario for $\tan\beta = 5$ (left) and $\tan\beta = 30$ (right). From Ref. [58].*

As shown in section 1.3.4, the main decay modes of the MSSM Higgs bosons are the decays to $b\bar{b}$ and to $\tau\tau$. This two channels have been exploited to explore the MSSM phase space at the LHC and in previous experiments, as I will detail in the next section.

2.2.3 Searches at the LHC

At the LHC, the main channel used in the search for the MSSM Higgs bosons is the di-τ final state. The previous MSSM Higgs bosons searches using the data collected with $\sqrt{s} = 7$ TeV at the LHC are reported in Refs. [105, 106]. In this section I will refer to the updated published results by the ATLAS and CMS collaborations [107, 108].

Search at ATLAS

The search for the neutral Higgs bosons predicted in the MSSM framework using the $\tau\tau$ final state was done using the data recorded by the ATLAS detector at $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of $19.5 - 20.3$ fb^{-1}. The final states considered in the search depend on the τ-lepton decay modes: $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$.

In the $e\mu$ channel, two event categories are defined in order to tag the to production modes. Their definition is based in the presence ("tag" category) or not ("veto" category) of a b-tagged jet in the final state. In the semi-leptonic channels ($\mu\tau_h$ and $e\tau_h$) a different selection is done for events with high mass ($m_A > 200$ GeV) in order to reduce the background contamination. The electron and muon channels are treated separately in the low mass categories and they are treated as a single channel in the high mass category. The $\tau_h\tau_h$ channel events are selected with a single τ_h trigger or a $\tau_h\tau_h$ trigger. The events are categorized into two event categories following the triggers they have fired.

All the channels are combined in specific MSSM Higgs boson mass regimes in order to improve the sensitivity of the analysis. The $e\mu$ channel and the low-mass categories of the semi-leptonic channels are used for the mass range $90 < m_A < 200$ GeV. The high-mass category of the semi-leptonic channels and the $\tau_h\tau_h$ channel are combined to compute the results for $m_A > 200$ GeV.

In figure 2.19, the 95% CL upper limits on the cross section times branching ratio are shown for the search of a single scalar boson ϕ with narrow width relative to the experimental resolution produced via gluon fusion or b-associated production.

In figure 2.20 the combined expected and observed limits interpreted in the ($m_A - \tan\beta$) plane in the m_h^{max} scenario are presented. No significant excess of observed data over the expected background is found.

Additional MSSM benchmark scenarios (see appendix A for their description) are used to interpret these results.

Search at CMS

The search for a neutral Higgs boson in the MSSM framework decaying to τ-lepton pairs is performed at CMS using the full integrated luminosity recorded during the LHC Run I, corresponding to 4.9 fb^{-1} at 7 TeV and 19.7 fb^{-1} at 8 TeV. The analysis channels depend on the τ-lepton decay mode. Five channels are considered: $\mu\mu$, $e\mu$, $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$.

Two event categories are used to enhance the sensitivity to a neutral MSSM Higgs boson produced in association with a b-quark jet or via gluon fusion. The "b-tag" and "No-btag" categories are defined with the presence or not of a b-tagged jet in the final state.

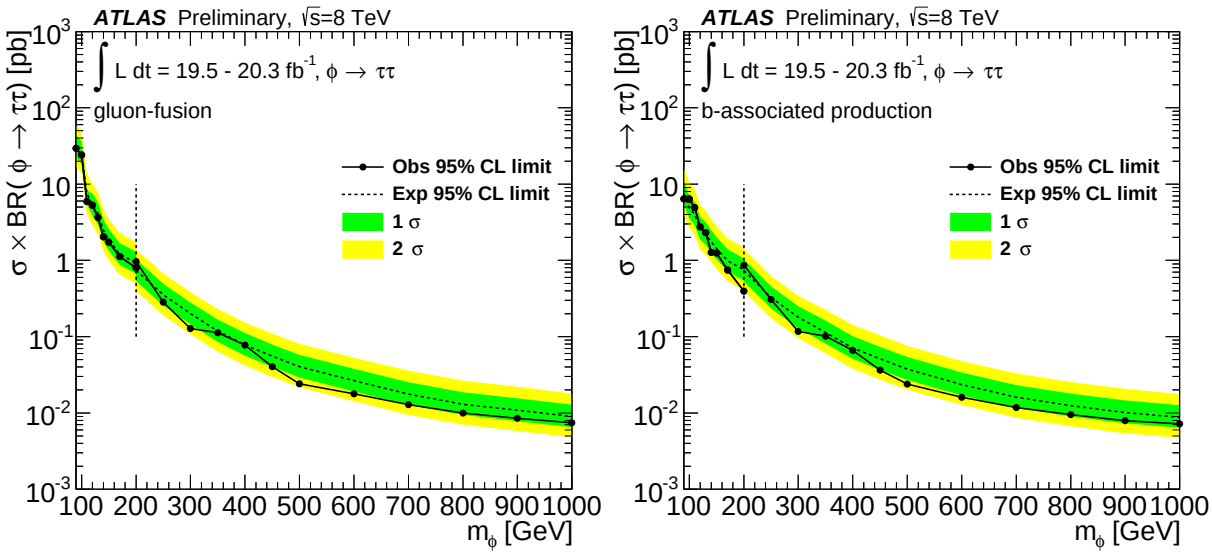


Figure 2.19: *Expected (dashed bold line) and observed (solid bold line) 95% CL upper limits on the cross section for gluon-fusion (left) and b-associated Higgs boson production (right) times the branching ratio into τ pairs. From Ref. [107].*

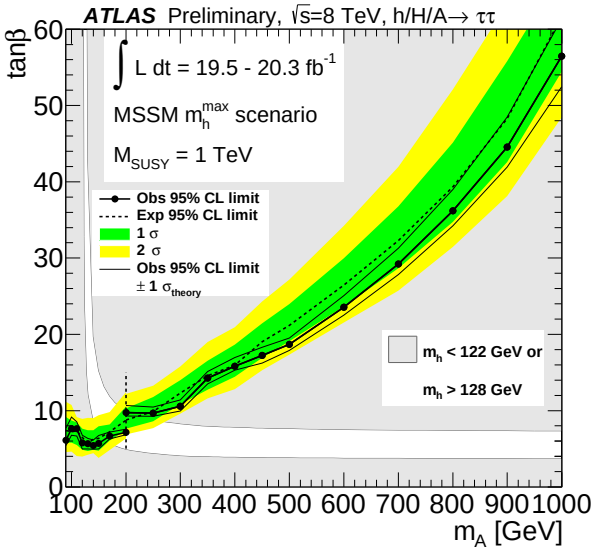


Figure 2.20: *Expected (dashed line) and observed (solid line with markers) 95% CL upper limits on $\tan\beta$ as a function of m_A. Values of $\tan\beta$ above the lines are excluded. The upper limits are shown for the statistical combination of all di-τ channels. The vertical dashed line at 200 GeV indicates the transition point between low and high mass categories. The $(m_A - \tan\beta)$ plane region which is incompatible with a light, CP-even Higgs boson with mass in the range 122 - 128 GeV is shown as a gray area. From Ref. [107].*

The observable used in the analysis is the reconstructed di-τ mass using the *SVfit* algorithm described in section 6.5. A simultaneous maximum likelihood fit is done over

all the channels and categories in order to extract limits on the cross sections or to quantify potential excesses.

In figure 2.21 the upper limits on the cross section of the production of a single resonance ϕ with a narrow width compared to the experimental resolution is shown. This model independent limits are extracted for the product of the production cross section times branching fraction to τ pairs as a function of the Higgs boson mass, separately for the gluon fusion and the b-associated production modes. For these results, only the data recorded at $\sqrt{s} = 8$ TeV have been used. No clear evidence for the presence of a MSSM Higgs boson signal is found. In figure 2.22, 95% CL upper bounds on $\tan\beta$ as a function of the pseudoscalar Higgs boson mass m_A are set for the traditional MSSM benchmark scenario m_h^{max}, and the recently proposed benchmark scenarios: m_h^{mod+}, m_h^{mod-}, light-stop, light-stau, and τ-phobic (see appendix A for their description).

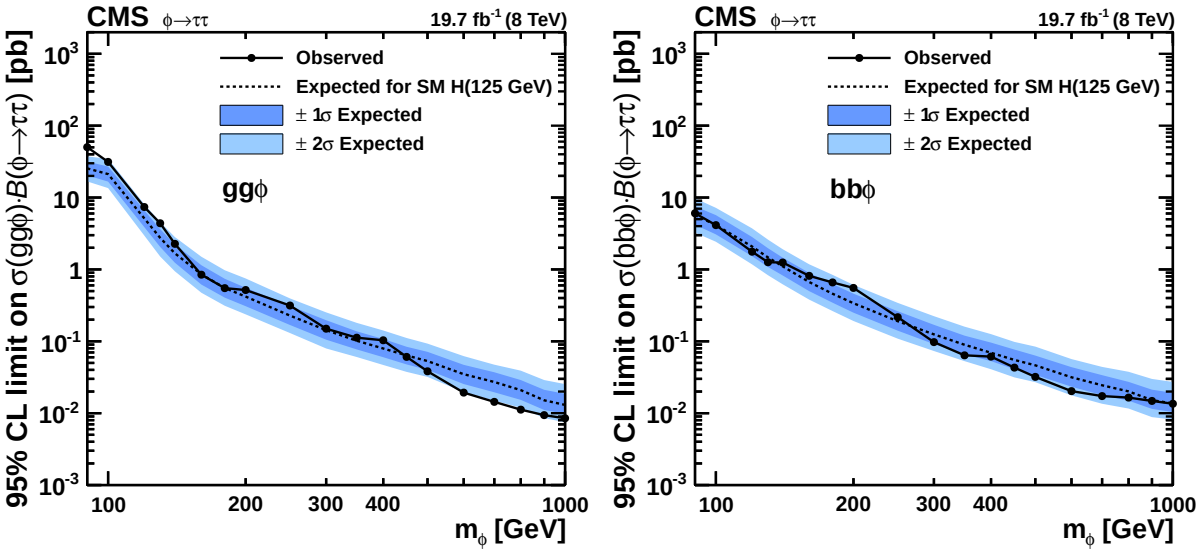


Figure 2.21: *Asymptotic CLs limit on $\sigma(gg\phi) \cdot B(\tau\tau)$ (left) and $\sigma(bb\phi) \cdot B(\tau\tau)$ (right). The associated Higgs production with b-quarks $\sigma(bb\phi) \cdot B(\tau\tau)$ (left) and gluon fusion $\sigma(gg\phi) \cdot B(\tau\tau)$ (right) is profiled. Expected limit is shown for background contribution including a SM Higgs boson at 125 GeV. Calculated for 8 TeV data only. From Ref. [108].*

This analysis is the starting point to the one presented in the thesis in chapters 6 and 7, my contributions are compared to this baseline analysis in section 7.5.

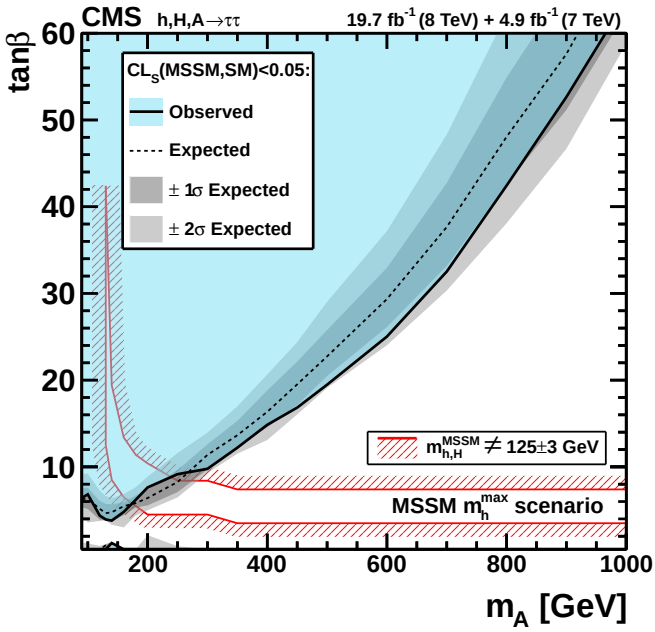


Figure 2.22: *MSSMvsSM limit in the MSSM m_h^{max} scenario. At each $(m_A - \tan\beta)$ point a Hypothesis test is performed testing the MSSM (A+H+h+BKG) hypothesis against the SM (hSM+BKG) hypothesis. From Ref. [108].*

Chapter 3

Experimental setup : Large Hadron Collider and CMS apparatus

Chapter content

In this chapter I introduce the experimental setup: the Compact Muon Solenoid (CMS) experiment [12] hosted at the Large Hadron Collider (LHC) [13] at CERN. First, a brief description of the LHC machine is done in section 3.1. The CMS sub-detectors are described in section 3.2, followed by the reconstruction methods of the objects used in the physics analyses in section 3.3.

3.1 The Large Hadron Collider

The LHC is an acceleration and collision complex located at CERN near Geneva, in the border between France and Switzerland. It is installed in a 26.7 km long circular tunnel, placed between 50 and 170 m underground, which hosted the former LEP (Large Electron-Positron) collider. In order to accelerate and collide hadrons (protons or heavy ions) the LHC is equipped with high frequency accelerating cavities, quadrupole magnets and 1232 superconducting dipole magnets, maintained at 1.9 K with superfluid Helium, operating at 8.3 Tesla to bend the hadron beams trajectory. Two beam pipes are installed inside the tunnel, the hadrons circulating in opposite directions. Four main experiments are located in different collision points at the LHC : ALICE, ATLAS, CMS and LHCb.

Two general and multipurpose experiments, ATLAS ("A Toroidal LHC ApparatuS") [11] and CMS ("Compact Muon Solenoid") [12], study the Standard Model physics and one of their main goals in their physics program is to discover and study the properties the Higgs boson and the physics beyond Standard Model. ALICE ("A Large Ion Collider Experiment") [109] is a heavy ions and quark-gluon plasma physics dedicated experiment. And LHCb [110] main goal is the study of the b-physics and the CP-violation. Figure 3.1 shows a general view of the LHC with the 4 experiments.

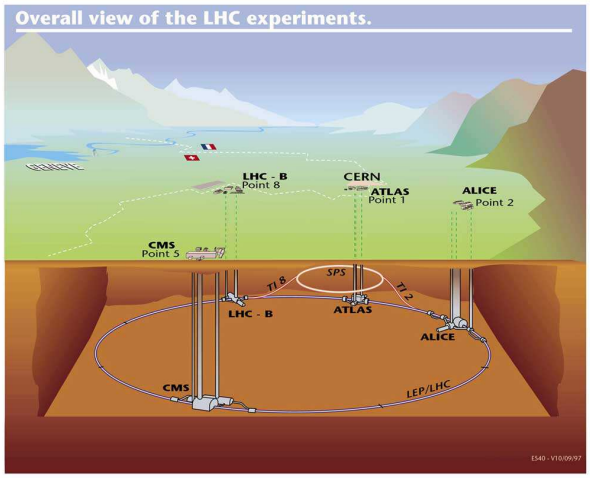


Figure 3.1: *General view of the LHC.*

3.1.1 General presentation

Before entering the LHC, the protons or heavy ions should be progressively accelerated in the acceleration complex showed in figure 3.2 . The first accelerator system is a linear accelerator (LINAC) that injects the hadron bunches in the Proton Synchrotron (PS) which raises their energy up to 26 GeV. Then the bunches are injected into the Super Proton Synchrotron (SPS) where they reach 450 GeV. The beams are then injected into the LHC up to their maximal energy.

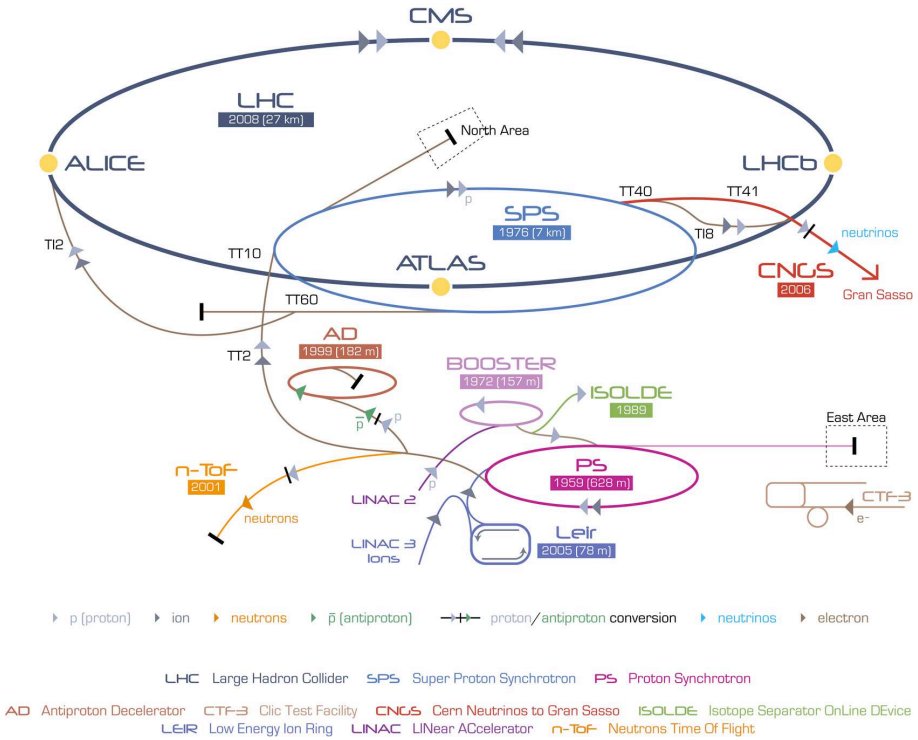


Figure 3.2: *LHC acceleration systems.*

In the nominal "high luminosity" setup (namely the phase preceding the High Luminosity phase HL-LHC starting in early 2020), the LHC should collide bunched beams spaced by 25 ns, each bunch containing about 10^{11} protons with an energy of 7 TeV. The nominal bunch instantaneous luminosity is $10^{34}\ cm^{-2}s^{-1}$. The instantaneous luminosity is given by :

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n \beta^*} F \tag{3.1}$$

where $\gamma = E_{beam}/m_p$ is the Lorentz factor of the accelerated hadrons, f the bunch revolution frequency, k_B the number of bunches per beam, N_p the number of protons per bunch and F a reduction factor due to a non flat crossing angle. β^* and ϵ_n are beam physics related parameters : β^* is the value of the betatron function relative to the transverse beam size at the interaction point, and ϵ_n is the normalized transverse beam emitance. In the ATLAS and CMS case $\beta^* \approx 0.5\ m$ and $\epsilon_n = 3.75\ \mu m$. In the table 3.1 some nominal

as well as the maximum reach high luminosity LHC configuration parameters are listed for proton-proton collisions.

Parameter	Symbol	Nominal	25 ns	50 ns
Energy per nucleon	E (TeV)	7	7	7
Luminosity	$\mathcal{L}(cm^{-2}s^{-1})$	1×10^{34}	9×10^{34}	9×10^{34}
Bunch separation	ns	25	25	25
Number of bunches	k_B	2808	2808	1404
Number of particles per bunch	N_p	1.15×10^{11}	2.2×10^{11}	3.5×10^{11}
β^* value at IP	β^* (m)	0.55	0.15	0.15
Number of collisions per bunch crossing	n_c	19	169	344

Table 3.1: *Some LHC machine nominal beam parameters ant the corresponding maximum HL-LHC performance reach for the configurations with 25 ns and 50 ns bunch spacing. From Ref. [111].*

The rate of events dN/dt (s^{-1}) of a given physical process is related to the luminosity $\mathcal{L}$ $(cm^{-2}s^{-1})$ and the cross section σ (cm^2) of the given process by :

$$dN/dt = \mathcal{L} \cdot \sigma \tag{3.2}$$

So for a given data taking period, the number of physical processes is then proportional to the integrated luminosity $L = \int \mathcal{L}$. The usual physical processes produced in the LHC collisions have a cross section 12 orders of magnitude larger than the Higgs boson production. Figure 3.3 shows the different cross sections for the processes involved in proton-proton collisions. A large instantaneous luminosity is then needed in order to produce enough interesting physical events. Therefore additional soft scattering collisions at the collision time, called pile-up, represents one main challenge in the trigger, reconstruction and data analysis.

The LHC start up was scheduled in September 2008, but after an accident due to a failure between two magnets, the LHC operations stopped until end of 2009. At that time the *pp* beams circulated with an energy of 450 GeV and collided at the center-of-mass energy $\sqrt{s} = 900$ GeV. Few weeks later, another data taking period followed at $\sqrt{s} = 2.36$ TeV, which was mainly used for detector commissioning tasks. In 2010, collisions with the record center-of-mass energy of $\sqrt{s} = 7$ TeV were achieved (3.5 TeV per beam), and during 2010 and 2011 around $L = 5.7fb^{-1}$ was delivered by the LHC with an instantaneous luminosity reaching 3.3×10^{33} $(cm^{-2}s^{-1})$. In 2012, the energy of the collisions was increased to $\sqrt{s} = 8$ TeV with an instantaneous luminosity reaching a maximum of 7.73×10^{33} $(cm^{-2}s^{-1})$. During the first three years of data taking the LHC bunch spacing was gradually reduced from 150 ns in 2010 to 50 ns in 2012. At the end of 2012 the first long shutdown (LS1) started, where accelerator and detector upgrades take place to achieve nominal data taking conditions at the start up in 2015 with a high

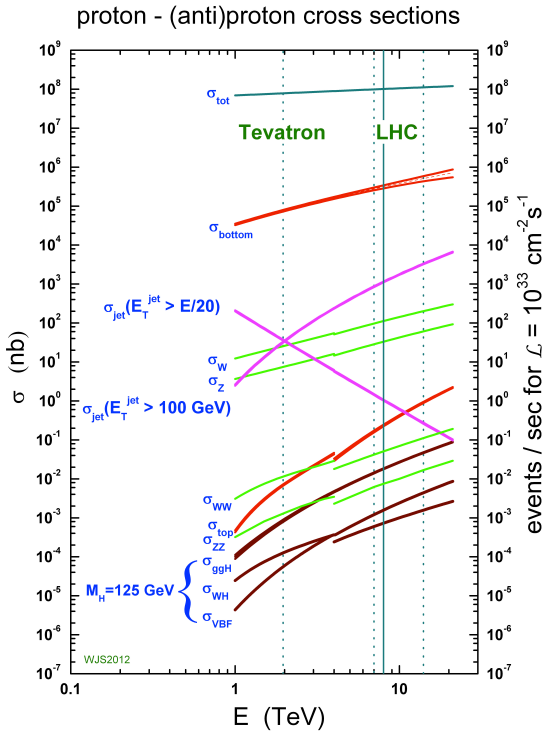


Figure 3.3: *Cross sections for typical processes in proton-proton collisions.*

luminosity configuration [111]. The total recorded luminosity by CMS during the Run 1 (before LS1) is showed in the figure 3.4. The analysis described in this thesis uses the data taken during 2011 and 2012 corresponding to a total integrated luminosity of about $25 fb^{-1}$.

Some dedicated data taking periods (1 month per year) are reserved for heavy ions (Pb-Pb) collisions operating at $\sqrt{s} = 2.76$ TeV. These collisions covers the quark-gluon plasma physics program at the LHC.

3.2 The Compact Muon Solenoid experiment

3.2.1 General presentation

The Compact Muon Solenoid (CMS) detector is one of the four big experiments at the LHC. The physics program of CMS covers a wide range: the search of the Standard Model Higgs boson, precision measurements of the known Standard Model physics, top quark physics, search for new physics beyond the Standard Model and heavy ions and quark-gluon plasma physics. CMS was built in order to perform such polyvalent quests. The CMS construction started in 2006 in the collision point number 5 located in Cessy (France). All the detector pieces were transported around 100 m underground in a cavern previously designed to host CMS.

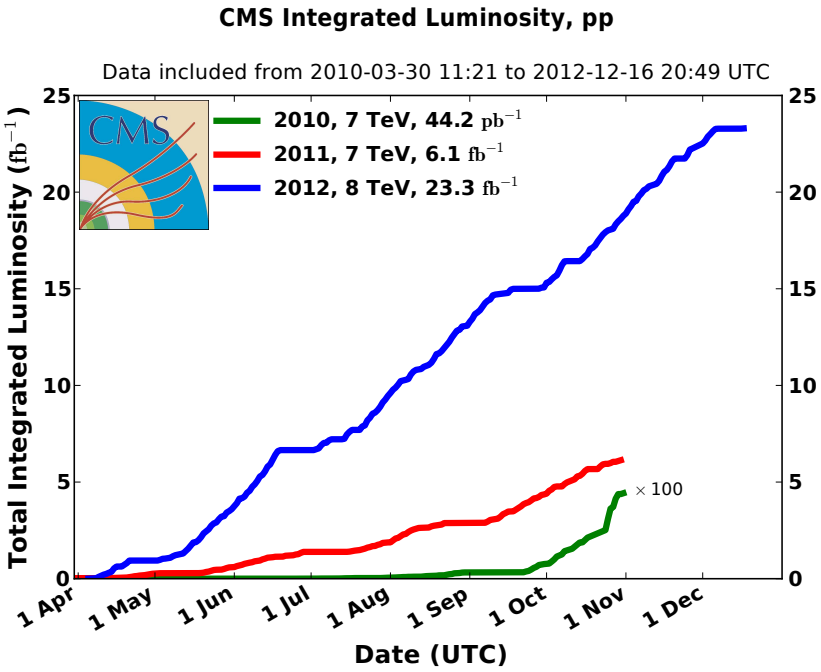


Figure 3.4: *Integrated luminosity delivered to CMS for p-p collisions during Run 1, for 2010 (green), 2011(red) and 2012 (blue).*

The design of the CMS detector is driven by the LHC physics program. The search and properties measurement of the Higgs boson play one of the main roles. As described in the previous chapter the newly found Higgs boson with a mass of 125 GeV can decay in a wide variety of final states. The main channels for the discovery being $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$, an electromagnetic calorimeter with high resolution and good granularity is required. For the searches in the channels $H \to b\bar{b}$ and $H \to \tau\tau$, a performant tracking system is needed to have an efficient b-tagging and hadronic tau reconstruction. A good lepton identification and isolation is needed in all the channels with final state leptons, specially for the channel $H \to WW^* \to 2l2\nu$.

The CMS program includes the search for new physics at the TeV scale. Many of the signatures could lead to very energetic muons or electrons, in a high instantaneous luminosity environment, so a good separation of the charged particles is required and a performant muon system is needed. To cope with these requirements, a high granularity tracker with pixel detectors near the beam was designed, ensuring high momentum and space resolution. In order to separate the charged particles, CMS solenoid generates an intense magnetic field of 3.8 Tesla, which provides with the muon systems a good muon momentum resolution and charge identification up to $p \sim$1 TeV.

Transverse missing energy originates from the presence of neutrinos in the final state and from instrumental limitations, but in many new physics scenarios, it can be created by the presence of new particles. Consequently a good jet reconstruction and performant missing energy estimation is required to be sensitive to these signatures. CMS is equipped with hermetic calorimeters able to reconstruct the jets and performant in the transverse

missing energy estimation.

Also the LHC instantaneous luminosity imposes stringent requirements on the detector performances during the data taking periods. The detector response and trigger have to cope with the LHC collisions rate, 40 MHz for a bunch spacing of 25 ns. The short time between two bunch-crossing (as low as 25 ns) requires very fast readout and trigger systems.

With an instantaneous luminosity expected to reach 2.5×10^{34} or 5×10^{34} $cm^{-2}s^{-1}$ in the very high luminosity scenario, a peak pile up of approximatively 100 extra interactions is foreseen during the collisions. This extra activity increases the detector occupancy and degrades the isolation and detector performances. In order to avoid a superimposition of events coming from two following bunches, the readout electronics must be very fast and the tracker very segmented. The effect of pile up can be reduced also by using high granularity detectors with good time resolution. A great number of channels has to be used and has to be synchronized.

Moreover the detector is subject to radiation damage mostly in the forward regions due to the large flux of particles coming from the interaction region and it must be built accordingly. Another source of radiation damage is the beam-halo which provides a non negligible rate of particles hitting the sub-detectors.

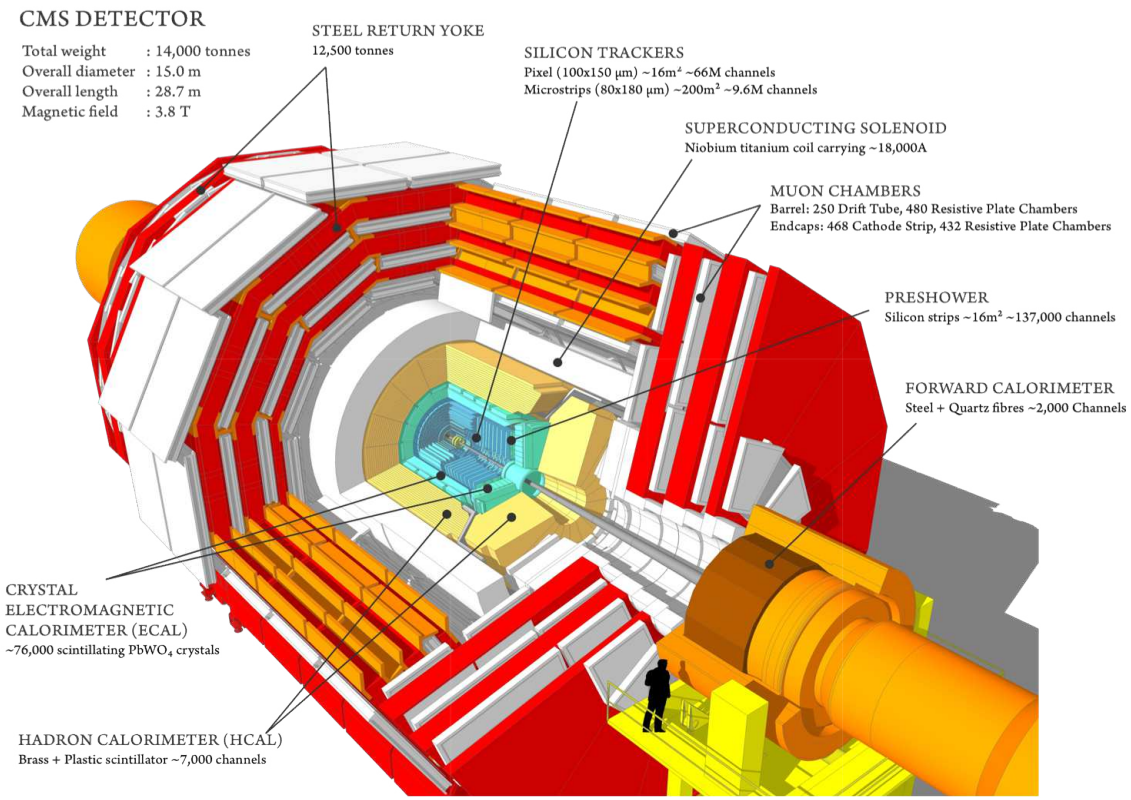


Figure 3.5: *General view of the CMS apparatus and its sub-detectors.*

As shown in figure 3.5, CMS geometry is cylindrical, featuring a 4π coverage around the interaction point. It is 21.6 m long and has 15 m of diameter and its weight is about

14000 tons.

The CMS detector is divided in 2 geometrical regions : the central part called Barrel corresponds to the region where $|\eta| < 1.5$ and the forward region called End-cap where $|\eta| > 1.5$. In order to reduce the bremsstrahlung radiation of the charged particles between the tracker and the calorimeters, the magnet encompasses the tracker system, the electromagnetic calorimeter and the hadron calorimeter. A Pre-Shower (PS) sampling calorimeter is placed in the Endcap region before the electromagnetic calorimeter. This choice implied a compact detector having calorimeters made with very dense materials and an intense magnetic field.

3.2.2 CMS coordinate system

Figure 3.6 shows the coordinate system adopted in CMS. The interaction point is in the center of the detector. The x axis points to the center of the LHC ring. The z axis is tangent to the beam trajectory pointing to the Jura, so the y axis is perpendicular to the x axis pointing upwards to complete the right handed coordinate system. We can transform this cartesian system into spherical coordinates: the distance r to the z-axis, the azimuthal angle ϕ and the polar angle θ. The rapidity ρ defined as :

$$\rho = \frac{1}{2} ln \left(\frac{E + p_z}{E - p_z} \right) \tag{3.3}$$

For a particle having the momentum $\vec{p}$, the angle θ is the angle between $\vec{p}$ and the z-axis. In the relativistic approximation where $\beta \to 1$ the rapidity converges to the pseudorapidity:

$$\rho \to \eta \equiv -ln \tan \left(\frac{\theta}{2} \right) = \frac{1}{2} ln \left(\frac{p + p_z}{p - p_z} \right) \tag{3.4}$$

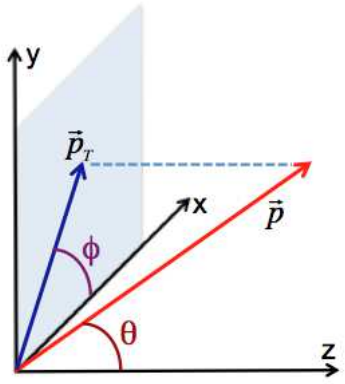


Figure 3.6: *CMS coordinate system.*

The transverse momentum $\vec{p_T}$ is defined as the projection of $\vec{p}$ to the (xy) plane. Its magnitude is then :

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{3.5}$$

The transverse energy is defined as $E_T = E \sin\theta$. The distance ΔR between two particles with pseudorapidities $\eta_{1,2}$ and azimuthal angle $\phi_{1,2}$ is defined as:

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2} \tag{3.6}$$

which is invariant under longitudinal boosts.

3.2.3 CMS solenoid

One of the main features of CMS is its powerful solenoid that delivers a 3.8 Tesla field providing the bending power needed to have a high momentum resolution of the charged particles coming from the collisions in a compact volume. This magnet, 13 m long and having 6 m of diameter, surrounds the tracker, the electromagnetic and hadron calorimeters described next. The main characteristics of the solenoid are given in table 3.2.

The return field up to 2T follows the return yoke made of 1.5 m thick iron slabs interposed between the muon chambers system. In this way, the muon momentum measurement can be combined from the trajectories in the tracker and the muon chambers. The bending power provided by the magnet ensures a good P_T resolution and separation of the charged particles required in the CMS physics programme.

Field	3.8 T
Inner radius	5.9 m
Length	12.9 m
Number of spires	2168
Current	19.5 kA
Stored energy	2.7 GJ

Table 3.2: *CMS solenoid parameters.*

3.2.4 Tracking system

The tracking system is used to reconstruct the charged particle trajectories as well as the primary and displaced vertices produced during the hadron collisions. The particle flux in the collisions increases near the interaction point. Given the beam collisions at the LHC, the main challenges in the tracker conception are a good granularity, a quick response and radiation hardness.

Driven by this requirements, the CMS tracker is entirely based on silicon detectors, where the charged particles are detected by the ionization of silicon cells. The association of many hits in different cells are used to build the particle trajectory and measure its momentum. The tracker consists on a hermetic detector of 5.8 m long and 1.1 m of diameter. It covers the region up to 2.5 in $|\eta|$. It is built in two main parts : the pixel

detector and the silicon strip tracker. In figure 3.7 a transverse view in the $(r-\varphi)$ plane is shown.

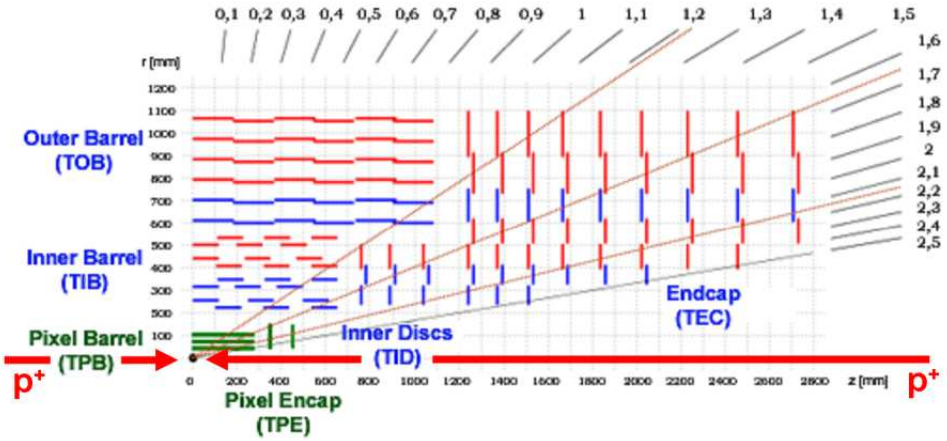


Figure 3.7: *Longitudinal view of the CMS tracking system. The pixel detector : Tracker Pixel Barrel (TPB) and Tracker Pixel Endcap (TPE) and the strip tracker : Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), Tracker Outer Barrel (TOB) and Tracker EndCap (TEC).*

Pixel detector

The pixel detector is made of three cylindrical layers (Tracker Pixel Barrel) 53 cm long at 4.4, 7.3 and 10.2 cm from the beam axis. They are complemented with two endcap disks (Tracker Pixel Endcap) on each side placed at $|z| = 34.5$ and 46.5 cm with a radius of 6 and 15 cm. Each pixel is sized $100 \times 150\ \mu m^2$ achieving a spatial resolution of about 10(20) μm in the $r-\varphi$ plane (z axis).

Strip tracker

The strip tracker surrounds the pixel detector, it is made of layers of 320 μm thick micro-strip sensors covering the region $r = 20$ to 116 cm. It is divided in various sub-systems as shown on figure 3.7. The Tracker Inner Barrel and Disks up to r=55 cm and $|z| < 65$ cm are composed by 4 layers and 3 disks in each side. They have variable strip pitch of 80-100 mm in the TIB and 100-141 μm mean pitch in the TID, giving a single point resolution of 23 μm to 35 μm.

Those 2 systems are surrounded by the Tracker Outer Barrel having this time 500 μm thick silicon sensors extending to the region r=116 cm and $|z| < 118$ cm. It is disposed in 6 layers having a variable pitch of 183 μm to 122 μm leading to a single point resolution of 35 μm to 53 μm.

In the region $120 < |z| < 280$ cm and $22.5 < |r| < 113.5$ cm, the Tracker EndCap is arranged in 18 outer disks having 7 rings of micro-strip detectors 320 μm thick on the inner 4 rings and 500 μm thick on rings 5 to 7 and with 97 μm to 184 μm average pitch.

The strip tracker have spacial resolution of about 40 to 60 μm in the $(r - \varphi)$ plane and 500 μm along z.

The tracker system is built following a compromise between granularity and reconstruction performances. The cooling system needed to deal with the hard radiations and the electronics add a non negligible material budget the charged particles have to go through, which leads to bremsstrahlung radiation and energy loss. In the figure 3.8, the tracker system material thickness in radiation lengths is shown for each of the subsystems described above.

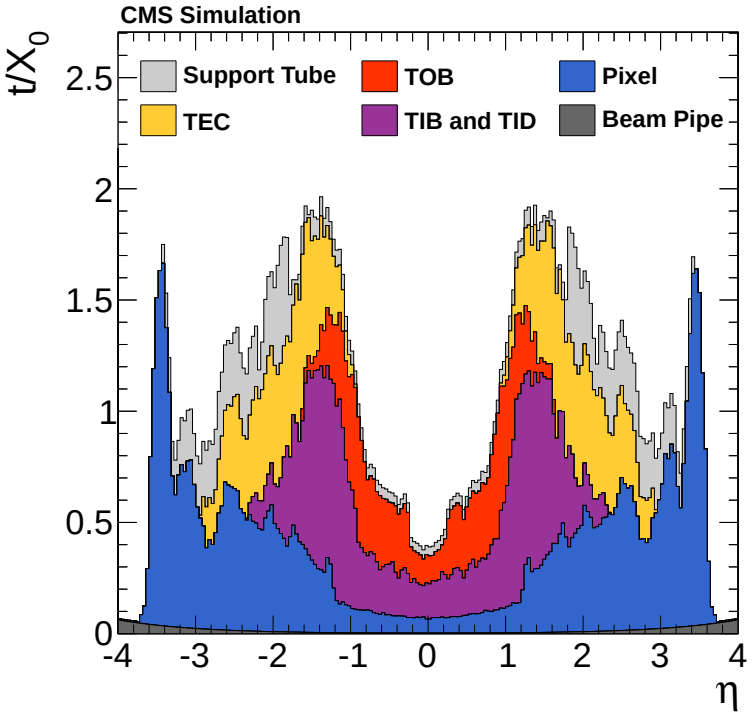


Figure 3.8: *CMS Tracker material budget, in units of radiation length, split in the different tracker subsystems. The beam pipe and support tube material budget is also shown.*

The tracker system is essential for the event reconstruction in CMS. The momentum of charged particles as well as the energy estimation of neutrals in the jets using the Particle Flow algorithm, combines the information from all sub-detectors. This technique relies essentially on the tracker measurements as we will see in sections 3 and 4. So the tracker system plays a major role on the charged particle reconstruction (including muon and electron momenta), the tau reconstruction in its different hadronic decays, as well as the estimation of the missing transverse energy.

3.2.5 Electromagnetic calorimeter

Two of the main channels in the CMS physics program in the context of the Higgs boson are the 4 leptons and the two photons final states. A very good resolution on

the electrons and photons energy is crucial. The typical resolution needed to perform this analysis is about 1% for a 100 GeV particle. Also the LHC collisions rate and instantaneous luminosity requires a radiation resistant, fine segmented as well as a quick response calorimeter.

The CMS electromagnetic calorimeter (ECAL) is a homogeneous, hermetic calorimeter made of 75848 lead tungstate ($PbWO_4$) scintillating crystals (61200 in the barrel and 7324 in the endcap region). The crystals are disposed in a semi-projective arrangement as we can see in figure 3.9. A Pre-Shower in placed in front of the endcap crystals in order to distinguish π^0 from γs.

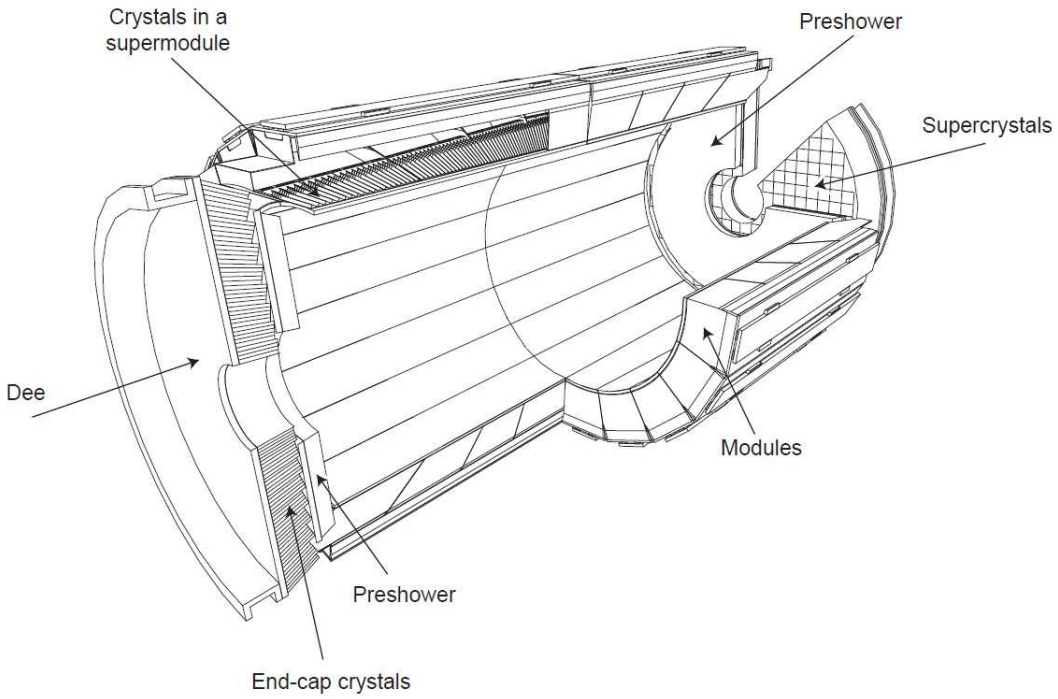


Figure 3.9: *CMS electromagnetic calorimeter.*

The lead tungstate crystals were chosen because of their high density (8.28 g/cm^3), their short radiation length ($X_0 = 0.89$ cm), their Molière radius ($R_M = 2.2$ cm), and the scintillation decay time of the same order of the LHC bunch crossing : about 80% of the light is emitted in 25 ns. The light yield in this material of 4.5 γ/MeV depends on the temperature so a cooling system ensures the crystals temperature to be of 18 ± 0.05^oC. The crystals have a trapezoidal shape covering an angle around 0.0174×0.0174 in (η, ϕ). They are 23 cm long corresponding to 25.8 X_0.

The barrel region covers $0 < |\eta| < 1.48$. In order to collect the crystal light, avalanche photo diodes (APD) are used. In the endcap region, $1.48 < |\eta| < 3$, high radiation resistant vacuum phototriodes are used to recover the signal photoelectrons.

The pre-shower is a sampling calorimeter covering the region $1.653 < |\eta| < 2.6$. It is made with two types of layers : lead radiators and silicon strip sensors. The silicon strip

sensors are oriented following the x and y axis, they measure the electromagnetic shower initiated by the lead radiators $2X_0$ and $3X_0$ thick. Thus about 95% of single incident photons start showering before the second sensor plane.

The electromagnetic calorimeter energy resolution is given by :

$$\left(\frac{\sigma(E)}{E}\right) = \left(\frac{S}{\sqrt{E}}\right) \oplus \left(\frac{N}{E}\right) \oplus C \tag{3.7}$$

Where the energy E is in GeV, S (2.8%) is the stochastic response term, N (0.125) is the noise term and C (0.3%) a constant term [112]. The fluctuation in the number of produced and collected electrons is included in the stochastic term, while the noise term is an overall term due to electronic noise and to pile-up events. The constant term is related to the calibration of the calorimeter. The achieved resolution is then less than 1% for electrons with an energy higher than 15 GeV, reaching 0.6% for 40 GeV electrons. In figure 3.10 the intrinsic ECAL energy resolution as a function of the electron energy is presented.

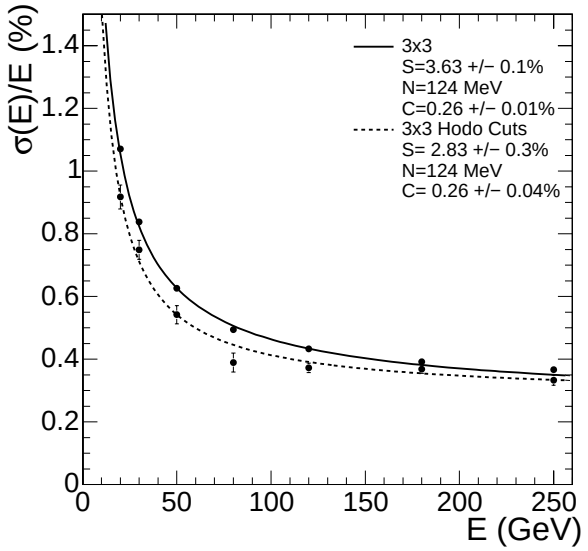


Figure 3.10: *ECAL energy resolution as a function of electron energy as measured from a beam test. From Ref. [112].*

Because of the high radiation environment during the LHC collisions, the ionizing radiation can create color centers due to oxygen vacancies and impurities in the lattice. This provokes a loss in transparency of the ($PbWO_4$) crystals. To correct for this, a monitoring of the crystals transparency is performed with laser light pulses injection asynchronous to the bunch crossing allowing the extraction of calibration constants.

3.2.6 Hadron calorimeter

The hadron calorimeter (HCAL) completes the calorimetric system of CMS. It measures the energy deposited by the neutral and charged hadrons that go through the tracker and

the electromagnetic calorimeter and it is used with the ECAL to estimate the missing transverse energy (MET). The HCAL sub detectors completely surround the ECAL and they are divided in 4 parts. A barrel region (HB) and two end-cap regions (HE) are placed inside the solenoid range, between 1.7 and 2.95 m from the beam axis and cover the rapidity range up to $|\eta| = 3$. In order to ensure the hermeticity and recover the long hadronic showers going through the inner part, an outer hadron calorimeter (HO) is placed outside the magnet in the region up to $|\eta| = 1.26$. To detect the hadronic showers in the very forward regions, a forward calorimeter (HF) using Cherenkov technology covers the region up to $|\eta| = 5$. Figure 3.11 shows an overview of the HCAL subsystems geometry.

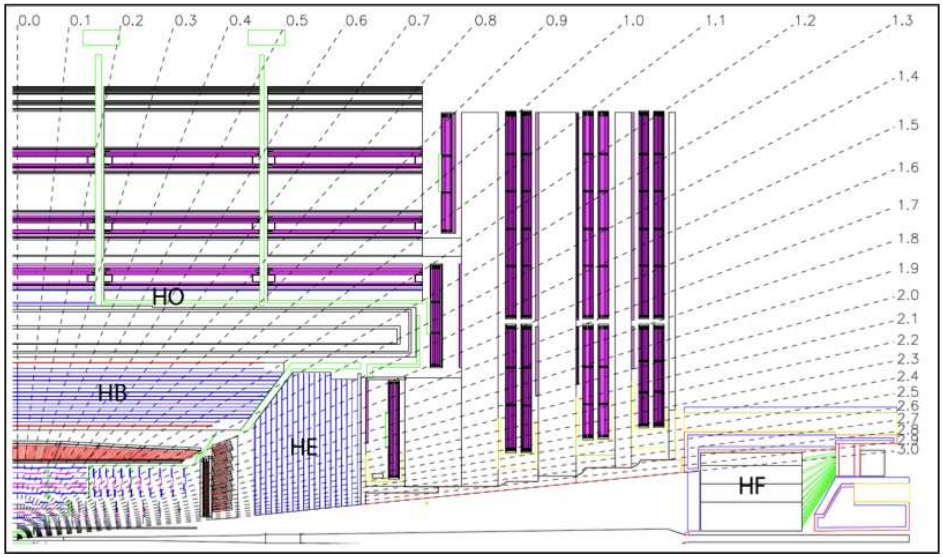


Figure 3.11: *Longitudinal view of the CMS hadron calorimeter and its sub detectors : Barrel (HB), Endcap (HE), Outer (HO) and Forward (HF) calorimeters.*

The hadron calorimeter is a sampling calorimeter alternating absorbers and active mediums. Due to the high magnetic field environment, non-ferromagnetic materials have been chosen : brass (70%Co, 30%Zn) and stainless steel absorbers are interposed with plastic tiles scintillators. The magnet acts as an additional absorber layer for the outer calorimeter.

Barrel (HB) and Endcaps (HE)

The active medium in the barrel calorimeter is Kuraray SCSN81 plastic scintillator tiles, 3.7 mm width, except for the first layer made of 9 mm thick Bicron BC408. 17 layers of tiles are placed in a quasi-projective direction to the interaction point, alternating absorber plates 50,5 mm to 75 mm width. The emitted light is collected by wavelength-shifting optic fibres and then collected and read out with hybrid photodiodes (HPD). The barrel region extends to the region $|\eta| < 1.4$.

The endcap part covers the region $1.3 < |\eta| < 3$, avoiding dead zones between the barrel and endcap. In this part there are 19 layers of absorber plates 78 mm width and scintillator tiles 3.7 mm width.

The tiles of scintillator are grouped into variable size towers with a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ at central rapidity $|\eta| < 1.74$ and 0.09×0.174 to 0.5×0.174 at forward rapidity $1.74 < |\eta| < 3$.

Outer calorimeter (HO) and Forward calorimeter (HF)

The outer calorimeter is placed outside the magnet, acting as a "tail catcher" of the hadronic showers in the region $|\eta| < 1.3$. It is placed within the muon system. The magnet is used as the additional absorber with a thickness equal to $1.4/sin\theta\ \lambda_I$[1]. The total detector depth is then $\approx 11\lambda_I$.

The forward calorimeter (HF) is placed at 11.2 m from the interaction point along the z axis, it covers the region $3 < |\eta| < 5$. It is made with stainless steel absorbers (78 mm width) and radiation resistant quartz fibers (3.7 mm width) that collect Cherenkov light through radiation hard photomultiplier tubes (PMT).

The hadronic calorimeter energy resolution is given by [112]:

$$\left(\frac{\sigma(E)}{E}\right) = \left(\frac{111.5\%}{\sqrt{E}}\right) \oplus 8.6\% \tag{3.8}$$

As we will see later, the energy resolution of hadrons is improved significantly using the Particle Flow algorithm.

3.2.7 Muon detector

Muons are the only particles able to go through the previous detectors (neutrinos are not detectable in CMS), their signature is then very clean. Muons are present in the final state of many important analysis ($H \to ZZ^* \to 4\mu, H \to \tau\tau \to \tau_h\mu$, ...). The CMS muon system is designed to detect and measure the muon momentum in the kinematic range of the LHC within all the geometrical acceptance. It plays also an important role in the trigger setup.

The muon system is made of 3 types of gaseous detectors in order to cover all the surface and to deal with the different radiation environments : Drift Tube chambers (DT), Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC). They cover the region $|\eta| < 2.4$. In figure 3.12 a longitudinal view of CMS muon system is shown.

The barrel region, covering the pseudo rapidity range up to $|\eta| < 1.2$ is placed within the magnet return yoke. In this region the muon flux is smaller and the magnetic field is weaker, the Drift Tube chambers (DT) technology has been chosen. The hit position in each DT is reconstructed by measuring the drift time of the avalanche electrons originated from the muon crossing. The iron return yoke acts as an absorber of particles and reduces the muon signal background. Four concentric stations are equipped with 250 muon chambers. The first muon station is at a radial distance of about 4.5 m and the last is at about 7 m. Each station is divided in 5 rings following z axis, each ring being split in 12 sectors of 30^0 in ϕ. The drift tubes measure the muon position in the (r, ϕ)

[1] nuclear interaction length, X_0 equivalent for strong interaction.

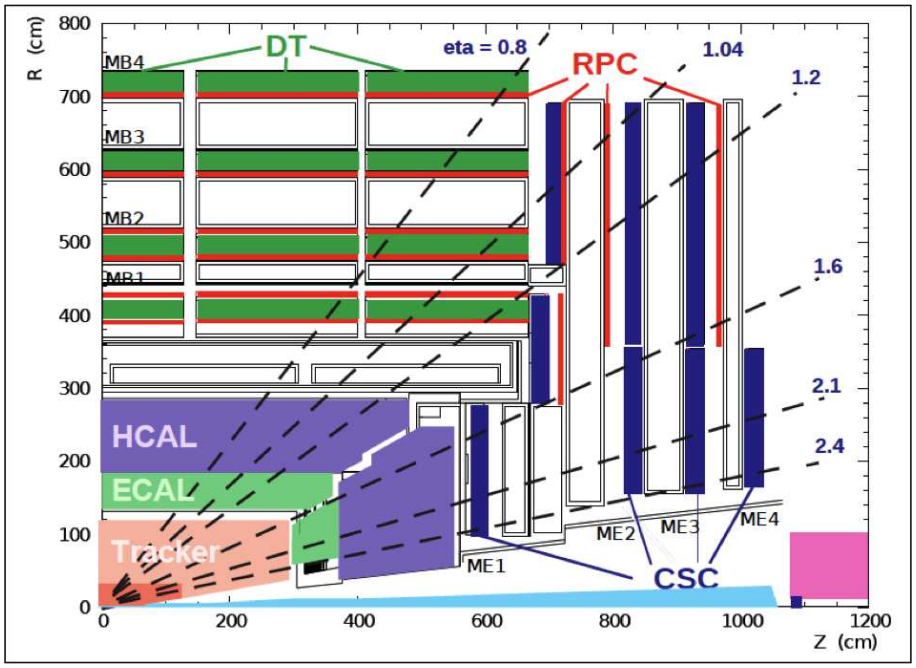


Figure 3.12: *Longitudinal view of CMS muon detectors. Drift Tube chambers (DT), Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC) are shown.*

plane and in the z direction. The last station does not contain a z measuring plane. The design chosen for the barrel provides a single point resolution of about 200 μm and a ϕ precision better than about 100 μm in position and about 1 mrad in direction.

In the endcap region the activity is more important and the magnetic field is stronger and non homogeneous. Radiation robust Cathode Strip Chambers are then used in the region $0.9 < |\eta| < 2.4$. Each CSC consists of closely spaced anode wires placed between two cathodes. The avalanche electrons from the gaz ionization due to the muon crossing is collected by the anode wire, originating an image charge on the cathode strips. Each endcap is equipped with 468 CSCs with trapezoidal shape distributed in 4 stations installed perpendicularly to the beam line. Each station is composed by six layers of CSCs, with a point resolution of about 75-150 μm in the ϕ coordinate and 200 μm in the r coordinate.

Because of the uncertainty in the eventual background rates and in the ability of the muon system to measure the correct beam-crossing time when the LHC reaches full luminosity, a complementary dedicated trigger system, consisting of Resistive Plate Chambers (RPC), was added in both the barrel and endcap regions. The RPCs are double-gap chambers, operated in avalanche mode to ensure good operation at high rates. They are disposed in 6 layers in the barrel region and 3 in the endcap region. Their fast response, shorter than 25 ns, and good time resolution provide a fast, highly segmented trigger and are also used to solve ambiguities when reconstructing tracks from the hits in the muon chambers.

For transverse momentum lower than 100 GeV, the resolution is tracker dominated. At higher momentum the resolution is dominated by the muon system response. The momentum resolution for a muon with an energy of 1 TeV is about 10%.

3.2.8 Data acquisition and trigger system

The first step of any physics analysis in CMS is the trigger selection. At this step the final reconstructed objects are preselected. As explained above, the LHC nominal design provides proton bunch crossings separation of 25 ns, leading a collision rate of 40 MHz. It leads to around 10^9 events per second, or an interaction rate of 1 GHz. Each event having a size of 1 MB, it is impossible to store and treat all the data. CMS storage capacity can accept a maximal output rate of $\mathcal{O}(50)$ Hz. The task of the trigger system is then to filter online the most interesting ~ 300 events over the 40 million per second. This rate reduction is done in two steps in CMS, the *Level 1* (L1) trigger, and the *High Level Trigger* (HLT).

The first level trigger (L1)

The Level 1 trigger does a quick filtering (3.2 μs) of the events with the information provided by the data readout from the front-end electronics. It is housed in a service cavern near the experiment. It provides a rate reduction of $\mathcal{O}(10^3)$, leading to an output rate of about 100 kHz.

As the time to take the L1 trigger decision is around a couple of μs, it is not possible to use the total raw data information. The L1 trigger decision is based on calorimetric (Calorimeter trigger) and muon system (Muon trigger) information running in parallel as well as the correlations between both (Global trigger). The L1 trigger generates, with this information, Trigger primitives (a coarse–granularity and low–resolution estimation of the momentum of particles) above some threshold on transverse energy E_T. Object candidates are then made within a $(\eta - \phi)$ phase space : electromagnetic (L1 EG), muons (L1 Mu), jets (L1 Jets) and taus (L1 Taus). Also some global quantities are constructed as the global electromagnetic energy (in the ECAL and in the HCAL) and the missing transverse energy (L1 ETM). The tracker and the pre-shower detectors are not used in the level 1 trigger. The architecture of the first level trigger is shown in figure 3.13

Being more performant in terms of computing time, the Calorimeter trigger is divided in regions. The Regional Calorimeter Trigger (RCT) builds level 1 candidates from the Trigger primitives in regions of 4×4 towers in η, ϕ, each trigger tower grouping 5×5 crystals of the electromagnetic calorimeter. A card for each trigger tower saves the information in a buffer and starts the trigger primitives reconstruction. The same procedure is done in the hadron calorimeter, so each trigger tower has a total transverse energy associated. ECAL primitives have additional information about the shape of the energy deposit.

L1 Muon candidates are also created from tracks built by a regional algorithm combining local hits in the DT and CSC. The constructed tracks transverse momenta and quality is then estimated. All the hits in the RPC are combined to build additional tracks. Then all tracks are sent to the Global Muon Trigger (GMT) which combines all the information. Additional information from the RCT is used to identify isolated Level 1 Muon candidates.

All candidates are transmitted to the global trigger which applies a menu of L1 trigger algorithms. L1 Jets and L1 Taus will be described in section 4.2.1. Some of the algorithms in the menu can be pre-scaled : just one over n events passing the algorithm will be

randomly chosen to trigger the event. The trigger menu is constantly changing with the increase of the LHC instantaneous luminosity. If the event is accepted by any of the algorithms, a signal is sent to the sub detectors front-end, the data being then transferred to the processors farm where sits the next trigger level : the High-Level Trigger (HLT).

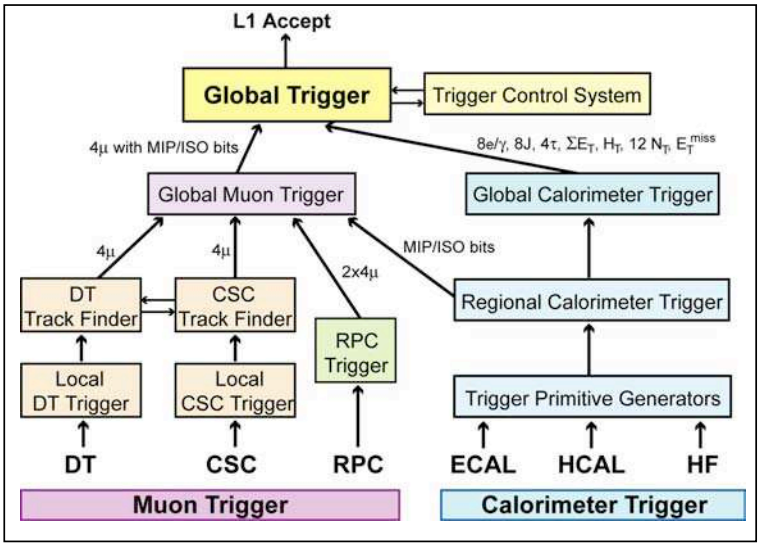


Figure 3.13: *Architecture of the CMS Level 1 Trigger.*

The High-Level Trigger system (HLT)

The High-Level Trigger system uses more sophisticated algorithms, similar to the ones used for the offline reconstruction. It reduces the event rate by another factor of $\mathcal{O}(10^3)$, leading to an output rate of about 100 Hz. The data coming from the readout buffers are transferred to a processor farm with 16000 CPUs. This trigger level is composed by a list of filters increasing in complexity. In the first filters, also called level 2 filters, a "local reconstruction" is made initially using the full information of the muon system and the calorimeters. Then, level 2.5 filters introduce algorithms using the tracker pixels information and finally level 3 filters use the full detector information, including the pre-shower and the tracker strips data.

At this step, high level objects, hadronically decaying taus and b-tagged jets, can be searched for. Also regional reconstruction around L1 seeds allows for a maximal exploitation of the sub-detector granularity while keeping the timing low. Eventually, the HLT uses the full event data for the decision to keep an event or not. The last stage of HLT processing does reconstruction and event filtering in order to make datasets of different physics signatures. In figure 3.14 an overview of the CMS data acquisition system is shown.

Unlike the Level 1 trigger, the time to process an event is not constant and varies depending on the algorithms applied. The mean time per event is around 60 ms, some events taking up to 1 second.

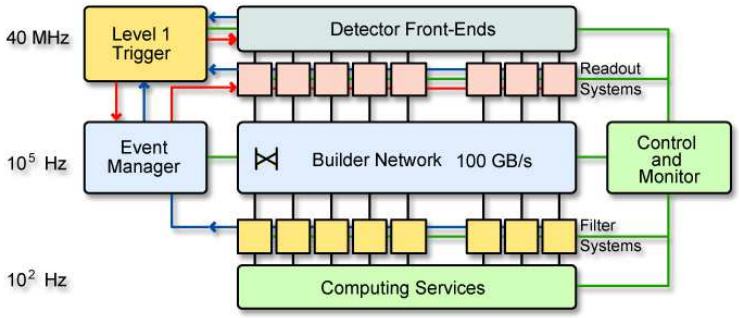


Figure 3.14: *Architecture of the CMS data acquisition system.*

3.3 Object reconstruction in CMS

In order to identify the physics objects produced in the collisions one technique in particular is widely used within CMS, the Particle Flow algorithm. The particle flow approach makes best use of the information from the sub-detectors to identify the particles arising from the collision.

3.3.1 Tracks and interaction point

Track reconstruction

The track reconstruction in CMS is of primary importance. It is used to measure the momentum of the charged particles and to determine the production vertex of each of them. Moreover a good separation between reconstructed tracks is required to separate the charged constituent inside a jet, a key feature within the Particle Flow algorithm. This is achieved via the good performance of the track reconstruction and the considerable magnetic field. A detailed description of the track reconstruction in CMS can be found in Ref. [113].

First, hits in the pixel and strip detectors are reconstructed locally. The track reconstruction uses these hits to obtain an estimation of the momentum and position parameters of the charged particles. This is done by the tracking software called Combinatorial Track Finder (CTF) which is an extension of the combinatorial Kalman filter algorithm [114,115]. Various iterations are done in order to produce the final collection of reconstructed tracks. In this *iterative tracking* procedure, the tracks having the best quality (see later) are removed in each iteration which simplifies the combinatorial complexity of the next iteration.

The CTF reconstruction software applies a series of six iterations. In the first iteration the tracks with $p_T > 8$ GeV having three pixel hits originating near the proton-proton interaction point are considered. The next iteration recovers the tracks with only two pixel hits and the subsequent iterations are used to find tracks originating far from the beam spot and additional track that were not found previously. At the beginning of each iteration the hits of the tracks found in the previous iterations are not considered anymore. Each iteration is executed in four steps:

- Seed generation: determines the initial track candidates with few hits (2 or 3). It defines the initial estimate of the trajectory and its uncertainty.
- Track finding: based on a Kalman filter, it extrapolates the seed trajectories onto the expected flight path of a charged particle and associates the subsequent found hits to the track.
- Track fitting: a fit is performed to obtain the best estimate of the trajectory parameters and to smooth the track path.
- Track selection: quality criteria (χ^2 of the track fit, number of hits and missing hits in the track) is used to discard badly reconstructed tracks.

The tracking efficiency can be measured directly from data using the Tag-and-Probe technique in $Z/\gamma^* \to \mu\mu$ events. One Tag muon is selected by requiring it to be reconstructed both in the muon system and associated to a track in the tracker. The probe muon is required to be reconstructed in the muon system, and the invariant mass of the di-muon system is required to be in the range $50 - 130$ GeV compatible with the mass of the Z boson. The measured efficiency of the probe track reconstruction in the tracker is presented in figure 3.15: it is higher than 99% in data as in the simulation, the loss in certain regions in η is explained by the crack regions in the tracker system. It is pretty flat as function of the number of reconstructed vertices, with a slight trend in data, meaning that the tracking is relatively pile-up insensitive.

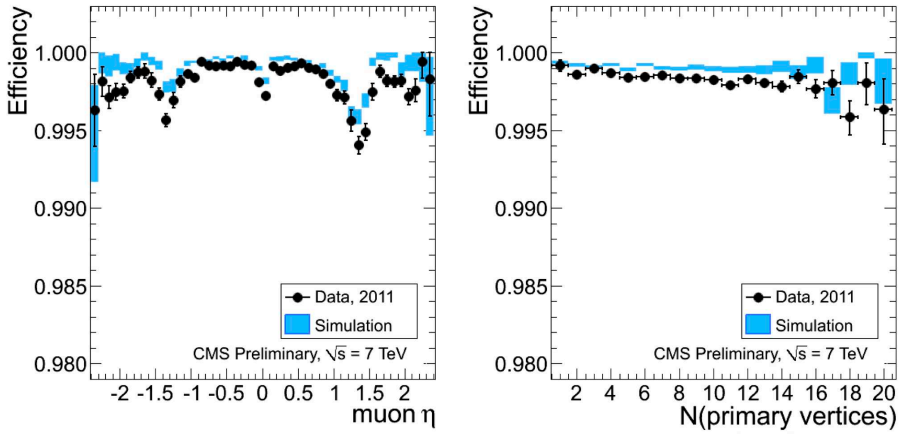


Figure 3.15: *Tracking reconstruction efficiency measured with the Tag-and-Probe technique using muons from Z boson decays, as a function of the muon η (left) and the number of reconstructed primary vertices (right). From [113].*

Interaction point reconstruction

The interaction point reconstruction measures the position of all the proton-proton collision interaction vertices whether it be the primary vertex or vertices from pile-up collisions.

Reconstructed tracks are clustered using the *deterministic annealing* (DA) algorithm [116], which finds the global minimum for a problem with many degrees of freedom. The DA algorithm is comparable to finding a minimum energy state of a physical system through a series of gradual temperature reductions.

Reconstructed vertices should be identified from a fit with at least 4 degrees of freedom, and the distance between its position and the nominal detector center should be smaller than 24 *cm* in z and 2 *cm* in the transverse coordinate r. Out of the selected vertices, the one with the largest $\sum_{tracks} P_T^2$, the sum of the squares of the transverse momenta of tracks associated to that vertex, is chosen as the production vertex of the signal event. All the other reconstructed vertices are considered to come from pile-up.

The primary vertex reconstruction resolution in x and z measured for minimum bias events is less than 20 μm and 25 μm respectively for primary vertices reconstructed with at least 50 tracks. The efficiency of the primary vertex reconstruction is close to 100% when more than two tracks are used to reconstruct the vertex. The estimated loss of efficiency due to pile-up is found to be $< 0.1\%$ for a pile-up mean value of 8.

3.3.2 Particle Flow Algorithm

The particle flow paradigm aims at reconstructing all stable particles individually arising from the collision by combining in an optimal way the information from the different CMS detectors to determine their type and energy [117, 118]. This provides a maximal description of the final state.

The jet reconstruction and energy measurement was typically assigned to calorimetric algorithms. Within the particle Flow approach, the reconstructed stable particles are electrons, muons, photons, charged hadrons and neutral hadrons. The list of individual particles are then used as if they came from a Monte Carlo event generator to built jets. The energy of a jet is on average shared among charged, photons and neutral hadrons in the approximate proportion of 65:25:10. The charged component in the jets are more precisely measured in the tracker, the neutral particles energy is then provided by the calorimeters. However the separation between neutral and charged components in the calorimeter can sometimes be non-trivial.

The Particle Flow particles are also used in the determination of the missing transverse energy (see section 3.3.7), giving an estimation of the energy and direction of the neutrinos and other invisible particles. It also represents a key element in the τ-lepton reconstruction (see chapter 4), as the decay products of the τ-lepton can be looked for in narrow jets.

Particle Flow key elements

The CMS detector is well suited for the implementation of the Particle Flow approach because of the good performance of the CMS tracker, ECAL and Muon systems.

The large silicon tracker is well adapted to separate the different tracks from the charged particles thanks to the uniform 3.8 Tesla magnetic field provided by the superconducting solenoid. This is a key piece in the charged particles disambiguation task. The tracker also provides a precise measurement of the charged particle direction at the production vertex with a very high efficiency, larger than 90% for charged hadrons in jets and

larger than 99% for isolated muons, with a fake rate at the percent level, obtained with the CTF algorithm (see section 3.3.1). The momentum of charged hadrons is measured in the tracker with a higher resolution than in the calorimeters up to several hundreds of GeV. The performant tracking and the magnetic field are then the cornerstones of the Particle Flow approach.

Another key element in the Particle Flow feasibility is the granularity of the electromagnetic calorimeter surrounding the tracker. It allows separating reconstructed photons from charged particles energy deposits. A topological clustering algorithm developed specially for the Particle Flow event reconstruction groups the ECAL crystals. A similar clustering is done for the HCAL towers. First, cells with a local maximum of energy above a given threshold (230 MeV in the ECAL and 800 MeV in the HCAL) provide a seed for growing topological clusters; neighboring cells with energy above twice the standard deviation of electronic noise ($>$ 80 MeV in the ECAL, $>$ 400 MeV in HCAL) are clustered. The calorimeter granularity is exploited by sharing the energy of a crystal between all particle flow clusters according to the crystal-cluster distance.

Linking algorithm

A single particle is expected to have a specific signature in the different sub-detectors: tracks, clusters or muon tracks. These different elements are connected between them by a linking algorithm in order to fully reconstruct a single particle. The linked elements constitutes "blocks" that are then inspected to sort out ambiguities. Each pair of elements in the block have a *link measure* that estimates the compatibility between them (e.g. a distance ΔR or a χ^2 of a combined fit). The compatibility variable depends on the considered elements as follows:

- The track is linked to any cluster in which the extrapolated position is in the cluster boundaries.
- In order to collect the Bremsstrahlung photons emitted by electrons, a link between an ECAL cluster and a track is created if the tangent to the track, extrapolated from any of the crossing with the tracker layers, points within the cluster boundaries.
- The link is established when the position of the more granular calorimeter lays within the envelope of the coarser calorimeter.
- The link between a charged particle track and a track in the muon system is done when the global fit between the two tracks returns an acceptable χ^2. When multiple combinations are found the one with the minimum χ^2 is retained. The link measure is represented by the χ^2.

All the elements are then grouped by the linking algorithm in the smallest possible number of independent blocks. The Particle Flow algorithm then solves the ambiguities in each of the blocks by assigning the elements in each block to a unique particle from the ones described in the following sections. Elements assigned to a particle are removed from the block to reduce the combinatorics and the algorithm is iterated until no more elements are left.

3.3.3 Muons

The muon reconstruction in CMS [119] is done with very loose quality criteria before the Particle Flow event reconstruction using the tracks reconstructed independently in the silicon tracker and in the muon system. Different reconstruction approaches are used:

- **Global muon reconstruction** (outside-in): From a standalone muon reconstructed in the muon system a matching to a track in the tracker is looked for and a global muon track is built combining the hits from the tracker and muon system. The global muon track fit can improve the momentum resolution at high values ($p_T > 200$ GeV).
- **Tracker muon reconstruction** (inside-out): All tracks reconstructed in the tracker with $p_T > 0.5$ GeV and $p > 2.5$ GeV are considered as muon candidates. They are extrapolated to the muon system taking into account the expected energy loss and the uncertainty from multiple scattering. If at least one muon segment made of either DT or CSC hits is matched, the track is classified as tracker muon track. At low momentum ($p_T < 5$ GeV) this approach brings better resolution than the global muon reconstruction and is more efficient as it requires only one muon segment in the muon system.
- **Standalone muon**: A third category of muons comes from the cases where the two previous approaches fail but a standalone muon is reconstructed in the muon system. Only 1% of reconstructed muons from the collisions enter this category, but the contamination from cosmic-ray muons leads to a collision muon to cosmic-ray muon ratio that is a factor 10^4 to 10^5 less favorable than for the previous two muon categories.

The resulting collection of candidates is used as the main input for the identification of Particle Flow muons. The Particle Flow algorithm starts by muons because of the high purity in the muon reconstruction. The Particle Flow muon reconstruction has been optimized to identify muons in jets with high efficiency, keeping fake rate from misidentified charged hadrons low [120]. If a muon is not well identified it can be eventually treated as a charged hadron, and any deposit in the HCAL from an overlapping neutral hadron will be wrongly assigned to the muon track, leading to a degradation of the jet resolution an biasing the momentum to lower values.

- **Particle Flow muon**: The global muons having at least one valid hit in the muon system and for which the transverse energy sum of all the neighboring tracks and calorimetric cells around the muon inside a surrounding cone of $\Delta R < 0.3$ is less than 10% are classified as Particle Flow muons. This requirements selects a very high purity sample of real muons.

Muon quality:

Two sets of quality criteria, a loose and a tight, identify genuine muons with increasing purity. The tight identification is used to select muon candidates coming from the leptonic decay of the tau lepton:

- To be reconstructed as *Global* and Particle Flow muon.
- ≥ 1 pixel hits associated to the muon track.
- ≥ 6 tracker layers with hits.
- ≥ 1 hits in the muon system.
- ≥ 2 matched segments.
- $\chi^2/N_{DoF} < 10.0$ for the global track fit.
- To have a transverse impact parameter of the track reconstructed in the pixel plus strip detectors $d_{IP} < 2\ mm$ wit respect to the primary vertex.

Loose muons are required to be reconstructed as *Global* muons.

The muon reconstruction performance in CMS allows to reconstruct the different resonances in the di-muon invariant mass spectra in the range 1 à 150 GeV as shown in figure 3.16.

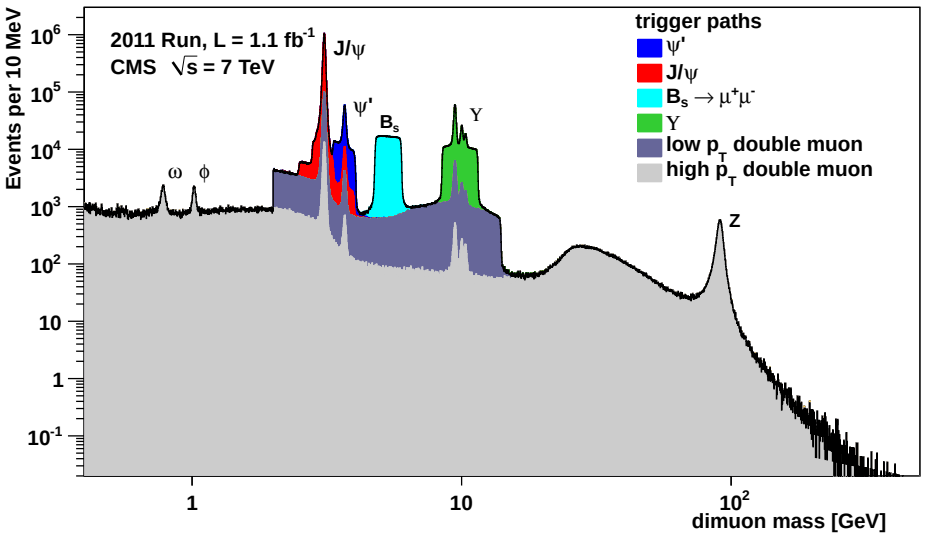


Figure 3.16: *Invariant mass spectra of opposite-sign muon pairs collected with different di-muon triggers.*

3.3.4 Electrons

Electron reconstruction is critical within the Particle Flow event reconstruction. The main challenge in the electron reconstruction comes from the fact that electrons tend to radiate bremsstrahlung photons in the tracker volume and their energy deposits can be spread over a large area. If some elements are not properly linked together, the energy measurement can be spoiled and as a consequence degrades the jet and $\not\!\!E_T$ resolution.

A specific algorithm recovers the bremsstrahlung photons among the clusters reconstructed in the ECAL. The standard electron reconstruction in CMS [121, 122] uses the reconstructed ECAL clusters to seed the electron track finding (so-called *ECAL-driven* method). It is well suited for high-p_T and isolated electrons. A complementary *tracker-driven* method takes into account the low-p_T and non isolated electrons. Additional bremsstrahlung deposits are recovered by linking ECAL clusters compatible with the track tangent from any of the crossing with the tracker layers.

Electron ECAL cluster reconstruction

The effect of photon radiation through bremsstrahlung can be large for electrons in CMS. The radiated energy before reaching the ECAL surface can be more than 80% of the initial electron energy in the regions where the material budget is the largest ($\eta \sim 1.4$). For this reason the recovery of the energy carried by the radiated photons is necessary: it is the so-called super-clustering procedure.

In the barrel region, the ECAL clusters are built within a window narrow in η and broad in ϕ around the cell with the maximal energy (seed). An iterative procedure groups dominoes made of an array of 5 crystals along ϕ and 1 crystal along η. The central crystal of a domino is at the same η than the seed in a region extended over 0.3 radians in ϕ. In the end-cap regions a more complex algorithm is used due to the ECAL geometry. Matrices of 5×5 crystals are clustered around the seed within a window of 0.3 radians in ϕ and 0.7 units in η.

Electron track reconstruction

The standard track reconstruction in CMS was presented in section 3.3.1. It is based on the Kalman filter algorithm. This algorithm is not perfect for electron track reconstruction, due to the large loss of energy in the tracker material and consequently missing hits when the change in curvature is too high, leading to a loss of efficiency. An alternative reconstruction method based on a Gaussian Sum Filter (GSF) pattern recognition algorithm is used instead [123, 124], started only on pre-selected track seeds, as it is time consuming.

The *ECAL-driven* method is initiated from the ECAL supercluster. As it contains all the energy lost by the electron, the *supercluster* barycenter corresponds to the position of an electron on the ECAL surface that would not have radiated photons (see figure 3.17). Therefore, an helix is propagated from this barycenter back to the interaction point. The two charge hypotheses are tested and their intersection with the innermost layers or disks predicts the position of the track seeds. Only superclusters passing a transverse energy cut of $E_{T,SC} > 4$ GeV and passing the hadronic veto cut $H/E_{SC} < 0.15$ are considered.

The electrons in jets may not be covered in the *ECAL-driven* track seeding due to the hadronic veto. Also low-p_T electrons may fail the reconstruction criteria due to the size of the supercluster which cannot recover all the bremsstrahlung photons in this case. In addition, electrons falling in the ECAL crack regions can induce a supplementary loss of efficiency. The *tracker-driven* (in to out) seeding developed within the Particle Flow reconstruction aims to recover these limitations [125].

The two collections of seeds (*ECAL-driven* and *tracker-driven*) are then combined and used to seed the GSF algorithm. Each GSF track is linked to the Particle Flow cluster that matches its extrapolation to the calorimeters. The GSF algorithm takes into account the energy loss at each layer and the bremsstrahlung photons are recovered by extrapolating the track tangents to the ECAL. Moreover tracks from converted bremsstrahlung photons are also recovered in this procedure by a dedicated algorithm. In figure 3.17 a cartoon imaging the electron reconstruction is displayed.

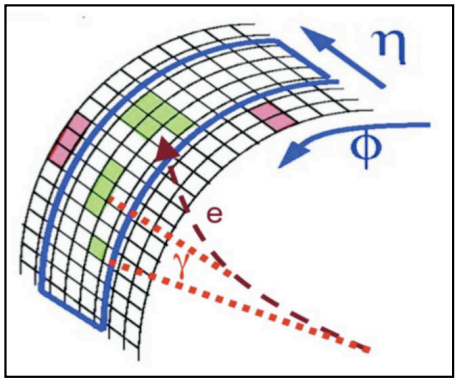


Figure 3.17: *Electron radiating bremsstrahlung photons in CMS.*

E-p combination

The electron momentum is estimated by combining the ECAL super-cluster energy and the associated track momentum. The track momentum estimation becomes more precise than the single ECAL measurement at low p_T (typically lower than 15 GeV). A regression technique is used to obtain the best estimate of the momentum by linearly combining the supercluster energy and the estimated track momentum. The resolution is improved typically by 25% for electrons with p_T around 15 GeV in the barrel region.

Electron identification

Different types of variables are used to identify electrons in CMS: variables related to the ECAL-track matching, shower shape variables and purely tracking variables. A cut-based and an MVA-based identification have been deployed in order to identify electrons.

The MVA discriminator is based on a Boosted Decision Tree (BDT) (see section 4.4) which has been trained to separate electrons from $jet \rightarrow e$ fakes. The training was performed on data by selecting candidates using a sample of $Z \rightarrow ee$. The pair of oppositely charged electrons closest to the Z mass are considered as "signal" while the other electron candidates reconstructed in the event (likely to be due to $jet \rightarrow e$ fakes) as "background". The BDT was trained using the following quantities:

ECAL-track matching variables:

- The distance in η and in ϕ between the reconstructed super cluster and the track position evaluated at the primary vertex.

- The distance in η between the reconstructed super cluster and the track position evaluated at calorimeter surface.

- The ratios: E_{SC}/p_{in}, E_{SC}/p_{out}, E_C/p_{out}, where p_{in} (p_{out}) indicates the track momentum evaluated at the primary vertex (calorimeter surface), and E_C the energy of the seed cluster.

- The variable $1/E - 1/p$ where E and p are the energy and the momentum of the reconstructed electron.

Shower shape variables:

- The cluster covariance matrix in the η and in ϕ directions , $\sigma_{\eta\eta}$ and $\sigma_{\phi\phi}$.

- The ratio $R_9 = E_{3\times3}/E_{SC}$ where $E_{3\times3}$ indicates the energy in an array of 3×3 cells on the vicinity of the super cluster seed and E_{SC} indicates the raw energy of the reconstructed super cluster.

- The variable $f_e = 1 - E_{1\times5}/E_{5\times5}$, where $E_{1\times5}$ ($E_{5\times5}$)is the energy deposition in an array of 1×5 (5×5) cells in the vicinity of the super cluster seed.

- The ratio between the hadronic energy and the electromagnetic energy (H/E).

- The ratio between E_{PS}, the energy reconstructed in the pre-shower detector, over E_{SC}.

Tracking variables:

- The momentum and η coordinate of the reconstructed electron.

- The normalized χ^2 of the common track fit, the number of valid hits in the track fit, the normalized χ^2 of the GSF track fit.

The training was performed in two bins of p_T and three bins in η. Loose and Tight working points are extracted and defined by corresponding cuts on the BDT output for each bin in p_T and η.

The cut-based identification criteria [126] based on five of the identification variables are also used in physics analysis. Rectangular cuts are applied on: $\Delta\eta$ and $\Delta\phi$ between the electron supercluster position and the direction of the track at the vertex, the H/E_{SC} ratio, $\sigma_{\eta\eta}$ and $1/E - 1/p$. The set of cuts on these variables are presented in table 3.3. The very loose "Veto" cut-based identification criteria is defined and used to veto events containing loosely identified electrons.

Variable	Veto WP	Loose WP	Medium WP	Tight WP
Barrel ($\lvert\eta_{SC}\rvert \leq 1.479$)				
$\lvert\Delta\eta\rvert <$	0.007	0.007	0.004	0.004
$\lvert\Delta\phi\rvert <$	0.8	0.15	0.06	0.03
$\sigma_{\eta\eta} <$	0.01	0.01	0.01	0.01
$H/E_{SC} <$	0.15	0.12	0.12	0.12
$1/E - 1/p <$	N/A	0.05	0.05	0.05
Endcap ($1.479 < \lvert\eta_{SC}\rvert < 2.5$)				
$\lvert\Delta\eta\rvert <$	0.001	0.009	0.007	0.005
$\lvert\Delta\phi\rvert <$	0.7	0.10	0.03	0.02
$\sigma_{\eta\eta} <$	0.03	0.03	0.03	0.03
$H/E_{SC} <$	N/A	0.10	0.10	0.10
$1/E - 1/p <$	N/A	0.05	0.05	0.05

Table 3.3: *Cut-based electron working points definition. From Ref. [126].*

Performance

The MVA-based electron identification and the cut-based criteria are compared thanks to ROC (Receiver Operating Characteristic) curves, presenting the fake-rate as a function of the efficiency, separately for electrons reconstructed in the barrel and in the end-cap regions, as presented in figure 3.18. The MVA-based criteria achieves better performance ($\sim$ 10% higher efficiency) as the discriminating variables are combined in the Boosted Decision Trees.

Non-isolated electrons identification

A specific identification for non-isolated electrons [125] discriminates electrons from charged hadrons (mostly pions) in jets. A multivariate analysis (see section 4.4) was developed to do this task. A single Boosted decision Trees discriminator was trained with the previously described identification variables. A set of isolated electrons (from $Z/\gamma^* \rightarrow ee$ events) and non-isolated electrons from jets are used as signal for the MVA training. Charged pions in jets are used as background. A single working point is chosen, leading a 65% efficiency on electrons and 1% of pion contamination in b-jets enriched events.

This discriminator has been used within the τ-lepton reconstruction as an anti-electron discriminator by reversing the cut on the BDT score (see chapter 5).

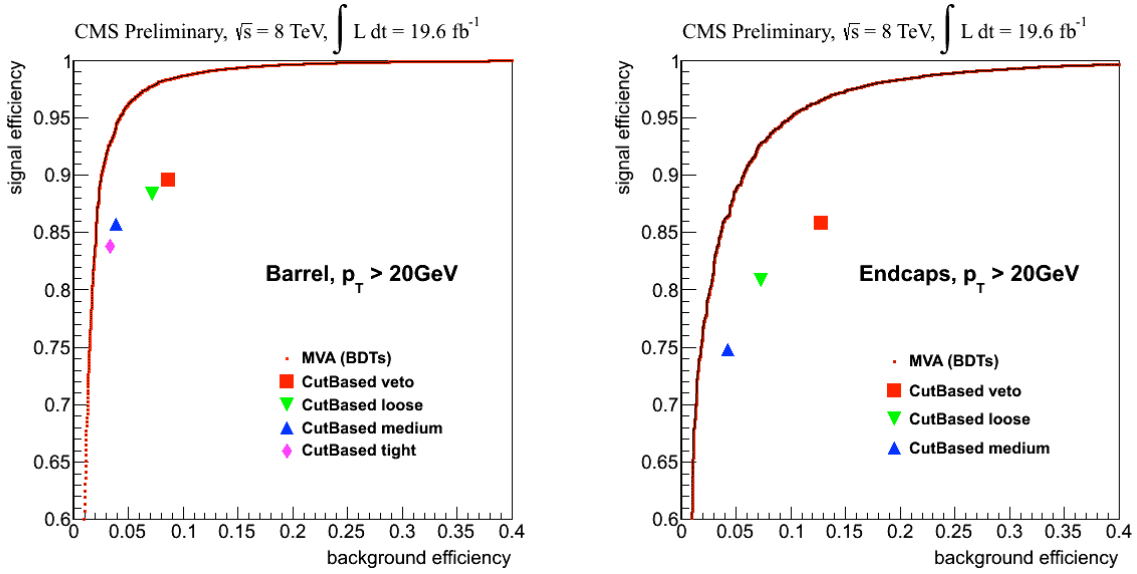


Figure 3.18: *ROC curves for the electron multivariate identification (Boosted Decision Trees) compared with the cut-based selection working points. Signal from Drell-Yan Monte Carlo simulation sample. Background from jets faking electrons in a data sample dominated by $Z + jets$. Electron candidates with $p_T > 20$ GeV are shown. Left: Barrel. Right: End-cap . From Ref. [122].*

3.3.5 Charged Hadrons, neutral hadrons and photons

The remaining reconstructed tracks are then treated by the Particle Flow algorithm and linked to the ECAL and HCAL clusters. In order to separate the neutral components (neutral hadrons and photons) and the charged particles in the block, a comparison between the momentum of the tracks and the calorimetric energy is done. The comparison is done after performing the cluster calibration detailed in [117], correcting for the non linearity of the calorimeter response and threshold effects.

Charged hadrons

In the case where the total calibrated cluster energy is smaller than the total charged particle momenta, a search for loose muons is performed. The remaining tracks are removed progressively ordered by their p_T uncertainty. Tracks with an uncertainty higher than 1 GeV are removed from the list until the total track momentum reaches the cluster energy or all this tracks have been examined. Each remaining track gives rise to a charged hadron with a momentum taken from the tracker measurement and a mass corresponding to a charged pion $\pi^{\pm}$.

Neutral hadrons and photons

Conversely, the calibrated calorimetric clusters can be higher than the total charged particle momentum. If the difference is higher than one standard deviation of the cluster energy the algorithm will create additional neutral particles. The preference in the ECAL

is given to photons. This is justified by the observation than in jets 25% of the energy is carried by photons, while neutral hadrons leave only 3% of the jet energy in the ECAL. Also, as we will see in next chapter, for τ decays neutral hadrons decay mainly to photons ($\pi^0 \to \gamma\gamma$), which reduces the neutral hadron contribution in τ jets by an order of magnitude.

If the calorimetric excess is larger than the total ECAL cluster energy, the ECAL cluster gives raise to a photon and the remaining part of the excess is associated to a neutral hadron. Otherwise, the energy excess is associated to a photon and the remaining ECAL energy is interpreted as an early shower of the charged hadron. The remaining ECAL and HCAL clusters not linked to any track give rise to photons or neutral hadrons respectively. The neutral hadron energy is determined by the calibrated HCAL cluster.

When photons are isolated, a simple algorithm [127–129] based on electromagnetic deposits in the ECAL is used to define Particle Flow photons.

3.3.6 Jets

In CMS the jets are reconstructed within the geometric acceptance $|\eta| < 4.7$ by using the anti-k_T clustering algorithm [130] with a distance parameter $R = 0.5$. The Particle Flow jets are reconstructed using the total set of particles reconstructed by the Particle Flow algorithm [131,132]. The jet momentum and spacial resolution are improved with respect to the so-called *calo-jets* reconstructed as a cluster of ECAL and HCAL cells arranged into projective towers. This is due to the use of the combination of the tracker and the granularity of the ECAL to distinguish and precisely measure the energy of charged hadrons and photons, both carrying 90% of the jet energy. The jet energy response is measured in $Z/\gamma^* \to \mu\mu$ events using the Tag-and-probe technique where the jet recoiling against the Z boson is taken as a probe. Simulation to data correction factors are extracted. In figure 3.19 the jet transverse momentum resolution is presented. Data is well reproduced by the simulation and the resolution is improved specially at low transverse momenta in the Particle Flow jets case.

In collisions with a large number of pile-up interactions, fake jets can be reconstructed from the accidental clustering of many neighboring particles, or from the superimposition of soft jets from different vertices. In order to distinguish the jets coming from the production vertex from the ones originating from soft interactions a multivariate Boosted Decision Tree is used [133]. The input variables for this pile-up jets discriminator are the compatibility of the tracks in the jet with the primary vertex, jet shape variables and the multiplicity of the charged and neutral components within the jets. The training is performed using jets with $p_T > 20$ GeV in different η regions. The efficiency is evaluated on $Z/\gamma^* \to \mu\mu$ events. The results as a function of the number of reconstructed vertices and the jet transverse momenta are presented in figure 3.20.

The efficiency and data to simulation scale factors shown are not dependent on the pile-up and are robust in all the transverse momenta range.

A dedicated algorithm is used to separate jets coming from the fragmentation of b/c quarks. The analysis of b-jets properties as the displaced vertex and large impact parameter due to the long lifetime of the b hadrons and the production of leptons can be

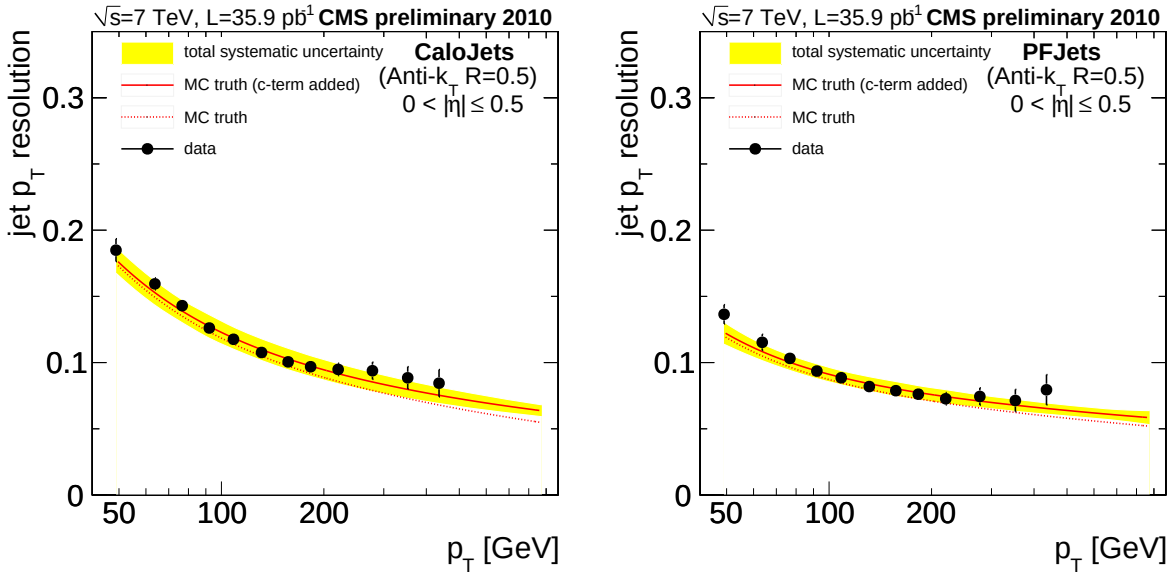


Figure 3.19: *Jet transverse momentum resolution measured with the data recorded in 2010 at $\sqrt{s} = 7$ TeV (dots) and in Monte Carlo simulation (red line) as function of the jet transverse momenta for calo-jets (left) and Particle Flow jets (right). From Ref. [132].*

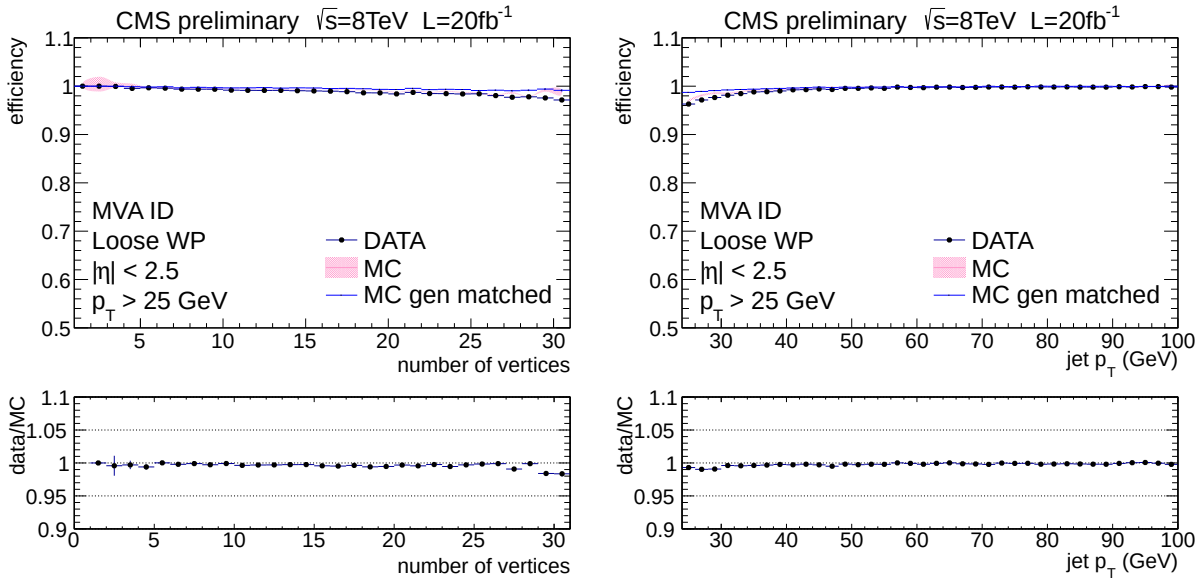


Figure 3.20: *Data to simulation comparison of the MVA (loose working point) pileup jet identification efficiency versus the number of primary vertices (left) and the jet transverse momentum (right) on the $Z/\gamma^* \to \mu\mu$+jets sample for PF jets with $p_T > 25$ GeV. From Ref. [133].*

useful to create performant b-tagging algorithms [134]. The so-called Combined Secondary Vertex (CSV) algorithm, based on a likelihood discriminator is used [135] to distinguish jets from b-quarks and those from charm or light quarks and gluons.

3.3.7 Missing transverse energy

The CMS detector is designed to reconstruct almost all the stable particles produced in the *pp* collisions delivered by the LHC, notably using the *Particle Flow* algorithm described above. The particles that cannot be reconstructed are typically the neutrinos produced in the decay chains, the particles falling out of the detector acceptance and hypothetical neutral weakly interacting particles not discovered yet. The longitudinal component of the momentum of the interacting partons during the collision stays an unknown quantity, as being a fraction of the proton momentum. While the transverse momentum before and after the collision is conserved, the total sum of all the momenta of the produced particles is null.

The missing transverse momentum $\overrightarrow{\not{E}_T}$ is defined as the imbalance in the transverse momentum of all the visible particles in the final state of the *pp* collisions. Its magnitude is denoted $\not{E}_T$. It represents the momentum carried by the neutral weakly interacting particles as the neutrinos.

CMS has developed several algorithms to reconstruct $\overrightarrow{\not{E}_T}$ [136]: tracker based, calorimetric, depending on the objects used to compute it (particle candidates, calorimeter clusters, etc). Typically, two types of $\overrightarrow{\not{E}_T}$ reconstruction can be pointed out: PF $\overrightarrow{\not{E}_T}$ and Calo $\overrightarrow{\not{E}_T}$ [137]. The PF $\overrightarrow{\not{E}_T}$ is the vectorial sum of all the visible particle flow particles momenta. While the Calo $\not{E}_T$ is computed using the energies collected by the calorimeter towers and its direction is taken from their directions with respect to the center of CMS.

Type-0 and type-1 corrections

The missing transverse energy computation depends strongly on the other particle reconstruction, as well as on detector calibration and malfunctions. The $\overrightarrow{\not{E}_T}$ can be mismeasured because of calorimeter thresholds and non linearity, tracker inefficiency and the presence of neutrinos in leptonic decays. To cope with this, the so-called "type-1" corrections are applied. The corrected $\overrightarrow{\not{E}_T}$ is then:

$$\overrightarrow{\not{E}_T}^{corr} = \overrightarrow{\not{E}_T} - \sum_{jets} (\overrightarrow{p}^{corr}_{T,jet} - \overrightarrow{p}_{T,jet}) \tag{3.9}$$

where $\overrightarrow{p}^{corr}_{T,jet}$ is an evaluation of the $\overrightarrow{p}_T$ that tries to correct for all the previous mentioned experimental defects as well as corrections on the energy scale.

A second type of corrections, introduced for the 2012 data takes into account the calorimeter nonlinearity and the minimum thresholds used in the *Particle Flow* reconstruction. The $\overrightarrow{\not{E}_T}$ direction is driven by the direction of the sum of the vectorial momenta of the neutral particles. The "type-0" correction is applied to correct for the direction of $\overrightarrow{\not{E}_T}$, taking the direction of the charged particles associated to pile-up vertices to build an estimator of the induced $\overrightarrow{\not{E}_T}$. This type of correction is then applied only to the PF $\not{E}_T$.

Missing transverse momentum scale and resolution

The performance of the $\overrightarrow{\not{E}_T}$ is determined using $Z \to \mu\mu$, $Z \to ee$ or events triggered by photons. In these events the resolution of the $\overrightarrow{\not{E}_T}$ is dominated by the hadronic activity as the photon and lepton resolutions are very good ($\sigma_{p_T}/p_T \sim 1-6\%$ for muons and $\sigma_E/E \sim 1-6\%$ for electrons and photons). In these events there is no real missing transverse energy, but it is induced by removing the reconstructed Z or photon from the event.

Then the scale and resolution of the $\overrightarrow{\not{E}_T}$ are measured by comparing the response of the hadronic recoil system $\overrightarrow{u_T}$ to the transverse momentum of the vector boson $\overrightarrow{q_T}$. We have then the momentum conservation equation:

$$\overrightarrow{q_T} + \overrightarrow{u_T} + \overrightarrow{\not{E}_T} = 0 \tag{3.10}$$

In figure 3.21 the different vectors are illustrated.

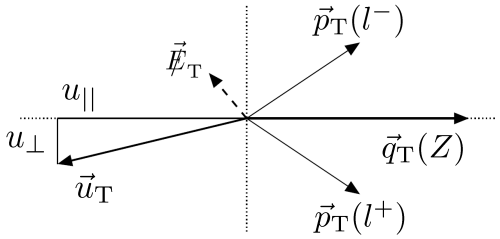


Figure 3.21: *$Z \to ll$ kinematic system in the transverse plane. The vector $\vec{q_T}$ is the Z transverse momentum, $\vec{u_T}$ the vectorial sum of all particles except for the two leptons from the Z decay and $\vec{\not{E}_T}$ is the transverse missing momentum.*

The hadronic recoil is then decomposed to the parallel $u_\parallel$ and perpendicular $u_\perp$ components with respect to the axis defined by $\overrightarrow{q_T}$. The response is thus defined as $-u_\parallel/q_T$ (because $\vec{\not{E}_T}$ is supposed to be null in these events) and the resolutions following the two projections $\sigma(u_\parallel)$ and $\sigma(u_\perp)$ are taken from the full width at half maximum of the Voigtian (convolution of a Breit-Wigner and a Gaussian distributions) $u_\parallel + q_T$ and $u_\perp$ distributions.

Figure 3.22 shows the response curves for the PF $\overrightarrow{\not{E}_T}$ as a function of the vector boson transverse momentum. The response converges to unity for boosts higher than ~ 40 GeV for Z events and the data is well modeled by the simulation. The photon events have slower response due to the hadronic activity in the recoil because of the presence of a sizable background from QCD multi-jets events.

In figure 3.23 the comparison between the Calo $\overrightarrow{\not{E}_T}$ and PF $\overrightarrow{\not{E}_T}$ resolution of the perpendicular and parallel components is showed as a function of the reconstructed vertices. The resolution improves when using the *Particle Flow* particles. Anyhow, the dependence on the pile-up is high which justifies the need of pile-up mitigation algorithms.

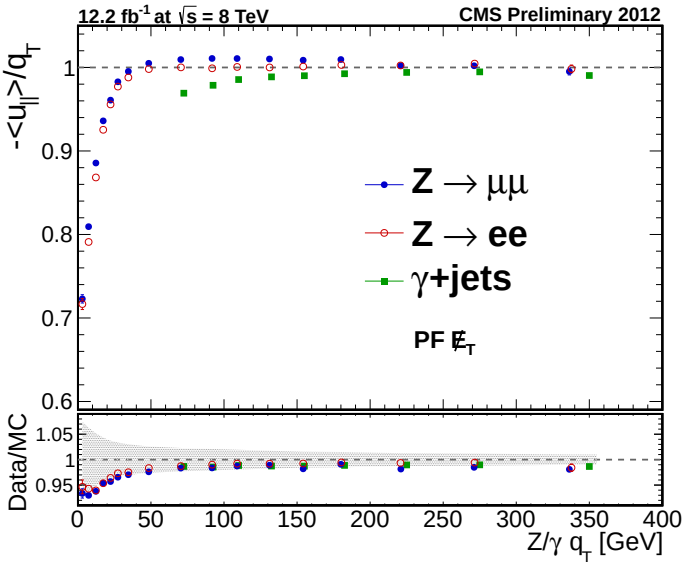


Figure 3.22: *Response curves for PF $\vec{\not{E}}_T$ for events with a Z or γ boson. From [137].*

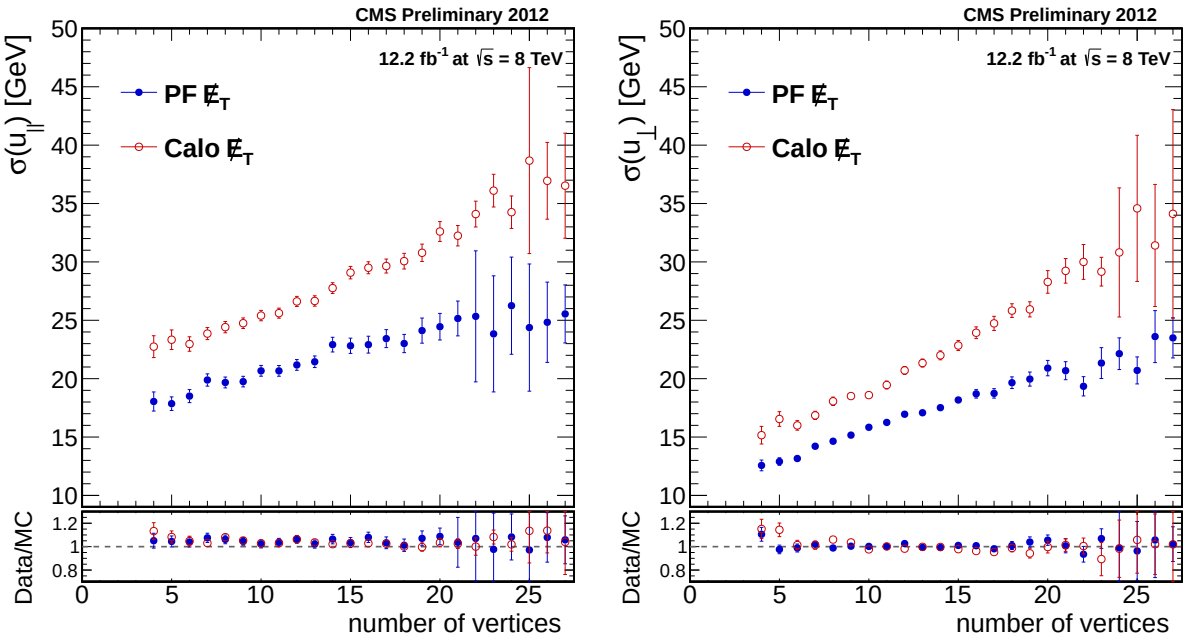


Figure 3.23: *Parallel (left) and perpendicular (right) recoil component resolution for the PF $\vec{\not{E}}_T$ and Calo $\vec{\not{E}}_T$ algorithms as a function of the number of reconstructed vertices for $Z \to \mu\mu$ events.From [137].*

Pile-up effect reduction

The presence of pile-up interactions has small effect on the $\overrightarrow{\not{E}}_T$ response but strongly degrades the resolution inducing a smearing of few GeV in the parallel and perpendicular components. Two algorithms have been developed to mitigate this effect on the PF $\overrightarrow{\not{E}}_T$.

The strategy is to separate the PF $\overrightarrow{\not{E}_T}$ in different contributions: particles originating from the production vertex (PV) and particles originating from pile-up interactions (PU). The association to the PV is either geometrical for the charged particles, or using the MVA pile-up jet identification output described in section 3.3.6.

The first algorithm so-called No-PU PF $\overrightarrow{\not{E}_T}$ is described in detail in [137]. It applies a reduction factor on the missing transverse momentum sum for the particles originating from the soft scatter interactions:

$$S_F = \frac{\sum_{PV,charged} p_T}{\sum_{PV,charged} p_T + \sum_{PU,charged} p_T} \tag{3.11}$$

The second algorithm denoted MVA PF $\not{E}_T$ is based on a set of multivariate regressions computed as corrections to the hadronic recoil $\overrightarrow{u_T}$ of the PF $\overrightarrow{\not{E}_T}$. First a correction on the recoil direction computed by a boosted decision tree is done in order to match the azimuthal angle of $-\overrightarrow{q_T}$. It is trained using simulated $Z \to \mu\mu$ events. Then a second multivariate analysis is used to predict the true magnitude of the hadronic recoil. It is trained on simulated $Z \to \mu\mu$ events after the first correction. It uses as input for the regression a set of five $\overrightarrow{\not{E}_T}$ variables computed from the *Particle Flow* particles:

1. the negative vectorial sum of all the *Particle Flow* particles: PF $\overrightarrow{\not{E}_T}$.
2. the negative vectorial sum of all the charged *Particle Flow* particles associated to the selected hard-scatter vertex.
3. the negative vectorial sum of all the charged *Particle Flow* particles associated to the selected hard-scatter vertex and all the neutral *Particle Flow* particles within jets that have passed the MVA pile-up jet ID.
4. the negative vectorial sum of all the charged *Particle Flow* particles not associated to the selected hard-scatter vertex and all the neutral *Particle Flow* particles within jets that have failed the MVA pile-up jet ID.
5. the negative vectorial sum of all the charged *Particle Flow* particles not associated to the selected hard-scatter vertex and all the neutral *Particle Flow* particles plus the positive vectorial sum of all the neutral *Particle Flow* particles within jets that have failed the MVA pile-up jet ID.

The boosted decision tree is trained having as input variables the different angles and magnitude of the hadronic recoils $\overrightarrow{u_T}$, the scalar sum of the transverse momenta of all the *Particle Flow* particles for each $\overrightarrow{\not{E}_T}$, the momentum vectors of the two highest p_T jets in the event and the number of primary vertex. It takes advantage of each missing transverse momentum estimation: (2) is almost pile-up insensitive, (3) is not much pile-up dependent and gives a better estimation of the energy scale, (4) estimates the pile-up contribution and (5) takes into account unclustered particles.

The MVA based regression minimizes the pile-up effects and improves the resolution by 40% and is almost pile-up independent. In figure 3.24 the resolution for the different

$\overrightarrow{\not{E}_T}$ described above are displayed as a function of the number of reconstructed primary vertex.

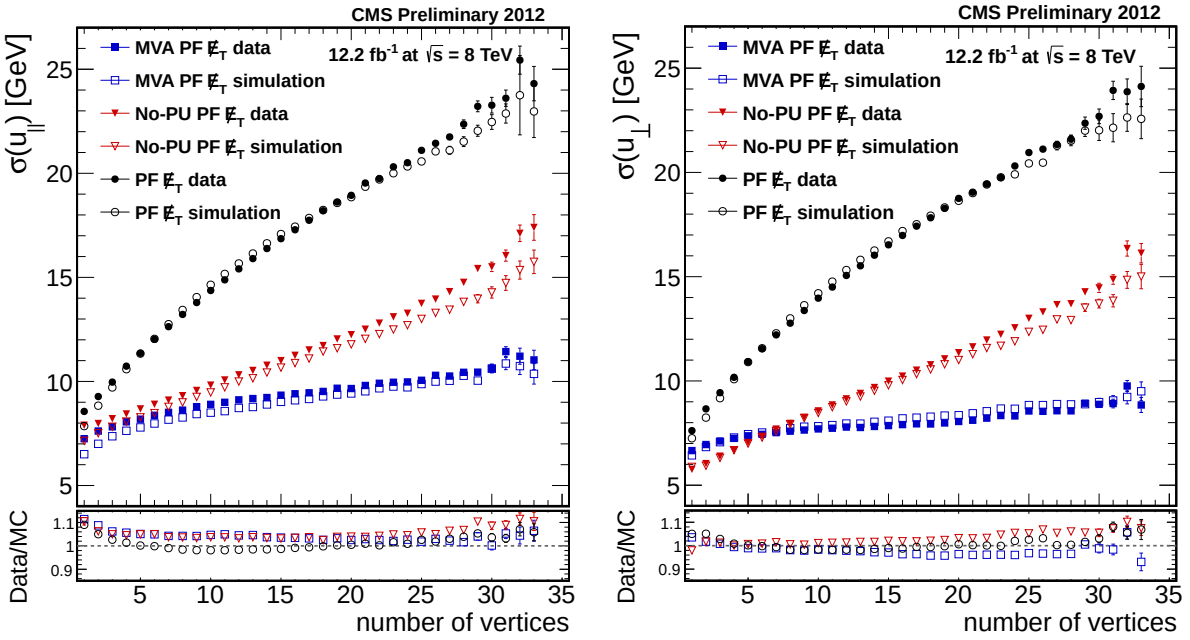


Figure 3.24: *Parallel (left) and perpendicular (right) recoil component resolution as a function of the number of reconstructed vertices for the PF $\vec{\not{E}}_T$, No pile-up $\vec{\not{E}}_T$ and MVA $\vec{\not{E}}_T$ using $Z \to \mu\mu$ events. From [137].*

In figure 3.25 the MVA PF $\not{E}_T$ distributions in events with $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to ee$ are displayed. Data is very well reproduced by simulation.

The MVA PF $\not{E}_T$ is very important in the analysis presented in chapters 6 and 7 as it is an input driving the resolution of the di-τ mass reconstruction and the transverse mass variables.

3.3.8 Taus

τ leptons can decay either to leptons, muons or electrons, or hadronically. Electrons and muons from a τ decay are not distinguishable from electrons or muons originating in the primary proton-proton interaction. The τ reconstruction algorithms refer to hadronically decaying τ leptons that will be denoted in the following by τ_h. The next part of this thesis is dedicated specifically to the τ_h reconstruction in CMS in which I have contributed during my thesis.

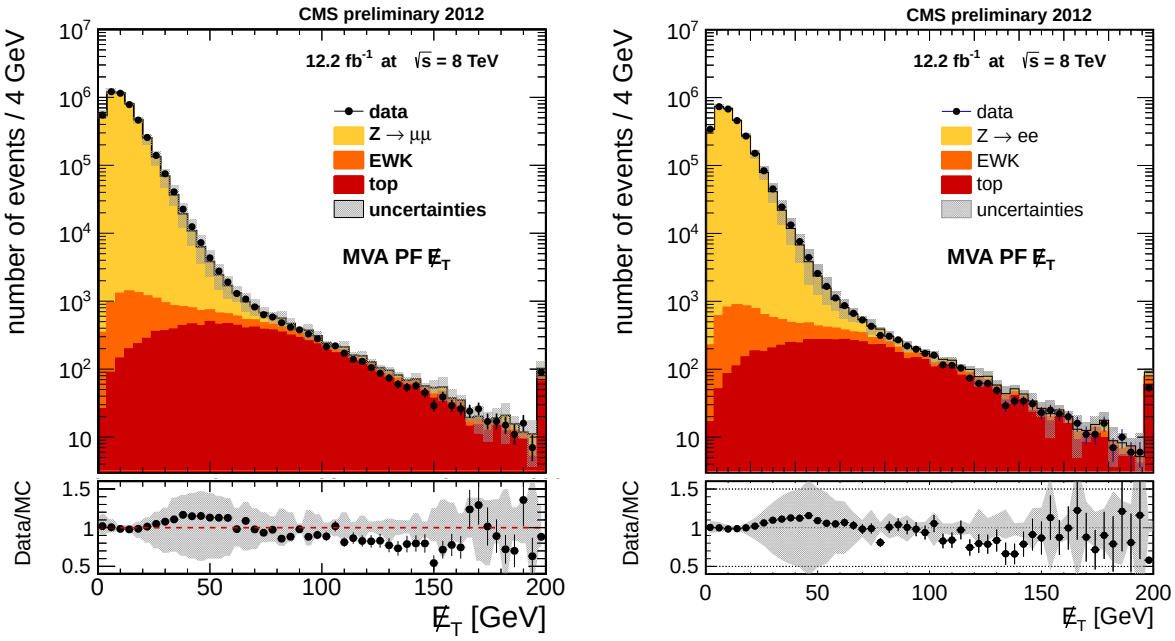


Figure 3.25: *MVA PF $\not{E}_T$ distributions using $Z \to \mu\mu$ (left) and $Z \to ee$ (right) events. The bottom part represents the ratio between data and simulation, including the statistical uncertainties for data and the systematic uncertainty on the simulation (grey error band). From [137].*

Part II

τ-lepton reconstruction in CMS

Chapter 4

τ-lepton trigger and reconstruction

Chapter content

In this chapter I present the τ-lepton reconstruction in CMS and report on its performance [138]. First I introduce the τ-lepton physics and motivations for physics analyses involving them. Then I concentrate on the trigger and reconstruction of hadronically decaying taus (τ_h). I detail the so-called HPS (Hadron Plus Strips) algorithm [139] and show its performance. A specific section is dedicated to the τ_h isolation, which constitutes the main handle to reject jets mis-identified as τ_h. I concentrate on the MVA based isolation which I have validated in the context of the Higgs boson searches decaying to two τ-leptons (see chapters 6 and 7). Finally, I present the discrimination of hadronically decaying taus from muons. A specific chapter of this thesis (chapter 5) reports my work on the rejection against electrons. The τ-lepton reconstruction and performance are detailed in Ref. [140], where my contributions are documented.

4.1 τ-lepton physics

The τ is the heaviest lepton and has a mass of 1776.82 ± 0.16 MeV, a charge of -1 and a lifetime times $c\tau \sim 87\mu m$ [141]. As it is shown in figure 4.1, it decays into a τ-neutrino and a virtual W boson that subsequently decays into leptons (electron or muon plus neutrino) or into up and down type quarks. The high mass of the τ makes it the only lepton that can decay into hadrons.

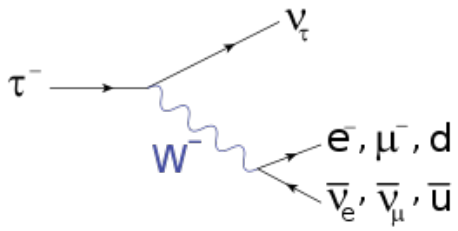


Figure 4.1: *τ decay Feynman diagram.*

Table 4.1 shows the different hadronic τ decays, their corresponding branching ratios and the different intermediate resonances. A τ decays into hadrons in about 2/3 of the cases and into lighter leptons (electron or muon) in about 1/3 of the cases. In the CMS experiment neutrinos are detected indirectly through the measurement of the transverse missing energy. Electrons or muons from a τ decay are hardly distinguishable from other electrons or muons from the pp collision. Therefore τ reconstruction refers specifically to the reconstruction of **hadronically decaying τ only (τ_h)**.

The hadronic decays of the τ-lepton leads to a final state containing either one or three charged mesons, typically $\pi^{\pm}$ and a presence of none, one or two neutral pions (π^0). The neutral pions decay instantaneously into photons via the decay $\pi^0 \to \gamma\gamma$.

The τ-lepton is used in many analyses in the CMS physics program. It constitutes an essential signature in the analysis presented in this thesis: the search for the Standard Model Higgs boson decaying into tau pairs and the search for the Minimal Supersymmetric Higgs bosons in the same final state. Improvements in the τ_h reconstruction have

Decay Mode	Resonance	Mass (MeV)	Branching Ratio (%)
$\tau^- \to \mu^- \bar{\nu_\mu} \nu_\tau$	-	-	17.36 %
$\tau^- \to e^- \bar{\nu_e} \nu_\tau$	-	-	17.85 %
$\tau^- \to h^- \nu_\tau$	π	139.6	11.6 %
$\tau^- \to h^- \pi^0 \nu_\tau$	ρ	770	26.0 %
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	a_1	1200	9.5 %
$\tau^- \to h^- h^+ h^- \nu_\tau$	a_1	1200	9.8 %
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$			4.8 %

Table 4.1: *Branching ratio of the main τ decay modes. The symbol h^- refers to a charged meson. The decays $\tau^- \to h^- \nu_\tau$, $\tau^- \to h^- \pi^0 \nu_\tau$, $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ and $\tau^- \to h^- h^+ h^- \nu_\tau$ pass trough an intermediate meson resonance which mass is displayed.*

a major impact in the sensitivity of these analysis (see section 7.5.1). It also plays a major role in the searches for charged Higgs bosons [142], for Supersymmetry [143–145], Z' [146], W' [147] and doubly charged Higgs bosons [148]. Many of the Standard Model measurements have a τ-lepton in the final state; then the τ_h reconstruction is a main ingredient in the analysis of Drell-Yann production [149], W boson measurements [150], and top quark pair production analysis [151].

The analysis of the τ_h decay products can be used to extract the τ polarization, for example by analyzing the angles between the decay products and the original τ. This represents a unique property of the hadronic decays (leptonically decaying taus loose sensitivity because of the presence of an additional neutrino in the decay). The analysis of the distributions of variables in particles decaying to τ_h pairs can reveal spin correlations allowing to access the original resonance quantum numbers (J, C and P) [152].

4.2 τ_h trigger

The τ trigger is important for a wide variety of physics analyses which contain final states with τ-leptons. The tau triggers collect events containing taus decaying hadronically. Leptonic tau decays are collected by electron and muon triggers. In this section I will describe briefly the τ_h trigger at the Level-1 [153] and High Level Trigger [154].

4.2.1 Level-1 Jets and Level-1 Taus

The τ_h trigger algorithms, both at Level-1 or at High Level Trigger, are seeded by candidates reconstructed by the Level-1 Jet algorithm. In this section the Level 1 Jet and Level 1 Tau algorithms are presented.

L1 Jets

In order to build the Level-1 Jets the information from the Regional Calorimeter Trigger (RCT) is used. The transverse energy computed in the trigger regions from the calorimeters (hadronic and electromagnetic) is considered. Each region is made of 4×4 trigger towers (see section 3.2.8). The Level-1 Jet trigger uses a sliding window of 3×3 calorimeter region (144 trigger towers) able to cover the full (η, ϕ) CMS calorimeter geometry. In figure 4.2 the geometry of the sliding window and the RCT are shown.

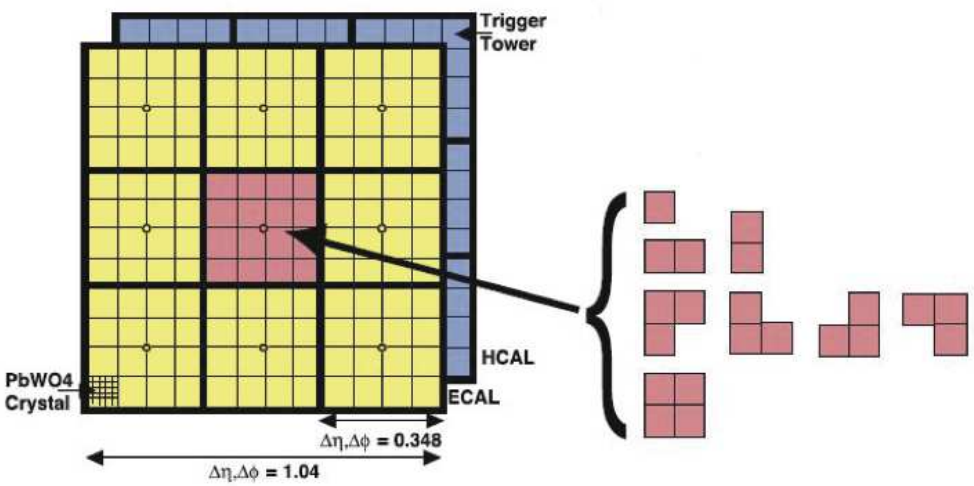


Figure 4.2: *Level-1 Jet finding algorithm illustration with a sketch of the τ veto bits (see text). From [153].*

The central region in the window is required to have a higher E_T than the eight neighboring regions and to pass a minimum threshold in order to suppress noise. The transverse energy of the Level-1 Jets is computed by summing the deposit in the 12×12 trigger towers in the barrel and end-cap regions, or the 3×3 larger towers in the HF calorimeter. The jets having $|\eta| > 3.0$ ($|\eta| < 3.0$) are classified as forward (central) jets. A Level-1 jet is classified as τ-jet if any of the nine RCT in the window is compatible with the τ-veto bits (see next section). The four central jets, forward jets and central taus with the highest energy are selected and sent to the Global Trigger. After jets are found, Look-Up Tables are used to apply a η-dependent jet energy scale correction.

L1 Taus

The common approach to separate τ_h and quark/gluon jets is to use the isolation of the τ_h candidate. Given the coarse granularity of the Level-1 system, this is a very challenging task.

The Level-1 tau algorithm is seeded by the reconstructed Level-1 jets described previously. The Level-1 tau algorithm requires additional isolation criteria plus a tau veto bit. Seven out of the eight non-central trigger regions have to contain energy deposits such that $E_T < 2$ GeV. In addition, for each trigger region a τ-veto bit is set if the energy

deposit is spread over more than 2×2 trigger towers as shown in figure 4.3. The deposited energy should then be contained in a limited region, as expected for τ narrow jets.

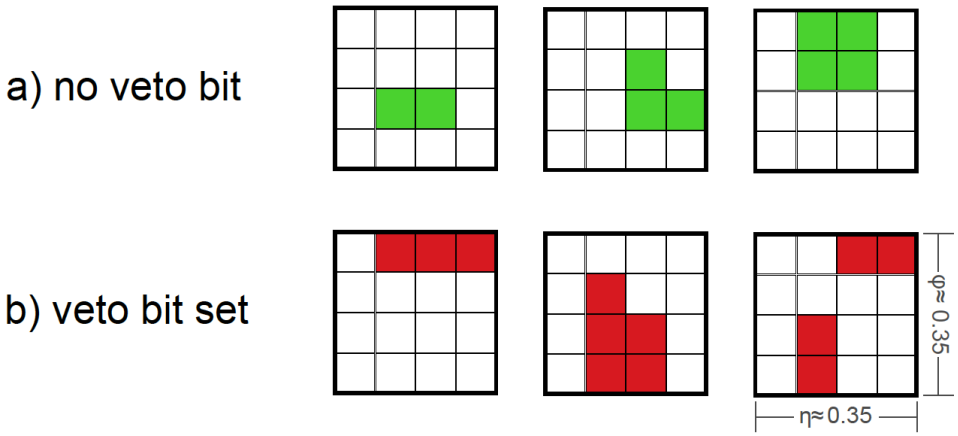


Figure 4.3: *Examples of trigger regions, where trigger towers with energy deposits ($E_T^{ECAL} > 4$ GeV or $E_T^{HCAL} > 4$ GeV) are shown as shaded squares. The Level-1 τ-veto bit is not set, if the energy is contained in a square of 2×2 trigger towers (a). Otherwise, the τ-veto bit is set (b). From [153].*

The Level-1 taus are required to have no τ-veto bit set in all the nine trigger regions. If any of the isolation or the τ-veto bit requirements fails, the object is considered as a simple Level-1 jet.

4.2.2 High Level Trigger Taus

The High Level Trigger identification of the τ_h uses much more sophisticated algorithms, similar to the offline τ_h reconstruction described later in section 4.3. The HLT τ_h algorithm is divided in three sequences, each one aiming to reduce the rate before running the next step.

The Level-2 τ trigger reconstruction is entirely based on calorimeter information. CaloJets are built with a cone of radius equal to $\Delta R = 0.2$ within $|\eta| < 2.1$ and a p_T cut on the jet transverse energy is applied.

The Level-2.5 algorithm applies a pixel track based isolation on the Level-2 τ candidates passing the transverse momenta requirement. The pixel tracks are reconstructed and the ones coming from the primary vertex and nearby the Level-2 candidates are selected. The candidate is considered isolated if there no pixel track from the same vertex with transverse momentum greater than 1.2 GeV is found in an isolation cone between $0.2 < \Delta R < 0.4$ around the tau candidate.

The Level-3 τ reconstruction is based on the Particle Flow algorithm. The algorithm is seeded by Particle Flow jets. The leading track is defined as the highest momenta track in a cone with $\Delta R < 0.2$ around the jet axis and compatible with the jet production vertex. The signal cone, defined around the leading track by $\Delta R < 0.18$ is used to

compute the τ_h momenta by summing the momenta of the charged particles and photons inside. An isolation cone between $0.18 < \Delta R < 0.45$ around the leading track is defined. The algorithm uses tracker-only isolation using tracks from a vertex compatible with the production vertex of the τ_h in order to minimize the pile-up dependence. The isolation is defined by an occupancy veto: no tracks above a given threshold. Two isolation working points are defined, the Loose (Medium) working point requires that no tracks with transverse momentum greater than 1.5 GeV (1.0 GeV) are found in the isolation cone.

HLT $\tau_h + X$ cross triggers

At HLT, the τ_h trigger can also be based on a cross-trigger between a light lepton and a τ_h, or a cross-trigger between missing transverse energy and a τ_h. This cross-triggers are seeded respectively by a Level-1 lepton or missing transverse energy. These triggers have a HLT electron, muon or missing energy selections to reduce the rate before running the Level-3 τ_h algorithm with the Loose isolation working point. At the end the HLT is called a cross-trigger fired by a τ_h and an additional object.

HLT double τ_h triggers

The double τ_h trigger is seeded by a logical or between Level-1 Jets and Level-1 taus in order to maintain a high efficiency with manageable trigger rates. The leading track of the τ_h candidates is required to pass a certain threshold (> 1 or > 5 GeV) and to be restricted to the central region of CMS ($|\eta| < 2.1$). The τ_h candidates are also required to pass the Medium isolation at Level-3.

HLT paths for $H \to \tau\tau$ analysis

In table 4.2 the Level-1 and HLT trigger paths used in the Higgs searches decaying to tau pairs containing τ_h objects in the final state are presented. The corresponding efficiencies are typically as high as 90% (*c.f.* section 6.2.1).

For the semi-leptonic channels in the $H \to \tau\tau$ analysis, a lepton plus τ_h cross trigger is used. It requires a lepton (sometimes restricted geometrically) with a reconstructed HLT τ_h passing a loose isolation criteria. In the double hadronic final state, HLT double τ_h triggers are used.

4.3 τ_h reconstruction

The τ_h lepton reconstruction aims at analyzing the τ_h footprint in a narrow jet in order to best identify the real τ-leptons from the fakes coming from jets, electrons and in a lower scale, muons.

The decay products of the τ_h are produced with an energy that usually exceeds the mass of the τ, m_τ, leading a signature in the detector made of an isolated, very collimated jet with low multiplicity compared to the quark/gluon initiated jets. The main challenge in the τ_h identification is to separate the real τ_h jets from the QCD multi-jet production which has a rate orders of magnitude higher. A granular detector able to

L1	HLT
$\tau_h\tau_h$ channel	
DoubleTauJet44er OR DoubleJetC64	DoubleMediumIsoPFTau25_Trk5_eta2p1_Jet30
DoubleTauJet44er OR DoubleJetC64	DoubleMediumIsoPFTau30_Trk5_eta2p1_Jet30
DoubleTauJet44er OR DoubleJetC64	DoubleMediumIsoPFTau30_Trk1_eta2p1_Jet30
DoubleTauJet44er OR DoubleJetC64	DoubleMediumIsoPFTau35_Trk5_eta2p1
DoubleTauJet44er OR DoubleJetC64	DoubleMediumIsoPFTau35_Trk1_eta2p1
$e\tau_h$ channel	
IsoEG18er or EG20	Ele20_Tight_LooseIsoPFTau20
IsoEG18er or EG20er or EG22	Ele22_eta2p1_WP90Rho_LooseIsoPFTau20
$\mu\tau_h$ channel	
Mu16er	IsoMu18_eta2p1_LooseIsoPFTau20
Mu14er	IsoMu17_eta2p1_LooseIsoPFTau20

Table 4.2: *Trigger paths containing HLT τ_h used in the SM and MSSM analysis described in this thesis using the data collected in 2012.*

separate the charged and neutral particles inside a jet is then a key handle to reconstruct hadronic taus. In CMS, the τ_h reconstruction is based on the Particle Flow algorithm (see section 3.3.2) thanks to the good separation achieved on the jet constituents. Also, lighter leptons as muons or electrons can be misidentified as some of the decays of τ_h. Dedicated discriminators rejecting these fakes are then developed, and they are included in the τ_h reconstruction protocol. The presence of neutrinos in the final state represents an additional challenge in reconstructing tau leptons. The p_T of the visible particles from the τ_h decay is lower than the original τ-leptons.

4.3.1 Hadron Plus Strips algorithm

The τ_h reconstruction makes use of the Particle Flow particles: individual charged hadrons, neutral hadrons, electrons, muons and photons (*c.f.* section 3.3.2). The Particle Flow objects are clustered in jets using the anti-k_T algorithm [130] with radius $R = 0.5$ that will then seed the τ_h reconstruction algorithm. The Hadron Plus Strips (HPS) reconstruction algorithm is done in two steps:

- **Reconstruction:** The Particle Flow particles are recombined to match one of the decay channels in the hadronic modes from table 4.1. The most compatible of the τ_h candidates is retained and its four momentum is computed.

- **Identification:** The τ_h candidate isolation is computed, this will serve as handle to reduce significantly the $jet \rightarrow \tau_h$ fake rate. Then the specific discriminators

against muons and electrons are built, in order to reduce the rate of $\mu \to \tau_h$ and $e \to \tau_h$ fakes. At the analysis level, only τ_h passing the isolation requirement are used. Hence by definition, an identified τ_h is necessarily isolated.

In this section I present the reconstruction stage and its performance. The identification discriminators will be detailed in the next sections.

Most of the hadronic decay modes of the tau present final states with at least one π^0. In order to reconstruct the neutral pions from the τ_h decay, the HPS algorithm is designed to extract the footprint of converted photons from the dominating $\pi^0 \to \gamma\gamma$ decays which convert in the tracker material (up to 1.5 X_0). This is done by clustering photons and electrons from the seed jet in $\eta \times \phi$ "strips" broad in the azimuthal direction to take into account the bending induced by the 3.8 Tesla magnetic field on the e^+e^- electron pairs from the photon conversions.

The strips are built starting with the most energetic electromagnetic particle in the seed jet (pivot). A rectangular window of ($\Delta\eta \times \Delta\phi = 0.05 \times 0.20$) centered around it is explored and the second most energetic particle is clustered. The sum of the two particles four-momenta is computed and defines the four momentum of the new pivot. In this clustering only photons with $P_T > 0.5$ GeV are considered. The procedure is repeated until no new particle is clustered. To be retained as π^0 candidates the strips should contain one or more photons and the sum of the photon transverse momentum have to be such that $p_T > 2.5$ GeV.

The strips are then combined with the charged particles in the seed jet, candidates of the charged hadrons in the τ_h decays. Hence the name of the algorithm. Quality cuts on the tracks associated to the charged particles are required and are listed in table 4.3[1]. No requirement on the type of the charged particles is required.

Variable	Cut
p_T	> 0.5 GeV
χ^2	< 100
d_0	$< 0.03\ cm$
d_Z	$< 0.4\ cm$
N_{hit}	≥ 3

Table 4.3: *Track selection criteria applied to charged constituents when building the τ_h candidates. Quality cuts are applied on the track transverse momentum p_T, its χ^2, its transverse impact parameter d_0 and its longitudinal impact parameter d_Z with respect to the primary vertex and the total number of hits in the pixel and silicon strip tracking detectors N_{hits}.*

All the reconstructed strips and charged particles (muons, electrons and charged hadrons not contained into strips) in the jet are combined into τ_h candidates following the hadronic decay modes from table 4.1. The combinations consist of either one or

[1]The impact parameter of a track is defined as the transverse distance of closest approach (DCA) of the track to the primary vertex point.

three charged particles plus zero, one or two strips. Multiple hypotheses are built at the same time[2] and they are tested for the considered decay modes:

1. **Single Hadron (One-prong)**: One charged particle with no strips.

2. **Single Hadron plus one Strip (One-prong plus one strip)**: One charged particle combined with one strip, having a mass such that $0.3 < M < 1.3 \cdot \sqrt{p_T(GeV)/100\ \text{GeV}}$. The size of the mass window is enlarged for high p_T candidates to take into account resolution effects. The mass window upper limit is set to 1.3 GeV (4.2 GeV) for τ_h candidate having $p_T < 100$ GeV (> 1044 GeV). This ensures the compatibility of the candidate to the $\rho(770)$ resonance.

3. **Single Hadron plus two Strips (One-prong plus two strips)**: One charged particle combined with two strips. The combined system should fit a mass window of $0.4 < M < 1.2 \cdot \sqrt{p_T(GeV)/100\ \text{GeV}}$. The mass window upper limit is set to 1.2 GeV (4.0 GeV) for τ_h candidate having $p_T < 100$ GeV (> 1111 GeV).

4. **Three Hadrons (Three-prongs)**: Three charged particles having a mass such that $0.8 < M < 1.5$ GeV, ensuring the compatibility with the $a_1(1200)$ resonance. The tracks are required to be compatible in the event vertex within $\Delta z < 0.4\ cm$ and the combined charge should match unity.

In figure 4.4 a the reconstructed decay modes of the τ_h are presented.

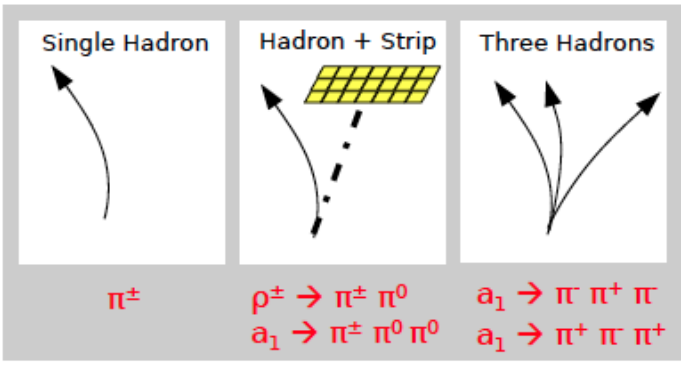


Figure 4.4: *Sketch showing the τ_h reconstructed decay modes using the HPS algorithm.*

These decay modes reconstruction target the $h^-\nu_\tau$, $h^-\pi^0(\pi^0)\nu_\tau$ and $h^-h^+h^-\nu_\tau$ decays that account for $\sim 75\%$ of the hadronic part of the τ decay branching ratio. The decays $h^-h^+h^-\pi^0\nu_\tau$, representing 4.8% of the branching ratio, are not considered due to the high contamination of $jet \to \tau_h$ fakes but can experimentally fall into one of the previous categories.

Candidates that fail the mass window cut are rejected. The four momenta of the τ_h candidate is then computed as the sum of the momenta of the charged particles plus strips

[2]The set of input objects is limited to the 6 highest p_T charged particles and the 6 highest p_T strips in order to limit the computing time

considered in the decay mode reconstruction. An additional requirement is applied to the τ_h candidates of the second, third and fourth decay mode. The charged particle and strips entering the combination should lie inside a cone around the τ_h candidate of size:

$$\Delta R = \begin{cases} 0.05 & \text{if } p_T > 60 \text{ GeV} \\ 3.0/p_T(GeV) & \text{if } 30 < p_T \leq 60 \text{ GeV} \\ 0.10 & \text{if } p_T \leq 30 \text{ GeV} \end{cases} \quad (4.1)$$

The shrinking cone takes into account that the decay products of a high p_T tau are more collimated.

If many hypothesis are retained after the mass window and shrinking cone requirements, the τ_h candidate with the highest p_T is kept.

4.3.2 Performance

The performance of the HPS reconstruction algorithm is evaluated in terms of decay modes reconstruction efficiency and energy resolution using simulated Monte Carlo events. The performance is evaluated using $Z/\gamma^* \to \tau\tau$ events which covers a big part of the phase space of interest for physics analysis.

Reconstructed energy

The energy reconstruction quality is determined by two quantities: the response and resolution, *i.e.* the mean and width of the ratio between the reconstructed and the generator level transverse momentum of the visible decay products of the τ.

In figure 4.5 the distributions of this ratio are given for the different generated decay modes in many pile-up scenarios, using $Z/\gamma^* \to \tau\tau$ events. The energy reconstruction in the *Single Hadron* and *Three Hadrons* decay modes (*Single Hadron plus Strips*) has a good resolution of $\lesssim 3\%$ ($\lesssim 10\%$) with a good stability against pile-up.

Reconstructed decay modes

The correlation between the generated and the reconstructed decay modes is the figure of merit to quantify the decay mode reconstruction performance. It is represented in the form of a 3×3 matrix, each entry (i, j) of the matrix represents the fraction of taus generated in the decay channel i,

$$i \in \{h^-\nu_\tau, \quad h^-\pi^0(s)\nu_\tau, \quad h^-h^+h^-\nu_\tau\}$$

that are reconstructed by the HPS algorithm in the decay mode j,

$$j \in \{\text{one-prong}, \quad \text{one-prong plus one or two strips}, \quad \text{three prongs}\}$$

In figure 4.6, the correlation matrices obtained from simulated $Z/\gamma^* \to \tau\tau$ and $Z'(2.5 \text{ TeV}) \to \tau\tau$ events are presented for different pile-up conditions. The results

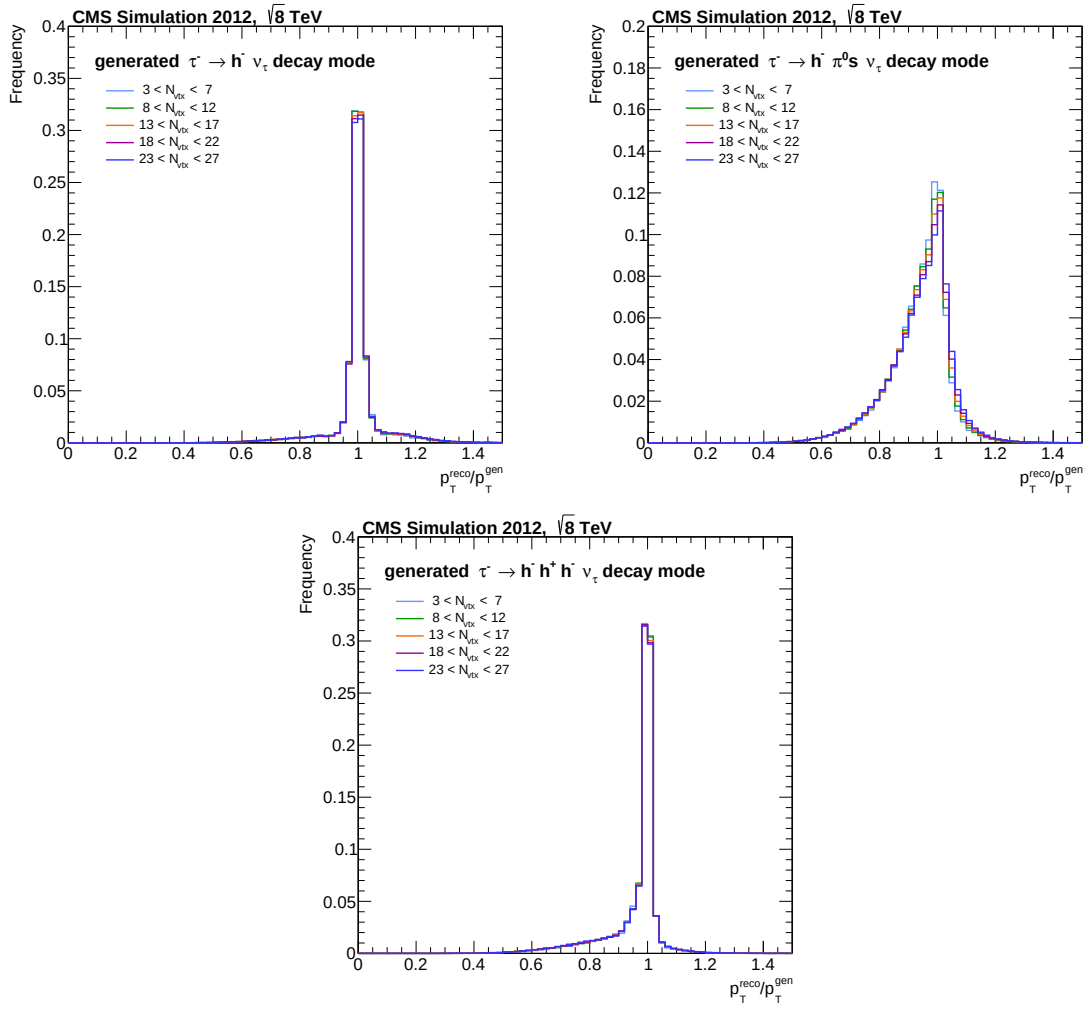


Figure 4.5: *Ratio between the reconstructed and the generated transverse momentum of the visible p_T of the τ_h using simulated $Z/\gamma^* \to \tau\tau$ events for different generated decay modes: $\tau^- \to h^-\nu_\tau$ (top left), $\tau^- \to h^-\pi^0\nu_\tau$ plus $\tau^- \to h^-\pi^0\pi^0\nu_\tau$ (top right) and $\tau^- \to h^-h^+h^-\nu_\tau$ (bottom). 5 distributions are displayed for different pile-up scenarios. N_{vtx} denotes the number of reconstructed vertices in the event. From [140].*

show that the correct decay mode is reconstructed in $\sim 90\%$ of the cases for taus from $Z/\gamma^* \to \tau\tau$ decays. For high p_T taus, the probability to reconstruct 3-prong decay modes decreases. This is due to three effects: the reconstructed hits in the pixel detector are merged due to the high momenta of the system, tracks are lost due to the cuts applied in the Particle Flow algorithm and tracks are rejected by the $d_0 < 0.03$ *cm* quality cut applied in the τ_h reconstruction (*c.f.* table 4.3).

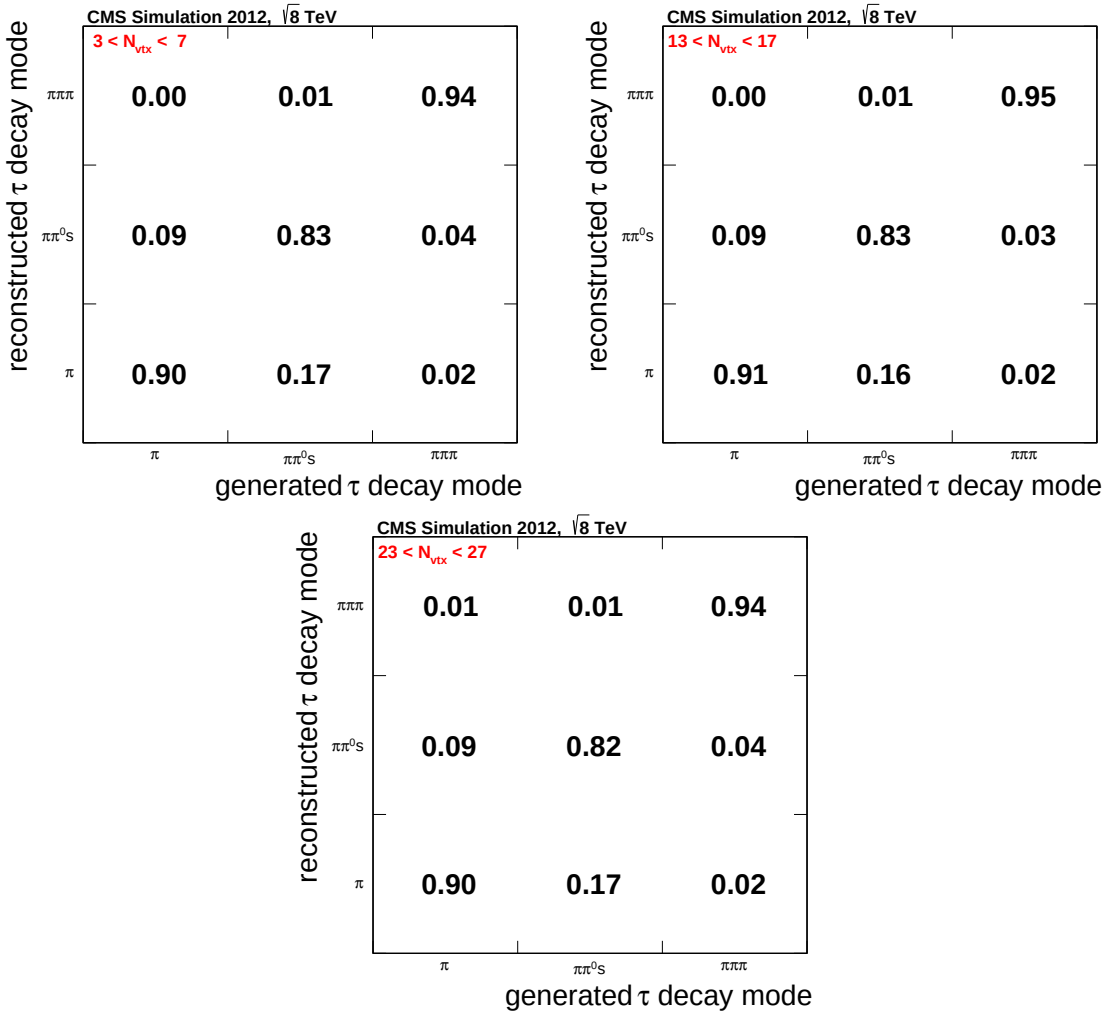


Figure 4.6: *Correlation between generated and reconstructed decay modes of τ_h in simulated $Z/\gamma^* \to \tau\tau$ events for different pile-up scenarios. From [140].*

4.4 Introduction to multivariate analyses

A multivariate analysis (MVA) is a statistical tool able to simultaneously use the information provided by many variables. It is based on machine learning methods able to

separate a given signal from backgrounds (classification) or to estimate the most compatible value of a given observable given a set of input variables correlated with the observable of interest (regression). An example of regression is the MVA $\not\!\!E_T$ regression described in section 3.3.7 which combines the information given by different estimation of the missing transverse energy to reduce the pile-up effects. The MVA-based discriminators described in this section, as well as the MVA-based isolation presented in section 4.5.2 are classification multivariate analyses.

The simplest classification technique is the so-called cut-based analysis, in which several rectangular cuts are done in the phase space of the $(X_1, \ldots, X_N)$ input variables in order to separate a given signal from background. This method is the easiest to implement and the most reliable in terms of understanding, as the signal to background ratio can be obtained at each step of the cuts sequence. But this technique has intrinsically limited performance, as it does not take advantage of the correlations between the input variables and partially includes the shape information. As a result the purity in signal of the selected sample is not optimal. In the left plot of figure 4.7 an example in two dimensions is given for a cut-based analysis.

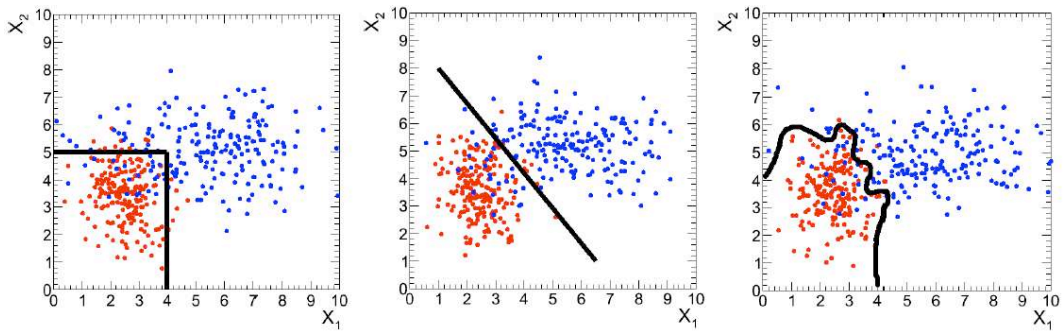


Figure 4.7: *Different types of MVA methods: rectangular cuts (left), linear cuts (center) and non-linear discriminator (right) for a two dimensional example. The signal (background) events are represented by the blue (red) dots. From Ref. [155].*

More complex classifiers can then be used in order to include the correlations between variables. For example linear combinations of the input variables (shown in the center plot of figure 4.7), or even more complex non linear discriminators (right plot of figure 4.7). Usually those types of MVA classifiers combine the discriminating power of the input variables into a single one dimensional output, called the MVA output or MVA score. Then a single or various cuts on it are performed in order to achieve different signal purity and background contamination.

Several MVA techniques were introduced in the past and in High Energy Physics the most successful ones are the *Likelihood, Neutral Networks* and *Boosted decision Trees* [156]. All rely in the full information given by the input variables, their shapes and correlations.

4.4.1 MVA training and testing

The first step in the development of a multivariate analysis is the choice of the input variables. They are chosen in order to maximize the discriminating power of the algorithm, but some of them can also be included (sometimes as spectator variables) to give more information to the MVA. For example the kinematical variables as the η and p_T in a reconstruction algorithm can provide information about the typical phase space where the signal or background would typically lay. Therefore the choice of the input variables have to be physically justified so the performance of the MVA will strongly depend on the discriminating power and modeling of the inputs.

After defining the input variables, the next step is the *training* of the multivariate analysis. During this step, the MVA builds the discriminator based on the information available. The training is done on two classes, one for signal events and a second containing the background events we want to reject. This step introduces machine learning techniques in which the algorithm maximizes the separation between the two samples and creates the MVA score.

Finally, after the training is done, another important step is the *testing* phase. An independent set of signal and background samples is used to evaluate the MVA response and to calibrate the MVA training. An overtraining check is usually done to make sure that the MVA is not biased and avoid that the training step was specific to the input sample. Hence this validates the expected performance of the MVA classifier and its reliability. Additional checks on the data to simulation agreement are needed in order to avoid potential biases on the MVA performance in data.

4.4.2 Boosted Decision Trees

In High Energy Physics analysis, the Boosted Decision Trees (BDT) [156] were introduced recently [157,158] and are now widely used. It is a type of multivariate analysis classifier in which a binary tree structure is built in order to separate signal from background events. It starts with a root node and the sequential and successive binary cuts are applied in the different nodes until the last ones called leafs which classify the events as background or signal. The BDT structure is displayed in figure 4.8.

The growing or building of the Boosted Decision Tree is done during the training step where the split criteria of each node is defined. From the root node an initial splitting criteria is determined and the events are split into two other nodes depending if they pass the criteria or not. The two subsets of training events go through the same algorithm and the next splitting iteration is determined. This procedure is repeated until a certain node has reached either a minimum number of events, or a minimum or maximum signal purity. The node purity is defined as:

$$P = \frac{\sum_S W_S}{\sum_S W_S + \sum_B W_B}$$

where W_i is the weight of the i^{th} event and $\sum_S$ ($\sum_B$) is the sum over the signal (background) events. The signal and background mixing within a node can be defined in order

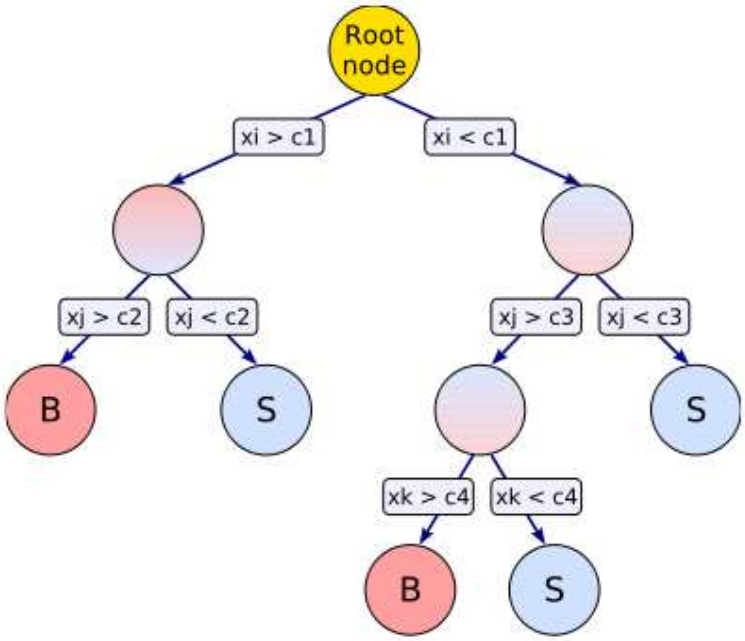


Figure 4.8: *Illustration of a Boosted Decision Tree structure . Starting from the Root node, a cut is performed on the variable x_i (maximizing the signal to background separation at this node) and the events are split into two other nodes depending whether they pass the cut or not. Each node is then split again using the same variable or a different one. This operation is repeated until the parametrizable minimal number of events at a node, the maximum number of nodes (here 8) or the maximum number of layers (depth, here 3) are reached. The final nodes are classified into signal-type (S) or background-type (B) depending on the majority of events of each kind populating the node.*

to characterize the performance of a variable or a cut criteria inside a node. One common way is by using the Gini index [159] defined as:

$$Gini = \sum_i W_i P(1-P)$$

The training procedure selects the variable and cut value that maximize the gain in the Gini Index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events. The separation gain of such split in a Decision Tree can be quantified using the formula:

$$G = Gini_{parent} - (Gini_1 + Gini_2)$$

The *growing* process iterates the splitting procedure by maximizing the gain.

Decision Trees are known to be sensitive to the statistical fluctuations in the input variables used to create the tree structure. For example if two variables have similar discrimination power, a statistical fluctuation of one of the two may cause the tree growing algorithm to chose it instead of the other one. The whole tree structure is then biased

by this choice, leading to a potential difference of the classifier score if the other variable had been chosen.

This problem can be avoided by applying the so-called *boosting* procedure. A forest of decision trees is grown from the same input samples with the events being weighted differently. The idea behind the boosting is that the events that were misclassified during the training of a decision tree are given a higher common event weight w_i in the training of the following tree. This weight quantifies the event misclassification. The boosting can be done several times (typically between 100 and 500). Boosting increases the statistical stability of the classifier and improves the separation performance compared to a single decision tree. The new type of classifier is called boosted decision tree (BDT). In the so-called AdaBoost method the weight takes the form $w_i = (1 - f_{err})f_{err}$ where f_{err} is the fraction of misclassified events in the final nodes of the previous tree. The BDT response is a weighted average of the error-rate of the individual tree responses $C(i)(x)$ for a given vector of input variables $\overrightarrow{x}$:

$$y(\overrightarrow{x}) \propto \sum_{i}^{N_{Trees}} \ln w_i \cdot C(i)(\overrightarrow{x}) \tag{4.2}$$

The *boosting* procedure gives more weight to events that are difficult to categorize as signal or background, and the final score is an average of all the scores obtained with different weights.

4.4.3 Toolkit for Multivariate Analysis (TMVA)

The Toolkit for Multivariate Analysis (TMVA) [155] provides a machine learning environment, integrated in ROOT, convenient in the context of high energy physics. This toolkit is used for the processing and parallel evaluation of multivariate classification techniques. TMVA consists of object-oriented implementations in C++ of a number of multivariate methods and provides training, testing and performance evaluation algorithms and visualization scripts. The training and testing steps are performed with datasets where the true event classification is known, for example samples created by Monte Carlo simulation or data enriched regions.

Several multivariate methods are embedded in TMVA. The most known being *Rectangular cuts, Fisher, Likelihoods, Boosted Decision Trees and Neural Networks.* The multivariate analysis presented in this thesis are all trained and tested within the TMVA framework.

4.5 τ_h identification through τ_h isolation

The main handle to keep a good identification efficiency while rejecting the large $jet \to \tau_h$ fake contribution is the isolation of the τ_h candidates. The τ-lepton is colorless and its decay occurs purely through weak interaction. Therefore the reconstructed τ_h are typically produced in a relatively clean environment, isolated from other reconstructed particles, contrary to the quark and gluon jets. In CMS two types of isolation discriminators have

been developed in order to keep the fake rate from QCD jets smaller than the percent level [140]. They use either a cut-based selection or a multivariate approach (MVA) which utilizes the information in the event due to the small but non negligible [3] τ lifetime. I will detail both in the following sections.

4.5.1 Cut-based isolation

In order to compute the isolation of the τ_h, the electromagnetic or charged particles found within $\Delta R < 0.5$ from the τ_h candidate direction are treated as "isolation particles". A charged particle is considered as an "isolation particle", if its $p_T > 1.0$ GeV and passes the track requirements listed in table 4.4. For the photons, they should have $E_T > 1.5$ GeV.

Variable	Cut
p_T	> 0.5 GeV
χ^2	< 100
d_0	$< 0.03\ cm$
d_Z	$< 0.2\ cm$
N_{hit}	≥ 3

Table 4.4: *Track selection criteria applied to charged particles when computing the τ_h isolation.*

The impact parameter is computed with respect to the vertex closest to the "leading" track (the one with the highest p_T) of the τ_h candidate.

In order to maintain an isolation robust against the pile-up rate, only the charged particles with their track matched to the τ_h production vertex are taken into account. As the photons are difficult to associate to the τ_h production vertex their contribution have to be estimated otherwise. An event-by-event estimation of the pile-up photon energy (also called $\Delta\beta$) is computed by summing the transverse momenta of all the charged particles whose direction match the τ_h candidate by $\Delta R < 0.8$ and which are not matched to the τ_h production vertex ($d_Z > 0.2\ cm$), so they are coming from pile-up vertices. This sum is scaled by a factor ~ 0.46 which represents the ratio between the neutral and the charged pile-up contribution and was optimized to make the τ_h identification efficiency less sensitive to pile-up.

The isolation is computed as the sum of transverse momenta of the Particle Flow particles in the isolation cone, after subtracting $\Delta\beta$. All the constituents of the τ_h candidate (charged particles and photons) are excluded from the sum:

$$I_\tau = \sum p_T^{charged}(d_Z < 0.2\ cm) + \max(\sum p_T^{\gamma} - \Delta\beta, 0) \tag{4.3}$$

with

$$\Delta\beta = 0.4576 \cdot \sum p_T^{charged}(d_Z > 0.2\ cm)$$

[3] with respect to the transverse impact parameter resolution of the detector.

Thresholds on the isolation value of 2.0, 1.0 and 0.8 GeV define the loose, medium, and tight $\Delta\beta$-corrected working points for the cut-based discriminator (also called *HPS 3-Hit*).

τ_h identification efficiency and expected performance

The τ_h identification efficiency is defined as the efficiency to pass the reconstruction plus the isolation requirement for the reconstructed τ_h with generated $p_T > 20$ GeV and $|\eta| < 2.3$:

$$\epsilon = \frac{p_T^{rec} > 20 \text{ GeV } \& \ |\eta^{rec}| < 2.3 \ \& \text{ Reconstruction } \& \text{ Isolation}}{p_T^{gen} > 20 \text{ GeV } \& \ |\eta^{gen}| < 2.3} \tag{4.4}$$

The expected τ_h identification efficiency for the cut-based isolation is displayed in figure 4.9 as a function of the transverse momentum and the number of reconstructed vertices in the event. The τ_h identification efficiency increases with the transverse momentum of the visible decay products and is pile-up insensitive.

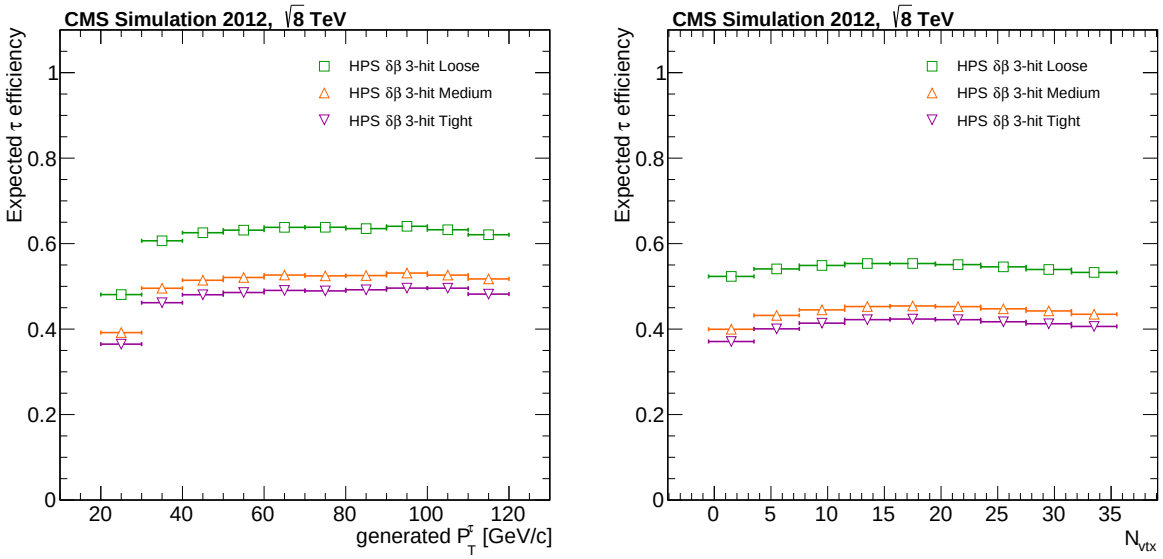


Figure 4.9: *τ_h identification efficiency for the cut-based isolation in simulated $Z/\gamma^* \to \tau\tau$ events as a function of the τ_h candidate transverse momenta (left) and the number of reconstructed vertices (right). From [140].*

4.5.2 MVA-based isolation

A novel isolation was developed based on a multivariate analysis discriminator (*c.f.* section 4.4). The analysis presented in this thesis is the first one using this approach and constitutes a baseline for the following analysis in CMS. The τ leptons have a non negligible lifetime, consequently their signature in the detector can be analyzed using the lifetime information, especially in the boosted τ_h topologies. The MVA-based isolation aims at tagging the τ_h-jets.

The multivariate approach makes use of the correlation between the isolation variables in order to maximize the rejection power against jet misidentified as hadronic taus. On top of the quantities used in the cut-based isolation, the MVA approach makes use of the transverse impact parameter of the "leading track" of the τ_h candidate, and the secondary vertex distance to the primary vertex when the τ_h candidate is reconstructed in the $3-prong$ decay mode. These additional variables introduce additional information that has discriminating power against jet fakes, and is more suited for boosted τ_h. In figure 4.10 the correlation between the τ_h flight distance versus transverse momenta for fakes and real taus and in figure 4.11 the τ_h flight distance distribution is shown for generated taus from $\phi(m_H = 1000\ \text{GeV}) \rightarrow \tau\tau$ events and for fake taus from $W + jets$ events.

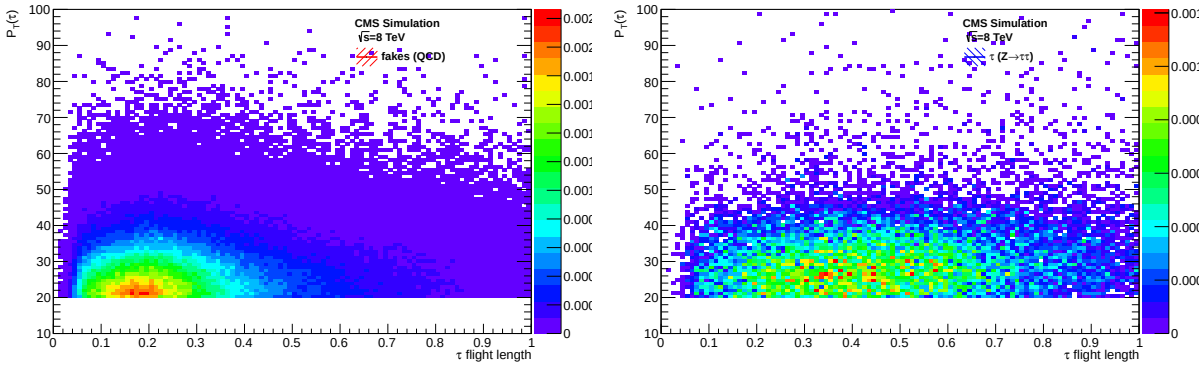


Figure 4.10: *τ_h flight distance versus transverse momenta for fakes (left) and real (right) taus.*

The real τ-leptons have a bigger reconstructed flight distance than the misidentified jets. This handle is taken into account in the MVA-based isolation.

Input variables

The multivariate discriminator is based on a Boosted Decision Tree (BDT) [156] (see section 4.4 for an introduction to multivariate analysis) trained to separate hadronic tau decays ("signal") from quark and gluon jets misidentified as taus ("background"). It is trained using the 11 following input variables:

- The τ_h candidate p_T and η.

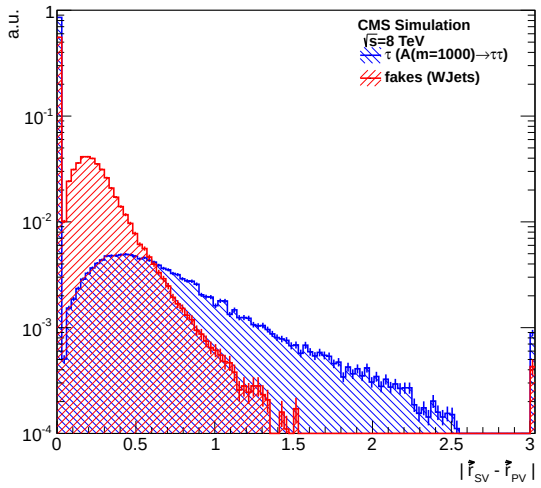


Figure 4.11: *τ_h flight distance for fakes (red) and real (blue) taus from QCD multijets and $Z/\gamma^* \to \tau\tau$ events respectivelly.*

- The three contributions to the isolation sum: the charged particles isolation $\sum p_T^{charged}(d_Z < 0.2\ cm)$, the neutral particles contribution $\sum p_T^{\gamma}$ and the $\Delta\beta$ correction.
- The reconstructed decay mode.
- The transverse impact parameter d_0 of the leading track of the τ_h candidate and its significance d_0/σ_{d_0}.
- The distance between the production and decay vertex of the τ_h candidate $|\vec{r}_{SV} - \vec{r}_{PV}|$ and its significance $|\vec{r}_{SV} - \vec{r}_{PV}|/\sigma_{|\vec{r}_{SV}-\vec{r}_{PV}|}$, plus a flag indicating that a decay vertex has been reconstructed for a given τ_h candidate.

Figures 4.12 and 4.13 shows the input variables distributions for simulated events, $Z/\gamma^* \to \tau\tau$ for signal and $W + jets$ for background. The discriminatory power of all the input variables is shown.

The kinematical input variables p_T and η parametrize the other input variables in order to take into account possible dependency on the kinematical regimes. The events used for the training are reweighted as to ensure that the p_T and η distributions are identical for "signal" and "background". This ensures that the MVA will not depend on the events pseudo-rapidity or transverse momenta.

Training of the Boosted Decision Trees

The Boosted Decision Trees are trained on Monte Carlo simulated samples amounting to a statistics of $\mathcal{O}(10^6)$ events. "Signal" containing real generated τ-leptons samples are chosen: Drell-Yann production of $Z/\gamma^* \to \tau\tau$, Standard Model and MSSM Higgs $H/\phi \to \tau\tau$, $Z' \to \tau\tau$ and $W' \to \tau\nu$. The reconstructed τ_h matched to generated τ-leptons within $\Delta R < 0.3$ are considered for the training.

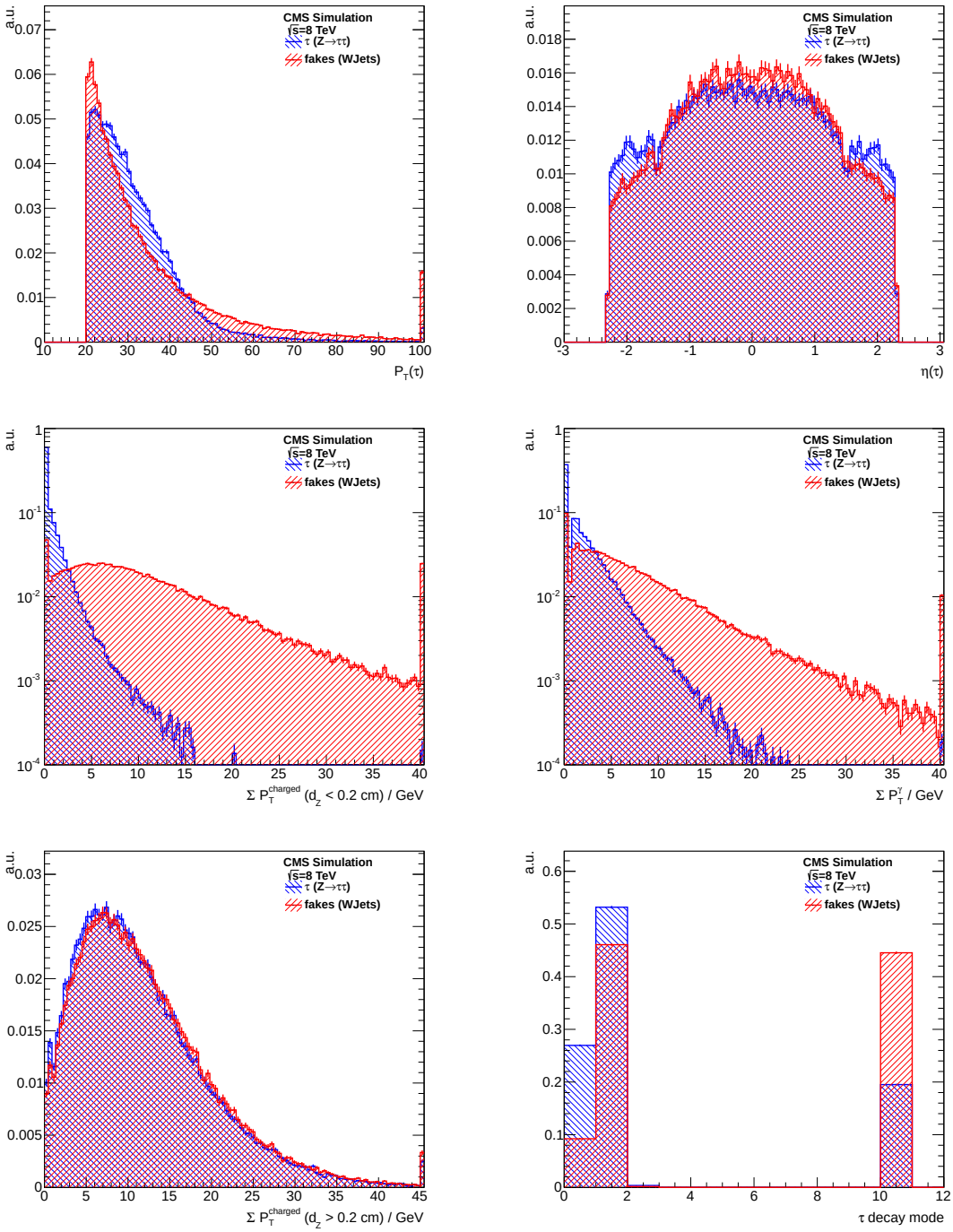


Figure 4.12: *Input variables distributions, normalized to unity, for the MVA-based isolation for simulated $Z/\gamma^* \to \tau\tau$ (blue) and $W + jets$ (red).*

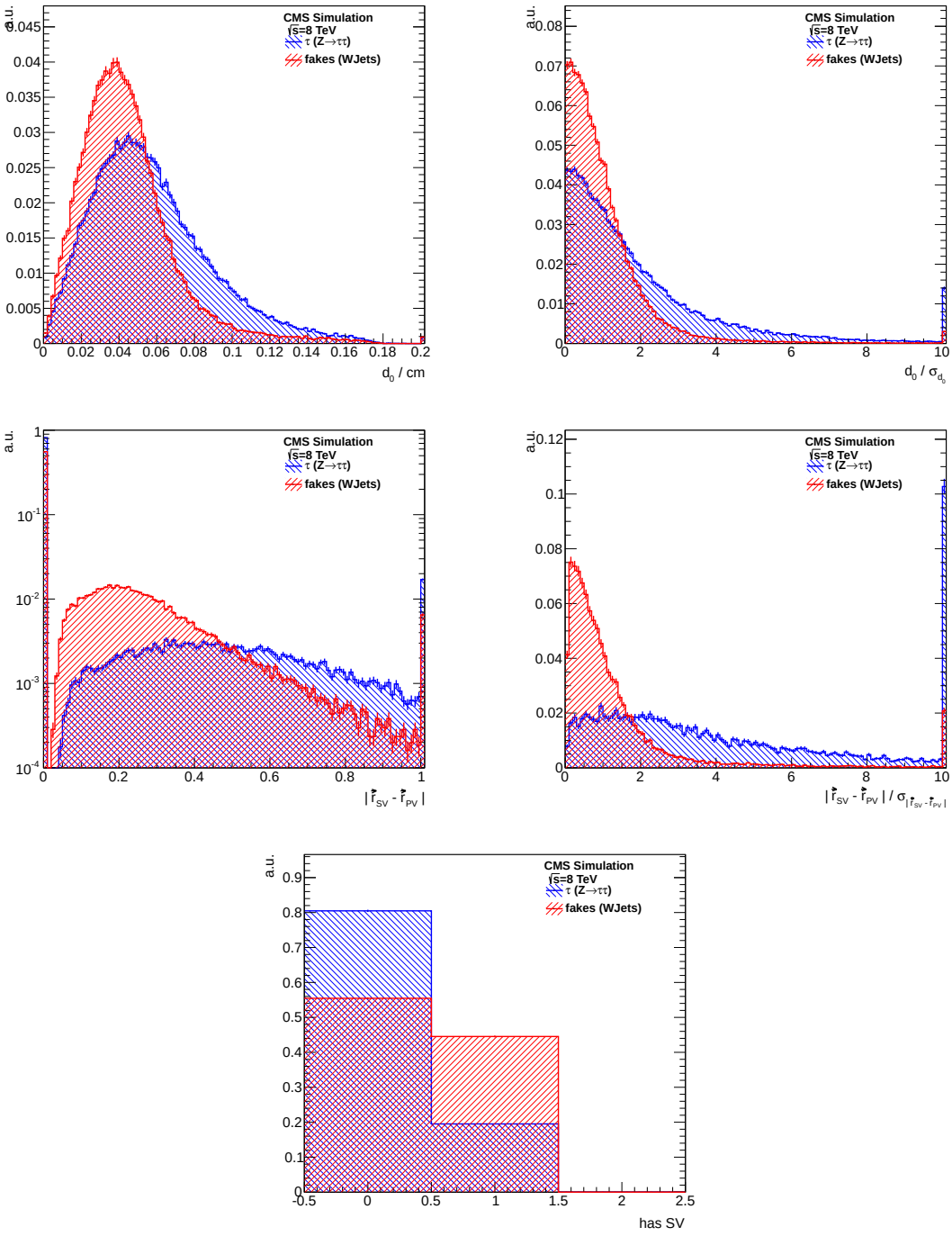


Figure 4.13: *Input variables distributions, normalized to unity, for the MVA-based isolation for simulated $Z/\gamma^* \to \tau\tau$ (blue) and $W + jets$ (red).*

For the "background" samples QCD-multijets and $W + jets$ simulated Monte Carlo are used. To be considered in this category, the reconstructed τ_h is required to not match the leptons originating from the W decays.

The sum of the samples cover the range in $p_T(\tau_h)$ between 20 to 2000 GeV. Half of the samples are used to train the MVA, the other half is used to evaluate the performance and perform overtraining checks. The training options used are the same as the ones of the anti-electron discriminator training that we will detail in section 5.3.3. The training is validated by comparing data to simulation agreement in the input variables and by comparing the expected ROC curves for the training dataset and the testing dataset. An overall good agreement is found [140]. In figure 4.14 the BDT output distribution is presented. A good separation of the "signal" and "background" species is seen.

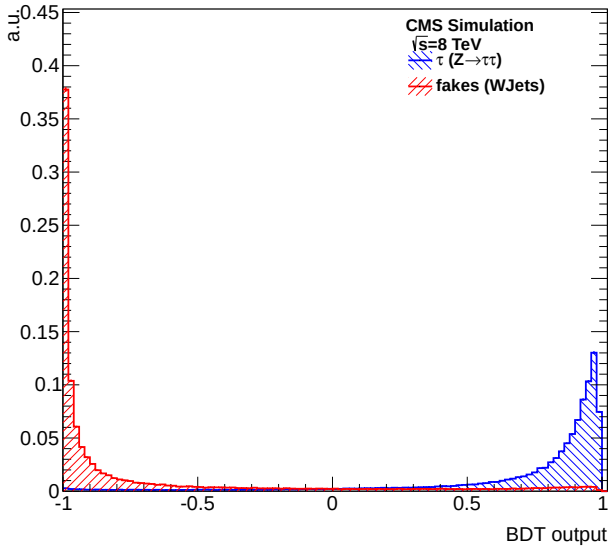


Figure 4.14: *BDT output for fakes (red) and real (blue) taus.*

A separate training is done without including the tau lifetime information in order to determine the gain on performance coming from the use of the additional information (see figure 4.16).

Working points

The working points of the MVA discriminator are chosen such that the τ_h identification efficiency is flat versus the τ_h transverse momentum. A target efficiency in the test sample is chosen for each working point (40% to 90% in steps of 10%).

The transverse momenta phase space is divided in 31 bins in the range $20 - 5000$ GeV [4], and for each bin the cut on the MVA output is chosen as to give the target efficiency

[4]The following binning is used: 20, 22.5, 25, 27.5, 30, 32.5, 35, 37.5, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 1000, 1200, 1500, 5000 GeV.

(*e.g.* 60% for the Tight working point). This ensures that the τ_h identification efficiency is flat in all the p_T range.

τ_h identification efficiency and expected performance

As for the cut-based isolation we can draw the τ_h identification efficiency (see equation 4.4) as a function of the τ_h transverse momentum and the number of reconstructed vertices (see figure 4.15). The same trend for the MVA-based isolation is seen: the τ_h identification efficiency is reasonably flat with the transverse momenta and remains reasonably pile-up insensitive.

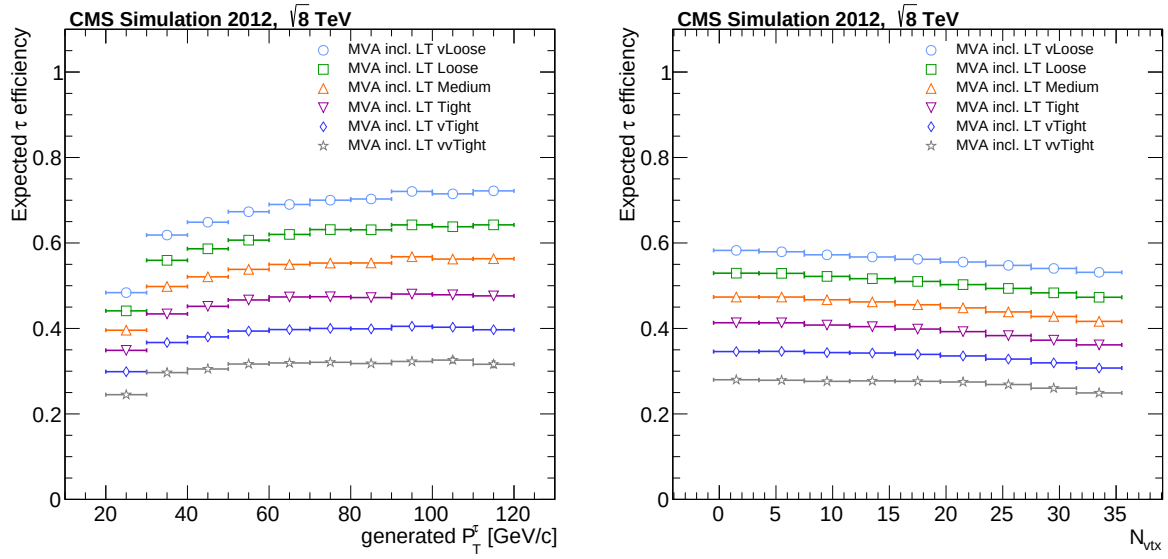


Figure 4.15: *τ_h identification efficiency for the MVA-based isolation in simulated $Z/\gamma^* \to \tau\tau$ events as a function of the τ_h transverse momentum (left) and the number of reconstructed vertices (right). From [140].*

We can define the $jet \to \tau_h$ fake rate as the rate in which the quark/gluon jets are misidentified as hadronic tau decays with respect to jets with $p_T > 20$ GeV and $|\eta| < 2.3$:

$$f_r = \frac{p_T^\tau > 20 \text{ GeV } \& \ |\eta^\tau| < 2.3 \ \& \text{ Reconstruction } \& \text{ Isolation}}{p_T^{jet} > 20 \text{ GeV } \& \ |\eta^{jet}| < 2.3} \tag{4.5}$$

In table 4.5 the expected τ_h identification efficiencies and $jet \to \tau_h$ fake-rates are reported. The τ_h identification efficiency typically leans between 40 to 70% for $jet \to \tau_h$ fake-rates between 10^{-3} and 10^{-2}.

A common way to compare the performance of different isolation discriminators is to draw the ROC curves, the curve of the fake-rate (f_r) as a function of the efficiency (ϵ). In figure 4.16 the ROC curves for the cut-based and the MVA-based isolation are displayed for simulated $Z/\gamma^* \to \tau\tau$ and $Z'(2.5 \text{ TeV}) \to \tau\tau$ events[5].

[5]Two sets of cut-based isolation discriminators are computed depending on the number of hits in the

Working point	τ_h identification efficiency		$jet \to \tau_h$ fake-rate	
	$Z/\gamma^* \to \tau\tau$	$Z'(2.5\ \text{TeV}) \to \tau\tau$	$W + jets$	QCD multi-jet
Cut-based isolation				
Loose	56.1%	67.2%	$9.82 \cdot 10^{-3}$	$6.92 \cdot 10^{-3}$
Medium	44.9%	57.9%	$5.59 \cdot 10^{-3}$	$3.76 \cdot 10^{-3}$
Tight	41.9%	54.8%	$4.78 \cdot 10^{-3}$	$3.19 \cdot 10^{-3}$
MVA-based isolation				
Very Loose	56.1%	71.7%	$7.36 \cdot 10^{-3}$	$5.63 \cdot 10^{-3}$
Loose	50.9%	64.8%	$4.28 \cdot 10^{-3}$	$2.92 \cdot 10^{-3}$
Medium	45.4%	57.9%	$2.84 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$
Tight	39.7%	51.0%	$1.98 \cdot 10^{-3}$	$1.18 \cdot 10^{-3}$
Very Tight	33.7%	44.1%	$1.41 \cdot 10^{-3}$	$7.83 \cdot 10^{-4}$
Very Very Tight	27.4%	36.6%	$9.61 \cdot 10^{-4}$	$4.96 \cdot 10^{-4}$

Table 4.5: *τ_h identification efficiencies and $jet \to \tau_h$ fake-rate for the cut-based and MVA-based isolation discriminators evaluated in simulated $Z/\gamma^* \to \tau\tau$, $Z'(2.5\ TeV) \to \tau\tau$, $W + jets$ and QCD multi-jet events. From [140].*

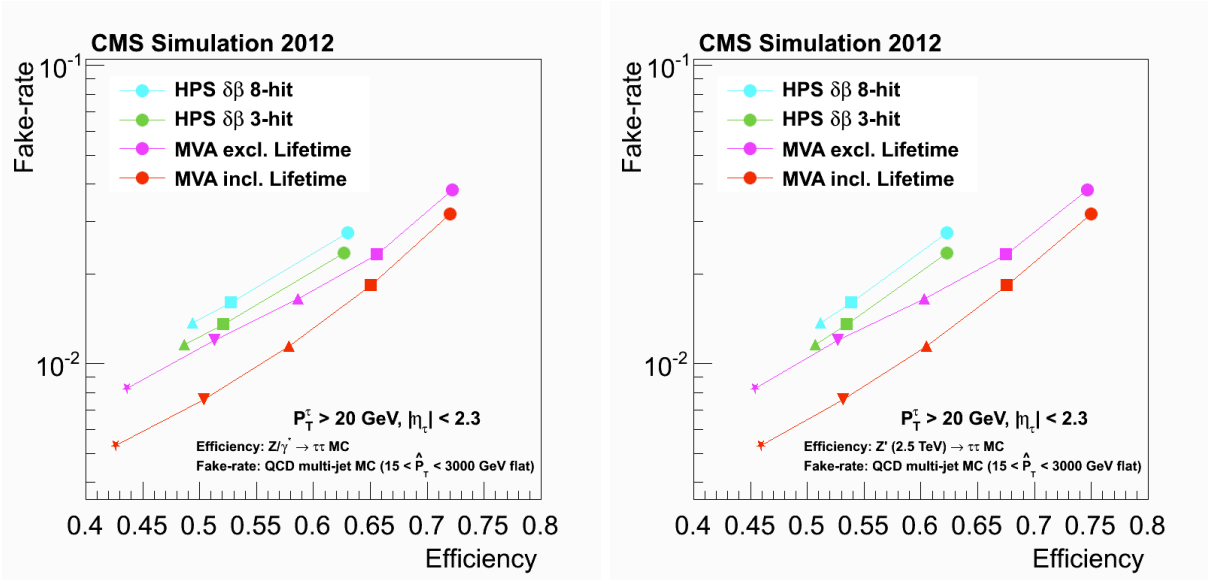


Figure 4.16: *ROC curves for the different isolation discriminators in $Z/\gamma^* \to \tau\tau$ (left) and $Z'(2.5\ TeV) \to \tau\tau$ (right) events. The cyan and green lines correspond to the cut-based isolation discriminators. The magenta and red lines to the MVA-isolation without and with τ_h lifetime information. From Ref. [160].*

pixel and silicon strip tracking detectors ($N_{hits} > 3$ or $N_{hits} > 8$). The one with the looser requirement is the one used in physics analyses.

The MVA strategy reduces the $jet \to \tau_h$ fake-rate by $40 - 50\%$ with respect to the cut-based approach for a given τ_h identification efficiency. We can note also that the improvement is bigger for the $Z^{'}(2.5\ \text{TeV}) \to \tau\tau$ sample, indicating that the MVA-isolation is more suited for boosted topologies of the τ. Finally we can deduce that the gain is mainly due to the τ_h lifetime information, as the performance of the MVA-isolation trained without this information is close to the cut-based isolation performance. This correspond to the intrinsic expected performance of the τ_h reconstruction and isolation discriminators. In the next sections I will present the measurements of the τ_h identification and $jet \to \tau_h$ fake rate measured in data.

4.5.3 τ_h identification efficiency measured in data

The τ_h identification efficiency is measured in data using the so-called Tag-and-probe technique using either $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ or $t\bar{t} \to bb\mu\tau_h$ events from the 2012 dataset. The two measurements are redundant in order to have a cross checked measurement and we present here the procedure for $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events.

Tag-and-probe using $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events

The events considered for this measurement are the ones firing the single muon trigger. The tag is an isolated triggering muon passing the Tight Particle Flow identification working point and with transverse momenta greater than 25 GeV. The probe is a τ_h candidate associated with a jet containing at least one charged particle with $p_T > 5$ GeV. A loose preselection on the probe is applied in order to suppress background events: the τ_h candidate "leading track" is required to be of opposite charge to the muon and coming from the same production vertex as the muon, the candidate should pass the Loose anti-electron discriminator (see chapter 5) and should not overlap with the tag. Additional third lepton veto is applied and events with b-Tagged jets are rejected also. A cut on the transverse mass between the lepton and the missing energy is also applied to reject the $W + jets$ background (see section 6.3.5 for the definition).

The τ_h identification efficiency can then be measured on the latter. The τ_h identification efficiency is estimated by measuring the number of probes that pass or fail the tau identification and isolation criteria:

$$\epsilon_\tau = \frac{N^\tau_{pass}}{N^\tau_{pass} + N^\tau_{fail}} \tag{4.6}$$

The Pass and Fail regions are fitted simultaneously using a maximum likelihood fit. The fit is constrained by nuisance parameters as for the analysis presented in this thesis (see section 6.6). The observables used for the fit is N_{tracks}, the total number of tracks in the signal plus the isolation cones of the τ_h candidate. N_{tracks} is lower for real taus and higher for jets being misidentified, this variable is chosen as it is weakly correlated to the transverse momentum of the candidate. In figure 4.17 an example of templates fit is displayed. A cross-check is done by fitting the visible mass of the tag and the probe, m_{vis}.

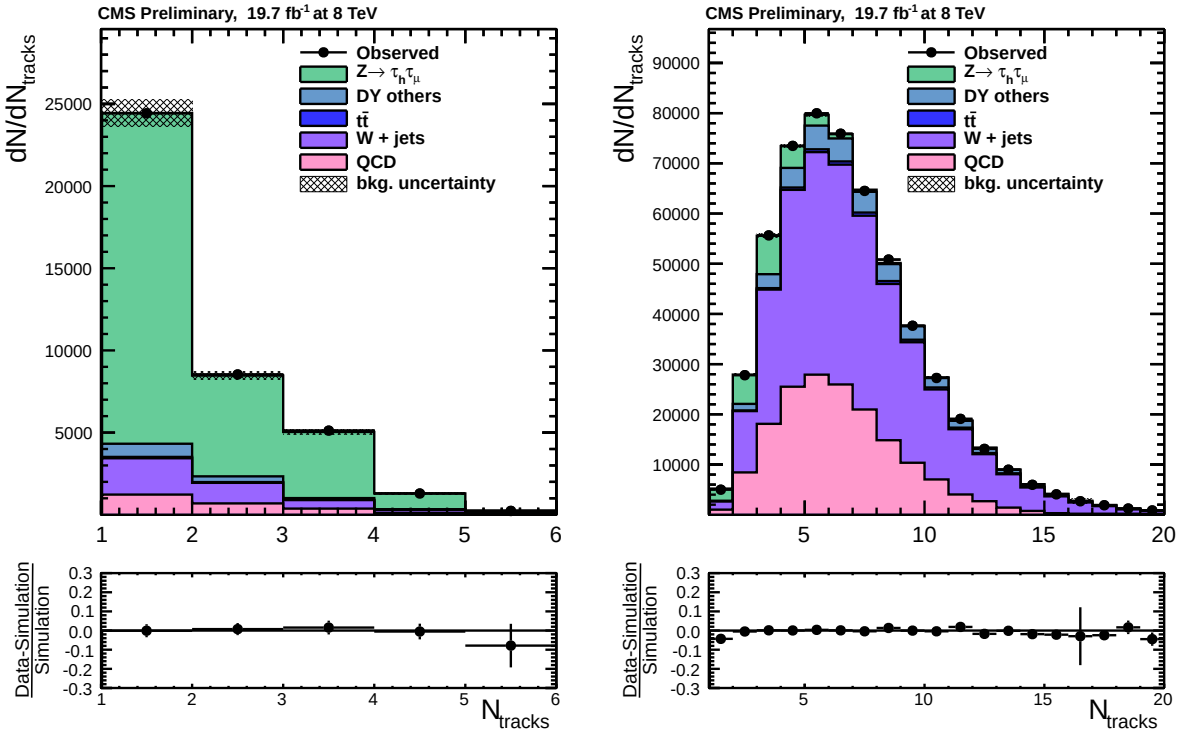


Figure 4.17: *Distribution of N_{tracks} in the Pass (left) and Fail (right) regions for the Tight working point of the MVA-based isolation after the maximum likelihood fit is performed. From [140].*

This measurement using the full $19.7 fb^{-1}$ at $\sqrt{s} = 8$ TeV data allows to measure the τ_h identification efficiency in data as presented in figure 4.18 and to extract data to simulation scale factors. The fits to N_{tracks} and m_{vis} distributions give compatible results and the data to simulation scale factors are compatible with unity within uncertainties from the maximum likelihood fit reading typically 5% (see table 4.6).

4.5.4 jet $\rightarrow \tau_h$ fake rate measured in data

As discussed previously, the τ_h reconstruction is highly contaminated by quark/gluon induced jets, an overwhelming background in hadron colliders as the LHC. The quarks and gluons fragmentation into multiple soft hadrons may spoil the isolation. Due to the high production cross-section of the QCD multi-jets production the tiny fraction of quarks and gluons which fragment into few hard hadrons can be misidentified as τ_h and constitutes the biggest background for the τ_h reconstruction and identification.

The measurement of the fake rate is performed in two types of samples: QCD multi-jets enriched and $W + jets$ enriched samples. The fake-rate is different for quark and gluon induced jets. The gluon induced jets have a bigger low p_T particle multiplicity and tend to be broader, leading to a lower fake rate. The quark induced jets are more

Working point	SF(N_{tracks})	SF (m_{vis})
Cut-based isolation		
HPS 3Hits Loose	0.991 ± 0.049	1.006 ± 0.044
HPS 3Hits Medium	0.978 ± 0.044	0.984 ± 0.044
HPS 3Hits Tight	0.990 ± 0.047	0.982 ± 0.044
MVA-based isolation		
MVA Very Loose	0.967 ± 0.048	1.034 ± 0.044
MVA Loose	0.968 ± 0.042	1.017 ± 0.044
MVA Medium	0.979 ± 0.042	1.018 ± 0.044
MVA Tight	0.979 ± 0.048	1.014 ± 0.044
MVA Very Tight	0.988 ± 0.060	1.012 ± 0.045
MVA Very Very Tight	1.048 ± 0.044	1.015 ± 0.045

Table 4.6: *Data-to-Monte Carlo scale factors (SF) for hadronic tau decays in $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events to pass different tau identification discriminators, measured using m_{vis} and N_{tracks} as observable in the TnP fit. From Ref. [140].*

collimated and they tend to pass the τ_h isolation requirements. The selection can bias the jet composition. The gluon initiated jets are dominating in the QCD multi-jets events after requiring two jets in the event. A quark type jets enriched sample can be obtained by requiring an isolated muon and high transverse mass between the muon and the missing energy (see section 6.3.5 for the definition). In this way, the event sample is enriched in $W + jets$ process: most of the events contain quark jets produced in association with a W boson. On the contrary, a sample of events with a muon but low values of the transverse mass is enriched in mainly b/c quarks.

The fake rate from jets is measured as a function of the jet p_T and compared to its Monte Carlo expectation as shown in figure 4.19. The fake rate at a given jet p_T is defined as the fraction of jets passing the τ_h reconstruction and identification criteria. The results show an agreement with expectation within $\sim 20\%$.

4.6 τ_h energy scale

The HPS algorithm reconstructs the τ_h by finding the corresponding decay modes inside the signal cone using the Particle Flow reconstructed particles as described in section 4.3.1. The HPS algorithm can sometimes miss some of the visible decay products of the tau decay and not be treated as a signal particle, in this case the final reconstructed momenta will be underestimated. The opposite scenario is also possible, an additional particle typically coming from pile-up can be mis-identified as being part of the τ_h. Then the reconstructed momenta will be overestimated. Other possible sources of mis-modeling

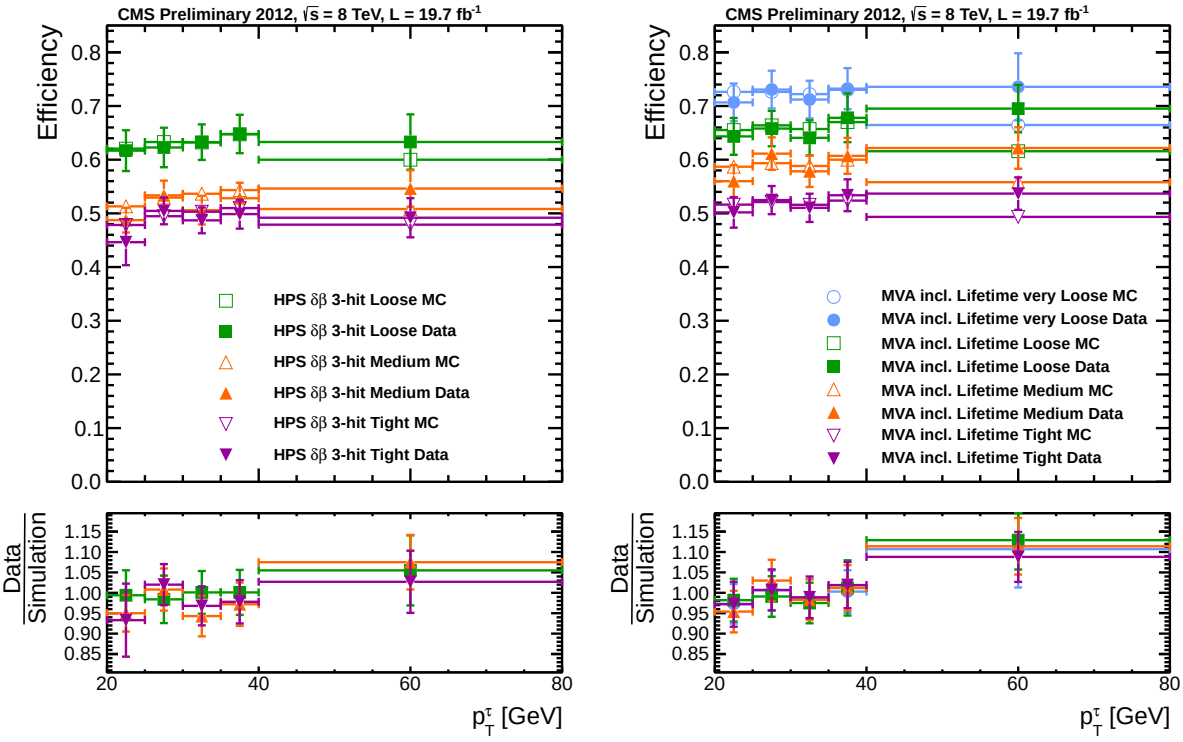


Figure 4.18: *τ_h identification efficiency as a function of the candidate transverse momenta measured in data and in Monte Carlo simulation for the cut-based isolation (left) and the MVA-based isolation (right). From [140].*

of the reconstructed tau energy can be caused by mis-calibration of the tracker and the calorimeters energy response. All these effects need to be well modeled by the simulation, then a correction to the tau energy scale (τ-ES) is needed.

As in section 4.5.3, $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events are used in order to access these corrections. The measurement is done via the modeling of the visible τ_h mass shape or via the visible $\mu\tau_h$ pair mass shape. A data to Monte Carlo simulation scale factor is determined by comparing those shapes in a $Z/\gamma^* \to \tau\tau$ dominated sample.

The scale factors are obtained for the different reconstructed decay modes from the HPS algorithm. The visible τ_h mass is defined only for the one prong plus strips or three prongs decay modes. The τ-ES for single hadron candidates is measured only by using the $m_{\mu\tau_h}$ observable as the visible τ_h mass is the one of the π resonance.

The shape templates for the samples containing simulated τ-leptons are produced for different values of the τ-ES varying between -6% and $+6\%$ in steps of 0.1%. In figure 4.20 the visible τ_h mass distribution is presented for the *Three Hadrons* decay mode with the central and extreme variations on the τ-ES. The best value of the τ-ES is extracted via a likelihood ratio comparing two hypotheses: the null hypothesis where the τ-ES=0% (meaning no correction of the simulation) and the tested hypothesis where the τ-ES=X%. The likelihood for each hypothesis is computed by comparing the data to the predicted

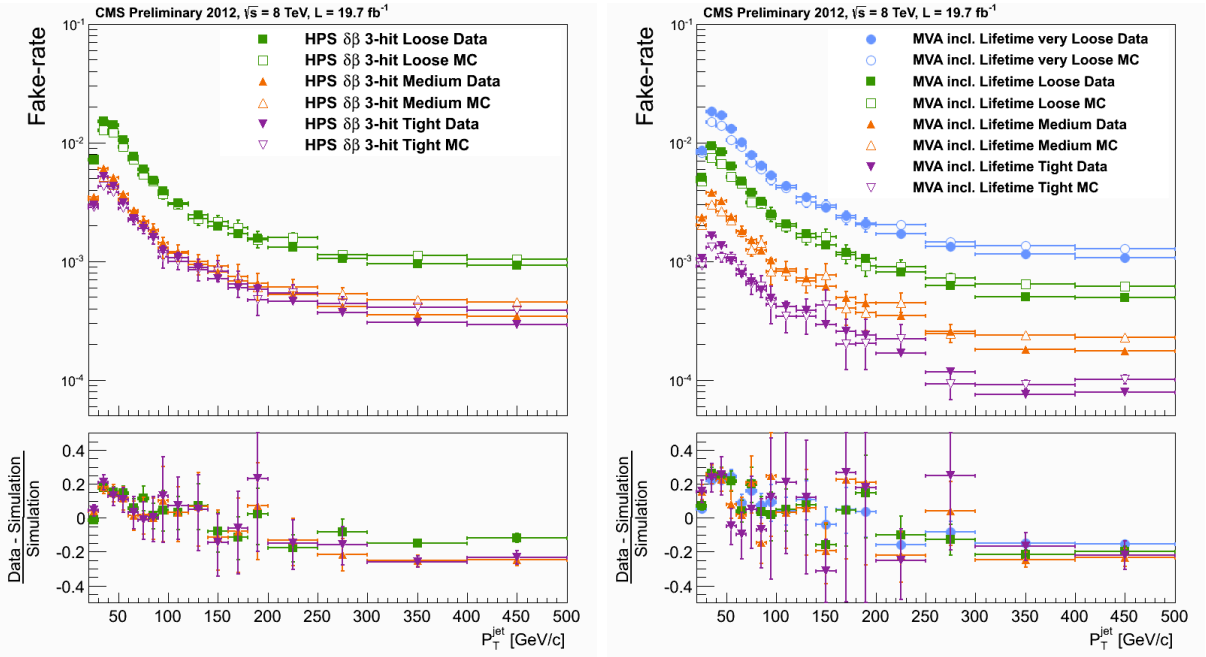


Figure 4.19: *Probabilities for quark and gluon jets to pass the cut-based isolation discriminator (left) and the MVA-based isolation discriminator (right), as function of jet p_T. The fake-rates measured in data are compared to the Monte Carlo expectation. From [140].*

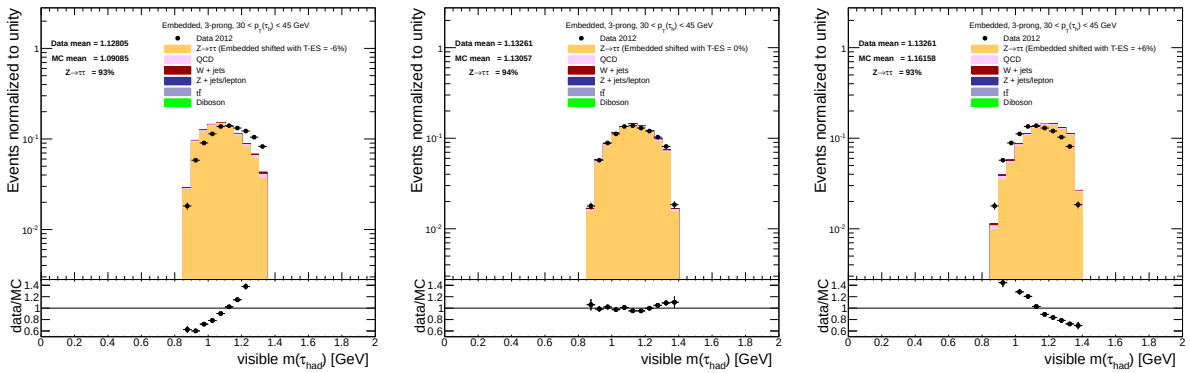


Figure 4.20: *Comparison between data and $Z/\gamma^* \to \tau\tau$ Embedded sample plus backgrounds in events containing τ_h candidates reconstructed in the* Three Hadrons *decay mode and $30 < p_T < 45$ GeV. The $m(\tau_h)$ shape templates for the $Z/\gamma^* \to \tau\tau$ Embedded sample are shown for τ-ES variations of -6% (left), 0% (center) and +6% (right).*

distributions of backgrounds:

$$q_X = -2 \ln \frac{\mathcal{L}(\tau\text{-ES=X\%})}{\mathcal{L}(\tau\text{-ES=0\%})} \tag{4.7}$$

The fit is performed taking the τ-ES as the parameter of interest and the systematic

uncertainties other than the τ-ES are accounted for by nuisance parameters. The best-fit value is the one minimizing the q_X distribution and the uncertainty interval ($\pm 1\sigma$) is taken as the τ-ES values in which q_X exceeds the minimum by $\Delta q_X = +1$.

The τ-ES correction is measured for each decay mode separately and the measurement is performed in three bins of transverse momentum $p_T(\tau_h) \in [20-30; 30-45; > 45$ GeV]. The result is shown in figure 4.21 for the two observables.

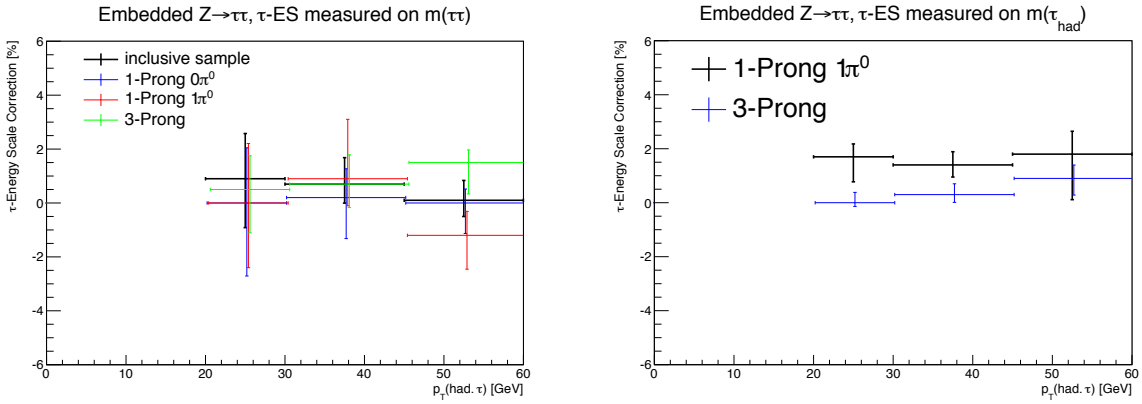


Figure 4.21: *τ energy scale measured using the $m_{\mu\tau_h}$ observable (left) and the m_{τ_h} observable (right). The τ-ES shows no dependence on the τ_h candidate transverse momenta. From [140].*

The curves show no indication of dependence of the τ-ES on the decay mode or transverse momenta. The correction is set globally to +1% to τ_h candidates for all decay modes and all p_T. The assigned uncertainty to the measured value is estimated to be of $\pm 3\%$.

The τ_h energy scale is an important shape systematic for many physics analyses with tau leptons in the final state. In particular for the Standard Model and MSSM $H \to \tau\tau$ searches, the τ-ES has a significant effect on the potential to discover a signal in the tau pair mass distribution close to the dominant irreducible background, Drell-Yan $Z/\gamma^* \to \tau\tau$ production.

4.7 Muon discriminators

In some rare cases a muon can be mis-reconstructed as a τ_h. Usually muons are detected by the muon chambers in the outer part of CMS, while τ_h decay products are stopped at the calorimeters. But cracks in the muon system and detector noise can cause a mis-reconstruction in some fraction of the muons, which adding the pile-up conditions can be reconstructed as one prong τ_h. The muon mis-identification is on the level of the permil for a τ_h identification efficiency of about 99%. Two kinds of muon discriminators have been developed, a cut-based and an MVA-based anti-muon discriminators are plugged to the τ_h reconstruction chain.

Cut-based muon discriminator

The cut-based muon discriminator vetoes the τ_h candidates in case of some muon detectors are fired around the τ_h candidate direction. Two working points are defined as function of the quality of the reconstructed muon depending on the energy deposits in the muon chambers surrounding the τ_h candidate direction.

MVA-based muon discriminator

A Boosted Decision Trees discriminator is trained to separate muons from hadronically decaying taus. The multivariate inputs are based on calorimetric information and muon system reconstructed hits and segments. Loose, Medium and Tight working points are set for different cuts on the MVA output.

$\mu \rightarrow \tau_{\mathbf{h}}$ fake rate

The anti-lepton discriminators efficiency is determined with respect to the number of τ_h candidates matched to a generated τ_h lepton ($\Delta R < 0.3$) passing the offline selection $p_T > 20$ GeV, $|\eta| < 2.3$, passing the decay mode finding from the HPS algorithm and the Loose working point of the cut-based isolation:

$$\epsilon_{anti-lep} = \frac{anti - lep}{P_T^{rec} > 20 \text{ GeV } \& \ |\eta_{rec} < 2.3| \ \& \ HPS_{decay\ mode} \ \& \ LooseIso_{Cut}} \tag{4.8}$$

As well we can define the lepton-to-tau fake rate as the rate of muons with $p_T > 20$ GeV and $|\eta < 2.3|$ passing the τ_h reconstruction chain:

$$f_r^{anti-lep} = \frac{P_T^{\tau} > 20 \text{ GeV } \& \ |\eta_{\tau} < 2.3| \ \& \ HPS_{decay\ mode} \ \& \ LooseIso_{Cut} \ \& \ anti - lep}{P_T^{lep} > 20 \text{ GeV } \& \ |\eta_{lep} < 2.3|} \tag{4.9}$$

The expected performance of the MVA-based anti-muon discriminators are displayed in figure 4.22 as a function of the τ_h candidate transverse momentum. The discriminators are very efficient ($\epsilon_{anti-mu} > 90\%$) and independent of the p_T. The fake-rate is higher at low moment due to the contamination of quark/gluon jet from the QCD multi-jets activity, ranging between 10^{-4} to the permil level.

Data to Monte Carlo simulation scale factors are derived using the Tag-and-probe technique in $Z/\gamma^* \rightarrow \mu\mu$ events as reported in ref. [140]. The scale factors are compatible with unity within uncertainties. Due to the lack of statistics this measurement suffers from big uncertainties that lie typically around 30%.

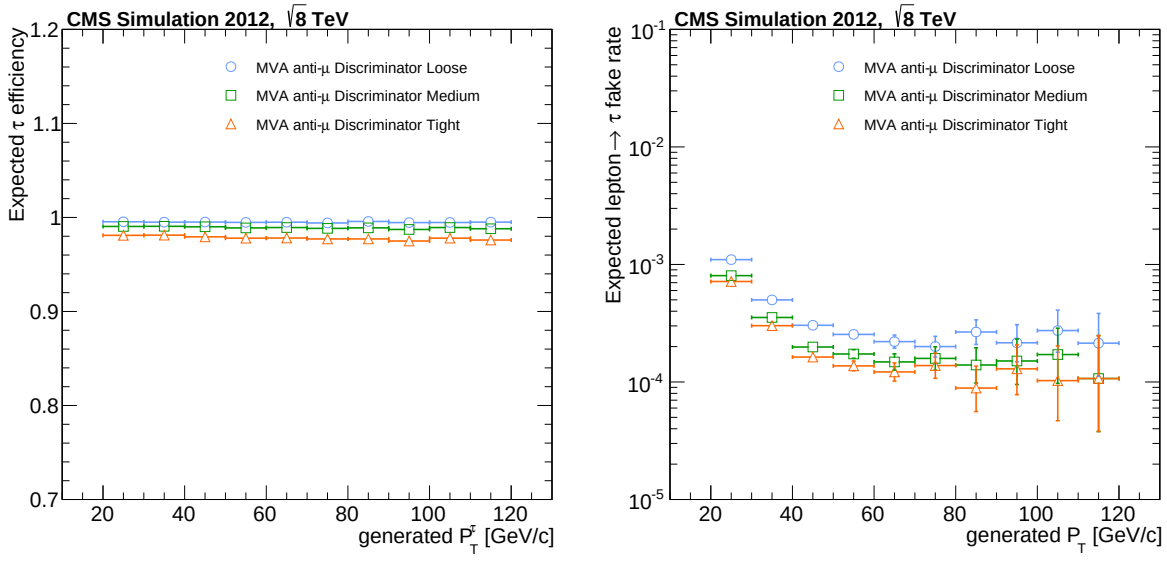


Figure 4.22: *τ_h anti-muon efficiency (left) and $\mu \to \tau_h$ fake-rate (right) as a function of the τ_h candidate transverse momentum in simulated $Z/\gamma^* \to \tau\tau$ and $Z/\gamma^* \to \mu\mu$ events. From [140].*

Chapter 5

Anti-electron discriminators

Chapter content

Electrons can be mis-identified as hadronically decaying taus, and represent an important source of background to many analyses with τ-leptons in the final state. Electrons radiate brehmsstrahlung photons, which in turn can convert in electron/positron pairs while crossing the tracker volume.

As described in section 3.3.4, the electron reconstruction algorithm is designed to recover the bremsstrahlung photons deposits in the ECAL via the track tangent extrapolation. However, the photon recovery fails and the electron becomes mis-identified. In this case, extra particles will be reconstructed and clustered around the electron direction leading to a reconstructed Particle Flow jet. Additional particles from the underlying event shall also be clustered in the jet. The reconstructed jet will then be narrow and have low multiplicity, mimicking a one-prong τ_h candidate. In figure 5.1 a schematic view of an electron being mis-reconstructed as a τ_h is shown.

More precisely, isolated electrons with high transverse momentum, like those coming from the decay of the gauge bosons, are likely to pass the identification and isolation requirements, and then can constitute a non-negligible background in the analysis with hadronic taus and electrons in the final state. Electrons can then be reconstructed as a τ_h candidate in the *Hadron plus one strip* decay mode. An anti-electron discriminator is then necessary to remove the $e \to \tau_h$ fakes.

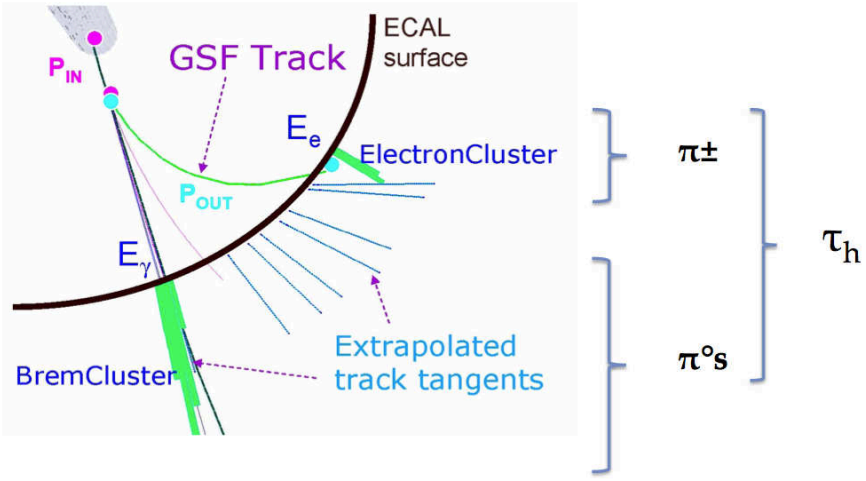


Figure 5.1: *Electron mis-reconstructed as a 1-prong τ_h.*

The electrons originating from typical Standard Model processes (as massive boson production) are boosted and then can more easily pass the τ_h reconstruction and isolation criteria. The benchmark analysis, where a performant anti-electron discriminator is needed, is the Higgs search in the di-τ final state where $H \to \tau\tau \to e\tau_h$. In this channel Drell Yan $Z/\gamma^* \to ee$ is a considerable background with larger cross-section and higher acceptance than the signal. We will show the impact of different anti-electron discriminators scenarios in the analysis section 7.5.1.

Different strategies have been implemented in CMS to reject electrons faking τ_h. In this chapter, I describe my personal contribution on the development of a multivariate

electron rejection discriminator. First I present the cut-based approach anti-electron discriminator in section 5.1. Then I describe some generalities about the multivariate methods, section 4.4. I then describe a first version of a multivariate anti-electron discriminator in section 5.2. An improved version of the discriminator resulting from my contributions is presented in section 5.3. Finally the expected performance is compared in section 5.4 and $e \to \tau_h$ fake rate measurement is presented in section 5.5. The improved discriminator was part of the latest analysis of the Higgs boson search in the tau pairs final state [161, 162] and is now part of the official tau reconstruction in CMS [140].

5.1 Cut-based electron discriminator

The cut-based[1] anti-electron discriminator [140] uses the output of the multivariate classifier used for electron identification in the Particle Flow algorithm (see section 3.3.4). This algorithm is trained to separate electrons from charged pions and then can be used to reject electrons by inverting the cut on the MVA score.

Three working points are defined for different τ_h identification efficiencies and $e \to \tau_h$ fake rates:

- Loose working point: the τ_h candidate passes the discriminator if the output of the Particle Flow electron MVA is less than 0.6.
- Medium working point: the τ_h candidate passes the discriminator if the output of the Particle Flow electron MVA is less than -0.1.
- Tight working point: the τ_h candidate passes the discriminator if it passes the Medium working point and if it is reconstructed outside the "crack" regions in the ECAL barrel modules ($|\eta| < 0.018$, $0.423 < |\eta| < 0.461$, $0.770 < |\eta| < 0.806$, $1.127 < |\eta| < 1.163$) and outside the transition between the barrel and end-cap transition region ($1.460 < |\eta| < 1.558$). The Particle Flow electron MVA does not perform well in these "crack" regions due to the material budget in these zones leading to a distortion or missing shower shape variables.

The anti-electron efficiency and fake-rate can be estimated using the equations 4.8 and 4.9. The expected performance of the cut-based anti-electron discriminators are displayed in figure 5.2 as a function of the τ_h candidate η.

Due to the vetoes of ECAL regions, the loss in efficiency on the Tight working point is estimated to be around 4-5%. Therefore, it is suited for analyses where the contamination from real electrons is important. Also, the cut-based discriminator is not specifically tuned to separate electrons from τ_h as it uses the BDT output of a MVA trained to separate electrons from charged pions. We will see in the next section how the quality of the electron rejection can be improved by using a dedicated multivariate analysis. The cut-based anti-electron discriminator provides an electron rejection leading typically to a fake-rate of the order of 10^{-2} for an efficiency around $80 - 90\%$.

[1] "Cut-based" is a misnomer as the "cut" is done on the output of an MVA.

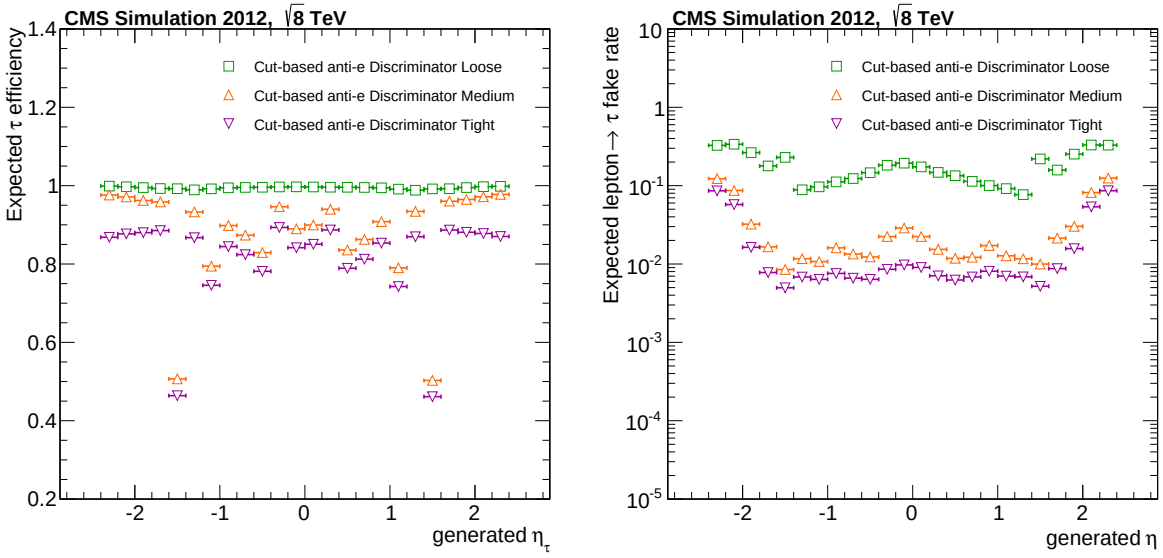


Figure 5.2: *τ_h anti-electron efficiency (left) and $e \to \tau_h$ fake-rate (right) as a function of the τ_h candidate transverse momentum in simulated $Z/\gamma^* \to \tau\tau$ and $Z/\gamma^* \to ee$ events. From Ref. [140].*

5.2 First version of the MVA-based anti-electron

A tighter discrimination against electrons with respect to the cut-based working points in section 5.1 is achieved by using a BDT-based discriminator specially tuned to discriminate electrons from τ leptons. This version of the discriminator is reported in Ref. [163] and constituted a first attempt of improvement before my contributions.

This training provides a Tight working point with better $e \to \tau_h$ fakes rejection than the cut-based one. The τ_h candidates seeding the MVA training are required to pass the Tight cut-based Working point. Only the τ_h reconstructed in the *Single Hadron* and *Single Hadron plus Strips* decay modes are considered, the electrons having a very low probability to be reconstructed in the *Three Hadrons* decay mode.

The τ_h candidates are split into three exclusive categories depending on their reconstructed decay mode:

1. One charged hadron without strips.
2. One charged hadron plus strips and the "leading" charged particle of the τ_h candidate is associated to a track reconstructed by the Gaussian-sum-filter (GSF) algorithm [164].
3. One charged hadron plus strips and the "leading" charged particle of the τ_h candidate is not associated to any track reconstructed by the Gaussian-sum-filter (GSF) algorithm.

The three categories are split into two sub-categories, depending on the pseudo-rapidity of the reconstructed τ_h candidate: one for the barrel region ($|\eta| < 1.5$) and

one for the end-cap region ($|\eta| > 1.5$). For each category, a multivariate analysis based on Boosted Decision Trees are trained using Drell-Yan $Z/\gamma^* \to \tau\tau$ and $Z/\gamma^* \to ee$ simulated Monte Carlo samples in order to discriminate real τ leptons (signal) from electrons (background).

5.2.1 Input variables

The input variables for each category with the highest discriminating power are chosen to perform the training. For the first category only two variables are used:

- E/P , defined as the ratio between the ECAL energy associated to the "leading" charged particle of the τ_h candidate and the momentum of its track.

- H/P, defined as the ratio between the HCAL energy associated to the "leading" charged particle of the τ_h candidate to the momentum of its track.

The second and third categories use the information from the shape of the photons energy deposits in the calorimeter around the direction of the charged particles: the electrons typically leave a broad pattern due to the bremsstrahlung photons along the azimuthal direction, while the τ lepton photons are usually more symmetric around the τ_h direction. Strip variables are then accessible and used in this category:

- $\sqrt{P_T^\gamma \cdot (\Delta\eta)^2}$ and $\sqrt{P_T^\gamma \cdot (\Delta\phi)^2}$, the P_T-weighted RMS of distances in η and ϕ between all photons included in any Strip to the "leading" charged particle, respectively.

- Fraction of τ_h energy carried by photons.

The second and third categories are trained separately due to one more variable that is accessible only in the second category related to the presence of the 'leading" charged particle GSF track:

- the PF electron MVA output for the "leading" charged particle used to define the cut-based working points.

5.2.2 Anti-electron discriminator working point

All the τ_h candidates being reconstructed in the three-hadrons decay mode pass this MVA tight anti-electron discriminator. The three BDT are trained in samples with different signal purities, hence the cuts applied in each of them are chosen using the signal efficiency as figure of merit. The working point was chosen to have similar signal efficiency as the Tight working point of the cut-based discriminator. The MVA anti-electron discriminator reduces the $e \to \tau_h$ fake rate of about 50% with respect to the cut-based discriminator.

5.2.3 Limitations

The MVA strategy leads to a considerable reduction of the mis-reconstructed τ_h from real electrons. However, this version of the anti-electron discriminator presents some limitations.

The training is not optimal and done in a very analysis oriented way, using simulated τ leptons from $Z/\gamma^* \to \tau\tau$ for the signal events and electrons from $Z/\gamma^* \to ee$ for the background. This is specifically tuned to reject the $Z/\gamma^* \to ee$ background contamination in the Higgs boson search specially in the $H \to \tau\tau \to e\tau_h$ channel, as we require an electron in the final state. Moreover, a single working point is provided, tuned to a high discrimination power also corresponding to this channel.

The categories for the training were made in a simple way, depending on the reconstructed particles in the τ_h candidate. As we will see in the next section, a more complex categorization can add more input variables with high discrimination power, enhancing the performance of the final MVA.

Another important limitation is the fact that the MVA is trained on top of the Tight cut-based discriminator. This reduces the phase space covered by the training and limits the usage of the discriminator to only analyses needing a high $e \to \tau_h$ rejection.

Finally, the Tight cut-based working point includes vetoes on the τ_h candidates which are reconstructed in the cracks on the ECAL. This corresponds to a loss in efficiency of $\sim 4\%$.

My specific work consisted in optimizing the MVA-based discriminator in order to circumvent the previous limitations. I contributed to the development of a new optimized version of the MVA anti-electron discriminator that I detail in the next section.

5.3 Improved version of the MVA-based anti-electron

The improved version of the anti-electron discriminator is detailed in Ref. [165]. This new MVA based anti-electron discriminator is trained by using the maximal amount of information thanks to a refined categorization of the τ_h candidates. In order to recover the τ_h reconstructed in the ECAL cracks, vetoed in the previous versions, the new anti-electron suppresses this veto. Additional input variables are then used to discriminate electrons in these critical regions.

As in the previous discriminator, all τ_h candidates reconstructed in the *Three Hadrons* decay mode pass the MVA-based anti-electron discriminator.

5.3.1 Categories of τ_h

The τ_h candidate categorization is based on the reconstructed particles around the τ_h and on the regions of the detector where it is reconstructed (barrel or end-cap). In addition with respect to the previous discriminator new categories including the cases where a GSF electron is matched with the τ_h candidate are taken into account.

The reconstructed τ_h candidates are classified into 16 categories, depending on:

- Whether the "leading" charged particle of the τ_h candidate is associated to a track reconstructed by the GSF algorithm.

- Whether an electron is reconstructed within a cone of size $\Delta R = 0.3$ around the τ_h direction.

- Whether the τ_h is reconstructed in either the *Single hadron* decay mode or in one of the two decays modes: *Hadron plus one Strip*, *Hadron plus two Strips*.

- Whether the τ_h is reconstructed in the ECAL barrel ($|\eta| < 1.479$) or in the end-caps ($|\eta| > 1.479$).

Each of the 16 categories is uniquely associated to one BDT. The motivation for organizing the MVA-based anti-e discriminator into 16 categories is to pass as input to each BDT the maximal amount of information available for each τ_h candidate in order to obtain the lowest possible $e \to \tau_h$ fake-rate.

5.3.2 Input variables

The previous categories depend on different input variables depending on the kind of reconstructed τ_h. The input variables are then classified in several sets:

1. A generic set of τ_h variables, called "generic variables" and common to all categories.

2. A set of "strip variables" related to τ_h reconstructed in the *Hadron plus one Strip* and *Hadron plus two Strips*.

3. A set of "GSF track variables" valid when the leading charged particle of the τ_h is associated to a GSF track.

4. A set of "electron variables" valid when the τ_h candidate is matched to an electron within $\Delta R < 0.3$ around the τ_h direction.

The variables of the four sets are:

Generic variables: As explained above, 4% loss of efficiency was due to the veto on ECAL crack regions. In this version of the discriminator, only the crack region between the ECAL barrel and end-cap is kept, and two additional variables are added to the BDT training: the distance in η and in ϕ to the other crack regions. The set of τ_h generic variables is then:

- P_T and η of the τ_h candidate.

- $E/E+H$, the electromagnetic energy fraction, defined as the ratio of ECAL to the sum of ECAL plus HCAL energy associated to the charged particles and photons that constitute the τ_h candidate.

- E/P and H/P, defined as ratios of ECAL and HCAL energy associated to the "leading" charged particle of the τ_h candidate to the momentum of its track.
- Mass of the τ_h.
- Distance in η between τ_h direction and nearest ECAL crack.
- Distance in ϕ between τ_h direction and nearest ECAL crack (computed for τ_h candidates in the ECAL barrel only).

Strip variables:

- $\sqrt{P_T^\gamma \cdot (\Delta\eta)^2}$ and $\sqrt{P_T^\gamma \cdot (\Delta\phi)^2}$, the P_T–weighted RMS of distances in η and ϕ between all photons included in any Strip to the "leading" charged particle, respectively.
- Fraction of τ_h energy carried by photons.

GSF track variables:

- PF electron MVA output for the "leading" charged particle.
- Normalized χ^2 of the GSF track.
- $(N_{hits}^{GSF} - N_{hits}^{KF})/(N_{hits}^{GSF} + N_{hits}^{KF})$, with N_{hits}^{GSF} (N_{hits}^{KF}) representing the numbers of hits in silicon pixel plus strip tracking detector associated to the track reconstructed by the GSF (Kalman filter) track reconstruction algorithm.
- $\ln(P_T)$ of the GSF track.
- η of the GSF track.

Electron variables:

- $(E_e + \sum E_\gamma)/P_{in}$, the ratio between the total ECAL energy and the inner track momentum.
- $\sum E_\gamma/(P_{in} - P_{out})$, the ratio between the Bremsstrahlung photon energy as measured by the ECAL and by the track.
- $F_{brem} = (P_{in} - P_{out})/P_{in}$, the ratio between the difference between the GSF track momentum at the outermost state and the innermost state and the innermost state track momentum.
- N_{hits}^{GSF}, the numbers of hits in silicon pixel plus strip tracking detector.
- Normalized χ^2 of the electron track.
- $\ln(P_T)$ of the electron track.

- η of the electron track.
- σ_{P_T}/P_T, the resolution of the electron track.

The kinematical p_T and η input variables parametrize the other input variables in order to take possible dependency on the kinematical regimes into account. The events used for the training are reweighted as to ensure that the p_T and η distributions are identical for "signal" and "background". This aims at preventing the MVA to depend on the event kinematics. The discriminatory power of the input variables is illustrated in Fig. 5.3, 5.4, 5.5 and 5.6.

An example of linear correlation coefficients between the input variables for "signal" and "background" is shown in appendix B.

5.3.3 BDT Training

The Boosted Decision Trees are trained on Monte Carlo samples of simulated events: $Z/\gamma^* \to \tau\tau$, $Z/\gamma^* \to ee$, $W \to \tau\nu$, $W \to e\nu$, $t\bar{t}$, Higgs $\to \tau\tau$, $Z' \to \tau\tau$, $Z' \to ee$, $W' \to \tau\nu$ and $W' \to e\nu$. Reconstructed τ_h candidates are considered as "signal" ("background") in case they are matched to a hadronic tau decay (electron) at generator level within $\Delta R < 0.3$. The MVA training is customized by choosing the value of the following parameters:

- N_{Trees}, the total number of trees in the forest.
- BoostType, boosting type for the trees in the forest[2].
- Shrinkage, the learning rate of the BDTs.
- GradBaggingFraction, the fraction of events to be used at each iteration.
- Separation Type, the separation criterion for node splitting.
- nCuts, the number of steps for the node cuts optimization.
- PruneMethod, the method used for pruning (removal) of statistically insignificant branches.
- NNodesMax, the maximal number of nodes in the trees.

The training was done using TMVA with the training parameters presented in table 5.1. Half of the samples are used to train the MVA, the other half is used to evaluate the performance and perform overtraining checks.

An example of individual BDT output, build using equation 4.2, for one of the 16 categories is presented in appendix B. The combined BDT output of the 16 categories is shown in figure 5.7 for signal and background, a good separation is achieved.

[2]for details about Gradient boosting please refer to Ref. [155].

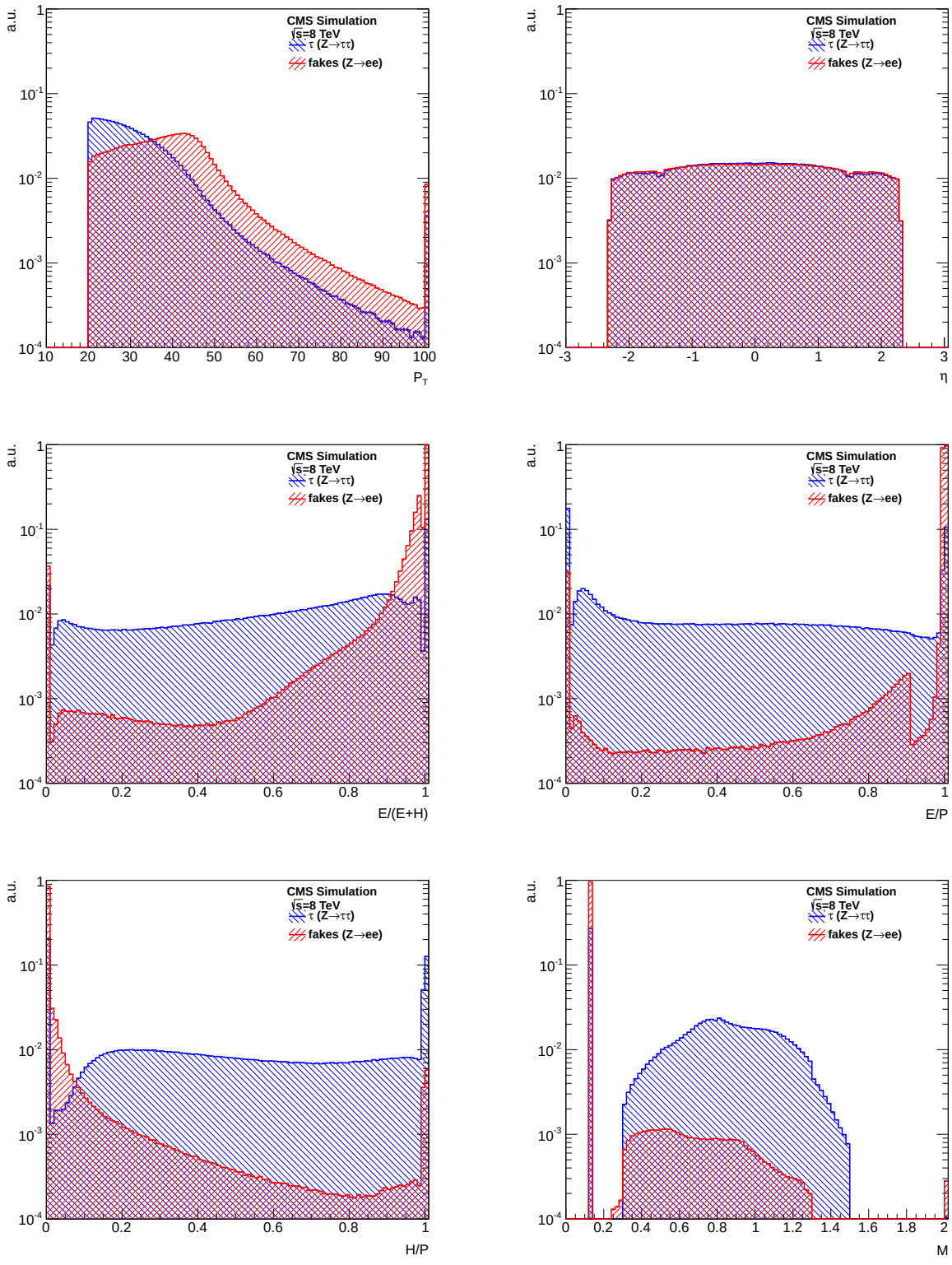


Figure 5.3: *Input variables distributions described in the text, normalized to unity, for the MVA based anti–e discriminator for simulated signal and background in $Z/\gamma^* \to \tau\tau$ (blue) and $Z/\gamma^* \to ee$ (red) events respectively.*

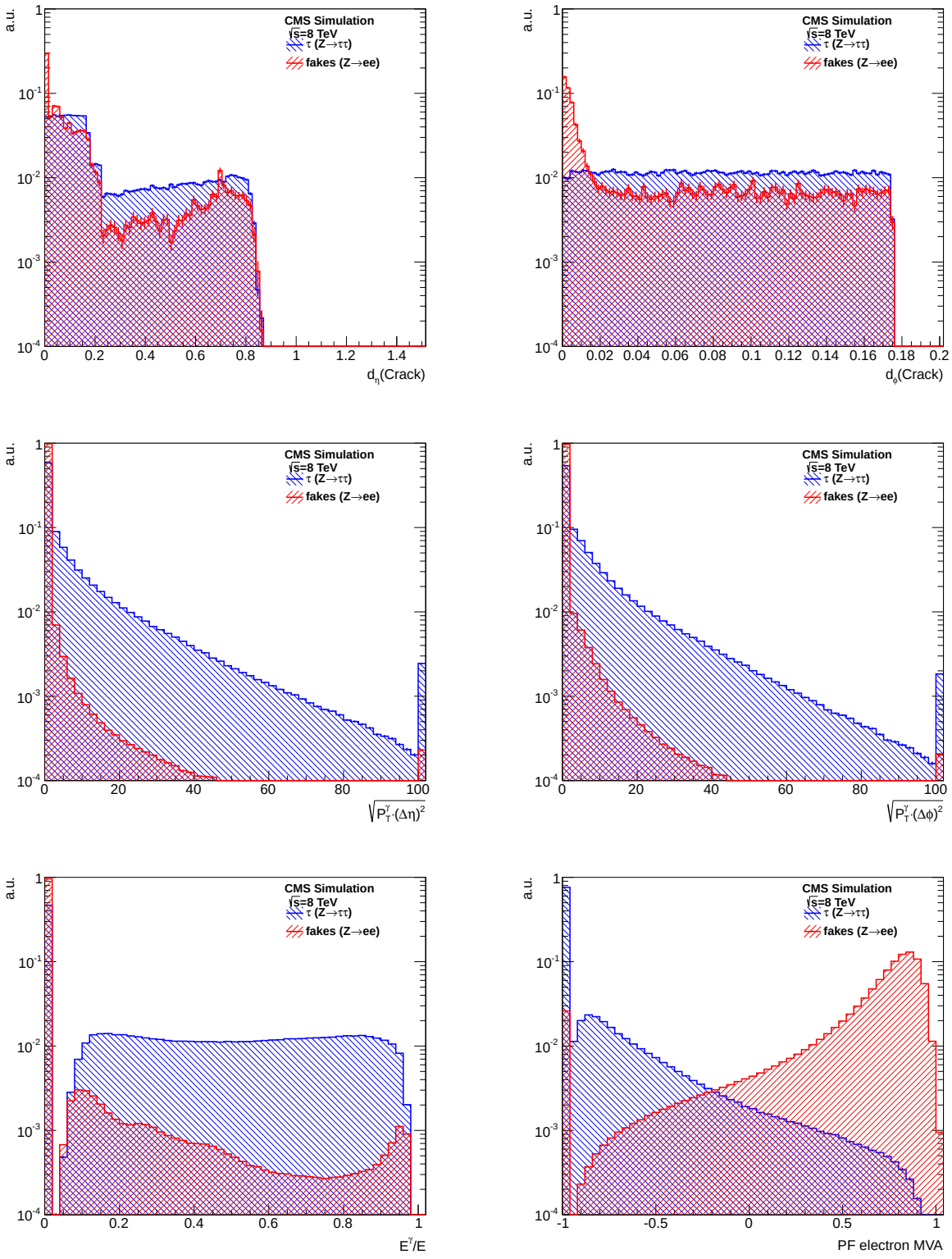


Figure 5.4: *Input variables distributions described in the text, normalized to unity, for the MVA based anti–e discriminator for simulated signal and background $Z/\gamma^* \to \tau\tau$ (blue) and $Z/\gamma^* \to ee$ (red) events respectively.*

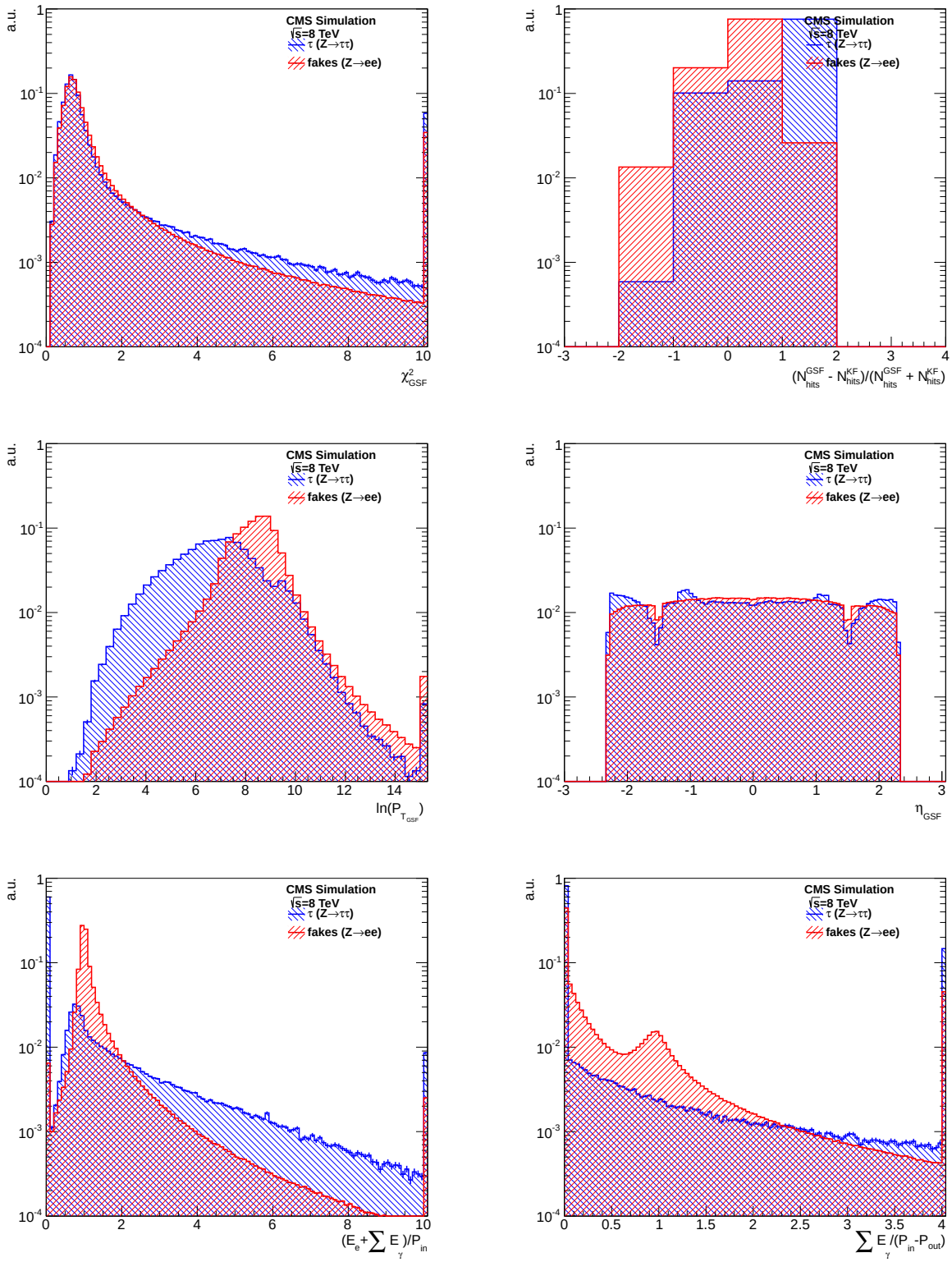


Figure 5.5: *Input variables distributions described in the text, normalized to unity, for the MVA based anti–e discriminator for simulated signal and background $Z/\gamma^* \to \tau\tau$ (blue) and $Z/\gamma^* \to ee$ (red) events respectively.*

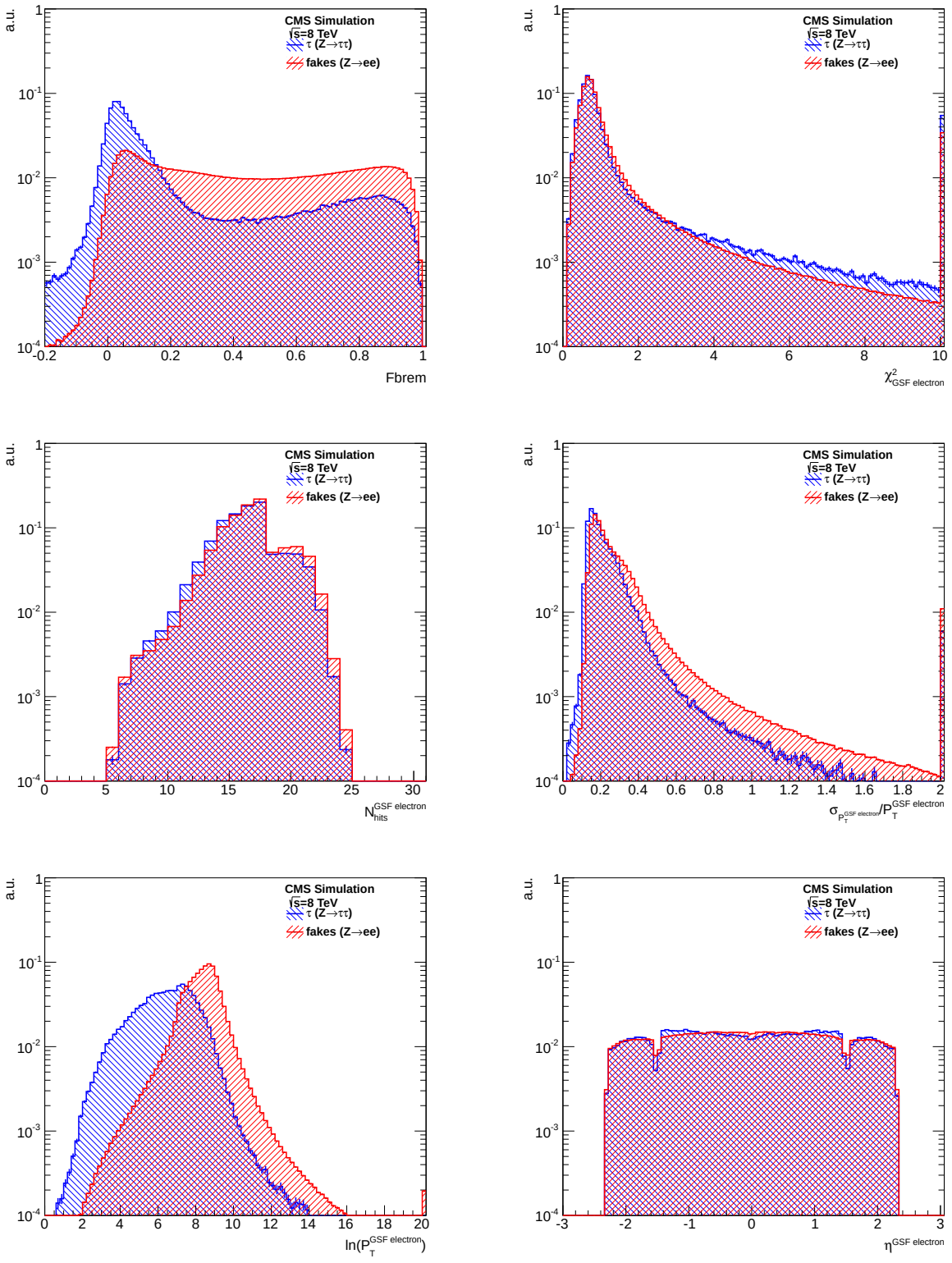


Figure 5.6: *Input variables distributions described in the text, normalized to unity, for the MVA based anti–e discriminator for simulated signal and background $Z/\gamma^* \to \tau\tau$ (blue) and $Z/\gamma^* \to ee$ (red) events respectively.*

Option	Value
TMVA::Types::kBDT	BDTG
NTrees	600
BoostType	Grad
Shrinkage	0.30
UseBaggedGrad	enabled
GradBaggingFraction	0.6
Separation Type	GiniIndex
nCuts	20
PruneMethod	CostComplexity
PruneStrength	50
NNodesMax	5

Table 5.1: *List of TMVA option used to train the 16 BDTs. See [155] for a description of the different options.*

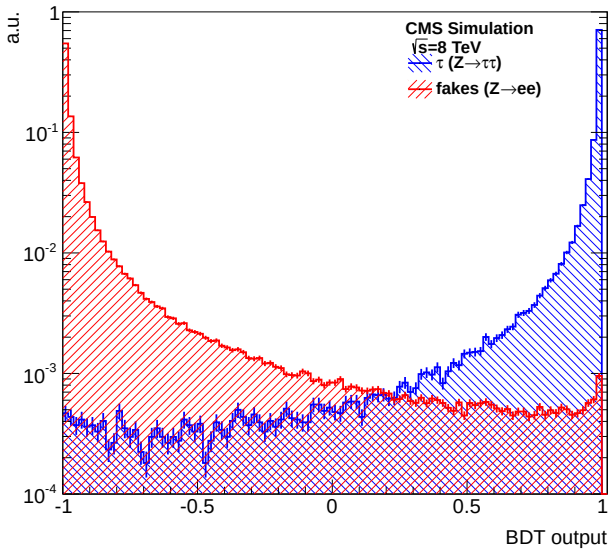


Figure 5.7: *BDT output of the anti-electron discriminator combined for all the categories.*

5.3.4 BDT Testing

The training is validated by comparing data to simulation agreement in the input variables and by comparing the expected ROC curves for the training dataset and the testing dataset. The comparison between the background model and the data using an inclusive

di-τ selection (see section 7.1) relaxing the anti-electron requirement, is shown in figure 5.8. A good agreement in the region of interest (BDT output$>$ 0.9) is observed. This shows that the MVA performance is not biased.

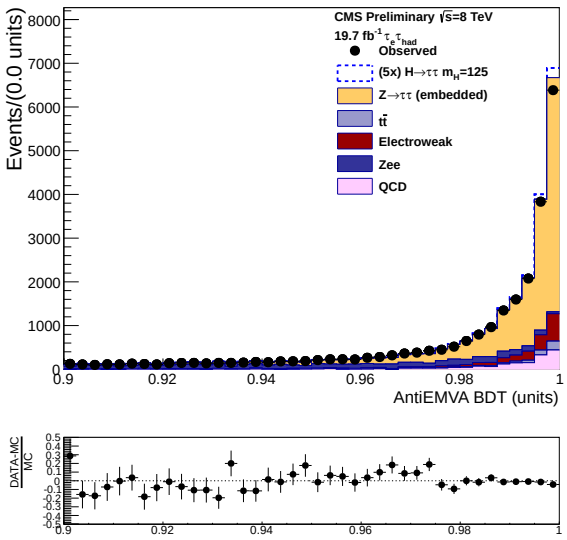


Figure 5.8: *BDT output of the anti-electron discriminator.*

5.3.5 Working points optimization

Loose, Medium, Tight and Very-Tight working points (WP) are defined, corresponding to different τ_h identification efficiencies and $e \to \tau_h$ fake rates. Each WP is tuned to target an efficiency between 75% and 90% in steps of 5%. They are defined by a set of 16 thresholds, representing cuts on the outputs of the BDTs that have been trained for each category. The thresholds have been optimized to yield the lowest possible $e \to \tau_h$ fake-rate for given τ_h identification efficiency. The optimization is performed using an iterative procedure. The procedure is started by setting each of the 16 thresholds to 1.0 (corresponding to zero efficiency and zero fake rate). These 16 thresholds are represented by a 16-dimensional vector. Each step of the iteration proceeds by varying one component of the vector by 0.001 units at a time and determining the change in signal efficiency, ΔS, and fake-rate, ΔB, for that variation. The component with the maximum ratio $\Delta S/\Delta B$ is taken to be the best performing category of the current iteration step. The threshold corresponding to this category is lowered by 0.001 units. The iterative procedure continues until a given target τ_h identification efficiency is reached.

The 16-dimensional cut vectors defining the Loose, Medium, Tight and Very-tight WP are given in table 5.2. In order for a τ_h candidate reconstructed in category i to pass a certain WP the output of the i-th BDT is required to exceed the i-th component of the corresponding cut vector.

Category	Loose	Medium	Tight	VTight
0	0.835	0.937	0.974	0.986
1	0.831	0.949	0.976	0.986
2	0.849	0.955	0.978	0.986
3	0.859	0.956	0.978	0.990
4	0.873	0.962	0.971	0.983
5	0.823	0.934	0.969	0.977
6	0.850	0.946	0.982	0.992
7	0.855	0.948	0.972	0.981
8	0.816	0.959	0.982	0.989
9	0.861	0.950	0.977	0.989
10	0.862	0.954	0.981	0.987
11	0.847	0.954	0.978	0.987
12	0.893	0.897	0.897	0.976
13	0.820	0.951	0.976	0.991
14	0.845	0.948	0.975	0.984
15	0.851	0.953	0.977	0.986

Table 5.2: *Loose, Medium, Tight and Very-tight working points cuts on the 16 BDT scores.*

In addition to passing a cut on the BDT output, τ_h candidates reconstructed in the tau decay modes *Single hadron*, *Hadron plus one Strip* or *Hadron plus two Strips* are required not to be within the crack between ECAL barrel and end-cap, defined by $1.460 < |\eta| < 1.558$.

5.4 Performance of the new MVA-based anti-electron discriminator

In physics analyses with τ_h in the final state and requiring and requiring a performant anti-electron discriminator, a veto is usually applied on a second electron in the event. Indeed, the collection of τ_h candidates are "cleaned" with respect to loosely identified electrons: τ_h candidates are required not to overlap with electrons passing the veto-electron identification criteria (*c.f.* section 6.3.2) within $\Delta R < 0.3$. Such a cleaning is applied in the $H \to \tau\tau$ analysis and correspond to a baseline event selection (see section 7.1). Therefore, the efficiencies and fake-rates are reported separately for the case where the cleaning is applied.

The efficiencies and $e \to \tau_h$ fake-rate are reported in table 5.3. The values obtained are the expected contribution from Monte Carlo simulated samples $Z/\gamma^* \to \tau\tau$ and $Z/\gamma^* \to ee$.

Working Point	Efficiency (%)	Fake-rate (%)
Cut-based		
Loose	99.6 (99.9)	24.8 (99.7)
Medium	87.8 (89.7)	4.73 (37.9)
Tight	80.9 (84.5)	3.06 (18.0)
Previous MVA		
Tight	80.8 (84.4)	2.02 (13.7)
This MVA		
Loose	89.6 (95.0)	0.56 (16.7)
Medium	84.3 (91.2)	0.24 (7.12)
Tight	77.2 (85.2)	0.12 (3.39)
Very Tight	72.6 (80.8)	0.08 (2.2)

Table 5.3: *Efficiency and fake rate for different working points of the anti-electron discriminators. The quoted efficiency values refer to $Z/\gamma^* \to \tau\tau$ events while fake rate refers to $Z/\gamma^* \to ee$ events both produced by Monte Carlo simulation. Numbers given in brackets represent the efficiency and fake rate obtained when τ_h candidates are "cleaned" (see text).*

The new version of the MVA-based anti-electron discriminator reduces significantly the $e \to \tau_h$ fake-rate. The performance of the new MVA anti-electron discriminator

is compared to the one of the previous MVA discriminator on figures 5.9 showing the efficiency and the fake rate as function of p_T and η. On figure 5.10 the efficiency and the fake rate as function of ϕ and N_{vtx} are presented. The $e \to \tau_h$ fake-rate is minimal for p_T values of approximately half the Z boson mass, the region in which most of the electron background is expected. The efficiency (fake-rate) decreases (increases) in the cracks between ECAL super-modules because the separation of electrons from hadronic tau decays is more difficult in these regions, as shower shape variables get distorted and energy resolution degrades near the cracks. The efficiency recovery from the previously vetoed crack regions in the ECAL barrel is also visible. The dependence of the efficiency as well as of the fake rate on pile-up, parametrized by N_{vtx}, is seen to be small as shown in figure 5.10. The fake rate is almost insensitive to pile-up for the new MVA.

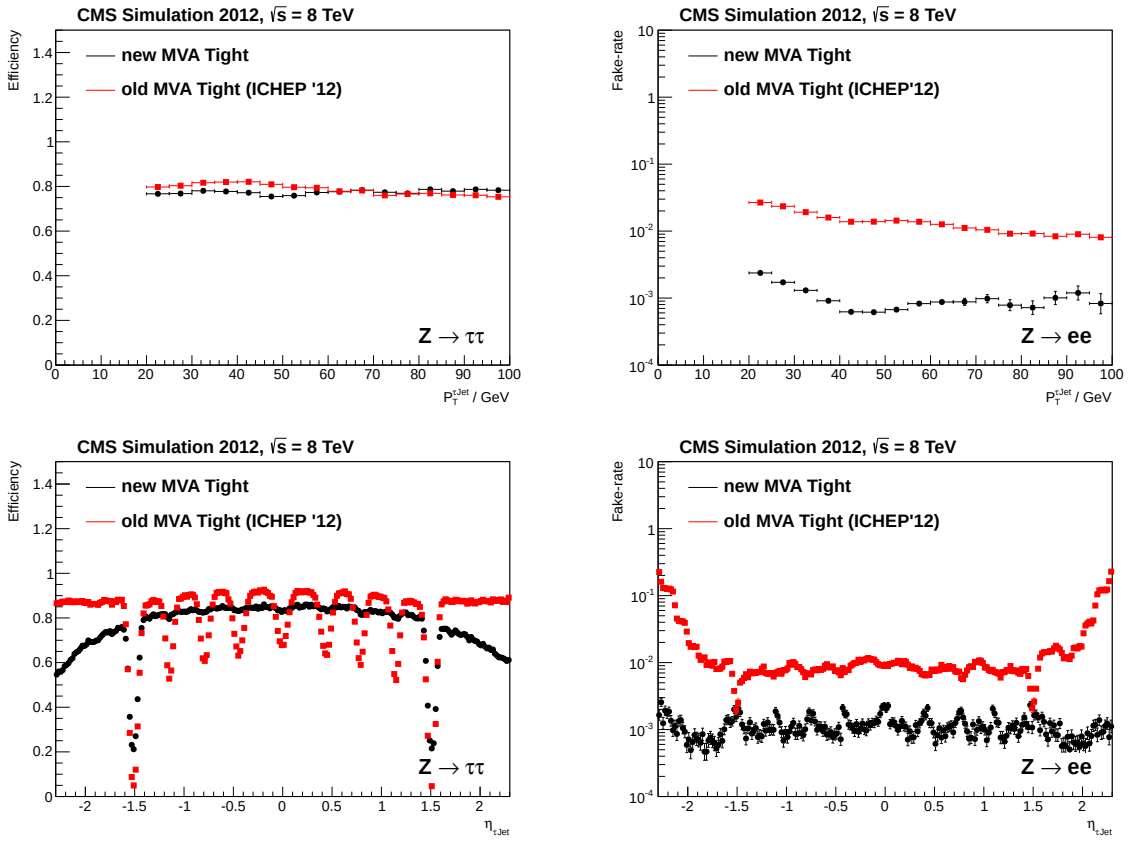


Figure 5.9: *Efficiency (left) and fake-rate (right) of the Tight WP of the anti-electron discriminators as function of $p_T(\tau_h)$ (top) and η (bottom). The hadronic tau identification efficiency ($e \to \tau_h$ fake-rate) is obtained using samples of simulated $Z/\gamma^* \to \tau\tau(Z/\gamma^* \to ee)$ events.*

For a given τ_h identification efficiency the new MVA-based anti-electron discriminator developed in this thesis achieves $e \to \tau_h$ fake-rates which are about one order of magnitude lower compared to previously used tau identification discriminators against electrons. Due to its superior performance, it has been adopted in the most recent CMS analyses [161, 162]. In particular, in the analysis presented in this thesis. The improvement brought for

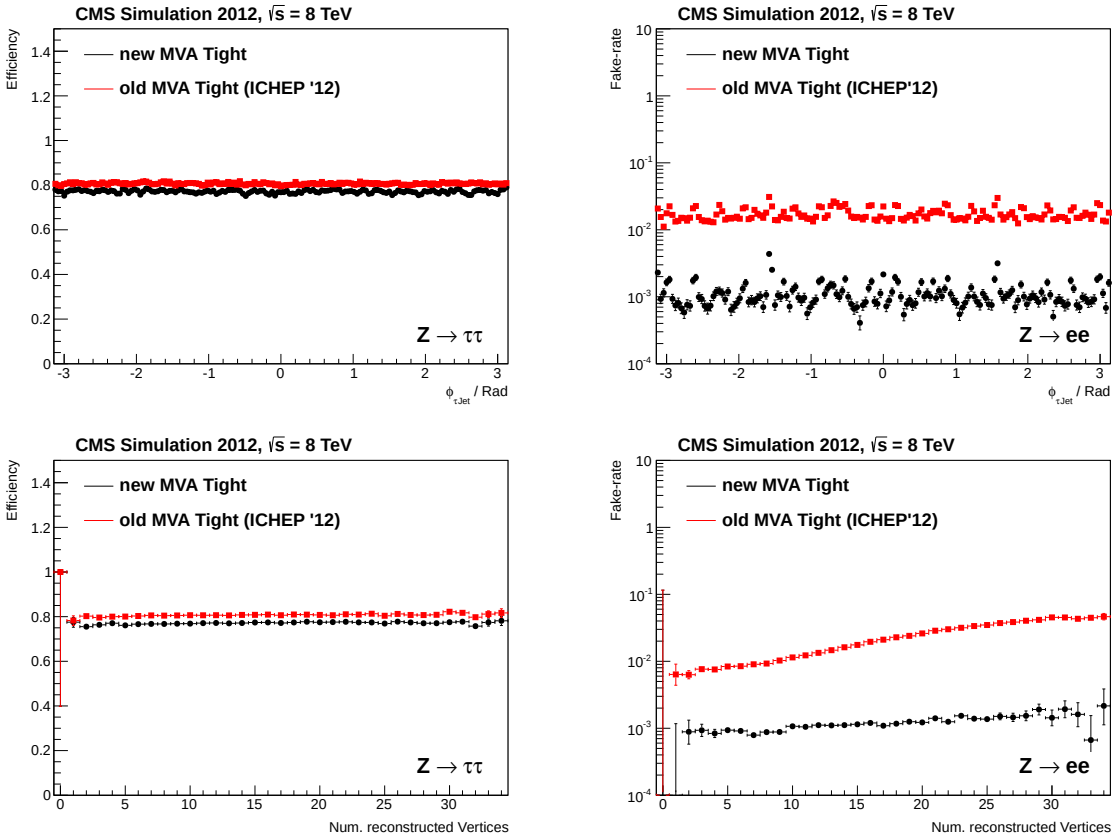


Figure 5.10: *Efficiency (left) and fake-rate (right) of the Tight WP of the anti-electron discriminators as function of $\phi(\tau_h)$ (top) and as function of reconstructed vertex multiplicity (bottom). The hadronic tau identification efficiency ($e \to \tau_h$ fake-rate) is obtained using samples of simulated $Z/\gamma^* \to \tau\tau(Z/\gamma^* \to ee)$ events. The number of reconstructed vertices is used as a measure of pile-up.*

the Higgs searches will be evaluated by comparing the expected sensitivity in chapter 7.

5.5 $e \to \tau_h$ fake-rate measurement with real data

The $e \to \tau_h$ fake-rates are measured in data using the Tag-and-Probe method on $Z/\gamma^* \to ee$ events. The measurement, in which I contributed, on the full 2012 data recorded by CMS is detailed in [140], and a short summary is given in this section.

The events are selected with a triggering electron passing the Tight cut-based identification and isolation criteria. This defines the Tag electron. The probe is required to be a reconstructed τ_h candidate passing the decay mode reconstruction (*Single Hadron, Hadron plus one Strip, Hadron plus two Strips or Three Hadrons*). The events selected are the ones containing a Tag-and-probe pair with a visible mass such that $60 < m_{vis} < 120$ GeV. Additional cuts are applied in order to reduce background contamination: a cut in the

transverse mass [3] is applied ($M_T < 25$ GeV) reducing the $W + jets$ contribution and the events are required to have $\not{E}_T < 25$ GeV reducing also the $t\bar{t}$ contamination.

As described in the previous chapter, using the equations 4.8 and 4.9 we can define the Pass and Fail regions and then define the fake-rate as:

$$f_r = \frac{N^{\tau}_{pass}}{N^{\tau}_{pass} + N^{\tau}_{fail}} \tag{5.1}$$

The measurement is done by fitting the tag-and-probe pair visible mass m_{vis} with shape templates. The $e \to \tau_h$ fake-rate is measured for the different anti-electron discriminators working points separately in the barrel and end-cap regions. Two examples of fit templates are presented in figure 5.11. The Tag-and-probe procedure can also be applied to the simulation allowing the derivation of data to Monte Carlo scale factors. The measured fake rates are displayed in table 5.4.

Working Point	Simulation (%)	Data (%)
Barrel		
Loose	0.40 ± 0.01	0.57 ± 0.02
Medium	0.13 ± 0	0.23 ± 0.01
Tight	0.06 ± 0	0.13 ± 0.01
Very Tight	0.03 ± 0	0.01 ± 0.01
End-cap		
Loose	0.40 ± 0.01	0.04 ± 0.01
Medium	0.12 ± 0.01	-
Tight	0.05 ± 0	0.03 ± 0.02
Very Tight	0.03 ± 0	0.02 ± 0.01

Table 5.4: *The $e \to \tau_h$ fake-rates measured in data and in Monte Carlo simulation, separately for electrons in the barrel and in the end-cap regions. The fit on data for the Medium WP in the end-cap region did not converge.*

For electrons in the barrel region the measured fake-rate exceeds the Monte Carlo prediction, while in the end-cap region the measured fake rate is below the Monte Carlo expectation. The difference between data and Monte Carlo simulation is higher for the Tight and Very Tight working points. The differences between data and simulation are significant in terms of the estimated uncertainties and need to be taken into account by physics analyses that are sensitive to $e \to \tau_h$ fakes. As we will see in the analysis chapter in section 6.4.10, a specific correction on $e \to \tau_h$ fakes is performed, taken into account a systematic error of 30%.

[3] $M_T = \sqrt{(p_T^{probe} + \not{E}_T)^2 - ((p_x^{probe} + \not{E}_x)^2 + (p_y^{probe} + \not{E}_y)^2)}$

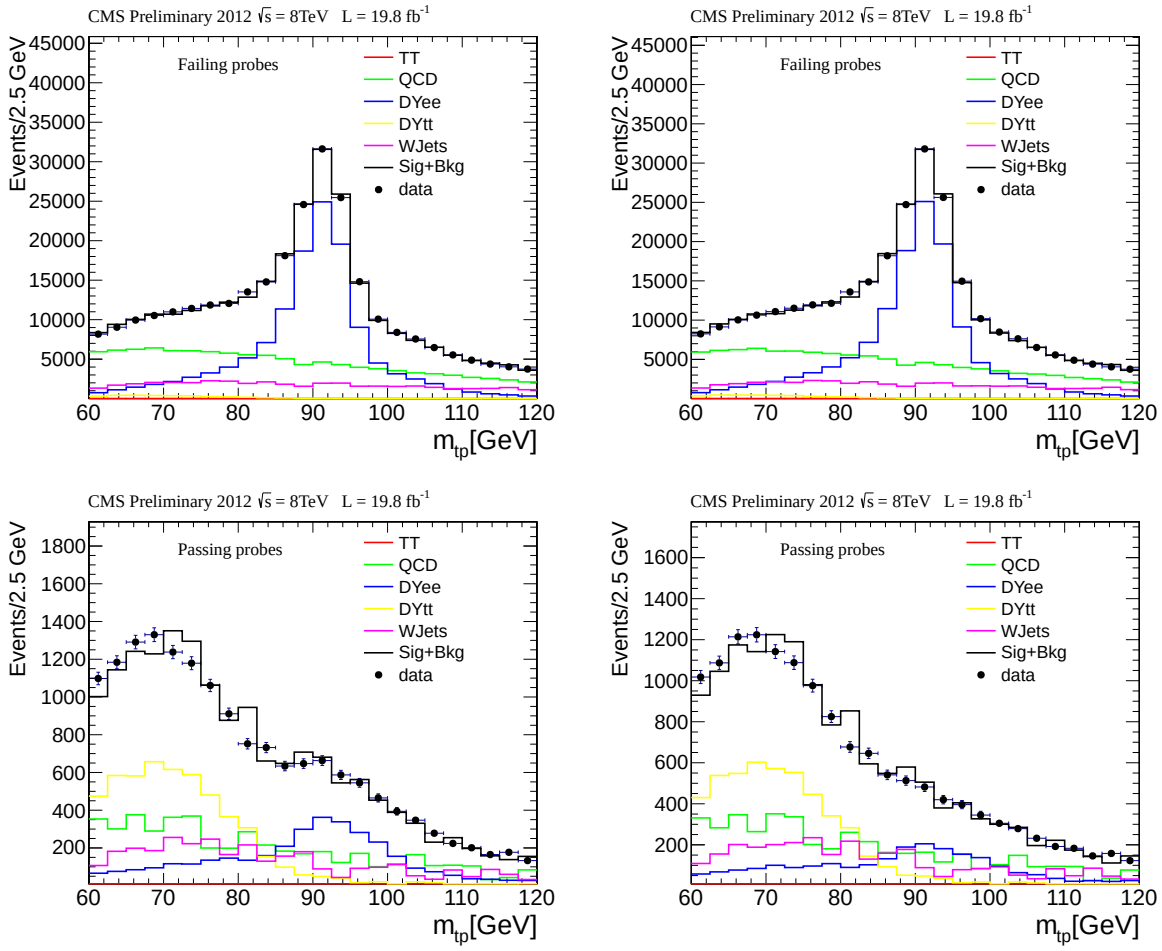


Figure 5.11: *Post-fit visible mass distribution of the tag and probe pair for the Pass (up) and Fail (bottom) regions compared to the Monte Carlo expectation, for the Medium (left) and Tight (right) WP of the new MVA-based discriminator against electrons in the barrel region. From Ref. [140].*

Part III

Standard Model and MSSM $H \to \tau\tau$ analysis in CMS

Chapter 6

$H \to \tau\tau$ analysis strategy

Chapter content

6.1 Introduction

As shown in the first chapter, the Standard Model Higgs boson coupling to fermions is derived by introducing *ad hoc* Yukawa couplings. A way to study in detail these couplings at the LHC is via the study of the fermionic decays $H \to b\bar{b}$ and $H \to \tau\tau$. And in particular the study of the leptonic decays is specially interesting. The only channel available to test this couplings within the Run 1 is the search $H \to \tau\tau$, due to the great mass of the τ-leptons, the other leptonic channels $H \to \mu\mu$ need much more integrated luminosity to be statistically significant.

The observable for both analyses is the reconstructed mass of the di-τ system estimated using a likelihood approach. An excess of events is searched by performing a simultaneous maximum likelihood fit over all the categories and channels. The statistical interpretation is done by performing a CL_S ratio in order to draw exclusion limits and statistical significance in the case of an excess.

In this chapter, I will introduce the tools and techniques common to the Standard Model and the MSSM Higgs boson searches in the channel where they decay to tau pairs in the semi-leptonic channels. The trigger, data acquisition and simulated samples are presented in section 6.2. I then describe the objects of the analysis in section 6.3. The corrections that are applied to Monte Carlo simulated events in order to improve the modeling of the data are detailed in section 6.4. In section 6.5, I describe the di-τ mass reconstruction, based on a likelihood based algorithm. And finally the statistical tools used to compute the exclusion limits and statistical significance are described in section 6.6.

6.2 Data acquisition and simulated samples

6.2.1 Trigger and data acquisition

The trigger constantly evolved during the data taking periods in order to adapt to the increase in instantaneous luminosity. The trigger system at level 1 (L1), described in 3.2.8, accepts an event if coarse identification, isolation and kinematic criteria are verified. As the reconstruction of a τ_h is very difficult at level 1, the single lepton seeds electron (EG) for the $e\tau_h$ channel or muon (Mu) for the $\mu\tau_h$ channel are used in the analysis to feed the high level trigger (HLT) algorithms.

The HLT reconstructs a particle around the L1 object, and a particle flow like reconstruction, described in 4.2 is performed around the L1 jet candidates. In table 6.1, the HLT trigger paths used in the analysis with their corresponding L1 seeds are listed. p_T thresholds above 20 GeV on the τ_h candidate, above 20 or 22 GeV on the electron and above 17 or 18 GeV on the muon were used during the data taking period. The mean output trigger rate for the triggers used in the $H \to \tau\tau$ search was around 16 Hz for the $\mu\tau_h$ trigger path and 25 Hz for the $e\tau_h$ trigger path, which represents a quite considerable part of the Higgs searches bandwidth of $\sim$ 100 Hz and of the total CMS bandwidth (300 Hz).

Single lepton trigger paths (*SingleMu* and *SingleElectron*) are also used to record data used to measure the trigger and reconstruction efficiencies. Events triggered by two muons (*DoubleMu*) are used to create the so-called embedded datasets used to model the irreducible $Z/\gamma^* \to \tau\tau$ background, where the reconstructed muon is replaced by a simulated τ (see section 6.2.3).

L1	HLT	Recorded luminosity
$e\tau_h$ **channel**		
IsoEG18er or EG20	Ele20_Tight_LooseIsoPFTau20	0.7 fb^{-1}
IsoEG18er or EG20er or EG22	Ele22_eta2p1_WP90Rho_LooseIsoPFTau20	18.4 fb^{-1} + MC
$\mu\tau_h$ **channel**		
Mu16er	IsoMu18_eta2p1_LooseIsoPFTau20	0.7 fb^{-1}
Mu14er	IsoMu17_eta2p1_LooseIsoPFTau20	18.4 fb^{-1} + MC

Table 6.1: *Trigger paths used in the analysis.*

A full reconstruction of the events is done several weeks after the data taking periods, taking into account the alignment and calibration of all CMS sub-detectors. The events used for the search presented here are stored in the *TauPlusX* datasets. They contain events firing at least one of the High Level Trigger (HLT) paths based on the online coincidence of a τ_h and an extra object (electron, muon, tau, or $\not\!\!E_T$). The datasets used in the analysis are summarized in table 6.2.

dataset name	Run range	Luminosity (fb^{-1})
/TauPlusX/Run2012A*	190456-193621	0.887
/TauPlusX/Run2012B*	193833-196531	4.446
/TauPlusX/Run2012C*	198022-203742	7.153
/TauPlusX/Run2012D*	203777-208686	7.318
Total	-	19.804

Table 6.2: *datasets used in the analysis.*

6.2.2 Simulated datasets

Simulated Monte Carlo is generated in order to estimate the background and signal yields kinematics in the analysis.

The signal simulation is performed separately for the different Higgs production modes and mass hypotheses. The samples of the Standard Model Higgs boson produced via gluon-gluon fusion and vector boson fusion are generated using POWHEG [166]. For the

associated production with a vector boson or with a top anti-top quarks pair as well as the MSSM Higgs boson signal events PYTHIA [167] is used. For the Standard Model, the tested mass hypothesis run from 110 GeV to 145 GeV, in 5 GeV steps. While for the MSSM, the mass range extends from 90 GeV to 1000 GeV with increasing steps. The cross sections and branching ratios as well as the corresponding corrections and uncertainties for the different production modes follow the recommendations from the LHC cross section working group [58–60].

The Standard Model Higgs boson p_T spectrum produced via gluon fusion is reweighed to the next-to-next-to-leading-order (NNLO) calculation from HERS [168]. This reweighting increases the signal yields by 3% for the Standard Model Higgs boson with $m_H = 125$ GeV for $p_T(H) > 100$ GeV. Details about the correction on the Higgs boson p_T simulated signal are given later in section 6.4.13.

As shown in section 2.2.2, the main production mode of the MSSM Higgs boson is the gluon-gluon fusion mechanism. This mechanism relies predominantly on top and bottom quark loops, but the contribution of stop and sbottom loops can be significant if their masses are low. QCD NLO corrections using HIGLU [169] are applied and they can increase the cross section up to 100% (50%) at small (large) $\tan\beta$. The NNLO corrections to the contribution of the top quark in the finite mass limit are taken from GGH@NNLO [170]. Electroweak, squark contributions and SUSY QCD corrections have been neglected in the calculation as they are small (less than 10%), and they are taken into account through an additional systematic uncertainty. The branching ratio of the neutral MSSM Higgs boson decaying to τ pairs is computed using FEYNHIGGS [171].

The $Z \to ll$ $(l = e, \mu)$, $W + jets$, $t\bar{t}$ and di-boson processes are generated using MADGRAPH [172] and the single top processes using POWHEG [166]. Exclusive samples binned in jet multiplicities are also generated and added to the inclusive $Z \to ll + jets$ and $W + jets$ samples in order to increase the background statistics in regions of high signal purity. $Z \to ll$ $(l = e, \mu)$, $W + jets$ and $t\bar{t}$ are normalized to their corresponding NNLO cross sections [173,174], while the single top and di-boson processes are normalized to their next-to-leading-order (NLO) cross sections [175].

The MADGRAPH and POWHEG generators are interfaced with PYTHIA in order to include the parton shower and fragmentation. The pile-up is included by simulating additional pp collisions using PYTHIA. PYTHIA also takes care of the simulation of the underlying event. All generators are interfaced with TAUOLA [176], which simulates the tau decays (leptonically or hadronically). Tau polarization effects are modeled using the TAUSPINNER [177] framework.

The full detector response is simulated using GEANT4 [178], and the events are reconstructed using the CMS reconstruction software [12]. The complete Monte Carlo simulated datasets are reported in table 6.3

The simulated events are weighted according to their respective cross sections σ, the integrated luminosity $\mathcal{L}$ which is measured in data, the MC filter efficiency ϵ[1] and the

[1] *i.e.* the acceptance of the baseline analysis cuts.

dataset description	MC generator	8 TeV generator cross section (pb)	Number of events
Standard Model Higgs signal processes (mass hypothesis between 110 GeV and 145 GeV)			
$gg \to H$	POWHEG	19.6	968134 ($m_H = 125$ GeV)
$qq \to qqH$	POWHEG	1.55	998836 ($m_H = 125$ GeV)
$gg \to t\bar{t}/VH$	PYTHIA	1.14	200124 ($m_H = 125$ GeV)
MSSM Higgs signal processes (mass hypothesis between 90 GeV and 1000 GeV)			
$gg \to \phi$	PYTHIA	-	990976 ($m_\phi = 300$ GeV)
$gg \to \phi b$	PYTHIA	-	999900 ($m_\phi = 300$ GeV)
Background processes			
$t\bar{t}$	MADGRAPH	225.2	68188700
$t \to X(bW)$	POWHEG	11.1	497658
$\bar{t} \to X(bW)$	POWHEG	11.1	493460
$Z \to ll \ + \ jets$ (inclusive)	MADGRAPH	3503.7	30459503
$Z \to ll \ + \ 1 \ jet$	MADGRAPH	666,3	24045248
$Z \to ll \ + \ 2 \ jets$	MADGRAPH	215,0	21852156
$Z \to ll \ + \ 3 \ jets$	MADGRAPH	27,3	11015445
$Z \to ll \ + \ 4 \ jets$	MADGRAPH	3503.7	6402827
$W \to l\nu \ + \ jets$ (inclusive)	MADGRAPH	36257	76102995
$W \to l\nu \ + \ 1jet$	MADGRAPH	6381	52926398
$W \to l\nu \ + \ 2jets$	MADGRAPH	2030	64738774
$W \to l\nu \ + \ 3jets$	MADGRAPH	616	30780647
$W \to l\nu \ + \ 4jets$	MADGRAPH	254	13382803
$WW \to 2l2\nu$	MADGRAPH	5.82	1933235
$WZ \to 3l\nu$	MADGRAPH	1.06	2017979
$WZ \to 2l2q$	MADGRAPH	2.21	1936727
$ZZ \to 2l2q$	MADGRAPH	1.25	1936727
$ZZ \to 4l$	MADGRAPH	0.2	4807893

Table 6.3: *Simulated datasets used in the analysis.*

number of processed MC events N_{proc} following:

$$w_{MC} = \frac{\sigma \cdot \mathcal{L} \cdot \epsilon}{N_{proc}} \tag{6.1}$$

6.2.3 Embedded samples

In order to simulate accurately the most important $Z \to \tau\tau$ irreducible background, a technique called embedding is used [179–181]. $Z \to \mu\mu$ events are selected in data and the reconstructed muons are replaced by generator level τ-leptons which decays are simulated by TAUOLA. The GEANT detector simulation is used to produce the detector response to the τ decay products. The idea behind this approach is to consider that the τ_h decays are very well described by generators while the $pp \to Z/\gamma^* \to \mu\mu$ can suffer from large uncertainties. Using real events of the latter minimizes considerably these uncertainties.

$Z/\gamma^* \to \mu\mu$ events are selected by requiring two well identified and isolated muons. Each of the *Particle Flow* muons is then replaced by a Monte Carlo generated τ-lepton decaying to electron, muon or to hadrons. Also, only events where the invariant τ-pair mass exceeds 50 GeV are kept. Tracks in the inner tracker detectors, energy deposits in the calorimeters plus hits in the muon system from the tau decay products are mixed with the rest of the $Z \to \mu\mu$ event after the two muons from the Z are removed.

This samples are produced for all the possible combinations of tau decays in order to make use of the maximum available event statistics limited by the number of reconstructed $Z \to \mu\mu$ events. They have the advantage to be treated as real data: pile-up content, jet content and kinematics, missing transverse energy, etc.

Specific $t\bar{t}$ embedded samples are used in the analysis in order to correct for its contamination into the $Z \to \tau\tau$ embedded samples (*c.f.* 6.4.9).

6.3 Analysis objects, missing transverse energy and transverse mass

6.3.1 Hadronic taus

The reconstructed tau (see chapters 4 and 5) is required to pass the Tight working point of the multivariate discriminator described in section 4.5. The τ_h isolation and anti-lepton discriminators are chosen in each analysis final state by using the one giving the best expected significance or exclusion limit, depending on the analysis. The anti-lepton discriminators and the working points are chosen depending on whether the dominant background contributions arises from $Z/\gamma^* \to ee$ ($e\tau_h$ channel) or from $Z/\gamma^* \to \mu\mu$ ($\mu\tau_h$ channel).

6.3.2 Electrons and muons

Electrons

The electron reconstruction in CMS is detailed in section 3.3.4. Electrons are identified using a multivariate analysis. The Tight working point is chosen for the electrons in the $e\tau_h$ final state.

Veto electrons are also used for the second electron veto of the baseline analysis selection described in section 7.1.

Muons

For this analysis, muons are required to be reconstructed by the *Tracker* and *Global* muon reconstruction algorithms, as well as to be identified as real muons by the Particle Flow algorithm, as described in 3.3.3.

Electron and muon isolation

In order to reduce the contamination from jets faking leptons, the selected muons and electrons are required to be isolated. The isolation is computed summing over the p_T of the charged particles, neutral hadrons and photons reconstructed by the Particle Flow algorithm within a cone of size $\Delta R_{iso} = 0.4$ around the electron or muon candidate and associated to the primary vertex ($\Delta z < 2\ mm$):

$$I_{rel} = \frac{\sum p_T^{charged}(\Delta z < 2\ mm) + \sum E_T^{neutral}(\Delta z < 2\ mm) + E_T^{photons}(\Delta z < 2\ mm)}{p_T(\mu \text{ or } e)} \tag{6.2}$$

The contribution of pile-up to the isolation of leptons with respect to neutral hadrons and photons is accounted for by applying corrections computed by summing the transverse momenta of the charged particles with longitudinal impact parameters $\Delta z > 2\ mm$ with respect to the primary vertex and scaling the sum by a factor 0.5. The factor 0.5 is coming from the fact that pile-up interactions at LHC produce two times more charged particles than neutrals. The neutral contribution from pile-up interactions, called $\Delta\beta$ contribution is estimated as follows:

$$\Delta\beta = 0.5 \sum p_T^{charged}(\Delta z > 2\ mm) \tag{6.3}$$

Then the pile-up corrected neutral contribution from the hard scattering process is:

$$\sum E_T^{neutral}(\Delta z < 2\ mm) + E_T^{photons}(\Delta z < 2\ mm) - \Delta\beta \tag{6.4}$$

So equation 6.2 becomes:

$$I_{rel} = \frac{\sum p_T^{charged}(\Delta z < 2\ mm) + max(0, \sum E_T^{neutral}(\Delta z < 2\ mm) + E_T^{photons}(\Delta z < 2\ mm) - \Delta\beta)}{p_T(\mu \text{ or } e)} \tag{6.5}$$

The inner veto cone ΔR_{veto} is excluded from the sum in order to avoid contributions from energy deposits in the electromagnetic and hadronic calorimeters that are due to the lepton. In table 6.4 the selection criteria for the candidates entering the isolation computation are displayed.

Type of particle	ΔR^{μ}_{veto}	ΔR^{e}_{veto}	Δz	p_T or E_T
Charged (IP)	0.0001	0.01 (Barrel) / 0.015 (end-cap)	< 2 mm	-
Photons	0.01	0.08	-	> 0.5 GeV
Neutral hadrons	0.01	-	-	> 0.5 GeV
Charged (PU)	0.01	-	> 2 mm	> 0.5 GeV

Table 6.4: *Selection of particles entering the calculation of the relative isolation I_{rel} from equation 6.5. The symbol "-" indicates cuts that are not applied.*

The relative isolation is required to be $I_{rel} < 0.1$ in the semi-leptonic channels of the analysis.

6.3.3 Jets and b-tagging

In order to classify the events in the analysis corresponding to the different production mechanisms, the number of jets and b-tagged jets are used. The jet reconstruction is detailed in section 3.3.6. The MVA based jet identification discriminator is used to improve the discrimination between jets from the production vertex and jets due to pile-up. The loose working point of the so-called MVA "full jet-id" training is used.

The jets are required to be in the kinematic acceptance region $|\eta| < 4.7$, to have a corrected momentum $p_T > 30$ GeV and to be separated from the leptons and taus considered in the analysis by a distance $\Delta R > 0.5$. The cut on $p_T > 30$ GeV applied in the analysis leads to an identification efficiency higher than 99% in the tracker acceptance and around 95% outside.

In the MSSM analysis the number of b-tagged jets is used to define event categories aiming at tagging the associated production of the Higgs boson with b-quarks. The Combined Secondary Vertex (CSV) algorithm is used. The Medium working point is chosen, defined as a threshold on the discriminator output above 0.679, and leads to a b-tag efficiency around 68% and a mis-tag rate of 1,42%. A jet is considered to be b-tagged if it passes the medium working point of the CSV output and enters the kinematical region $|\eta| < 2.4$ and $p_T > 20$ GeV.

6.3.4 Missing transverse energy

The missing transverse energy $\not{E}_T$ is used in the analysis in order to discriminate the signal events containing Higgs bosons from background events with $W^{\pm}$ bosons which decays produce neutrinos and therefore have significant missing energy. The $\not{E}_T$ corresponds

ideally to the sum of transverse momenta of the invisible particles (neutrinos) in the event, used as input of the di-τ mass estimation algorithm. In section 3.3.7 the $\not\!\!E_T$ computation is presented.

In the analysis, the multivariate regression based $\not\!\!E_T$ (so-called MVA $\not\!\!E_T$) is used. The algorithm uses the fact that pile-up predominantly produces jets with low momentum and the energy is unclustered (the hadrons are not reconstructed within jets), while leptons and high momenta jets are almost exclusively produced by the hard scattering, even in high pile-up conditions.

6.3.5 Transverse mass

The transverse mass $M_T(lep, \not\!\!E_T)$ of the lepton (electron or muon) and the missing transverse energy is:

$$M_T(lep, \not\!\!E_T) = \sqrt{(p_T^{lep} + \not\!\!E_T)^2 - ((p_x^{lep} + \not\!\!E_x)^2 + (p_y^{lep} + \not\!\!E_y)^2)} \tag{6.6}$$

A cut on this variable is applied to reject and select events with $W + jets$ contribution.

6.4 Corrections to the simulation

Simulation only imperfectly models real collisions and CMS's response. In order to improve the quality of the simulation, several corrections are applied to the Monte Carlo and embedded samples.

6.4.1 Pile-up

As explained in section 6.2.2, pile-up interactions are simulated by superimposing minimum bias events, generated by PYTHIA, to the hard scattering event. The multiplicity of pile-up interactions in the simulation is sampled from a Poisson distribution. It does not exactly reproduce the one observed in data as seen in figure 6.1, describing the pile-up distributions for the Monte Carlo simulation and the data during the 2012 data taking period. In order to match the conditions in data, every event simulated with N in time pile-up interactions is assigned a weight equal to the ratio between the probability of observing N in data and in the simulation (pile-up reweighting):

$$w_{PU} = \frac{N_{PU}^{data}}{N_{PU}^{MC}} \tag{6.7}$$

6.4.2 Trigger efficiency

The ratio of the trigger efficiencies measured in data and in Monte Carlo samples using the Tag-and-Probe technique [182] is used to correct simulation. In good approximation the efficiency of each cross-trigger, involving two kind of trigger objects (for instance an

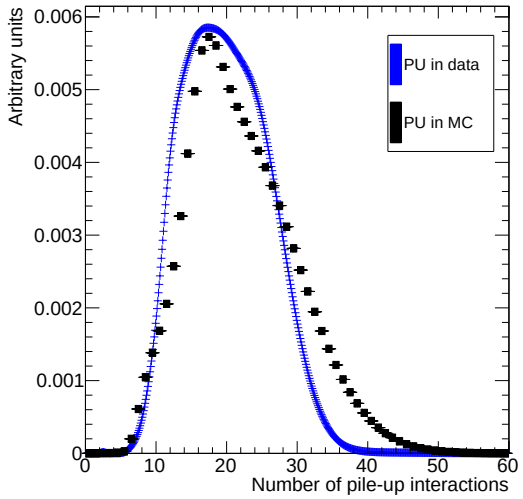


Figure 6.1: *Pile-up distributions for simulation before reweighting and data.*

electron and a τ_h), is given by the product of efficiencies of each trigger object. The correction is applied as a function of the transverse momentum of the object. The trigger efficiency corrections are applied to all simulated samples:

$$w_{trig} = \frac{\epsilon^{data}}{\epsilon^{MC}} = \frac{\epsilon^{data}(p_T^{lep}) \times \epsilon^{data}(p_T^{\tau})}{\epsilon^{MC}(p_T^{lep}) \times \epsilon^{MC}(p_T^{\tau})} \tag{6.8}$$

The p_T dependence of the trigger efficiencies is modeled by the integral of a Crystal-Ball function [183]:

$$F(P_T) = \int_{-\infty}^{P_T} f(t)\, dt \tag{6.9}$$

where the integrand is defined by:

$$f(t;\alpha,n,m_0,\sigma,norm) = \frac{norm}{\sigma(C+D)} \cdot \begin{cases} e^{-\left(\frac{t-m_0}{\sqrt{2}\sigma}\right)^2} & \text{if } \frac{t-m_0}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{t-m_0}{\sigma}\right)^{-n} & \text{if } \frac{t-m_0}{\sigma} \leq -\alpha \end{cases} \tag{6.10}$$

with

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{|\alpha|^2}{2}}, \qquad B = \frac{n}{|\alpha|} - |\alpha|$$

$$C = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot e^{-\frac{|\alpha|}{2}}, \qquad D = \sqrt{\frac{\pi}{2}}\left(1 + \mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right)$$

and

$$\mathrm{erf}(x) = \frac{2}{\pi}\int_0^x e^{-t^2}\, dt. \tag{6.11}$$

Electron and muon leg trigger efficiency $\epsilon(p_T^{lep})$

The efficiency of the electron and muon legs of the $e\tau_h$ ($\mu\tau_h$) trigger is measured using $Z/\gamma^* \to ee$ ($Z/\gamma^* \to \mu\mu$) events. Events passing a single electron (muon) trigger are selected in order to not bias the measurement. Two electrons (muons) of opposite charge within the mass window $60 < M_{ee} < 120$ GeV ($60 < M_{\mu\mu} < 120$ GeV) are required. The tag electron (muon) is required to satisfy $P_T > 30$ GeV($P_T > 25$ GeV), $|\eta| < 2.1$ and to to pass the single electron (muon) trigger. The probe electron (muon) is required to satisfy $P_T > 24$ GeV($P_T > 20$ GeV) and $|\eta| < 2.1$. Both electrons (muons) are required to pass the particle identification and isolation criteria described in section 6.3.2.

The efficiency is then computed as:

$$\epsilon = \frac{N(\text{pass HLT \&\& matched to HLT object})}{N(\text{pairs})} \tag{6.12}$$

measured as function of P_T in bins of η, separately for data and Monte Carlo simulated events.

The coefficients of the fitted Crystal-Ball function (also called "turn-on curve") obtained for data and Monte Carlo simulated events are given in table 6.5 for the electron leg and table 6.6 for the muon leg. The measured trigger efficiency as function of the transverse momentum of the lepton ($p_T(e \text{ or } \mu)$) is shown in figures 6.2 and 6.3.

Fit parameters for 8 TeV data						
Trigger Path	Region	α	n	m_0	σ	norm
Ele22	$\|\eta\| < 1.479$	1.26889	1.31024	22.9704	1.0258	1.06409
	$\|\eta\| > 1.479$	0.978597	2.33144	21.9816	1.40993	0.937552
Fit parameters for 8 TeV Monte Carlo simulation						
Trigger Path	Region	α	n	m_0	σ	norm
Ele22	$\|\eta\| < 1.479$	0.739301	1.34903	21.7243	0.619015	1.02594
	$\|\eta\| > 1.479$	1.8885	1.01855	22.1217	1.34054	4.7241

Table 6.5: *Coefficients of the Crystal Ball function parametrizing the efficiency of the electron leg of* $e\tau_h$ *cross-trigger (used in the* $e\tau_{had}$ *channel) for data taken in* 2012 *compared to the Monte Carlo simulation.*

The simulation is corrected by the ratio of the efficiencies measured in data and Monte Carlo as function of the p_T of the electron. For high p_T electrons, the ratio is of the order of 98%.

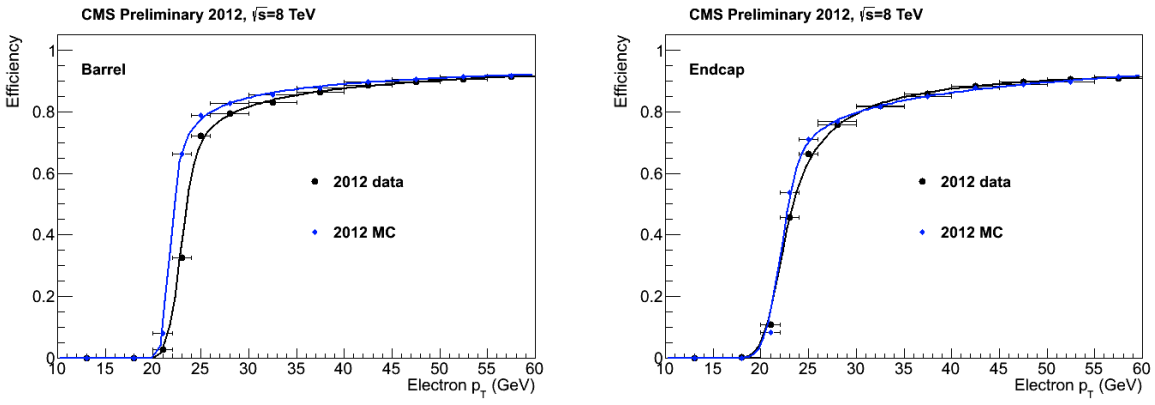


Figure 6.2: *Trigger "turn-on" curves measured in data and in simulation for the electron leg of $e\tau_h$ cross-trigger (Ele22 trigger path) as a function of the electron transverse momentum in the Barrel (left) and end-cap (right) regions. The efficiency is fitted by the integral of a Crystal-Ball function. From Ref. [184].*

Fit parameters for 8 TeV data						
Trigger Path	Region	α	n	m_0	σ	norm
IsoMuX	$\eta < -1.2$	6.4951e-08	1.57403	15.9977	7.64004e-05	0.865325
	$-1.2 < \eta < -0.8$	0.804001	1.24295	17.3974	0.804001	0.928198
	$-0.8 < \eta < 0$	0.226312	1.55756	16.4307	0.226312	0.974462
	$0 < \eta < 0.8$	0.662731	1.05778	17.313	0.662731	1.26624
	$0.8 < \eta < 1.2$	0.550532	1.55402	16.9966	0.550532	0.885134
	$\eta > 1.2$	0.000106195	1.9991	15.9962	0.000106195	0.851294
Fit parameters for 8 TeV Monte Carlo simulation						
Trigger Path	Region	α	n	m_0	σ	norm
IsoMu17	$\eta < -1.2$	4.3335e-09	1.66134	16.0051	2.45144e-05	0.87045
	$-1.2 < \eta < -0.8$	1.21803	1.40611	17.3135	0.747636	0.934983
	$-0.8 < \eta < 0$	0.00589832	1.75409	15.9556	0.0236127	0.981338
	$0 < \eta < 0.8$	0.00448573	1.92101	15.9289	0.0271317	0.978625
	$0.8 < \eta < 1.2$	0.354533	1.67085	16.5678	0.328333	0.916992
	$\eta > 1.2$	4.40036e-08	1.66272	15.997	7.90069e-05	0.884502

Table 6.6: *Coefficients of the Crystal Ball function parametrizing the efficiency of the muon leg of $\mu\tau_h$ cross-trigger (used by the $\mu\tau_{had}$ channel) for data taken in* 2012 *compared to the Monte Carlo simulation. The label IsoMuX refers to the mix of IsoMu17 and IsoMu18 trigger paths that were used during different data-taking periods in* 2012 *(cf. Tab. 6.1).*

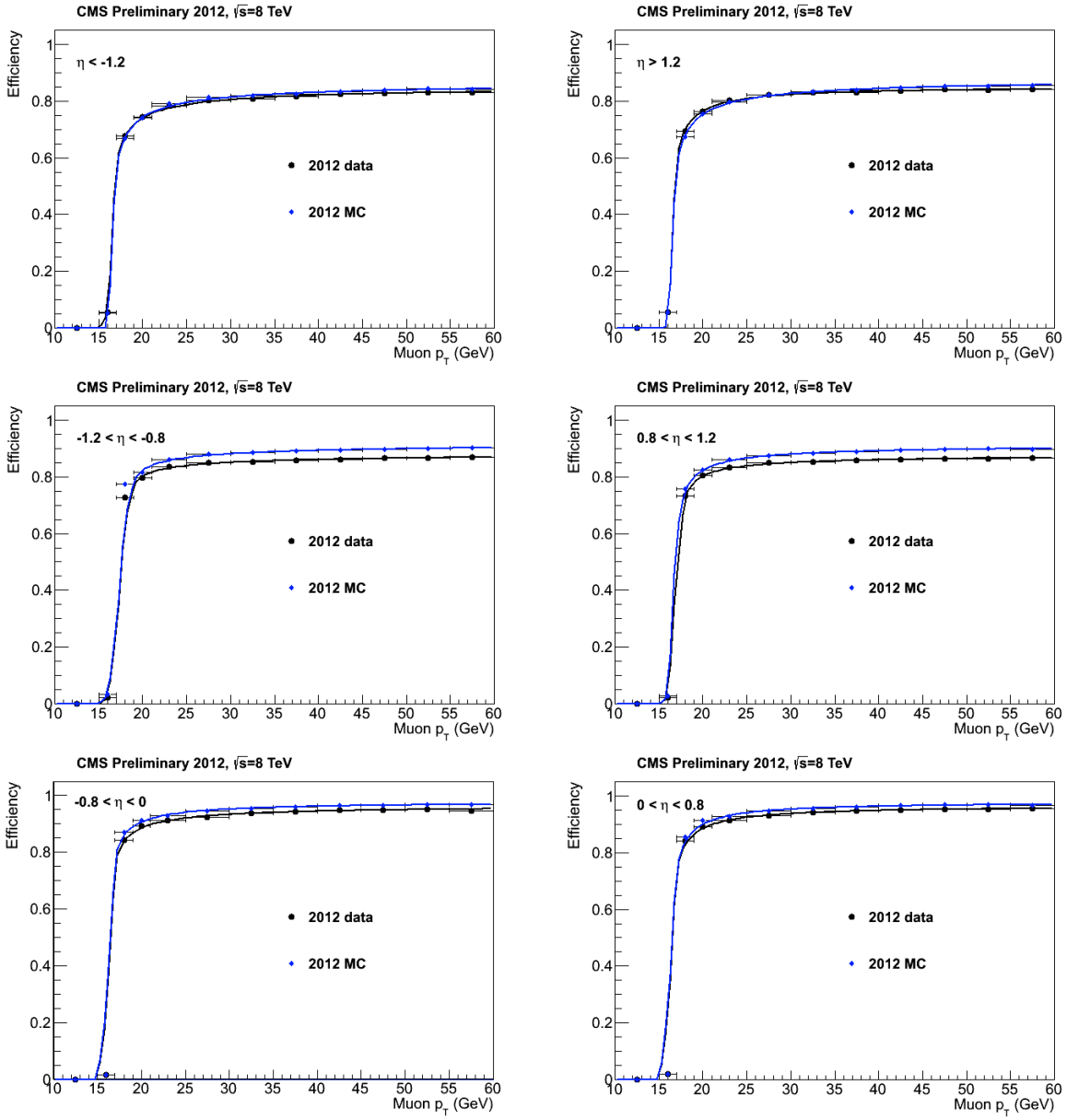


Figure 6.3: *Trigger "turn-on" curves measured in data and in simulation for the muon leg of $\mu\tau_h$ cross-trigger (IsoMu17 trigger path) as a function of the muon transverse momentum in different detector geometrical regions. The efficiency is fitted by the integral of a Crystal-Ball function. From Ref. [184].*

τ_h leg trigger efficiency $\epsilon(p_T^\tau)$

The efficiency of the hadronic tau leg of the trigger is measured also with the Tag-and-Probe technique using $Z \to \tau\tau \to \mu\tau_h$ events. The events are required to fire the $HLT\_IsoMu12\_eta2p1\_ETM20$ trigger path. The events are selected with one tight isolated muon and one hadronic decaying tau satisfying $I_\tau < 1.5$ GeV (see equation 4.3 from section 4.5) with a visible mass reconstructed in the window range $[45 - 70$ GeV$]$.

The offline selected muon is required to match the muon leg of trigger that fires the $HLT\_IsoMu12\_eta2p1\_ETM20$ path as well as the analysis HLT path (*c.f.* table 6.1). The efficiency of the τ_h leg is computed as the ratio of the number of selected probes passing the corresponding HLT path as well as the offline selected τ_h matched to the tau leg of the corresponding trigger and the total number of the selected events.

For trigger paths used in the $\mu\tau_h$ and $e\tau_h$ final states, the measurement is performed separately for the barrel and end-cap regions of the CMS detector. The measured trigger efficiency as function of the $p_T(\tau_h)$ is shown in figure 6.4. The turn on is parametrized using an integral Crystal-Ball function. The parameters of the fitted function is provided in table 6.7.

Fit parameters for 8 TeVdata						
Trigger Path	Region	α	n	m_0	σ	norm
LooseIsoPFTau20	$\|\eta\| < 1.5$	3.87275	1.06881	18.3508	1.94454	1.24298
	$\|\eta\| > 1.5$	1.15112	11.206	17.918	1.34882	0.891967
Fit parameters for 8 TeVMonte Carlo simulation						
Trigger Path	Region	α	n	m_0	σ	norm
LooseIsoPFTau20	$\|\eta\| < 1.5$	3.25596	1.11202	17.9957	-1.54784	1.07779
	$\|\eta\| > 1.5$	1.15562	11.3419	18.205	1.06276	0.898667

Table 6.7: *Coefficients parametrizing the efficiency of the hadronic tau leg of the $e\tau_h$ and $\mu\tau_h$ cross-trigger (used by the $e\tau_{had}$ and $\mu\tau_h$ channels) for data taken in* 2012 *compared to the Monte Carlo simulation. The efficiency is measured as described in the text and fitted by the integral of a Crystal-Ball function.*

In the context of the analysis presented in this thesis, it was noticed that, due to a bug in the particle-flow reconstruction in the HLT, the τ_h reconstruction efficiency for high p_T τ_h ($\gtrsim$ 140 GeV) decreases rapidly. It was corrected in the HLT in the middle of 2012, thus affecting part of the 8 TeV data recorded during early 2012. However, the MC was produced with a HLT menu before the correction was implemented and hence is affected. Figure 6.5 shows the trigger efficiency in the MC events at high τ_h p_T obtained from the combination of all MSSM $H \rightarrow \tau\tau$ samples. In order to correct the effect of this loss in efficiency at high p_T, the analysis now takes into account this effect by applying additional p_T-dependent scale factors which is parametrized as a second order polynomial. In barrel, the polynomial used in the fit is:

$$F(x) = 1 - \alpha \times (x - 140) + \beta \times (x - 140)^2 \tag{6.13}$$

while in end-cap, it is:

$$F(x) = 1 - \alpha \times (x - 60) + \beta \times (x - 60)^2 \tag{6.14}$$

It can be noted that, the drop in efficiency starts earlier in end-cap region as compared to the barrel. The fit parameters are provided in table 6.8.

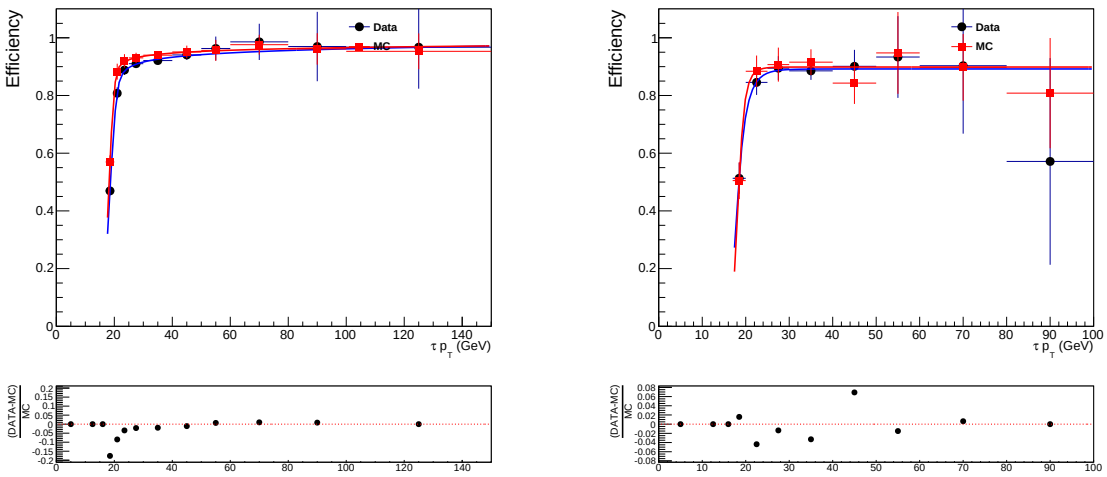


Figure 6.4: *Trigger "turn-on" curves measured in data and in simulation for the hadronic tau leg of $e\tau_h$ and $\mu\tau_h$ cross-trigger (LooseIsoPFTau20 trigger path) as a function of the τ transverse momentum in the Barrel (left) and end-cap (right) regions. The efficiency is measured as described in the text and fitted by the integral of a Crystal-Ball function. From Ref. [185].*

Fit parameters for 8 TeVMonte Carlo simulation			
Trigger Path	Region	α	β
LooseIsoPFTau20	$\lvert\eta\rvert < 1.5$	9.0128010^{-4}	4.8159210^{-7}
	$\lvert\eta\rvert > 1.5$	1.8114810^{-3}	5.4433510^{-7}

Table 6.8: *The efficiency of the τ_h leg of the $e\tau_h$ and $\mu\tau_h$ cross-trigger (used by the $e\tau_{had}$ and $\mu\tau_h$ channels) for the sum of all MSSM signal samples. The efficiency at $p_T > 140$ GeV for barrel and $p_T > 60$ GeV for end-cap are fitted with a second order polynomial.*

6.4.3 τ_h decay mode reweighting

The τ_h decay modes are reconstructed using the "Hadron Plus Strip" algorithm as explained in section 6.3.1. Differences in the reconstruction of each decay mode are observed in data and in simulation, as can be seen in figure 6.6 for the baseline selection in the $\mu\tau_h$ channel. To account for this, six scale factors are derived: one for each decay mode and for each region of the detector (barrel, end-cap). The results are summarized in table 6.9. Simulated events are weighted by this data-to-Monte Carlo scale factors per reconstructed τ_h candidate that is matched to a genuine hadronic tau decay on generator level in Monte Carlo simulated events (signal and background) and in the $Z/\gamma^* \to \tau\tau$ embedded samples.

The scaling is done such that the total yield of taus reconstructed in the sum of decay modes *Single hadron*, *Hadron plus one Strip*, *Hadron plus two Strips* and *Three Hadrons* is kept constant.

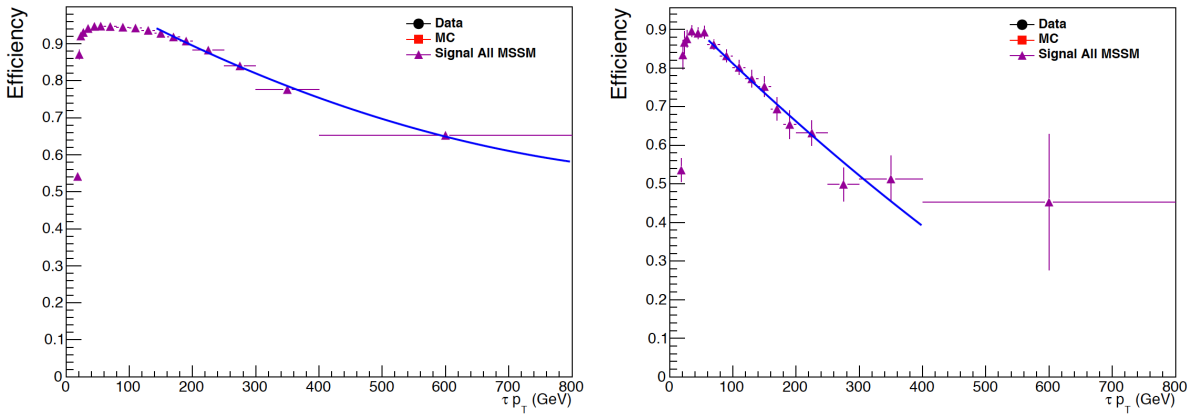


Figure 6.5: *The efficiency of the hadronic tau leg of the $e\tau_h$ and $\mu\tau_h$ cross–trigger (used by the $e\tau_{had}$ and $\mu\tau_h$ channels) for sum of all MSSM signal samples. The efficiency at $p_T > 140$ GeV for barrel and $p_T > 60$ GeV for end-cap are fitted with a second order polynomial. From Ref. [185].*

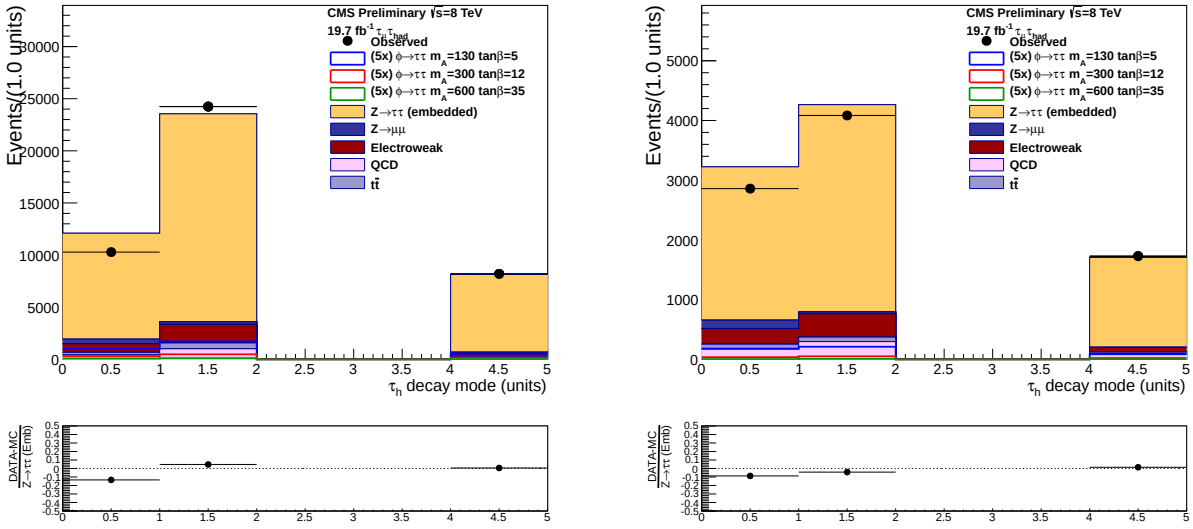


Figure 6.6: *Distribution of the decay mode (0=Single Hadron, 1=Single Hadron plus Strips, 4=Three Hadrons) in the inclusive event category for τ_h candidates in the barrel (left) and the end-caps (right).*

6.4.4 τ_h energy scale

The τ_h energy scale measurement is detailed in section 4.6. The corrections applied in the analysis are:

- **Three Hadrons:** The energy of taus reconstructed in this decay mode is scaled by a factor 1.01.

Decay mode	Inclusive	Barrel	Endcaps
1-prong	0.89	0.87	0.96
1-prong + π^0(s)	1.05	1.06	1.00
3-prong	1.02	1.02	1.06

Table 6.9: *Decay mode scale factors, to be applied to the $Z \to \tau\tau$ simulation, in order to correct for the difference seen with respect to data. The highest correction is for the 1-prong decays in the barrel.*

- **Hadron plus one Strip** or **Hadron plus two Strips:** The energy of taus reconstructed in this decay mode is scaled by a factor 1.01.
- **Single hadron:** The energy of taus reconstructed in this decay mode is scaled by a factor 1.01.

The tau energy scale correction is applied to reconstructed τ_h candidates that are matched to hadronic tau decays at generator level in Monte Carlo simulated events and in the $Z/\gamma^* \to \tau\tau$ embedded samples.

6.4.5 Jet energy scale

The raw jet momentum is equal to the sum of the momenta of all particles in the jet. A calibration of the raw momentum is performed (using the balance between a jet and a photon or a Z boson) as a function of the raw jet kinematic to match, on average, the initial parton momentum. Three levels of corrections are sequentially applied to the raw jet momentum to account for particle contamination and to correct for the response, non-linearity, and inhomogeneity of the calorimeters [186].

The contribution to the jet momentum coming from pile-up interactions and from hadrons produced in the underlying event is estimated using the *FastJet* technique [187]. First, the transverse energy density (ρ) of the extra hadronic activity is determined by the median energy density obtained from all particle-flow jets reconstructed in the event with p_T greater than 5 GeV. An amount equal to ρ times the jet area [188] is then subtracted from the jet transverse momentum.

6.4.6 b-tag efficiency and mis-tagging rate

The b-tag efficiency is measured in data and simulated samples using jets containing soft muons or in $t\bar{t}$ events. The mis-tagging rate is estimated by using negative tagging algorithms, identical to the default algorithms, except that they are used only on tracks with negative IP values or on secondary vertices with negative decay lengths. The differences are corrected by applying efficiency and mis-tag scale factors (SF) [134] to simulation. The simulated samples are then corrected by reclassifying the jets as tagged or un-tagged. The promotion probability (to tag a previously un-tagged jet) or demotion (to un-tag a

previously tagged jet) are defined as a function of p_T, η, the jet flavor and the efficiency in simulated samples (Eff). They are defined as:

$$\begin{aligned} P(denote) &= 1 - SF \qquad \text{when } SF < 1 \\ P(promote) &= \frac{SF - 1}{\frac{SF}{Eff} - 1} \qquad \text{when } SF > 1 \end{aligned} \tag{6.15}$$

The scale factors uncertainties are propagated by varying them by $\pm 1\sigma$. The b-tagging and mis-tagging uncertainties are then treated as independent nuisance parameters in the analysis maximum-likelihood fit.

6.4.7 Lepton Identification and isolation efficiencies

The mis-modeling of the electron and muon identification and isolation efficiencies are taken into account by applying data-to-simulation scale factors. The efficiency of the electron and muon identification and isolation are measured by using the Tag-and-Probe technique on $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ events.

The simulated events are weighted in order to match the data identification and isolation efficiencies in (p_T, η) bins. The scale factors for the electrons and muons used in the analysis are reported in table 6.10 and 6.11.

Electrons		
Identification	Barrel	end-cap
24 GeV $< p_T <$ 30 GeV	0.8999 ± 0.0018	0.7945 ± 0.0055
$p_T >$ 30 GeV	0.9486 ± 0.0003	0.8866 ± 0.0001
Isolation	Barrel	end-cap
24 GeV $< p_T <$ 30 GeV	0.9417 ± 0.0019	0.9471 ± 0.0037
$p_T >$ 30 GeV	0.9804 ± 0.0003	0.9900 ± 0.0002

Table 6.10: *Identification and isolation data-to-simulation scale factors for electrons. From Ref. [184].*

6.4.8 Missing transverse energy recoil corrections

The MVA $\not{E}_T$ used in the analysis reduces considerably the pile-up effects as shown in section 3.3.7. However, large discrepancies in the scale and resolution are present between data and simulated samples due to the mis-modeling of the detector and of the hadronic activity. In order to take this effects into account, a data-driven "recoil" correction is derived from $Z/\gamma^* \to \mu\mu$ events.

The $u_\perp$ and $u_\parallel + q_T$ components of the hadronic recoil (see section 3.3.7) are parametrized as a function of the Z boson transverse momentum q_T. A fit is done on the resolution

Muons			
Identification	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.2$	$1.2 < \lvert\eta\rvert < 2.1$
20 GeV $< p_T <$ 30 GeV	0.9818 ± 0.0005	0.9829 ± 0.0009	0.9869 ± 0.0007
$p_T >$ 30 GeV	0.9852 ± 0.0001	0.9852 ± 0.0002	0.9884 ± 0.0001
Isolation	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.2$	$1.2 < \lvert\eta\rvert < 2.1$
20 GeV $< p_T <$ 30 GeV	0.9494 ± 0.0015	0.9835 ± 0.0020	0.9923 ± 0.0013
$p_T >$ 30 GeV	0.9883 ± 0.0003	0.9937 ± 0.0004	0.9996 ± 0.0005

Table 6.11: *Identification and isolation data-to-simulation scale factors for muons. From Ref. [184].*

and scale of the recoil using a third-order and a first order polynomial, respectively. The fits are performed separately on data and simulation. The parametrization on the data is used to correct the Monte Carlo simulation by rescaling the response and smearing the resolution by the difference in quadrature. Separate measurements and corrections as a function of the q_T are derived for events having 0, 1 and ≥ 2 jets with $p_T > 30$ GeV.

The $\overrightarrow{\not{E}_T}$ in simulation is recomputed using equation 3.10 by replacing the corrected recoil and adding back the q_T for each event. The recoil correction improves the simulation agreement to data by typically decreasing the $\not{E}_T$ response by $\sim 4\%$ and degrading the resolution by $\sim 6\%$ in simulated events with q_T around $10 - 15$ GeV.

In figure 6.7 the response of the MVA $\not{E}_T$ as a function of q_T is showed. The values are around 0.9 even at large q_T due to the BDT training. The optimization was done aiming for the best performance in terms of improved resolution rather than for the unity response. Very good agreement between data and simulation is achieved.

The parallel ($u_\parallel$) and perpendicular ($u_\perp$) resolution components as a function of q_T are displayed in figure 6.8. As for the response, simulation shows very good agreement with data after the corrections.

6.4.9 Embedded samples

Embedded $Z \to \tau\tau$ samples can be polluted by $t\bar{t}$ events. Leptonically decaying $t\bar{t}$ events can be selected and embedded with two simulated *tau* leptons. In order to take this contamination into account, a specific Monte Carlo $t\bar{t}$ embedded simulation is produced by replacing reconstructed muons by simulated τ-leptons following the same treatment as the usual embedded data.

These embedded $t\bar{t}$ samples are analyzed and their contribution is directly subtracted from the $Z \to \tau\tau$ samples. The contribution of the $t\bar{t}$ in the embedded samples amounts to $5-9\%$ on the $Z \to \tau\tau$ yield in the analysis categories (*c.f.* 7.2.2) where a jet is b-tagged (b-tag categories).

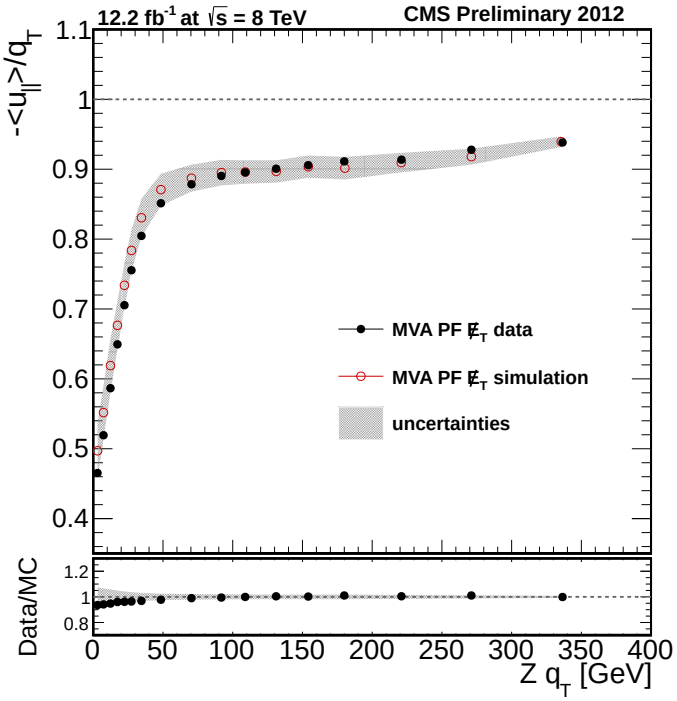


Figure 6.7: *Response as a function of q_T for the MVA $\not\!\!E_T$ using $Z \to \mu\mu$ events. From Ref. [137].*

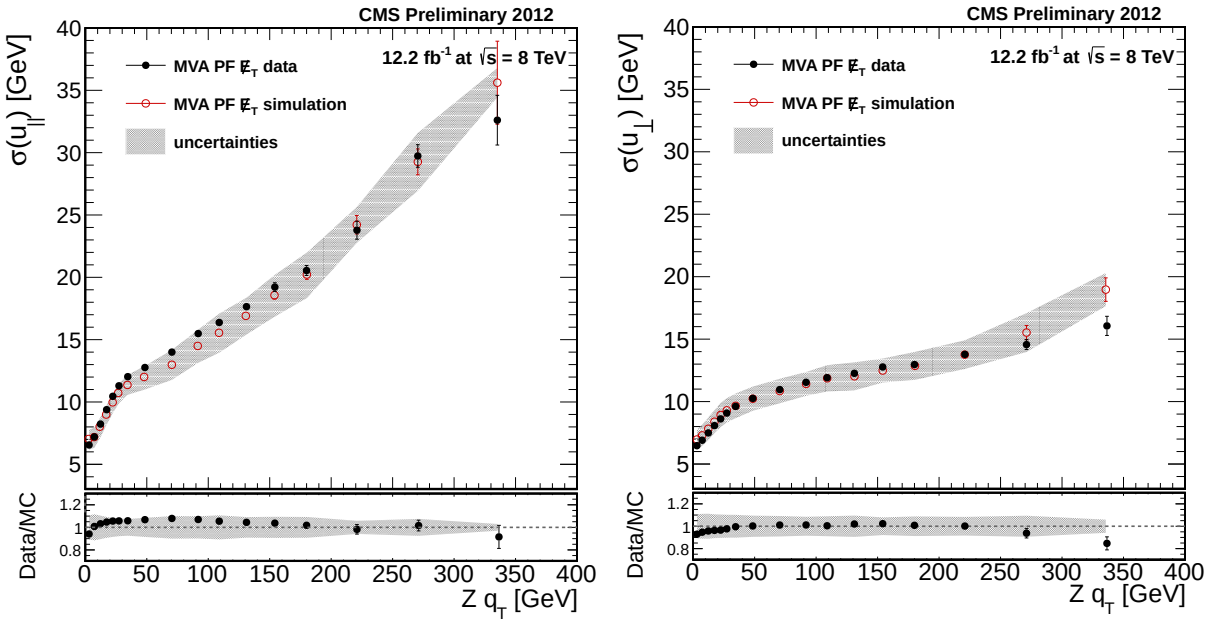


Figure 6.8: *MVA $\not\!\!E_T$ parallel (left) and perpendicular (right) resolution as a function of q_T using $Z \to \mu\mu$ events. From Ref. [137].*

6.4.10 Rate of $e \to \tau_h$ fakes

The rate of $e \to \tau_h$ fakes is measured in the context of this analysis in order to have corrections as a function of the τ_h decay modes. The measurement is done by fitting the distribution of $e + \tau_h$ visible mass in events passing the following criteria in order to have an unbiased sample of $Z/\gamma^* \to ee$ background events:

- $M_T(e, \not{E}_T) < 40$ GeV
- The event is required to have less than 2 jets of $P_T > 20$ GeV
- The hadronic τ is required to have $P_T > 20$ GeV, to pass the Medium WP of the MVA anti–e discriminator plus the Loose working point of the anti–mu discriminator and to pass the the Tight working point (WP) of the MVA based tau isolation discriminator that includes tau lifetime information (see section 4.5).

Separate fits are performed for $e \to \tau_h$ fakes reconstructed in the tau decay modes *Single hadron* and *Hadron plus* ≥ 1 *Strips* and in the ECAL barrel ($|\eta| < 1.479$) and end-cap ($|\eta| > 1.479$) regions. The data/MC scale factors are applied as event weight to simulated $Z/\gamma^* \to ee$ background events in the $e\tau_h$ channel. In figures 6.9 and 6.10 the prefit and postfit distributions of the visible di-τ mass are displayed for the different regions and hadronic τ decay modes.

$e \to \tau_h$ fake rate correction factors for 8 TeVdata		
Region	*Single hadron*	*Hadron plus* ≥ 1 *Strips*
$\|\eta\| < 1.479$	1.37 ± 0.10	1.86 ± 0.22
$\|\eta\| > 1.479$	0.72 ± 0.30	0.85 ± 0.10

Table 6.12: $e \to \tau_h$ *fake-rate correction factors for* τ_h *candidates reconstructed in the tau decay modes* Single hadron *and* Hadron plus ≥ 1 Strips *and in the ECAL barrel (*$|\eta| < 1.479$*) and end-cap (*$|\eta| > 1.479$*) regions.*

6.4.11 Reweighting of the τ_h transverse momentum in W+jets background events

Differences between data and Monte Carlo simulation in the rates of $jet \to \tau_h$ fakes have been observed. This difference do not represent a problem in terms of W + jets background normalization, as the yield of W + jets background in the signal region is determined from data, as described in section 7.3.

The shape of the $p_T(\tau_h)$ spectrum in the sideband region differs quite substantially between data and simulation, as shown in the left side of figure 6.11. To account for this, a corrective weight is applied to Monte Carlo. It is derived from the data to Monte Carlo ratio in a region dominated by W+jets, the high-M_T sideband ($M_T > 50$ GeV).

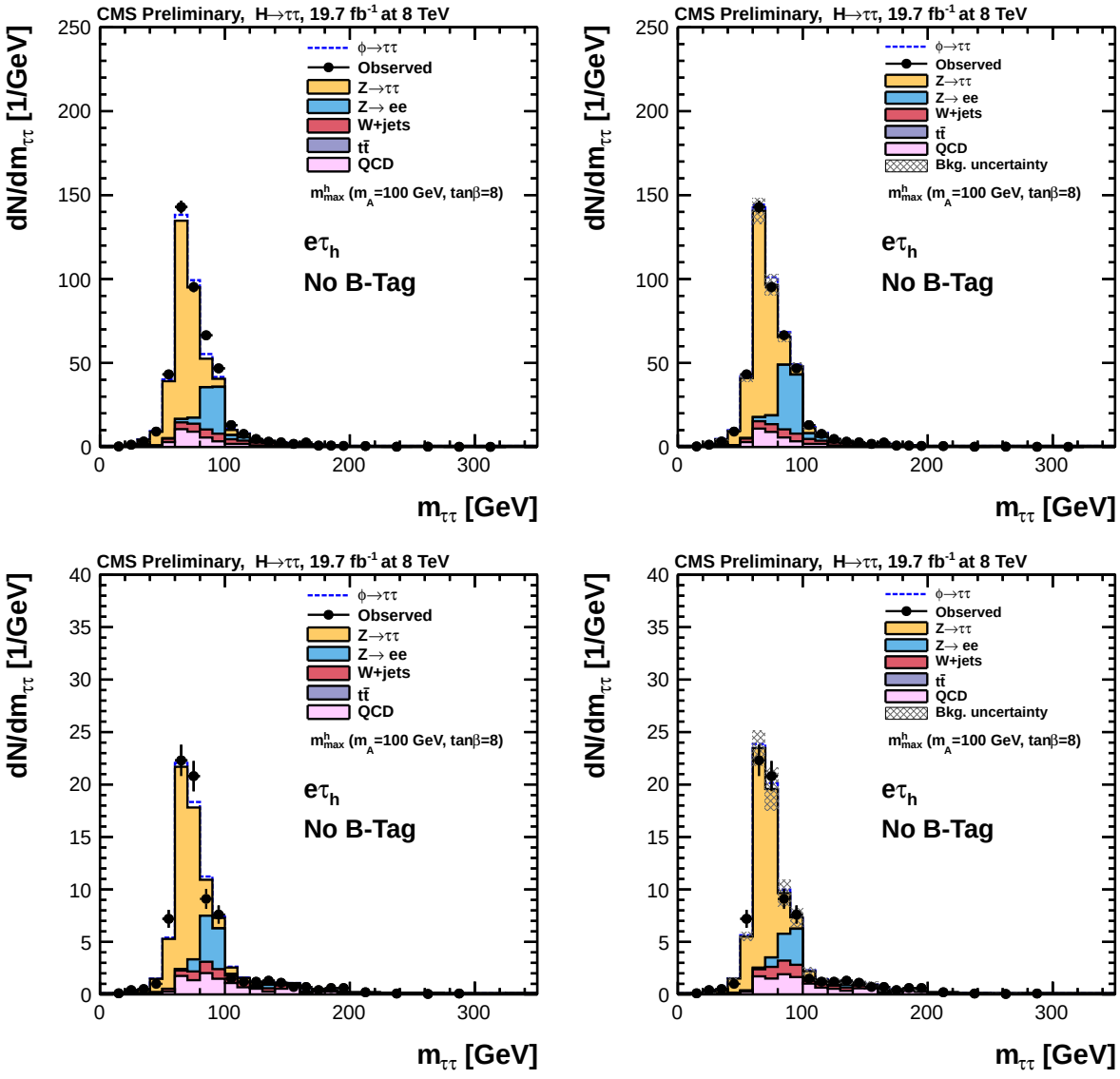


Figure 6.9: *Prefit (left) and post-fit (right) distributions of $M_{\tau\tau}$ in the barrel region (up) and endcap region (bottom) for taus reconstructed in the* Single hadron *decay mode.*

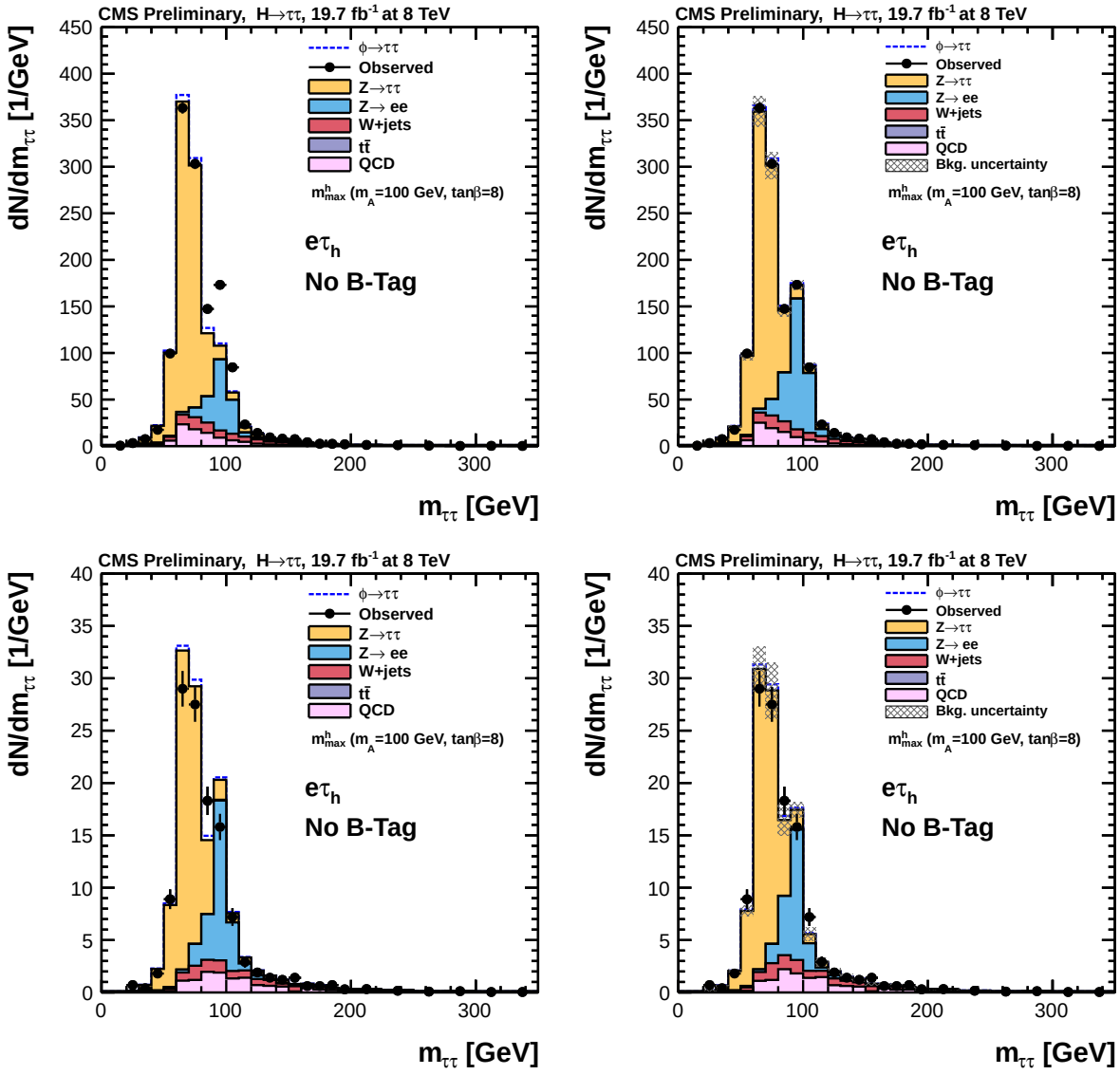


Figure 6.10: *Prefit (left) and post-fit (right) distributions of $M_{\tau\tau}$ in the barrel region (up) and end-cap region (bottom) for taus reconstructed in the* Hadron plus ≥ 1 Strips *decay mode.*

The difference is accounted for by subtracting the contribution of backgrounds other than W + jets from the data and deriving the ratio with respect to the W + jets background prediction, obtained from the simulation.

The ratio is fitted using a 3$^{\text{rd}}$ degree polynomial function, and the function is applied as an event weight to the Monte Carlo simulated W+jets sample. The result of the fit is shown in the right side of figure 6.11.

The resulting fit function, used to reweight the W+jets Monte Carlo events in the signal region, takes the following form:

$$\begin{aligned} F[X] &= 0.79 - 0.15X - 0.03X^2 - 0.08X^3, \\ \text{where } X &= (p_T(\tau_h)[GeV] - 150)/100 \end{aligned} \tag{6.16}$$

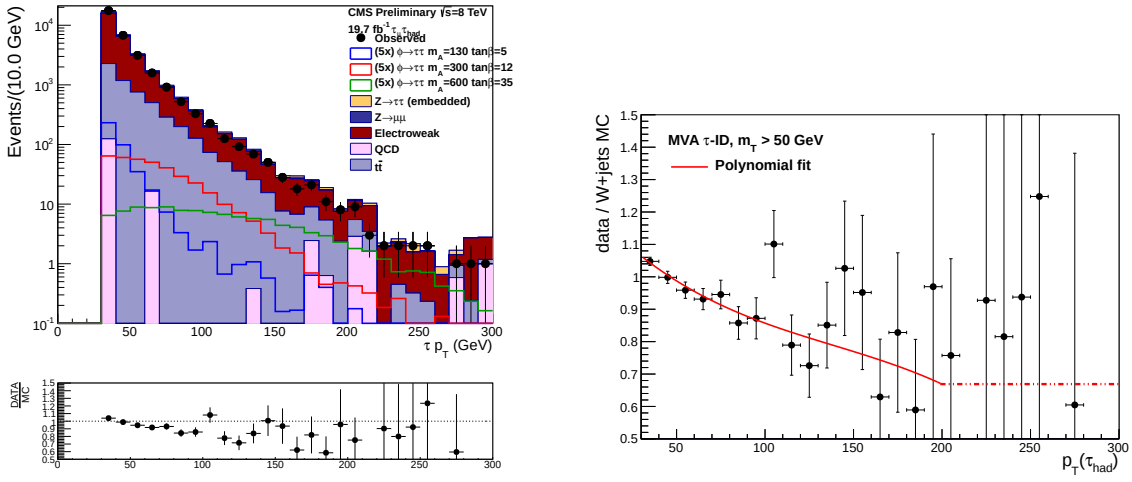


Figure 6.11: *$p_T(\tau_h)$ distribution of data and backgrounds in the high-M_T sideband (left) dominated by W+jets events. $p_T(\tau_h)$ data to Monte Carlo ratio in the high-M_T sideband and fitted function of this ratio (red line using a 3$^{\text{rd}}$ polynomial fit in the 30 to 200 GeV region) (right). Above 200 GeV, the fit is not used but the correction for $p_T(\tau_h) = 200$ GeV is kept.*

6.4.12 $t\bar{t}$ p_T-reweighting

In the normalized differential top-quark-pair cross section analysis, the shape of the p_T spectrum of the individual top quarks in data was found to be softer than predicted by the various simulations while the available NNLO prediction [189] delivers a reasonable description. Based on these measurements, $t\bar{t}$ simulated events are weighted in order to correct for the mis-modeling of the top quark p_T distribution.

The shape correction is applied as an event weight to the Monte Carlo simulated events. The weights are derived as a function of the generated top-quark p_T up to 400 GeV for both top quarks from the $t\bar{t}$ pair. The event weight is then *weight* =

$\sqrt{SF(top) \times SF(anti-top)}$ where the SF function is $SF(x) = \exp(a+bx)$. In table 6.13, the values obtained for the parameters a and b are shown.

$t\bar{t}$ **event weight for 8 TeV data**		
Channel	a	b
All combined	0.156	-0.00137
lep+jets	0.159	-0.00141
di-lepton	0.148	-0.00129

Table 6.13: *a and b coefficients for the event weights applied to the $t\bar{t}$ simulated events. From Ref. [190].*

6.4.13 Higgs boson p_T-reweighting in gluon-gluon fusion process

As described in section 6.2.2, the Higgs boson transverse momentum is simulated using a fixed-order Monte Carlo generator (POWHEG and PYTHIA). In the gluon fusion process, large logarithmic terms spoil the computation of the Higgs p_T distribution for low values. The effects of the different orders in the development on the Higgs boson p_T spectra are shown in figure 6.12. At the analysis level, due to the kinematical acceptance cuts on the transverse momenta, the signal yield will depend on the Higgs p_T.

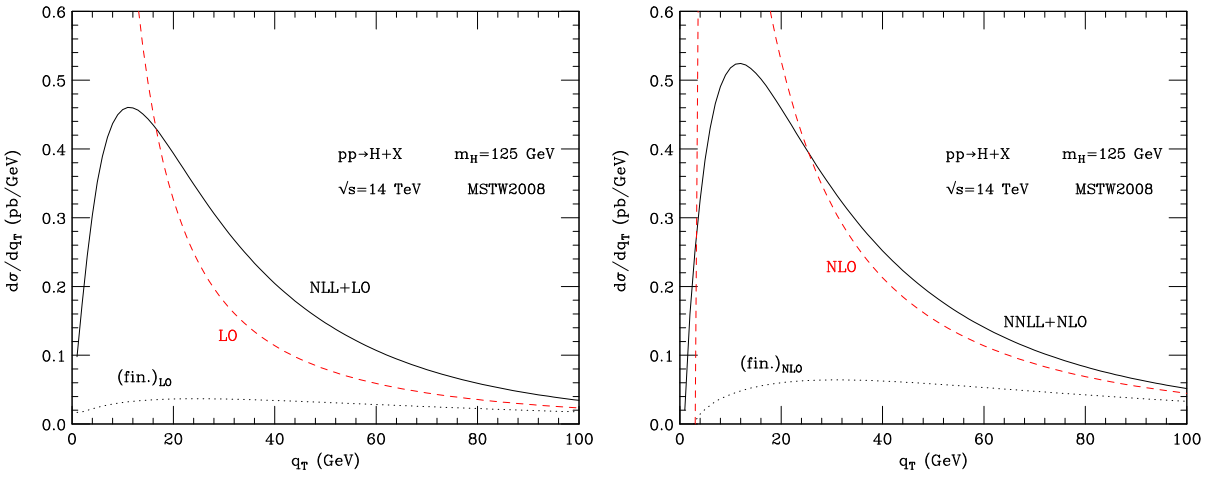


Figure 6.12: *Higgs boson gluon fusion cross section differential distributions as a function of the transverse momentum from the* HQT2 *program at $\sqrt{s} = 14$ TeV and $m_H = 125$ GeV. (left) LO fixed-order computation (red dashed) and computation at LO with next-leading-logs NLL corrections (solid black). (right) NLO fixed-order computation (red dashed) and computation at NLO with NNLL corrections (solid black). From Ref. [191].*

Therefore, a correction on the p_T spectra is needed in order to take into account the missing terms in the simulation. The signal events are reweighed to the Higgs p_T

spectrum computed at NLO plus next-to-next-to-leading logarithmic (NNLL) precision by the HQT [191,192] program using the heavy top quark mass approximation ($m_t = \infty$) and assuming that they solely contribute to the gluon fusion loop. Figure 6.13 shows the Higgs p_T distributions comparison between POWHEG and HQT.

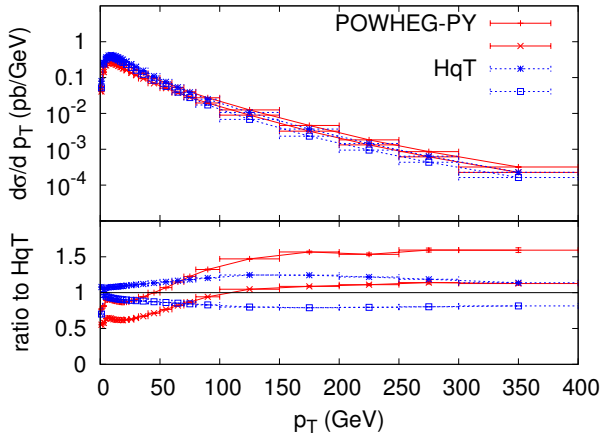


Figure 6.13: *Uncertainty bands of the Higgs boson transverse momentum spectrum in* POWHEG + PYTHIA *(red) compared to the* HQT *(blue), normalized to the* HQT *central value. From Ref. [59].*

In the current MSSM Higgs boson searches (see Ref. [162]), the uncertainty on the transverse momentum of h, H and A bosons produced via the gluon fusion process $gg \to \phi$ is not taken into account. The contributions from top, bottom, stop and sbottom quarks to the loop are sensitive to the Higgs coupling to up and down type quarks. This leads to an uncertainty on the $p_T(\tau)$ which can induce a model dependence through the acceptance cuts applied in the analysis. We will see in section 7.5.3 how this dependence can be treated. This constitutes a personal contribution to the consistency of the MSSM analysis.

6.5 Di-τ mass reconstruction

The observable used in the analysis is the reconstructed τ-pair mass. This variable allows to discriminate the Higgs signal ($\phi \to \tau\tau$) from the main irreducible $Z \to \tau\tau$ background. The inclusive cross section times branching fraction of the signal being orders of magnitude smaller then the $Z \to \tau\tau$ process, the effectiveness of the discrimination power leans on the separation of the two mass peaks.

This poses a real challenge due to the presence of neutrinos from the hadronic and leptonic τ-lepton decays. This leads to a kinematically under-constrained system. Two approaches have been used to deal with this. In a first approach, some experimental uncertainties are neglected from the calculation and some assumptions are made to simplify the problem. For example, the collinear approximation (see *e.g.* [193]) that we will describe in 6.5.1. The second type of approach consist on di-τ mass estimators capable to favor the best solution over the infinite (or multiple) cases. Every point in the phase space is

weighted by its *a priori* likelihood, which is evaluated from analytical parametrizations. For every event, the *a posteriori* likelihood of the di-τ mass is computed. The τ pair mass $M_{\tau\tau}$ in the analysis is reconstructed by a likelihood based algorithm called *SVfit* described in detail in [194]. Contrary to the collinear approximation, the *SVfit* algorithm is part of the so-called *Dynamical Likelihood Methods* [195,196], given that it reconstructs a kinematic quantity with likelihood approach in an event-by-event basis.

6.5.1 Collinear approximation

In this approach the τ pair kinematics are simplified assuming that all the decay products are aligned with the τ-lepton momentum and that all the missing energy originates from the neutrinos of the τ decay. This approximation is indeed good in the limit where the τ-lepton energy in the laboratory frame ($E_\tau = \gamma m_\tau$) is much larger than the τ mass (1.777 GeV), because the angle between the visible decay products in the laboratory is shrunk by a factor of $1/\gamma \ll 1$.

The total invisible momentum carried by the neutrinos from each of the τ-leptons are given by the equation system:

$$\begin{aligned} \not{E}_{Tx} &= p_1^{miss} \sin\theta_1^{vis} \cos\phi_1^{vis} + p_2^{miss} \sin\theta_2^{vis} \cos\phi_2^{vis} \\ \not{E}_{Ty} &= p_1^{miss} \sin\theta_1^{vis} \sin\phi_1^{vis} + p_2^{miss} \sin\theta_2^{vis} \sin\phi_2^{vis} \end{aligned} \tag{6.17}$$

where $\not{E}_{Tx}$ and $\not{E}_{Ty}$ are the x and y components of the $\not{E}_T$ vector, p_1^{miss} and p_2^{miss} are the combined invisible momenta of each τ decay, in the case where two neutrinos are produced, and $\theta_{1,2}$ and $\phi_{1,2}$ are the polar and azimuthal angles of the visible products of each τ decay.

If the matrix $A_{1,2}$ given by:

$$A_{1,2} = \begin{pmatrix} \sin\theta_1^{vis} \cos\phi_1^{vis} & \sin\theta_2^{vis} \cos\phi_2^{vis} \\ \sin\theta_1^{vis} \sin\phi_1^{vis} & \sin\theta_2^{vis} \sin\phi_2^{vis} \end{pmatrix} \tag{6.18}$$

can be inverted then the di-τ mass in the collinear approximation $M_{\tau\tau}^{CA}$ is given by:

$$\begin{aligned} M_{\tau\tau}^{CA} = [&2m_\tau^2 + 2\frac{p_1^{vis} p_2^{vis}}{x_1 x_2} \\ &\times (1 - (\cos\theta_1^{vis} \cos\theta_2^{vis} + \sin\theta_1^{vis} \cos\phi_1^{vis} \sin\theta_2^{vis} \cos\phi_2^{vis} \\ &+ \sin\theta_1^{vis} \sin\phi_1^{vis} \sin\theta_2^{vis} \sin\phi_2^{vis}))]^{1/2} \end{aligned} \tag{6.19}$$

where $x_{1,2} = p_{1,2}^{vis}/(p_{1,2}^{\nu} + p_{1,2}^{vis})$ are the fractions of the τ energy carried by the visible decay products. The equation system 6.17 cannot be solved if $\det|A_{1,2}| = 0$, so if:

$$\tan\phi_1^{vis} = \tan\phi_2^{vis} \Leftrightarrow (\phi_1^{vis} - \phi_2^{vis}) = k\pi, k \in Z \tag{6.20}$$

This means that the collinear approximation has no solution for back-to-back topologies. This is never exactly verified, because either the detector resolution or the presence of

extra radiation ($pp \to \phi + X$), always leads to $\phi_1^{vis} - \phi_2^{vis} \neq 0$. Unfortunately, the majority of $\phi \to \tau\tau$ events are produced with τ-leptons in nearly the back-to-back topology. And when $\phi_1^{vis} - \phi_2^{vis} \approx 0$ the solutions are biased giving unphysical large values, raising long tails in the $M_{\tau\tau}^{CA}$ distribution. This is clearly undesirable for low mass Higgs boson searches, where the tails of the much larger $Z \to \tau\tau$ irreducible background dominates the signal region. Therefore, this technique is applicable only to a relatively small fraction of events (50-60%).

6.5.2 τ-lepton decay kinematics parametrization

The τ pair mass is totally determined by the two four momenta of the visible and invisible tau decay products. The visible quantities as the energy, the polar and azimuthal angles as well as the visible decay products mass are measured precisely. The main experimental challenge faced is to reconstruct the four-momenta of the invisible particles. However it can be constrained by the $\not\!\!E_T$ and the τ-lepton mass m_τ obtained from the sum of the momentum of the visible and the invisible decay products.

As shown in section 4.1, in the leptonic τ decays two neutrinos are produced, while the hadronic decays leads to one neutrino. This implies that the invisible system is in general massive for the leptonic decays (resulting from $m^2 = (p_{\nu 1} + p_{\nu 2})^2$) and massless for the hadronic decays ($m^2 = p_\nu^2$). The τ decays can then be parametrized via:

- the opening angle θ defined as the angle between the boost direction of the τ-lepton and the momentum vector of the visible decay products in the rest frame of the τ.
- $\bar{\phi}$, the azimuthal angle of the τ in the CMS detector frame[2], i.e. the angle between $(\overrightarrow{p_{vis}}, \overrightarrow{p_\tau})$ and $(\overrightarrow{p_{vis}}, \overrightarrow{u_z})$[3].
- the invariant mass of the invisible momentum system $m_{\nu\nu}$, specific to the leptonic τ decays.

Then the energy and momentum of the visible decay products in the rest frame of the τ are given by:

$$E^{vis} = \frac{m_\tau^2 + m_{vis}^2 - m_{\nu\nu}^2}{2m_\tau}, \qquad p^{vis} = \sqrt{(E^{vis})^2 - m_{vis}^2} \tag{6.21}$$

with $m_{\nu\nu} = 0$ for the hadronic decays.

The opening angle in the CMS detector frame $\bar{\theta}$ are related to the corresponding quantities in the τ rest frame via the Lorentz invariant component of the visible momentum perpendicular to the τ-lepton direction:

$$p_\perp^{vis} = p_\perp^{\bar{vis}} \Rightarrow sin\bar{\theta} = \frac{p^{vis} \sin\theta}{p^{\bar{vis}}} \tag{6.22}$$

[2]Symbols with an overline refer to quantities defined in the laboratory frame.

[3]u_z is the unitary vector aligned with the proton beam axis.

And the energy in the laboratory frame $\bar{E}_\tau$ is given by $\bar{E}_\tau = \gamma m_\tau$, by determining the Lorentz boost factor γ of the component of the visible momentum parallel to the τ direction from the equations 6.21 and 6.22:

$$\begin{aligned} \bar{p}^{vis}\cos\bar{\theta} &= \gamma\beta E^{vis} + \gamma p^{vis}\cos\theta \\ \Rightarrow \gamma &= \frac{E^{vis}[(E^{vis})^2 + (\bar{p}^{vis}\cos\bar{\theta})^2 - (p^{vis}\cos\theta)^2]^{1/2} - p^{vis}\cos\theta\bar{p}^{vis}\cos\bar{\theta}}{(E^{vis})^2 - (p^{vis}\cos\theta)^2} \end{aligned} \tag{6.23}$$

So the energy of the τ in the laboratory frame as a function of the visible momentum $\bar{p}^{vis}$ depends only on two of the three parameters: the angle θ and the invariant mass of the neutrino system $m_{\nu\nu}$. The τ-lepton direction is within a cone of opening angle θ around the axis defined by the visible momentum. The direction of the four momentum is determined by the third parameter $\bar{\phi}$, the azimuthal angle of the τ-lepton with respect to the visible momentum vector. Figure 6.14 shows the relation between p_τ, p^{vis}, $\bar{\theta}$ and $\bar{\phi}$.

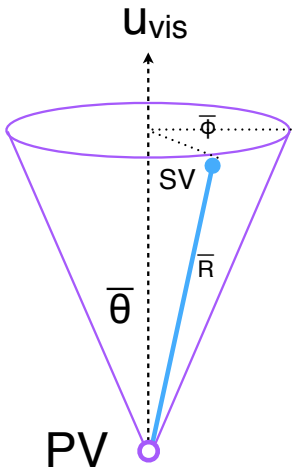


Figure 6.14: *Parametrization of the τ decay used in the SVfit algorithm.*

It is useful to make a transformation of the parameters space $(\theta, \bar{\phi}, m_{\nu\nu})$ into the alternative $(\bar{x}, \bar{\phi})$ parametrization, where $\bar{x}$ is the energy fraction carried by the visible decay products. They are related by the visible energy in the laboratory frame $\bar{E}_{vis}$, as a function of $p_{vis} = \sqrt{E_{vis}^2 - m_{vis}^2}$ and $\beta = \sqrt{\gamma^2 - 1/\gamma^2}$:

$$\cos\theta = \frac{\bar{E}_{vis} - \gamma E_{vis}}{\gamma\beta p_{vis}} \xrightarrow{\beta\to 1} \cos\theta = \frac{2\bar{x} - 1 - \frac{m^2_{vis(\nu\nu)}}{m^2_\tau}}{1 - \frac{m^2_{vis(\nu\nu)}}{m^2_\tau}} \tag{6.24}$$

for leptonic (hadronic) τ-lepton decays.

6.5.3 *SVfit* algorithm likelihood approach

The mass of the di-τ system is under constrained and its kinematics is controlled by 4 to 6 parameters depending on the leptonic or hadronic decays: the angles $\theta_1, \bar{\phi}_1$ and $\theta_2, \bar{\phi}_2$

as well as the masses $m_{\nu\nu}^{1,2}$ in the leptonic decays. The two components of the missing transverse momentum $\not{E}_{Tx}$ and $\not{E}_{Ty}$ provide two further constraints. The total number of free parameters is then $3 \times N(\tau_{lep}) + 2 \times N(\tau_h) - 2$.

We can then construct the likelihood function $f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2})$ describing the probability to reconstruct $\vec{z} = (\not{E}_{Tx}, \not{E}_{Ty})$, given the unknown parameters of the two τ decays taking the values $\vec{a_1} = (\theta_1, \bar{\phi}_1, m_{\nu\nu}^1)$, $\vec{a_2} = (\theta_2, \bar{\phi}_2, m_{\nu\nu}^2)$ and the four momenta of the visible decay products corresponding to the reconstructed values $\vec{y} = (p_1^{vis}, p_2^{vis})$. The mass of the di-tau system $m_{\tau\tau}(\vec{y}, \vec{a_1}, \vec{a_2})$ is a well defined function of $\vec{a_2}, \vec{a_2}$ and $\vec{y}$.

The *SVfit* algorithm strategy to find the best estimation of $\vec{a_1}$ and $\vec{a_2}$ is to test a series of di-τ mass hypothesis by computing the probability (so called *marginalization*):

$$P(m_{\tau\tau}^i) = \int \delta\left(m_{\tau\tau}^i - m_{\tau\tau}(\vec{y}, \vec{a_1}, \vec{a_2})\right) f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2}) d\vec{a_1} d\vec{a_2} \tag{6.25}$$

The integration is performed numerically using the VEGAS [197] algorithm. The values tested for $m_{\tau\tau}$ runs from m_τ to 2 TeV by steps of 2.5 GeV in the range [0, 100] GeV, followed by 2.5% spaced values up to 2 TeV. The likelihood value is evaluated for each hypothesis. The best estimate of $M_{\tau\tau}$ corresponds to the one which maximizes $P(m_{\tau\tau}^i)$.

The likelihood function $f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2})$ is the product of three likelihood functions. Two of them describe the decay of the two τ-leptons, the third one quantifies the compatibility of a τ decay hypothesis with the reconstructed $\not{E}_T$.

τ-lepton decay widths

It was found that the selection of the likelihood model makes little difference in the distribution of the reconstructed τ pair mass. The matrix element model for unpolarized τ-leptons decays [198] are used to model the leptonic decays:

$$\mathcal{L}_{\tau,lep} = \frac{d\Gamma}{d\bar{x} dm_{\nu\nu} d\bar{\phi}} \propto \frac{m_{\nu\nu}}{4m_\tau^2}[(m_\tau^2 + 2m_{\nu\nu}^2)(m_\tau^2 - m_{\nu\nu}^2)] \tag{6.26}$$

within the physically allowed region $0 \leq \bar{x} \leq 1$ and $0 \leq m_{\nu\nu} \leq m_\tau\sqrt{1-\bar{x}}$, taken from equation 6.24 ($-1 \leq \cos\theta \leq 1$ condition). While a model based on the two-body phase space for the hadronic decays [141] gives:

$$\mathcal{L}_{\tau,had} = \frac{d\Gamma}{d\bar{x} d\bar{\phi}} \propto \frac{1}{1 - m_{vis}^2/m_\tau^2} \tag{6.27}$$

within the physically allowed region $m_{vis}^2/m_\tau^2 \leq \bar{x} \leq 1$ from equation 6.24. The comparison of the kinematic distributions with respect to the detailed simulation using TAUOLA [176] shows that the two-body phase space model is well suited to describe the hadronic τ decays.

Missing transverse energy

Assuming that the only source of missing transverse energy are the neutrinos from the τ decays, then the sum of the neutrino momenta should match the reconstructed $\not{E}_T$ obtained from the particle flow reconstructed objects:

$$\sum p^{\nu}_{x,y} = \not{E}_{x,y} = -\sum p^{reco}_{x,y} \tag{6.28}$$

Each of the reconstructed particles has an uncertainty on the transverse momentum σ_{E_T} and its direction σ_ϕ. For the i^{th} particle reconstructed by the particle flow algorithm the individual covariance matrix is then:

$$U_i = \begin{pmatrix} \sigma_{E_T,i} & 0 \\ 0 & \sigma_{\phi,i} \end{pmatrix} \tag{6.29}$$

evaluated in the reference frame with axis parallel and perpendicular to $\overrightarrow{E_{T,i}}$. After a rotation into the global CMS frame the covariance matrices are summed together to built the global covariance matrix: $V = \sum_i R(\phi)U_iR(\phi)$.

To take into account the resolution effects that cause the differences between $\sum p^{\nu}_{x,y}$ and $\not{E}_{x,y}$, a Gaussian resolution model is taken for each of the reconstructed particle flow objects. The likelihood for observing $\not{E}_{x,y}$ given a true missing energy $\sum p^{\nu}_{x,y}$ is given by:

$$\mathcal{L}_\nu(\not{E}_x, \not{E}_y) = \frac{1}{2\pi\sqrt{|V|}} exp\left[-\frac{1}{2}\begin{pmatrix} \not{E}_x - \sum p^{\nu}_x \\ \not{E}_y - \sum p^{\nu}_y \end{pmatrix}^T V^{-1} \begin{pmatrix} \not{E}_x - \sum p^{\nu}_x \\ \not{E}_y - \sum p^{\nu}_y \end{pmatrix}\right] \tag{6.30}$$

where the expected $\not{E}_T$ resolution is represented by the covariance matrix V, estimated on an event-by-event basis using $\not{E}_T$ significance algorithm and $|V|$ is the determinant of this matrix.

Full likelihood

The full likelihood is then written by grouping the previous terms:

$$f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2}) \propto \sum_{i,j} \mathcal{L}_{\tau,i} \times \mathcal{L}_{\tau,j} \times \mathcal{L}_\nu(\not{E}_x, \not{E}_y) \tag{6.31}$$

where i,j denote the different possibles decay modes. Injecting this likelihood in equation 6.25, $P(m_{\tau\tau})$ is computed for all the mass hypothesis and the value which maximizes the probability is kept as the di-τ mass in an event-by-event basis.

6.5.4 Performance

In figures 6.15 and 6.16 the distributions of the visible mass (mass of the visible reconstructed decay products) and the *SVfit* mass are shown for simulated $Z \rightarrow \tau\tau$ and

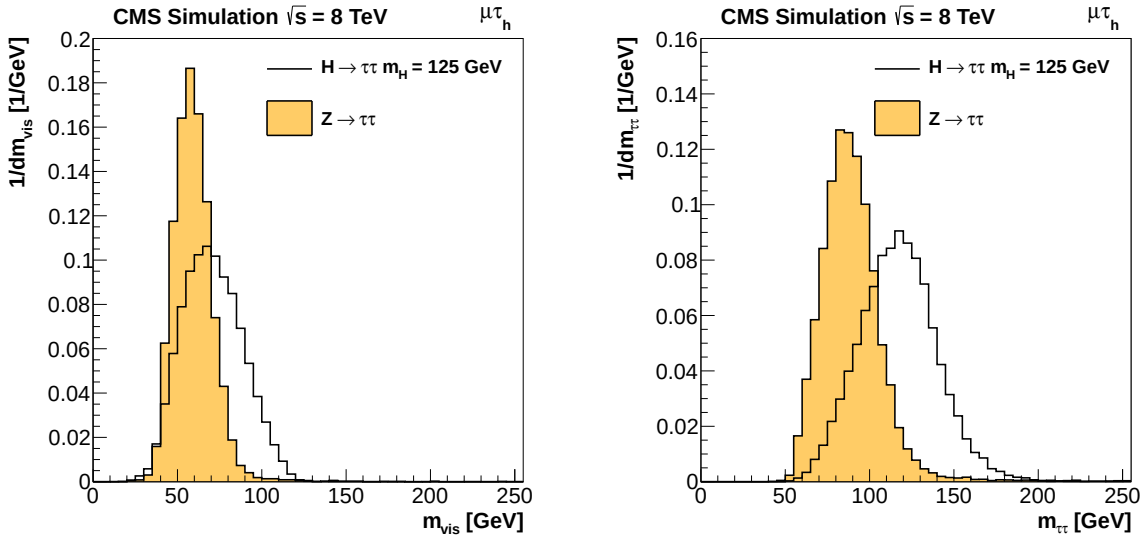


Figure 6.15: *Distribution of the τ pair visible mass (left) and the mass reconstructed by the SVfit algorithm (right) for simulated $Z \to \tau\tau \to \mu\tau_h$ background events and for the Standard Model Higgs boson signal $H \to \tau\tau \to \mu\tau_h$ with $m_H = 125$ GeV.*

different Higgs signals (Standard Model Higgs with $m_H = 125$ GeV on figure 6.15 and different mass hypothesis of MSSM Higgs signals on figure 6.16). This allows to compare the separation power of the two observables.

A better separation is achieved by the *SVfit* algorithm, increasing the analysis capabilities to find resonances on the background tails and to discriminate the overwhelming $Z \to \tau\tau$ background. It brings an improvement in the final expected significance of $\sim 40\%$ with respect to the visible di-τ mass observable.

The relative resolution on $m_{\tau\tau}$ has been estimated from simulation to be about 10% for double hadronic $\tau_h\tau_h$ decays, 15% in the semi-leptonic $\tau_h\tau_l$ decays, and 20% in the fully leptonic $\tau_l\tau_l$ decays, increasing with the number of neutrinos in the τ decay final states. Also the *SVfit* algorithm ensures a very high efficiency, with a failure rate at the per-mill level, even if it is a CPU-intensive algorithm with an average computing time on the order of $\mathcal{O}(1s)/event$.

6.6 Statistical tools

In this section, I will present the statistical procedure used in both the Standard Model and MSSM Higgs boson searches. The signal extraction is performed by a binned maximum likelihood fit on the *SVfit* mass distribution previously described for the *signal plus background* (H_1) and *background only* (H_0) hypothesis. The fit is done simultaneously over all the categories and channels. The statistical approach used is a modified frequentist CL_S method [199], as recommended by the LHC Higgs Combination Group [200,201]. This approach is different from the ones used before for Higgs searches at LEP and Teva-

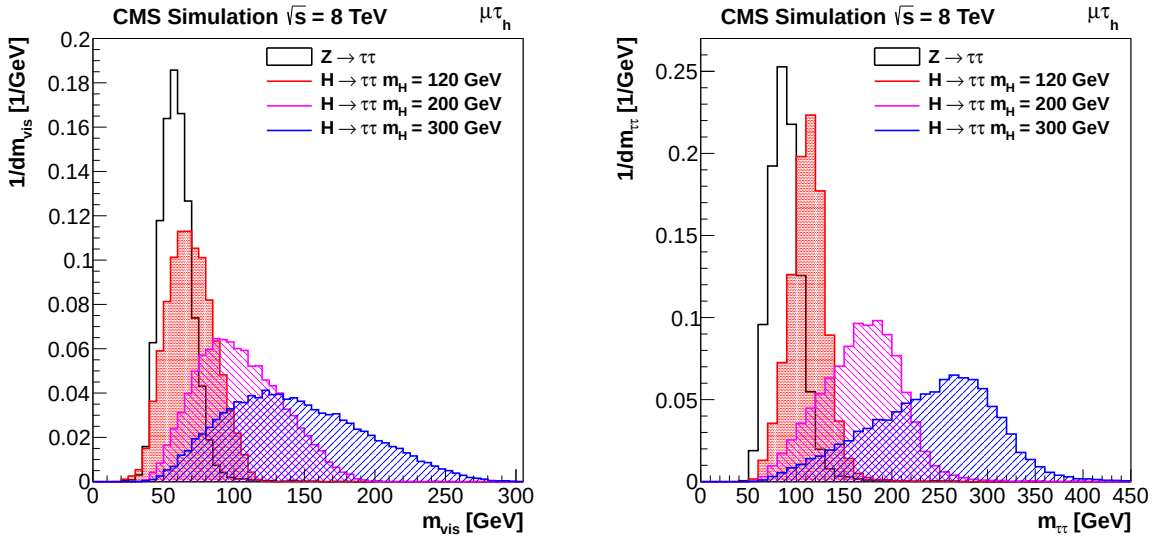


Figure 6.16: *Distribution of the τ pair visible mass (left) and the mass reconstructed by the SVfit algorithm (right) for simulated $Z \to \tau\tau \to \mu\tau_h$ background events and hypothetical MSSM $\phi \to \tau\tau \to \mu\tau_h$ signals of $m_A = 120$, 200 and 300 GeV.*

tron. LHC modified frequentist limits are obtained with a test statistic based on a profile likelihood ratio. The combination tool used is implemented in COMBINE [202], the CMS Higgs combination tool implemented using the RooStats [203] framework.

Likelihood model

The result of an observation is an array of measured data yields over the bins of all mass histograms. Data can be the real data recorded by the experiment but also may refer to pseudo-data simulated in a "toy" experiment. The expected number of events in the i^{th} bin of the distribution can be written as:

$$\nu_i(\mu, \vec{\theta}) = \mu s_i(\vec{\theta}) + b_i(\vec{\theta})$$

where s_i is the expected signal yield and b_i the expected background yield. Both depend on a set of nuisance parameters $\vec{\theta}$ (for example the systematic errors). The parameter μ is called the *signal strength modifier* which scales the signal rate of all the production modes by the same amount. It is a free parameter in the fit. The *background only* hypothesis then corresponds to $\mu = 0$ and the *signal plus background* hypothesis to $\mu = 1$.

The likelihood function $\mathcal{L}(n|\mu s(\vec{\theta}) + b(\vec{\theta}))$, the probability of observing n events in this model, is given by the product of the $Poisson(n_i|\nu_i(\mu, \vec{\theta}))$ probabilities for each bin:

$$\begin{aligned} \mathcal{L}(n|\mu s(\vec{\theta}) + b(\vec{\theta})) &= \prod_i Poisson(n_i|\nu_i(\mu, \vec{\theta})) \cdot \prod_j \rho(\theta_j, \hat{\theta}_j) \\ &= \prod_i \frac{\mu s_i(\vec{\theta}) + b_i(\vec{\theta})}{n_i!} e^{-[\mu s_i(\vec{\theta}) + b_i(\vec{\theta})]} \cdot \prod_j \rho(\theta_j, \hat{\theta}_j) \end{aligned} \tag{6.32}$$

where the component θ_j of the vector $\vec{\theta}$ represents one of the nuisance parameters assigned to one systematic uncertainty of the analysis that we will discuss in the next chapter (*c.f.* section 7.4). The nuisance parameters $\vec{\theta}$ are constrained by auxiliary measurements that restrict their values within confidence intervals. The probability $\rho(\theta_j, \hat{\theta}_j)$ represents the probability for the true value of the nuisance parameter to be equal to θ_j given that the best estimate for this parameter is $\hat{\theta}_j$. Hence, the second product in the equation estimates the knowledge we have on the value of the nuisance parameters. The nuisance parameters are then incorporated as free parameters in the likelihood ratio, and are fitted on data simultaneously to μ (the parameter of interest). Their values after fit can indicate a mis-modeling of a nuisance, a background or a sub-estimation of a systematic if they are away of the initial value. Let us discuss the different types of nuisance parameters entering in the likelihood.

Yield uncertainties

A first type of nuisance parameters are the ones that affect the event yields. They can be modeled by two kinds of probability density functions: log-normal or Gamma distributions.

The uncertainties corresponding to normalization nuisances affecting the signal or background yields are modeled with a log-normal constraints:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\hat{\theta})^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta} \tag{6.33}$$

with $\kappa = 1+\epsilon$ (ϵ being relative scale of the uncertainty). For example, a value of $\kappa = 1.10$ is chosen in the case where the value of the nuisance is known with an uncertainty of 10% with respect to its nominal value. Examples of this type of nuisance can be the cross sections uncertainties, the identification and isolation efficiencies, the contamination of fake rates or the uncertainties on the multiplicative extrapolation factors used for determining background contributions. A single nuisance parameter θ_j can affect many processes, the signal and/or background yields in many channels and categories, and in general with different scales depending on the category and the channel. The errors originating from the same nuisance parameter are taken as fully correlated or fully anti-correlated by choosing the relative sign on the definition of κ ($\kappa_j = 1 \pm \theta_j$).

The second type of normalization nuisances can come from auxiliary measurements, for example event counting in sidebands which constrain the background rates in the signal region. The uncertainties have a statistical origin and are modeled by a Gamma distribution:

$$\rho(n) = \frac{1}{\alpha}\frac{(n/\alpha)^N}{N!}\exp(-n/\alpha) \tag{6.34}$$

where N is the number of events in the control region which is used to estimate the number n of events in the signal region using the extrapolation factor α by the relation $n = N\alpha$. The α parameter can be at the same time affected by a nuisance with an uncertainty modeled by a log-normal constraint.

Shape uncertainties

Another type of nuisance parameters taken in consideration in the analysis are sensitive to the scale of a variable correlated to the di-τ mass leading to a shape template uncertainty on the *SVfit* mass distribution. They are taken into account by using a "vertical template morphing" technique. For each shape nuisance, the templates (histograms) corresponding to ± 1 standard deviations are generated by scaling down or up the given nuisance. They are associated to the values $\lambda = +1$ (up) and $\lambda = -1$ (down). The parameter λ is added in the likelihood model to be interpolated between the nominal template ($\lambda = 0$) and the two deviated templates. The effect of all the shape nuisances are additive. By noting h_0 the nominal template and $h_j^{\pm}$ the shifted (up/down) templates after varying the nuisance θ_j by $\pm 1\sigma$, we have the final template formula:

$$h(\lambda) = h_0 + \sum_j \left(a(\lambda_j) h_j^+ + b(\lambda_j) h_0 + c(\lambda_j) h_j^- \right) \tag{6.35}$$

The set of templates are defined by a quadratic interpolation between the up and down template for $|\lambda| < 1$, and linear beyond:

$$a(\lambda_j) = \begin{cases} \lambda_j \cdot (\lambda_j + 1)/2 & \text{if} \quad |\lambda_j| \leq 1 \\ 0 & \text{if} \quad \lambda_j < -1 \\ \lambda_j & \text{if} \quad \lambda_j > +1 \end{cases} \quad b(\lambda_j) = \begin{cases} -\lambda_j^2 & \text{if} \quad |\lambda_j| \leq 1 \\ -(|\lambda_j| - 1) & \text{if} \quad |\lambda_j| > 1 \end{cases} \quad c(\lambda_j) = \begin{cases} \lambda_j \cdot (\lambda_j - 1)/2 & \text{if} \quad |\lambda_j| \leq 1 \\ 0 & \text{if} \quad \lambda_j > +1 \\ \lambda_j & \text{if} \quad \lambda_j < -1 \end{cases}$$

The *pdf* associated to λ is taken to be the normal distribution with mean 0 and $\sigma = 1$.

6.6.1 Test statistic

In order to test the compatibility between the result of an experiment and the H_0 or H_1 hypothesis, we can build the test statistic q_μ from the profile likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \tag{6.36}$$

with $\hat{\mu}$ and $\hat{\vec{\theta}}$ the combination of values maximizing $\mathcal{L}(\mu, \vec{\theta})$, μ the tested signal strength and $\hat{\hat{\vec{\theta}}}_\mu$ the value of $\vec{\theta}$ maximizing $\mathcal{L}$ for this value of μ. This equation is complemented by the constraint:

$$0 \leq \hat{\mu} \leq \mu \tag{6.37}$$

ensuring a physical meaning of the result, the signal rate is then positive ($\hat{\mu} \geq 0$), and that one-sided confidence intervals are obtained ($\hat{\mu} \leq \mu$) when setting limits on μ. The test statistic q_μ is then defined as:

$$q_\mu = -2 \ln \lambda(\mu) = \begin{cases} -2 \ln \dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(0, \hat{\hat{\vec{\theta}}}_0)} & \text{if} \quad \hat{\mu} < 0 \\ -2 \ln \dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} & \text{if} \quad 0 \leq \hat{\mu} \leq \mu \\ 0 \quad \text{if} \quad \mu < \hat{\mu} \end{cases} \tag{6.38}$$

By construction, q_μ is non-negative. It is a measure of the incompatibility between data and the $(\mu s + b)$ hypothesis. The higher it is, the less data is compatible with the hypothesis on μ.

6.6.2 Limit setting procedure

In order to establish confidence levels on μ, the probability density functions $f(q_{\mu|\mu' s+b})$ as a function of μ', or sampling distribution of q_μ is used. It is computed for both H_0 and H_1 hypothesis using the distribution of the test statistic q_μ^{toy} computed using a set of pseudo-datasets generated for the corresponding hypothesis. The p-value is then defined as the probability of observing data of equal or greater incompatibility with the predictions of the given hypothesis. It can be evaluated for the signal plus background hypothesis (p_μ) and in the background only hypothesis $(1-p_b)$ given the actual observed value of the test statistic q_μ^{obs}:

$$\begin{aligned} p_\mu &= \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\mu)\mathrm{d}q_\mu \\ 1-p_b &= \int_{q_\mu^{obs}}^{\infty} f(q_\mu|0)\mathrm{d}q_\mu \end{aligned} \tag{6.39}$$

In order to set limits on the cross section, we can use the frequentist CL_{s+b} method, where the confidence level is defined as $CL_{s+b} = 1-p_\mu$. If $CL_{s+b} > 0.95$ the cross section arising from the signal strength μ is excluded at 95% confidence level. In the case where the signal yield is very small compared to the background one, CL_{s+b} can become very small due to a negative background fluctuation, which could lead to an exclusion of a μ value at 95% C.L. not originated by the absence of signal. The alternative used is the CL_s method defined as:

$$CL_s = \frac{p_\mu}{1-p_b} = \frac{CL_{s+b}}{CL_b} \tag{6.40}$$

The CL_s method is more conservative than CL_{s+b}. As $CL_b = 1-p_b < 1$ it will reject less μ hypothesis $(CL_s < CL_{s+b})$. It is also more robust against background fluctuations. The signal strength modifier μ is said to be excluded at a confidence level $1-\alpha$, if CL_s is equal to α. The sensitivity of the experiment is reported by the median expected exclusion limit on μ in the background only hypothesis, μ_{exp}^{95} , together with the intervals where μ_{obs}^{95} is expected to lie in the 68% (1σ) and 95% (2σ) of the cases.

6.6.3 Discovery significance

Observation can be found to be incompatible with the background only hypothesis due to an excess of data. In order to establish a discovery, we have to rule out the null hypothesis. The test statistic used to measure the deviation to the background only hypothesis is q_0:

$$q_0 = \begin{cases} -2\ln\dfrac{\mathcal{L}(0,\hat{\vec{\theta}}_0)}{\mathcal{L}(\hat{\mu},\hat{\vec{\theta}})} & \text{if} \quad \hat{\mu} \geq 0 \\ 0 \quad \text{if} \quad \hat{\mu} < 0 \end{cases} \tag{6.41}$$

In the case where $\hat{\mu} < 0$ the test statistic is zero as this case happens when the background fluctuates negatively, as the aim is not to characterize data deficits but to discover signal. Large values of q_0 will be found in the case of large deviations from the H_0 hypothesis. In order to quantify this deviation we introduce the p-value defined as:

$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0|\mu = 0)\mathrm{d}q_0 \tag{6.42}$$

where q_0^{obs} is the observed value of q_0 in data and $f(q_0|\mu = 0)$ the *pdf* of q_0 given the background only hypothesis. p_0 is the probability that the observation is provoked by a fluctuation of the background only hypothesis. In high energy physics, the significance of p_0 is often reported in the form of a gaussian probability $Z = \phi^{-1}(1 - p_0)$, where ϕ is the inverse cumulative function of the normal distribution[4]. By convention, an evidence is claimed when $Z > 3\sigma$ and a discovery when $Z > 5\sigma$. The relation between p_0 and Z for some typical values is displayed in table 6.14.

Significance (Z)	p-value (p_0)
1σ	$1.586 \cdot 10^{-1}$
2σ	$2.228 \cdot 10^{-2}$
3σ	$1.350 \cdot 10^{-3}$
4σ	$3.167 \cdot 10^{-5}$
5σ	$2.867 \cdot 10^{-7}$
6σ	$9.866 \cdot 10^{-10}$
7σ	$1.280 \cdot 10^{-12}$

Table 6.14: *Relation between the p-value and the significance.*

[4] $\phi^{-1} = \sqrt{2} \cdot \mathrm{erf}^{-1}(2x - 1), \quad x \in [0, 1]$

Chapter 7

Standard Model and MSSM H $\rightarrow \tau\tau$ searches in CMS

Chapter content

As described in chapter 4, the τ-lepton can decay either leptonically or to hadrons (τ_h decaying usually to pions). CMS has covered all the six possible channel combinations: leptonic $ee, \mu\mu, e\mu$, semi-leptonic $\mu\tau_h$ and $e\tau_h$ and double hadronic $\tau_h\tau_h$ [161]. In this thesis I describe the analysis of the complete pp collisions data recorded by CMS in 2012 at $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of 19.4 fb^{-1} for the semi-leptonic channels $\mu\tau_h$ and $e\tau_h$. These are the most sensitive channels thanks to the efficient electron and muon reconstruction and the branching ratio of the hadronic tau decay. The analysis in the other decay channels and using data collected in 2011 $\sqrt{s} = 8$ TeV is detailed in Ref. [161].

The production modes studied are the VBF and gluon fusion that are exploited using an event categorization aiming to tag them. The production of the Higgs boson in association with a W, Z or a pair of heavy quarks, is an exclusive analysis described elsewhere [204] and combined in the final result.

In section 2.2 we pointed out a good motivation for the search of the MSSM Higgs boson in the di-τ final state due to its enhanced couplings to down-type fermions. As the search of the MSSM Higgs boson extends to higher masses (up to 1 TeV), the analysis is expected to be strongly enhanced using the novel τ_h isolation described in section 4.5 due to the use of the τ lifetime information. The analysis uses also event categories aiming at tagging the two production modes: gluon fusion and associated production with b-quarks.

In this chapter, I present some original contributions leading to a significant improvement of the SM and MSSM Higgs bosons searches. The current MSSM analysis is reported here as well [162]. In section 7.1 the baseline selection for both analyses is presented. An analysis with specific categorization aiming at tagging the Higgs boson production modes is done. Exclusive categories are defined to increase the signal purity, as described in section 7.2. Then, reducible and irreducible backgrounds are modeled and estimated using simulation or data-driven methods, as described in section 7.3. The theoretical and experimental systematic uncertainties are presented in section 7.4. Section 7.5 is dedicated to summarizing my personal contributions to the analysis. Finally the last two sections, 7.6 and 7.7, show the yields, the statistical interpretation and the results of the Higgs boson searches in the Standard Model and the MSSM sectors.

7.1 Baseline event selection

The analysis baseline di-τ selection selects one signal candidate per event. It ensures that all the selected events pass a quality criteria and suppresses the reducible background. The objects used to apply the pre-selection are the ones as described in 6.3. Then channel-specific and common selection criteria are applied.

$\mathbf{e\tau_h}$ channel

The events containing a reconstructed $e\tau_h$ pair are selected requiring the following criteria:

- The events have to be recorded after firing any of the $e\tau_h$ trigger paths listed in table 6.1;

- One electron passing the Tight working point of the MVA-based identification criteria(*c.f.* section 6.3.2), matching the HLT electron filter object of the corresponding trigger path by $\Delta R = 0.5$ is required. It should be such that $p_T(e) > 24$ GeV and $|\eta(e)| < 2.1$ and pass the isolation cut $I_{rel} < 0.1$ (*c.f.* equation 6.5);

- One hadronic decaying τ lepton passing the Tight working point of the MVA-based isolation criteria (*c.f.* section 6.3.1), matching the HLT tau filter object of the corresponding trigger path by $\Delta R = 0.5$, is required. It should be such that $p_T(\tau_h) > 30$ GeV and $|\eta(\tau_h)| < 2.1$. It is also required to pass the Medium working point of the MVA-based anti-electron discriminator and the Loose working point of the cut-based anti-muon discriminator in order to reduce the $e \to \tau_h$ and $\mu \to \tau_h$ fake rates;

- The event is vetoed if it contains a second electron with $p_T(e) > 15$ GeV and $|\eta(e)| < 2.4$ passing the "Veto" (*c.f.* section 6.3.2) working point identification criteria and passing the isolation cut $I_{rel} < 0.3$. This requirement is used in order to further suppress the $Z/\gamma^* \to ee$ background.

$\mu\tau_\mathbf{h}$ channel

The events containing a reconstructed $\mu\tau_h$ pair are selected requiring the following criteria:

- The event have to be recorded after firing any of the $\mu\tau_h$ trigger paths listed in table 6.1;

- One muon passing the Tight working point identification criteria (*c.f.* section 6.3.2), matching the HLT muon filter object of the corresponding trigger path by $\Delta R = 0.5$, is required. It should be such that $p_T(\mu) > 20$ GeV and $|\eta(\mu)| < 2.1$ and pass the isolation cut $I_{rel} < 0.1$ (*c.f.* equation 6.5);

- One hadronically decaying τ lepton passing the Tight working point of the MVA-based isolation criteria (*c.f.*section 6.3.1), and matching the HLT tau filter object of the corresponding trigger path by $\Delta R = 0.5$ is required. It should be such that $p_T(\tau_h) > 30$ GeV and $|\eta(\tau_h)| < 2.1$. It is also required to pass the Tight working point of the MVA-based anti-muon discriminator and the Loose working point of the cut-based anti-electron discriminator in order to reduce the $\mu \to \tau_h$ and $e \to \tau_h$ fake rates;

- The event is vetoed if it contains a second muon with $p_T(\mu) > 15$ GeV and $|\eta(\mu)| < 2.4$ passing the passing the Loose working point identification criteria and passing the isolation cut $I_{rel} < 0.3$. This requirement is used in order to further suppress the $Z/\gamma^* \to \mu\mu$ background.

Common criteria

For both channels, further requirements have to be checked for:

- The primary vertex has to pass the quality criteria described in section 3.3.1;
- $lep + \tau_h$ ($lep = e$ *or* μ) pairs are required to be of opposite charge;
- If many pairs pass the previous criteria, the one with the highest scalar sum of the lepton and τ_h momenta, $p_T(lep) + p_T(\tau_h)$, is chosen as the signal candidate;
- In order to reduce the $W + jets$ background, the transverse mass of the lepton and the missing transverse energy is required to be such that $M_T(lep, \not{E}_T) < 30$ GeV. For simulated events, the MVA-based recoil corrected missing transverse energy is used (*c.f.* sections 6.3.4 and 6.4.8);
- Additionally to the second electron or muon veto, a third lepton veto is applied in both channels requiring the selected events to contain no additional loosely identified taus, electrons or muons. This requirement suppresses background events and avoids the overlap with the VH analysis [204].

7.2 Event categorization

In order to exploit the signatures of the main Higgs production mechanisms at the LHC and to further suppress Standard Model background processes, events passing the baseline selection are classified into exclusive event categories. The categorization of the events into exclusive categories aims at improving the statistical sensitivity to a Higgs boson. The gain is maximized when the events are sorted in categories with very different signal over background ratios as:

$$\frac{S}{\sqrt{B}} < \sqrt{\left(\frac{S_1}{\sqrt{B_1}}\right)^2 + \left(\frac{S_2}{\sqrt{B_2}}\right)^2}$$

with

$$S = S_1 + S_2$$
$$B = B_1 + B_2$$

for $S_1/B_1 \neq S_2/B_2$. The categories with high S/B will then be very sensitive to the Higgs boson and the categories with low S/B will serve to constrain the backgrounds and systematic uncertainties.

The categories definition are tuned by maximizing the sensitivity of the analysis based on the expected exclusion limit on the Higgs cross section. A simultaneous maximum

likelihood fit is performed over all the categories and channels. The categories populated mostly by background events strongly constrain the nuisance parameters on the fit, while the signal is primarily extracted from the most sensitive ones.

The event jet content and topology is used to tag different Higgs production modes. Also the categorization aims at separating events with good resolution on the *SVfit* mass from events with poor di-τ mass resolution. Therefore, requirements on the $p_T(\tau_h)$ and the presence of an additional jet, or even requirements on the Higgs candidate p_T (in the Standard Model case) are used.

7.2.1 Standard Model categories

In the Standard Model case, the categories are defined in order to maximize the sensitivity to the presence of a Standard Model Higgs boson with a mass m_H laying between 110 and 145 GeV.

The two main production modes of the Higgs boson at the LHC is the gluon fusion process (GGF) and the vector boson fusion (VBF). In order to classify the events by the number of reconstructed jets, jets are required to satisfy:

- $p_T^{jet} > 30$ GeV;
- $|\eta^{jet}| < 4.7$;
- jets well apart from leptons: $\Delta R(jet - (lep/\tau_h)) > 0.5$.

In order to reduce the $t\bar{t}$ contamination, events with b-tagged jets with $p_T^{jet} > 20$ GeV are rejected in all the categories. The categories used in the Standard Model analysis are defined as follows:

- **VBF or 2-jets categories**

  The vector boson fusion process (VBF) is exploited in the analysis by its specific signature in the detector. In this mechanism, the two quark induced jets ("tag" jets) have a large invariant mass and are very separated in pseudorapidity (very forward jets). As it is a pure electroweak process, the QCD activity is suppressed in the central region. The kinematical cuts on the "tag" jets and a central jet veto is used to strongly reduce the background contamination coming mainly from gluon-initiated jets from initial state radiation.

  Events are required to contain at least two "tag" jets with $p_T(jet) > 30$ GeV and $|\eta| < 4.7$. A jet veto is applied, the event is rejected if a third jet with $p_T(jet) > 30$ GeV is reconstructed between the two "tag" jets. A dedicated sub-set of categories is defined to tag the most VBF-like events (tight VBF tag).

  tight VBF tag: $m_{jj} > 700$ GeV, $|\Delta\eta| > 4.0$ and $p_T(\tau\tau) > 100$ GeV
  loose VBF tag: Not tight VBF tag, $m_{jj} > 500$ GeV and $|\Delta\eta| > 3.5$

  Events failing the cuts on the VBF tag categories are recovered in the 1-jet or 0-jet categories depending on the number of reconstructed jets in the event.

- **1-jet categories**

  Events not entering the VBF category and containing at least one jet with $p_T(jet) > 30$ GeV and $|\eta| < 4.7$ and no b-tagged jets with $p_T^{jet} > 20$ GeV enter the 1-jet category[1]. The resolution on the reconstructed di-τ mass is enhanced when the τ_h has a significant transverse momentum as its decay products become more collimated. This category is split in two sub-categories in function of the τ_h transverse momentum:

  $$\textbf{1-jet medium: } 30 \leq p_T(\tau_h) \leq 45 \text{ GeV}$$
  $$\textbf{1-jet high: } p_T(\tau_h) > 45 \text{ GeV}$$

  Also the p_T of the reconstructed Higgs boson candidate can be used to improve the signal-to-background separation. It can be reconstructed by summing the momenta of the visible tau decays products (electron, muon and τ_h) plus the missing missing transverse energy reconstructed in the event. The Higgs candidate p_T is defined by:

  $$p_T(\tau\tau) = |\overrightarrow{p_T}(lep) + \overrightarrow{p_T}(\tau_h) + \overrightarrow{\not{E}_T}| \tag{7.1}$$

  In boosted topologies, the reduction of the backgrounds plus the improvement of the resolution on the di-τ mass purity the selection and enhance the expected significance. The 1-jet high category is also divided in two sub-categories (denoted with the arrow below) depending on the boost of the Higgs candidate, thus enhancing the signal purity.

  $$\textbf{1-jet high } \downarrow\textbf{: } p_T(\tau\tau) \leq 100 \text{ GeV}$$
  $$\textbf{1-jet high } \uparrow\textbf{: } p_T(\tau\tau) > 100 \text{ GeV}$$

  An additional cut on the transverse missing energy $\not{E}_T > 30$ GeV considerably reduces the $Z/\gamma^* \to ee$ contamination in the $e\tau_h$ channel. This extra requirement makes it difficult to predict the *SVfit* distribution for the $Z/\gamma^* \to \tau\tau$ background events in the 1-jet high $\uparrow$ category. Therefore this category is not used in this channel.

- **0-jet categories**

  Events not entering the VBF neither the 1-jet categories or containing no jets with $p_T(jet) > 30$ GeV and $|\eta| < 4.7$ and no b-tagged jets with $p_T^{jet} > 20$ GeV. Similarly, this category is split in two sub-categories as a function of the τ_h transverse momentum:

  $$\textbf{0-jet medium: } 30 \leq p_T(\tau_h) \leq 45 \text{ GeV}$$
  $$\textbf{0-jet high: } p_T(\tau_h) > 45 \text{ GeV}$$

  This categories are dominated by background events. They are used to constrain the nuisance parameters of the simultaneous fit. As we will see in section 7.4, most of the nuisance parameters are correlated between categories (trigger efficiencies, lepton identification and isolation, τ_h energy scale, etc.). This category being the most populated it drives the constrains.

[1] Boosted Higgs is also a feature of the VBF mechanism, VBF events failing the VBF category cuts will then be recovered here.

In total, in the $\mu\tau_h$ channel 7 categories and in the $e\tau_h$ channel 6 categories are fitted to extract the signal. In figure 7.1 the event categories are summarized. An event display of a selected candidate entering the VBF tagged categories is presented in figure 7.2.

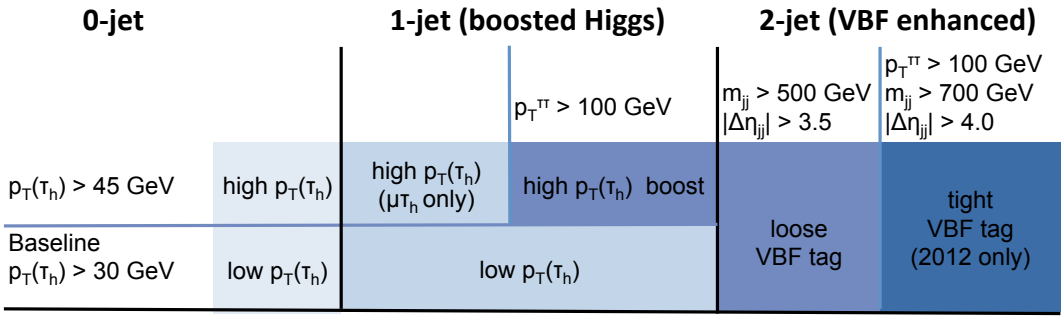


Figure 7.1: *Event categories for the semi-leptonic channels. In the 1-jet category of the* $e\tau_h$ *channel the events are required to have* $\not{E}_T > 30$ *GeV.*

7.2.2 MSSM categories based on the number of b-tagged jets

In the case of the search for MSSM Higgs bosons, the two main production modes at the LHC are gluon fusion process and the associated production with bottom quarks. The associated production with b-quarks leads to a final state with b-tagged jets. This can be used to enhance the signal purity. The categories are defined as follows using the output of the CSV b-Tag discriminator (see section 6.3.3).

- **b-Tag** Events containing at most one jet with $p_T(jet) > 30$ GeV and $|\eta| < 4.7$ and at least one b-tagged jet (Medium working point CSV output above 0.679) with $p_T(jet) > 20$ GeV.
- **No b-Tag** Events not containing a b-tagged jet (Medium working point CSV output above 0.679) with $p_T(jet) > 20$ GeV.

Here the "b-tag" category is the most sensitive and the "no b-tag" category is used mainly to constrain nuisance parameters of the fit. In section 7.5.2 we will show how this categorization can be improved.

7.3 Background estimation

The estimation of the contribution of the main backgrounds in each channels is done using *data-driven* methods. The most important backgrounds that we will describe next are the irreducible $Z/\gamma^* \to \tau\tau$ Drell-Yan production, QCD multi-jets, $W + jets$ production, Drell-Yan $Z/\gamma^* \to ee/\mu\mu$, di-boson production, single top production and $t\bar{t}$ production.

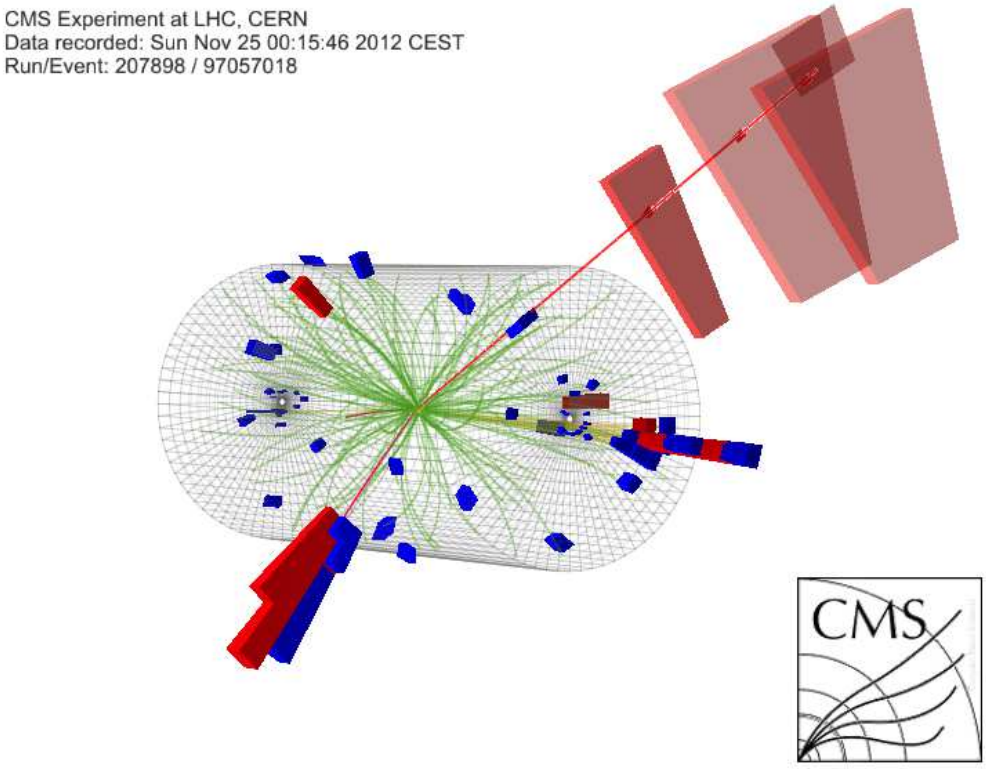


Figure 7.2: *3D view of a $H \to \tau\tau$ VBF event candidate in the $\mu\tau_h$ channel, recorded by CMS during 2012. The muon, with hits visible in the end-cap muon chambers, has a p_T of 31.9 GeV. The hadronically-decaying tau candidate, indicated by the red and blue towers in this figure, has a pt of 44.0 GeV. The di-τ mass, calculated using the SVfit algorithm, is 120.3 GeV. The two jets passing the VBF selection have p_T 80.5 and 36.3 GeV, are separated in eta by 6.2 and have a mass of 1.2 TeV. There are no additional jets in the eta range between these tagging jets with a pt above 30 GeV.*

7.3.1 Irreducible $Z/\gamma^* \to \tau\tau$

The largest source of background in the analysis is the irreducible Drell-Yan production of $Z/\gamma^* \to \tau\tau$. The final state contains two real τ leptons. This contribution is strongly reduced by the analysis categories selection: the jet multiplicity in the Drell-Yan production decreases steeply, so requiring jets or requiring the VBF topology are good handles to reject this background.

Shape

The mass shape is modeled using the embedded technique described in section 6.2.3. This background normalization, efficiency and acceptance is derived using simulated Drell-Yan $Z/\gamma^* \to \tau\tau$ Monte Carlo. The full set of data to simulation corrections (*c.f.* 6.4) and the baseline selection described in 7.1 are applied.

Normalization

In order to extract the normalization in each category, an extrapolation factor $\epsilon^{Emb,cat}_{Z\to\tau\tau}$ from the baseline to the category selection is applied. The efficiency to pass the category cuts is taken from the $Z/\gamma^* \to \tau\tau$ embedded samples, so the normalization for a given category is:

$$N^{cat}_{Z\to\tau\tau} = N^{MC,incl}_{Z\to\tau\tau} \times \epsilon^{Emb,cat}_{Z\to\tau\tau}$$
$$\epsilon^{Emb,cat}_{Z\to\tau\tau} = \frac{N^{Emb,cat}_{Z\to\tau\tau}}{N^{Emb,incl}_{Z\to\tau\tau}}$$

with $N^{MC,incl}_{Z\to\tau\tau}$ the normalization taken from simulation with the baseline selection without $m_T(lep, \not{E}_T)$ cut, $N^{Emb,incl}_{Z\to\tau\tau}$ the number of events passing the same baseline selection using the embedded sample and $N^{Emb,incl}_{Z\to\tau\tau}$ the number of events passing the category selection using the embedded sample.

7.3.2 $W + jets$

The $W + jets$ production where the W boson decays leptonically constitutes a major background in the semi-leptonic channels. Typically, the reconstructed electron or muon comes from the W and the reconstructed τ_h is due to a jet misidentified as a τ_h.

Shape

The *SVfit* mass shape is modeled using Monte Carlo simulation for each category. The shapes are corrected for the difference in the $p_T(\tau_h)$ distributions between simulation and data as described in 6.4.11. Smooth shapes are obtained in the VBF tagged, 1-jet, and b-tagged categories by relaxing some selection criteria. The criteria depends on the category:

- VBF tight tag: the opposite sign requirement between the lepton and the τ_h is relaxed and the Loose working point for the τ_h isolation criteria is used.
- VBF loose tag: the VBF selection criteria is relaxed and simply requires $M_{jj} >$ 200 GeV, $|\Delta\eta| > 2.0$ and the central jet veto. The opposite sign requirement between the lepton and the τ_h is also removed.
- 1-jet low: the opposite sign requirement between the lepton and the τ_h is removed.
- 1-jet high $\downarrow$: the opposite sign requirement between the lepton and the τ_h is removed.
- 1-jet high $\uparrow$: the opposite sign requirement between the lepton and the τ_h is removed and the Loose working point for the τ_h isolation criteria is used.
- b-Tag: the b-tag selection is relaxed to the Loose working point.

The impact of relaxing the cuts is found to be small on the *SVfit* distribution.

Normalization

The normalization is done using a data-driven method. A data control region enriched in $W + jets$ background is obtained by requiring a "high-m_T" selection: $m_T(lep, \not{E}_T) > 70$ GeV. Specifically in the VBF categories, the requirement $m_T(lep, \not{E}_T) < 120$ GeV is also applied to reduce the non-W contamination in the tails. The contribution of other background processes are taken into account by subtracting their predicted contributions from simulation. For each category, the $W+jets$ yield is estimated in the signal region by applying a "high-m_T" to "low-m_T" ($m_T(lep, \not{E}_T) < 30$ GeV) extrapolation factor obtained from simulation:

$$r_W = \frac{N^{MC,low-m_T}_{W+jets}}{N^{MC,high-m_T}_{W+jets}} \tag{7.2}$$

The total yield is then given by:

$$N_{W+jets} = r_W \times [N^{Data} - N^{MC}_{Z\to\tau\tau} - N^{MC}_{Z\to ll} - N^{MC}_{t\bar{t}} N^{MC}_{VV}]^{high-m_T}$$

In figure 7.3, the observed and predicted $m_T(lep, \not{E}_T)$ distributions are shown for the $\mu\tau_h$ and $e\tau_h$ channels, using the 8 TeV data, after the baseline selection but relaxing the cut on the transverse mass.

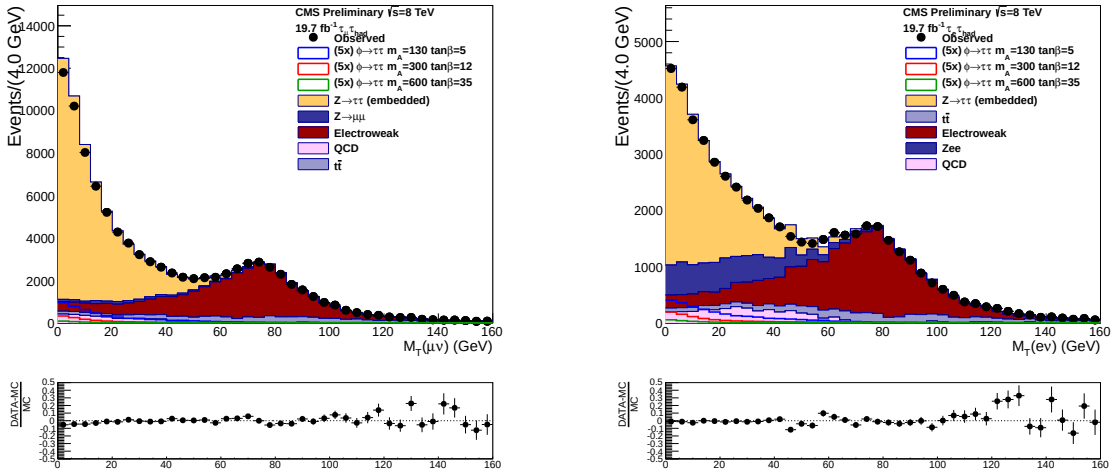


Figure 7.3: *Transverse mass variable for the $\mu\tau_h$ channel (left) and the $e\tau_h$ channel (right).*

7.3.3 QCD multi-jets

The QCD multi-jets background is another major background due to its overwhelming production cross-section at the LHC ($\sigma_{jets} \sim 10$ mb, $\sigma_W \sim 100$ nb, $\sigma_Z \sim 30$ nb). In these events the reconstructed lepton comes from a heavy flavor decay or a misidentified jet and the τ_h arises from another misidentified jet. This contribution is estimated using a data-driven technique, by counting the observed events in a QCD enriched control region,

subtracting the other background components, and then extrapolating the yield to the signal region.

Shape

The reconstructed di-τ mass shape are obtained from events where the lepton and the τ_h have the same charge (referred as SS region) and the lepton is anti-isolated:

$$0.2 < I_{rel} < 0.5.$$

Studies based on control regions confirm that the mass distributions in the sideband and signal regions are compatible [205]. The other background contributions estimated from Monte Carlo simulation are then subtracted. In order to obtain a smooth template, for the VBF tag categories, the 1-jet $\uparrow$ and b-tag categories the selection criteria are relaxed as follows:

- VBF loose and tight tag: by relaxing the VBF selection criteria requiring $p_T(jet) >$ 20 GeV, $M_{jj} > 200$ GeV, $|\Delta\eta| > 2.0$ and the central jet veto.
- 1-jet high $\uparrow$: by relaxing the τ_h isolation criteria to the Loose working point.
- b-Tag: by relaxing the b-tag selection to the Loose working point.

Normalization

The normalization is obtained from a data SS control region enriched in QCD multi-jets events. The contributions from other backgrounds in this sideband are estimated using Monte Carlo predictions except for the $W + jets$ component which is estimated using the data-driven method described above. The yield in the signal region (also called OS region) is then obtained by using an OS to SS extrapolation factor measured in a pure QCD multi jet sample obtained by relaxing the τ_h isolation and inverting the lepton isolation:

$$r_{OS/SS} = \frac{N_{QCD}^{Data,OS,AntiIso}}{N_{QCD}^{Data,SS,AntiIso}} = 1.06 \pm 0.05 \tag{7.3}$$

A systematic uncertainty of 10% is attributed to this extrapolation factor accounting for the statistical uncertainty of the measurement and a small dependence on $p_T(\tau_h)$. The mass shapes between the OS and SS regions are found to be compatible. The asymmetry between the two production cross-sections can be due to charge asymmetric processes as heavy flavor pair production ($b\bar{b}$, $c\bar{c}$). The QCD multi-jets contribution is then estimated as:

$$N_{QCD} = r_{OS/SS} \times [N^{Data} - N_{Z\to\tau\tau}^{MC} - N_{Z\to ll}^{MC} - N_{t\bar{t}}^{MC} - N_{VV}^{MC} - N_{W+jets}^{Data-driven}]^{SS}$$

A specific estimation of this background is done for the VBF tagged and 1-jet $\uparrow$ categories. In this category the number of events in the SS control region is too small. Therefore, the yield is estimated with the efficiency of the baseline to the category selection extrapolation factor. The extrapolation factor is derived from the anti-isolated lepton and relaxed τ_h control region.

7.3.4 $Z/\gamma^* \to ll + jets$

The Drell-Yan $Z/\gamma^* \to ll + jets$ production processes can be selected in the analysis when the τ_h is a mis-reconstructed electron or muon (ZL component), or jet (ZJ component).

The $Z/\gamma^* \to ee + jets$ production constitutes an important background in the $e\tau_h$ channel due to the $2-3\%$ $e \to \tau_h$ fake rate. Also the reconstructed di-τ mass for this background peaks in the Standard Model Higgs boson search range. An effective anti-electron discriminator is then needed. The $e \to \tau_h$ fake rate is corrected in simulation by the data to simulation scale factors described in section 6.4.10. The dominant systematic uncertainty for this background comes from this misidentification rate. The $Z/\gamma^* \to \mu\mu + jets$ background is less important due to the smaller $\mu \to \tau_h$ fake-rate.

The shape and yield of this background are taken from the Monte Carlo simulation. For the VBF tag and b-tag categories the shape is taken from a relaxed VBF selection as in the previous section 7.3.3. The normalization of the ZL component in the VBF loose category is estimated in a relaxed selection (baseline selection and 2 jets requirement) and extrapolated to the signal region using the efficiency of the VBF selection with respect to the relaxed selection. This efficiency is evaluated in the $Z/\gamma^* \to \tau\tau$ embedded samples in order to have an unbiased definition of the VBF loose category.

7.3.5 Di-boson and single top

The diboson and single top quark production in the semi-leptonic channels is very small and is taken directly from Monte Carlo simulation. The shape of the di-τ mass distribution in the VBF tag categories is taken using the relaxed VBF selection described in section 7.3.3.

7.3.6 $t\bar{t}$

The $t\bar{t}$ production is modeled by simulation. The kinematics of the $t\bar{t}$ production are shown to be in general well reproduced by the Monte Carlo simulation [206]. In order to improve the agreement between data and simulation in the top p_T distribution, the $t\bar{t}$ simulated events are reweighted as described in 6.4.12. The $t\bar{t}$ contamination is strongly reduced by asking no b-tagged jets in the categories of the Standard Model case, and strongly enhanced in the b-tag categories of the MSSM Higgs boson search. The b-tagged jet multiplicity is shown in the figure 7.4 for both channels.

The mass shape is taken from simulation. In the VBF tag categories it is obtained by using the relaxed VBF selection described in section 7.3.3. The $t\bar{t}$ contribution is normalized to the NNLO cross section [207].

7.3.7 Background yields summary and *SVfit* mass distributions

Tables 7.1, 7.2, 7.3 and 7.4 report the background yields in the Standard Model and MSSM categories estimated using the methods described above. The analysis acceptance is higher in the $\mu\tau_h$ channel due to the higher efficiency in the muon trigger and reconstruction.

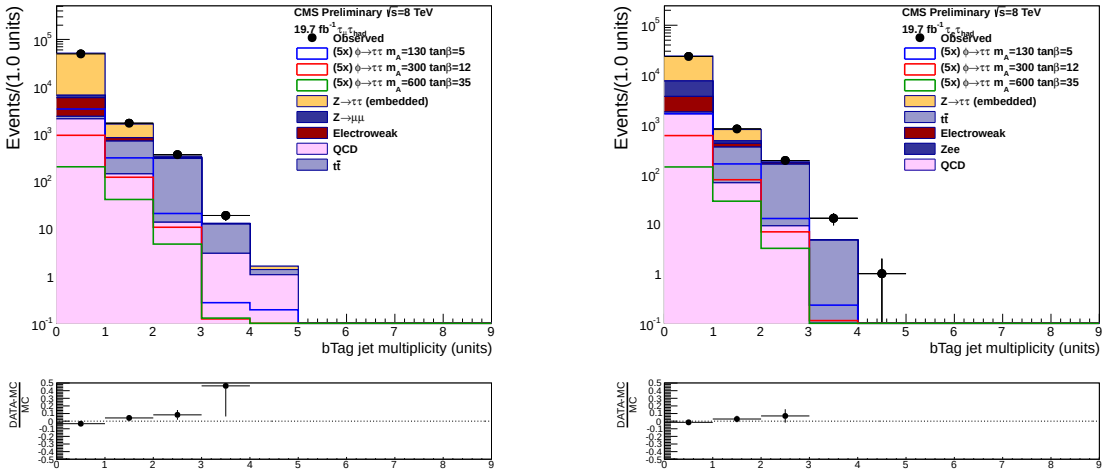


Figure 7.4: *b-tagged jet multiplicity for the $\mu\tau_h$ channel (left) and the $e\tau_h$ channel (right).*

In figures 7.5, 7.6, 7.7, 7.8 and 7.9, the di-τ mass distributions are shown, before the maximum likelihood fit is performed (pre-fit).

Process	*0-Jet*	*1-Jet*	*VBF*
Z→ $\tau\tau$	12603 ± 1180	1306 ± 101	29 ± 3
QCD	1265 ± 110	46 ± 4	0.0 ± 0.0
W+jets	903 ± 132	148 ± 19	9 ± 2
Z+jets (l/jet faking τ)	3464 ± 487	20 ± 4	2 ± 0.8
$t\bar{t}$	4 ± 0.7	62 ± 7	1 ± 0.4
Dibosons	41 ± 6	43 ± 5	1 ± 0.4
Total Background	18280 ± 1287	1625 ± 104	42 ± 4
H→ $\tau\tau$	71 ± 8	20 ± 1	4 ± 0.4
Data	18036	1404	32

Signal Eff.

gg→ H (ggF)	$2.84 \cdot 10^{-3}$	$5.79 \cdot 10^{-4}$	$3.73 \cdot 10^{-5}$
qq→ H (VBF)	$3.58 \cdot 10^{-4}$	$1.85 \cdot 10^{-3}$	$1.66 \cdot 10^{-3}$
qq→ H$t\bar{t}$ or VH	$3.44 \cdot 10^{-4}$	$1.28 \cdot 10^{-3}$	-

Table 7.1: *Standard Model $H \rightarrow \tau\tau$ search: background yields in each category for the $e\tau_h$ channel as well as the observed data event yields and expected Higgs boson contribution with $m_H = 125$ GeV. The reconstruction and selection signal efficiency for the various production mechanisms considered are also presented in the small table. The di-boson and single top production processes are grouped. The categories of high and low pT are summed. Combined statistical and systematic uncertainties on each estimate are reported.*

Process	*0-Jet*	*1-Jet*	*VBF*
Z→ $\tau\tau$	35762 ± 3395	8652 ± 640	56 ± 6
QCD	1585 ± 85	507 ± 39	6 ± 2
W+jets	1701 ± 254	1022 ± 104	17 ± 4
Z+jets (l/jet faking τ)	906 ± 175	260 ± 33	0.0 ± 0.0
$t\bar{t}$	10 ± 1	259 ± 23	2 ± 0.6
Dibosons	110 ± 14	191 ± 21	1 ± 0.5
Total Background	40074 ± 3410	10891 ± 651	82 ± 7
H→ $\tau\tau$	153 ± 17	102 ± 8	7 ± 0.6
Data	38392	10732	81

Signal Eff.

gg→ H (ggF)	$6.12 \cdot 10^{-3}$	$3.28 \cdot 10^{-3}$	$5.48 \cdot 10^{-5}$
qq→ H (VBF)	$7.46 \cdot 10^{-4}$	$6.96 \cdot 10^{-3}$	$2.87 \cdot 10^{-3}$
qq→ H$t\bar{t}$ or VH	$6.63 \cdot 10^{-4}$	$4.77 \cdot 10^{-3}$	$1.93 \cdot 10^{-5}$

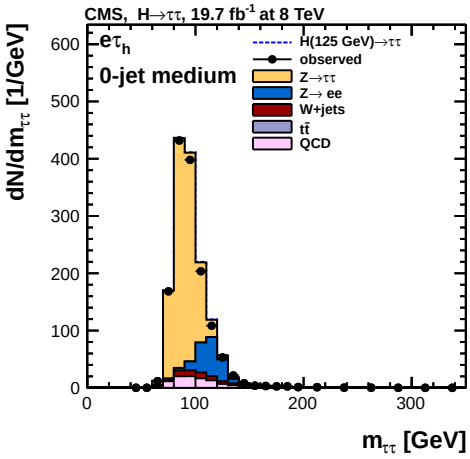
Table 7.2: *Same as table 7.1 but for the $\mu\tau_h$ channel.*

Process	*B-Tag*	*No B-Tag*
Z→ $\tau\tau$	216 ± 15	16288 ± 1196
QCD	37 ± 7	1651 ± 142
W+jets	15 ± 4	1594 ± 104
Z+jets (l/jet faking τ)	48 ± 7	4101 ± 523
$t\bar{t}$	57 ± 8	141 ± 12
Dibosons	19 ± 3	138 ± 15
Total Background	391 ± 20	23913 ± 1317
H→ $\tau\tau$	15 ± 1	113 ± 5
Data	382	23470

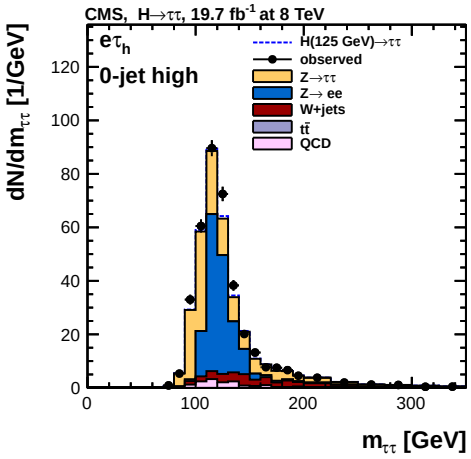
Signal Eff.

gg→ H (ggF)	$7.13 \cdot 10^{-5}$	$8.33 \cdot 10^{-3}$
bb→ H	$1.26 \cdot 10^{-3}$	$7.10 \cdot 10^{-3}$

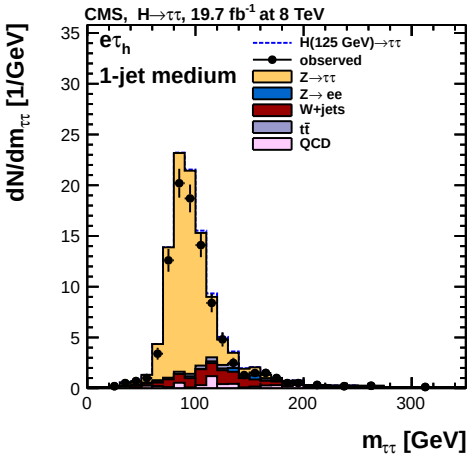
Table 7.3: *MSSM $H \to \tau\tau$ search: background yields in each category for the $e\tau_h$ channel as well as the observed data event yields and expected Higgs boson contribution with $m_\phi = 160$ GeV and $\tan\beta = 8$. The reconstruction and selection signal efficiency for the various production mechanisms considered is also presented in the small table. The di-boson and single top production processes are grouped. Combined statistical and systematic uncertainties on each estimate are reported.*



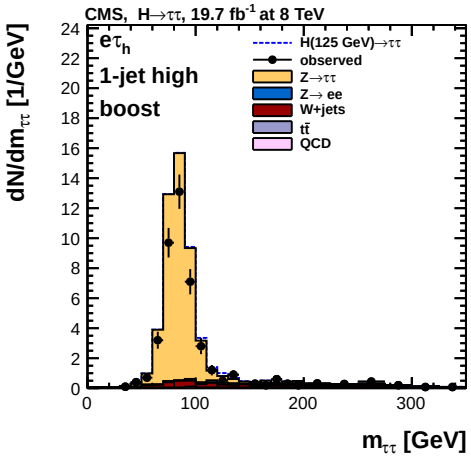
(a) *SVfit* Mass : $e\tau_h$ 0-jet medium



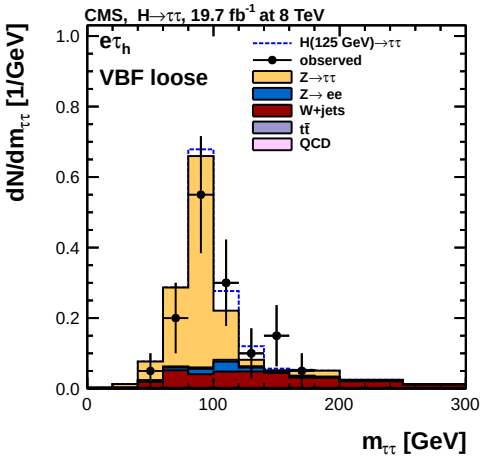
(b) *SVfit Mass* : $e\tau_h$ 0-jet high



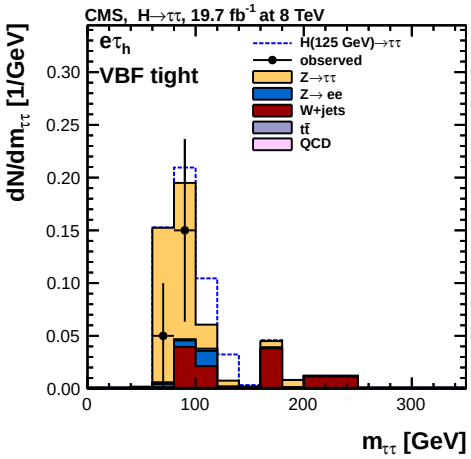
(c) *SVfit Mass* : $e\tau_h$ 1-jet medium



(d) *SVfit Mass* : $e\tau_h$ 1-jet high ↑

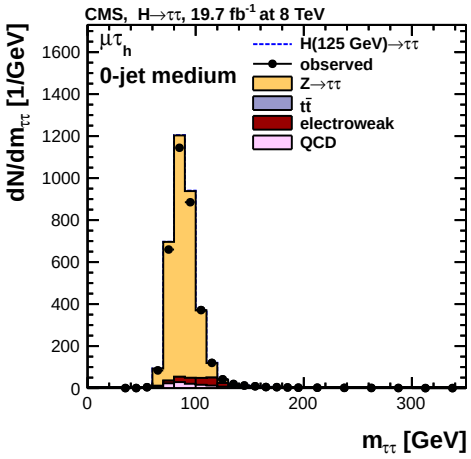


(e) *SVfit Mass* : $e\tau_h$ vbf loose

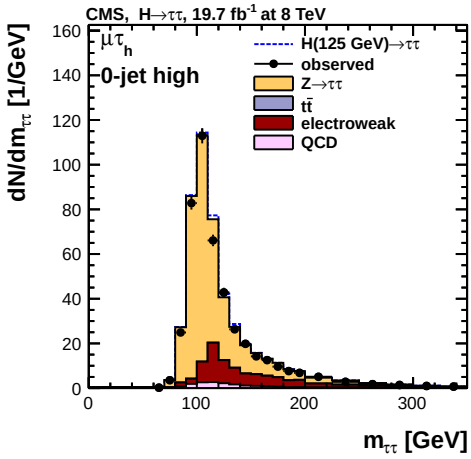


(f) *SVfit Mass* : $e\tau_h$ vbf tight

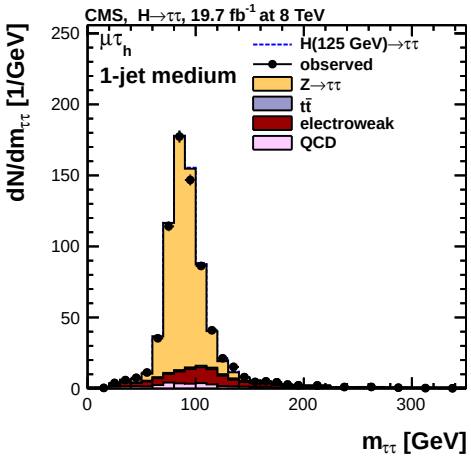
Figure 7.5: *Di-τ mass distribution for the categories used in the Standard Model analysis in the* $e\tau_h$ *channel before the maximum likelihood fit is performed.*



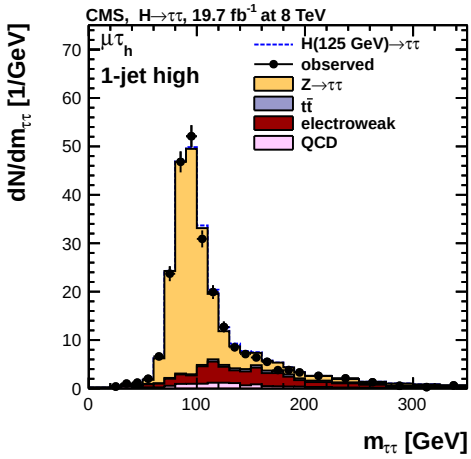
(a) *SVfit* Mass : $\mu\tau_h$ 0-jet medium



(b) *SVfit Mass* : $\mu\tau_h$ 0-jet high

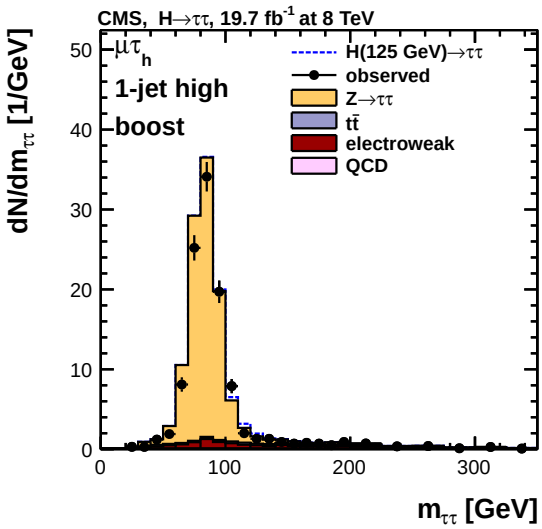


(c) *SVfit Mass* : $\mu\tau_h$ 1-jet medium

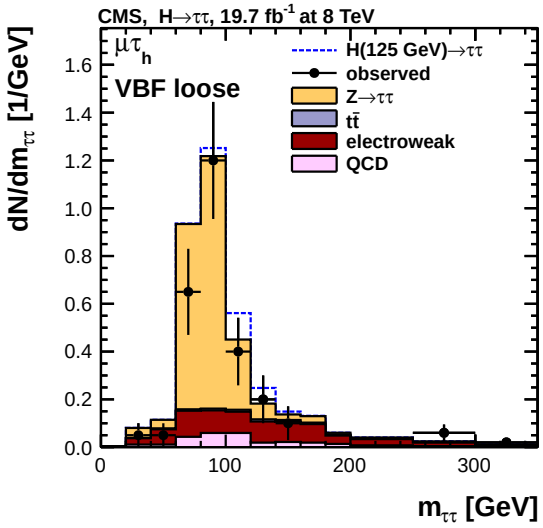


(d) *SVfit Mass* : $\mu\tau_h$ 1-jet high ↓

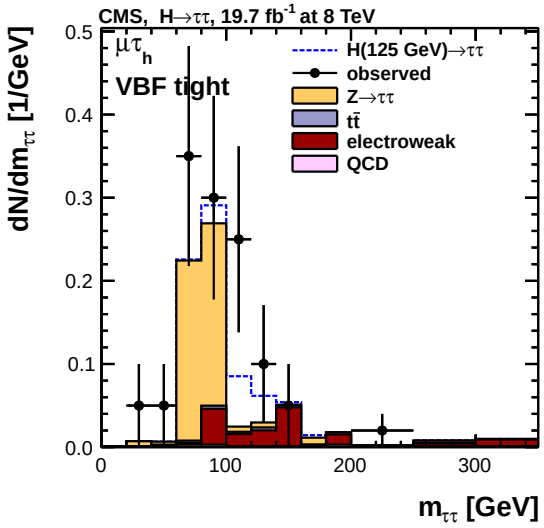
Figure 7.6: *Di-τ mass distribution for the categories used in the Standard Model analysis in the $\mu\tau_h$ channel before the maximum likelihood fit is performed.*



(a) *SVfit Mass* : $\mu\tau_h$ 1-jet high ↑

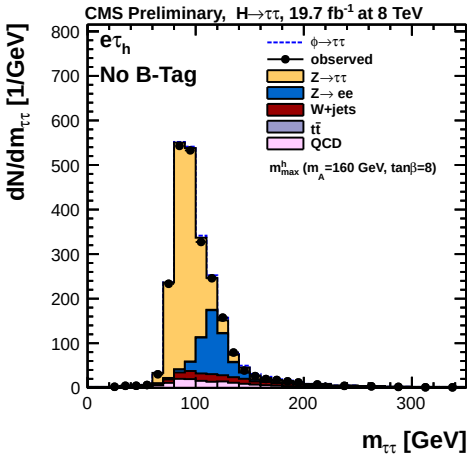


(b) *SVfit Mass* : $\mu\tau_h$ vbf loose

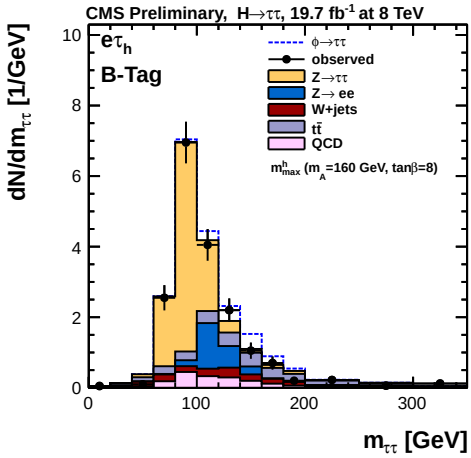


(c) *SVfit Mass* : $\mu\tau_h$ vbf tight

Figure 7.7: *Di-τ mass distribution for the categories used in the Standard Model analysis in the* $\mu\tau_h$ *channel before the maximum likelihood fit is performed.*

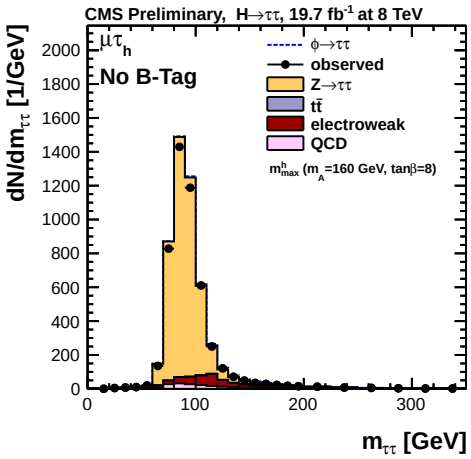


(a) *SVfit Mass* : $e\tau_h$ no b-Tag

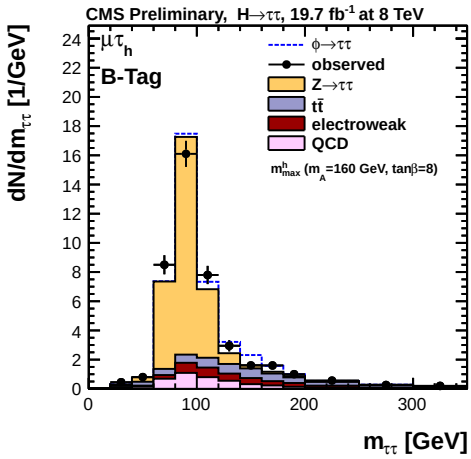


(b) *SVfit Mass* : $e\tau_h$ b-Tag

Figure 7.8: *Di-τ mass distribution for the categories used in the current MSSM analysis (Ref. [162]) in the $e\tau_h$ channel before the maximum likelihood fit is performed.*



(a) *SVfit Mass* : $\mu\tau_h$ no b-Tag



(b) *SVfit Mass* : $\mu\tau_h$ b-Tag

Figure 7.9: *Di-τ mass distribution for the categories used in the current MSSM analysis (Ref. [162]) in the $\mu\tau_h$ channel before the maximum likelihood fit is performed.*

Process	*B-Tag*	*No B-Tag*
Z→ $\tau\tau$	549 ± 39	44405 ± 3266
QCD	88 ± 16	2097 ± 179
W+jets	29 ± 6	2858 ± 195
Z+jets (l/jet faking τ)	14 ± 2	1075 ± 174
$t\bar{t}$	113 ± 15	277 ± 24
Dibosons	34 ± 4	309 ± 32
Total Background	827 ± 45	51021 ± 3282
H→ $\tau\tau$	28 ± 2	213 ± 9
Data	874	49315
Signal Eff.		
gg→ H (ggF)	$1.71 \cdot 10^{-4}$	$1.54 \cdot 10^{-2}$
bb→ H	$2.29 \cdot 10^{-3}$	$1.34 \cdot 10^{-2}$

Table 7.4: *Same as table 7.3 but for the* $\mu\tau_h$ *channel.*

7.4 Systematic uncertainties common to the SM and MSSM Higgs boson searches

As explained in section 6.6, the systematic uncertainties are treated in the likelihood model as nuisance parameters in the fit. They can modify the norm and the shape of the mass distributions. All the nuisance parameters are profiled in the simultaneous maximum likelihood fit over all the categories. The most important sources of systematic uncertainties are listed in the next section. They can be grouped into experimental and theory uncertainties. The experimental ones can be related to the reconstruction of physics objects or to the background estimation methods. Additional systematic uncertainties are added to the regions and categories with low statistics.

7.4.1 Experimental systematics

Luminosity

The uncertainty on the integrated luminosity recorded by CMS is 2.6% for the data taken at 8 TeV [208]. It affects the yields of signal as well as backgrounds which normalization is based on simulation: $Z/\gamma^* \to ll$, $Z/\gamma^* \to \tau\tau$, $t\bar{t}$, di-boson and single top.

Muon/electron trigger, identification and isolation efficiency

The uncertainties on the muon and electron trigger, identification and isolation efficiencies affect the yield of the signal and all the backgrounds. It is taken from the uncertainties on the data-to-simulation scale factors described in 6.4.2 and 6.4.7. The

uncertainties are added in quadrature, leading to a value of 2% for electrons and muons in all the p_T and η ranges.

τ_h identification efficiency and τ_h trigger

The main experimental uncertainty comes from the hadronic tau reconstruction. The τ_h identification efficiency has been measured in data using the Tag-and-Probe technique on $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events [209] as presented in section 4.5.3. The uncertainty on this measurement is 6%, as the τ_h in these events have a transverse momenta of about $20-50$ GeV. An uncorrelated uncertainty of $\frac{p_T}{1000\ \text{GeV}} \cdot 20\%$ is added for taus with higher momenta. The trigger efficiency of the τ_h leg is measured also using the Tag-and-Probe technique using events firing the single muon trigger path as described in 6.4.2. The uncertainty of 3% is added in quadrature to the identification uncertainty affecting the yields of the signal and $Z/\gamma^* \to \tau\tau$ and is completely correlated between the categories in a given channel.

$e \to \tau_h$ and $\mu \to \tau_h$ fake rate

The uncertainty on the $e \to \tau_h$ fake rate is taken from the measurement of the scale factors described in section 6.4.10. As shown in figure 7.10, the corrections only affect the event yield of the ZL background component (the τ_h being a mis-reconstructed $e^{\pm}$ or $\mu^{\pm}$) and the shape is conserved. A systematic uncertainty of 30% affects this background totally correlated between the different reconstructed decay modes.

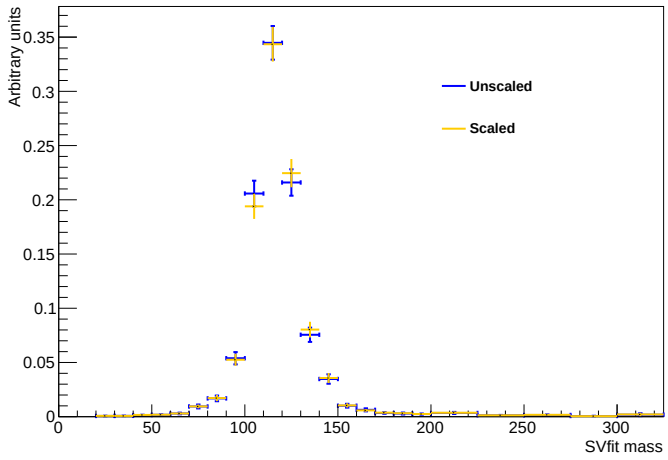


Figure 7.10: *Normalized SVfit mass distribution before (blue) and after (yellow) scaling the ZL component by the scale factors measured in section 6.4.10. We see here that the shape is not affected by the scale factors.*

The uncertainty on the $\mu \to \tau_h$ fake rate amounts to 30%, and affects the yield of the ZL background component in the $\mu\tau_h$ channel.

Electron and τ_h energy scale

The electron energy scale is attributed an uncertainty of 1% in the barrel region and 2.5% in the end-cap regions [161]. The muon momentum uncertainty is assumed to be negligible. An uncertainty on the τ_h energy scale of 3% is estimated from template fits to the visible τ_h mass as described in section 4.6. The 3% uncertainty covers the differences between data and simulation for all the τ_h decay modes, transverse momenta and regions (*c.f.* section 6.4.4). The τ_h energy scale uncertainty affects the normalization due to the cuts on the p_T and the shape of the reconstructed mass template. Two shape templates are obtained after varying up and down the energy scales in order to feed the shape interpolation done during the fit.

Jet energy scale

The uncertainty on the jet energy scale depends on the jet p_T and η [210]. As no requirement on the jet content is done in the baseline selection (7.1), the jet energy scale does not affect the event yield in the baseline sample. However, it leads to uncertainties on the rate and shape of the mass distribution in each category. As for the τ_h energy scale, two shape templates are obtained after varying up and down the energy scale. The jet energy scale is not propagated to the $\not\!\!E_T$, because the uncertainties on the missing transverse energy and $m_T(lep, \not\!\!E_T)$ are covered by the recoil correction uncertainties.

$\not\!\!E_T$ scale

The $\not\!\!E_T$ scale uncertainty affects the yield and the shape of all distributions due to the $m_T(lep, \not\!\!E_T)$ cut or the direct $\not\!\!E_T$ cut in the 1-jet category of the $e\tau_h$ channel. A relative scale uncertainty of 5% comes from the error on the resolution and response of the hadronic recoil measurement described in section 6.4.8, fully correlated between all categories. The $\not\!\!E_T$ and all the variables related to the $\not\!\!E_T$ are recomputed varying up and down the recoil correction parameters.

b-Tag scale factors

The uncertainties on the b-tag efficiencies and mis-tagging rates [211] are evaluated by varying the data-to-simulation scale factors presented in 6.4.6 within their uncertainties and reanalyzing the events. This results in yield uncertainties up to 8%.

7.4.2 Background evaluation systematics

$Z/\gamma^* \to \tau\tau$ normalization

An uncertainty of 3% is attributed to the baseline yield of $Z/\gamma^* \to \tau\tau$. The extrapolation factors evaluated from the embedded samples for each category lead to additional systematic uncertainties. They correspond to the statistical error on the extrapolation factors and they range from 2 to 14%.

W+jets normalization

The W+jets normalization is taken from observed data high-m_T sideband. Two sources of systematic uncertainties arise, the limited statistics on the sideband region and the uncertainty on the extrapolation factor r_W (see equation 7.2). The total systematic is obtained by adding both in quadrature and is dominated by the uncertainties on the extrapolation factor.

The systematic uncertainty on r_W takes into account the $m_T(lep, \not\!\!E_T)$ distribution data to simulation differences. It is evaluated differently in the Standard Model and MSSM analyses. In the Standard Model search, it is estimated by comparing simulated and observed $Z \to \mu\mu + jets$ events where a reconstructed muon is removed from the event to emulate $W + jets$ events. The systematic uncertainty is estimated to be 10% to 25% depending on the category. For the MSSM case, the search is extended to high masses. Hence, the $Z \to \mu\mu + jets$ in data are not suited to evaluate this systematic error. Instead, the uncertainty is obtained by varying the hadronic recoil correction parameters within their uncertainties as for the $\not\!\!E_T$ scale systematic. The systematic uncertainty ranges from 10% in the no b-Tag categories to 30% in the b-Tag categories. It is fully de-correlated between categories.

QCD multi-jets normalization

The QCD multi-jets normalization systematic uncertainty comes from two sources: the statistical uncertainty on the observed events selected in the sideband enriched in QCD and the systematic uncertainty on $r_{OS/SS}$ (see equation 7.3). The total uncertainty ranges from 6 to 35% depending on the category and is fully de-correlated.

$Z/\gamma^* \to ll$ + jets

The uncertainty on the Drell-Yan $Z/\gamma^* \to ll$ + jets production comes from the uncertainty on the jet$\to \tau_h$, $e \to \tau_h$ and $\mu \to \tau_h$ fakes. It ranges between 6 to 40% depending on the mis-identified object and the category.

Di-boson and top normalization

The uncertainty on the di-boson, single top and $t\bar{t}$ normalization is 15%. An additional uncertainty takes into account the top quark p_T reweighting applied to simulation (*c.f.* 6.4.12) by varying the reweighting between no and twice the correction. Due to the low event statistics in the VBF tag categories, additional uncertainties are added in quadrature: an extra uncertainty of 33% is added to the $t\bar{t}$ normalization while the total uncertainty of the di-boson normalization amounts to 100% in the $e\tau_h$ channel.

7.4.3 Theoretical systematics

The theoretical systematics are related to the uncertainty on the theoretical calculation of the cross section of the signal processes. Uncertainties on the parton distribution functions (PDF), the strong coupling constant α_S, the renormalization and factorization scales, as

well as the simulation of the parton showering, hadronization and underlying event are taken into account.

Following the recommendations in Ref. [58], the normalization uncertainty on the signal yield in the VBF tag categories rises to 4%. For the gluon fusion process they amount to 12-25%. Due to the missing higher-order corrections, as detailed in section 6.4.13, in the gluon fusion process additional uncertainties of 10 to 41%, depending on the category, are included. For the MSSM analysis, the corresponding uncertainties for the gluon fusion process amounts to 10-15% and 15-20% for the associated production with b quarks.

A yield systematic of 8% is assigned to the PDFs and α_S uncertainty. Finally, the corresponding uncertainty on the parton showering and underlying event modeling in simulation raise to 4%.

In the current MSSM analysis [162], the uncertainty on the Higgs boson p_T is not taken into account and is considered to be covered by the other theoretical uncertainties. In section 7.5.3 a specific systematic uncertainty is proposed, taken as a signal shape systematic. The Higgs boson p_T spectra is affected by this nuisance and the yield in the analysis categories is modified due to the change in the acceptance. The dependence of the Higgs momentum on the up and down type quarks and stop and sbottom masses is estimated by varying $\tan\beta$. The systematic uncertainty is chosen as the difference between the two extreme values $\tan\beta = 1$ (pure top loop) and $\tan\beta = 60$ (pure bottom loop). This systematic then takes the possible model dependence on the analysis selections into account.

7.4.4 Statistical systematics

Additional shape uncertainties on the background modeling are applied to the categories where the number of events in the data control regions is low.

Bin-by-bin uncertainties

These uncertainties, so-called bin-by-bin uncertainties, are added in the VBF tagged categories in the Standard Model analysis and b-Tag categories in the MSSM analysis to take into account the statistical fluctuations on the shape templates due to the low statistics. They are applied to all the background processes. They are un-correlated across the bins in each template. They can float within statistical uncertainties during the fit by the addition of one nuisance parameter per bin constrained by log-normal distributions. They are applied to all the bins where the background has an important contribution ($m_H < 150$ GeV). In order to reduce the computing time after the inclusion of this extra parameters to the fit, some simplifications are done:

- The statistical uncertainty of all the backgrounds present in a given bin are added in quadrature and associated to the dominant background.

- The uncertainties smaller than 10% times the background contribution expected in a given bin are neglected (*pruning*).

Fit of the di-τ mass tail in the MSSM analysis

In the MSSM analysis, in the high mass region the event statistics is very limited. The background contributions are expected to fall exponentially, but statistical fluctuations of this tail region can bias the fit result. In order to have a better background expectation in the region of large *SVfit* mass, a fit of the QCD multi-jets, W+jets and $t\bar{t}$ is performed individually with exponential functions:

$$f = \exp -\frac{M_{\tau\tau}}{c_0 + c_1 \cdot M_{\tau\tau}}$$

where c_0 and c_1 are two fit parameters. The eigenvectors $v_{i=1,2}$ and eigenvalues λ_i of the fit covariance matrix are determined. The uncorrelated transformed parameters $k_i = v_{i0}c_0 + v_{i1}c_1$ are computed. The eigenvalues represent the uncertainties on the transformed parameters. The shape uncertainties on the fit function are taken into account by the nuisance parameters constrained by log-normal distributions (*c.f.* equation 6.33) with the parameters $\theta_i = k_i$ and $\kappa_i = 1 + \lambda_i$. An example of fit is presented in figure 7.11.

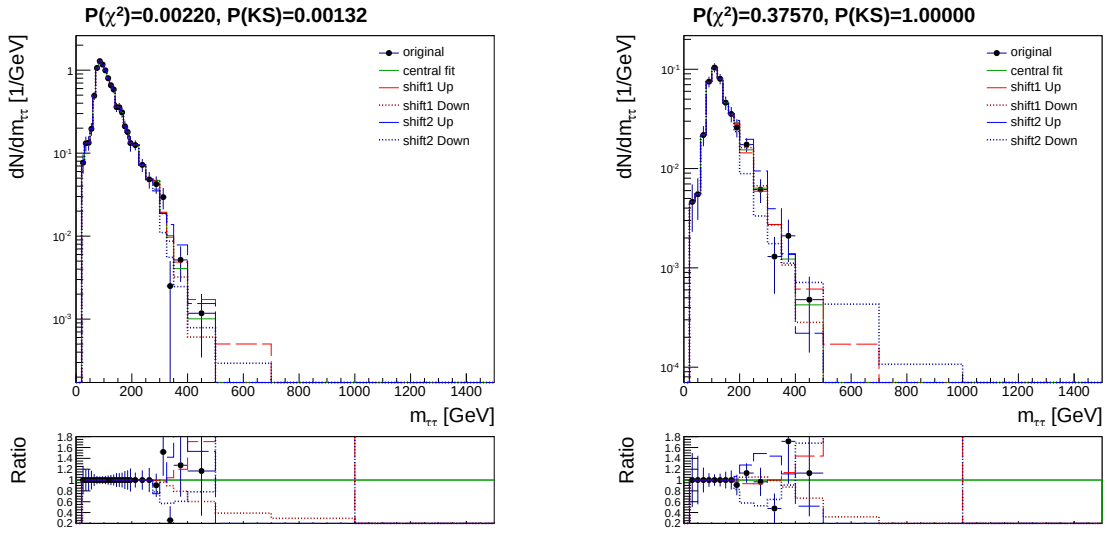


Figure 7.11: *Example of analytic fit of the tail of the di-τ mass distributions for the QCD background in the no-bTag high (left) and b-Tag high (right) categories in the $\mu\tau_h$ channel. The fitted function is represented by the central fit binned function, the fit functions under the variations of the fit parameters individually are also displayed (ShiftUp and ShiftDown curves).*

7.4.5 Summary

In table 7.5 the full list of systematic uncertainties is presented. The most important systematic affecting both Standard Model and MSSM analysis are the uncertainties related to the τ_h object (identification (8%) and energy scale (3%)) due to their impact in the large irreducible background.

Name of systematic uncertainty	Affected samples	Uncertainty
τ_h Energy Scale (†)	Signal and MC Backgrounds	3%
τ_h Identification & Trigger (†)	Signal and MC Backgrounds	8%
$e \to \tau_h$ rate (†∗)	$Z \to ee$	20-36%
$\mu \to \tau_h$ rate (†∗)	$Z \to \mu\mu$	30%
$jet \to \tau_h$ rate (†∗)	$Z + jets$	20-40%
e Identification & Trigger (†∗)	Signal and MC Backgrounds	2%
μ Identification & Trigger (†∗)	Signal and MC Backgrounds	2%
e Energy Scale (†)	Signal and MC Backgrounds	1%
Jet Energy Scale (†∗)	Signal and MC Backgrounds	1-20%
$\not\!\!E_T$ (†∗)	Signal and MC Backgrounds	2-8%
b-Tag Efficiency (†∗)	Signal and MC Backgrounds	2-3%
Mis-Tagging (†∗)	Signal and MC Backgrounds	2-3%
$\mathcal{L}_{int}$	Signal and MC Backgrounds	2.6%
Norm Z (†∗)	Z	3%
Extrapolation Z	$Z \to \tau\tau$	0-13%
Norm W	$W + jets$	10-33%
Norm $t\bar{t}$ († ∗ ex. vbf)	$t\bar{t}$	10-33%
Norm diboson († ∗ ex. vbf)	VV	15-100%
Norm QCD	QCD multi-jet	6-35%
PDF+α_S (qq)	Signal and MC Background	8%
PDF+α_S (gg)	Signal and MC Background	8%
Scale variation	Signal	4-25%
UE & PS	all	4%
Limited statistics	all	bin-by-bin
Limited statistics (MSSM)	$W + jets$,QCD,$t\bar{t}$	tail fits

Table 7.5: *Main systematic uncertainties entering the analysis. The (∗) symbol indicates correlation between separate channels. The (†) symbol indicates correlation between separate categories. In the instances where "ex. vbf" is indicated, an additional uncorrelated nuisance is added to account for statistical uncertainties.*

7.5 Improvements brought to the analysis

In this section, I will detail the main aspects in which I brought original contributions to the analysis. Three main contributions will be detailed:

- First, at the object level, the improved τ_h reconstruction has shown a big improvement on the expected sensitivity by reducing the τ_h fakes contamination. This improvements will be detailed in section 7.5.1. The anti-electron discriminator described in chapter 5 brings a better control of the $Z/\gamma^* \to ee$ background in the $e\tau_h$ channel and reduces it considerably. The novel MVA-based τ_h isolation (see section 4.5.2) brings better fake jets rejection specially for boosted τ_h leptons. It brings considerable improvement in the sensitivity in the MSSM Higgs boson searches.

- As will be described in section 7.5.2, an improved categorization brings also a substantial gain in sensitivity for the MSSM analysis. Following a strategy similar to the Standard Model search, the events can be categorized by separating the phase space as functions of the τ_h transverse momentum. Boosted τ_h events have lower background contamination and their *SVfit* mass resolution is improved.

- Finally section 7.5.3 is devoted to the studies I have carried to introduce a better method to treat the model dependence related to the Higgs boson p_T in the MSSM Higgs bosons searches.

Following the blinding policy, the analysis was optimized without looking at the data in the signal region of the di-τ mass distribution, and the choices were based on the best expected sensitivity.

7.5.1 Improved τ_h reconstruction

MVA based anti-electron discriminator

The anti-electron discriminator described in chapter 5 is designed to separate τ_h candidates from fake electrons. At the analysis level, one important background is the Drell-Yan $Z/\gamma^* \to ee$ production, mainly in the $e\tau_h$ channel. In figure 7.12, the di-τ mass distributions after the baseline selection are presented for the previous and the current anti-electron discriminators. The $Z/\gamma^* \to ee$ background in blue is reduced by a factor ~ 4 while keeping the same τ_h identification efficiency as the $Z/\gamma^* \to \tau\tau$ stays at the same level.

The background contamination in the bins near $m_H = 120$ GeV is reduced due to the reduction of the $Z/\gamma^* \to ee$ background. This leads to an improvement in the analysis sensitivity in this mass range, and the expected significance to a Higgs boson is enhanced. In figure 7.13 the comparison of the expected significance for the previous and new anti-electron discriminators is presented. An improvement of the expected significance in the $e\tau_h$ channel in all the search mass range is clearly visible, reaching about 15% at $m_H = 120$ GeV.

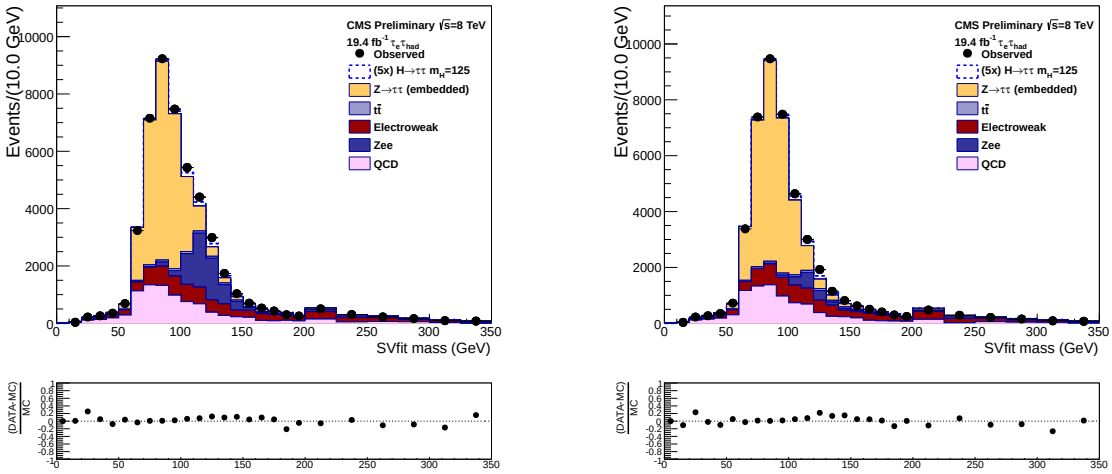


Figure 7.12: *Pre-fit SVfit mass distribution for the previous anti-electron discriminator and the MVA based anti-electron discriminator described in this thesis (c.f. chapter 5). A big reduction of the $Z/\gamma^* \to ee$ background is observed.*

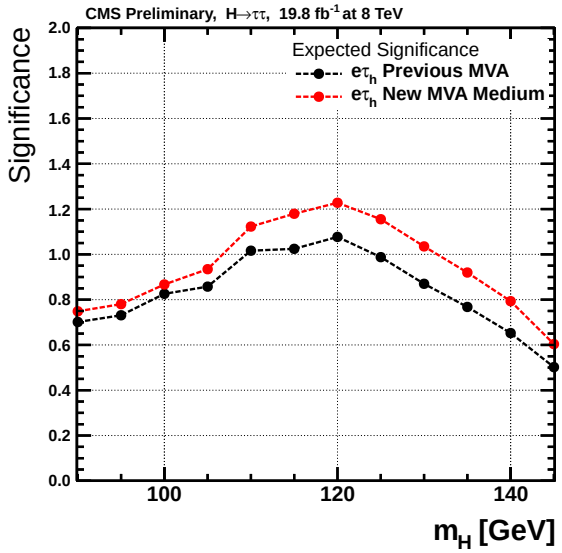


Figure 7.13: *Expected significance to the Standard Model Higgs boson as a function of m_H in the $e\tau_h$ channel using different anti-electron discriminators in the τ_h identification. In red, the discriminator developed in this thesis.*

MVA based isolation

As presented in chapter 4, the main handle to reject jets misidentified as hadronic τ leptons is the isolation of the tau candidate. A novel MVA based isolation making use of the τ lifetime has been described and validated for the analyses presented in this thesis.

In figure 7.14, the effect of the MVA based isolation with respect to the cut based in the Standard Model analysis is shown for the $\mu\tau_h$ channel. The expected performance is similar at low transverse momenta with respect to the cut based isolation but the new approach brings a small improvement ($\sim 4\%$ in the $\mu\tau_h$ channel) in sensitivity.

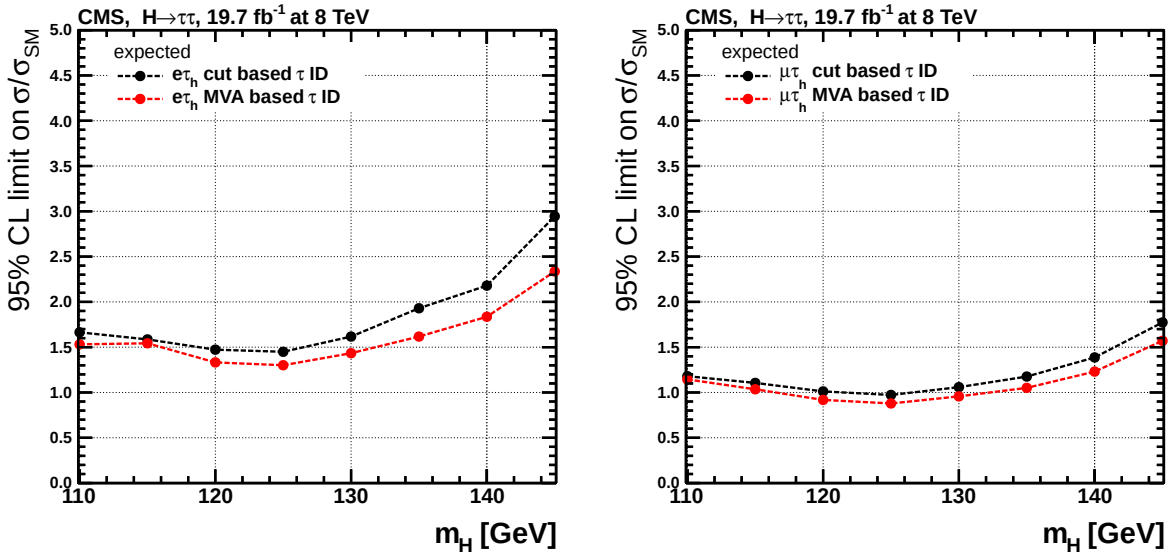


Figure 7.14: *Expected exclusion limits of a Standard Model Higgs boson for the cut-based isolation (black) and for the MVA-based isolation (red) for the $e\tau_h$ channel (left) and the $\mu\tau_h$ channel (right).*

The performance of the MVA is expected to be more significant for boosted topologies of the τ lepton decays where the displaced vertex variables are reconstructed. Therefore the MSSM analysis in which the signal hypothesis masses reach 1 TeV is expected to be greatly improved due to the $jet \to \tau$ fake reduction for the same efficiency.

The improvement in the analysis sensitivity obtained from the MVA isolation algorithm is quantified by comparing the model independent limits on the $gg \to \phi$ and $gg \to b\phi$ cross-sections that we expect to set in the absence of a MSSM Higgs signal. The other production process is profiled[2] in the fit. The results are shown for the $e\tau_h$ and $\mu\tau_h$ channels in figure 7.15. The reduction in $jet \to \tau_h$ fake rate provided by the new algorithm improves the expected limit by 20-50% in the $e\tau_h$ and $\mu\tau_h$ channels, corresponding to an increase in equivalent luminosity by a factor 1.5 to 2. The performance was also evaluated in the $\tau_h\tau_h$ channel [3] as well as for the combination of all channels and all CMS run 1 data in figure 7.16. The improvement is larger in the $\tau_h\tau_h$ channel, due to the presence of two reconstructed τ_h, equivalent to an increase in luminosity by a factor 2. The gain in analysis sensitivity is present over the full range of Higgs masses considered.

The MVA-based τ_h isolation will be then considered in the results for combination of the final states of the MSSM analysis presented in this thesis (see section 7.7).

[2]The nuisance parameters for the other production process are set to the constrained value.
[3]Inputs for the comparison were kindly provided by Dr. Christian Veelken.

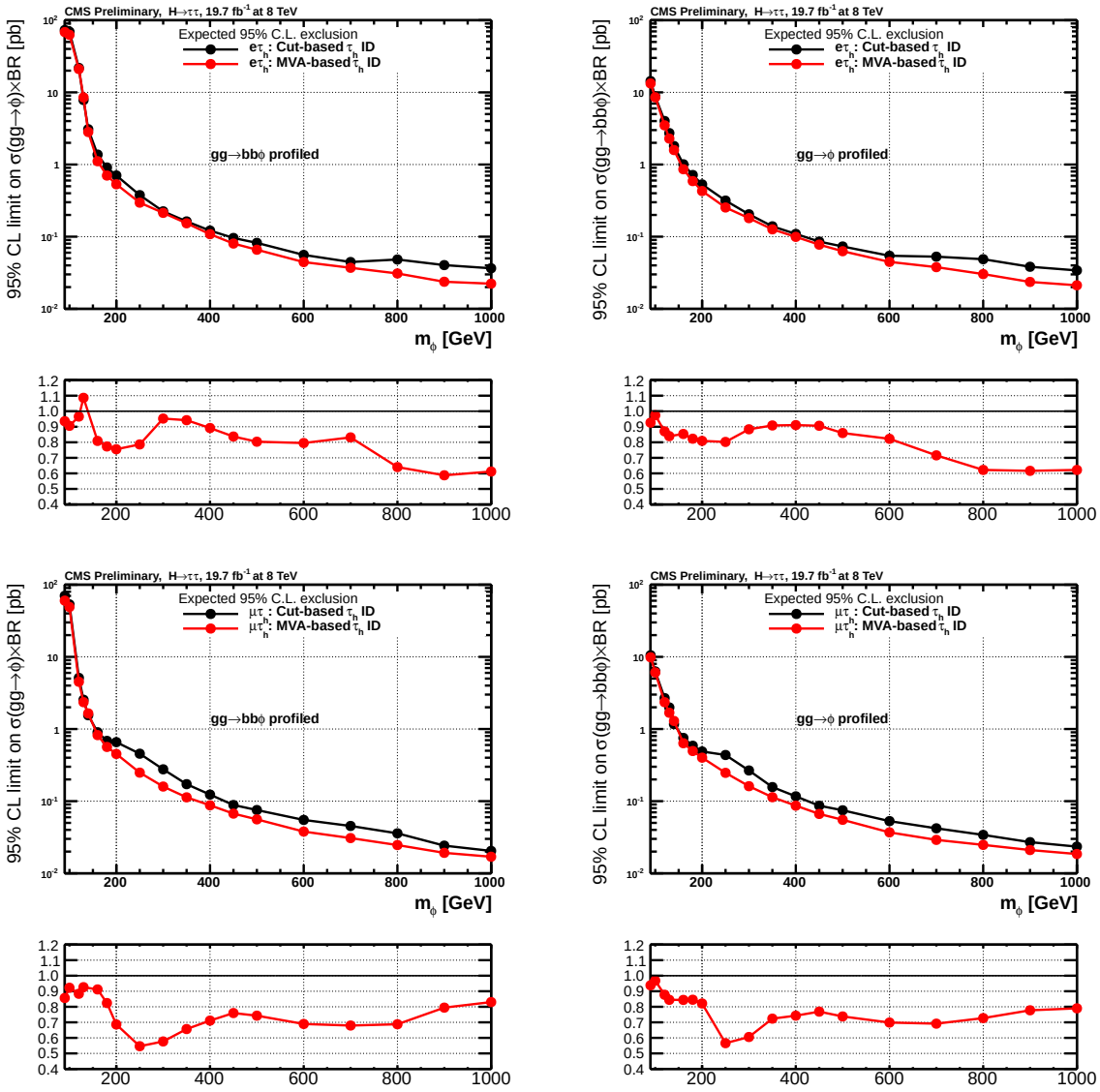


Figure 7.15: 95% *CL upper limits on cross section times branching ratio for* $gg \to \Phi \to \tau\tau$ *(left) and* $gg \to \Phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal. The sensitivity of the* $e\tau_h$ *channel (top) and* $\mu\tau_h$ *channel (bottom) in the* $\sqrt{s} = 8$ *TeV data is compared for two cases: using the cut-based* τ_h *isolation algorithm used in [162] (denoted by "Cut-based* τ_h *ID" in the legend) and using the MVA-based* τ_h *isolation algorithm described in section 4.5.2 (denoted by "MVA-based* τ_h *ID" in the legend). The ratio between the two curves is displayed at the bottom of each plot.*

7.5.2 New categorization in MSSM analysis

Similarly to the Standard Model case, the resolution of the di-τ mass improves for taus with higher transverse momentum. Typically the transverse momentum of the taus from the decays of the MSSM Higgs bosons leans around $\sim 25\%$ of the Higgs boson mass, as

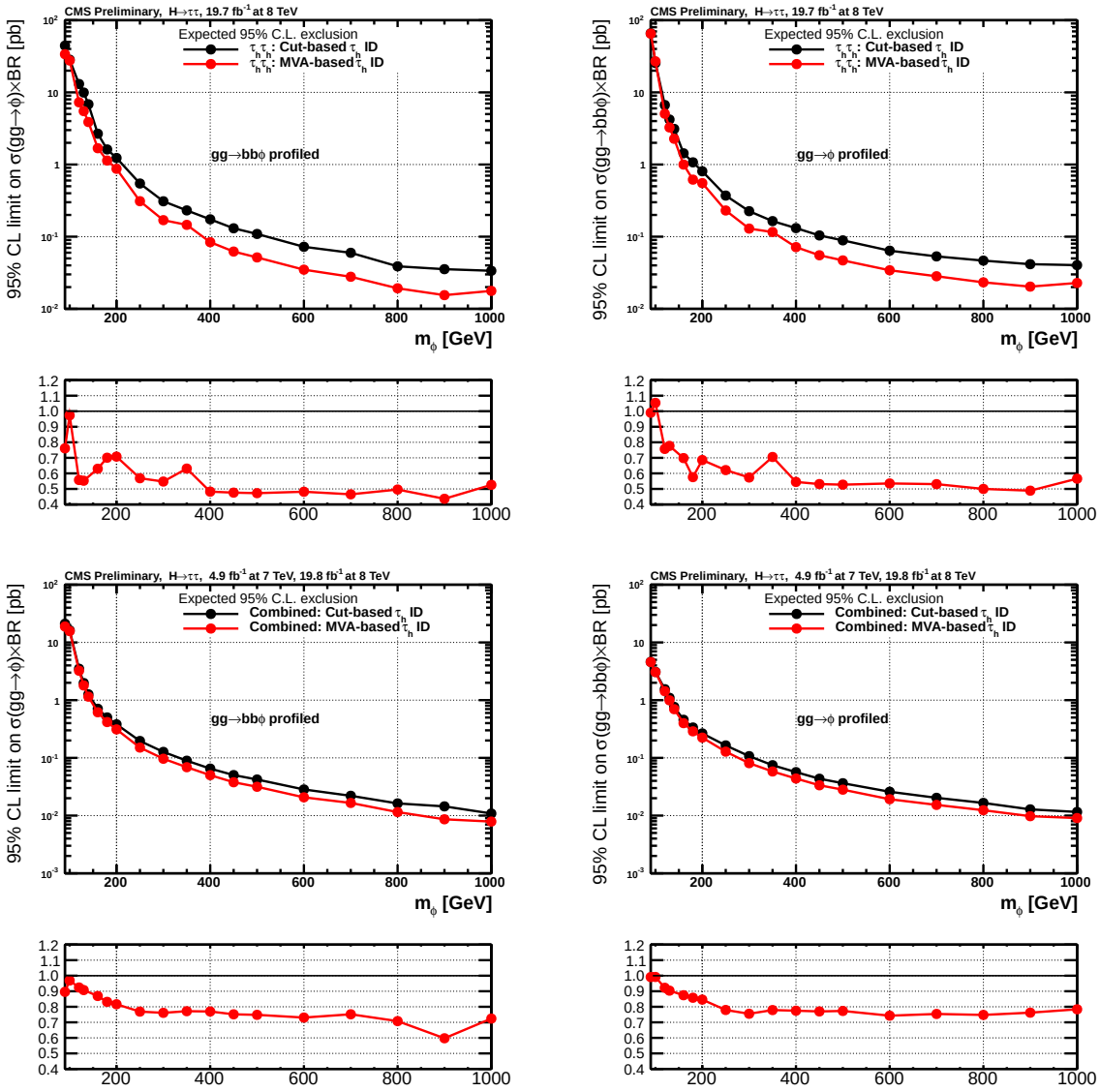


Figure 7.16: 95% *CL upper limits on cross section times branching ratio for* $gg \to \Phi \to \tau\tau$ *(left) and* $gg \to \Phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal. The sensitivity of the* $\tau h \tau_h$ *channel in the* $\sqrt{s} = 8$ *TeV data (top) and of the combination of all channels and all CMS run 1 data (bottom) is compared for two cases: using the cut–based* τ_h *identification algorithm used in Ref. [162] (denoted by "Cut-based* τ_h *ID" in the legend) and using the new MVA-based* τ_h *isolation algorithm described in section 4.5.2 (denoted by "MVA-based* τ_h *ID" in the legend). The ratio between the two curves is displayed at the bottom of each plot.*

shown in figure 7.17.

Also, the distribution of the tau transverse momentum steeply falls in the case of the most important backgrounds containing τ fakes. Therefore, separating events with high

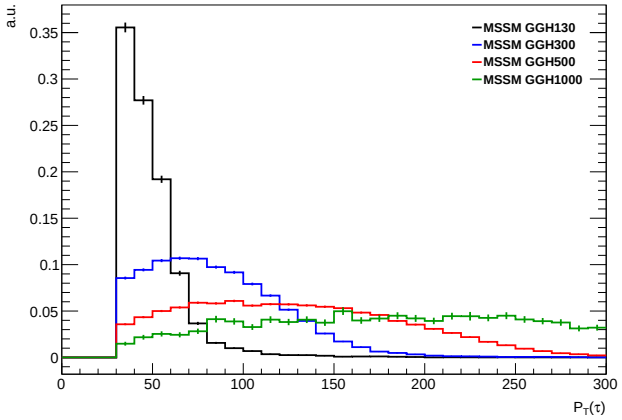


Figure 7.17: *Distribution of the τ_h transverse momentum for different MSSM Higgs signal mass hypothesis: $m_\phi = 130$ (black), $m_\phi = 300$ (blue), $m_\phi = 500$ (red) and $m_\phi = 1000$ (green).*

$p_T(\tau_h)$ from the low $p_T(\tau_h)$ ones leads to categories with low and high signal purity, enhancing the analysis sensitivity. The reason is that the irreducible $Z/\gamma^* \to \tau\tau$ background predominantly produces hadronic taus of $p_T \lesssim 45$ GeV. The other reason is that the jet p_T spectra in the W + jets and QCD multi-jet backgrounds decreases with p_T faster than for signal, in addition the jet $\to \tau_h$ fake-rate decrease for higher p_T jets (see chapter 4). In particular for the region $m_\phi \gtrsim 200$ GeV in which the channels with a hadronically decaying τ-lepton ($e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$) contribute mostly to the sensitivity.

As the explored values in the MSSM Higgs boson search varies from 90 to 1000 GeV we split the categories in 3 bins of $p_T(\tau_h)$ in the most populated "no b-tag" category and 2 bins for the statistically limited "b-tag" category. Those categories are defined as follows.

- **b-Tag**

  Events containing at most one jet with $p_T(jet) > 30$ GeV and $|\eta| < 4.7$ and at least one b-tagged jet (Medium working point CSV output above 0.679) with $p_T(jet) > 20$ GeV. This category is split in two sub-categories depending on the hadronic tau transverse momentum:

  b-Tag low: $30 \leq p_T(\tau_h) \leq 45$ GeV
  b-Tag high: $p_T(\tau_h) > 45$ GeV

- **No b-Tag**

  Events not containing a b-tagged jet (Medium working point CSV output above 0.679) with $p_T(jet) > 20$ GeV. This category is split in three sub-categories depending on the hadronic tau transverse momentum:

  No b-Tag low: $30 \leq p_T(\tau_h) \leq 45$ GeV
  No b-Tag medium: $45 < p_T(\tau_h) \leq 60$ GeV
  No b-Tag high: $p_T(\tau_h) > 60$ GeV

The *SVfit* mass distributions before performing the maximum likelihood fit in the new MSSM categories are displayed in figures 7.18 and 7.19.

The simultaneous maximum likelihood fit is then performed over the five categories in this analysis. The "b-tag" categories are the most sensitive and the "no b-tag" categories, specially the low and medium, are used mainly to constrain the background estimation and nuisance parameters.

The improvement in the analysis sensitivity obtained by categorizing the events using $p_T(\tau_h)$ are shown in figure 7.20 for the $e\tau_h$ and $\mu\tau_h$ channels and in figure 7.21 for the $\tau_h\tau_h$ channel and for the combination of all channels. A significant gain in sensitivity is seen for all channels: about 20% for $e\tau_h$, 30% for $\mu\tau_h$ and 40% for $\tau_h\tau_h$, over the full Higgs mass range considered. The improvement in sensitivity corresponds to an increase in equivalent increase in luminosity by a factor 1.5 to 2.

7.5.3 Better treatment of the model dependence related to the uncertainty on the Higgs boson p_T

In the MSSM framework, the transverse momentum of h, H and A bosons produced via the gluon fusion process $gg \to \phi$ depends on the relative contributions of top, bottom, stop and sbottom quarks to the loop. These contributions are then sensitive to the Higgs couplings to up and down type quarks and on the masses of the stop and sbottom particles. The analysis is sensitive to the predicted Higgs p_T spectrum through the acceptance cuts and the categorization based on the transverse momentum of the τ_h p_T. The signal yields after the τ p_T acceptance cuts are expected to have a dependance on the MSSM model parameters, in particular on $\tan\beta$, because the τ p_T is correlated with the Higgs transverse momentum. At small (large) $\tan\beta$, the Higgs boson is mainly produced via top (bottom) loops in the gluon fusion process. The consequence of the mass difference between top and bottom quarks is that the Higgs p_T distribution is softer (harder) if $\tan\beta$ is large (small). Events with high Higgs p_T have a higher probability to pass the visible p_T cuts.

The modeling of the Higgs p_T distribution is improved firstly by reweighting the signal events generated with PYTHIA [167] to the spectrum computed at NLO plus next-to-next-to-leading logarithm (NNLL) by the HQT [191, 192] program.

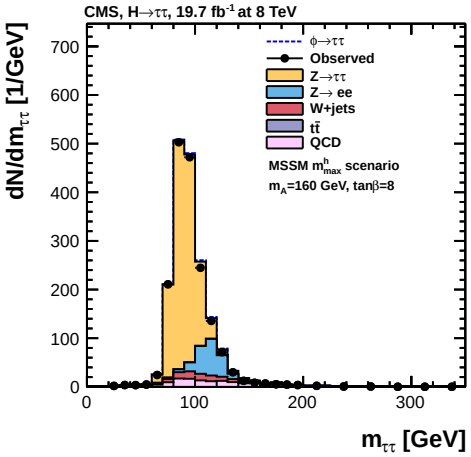
In order to take the finite top and bottom mass approximation (used in the previous computation) and the stop and sbottom contributions to the loop into account, an additional reweighting factor is computed:

$$\left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t,m_b} \Big/ \left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t=\infty}$$

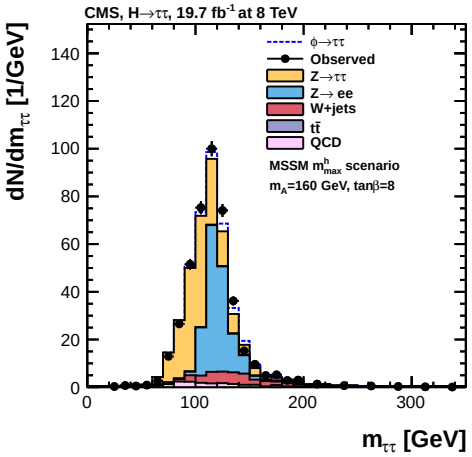
computed at LO accuracy using the program HIGLU [169].

Then, the full weight for a $gg \to \Phi$ signal event of given P_T^Φ at generator level is applied:

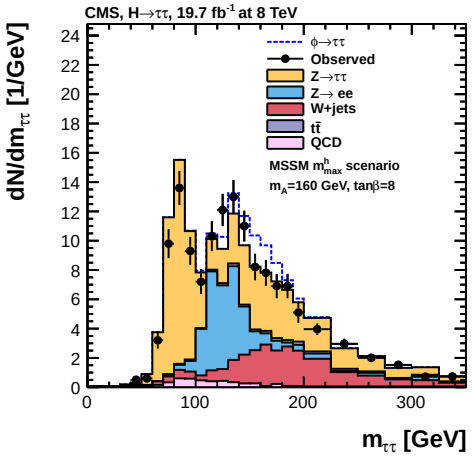
$$w_{Higgsq_T} = \left(\left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t=\infty}^{HqT} \Big/ \left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t=\infty}^{\text{PYTHIA}} \right) \cdot \left(\left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t,m_b}^{\text{HIGLU}} \Big/ \left.\frac{d\sigma}{dP_T^\Phi}\right|_{m_t=\infty}^{\text{HIGLU}} \right) \tag{7.4}$$



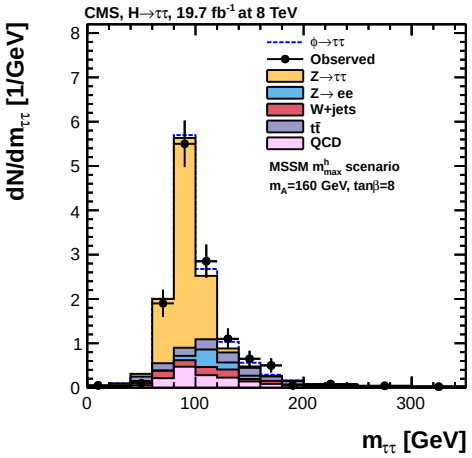
(a) *SVfit* Mass : $e\tau_h$ no b-Tag low



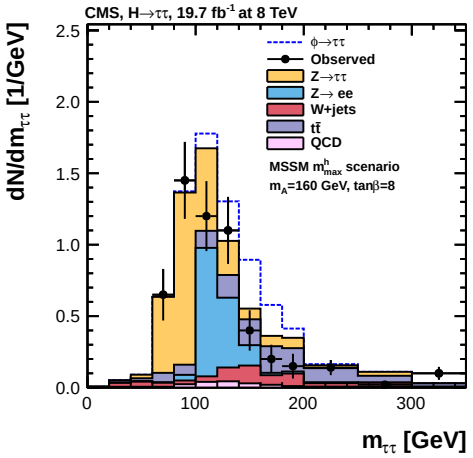
(b) *SVfit Mass* : $e\tau_h$ no b-Tag medium



(c) *SVfit Mass* : $e\tau_h$ no b-Tag high

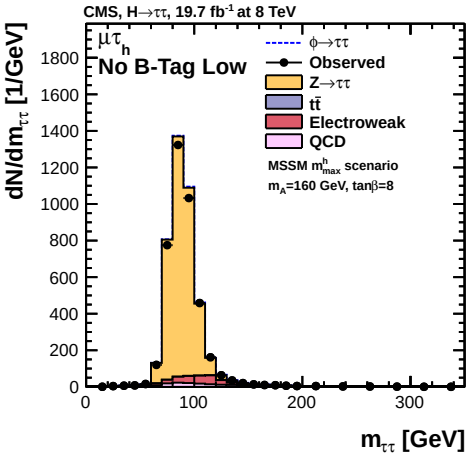


(d) *SVfit Mass* : $e\tau_h$ b-Tag low

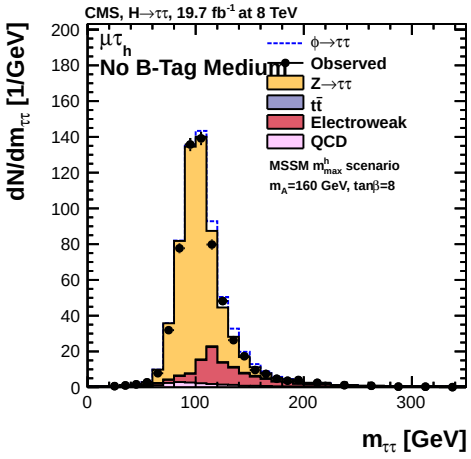


(e) *SVfit Mass* : $e\tau_h$ b-Tag high

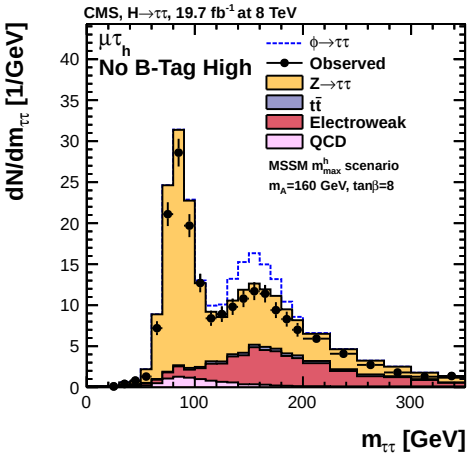
Figure 7.18: *Di-τ mass distribution for the 5 categories used in the MSSM analysis in the* $e\tau_h$ *channel before the maximum likelihood fit is performed.*



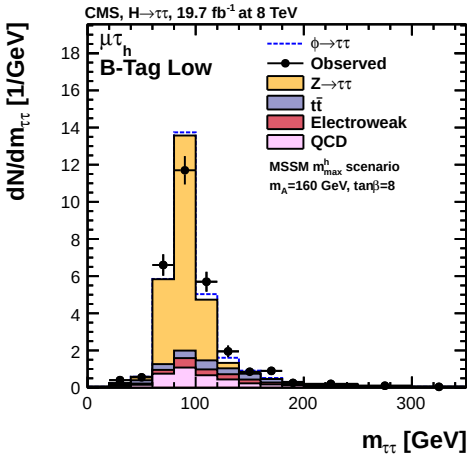
(a) *SVfit* Mass : $\mu\tau_h$ no b-Tag low



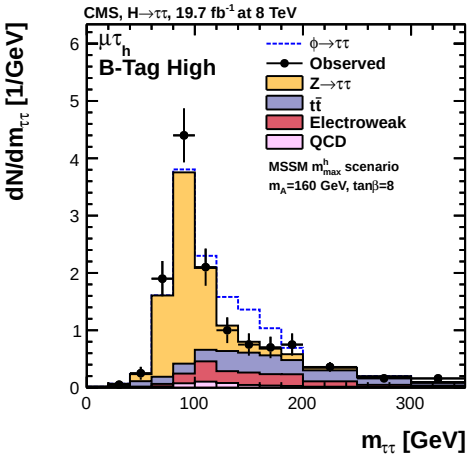
(b) *SVfit Mass* : $\mu\tau_h$ no b-Tag medium



(c) *SVfit Mass* : $\mu\tau_h$ no b-Tag high



(d) *SVfit Mass* : $\mu\tau_h$ b-Tag low



(e) *SVfit Mass* : $\mu\tau_h$ b-Tag high

Figure 7.19: *Di-τ mass distribution for the 5 categories used in the MSSM analysis in the $\mu\tau_h$ channel before the maximum likelihood fit is performed.*

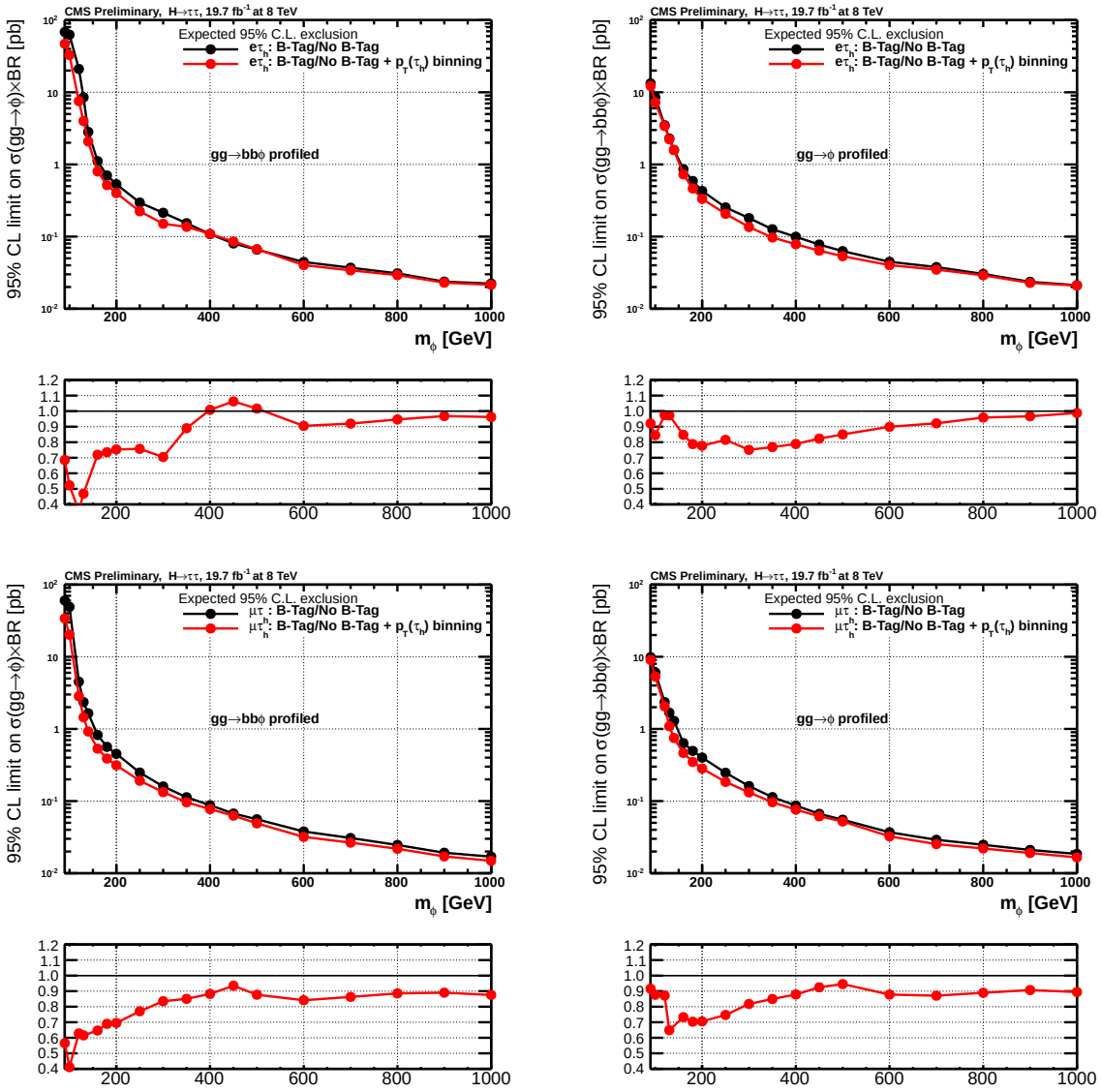


Figure 7.20: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal. The sensitivity of the* $e\tau_h$ *channel (top) and* $\mu\tau_h$ *channel (bottom) in the* $\sqrt{s} = 8$ *TeV data is compared for two cases: using the event categories of Ref. [162] (denoted by "B-Tag/No B-Tag" in the legend) and using the new categorization based on tau* p_T *described in the text (denoted by "B-Tag/No B-Tag +* $p_T(\tau_h)$ *binning" in the legend). The ratio between the two curves is displayed at the bottom of each plot.*

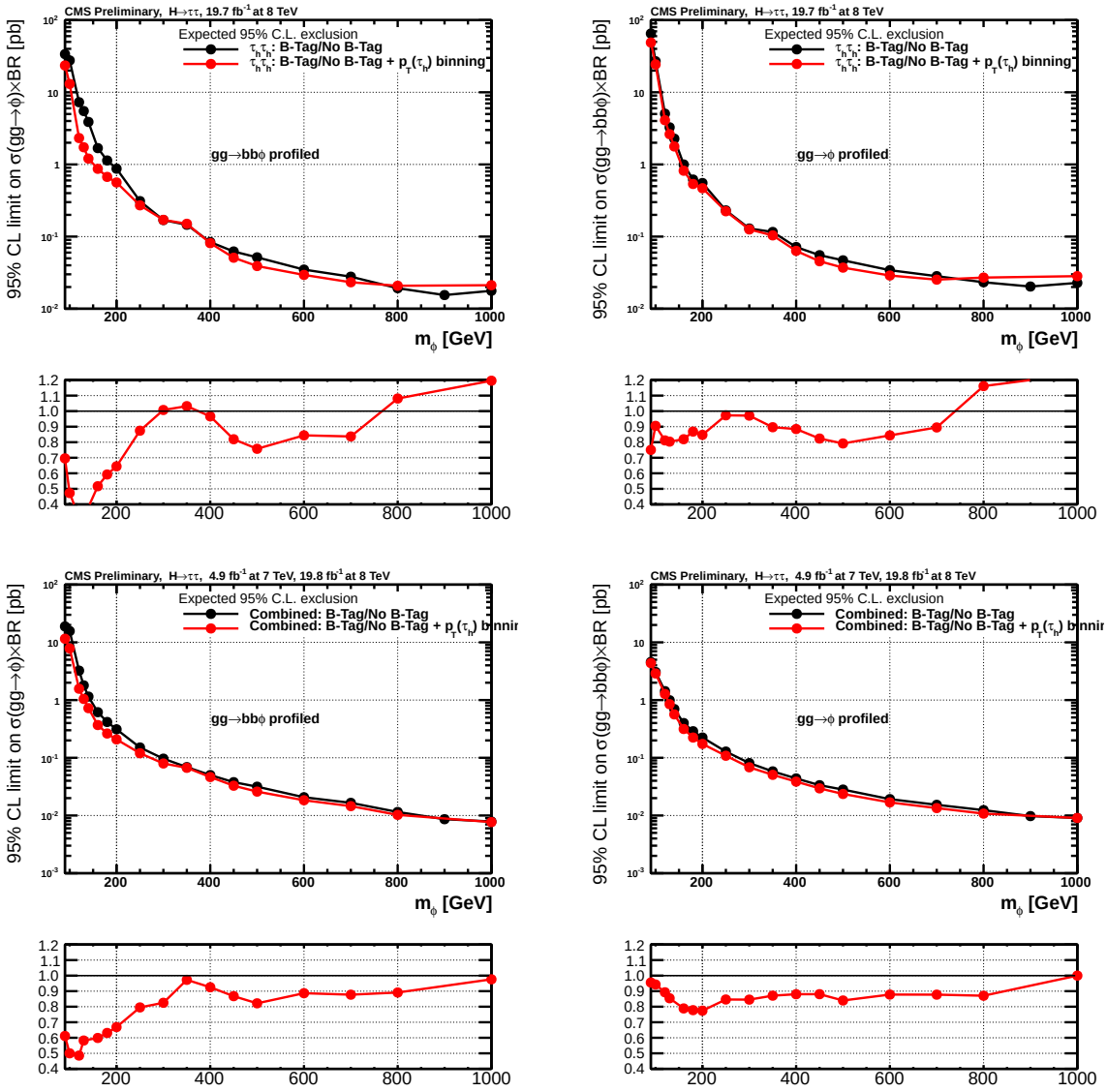


Figure 7.21: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal. The sensitivity of the* $\tau_h\tau_h$ *channel in the* $\sqrt{s} = 8$ *TeV data (top) and of the combination of all channels and all CMS run 1 data (bottom) is compared for two cases: using the event categories of Ref. [162] (denoted by "B-Tag/No B-Tag" in the legend) and using the new categorization based on tau* p_T *described in the text (denoted by "B-Tag/No B-Tag +* $p_T(\tau_h)$ *binning" in the legend). The ratio between the two curves is displayed at the bottom of each plot.*

The reweighting factor taken is the average of the differential cross-sections computed for $\tan\beta = 1$ (pure top loop) and $\tan\beta = 60$ (pure bottom loop). In figure 7.22, the effect of applying the Higgs p_T reweighting on the Higgs p_T and τ_h p_T distributions is presented. As expected, the Higgs p_T distribution becomes softer and the transverse momenta of the τ_h decay products is mostly affected for the values of $p_T \sim \frac{1}{2}m_\phi$.

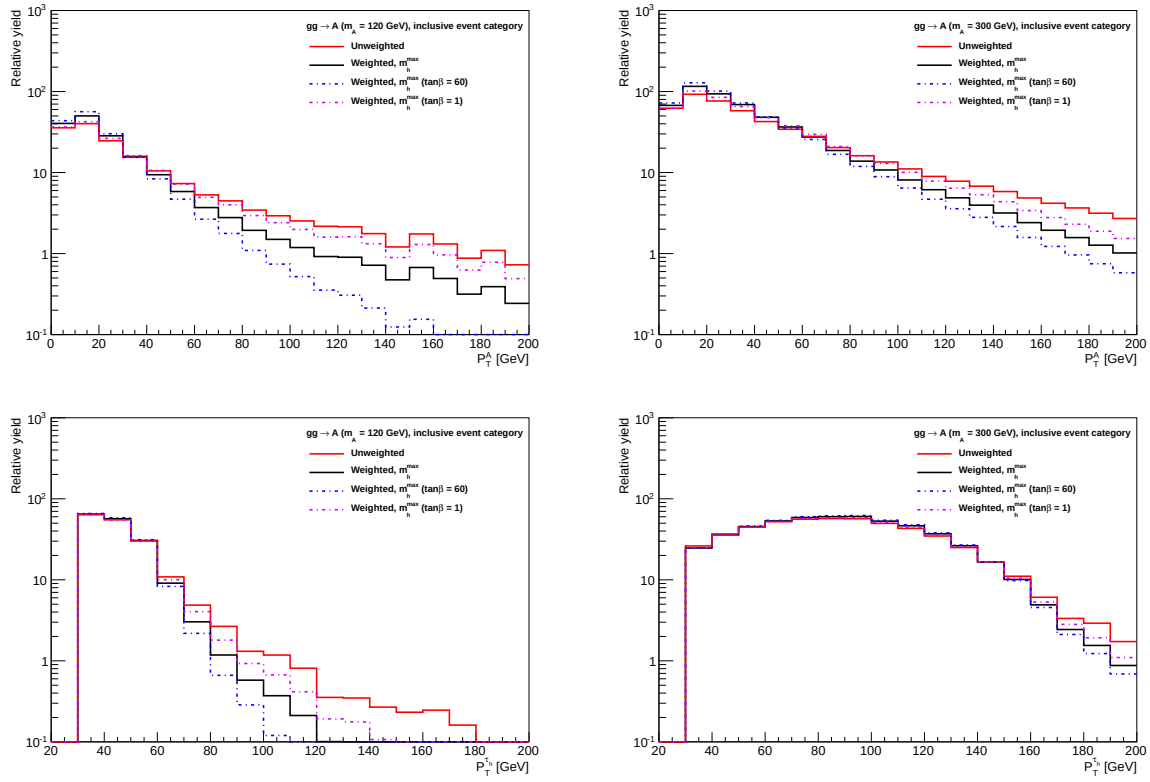


Figure 7.22: *Distribution in transverse momentum of the pseudoscalar Higgs boson A (top) and hadronic tau (bottom) in* $gg \to A \to \tau\tau \to \mu\tau_h$ *signal events before (red line) and after (black line) the reweighting factor, Eq. 7.4, is applied. Signal events are generated by* PYTHIA *for* $m_A = 120$ *GeV(left) and* $m_A = 300$ *GeV(right). The systematic uncertainty associated with the Higgs* P_T *reweighting is represented by dashed lines.*

In this section, the impact of the reweighting on the sensitivity of the analysis is studied. Altering the Higgs p_T spectrum affects the number of signal events that pass the selection cuts (*acceptance*) and that enter the different categories (*migrations*). The effect of migrations is expected to be larger in case events are analyzed in categories based on $p_T(\tau_h)$. It will be demonstrated, however, that the Higgs p_T reweighting has a sizable effect also when no categorization by $p_T(\tau_h)$ is used.

As discussed in section 6.4.13 and seen in figure 7.22 the reweighting makes the Higgs p_T spectrum softer. As a consequence, the sensitivity of the analysis is expected to be reduced since the reweighted spectrum is more background-like than the unweighted one. The effect of the p_T reweighting, and the associated uncertainty, is studied in two cases:

- Using the MVA τ_h isolation and the old categorization. The results are shown in figure 7.23.
- Using the MVA τ_h isolation and the new categorization using $p_T(\tau_h)$. The results are shown in figure 7.24.

In both cases the reweighting factors are computed for the m_h^{max} MSSM benchmark scenario (*cf.* appendix A). The central value is obtained by averaging the predictions for $\tan\beta = 1$ and $\tan\beta = 60$. Plus and minus one sigma variations around the central value are obtained by taking the spectrum computed for $\tan\beta = 1$ and $\tan\beta = 60$ respectively.

As expected, the effect of the reweighting is more pronounced for small Higgs masses, in particular in the region where $m_\phi \lesssim 200$ GeV. Comparing Figs. 7.23 and 7.24, it can be seen that the effect of the Higgs p_T reweighting is larger in case that tau p_T categories are used, but the effect is in fact sizable in both cases (old and new categories). For the old (new) categorization, in the $e\tau_h$ and $\mu\tau_h$ channels, the maximal effect on the limit is $\sim 5\%$ ($\sim 10\%$), while it is $\sim 40\%$ (up to $>> 100\%$) in the $\tau_h\tau_h$ channel. The effect of the systematic uncertainty associated with the reweighting is only visible in the $\tau_h\tau_h$ channel, and remains small ($< 5\%$) with respect to the effect of the reweighting itself.

7.5.4 Summary of the changes and their effects

The summary of the improvements in analysis sensitivity obtained by categorizing the events using $p_T(\tau_h)$ and by using the MVA based τ isolation are shown in figure 7.25 for the $e\tau_h$ and $\mu\tau_h$ channels, and in figure 7.26 for the $\tau_h\tau_h$ channel and for the combination of all channels.

The analysis presented in this thesis includes three major updates with respect to previous MSSM $\to \tau\tau$ analyses by CMS. The improvement in the hadronic tau isolation algorithm as well as the categorization of events selected in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels each enhances the sensitivity of the analysis corresponding to an increase by a factor 1.5-2 in equivalent luminosity. The combined effect of the two improvements taken together[4] is shown in figure 7.27.

The total improvement in analysis sensitivity is seen to be $40 - 50\%$, compatible with taking the product of the improvement factors given in sections 7.5.1 and 7.5.2, demonstrating that the improvements due to the MVA hadronic tau isolation algorithm and categorizing selected events in tau p_T factorize. Over the whole range of Higgs mass values considered the analysis presented in this thesis improves the sensitivity of the search for neutral MSSM Higgs bosons in the $\phi \to \tau\tau$ decay channel with respect to previous publications [162] by an amount that is equivalent to increasing the luminosity by a factor 3-4.

The Higgs p_T reweighting is seen to have a significant effect on the expected limits. The effect is more sizable in case $p_T(\tau_h)$ categories are used, but even without using the categorization in $p_T(\tau_h)$ the Higgs p_T reweighting plus the systematic uncertainty

[4]The comparison is done using the so-called m_H^{max} scenario: for a given value of the common SUSY-breaking sfermion masses (M_{SUSY}), and for a maximal value of $M_A = 1$ TeV, the other SUSY parameters are tuned to maximize M_h as a function of $\tan\beta$. Other scenarios are detailed in appendix A.

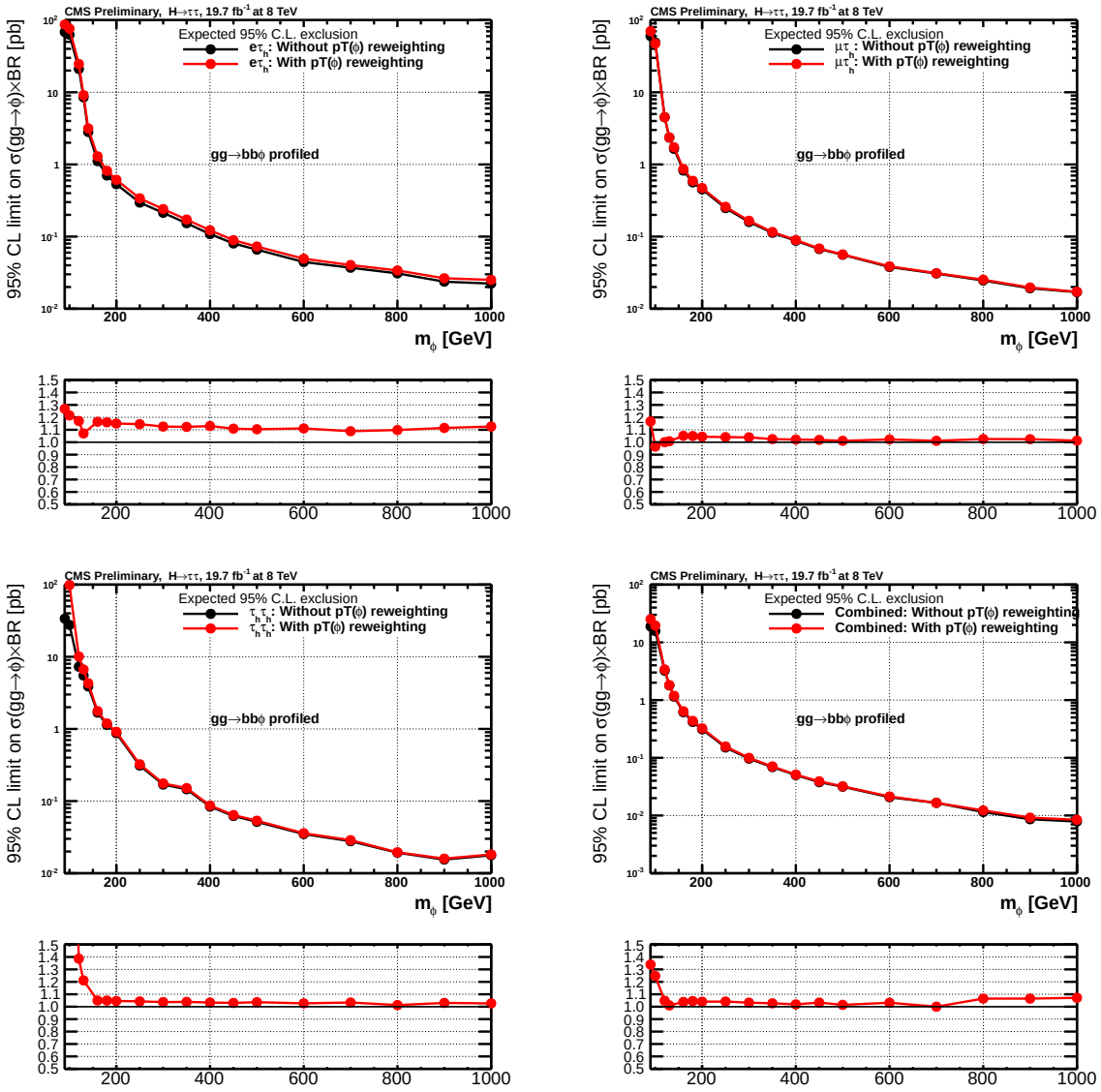


Figure 7.23: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *that we expect to set in the absence of a MSSM Higgs signal, for the* $e\tau_h$ *(top left),* $\mu\tau_h$ *(top right) and* $\tau_h\tau_h$ *(bottom left) channels in the* $\sqrt{s} = 8$ *TeV data and for the combination of all channels and all CMS run 1 data (bottom right). Expected limits are computed for the case that the event categories of Ref. [162] are used and are computed in two cases: with no Higgs* P_T *reweighting or with the reweighting applied. The ratio between the two curves is displayed at the bottom of each plot.*

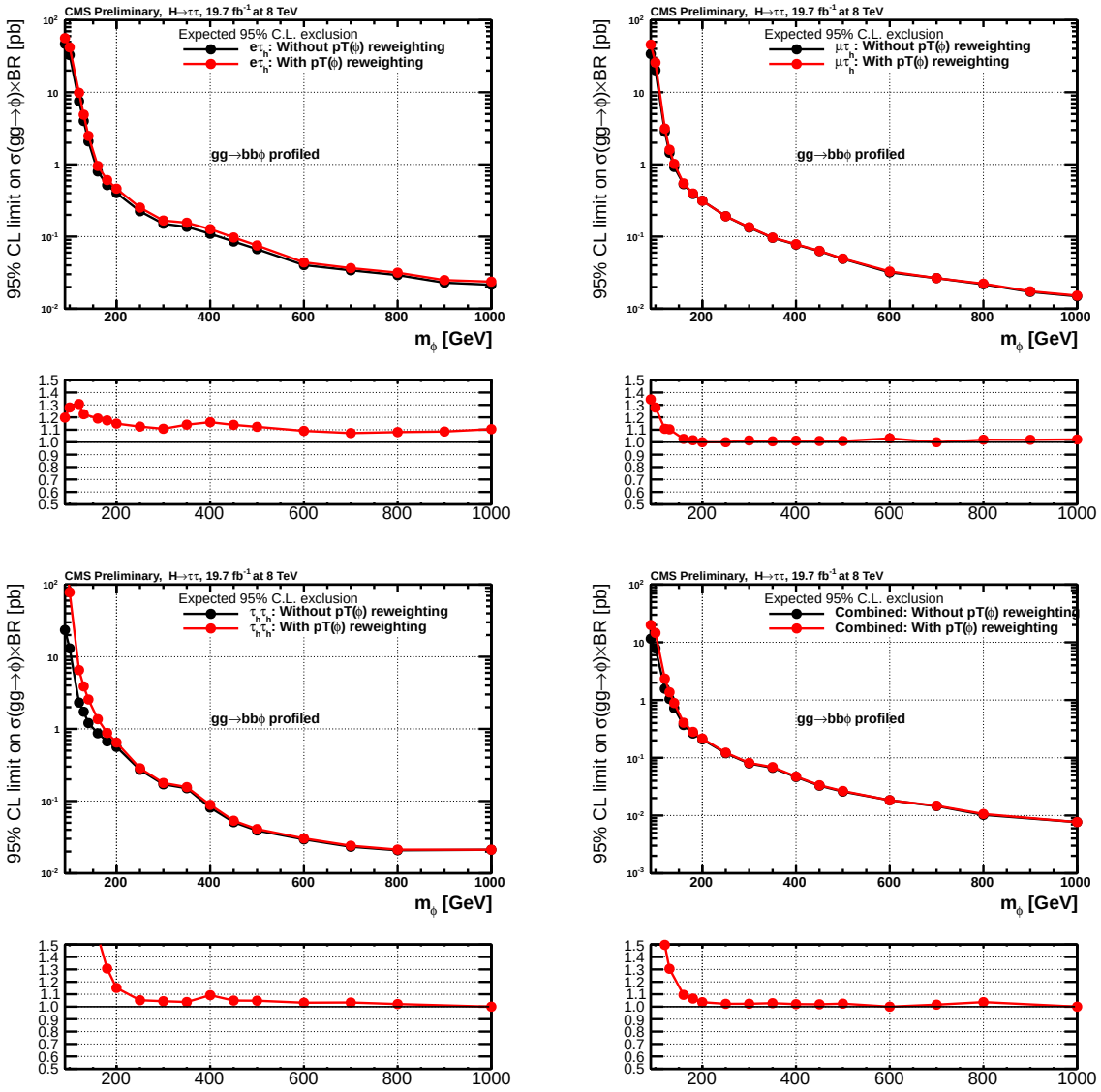


Figure 7.24: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *that we expect to set in the absence of a MSSM Higgs signal, for the* $e\tau_h$ *(top left),* $\mu\tau_h$ *(top right) and* $\tau_h\tau_h$ *(bottom left) channels in the* $\sqrt{s} = 8$ *TeV data and for the combination of all channels and all CMS run 1 data (bottom right). Expected limits are computed for the case that the event categories based on tau* p_T *are used and are computed in two cases: with no Higgs* p_T *reweighting or with the reweighting applied. The ratio between the two curves is displayed at the bottom of each plot.*

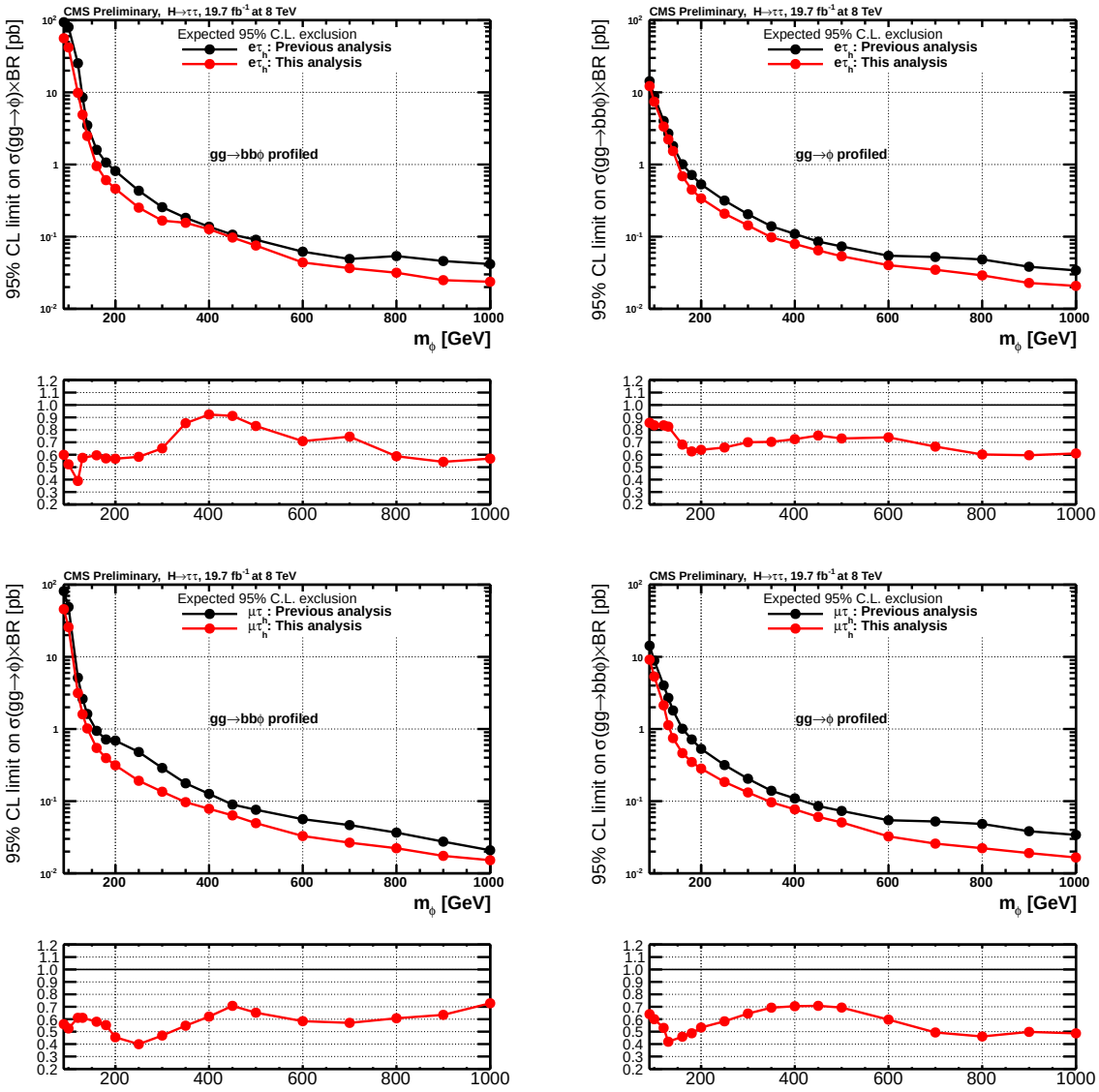


Figure 7.25: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal, for the* $e\tau_h$ *(top) and* $\mu\tau_h$ *(bottom) channels in the* $\sqrt{s} = 8$ *TeV data. Expected limits are computed using the cut-based* τ_h *isolation and old categorization and using the MVA* τ_h *isolation and new categorization with Higgs* p_T *reweighting applied. The ratio between the two curves is displayed at the bottom of each plot.*

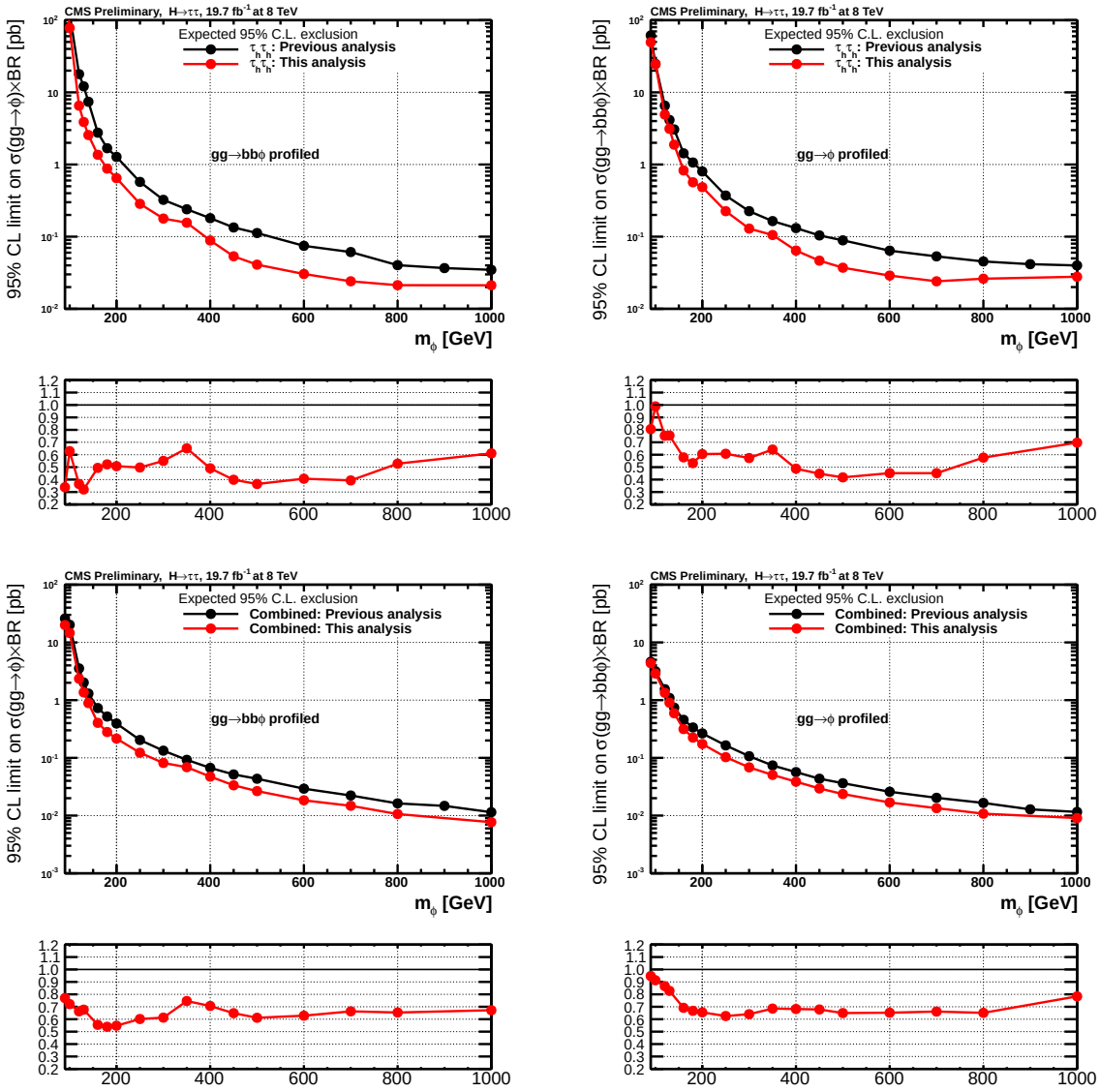


Figure 7.26: 95% *CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi bb \to \tau\tau$ *(right) that we expect to set in the absence of a MSSM Higgs signal, for the* $\tau_h\tau_h$ *(top) and for the combination of all channels and all CMS run 1 data. Expected limits are computed using the cut-based* τ_h *isolation and old categorization and using the MVA* τ_h *isolation and new categorization with Higgs* p_T *reweighting applied. The ratio between the two curves is displayed at the bottom of each plot.*

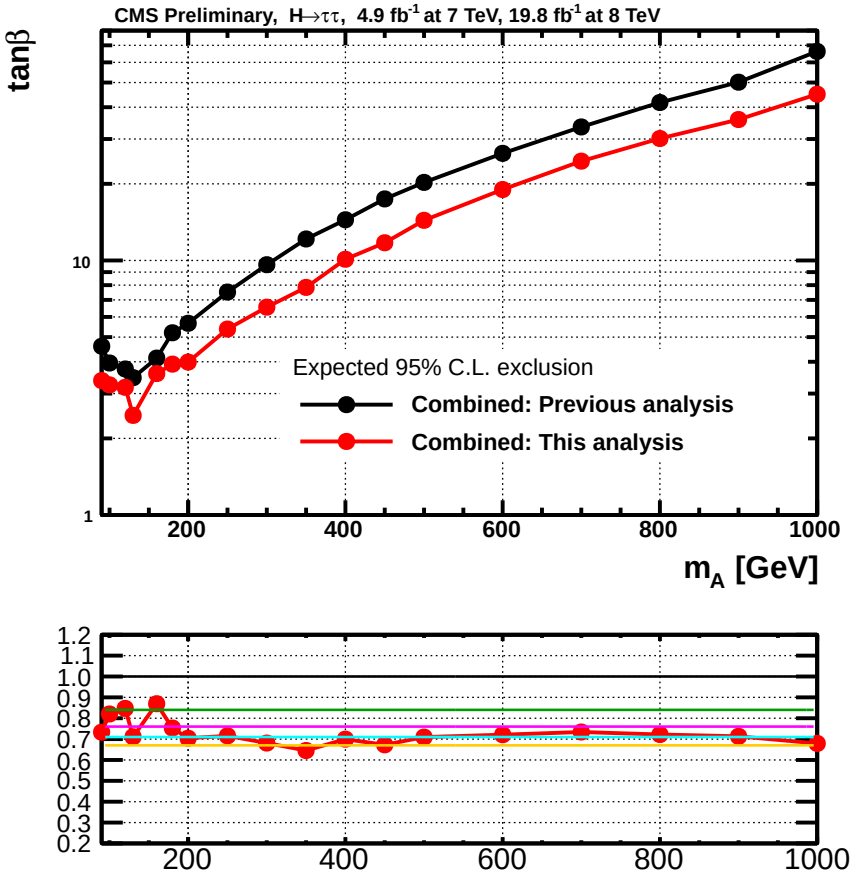


Figure 7.27: *Region in* m_A, $\tan\beta$ *parameter space expected to be excluded at* 95% *CL by the search, in the the combination of all the channels. Expected limits are compared in two cases (left): using the Cut-based* τ_h *isolation and old categorization and using the MVA* τ_h *isolation and new categorization. In the ratio plots at the bottom, respectively in terms of improvement and equivalent luminosity increase, the horizontal green line corresponds to 16% and a factor 2, the magenta line to 24% and a factor 3, the cyan line to 29% and a factor 4, and the yellow line to 33% and a factor 5.*

associated to it needs to be taken into account for the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels in order to make the analysis model independent.

7.6 Evidence in the search for the SM Higgs decaying to τ leptons

In 2013, the CMS collaboration decided to update the results of the search for the Standard Model Higgs boson decaying in two τ-leptons. The improvement brought by the new anti-electron discriminator exposed in this thesis (chapter 5) was included (among others). Therefore, I present in this section the results obtained and published in Ref. [161] following the Moriond 2013 International conference.

The di-τ mass distributions after the maximum likelihood fit are presented and the combination with the other final states channels is shown.

7.6.1 Post-fit di-τ mass distributions

The maximum likelihood fit procedure takes into account the nuisance parameters and their correlation and is used to extract the signal strength modifier parameter μ. The pull of a nuisance parameter is defined as:

$$pull = \frac{postfit - prefit}{prefit}$$

In figures 7.28 and 7.29 the biggest pulls for the $e\tau_h$ and $\mu\tau_h$ channels are displayed. The post-fit values of the nuisance parameters are all within one σ of the expected value showing a good control over the fit. In figures 7.30, 7.31 and 7.32 the di-τ mass distribution are shown, after the maximum likelihood fit is performed. The modeling of the background is improved after the fit.

7.6.2 Combination of all $H \to \tau\tau$ final states

The search of the Standard Model Higgs boson decaying to two tau leptons is performed in CMS in all the six possible decay channels ($\mu\tau_h$, $e\tau_h$, $e\mu$, $\tau_h\tau_h$, ee and $\mu\mu$) [161]. This search makes use of the whole Run 1 data collected by CMS, 4.9 (19.7) fb^{-1} at $\sqrt{s}$ = 7 TeV ($\sqrt{s}$ = 8 TeV). In addition, an independent search looking for Higgs produced in association with a vector boson Z or W [204] is combined in the final result.

The most sensitive channels are the semi-leptonic $e\tau_h$ and $\mu\tau_h$ and the most sensitive categories the 1-jet and VBF tagged. In figure 7.33, the sensitivity of the analysis split in channels and categories is shown. We can note that the analysis is sensitive to a Higgs boson produced with the rates predicted by the Standard Model in all the mass range considered.

The di-τ mass distributions can be combined between the $\mu\tau_h$, $e\tau_h$, $e\mu$ and $\tau_h\tau_h$ channels. The mass distribution of the combined $e\tau_h$,$\mu\tau_h$,$e\mu$ and $\tau_h\tau_h$ channels is displayed in figure 7.34. The distribution is obtained by weighting each category of each channel by the $S/(S+B)$ ratio, with S the expected signal yield of a Standard Model Higgs boson with $m_H = 125$ GeV ($\mu = 1$) and B the expected background contribution resulting from the maximum likelihood fit, obtained in the central $m_{\tau\tau}$ interval containing 68% of the signal events. An excess of events with respect to the background expectation is visible.

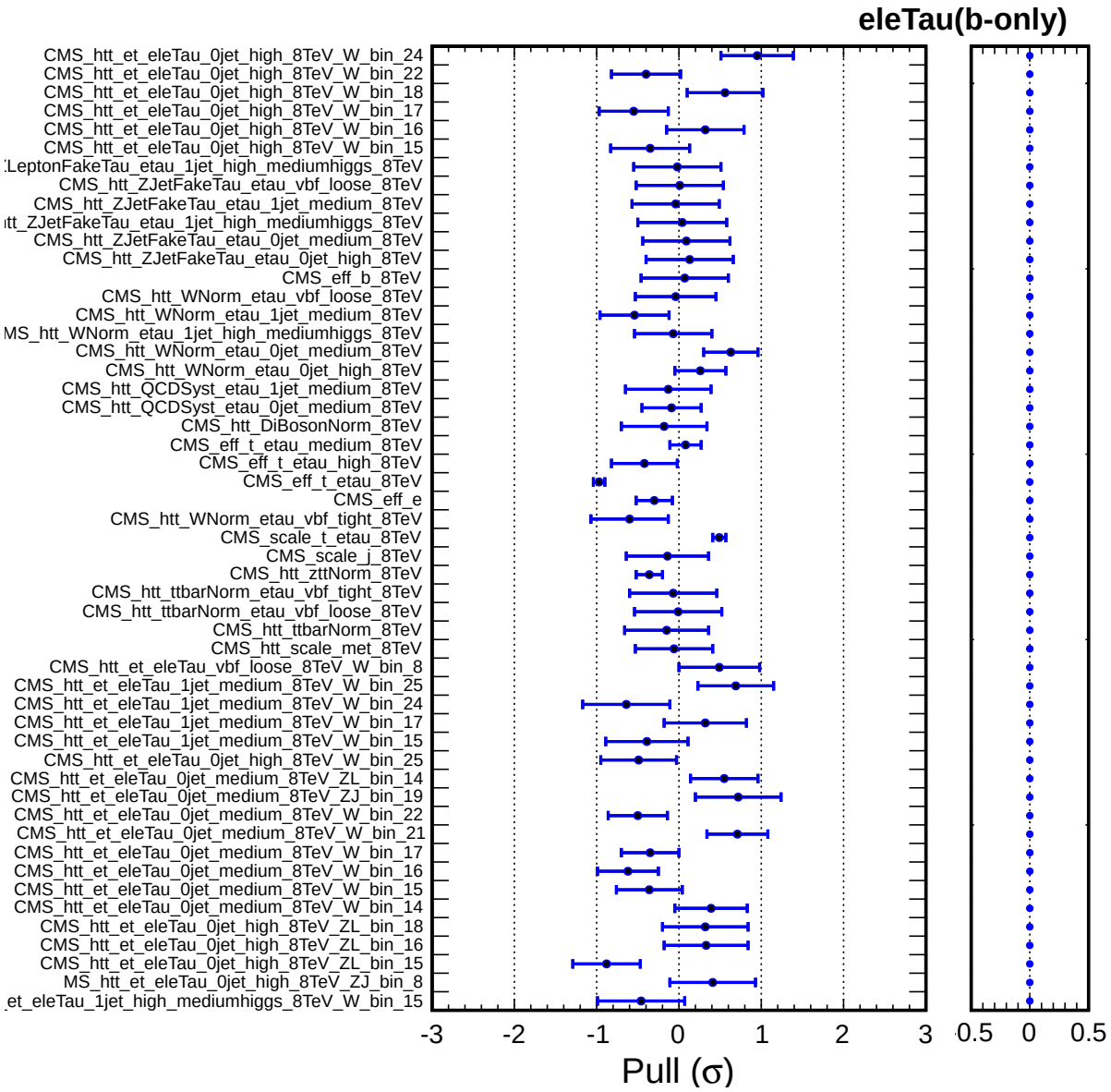


Figure 7.28: *Pulls on the nuisance parameters for the* $e\tau_h$ *channel.*

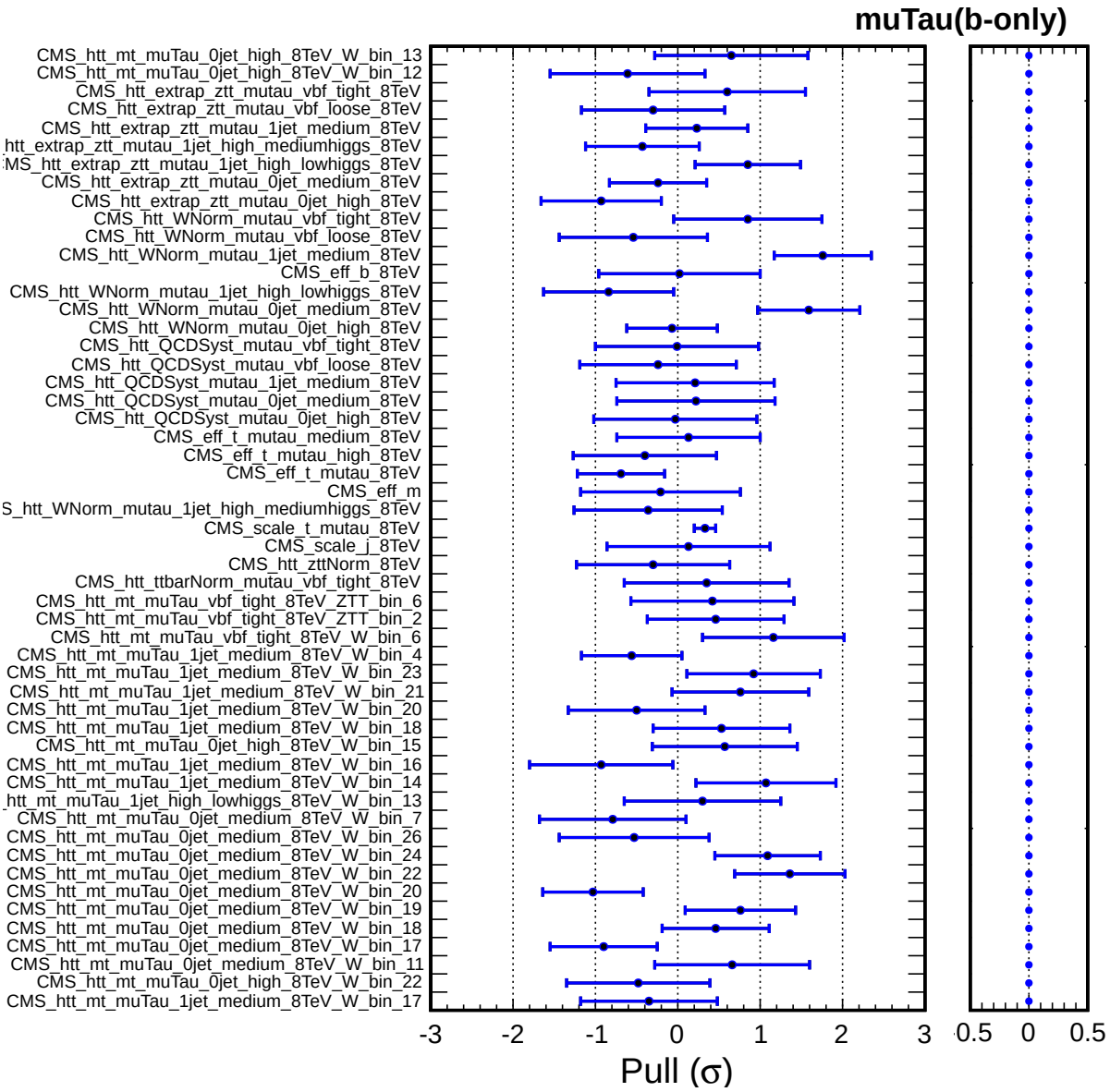
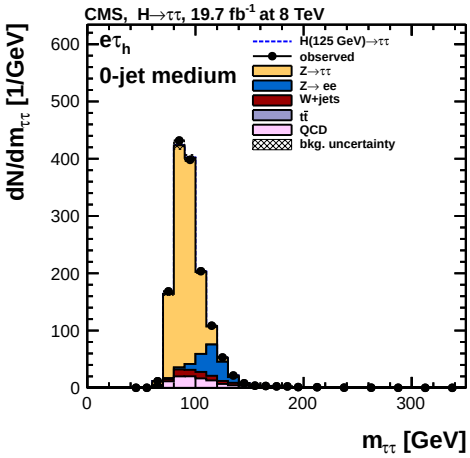
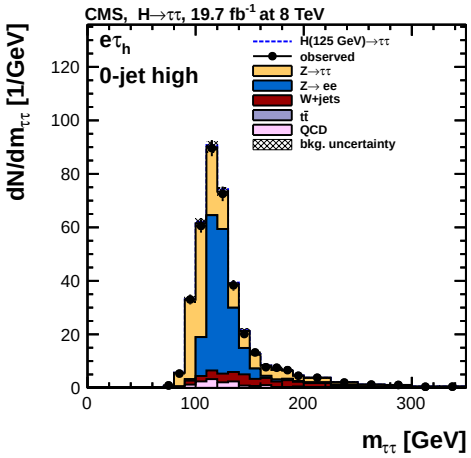


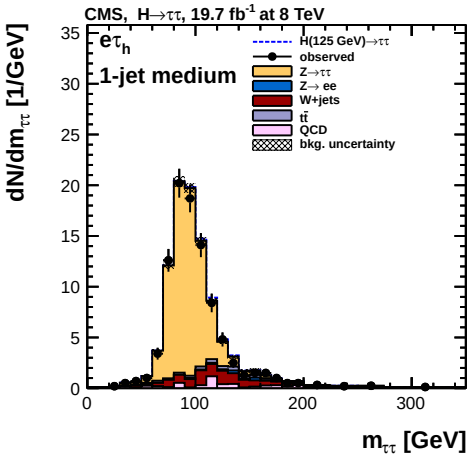
Figure 7.29: *Pulls on the nuisance parameters for the $\mu\tau_h$ channel.*



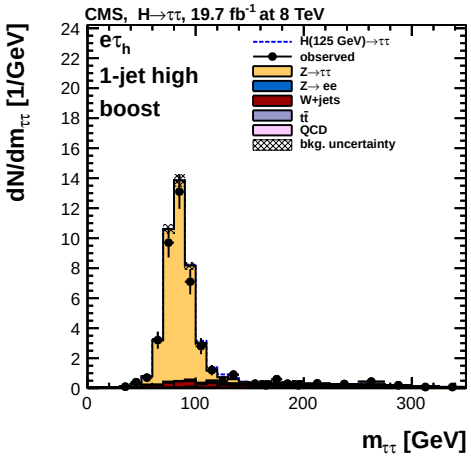
(a) *SVfit* Mass : $e\tau_h$ 0-jet medium



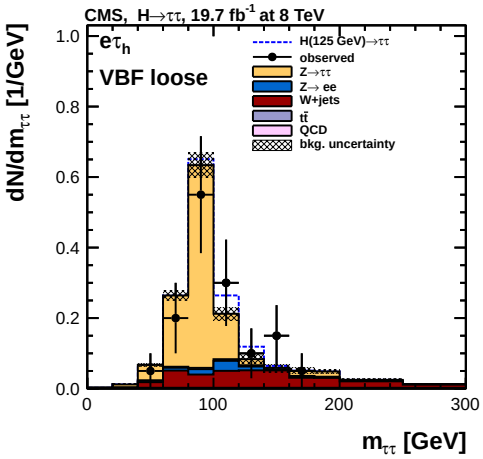
(b) *SVfit Mass* : $e\tau_h$ 0-jet high



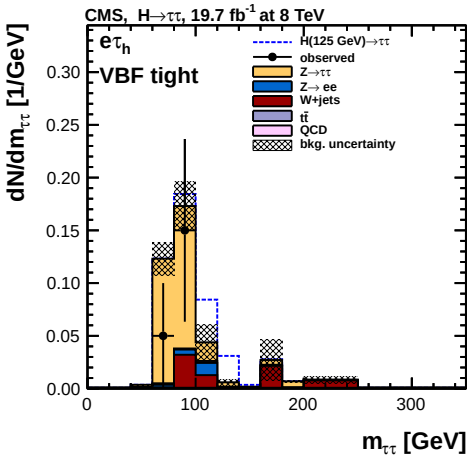
(c) *SVfit Mass* : $e\tau_h$ 1-jet medium



(d) *SVfit Mass* : $e\tau_h$ 1-jet high ↑

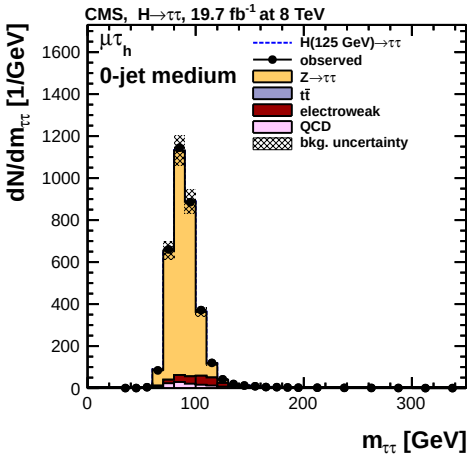


(e) *SVfit Mass* : $e\tau_h$ vbf loose

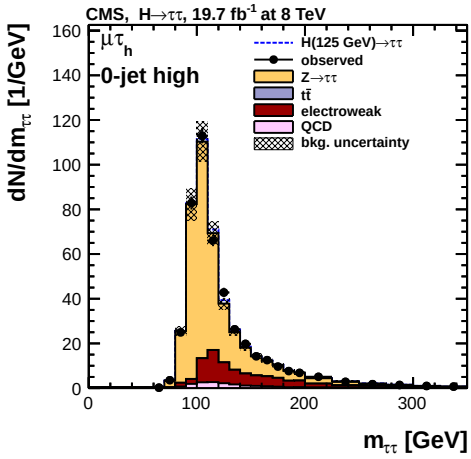


(f) *SVfit Mass* : $e\tau_h$ vbf tight

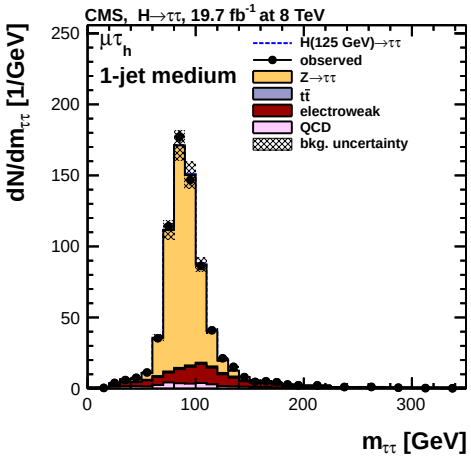
Figure 7.30: *Di-τ mass distribution for the categories used in the Standard Model analysis in the* $e\tau_h$ *channel after the maximum likelihood fit is performed.*



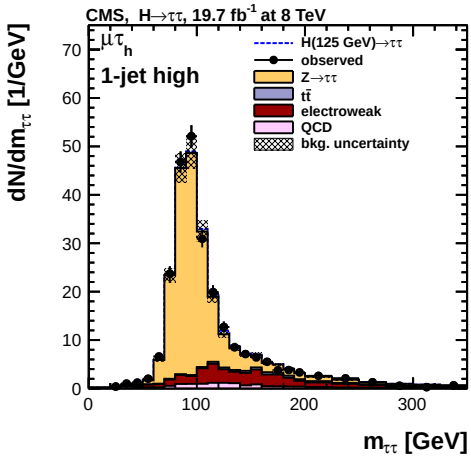
(a) *SVfit* Mass : $\mu\tau_h$ 0-jet medium



(b) *SVfit Mass* : $\mu\tau_h$ 0-jet high

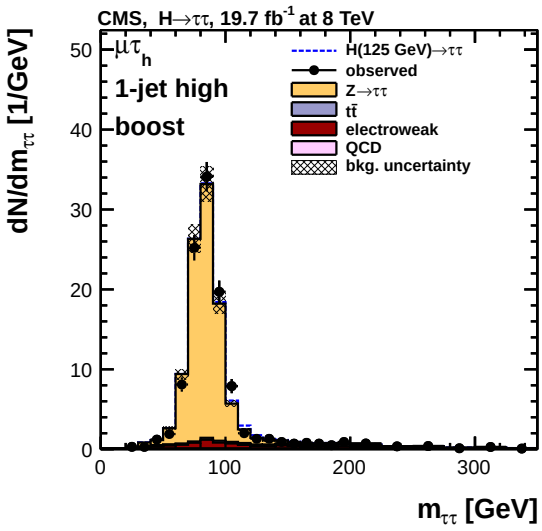


(c) *SVfit Mass* : $\mu\tau_h$ 1-jet medium

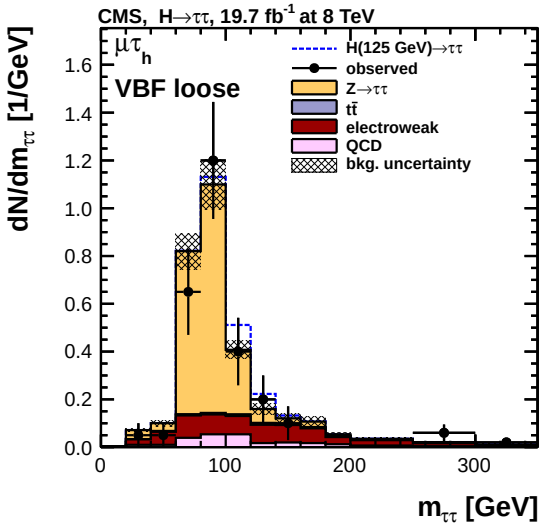


(d) *SVfit Mass* : $\mu\tau_h$ 1-jet high ↓

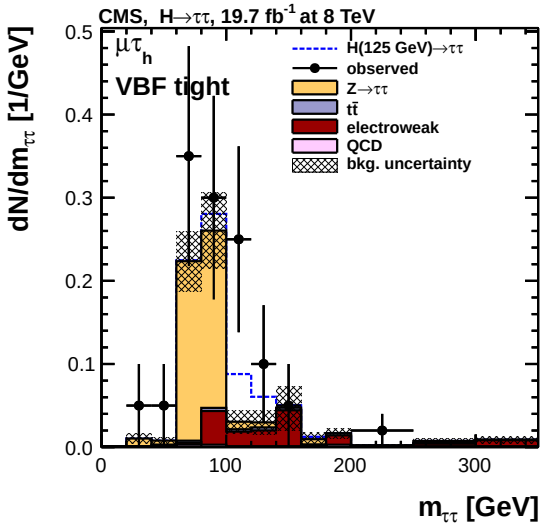
Figure 7.31: *Di-τ mass distribution for the categories used in the Standard Model analysis in the $\mu\tau_h$ channel after the maximum likelihood fit is performed.*



(a) *SVfit Mass* : $\mu\tau_h$ 1-jet high ↑



(b) *SVfit Mass* : $\mu\tau_h$ vbf loose



(c) *SVfit Mass* : $\mu\tau_h$ vbf tight

Figure 7.32: *Di-*τ *mass distribution for the categories used in the Standard Model analysis in the* $\mu\tau_h$ *channel after the maximum likelihood fit is performed.*

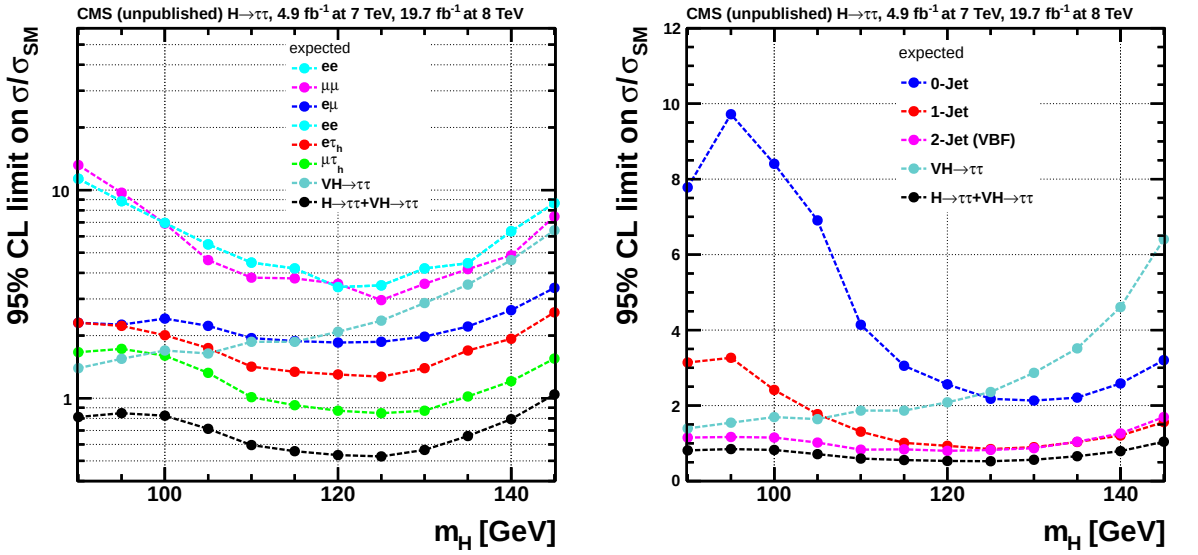


Figure 7.33: *Expected 95% CL upper limit on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ in the background only hypothesis, shown separately for the seven channels (left) and separately for 0-Jet, 1-Jet, VBF, VH categories (right). The black curve represents the combination of all decay channels (left) or categories (right). It is lower than 1 in the whole mass range, meaning that the analysis is sensitive to the SM Higgs boson.*

Using the same limit setting procedure as in the previous section, the combined expected and observed upper limits on the signal strength modifier μ at 95% CL level are set using the modified frequentist construction CL_S described in section 6.6.2. In figure 7.35 the expected and observed limits are displayed in the background-only hypothesis and in the signal-plus-background hypothesis for a Standard Model Higgs boson with $m_H = 125$ GeV. An excess of observed data prevents the exclusion of the Standard Model Higgs boson in the whole mass range, and it is compatible with the presence of a Standard Model Higgs boson with $m_H = 125$ GeV.

To quantify the excess, the local p-value is computed as a function of m_H. In figure 7.36, the expected and observed local p-values and corresponding significance are displayed. The minimum expected local p-value is found at $m_H = 125$ GeV corresponding to a significance of 3.7 standard deviations and the observed p-value is minimal for $m_H = 120$ GeV with a significance of 3.3 standard deviations. The observed significance is larger than 3 standard deviations in the mass range between 115 and 130 GeV, and equal to 3.2 standard deviations for $m_H = 125$ GeV. This represents a direct evidence for a Standard Model Higgs boson.

The best-fit value for μ, the value maximizing the likelihood of the combined fit is $\hat{\mu} = 0.78 \pm 0.27$ at $m_H = 125$ GeV. In figure 7.37, the best-fit value is shown separately for each final state and category. The figures show a good compatibility between the channels and categories with the expectation for a Standard Model Higgs boson with a

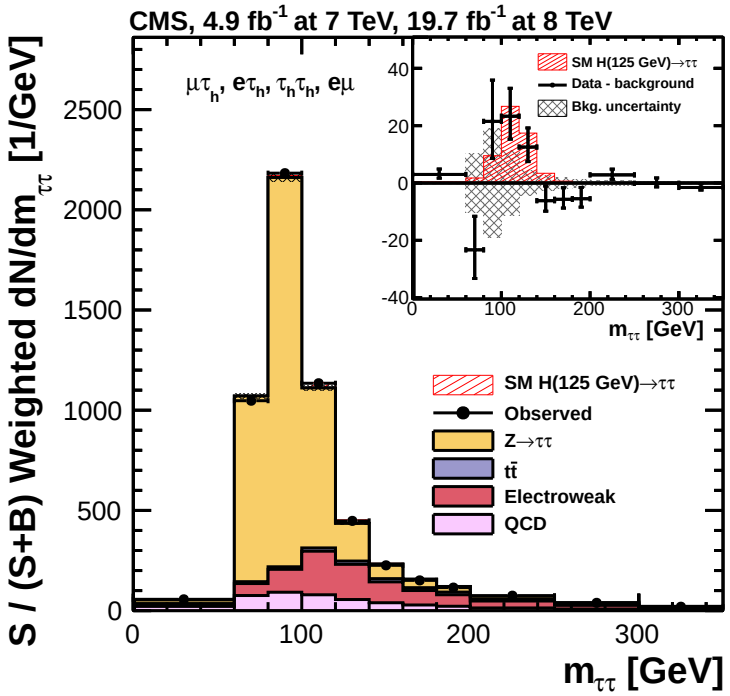


Figure 7.34: *Combined $m_{\tau\tau}$ distribution for the $\mu\tau_h$, $e\tau_h$, $e\mu$ and $\tau_h\tau_h$ channels. The distributions obtained in each category of each channel are weighted by the $S/(S+B)$ ratio. The inset shows the corresponding difference between the observed data and expected background distributions, together with the signal distribution for a SM Higgs boson at $m_H = 125$ GeV.*

mass of 125 GeV.

Despite the coarse granularity of the *SVfit* distribution (15-20% mass resolution), a mass measurement can be done by performing a scan of the negative log-likelihood difference $-2\ln\mathcal{L}$ as a function of m_H. For each point in the parameter space all nuisance and μ parameters are profiled. The best-fit mass is $m_H = 122 \pm 7$ GeV.

The scan on the effective Higgs boson couplings (κ_V, κ_f) [60] to vector bosons and fermions respectivelly is also performed. The effective Higgs boson couplings are defined as:

$$\kappa_i^2 \sim \frac{\Gamma_i}{\Gamma_{i(SM)}} \text{ and } \kappa_H^2 \sim \frac{\Gamma_{tot}}{\Gamma_{tot(SM)}} \tag{7.5}$$

The number of signal events observed is:

$$N(XX \to H \to YY) \sim \sigma(XX \to H).BR(H \to YY) \sim \frac{\Gamma_{XX} \cdot \Gamma_{YY}}{\Gamma_{tot}} \tag{7.6}$$

then:

$$n_{signal} \sim \frac{\kappa_{XX}^2 \cdot \kappa_{YY}^2}{\kappa_H^2} \tag{7.7}$$

The measurement is performed assuming the single narrow resonance hypothesis, using the zero-width approximation, neglecting the contribution of BSM particles in the loops

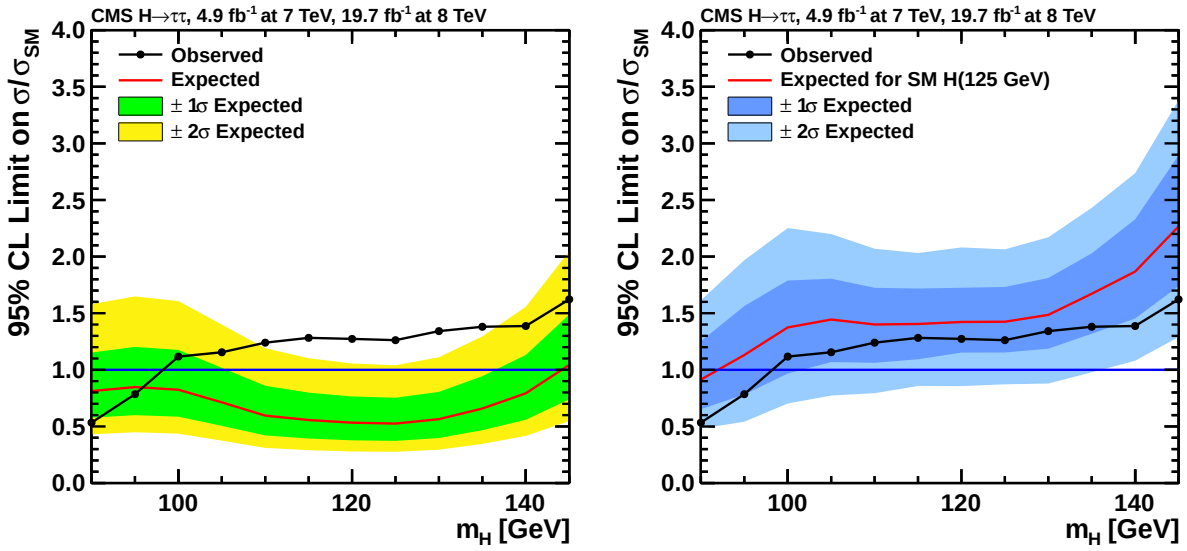


Figure 7.35: *Combined observed 95% CL upper limit on the signal strength parameter $\mu = \sigma/\sigma_{SM}$, together with the expected limit obtained in the background-only hypothesis (left) and the signal-plus-background hypothesis for a SM Higgs boson with $m_H = 125$ GeV (right). The bands show the expected one- and two-standard-deviation probability intervals around the expected limit.*

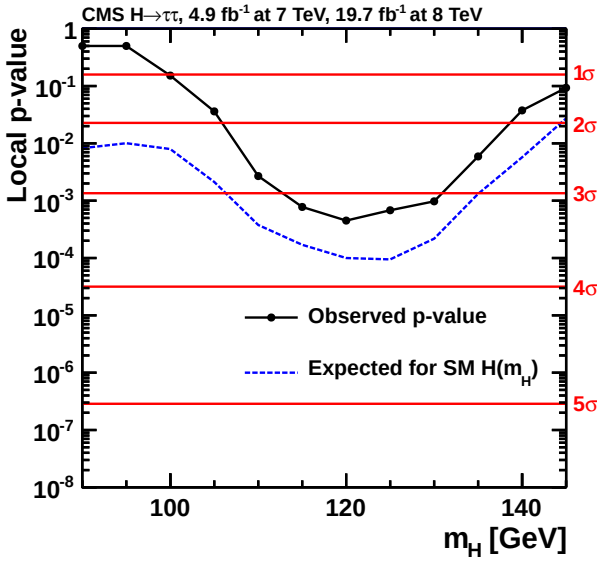


Figure 7.36: *Local p-value and significance in number of standard deviations as a function of the SM Higgs boson mass hypothesis for the combination of all decay channels. The observation (solid line) is compared to the expectation (dashed line) for a SM Higgs boson with mass m_H.*

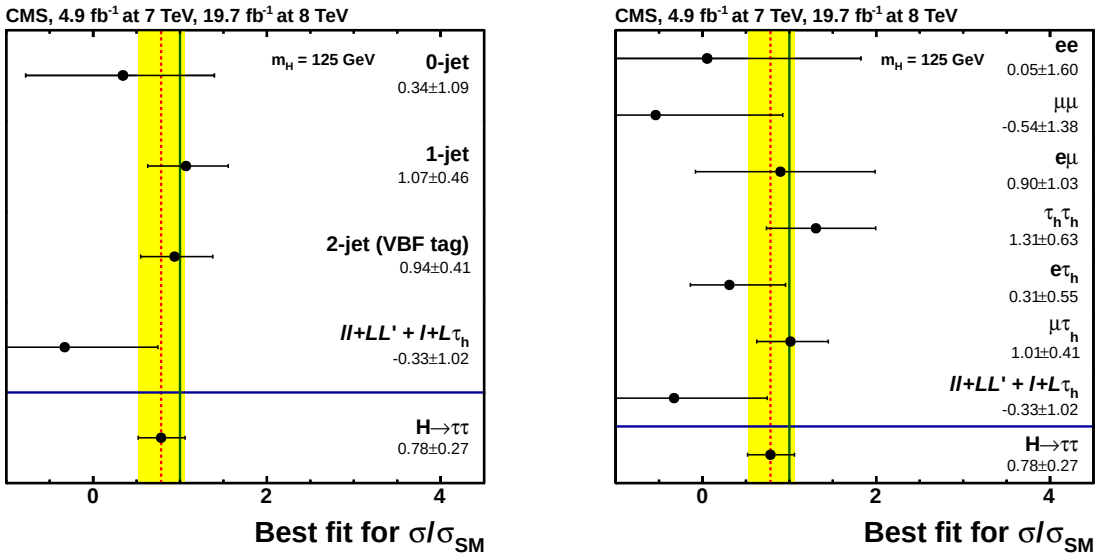


Figure 7.37: *Best-fit signal strength values, for independent channels, for mH=125 GeV. The combined value for the* $H \to \tau\tau$ *analysis in both plots corresponds to* $\hat{\mu}$*, obtained in the global fit combining all categories of all channels. The dashed line corresponds to the best-fit* μ *value. The* $ll + LL' + l + L\tau_h$ *represents the results for the VH analysis.*

and widths and assuming the effective couplings within fermions and within vectors are identical. The scan in the parameter space (κ_V, κ_f) characterizes the ratio between the measured and the expected Standard Model coupling of the Higgs boson to vector bosons and fermions. The two likelihood scans are presented in figure 7.38. The observed likelihood contour is compatible with the Standard Model expectation ($\kappa_V = \kappa_f = 1$) within two sigmas.

The results are the first evidence for the coupling of the 125 GeV Higgs boson discovered in 2012 by the ATLAS and CMS collaborations [71, 72]. The measurements of the observed 95% CL upper limits on the signal strength modifier, the best fit value and the likelihood scans are compatible with the Higgs boson predicted by the Standard Model. A combination of the $H \to \tau\tau$ and $H \to b\bar{b}$ have also been performed and it is reported here [212]. The result shows an evidence of the coupling of the discovered Higgs boson to fermions with an observed (expected) local significance of 3.9 (4.3) standard deviations for $m_H = 125$ GeV.

More data will be recorded during the Run 2, aiming as a first step to make an standalone discovery (observed significance in excess of 5 standard deviations) of the Higgs boson decaying to tau leptons. Measurement of the CP quantum numbers of the discovered particle can be attempted by studying the topology of the τ decay products [213]. In addition, a first measurement of the Higgs self coupling via the $HH \to b\bar{b}\tau\tau$ decay channel is in the CMS physics program. Another important quest in the Higgs sector is the search for Supersymmetry, and the $\tau\tau$ final state provides a strong sensitivity to the

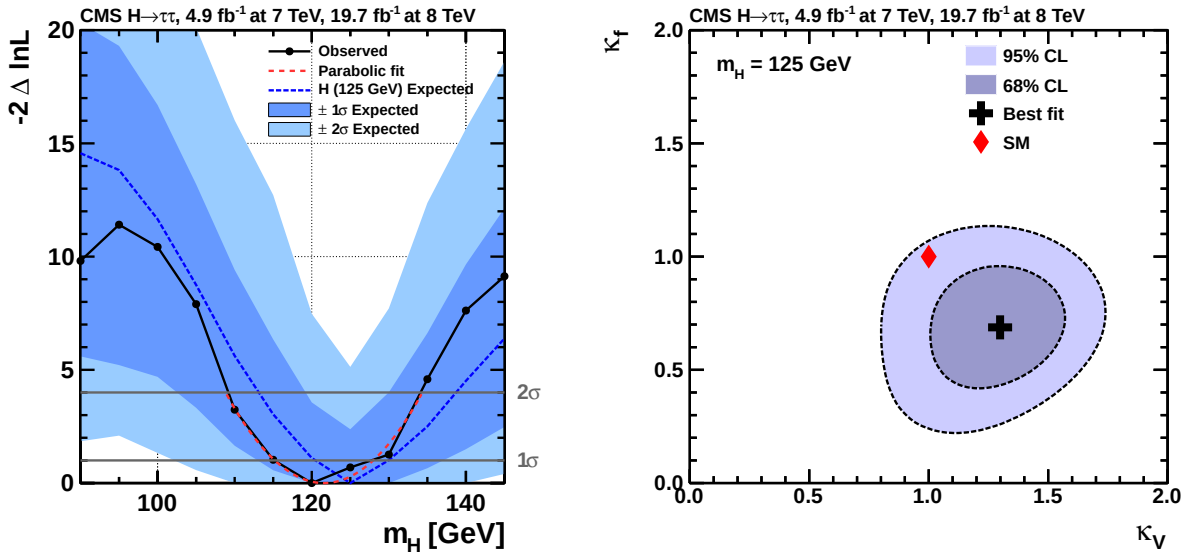


Figure 7.38: *Scan of the negative log-likelihood difference, $-2\ln\mathcal{L}$, as a function of mH (left). For each point, all nuisance parameters are profiled. The observation (solid line) is compared to the expectation (dashed line) for a SM Higgs boson with mass mH=125 GeV. Likelihood scan as a function of κ_V and κ_f (right). For each point, all nuisance parameters are profiled. The observation (black cross) is compared to the expectation (red lozenge) for a SM Higgs boson with mass $m_H = 125$ GeV.*

MSSM ϕ bosons ($\phi = h, A, H$) as we will present in next section.

7.7 Improved search for MSSM Higgs decaying to τ leptons

In this section, the results of the Minimal Supersymmetric Standard Model Higgs boson search are reported, including for the first time the analysis improvements[5] that were presented in section 7.5. The di-τ mass distributions after the maximum likelihood fit are presented as well as the corresponding statistical interpretation. The model independent 95% CL upper limits on cross section times branching ratio for the $gg \to \phi$ and $gg \to b\phi$ production modes are presented. Finally, the combination with the other final states channels is shown.

7.7.1 Post-fit di-τ mass distributions

The maximum likelihood fit procedure takes the nuisance parameters and their correlation into account and is used to extract the signal strength modifier parameter μ. In figures 7.39

[5]the improvements are used in the 8 TeV data only.

and 7.40 the biggest pulls for the $e\tau_h$ and $\mu\tau_h$ channels are displayed. The post-fit values of the nuisance parameters are almost all within one σ of the expected value showing a good control over the fit. In figures 7.41 and 7.42 the di-τ mass distribution are shown, after the maximum likelihood fit is performed. The modeling of the background is improved after the fit and no excess of data is observed in any of the categories.

7.7.2 Results

Model independent 1D cross section limits

Similarly to the Standard Model case, the test statistics defined in 6.6.1 are used. In the absence of a clear evidence of a signal in the mass distributions, we set limits on the production cross section times branching ratio for the two production mechanisms: the gluon fusion process $gg \to \phi \to \tau\tau$ and the b-associated production process $gg \to \phi b \to \tau\tau$. The upper limits for the gluon fusion process (b-associated production) are obtained by profiling the production rate for b-associated production (gluon fusion process). Therefore this computation is model independent as the production processes are treated independently. The results are presented in figure 7.43 for the $e\tau_h$ channel and figure 7.44 for the $\mu\tau_h$ channel. The observed limit is compatible with the expectation within two sigmas for all the masses considered for the search.

7.7.3 Combined results

The five final states covered by the CMS search for the MSSM Higgs boson decaying to tau pairs are $\mu\tau_h$, $e\tau_h$, $e\mu$, $\tau_h\tau_h$ and $\mu\mu$ [162]. The combined limit takes the results from the data recorded by CMS in 2011 and 2012, representing $4.9 fb^{-1}$ recorded at $\sqrt{s} = 7$ TeV and $19.7 fb^{-1}$ recorded at $\sqrt{s} = 8$ TeV. The MVA based isolation improves the sensitivity of the analysis in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels and is considered only in the 2012 data. For the data taken at $\sqrt{s} = 7$ TeV the categories considered are just the b-Tag and no b-Tag without splitting in $p_T(\tau_h)$.

The model independent 1D cross section limits are displayed in figure 7.45. The observed data is consistent with the background-only expectation within two sigmas for the whole mass range and for both processes. For comparison, the 1D cross section limits from the latest public result [162] are presented in figure 7.46. The analysis presented in this thesis explores a bigger phase space, leading to improved exclusion limits (see the details of the improvements in section 7.5.4).

This analysis puts stringent limits in the production cross-section times branching ratios for the signal processes $gg \to \phi \to \tau\tau$ and $gg \to \phi b \to \tau\tau$. More data from the next Run is needed to push the limits in the MSSM Higgs bosons search. The analysis presented in this thesis will be the baseline starting point adopted by CMS for analyzing the new data to come starting from 2015.

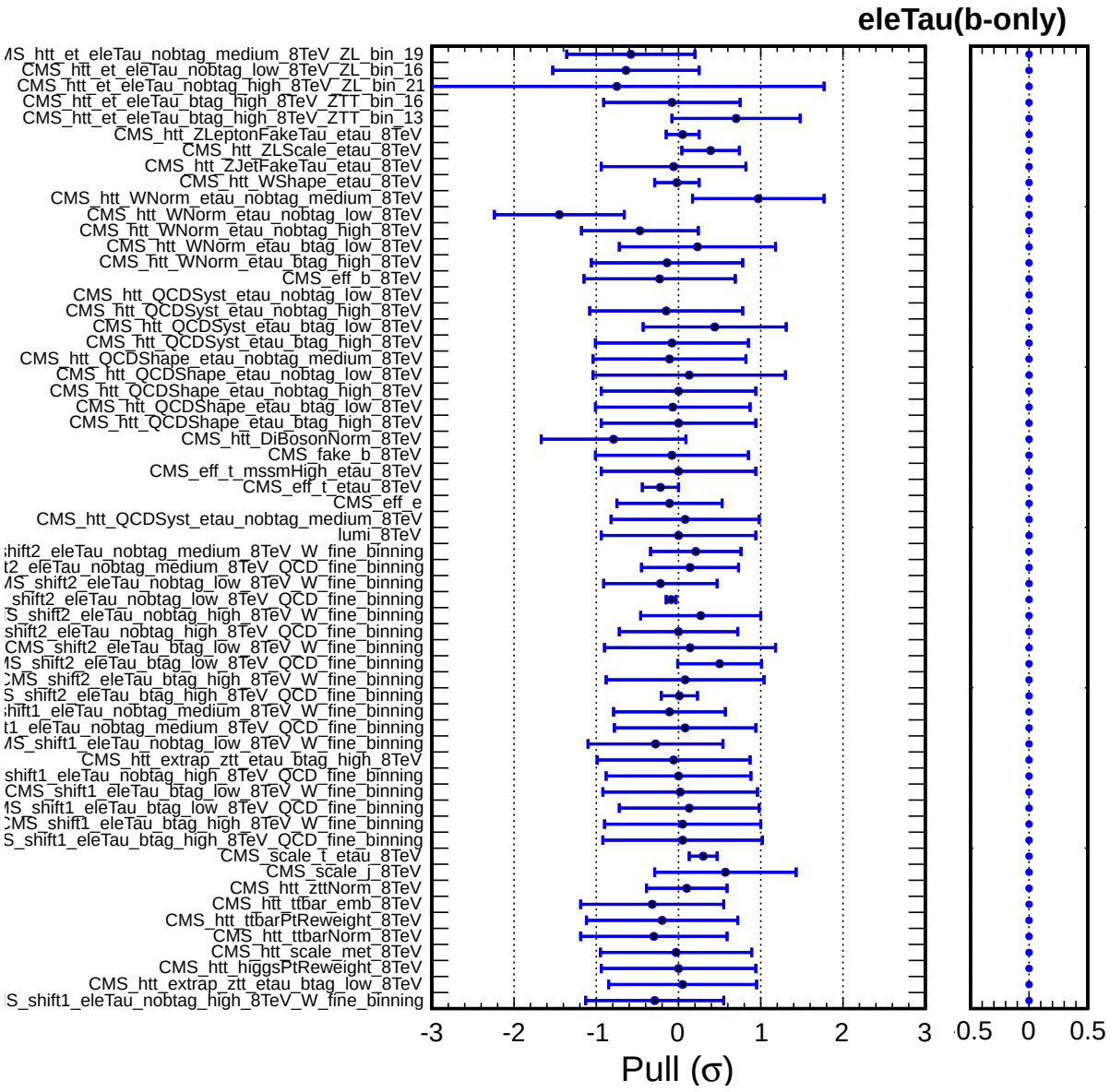


Figure 7.39: *Pulls on the nuisance parameters for the $e\tau_h$ channel.*

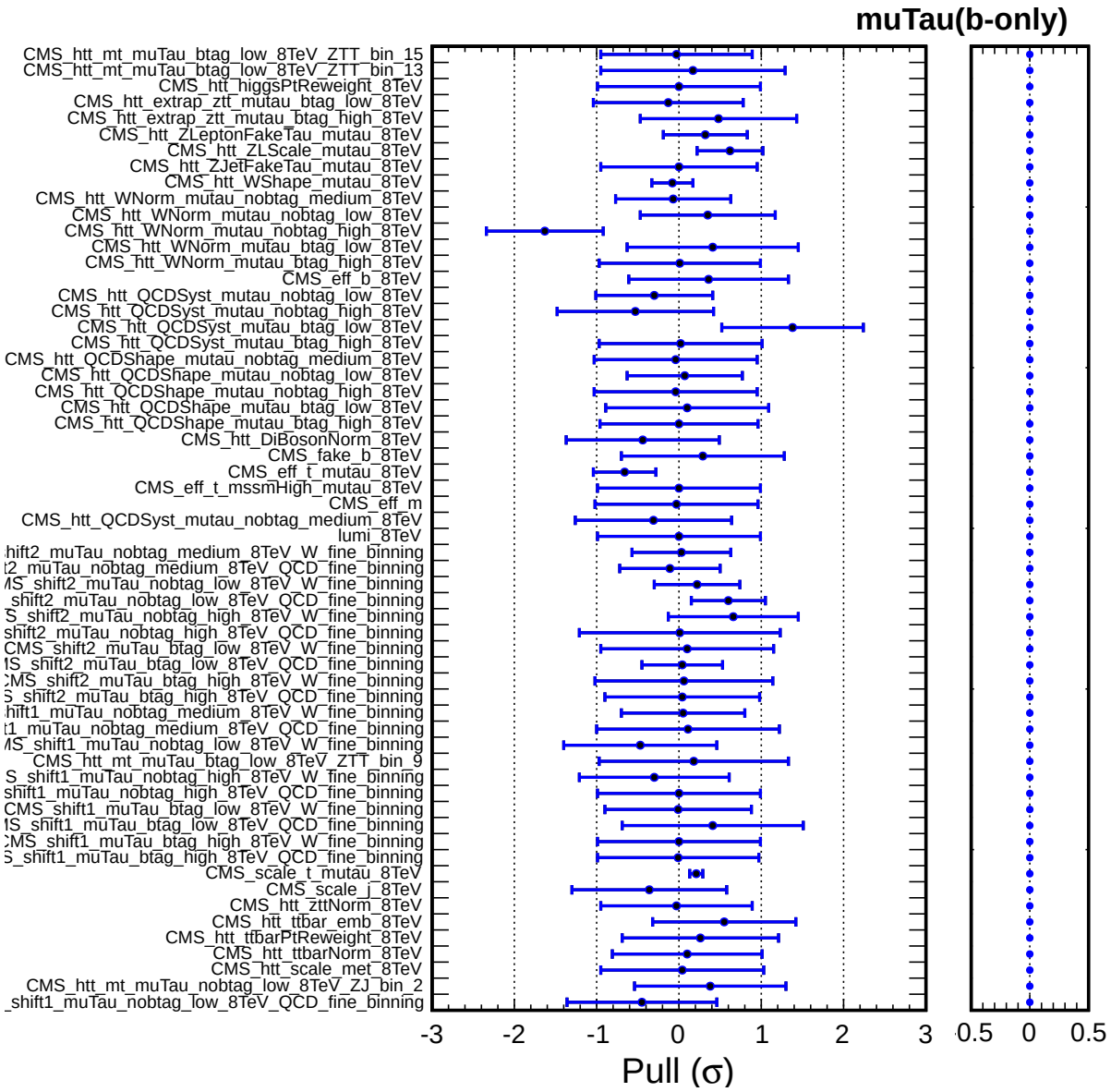
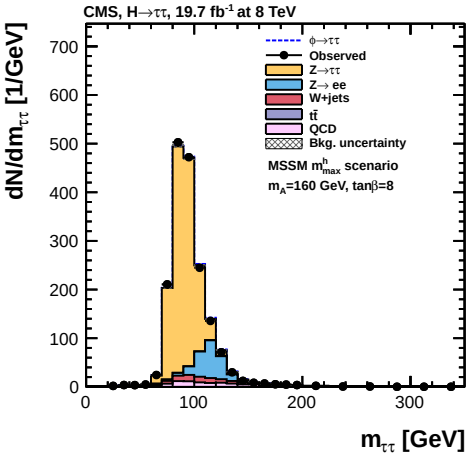
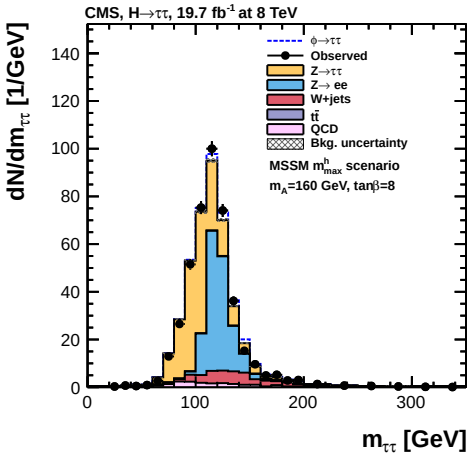


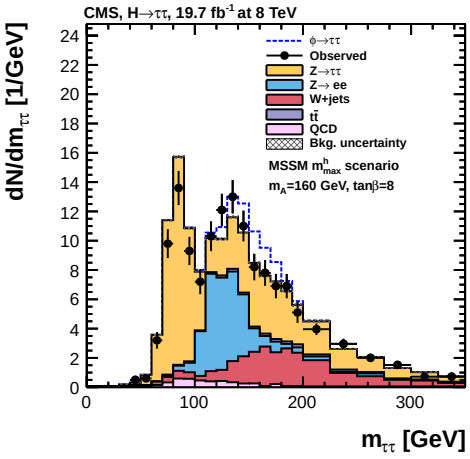
Figure 7.40: *Pulls on the nuisance parameters for the* $\mu\tau_h$ *channel.*



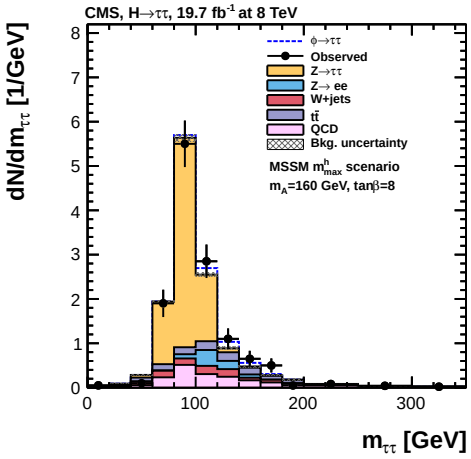
(a) *SVfit* Mass : $e\tau_h$ no b-Tag low



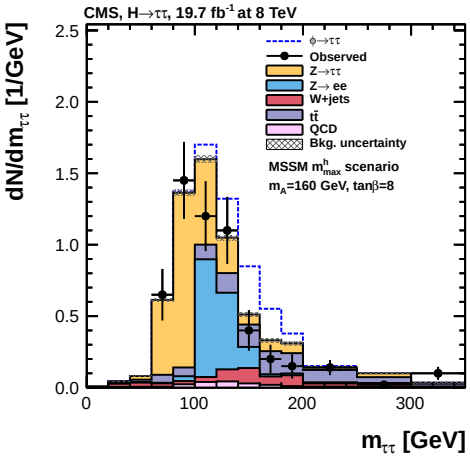
(b) *SVfit Mass* : $e\tau_h$ no b-Tag medium



(c) *SVfit Mass* : $e\tau_h$ no b-Tag high

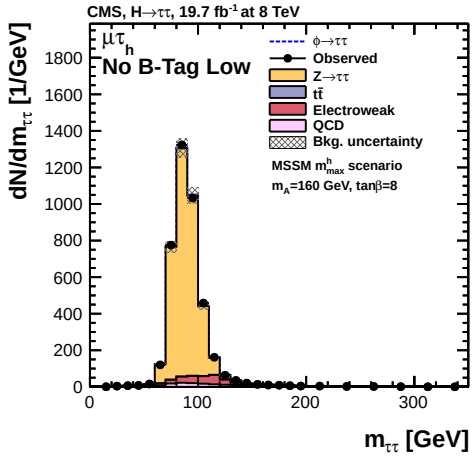


(d) *SVfit Mass* : $e\tau_h$ b-Tag low

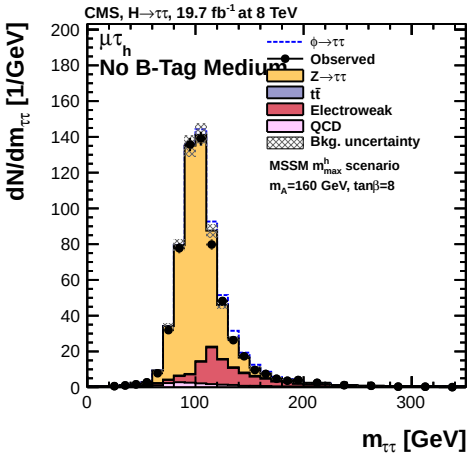


(e) *SVfit Mass* : $e\tau_h$ b-Tag high

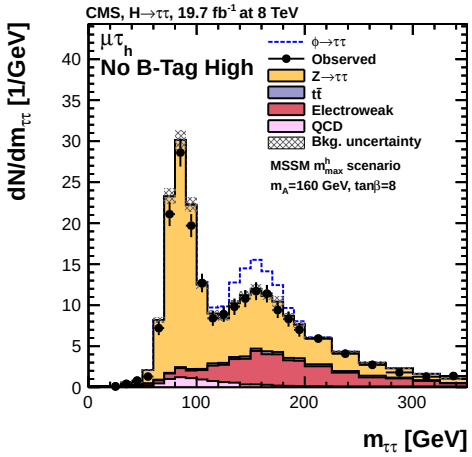
Figure 7.41: *Di-τ mass distribution for the 5 categories used in the MSSM analysis in the $e\tau_h$ channel before the maximum likelihood fit is performed.*



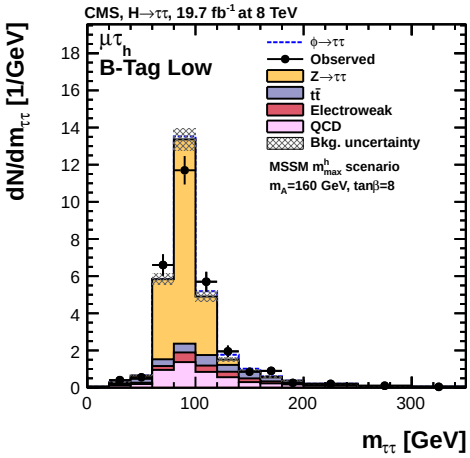
(a) *SVfit* Mass : $\mu\tau_h$ no b-Tag low



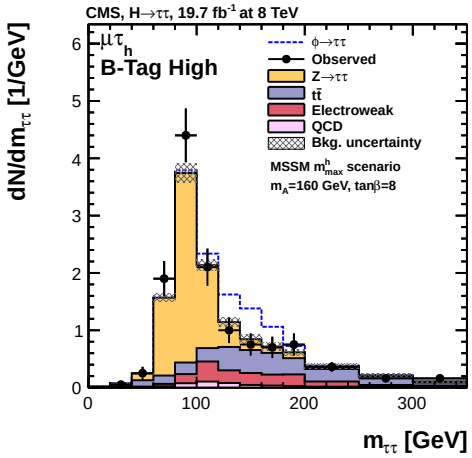
(b) *SVfit Mass* : $\mu\tau_h$ no b-Tag medium



(c) *SVfit Mass* : $\mu\tau_h$ no b-Tag high



(d) *SVfit Mass* : $\mu\tau_h$ b-Tag low



(e) *SVfit Mass* : $\mu\tau_h$ b-Tag high

Figure 7.42: *Di-τ mass distribution for the 5 categories used in the MSSM analysis in the $\mu\tau_h$ channel before the maximum likelihood fit is performed.*

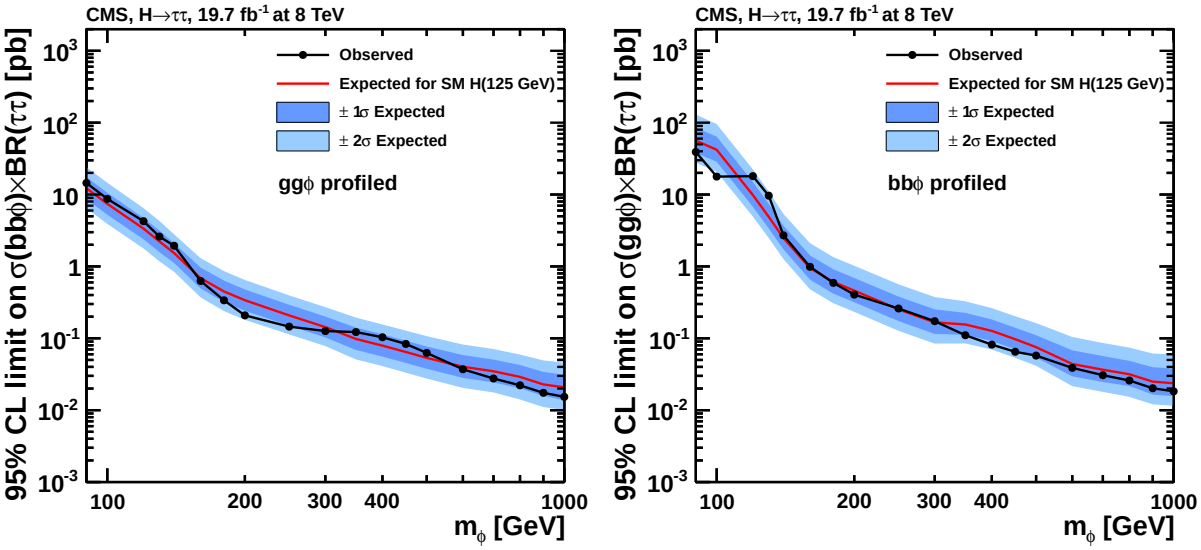


Figure 7.43: *Observed and expected 95% CL upper limits on cross section times branching ratio for $gg \to \phi \to \tau\tau$ (left) and $gg \to \phi b \to \tau\tau$ (right) for the $e\tau_h$ channel in the $\sqrt{s} = 8$ TeV data.*

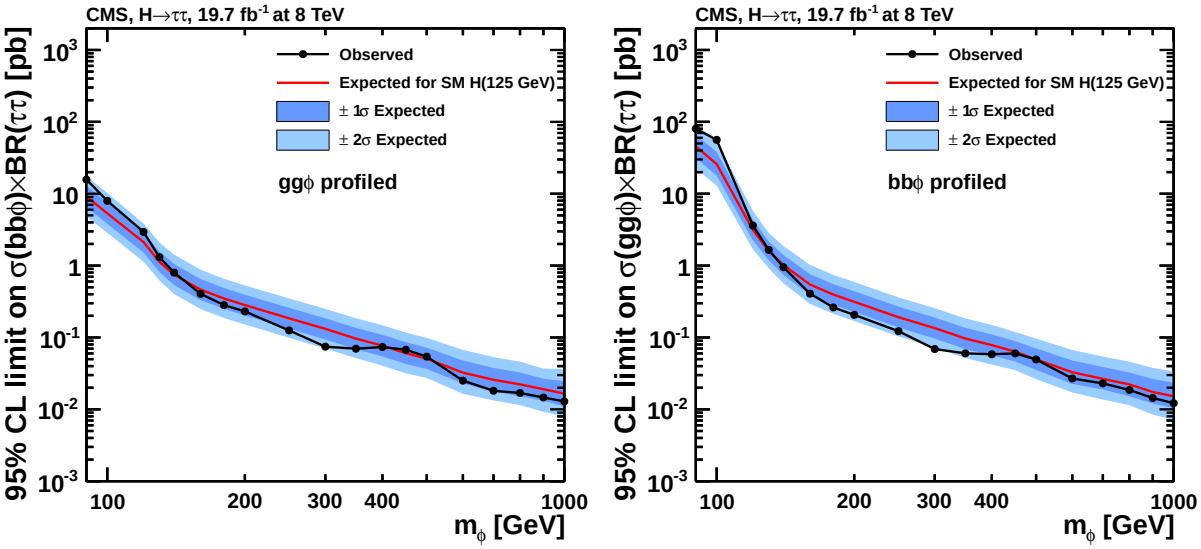


Figure 7.44: *Observed and expected 95% CL upper limits on cross section times branching ratio for $gg \to \phi \to \tau\tau$ (left) and $gg \to \phi b \to \tau\tau$ (right) for the $\mu\tau_h$ channel in the $\sqrt{s} = 8$ TeV data.*

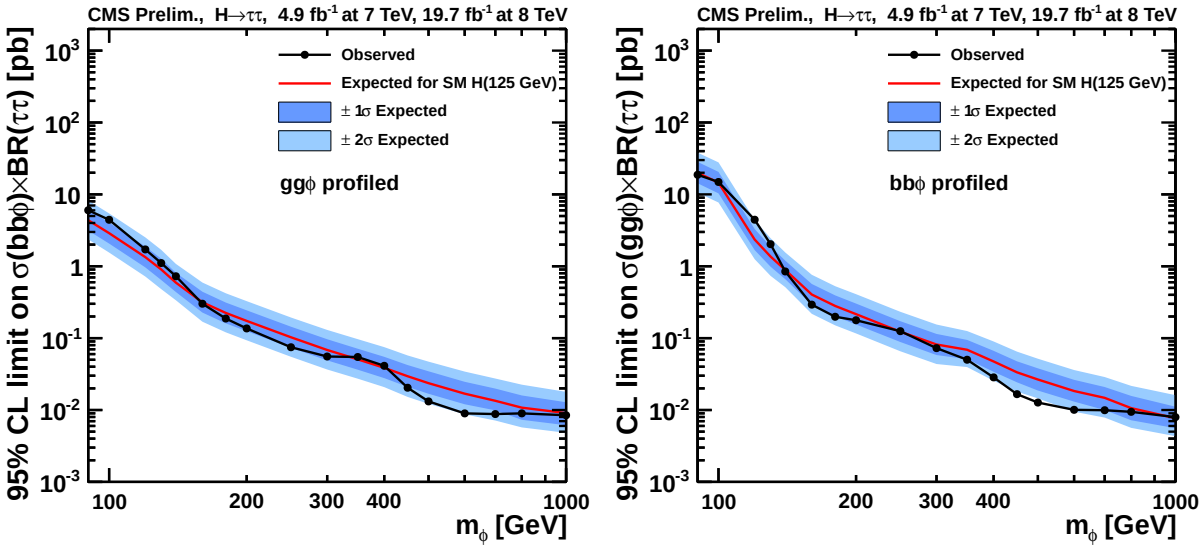


Figure 7.45: *Observed and expected 95% CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi b \to \tau\tau$ *(right) for the combination of all the final states.*

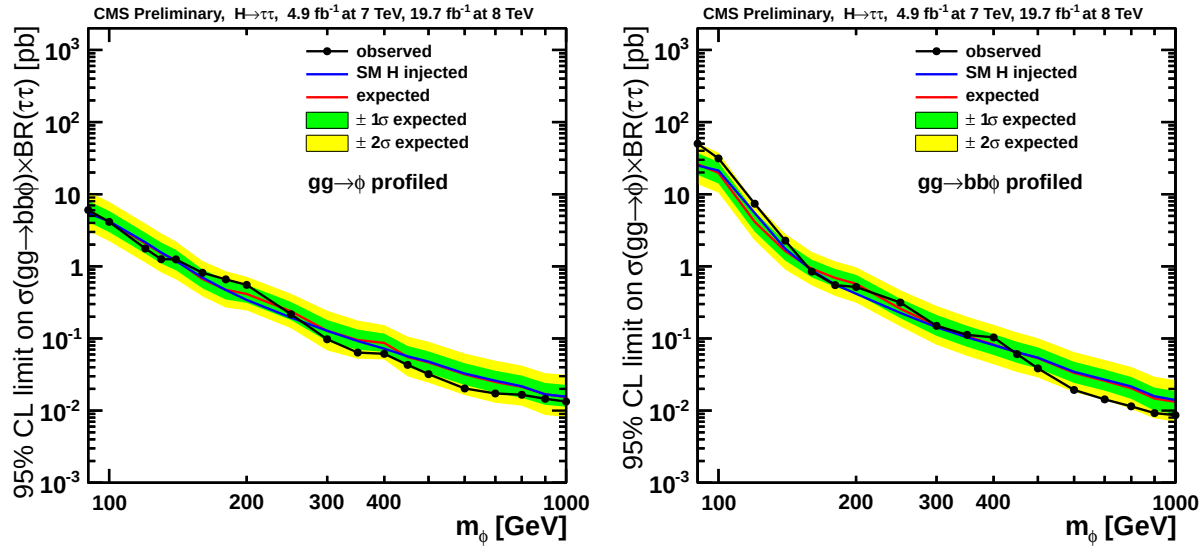


Figure 7.46: *Observed and expected 95% CL upper limits on cross section times branching ratio for* $gg \to \phi \to \tau\tau$ *(left) and* $gg \to \phi b \to \tau\tau$ *(right) for the combination of all the final states. From the latest public result (Ref. [162]).*

Conclusion

The discovery of the first fundamental scalar particle by the ATLAS and CMS collaborations in 2012 opened a new era of particle physics. This leap forward was possible thanks to the remarkable performance of the LHC, as well as the detectors themselves. During the first LHC Run, the CMS experiment recorded a total integrated luminosity of 24.6 fb^{-1}, with 4.9 fb^{-1} at 7 TeV and 19.7 fb^{-1} at 8 TeV.

The coupling of the scalar boson to fermions must be measured in order to prove that it is the origin of fermions' masses. The direct evidence of the corresponding Yukawa interactions in the fermionic final states ($H \to b\bar{b}$ and $H \to \tau\tau$) have to be found in order to validate the Higgs boson hypothesis in the Standard Model. In particular, the $H \to \tau\tau$ channel is the only exploitable channel with the available data to prove that the scalar boson couples to leptons. In this thesis, the analysis of the semi-leptonic final states ($H \to \tau\tau \to \ell\tau_h$ where ℓ denotes a muon or an electron and τ_h a τ decaying hadronically) has been detailed. These channels are the most sensitive ones due to the high electron and muon identification efficiencies and the high branching ratio of the hadronic τ decays.

The hadronic τ-lepton reconstruction is a key element in these Higgs search analyses. One important aspect revealed in this thesis is the fake rate due to mis-identified electrons affecting the channel $e\tau_h$. A specific anti-electron discriminator based on a multivariate analysis has been developed in order to mitigate this fake rate. It reduces the $e \to \tau_h$ fake rate by an order of magnitude while preserving the same τ_h identification efficiency. The analysis sensitivity is increased by about 15% in the $e\tau_h$ channel. Combining this channel with the other final states of $H \to \tau\tau$, an excess of events has been observed, with a significance larger than 3 standard deviations in the boson's mass range between 115 and 130 GeV. The maximal significance, 3.3 standard deviations, is reached at 120 GeV. In addition, other studies [214] have shown no evidence of the $H \to \mu\mu$ decay. Therefore we can conclude that the newly discovered boson has non-universal family couplings. The study of the Higgs sector will likely play a major role in the understanding of the fermion families hierarchy.

The Minimal Super-Symmetric Model (MSSM) is an interesting extension of the SM. A search for MSSM Higgs bosons is reported here. In this framework, the coupling of the Higgs bosons to τ-leptons is enhanced, and the analysis is sensitive to bosons with higher masses, up to 1000 GeV. In this thesis, the sensitivity of this analysis has been improved thanks to several contributions:

- A novel τ_h isolation using lifetime variables in a multivariate analysis based on Boosted Decision Trees. It results in a performant rejection of jets misidentified as τ_h, the main background component, reducing the fake rate by $40-50\%$ with respect to the previous approach. Consequently, the MVA-based isolation brings an improvement of the order of $20-30\%$ in the MSSM analysis sensitivity since the lifetime's discriminant power increases in topologies where the τ_h is boosted.

- An improved event categorization based on the τ_h transverse momentum allows to separate high and low signal purity categories. The sensitivity of the analysis is enhanced by about 20% in the $e\tau_h$ channel, 30% in the $\mu\tau_h$ channel and 40% in the $\tau_h\tau_h$ channel over the full mass range considered.

- A better treatment of the uncertainty on the Higgs boson transverse momenta was implemented by reweighting the simulated signal events. The impact in the analysis sensitivity has been studied showing a large effect at lower masses with a maximal effect in the expected limit of the order of 10% (100%) in the $\ell\tau_h$ ($\tau_h\tau_h$) channels.

The total improvement in the MSSM analysis reaches $30-40\%$ after the combination with the other decay channels[1]. The improved MSSM analysis presented here has shown no significant excess in the observed data with respect to the background-only hypothesis in the mass range considered ($90-1000$ GeV). The interpretation is done by computing the generic model independent upper limits on the cross section times branching ratio for the two production modes. They show more stringent limits than those published previously by CMS [**?**].

The next LHC Run will provide collisions at a center-of-mass energy of 13 TeV, increasing the production cross section of the Higgs boson. A major goal in the LHC physics program will be to determine if the discovered scalar boson is indeed the one of the Standard Model. In the particular case of the di-τ final states searches, a standalone discovery should be achieved, including a precise measurement of the coupling between the scalar boson and the τ-lepton. The τ-lepton, due to its high mass, is endowed with particular properties. The increase of statistics should allow to profit from the τ_h polarization in order to test the Higgs boson properties [152].

The MSSM analysis presented in this thesis constitutes a milestone of the analysis to come. The additional new data will allow to test extensively a larger fraction of the MSSM phase space. The di-τ final state is promising and might well be the next discovery channel.

In a longer term, the precision measurements on the Higgs boson will provide an excellent test of the Standard Model, thus unravelling the nature of the electroweak symmetry breaking. The measurement of the Higgs boson self-coupling might give hints on the stabilization of the Higgs boson mass. Data of the next run might even lead to a new discovery via the decays: $A/H^* \to HH$. One attractive channel for this is the $b\bar{b}\tau\tau$ final state due to its high branching ratio. The door to the scalar sector has just been

[1] Combination of $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$, $e\mu$ and $\mu\mu$.

opened. More data at higher energy will allow precision measurements and possibly new discoveries in the years to come.

Appendices

Appendix A

MSSM Benchmark scenarios

Different MSSM benchmark scenarios can be defined to perform the searches for MSSM Higgs bosons. They are defined without allowing CP violation and they depend on the following parameters:

- m_t the mass of the top quark.
- m_b the mass of the bottom quark.
- M_{SUSY} the mass of the third generation squarks, stop and sbottom.
- μ the higgsino mass parameter.
- $M_{\tilde{l}_3}$ the mass of the third generation sleptons, the staus.
- M_1 the $U(1)$ gaugino mass parameter.
- M_2 the $SU(2)$ gaugino mass parameter.
- A_t, A_b and A_τ the trilinear coupling of the stops, bottoms and staus.
- χ_t, χ_b and χ_τ the mixing parameters of the stops, bottoms and staus.

Historically the results of searches for neutral MSSM Higgs bosons have been shown in the context of the m_H^{max} benchmark scenario. In this scenario the parameters of SUSY are chosen such that the mass of the CP-even Higgs boson h becomes maximal in the case of $M_A \gg M_Z$ reaching the theoretical limit of $m_h \sim 140$ GeV (given by the equation 1.60). The parameters defining this scenario are defined in the first column of table A.1.

The observation of a Standard Model like Higgs boson with a mass around 125 GeV excludes most of the $m_A, \tan\beta$ parameter space of the m_h^{max} scenario. Hence, new scenarios have been proposed to interpret the results in a more plausible way in the context of MSSM.

Two modified m_h^{max} scenarios called m_h^{mod+} and m_h^{mod-} depending on the sign of the stop mixing parameter have been proposed. The modification mainly deviates the value of the stop mixing parameter in order to match a light scalar compatible with a mass of $m_h = 125.5 \pm 3$ GeV within experimental uncertainties from the latest Higgs boson

mass measurements and uncertainties on the theoretical prediction on the MSSM Higgs boson mass. For the m_h^{mod+} (m_h^{mod-}) scenario the value of the stop mixing parameter reads $X_t = 1.5 M_{SUSY}$ ($X_t = -1.9 M_{SUSY}$).

Other benchmark scenarios have also been proposed motivated by the experimental discovery of the Standard Model like particle: *light stop*, *light stau*, *τ-phobic* and the *low-M_H* scenarios are detailed in [215, 216]. In table A.1 the corresponding set of SUSY parameters defining the benchmark scenarios are displayed.

Parameter	m_h^{max}	m_h^{mod+}	m_h^{mod-}	light stop	light stau	τ-phobic	low m_H
M_{SUSY}	1000 GeV	1000 GeV	1000 GeV	500 GeV	1000 GeV	1500 GeV	1500 GeV
μ	200 GeV	200 GeV	200 GeV	350 GeV	500 GeV	2000 GeV	-
M_2	200 GeV	200 GeV	200 GeV	350 GeV	200 GeV	200 GeV	200 GeV
X_t	$2M_{SUSY}$	$1.5M_{SUSY}$	$-1.9M_{SUSY}$	$2M_{SUSY}$	$1.6M_{SUSY}$	$2.45M_{SUSY}$	$2.45M_{SUSY}$
A_b	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$
A_τ	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	0	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$
A_t	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$	$X_t + \mu/\tan\beta$
$m_{\tilde{g}}$	1500 GeV	1500 GeV	1500 GeV	1500 GeV	1500 GeV	1500 GeV	1500 GeV
$m_{\tilde{l_3}}$	1000 GeV	1000 GeV	1000 GeV	1000 GeV	245 GeV	500 GeV	1000 GeV

Table A.1: *Values of the SUSY parameters in the different MSSM benchmark scenarios. M_{SUSY} corresponds to the soft-SUSY breaking third generation squark mass, μ the Higgsino mass parameter, M_2 the gauging mass parameter, A_t, A_b and A_τ the trilinear Higgs-stop, Higgs-sbottom and Higgs-stau couplings respectively, $m_{\tilde{g}}$ the gluino mass and $m_{\tilde{l_3}}$ the third generation slept on mass. The value of M_1 is given by the GUT relation $M_1 = (5/3)M_2 \tan^2\theta_W$. In the low-$m_H$ scenario μ is varied with $M_A = 110$ GeV fixed.*

Appendix B

MVA-based anti-electron: example of correlation matrix for the input variables and BDT output

An example of linear correlation matrix is shown in figure B.1. In this category the τ_h candidate is not matched to a reconstructed electron, reconstructed in the *Single Hadron* decay mode which "leading" charged particle is associated to a GSF track in the barrel region. The corresponding BDT output is displayed in figure B.2. Similar distributions were obtained to validate the training of each one of the 16 categories.

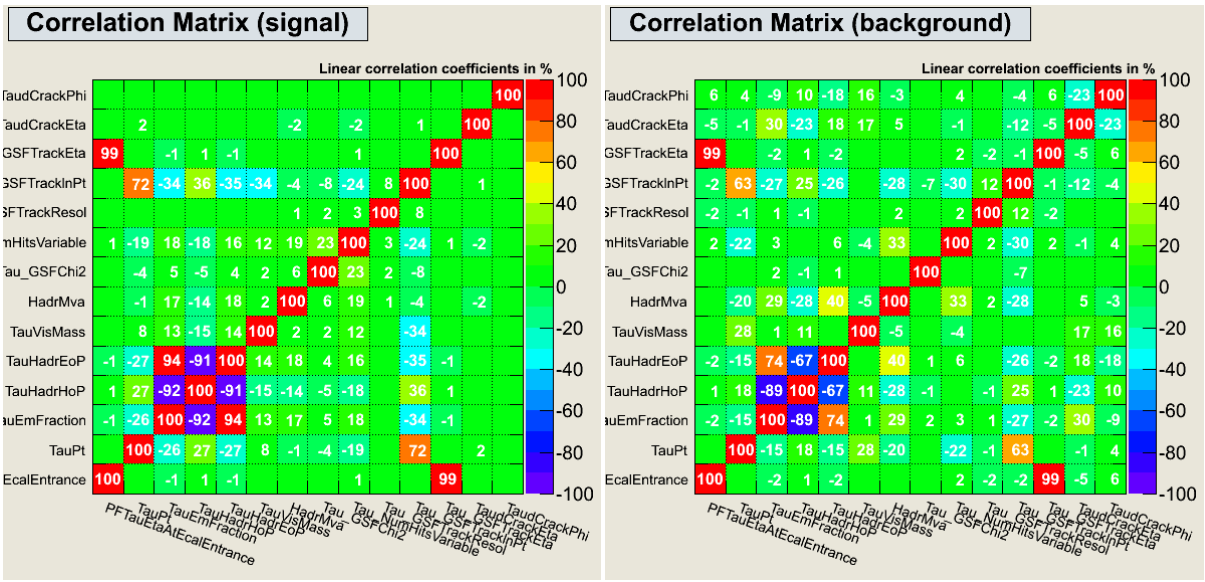


Figure B.1: *Linear correlation matrices of the input variables of the anti-electron discriminator for τ_h candidates not matched to a reconstructed electron and reconstructed in the* Single Hadron *decay mode which "leading" charged particle is associated to a GSF track in barrel, for signal (left) and for background (right).*

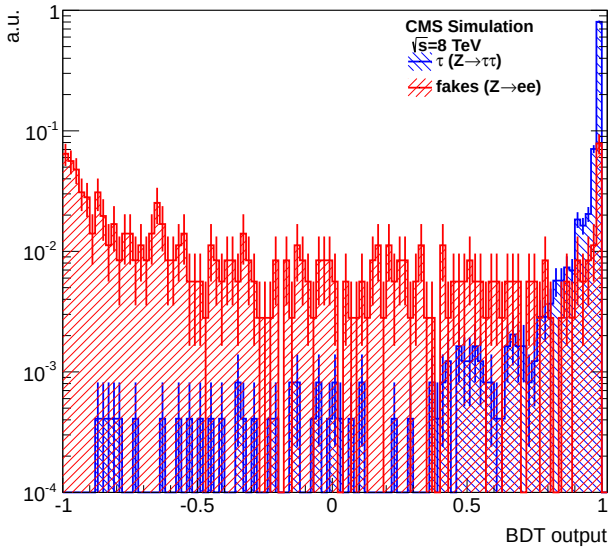


Figure B.2: *BDT output of the anti-electron discriminator for τ_h candidates not matched to a reconstructed electron and reconstructed in the* Single Hadron *decay mode which "leading" charged particle is associated to a GSF track for the barrel region.*

Appendix C

Control plots

In this Appendix, control plots for the kinematical variables of objects used in the analysis are presented.

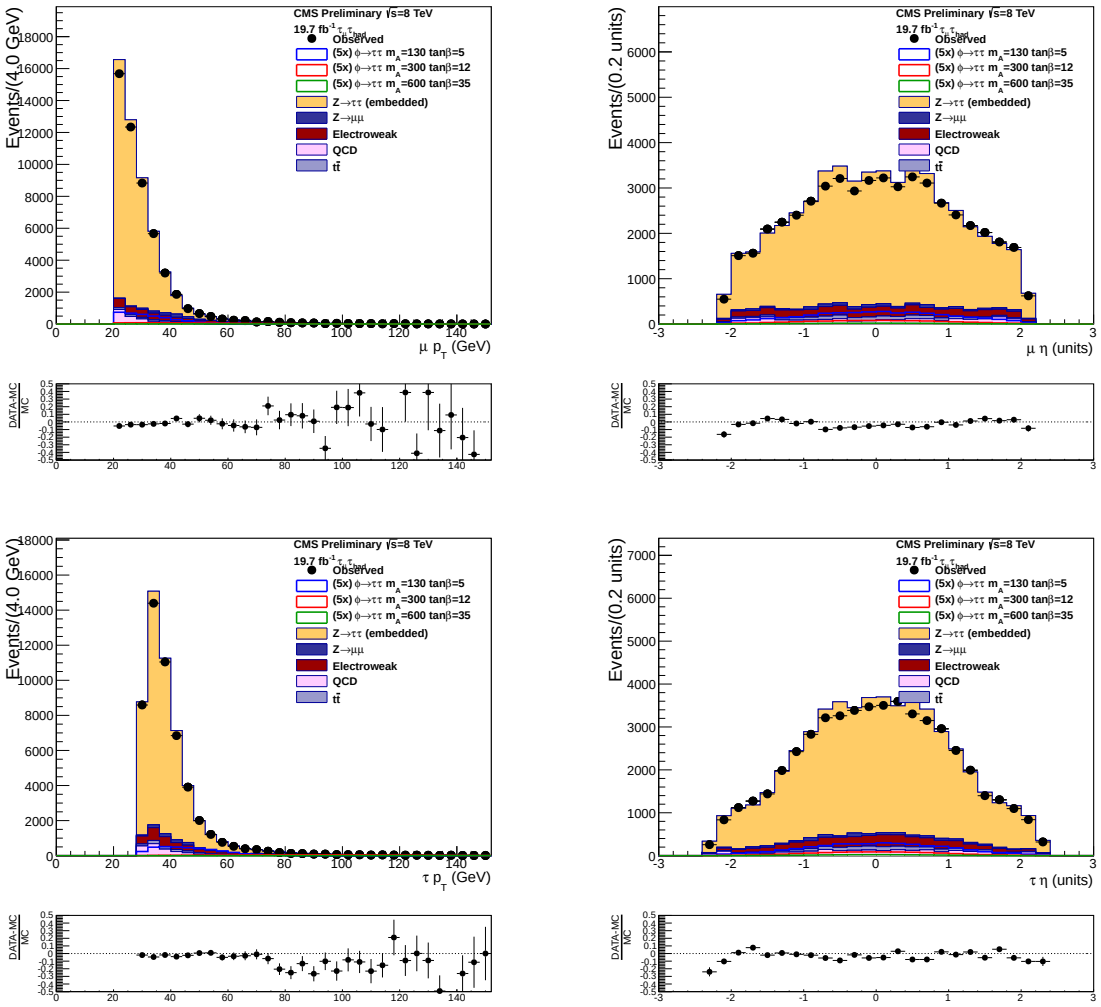


Figure C.1: *Distribution of the muon (up) and τ_h (bottom) transverse momenta (left) and η (right) in the $\mu\tau_h$ channel after the baseline selection.*

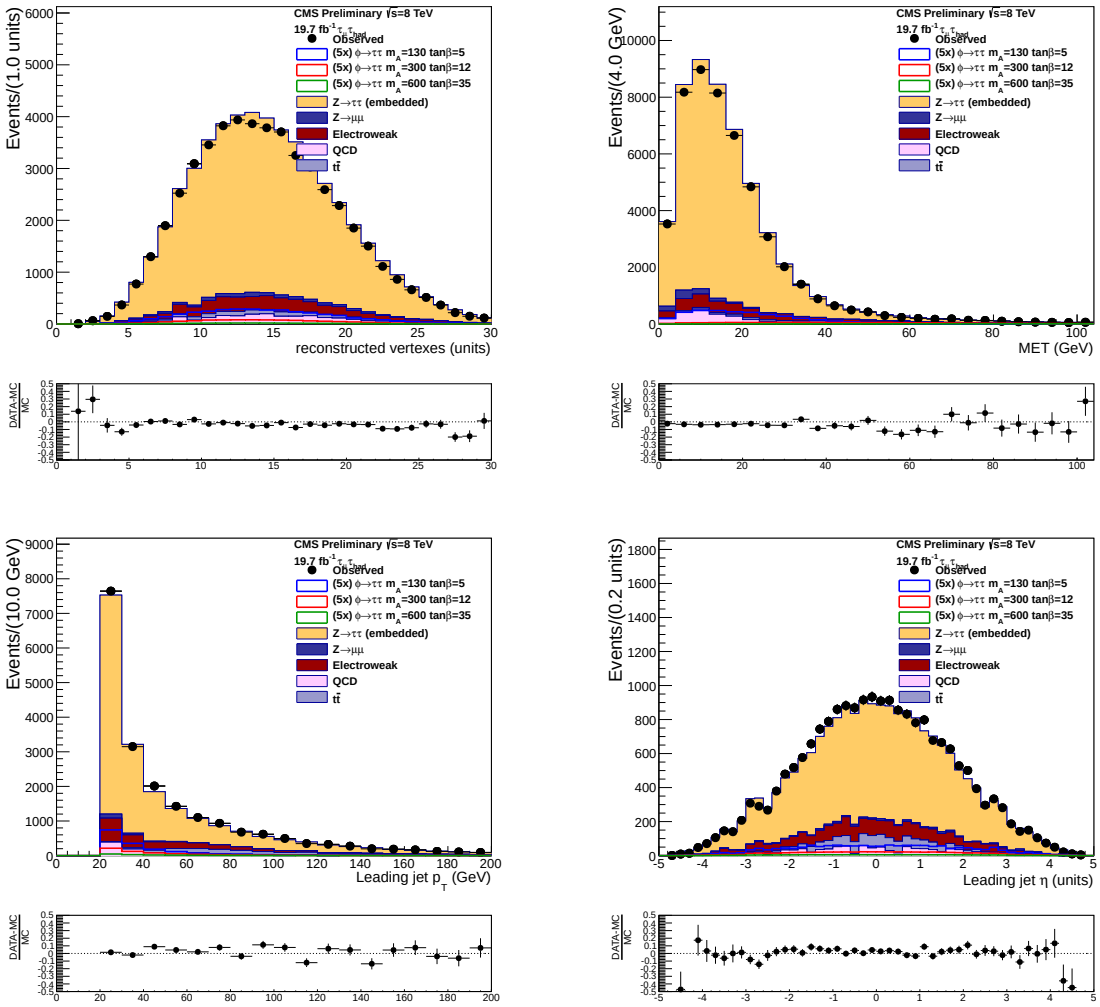


Figure C.2: *Distributions of the number of reconstructed vertices (up left), the transverse missing energy (up right), the leading jet transverse momentum (down left) and η (down right) in the $\mu\tau_h$ channel after the baseline selection.*

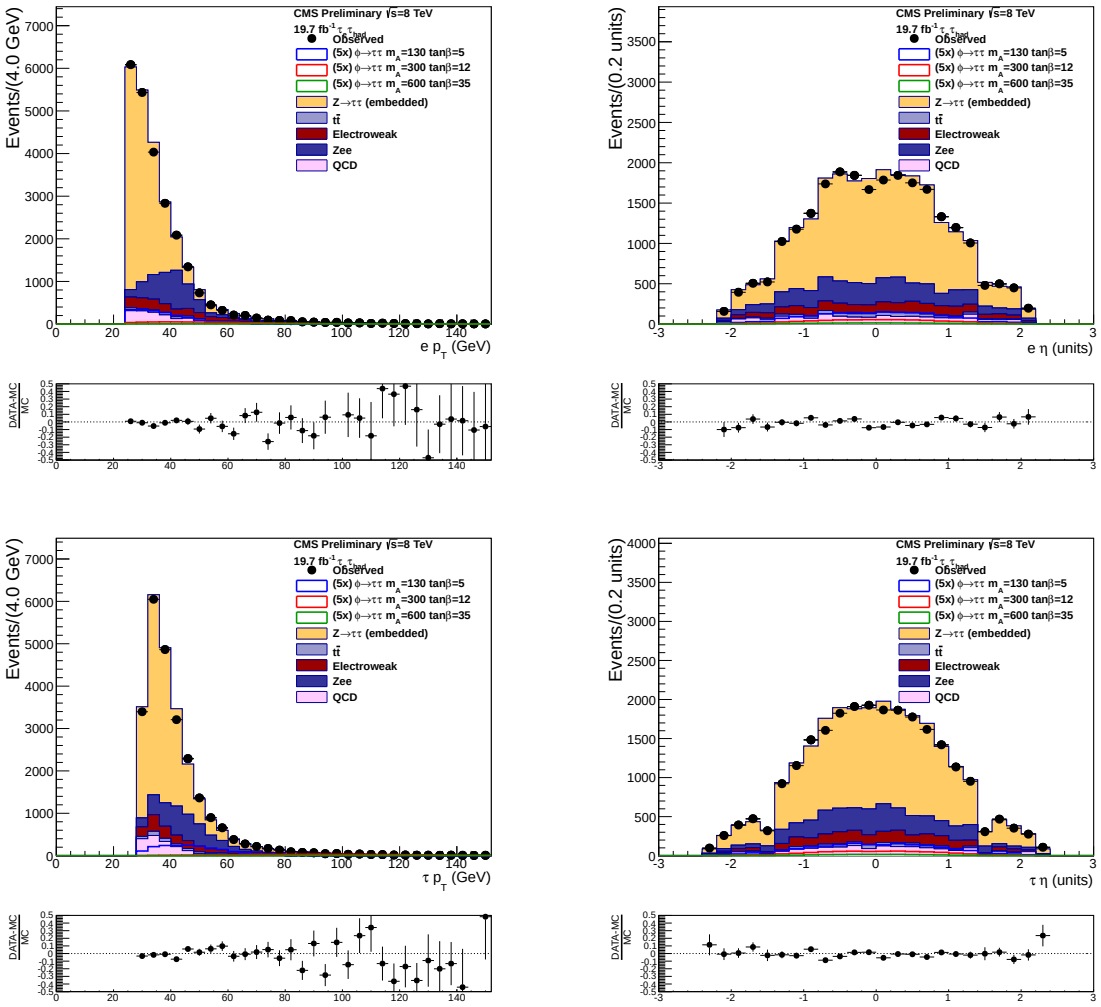


Figure C.3: *Distribution of the electron (up) and τ_h (bottom) transverse momenta (left) and η (right) in the $e\tau_h$ channel after the baseline selection.*

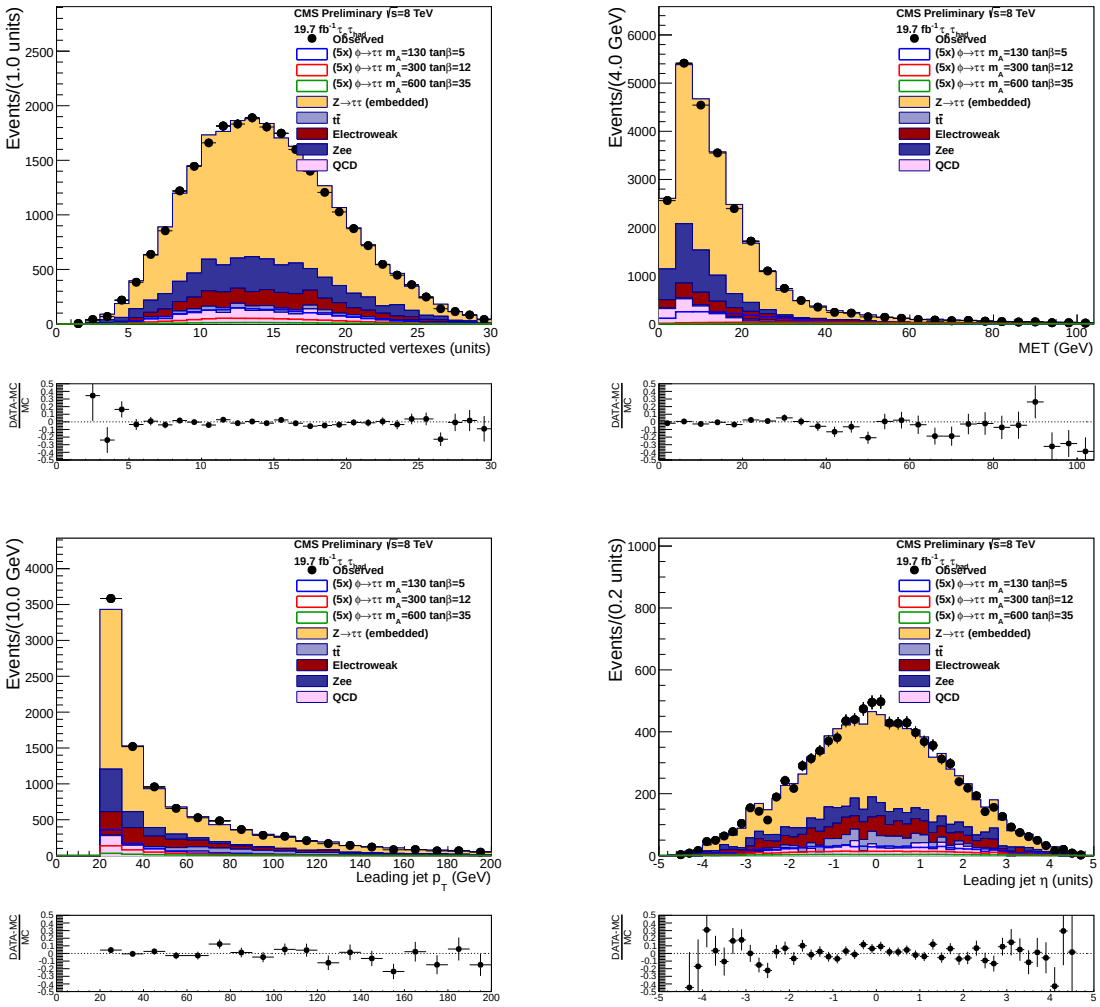


Figure C.4: *Distributions of the number of reconstructed vertices (up left), the transverse missing energy (up right), the leading jet transverse momentum (down left) and η (down right) in the $e\tau_h$ channel after the baseline selection.*

Bibliography

[1] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl.Phys.*, vol. 22, pp. 579–588, 1961.

[2] A. Salam and N. Svartholm, "Elementary particle theory," *Almquist and Wiksell, Stockholm*, p. 367, 1968.

[3] S. Weinberg, "A Model of Leptons," *Phys.Rev.Lett.*, vol. 19, pp. 1264–1266, 1967.

[4] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl.Phys.*, vol. 22, pp. 579–588, 1961.

[5] A. Salam and J. C. Ward, "Electromagnetic and weak interactions," *Phys.Lett.*, vol. 13, pp. 168–171, 1964.

[6] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys.Rev.Lett.*, vol. 13, pp. 508–509, 1964.

[7] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," *Phys.Rev.*, vol. 145, pp. 1156–1163, 1966.

[8] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys.Rev.Lett.*, vol. 13, pp. 321–323, 1964.

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[10] T. Kibble, "Symmetry breaking in nonAbelian gauge theories," *Phys.Rev.*, vol. 155, pp. 1554–1561, 1967.

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

[12] S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[13] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[14] J. Wess and B. Zumino, "Supergauge Transformations in Four-Dimensions," *Nucl.Phys.*, vol. B70, pp. 39–50, 1974.

[15] P. Ramond, "Dual Theory for Free Fermions," *Phys.Rev.*, vol. D3, pp. 2415–2418, 1971.

[16] A. Neveu and J. Schwarz, "Factorizable dual model of pions," *Nucl.Phys.*, vol. B31, pp. 86–112, 1971.

[17] J.-L. Gervais and B. Sakita, "Field Theory Interpretation of Supergauges in Dual Models," *Nucl.Phys.*, vol. B34, pp. 632–639, 1971.

[18] Y. Golfand and E. Likhtman, "Extension of the Algebra of Poincare Group Generators and Violation of p Invariance," *JETP Lett.*, vol. 13, pp. 323–326, 1971.

[19] D. Volkov and V. Akulov, "Is the Neutrino a Goldstone Particle?," *Phys.Lett.*, vol. B46, pp. 109–110, 1973.

[20] J. Bagger and J. Wess, "Supersymmetry and supergravity," 1990.

[21] G. L. Kane, "Perspectives on supersymmetry," 1998.

[22] P. Fayet, "Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino," *Nucl.Phys.*, vol. B90, pp. 104–124, 1975.

[23] P. Fayet, "Supersymmetry and Weak, Electromagnetic and Strong Interactions," *Phys.Lett.*, vol. B64, p. 159, 1976.

[24] P. Fayet, "Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions," *Phys.Lett.*, vol. B69, p. 489, 1977.

[25] P. Fayet, "Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents," *Phys.Lett.*, vol. B84, p. 416, 1979.

[26] K. Inoue, A. Kakuto, H. Komatsu, and S. Takeshita, "Aspects of Grand Unified Models with Softly Broken Supersymmetry," *Prog.Theor.Phys.*, vol. 68, p. 927, 1982.

[27] P. Fayet and S. Ferrara, "Supersymmetry," *Phys.Rept.*, vol. 32, pp. 249–334, 1977.

[28] H. E. Haber and G. L. Kane, "The Search for Supersymmetry: Probing Physics Beyond the Standard Model," *Phys.Rept.*, vol. 117, pp. 75–263, 1985.

[29] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2008.

[30] P. A. Dirac, "Quantum theory of emission and absorption of radiation," *Proc.Roy.Soc.Lond.*, vol. A114, p. 243, 1927.

[31] R. Feynman, "The Theory of positrons," *Phys.Rev.*, vol. 76, pp. 749–759, 1949.

[32] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys.Lett.*, vol. 8, pp. 214–215, 1964.

[33] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 2," 1964.

[34] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?," *Phys.Rev.Lett.*, vol. 30, pp. 1346–1349, 1973.

[35] G. 't Hooft and M. Veltman, "Proceedings of the Colloquium on Renormalization of Yang-Mills Fields, Marseille, June 19-23, 1972," 1972.

[36] E. Fermi, "An attempt of a theory of beta radiation. 1.," *Z.Phys.*, vol. 88, pp. 161–177, 1934.

[37] R. Feynman and M. Gell-Mann, "Theory of Fermi interaction," *Phys.Rev.*, vol. 109, pp. 193–198, 1958.

[38] G. Arnison *et al.*, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider," *Phys.Lett.*, vol. B126, pp. 398–410, 1983.

[39] P. Bagnaia *et al.*, "Evidence for $Z^0 \to e^+e^-$ at the CERN anti-p p Collider," *Phys.Lett.*, vol. B129, pp. 130–140, 1983.

[40] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012.

[41] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys.Lett.*, vol. B716, pp. 30–61, 2012.

[42] A. Djouadi, "The anatomy of electro-weak symmetry breaking. i: The higgs boson in the standard model," *Phys.Rept.*, vol. 457, pp. 1–216, 2008.

[43] P. Paganini, "An introduction to the Standard Model of particle physics," *Cours du Master HEP Ecole Polytechnique.*

[44] E. Noether, "Invariant Variation Problems," *Gott.Nachr.*, vol. 1918, pp. 235–257, 1918.

[45] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys.Rev.Lett.*, vol. 10, pp. 531–533, 1963.

[46] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog.Theor.Phys.*, vol. 49, pp. 652–657, 1973.

[47] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys.Rev.*, vol. 127, pp. 965–970, 1962.

[48] J. Goldstone, "Field Theories with Superconductor Solutions," *Nuovo Cim.*, vol. 19, pp. 154–164, 1961.

[49] Y. Nambu and G. Jona-Lasinio, "Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.," *Phys.Rev.*, vol. 122, pp. 345–358, 1961.

[50] Y. Nambu, "Axial vector current conservation in weak interactions," *Phys.Rev.Lett.*, vol. 4, pp. 380–382, 1960.

[51] C. Llewellyn Smith, "High-Energy Behavior and Gauge Symmetry," *Phys.Lett.*, vol. B46, pp. 233–236, 1973.

[52] J. Bell, "High-energy behavior of tree diagrams in gauge theories," *Nucl.Phys.*, vol. B60, pp. 427–436, 1973.

[53] J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, "Uniqueness of spontaneously broken gauge theories," *Phys.Rev.Lett.*, vol. 30, pp. 1268–1270, 1973.

[54] J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, "Derivation of Gauge Invariance from High-Energy Unitarity Bounds on the s Matrix," *Phys.Rev.*, vol. D10, p. 1145, 1974.

[55] F. Dyson, "The Radiation theories of Tomonaga, Schwinger, and Feynman," *Phys.Rev.*, vol. 75, pp. 486–502, 1949.

[56] J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, A. Riotto, *et al.*, "Higgs mass implications on the stability of the electroweak vacuum," *Phys.Lett.*, vol. B709, pp. 222–228, 2012.

[57] "https ://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections,"

[58] S. Dittmaier *et al.*, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," 2011.

[59] S. Dittmaier, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, *et al.*, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions," 2012.

[60] S. Heinemeyer *et al.*, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," 2013.

[61] J. R. Ellis, S. Kelley, and D. V. Nanopoulos, "Probing the desert using gauge coupling unification," *Phys.Lett.*, vol. B260, pp. 131–137, 1991.

[62] U. Amaldi, W. de Boer, and H. Furstenau, "Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP," *Phys.Lett.*, vol. B260, pp. 447–455, 1991.

[63] P. Langacker and M.-x. Luo, "Implications of precision electroweak experiments for M_t, ρ_0, $\sin^2\theta_W$ and grand unification," *Phys.Rev.*, vol. D44, pp. 817–822, 1991.

[64] C. Giunti, C. Kim, and U. Lee, "Running coupling constants and grand unification models," *Mod.Phys.Lett.*, vol. A6, pp. 1745–1755, 1991.

[65] M. Veltman, "Second Threshold in Weak Interactions," *Acta Phys.Polon.*, vol. B8, p. 475, 1977.

[66] A. Djouadi, "The anatomy of electro-weak symmetry breaking. ii. the higgs bosons in the minimal supersymmetric model," *Phys.Rept.*, vol. 459, pp. 1–241, 2008.

[67] S. Weinberg, "Gauge Hierarchies," *Phys.Lett.*, vol. B82, p. 387, 1979.

[68] M. Veltman, "The Infrared - Ultraviolet Connection," *Acta Phys.Polon.*, vol. B12, p. 437, 1981.

[69] C. Llewellyn Smith and G. G. Ross, "The Real Gauge Hierarchy Problem," *Phys.Lett.*, vol. B105, p. 38, 1981.

[70] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry," *Phys.Lett.*, vol. B76, pp. 575–579, 1978.

[71] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012.

[72] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys.Lett.*, vol. B716, pp. 30–61, 2012.

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

[74] R. M. Barnett *et al.*, "Review of particle physics. Particle Data Group," *Phys.Rev.*, vol. D54, pp. 1–720, 1996.

[75] R. Barate *et al.*, "Search for the standard model Higgs boson at LEP," *Phys.Lett.*, vol. B565, pp. 61–75, 2003.

[76] T. Aaltonen *et al.*, "Higgs Boson Studies at the Tevatron," *Phys.Rev.*, vol. D88, no. 5, p. 052014, 2013.

[77] T. Han and S. Willenbrock, "QCD correction to the p p $\rightarrow$ W H and Z H total cross-sections," *Phys.Lett.*, vol. B273, pp. 167–172, 1991.

[78] O. Brein, A. Djouadi, and R. Harlander, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders," *Phys.Lett.*, vol. B579, pp. 149–156, 2004.

[79] M. Ciccolini, S. Dittmaier, and M. Kramer, "Electroweak radiative corrections to associated WH and ZH production at hadron colliders," *Phys.Rev.*, vol. D68, p. 073003, 2003.

[80] G. Altarelli, B. Mele, and F. Pitolli, "Heavy Higgs Production at Future Colliders," *Nucl.Phys.*, vol. B287, pp. 205–224, 1987.

[81] M. Ciccolini, A. Denner, and S. Dittmaier, "Strong and electroweak corrections to the production of Higgs + 2jets via weak interactions at the LHC," *Phys.Rev.Lett.*, vol. 99, p. 161803, 2007.

[82] M. Ciccolini, A. Denner, and S. Dittmaier, "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC," *Phys.Rev.*, vol. D77, p. 013002, 2008.

[83] K. Arnold, M. Bahr, G. Bozzi, F. Campanario, C. Englert, *et al.*, "VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons," *Comput.Phys.Commun.*, vol. 180, pp. 1661–1670, 2009.

[84] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, "Higgs production via vector-boson fusion at NNLO in QCD," *Phys.Rev.Lett.*, vol. 105, p. 011801, 2010.

[85] A. Djouadi, M. Spira, and P. Zerwas, "Production of Higgs bosons in proton colliders: QCD corrections," *Phys.Lett.*, vol. B264, pp. 440–446, 1991.

[86] R. V. Harlander, "Virtual corrections to g g $\to$ H to two loops in the heavy top limit," *Phys.Lett.*, vol. B492, pp. 74–80, 2000.

[87] S. Dawson, "Radiative corrections to Higgs boson production," *Nucl.Phys.*, vol. B359, pp. 283–300, 1991.

[88] S. Catani, D. de Florian, M. Grazzini, and P. Nason, "Soft gluon resummation for Higgs boson production at hadron colliders," *JHEP*, vol. 0307, p. 028, 2003.

[89] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, "Two loop light fermion contribution to Higgs production and decays," *Phys.Lett.*, vol. B595, pp. 432–441, 2004.

[90] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, "NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders," *Phys.Lett.*, vol. B670, pp. 12–17, 2008.

[91] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, *et al.*, "Higgs radiation off top quarks at the Tevatron and the LHC," *Phys.Rev.Lett.*, vol. 87, p. 201805, 2001.

[92] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, *et al.*, "NLO QCD corrections to t anti-t H production in hadron collisions," *Nucl.Phys.*, vol. B653, pp. 151–203, 2003.

[93] S. Dawson, L. Orr, L. Reina, and D. Wackeroth, "Associated top quark Higgs boson production at the LHC," *Phys.Rev.*, vol. D67, p. 071503, 2003.

[94] S. Schael *et al.*, "Search for neutral MSSM Higgs bosons at LEP," *Eur.Phys.J.*, vol. C47, pp. 547–587, 2006.

[95] T. Aaltonen *et al.*, "Search for Neutral Higgs Bosons in Events with Multiple Bottom Quarks at the Tevatron," *Phys.Rev.*, vol. D86, p. 091101, 2012.

[96] D. Benjamin *et al.*, "Combined CDF and D0 Upper Limits on MSSM Higgs Boson Production in tau-tau Final States with up to 2.2 fb-1," 2010.

[97] A. Djouadi and M. Spira, "SUSY - QCD corrections to Higgs boson production at hadron colliders," *Phys.Rev.*, vol. D62, p. 014004, 2000.

[98] S. Dawson, A. Djouadi, and M. Spira, "QCD corrections to SUSY Higgs production: The Role of squark loops," *Phys.Rev.Lett.*, vol. 77, pp. 16–19, 1996.

[99] R. P. Kauffman and W. Schaffer, "QCD corrections to production of Higgs pseudoscalars," *Phys.Rev.*, vol. D49, pp. 551–554, 1994.

[100] S. Dawson and R. Kauffman, "QCD corrections to Higgs boson production: nonleading terms in the heavy quark limit," *Phys.Rev.*, vol. D49, pp. 2298–2309, 1994.

[101] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, "Higgs-Boson production in association with a single bottom quark," *Phys.Rev.*, vol. D67, p. 095002, 2003.

[102] S. Dawson, C. Jackson, L. Reina, and D. Wackeroth, "Higgs boson production with one bottom quark jet at hadron colliders," *Phys.Rev.Lett.*, vol. 94, p. 031802, 2005.

[103] M. Beccaria, G. Dovier, G. Macorini, E. Mirabella, L. Panizzi, *et al.*, "Semi-inclusive bottom-Higgs production at LHC: The complete one-loop electroweak effect in the MSSM," *Phys.Rev.*, vol. D82, p. 093018, 2010.

[104] S. Dawson and C. Jackson, "SUSY QCD Corrections to Associated Higgs-bottom Quark Production," *Phys.Rev.*, vol. D77, p. 015019, 2008.

[105] G. Aad *et al.*, "Search for neutral MSSM Higgs bosons decaying to $\tau^+\tau^-$ pairs in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Phys.Lett.*, vol. B705, pp. 174–192, 2011.

[106] S. Chatrchyan *et al.*, "Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys.Rev.Lett.*, vol. 106, p. 231801, 2011.

[107] The ATLAS collaboration, "Search for neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," 2014.

[108] V. Khachatryan *et al.*, "Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions," 2014.

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

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

[111] C. Carli, "Proceedings, Workshop on LHC performance," 2012.

[112] *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.*

[113] S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," 2014.

[114] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method," *Nucl.Instrum.Meth.*, vol. A294, pp. 219–228, 1990.

[115] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl.Instrum.Meth.*, vol. A262, pp. 444–450, 1987.

[116] E. Chabanat and N. Estre, "Deterministic Annealing for Vertex Finding at CMS," 2005.

[117] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," 2009.

[118] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV," 2010.

[119] CMS Collaboration, "Performance of muon identification in pp collisions at s**0.5 = 7 TeV," 2010.

[120] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV," 2010.

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

[122] CMS Collaboration, "Performance of electron reconstruction and selection with the cms detector at $\sqrt(s) = 8$ *tev* (in preparation)," *EGM-13-001.*

[123] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC," *eConf*, vol. C0303241, p. TULT009, 2003.

[124] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC," 2005.

[125] F.Beaudette, D. Benedetti, P. Janot, and M. Pioppi, "Electron reconstruction within the particle flow algorithm," *CMS Analysis Note*, no. AN-2010-034, 2010.

[126] CMS EGamma POG, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/ EgammaCutBasedIdentification,"

[127] CMS Collaboration, "Photon reconstruction and identification at sqrt(s) = 7 TeV," *EGM-10-005*, 2010.

[128] CMS Collaboration, "Isolated Photon Reconstruction and Identification at sqrt(s)=7 TeV," *EGM-10-006*, 2011.

[129] CMS Collaboration, "Performance of photon reconstruction and selection in the cms detector at sqrt(s)=8tev (in preparation)," *EGM-14-001*.

[130] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP*, vol. 0804, p. 063, 2008.

[131] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[132] CMS Collaboration, "Jet Energy Resolution in CMS at sqrt(s)=7 TeV," *JME-10-014*, 2011.

[133] CMS Collaboration, "Pileup Jet Identification," *JME-13-005*, 2013.

[134] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events," 2013.

[135] CMS Collaboration, "Commissioning of b-jet identification with pp collisions at sqrt(s) = 7 TeV," Tech. Rep. CMS-PAS-BTV-10-001, CERN, 2010. Geneva, 2010.

[136] S. Chatrchyan *et al.*, "Missing transverse energy performance of the CMS detector," *JINST*, vol. 6, p. P09001, 2011.

[137] CMS Collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at sqrt(s) = 8 TeV with the CMS Detector," 2012.

[138] S. Chatrchyan *et al.*, "Performance of tau-lepton reconstruction and identification in CMS," *JINST*, vol. 7, p. P01001, 2012.

[139] "Tau identification in CMS," Tech. Rep. CMS-PAS-TAU-11-001, CERN, Geneva, 2011.

[140] The Tau Physics Object Group, "Performance of tau reconstruction and identification in pp collisions at $\sqrt{s}$ = 8 tev (in preparation)," *CMS Analysis Note*, no. AN-14-008, 2014.

[141] J. Beringer *et al.*, "Review of Particle Physics (RPP)," *Phys.Rev.*, vol. D86, p. 010001, 2012.

[142] CMS Collaboration, "Updated search for a light charged Higgs boson in top quark decays in pp collisions at sqrt(s) = 7 TeV," *HIG-12-052*, 2012.

[143] S. Chatrchyan *et al.*, "Search for physics beyond the standard model in events with τ leptons, jets, and large transverse momentum imbalance in pp collisions at $\sqrt{s}$ = 7 TeV," *Eur.Phys.J.*, vol. C73, p. 2493, 2013.

[144] CMS Collaboration, "Search for RPV supersymmetry with three or more leptons and b-tags," *CMS-PAS-SUS-12-027*, 2012.

[145] CMS Collaboration, "A search for anomalous production of events with three or more leptons using 19.5/fb of sqrt(s)=8 TeV LHC data," *CMS-PAS-SUS-13-002*, 2013.

[146] S. Chatrchyan *et al.*, "Search for high mass resonances decaying into τ^- lepton pairs in *pp* collisions at $\sqrt{s} = 7$ TeV," *Phys.Lett.*, vol. B716, pp. 82–102, 2012.

[147] K. Hoepner, S. Knutzen, and K. Padeken, "Search for W' in the decay channel tau + MET with LHC data at $\sqrt{s} = 8$ TeV," *CMS Analysis Note*, no. AN-13-080, 2013.

[148] S. Chatrchyan *et al.*, "A search for a doubly-charged Higgs boson in *pp* collisions at $\sqrt{s} = 7$ TeV," *Eur.Phys.J.*, vol. C72, p. 2189, 2012.

[149] S. Chatrchyan *et al.*, "Measurement of the Inclusive Z Cross Section via Decays to Tau Pairs in *pp* Collisions at $\sqrt{s} = 7$ TeV," *JHEP*, vol. 1108, p. 117, 2011.

[150] CMS Collaboration, "Measurement of W to tau-nu," *CMS-PAS-EWK-11-019*, 2011.

[151] CMS Collaboration, "Top pair cross section in tau+jets," *CMS-PAS-TOP-11-004*, 2012.

[152] B. Bullock, K. Hagiwara, and A. D. Martin, "Tau polarization and its correlations as a probe of new physics," *Nucl.Phys.*, vol. B395, pp. 499–533, 1993.

[153] CMS Collaboration, "The CMS Trigger," *TRG-12-001*, no. TRG-12-001, 2012.

[154] V. Gori, "The CMS High Level Trigger," *Int.J.Mod.Phys.Conf.Ser.*, vol. 31, p. 1460297, 2014.

[155] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," *PoS*, vol. ACAT, p. 040, 2007.

[156] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, *et al.*, "Boosted decision trees, an alternative to artificial neural networks," *Nucl.Instrum.Meth.*, vol. A543, pp. 577–584, 2005.

[157] V. Abazov *et al.*, "Evidence for production of single top quarks and first direct measurement of $|V_{tb}|$," *Phys.Rev.Lett.*, vol. 98, p. 181802, 2007.

[158] T. Aaltonen *et al.*, "Observation of Single Top Quark Production and Measurement of $|V_{tb}|$with CDF," *Phys.Rev.*, vol. D82, p. 112005, 2010.

[159] C. Gini, *Variabilità e mutabilità*. C. Cuppini, Bologna, 156 pages. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955)., 1912.

[160] "Tau ID Performance Plots," *CMS-DP-2014-015*, Apr 2014.

[161] S. Chatrchyan *et al.*, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons," *JHEP*, vol. 1405, p. 104, 2014.

[162] CMS Collaboration, "Higgs to tau tau (MSSM)," *HIG-13-021*, 2013.

[163] L. Bianchini and Y. Sirois, *Search for the Standard Model Higgs Boson decaying to tau leptons with the CMS experiment at LHC*. PhD thesis, Zurich, ETH, 2013. presented 18 Sep 2012.

[164] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC," 2005.

[165] L. Bianchini, I. Naranjo, and C. Veelken, "Development of a new tau identification discriminator against electrons," *CMS Analysis Note*, no. AN-2012-417, 2012.

[166] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 0711, p. 070, 2007.

[167] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 0605, p. 026, 2006.

[168] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, "Higgs boson production at the LHC: transverse momentum resummation effects in the H $\rightarrow$ 2gamma, H $\rightarrow$ WW$\rightarrow$ lnulnu and H$\rightarrow$ ZZ $\rightarrow$ 4l decay modes," *JHEP*, vol. 1206, p. 132, 2012.

[169] M. Spira, "HIGLU: A program for the calculation of the total Higgs production cross-section at hadron colliders via gluon fusion including QCD corrections," 1995.

[170] R. V. Harlander and W. B. Kilgore, "Production of a pseudoscalar Higgs boson at hadron colliders at next-to-next-to leading order," *JHEP*, vol. 0210, p. 017, 2002.

[171] S. Heinemeyer, W. Hollik, and G. Weiglein, "FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM," *Comput.Phys.Commun.*, vol. 124, pp. 76–89, 2000.

[172] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP*, vol. 0302, p. 027, 2003.

[173] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)," *Phys.Rev.*, vol. D74, p. 114017, 2006.

[174] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\text{Œ} \pm \frac{4}{S})$," *Phys.Rev.Lett.*, vol. 110, no. 25, p. 252004, 2013.

[175] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC," *JHEP*, vol. 1107, p. 018, 2011.

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

[177] Z. Czyczula, T. Przedzinski, and Z. Was, "Tauspinner program for studies on spin effect in tau production at the lhc," *The European Physical Journal C*, vol. 72, no. 4, pp. 1–8, 2012.

[178] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl.Instrum.Meth.*, vol. A506, pp. 250–303, 2003.

[179] M. Bluj, A. Burgmeier, T. Fruboes, G. Quast, and M. Zeise, "Modelling of $\tau\tau$ final states by embedding τ pairs in $Z \to \mu\mu$ events," *CMS Analysis Note*, no. AN-2011-020, 2011.

[180] T. Fruboes, "Determination of drell-yan $\tau\tau$ background using embedding of simulated τ decays in drell-yan $\mu\mu$ events," *CMS Analysis Note*, no. AN-2012-495, 2012.

[181] A. Burgmeier, T. Fruboes, and C. Veelken, "Modelling backgrounds via embedding technique applied on rechit level," *CMS Analysis Note*, no. AN-2013-073, 2013.

[182] N. Adam, J. Berryhill, V. Halyo, A. Hunt, and K. Mishra, "Generic tag and probe tool for measuring efficiency at cms with early data," *CMS Analysis Note*, no. AN-2009-111, 2009.

[183] "http://en.wikipedia.org/wiki/Crystal_Ball_function," *Wikipedia*.

[184] "https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsToTauTauWorkingSummer2013," *CMS HiggsTauTau PAG*.

[185] Courtesy of A. Nayak.

[186] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[187] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets," *JHEP*, vol. 0804, p. 005, 2008.

[188] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys.Lett.*, vol. B659, pp. 119–126, 2008.

[189] N. Kidonakis, "NNLL threshold resummation for top-pair and single-top production," 2012.

[190] "https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting," *CMS TOP POG*.

[191] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, "Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC," *JHEP*, vol. 1111, p. 064, 2011.

[192] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, "Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC," *Nucl.Phys.*, vol. B737, pp. 73–120, 2006.

[193] D. L. Rainwater, D. Zeppenfeld, and K. Hagiwara, "Searching for $H \to \tau^+\tau^-$ in weak boson fusion at the CERN LHC," *Phys.Rev.*, vol. D59, p. 014037, 1998.

[194] L. Bianchini, J. Conway, E. Friis, and C. Veelken, "New SVfit developments," *CMS Analysis Note*, no. AN-2012-124, July 2012.

[195] K. Kondo, "Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models," *J.Phys.Soc.Jap.*, vol. 57, pp. 4126–4140, 1988.

[196] K. Kondo, "Dynamical likelihood method for reconstruction of events with missing momentum. 2: Mass spectra for $2 \to 2$ processes," *J.Phys.Soc.Jap.*, vol. 60, pp. 836–844, 1991.

[197] G. P. Lepage, "A New Algorithm for Adaptive Multidimensional Integration," *J.Comput.Phys.*, vol. 27, p. 192, 1978.

[198] B. Bullock, K. Hagiwara, and A. D. Martin, "Tau polarization and its correlations as a probe of new physics," *Nucl.Phys.*, vol. B395, pp. 499–533, 1993.

[199] A. L. Read, "Presentation of search results: The CL(s) technique," *J.Phys.*, vol. G28, pp. 2693–2704, 2002.

[200] "Procedure for the LHC Higgs boson search combination in summer 2011," 2011.

[201] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur.Phys.J.*, vol. C71, p. 1554, 2011.

[202] "https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit," *CMS Higgs PAG*.

[203] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, *et al.*, "The RooStats Project," *PoS*, vol. ACAT2010, p. 057, 2010.

[204] HTauTau Group, "Search for a Standard Model Higgs boson decaying to tau pairs produced in association with a W or Z boson," *CMS Analysis Note*, no. AN-2013-187, December 2013.

[205] "https://indico.cern.ch/event/252865/session/14/contribution/58/material/slides/0.pdf," *CMS HTauTau group.*

[206] CMS Collaboration, "Measurement of Top Quark Pair Differential Cross Sections at sqrt(s) = 7 TeV," 2012.

[207] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$," *Phys.Rev.Lett.*, vol. 110, no. 25, p. 252004, 2013.

[208] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update," 2013.

[209] "https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendation," *CMS TAU POG.*

[210] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[211] "https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG," *CMS BTV POG.*

[212] S. Chatrchyan *et al.*, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions," *CMS-HIG-13-033*, 2014.

[213] S. Berge, W. Bernreuther, B. Niepelt, and H. Spiesberger, "How to pin down the CP quantum numbers of a Higgs boson in its tau decays at the LHC," *Phys.Rev.*, vol. D84, p. 116003, 2011.

[214] CMS Collaboration, "Search for the standard model Higgs boson in the dimuon decay channel in pp collisions at sqrt(s)= 7 and 8 TeV," 2013.

[215] M. Carena, S. Heinemeyer, O. Stål, C. Wagner, and G. Weiglein, "MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle," *Eur.Phys.J.*, vol. C73, p. 2552, 2013.

[216] M. S. Carena, S. Heinemeyer, C. Wagner, and G. Weiglein, "Suggestions for benchmark scenarios for MSSM Higgs boson searches at hadron colliders," *Eur.Phys.J.*, vol. C26, pp. 601–607, 2003.