Search for the Higgs boson decaying into tau lepton pairs with the Matrix Element Method and tau trigger optimization in the CMS experiment at the LHC

Luca Mastrolorenzo

ÉCOLE POLYTECHNIQUE

THÈSE DE DOCTORAT

Search for the Higgs boson decaying into τ lepton pairs with the Matrix Element Method and τ trigger optimization in the CMS experiment at the LHC

Présentée par

Luca Mastrolorenzo

en vue d'obtenir le grade de

DOCTEUR ÈS SCIENCE

Soutenue le 23 Octobre 2015 devant le jury composé de:

Dr.	Arnd	SPECKA	Président du Jury
Dr.	Emmanuel	PEREZ	Rapporteur
Dr.	David	ROUSSEAU	Rapporteur
Dr.	Florian	BEAUDETTE	Directeur de thèse
Dr.	Laurent	DUFLOT	Examinateur
Dr.	Roger Manfred	WOLF	Examinateur

Summary

I performed my thesis work in Particle Physics at the laboratoire Leprince-Ringuet of the Ecole Polytechnique. I have been involved in the analysis of the data produced in the proton-proton collisions at the Large Hadron Collider (CERN) and collected by the CMS experiment. Particle physics is a scientific field which is currently undergoing very important breakthroughs. The discovery of the Higgs boson is a major step forward as the mass of vector bosons are explained through their interactions with the corresponding field. I worked on the newly discovered heavy boson analysis. As its direct coupling to fermions remained to be exhibited, I focused on the search for the Higgs boson decaying in tau lepton pairs. The Higgs decay into tau pairs is the most promising decay channel to measure the couplings between the Standard Model Higgs boson and the fermions. Indeed, this decay channel benefits from a large expected event rate compared to the other leptonic decay modes. The Higgs boson decaying to tau lepton analysis is particularly challenging at the trigger level because the selection of tau leptons relies on its decay into electron and muon whose energy spectrum is relatively soft because of the two neutrinos in the decay chain. The higher threshold on single physics objects has thus a severe impact on the signal acceptance. I investigated this crucial aspect and I worked to implement a cross trigger using the missing transverse energy to lower the threshold on the single lepton. This approach allows the recovery of 41% of the signal events. Events with large missing transverse momentum were selected in order to control the trigger rate. My personal contribution consisted in a thorough characterization of such a trigger and the evaluation of the associated uncertainty. The results of this approach lead to an amelioration of 2% in the exclusion limits computed in the Higgs to taus semileptonic channel. In the Run 2, the centre-of-mass energy of the LHC collisions has been increased to 13 TeV and the instantaneous luminosity will reach $2 \cdot 10^{34}\ cm^{-2}s^{-1}$. To guarantee a

successful and ambitious physics program under this intense environment, the CMS Trigger and Data acquisition system must be consolidated. During the first long shutdown of the LHC, the L1 Calorimeter Trigger hardware and architecture have been upgraded, benefiting from the recent microTCA technology allowing the calorimeter granularity to be better exploited with more advanced algorithms. Thanks to the enhanced granularity provided by the new system, an innovative dynamic clustering technique has been developed to obtain an optimized tau selection algorithm. I took the responsibility to develop a complete new tau trigger algorithm at Level-1 (L1, hardware based first level of the CMS trigger system). This original approach is aiming at producing the first hardware tau lepton trigger efficient at a hadron collider. I had the opportunity to present the results of my work at the ICHEP-2014 conference, in Valencia and the proceedings were published in Nuclear Physics B afterwards. During my last year of PhD I focused again on the Higgs decays into di-taus analysis, initiating the very first matrix element (ME) approach in this channel, starting with the most sensitive final state: the semileptonic decay mode of the Higgs boson produced through Vector Boson Fusion (VBF) mechanism. No ME-based analysis using tau leptons has ever been published. The aim is to increase the sensitivity of the analysis to the SM Higgs boson, with respect to the conventional methods such as cut-based and multi-variate analyses. The novelty of my work is the treatment of the tau decay through transfer functions. In addition to the computation of the transfer functions, my contribution consisted also in the full characterization of the method and its validation using Monte Carlo samples. Subsequently, I evaluated its performance in the context of the CMS Higgs into di-taus search. The application of the ME method lead to an amelioration of the analysis significance of $\sim 8\%$ in the semileptonic channel ($\sim 30\%$ considering only the VBF-tag categories) and to an observation of the decay $H \to \tau\tau \to \mu\tau_h$ with a significance of 3.1σ.

Résumé

J'ai effectué ma thèse en physique des particules au laboratoire Leprince-Ringuet à l'École Polytechnique. J'ai participé à l'analyse des collisions proton-proton à 8 TeV produites par le Large Hadron Collider (LHC) et collectées par le détecteur CMS. La découverte du boson de Higgs a été un événement majeur pour la physique des hautes énergies car la masse des bosons vecteurs provient de leur interaction avec le champ de Higgs. Le couplage du boson de Higgs avec les leptons n'ayant pas encore mis en évidence, mon travail de thèse s'est articulé autour du canal de désintégration en pair de taus, car c'est le seul qui permet de mettre ce couplage en évidence. Le ddésintdégration du boson de Higgs en paires de tau constitue le canal le plus favorable pour mesurer le couplage du boson de Higgs aux fermions. En effet, ce canal bdéndéficie d'un rapport d'embranchement important compardé aux autres leptons.

Cette analyse est difficile au niveau du déclenchement. En effet, l'important bruit de fond impose d'appliquer des seuils en énergie élevés. Un critère de déclenchement utilisant l'énergie transverse manquante pour abaisser le seuil sur les leptons a été introduit à la fin de la prise de données. Cette approche permet de récupérer 41% d'événements en plus. J'ai effectué une caractérisation exhaustive de ce déclenchement, et en particulier étudié les erreurs systématiques associées au niveau de l'analyse. Ce déclenchement a permis d'améliorer de 2% la limite d'exclusion dans le canal Higgs en tau-tau semi-leptonique. Au Run 2, l'énergie des collisions a été portée à 13 TeV et la luminosité va augmenter. Pour garantir un déclenchement efficace pour la physique, le premier niveau du système de déclenchement (L1) a été remplacé en se basant sur la technologie microTCA. Ce nouveau système permet un accès à des informations plus détaillées en provenance du détecteur. J'ai développé, pour ce nouveau système, un algorithme de

sélection des taus basé sur un algorithme d'agrégation dynamique exploitant au maximum la granularité du calorimètre. J'ai mesuré ses performances en terme d'efficacité et de taux de déclenchement, et pour la première fois auprès dun collisionneur hadronique, CMS disposera en 2016 d'un déclenchement efficace sur les taus au L1 avec un taux raisonnable. J'ai présenté ce travail sous forme de poster à la conférence ICHEP 2014. Dans la dernière partie de ma thèse, je me concentre sur la recherche du boson de Higgs se désintégrant dans une paire de leptons taus, en me concentrant sur le canal semi-leptonique et en introduisant pour la première fois la méthode des éléments de matrice (MEM) qui permet d'améliorer la sensibilité par rapport aux méthodes traditionnelles. Aucune analyse utilisant la MEM avec des taus n'a jamais été publiée auparavant. L'innovation dans ce travail consiste notamment dans le traitement de la désintégration des taus. La réponse du détecteur est quant-à-elle modélisée par des fonctions de transfert. Puis, j'ai du tenir compte des limitations numériques intervenants dans l'évaluation des intégrales multidimensionnelles à la base de cette méthode. Enfin, j'ai mesuré le gain en sensibilité apporté par la méthode sur les données à 8 TeV. J'ai évalué sa performance dans le contexte de la recherche du boson de Higgs produit par fusion de bosons vecteurs que se désintègre en paires de taus. L'application de la méthode conduit à une amélioration de la significance de l'analyse d'environ $\sim 8\%$ dans le canal semileptonic ($\sim 30\%$ considérant que les catégories VBF) et à une observation de la désintégration $H \to \tau\tau \to \mu\tau_h$ avec une signification de 3.1σ.

Remerciements

Paris, 23 Octobre 2015

Je voudrais sincèrement remercier Florian Beaudette. Je peux affirmer que je n'aurais pas pu avoir un meilleur directeur de thèse. Grâce à sa constante bonne humeur, à sa patience infinie et sa pédagogie, il m'a encadré de façon efficace. Il a su se rendre disponible pour échanger des idées et me guider tout au long de ma thèse. Il a également su me motiver notamment en valorisant mon travail. Je lui suis reconnessant d'avoir toujours mis en avant les points positifs du travail que j'effectuais. Il m'a permis de sortir des résultats très intressants et innovants pendant ces trois années. Ma gratidute envers Florian restera infinie.

Une pense chaleureuse à celui qui fut, pour presque la moitié de ma thèse, comme mon deuxieme encadrant. Je parle bien évidement d'Alexandre Zabi, appelé communment Alex! Grâce à sa attitude et son professionnalisme il m'a appris à ne jamais laisser tomber. Grâce aux responsabilités qu'il m'a données, j'ai dépassé mes limites, j'ai appris à travailler dure et à m'adapter à toutes les situations qu'on peut rencontrer dans CMS. Je n'oublierai jamais "la conference internationale de nuit" et tous ces moments passés à "fabriquer le systeme". Je peux aussi affirmer que j'ai trouvé un veritable ami, avec lequel je me suis beaucoup amusé pendant ce trois années.

Je voudrais aussi remercier Jean-Baptiste et Olivier, ils m'ont été d'une aide précieuse dans les aspects plus techniques de mon travail de thèse. Sans eux, il ne fait aucun doute, que je n'aurais jamais pu comprendre dans un temps aussi bref tout les details des sujets complexes et sourtout m'améliorer en informatique aussi rapidement. Je remercie en particulier Olivier pour tout le temps qu'il ma dédié. J'ai bien conscience que tu étais sur le

point de craquer, et pourtant tu m'as soutenu jusqu'au bout, et je t'en suis reconnaissant. Je voudrais remercier aussi Sthéphanie. Cela a été une plaisir de partager le bureau avec toi ainsi que les pauses. Tu m'as également soutenu et prodigué de precieux conseils, sur le travail et pas seulement...

Lors que ma dernière année de thèse, il a été plus qu'un plaisir de pouvoir travailler avec les "MEMbers": Pascal, Thomas et Gilles. Chaque d'eux a apporté son extraordinaire contribution au développement de mon dernier sujet de thèse. Grâce à Pascal, à ses intuitions et ses conseils très pertinents, j'ai pu avancer rapidement dans les aspects plus théoriques. Avec Thomas, avec lequel je partage ma passion, je me suis intéressé aux calculs nécessaires au développement de la méthode de l'élément de matrice, et j'ai réussi à converger directement au but. Enfin, un très gros merci à Gilles, il s'est occupé avec brio des aspect informatiques de mon analyse.

Trois ans ont passé depuis le jour que j'ai fait leur connaissance et je voudrais les remercier de manière très spéciale: Ivo, Nadir et Yacine. Tous les trois m'ont accueilli très chaleureusement lors de mon arrivée en France et ils m'ont guidé avec de précieux conseiles pendant mes premiers deux ans de thèse. J'ai trouvé en eux aussi des amis très fiables, sur lesquels je peux compter en toutes situations. Pendant ma deuxieme année de thèse Nadir a nous laissé pour la Belgique et, Xavier, alors stagiaire au LLR, il s'est ajouté au groupe. Je voudrais le remercier lui aussi pour sa contribution à la bonne réussite de l'approval que m'a permis de participer a ICHEP-2014. Je me souviendrais toujour de notre bières au R1 et des nos discours sur la politique, l'histoire... on paralait pas vraiment de physique eh!.

Un merci aussi à Yves Sirois, en tant que chef du groupe CMS au LLR, pour ces conseils plus que profitables pendant le meeting au LLR et nos déjeuners. Je voudrais aussi le remercier pour ses inestimables encouragements donnés lors des mes présentations au CERN et pour son soutien. Enfin, Yves fait parti des responsables de la bonne réussite de ma soutenance: en fait c'est grâce à lui que j'ai commencé finalement à modifier l'introduction de ma présentation de thèse deux jours avant la soutenance!!

Enfin, une forte pensée va aussi à un stagiaire spécial, dénommé Valdo, que est arrivé en stage lors de mes derniers mois de thèse. Il m'a aidé à survivre pendant la fermeture d'été de l'Ecole Polytechnique, et je me souviendrai toujours de nos discussions quasi philosophiques sur la physique théorique. Dans des moments où la motivation pouvait manquer, son enthousiasme m'a fait revenir à la memoire la raison pour laquelle j'ai choisi

de faire de la physique.

Many thanks also to Michal Bluj and Christian Velkeen. Michal's help in the very beginning of my thesis was very precious, as also his unlimited calm and patience in explaining physics to newcomers like me. It was for me really an honour to share the office with you. Christian helped me in different occasions during this three year, he represented for me the "tau reference" and it has been thanks to him if I got more acquainted with the analysis software. His precious helps in different occasion and our discussion at CERN, helped me (together with the precious helps of Lorenzo Bianchini, that I wish also to thank) to define the base ingredient to develop the matrix element method in the Higgs tau tau analysis.

Vorrei inoltre ringraziare due persone che, in modo e in periodi diversi, mi sono state molto vicine durante gli ultimi mesi della mia tesi. Grazie a Mikaela, di cui ricordo volentieri la sua simpatia e solaritá, nonché le sue doti culinarie! Grazie anche a Adrien, che sin dal primo momento si é rivelato un amico fedele su cui poter contare. Grazie per avermi fatto scoprire Parigi e i suoi segreti, e per avermi in qualche modo ispirato a ricominciare a suonare il pianoforte, cosa che si é rivelata molto utile nei momenti di stress prossimi alla discussione!

Un grazie un po particolare, va a te, Iris. Tu che hai accettato senza esitazione i miei ritmi di lavoro serrato durante quasi due anni, tu che mi sei stata allo stesso tempo cosi vicina e cosi lontana, contribuendo però a definire meglio i miei obiettivi non solo in campo lavorativo. Abbiamo passato momenti difficili, ma la fine di questa tesi e il sostegno che mi hai dato nonostante tutto, testimoniano come alla fine ci siamo ritrovati. A te che hai saputo sostenermi nei momenti difficili di questo lavoro, soprattutto nell'ultimo periodo prima della difesa, sempre con il sorriso. Sei sempre stata molto curiosa rispetto al mio lavoro al punto da raggiungermi nella sessione poster di ICHEP-2014 a Valencia! Grazie a te, con cui ho vissuto nel marais di Parigi, a due passi dal Louvre e da Notre Dame, a te che mi hai fatto innamorare non solo della tua persona ma anche di questa splendida cittá in cui ho avuto il piacere di vivere e di questa nazione, la Francia, in cui è stato un onore lavorare.

Per concludere, vorrei infine ringraziare tutti i membri della mia famiglia, mio papá, mia mamma e mia sorella Giulia (e pure i nonni!). Mi avete sostenuto in ogni modo, incessantemente, moralmente (e a volte anche economicamente) durante tutti questi tre anni. Avete ben sopportato la mia lontananza da Milano e avete compreso a fondo

l'importanza dell'impegno in cui mi ero profuso. Grazie per le vostre visite a Parigi, e per i bellissimi week end passati insieme. Mi avete fatto veramente capire cosa vuol dire essere una famiglia e penso che voi siete la piu bella che potessi avere. Grazie mamma perche ogni volta che parlavo una parola in piu di francese mi hai sempre ricordato di non dimenticare mai l'italiano. Grazie infinite. Un grazie speciale é dedicato anche a tutti i miei amici, in particolare Federico e Francesco, che mi sono stati vicini durante la tesi e hanno sempre saputo sostenermi durante questi anni.

Introduction

"Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza."

D. Alighieri, La Divina Commedia - Inferno, XXVI canto (vv. 118-120)

The particle physics represents our deeper, or at least our more fundamental, understanding of Nature by studying the elementary constituents of the matter and the fundamental forces acting between them.

The story of particle physics begun with the identification of the electron by J.J. Thomson et al. in 1897. During the last century, up to the recent discovery of the Higgs boson in 2012, many other fundamental particles have been found with increasing masses. In order to correctly describe very small objects moving very fast, both Quantum Mechanic and Special Relativity are needed. This fusion has been carried out successfully by a more general theory: the Quantum Field Theory (QFT). Within this theory a particles is identified as the local excitation of a quantized field, and forces are described through the exchanges of mediator-particles. The association of a conserved current to a symmetry of a physical system is the important thesis of the Noether's theorem, and it is largely used to describe the interactions between the elementary particles. In particular, it turned out that the proper geometrical symmetries were found to be local gauge symmetries.

So far, three of the four forces observed in Nature have been successfully described by local gauge-invariant QFTs: the strong interaction described by the Quantum Chromo-

dynamic (QCD, responsible of the strong force among the quarks, the constituents of the protons and neutrons), the weak interaction (responsible of the decays of the atomic nuclei) and electromagnetic interaction (QED). The gravitational force represents a story apart, due to difficulties in the quantization of the gravitational field in the context of the General Relativity. The whole theoretical framework used to describe these forces is called the Standard Model (SM) of Particle Physics. The SM has been developed taking care of the main experimental discoveries performed in the last century and it is very successful to describe the large variety of phenomena observed at the subatomic scale.

Searching for new particles often means increasing the energy at which the interactions between the elementary particles are probed. To do so, during the last century, particles accelerators have been invented. Such experimental devices, used to test the theoretical predictions, have also considerably evolved, allowing to explore increasingly high energies. The Large Hadron Collider (LHC), at CERN, enables to probe physics at unprecedented energy scales and luminosity regimes. It is designed to collide protons at a center-of-mass energy of $13\,TeV$ and at an instantaneous luminosity of $10 \times 34\ cm^2 \cdot s^{-1}$. Four experiments of high complexity have been designed and installed to collect and analyse the corresponding collision data; two of them, ATLAS and CMS, are general-purpose experiments.

However, until the 2012, a mystery of Nature was how elementary particles gain their mass which is observed but cannot easily be explained by theory. In the frame of the SM the solution proposed by Higgs, Brout, Englert, Guralnik, Hagen and Kibble in 1964 is that elementary particles gain mass through a dynamical mechanism. This leads to the prediction of a new elementary particle, called the Higgs boson, then accountable for elementary particles masses. This particles has been discovered at the LHC on July 4th 2012. Therefore, given the observation of the Higgs boson, to precisely measures its properties represents one of the LHC major goal in the future. The aura of mystery around the Higgs boson is far from being dissolved. Indeed,the Higgs boson is, up to now, the only discovered boson that is not a gauge boson. Moreover, what triggers the spontaneous symmetry breaking? Is the Higgs boson a fundamental particle or is it a composite one? Is there are more than one Higgs doublet? All of these are still unanswered questions up to these days. Recently, the observation of the direct decay of the Higgs boson into a pair of τ-lepton, represents the first observation of the coupling between the Higgs boson and the fermions, in particular into a pair of lepton τ.

The $H \rightarrow \tau\tau$ decay is the most promising channel to measure the couplings between

the Standard Model (SM) Higgs boson and the fermions. During my thesis work I have been concentrating on various keypoints of this analysis. I have carried out studies related to both the τ lepton identification at trigger level and on the $H \to \tau\tau$ signal extraction. During the shutdown of the LHC, I have worked in the context of the Run 2 CMS upgrade. This upgrade is needed to face the new intense hadronic environment generated by an higher luminosity ($2 \cdot 10^{34}\ cm^{-2}s^{-1}$) and an higher collision energy (13 TeV). Indeed, a large part of my thesis work was devoted to the development of a complete new τ trigger algorithm at L1. This project is part of the Phase-1 CMS calorimeter trigger upgrade, which will be setup by 2015-2016. This original approach is aiming at producing the first hardware τ lepton trigger efficient at a hadron collider. These important results were presented at the ICHEP-2014 conference, and published afterwards in Nuclear Physics B. The τ physics represents also an excellent testbed to probe the physics beyond the standard model. This represents one of the LHC major goals in the already started Run 2. In this context the new τ trigger can play an important role thanks to its large dynamical range in term of signal efficiency for a sustainable trigger rate, allows for the usage of multiple working points suited for different possible analysis. The studies performed for the development of the stage 2 τ trigger, represent, in addition, a full scale test-bench for the upcoming HGCAL trigger.

Despite the discovery of the $H \to \tau\tau$ obtained combining the results from ATLAS and CMS, no standalone observation, by none of the two collaborations, has been claimed yet. Thus one of the next priority for CMS will be the standalone observation of the Higgs boson coupling to τ leptons during the Run 2. The vector boson fusion (VBF) Higgs boson production mode plays a particular role in the analysis. Exploiting the kinematic of the high energy jets allows the background to be reduced. Thus, the VBF category contributes significantly to the sensitivity of the analysis. I therefore concentrated to the VBF production mode and looked for improvement of the signal extraction procedure. It is in this context that I proposed the very first matrix element (ME) approach in the $H \to \tau\tau$ channel, starting with the most sensitive final state: the semileptonic one. The aim is to increase the sensitivity of the analysis to the SM Higgs boson, with respect to the conventional methods such as cut-based and multi-variate analyses. The results obtained are really promising and the application of such a method can be extended to all the possible di-τ decay channel considered in the $H \to \tau\tau$ analysis. The Run 1 saw the transition between cut-based analyses and multivariate analyses, and the Matrix Element Method is one of them. It represents, among all the other multivariate analysis techniques, the most promising when a search for a physical process is performed within a defined theoretical model. At the same time, the Matrix Element Method is one of the most challenging analysis approaches as it requires advanced computing techniques such

as those I had the opportunity to use during my thesis work. During the LHC Run 2 more and more studies related to the different exclusive Higgs boson production mode will be performed. Also in this context, the Matrix Element Method can be applied to successfully target a particular production mechanism, as it has been done in my thesis work for the higgs produced through the vector boson fusion.

Contents

Chapter 1

The Standard Model of Particle Physics and the Higgs Boson

1.1 Introduction

The Standard Model (SM) of electroweak (EWK) and strong interactions is the outcome of one century of interplay between experimental measurements and theoretical developments. Proposed by Glashow, Salam and Weinberg in the middle sixties [1] [2], it has been later extensively tested over a wide range of energies. The discovery of the neutral current interactions [3] [4] and the production of intermediate vector bosons ($W^{\pm}$ and Z^0) at SPS, with the expected properties, increased the confidence in the model. The EWK parameters have been measured with high precision at the Large ElectronPositron Collider (LEP), contributing to successfully test the model with sensitivity to EWK radiative corrections [5]. However, a cornerstone of the model, i.e. the invariance of the theory under gauge transformations, manifestly clashes against the experimental evidence that some of the gauge bosons (the force-mediators) are massive. Indeed, the gauge symmetries of the Standard Model forced all the elementary particles to be massless. In the '60 a solution was proposed to solve this apparent ambiguity, taking inspiration from symmetry-breaking phenomena occurring in condensed matter: the gauge symmetry of the theory is spontaneously broken by the vacuum state being no longer invariant under an arbitrary gauge transformation. This mechanism has gone down in history as the spontaneous EWK symmetry breaking mechanism, known also as the *Brout-Englert-Higgs-Hagen-Guralnik-Kibble mechanism* (BEH) [6][7][8][9][10][11]. The BEH (often simply named *Higgs mechanism*) is a fundamental ingredient of the theory, that requires the existence of a neutral scalar field, generating all the elementary particles masses and guaranteeing the renormalizability of the theory. The excitations of such a field are inter-

preted as a new elementary particle, namely the Higgs boson that has been experimentally discovered on July 4th, 2012 by the CMS and ATLAS collaborations [12][13] confirming the BEH mechanism. One of the historical plot is presented in Fig. 1.1. The study of its properties represents now one of the main goals of the Large Hadron Collider (LHC) in the near future, which, colliding protons at an energy in the center-of-mass of the collisions of $13\,TeV$, will have direct access to unexplored energies. Both CMS and ATLAS are general purpose experiments and they have been designed to explore a great variety of phenomena beyond the SM [14] [15]. The LHC schedule presents many physics goals that have to be achieved during the machine operation time. According to our current understanding there seems to be indications pointing towards a unification of the strong and EWK forces. We know, however, that our picture of the observed forces and particles is incomplete. The Higgs boson was the last missing ingredient of the SM of particle physics. However, even if a SM-like Higgs boson has been already found, the SM cannot be the ultimate theory [16], which is obvious already from the fact that it does not contain gravity. Another problem in the SM is represented by the so called *hierarchy problem*. It is one of the great unsolved fundamental questions of modern physics. It arises from the fact that there are 16 orders of magnitude between EWK unification scale ($\sim 100\,\text{GeV}$) and the Plank's one ($10^{19}\,\text{GeV}$), where the strengths of gravity and the other interactions become comparable. In the last decades, many theories tried to solve this question, and an ideal candidate might be a string theory. Nevertheless, no experimental evidence of such a theory has been found up to now. Also Dark Matter represents an experimental evidence that goes behind the SM. A very attractive possibility of new physics that stabilizes the hierarchy between the EWK and the Planck scale is supersymmetry (SUSY) [17][18][19][20]. Supersymmetric models predict the existence of superpartners; to each fermion of the SM associates a scalar particle, to account for the degree of freedom present in the super-symmetric lagrangian. Supersymmetric theories allow the unification of the strong, electromagnetic and weak interactions at a scale of about $10^{16}\,\text{GeV}$. In such a theory, the strong, electromagnetic and weak interactions can be understood as being just three different manifestations of a single fundamental interaction. Probing SUSY scale represent another goal of LHC [16]. In the same unifying framework lies extra-dimensions research, that could give a test for string theory answering the question of why the gravitational interaction is so much weak compared to other forces. At the Planck scale, in models that predict the existence of extra-dimensions, the gravitational force shows its quantum nature and it can be probed at LHC by searching for black holes [21][22]. LHC represents a unique instrument also to explore cosmological issues, such as the observed asymmetry of matter-antimatter in the Universe, investigating beauty quark physics and probing the existence of the quark-gluon plasma. In order to do that LHC is capable to

collide lead ions to recreate the conditions just after the Big Bang in laboratory: ALICE is the detector designed to analyse this type of events while LHCb perform b quark physics analysis.

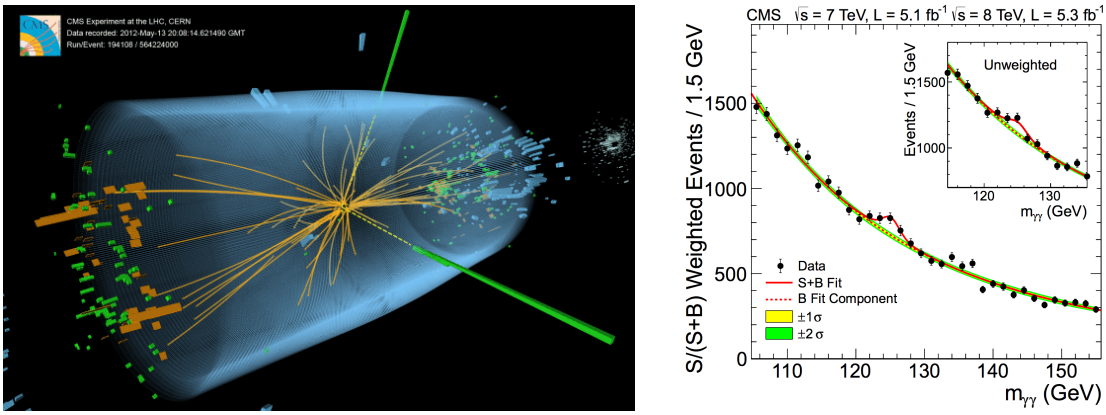


Figure 1.1: *Left: a candidate event for* $H \to \gamma\gamma$ *as it is shown in the CMS detector. Right:* $\gamma\gamma$ *invariant mass plot: to notice, the excess around* $m_{\gamma\gamma}$ *related to the Higgs boson events over the SM background [12].*

1.2 The Standard Model of particle physics

The Standard Model of particle physics (SM) is the theory that describes our current understanding of the matter structure and the fundamental interactions occurring in Nature. The SM is developed as a mathematical theory based on the least number of axiomes: the principle of least action and gauge symmetries.

- To each physical system is associated a scalar function of its generalized coordinates (degrees of freedom of the system): the lagrangian. The evolution of the system will follow the trajectory in the phase-space of the generalized coordinates that minimizes the variation of the action. The action is defined as the time integral of the lagrangian.

- The Noether theorem[23] proves that the invariance of the lagrangian of a system under a given transformation (symmetry) implies a conserved physical quantity.

- In order to describe particles, microscopic objects moving almost at the speed of light, the simultaneous use of quantum mechanic and special relativity is mandatory. Quantum field theory successfully realizes the merging of these two theories. It

describes particles as excitations of fields. Moreover, these fields are operators in the sense of quantum mechanic satisfying the commutation rules.

The most relevant example of quantum field theory used to describe physics phenomena is Quantum Electrodynamic (QED). This theory successfully describes the electromagnetic interaction that occurs between charged elementary particles. The QED lagrangian is invariant under the Lorentz transformations and global phase transformations; this invariance translates into the conservation of the electric charge. The interaction between charged particles is described as the exchange of a virtual particle that acts as mediator of the force (a boson, since it is demonstrated to have spin integer). The possibility to infer the electromagnetic interactions from prime principles, in addition to the huge amount of experimental confirmations of QED predictions, inspired a generalization of this approach in order to describe other interactions.

Weak interaction is the mechanism responsible for the weak force, one of the four known fundamental interactions of nature together with the strong interaction, electromagnetism, and gravitation. The weak interaction is responsible for the radioactive decay of subatomic particles, and it plays an essential role in nuclear fission. In the Standard Model of particle physics, the weak interaction is described by the exchange of a W and Z bosons. The mass of the W and Z gauge boson is such that the range of the weak force is very short ($\sim 10^{-18}$ m). All known fermions interact through the weak interaction. The force is termed weak because its field strength over a given distance is typically several orders of magnitude less than that of the strong nuclear force and electromagnetic force. The model of the EWK interactions [1, 2, 24] is a quantum field theory based on the gauge symmetry group $SU(2) \otimes U(1)$ of weak left-handed isospin and hypercharge. The gauge bosons W_{μ_i} , $i = 1, 2, 3$ and B_μ are associated to the $SU(2)$ and $U(1)$ factors respectively, and the corresponding gauge coupling constants are named g and g'. This theory, together with Quantum Chromodynamics (QCD) [25, 26], the theory explaining the strong forces between quarks, described in terms of a $SU(3)$ gauge theory, yields to the Standard Model of particle physics: a quantum field theory having as a group of symmetry: $SU(3) \bigotimes SU(2) \bigotimes U(1)$. The SM provides a unified framework to describe three of the four fundamental forces currently known.

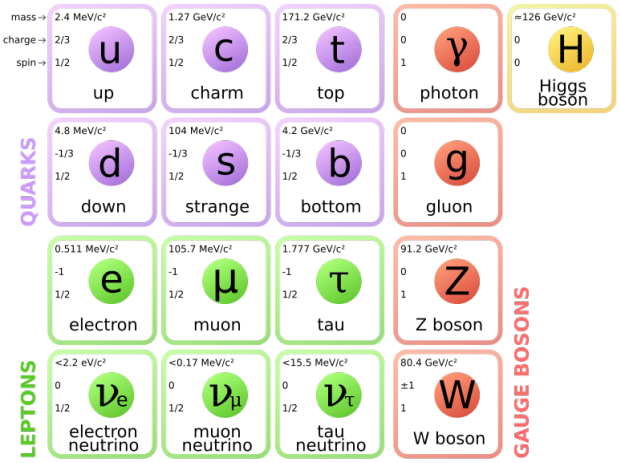


Figure 1.2: *The constituents of the SM, both for matter and forces fields. The SM Higgs boson is represented. The graviton is put aside as it is not included in the SM, and has not been observed.*

The SM has two kinds of fields. The fundamental matter fields are quarks and leptons. They have spin 1/2 and appear in three generations. Gauge fields correspond to the spin-1 bosons that mediate the interactions (Fig. 1.2). In this section, we will go a bit more into the details of the EWK unification. Starting from the Fermi lagrangian and taken as done all the experimental milestone about weak interaction experienced in the last century, it is possible to describe the weak interaction as a non-abelian gauge theory. This is also enforced by the fact that only a non-abelian gauge theory could resolve the problem of the non-unitarity of the scattering matrix and at the same time ensure the renormalizability of the theory as requested by a correct quantum field theory. The starting point is the experimentally observed charged current:

$$J_\mu^+ = \bar{e}\gamma^\mu \left(\frac{1-\gamma^5}{2}\right)\nu \qquad\qquad J_\mu^- = \bar{\nu}\gamma^\mu \left(\frac{1-\gamma^5}{2}\right)e$$

After defining the isospin doublet in the following way:

$$\Psi_L = \begin{pmatrix} \nu \\ e \end{pmatrix}_L \tag{1.1}$$

it is possible to write the expression above in a unique way, similarly to that of electromagnetic current:

$$\boxed{\boldsymbol{J}^\pm = \bar{\Psi}_L \gamma^\mu \boldsymbol{\tau}_\pm W_\mu^\mp \Psi_L} \tag{1.2}$$

where $\boldsymbol{\tau}^+ = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\boldsymbol{\tau}^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. If we want to identify the matrix $\boldsymbol{\tau}$ as the generator of the transformation of the SU(2) Lie group, $\boldsymbol{\tau}$ must obey the commutator relations of the corresponding SU(2) Lie algebra:

$$\left[\boldsymbol{\tau}^+, \boldsymbol{\tau}^-\right] = \tau^3 \tag{1.3}$$

where $\boldsymbol{\tau^3} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is diagonal in this representation. If two generators of a non-abelian group exist, the commutator of them is also a generator of the same group. The importance of this fact is that the third matrix was not requested a priori but is a direct consequence of imposing a description in term of a gauge theory. In addition, a conserved current is associated to τ^3: $J^3 = \bar{\nu_L}\gamma^\mu\nu_L W^3_\mu - \bar{e_L}\gamma^\mu e_L W^3_\mu$. If a conserved current exists, it implies that the spatial integral of this current over all the 3D space is also a conserved charge. For the current J^3, the quantum number that is associated to the conserved charge is defined as the hypercharge and its symbol is Y. If the generator are written in an Hermitian mode, the τ matrices are exactly the Pauli matrices and the gauge fields become $\boldsymbol{W^1} = \frac{\boldsymbol{W^+ + W^-}}{\sqrt{2}}$, $\boldsymbol{W^2} = i\frac{\boldsymbol{W^+ - W^-}}{\sqrt{2}}$, $\boldsymbol{W^3}$. The $\boldsymbol{J^3}$ is a neutral current but could not be identified with the electromagnetic one mainly because of two reasons:

- The electromagnetic current couples left-handed and right-handed spinors;
- The electromagnetic current does not couple neutrinos.

A possible solution to recover the electromagnetic current is to extend the gauge symmetry $SU(2) \longrightarrow SU(2) \otimes U(1)$, leaving the EWK lagrangian invariant. The gauge field associated to new symmetry group is B_μ and it must couple with the same coupling constant to each member of the isospin doublet. Also introducing the right-handed component, the isospin singlet: e_R, ν_R, the lagrangian becomes:

$$\mathcal{L}_{EW-free} = \bar{\Psi}_L \not{\partial} \Psi_L + \bar{e}_R \not{\partial} e_R + \bar{\nu}_R \not{\partial} \nu_R + \frac{g}{2}\bar{\Psi}_L \gamma_\mu \tau_i \Psi_L W^\mu_i + \frac{g}{2}\bar{\Psi}_L \gamma_\mu \tau^3 \Psi_L W^\mu_3 + \frac{g'}{2} B^\mu \left[Y_L \bar{\Psi}_L \gamma_\mu \Psi_L + Y_{\nu_R} \bar{\nu}_R \gamma_\mu \nu_R + Y_{e_R} \bar{e}_R \gamma_\mu e_R \right]$$

in this case, the index $i = 1, 2$. This interaction term arises after replacement of the derivative operator with the covariant derivative operator:

$$\not{\partial} \Longrightarrow \not{\mathcal{D}} = \gamma_\mu \left[\partial_\mu - \frac{i}{2} g W^\mu_i \tau^i - \frac{i}{2} g' Y B^\mu \right] \tag{1.4}$$

The lagrangian can take the more concise form:

$$\mathcal{L} = \bar{\Psi}_L \not{\mathcal{D}} \Psi_L + \bar{e}_R \not{\mathcal{D}} e_R + \bar{\nu_R} \not{\mathcal{D}} \nu_R \tag{1.5}$$

It is possible to further reduce the form of the lagrangian introducing the multiplet:

$$\xi = \begin{pmatrix} \Psi_L \\ \nu_R \\ e_R \end{pmatrix} \tag{1.6}$$

So that:

$$\mathcal{L} = \bar{\xi} \not{\mathcal{D}} \xi \tag{1.7}$$

To obtain the fundamental relationship of the SM, it is mandatory to consider the neutral current interaction part of the lagrangian:

$$\mathcal{L}_{\mathcal{NC}} = \frac{g}{2} \bar{\xi} \gamma_\mu \tau_3 \xi W_3^\mu + \frac{g'}{2} \bar{\xi} \gamma_\mu Y \xi B_\mu. \tag{1.8}$$

Since the τ_3 matrix acts in a different way on the *Left-Handed* (LH) and *Right-Handed* (RH) spinors (in fact it assumes the values ± 1 on the doublet elements but 0 on right handed fermions), it is natural that also Y takes different values when acting on different type of spinors. Since Y it is always coupled with g' is possible to arbitrarily fix its values on one field. The common practice consists in fixing Y in order to recover the electromagnetic current. The first step is to perform the Weinberg rotation (θ_W is called the Weinberg angle):

$$\begin{cases} B^\mu = A^\mu \cos\theta_W - Z^\mu \sin\theta_W \\ W_3^\mu = A^\mu \sin\theta_W + Z^\mu \cos\theta_W \end{cases} \tag{1.9}$$

Making this substitution, in order to recover the electromagnetic current, one should have:

$$\frac{\tau_3}{2} g \sin\theta_W + \frac{Y}{2} g' \cos\theta_W = eQ \tag{1.10}$$

Thus, it is possible to fix the hypercharge value on the left-handed electron field:
$Y(e_L) = -1$

$$\boxed{g \sin\theta_W = g' \cos\theta_W = e} \tag{1.11}$$

From Eq. 1.11 and 1.10 results an other important relation of the SM known as the Gell-Mann-Nishijima formula, which relates the electric charge to the isospin and the hypercharge:

$$\boxed{T_3 + \frac{Y}{2} = Q} \tag{1.12}$$

In Tab. 1.1 there are summarized all the relevant information in terms of the EWK quantum numbers.

As already pointed out, the electromagnetic interaction is only a part of the neutral

field	Q	T_3	Y
ν_L	0	$+\frac{1}{2}$	-1
e_L	-1	$-\frac{1}{2}$	-1
ν_R	0	0	0
e_R	-1	0	-2
u_L	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$
d_L	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$
u_R	$+\frac{2}{3}$	0	$+\frac{4}{3}$
d_R	$-\frac{1}{3}$	0	$-\frac{2}{3}$

Table 1.1: *The most relevant quantum number of the SM for the matter fields. Note that are presents also the quarks.*

current interaction lagrangian. The missing term is made of weak neutral current:

$$\bar{\xi}\gamma_\mu \left[g\frac{\tau_3}{2}\cos\theta_W - g'\frac{Y}{2}\sin\theta_W \right] \xi Z^\mu. \tag{1.13}$$

Using again equations 1.11 and 1.12, it is possible to obtain the third important relation of the SM, that express the strength of the coupling to Z^0 intermediate boson:

$$\boxed{Q_Z = \frac{1}{\sin\theta_W \cos\theta_W}\left[T_3 - Q\sin^2\theta_W\right]}. \tag{1.14}$$

The last part concerning the electroweak unification topic regards the auto-interaction between the gauge bosons. As a matter of fact this peculiarity is typical of non-abelian gauge theories and directly derives from the non-vanishing commutation rules between the vector gauge bosons present in the theory. As a consequence, the kinetic therm in the lagrangian has got an extra term:

$$\mathcal{L}_\mathcal{B} = -\frac{1}{4}B_{\mu\nu}(x)B^{\mu\nu}(x) \quad \text{where} \quad B_{\mu\nu}(x) = \partial_\mu B_\nu(x) - \partial_\nu B_\mu(x) \tag{1.15}$$

$$\mathcal{L}_\mathcal{G} = -\frac{1}{4}G^i_{\mu\nu}(x)G^{\mu\nu}_i(x) \quad \text{where} \quad G^i_{\mu\nu}(x) = \partial_\mu W^i_\nu(x) - \partial_\nu W^i_\mu(x) + g\varepsilon_{ijk}W^j_\mu(x)W^k_\nu(x) \tag{1.16}$$

leads to the couplings shown in Fig. 1.3

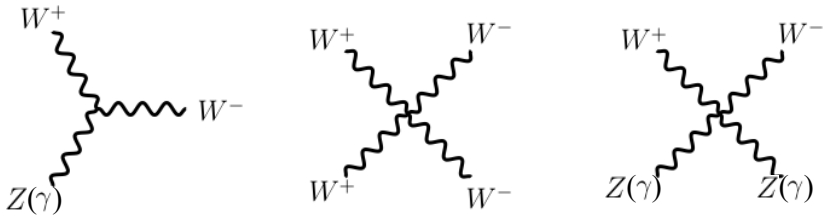


Figure 1.3: *Gauge bosons autointeraction: trilinear and quadrilinear terms.*

To conclude this section, the full EWK lagrangian is thus summarized:

$$\begin{aligned}\mathcal{L}_{EW} &= \bar{\Psi}_L \not{\partial} \Psi_L + \bar{e}_R \not{\partial} e_R + \bar{\nu}_R \not{\partial} \nu_R \frac{g}{2} \bar{\Psi}_L \gamma_\mu \tau_i \Psi_L W_i^\mu + \frac{g}{2} \bar{\Psi}_L \gamma_\mu \tau^3 \Psi_L W_3^\mu \\ &+ \frac{g'}{2} B^\mu \left[Y_L \bar{\Psi}_L \gamma_\mu \Psi_L + Y_{\nu_R} \bar{\nu}_R \gamma_\mu \nu_R + Y_{e_R} \bar{e}_R \gamma_\mu e_R \right] - \frac{1}{4} B_{\mu\nu}(x) B^{\mu\nu}(x) \\ &- \frac{1}{4} F_{i\mu\nu}(x) F_i^{\mu\nu}(x) + g \varepsilon_{ijk} W_{i\mu}(x) W_{j\nu}(x) \partial^\mu W_k^\nu(x) \\ &- \frac{1}{4} g^2 \varepsilon_{ijk} \varepsilon_{i\ell m} W_j^\mu(x) W_k^\nu(x) W_{\ell\mu}(x) W_{m\nu}(x)\end{aligned} \tag{1.17}$$

1.2.1 Generation of masses and Higgs mechanism

Gauge theories are incompatible with gauge boson mass terms because these will violate the local gauge invariance and the renormalizability of the theory. The most accepted model that can confer mass on the gauge bosons without violating the local gauge invariance and at the same time preserving the renormalizability is known as Higgs Mechanism [8]. In the SM the Higgs mechanism consists in applying the already known phenomena of the spontaneous symmetry breaking [27] [28] to the $SU(2)_L \bigotimes U(1)_Y$ local gauge symmetry of the SM. A spontaneous symmetry breaking can occur if there exists a field with a degenerate vacuum state. That is, a field that shows a potential having multiple ground states (state with the minimum of the potential energy) with a vacuum expectation value different from zero (see Fig. 1.4).

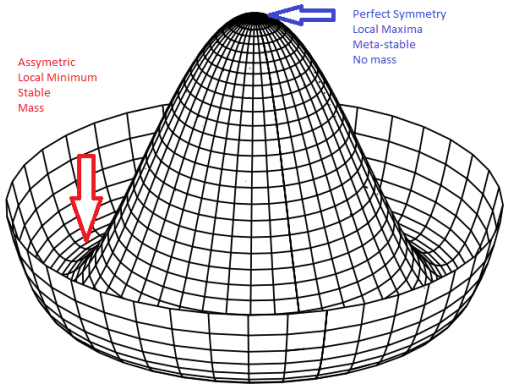


Figure 1.4: *Higgs potential representation. The infinite-degeneracy of the vacuum state describe a circle of local minima.*

Moreover, if a symmetry exists in the theory, this state is necessary degenerate. The existence of a quantum vacuum state is very important because it allows a perturbation theory to be carried on. Small excitations of the field can be possible in the neighbourhood of the ground state (identifiable with the vacuum state). Once the system is forced to assume a ground state there exist multiple possibilities (due to the symmetry of the theory the ground state is, indeed, degenerate). Once the system takes the state with the lowest possible potential energy a *choice* is performed and the symmetry is naturally broken. As a consequence, Goldstone bosons arise in the theory, so many as the degree of freedom that preserve their original symmetry. The Higgs mechanism adds something more: if the symmetry is a local gauge one, then it is possible, performing a gauge transformation, to remove all Goldstone bosons. This particular choice of the gauge is known as the unitary gauge. Moreover, in a lagrangian that is invariant under local gauge transformation the covariant derivative operator replace the usual one. This peculiarity allows the mass terms for the vector bosons occurring from the coupling between the covariant derivative and the Higgs field. In some way the Goldstone bosons degree of freedom are replaced by the vector bosons mass terms. In particular, the Higgs mechanism in the SM should fulfill the following requirements:

- It breaks the $SU(2)_L \bigotimes U(1)_Y$ symmetry but preserves the $U(1)_{EM}$ one (the gauge group of the electromagnetism);
- It preserves the Lorentz invariance of the Lagrangian;

From these requirements follows that a complex isoscalar doublet with one neutral component has to be introduced in the theory. Moreover, from Gell-Mann's formula 1.11, two choices are possible for the Higgs doublet hypercharge: ± 1:

$$Y = 1 \qquad \Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L \tag{1.18}$$

$$Y = -1 \qquad \tilde{\Phi} = \begin{pmatrix} \phi^0 \\ -\phi^- \end{pmatrix}_L \tag{1.19}$$

Now it is possible to insert in the EWK lagrangian the Higgs field terms, kinetic $|\not{\mathcal{D}}\Phi|^2$ and potential $V(\Phi) = -\lambda\left(\frac{\nu^2}{2} - |\Phi|^2\right)$, and perform the choice of the vacuum state:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{1.20}$$

Performing small oscillations $\eta_i(x)$ $i = 1, 2, 3, 4$ around this ground state is possible to identify the Higgs field with:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ \nu + h(x) + i\eta^3(x) \end{pmatrix} \Longrightarrow \Phi = \frac{1}{\sqrt{2}}(\nu + h(x)) \cdot e^{\frac{i}{\nu}\bar{\xi}\cdot\bar{\tau}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.21}$$

The above equation is true for $\xi_1 = \eta_2$, $\xi_2 = \eta_1$ and $\xi_3 = -\eta_3$ and the $\bar{\tau}$ are the isospin matrices. The next step consists in performing a local gauge transformation of the type $e^{-\frac{i}{\nu}\bar{\xi}\cdot\bar{\tau}}$ in order to recover the unitary gauge. In this gauge the Higgs doublet takes the simpler form:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \tag{1.22}$$

Considering the covariant derivative applied to the Higgs field:

$$\mathcal{D}_\mu \Phi = \left[\partial_\mu - \frac{i}{2} g \boldsymbol{\tau} \boldsymbol{W_\mu} - \frac{i}{2} g' Y B_\mu\right] \begin{pmatrix} 0 \\ \dfrac{\nu + h\,(x)}{\sqrt{2}} \end{pmatrix} =$$

$$= \frac{1}{\sqrt{2}} \left[\partial_\mu - \frac{i}{2} g \left(\tau_1 W_\mu^1 + \tau_2 W_\mu^2 + \tau_3 W_\mu^3\right) - \frac{i}{2} g' Y B_\mu\right] \begin{pmatrix} 0 \\ \nu + h\,(x) \end{pmatrix} =$$

$$= \frac{1}{\sqrt{2}} \partial_\mu \begin{pmatrix} 0 \\ \nu + h\,(x) \end{pmatrix} - \frac{i}{2\sqrt{2}} \begin{pmatrix} g W_\mu^3 + g' B_\mu & g\left(W_\mu^1 - i W_\mu^2\right) \\ g\left(W_\mu^1 + i W_\mu^2\right) & g' B_\mu - g W_\mu^3 \end{pmatrix} \begin{pmatrix} 0 \\ \nu + h\,(x) \end{pmatrix}$$

Computing the kinetic term $|\mathcal{D}_\mu|^2$:

$$\frac{1}{2}\{\partial_\mu \mathrm{H}\}^2 + \frac{1}{8}\left[\begin{pmatrix} gW_\mu^3 + g'B_\mu & g\sqrt{2}W_\mu^+ \\ g\sqrt{2}W_\mu^- & g'B_\mu - gW_\mu^3 \end{pmatrix}\begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix}\right]^2 =$$

$$\frac{1}{2}\{\partial_\mu \mathrm{H}\}^2 + \frac{1}{8}\begin{pmatrix} \left(\sqrt{2}gW_\mu^+\right)(\nu + h(x)) \\ \left(g'B_\mu - gW_\mu^3\right)(\nu + h(x)) \end{pmatrix}^2 =$$

$$\frac{1}{2}\{\partial_\mu \mathrm{H}\}^2 + \frac{1}{8}\left[\left(2g^2 W_\mu^- W_\mu^+\right)(\nu + h(x))^2 + \left(g'B_\mu - gW_\mu^3\right)^2(\nu + h(x))^2\right]$$

$$\frac{1}{2}\{\partial_\mu \mathrm{H}\}^2 + \frac{1}{4}\left(g^2 W_\mu^- W^{+\mu}\right)(\nu + h(x))^2 + \frac{1}{8}\bar{g}^2 Z_\mu Z^\mu (\nu + h(x))^2$$

Comparing the quadratic terms, the masses of the vectors bosons can be deduced:

$$\boxed{m_{W^+} = m_{W^-} = \frac{\nu g}{2}} \tag{1.23}$$

$$\boxed{m_{Z^0} = \frac{\nu\sqrt{g^2 + g'^2}}{2}} \tag{1.24}$$

This quantities yield to another fundamental relation in the SM:

$$\boxed{\frac{m_W^2}{m_{Z^0}^2} = \cos^2(\theta_W)} \tag{1.25}$$

The Weinberg angle θ_W, encountered already in Eq. 1.9, is the angle that parametrises the rotation of the original W and B vector boson plane, producing as a result the physical Z^0 boson, and the photon γ. From the potential term is obtained the mass of the Higgs boson:

$$\boxed{m_{\mathrm{H}} = \sqrt{2\lambda\nu^2}} \tag{1.26}$$

In the lagrangian other terms are present and are represented in Fig. 1.5 and 1.6 respectively:

$$\underbrace{\frac{m_W^2}{\nu^2}W_\mu^- W^{\mu+}h^2 + \frac{m_Z^2}{2\nu^2}Z_\mu Z^\mu h^2}_{\text{quadrilinear terms}} + \underbrace{\frac{2m_W^2}{\nu}W_\mu^- W^{\mu+}h + \frac{m_Z^2}{\nu}Z_\mu Z^\mu h}_{\text{trilinear terms}} \tag{1.27}$$

$$- \underbrace{\lambda\nu h^3}_{\text{trilinear autointeraction}} - \underbrace{\frac{\lambda}{4}h^4}_{\text{quadrilinear autointeraction}} \tag{1.28}$$

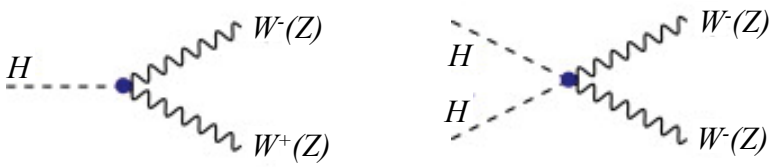


Figure 1.5: *Higgs coupling to vector boson in trilinear and quadrilinear modes.*

Figure 1.6: *Higgs autointeractions.*

From the above equations, it is possible to deduce the vacuum expectation value from the mass of the vector boson W:

$$\boxed{\nu \simeq \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246.2\,\text{GeV}} \tag{1.29}$$

where $G_F = 1.16637 \times 10^{-5} GeV^{-2}$ is the Fermi constant. This relation does not give the mass of the boson as λ is a free parameter of the Model.

Fermions masses:

The mass of the fermions of the type $m\bar{\psi}_R\psi_L + m'\bar{\psi}_L\psi_R$ could not appear in the lagrangian without breaking the gauge invariance. This is due to the different transformation properties of the Left-Handed (LH) and Right-Handed (RH) spinor fields under gauge transformations. Nevertheless, it is possible for the fermions to gain mass through their coupling to the Higgs field. The couplings between fermions and Higgs field are known as Yukawa couplings. Starting from the double allowed possibilities for the Higgs doublets, both shown in Eq. 1.18 and in Eq. 1.19, corresponding to the ± 1 values of the hypercharge, it is possible to show that

$$\tilde{\Phi} = \begin{pmatrix} \phi^0 \\ -\phi^- \end{pmatrix}_L \tag{1.30}$$

transforms exactly in the same way as Φ, thanks to the properties of the *Pauli* matrices. At this point is possible to introduce the Yukawa coupling as follows:

$$\boxed{\lambda\bar{\Psi}_L\Phi e_R + \lambda\bar{e}_R\Phi^\dagger\Psi_L + \tilde{\lambda}\bar{\Psi}_L\tilde{\Phi}\nu_R + \tilde{\lambda}\bar{\nu}_R\tilde{\Phi}^\dagger\Psi_L} \tag{1.31}$$

These terms are allowed in the lagrangian because they are singlets under both $SU(2)_L$ and $U(1)_Y$ transformations. The final step is represented by the spontaneous symmetry breaking (described in the unitary gauge) that leads to the fermion mass terms and to the fermion-Higgs field interaction terms:

$$\frac{\lambda\nu}{\sqrt{2}}\left[\bar{e}_L e_R + \bar{e}_R e_L\right] + \frac{\tilde{\lambda}\nu}{\sqrt{2}}\left[\bar{\nu}_L\nu_R + \bar{\nu}_R\nu_L\right] + \frac{\lambda}{\sqrt{2}}\bar{e}_L e_R h(x) + \frac{\lambda}{\sqrt{2}}\bar{e}_R\nu_L h(x) + \frac{\tilde{\lambda}}{\sqrt{2}}\bar{\nu}_R\nu_L h(x) + \frac{\tilde{\lambda}}{\sqrt{2}}\bar{\nu}_L\nu_R h(x) + \tag{1.32}$$

Thus the fermion masses are:

$$\boxed{m_e = \frac{\lambda\nu}{\sqrt{2}} \qquad m_\nu = \frac{\tilde{\lambda}\nu}{\sqrt{2}}} \tag{1.33}$$

Moreover, from Eq. 1.34 it is shown that also fermion couplings with the Higgs field are proportional to the fermion mass itself.

$$\boxed{\lambda = \frac{\sqrt{2}m_e}{\nu} \Longrightarrow \frac{m_e}{\nu}\left[\bar{e}_L e_R h(x)\right]...} \tag{1.34}$$

This mechanism represents one possible way to introduce neutrino masses into the SM. Other models are able to explain theoretically the existence of non-vanishing neutrino masses [29]. Instead, another question regarding fermion masses arises from the quark sector. In fact, in the lagrangian it is possible to include quark mass terms:

$$(\bar{u}_L, \bar{d}_L)\begin{pmatrix} \nu + h(x) \\ 0 \end{pmatrix} u_R \qquad (\bar{u}_L, \bar{d}_L)\begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} d_R \tag{1.35}$$

A way to give different mass to the quarks and allowing flavour-changing through weak interactions has to be found. From the equation 1.35, it is possible to infer that m_u and m_d can be different. The mass terms are not diagonal. Thus, it is possible to include u_L and d_L in a left-handed doublet and u_R and d_R in right-handed singlets, for all the quark flavours:

$$\begin{gathered} \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \equiv Q_L^i \\ u_R,\, c_R,\, t_R \equiv U_R^i \\ d_R,\, s_R,\, b_R \equiv D_R^i \end{gathered} \tag{1.36}$$

Thus, the most general terms (without Majorana masses) are summarized as follow:

$$\bar{U}_L^i M_{ij}^U U_R^i + \bar{D}_L^i M_{ij}^D D_R^j \tag{1.37}$$

where the labels D, U represents the flavour eigenstates. Nevertheless, in the weak interactions with the gauge bosons, the eigenstates are the left or right components of the fields:

$$\bar{U}_L^i \not{V} U_L^i \;\; \bar{U}_R^i \not{V} U_R^i \;\; \bar{D}_L^i \not{V} D_L^i \;\; \bar{D}_R^i \not{V} D_R^i \;\; \bar{U}_L^i \not{W} D_L^{i\,1}, \tag{1.38}$$

where V denotes the field associated either to the charger or neutral vector bosons, while W denotes the field associated only to the charged vector boson. If the attention is put on the

[1]The fields U_L and D_L are independents and transform differently from each others.

charged current, it is remarkable that interaction eigenstates do not coincide with mass eigenstates. If the theory requires to diagonalize the masses matrix it is necessary to transform the fields in the charged current interaction terms:

$$\bar{U}_L^i M_{ij}^U U_R = (A_U \bar{U}_L) m_U \, (B_U U_R) \tag{1.39}$$

where left-handed and right-handed fields transform in different ways. A and B are unitary matrices. Defining the Cabibbo-Kobayashi-Maskawa[30] matrix as:

$$\boxed{\mathcal{M}_{\mathcal{CKM}} \equiv \mathcal{A}_{\mathcal{U}}^{\dagger} \mathcal{B}_{\mathcal{D}}} \tag{1.40}$$

the charged current flavour-changing the quark masses from the Yukawa couplings become:

$$m_U = \sum_i \bar{U}_L^i m_u^i U_R^i \;\; and \;\; m_D = \sum_i \bar{D}_L^i m_d^i D_R^i \tag{1.41}$$

1.3 The Higgs boson of the Standard Model

Theoretical constraints on the Higgs boson mass

Comparing the prediction of the SM to the effective lagrangian describing the Fermi interaction, it is possible to link the coupling constant to the Fermi constant:

$$\text{From } \frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8 m_w^2} \text{ and } m_w = \frac{g_W \nu}{2} \Rightarrow \nu^2 = \frac{1}{\sqrt{2} G_F} \Rightarrow \nu \simeq 246\,\text{GeV} \tag{1.42}$$

The intensity of the couplings between the Higgs boson and the fermions (resp. vector bosons) are given by:

$$g_{Hff} = \frac{m_f}{\nu} = \left(\sqrt{2} G_F\right)^{1/2} m_f \tag{1.43}$$

$$g_{HVV} = \frac{2 M_V^2}{\nu} = 2 \left(\sqrt{2} G_F\right)^{1/2} M_V^2 \tag{1.44}$$

Study of the vector boson scattering gives constraints on the value of the Higgs boson mass [31]. The Feynman diagrams that describe the processes $VV \to VV$ must show at LO the presence of the Higgs boson exchange. In that case the unitarity of the scattering matrix impose $M_{\text{H}}^2/(8\pi\nu^2) < 1/2 \Rightarrow M_{\text{H}} \lesssim 870\,\text{GeV}$. In addition the renormalization group equations impose that the parameter λ that appears in the Higgs potential $V(\phi)$ evolves with the energy scale:

$$\lambda(Q^2) = \lambda(\nu^2) \left[1 - \frac{3}{4\pi^2} \lambda(\nu^2) log \frac{Q^2}{\nu^2}\right]^{-1} \tag{1.45}$$

From this equation emerges that a Landau pole at $\Lambda = \nu \cdot exp\left(\frac{2\pi^2\nu^2}{3M_{\text{H}}^2}\right)$ must exist and represents the limit of validity of the theory. The parameter Λ is defined as the value of the energy scale for which $\lambda(\Lambda) \to \infty$. Thus the Λ as function of the Higgs boson mass M_{H} defines a region in the $\Lambda - M_{\text{H}}$ plane where the SM is valid. This condition is named *triviality condition*. The

value of λ at the electroweak scale must be insensitive to the corrections coming from diagrams involving heavy fermions. This condition is named *vacuum stability* because if it is not fulfilled the minimum of the Higgs potential could assume non-finite values. In Fig. 1.7 is shown the region in the $\Lambda - M_{\mathrm{H}}$ plane where the SM is valid.

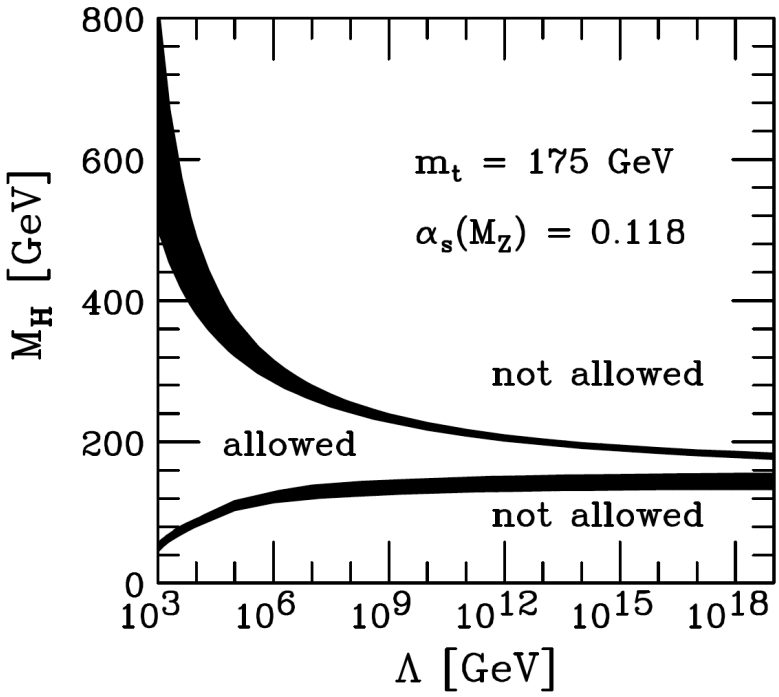


Figure 1.7: *Region of validity of the Standard Model in the* $\Lambda - M_{\mathrm{H}}$ *plane.* Λ *represents the energy scale beyond which the Standard Model is not longer valid.* M_{H} *represents the mass of the Higgs boson. The lower limit corresponds to the vacuum stability condition, while the upper limits is fixed by the triviality condition. These values are obtained for a top quark mass of* 175 ± 6 *GeV and a strong coupling constant equal to* $\alpha_S = 0.118 \pm 0.002$ *[32].*

Higgs boson decays

The behaviour of the total SM Higgs boson width Γ_{H} as a function of the Higgs boson mass M_{H} is illustrated in Fig. 1.8. Since the mass of the Higgs boson is measured $M_{\mathrm{H}} = 125.09 \pm 0.21(stat.) \pm 0.11(scale) \pm 0.01(theor.) \pm 0.02(other)\,\mathrm{GeV}/c^2$ [33], Γ_{H} is expected to be very small ($\Gamma_{\mathrm{H}} \sim 4\,\mathrm{MeV}/c^2$). Recently, constraints have been set on Γ_{H} by the CMS and ATLAS collaboration [34][35], using the off-shell production of the Higgs boson through gluon-gluon fusion and its decay to ZZ^* in the four-lepton, or two-lepton + two-neutrinos final state. The result of this analysis, performed over the whole CMS recorded luminosity, consists in an upper limit on the Higgs boson width of $\Gamma_{\mathrm{H}} < 22\,\mathrm{GeV}$ at the 95% confidence level [34].

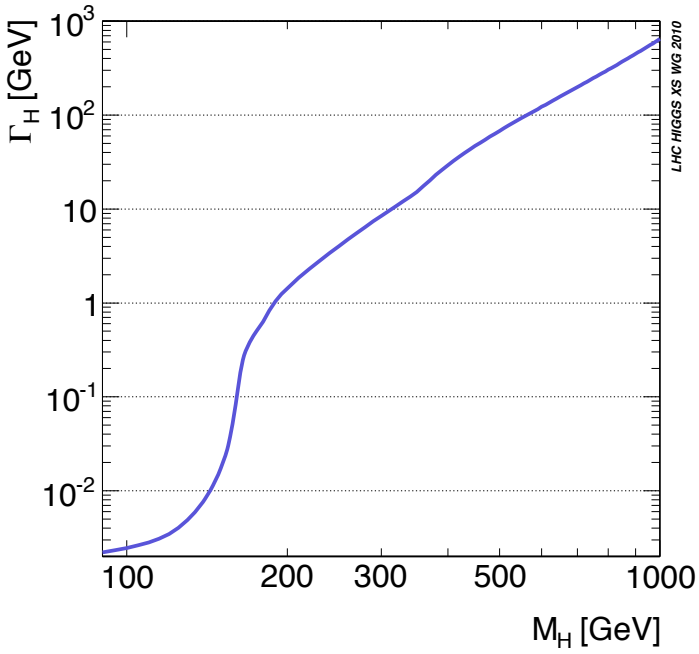


Figure 1.8: *Total SM Higgs boson width* $\Gamma_{\mathrm{H}} = \Gamma(\mathrm{H} \to anything)$ *as function of the Higgs boson mass* M_{H} *[36].*

The branching fraction of the channel $\mathrm{H} \to X_f$, where X_f denotes any of the possible final states, is given by:

$$BR(\mathrm{H} \to X_f) = \frac{\Gamma(\mathrm{H} \to X_f)}{\Gamma_{\mathrm{H}}} \tag{1.46}$$

The prediction of the SM Higgs boson decay branching ratios is one of the main topics that drive the choice of the major physics analyses performed at the LHC. In the following, the Higgs boson decays are analysed according to the nature of the final state, either in a pair of fermions or of bosons [31].

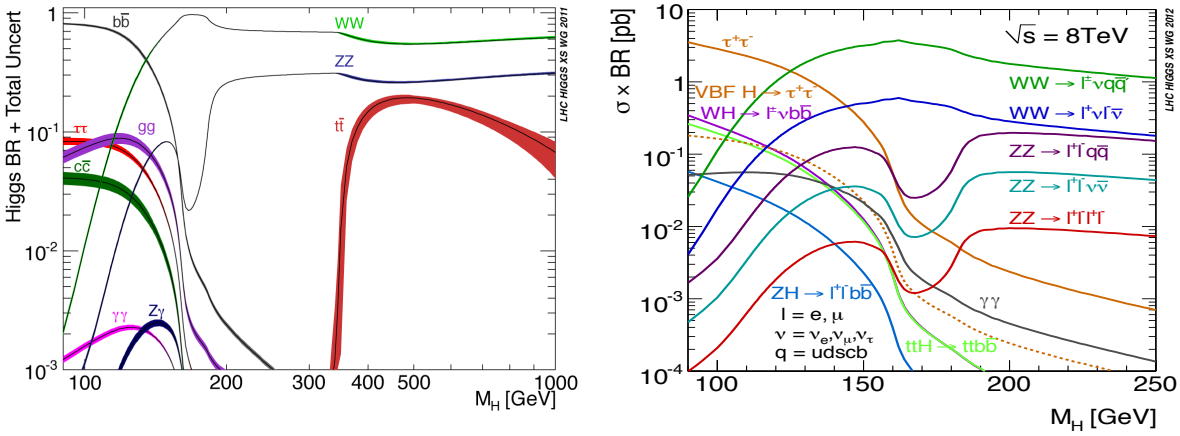


Figure 1.9: *Left: Branching ratios for Higgs boson decays; Right: Higgs cross section times branching ratios for pp collision at* $\sqrt{s} = 8$ *TeV.*

- **Decay into fermions**

  At Born level, the partial decay width of the SM Higgs boson into fermions is:

$$\Gamma_{Born}(\mathrm{H} \to f\bar{f}) = \frac{G_\mu N_C}{4\sqrt{2}\pi} M_\mathrm{H} m_f^2 \beta_f^3 \textbf{ with } \beta_f = \left(1 - \frac{4m_f^2}{M_\mathrm{H}^2}\right)^{1/2} \tag{1.47}$$

  where N_C is the number of colors (3 for the quarks and 1 for the leptons) and β_f is the fermion velocity evaluated in the Higgs boson frame. The decay width is proportional to m_f^2 at Leading Order (L.O.) for this reason, the decay of the Higgs boson into a di-τ or $b\bar{b}$ pair is largely enhanced respect to the $\mathrm{H} \to \gamma\gamma$ or other fermionics Higgs boson decays (for intermediate-low value of the Higgs boson mass, see Fig.1.9).

- **Decay into bosons**

  At Born level the partial decay width of the Higgs boson into a couple of massive vector boson (both of them on-shell) is represented by:

$$\Gamma_{Born}(\mathrm{H} \to VV) = \frac{G_\mu M_\mathrm{H}^3}{16\sqrt{2}\pi} \delta_V \sqrt{1-4x}(1 - 4x + 12x^2) \text{ where } x \equiv \frac{M_V^2}{M_\mathrm{H}^2} \tag{1.48}$$

  The decay width is thus proportional to the 3rd power of the Higgs boson mass.

- **bosonic Higgs decays through a loop**

  The Higgs boson does not couple to massless particles such as photons and gluons. Nevertheless, it is possible that processes like $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to \gamma Z$ and $\mathrm{H} \to gg$ proceed through loops. In that case, the presence of higher orders in the coupling constant has the effect

to suppress the partial decay width relative to these processes. In particular, in the case of $H \to \gamma\gamma$, the loop involves only charged particles:

$$\Gamma_{Born}(H \to \gamma\gamma) = \frac{G_\mu M_H^3 \alpha^2}{128\sqrt{2}\pi^3} |\sum_f N_C Q_f^2 A_{1/2}^H(\tau_f) + A_1^H(\tau_W)|^2 \quad (1.49)$$

The parameter τ_i is defined as $M_H^2/4M_i^2$ where M_i is the mass of the fermion or vector boson involved and the functions $A_{1/2}$ and A_1 are the form factors associated to fermions and bosons, respectively. The shapes of the form factors are shown in Fig. 1.10 for the W boson and the top quark. The Higgs boson decay into γZ proceeds through the same loop as $H \to \gamma\gamma$, but a factor proportional to $(1 - M_Z^2/M_H^2)$ appears in the partial decay width expression. Finally, the decay into gluons proceeds through quark loops (mainly involving the top quark with a weaker contribution of the b quark):

$$\Gamma_{Born}(H \to gg) = \frac{G_\mu M_H^3 \alpha_S^2}{36\sqrt{2}\pi^3} |\sum_Q A_{1/2}^H(\tau_Q)|^2 \quad (1.50)$$

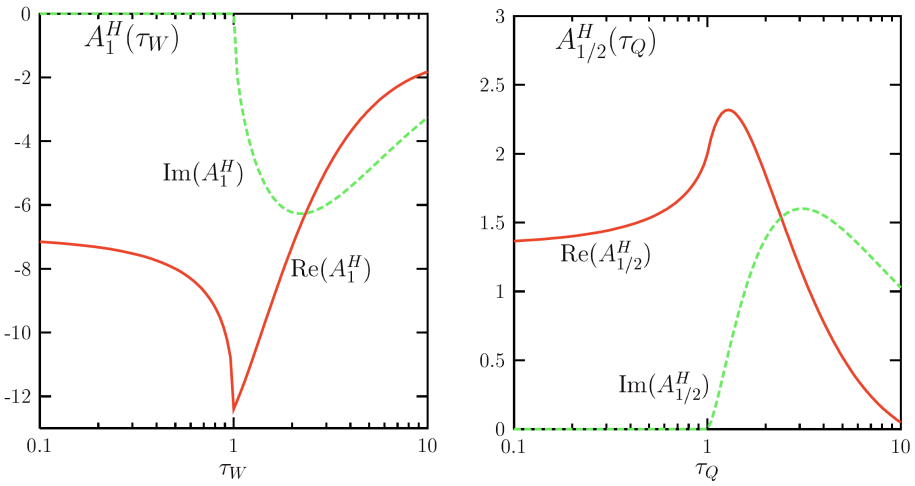


Figure 1.10: *Form factor $A_1^H(\tau_W)$ (Left) and $A_{1/2}^H(\tau_Q)$ (Right).*

1.4 Higgs boson searches before the LHC

1.4.1 LEP exclusion limits

The Higgs boson does not couple directly to photons, and the couplings to electrons, muons and light quarks are very small. Hence, although it is possible to produce resonantly Higgs bosons in reactions initiated by these particles, the production rates are very small. Nevertheless, considering heavier particles, such gauge bosons $W^\pm$, Z^0 or top quark, the coupling with Higgs

bosons becomes stronger. A search based on this concept has been performed at electron-positron collider (LEP II) accelerator at CERN. In order to obtain constraints on the mass of the SM Higgs boson, the analysis is performed exploiting the hadronic vacuum polarization, at the $Z - pole$ energy [5]. Through radiative corrections evaluated in the framework of the SM, the Z^0 resonance is also used to predict the mass of the top-quark, $m_t = 173^{+13}_{-10}\,\mathrm{GeV}/c^2$, while the mass of the W boson has been accurately measured: $m_W = 80.363 \pm 0.032\,\mathrm{GeV}/c^2$ [5]. These indirect constraints are compared to the direct measurements, providing a stringent test of the SM. In fact, LEP II machine reached the maximum energy of 206 GeV, and this energy is high enough to produce Higgs bosons with masses up to almost $120\,\mathrm{GeV}/c^2$ in the associated production mechanism (Higgsstrahlung, illustrated in Fig. 1.12 (c). This mechanism, at the LEP II energies, is the dominant one [5]. The data collected by the four LEP collaborations prior to year 2000 gave no direct indication of the production of the SM Higgs boson [37] and allowed a lower bound of $107.9\,\mathrm{GeV}/c^2$ to be set, at the 95% confidence level, on the mass. During the last year of the LEP programme (the year 2000), substantial data samples were collected at centre-of-mass energy exceeding 206 GeV, extending the search sensitivity to Higgs boson masses of about $115\,\mathrm{GeV}/c^2$.

1.4.2 Tevatron exclusion limits

Higgs boson searches have also been carried out at the Tevatron, which is a $p\bar{p}$ collider with energy of about $1\,TeV$ per beam, situated at the FNAL laboratory, near Chicago (USA). At the Tevatron, the searches concentrate on the associated production mechanism. Relationships between measurable EWK parameters within the SM, in conjunction with direct searches performed at LEP, constrained the Higgs boson mass to be between $114\,\mathrm{GeV}/c^2$ and $185\,\mathrm{GeV}/c^2$ at 95% CL. [38]. With enough data, this entire mass range is accessible to the 1.96 TeV center-of-mass Tevatron, for either observation or exclusion of the SM Higgs boson. The two multi-purpose detectors at the Tevatron, CDF and DØ, are able to reconstruct all the final state particles and topologies resulting from SM Higgs boson production and decay. The Tevatron has delivered $9 fb^{-1}$ of luminosity to CDF and DØ. Data collection efficiencies were $85 - 90\%$ for this data, and an integrated luminosity of up to 6.7 fb^{-1} has been analyzed for the Higgs boson searches.

- **Low Mass Higgs** Searches for a low mass ($m_{\mathrm{H}} < 135\,\mathrm{GeV}/c^2$) SM Higgs boson are performed for $\mathrm{H} \to b\bar{b}$, $\mathrm{H} \to \tau^+\tau^-$, and $\mathrm{H} \to \gamma\gamma$ [39]. In this mass region the stronger coupling of the Higgs is with the b-quarks. Given the high background event rate in $gg \to \mathrm{H} \to b\bar{b}$, the couple of $b\bar{b}$ is most easily identified in events produced via associated Higgs production, WH and ZH, when the W and Z decay leptonically, into final states of $WH \to \ell\nu b\bar{b}$, $ZH \to \nu\nu b\bar{b}$, and $ZH \to \ell\ell b\bar{b}$ where $\ell = e$ or μ. In events with a reconstructed W or Z boson and two or more additional jets, the di-jet invariant mass is used to search for a resonance originating from $Z\mathrm{H}(\mathrm{H} \to b\bar{b})$ and $W\mathrm{H}(\mathrm{H} \to b\bar{b}$ [39]

- **High Mass Higgs** The most sensitive channel in this mass range at the Tevatron is

$gg \rightarrow \mathrm{H} \rightarrow WW \rightarrow \ell\nu\ell\nu$, due to the high cross section and well identified final state. The high mass analysis benefits from separating events into categories according to the number of jets and the number of leptons in the final state, because of the difference in the topology and kinematics of the signal production mechanisms and the main background processes [39]. The exclusion limit as reported by the CDF and DØ collaborations at the startup of the LHC is shown in Fig. 1.11.

The results of LEP and Tevatron were of fundamental importance to constrain the mass range of the Higgs boson. The discovery performed at the LHC shows the complete compatibility between all experiments (even if two of them provide only an exclusion limit).

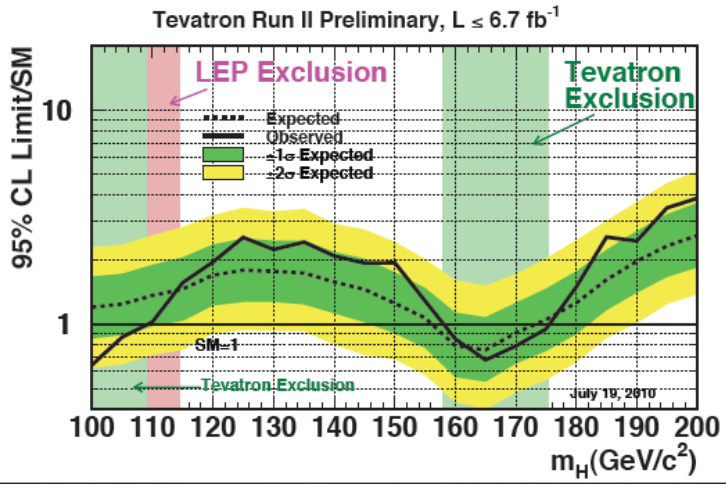


Figure 1.11: *Observed and expected 95% C.L. upper limits on the ratios to the Standard Model cross section, as functions of the Higgs boson mass for the combined CDF and DØ analyses (2010) [40]. The limits are expressed as a multiple of the SM prediction for test masses (every 5 GeV/c^2). The black line (dashed line) represents the observed (expected) limit. [40] The green and yellow bands represent the incertitude on the expected limit: 1 sigma and 2 sigma respectively. The mass interval excluded by LEP and Tevatron are represented by the vertical green and red bands.*

1.5 Higgs boson searches at the LHC

1.5.1 Higgs boson production

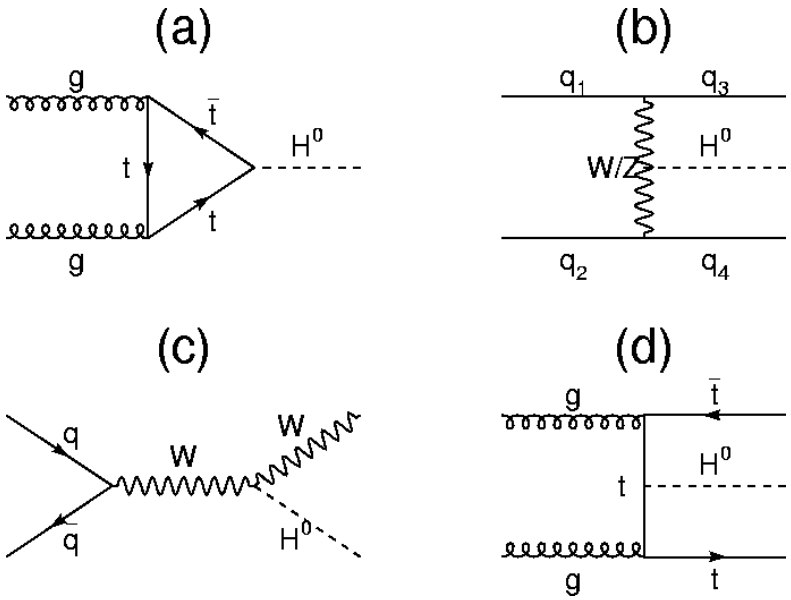


Figure 1.12: *(a) Gluon-gluon fusion; (b) Vector Boson Fusion (VBF); (c) higgs-strahlung; (d) production in association with a heavy quark-antiquark pair.*

The coupling between the Higgs boson and the fermions (resp. bosons) is proportional to the mass of the fermion (resp. squared mass of the boson). Thus the most important production modes involves heavy particles like the vector bosons $W^{\pm}$, Z^0 and the top quark. In what follows, the four main Higgs boson production processes are explained in more details:

- gluon-gluon fusion through a heavy quark loop : $gg \rightarrow \mathrm{H}$;
- vector boson fusion (VBF): $q_1 q_2 \rightarrow V^* V^* \rightarrow q_1^{'} q_2^{'} + \mathrm{H}$
- associate production of the Higgs boson with a massive boson (*Higgsstrahlung*): $q\bar{q} \rightarrow V^* \rightarrow V + \mathrm{H}$
- associate production of the Higgs boson with a pair of top quarks: $gg \rightarrow t\bar{t} + \mathrm{H}$

The Higgs boson production modes that dominate at the LHC energies are shown in Fig. 1.12. Despite a smaller (factor 1/10) production cross section with respect the gluon-gluon process, the VBF mechanism can be very interesting because of its distinctive topology (more details on the VBF caracteristic are given in 1.5.1). There exist other minor production modes, like the double Higgs bosons production.

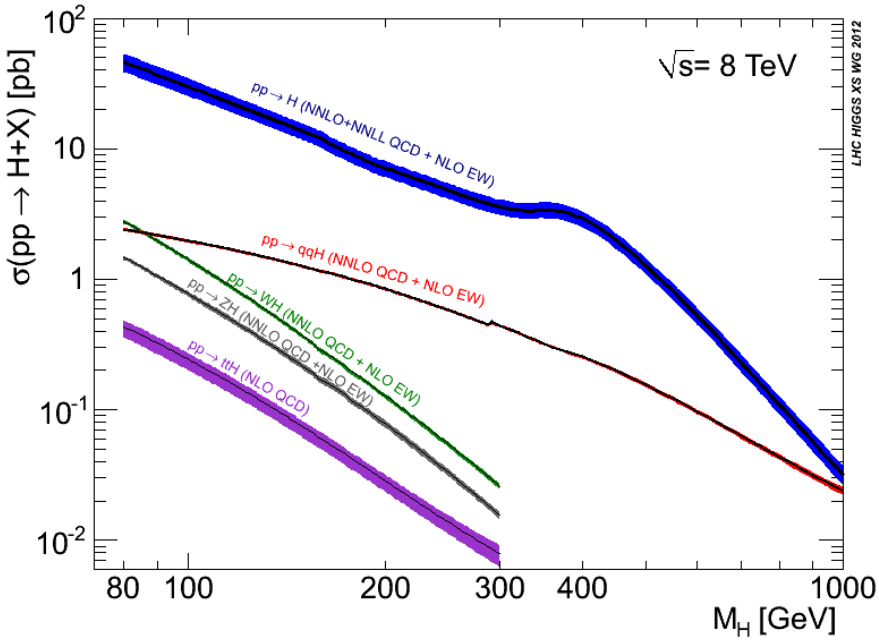


Figure 1.13: *Higgs boson cross section production as a function of the Higgs boson mass for* $\sqrt{(s)} = 8\,TeV$ *[36].*

Gluon-gluon fusion

For $m_H = 125\,\mathrm{GeV}/c^2$, the dominant production mechanism at the LHC consists in the gluon-gluon fusion process $gg \to \mathrm{H}$ (see Fig. 1.13). The cross section of this particular Higgs boson production mechanism is enhanced by the high Yukawa coupling between the Higgs boson and the heavy quarks present in the loop and by the gluon distribution inside the protons. At leading order (LO) the gluon-gluon fusion is a QCD process with a cross section proportional to α_S^2 [41]

$$\hat{\sigma}_{LO}\left(gg \to \mathrm{H}\right) = \sigma_0^{\mathrm{H}} M_{\mathrm{H}}^2 \delta(\hat{s} - M_{\mathrm{H}}^2) \tag{1.51}$$

with:

$$\sigma_0^{\mathrm{H}} = \frac{G_\mu \alpha_S^2\left(M_{\mathrm{H}}^2\right)}{288\pi\sqrt{2}} |\frac{3}{4}\sum_q A_{1/2}^{\mathrm{H}}\left(\tau_q\right)|^2 \tag{1.52}$$

where the function $A_{1/2}^{\mathrm{H}}\left(\tau_q\right)$ is the form factor reported in Fig. 1.10 with $\tau_q = M_{\mathrm{H}}^2/4m_q^2$. The variable $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$ corresponds to the center-of-mass energy of the interacting partons. The Dirac delta $\delta(\hat{s} - M_{\mathrm{H}}^2)$ represents the energy conservation between the incoming gluons and the produced Higgs boson.

As visible in Fig.1.13, the Higgs boson production cross section through gluon-gluon fusion decreases rapidly with the Higgs boson mass because of the partonic distribution function for the gluon which decreases hyperbolically for increasing Bjorken fraction, x. The cross section is computed at the next-to-next-to-leading order (NNLO) accuracy in QCD, using the approxima-

tion for the infinite top quark mass [42][43][44]. The calculation of the production cross section takes the QCD radiative corrections at higher orders in the perturbation expansion into account: the next-to-leading order (NLO) and NNLO evaluated with a re-summation of the logarithmic divergences appearing with the soft-gluons emissions (NNLL) [45]. Additional improvements in the calculation include NLO EWK corrections [46][47]. While the QCD radiative corrections are independent on the mass of the Higgs boson (they mostly depend on the mass of the heavy quarks), the EWK ones strongly depend on the Higgs boson mass. The major source of theoretical error in the $gg \rightarrow \mathrm{H}$ production cross section computation arises from the QCD correction terms not included in the calculation itself and for the missing of the higher orders terms. The level of incertitude corresponds to a value of $\sim 10\%$ for $100\,\mathrm{GeV}/c^2 < M_{\mathrm{H}}^2 < 300\,\mathrm{GeV}/c^2$ and $\sim 7\%$ for $M_{\mathrm{H}} > 300\,\mathrm{GeV}/c^2$ [48].

Vector Bosons Fusion

For very large Higgs masses ($\sim 800\,\mathrm{GeV}/c^2$ Fig. 1.13), the VBF $q_1 q_2 \rightarrow V^* V^* \rightarrow q_1^{'} q_2^{'} + \mathrm{H}$ cross section and the $gg \rightarrow \mathrm{H}$ one have the same order of magnitude. For this reason, first calculations of the VBF Higgs boson production cross section have been carried out in the context of searches for a very massive Higgs boson [49]. The VBF process, illustrated in Fig. 1.12, involves the radiation of a heavy vector boson from each incoming parton. Subsequently, the two vector bosons "fuse" to produce a Higgs boson. The partonic cross section evaluated at LO, assuming that the incoming quarks give a small fraction of their energy to the vector bosons, is:

$$\hat{\sigma}_{LO}\left(\hat{s}\right) = \frac{1}{16 M_W^2}\left(\frac{\alpha}{sin^2\theta_W}\right)^3 \left[\left(1 + \frac{M_{\mathrm{H}}^2}{\hat{s}}\right) log\left(\frac{\hat{s}}{M_{\mathrm{H}}^2}\right) - 2 + 2\frac{M_{\mathrm{H}}^2}{\hat{s}}\right] \tag{1.53}$$

The VBF cross section increases as the energy in the center-of-mass increases thanks to the contribution from the longitudinally polarized W's. The behaviour of the cross section as function of the Higgs boson mass, on the contrary, decreases as the Higgs boson mass achieves higher values. This feature is, however, less pronounced than in the gluon-gluon-gluon fusion process. The VBF production is also of major interest in the region of medium Higgs boson masses because of its relative high cross section (second only to the gluons fusion production one) and of the special signatures it can provide for the identification of the Higgs boson. Indeed, the VBF mechanism presents a very particular topology that can be used as handle to optimize the background rejection. A very important role is taken by the quarks. After the radiation of an heavy boson, the quarks participating to the interaction scatter with a small angle with respect to the incident direction of the colliding beams. This fact translates into the presence of a forward/backward jet pair. As a result, technique exploiting the forward/backward jet tagging is used to drastically suppress the background. Another important feature typical of the VBF Higgs boson produced events, is the lack of color exchange between the initial state quarks. This color coherence between initial and final state quarks leads to a suppression of the hadronic activity in the central rapidity region and it opens the possibility to use Central Jet

Veto to select the signal events. In Fig. 1.14 the distributions of the invariant mass of the two VBF quarks as well as their pseudorapidity are shown.

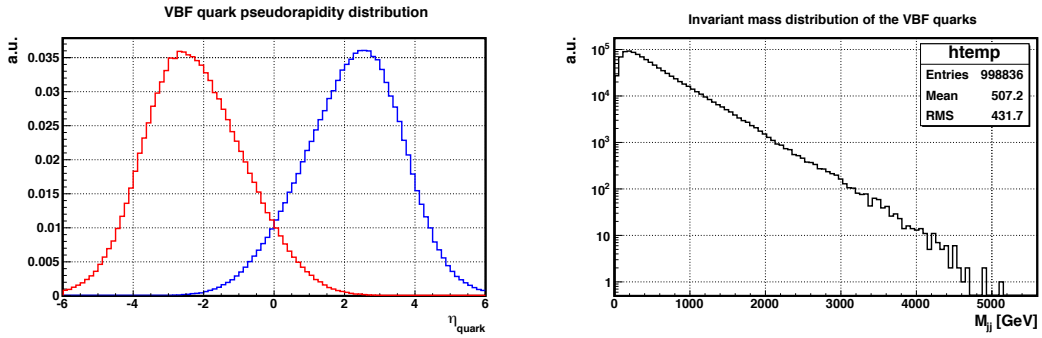


Figure 1.14: *Distribution of the pseudorapidity of the two outgoing VBF quark (Left) and the distribution of their invariant mass (Right) obtained from MC simulated VBF $H \to \tau\tau$ events.*

The cross section in Eq. 1.53 is evaluated at first order in the EWK coupling constant expansion. In order to increase the precision in the prediction of the VBF cross section, QCD and EWK corrections have been considered. The NLO corrections in QCD and EWK are applied to the inclusive VBF cross section, leading to modifications between 5% and 10%. Finally, QCD NNLO corrections reduce the uncertainty coming from the renormalization and factorization scale to below 1% [50]. However, from an experimental point of view, the search for the Higgs boson produced through a VBF mechanism is restricted to a subset of the phace space. In particular, it is possible to imagine that the application of cuts based on the transverse momenta of the forward-tagging jets could be subject to important NNLO corrections that are otherwise non-visible in the fully inclusive cross section [51]. Recently, an important achievement was reached in the understanding of the NNLO correction in the VBF fully differential cross section. This was possible thanks to a new technique called *projection-to-Born* that combines an inclusive NNLO calculation (using the structure function method) and, an exclusive NLO order calculation based on the VBF Higgs + 3 Jet process. The results achieved in term of differential cross section as function of the transverse momenta of the two forward-tagging jets and the rapidity difference between the same jets can be seen in Fig. 1.15.

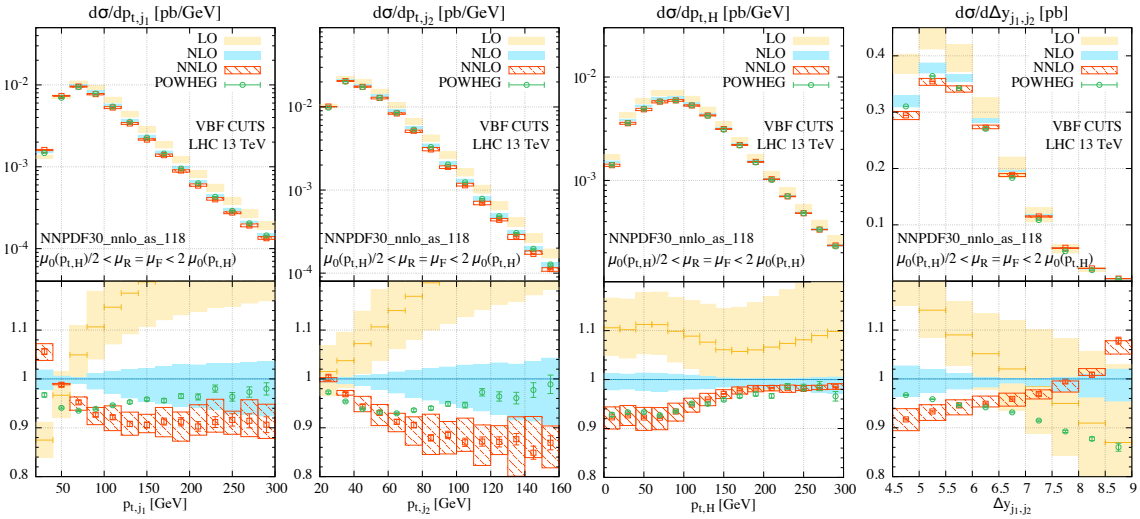


Figure 1.15: *Differential cross section distribution for the VBF higgs boson production evaluated at the NNLO as function of the variables:* $p_{T,j_1}, p_{T,j_2}, p_{T,\mathrm{H}}, \Delta\eta_{j_1,j_2}$ *[51].*

With these recent results, the VBF Higgs production benefits from a precision on the differential cross-section similar to that achieved for the other production modes [51].

Associate production with a heavy vector boson

This production mode is based on the annihilation of a quark couple into a virtual vector boson (off-shell) and the subsequent emission of an Higgs boson and of a real vector boson. The partonic cross section is thus factorizable and it can be written as:

$$\hat{\sigma}\left(\hat{s}\right) = \int_0^{\hat{s}} dp_{V^*}^2 \sigma\left(p_{V^2}^2\right) \frac{d\Gamma\left(V^* \to VH\right)}{dp_{V^*}^2} \tag{1.54}$$

This production mode proceeds through the s-channel. The energy in the centre of mass of the interacting partons must fulfil $\sqrt{\hat{s}} \geq M_V + M_{\mathrm{H}}$. As a consequence, the cross section as function of M_{H} decreases much more rapidly than in the gluon-gluon fusion or VBF cases. Despite the smaller cross section compared to other production mechanism, the *Higgsstrahlung* can be exploited in the searches for the Higgs boson thanks to its clear signature. Indeed, the vector boson present in the final state can decay into leptons that are reconstructed very efficiently.

Associate production with a heavy quarks pair

The measure of the production cross section of an Higgs boson in association to a couple of heavy quarks (mainly with the top quark) can represent en excellent test of the Yukawa couplings. At LO this mechanism proceeds through a quark-antiquark annihilation into a couple of top-antitop quarks (or heavy quark-antiquark pair), where the Higgs boson is radiated from a top quark in

the final state. As the energy in the center-of-mass increases, the gluon-gluon fusion mechanism (through a t-channel) becomes also significant.

1.5.2 The Higgs boson discovery

The Higgs boson production cross section as a function of the hypothesized SM Higgs boson mass for a fixed energy in the center-of-mass of the collisions equal to $8\,TeV$ is illustrated in Fig. 1.13. At high mass, the relative fraction of the vector boson fusion mechanism is increasing. Before the discovery of the Higgs boson, a wide range of possible decay modes has been taken into consideration, since the decay modes of the SM Higgs boson strongly depend on its mass m_{H}. For a low mass Higgs boson ($110\,\mathrm{GeV}/c^2 < m_{\mathrm{H}} < 150\,\mathrm{GeV}/c^2$) its natural width is only few MeV/c^2. In this context, the main channels that play an important role are $\mathrm{H} \to ZZ^* \to 4\ell$, $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to WW^* \to 2\ell 2\nu$, $\mathrm{H} \to b\bar{b}$ and $\mathrm{H} \to \tau^+\tau^-$. For a given hypothesis for a Higgs boson mass, the sensitivity of the search depends on the Higgs boson production cross section (reported in Fig. 1.13), its decay branching ratios into the chosen final state, the signal selection efficiencies, the reconstructed mass resolution, and the level of SM backgrounds in the final state [12]. The $\mathrm{H} \to ZZ \to 4\ell$ and $\mathrm{H} \to \gamma\gamma$ channels, providing an excellent mass resolution for the reconstructed 4-lepton and di-photon final states, played a crucial role in the Higgs boson discovery. Also $\mathrm{H} \to WW \to 2\ell 2\nu$ is characterized by a high sensitivity [12], but also by a poor mass resolution due to the presence of neutrinos in the final state. Finally, the $b\bar{b}$ and $\tau^+\tau^-$ channels provide a lower significance than the channels mentioned previously because of the presence of large backgrounds and a worse energy resolution [12]. The results presented in July 2012 were obtained with data collected by CMS corresponding to an integrated luminosity of $5.1\,fb^{-1}$ at $7\,TeV$ and $5.3\,fb^{-1}$ at $8\,TeV$. The individual results for the five decay modes mentioned above are combined. The combination assumes the relative branching fractions predicted by the SM and takes into account the experimental statistical and systematic uncertainties as well as the theoretical ones. The 95% confidence level exclusion limits and the corresponding p-value obtained in the final combination are illustrated in two historical plots in Fig. 1.16.

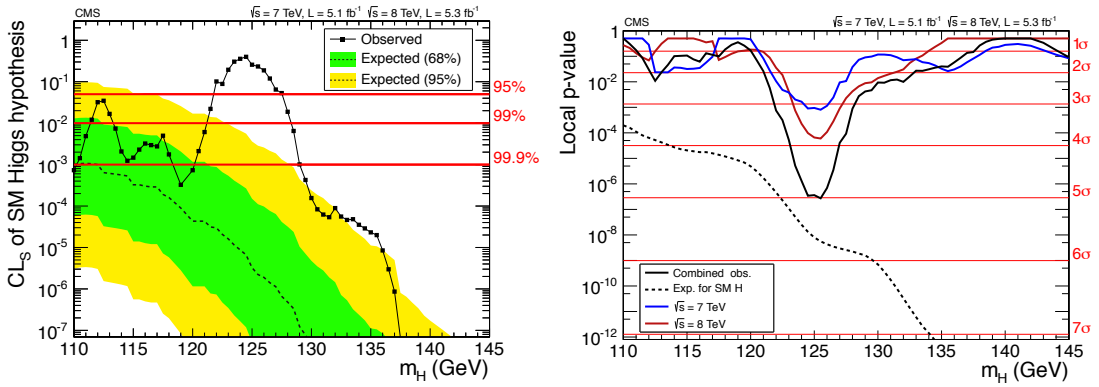


Figure 1.16: *The expected and observed 95% CL exclusion limits for the SM Higgs boson hypothesis as a function of the Higgs boson mass (m_{H}) in the range* $[110, 145]$ *GeV/c^2, obtained combining all the decay channel analysed (Left). The expected (black dashed line) and observed (continuous lines) local p-value as a function of the Higgs boson mass (m_{H}). The observed p-value are shown for the* $7\,TeV$ *data (red),* $8\,TeV$ *data (blue) and the their combination (black). These plots have been officially shown on the July* 4^{th}*, 2012. [12]*

Even if the discovery of the Higgs boson has been announced, the CMS experiment continued to record pp collision data until the end of the LHC Run 1, in the beginning of 2013. The total integrated luminosity collected by CMS amounts to $\sim 20\,fb^{-1}$. In what follows, a general overview concerning the state of the art of the analyses carried out by the CMS collaboration in the five main Higgs boson decay channels is provided.

1.5.3 $\mathrm{H} \to ZZ^*$

- $\mathrm{H} \to ZZ^{(*)} \to 4\ell$ [52]: In this *golden* channel, a search has been performed for a four-lepton mass peak over a small continuum background. To further distinguish signal from background, a particular kinematic discriminant based on Matrix-Elements has been developed. It is used to disentangle the signal from the main ZZ continuum background. Five observables are sufficient to describe the kinematics of the decay: the masses of di-lepton pairs and fives angles that completely define their configuration in the center-of-mass frame. The $4e$, 4μ, and the $2e2\mu$ sub-channel are analysed separately since the 4ℓ-mass resolution for each sub-channel is different [12].

- $\mathrm{H} \to ZZ \to 2\ell 2\nu$ [53]: the events selection consists in requiring events with a lepton pair (ee, $\mu\mu$), with invariant mass consistent with that of the Z boson, and large missing transverse energy.

- $\mathrm{H} \to ZZ^{(*)} \to 2\ell 2q$ [54]: The basic request for selecting signal events is the presence of 2ℓ and two jets. The jets are required to form an invariant mass consistent with that of

the Z boson. The purpose is to search for a peak in the invariant mass distribution of the di-lepton + di-jet system. The background is estimated using data-driven techniques.

- $H \to ZZ \to 2\ell 2\tau$ [52]: One of the two Z bosons is required to be on-shell and to decay to a lepton pair (ee or $\mu\mu$). The other Z boson is required to decay through a $\tau\tau$ pair to one of the four final state $e\mu$, $\mu\mu$, $e\tau_h$, $\mu\tau_h$. In this channel analysis, the goal is to search for an excess in the distribution of the di-lepton + di-tau mass. The dominant background originates from non-resonant ZZ production [12]

In order to further discriminate the Higgs boson signal lfrom the dominant non-resonant ZZ^* background, CMS implied a likelihood approach based on the matrix element method [55]. The CMS experiment observe an excess at $m_H = 125.8\,\mathrm{GeV}/c^2$ with a observed (expected) significance of 6.7σ (7.2σ) [56]. The four-lepton invariant mass distribution obtained only for the 4ℓ final state is illustrated in Fig. 1.17.

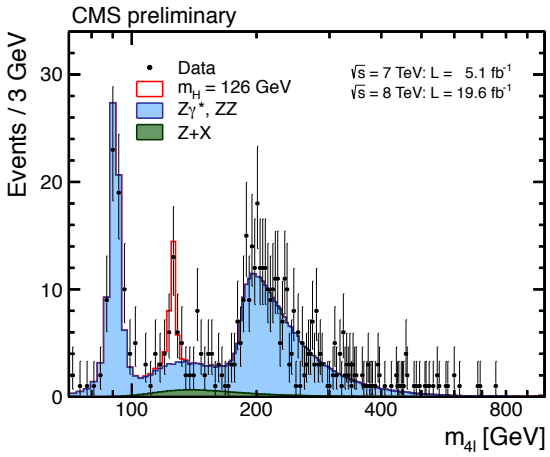


Figure 1.17: *Distribution of the four-lepton reconstructed mass in the full mass range for the sum of the $4e$, 4μ, and $2e2\mu$ channels. Points represent the data, shaded histograms represent the background and the unshaded histogram the signal expectation. The expected distributions are presented as stacked histograms. The measurements are presented for the sum of the data collected at $\sqrt{s} = 7\,TeV$ and $\sqrt{s} = 8\,TeV$ [56].*

1.5.4 $H \to \gamma\gamma$

The $H \to \gamma\gamma$ analysis [57] is focused on the search for a narrow peak in the invariant mass distribution of two high p_T photons. The main background in this channel originates from prompt $\gamma\gamma$, γ+jet and di-jet processes. In order to increase the sensitivity of the analysis, the event sample is split into several mutually exclusive categories [58]:

- di-γ events with high p_T leptons, di-jet, missing transverse energy compatible with the W or Z bosons decay enters in the VH production category;

- di-γ events with two energetic jets separated by a high pseudorapidity gap enter in the VBF production category;
- all the remaining events ($\sim 99\%$ of the total) are grouped in the gluon-gluon production category.

The events remaining untagged are further subdivided into classes based on the output of a multivariate (MVA) discriminant that assigns a high score to signal-like events. The MVA is trained using the following discriminating variables:

1. an event-by-event estimate of the di-photon mass resolution;
2. a photon identification score for each photon;
3. kinematic information about the photons and the di-photon system

The photon identification also follows a multivariate approach that uses shower information in order to discriminate prompt photons from those coming from a jet. The functional form of the background has been estimated through a fit to the full di-photon invariant mass distribution ($m_{\gamma\gamma}$) in each category. The signal extraction has been performed through a simultaneous fit to $m_{\gamma\gamma}$ in all the categories. The CMS experiment observed an excess at $m_{\mathrm{H}} = 124.70\,\mathrm{GeV}/c^2$ with an observed (expected) significance of 5.7σ (5.2σ) [59]. The distribution of $m_{\gamma\gamma}$ for the combination of all the categories is illustrated in Fig. 1.18.

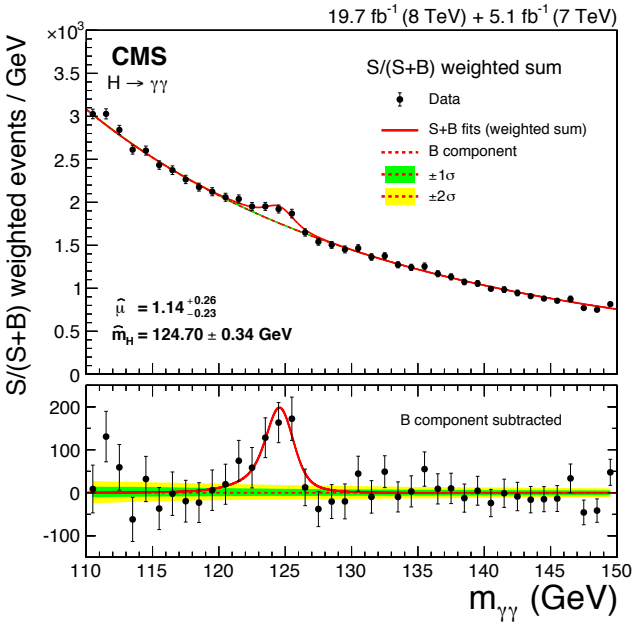


Figure 1.18: *The diphoton mass spectrum weighted by the ratio S/(S + B) in each event class, together with the background subtracted weighted mass spectrum. The black dots represent the data, the red dashed line represent the background and the continuous line represent the fit performed in the signal + background hypothesis. The measurements are presented for the sum of the data collected at* $\sqrt{s} = 7\,TeV$ *and* $\sqrt{s} = 8\,TeV$ *[59].*

1.5.5 $H \to WW^*$

This channel offers the possibility to study directly the Higgs boson couplings to the W vector bosons. The study of the angular distributions are good handles to measure the spin and CP quantum numbers of the Higgs boson. Even if the branching fraction for this SM Higgs boson decay channel is large, due to the large mass of the W vector boson, the search for the $H \to W^+W^-$ is complicated by the presence of neutrinos in the W decay products. For this reason the mass resolution is quite poor, and the search is reduced to a counting experiment of the event yield in broad bins of m_H

- $H \to WW^* \to \ell\nu\ell\nu$ [60]: This analysis searches for an excess of events with two leptons of opposite charge, large $\not{E}_T$ and up to two jets. For events without jets, the main background arises from non-resonant WW production; for events with 1 jet, the dominant background comes from WW and top-quark production. The 2-jets category is optimized to exploit the clearer signature of VBF production and the main background for this channel originates from the top quark production. All background rates are estimated using data-driven techniques, with the exception of the small contributions from WZ, ZZ, and $W\gamma$.

- H $\rightarrow WW^* \rightarrow \ell\nu 2q$ [61] [62]: This analysis searches for an excess of events with one electron (or muon), $\not\!\!E_T$ and two or three jets. The dominant background is from $W + jets$ events and has been estimated from data. Because only one neutrino is produced in this channel, both W bosons can be fully reconstructed, and a four-body mass peak can be reconstructed for WW and WZ pairs. A kinematic fit is performed to improve the resolution. In this fit procedure, the $\not\!\!E_T$ of the lepton is constrained to the W on-shell mass [12].

- $WH \rightarrow WWW^* \rightarrow 3\ell 3\nu$[63]: In this channel, a search for an excess of events with three leptons, electrons or muons, large missing transverse energy, and low hadronic activity, is performed. The main background is from $WZ \rightarrow 3\ell 3\nu$ production. This can be reduced requiring that all same-flavour oppositely charged lepton pairs have a di-lepton mass away from m_Z [12], and oppositely charged leptons are required not to be back-to-back.

- $VH \rightarrow WW \rightarrow 2\ell 2\nu 2q$: The research for this type of events produced in association with a vector boson (V) decaying into a couple of jets [64] is carried out by selecting events with two oppositely charged leptons, large $\not\!\!E_T$ and two jets with an invariant mass around the W/Z pole. The most important backgrounds are the top quark and $Z + Jets$ production [12].

Thanks to the spin quantum number conservation in the H $\rightarrow WW^*$ decays, the angular correlation between the two leptons direction and the direction of the missing transverse energy can be exploited as handle to improve the signal to background separation. In addition, different categories are defined in function of the number of reconstructed jets in the event in order to further increase the sensitivity of the analysis: 0-jet, 1-jets and 2-jets. In particular, the analysis performed by the CMS collaboration in the 0-jet and 1-jet categories, using all flavour leptons combinations, brought evidence of an excess at $m_{\mathrm{H}} = 125.4\,\mathrm{GeV}/c^2$ with an observed (expected) significance of 4σ (5.2σ) [60]. The transverse mass distribution m_T for data, simulated background and simulated signal events is illustrated in Fig. 1.19.

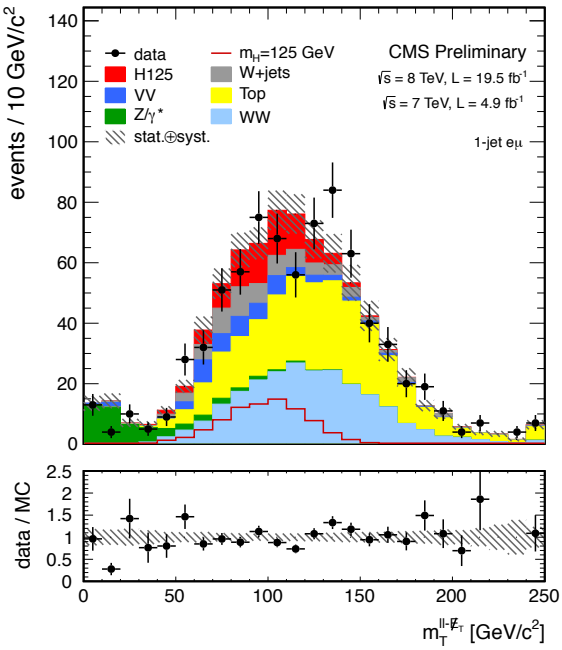


Figure 1.19: *The transverse mass (m_T) distribution in the 1-jet category summed over all lepton flavours for all the simulated SM background and Higgs boson signal with $m_H = 125\ GeV/c^2$, as well as for the data (black dots). The bottom plot shows the agreement between the data and the simulation [60].*

1.5.6 $H \to b\bar{b}$

This channel is the only one that gives direct access to the Higgs boson coupling to down-type quarks. The dominant Higgs boson production mode $gg \to H$ and its decay into a $b\bar{b}$ pair is overwhelmed by the inclusive production of $p\bar{p} \to b\bar{b} + X$ via the strong interaction. The Higgs boson associate production allows use of the leptonic decays of the vector bosons to reject the QCD background and, thus, purify the signal. For this reason, the $H \to b\bar{b}$ search concentrates on Higgs boson production in association with a W or a Z boson. Attention is put, in particular, on the following decay modes: $W \to e\nu$ and $Z \to ee/\mu\mu/\nu\nu$. In order to improve the sensitivity, the events are categorized in function of the p_T of the Higgs boson candidate(low- and high-boost topologies). The presence of missing transverse energy and/or leptons in the final state is used to select the events. In particular, $Z \to \nu\nu$ is identified by requiring large $\not{E}_T$. The Higgs boson candidate is reconstructed by requiring two b-tagged jets [65] and the search is divided into events where the vector bosons have medium or large transverse momentum and recoil away from the candidate Higgs boson. Events with higher transverse momentum bosons have smaller backgrounds and a better di-jet mass resolution. A multivariate analysis technique, based on Boosted Decision Trees (BDTs), trained on simulated signal and background events for several different values of the Higgs boson mass, is used to separate signal and background events. The main background consists of $W/Z + jets$ and top-quark events. Results are obtained through a

binned likelihood fit to the shapes of the BDT discriminants. An example of such a distribution is given in Fig. 1.20, where the distribution of the BDT output is shown in the case of the $Z \to \mu\mu$ decay channel. The CMS experiment observed an excess of events above the expected background with a significance of 2.6σ [65].

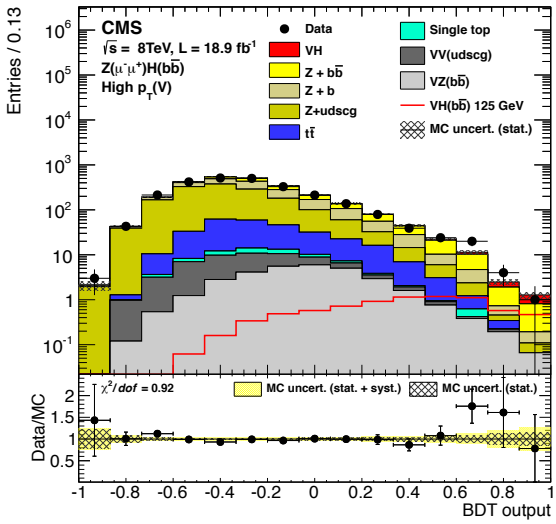


Figure 1.20: *Post-fit BDT output distribution for $Z \to \mu\mu$ events in the high-boost region for $\sqrt{s} = 8\,TeV$ data [65].*

1.5.7 $H \to \tau^+\tau^-$

The results provided by the ATLAS and CMS experiments strongly suggest that there is at least one isospin doublet Higgs field. Even if this fact can be explained in the context of the SM, there still a number of open questions regarding the Higgs field. In particular, the structure of the Higgs sector is not completely understood. The recent observation of the coupling between the Higgs boson and the τ leptons [66], obtained through the combination of the ATLAS and CMS results in the $H \to \tau\tau$ searches, constitutes another step forward in the understanding of the nature of the Higgs boson. However, at the present time, there are not independent observations by either of the two experiments. Thus, the $H \to \tau\tau$ search will continue to play a leading role in the future.

Once observed that the Higgs boson decays into fermions, precise measurements of the Higgs-fermions couplings must be carried out in order to verify that they are proportional to the fermion masses, as predicted by the SM. An hint in this direction is provided by the fact that no excesses have been observed in the $H \to \mu\mu$ and $H \to ee$ decay channels [67]. Indeed, any deviation of the Higgs-fermions couplings from the values predicted by the SM could represent an hint of new physics. In addition, a complete characterization of the tensor structure of the

trilinear coupling [68][69] still to be performed. In particular, among the various analyses that can have access to the spin-parity properties of the Higgs boson, there is the possibility to probe these quantum numbers via the jet kinematics in the vector boson fusion Higgs production processes [69]. Also in this case the H $\rightarrow \tau\tau$ will play a central role, since the VBF categories largely drive the H $\rightarrow \tau\tau$ analysis sensitivity. For the reasons mentioned above, the H $\rightarrow$ decay channel represents a major testbed to investigate the properties of the Higgs sector.

A Higgs boson with a mass of 125 GeV/c^2 decays almost entirely to $b\bar{b}$ pairs (see Fig. 1.9) and, secondly, to $\tau^+\tau^-$ pairs. Much of the phenomenology of the low mass Higgs boson search is focused on identifying b quarks and τ lepton. Moreover, an advantage of the H $\rightarrow \tau^+\tau^-$ channel, in particular if compared to the dominant H $\rightarrow b\bar{b}$ mode, is the lower background from QCD processes. In fact, the H $\rightarrow \tau^+\tau^-$ channel offers the best prospects for a direct measurement of the Higgs boson's couplings to fermions [70], and the fact that an excess have been seen in this channel [71], increases the importance of the results coming from the analyses on this decay mode. As already pointed out in the previous subsection, the most important Higgs boson production modes are $gg \rightarrow$ H, $qq \rightarrow qqVV \rightarrow qqH$, and VH. The first one has the largest cross section by almost one order of magnitude, there are substantial QCD backgrounds but few handles to distinguish them from the signal. Events coming from VBF produced Higgs contain additional information in the observable jets. In fact these jets are separated by a big pseudorapidity gap [72]. Techniques like forward jet tagging can then be exploited to reduce the backgrounds. Another feature of the VBF signal is the lack of color exchange between the initial-state quarks. Color coherence between initial- and final-state gluon bremsstrahlung leads to suppressed hadron production in the central region, between the two tagging-jet candidates of the signal [73]. This is in contrast with most of background process that contain large hadronic activity in the central region.

Also topics beyond the SM could be investigated with τ probe. In fact some attention has been given to $A/\mathrm{H} \rightarrow \tau^+\tau^-$ searches at the LHC in the framework of the Minimal Supersymmetric Standard Model (MSSM), where the increased couplings of A/H to τ predicted for $\tan\beta > 1$ lead to higher production rates. However, the great variety of τ decay modes represent, surely, a challenge for the correct reconstruction of the τ object in the CMS expertiment but at the same time allow a good compromise between trigger efficiencies and background rejection since also leptonic decay mode are allowed. For all these reasons the analysis of the H $\rightarrow \tau^+\tau^-$ decay channel is one of the most promising to determine the Higgs properties and couplings and either to discover if the new particle seen by LHC experiments [12] is a SM particle or a MSSM one.

In this context τ object characterization and identification plays a key rule and the improving of this represent a very challenge. Moreover, τ is the only lepton that could decay into hadrons because of its mass. The branching ratio for the $\tau \rightarrow$ hadrons is 64% [74] (for more

details see Tab. 3.1) and overcome that for all the other decay mode. Indeed, a study of the H $\rightarrow \tau\tau \rightarrow \tau_h\tau_h$ decay mode, in addition to the other di-τ final states initially exploited by CMS in the $H \rightarrow \tau\tau$ search, i.e. ee, $\mu\mu$, $e\mu$, $e\tau_h$ and $\mu\tau_h$, further increased the sensibility in the H $\rightarrow \tau\tau$ analysis. At the same time, it is a challenging path to follow by an experimental point of view because of both the variety of the different hadronic decay modes and the considerable QCD background activity present at LHC that could affect τ identification (see section 3.4.1).

Finally, the fact that the τ is the only lepton decaying into hadrons (as already pointed out) could be exploited in order to discriminate Higgs boson signal events from those coming from the main background ($Z \rightarrow \tau\tau$) using the τ polarization [75]. Moreover, this information may be used to identify charged and neutral Higgs bosons and can also provide information about the spin of the new particle [75]. In order to reach this goals the τ hadronic decay modes play a key role. In fact the decay of the tau involving hadronic final states allows an efficient measurement of its polarization [75].

In the H $\rightarrow \tau\tau$ search [71] τ lepton can decay into electrons, muons and hadrons τ_h. The di-τ invariant mass $m_{\tau\tau}$ is reconstructed through a kinematic fit of the visible decay products and the missing transverse energy measured in the event. Due to the presence of neutrinos in the final state the di-τ invariant mass resolution is quite poor, of the order of $10\% - 20\%$. The analysis is performed searching for an excess of events over the expected SM background in the $m_{\tau\tau}$ distribution. The main sources of background in the H $\rightarrow \tau\tau$ analysis arise from the Drell-Yan $Z \rightarrow \tau\tau$, $Z \rightarrow ee$, QCD multijet production, W+jets, di-boson and $t\bar{t}$ production. In all the channels analysed, a complex categorization, based on the number of energetic jets in the event and the p_{T} of the di-τ system, is exploited to improve the analysis sensitivity.

The most sensitive category is the VBF one, in which the topology of the VBF Higgs boson production mechanism is exploited to separate the Higgs boson signal from the various background, in particular the Drell-Yan $Z \rightarrow \tau\tau$. In the VBF categories at least two energetic jets with a large pseudorapidity gap and a high di-jet invariant mass are required. The observed boost of the Higgs boson candidate is used to improve the mass resolution, especially in the 1-jet categories. The 0-jet categories are used to constrain the background normalization and to measure the energy scales. The major background contributions are estimated through data-driven techniques. The final signal extraction is obtained performing a simultaneous binned maximum likelihood fit to the $m_{\tau\tau}$ distribution in all the categories and channels analysed (more details in Section 3.4). The search for H decays produced in association with a W or Z boson [76] is conducted in events with three or four leptons in the final state. The WH analysis selects events which have two electrons or muons with the same sign and a hadronically decaying τ ($e^{\pm}e^{\pm}\tau_h^{\mp}$ and $\mu^{\pm}\mu^{\pm}\tau_h^{\mp}$). The ZH analysis is performed in events with an identified $Z \rightarrow ee$ or $Z \rightarrow \mu\mu$ decay and a Higgs boson candidate with one of the following final states: $e\mu$, $\mu\mu$,

$e\tau_h$, $\mu\tau_h$. The main irreducible backgrounds to the WH and ZH searches are WZ and ZZ di-boson events, respectively [12]. The irreducible backgrounds are estimated using simulation, corrected by control samples in data. The $m_{\tau\tau}$ distribution combining all the non-VH categories is illustrated in Fig. 1.21. The CMS collaboration observed an excess of with a significance of 3.2σ at $m_{\mathrm{H}} = 125\,\mathrm{GeV}/c^2$ [71].

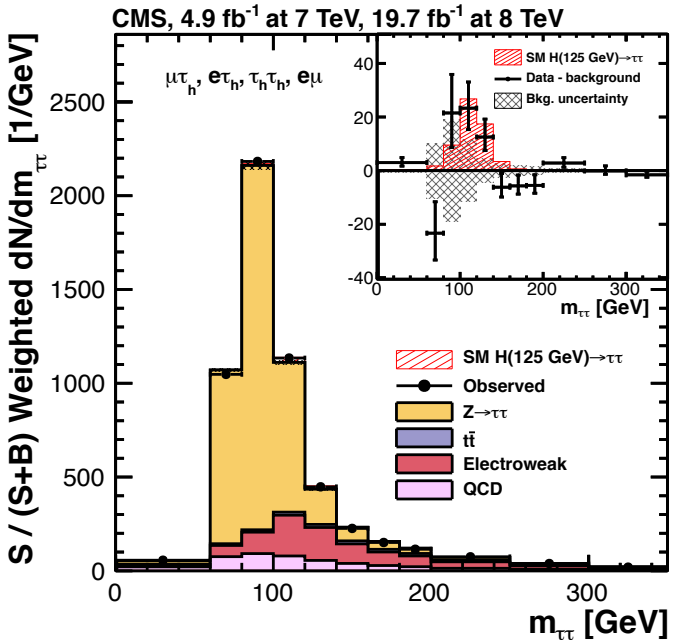


Figure 1.21: *Combined observed and predicted m distributions for the $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$, and $e\mu$ channels. The normalization of the predicted background distributions corresponds to the result of the global fit. The signal distribution, on the other hand, is normalized to the SM prediction. The distributions obtained in each category of each channel are weighted by the ratio between the expected signal and signal-plus-background yields in the category, obtained in the central $m_{\tau\tau}$ interval containing 68% of the signal events. 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 $mH = 125\ GeV/c^2$ [71].*

Chapter 2

Experimental apparatus

The Compact Muon Solenoid (CMS) experiment is a general purpose detector operating at the Large Hadron Collider (LHC) at the *Organisation Européenne pour la Recherche Nucléaire* (CERN). The LHC project was approved by the CERN Council in December 1994 and in December 1996 the same council approved the construction of a proton-proton collider capable to reach the centre-of-mass energy of 14 TeV. The decision to build the LHC at CERN was mostly influenced by the possibility to re-use the tunnel and the injection chain already built for the LEP. It was decided to equip four of the eight possible interaction regions. Out of the four chosen to host the experiments, two were equipped with new underground caverns, where the ATLAS and CMS experiments were placed [14]. In this chapter I give a brief description of the collider machine and a summary of its performance showed during the first period of data taking (LHC Run 1). In section 2.2, I present the CMS apparatus focussing on some subdetectors. In the final part of the chapter a brief description of the CMS data acquisition system is given.

2.1 The Large Hadron Collider

The *Large Hadron Collider* (LHC) is a two-ring-superconductive-hadron accelerator and collider installed at the European Organization for Nuclear Research - CERN (Geneve) and located $\sim$ 100 m underground inside a tunnel 27 km long. The main physics motivation for the development of the LHC project was the investigation of the fundamental interactions at the TeV scale and this requirement drove the LHC design. The target centre-of-mass energy of the collisions is reachable only through circular accelerators. This choice has a direct impact on the type of particles to be accelerated. Indeed, particles bended in a circular trajectory loose energy through synchrotron radiation. Since the power emitted by synchrotron radiation is proportional to $\sim 1/m^4$ (where m is the mass of the accelerated particle), the choice to accelerate protons (or heavy ions) instead of electrons follows as a natural consequence. The LHC uses 1232 superconductive dipole magnets (14 m long and about ~ 35 t) in order to bend the particle

trajectory and keep it in the 27 km circular orbit. Quadrupole magnets (392 along the whole accelerator ring) are used to collimate the beams and more than 7000 magnets of different nature are used to apply fine tuning corrections to the beam parameters. To bend ultra relativistic protons with a kinetic energy of $\sim 7\,\text{TeV}$ and keep them along the LHC circular trajectory, a magnetic field of $8.33\,T$ is required. The LHC dipoles magnets work at a temperature of $1.9\,K$ and are supplied through superconductive cables made of a particular alloy of NbTi [14] able to support a current intensity of $11850\,A$ [14]. The difficulties in producing antiprotons lead to the choice to build a pp collider instead of a $p\bar{p}$ one. For this reason, two distinct beams pipes are required to accelerate beams in opposite directions. Because of the space limitation inside the tunnel, the two proton beam pipes with the corresponding magnets are hosted together in the same cryostat as it is shown in Fig.2.1. One of the major difficulties in the conception of the LHC magnets structure relies, indeed, in the fact that the magnetic flux circulate in opposite side in the two beam channels.

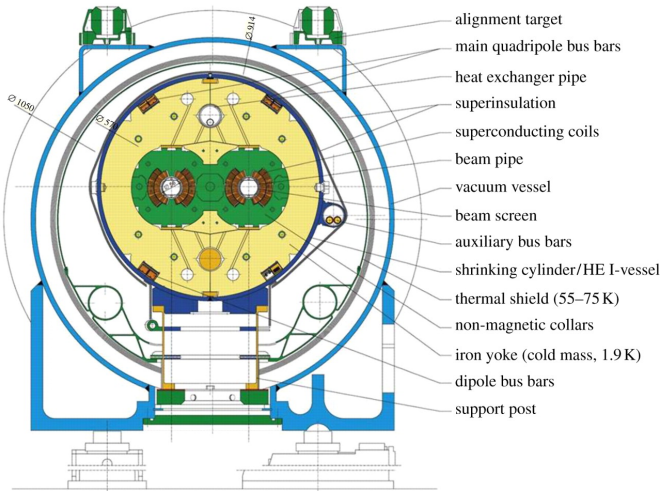


Figure 2.1: *LHC dipole section.*

The LHC parameters

The LHC beam is made of up to 2808 bunches. Each bunch contains $\sim 1.15 \cdot 10^{11}$ protons [14]. The LHC is designed to accept protons with an energy of 450 GeV and accelerate them up to the nominal collision energy. After a first operational period with a beam energy of 3.5/4 TeV, the LHC undertook a first long shut-down. Recently beams are back in the LHC with an energy that already reached the new record of 6.5 TeV. Bunches are spaced by $50\,ns$ and, at the interaction point, their longitudinal length is $7.55\,cm$ while their transversal size is designed to be $70.9\,\mu m$

[14]. One of the most important features of a collider is its instantaneous luminosity. This physical quantity is completely determined by the parameters of the machine:

$$\mathcal{L}_{inst} = \frac{\gamma f N_B N_p^2}{4\pi\varepsilon_T \beta^*} F \tag{2.1}$$

where, in the nominal proton beam operation conditions, $\varepsilon_T = 3.75\,\mu m$ is the transverse emittance, $\beta^* = 0.55\,m$ is the betatron function evaluated at the interaction point, $\gamma = E_{beam}/m_p$ is the Lorentz factor for the accelerated particle, f is the revolution frequency and $F = 0.836$[1] is a reduction factor that takes into account the crossing angle at the interaction point between the two beams. A summary of the main LHC parameters is provided in Tab. 2.1.

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

Table 2.1: Summary of the most relevant LHC parameters.

CERN acceleration chain

The acceleration up to such high energies could not be performed in one go. Protons are actually accelerated up to 6.5 TeV in different steps, accomplished in different accelerators, as it is shown in Fig. 2.2. Protons are firstly obtained from hydrogen atoms ionisation and accelerated linearly by the LINAC-2 until they reach an energy of $50\,MeV$. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons up to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated up to 450 GeV. The protons are finally transferred to the two beam pipes of the LHC, where they are accelerated to the maximal energy. The LHC makes use of 8 radio-frequency cavities operating at a frequency of $400\,MHz$ to accelerate protons in the final stage. The gradient of the electric field delivered is of the order of $5\,MV/m$. LHC is designed to collide protons versus protons (or protons versus ions or

[1]The geometric luminosity reduction factor depends on the total crossing angle at the interaction point: $F = 1/\sqrt{1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)}$. The quoted number in assumes a total crossing angle of $285\,\mu rad$ as it is used in experimental insertion region (IR) 1 and 5.

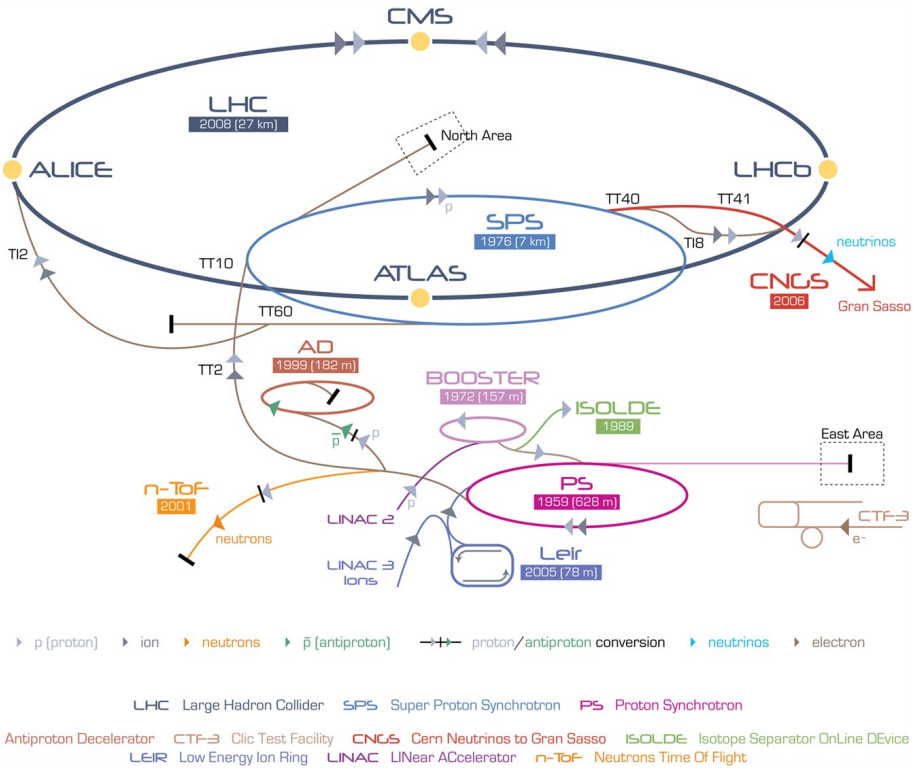


Figure 2.2: *The CERN accelerator complex.*

ions versus ions) with a maximum energy in the centre-of-mass of the collision equal to 14 TeV (2.76 TeV per nucleon [14]) and with a design luminosity of $1 \cdot 10^{34} cm^{-2} s^{-1}$. As a matter of facts, CERN changed the LHC machine operation parameters different times since its first collisions. In the very beginning of the LHC operations, the centre-of-mass energy of the collisions has increased from 900 GeV to 2.36 TeV in 2009. Subsequently it reaches $\sqrt{s} = 7$ TeV in 2010 and $\sqrt{s} = 8$ TeV in 2012. The instantaneous luminosity $\mathcal{L}_{inst}$ has also continuously increased over the running period. Indeed, $\mathcal{L}_{inst}$ passed from $10^{29}\, cm^{-2} s^{-1}$ in 2009 to $3 \cdot 10^{33}\, cm^{-2} s^{-1}$ at the end of 2011 and to a peak of $7 \cdot 10^{33}\, cm^{-2} s^{-1}$ during 2012. During the whole LHC Run 1 the total integrated luminosity delivered by the machine is about 30 fb^{-1}.

In April 2015, the LHC, after its first long shut-down, started the Run 2 operations, circulating beams at the energy of 6.5 TeV with a bunch spacing of $25ns$. Beams collided at the record-breaking energy of 13 TeV for the first time on the 21^{st} of May 2015.

2.2 The CMS detector

The prime motivation for the LHC is to elucidate the nature of the electroweak symmetry-breaking. The LHC physics programme is really wide and extends from the study of the property of b-mesons, to the search of new physics signals like supersymmetry, extra dimensions, Dark Matter particles [14]. Several detectors are positioned along the accelerator ring to exploit the LHC collisions and the *Compact Muon Solenoid* (CMS) is one of them. The CMS detector is a multi-purpose apparatus: it is a cylindrical, 21.6 m long detector, with a diameter of 15 m and weights about 14000 t. CMS is composed by different subdetector units disposed in a radially cylindrical geometry. The central part, called barrel, is completed by two endcaps that make the detector almost hermetic. A sketch of the CMS detector is illustrated in Fig. 2.3.

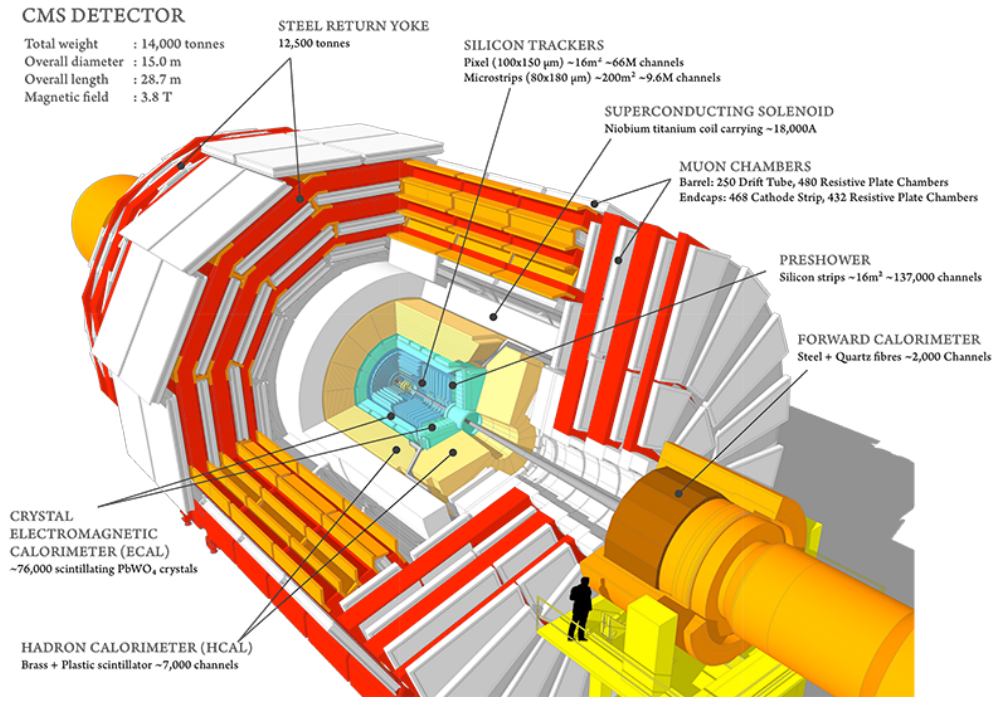


Figure 2.3: *Section of the CMS detector.*

The different CMS subdetectors provide complementary measurements as to identify and measure the momenta of different types of stable particles (as illustrated in Fig. 2.4):

- *Tracker*: identifies the charged particles, measure their momenta and their charge exploiting the bending of the trajectory in the magnetic field
- *Electromagnetic calorimeter*: measures the energy, the direction of the electrons and photons and provide information for their identification

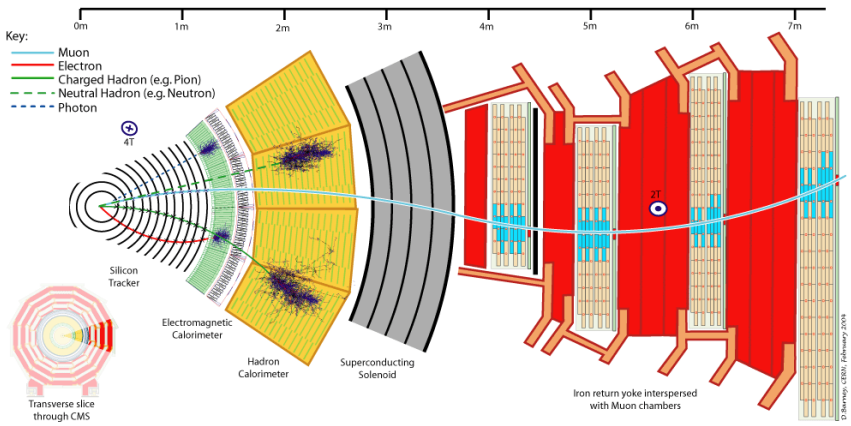


Figure 2.4: *A slice of the CMS detector. The different subdetectors can be seen. The typical behaviour of different particles in the various subdetectors is also represented.*

- *Hadronic calorimeter*: measures the energy of the hadrons
- *Muon detectors*: identifies muons, measure their momenta and their charge exploiting the bending of the trajectory in the magnetic field

One of the experimental challenge imposed by the LHC collisions lies in the event complexity generated by hadron collisions that can result in hundreds of particles propagating through the detector. Moreover, in the same bunch crossing, additional inelastic collisions may occur giving rise to pile-up interactions. Therefore, the detector and the reconstruction algorithms have the difficult task to detect, reconstruct and disentangle the particles produced in the hard process. In addition, this process should already be partially carried out online, in the trigger system.

The detector requirements for CMS to meet the goals of LHC physics programme can be summarized as follows:

- Good *muon* momentum identification and resolution (1% for muons up to 100 GeV in transverse momentum), and identification of the *muon* charge until muon $pt < 1$ TeV [15]
- Excellent *charged particle* momentum resolution and reconstruction efficiency in the inner tracker [15]. Efficient triggering and *offline* tagging of *b-jets* impose a pixel detector close to the interaction point [15]
- Excellent electromagnetic energy resolution, good *di-photon* and *di-electron* mass resolution ($\approx 1\%$ at 100 GeV), wide geometric coverage, efficient *lepton* and *photon* isolation at high luminosity [15]
- Good $\not{E}_T$ and *di-jet* mass resolution. This last requirement impose an *hadron calorimeter* with a wide geometric coverage [15]

One of the main performance criteria for CMS is to optimize the resolution for charged particles. The magnetic field bends the trajectory of charged particles: the curvature radius provides a value of the particles momenta. The bending, moreover, increases with the magnetic field intensity, allowing a larger separation between the different charged particles and an improved precision on their momentum measurement. The CMS detector is equipped with a powerful magnet. The presence of a magnet between the tracker and the calorimeters would mean a consequent layer of inactive material budget where particles could interact. Particles could interact with this material loosing part of their energy before reach the calorimeters. For this reason, CMS opted for a compact detector design, placing the calorimeters inside the magnet. This choice presented, however, big challenges because of the requirements that the magnet must satisfy. Indeed, the magnet should be big enough to host the tracker system and the calorimeters but at the same time must produce a very intense magnetic field both in the center of the detector as well as outside the magnet. Indeed the muon detectors are placed inside the return yoke of the magnet. The solution was to make use of the superconductive technology.

2.2.1 The solenoid magnet

The CMS magnet is a superconducting solenoid of cylindrical shape: $12.9\,m$ long with a diameter of $5.9\,m$. It is able to produce a magnetic field of $3.8\,T$ in the middle of the detector. This feature make the CMS solenoid the biggest solenoid in the world. It is capable to produce an axial magnetic field of 3.8T inside its volume [14]. The magnet is composed of a superconducting coil housed in a vacuum tank and of a *return yoke*. Such an intense magnetic field has the function to allow a good measure of the momentum of the charged particles until $|\eta| = 2.4$. Moreover the high magnetic field does not allow low energy charged particles to reach the ECAL detector, and, in some way, contribute to reduce the effect of the *pile-up*. In particular, given the magnetic field of $3.8\,T$, all particles with $p_{\mathrm{T}} \lesssim 0.75\,\mathrm{GeV}$ do not reach the ECAL surface. The superconducting coil is cooled down by liquid helium. It is supported by a steel structure that plays the role of a return yoke and simultaneously host the muon chambers. The *return yoke* magnetic field reaches $\sim 2\,T$, and is used to reconstruct the muon track momentum in dedicated detectors. Some of the most important operation parameters of the CMS solenoid are summarized in the Tab. 2.2.

Magnet parameters	
Magnetic field in the interaction point	3.8T (4T nominal)
Coil length	12.48 m
Stored energy	2.70×10^9 J
Circulating current	20000 A

Table 2.2: *Main properties of the CMS solenoid.*

2.2.2 The CMS coordinate system

The coordinate system of CMS is chosen with the z axis in the direction of the tangent to the ideal LHC circumference and the beam direction, with the y axis perpendicular to z one and pointing to the center of the ring, and the x axis perpendicular to both z and y and directed to the surface. Nevertheless, since the detector has a cylindrical geometry, it is more convenient to use the coordinate system (r, ϕ, η). Where r is the distance from z (beam axis), ϕ is the azimuthal angle and η is the *pseudorapidity*, which is the ultra-relativistic approximation of the rapidity and can be written as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.2}$$

where θ is the polar angle with respect the beam direction z. Using η instead of θ is justified by the two reasons:

- the production of particles is flat in pseudorapidity
- the differences of rapidity is invariant in a boost along z

A more detailed description of the CMS coordinate system can be found in [77]

2.3 The CMS subdetector

2.3.1 The tracking system

The tracker [14] [78] is the sub-detector nearest to the interaction point, positioned inside the $3.8\,T$ magnetic field generated by the superconductive solenoid. At the LHC luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ there will be on average ~ 1000 particles from more that 20 overlapping proton-proton interactions traversing the tracker for each bunch crossing (every $25-50$ ns). Therefore a detector featuring high granularity and fast response is required, such that trajectories can be identified reliably and attributed to the correct vertex. Searching for dilepton resonances requires good momentum resolution for transverse momenta (p_{T}) of up to 1 TeV. At the same time, efficient reconstruction of tracks with very low p_{T} of order $300\,MeV$ is needed to obtain optimal jet energy resolution with the particle flow reconstruction (see Section 3.1). In addition, the hard radiation environment will also cause severe damage to the tracking system. The main challenge in the design of the tracking system was to develop detector components able to operate in this harsh environment for an expected lifetime of 10 years. These requirements, together with the attempt to keep the amount of material minimal as to limit the particle interactions, lead to a tracker design entirely based on silicon detector technology. The tracking system is further divided into three main regions according to incident particle flux and to the distance to the beam axis r.

- **The pixel detector:** The pixel detector lies in the high flux region ($\sim 10^7 neutrons$ s^{-1}, $r \sim 10$ cm) and is composed by 3 concentric layers positioned at distances $r = 4.4$ cm, $r = 7.3$ cm, $r = 10.2$ cm respectively from the interaction point (red modulus in Fig.2.5). Each barrel layer is 53 cm long. This region is complemented by two endcaps made of two disks positioned respectively at $|z| = 34.5$ cm and $|z| = 46.5$ cm from the interaction point and having a 6 cm $< r <$ 15 cm radius. These elements, made of silicon pixel detectors with dimensions: 100 μm $\times$ 150 μm $\times$ 250 μm, compose the *inner pixel detector*. This region of the tracking system allows the detection from two to three hits per track inside the geometrical region $|\eta| < 2.2$ for particles coming from a region inside $2\sigma_z$ from the interaction point (σ_z represents the bunch longitudinal dimension at the interaction point).

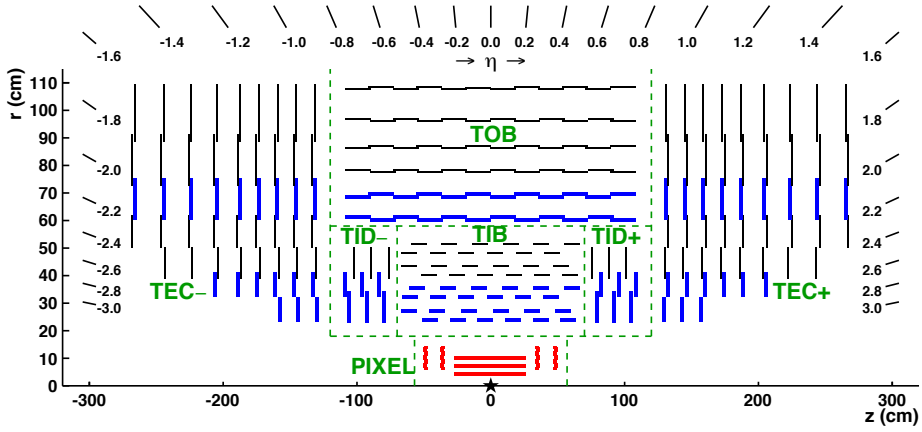


Figure 2.5: *Schematic cross section through the CMS tracker in the r-z plane. In this view, the tracker is symmetric about the horizontal line r = 0, so only the top half is shown here. The centre of the tracker, corresponding to the approximate position of the pp collision point, is indicated by a star [14].*

- **The inner tracker:** The inner tracker is positioned in the medium flux region (20 cm $< r <$ 55 cm) and it is composed by silicon microstrip detectors with dimensions: 10 cm $\times$ 80 μm $\times$ 300 μm. They are organized in 4 barrel layers (the most inner two are double-faced. All the barrel layer are named *TIB* in Fig.2.5.) and in 3 endcap for each side (*TID+* and *TID-* in Fig.2.5).

- **The outer tracker:** The outer tracker detector is positioned in the low flux region and it compounds also by silicon microstrip detectors of dimensions 25 cm $\times 80\,\mu m$ $\times 500\,\mu m$. They are organized in 6 barrel concentric layers (*TOB* in Fig.2.5) and in 9-layer endcap for each side (*TEC+* and *TEC-* in Fig.2.5).

Particle interaction with tracker material

The ensemble of tracker detectors represents a non negligible quantity of material budget if one considers not only the active volume of the sensors but also the read-out electronic, the electronic supplies, the cables that provide the data information flux with the outside, the mechanical support and cooling. In Fig. 2.6 the material budget as a function of $|\eta|$ is shown: the contribution coming from each single tracker subsystem (also the beam pipe) is detailed in units of radiation length, in Fig.2.6 (Left), as well in units of nuclear interaction length, in Fig.2.6 (Right). The particles interact with the material, loosing, as consequence, a fraction of their energy through Bremsstrahlung radiation, photon conversion and nuclear interaction [79].

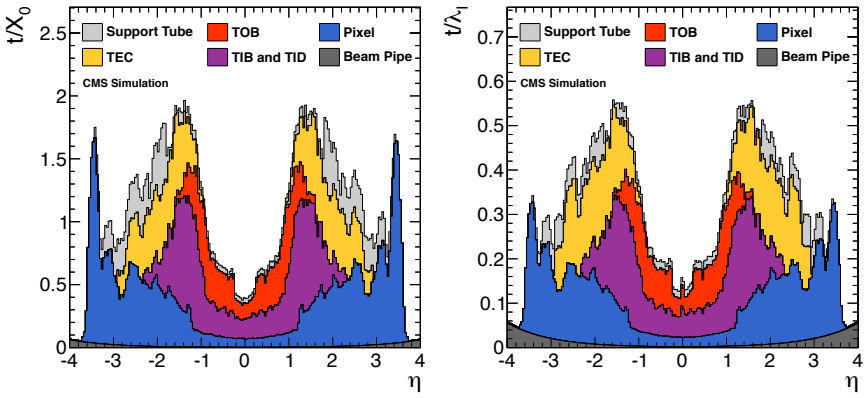


Figure 2.6: *Simulation of the tracker material budget (in addition also the beam pipe is material budget is shown) coming from the different tracker subsystem and expressed in units of radiation length (Left) as well in units of nuclear interaction length (Right) [80].*

Tracker performance

The CMS silicon tracker has been designed to provide precise hit measurements in order to allow for very efficient tracking and vertex reconstruction in the dense environment of the proton-proton interactions at the LHC. Here performance results of the pixel and strip trackers are summarized. The results are obtained with proton-proton collisions at $\sqrt{s} = 7$ TeV [80].

Resolution: The hit resolution in the pixel and strip barrel sensors has been studied by measuring residuals, defined as the difference between the measured and the expected hit position as predicted by the fitted track [80]. One can then translate the width of the residual distribution into the intrinsic resolution after unfolding it from the beam width [80]. The obtained result for the resolution of the transverse coordinate, for track with $p_T > 12$ GeV/c is 10.4 μm. The

resolution in the longitudinal direction depends on the angle between the track and the sensor. In the strip tracker the resolution in the barrel is measured using hits on tracks passing overlapping modules. The difference between the measured and expected (from the track) hit position are compared. The width of this difference is a measure of the hit resolution [80]. The measured values range between 14 μm (for a 80 μm sensor pitch) and 36 μm (for a 83 μm pitch).

Efficiency: The hit finding efficiency in the pixel detector is measured by using well reconstructed, isolated tracks with a $p_{\mathrm{T}} > 1$ GeV/c, originating from the primary vertex. The hit efficiency is calculated from the present and missing hits on and near the track (within 0.5 mm of the predicted position). The average hit efficiency is measured to be $> 99\%$. In case of the strip tracker the hit finding efficiency is measured using tracks reconstructed with at least eight hits and not passing near the sensors edges. The average hit finding efficiency is measured to be 99.7% [80].

Tracking and Vertexing Performance

Once the tracks are reconstructed with the approach described in the next Chapter 3, the vertices can be reconstructed as follows:

1. selection of the tracks to be used;
2. clustering of the tracks, i.e. decide the tracks which originate from the same interaction vertex;
3. fit of the position of each vertex using its associated tracks.

The estimation of the resolution in the main muons track parameters, the transverse (d_0) and longitudinal (Z_0) impact parameter and the track transverse momentum (p_{T}), as function of the reconstructed muon p_{T}, is performed using simulation and it is illustrated in Fig. 2.7.

The measured efficiency and resolution of the primary vertex reconstruction are shown in Fig. 2.8 where it can be seen that as soon as there are three tracks, the vertex reconstruction efficiency is about 99.8%.

2.3.2 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [81] measures the energy of electrons and photons. The energy resolution of photons and electrons constitutes a crucial aspect for many physics analysis. In particular for the analysis searching for the Higgs boson in the $H \to \gamma\gamma$ channel or in the $H \to ZZ^* \to 4\ell$ (with 2 or 4 electrons in the final state). In this context, the energy resolution for the di-photons and di-electron invariant mass should be of the order of $\sim 1\%$ at 100 GeV [14]. In addition, since ECAL is positioned inside the solenoidal magnetic field, the geometrical constraints force it to be very compact. ECAL is an homogeneous calorimeter,

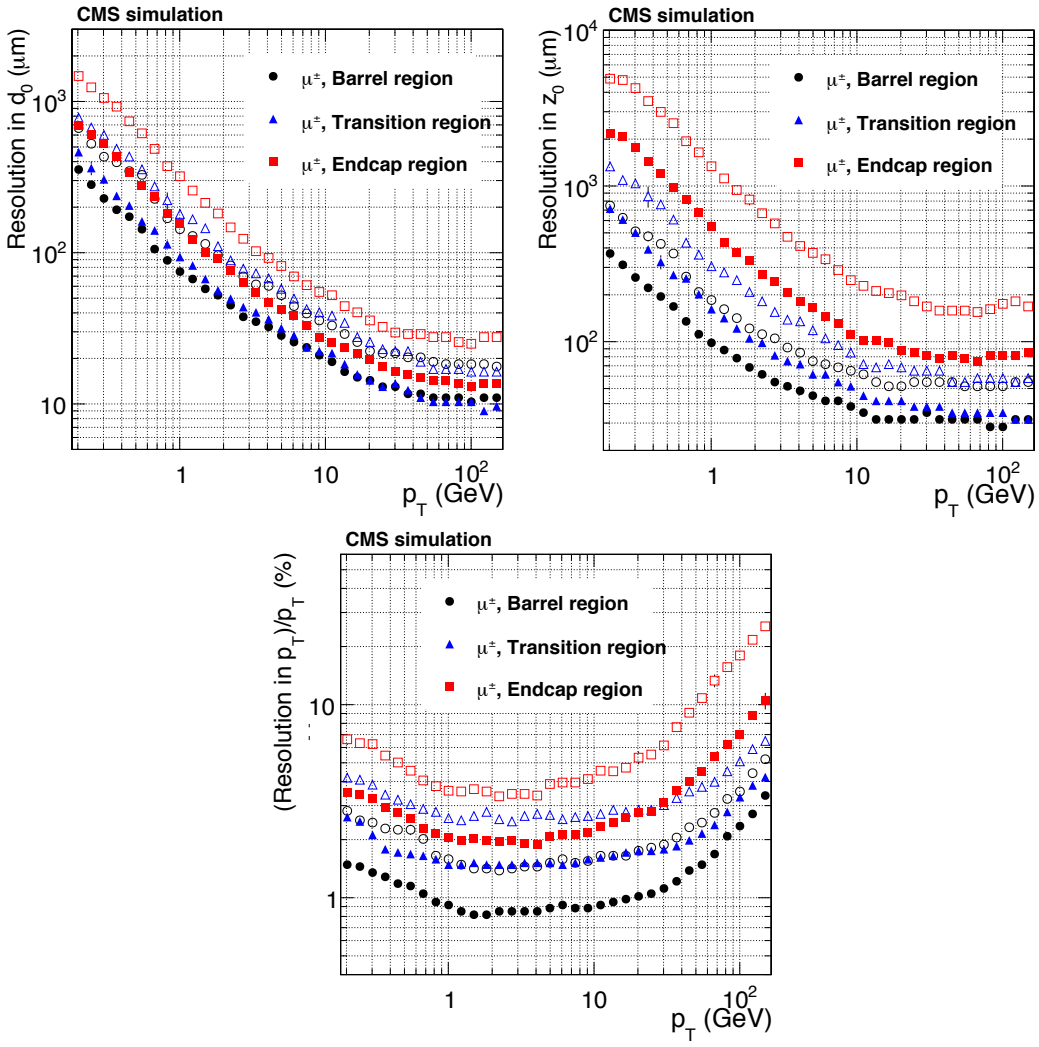


Figure 2.7: *Resolution, as a function of p_T, of the transverse (upper left) and longitudinal (upper right) track parameter as well as transverse momentum resolution (bottom) for single, isolated muons in the barrel, transition, and endcap regions, defined by η intervals of $[0, 0.9]$, $[0.9, 1.4]$ and $[1.4, 2.5]$, respectively [80].*

composed by 75848 lead tungstate ($PbWO_4$) crystals, 61200 of which are in the barrel and the others in the two endcaps. The ECAL geometrical coverage extends up to $|\eta| = 3$. Each crystal is a scintillator, and the choice of the $PbWO_4$ as scintillator material was guided mainly by the following needs [81]:

- longitudinal containment of electromagnetic showers: $PbWO_4$ density is approximately 8.28 g/cm^3 and its radiation length is $X_0 = 0.89$ cm [81]

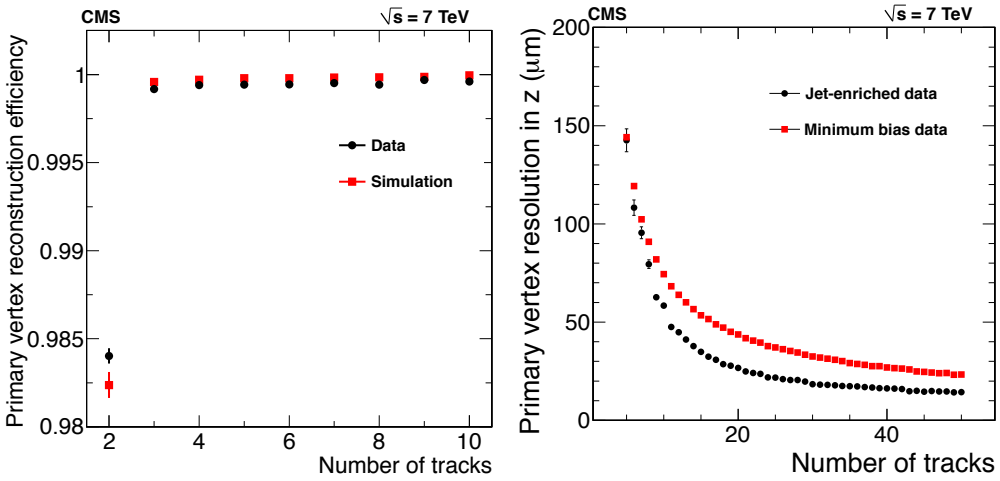


Figure 2.8: *Vertex reconstruction efficiency in data and simulated events (Left). Vertex resolution for clean (MinBias) and Jet ($p_T > 20\,GeV$) enriched event (Right) plotted as a function of the number of tracks used in the vertex reconstruction [80].*

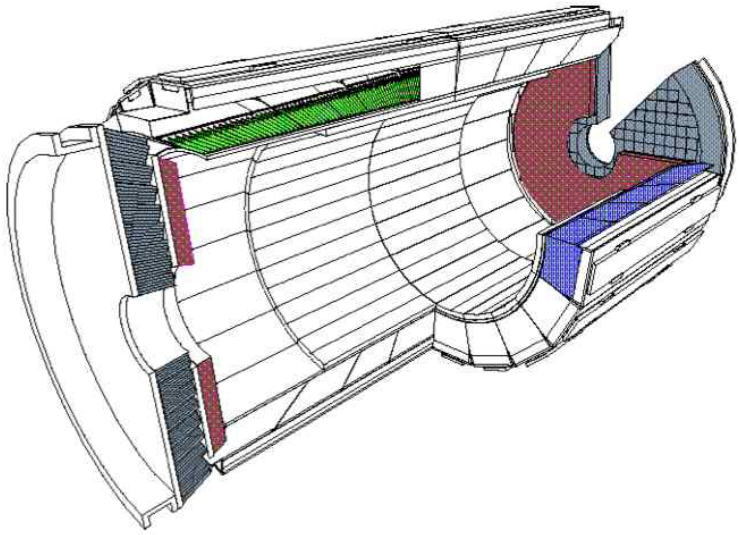


Figure 2.9: *Slice of the electromagnetic calorimeter - ECAL [81]*

- transversal containment of electromagnetic shower: $PbWO_4$ has got low values of Moliere radius $R_M = 21.9$ mm; [81]
- fast response: approximately 80% of the total scintillation light is emitted in less than 25 ns

- high resistance to radiations

The ECAL crystal layout follows a pseudo-projective geometry: the crystal points toward the center of the detector (that is also the nominal interaction point) with an additional angle of 3° (in the endcap this angle may vary between 2° and 8°). This arrangement allows the measurement of the energy of photons and electron coming from the interaction point avoiding that the particles could end up in inactive regions between the crystals. The ECAL barrel covers a region $|\eta| < 1.479$ and has an inner radius of 129 cm. It is structured in 36 *supermodules* (in Fig. 2.10 it is the crystals which are in blue), each one containing 1700 crystals and covering an angle of $\phi = 20°$. Each supermodule is subdivided along η in four modules that, are composed by different submodules, each one made of a group of 5×2 crystals. The barrel crystals show a frontal section of 22×22 mm^2 and have got a length of 220 mm, corresponding to $\sim 25.8 X_0$ [81]. The barrel *granularity* in the $\eta - \phi$ plan is of 0.0175×0.0175. The endcaps (highlighted in grey in Fig. 2.10) cover the region $1.48 < |\eta| < 3.0$ and allow precision measurements up to $|\eta| < 2.5$. The endcap crystals have dimensions $28.6 \times 28.6 \times 220$ mm^3 and are structured in groups of 5×5 units, named *supercrystal*.

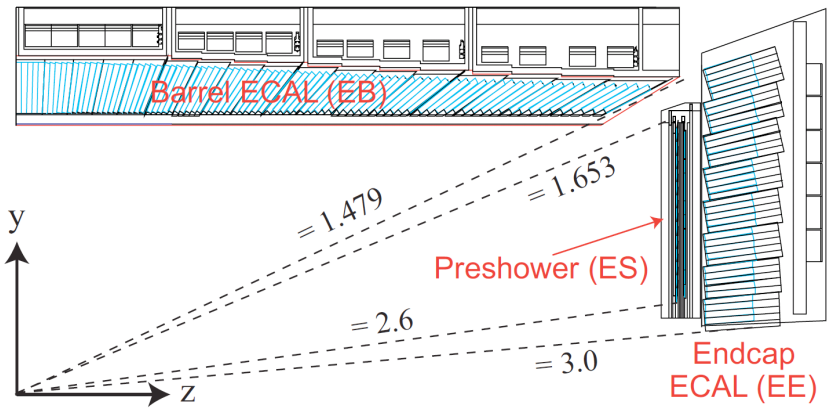


Figure 2.10: *Slice of the electromagnetic calorimeter - ECAL [81]*

The *preshower* detector (highlighted in red in Fig. 2.10) is a system of solid state detectors placed in front of the ECAL endcaps: the main purpose of the preshower is the prompt photon π^0 disambiguation in the region $1.653 < |\eta| < 2.6$. This could be of capital importance to obtain a correct isolation for *photon* coming from $\pi^0 \to \gamma\gamma$. The active elements of the *preshower* are two layers of silicon strips detectors lying between lead discs which are $2X_0$ and $1X_0$ thick respectively. The dimensions of the strips are $5\,cm \times 1.2\,mm$ and they are positioned in two orthogonal orientation in the two layers.

ECAL calibration

The calibration is a technical challenge for the operation of CMS ECAL. In fact many small effects need to be taken into consideration as the level of precision of a few *per mille* is sought [81]. The energy calibration must deal with both absolute energy scale and a channel-to-channel intercalibration. The essential issue are the energy response uniformity over the whole ECAL subdetector and in time. In this way, showers in different ECAL positions, at different times, could be accurately compared to each other. The main source of channel-to-channel response variation in the barrel is the crystal-to-crystal variation of scintillation light ($\approx$ 8% within supermodules and $\approx$ 15% in all barrel [82]). The final goal is to achieve the most accurate energy measurements for electrons and photons. In order to take into account conversions and bremsstrahlung radiation in the tracker material it has been very important to perform calibration with real data. The ECAL inter-calibration is carried out combining several methods [83]. The absolute ECAL calibration is obtained with $Z \to e^+e^-$ events and $Z \to \mu^+\mu^-\gamma$ events [82].

The laser-monitoring system

The ECAL $PbWO_4$ crystals feature a rapid loss of optical transmission under irradiation due to the production of colour centres which absorb a fraction of the transmitted light [84]. In the different LHC working condition, the result is a cyclic transparency behaviour between LHC collision run and machine refills [14]. The magnitude of the induced transparency change is dose-rate dependent (from 1% in the barrel at low luminosity to 30% in endcap at high luminosity [14]) and lead to unacceptably degraded performance. The evolution of crystal transparency is monitored using laser pulses injected into the crystals via optical fibres. Because of the different optical paths and spectra of the laser pulses and the scintillation light, the change of the transparency cause a change in the response to the laser light which is not necessary equal to the change in response to scintillation light. The relationship between the variations can be expressed by the following empirical power low:

$$\frac{S(t)}{S(t_0)} = \left[\frac{R(t)}{R(t)}\right]^{\alpha} \tag{2.3}$$

where $S(t)$ represent the response to scintillation light, $R(t)$ is the response to the laser pulses and α is a parameter characteristic of the crystal. The test-beam value of α is 1.52 [14]. The relationship expressed in Eq. 2.3 is valid in the barrel both for low and high luminosity. The relative response to laser light measured by ECAL is illustrated in Fig.2.11.

Energy Resolution

The ECAL energy resolution has been parametrized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.4}$$

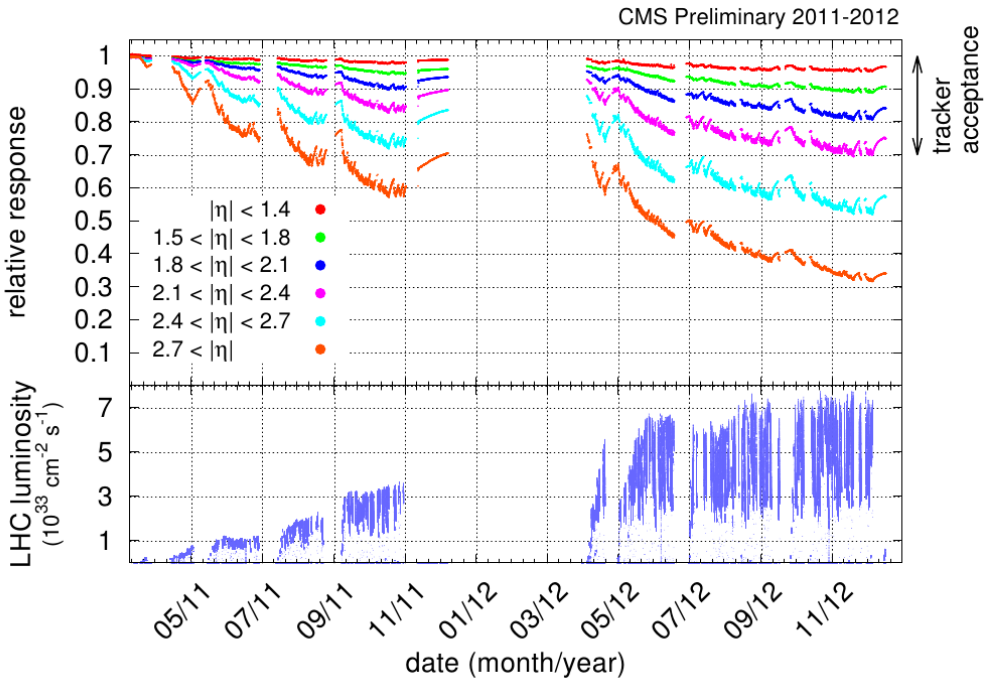


Figure 2.11: *Relative response to laser light (440 nm) measured by the ECAL laser monitoring system, averaged over all crystals in bins of pseudorapidity, for the 2011 and 2012 data taking periods [85].*

where the parameters S, N and C are respectively the *stochastic* term, the *noise* term and the constant term.

- **Stochastic term:** There are three main contribution to the *stochastic* term:
  1. event-to-event fluctuations in the lateral shower containment, expected to be $1.5 \div 2.0\%$
  2. photostatistics contribution of 2.1%
  3. fluctuations in the energy deposited in the preshower absorber (where present) with respect to what is measured in the preshower silicon detector
- **Noise term:** There are two contributions to the *noise* term:
  1. electronic noise and digitization noise ≈ 40 MeV/channel
  2. Pile-Up noise
- **Constant term:** The most important contribution to the *constant* term are:
  1. non-uniformity of the longitudinal light collection (affecting high energy particles)

2. intercalibration errors
3. rear leakage of showers

Taking into account all contribution, the fit to the energy resolution performed during the test beam with the parametric function reported in Eq. 2.4 yield the following values for each contributions in the barrel [86]

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12\%}{E}\right)^2 + (0.30\%)^2 \tag{2.5}$$

In the 7 TeV data, the calibration of the absolute energy is determined from $Z \to e^+e^-$ decays to a precision of 0.4% in the barrel and 0.8% in the endcaps [87].

2.3.3 The hadron calorimeter

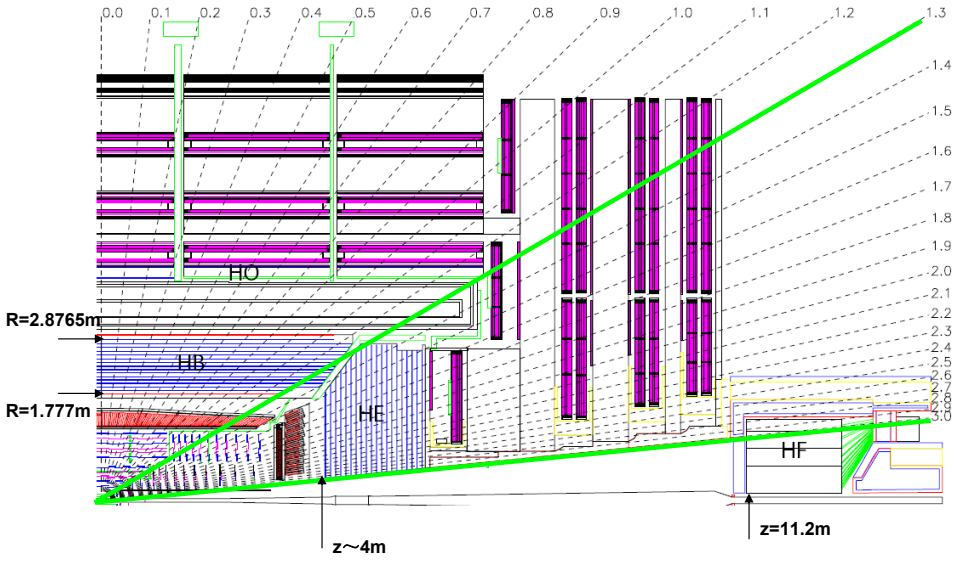


Figure 2.12: *Longitudinal size of HCAL detector*

The hadronic calorimeter (HCAL), positioned behind the ECAL, measures the energy deposited through *hadronic* interactions. In practice, it is the instrument used to measure the energy of hadrons. It is a sampling calorimeter and is made of different layers of brass alternated with plastic scintillator [88][14]. The HCAL consists of 4 main regions (Fig. 2.12):

- **Barrel Detector (HB)** ($|\eta| < 1.3$) - 1.806 m $\leqslant r \leqslant$ 2.950 m - cylindrical geometry
- **Endcap Detector (HE)** ($1.3 < |\eta| < 3.0$) - disk-shaped geometry

- **Outer Detector (HO)** ($|\eta| < 1.3$) 3.82 m (1° layer) $\leqslant r \leqslant$ 4.07 m (2° layer) - cylindrical geometry

- **Forward Detector (HF)** ($3.0 < |\eta| < 5.2$) - cylindrical geometry

The barrel is divided into 2 halves barrel each one consisting of 18 identical azimuthal wedges. A cross section of a wedge is reported in Fig. 2.13. Each wedge is segmented into 4 azimuthal angle (ϕ) sectors. Wedges are composed by brass absorber plates, combined in a staggered geometry. All this system of substructures have the purpose to minimize the dead material in the transverse direction. The absorbing material is made of brass plates with a density of 8.53g/cm^3, and a radiation length of $\lambda_l = 16.42cm$ [14]. The active medium consists exactly in 17 layers of 9 mm plastic scintillators [89] *Bicron BC408* and 3.7 mm *Kuraray SCSN81*. The scintillation light is carried along *WaveLenght Shifters* (WLS) to HPD's (Hibryd Photodiode) that perform the readout of the signal [89]. As shown in Fig. 2.12, the HCAL coverage is extended

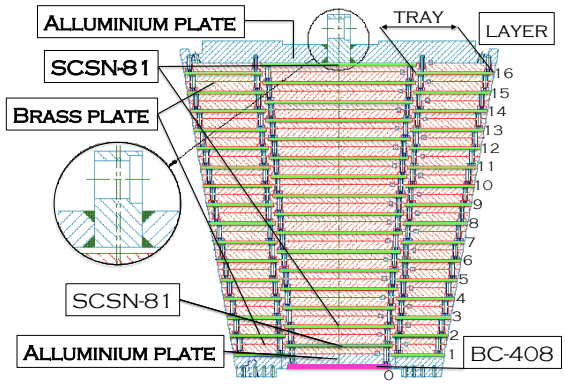


Figure 2.13: *Cross size of an HCAL wedge*

up to $|\eta| = 5$ by two *forward calorimeters*, capable to measure the electromagnetic and hadronic part of the showers initiated by interacting particles, to enable identification and reconstruction of very forward jets and improve the measurement of the missing transverse energy $\not{E}_T$. The forward hadronic calorimeter exploits the Cherenkov effect to detect particle. This kind of detector allows also a basic disambiguation between electromagnetic and hadron showers [89]. The most stringent requirement for the HF is the resistance to high fluence (10^{11} cm^{-2}s^{-1}) and to high absorbed radiation dose (100 Mrad/year). Another important goal of the HCAL is to minimize the non-Gaussian tails in the energy resolution. For these reasons HCAL design has to maximise the material inside the magnet in terms of radiation length. Nevertheless, in the barrel region, the radiation length allowed by the geometrical constraint is approximately 8 λ_I (ECAL+HCAL) [89], and this could be limiting in completely absorbing an hadronic shower started by very energetic particles or by particles that interact in the middle of the calorimeter. In order to correct this deficiency an additional layer of HCAL is positioned outside the coil:

the Outer Hadronic (OH) calorimeter. The design of this subdetector is very simple, in fact it is composed by sheets of scintillator material arranged in a cylindrical geometry with 12 identical ϕ-sectors. The dimensions and the position of OH is constrained by the *muon system*. In the end, the outer calorimeter improves the jet energy resolution acting as *tail catcher* in a way that is possible to see in graph reported in Fig. 2.14.

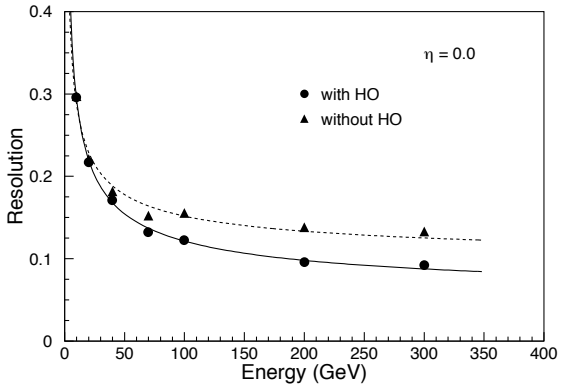


Figure 2.14: *Hadron calorimeter jet energy resolution without and with outer detector.*

According to the test-beam results, the expected energy resolution for single *pions* interacting in the central part of the calorimeter is:

$$\frac{\sigma_E}{E} = \frac{94\%}{\sqrt{E}} \oplus 4.5\% \tag{2.6}$$

where the energy is measured in GeV.
The jet energy resolution are further improved by using a *Particle Flow* technique, as explained further in Chapter 3.1. For this reason the jets and the $\not\!\!E_T$ are by default reconstructed by the Particle Flow algorithm.

2.3.4 The muon system

In the CMS detector, the particles produced in the interaction are stopped in the calorimeters, except the muons (and the neutrinos). The muons constitute a pure signal without particular physics background contamination. The muons play a major role in different research channels ($H \to ZZ^* \to 4\mu(2e2\mu)$, $H \to \tau\tau \to \mu\tau_h$, b physics,...). This is one of the main reason to develop a muon detector with the maximal η coverage. The return yoke houses the muon system, the outermost subdetectors, composed of various detectors, arranged according to four concentric layers hosted in special compartments of the detector structure. The main goals of the muon

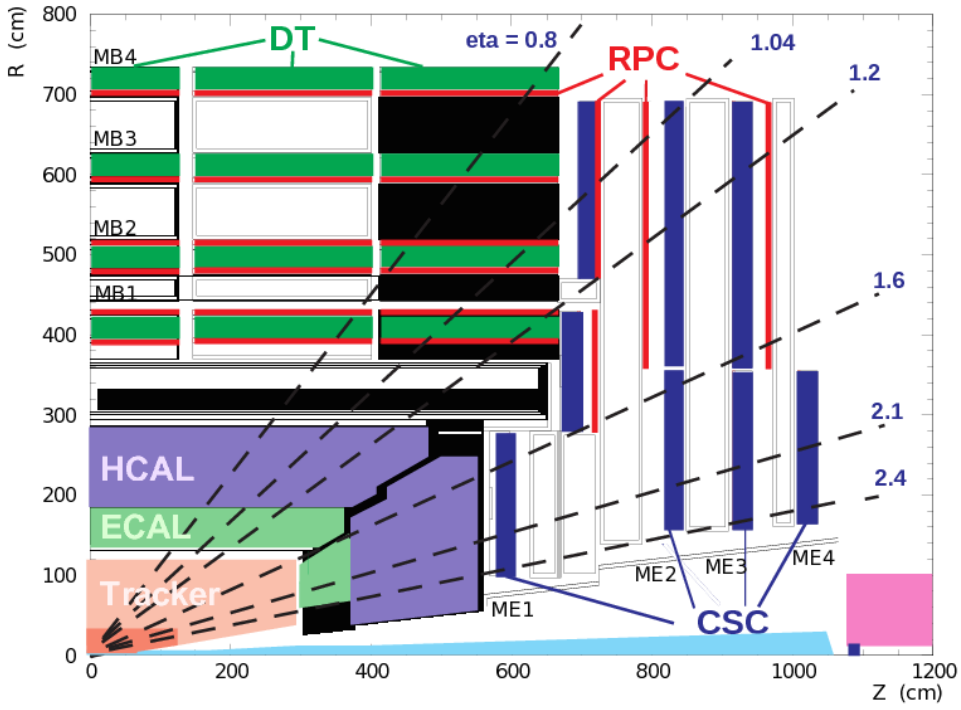


Figure 2.15: *View in the plane (y, z) of a fourth of the muon detector system of CMS. The barrel region extends up to $|\eta| < 1.2$ while the endcaps extend the detector coverage up to $|\eta| < 2.4$. Three different types of detectors are used to detect the passage of a muon: the drift tubes (DT, in green), the cathode strip chambers (CSC, in blue) and the resistive plate chambers (RPC, in red).*

detectors are the triggering and the identification of muons and a precise measurements of their momenta, with the help of the tracking system.

The muon detector is organised in a central part (the barrel) that covers a pseudorapidity region $|\eta| < 1.2$ and two endcaps that extend the pseudorapidity coverage until $|\eta| < 2.4$. In Fig. 2.15 the muon detectors system of CMS is shown in a (y, z) view. The barrel contains four concentric layers made of 250 muon chambers. These layers are further divided in 5 rings along the z axis of CMS: each ring contains 12 sectors covering 30° in ϕ. These sectors are equipped by drift tube detectors (DT) allowing the measurement of the (r, ϕ) position and the z one of the crossing muon. The endcaps cover a region where the activity, in term of particles crossing the surface unit per second, is higher than in the barrel and the magnetic field is not uniform (because of the bending of the magnetic flux lines from the center of the solenoid towards the return yoke). Each endcap contains 468 cathode strip chambers (CSC) detectors. These chambers are perpendicular to the beam axis and provide an excellent spatial resolution and are able to make precision measurement of the (r, ϕ) and η position of the crossing muon.

These detectors, thanks to their good timing and a finer granularity with respect to DT are suited to handle the higher activity in this region of the detector. A complementary set of subdetectors is placed in the barrel and in endcaps regions: the resistive plate chambers (RPC). They are arranged in 6 different layers in the barrel and in 3 layers in each endcap. Thanks to their excellent timing they are used as timing reference for the muon reconstruction and for the trigger. The large muon detector surface, with several layer of measurements, their redundancy and the combination with the tracker results in excellent muon reconstruction performance.

Muon momentum resolution

The resolution of the muon transverse momentum obtained using only the muon detectors, only the tracker and a combination of the tracker and muon detectors is shown in Fig.2.16. For a transverse momentum lower then 100 GeV the measure is dominated by the central tracker measurement. However, for muons with an higher transverse momentum, the most precise measurement comes from the muon detectors.

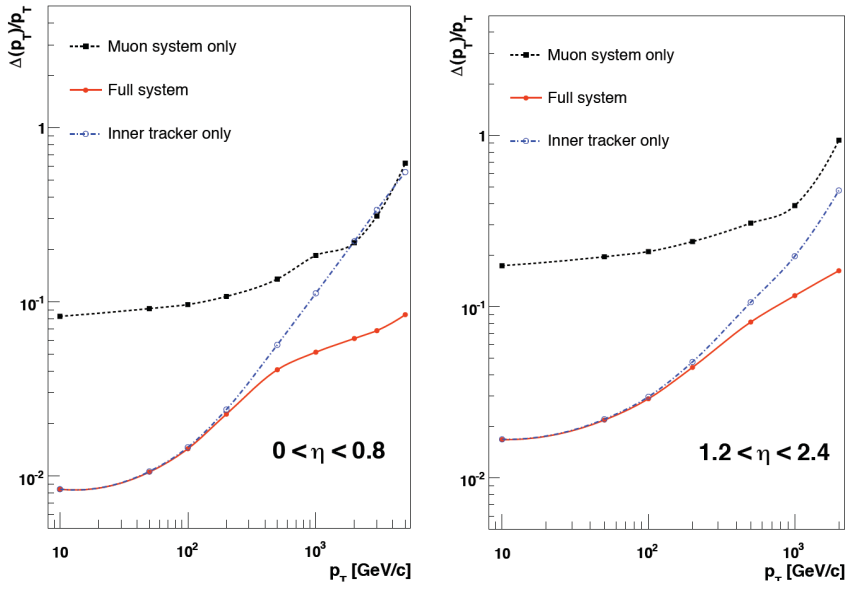


Figure 2.16: *Transverse momentum (p_T) resolution for muons detected in the barrel region (Left) and in the endcaps region (Right). The different lines indicates the different detectors used to measure the muon p_T: only the tracker (dashed blue line), only the muon detectors (dashed black line) a combination of the tracker and the muon detectors (red line) [90].*

2.3.5 The trigger system

The event trigger represents the first step of a physics analysis. The trigger, indeed, defines the nature of the objects and the events recorded, and defines also the phase-space available to study any final state. The search for rare signals at the LHC need both a high collision rate (delivered by the machine) and a very efficient and very fast online selection of the interesting events. At the nominal LHC luminosity, the expected event rate, limited by the crossing rate, is about 20 MHz. Given a typical raw event size of O(1) MB, it is not possible to store all collision events. In fact, the event rate is largely dominated by soft pp interactions (see Fig. 2.17), which are not interesting for the CMS Physics program. Therefore, a trigger system [91] has been devised with the purpose of providing a large rate reduction factor, whilst maintaining a high efficiency on potentially useful events. The total output rate is reduced by about five order of magnitudes to $O(100)\,Hz$ thanks to a two-level system: a Level-1 (L1) Trigger, which consists of custom-designed, largely programmable electronics, and a High-Level Trigger (HLT), which is a software system implemented in a farm of about one thousand commercial processors. These two levels are briefly described in the next of this section.

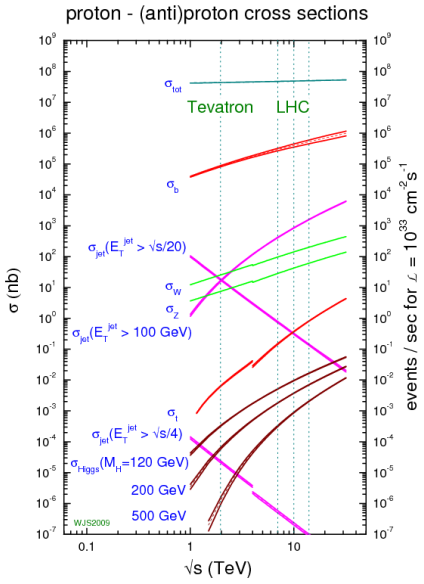


Figure 2.17: *Proton-(anti)proton cross section as function of the centre of mass energy [92].*

The Level-1 trigger

The L1 trigger reduces the rate of selected events down to $O(100)\,kHz$. The full data are stored in pipelined-memories of processing elements, while waiting for the trigger decision. The allowed

trigger latency, between a given bunch crossing and the distribution of the trigger decision to the detector front-end electronics, is 3.2 μs [91][14]. If the L1 accepts the event, the data are moved to be processed by the High Level Trigger. To deal with the LHC high bunch crossing rate, the L1 trigger has to take a decision in a time too short to read data from the whole detector, therefore it employs the calorimetric and muons informations only, since the tracker algorithms are too slow for this purpose. The L1 trigger is organized into a calorimeter and a muon trigger, whose information is transferred to the global trigger which takes the accept-reject decision. The basic purpose of the L1 trigger is to produce electromagnetic candidate (L1-EG), muon candidate (L1-Mu), jet candidate (L1 Jet) and tau candidate (L1 Tau) together with a measure of their transverse energy and isolation variables. In addition, global quantities are measured at L1 trigger level such as the $\not\!\!E_T$ and the total transverse energy deposited in ECAL and HCAL. The calorimeter trigger is based on the trigger towers (TT) read-out: arrays of 5×5 crystals in ECAL which match the granularity of the HCAL towers [14]. During the LHC Run 1 the L1 decisions were taken in three step. Firstly, the *trigger primitive* were identified starting from the TT. Then the trigger towers were grouped in calorimetric regions (RCT) of 4×4 TT. The calorimeter trigger identifies the best four candidates of each of the following classes: electrons and photons, central jets, forward jets and $\tau jets$, identified from the shape of the deposited energy. The information of these objects is passed to the global trigger (GT), together with the measured $\not\!\!E_T$. The muon trigger is performed separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the global trigger, which takes the accept-reject decision exploiting both the characteristic of the single objects and their combination. The scheme of the Run 1 L1 trigger logic adopted by CMS is represented in Fig. 2.18.

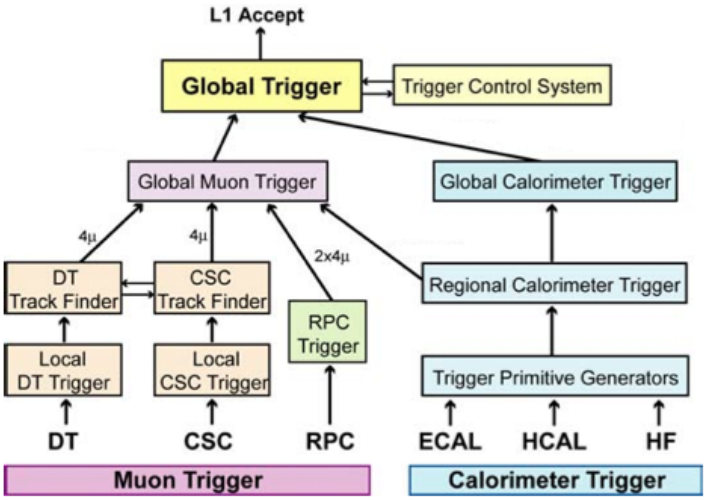


Figure 2.18: *The CMS Level-1 trigger architecture adopted during the LHC Run 1 data-taking period. The calorimeters and the muon detector constitute two parallel selection chains. Each subdetector provides the 4 best candidate and pass them to the Global Muon/Calorimeter Trigger. In a last step the candidates selected by the calorimeters and muon detector chain were merged in the Global Trigger to take the final decision to accept or reject the event.*

L1-τ calorimeter trigger

For the purpose of this thesis, more details on the Level 1 jet and τ trigger algorithms used during the LHC Run 1 are provided. The jet trigger uses the transverse energy (E_{T}) sum found in ECAL and HCAL computed in a calorimeter region. A calorimeter region is a set of 4×4 trigger towers and a trigger tower is defined as the output of 5×5 ECAL crystals and the corresponding HCAL output in the same $\eta - \phi$ area. The dimension of the trigger towers varies in function of the η, as illustrated in Fig.2.19.

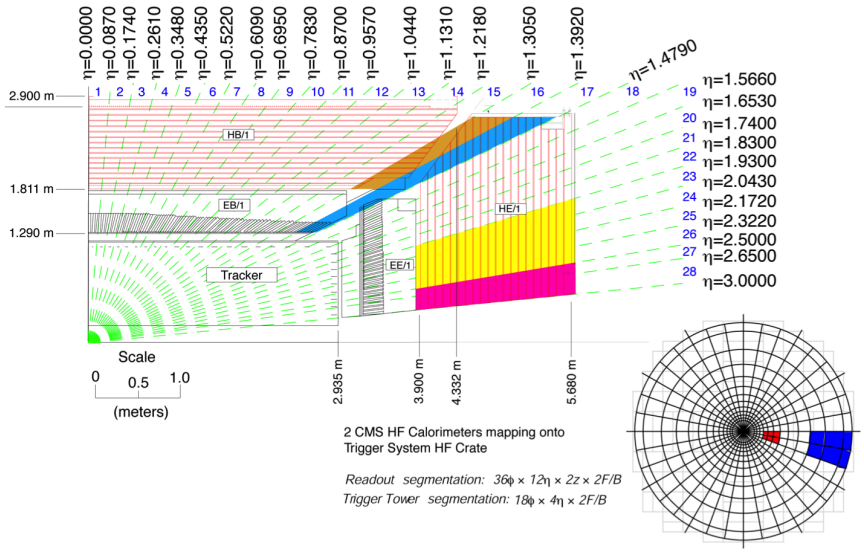


Figure 2.19: *A schematic view of the tower mapping of the CMS calorimeters in r-z plane.*

The jet trigger uses a 3×3 calorimeter region sliding window technique which spans the complete (η, ϕ) coverage of the CMS calorimeters. The E_{T} deposits in the central region is required to be higher than the 8 neighbour regions (see Fig. 2.20). The jets candidates are labelled by (η, ϕ) of the central calorimeter region. Jets in the forward and backward HF calorimeters are sorted and counted separately to prevent more background susceptible high η region from masking central jets [91].

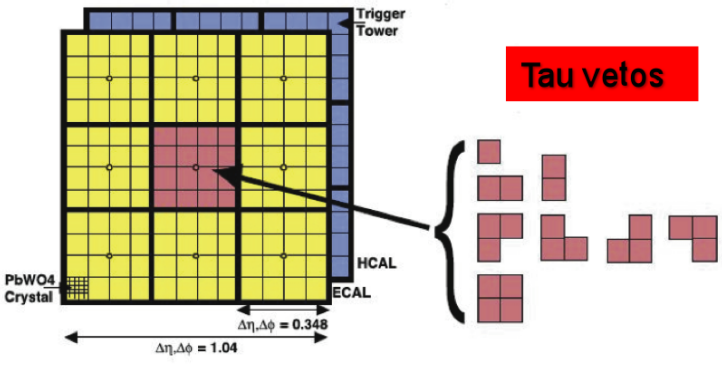


Figure 2.20: *Scheme of the CMS Level-1 jet and τ trigger algorithm.*

The Level 1 τ trigger exploits a generic jet trigger based only on the calorimetric information [91]. A loose isolation criteria is applied requiring active tower patterns to be made of neighbour towers as shown in Fig. 2.20 and called τ-veto. For each calorimeter region a τ-veto bit is set

ON if there are more than two active ECAL or HCAL towers in the 4×4 region whose shape is identified as one of those illustrated in Fig. 2.20. A jet is defined as τ-*like* if none of the 9 calorimeter region τ-veto bit is on. The desired rate at the Level 1 is reached with a further cut on the calorimetric energy requiring a transverse energy greater than 92 GeV for one τ jet and 64 GeV for two jets. The reconstructed jets that are not identified as τ's are labelled as central jets. The four highest energy central and forward jets, and central τs in the calorimeter are selected. Jet and τ's showing up in a calorimeter region where an electron is identified are not considered.

In order to decrease the E_{T} threshold keeping a good efficiency and an acceptable rate, a logic OR has been made between the τ and jet trigger. The performance in term of the turn on curves obtained during the Run 1 are illustrated in Fig. 2.21.

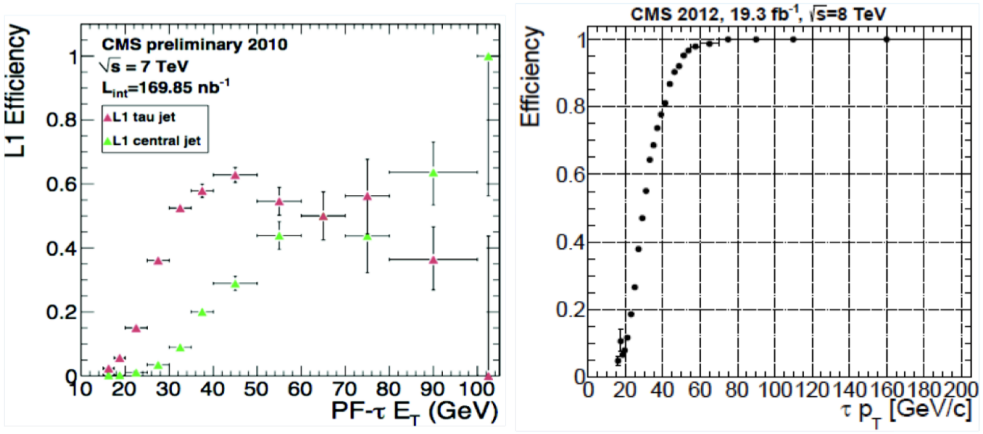


Figure 2.21: *Turn on curves for the L1 τ algorithm (Left) and the L1 τ algorithm in logic OR with the jet one.*

The Level 1 τ trigger shows an intrinsic limitation of its efficiency (60% at plateau) due to the τ veto shapes. Using the L1 τ algorithm combined with the jet one, the efficiency is restored to 100%.

The High-Level Trigger

The HLT reduces the output rate down to $O(100)\,Hz$ [14]. The idea of the HLT software is the regional reconstruction on demand, that is only the objects in the useful regions are reconstructed and the uninteresting events are rejected as soon as possible. This leads to the development of three virtual trigger levels: at the first level only the full information of the muon system and of the calorimeters is used, in the second level the information of the tracker pixels is added and in the third and final level the full event information is available. The use of a processor farm for all selections beyond Level-1 allows maximal benefit to be taken from the evolution of

computing technology. Flexibility is maximized since there is complete freedom in the selection of the data to access, as well as in the sophistication of the algorithms, usually referred to as HLT paths.

CMS performance during LHC Run 1

The CMS experiment has recorded 24.6 fb^{-1} of integrated luminosity during the whole LHC Run 1 data taking period. In particular, 4.9 fb^{-1} have been recorded during 2011 with a collision energy of $\sqrt{s} = 7\,\text{TeV}$ and 19.7 fb^{-1} have been recorded during 2012 with $\sqrt{s} = 8\,\text{TeV}$. From Fig. 2.22 and Fig. 2.23 it is possible to observe respectively the total integrated luminosity recorded by CMS during the Run 1 and the peak instantaneous luminosity recorded per day.

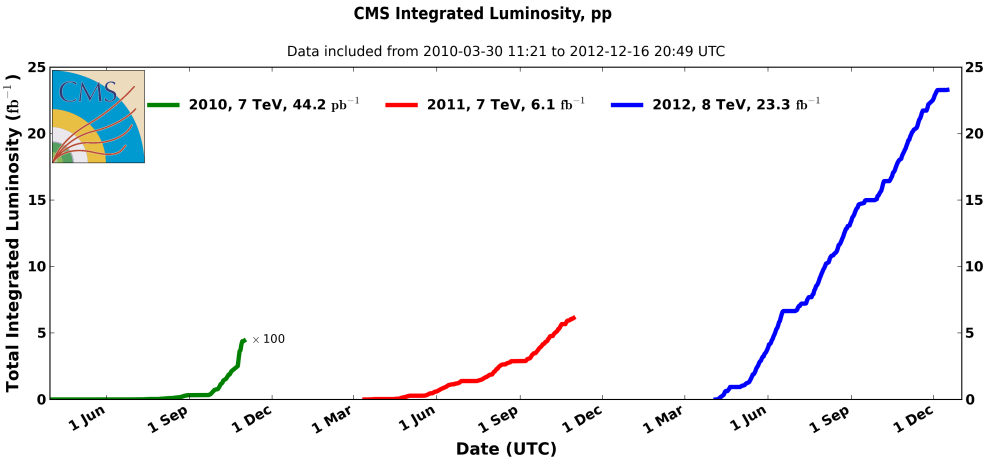


Figure 2.22: *The total integrated luminosity recorded by the CMS experiment during the whole LHC Run 1 data taking period [93].*

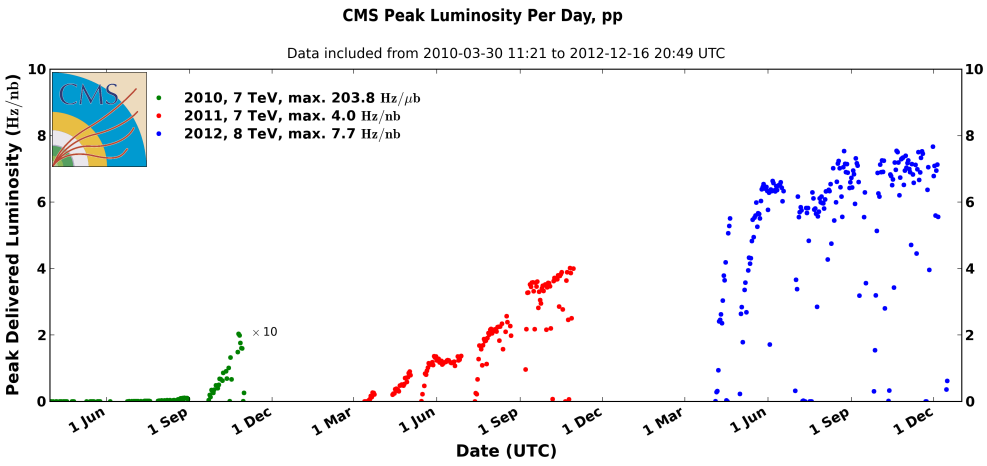


Figure 2.23: *The peak instantaneous luminosity per day recorded by the CMS experiment during the whole LHC Run 1 data taking period [93].*

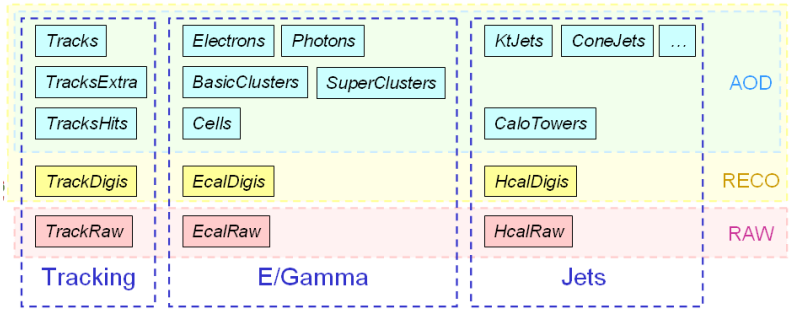
2.4 Data Organization: The CMS Data Hierarchy

CMS Data is organised into a hierarchy of data tiers. Each physics event is written into each data tier, where the tiers each contain different levels of information about the event. The three main data type written in CMS are:

1. **RAW:** Contains all the information from the detector. This kind of data are not useful for analysis;

2. **RECO:** Is the result of the first-pass processing of the raw data. At this level of reconstruction, the events remain too much "heavy";

3. **AOD:** Is a *"pruned"* version of the RECO data. This kind of data are used for most analysis.



Chapter 3

Object reconstruction in CMS

One challenge of a physics analyses at high-energy colliders often consists in reconstructing basic objects like tracks, electrons, muons, photons and higher level objects like jets, taus (τ) and missing transverse energy ($\vec{E}_T$), so as to achieve a description of the collision in terms of the original particles of the underlying physics process (namely quarks, gluons, leptons, photons, hadrons and neutrinos). Various reconstruction algorithms allow the identification of the physics objects used in the CMS analyses. Thanks to the detector design (strong magnetic field, large tracker, granular calorimetry), the CMS detector is well suited for Particle Flow (PF) reconstruction and is widely used in the analyses. In the following chapter a description of the PF, as it is implemented in CMS, is provided together with a more detailed description of the basic objects reconstruction like tracks, interaction vertices, muons and electrons. Afterwards, this chapter treats the reconstruction of the charged and neutral hadrons as well the photon identification performed using the PF algorithm and thus the jet and missing transverse energy reconstruction. The PF is also crucial to properly identify the lepton τ. Given the importance of the τ lepton in my whole thesis work, this topic is treated in more details at the end of the chapter in the Section 3.4. The performance of the PF algorithm in terms of reconstruction efficiency and energy resolution of the considered objects are evaluated both in the simulation and the data.

As described in Chapter 2 the organization of the CMS subdetectors, starting from the interaction point outwards, consists in *i)* a tracker, *ii)* an electromagnetic calorimeter, *iii)* a hadronic calorimeter, *iv)* a muon detector. In order to maximize the performance for a given amount of integrated luminosity, an experiment such CMS, must provide as much as possible a comprehensive description of the final state, i.e. to detect the largest fraction of particles produced in the hard interactions. The visible particles that are stable over the typical detector length scale ($\sim 1\,m$) are electrons, muons, photons, charged and neutral hadrons. The energy and direction of the stable particles coming from a collision can be traditionally measured in two ways: using the tracker that allows to reconstruct the track of the charged particles in the least invasive way, and through the calorimeters. In the latter case, the estimation of the

particle energy come from the energy released by the particle in its destructive interaction with the calorimeter medium.

3.1 Particle Flow

Many analyses rely on the measurement of the jets to assess the momenta of the quarks or gluons; and on the measurement of the missing transverse energy to access the neutrino momenta. Traditionally, these quantities have been measured by the calorimeters. The jet momentum measure could be enhanced by improve the calorimetry in the following ways:

- exploiting the calorimeter granularity to attempt a statistical separation between the hadronic and electromagnetic components of the shower. This approach, named *energy flow*, was historically adopted by the H1 experiment at DESY [94], and was proved to significantly improve the jet energy resolution (from 15% up to 30%);
- an other solution consists in changing of paradigm and reconstruct the particles of the jets individually and cluster them into jets. Considering the particles one-by-one allows the redundancy of the subdetectors to be exploited and to combine their measurements: charged particles measured by the tracker, photons measured by the electromagnetic calorimeter, neutral hadrons measured by the hadronic calorimeter. In this way $\sim 90\%$ of the jet energy can be measured with the superior resolution of the tracker and the electromagnetic calorimeter, leaving to the hadronic calorimeter, notably performing with a worse resolution, the estimation of the energy carried by the remaining $\sim 10\%$ of neutral hadrons. This approach to improve the jet energy resolution is called **particle flow** algorithm.

One of the first implementation of the particle flow algorithm was due to the Aleph collaboration at LEP [95].

3.1.1 The Particle Flow with the CMS detector

The particle-flow event reconstruction aims at reconstructing and identifying all stable particles in the event, with a thorough combination of all CMS sub-detectors towards an optimal determination of their direction, energy and type. The set of individual particles is then processed, as if it came from a Monte Carlo event generator, to build jets, to determine the missing transverse energy $\not\!\!E_T$ (which gives an estimate of the direction and energy of the neutrinos and other invisible particles), to reconstruct and identify τs from their decay products, to quantify charged lepton isolation with respect to other particles, to tag b jets, etc. The CMS detector appears to be almost ideally suited for this purpose. With its large silicon tracker immersed in a uniform axial magnetic field of 3.8 T provided by a superconducting solenoidal coil, charged-particle tracks can be reconstructed with large efficiency and enough small fake rate down to a p_{T} of $150\,MeV/c$, for pseudorapidities as large as ± 2.6 [96]. Most stable particles produced

in proton-proton collisions have a rather low p_{T}, even in collisions involving a large momentum transfer. As an example, in a quark or gluon jet with a total p_{T} of 500 GeV/c, the average p_{T} carried by the stable constituent particles is of the order of 10 GeV/c [96]. This value reduces to a few GeV/c in jets with a total p_{T} below 100 GeV/c. To be able to distinguish the production of particles coming from physically relevant events from the dominating SM background processes, it is therefore essential to accurately reconstruct and identify as many of the final stable particles as possible, some of them with small p_{T} and energies. The PF algorithm uses as basic ingredient tracks collected in the tracker, energy depositions measured in the calorimeter and sophisticated algorithms that link in the best way tracks to energy depositions.

3.1.2 Track reconstruction

The momentum of charged hadrons is measured in the tracker with a much better resolution respect to that of the calorimeters for p_{T} up to several hundreds of GeV/c [96]. For higher energies, however, the calorimeters (in particular ECAL), are more performing in measuring the energy of the particles. Indeed, the resolution on hadron energy measurement is $\sim 100\%/\sqrt{E}$ for calorimeters, while the relative resolution of the tracker for 100 GeV pions is of the order of a couple of percent, hence still better than the calorimeter measurement [97]. Tracks are used to measure the momentum of the charged particles and to determine the production vertex of each of them. Given the multiplicity of particles produced by an LHC pp collision, an efficient track reconstruction in CMS is of primary importance. Moreover a good separation between reconstructed tracks is required to separate the charged constituents inside a jet. In order to reach this goal the CMS experiment is equipped with a performant tracker, described in Chapter 2 and a superconducting coil providing with an intense magnetic field. To optimally exploit the performance of the tracker, a sophisticated track reconstruction algorithm has been conceived. CMS uses an iterative tracking algorithm with subsequent steps picking up inefficiencies from previous steps. The main tracking algorithm is based on pixel seeds and uses a *Kalman* filter method [80] for track finding. The basic idea of iterative tracking is that the initial iteration searches for tracks easy to find (e.g. tracks with relatively large p_{T}). After each iteration, the hits associated to tracks are removed, simplifying the subsequent iteration. The seeding criteria are made looser and looser across the iterations. Each iteration could be summarized in 4 steps:

- A track seed is generated (using 2 or 3 hits) and it defines the initial estimate of the track trajectory
- Track finding algorithm, based on the Kalman filter, extrapolates the seed trajectory along the expected path of a charged particle, searching for additional hits to associate with the track
- A fit of the track provide the best estimation of the track parameters
- Track selection sets quality flags and discards tracks that fail some criteria

The tracking efficiency is excellent for muons and exceeds 99%. For pions it is smaller because of hadronic interaction in the material of the detector: 95% in the barrel and 80-90% in the encaps, while the fake rate (also depending on η and p_{T} is around $\sim$ 1.5% [79]. The tracking efficiency numbers given above are measured directly from data using the tag-and-probe technique in $Z/\gamma^* \to \mu^+\mu^+$ events. The Tag is defined by a muon reconstructed both by in the tracker and in the muon chambers, while the Probe muon is required to be reconstructed only in the muon detectors. In addition, the dimuon invariant mass must be within 50-130 GeV range centred around the value of the Z boson mass [90]. In Fig. 3.1 the measured efficiency, defined as the fraction of the Probe muons associated with a track reconstructed in the tracker, is shown as a function of the muon pseudorapidity and the number of the reconstructed primary vertices, both for data and simulated events.

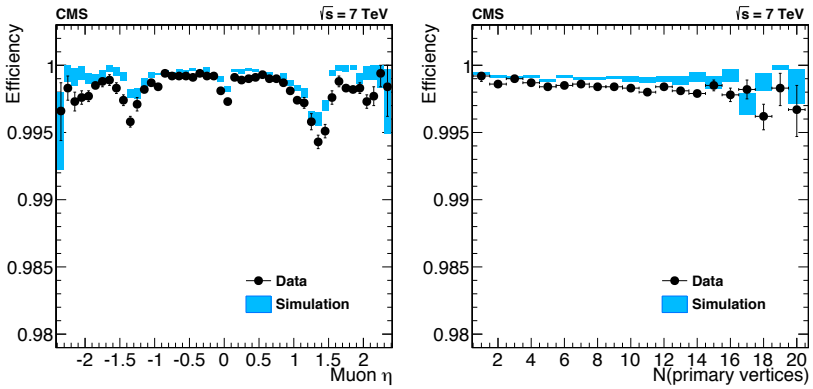


Figure 3.1: *Tracking efficiency measured with tag-and-probe technique, for muons coming from Z decays, as function of the muon η (Left) and the number of reconstructed primary vertices in the event (Right) for data (black dots) and simulation (blue bands) [79].*

The agreement between the data and the simulation is good and the small discrepancies observed are due to limitations in the tracking modelling. The measured efficiency is $\sim$ 99% over the whole pseudorapidity region and it is found to be almost insensitive to the pile-up activity.

3.1.3 Calorimeter Clustering

In addition to the tracks, another key ingredient of the PF algorithm is the energy deposition of the particles in the calorimeters. As the energy of the particles is always deposited in several crystals and most of the time in more than HF module, a clustering of the energy in the cells is needed. In the context of the PF, the purpose of the clustering is threefold [96]:

- detect and measure the energy and direction of stable neutral particles such as photons and neutral hadrons
- separate these neutral particles from energy deposits from charged hadrons
- help the energy measurement of charged hadrons for which the track parameters were not determined accurately, which is the case for low-quality, or high-p_{T} tracks

To reach this goal is possible only adopting calorimeters with enough granularity to provide an efficient energy collection in a wide pseudorapidity region together with an acceptable angular resolution. As a consequence, algorithms capable to exploit in the best way the calorimeters granularity are needed. In what follows are listed the most important steps adopted by the calorimeter clustering procedure for the particle flow algorithm developed in CMS [96]:

1. If an energy deposition in a calorimetric cell exceeds a fixed threshold it is identified as cluster seed. A seed is local maximum above a given threshold
2. topological clusters are grown from the seeds by aggregating cells with at least one side in common with a cell already in the cluster, and with an energy in excess of a given threshold: $2\sigma_{ECAL\ electronic\ noise} \simeq 80\,MeV$ (in the barrel) or $2\sigma_{HCAL\ electronic\ noise} \simeq 800\,MeV$;
3. A topological cluster gives rise to as many PF clusters as seeds
4. At the end, the calorimeters granularity is exploited to disentangle overlapping showers:
   - The energy of a cell can be shared between two clusters, the fraction of a cell energy contributing to a cluster is proportion to $\sim exp(-d^2/R^2)$ where d is the distance to the cluster barycentre and R is 5 cm (resp. 10 cm) in the ECAL (resp. HCAL).
   - As the position of the cluster is calculated depends on the energies (or fraction of energies) of the cells, an iterative procedure is applied. In practice, $\sim$ 4 iterations are sufficient.

3.1.4 Link algorithm

As mentioned earlier, the signatures of a particle in the different subdetectors should be combined for an optimal identification and measurement. It implies properly connecting together the so-called particle flow elements reconstructed in the different subdectors, i.e: tracks with clusters, ECAL clusters with HCAL clusters, etc. The link algorithm is tentatively performed for each pair of elements in the event and defines a distance between any two linked elements to quantify the quality of the link [1] [96]. The linking algorithm produces *blocks* of elements linked directly

[1] The link distance is defined as the distance in the (η, ϕ) plane between the extrapolated track position and the cluster position

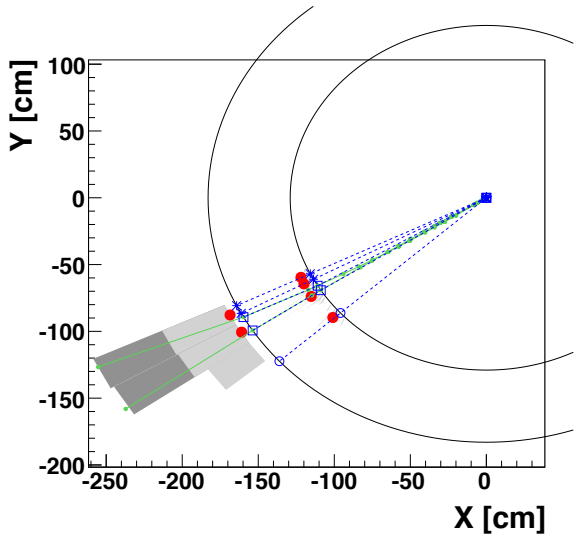


Figure 3.2: *An event display of a simple hadronic jet in the* (x, y) *plain [96].*

or indirectly. Thanks to the granularity of the CMS detectors, blocks typically contain only one, two or three elements, and constitute simple inputs for the particle reconstruction and identification algorithm. The smallness of the blocks ensures the algorithm performance to be essentially independent of the event complexity. The link algorithm proceeds as follow:

1. The track is first extrapolated from its last measured hit in the tracker to:
   - the preshower
   - the ECAL, to the expected barycenter of an electromagnetic shower
   - the HCAL, at a depth corresponding to one interaction length, typical of an hadron shower
2. The track is linked to any given cluster if the extrapolated position in the corresponding calorimeter is within the cluster boundaries
3. Similarly, a link between two calorimeter clusters, i.e., either between an HCAL and an ECAL cluster, or between an ECAL and a ES (preshower) cluster, is established when the cluster position in the more granular calorimeter (ES or ECAL) is within the cluster envelope in the less granular calorimeter (ECAL or HCAL)
4. link between a charged particle track in the tracker and a muon track in the muon system is established (and is called a global muon) when a global fit between the two tracks returns an acceptable χ^2 [2]. For deeper details see the analysis note [96] and [98] for

[2]This χ^2 defines the link distance in that case

analysis performed with real data.

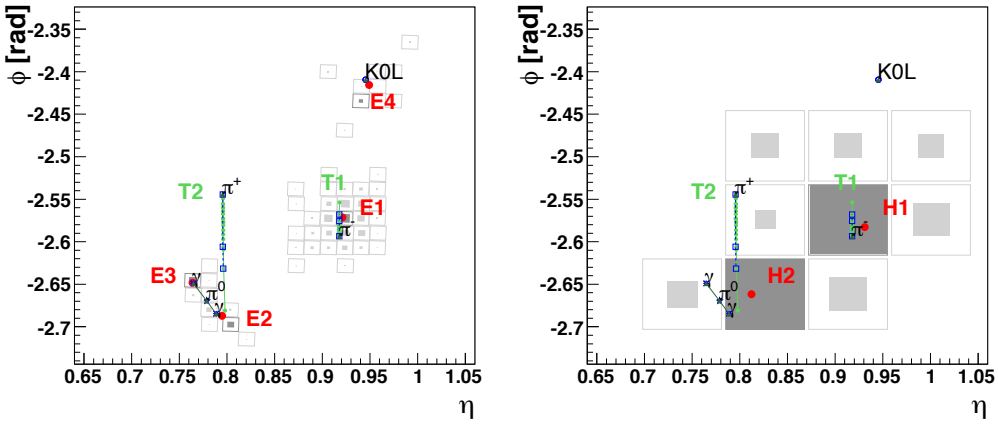


Figure 3.3: *An event display of a simple hadronic jet in the (η, ϕ) view, where η stands for pseudo-rapidity and ϕ for the azimuthal angle, on the ECAL surface (Left) and the HCAL surface (Right). (These two surfaces are represented as two circles centred around the interaction point in the Fig. 3.2). The K_L^0, the π^- and the two photons from the π^0 decay are detected as four well separated ECAL clusters (Left). The π^+ leaves no energy in the ECAL. The two charged pions are reconstructed as charged-particle tracks, appearing as vertical solid lines in the (η, ϕ) views and circular arcs in the view of Fig. 3.2. These tracks point towards two HCAL clusters (Right). In all three views, the cluster positions are represented by dots, the simulated particles by dashed lines, and the position of their impact on the calorimeter surfaces by various open markers [96].*

Once the links between all the elements have been established, one obtains blocks of elements which are connected one with each other. By definition, two blocks are completely independents. Ideally, one would have a bijection between a particle and a block. In practice, due to the overlaps, several particles can end up in a block.

The second part of the particle flow acts on one block at a time and is able to identify the following particle:

- **muons**: Each global muon fulfilling a set of optimized identification criteria [90] gives rise to a PF-Muon, if its combined momentum is compatible with that determined from the sole tracker within three standard deviations [96];

- **electrons**: The reconstruction of the electrons is a difficult task in CMS, and is even more difficult with the particle flow. Many Bremsstrahlung photons can be emitted because of the tracker material, and they can then convert. With the large magnetic field, they impinge the ECAL in large region along ϕ but narrow along η. Each of the particles of the

shower thus initiated in the tracker gives, most of the time, rise to an individual cluster. These clusters must absolutely be recollected and identified as coming from the electron, otherwise their energy will be double-counted with that of the track. A dedicated reconstruction, able to follow the trajectory of the electrons despite the changes of curvature, and making use of the Gaussian Sum Filter Algorithm is used and a Boosted Decision Tree (BDT) is finally used to disentangle electrons from pions

- **charged hadrons**: A track can be directly connected to a number of ECAL and HCAL clusters. The detection of the neutral particles in the block (photons and neutral hadrons) involves a comparison between the momentum of the tracks and the energy detected in the calorimeters [98]. For this comparison to be reliable, the ECAL and HCAL cluster energies, from which the expected muon energy deposits are subtracted, must undergo the calibration procedure, as reported in [96]. The energy of charged hadrons is obtained from a combination of the tracker and calorimeter measurements, when the two values are found to be compatible, in order to prevent the energy of charged hadrons from being counted twice (from the momentum of the track measured in the tracker and from the energy deposited in the calorimeters) [98]. If the energy measured in the calorimeters is small compared to the track momentum, a cleaning procedure to remove potential spurious or mis-reconstructed tracks is invoked [96]. If, instead, the calorimeter response is too large,at more than one sigma of the expected resolution, the particle-flow algorithm assigns the energy excess to a photon and possibly a neutral hadron

- **photons**: As mentioned, photons can be created when comparing the amount of calibrated calorimeter energy with the momentum of the corresponding track(s). But photons are also built from isolated ECAL clusters. Similarly, isolated clusters in the HCAL give rise to neutral hadrons.

The electron and muon reconstruction is of main importance in many physics analysis, such as standard model precision measurements, characterization of the Higgs sector and searches for physics beyond the standard model. In particular the electron reconstruction, as the muon one, played a leading role in the analysis of the $H \to ZZ^* \to 4\ell$ that provided one of the major contribution to the discovery of the new boson in 2012 [12]. The main requirement of such analysis is an excellent electron and muon reconstruction and selection efficiency over a large phase-space, an excellent electron and muon momentum resolution together with the smallest as possible misidentification probability.

3.1.5 Muon reconstruction

The CMS detector has been designed to have an excellent muon identification efficiency and momentum (for muons you measure tracks) resolution. Muons have a long life-time ($c\tau = 659$ m) and do not interact through nuclear force and, contrary to electrons, the bremsstrahlung

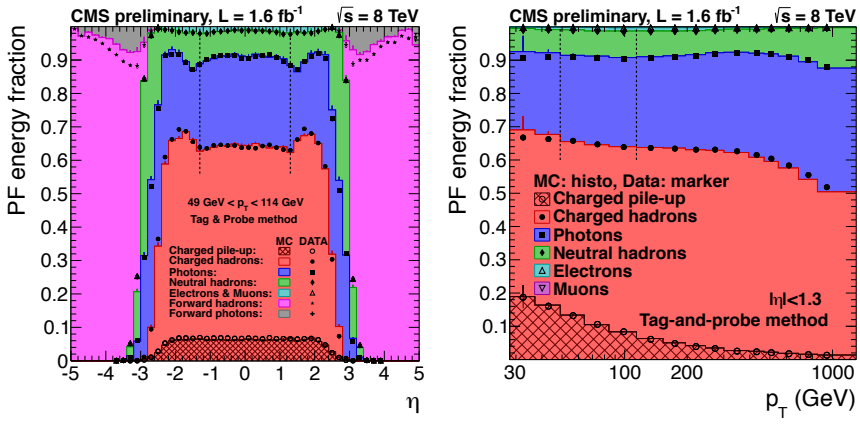


Figure 3.4: *Jet energy fraction as a function of η (Left) and p_T (Right). This plot are obtained with data collected by CMS at $\sqrt{s} = 8\,TeV$ [99].*

emission along their trajectory in the detector material is weak. Compared to electrons it is reduced by a factor $(m_e/m_\mu)^2$. For this reason, muons are capable to traverse the whole CMS detector before decaying via weak interaction. In CMS different approach to reconstruct the muons are adopted [90]:

- **Global muon reconstruction** (outside-in): From a standalone muon reconstructed in the muon chambers a matching to a track reconstructed in the tracker is required. If this track exists, a fit to the muon trajectory is performed using the hits in the tracker and the hits/segments in the muon system with the Kalman filter technique [100][101][102]. At large transverse momenta ($p_T > 200\,\text{GeV}/c$) the global muon fit can improve the momentum resolution respect to the sole tracker measurement

- **Tracker muon reconstruction** (inside-out): In this approach, all tracker tracks with $p_T > 0.5\,\text{GeV}/c$ and total momentum $p > 2.5\,\text{GeV}/c$ are considered as possible muon candidates. The selected tracks are thus extrapolated to the muon system taking into account the expected energy loss and the uncertainty coming from multiple scattering. If at least one muon segment made of DT or CSC hits is matched, the track is classified as tracker muon track. The advantage of this reconstruction algorithm is that it allows a higher muon identification efficiency (since it requires only one muon segment in the muon detectors) and a better energy resolution for low momentum ($p_T < 5\,\text{GeV}/c$) muons

- **Standalone muon**: This procedure is dedicated to reconstruct the muons that fails the global muon and tracker muon reconstruction. Only $\sim 1\%$ of the total reconstructed muons from collisions enter this category. However, higher contamination from cosmic rays leads to a misidentification rate (cosmic-muons/collision-muon) that is $10^4 - 10^5$ higher than the one observed using the other muon reconstruction techniques

Thanks to the different algorithms implemented in CMS to reconstruct muons, different categories are defined in order to identify genuine muons with increasing purity:

- **Soft Muon Selection**: The muon candidate is required to be reconstructed as tracker muon with the additional requirement that a muon segment is matched both in x and y coordinate with the extrapolated tracker track.
- **Tight Muon Selection** The muon candidate must be reconstructed as global muon with a track $\chi^2/d.o.f. < 10$ and at least one muon chamber hit included in the global-muon track fit. In addition, its correspondent tracker track is required to match a muon segment in at least two muon stations and using more than 10 inner-track hits (including at least one pixel hit) and has an impact parameter less then $2\,mm$ respect to the primary vertex.

The tight identification criteria are used in the analysis to select muon candidate coming from the leptonic decay of τ lepton. The muon reconstruction and identification efficiency is measured using real data applying the tag-and-probe technique on dimuon resonances (e.g. $J/\psi \rightarrow \mu\mu$ and Z$\rightarrow \mu\mu$ decays) [90]. Events are selected using tight requirements to define the tag and looser selection to identify the probe. In addition a dimuon invariant mass window, centred to the mass of the resonance, is imposed. Figure 3.5 shows the muon efficiency as function of the muon p_{T} and it is possible to see that the efficiency value at the plateau of the distributions for both soft muon and tight muon category is higher than 98% and a good agreement between data and simulation is observed.

The resulting collections of candidates are used as input for the identification of the PF muons. The PF algorithms starts by muons because of the high purity in the muon reconstruction. Thus an additional collection is available:

- **Particle Flow Muon**: This collection is made of all global muons reconstructed with at least one hit in the muon system and with the transverse sum of all the neighbouring tracks and calorimetric cells around a cone of $R = 0.3$ centred on the muon itself is less then 10% of the muon p_{T}. This additional requirement assures that candidates entering in the PF muon collection provide a sample of very high purity real muon

3.1.6 Electron reconstruction

One of the main challenge in the electron reconstruction lies in the fact that the electron trajectories, bended by the magnetic field, cross multiple layers of the tracker material with the consequent loss of an important fraction of their energy, irradiated through bremsstrahlung radiation. Electrons are reconstructed associating a track with a cluster of energy in the electromagnetic calorimeter [83]. The electron energy is spread out over several ECAL crystals. The number of crystals depends on the amount of bremsstrahlung and of the position of the photon emission. The fraction of energy loss could vary from $\sim 33\%$, in the central barrel region up to $\sim 86\%$ in the region where the material budget is maximal. To have a correct estimate

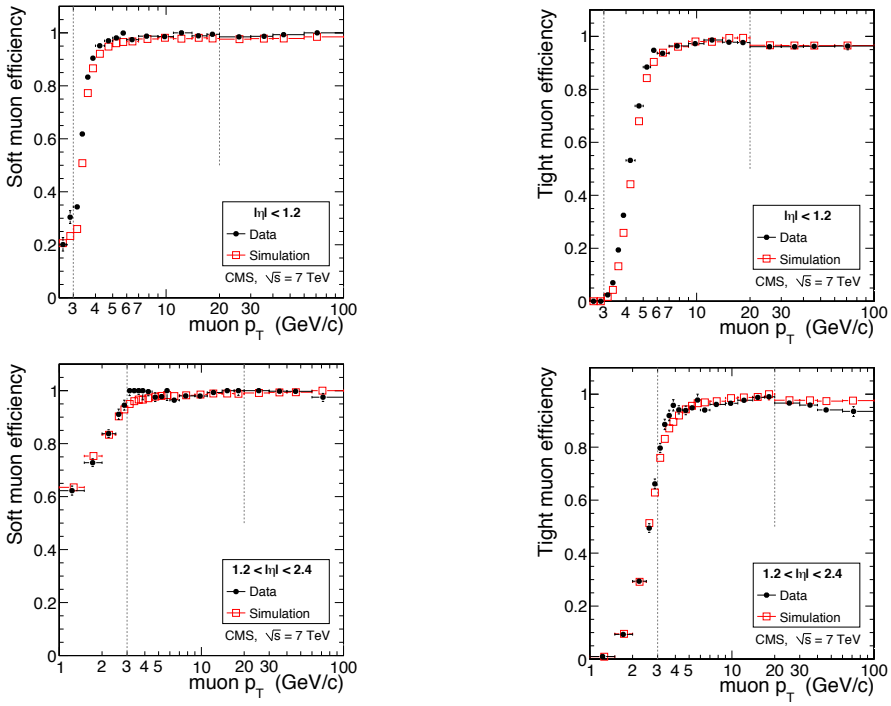


Figure 3.5: *Muon efficiency as function of the muon transverse momentum for soft (Left) and tight (Right) muon categories evaluated in data using the tag-and-probe-technique (black dots) and compared to simulation (red dots) for two different region (barrel and endcap) of the muon detectors [90].*

of the electron energy is thus mandatory to recover the energy carried by the bremsstrahlung photons, that mainly spread in the ϕ direction because of the bending of the electron in the CMS magnetic field.

Cluster reconstruction

The electron reconstruction algorithm exploits the ECAL granularity and properties relative to electron shower shape to collect the electron energy. As mentioned in Section 3.1, the energy deposits of the electrons located in a narrow band in η, and spread in ϕ must be recollected to form a super-cluster (see Fig. 3.6).

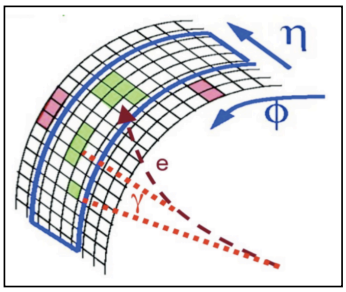


Figure 3.6: *Electron radiating bremsstrahlung photons in CMS.*

Two different approaches are used. The first one is purely ECAL-driven, and is progressively aggregating energy, starting from a side, and moving along phi in one direction and then in the other, thus forming a supercluster. The other approach is using the information of the electron track, and is emulating the emission of bremsstrahlung photon for each tracker layer by making a straight line extrapolation tangent to the trajectory and up to the ECAL inner surface. The individual clusters matching these extrapolations constitute a supercluster. The latter approach is done within the particle flow algorithm, but in the end the electron object results from a merging of the two.

Track reconstruction

The tracking plays an important role in the electron reconstruction and identification but the large radiative losses for electrons in the tracker material compromise the performance of the general track reconstruction algorithm described in section 3.1.2. A dedicated algorithm, the Gaussian Sum Filter [103], has been instead developed to reconstruct the electron tracks. Since this algorithm is very computing time-consuming, the pattern recognition procedure starts only on preselected track seeds. There exist two algorithm to select the tracks to be processed by the GSF:

- **ECAL driven method**: The ECAL-driven methods uses the superclusters reconstructed with the sole ECAL information. As these superclusters are supposed to collect the entire electron energy deposit, two back-propagations ($q = \pm e$) are carried out to find tracker seeds compatible with the supercluster energy and position. This algorithm is applied to superclusters fulfilling $E_{\mathrm{T}}^{SC} < 4\,\mathrm{GeV}$ and relatively isolated, i.e. $H/E < 0.15$ where H is the HCAL energy in a $\Delta R = 0.15$ cone behind the supercluster

- **Tracker driven method**: This method is developed as part of the PF algorithm and aims to recover the limitation of the ECAL-based seeding, especially for low electron p_{T}. Indeed, electrons inside jets could not be covered by the ECAL-driven track seeding since the hadronic veto. In addition, low p_{T} electrons are heavily deflected by the magnetic

fields and the supercluster may not be able to collect properly all its energy irradiated along the ϕ direction

The two collections are thus merged and used to seed the GSF algorithm. More details regarding the electron tracking procedure can be found in [83]. The track building provides electron trajectory that can be extrapolated to the ECAL surface. The fraction of the energy lost via bremsstrahlung radiation is then estimated using the formula: $f_{brem} = \frac{p_{in} - p_{out}}{p_{in}}$, where p_{in} and p_{out} are respectively the momentum of the electron measured by the tracker at the closest point to the beam spot and at the closest point to the ECAL. The fraction of bremsstrahlung (f_{brem}) is measured using $Z \to e^+e^-$ and $Z + jet$ enriched data and compared to the simulation for different η region as shown in Fig. 3.7. As it is possible to infer from the shape distributions, f_{brem} is a quantity that could be used to discriminate the pure electrons from the misidentified ones. Indeed, jets misidentified as electrons are composed by charged hadrons whose tracks fake electron ones, but, because of the greater hadrons masses compared to the electron one, the emission of bremsstrahlung radiation is highly suppressed.

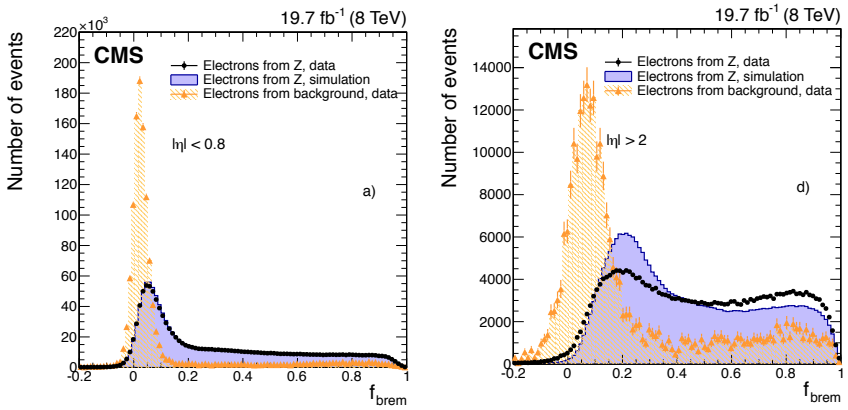


Figure 3.7: *Distribution of f_{brem} for electron coming from the Z boson decays in data (black dots) and in simulation (solid histograms) for different region in η: central barrel and forward endcap. The differences between data and simulation, arising especially in the high-η regions, are due to imperfect modelling of the material in the simulation [83].*

Estimation of the electron momentum

The electron momentum is estimated by combining the ECAL supercluster energy and the momentum measured by the tracker, in a formula that results:

$$p_{comb} = w \cdot p + (1 - w) \cdot E_{SC} \tag{3.1}$$

where E_{SC} is the energy of the supercluster, p is the momentum evaluated from the track and w is a weight that is computed using a multivariate analysis regression technique [104]. It is possible to see in Fig.3.8 the relative energy resolution from the simulation for two category of electrons: *Golden electron* and *Showering electron*. The first are electrons that irradiate a very small quantity of their energy via bremsstrahlung and the latter are electrons that irradiate a consistent fraction of their energy along their trajectory. The combination of the tracker and

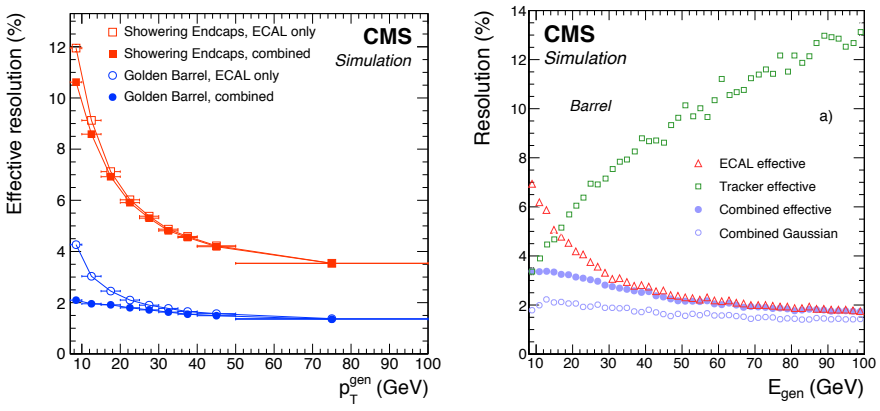
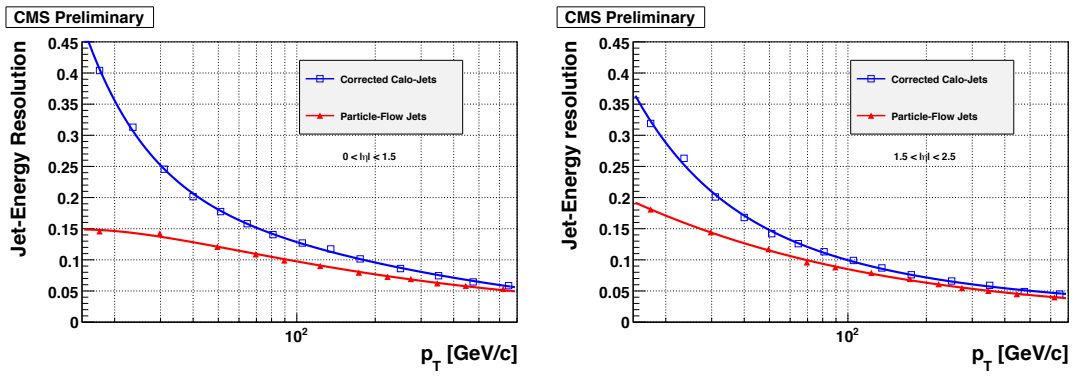


Figure 3.8: *Relative energy resolution as function of the generated electron p_T for Golden and Showering electron categories estimated with the ECAL only and after the combination between ECAL and track informations [83].*

ECAL information lead to an improvement of $\sim 25\%$ in the relative energy resolution for electron of $\sim 15\,\text{GeV}$.

3.2 PF jet reconstruction and performance

In CMS the jets are reconstructed using the $anti-k_T$ algorithm [105] [106] with different distance parameters, in a fiducial region that extends until $|\eta| = 4.7$. The PF jets are reconstructed using the total set of particle already reconstructed by the PF [96][98] and described in the previous section of this chapter. Using the PF algorithm the jet momentum and spatial resolution are significantly improved with respect to the so-called *calo-jet*, where only the information coming from the calorimeters are exploited. This is due to the use of the tracker and high granularity of the ECAL to reconstruct the charged constituents of the jet. The jet energy is carried by three main components: charged particles, photons and neutral hadrons, respectively components of the jet for the $65\%, 25\%$ and 10% [98], Fig. 3.4. These fractions ensure that 90% of the jet is reconstructed with good precision by the PF algorithm, both in energy and direction, while only 10% of the energy is affected by the poor hadron calorimeter resolution [14] [89].



(a) Barrel (b) End-Caps

Figure 3.9: *Jet energy resolution as a function of p_{T} for calo-jets (open squares) and for particle-flow jets (upwards triangles) in the barrel (a) and in the end-caps (b). The resolution curves are fit to the sum of a constant term, a stochastic term and a noise term [96]*

Jet reconstruction: The performance of PF jets are evaluated using a Monte Carlo QCD-multijet event sample and the reconstructed jets are then matched to the closest jet at generator level within a cone of $\Delta R < 0.5$ in the (η, ϕ) plane. The matching jet efficiency, i.e., the fraction of generated jets that give rise to a matched reconstructed jet, and the mismatched jet rate, i.e., the fraction of reconstructed jets that do not have a matched generated jet, are shown in Fig. 3.10 as a function of the jet p_{T}, in the barrel and in the endcaps. An efficiency larger than 80% is obtained for jets with a $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$. The 100% plateau is reached above $40\,\mathrm{GeV}/c$, at which point the mismatched jet rate is negligible.

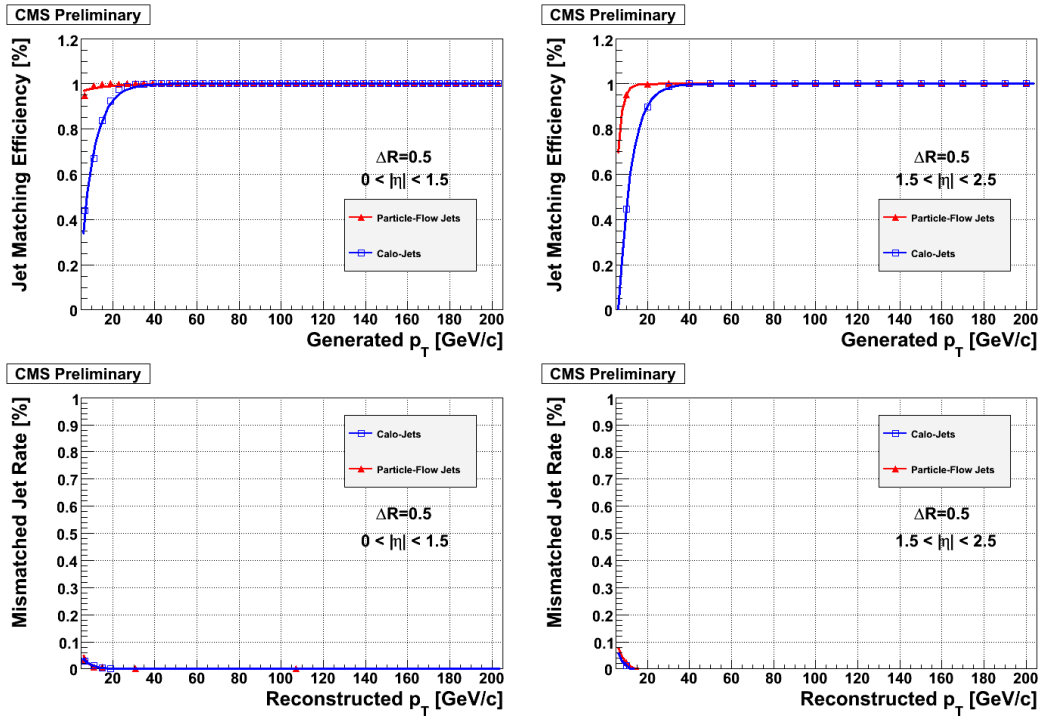


Figure 3.10: *Jet matching efficiency (a,b) and mismatched jet rate (c,d) as a function of the jet* p_{T}*, as obtained for calo-jets (open squares) and particle-flow jets (triangles) reconstructed in the barrel,* $|\eta| < 1.5$ *(a,c) and in the endcaps,* $1.5 < |\eta| < 2.5$ *(b,d), with a matching distance of 0.1 in the* $\eta - \phi$ *plane. Efficiencies and fake rates are fit to exponential functions of* p_{T} *[96].*

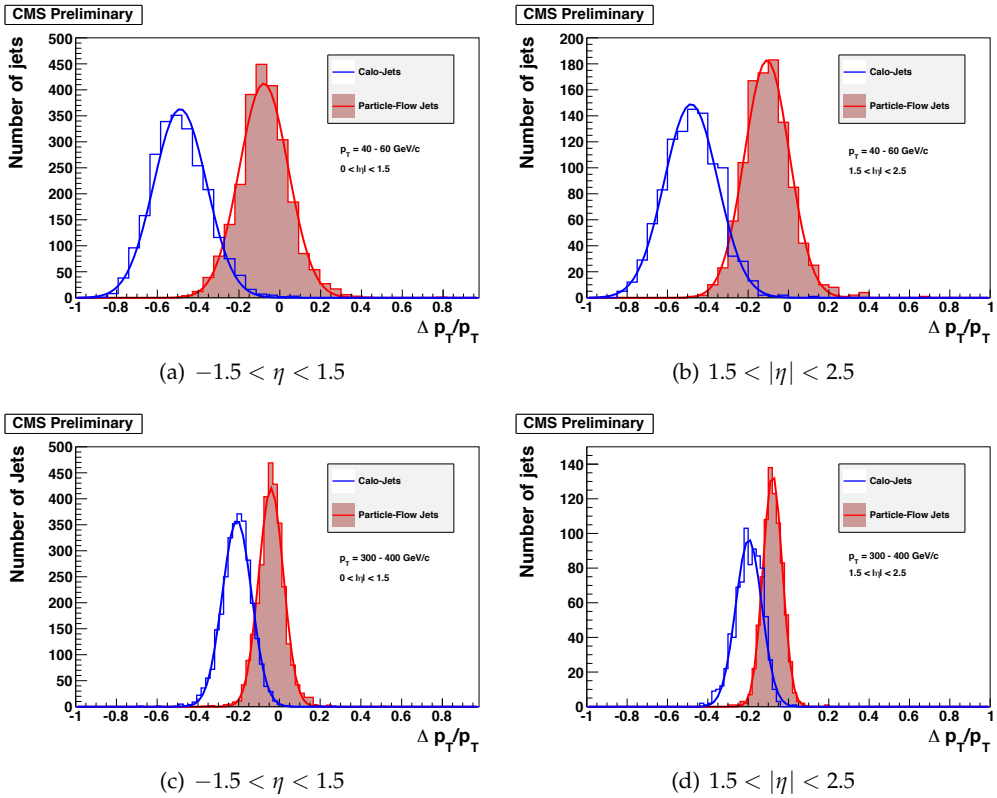
Jet energy resolution : A good jet energy and angular resolution is important to obtain a good resolution on the determination of the di-jet invariant mass:

$$m_{inv}^2 = 2p_{T_{jet1}}p_{T_{jet2}} \cdot (\cosh(\eta_{jet1} - \eta_{jet2}) - \cos(\phi_{jet1} - \phi_{jet2})) \tag{3.2}$$

The distributions presented in Fig. 3.16 compare the jet reconstructed momentum with the corresponding MC generated one:

$$\frac{\left(p_{\mathrm{T}}^{rec} - p_{\mathrm{T}}^{gen}\right)}{p_{\mathrm{T}}^{gen}} \tag{3.3}$$

and are fit by a Gaussian in each (η, p_{T}) bin, with central value μ and width σ. As it is possible to observe in Fig. 3.16 the response defined in equation (3.3) is much more close to zero using PF jets instead the calo jets, this means that the estimation of the jet energy performed with the PF algorithm is closer to the one expected from the simulation. Also the energy resolution of the PF jets as function of the true jet p_T is significantly improved compared to the one obtained using the calo-jets, as it is shown in Fig. 3.9.



(a) $-1.5 < \eta < 1.5$ (b) $1.5 < |\eta| < 2.5$

(c) $-1.5 < \eta < 1.5$ (d) $1.5 < |\eta| < 2.5$

Figure 3.11: *Energy resolution for $p_{T_{gen}}$ between 40 and 60 GeV/c (a,b) and between 300 and 400 GeV/c, (c,d) as obtained from calo-jets (open histogram) and from particle-flow jets (solid histogram) pointing to the barrel (a,c) and to the endcaps (b,d). A Gaussian distribution is fit to all the histograms, to determine the mean value of the response and its resolution [96].*

Jet energy corrections

By construction, the PF jets have thus an energy scale already very close to unity with respect to the corresponding jet made of the stable visible generated particles, hence they would need only small residual corrections, see Fig. 3.12. These corrections are derived, however, through a complex procedure [107]. The purpose of the jet energy calibration is to relate, on average, the energy measured for the reconstructed jet to the energy of the originating parton. corresponding true particle jet. The correction is applied as a multiplicative factor C to each component (μ) of the raw jet four-momentum vector:

$$p_\mu^{corr} = \mathcal{C} \cdot p_\mu^{raw} \tag{3.4}$$

The correction factor $\mathcal{C}$ is composed of the offset correction C_{offset} (or L1 correction), the Monte Carlo calibration factor C_{MC} (L2 and L3) , and the residual calibrations C_{rel} and C_{abs} for the

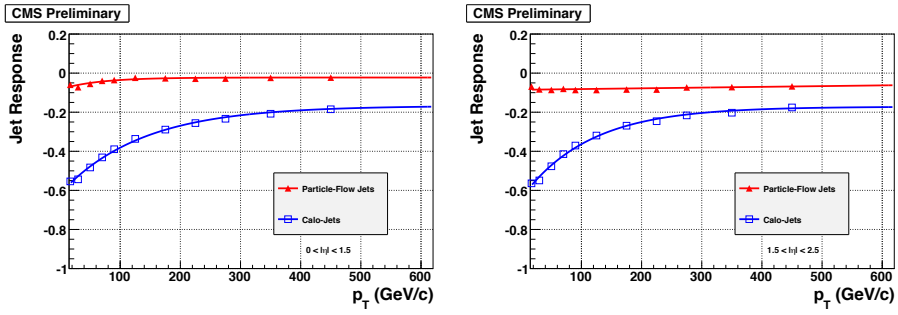


Figure 3.12: *Relative jet energy response for calo-jet (blu line) and PF jet (blue line) evaluated in the barrel (Left) and in the Endcap (right) [96].*

relative (L2res) and absolute (L3res) energy scales respectively [107]. A sketch of the jet energy corrections chain is illustrated in Fig. 3.13.

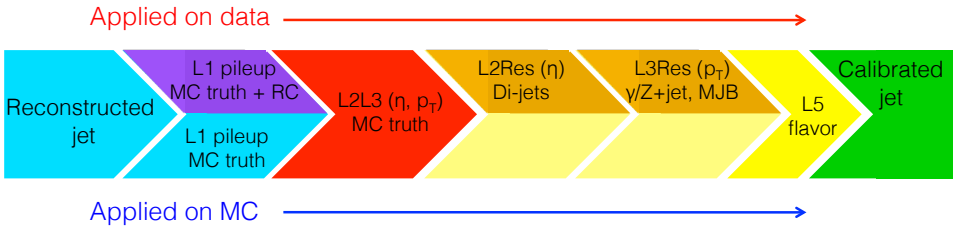


Figure 3.13: *Sketch of the various steps that contribute to the full jet energy correction. Different step were adopted to correct jet in data and in the simulations.*

The L1 correction removes the extra energy due to noise and pile-up, and the Monte Carlo correction removes the bulk of the non-uniformity in η (L2) and the non-linearity in p_{T} (L3). Finally, only in case of reconstructed jet in data, the residual corrections L2Res and L3Res account for the small differences between data and simulation. The various components are applied in sequence as described by the equation below:

$$\mathcal{C} = \mathcal{C}_{offset}(p_{\mathrm{T}}^{raw}) \cdot \mathcal{C}_{MC}(p_{\mathrm{T}}\prime, \eta) \cdot \mathcal{C}_{rel}(\eta) \cdot \mathcal{C}_{abs}(p_{\mathrm{T}}\prime\prime) \tag{3.5}$$

where $p_{\mathrm{T}}\prime$ is the transverse momentum of the jet after applying the offset correction and $p_{\mathrm{T}}\prime\prime$ is the p_{T} of the jet after all previous corrections. The L1 and the Monte Carlo L2L3 jet energy correction factor as function of the jet η is shown in Fig. 3.14 for different pile-up scenario and for different reconstructed jet p_{T} respectively [108].

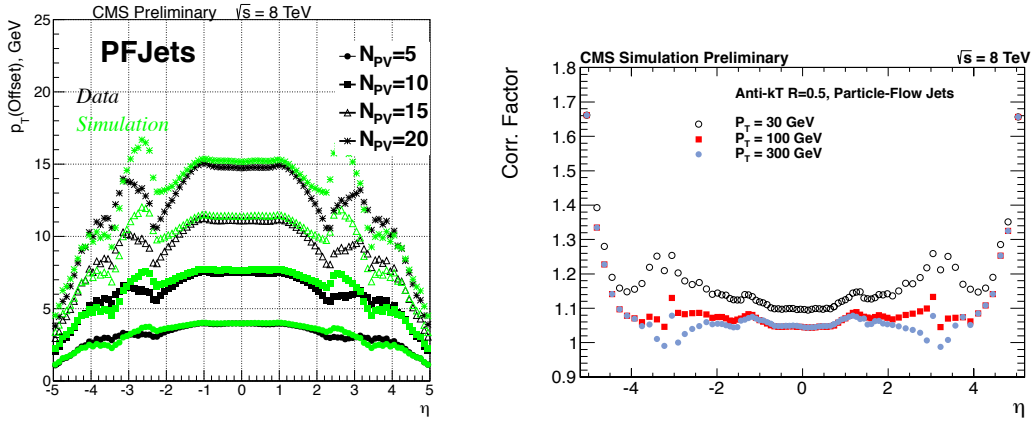


Figure 3.14: *L1 jet energy correction (JEC) factor as a function of jet η for different pile-up scenario, evaluated in data and in the simulation (Left). Monte Carlo L1L2 jet energy correction as a function of jet η for different value of the jet p_{T}: 30 GeV, 100 GeV and $p_{\mathrm{T}} = 300$ GeV (Right) [99].*

The uncertainties associated to the full energy scale correction shown in Eq. 3.5 have been evaluated on data and they are shown in Fig. 3.15.

In collisions with a large number of pile-up interactions, fake jets can be reconstructed from the accidental clustering of many neighbouring particles, or from the superposition of soft jets from different vertices. In order to distinguish between jets coming from the production vertex and those arising from soft interactions a multivariate analysis technique is applied. The input variables of the Boosted Decision Tree [104] implemented for this purpose are the compatibility of the tracks of the jet constituent with the primary vertex, the jet shape variables and the multiplicity of charged and neutral components within the jets. The jet p_{T} resolution as function of the jet p_{T} for jets reconstructed in the central barrel region is shown in Fig. 3.16. From the same plots it can be seen that a good agreement between the data and the simulation is achieved.

Also the jet reconstruction efficiency is evaluated directly from data selecting $Z/\gamma^* \rightarrow \mu\mu + jets$ events with a jet $p_{\mathrm{T}} > 25\,\mathrm{GeV}/c$ and it is shown in Fig. 3.17 as function of the number of reconstructed primary vertices and the jet transverse momentum. It is possible to see that the reconstruction efficiency is pile-up and p_{T} independent.

Finally, it should be mentioned that the PF jets are used as inputs to the b-tagging algorithms. In particular, b-hadrons are identified as jet presenting a reconstructed vertex shifted respect to the primary vertex because of the long lifetime of this particles.

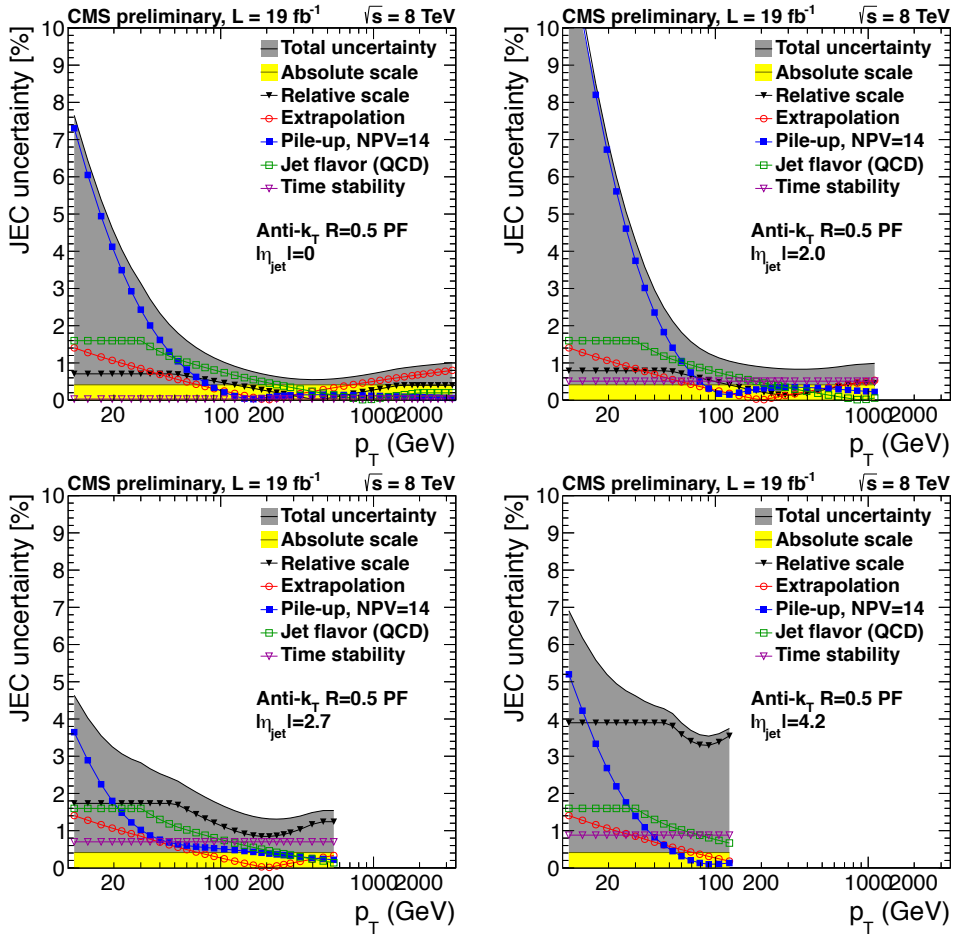


Figure 3.15: *Total jet-energy-scale uncertainty, as a function of jet p_T for various jet η values: η =0, 2, 2.7, 4.2. Different contributions are shown with markers of different colors, and the total uncertainty is shown with a grey band [107].*

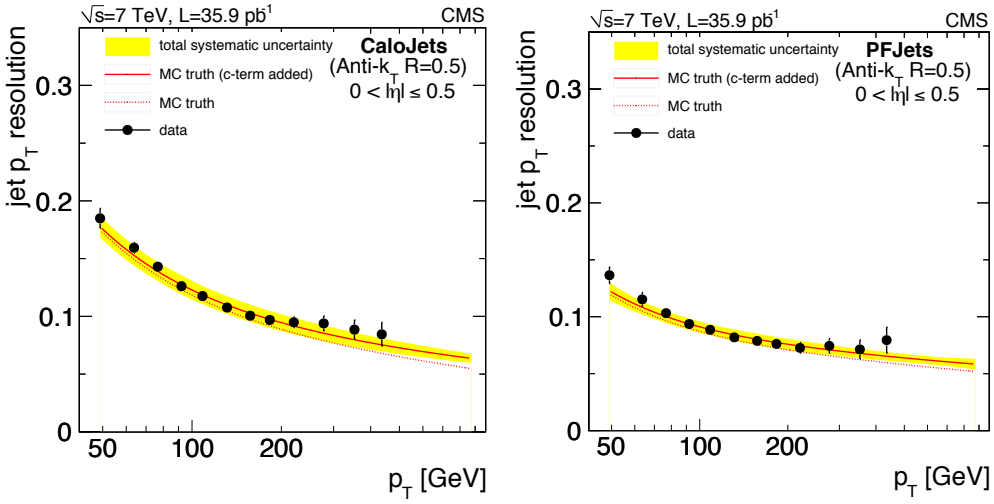


Figure 3.16: *Jet transverse momentum relative resolution measured with data recorder in 2010 with a* $\sqrt{s} = 7\,TeV$ *(black dots) compared to Monte Carlo simulation (red line) as function of jet* p_T *for calo-jets (Left) and PF jets (right) [107].*

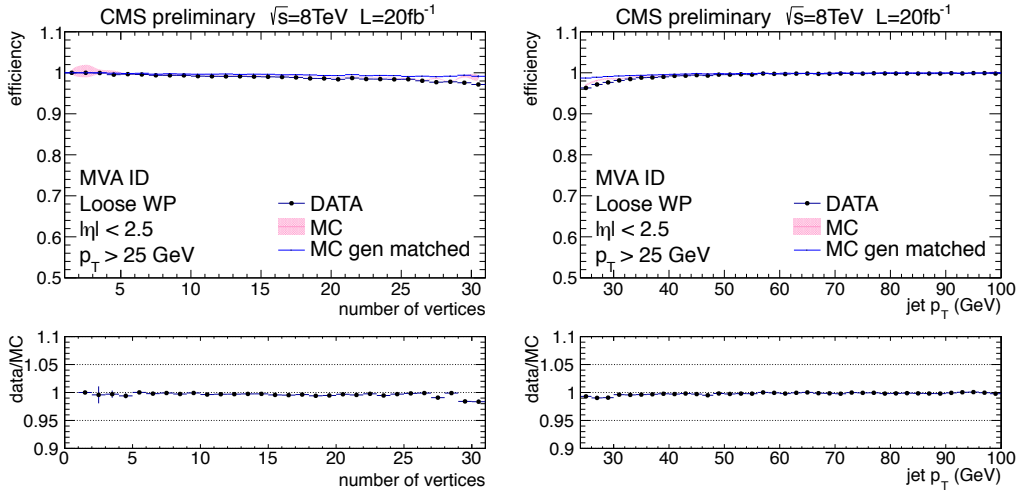


Figure 3.17: *Data to simulation comparison of the MVA (loose working point) pile-up jet identification efficiency versus the number of primary vertices (Left) and the jet* p_T *(Right) evaluated using* $Z/\gamma^* \rightarrow \mu\mu + jets$ *events for jets with* $p_T > 25\,GeV/c$ *[107].*

3.3 PF-$\vec{\not{E}}_T$ reconstruction and performance

The CMS detector has an almost hermetic structure (excluding the region close to the beam pipe and the gaps between the various subdetectors and between the different modules of each

subdetectors), but the detection of a particle depends largely on what kind of interaction this particle has with the material of the detector. There are some particles that interact so weakly with matter that can be practically considered invisible. Such particles can be neutrinos or even new neutral, weakly interacting, exotic particles. One important aspect in the complete understanding of an event produced in a *pp* collision is to have the most precise as possible measure of the particles escaping the detector, e.g. neutrinos, both in direction and momentum. The azimuthal angle and transverse momentum of all the invisible particles present in the final state can be inferred from the total momentum imbalance in the plane orthogonal to the beam axis (transverse plane): the so-called missing transverse momentum ($\vec{\not{E}}_T$) and its module ($\not{E}_T$), named missing transverse energy. Several algorithms to measure the missing transverse momentum presents in the event have been developed in CMS [109]. Among them, it worth to cite at least the PF-$\vec{\not{E}}_T$ used at the analysis level and the calo-$\vec{\not{E}}_T$ that is no longer used in the analyses but it is still used in some triggers. The PF-$\vec{\not{E}}_T$ is the opposite of the vectorial sum of all the visible momenta of the object reconstructed by the PF algorithm projected in the transverse plane. The calo-$\vec{\not{E}}_T$, instead, is computed using the energies collected by the calorimeter towers and the azimuthal angle is estimated from their direction with respect to the center of CMS.

The $\vec{\not{E}}_T$ is a complex object and it depends strongly on the other particles reconstruction, as well as on the detector calibration and malfunctions, the non-compensating nature of the calorimeters and the detector misalignment. Because of all this reasons, the measure of $\vec{\not{E}}_T$ could be not precise and differs from the real transverse momentum carried by all the invisible particles. To make $\vec{\not{E}}_T$ a better estimate of the true energy carried by undetected particles, different corrections have been studied [110] and applied:

- **Type-0 correction**: The Type-0 correction is a mitigation for the degradation of the $\vec{\not{E}}_T$ reconstruction due to the pile-up interactions. This correction can only be applied in the case of PF-$\vec{\not{E}}_T$. This correction aims to remove the charged hadrons identified with the PF algorithm originating from the vertices of pile-up interactions. In addition, the Type-0 correction removes an estimate of neutral pile-up contributions.

- **Type-I correction**: The Type-I correction is a propagation of the jet energy corrections (JEC) to the reconstructed $\vec{\not{E}}_T$ quantity. It consists in replacing the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied ($\vec{p}_T^{JEC}$):

$$\vec{\not{E}}_T^{\,corr} = \vec{\not{E}}_T - \sum_{jets} \left(\vec{p}_{T,jet}^{JEC} - \vec{p}_{T,jet}\right) \tag{3.6}$$

- **xy-shift correction**: The xy-Shift correction reduces the $\not{E}_T$ ϕ modulation. This correction is also a mitigation for the pile-up effects. In principle the distribution of true $\not{E}_T$ is independent of ϕ because of the rotational symmetry of the collisions around the

beam axis. However, a dependency of the reconstructed $\not\!\!E_T$ on ϕ is observed. The possible causes of the modulation include anisotropic detector responses, inactive calorimeter cells, the detector misalignment and displacements of the beam spot. The amplitude of the modulation increases roughly linearly with the number of the pile-up interactions.

Missing transverse momentum scale and resolution

The performance of the $\vec{\not\!\!E}_T$ are determined using $Z \to \mu\mu$ and $Z \to ee$ Monte Carlo generated events. In this events there is no real missing transverse energy and the $\not\!\!E_T$ resolution is dominated by the hadronic activity since the lepton resolution is very good ($\sigma_{p_\mathrm{T}}/p_\mathrm{T} \sim 1\% - 6\%$ fro muons and $\sigma_E/E \sim 1\% - 6\%$ for electrons).The response and resolution of the $\vec{\not\!\!E}_T$ are measured comparing the response of the hadronic recoil system $\vec{u_T}$ to the transverse momentum of the vector boson $\vec{q_T}$ (see Fig. 3.18)

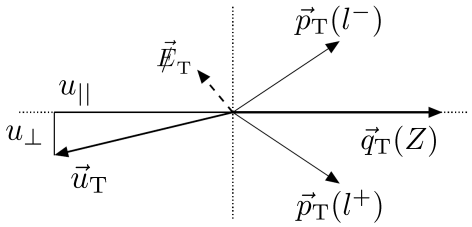


Figure 3.18: *$Z \to \ell\ell$ kinematic system in the transverse plane. The vector $\vec{q_T}$ represents the Z transverse momentum, $\vec{u_T}$ the vectorial sum of all particles except the two leptons from the Z decay and the $\vec{\not\!\!E}_T$ is the missing transverse momentum vector.*

Starting from the vectorial equation:

$$\vec{q_T} + \vec{u_T} + \vec{\not\!\!E}_T = 0 \tag{3.7}$$

and projecting the hadronic recoil along the parallel ($u_\parallel$) and perpendicular $u_\perp$ components with respect to the axis identified by the direction of $\vec{q_T}$, it is possible to define the response of the missing transverse energy as $-u_\parallel/q_T$ ($\vec{\not\!\!E}_T$ is supposed to be null in these events). The response of the PF-$\not\!\!E_T$ is illustrated in Fig. 3.19. It can be seen that after 50 GeV the response is quite flat, even if a small overestimation in the response of the order $O(1\%)$ is present in the $50 - 350$ GeV range. A good agreement is found between the data and the simulation. The resolution in the parallel ($\sigma(u_\parallel)$) and perpendicular ($\sigma(u_\perp)$) direction to q_T are thus evaluated as the Full Width at Half Maximum (FWHM) of the Voigtian function (Breit-Wigner convoluted with a Gaussian distribution) of the $u_\parallel + qT$ and $u_\perp$ distributions and shown in Fig. 3.20 where it can be seen that a good agreement is found between the data and the simulation.

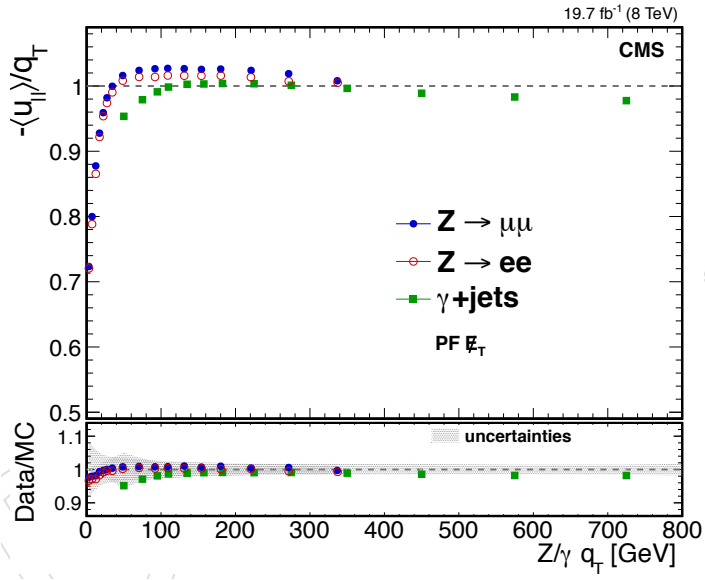


Figure 3.19: *Response for the PF-$\not\!\!E_T$ estimated from $Z \to \mu\mu$ (blue dots), $Z \to \ell\ell$ (red dots) and $\gamma + jets$ (green dots) events [110].*

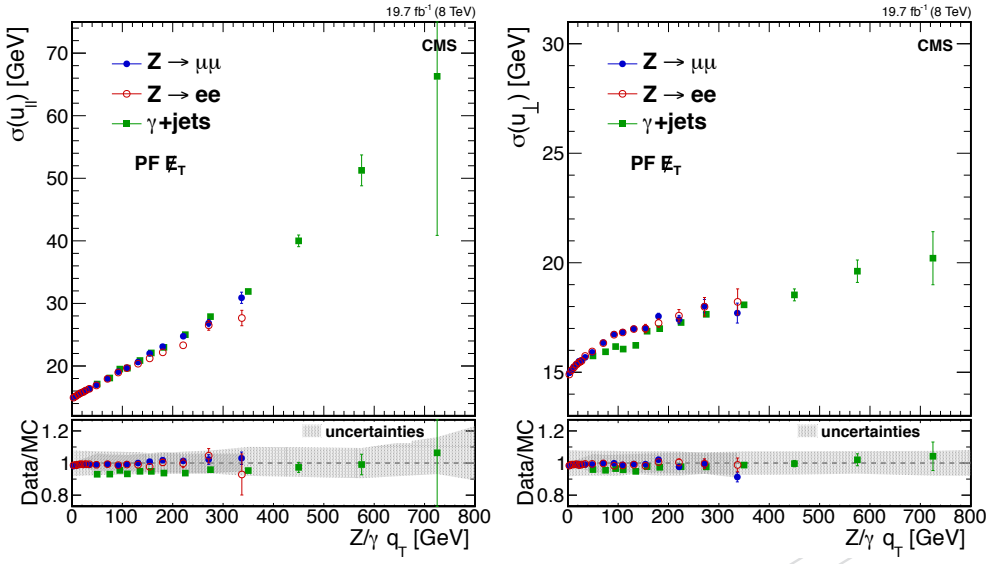


Figure 3.20: *Parallel (Left) and perpendicular (Right) recoil component resolution for the PF-$\not\!\!E_T$ as a function of the Z/γ boson transverse momentum for different processes selected from data: $Z \to \mu\mu$ (blue dots), $Z \to ee$ (red dots), and $\gamma + jets$ (green dots) [110].*

The corrections applied to the missing transverse energy reported below in this Section take care of the Pile-Up dependence. Since the majority of the pile-up events do not have significant $\not\!\!E_T$ and, in addition, the mean value of the $\not\!\!E_T$ projection along any axis is zero, the $\not\!\!E_T$ response is not significantly affected by the pile up. Contrarily, the resolution is strongly dependent on the number of reconstructed primary vertices in the event. CMS developed two algorithms to mitigate this effect on the PF-$\vec{\not\!\!E}_T$ and calo-$\vec{\not\!\!E}_T$. The general strategy is to separate the PF-$\vec{\not\!\!E}_T$ in different contributions: particles coming from the production vertex (PV) and particles originated by the pile-up interactions (PU). The first algorithm implemented is called No-PU PF-$\vec{\not\!\!E}_T$ and exploit the PF algorithm to recover the charged particles momentum coming from the production or pile-up vertices. Then it apply a reduction factor to the missing transverse momentum sum from the particles originated by the pile-up interactions:

$$S_F = \frac{\sum_{PV,charged} p_{\mathrm{T}}}{\sum_{PV,charged} p_{\mathrm{T}} + \sum_{PU,charged} p_{\mathrm{T}}} \tag{3.8}$$

The second algorithm developed is based on two consecutive multivariate regressions that provide an improved measurement of the $\not\!\!E_T$ in the presence of high pile-up. This algorithm, in particular, plays an important role in physics analyses sensitive to moderate missing transverse energy value, like the $H \to \tau\tau$. The MVA is used to evaluate two corrections. The first one is a correction to the direction of the $\vec{u}_T$ vector and it is obtained training a BDT to match the true hadronic recoil direction in Monte Carlo $Z \to \mu\mu$ simulated events. The second one returns the best value of the magnitude of the $\vec{u}_T$ vector. Finally, the corrected $\vec{u}_T$ is added to q_T in order to estimate the MVA PF-$\vec{\not\!\!E}_T$. More details about the algorithm implemented to compute No-PU PF-$\vec{\not\!\!E}_T$ and MVA PF-$\vec{\not\!\!E}_T$ can be found in [109]. The results on the resolution $\sigma_{\parallel}$ and $\sigma_{\perp}$ as function of the reconstructed primary vertices are shown in Fig. 3.21 where a significant improvement of the resolution using the MVA PF-$\vec{\not\!\!E}_T$ can be observed as well as a reduced dependency on the pile-up.

3.4 Tau lepton reconstruction

The τ is the heaviest lepton with a mass of $1776.82 \pm 0.16\, MeV$ and a lifetime $c\tau \sim 87\,\mu m$ [74]. Due to its large mass, the τ is the only lepton that can decay into hadrons. The τ leptons are expected to play a significant role in discovering new physics phenomena at LHC [14]. Many physics analyses are expected to benefit from an efficient reconstruction of τ leptons which should be complemented by a good performance in rejecting possible background contaminations. In about two thirds of cases τ leptons decay hadronically, typically into either one or three charged mesons (predominantly π^+ π^-) in presence of up to two neutral pions, decaying via $\pi^0 \to \gamma\gamma$ (see Tab. 3.1). More details on the τ decay modes are reported in Table 3.1. In Fig. 3.24 the different Feynman diagrams that contribute at L.0. to the most relevant τ hadronic decays are shown. The variety of the τ hadronic decay modes results in a complicated signature (the

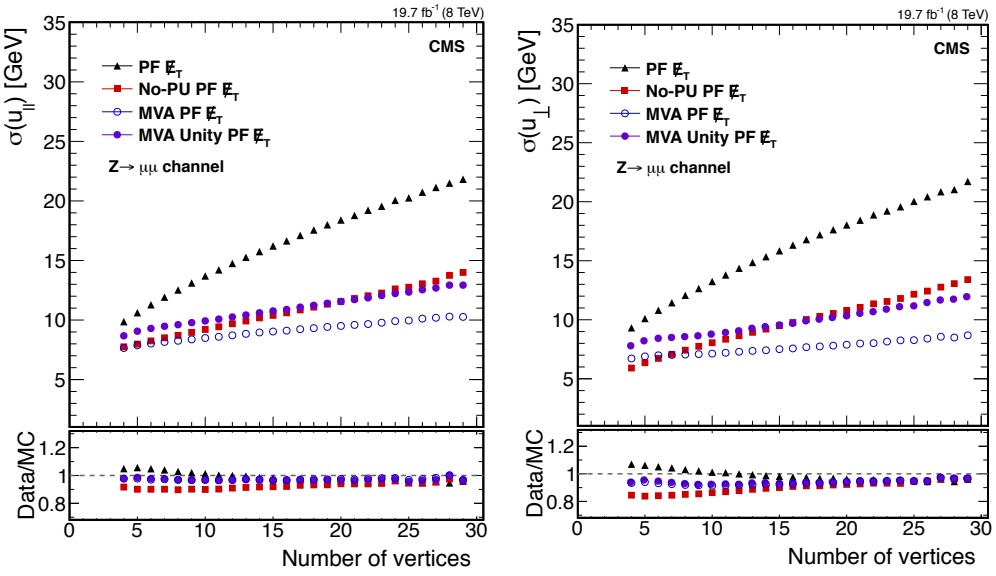


Figure 3.21: *Parallel (Left) and perpendicular (Right) recoil component resolution for the different corrections applied on top of the PF-$\vec{\not\!\!E}_T$ [110].*

final stable particles detectable can be up to 5 per τ giving rise to a low-multiplicity jet) for the final state. Similar experimental signature is expected for generic, quark and gluon, QCD jets production. Since the cross section of jet production exceeds the cross section of τ lepton production by several orders of magnitude, the experimental challenge in reconstructing and identifying hadronic τ decays is to discriminate efficiently between genuine τ lepton hadronic decays (τ-jets) and quark/gluon jets misreconstructed as τ candidates [111]. Several algorithms for reconstruction and identification of hadronic τ decays were studied in CMS and the *Hadron Plus Strip* algorithm (HPS) is found the one achieving the best performance and thus vastly used in the physics analysis. The *Hadron Plus Strips* (HPS) τ identification algorithm is based on charged hadrons and neutral electromagnetic objects (*photons*). The HPS algorithm starts from a particle-flow jet and searches for τ lepton decay products produced by any of the hadronic decay modes listed in Table 3.1. The last part of this chapter is completely dedicated to the description of this algorithm and to the most important results achieved in terms of its performance.

τ decay mode		resonance	Branching Ratio [%]
leptonic	$\tau^- \to e^- \nu_\tau \bar{\nu}_e$	-	17.83 ± 0.04
leptonic	$\tau^- \to \mu^- \nu_\tau \bar{\nu}_\mu$	-	17.41 ± 0.04
TOTAL leptonic			35.24 ± 0.11
1-Prong	$\tau^- \to \pi^- \nu_\tau$	-	10.83 ± 0.06
1-Prong + π^0s	$\tau^- \to \pi^- \pi^0 \nu_\tau$	$\rho^-(770)$	25.52 ± 0.09
1-Prong + π^0s	$\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$	$a1^-(1260) \to \rho^-(770)$	9.30 ± 0.11
3-Prongs	$\tau^- \to \pi^- \pi^- \pi^+ \nu_\tau$	$a1^-(1260) \to \rho^0(770)[\sigma^0(500)]$	9.31 ± 0.06
TOTAL hadronic			54.96 ± 0.82

Table 3.1: *Summary of the τ decay modes considered in the analysis.*

1-prong

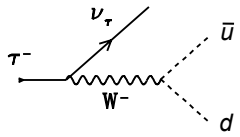


Figure 3.22: *Feymann diagram for τ 1-Prong + π^0 hadronic decay at LO.*

1-prong + π^0

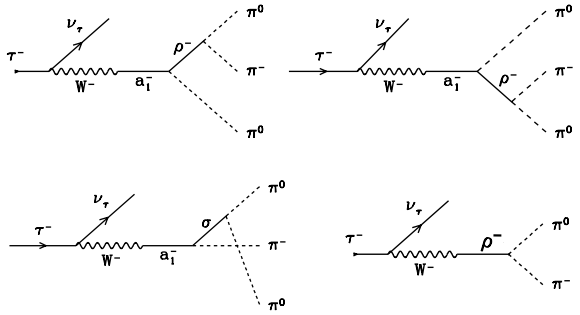


Figure 3.23: *Feymann diagrams for τ 1-Prong + π^0 hadronic decay at LO.*

3-Prongs

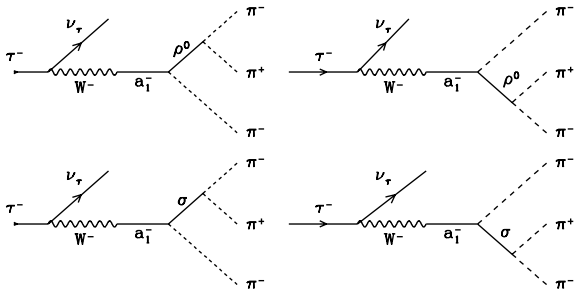


Figure 3.24: *Feymann diagrams for τ 3-Prongs hadronic decay at LO.*

3.4.1 Hadron Plus Strips (HPS) algorithm for τ lepton identification

This algorithm is the most performing, in terms of energy resolution and fake rate rejection [112], implemented by CMS as τ reconstruction algorithm. Since neutral pions appear regularly in hadronic τ decay modes, one of the main goal of the HPS algorithm is a correct reconstruction of the π^0. Even though a photon conversion reconstruction exists and used within the PF, its efficiency is limited. Therefore, in the HPS algorithm, photons, electrons and positrons are treated equally. Special attention is given to photons converting in the CMS tracker material [112]. A photon conversion typically results in broader calorimeter signatures for neutral pions in the azimuthal direction, due to the bending of the electron/positron trajectories. The possible broadening of calorimeter signatures by photon conversions is accounted for in the HPS algorithm by reconstructing photons in *strips*, objects which are built out of electromagnetic particles.

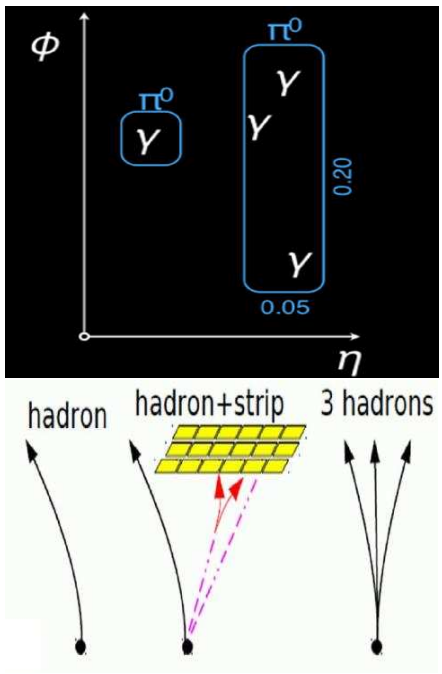


Figure 3.25: *Exemple of a strip in HPS algorithm in the* (η, ϕ) *plain.*

Algorithm steps:

- The strip reconstruction starts by centering one strip on the most energetic electromagnetic particle reconstructed by the PF algorithm within the jet

- The algorithm then searches for other electromagnetic particles within the window of size $\Delta\eta = 0.05$, $\Delta\phi = 0.20$ around the strip center, see Fig. 3.25
- In case other electromagnetic particles are found within that window, the most energetic one of these particles gets associated to the strip and the strip center position is recalculated to match the sum of four-vectors of all particle-flow particles associated to the strip;
- The next highest energetic electromagnetic particle is then searched for within the $\Delta\eta = 0.05$, $\Delta\phi = 0.20$ window around the new strip center until no further particles are found which can be associated to the strip. In this case the algorithm creates a new strip centred around the most energetic particle not associated to any strip
- Strips satisfying a minimum transverse momentum requirement of $E_{\mathrm{T}}^{strip} > 1\,\mathrm{GeV}/c$ are finally combined with the charged hadrons to reconstruct individual hadronic τ lepton decay modes

The decay modes which are considered by the HPS tau identification algorithm are:

1. **Single hadron:** This signature reconstructs $\tau^{\pm} \to h^{\pm}\nu$ decays and $\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau(\bar{\nu_\tau})$ decays in which the neutral pions have too little energy to be reconstructed as strips
2. **One hadron + one strip:** This signature aims to reconstruct the decay mode $\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau(\bar{\nu_\tau})$ in events in which the impact position of the photons from π^0 decays are close together on the calorimeter surface. The size of the strip in ϕ direction is large enough to account for the possibility that one or both of the photons have converted
3. **One hadron + two strips:** This signature aims to reconstruct the decay mode $\tau^{\pm} \to h^{\pm}\pi^0\pi^0\nu_\tau(\bar{\nu_\tau})$ and $\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau(\bar{\nu_\tau})$ in events in which the impact positions of photons from π^0 decays are well separated on the calorimeter surface;
4. **Three hadrons:** This signature aims to reconstruct the decay mode $\tau^{\pm} \to h^{\pm}h^{\mp}h^{\pm}\nu_\tau(\bar{\nu_\tau})$. The three charged hadrons are required to have charge $|\sum q| = 1$. In addition, the three charged tracks are refitted and required to be originated by the same secondary vertex, reconstructed by the Kalman vertex fit algorithm.

All charged hadrons and strips are required to be contained within a narrow cone of size $\Delta R = 2.8/p_{T_\tau}$ [112], where p_{T_τ} is computed by summing the four-vectors of reconstructed charged hadrons plus strips. The hypothetical visible τ momentum $\vec{p}_\tau$ is required to match the (η, ϕ) direction of the original jet within a maximum distance of $\Delta R = 0.1$ [112]. Particle-flow jets with different numbers of reconstructed charged hadrons or strips with respect to those listed above are not considered as potential τ candidates. The 4-momenta of the hadrons and strips are recomputed to assign masses compatible with the reconstructed decay channel. Besides, the total invariant mass of the visible decay products is required to match that of the intermediate

resonance (the intermediate resonances considered are listed in Tab. 3.1). In case of the one hadron + one strip signature the four-vectors reconstructed from the strip energy and position is set to match the nominal π^0 mass and the invariant mass of charged hadron plus strip is required to be in the range $0.3 - 1.3\,\mathrm{GeV}/c^2$. In case of the hadron + two strips signature a massless four-vectors is reconstructed for each strip and the invariant mass of the two strips is required to be within $50\,MeV/c^2$ and $200MeV/c^2$. The invariant mass of hadron plus strips is required to be in the range $0.4 - 1.2\,\mathrm{GeV}/c^2$. The signature of three charged hadrons is not affected by photon conversions and can be well identified. The invariant mass of the three tracks is expected to be in the range $0.8 - 1.5\,\mathrm{GeV}/c^2$. In case more than one hypothesis for possible τ decay signatures exist the hypothesis leading to the lowest E_T sum of jet constituents not associated to τ decay products is given preference.

3.4.2 HPS performance

The performance of the HPS algorithm are evaluated in terms of energy response, resolution and efficiency in the τ decay modes reconstruction. This study has been performed using Monte Carlo $Z/\gamma^* \to \tau\tau$ simulated events. The energy response and resolution are defined respectively as the mean value of the distribution of the ratio between the energy of the τ reconstructed by the HPS and the *true* energy at generator level (simulated τ) and the width of this distribution. From Fig.3.26 it is possible to see that the HPS provides a good energy resolution, between 3% and 10% depending on the decay mode and that these values are almost independent from the pile-up.

Another figure of merit used to evaluate the performance of the τ reconstruction consists in the capability of the HPS algorithm to identify the correct τ decay mode. In Fig. 3.27 the correlation between the reconstructed and Monte Carlo generated decay modes for three different pile-up scenario is shown. The τ decay channel is correctly reconstructed in a percentage that varies from $\sim 85\%$ up to 95% and it is stable with the pile-up.

As already mentioned in the introduction to the τ reconstruction, one of the main goal to provide a good identification of the tau lepton is to keep the contamination from QCD jets as low as possible. This goal is reached requiring that the τ lepton candidate is isolated. Indeed, τ lepton is colorless and its decay occurs purely through weak interaction. As a consequence, the τ decay products are emitted in a relative clean environment contrary to the quark/gluon jets. Two kind of isolation algorithm have been developed in CMS, a cut-based approach and an algorithm based on a multivariate analysis technique. In what follows a deeper description of the cut-based isolation algorithm and its performance is provided. This isolation method is, indeed, vastly used in the search for the SM $H \to \tau\tau$. In order to compute the isolation of the τ_h, the electromagnetic or charged particles found in a cone of $R = 0.5$ with $p_\mathrm{T} > 1.0\,\mathrm{GeV}$ ($> 1.5\,\mathrm{GeV}$ for photons) around the τ_h candidate are treated as *isolation particles*. In addition, the charged *isolation particles* must be associated to a track that pass the following requirements:

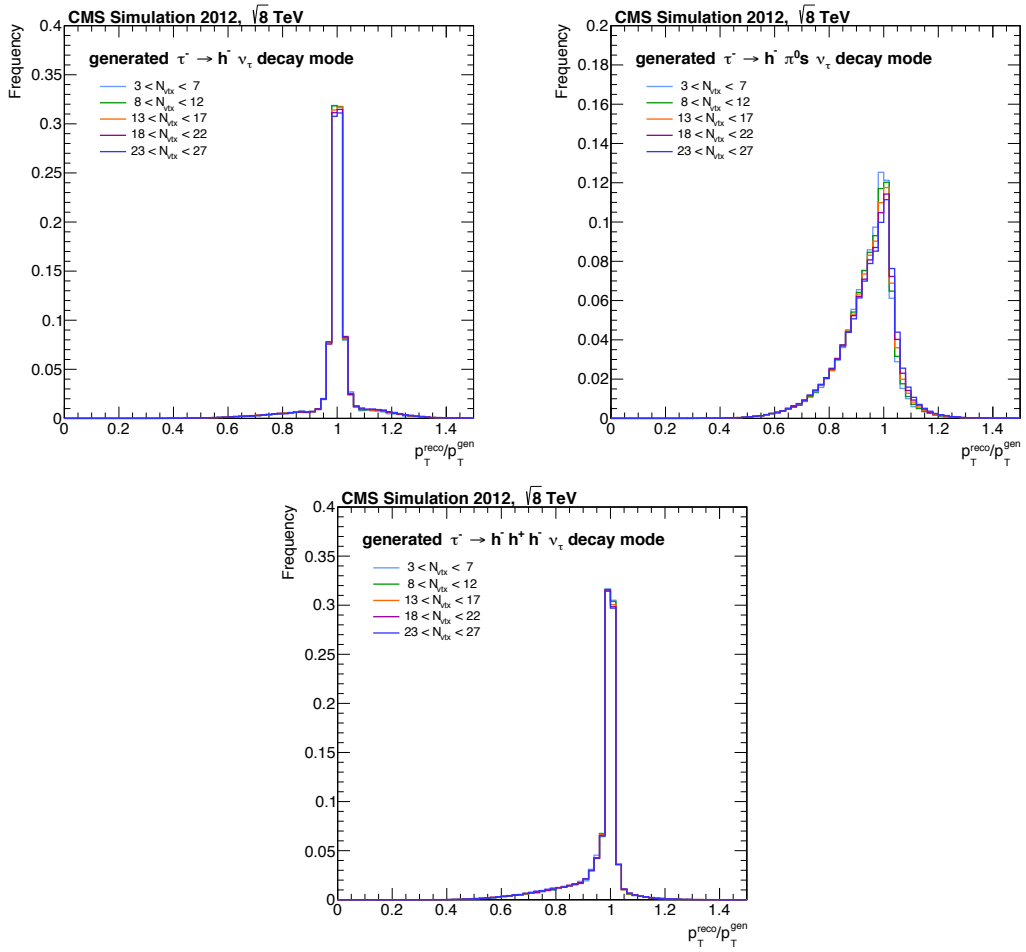


Figure 3.26: *The τ energy response distribution for different τ reconstructed decay mode 1-prong (top left), 1-prong+π^0s (top right) and 3-prongs (bottom) for 5 different pile-up scenarios [113].*

$p_{\mathrm{T}} > 0.5$, $\chi^2 < 100$, $d_0 < 0.03\,cm$, $d_Z < 0.2\,cm$ and $N_{hits} \geqslant 3$ (d_0 and d_Z are the transverse and longitudinal component of the impact parameter evaluated with respect to the vertex closest to the leading track of the τ candidate). In order to maintain the isolation performance stable with the pile-up activity, only the charged particles with their track matched to the τ_h production vertex are considered. An event by event estimation of the photons energy (also named $\Delta\beta$) is computed by summing the transverse momenta of all the charged particles whose direction match the τ_h candidate within $\Delta R < 0.8$ and which are not matched to the τ_h production vertex ($d_Z > 0.2\,cm$); they are therefore coming from pile-up vertices. This sum is then scaled by a constant mostly accounting for the ratio between the neutral and charged pile-up contribution

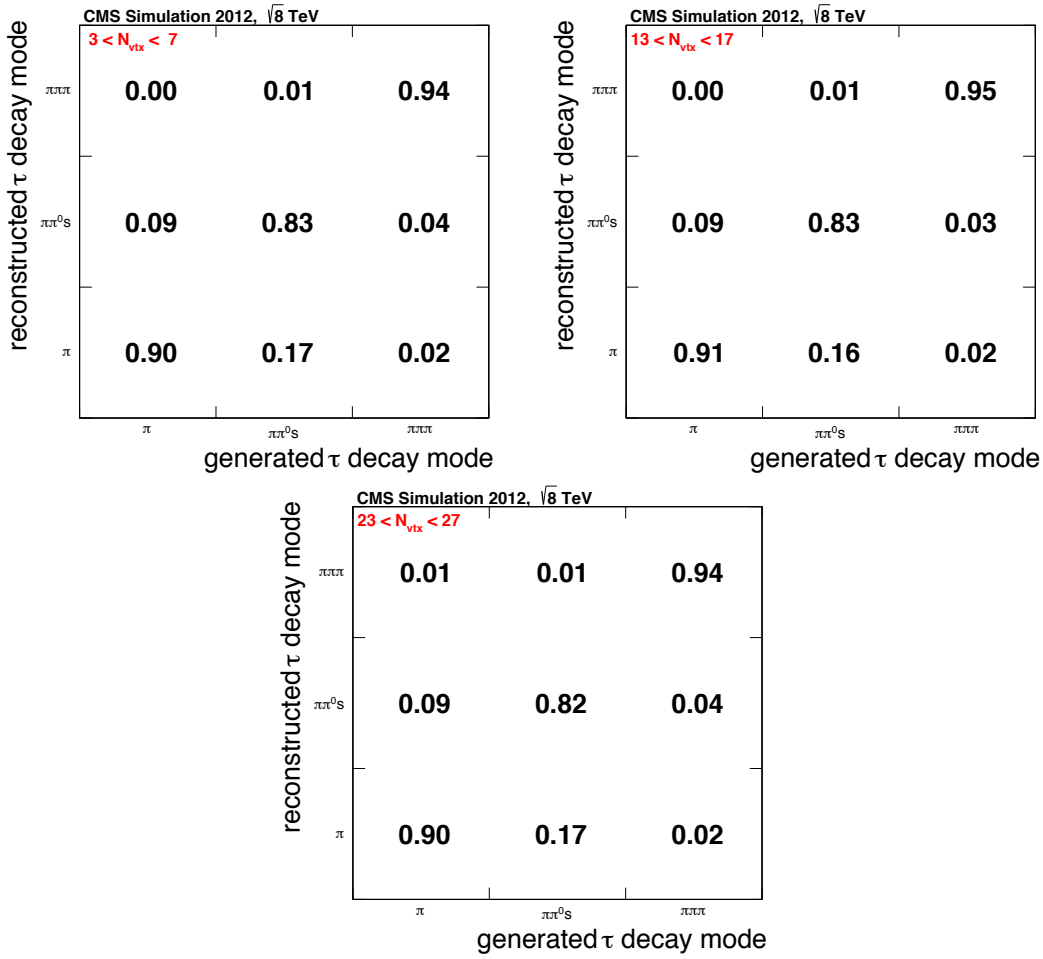


Figure 3.27: *Correlation between the generated and reconstructed τ_h decay modes in simulated $Z/\gamma^* \to \tau\tau$ events for different pile-up scenario [113].*

(~ 0.46). Finally, the isolation variable assumes the form:

$$I_\tau = \sum p_{\mathrm{T}}^{charged}(d_Z < 0.2\,cm) + max(\sum p_{\mathrm{T}}^{\gamma} - \Delta\beta, 0) \tag{3.9}$$

with

$$\Delta\beta = 0.46 \cdot \sum p_{\mathrm{T}}^{charged}(d_Z > 0.2\,cm) \tag{3.10}$$

Thresholds on the value of I_τ of 2.0, 1.0, and 0.8 define respectively the loose, medium and tight working point for the cut based isolation. Tau isolation efficiency as function of the p_{T}^{τ}, shown in Fig.3.29, is estimated from data, using the tag-and-probe technique and compared to simulation. A good agreement between data and simulation can be observed in the whole

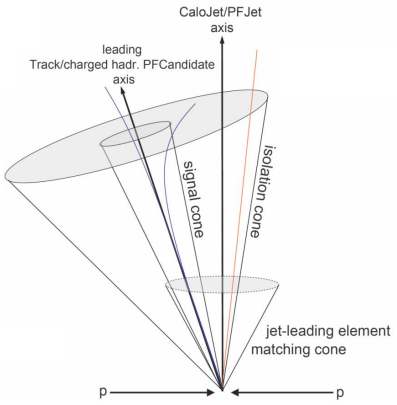


Figure 3.28: *Tau isolation cone structure used in the definition of the standard cut-based isolation variable.*

transverse momentum range considered and efficiencies of $\sim$ 45%, $\sim$ 55%, $\sim$ 63% are found respectively for the tight, medium and loose working point.

The τ energy scale

The HPS algorithm can sometimes miss some of the visible decay products of the τ decay and thus, produce a τ candidate with an underestimation of its true energy. The opposite scenario is also possible, where an additional particle coming from the pile-up can be mis-identified as being part of the τ_h. Then the reconstructed momenta will be overestimated. Other possible reasons that could lead to an energy mis-reconstruction are the tracker mis-calibration and the calorimeters energy response. In order to estimate the τ energy scale a likelihood fit is performed to the distribution of the the visible tau mass in $Z \to \tau\tau \to \mu\tau_h$ events from Monte Carlo sample. In Fig. 3.30 the distribution of the visible invariant mass, defined as the invariant mass of all the visible τ decay products, is shown for data, $Z \to \tau\tau \to \mu\tau_h$ simulated events and all the other SM backgrounds (also simulated with Monte Carlo). The distribution of the visible tau mass depends on the τ decay mode: for the 1-prong τ's it is peaked exactly at the value of the $\pi^{\pm}$ mass, while for 1-prong + π^0s and 3-prongs τ's it is spread out in a larger mass interval because of the presence of the intermediate resonances $\rho(770)$ and $a1(1260)$.

The likelihood fit adjusts the shape of the simulated visible mass distribution to match the one from data keeping the τ energy scale as a free parameter of the fit. The central value of the τ energy scale obtained from the fit is $\sim$ 1% for all the 3 decay modes. The systematic uncertainty related to this measure is found to be $\sim$ 3% and represent a very important shape systematic for many physics analysis that make use of the τ lepton in the final state, like the

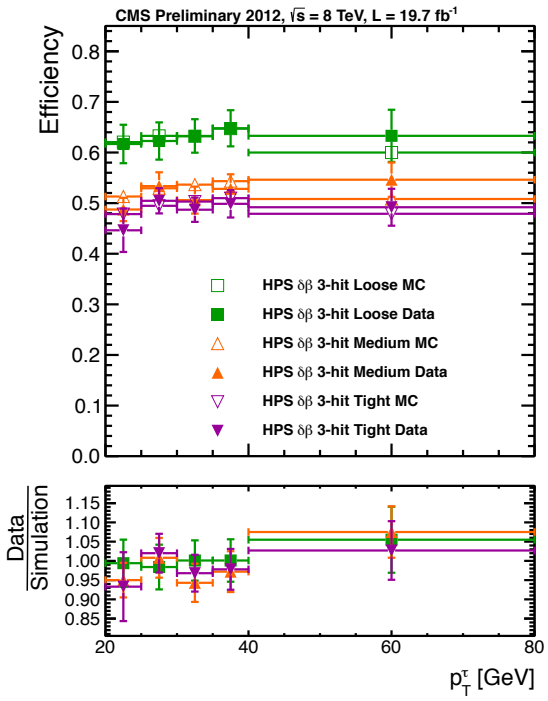


Figure 3.29: *Data and Monte Carlo simulation comparison for the τ isolation efficiency as function of the τ candidate transverse momentum for the three different working points: loose, medium and tight [113].*

searches for the SM $H \to \tau\tau$.

3.4.3 The di-τ invariant mass reconstruction

An important topic related to the τ object reconstruction is the evaluation of the di-τ invariant mass $m_{\tau\tau}$ in di-τ events. This quantity is used in most of the analyses that search for a resonance decaying into a τ lepton pair in order to discriminate the signal from the various backgrounds. In particular, in the search for SM $H \to \tau\tau$, $m_{\tau\tau}$ is used to discriminate the events coming from the $Z \to \tau\tau$ decay from those originating from the Higgs boson decays. It is also used to discriminate from other background. However, the estimation of the true $m_{\tau\tau}$ invariant mass is a real challenge since the hadronic and leptonic τ decays present neutrinos in the final state. For this reason the invariant mass evaluated only with the τ visible decay products four-momenta presents a limited discrimination power, since the exact amount of momenta carried by neutrinos is unknown. Among the various methods that have been developed by the CMS Collaboration to reconstruct the di-τ invariant mass, it worth to mention the most used:

- **Collinear approximation:** In this approach the τ lepton kinematics is simplified assum-

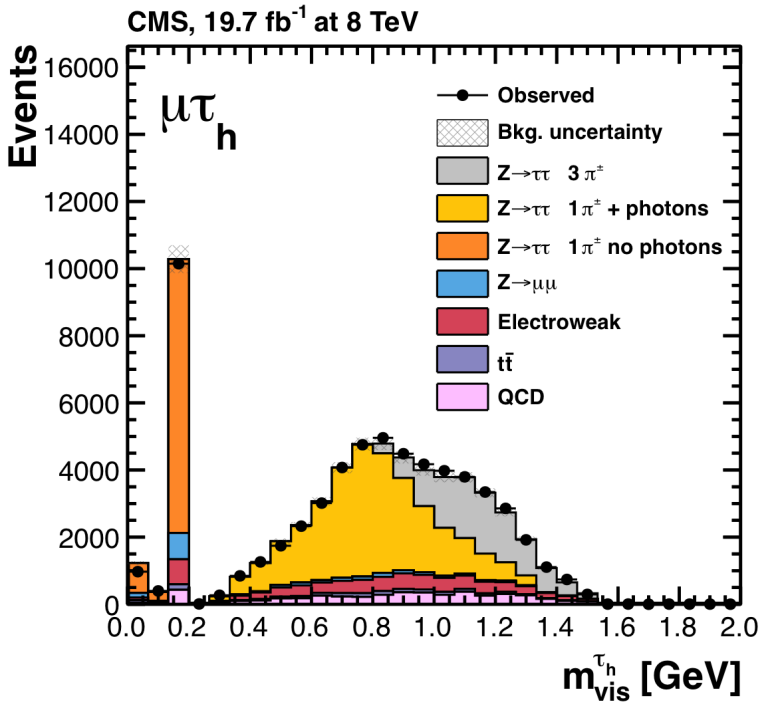


Figure 3.30: *Comparison between data and $Z \to \tau\tau$ + backgrounds simulated events containing τ_h candidates reconstructed in all the decay mode after the likelihood fit [71].*

ing that all its decay products are emitted collinearly with the τ-lepton momentum and that all the missing energy originates from the neutrinos. In the case of a di-tau event, this assumption implies that the projection of the $\vec{\not{E}}_T$ along each visible tau direction is interpreted as the associated neutrino(s) transverse momentum. This approximation works properly 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). The collinear approximation drawback is that it has no solution for topologies presenting two back-to-back τs [114]. Indeed, in this configuration, there is only one available projection for the $\vec{\not{E}}_T$, leading to an infinite number of solutions for the neutrinos transverse momenta.

- **Secondary Vertex Fit[3] (SVfit) algoritm [115]:** it is a likelihood-based mass reconstruction technique where all the allowed configuration for the neutrinos momenta are weighted by a probability density function obtained by a completely analytical approach.

[3]In the current implementation of the SVfit algorithm, the secondary vertices of the τ leptons are not used to determine the τ momenta, even if the algorithm was originally designed to with this functionality.

SVfit invariant mass reconstruction algorithm

The τ pair mass is totally determined by the two four momenta of the τ leptons. Each τ four momentum is, in turn, completely determined by the visible and invisible four momenta of its decay products. The main experimental challenge is to reconstruct the four momenta of the invisible particles. They can be constrained by the $\vec{E}_T$ and the τ-lepton mass m_τ which is known with great precision. As shown in Tab. 3.1 and in Fig. 3.24 (Section 3.4) the leptonic τ decays leads to two neutrinos in the final state, while a hadronic τ decay with the emission of one neutrino. This implies that the invisible system is massive for the leptonic decays (resulting from $m^2_{\nu\nu} = (p_{\nu 1} + p_{\nu 2})^2$) and massless for the hadronic decays ($m^2_{\nu\nu} = p^2_\nu = 0$). The τ decay 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[4], i.e. the angle between $(\vec{p}^{\,vis}_\tau, \vec{p}_\tau)$ and $(\vec{p}^{\,vis}_\tau, \vec{u}_z)$[5].
- the invariant mass of the invisible momentum system $m_{\nu\nu}$, in the case of leptonic τ decays.

The energy and momentum of the visible decay products in the rest frame of the τ are given by:

$$E^{vis}_\tau = \frac{m^2_\tau + m^2_{vis} - m^2_{\nu\nu}}{2m_\tau}\,, \;\; p^{vis}_\tau = \sqrt{(E^{vis}_\tau)^2 - m^2_\tau}, \tag{3.11}$$

where m_{vis} is the visible τ mass, defined as $m^2_{vis} = (\tau^{vis})^2$ (τ^{vis} is the four-momentum of the visible τ decay products) with $m_{\nu\nu} = 0$ for the hadronic decays. The opening angle in the CMS detector frame $\bar{\theta}$ is related to the corresponding quantity in the τ rest frame via the Lorentz invariant component of the visible momentum perpendicular to the τ-lepton direction:

$$p^{vis}_{\perp,\tau} = \bar{p}^{vis}_{\perp,\tau} \Rightarrow sin\bar{\theta} = \frac{p^{vis}_\tau \cdot sin\theta}{\bar{p}^{vis}_\tau}. \tag{3.12}$$

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 vector direction from equations 3.11 and 3.12:

$$\bar{p}^{vis}_\tau = \gamma\beta E^{vis}_\tau + \gamma p^{vis}_\tau cos\theta, \tag{3.13}$$

where:

$$\gamma = \frac{E^{vis}_\tau \left[(E^{vis}_\tau)^2 + (\bar{p}^{vis}_\tau cos\bar{\theta})^2\right]^{\frac{1}{2}} - p^{vis}_\tau cos\theta \cdot \bar{p}^{vis}_\tau cos\bar{\theta}}{(E^{vis}_\tau)^2 - (p^{vis}_\tau cos\theta)^2}. \tag{3.14}$$

[4]Symbols with an overline refer to quantities defined in the laboratory frame

[5]$\vec{u}_z$ is the unitary vector aligned with the beam axis

The energy of the τ in the laboratory frame, as function of the visible momentum $\bar{p}_\tau^{vis}$, only depends 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 3.31 shows the relation between p_τ, p_τ^{vis}, $\bar{\theta}$ and $\bar{\phi}$.

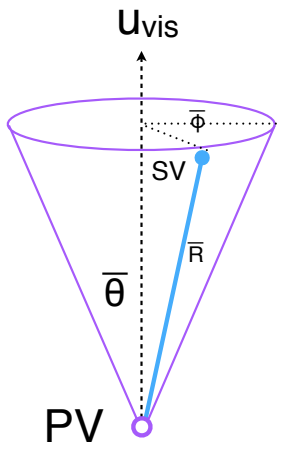


Figure 3.31: *Parametrization of the τ decay used in the SVfit algorithm. The parameter $\bar{\theta}$ represents the angle between the boost direction of the τ-lepton and the momentum vector of the visible decay products in the laboratory frame, $\bar{\phi}$, the azimuthal angle of the τ in the laboratory frame and y_{vis} is the unitary vector aligned with the visible τ direction in the laboratory frame.*

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 $\bar{E}_\tau^{vis}$, as a function of $p_\tau^{vis} = \sqrt{(E_\tau^{vis})^2 - (m_\tau^{vis})^2}$ and $\beta = \sqrt{\gamma^2 - 1/\gamma^2}$:

$$cos\theta = \frac{\bar{E}_\tau^{vis} - \gamma E_\tau^{vis}}{\gamma\beta p_\tau^{vis}} \xrightarrow{\beta\to 1} cos\theta = \frac{2\bar{x} - 1 - \frac{(m_\tau^{vis})^2}{m_\tau^2}}{1 - \frac{(m_\tau^{vis})^2}{m_\tau^2}} \tag{3.15}$$

for leptonic (hadronic) τ-lepton decays.

The mass of the di-τ system is under constrained and its kinematics are controlled by 4 to 6 parameters depending on the leptonic or hadronic decays: the angles θ_1, $\bar{\phi}_1$ and θ_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 $\not{E}_x$ and $\not{E}_y$ provide two further constraints. The total number of free parameters is then $3 \times N(\tau_{lep}) + 2 \times N(\tau_h) - 2$. It is possible to define the likelihood function $f\left(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2}\right)$ describing the probability to reconstruct $z = (\not{E}_x, \not{E}_y)$, given the unknown parameters of the two τ decays taking the values $a1 = (\theta_1, \bar{\phi}_1, m_{\nu\nu,1})$, $a2 = (\theta_2, \bar{\phi}_2, m_{\nu\nu,2})$ and the four momenta of the visible decay products corresponding to the reconstructed values $\vec{y} = \left(p_1^{vis}, p_2^{vis}\right)$. The mass of the di-$\tau$ system $m_{\tau\tau}\left(\vec{y}, \vec{a_1}, \vec{a_2}\right)$ is a well defined function of $\vec{y}, \vec{a_1}$ and $\vec{a_2}$. 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 hypotheses by computing the probability:

$$P\left(m_{\tau\tau}^i\right) = \int \delta(m_{\tau\tau}^i - m_{\tau\tau}\left(\vec{y}, \vec{a_1}, \vec{a_2}\right)) f\left(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2}\right) d\vec{a_1} d\vec{a_2} \tag{3.16}$$

The integration is performed numerically using the VEGAS [116][117] algorithm (see Section 6.3).

The τ lepton decay width

The matrix element model for unpolarized τ-leptons decays are used to model the leptonic decays:

$$\mathcal{L}_{\tau,\text{lep}} = \frac{d\Gamma}{d\bar{x} dm_{\nu\nu} d\bar{\phi}} \propto \frac{m_{\nu\nu}}{4m_\tau^2}\left[\left(m_\tau^2 + 2m_{\nu\nu}^2\right)\left(m_\tau^2 - m_{\nu\nu}^2\right)\right] \tag{3.17}$$

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 3.15 ($1 \leq cos\theta \leq 1$ condition). While a model based on the two-body phase space for the hadronic decays gives:

$$\mathcal{L}_{\tau,\text{had}} = \frac{d\Gamma}{d\bar{x} d\bar{\phi}} \propto \frac{1}{1 - \frac{m_{vis}^2}{m_\tau^2}} \tag{3.18}$$

within the physically allowed region $m_{vis}^2/m_\tau^2 \leq \bar{x} \leq 1$. The comparison of the kinematic distributions with respect to the detailed simulation, as illustrated in Ref. [118], 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 $\vec{\not{E}}_T$ obtained from the particle flow reconstructed objects. To take the resolution effects that cause the differences between the sum of the neutrino's p_T and the reconstructed missing transverse energy into account, a Gaussian resolution model is considered. It is possible to write a likelihood $\mathcal{L}_\nu(\not{E}_x, \not{E}_y)$ for observing $E_{\not{x},y}$ given a true value of missing energy, where $\not{E}_{x,y}$ are the component x, y of the measured missing transverse energy. This approach is better described in [115] and it is also used to model the response of the missing transverse energy in Chapter 6.

Full likelihood

The full likelihod in Eq 3.16 is the result of the product of the previous terms:

$$f(\vec{z},\vec{y},\vec{a_1},\vec{a_2}) \propto \sum_{i,j} \mathcal{L}_{\tau,i} \cdot \mathcal{L}_{\tau,j} \cdot \mathcal{L}_{\nu}(\not{E}_x, \not{E}_y) \tag{3.19}$$

where i,j denotes the possible τ decay mode. At this point, $P(m_{\tau\tau})$ (Eq. 3.16) is computed for all the mass hypothesis and the value of $m_{\tau\tau}$ that maximize the probability is interpreted as the di-τ invariant mass.

SVfit performance

In Fig. 3.32 the distributions of the visible invariant mass and the SVfit mass are shown for simulated $Z \to \tau\tau$ and SM $H \to \tau\tau$ events.

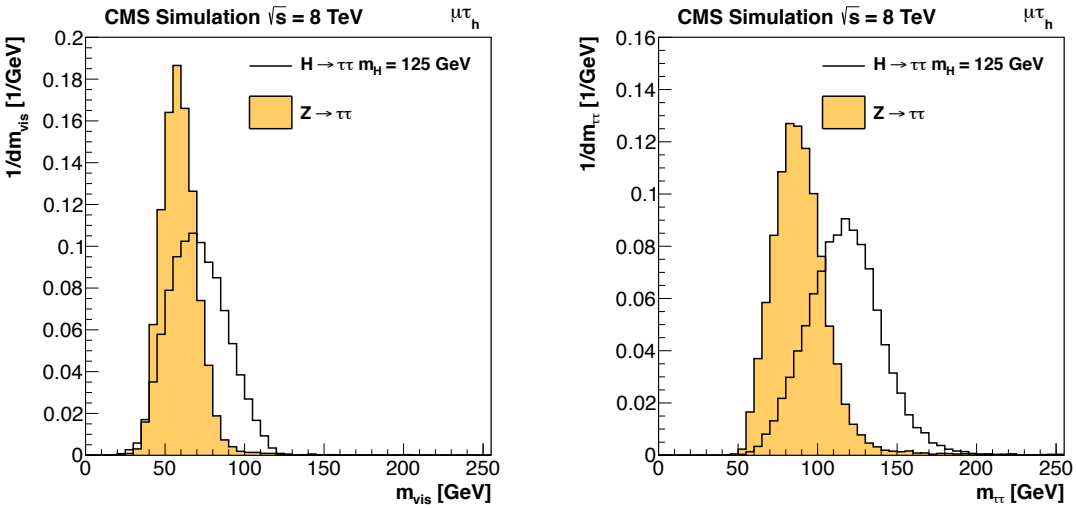


Figure 3.32: *Distribution of the τ pair visible invariant mass (Left) and the mass reconstructed by the SVfit algorithm (Right) for simulated $Z \to \tau\tau \to \mu\tau_h$ background events (white) and for the SM Higgs boson signal $H \to \tau\tau \to \mu\tau_h$ with $m_H = 125$ GeV [115].*

It is possible to appreciate the increase of the discrimination power obtained using the SVfit mass algorithm with respect to the visible invariant mass. The adoption of the SVfit algorithm brings an improvement in the final expected significance of the SM $H \to \tau\tau$ analysis of $\sim 40\%$ with respect to the expected significance obtained using, as observable, the visible di-τ mass. 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 $\tau_h\tau_\ell$ decays, and 20% in the fully leptonic $\tau_\ell\tau_\ell$ decays, increasing with the number of neutrinos in the τ decay final states [115]. The SVfit

algorithm also 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 O(1s)/event.

Chapter 4

Study of the Missing Transverse Energy reconstructed at L1 in the context of the $\mathrm{H} \to \tau\tau \to \ell\tau_{had}$ analysis with soft leptons

An overview on the main analyses that lead to the Higgs boson discovery, the measurement of its mass, followed by the bounding of its width has been presented in Chapter 1. During the LHC Run 1, the analyses performed in the main channels (ZZ, $\gamma\gamma$, WW, $b\bar{b}$ and $\tau\tau$) were aiming an optimal sensitivity to the SM Higgs boson. To do so, inclusive selections of the production mechanisms were adopted for each of the major final states. At the time I began my thesis work, the Higgs boson had just been discovered in the bosonic decay channels. An intensive work focussing on the measurements of its properties had started. In particular, first measurements of its spin-parity have been carried out [56]. All the measurements were (and still are) compatible with a Standard Model Higgs boson. The $\tau\tau$ channel plays a particular role, as it is the only way to check if the Higgs couples to leptons as predicted by the Standard Model. Indeed, the $H \to \mu\mu$ channel sensitivity is limited by the small branching ratio because of the weak coupling between the Higgs boson and the muon. In the context of the $H \to \tau\tau$ analysis, the main efforts concentrated in the increase of the analysis sensitivity in order to reach the first standalone observation of the Higgs boson decaying into fermions. Several approaches have been pursued,

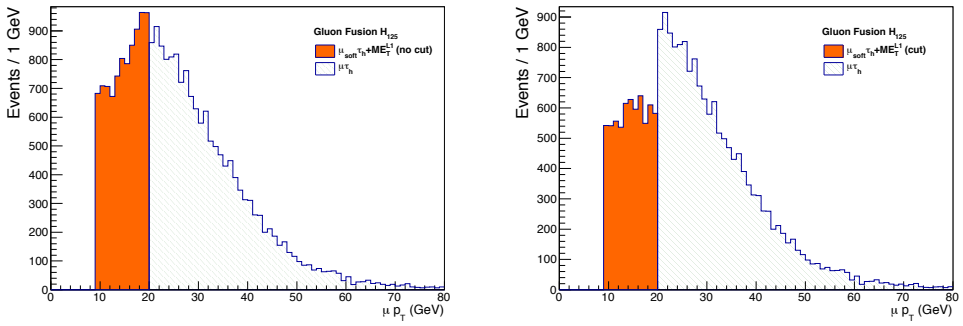
for example the subdivision into categories, the inclusion of new decay channels ($\tau_h\tau_h$) and the improvement of the object reconstruction. It has also been tried to improve the trigger strategy during the data taking. The work I have done at the very beginning of my thesis and that is described below took place in the context of this last approach and dealt with the characterisation of the new trigger; the analysis itself is described in 7. The τ lepton can decay

both into hadrons and leptons with the presence of neutrino(s) in the final state. The fraction of the τ momentum carried by the neutrinos translates into a missing transverse energy in the detector, especially for the semileptonic channels ($H \rightarrow \tau\tau \rightarrow \mu\tau_h$, $H \rightarrow \tau\tau \rightarrow e\tau_h$) and the fully leptonic ones ($H \rightarrow \tau\tau \rightarrow \mu\mu$, $H \rightarrow \tau\tau \rightarrow ee$). In these cases the p_{T} distribution of the leptons in the final state is characterized by a soft spectrum: e.g. the 53% of the $H \rightarrow \tau\tau$ events produced through gluon-gluon fusion have a $p_{\mathrm{T}} < 20\,\mathrm{GeV}$. With the triggers available during the largest period of the Run 1 data taking, the low p_{T} spectrum of the lepton represented a severe limitation. As already pointed out in the Section 2.3.5, a trigger system at a hadron collider is designed to provide the highest efficiency in selecting the most interesting events, keeping the rate of the events not interesting for physics (background events that could fake the signal) as low as possible. The hardware data and acquisition system impose severe limitation on the total bandwidth of triggered events. In order to control the trigger rate, relatively high thresholds ($\sim 20\,\mathrm{GeV}$) on the lepton p_{T} were mandatory at trigger level with a consequent loss in the signal acceptance ($\sim$ 53% of the signal events have leptons with $p_{\mathrm{T}} < 20$). Hence the idea to develop a new trigger capable to perform efficiently with lower thresholds on the lepton p_{T}. It allows improving the sensitivity of the analysis by increasing the signal acceptance. To control the trigger rate, a threshold on the L1 missing transverse energy was applied. The main challenge in this approach was to characterize the missing transverse energy reconstructed at Level 1, and it represented my direct contribution in the analysis. The application of this trigger has been fully analysed in the $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ channel [119]. Indeed, emphasis had been logically been put on this channel, as among the semileptonic channels, it is the most sensitive. It was not tested in the fully leptonic channel, as they represent decay modes with smaller signal yields because of the τ into leptons decay branching ratio and because di-lepton trigger had low threshold.

4.1 Introduction to the Run 1 $H \rightarrow \tau\tau$ trigger strategy

The search for the Higgs boson decaying $\tau\tau$ is carried out in five independent final states: $\mu\tau_h + X$, $e\tau_h + X$, $e\mu + X$,$\tau_h\tau_h + X$ and $\mu\mu + X$, where τ_h denotes a hadronic decay of a *tau*. Different categories in function of the number of jets in the final state and the p_{T} of the reconstructed hadronic τ are defined in order to exploit the different production modes. Despite the moderate branching ratio of the τ into electron or muon (17% each), the $\ell + \tau_h$ channel is particularly interesting as it allows a cleaner event selection, both online and offline. However, It should be noted that a fraction of the τ energy is taken by the two neutrinos. The electron or muon have a relatively low p_{T}. For a 125 GeV/c^2 Higgs boson produced by gluon fusion, only 47% of the muons have a p_{T} larger than $20\,\mathrm{GeV}/c$, as required in the "standard" analysis [71] and trigger selection and illustrated in Fig. 4.1. Hence, the idea to consider in addition muons

with $10 < p_{\mathrm{T}} < 20\,\mathrm{GeV}/c$ which represent 29% of the events.



(a) Sans coupure sur $\not{E}_T^{\mathrm{L1}}$ (b) Avec coupure sur $\not{E}_T^{\mathrm{L1}}$

Figure 4.1: *p_{T} distribution for muon coming from simulated $H \rightarrow \tau\tau$ events produced via gluon fusion: without (Left) and with (Right) the cut on the missing transverse energy. The orange area in the the plots represent the contribution given to the $\mu_{soft}\tau_h$ channel (p_{T} between 9 and 20 GeV) while the blue-green-dashed histogram represents the p_{T} region for the $\mu\tau_h$ channel.*

To do so, new triggers with a low p_{T} lepton threshold have been introduced. To keep the Level 1 trigger rate at a manageable level, it turned out to be necessary to introduce a cut on the missing transverse energy at Level 1. In this chapter, I will first describe the new triggers, before concentrating on the computation of the missing transverse energy at Level 1, hereafter called L1ETM. The data-simulation agreement for L1ETM is then studied demonstrating the need to apply corrections in the simulation. A procedure to correct L1ETM is then proposed and its performance are assessed. The systematic uncertainties on the trigger efficiency are then discussed before presenting the trigger turn-on curves. Finally, the impact on the analysis will be summarized.

4.2 New trigger for soft-lepton analysis

Two new triggers dedicated to the soft lepton analysis have been introduced in Fall 2012. For the muon channel, the p_{T} threshold for the muon is $8\,\mathrm{GeV}/c$ while that of the τ is $20\,\mathrm{GeV}/c$, leading to the `HLT_IsoMu8_eta2p1_LooseIsoPFTau20_L1ETM26` trigger path: an isolated muon with $p_{\mathrm{T}} > 8\,\mathrm{GeV}$ with $|\eta| < 2.1$ and a loose isolated PF τ with a $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are required together with a missing transverse energy larger then 26 GeV. This HLT filter is actually seeded by `L1_Mu7er_ETM26` or `L1_Mu7er_ETM20`, that require a muon with a $p_{\mathrm{T}} > 7\,\mathrm{GeV}$ and a missing transverse energy larger then 26 GeV or 20 GeV, depending on the instantaneous luminosity. Since the probability for a jet to fake an electron is larger than for

Trigger name	L1ETM threshold	Total luminosity
HLT_IsoMu8_eta2p1_LooseIsoPFTau20_L1ETM26	26 GeV	$4.9\,\text{fb}^{-1}$
HLT_IsoMu8_eta2p1_LooseIsoPFTau20_L1ETM20	20 GeV	$2.4\,\text{fb}^{-1}$
HLT_Ele13_eta2p1_WP90Rho_LooseIsoPFTau20_L1ETM36	36 GeV	$4.9\,\text{fb}^{-1}$
HLT_Ele13_eta2p1_WP90Rho_LooseIsoPFTau20_L1ETM30	30 GeV	$2.4\,\text{fb}^{-1}$

Table 4.1: *Integrated luminosity collected with each trigger and the different L1 seeds*

a muon, for the equivalent electron triggers, the p_{T} thresholds are higher, resulting in the HLT_Ele13_eta2p1_WP90Rho_LooseIsoPFTau20_L1ETM36 path, that require an electron with a $p_{\text{T}} > 13\,\text{GeV}$ and $|\eta| < 2.1$ and a loose isolated PF τ with a $p_{\text{T}} > 20$ together with a missing transverse energy smaller then 36 GeV. This HLT filter is seeded by L1_IsoEG13er_ETM36 or L1_IsoEG13er_ETM30, that require an electron with a $p_{\text{T}} > 13\,\text{GeV}$ and a missing transverse energy smaller then 36 GeV or 30 GeV. The integrated luminosity collected with the triggers is 7.3 fb^{-1}, the sharing between the different seeds is presented in Table. 4.1.

As these new triggers are not present in the simulation, the trigger efficiency is decomposed in the muon channel as the product of the tau-trigger, muon-trigger and L1ETM efficiencies as shown in Eq. 4.1. A similar decomposition can be done for the electron trigger. In the rest of the document, only the muon channel is considered.

$$\epsilon_{\text{HLT}} = \epsilon_{LoosePFTau20} \times \epsilon_{IsoMu8} \times \epsilon_{(\text{L1ETM}>20)\ (\text{resp. } 26)} \tag{4.1}$$

4.3 Samples, datasets and event selection

Since the muons that cross the calorimeters are at the minimum of the ionization, the response of the calorimeters does not depend on the muon transverse momentum. Therefore, the HLT_IsoMu17_eta2p1_LooseIsoPFTau20 path is seeded by a L1 muon requiring a muon with $p_{\text{T}} > 17\,\text{GeV}/c$ and requires a τ candidate with $p_{\text{T}} > 20\,\text{GeV}/c$ at HLT, and with no requirement on L1ETM, and it can be safely used to study the HLT_IsoMu8_eta2p1_LooseIsoPFTau20_L1ETM26 trigger even though the minimal muon p_{T} is higher.

The following samples simulated with a centre-of-mass energy of 8 TeV and with a pile-up scenario of ~ 20 reconstructed primary vertices in the event are used: $\text{Z} \rightarrow \tau^+\tau^-$, $W^{\pm}$+jets, $t\bar{t}$, multiboson and Higgs boson signal.

Unless specified in the text, all the events considered in the study described in this chapter pass the HLT_IsoMu17_eta2p1_LooseIsoPFTau20 trigger. In addition, offline selection criteria,

described in Ref. [71] are applied to the muon and the tau. Both the muon and the tau are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and $|\eta| < 2.1$; in addition they must have opposite charge.

With this simple event selection, the sample thus obtained is dominated by $\mathrm{Z} \rightarrow \tau\tau$, W+jets, QCD multijet events, and to a smaller extent of $t\bar{t}$ events. As it is illustrated in Fig. 4.2, where the muon p_{T} and tau p_{T} spectra are shown, the normalization obtained as in Ref. [71], is rather good.

The multijets background (QCD), which appears in several plots of this chapter, is evaluated from data by taking same-sign events (the τ's reconstructed for signal events is must be of opposite sign, while, in QCD events both τ's can have the same charge sign), relaxing the lepton isolation, imposing a transverse mass cut $m_{\mathrm{T}} < 40\,\mathrm{GeV}/c^2$; and subtracting the contribution of the $W^{\pm}$+jets background. More details on the QCD estimation through data-driven method can be found in Ref. 7.5.

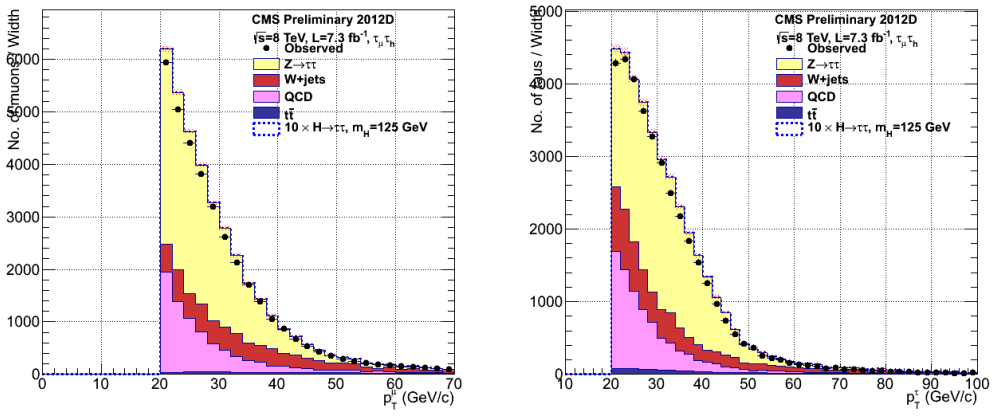


Figure 4.2: *Left: muon transverse momentum. Right: tau transverse momentum. The data are shown in black dots, and the various contributions of background and signal are represented in a stacked histogram.*

4.4 Calorimeter $\vec{E}_T^{miss}$ at L1

The Missing transverse energy reconstructed at the Level 1 is computed in the Global Calorimeter Trigger (GCT) from the Level 1 trigger regions [120], which roughly represents $\Delta\eta \times \Delta\varphi = 0.348$, i.e. 4×4 towers including the ECAL and the HCAL detectors. Since the GCT is emulated in simulation, the L1ETM quantity is computed during the digitization process of the simulation and available at the AOD level.

The trigger efficiency directly depends on the efficiency of the cut on L1ETM applied at Level 1, it is therefore of utmost of importance to study the agreement between the data and

the simulation for this quantity. Indeed, discrepancies between data and simulation can affect the event yields estimation. As can be seen in Fig.4.3, L1ETM tends to be overestimated in the simulation. This means, unfortunately, that the increase of acceptance brought by this new trigger will be smaller in the data with respect to what anticipated from the simulation. The understanding of the cause of the observed discrepancy is beyond the scope of this thesis. It is however interesting to study in parallel the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed offline, for which corrections improving the data-simulation agreement exist (see Chapter 3). To ensure apple-to-apple comparisons, the offline $E_{\mathrm{T}}^{\mathrm{miss}}$ is computed in the same acceptance as L1ETM, i.e. with ECAL and HCAL barrel and endcaps, without the HF detector and is therefore called CaloMetNoHF.

The same trend as for L1ETM can be observed for the CaloMetNoHF distribution in Fig.4.4-left. The strategy to correct the L1ETM in the simulation results from the assumption that the data-simulation differences for L1ETM and CaloMetNoHF have a common origin. In a first

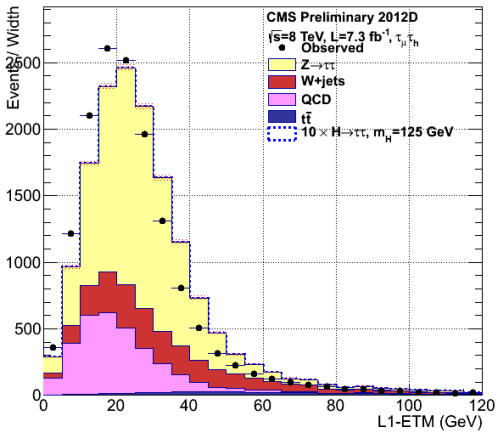


Figure 4.3: *Missing transverse energy as reconstructed with the calorimeters at Level 1 (L1ETM).*

step, the corrections applied to CaloMetNoHF are applied to L1ETM as explained below.

4.4.1 Leading Order correction

The large sample of Z $\rightarrow \mu\mu$+jets events allows, through the study of the p_{T}-balance, precise jet energy corrections to be derived for the data and the simulation, thus equalizing the response of the jets in the data and in the simulation. The absolute energy correction applied to each jet by these Jet Energy Scale Corrections (JEC) in the simulated samples is applied to the $E_{\mathrm{T}}^{\mathrm{miss}}$ (more details are given in the Section 3.4.3). As visible in Fig. 4.4 , the JEC corrections restore a good agreement between the simulation and data for CaloMetNoHF.

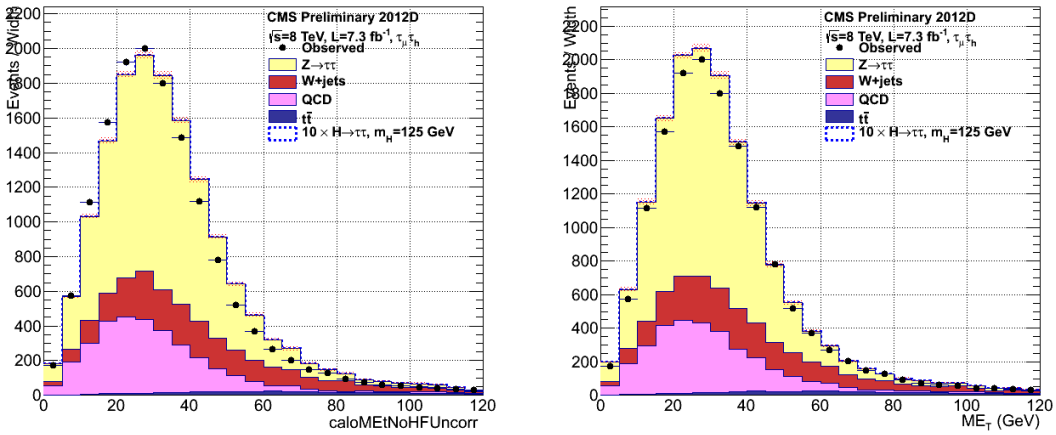


Figure 4.4: *Effect on caloMEtNoHF of the JEC corrections: before (Left) and after (Right).*

Even if the JEC corrections are applied on a jet-by-jet basis, the global factor applied to the $E_{\mathrm{T}}^{\mathrm{miss}}$ can be obtained by making the ratio of CaloMetNoHF after and before the JEC corrections. As explained above, this correction factor is then applied to L1ETM as follows:

$$\mathrm{L1ETM_{LOCorr}} = \mathrm{L1ETM} \cdot \left(\frac{\mathrm{CaloMETNoHFCorr}}{\mathrm{CaloMEtNoHF}} \right)$$

The results of this "leading-order" correction is presented in Fig.4.5 where the corrected L1ETM for simulation has been superimposed to data. As expected, and comparing with Fig.4.3, the agreement is improved but not at the level of that obtained with CaloMetNoHF. As a result, additional corrections are needed on top of $\mathrm{L1ETM_{LOCorr}}$ and are described in the next section.

4.4.2 Residual corrections

The strategy to further correct L1ETM in the simulation to improve the agreement with data is done two steps. In a first step, scale corrections are derived; and in a second step, a smearing is applied to adjust the resolution of L1ETM in the simulation to that of the data. The corrections are derived as function of CaloMetNoHF, but the dependency on the global event activity, measured by Sum_{ET}, i.e. the sum of the transverse momenta of all the Particle Flow candidates, is also investigated. Indeed, since the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is known to depend on Sum_{ET}, a possible dependency of data or simulation on the global activity cannot be neglected. In this section, the data-driven multijet component is subtracted from the data and the resulting histograms are compared to the sum of the simulated samples.

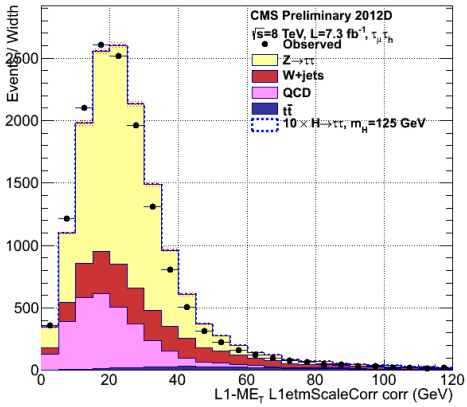


Figure 4.5: *The inclusive distribution of the L1ETM after the application of the leading order correction.*

The procedure consists in defining 10 bins in CaloMetNoHF between 0 and 80 GeV and and 9 bins in Sum_{ET} between 0 and 1300 GeV. The size of the bins in Sum_{ET} has been adjusted so that the statistics are similar among the different bins. For each bin, the $\text{L1ETM}_{\text{LOCorr}}$ distribution is fitted with a gaussian; the behavior of mean value and sigma, taken from the fit parameters, have been plotted as functions of Sum_{ET}. An example of the L1ETM distribution in a particular bin of CaloMetNoHF and Sum_{ET} together with the gaussian fit can be found in Fig.4.6. The procedure is repeated for all the bins of Sum_{ET}, and the result is shown in Fig.4.7 where the ratio of fitted mean values in the data and in the simulation is shown as function of Sum_{ET} for a given CaloMetNoHF bin. No dependency of the response nor of the resolution on Sum_{ET} can be observed. Concentrating on the left plot of Fig.4.7, one can see that the $E_{\text{T}}^{\text{miss}}$ in the data is on average about 7% lower than in the simulation. Repeating the procedure in the various bins of CaloMetNoHF results in the left plot of Fig. 4.8, where it can be observed that the difference on the scale does not depend on CaloMetNoHF either. Therefore a constant fit provides the numerical factor, R, of about 0.93 to be applied to $\text{L1ETM}_{\text{LOCorr}}$ to correct the scale. One thus gets $\text{L1ETM}_{ScaleCorr}$=$\text{L1ETM}_{\text{LOCorr}} \cdot R$.

Once this numerical multiplication factor, R, is applied and the procedure repeated, the right plot of Fig. 4.8 is obtained. As expected, the response is correctly brought back to unity. It becomes therefore relevant to compare the resolution between data and simulation. The same procedure as before, making bins of Sum_{ET} and CaloMetNoHF is carried out considering the Gaussian width. Similarly to what was observed with the scale, the data-MC discrepancy is found to be independent of Sum_{ET} but exhibits a non-negligible dependency on CaloMetNoHF as illustrated in Fig. 4.9-left. On the same plot, the log-shape function used to parametrize this dependency is shown. This function is used to adjust the resolution of L1ETM in the simulation

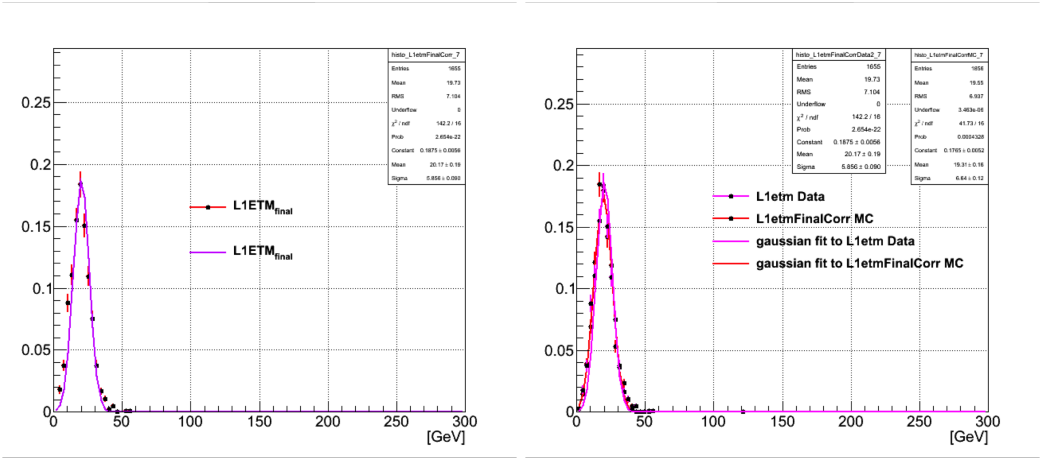


Figure 4.6: *Example of Gaussian fit to L1ETM distribution in the data (Left) and in the simulation (Right) for CaloMEtNoHF between [25-30 GeV] and in* Sum_{ET} *between [588-672 GeV].*

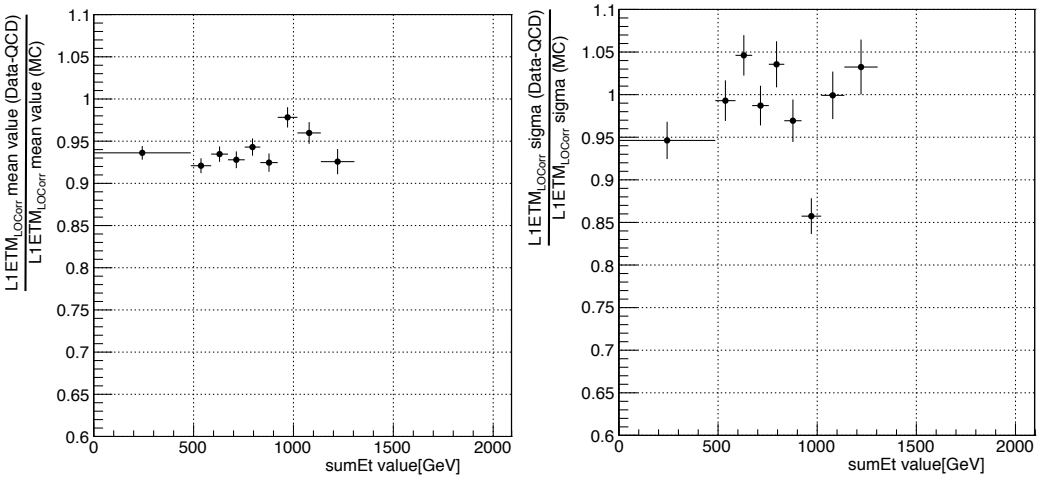


Figure 4.7: *Example of the behavior of the mean and sigma values as function of sumEt, obtained from the gaussian fit in the slice of CaloMEtNoHF between* $25 - 30\,GeV$ *as a function of* Sum_{ET}*.*

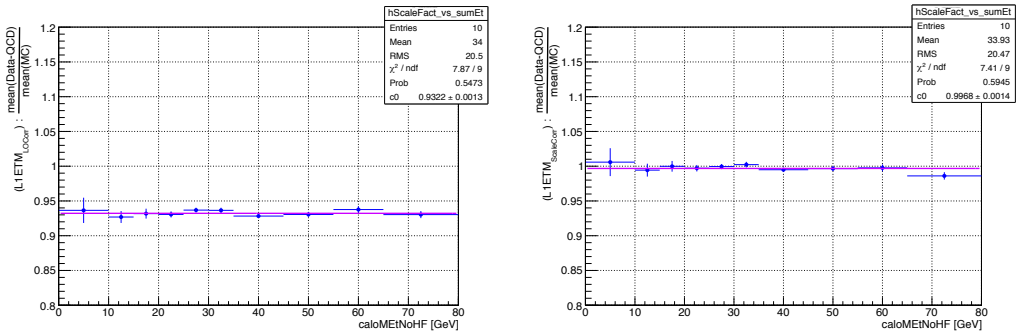


Figure 4.8: *Behavior of the average mean values ratio between data and MC as function of caloMEtNoHF. Left: before scaling; Right: after scaling.*

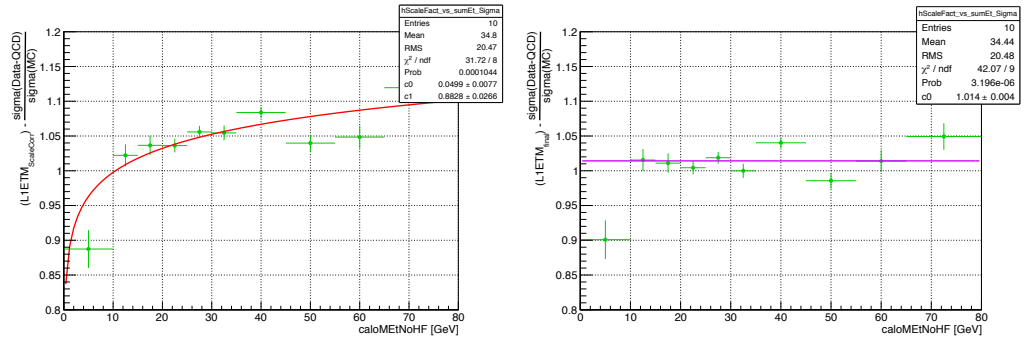


Figure 4.9: *Sigma ratio between data and MC distribution as function of caloMEtNoHF. Left: before resolution correction; Right: after resolution correction.*

according to the following formula:

$$\text{L1ETM}_{\text{final}} = \text{L1ETM}_{\text{ScaleCorr}} + \left[\text{L1ETM} \cdot \left(\frac{\text{CaloMEtNoHFCorr}}{\text{CaloMEtNoHF}}\right) \cdot R - K \cdot \text{CaloMEtNoHF}\right] \cdot H(\text{CaloMetNoHI} \tag{4.2}$$

where the R factor is the scale correction factor applied to L1ETM_{LOCorr}, H is the log-shape function parametrizing the dependency of the resolution disagreement. The term $K \cdot CaloMEtNoHF$ provides an expectation, in the simulation, of the mean value of L1ETM corresponding to a measured value CaloMetNoHF. Thus, the procedure will not affect the mean value, the mean value of $\text{L1ETM}_{\text{final}}$ will be close to that of $\text{L1ETM}_{\text{ScaleCorr}}$. The K factor is obtained from a linear fit to the 2D distribution of $\text{L1ETM}_{ScaleCorr-x}/\text{CaloMetNoHF}_x$ as a function of CaloMetNoHF (see Fig.4.10). The use of the projection on the x-axis allows the tails of the ratio to be equally distributed between the positive and negative values and to cancel when doing the fit.

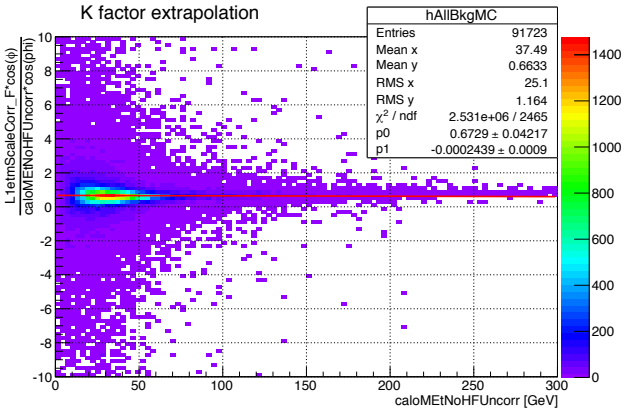


Figure 4.10: *Response of L1ETM relative to CaloMetNoHF as a function of CaloMetNoHF. The linear fit of the scatter plot is also shown.*

After applying the resolution correction, the ratio of the resolution in the data and in the simulation is changed from Fig. 4.9-left to Fig. 4.9-right where it can be observed that except in the first bin of CaloMetNoHF which, as far as L1 trigger efficiency is concerned, does not play any role, a fair agreement is obtained.

4.4.3 Summary and results

At this point, the final correction applied to Level-1 missing transverse energy is of the form:

$$\begin{cases} \mathrm{L1ETM_{final}} = \mathrm{L1ETM} \cdot \dfrac{\mathrm{CaloMEtNoHFCorr}}{\mathrm{CaloMEtNoHF}} \cdot R + \\ + H \cdot \left[\mathrm{L1ETM} \cdot \left(\dfrac{\mathrm{CaloMEtNoHFCorr}}{\mathrm{CaloMEtNoHF}} \right) \cdot \mathrm{R} - \mathrm{K} \cdot \mathrm{CaloMEtNoHF} \right] \\ R = 0.932 \pm 0.001 \\ H = -(0.117 \pm 0.004) + (0.050 \pm 0.008) \cdot log\,(\mathrm{CaloMEtNoHF}) \\ K = 0.67 \pm 0.02 \end{cases} \tag{4.3}$$

The full chain of corrections is summarized in the plot presented in Fig. 4.11 where the data and several steps of the corrections are presented for a given bin of CaloMetNoHF and Sum_{ET}. The comparison of the data (pink curve) and the L1ETM, as present in the simulation without any correction (blue curve) demonstrates the need for a correction. The Leading Order correction and the scale correction are applied together and give the red line. As can be seen from the parameters of the fit, the mean value of the corrected simulated L1ETM matches the data but the down-scaling simultaneously reduces the sigma of the distribution which is now too slightly small. The resolution correction gives the red curve which matches well the data curve both for the mean and the sigma. The performance of the full-fledged corrections can

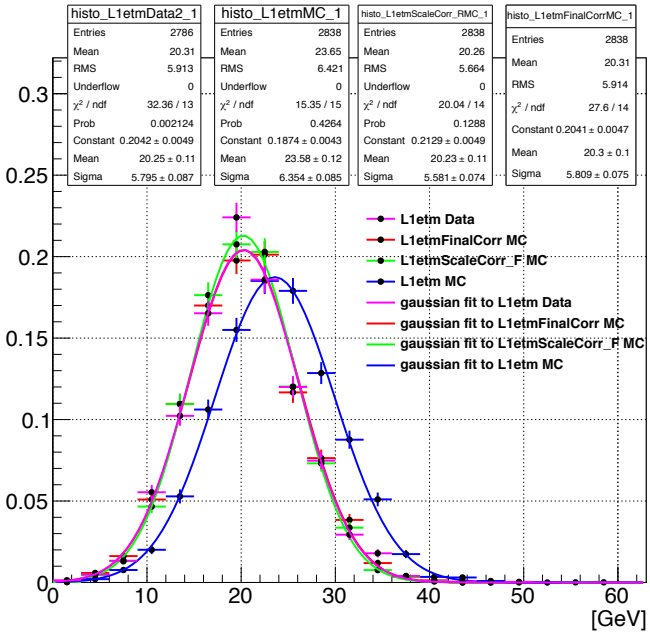


Figure 4.11: *Different steps of the L1ETM correction scheme for a given bin of CaloMEtNoHF [25-35 GeV] and Sum_{ET} [588-672 GeV].*

also be observed on the inclusive distribution, comparing the L1ETM distribution for data and simulation before correction (Fig.4.3) and after correction (Fig.4.12).

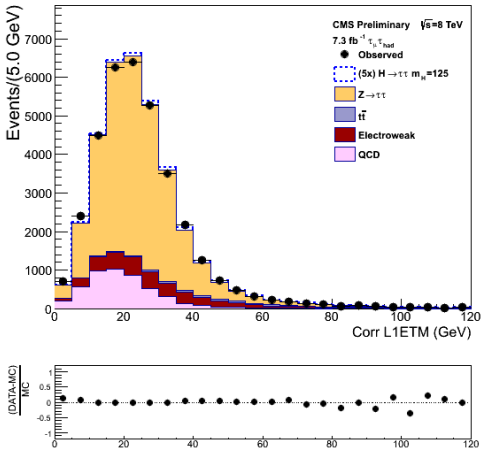


Figure 4.12: *Level-1 missing transverse energy inclusive distribution after full-fledged correction.*

4.5 Systematic uncertainties

Setting up a series of correction scheme for the simulated L1ETM is important because it has a direct impact on the trigger efficiency; it however implies also estimating the associated systematic error. Several sources of systematic errors have been identified and are discussed below.

The JECs come with an associated error in the form of an upper and a lower value for the correction factors. The impact on the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is obtained by applying the upper and lower corrections to the jets and propagating them to the $E_{\mathrm{T}}^{\mathrm{miss}}$. One thus get an upper and lower value of the $E_{\mathrm{T}}^{\mathrm{miss}}$. The $E_{\mathrm{T}}^{\mathrm{miss}}$ correction being the first ingredient of the L1ETM correction procedure, the JEC systematic is consequently a direct source of systematics. The effect is illustrated in Fig. 4.13 where the inclusive L1ETM distribution is plotted, in the simulation, after correction, for the central value, the upper and lower corrections. The impact on the trigger efficiency of the JEC corrections highly depends on the shape of the L1ETM distribution, and therefore on the sample composition. With the inclusive selection used throughout this chapter, the effect is of order 8% (resp. 5%) with a 26 GeV (resp. 21 GeV) threshold. These numbers cannot be used as such in the analysis; they however give an idea of the magnitude of the effect. The actual impact should be evaluated re-running the analysis feeding the formula 4.3 with the upper and lower values of the corrected CaloMetNoHF and quantify the impact on the event

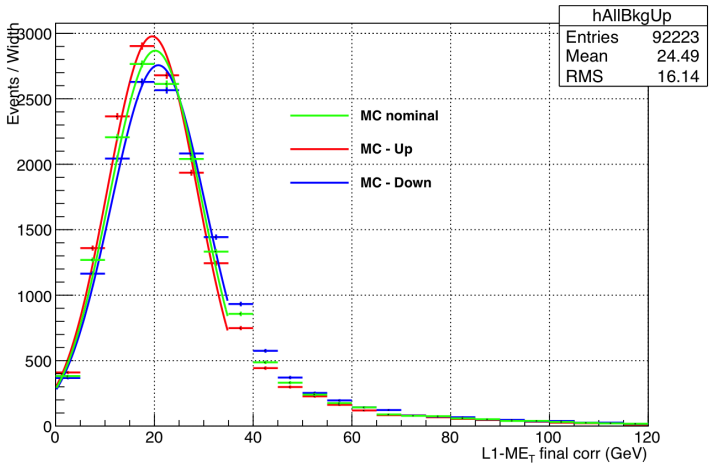


Figure 4.13: *Impact of the JEC uncertainties on L1ETM in the simulation.*

yield.

The other sources of systematic errors are found to be negligible compared to the one which originates from the JEC uncertainty. In particular, the propagation of the errors related to the fit procedure turns out to be more than one order of magnitude smaller than the error resulting from the JEC uncertainty. Similarly, the spread of the points obtained after the scale corrections (Fig. 4.8-right) above the region of interest of the trigger threshold, i.e. above 20 GeV, is at the percent level or below and is therefore negligible with respect to the error induced by the JEC.

4.6 Trigger efficiency curves

The proper behavior of the L1ETM corrections already demonstrated in Fig. 4.12 can be represented on the so-called turn-on curve showing the trigger efficiency as a function of an offline reconstructed quantity. In the present case, the CaloMetNoHF variable is used as it is the offline quantity which is most correlated with L1ETM. The turn-on curves for a 20 GeV and a 26 GeV threshold, respectively, are shown in Fig.4.14 and 4.15. In these plots, the same inclusive selection as before is used, but the QCD contribution (which is taken from data, and therefore without any correction) is counted together with the simulated samples and no longer subtracted from the data. The red dashed line represents the trigger efficiency obtained directly from the simulated L1ETM quantity. The efficiency after correction is represented by a continuous red line which is hard to distinguish from the data dashed black line. These curves are the fitted *erf* functions of the efficiency plots. The error associated to the fit is represented by a green band. To be visible, the height of this band has been magnified 7 times, which is a good indication

that the *erf* function is well adapted. Finally, the impact of the systematic errors described in Section 4.5, dominated by the error associated to the JEC, is represented by a yellow band.

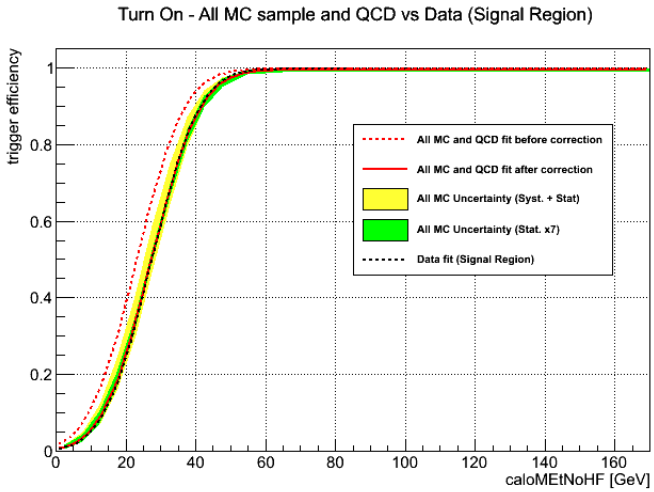


Figure 4.14: *Turn on as a function of corrected CaloMEtNoHF: comparison using corrected and uncorrected L1ETM in MC samples for a trigger threshold fixed at 20 GeV.*

4.7 Conclusion

The use of a L1ETM cut to control the rate in the new triggers introduced in the context of the H $\rightarrow \tau\tau$ analysis in the $\ell + \tau$-hadronic channel with low p_{T} leptons, requires the agreement between data and simulation to be studied. A set of corrections to be applied in the simulation has been proposed and was demonstrated to perform well. The associated systematic uncertainty has been studied and is found to be dominated by the JEC uncertainty. A method to evaluate this systematic error for given selections has been proposed. Finally, turn-on curves comparing the trigger efficiency in the data and in the simulation have been obtained and have proven the corrections to perform well. In order to investigate the contribution of the soft lepton analysis in the official $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ analysis we can compare the exclusion limits carried out with and without the contribution of the soft lepton analysis [119]. They are illustrated in Fig. 4.16. The $\mu\tau_h$ channel can exclude at $\frac{\sigma}{\sigma_{SM}} \sim 4.4$ at 120 GeV. Combining the soft-lepton with the standard $\mu\tau_h$ analysis, there is an improvement of the 2% in the expected 95% C.L. exclusion limit [119]. This result is very promising considering that only a fraction of the total luminosity has been used and it is one research line to improve the trigger strategy in τ channels for the Run 2. It is however not the only one. Indeed, as will be explained in the next Chapter, an important upgrade of the L1 system has been achieved and allows substantial progress on the τ

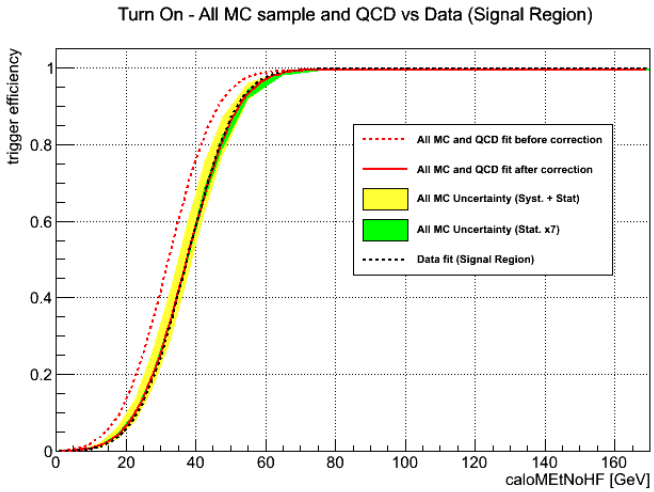


Figure 4.15: *Turn on as function of corrected CaloMEtNoHF comparison using corrected and uncorrected L1ETM in MC samples for a trigger threshold fixed at 26 GeV.*

identification at L1. It might thus turn out to be more efficient to design soft lepton + hadronic τ triggers that will be easier to commission.

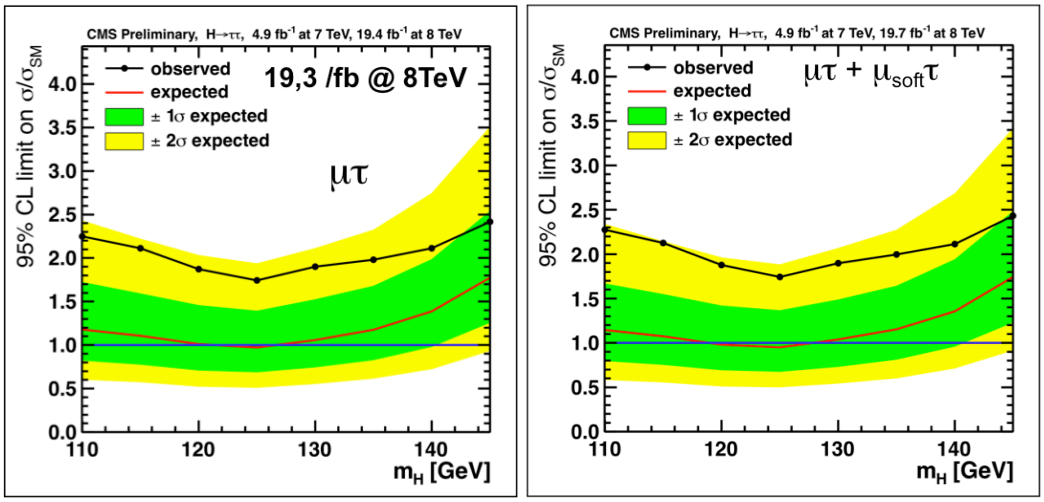


Figure 4.16: *Exclusion limits for the $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ evaluated for the "standard analysis" (Left) and for the combined "standard+Soft analysis" (Right). From [119].*

Chapter 5

A new L1-τ trigger algorithm for the CMS stage-2 upgrade

The CMS experiment implements a sophisticated two-level online selection system that achieves a rejection factor of nearly 10^5. For the Run 2, the centre-of-mass energy of the LHC collisions has been increased to 13 TeV and the instantaneous luminosity will increase reaching $2 \cdot 10^{34}$ $cm^{-2}s^{-1}$. To guarantee a successful and ambitious physics program under this intense environment, the CMS Trigger and Data acquisition system are being consolidated. In particular the L1 calorimeter Trigger hardware and architecture are upgraded, benefiting from the recent μTCA technology allowing sophisticated algorithms to be deployed, better exploiting the calorimeter granularity and opening the possibility of making correlations between different parts of the detector. Given the enhanced granularity provided by the new system, I have developed an optimized τ algorithm for the new, stage-2, L1 system. This algorithm is implementing an innovative dynamic clustering technique that has been primarily developed to trigger on electrons and photons. The selection of a hadronically decaying τ's, giving multiple decay channels, represents a real challenge for an electronic trigger system. In addition, to satisfy both physics performance, and hardware constraints, the L1 τ algorithm requires specific adaptation of the electron/photon algorithm, also completely redesigned for the CMS stage-2 upgrade. The performance of this τ trigger will be demonstrated, both in terms of efficiency and rate reduction. The different handles to control the rate in different pile-up scenario will be described. Finally, the plans for the commissioning with the first Run 2 data will be presented and the expected impact on the physics potential assessed.

5.1 Introduction

The search of new particles, in particular, in the context of the supersymmetry is a major goal of the LHC Run 2. However, so far no signs of supersymmetric particles have been found at the

LHC or anywhere else. The higher energies but also the higher luminosities will allow searching for processes with small cross sections and for particles of higher masses. The detailed study of the properties of the Higgs boson as well as precision measurement of the Standard Model will be carried on simultaneously and require to keep trigger energy thresholds as low as possible. The trigger is therefore a key item for the success of the Run 2. As cross sections generally rise with energy, this change by itself will result in an interaction rate increase of roughly a factor one to two (respectively for minimum bias events and for the Higgs boson production). At the same time, the luminosity will increase from the value of $0.7 \times 10^{34} cm^{-2} s^{-1}$ reached in 2012 to $2 \times 10^{34} cm^{-2} s^{-1}$ in 2015. With the expected LHC performance, the trigger rate will increase $\times 2$ due to the energy increase, and $\times 3$ due to the luminosity increase. Hence, for a given algorithm and given thresholds, the rate will increase by a factor of 6. In order to maintain an acceptable event rate for the $\mu + \tau$ cross trigger, letting the L1-μ algorithm unchanged, it is mandatory to reduce the rate by a factor $\sim O(6)$ acting on the L1-τ requirement. However, with the L1-τ trigger algorithm available during Run 1, the threshold on the τ leg would have to be set too high. Moreover, in these conditions, one of the additional challenge for the LHC experiments is to deal with the increased pile-up (see Fig. 5.1) even if a $25\,ns$ bunch spacing will mitigate the amount of pile-up.

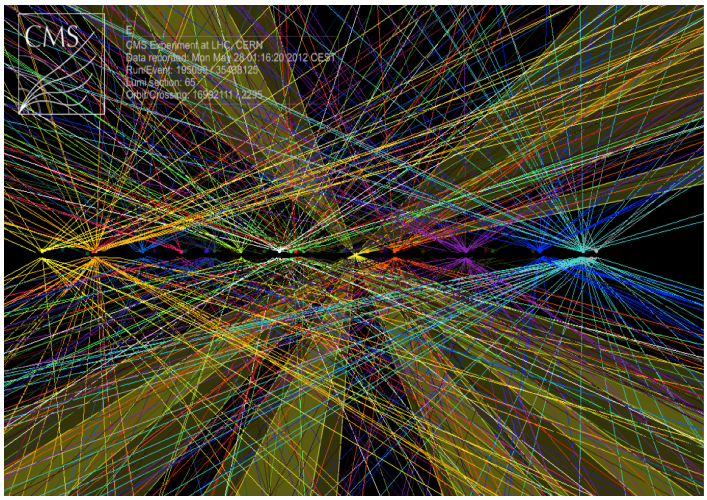


Figure 5.1: *Pileup the number of individual proton-proton collisions in each event is getting of growing concern as the LHC ramps up its luminosity. The figure shows an event with a large number of reconstructed vertices*

In Run 1, CMS had to deal with about 20 individual proton-proton collisions in each bunch crossing (each event). In the long term this number may go up to 100 or more individual collisions. Identifying a given type of particle within such a busy environment is a major challenge, and it should be stressed that the Run 1 Level 1 τ trigger could not be used as such in the Run 1 analyses. Indeed, as the energy of the τ was estimated from the energy collected in large

region, jets were easily giving rise to τ candidates and shape vetoes and isolation criteria had to be introduced to limit the rate. As a side-effect they significantly reduced the efficiency, especially at high energy. Therefore joint L1 τ+jet triggers had to be used to recover efficiency, in particular at high p_{T}.

5.2 Phase-1 L1-trigger upgrade

Nowadays, the LHC is running with a collision energy of 13 TeV and a luminosity constantly increasing in order to reach the nominal value of $2 \cdot 10^{34}\ cm^{-2}s^{-1}$. The LHC ended its Run 1 operations early 2013 when the "Long Shutdown 1" (LS1) started. The CMS collaboration has been using this period for initiating the first phase of the detector upgrade, named Phase-1 Upgrade. Certain detector parts, e.g. outer muon chambers in the forward region, have been added and parts of the trigger and data acquisition electronics have been replaced. This work will continue in parallel with the data taking. Some of the new electronics will be run in parallel for commissioning purposes through 2015 and CMS will switch over to using these new systems in production mode during LHCs annual technical stop at the beginning of 2016. The detailed upgrade plans for Phase 1 have been documented in a Technical Design Report [121].

With LHC collision rates rising, the goal of the upgraded trigger is to keep the rates for data to be recorded at a manageable level while conserving a high efficiency for the individual physics processes. The relatively low mass of the observed Higgs boson makes it mandatory to keep low energy thresholds in the trigger in order to remain sensitive to the Higgs decay products and be able to carry out precision studies of Higgs parameters. Several approaches will have to be combined to reach this goal:

- better resolution of the transverse momentum (p_{T}) or transverse energy (E_{T}) and better precision in the geometrical coordinates (azimuth ϕ and pseudorapidity η) of trigger objects; access to the trigger tower information and thus to a finer granularity compared to Run 1;
- more complex operations at an early level, such as pileup subtraction in the calorimeter trigger;
- combination of various subdetector data (such as data from redundant muon systems) at an early stage;
- more sophisticated trigger algorithms capable to exploit the correlations between different types of trigger data, calculation of invariant masses or transverse masses of pairs of trigger objects;

The present CMS trigger electronics consists largely of custom-built VME modules. The VME standard will be replaced by the new μTCA (Micro Telecommunications Computing Architecture, see Fig. 5.2). Optical links will provide increased bandwidth and less space will be needed

for bulky connectors. Use of new generations of electronics components, in particular of larger FPGAs (Field Programmable Gate Arrays) will allow to achieve higher performance with a smaller number of electronics modules. Figure 5.2 shows an example of a generic μTCA board using a single powerful FPGA.

Figure 5.2: *The MP7 μTCA board.*

5.2.1 Architecture of the upgraded calorimeter trigger

The calorimeter trigger architecture implemented during the LHC Run 1 is presented in Section 2.3.5 of Chapter 2. In the context of the phase-1 trigger upgrade, the calorimeter trigger upgrade is of crucial importance. A two-layer system has been built, adopting a strategy based on the time multiplexing. For a given bunch crossing, all the information from the detector are sent to a single FPGA wired on a main processor board (MP7) (see Fig. 5.3). The subsequent bunch crossing are processed in a similar way by others FPGA's [122] [123]. There are in total 8 MP7 boards processing events at this layer. This upgrade is known as phase-1 stage-2 calorimeter trigger upgrade. The stage-2 architecture will be used for the data-taken from 2016 and, until that time, it will be running in parallel to an upgraded version of the trigger system used during the Run 1 (phase-1 stage-1 trigger upgrade). The stage-1 algorithms are going to be used from August 2015 to the end of 2015 and it represents an intermediate step to get to the stage-2 architecture.

Layer 1

The first layer will consist of custom boards and a custom backplane to receive the trigger primitives (TPG) via optical fibre from the calorimeters and prepare data for Layer 2. At this stage it is possible to compute already some basic quantities, e.g. the ratio of the energy deposited in the ECAL and in HCAL, in addition to the energy and position of the TPGs.

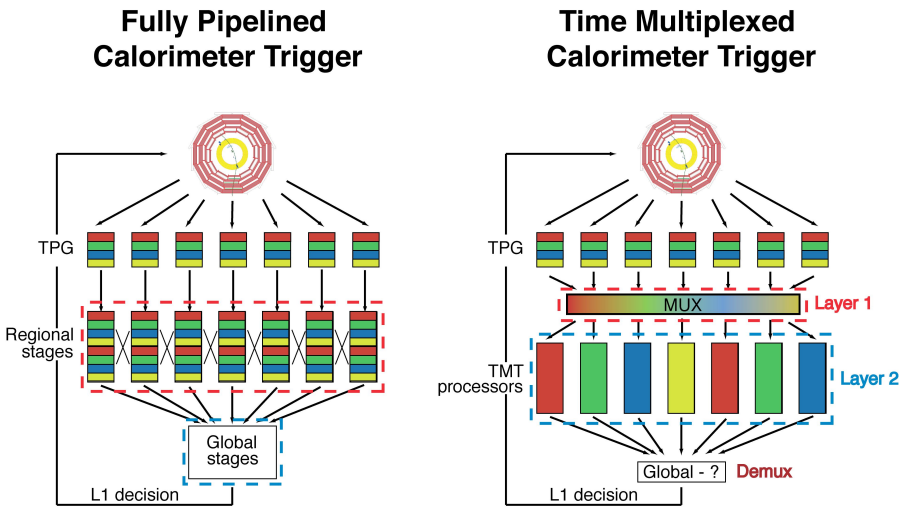


Figure 5.3: *Diagram of the time-multiplexed trigger*

The layer 1 implements 36 Calorimeter Trigger Processor (CTP7) boards. Each CPT7 covers a portion of the detector. The output of all the 36 CPT7 for a given bunch crossing number is sent to 1 MP7 board. Different MP7 boards process the information coming from all the CPT7 for the subsequent bunch crossing.

Layer 2

The second layer will consist of a set of custom multi-purpose cards (MP7) to receive the time-multiplexed trigger towers via optical fibre from the processing boards of the first layer. At this level the various algorithms run to reconstruct the different physics objects and the outputs are sent to the global trigger that takes the final decision.

5.3 Dataset and event selections

The new τ trigger algorithm development and its performance were obtained using both simulated samples and real data coming from the LHC Run 1 dataset. The Monte Carlo samples used for this study belong to the *Summer12* campaign, that means $\sqrt{s} = 8\,\text{TeV}$ and $< PU > \sim 20$. The process name and the event generator, are listed in Tab. 5.3.

The data come from the LHC Run 1 D-dataset during which an integrated luminosity of $7.3\,fb^{-1}$ of pp collisions at $\sqrt{s} = 8\,\text{TeV}$ has been collected by the CMS detector. For Monte Carlo studies, very loose request have been applied at the event selection level, just a $p_{\text{T}} > 5\,\text{GeV}$ and $|\eta| < 5$ of the reconstructed τ's. In addition, a matching between the MC reconstructed τ

process ($\sqrt{s} = 13\,\text{TeV}$)	subprocess	generator	N. of events
$gg \to H(M_H = 125) - AODSIM$	$H \to \tau^+\tau^-$	powheg+pythia6	478041
$VBF \to H(M_H = 125) - AODSIM$	$H \to \tau^+\tau^-$	powheg+pythia6	462681
$SUSYBBH(M_H = 160) - AODSIM$	$H \to \tau^+\tau^-$	pythia6+tauola	96611
$DYJets(M_H > 50) - AODSIM$	$Z \to \tau^+\tau^-$	pythia6	994899
$Neutrino p_T 2to 20_{gun} - GENSIM - RAW$	background	pythia6	45716756

Table 5.1: *MC sample used in the τ trigger studies.*

and the generated τ has been imposed. In the data I select the events triggered at L1 by the single μ trigger and at the HLT triggered by the presence of an hadronic τ with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.2$ and a global muon with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.1$. The events are required to contain an isolated hadronic τ and an isolated muon. Furthermore the hadronic τ is requested to pass the identification criteria and to be reconstructed in one of the possible decay modes recognized by the HPS algorithm. In order to identify the τ lepton used as reference to study the L1-cluster associated to a particular hadronic τ, a tag and probe technique has been used. In this technique, the tag is a reconstructed isolated muon with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.1$ and the probe a *medium isolated* HPS-τ with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.2$. In addition requirement on the di-τ visible invariant mass are applied ($m_{\tau\tau} \in [47.5, 72.5]$ GeV) and an anti-W cut is also applied on the transverse mass ($m_T < 40\,\text{GeV}$) in order to improve the purity of the probe sample. These requirements aim to select a sample enriched in $Z \to \tau\tau \to \mu\tau_h$ events. In order to emulate the stage-2 trigger tower reconstruction, the so-called RAW information are needed. The reconstructed quantities were accessed from the AODs.

5.4 Tau energy deposits calorimeter patterns

In order to develop en efficient algorithm for τ's exploiting the finer granularity provided by the new trigger architecture, the pattern of the τ energy deposits in the calorimeter should be studied. The variety of the τ hadronic decay modes (up to 5 particles in the final state - not counting the neutrinos) reflects in a complicated signature for the final state (see Section 3.4).

The different decay modes as well as the dependency on the initial τ p_T spectrum implies that the p_T of the charged hadrons is very spread as can be seen in Fig. 5.4. Thus the experimental challenge is to collect energy of objects that hit the calorimeters surface at different positions but coming from the same τ lepton. The goal of the new trigger architecture for τ lepton is to identify correctly the true τ leptons with the highest possible efficiency and provide criteria to distinguish them from fake τ's. To achieve this goal, it is mandatory to estimate the area of the calorimeter covered by the τ final decay products and to construct an algorithm able to deal with the different energy deposit patterns arising from the variety of the decay channels.

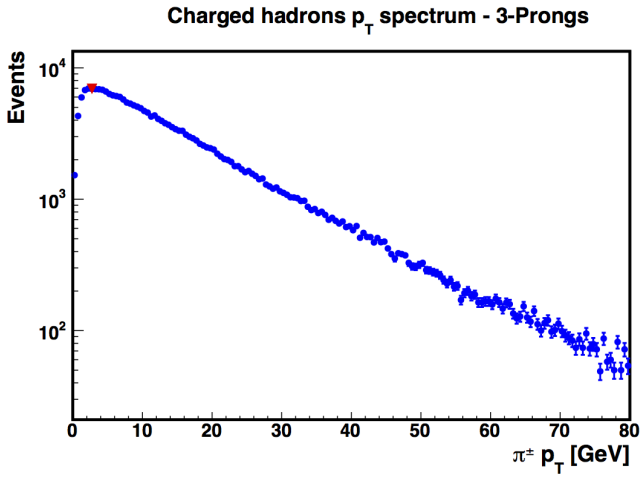


Figure 5.4: *Charged pions p_{T} distribution for pions coming from τ's decaying 3-prongs derived from MC sample of $H \to \tau\tau$ events produced through the gluon fusion process. The red triangle is the most probable value of the distribution ($\sim$ 2.7 GeV)*

Given the p_{T} spectrum, e.g. of the charged hadrons coming from the 3-prong τ decay, for τ's coming from H decays, it is possible to estimate the mean value ($\sim$ 15 GeV) and the most probable value ($\sim$2.7 GeV). Now if we consider the most probable value in the distribution, it is possible to evaluate the bending radius from the equation:

$$r = \frac{p_{\mathrm{T}}[\,\mathrm{GeV}]}{0.3 \cdot B[T]} \sim 2.36\,m \tag{5.1}$$

for a $p_{\mathrm{T}} = 2.7\,\mathrm{GeV}/c$ and a constant magnetic field of 3.8 T. Thus the distance between the charged hadron and the the straight-line extrapolation of the direction of the initial τ estimated at the ECAL entrance is:

$$d(h^{\pm},\tau) \sim \sqrt{\left(R_{ECAL} - \sqrt{R_{ECAL}^2 - \frac{R_{ECAL}^4}{4r^2}}\right)^2 + \left(\frac{R_{ECAL}^2}{2r}\right)^2} \sim 0.36\,m. \tag{5.2}$$

It means that the charged hadrons enter in the calorimeter at a distance greater then 3 trigger towers with respect to the initial direction of the τ. Now if we consider that in the 3 prong decay there are at least 2 opposite-sign charged hadrons, assuming that they have both the same p_{T} value fixed to the most probable one in according to the distribution in Fig.5.4, it follows that the maximum separation between the τ decay products (for the 3-prong decay) could exceed 6 trigger towers. Even though an extreme case has been considered, such a basic calculation justifies a deeper study concerning τ behaviours, in terms of energy deposition of its daughter particle(s), in the the calorimeters trigger towers.

5.4.1 Study of the τ footprint

From the trigger point of view, the basic unit is the trigger tower. In the barrel, each trigger tower consists in 5×5 ECAL crystal matrix in front of an HCAL cell. In the endcaps, the definition is intricate but the tower is about of the same size. A RCT region is defined as the region giving to the trigger system the output of a 4×4 trigger towers region. In Fig. 2.19 of Chapter 2 a slice of the CMS calorimeters with the definition of all the trigger towers in terms of η coverage and position is reported.

The map of the fractions of the tau energy deposited in each calorimeter trigger tower (considering both energy deposited in ECAL and in HCAL), hereafter called *footprint*, provides useful information on the global size of the clusters.

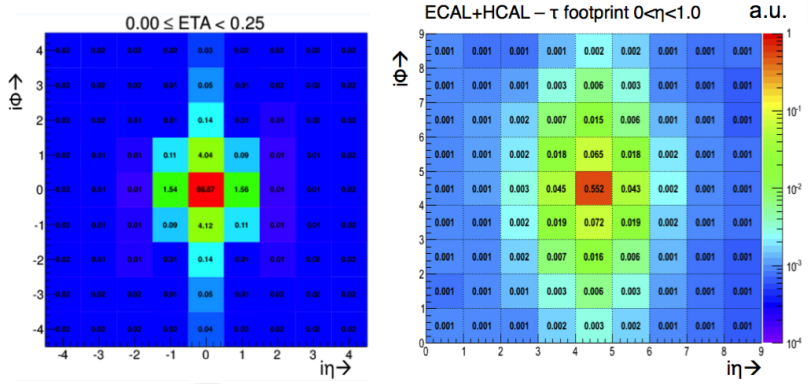


Figure 5.5: *Average electron footprint (Left) and average hadronic τ footprint (Right) obtained from 2012 data. For the electron footprint, only the energy deposited in the ECAL has been considered, while, for τ's the energy deposited in both the ECAL and the HCAL has been considered.*

In Fig.5.5 global footprints (averaged over several events) are shown in the case of an electron (left) and of a τ (right) for all reconstructed τ_h decay mode. For the τ footprint both the energy deposited in ECAl and HCAL are considered, while for the electron one only the energy depositions in ECAL are shown. As can be seen the energy fraction coming in the most energetic cell is $\sim 85\%$ of the total energy deposited in the 9×9 region in the case of the electron, while it is just $\sim 50\%$ in the case of the τ. This means that, on average, the τ footprint is larger and involves a higher number of trigger towers. Averaged footprints do not, however, give hints on the optimal shape and topology of the clusters to collect the energy deposited by the τ in the calorimeter. To get this information, the τ footprint should be obtained individually for each τ and searching for a common pattern, which will guide the design of the algorithm. Moreover, the clustering algorithm should dynamically be able to adapt to the different shapes and size of the τ energy deposits, which depend on the τ decay channel. The τ decay modes reconstructed offline with the *hadron plus strips* algorithm (HPS) are: 1-prong, 1-prong + π^0's and 3-prongs [113]. Nearly hundred of τ footprints have been visually scanned for different decay modes. A

few examples are presented in Fig.5.6.

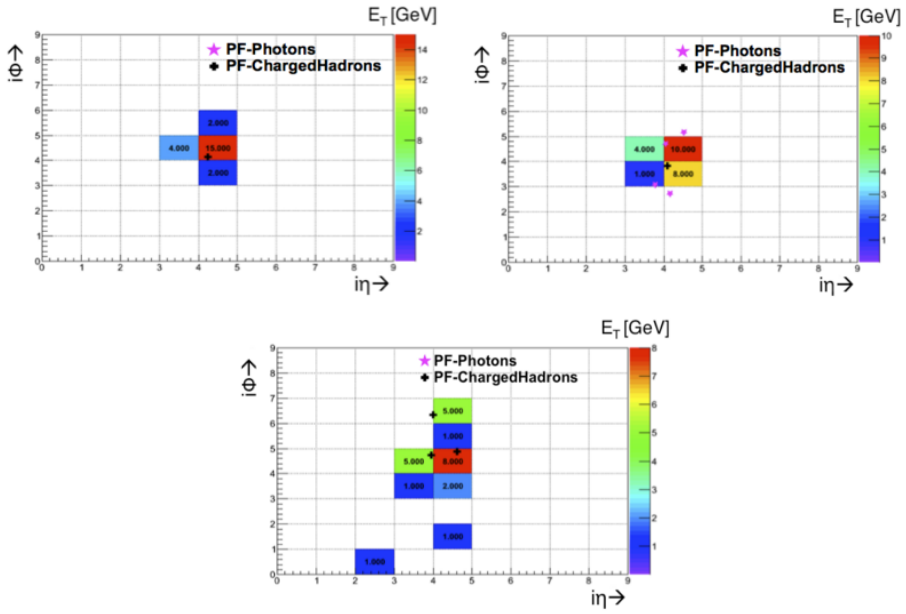


Figure 5.6: *Individual τ footprints for the different τ decay modes reconstructed by the HPS-algorithm: 1-prong(Up-Left), 1-prong + π^0's (Up-Right) and 3-prongs (Bottom-Center). The reconstructed Particle-Flow τ constituents are superimposed to the calorimeters trigger towers in order to emphasize the matching between the single reconstructed decay product of the τ with a consistent energy deposition in a trigger tower*

The conclusion we get from this study is that the clustering algorithm already applied for the new e/γ trigger for stage-2 [124] could be largely reused as building brick. This is an important observation in view of the hardware implementation. It implies that the clusterings for e/γ and τ could share resources. Some modifications are however needed in order to define the geometrical dimension and shape for a cluster of a 1-prong decaying τ. In the case of 1-prong + π^0's and 3-prongs, it was observed that the clustering used for the electrons and photons is not extended enough to contain the entire energy deposited in the calorimeter. More precisely: in $\sim 25\%$ of 3-prong decays and in $\sim 10\%$ of 1-prong + π^0's decays, it happens that a trigger tower with non-negligible quantity of the τ energy is found outside the 3×3 trigger tower region used as protocluster and a second cluster could be built. The treatment of these cases is more complicated and will be described in the Section 5.5.3 dedicated to the cluster merging.

5.5 A new clustering algorithm for the τ

5.5.1 The e/γ clustering algorithm

As explained in Section 5.4 the clustering algorithm already developed to identify the e/γ clusters could be adopted with some additional modifications to identify τ clusters in the majority of the cases.

The e/γ clustering algorithm starts searching a trigger tower with a local maximum energy above a given threshold (2 GeV) within a 3×3 trigger tower window. Once a local maximum has been found, it defines a "seed" to build a "protocluster". A protocluster is a 3×3 trigger tower window centred in the seed and collecting energy from all the trigger towers presenting a deposits above an energy threshold (1 GeV) in 3×3 region around the seed. Once the protocluster is built, it is *trimmed* according to the asymmetries in the η slices around the seed in the energy depositions. This procedure is needed in order to avoid collecting energy in a larger area, that would increase the trigger rate. The cluster obtained must pass some shape vetoes and a requirement on the hadronic energy fraction assigned to the cluster (H/E ratio). Finally, the energy spread in the ϕ direction because of the electrons bending in the CMS magnetic field, is recovered with an additional trigger tower positioned either at $\phi + 2$ or in the $\phi - 2$ with respect to the cluster (η, ϕ) trigger tower position, as illustrated in Fig.5.7. More detailed information regarding the e/γ algorithm can be found in Ref. [124].

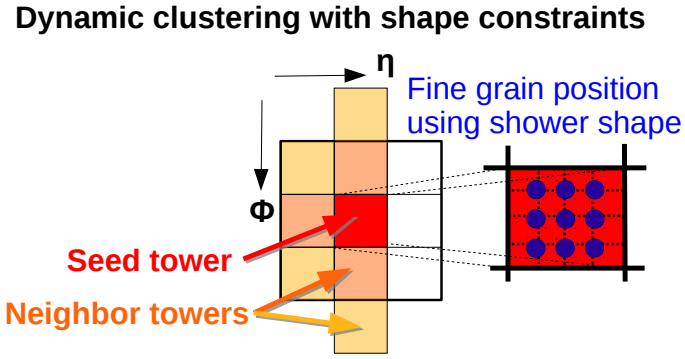


Figure 5.7: *Sketch of the e/γ dynamic clustering algorithm*

5.5.2 The τ clustering algorithm

The modifications required to use the e/γ algorithm for τ clusters are related to the presence of hadron(s) in the final state possibly accompanied by photons. An electron or a photon deposits almost the totality of their energy inside ECAL but it is not true for a τ. Indeed, the τ decay always involves at least one charged pion and possibly some π^0s. The charged pions interact either through electromagnetic or strong force with the detector and thus they

could deposit energy both in ECAL and in HCAL. The π^0, instead, decays into a couple of γ that could either arrive directly to the ECAL surface or convert into electrons through the tracker material. To estimate the fraction of the energy deposited in the two calorimeters, the $E_{HCAL}/(E_{ECAL}+E_{HCAL})$ ratio is plotted for the different decay mode and a clear dependency can be observed, see Fig. 5.8.

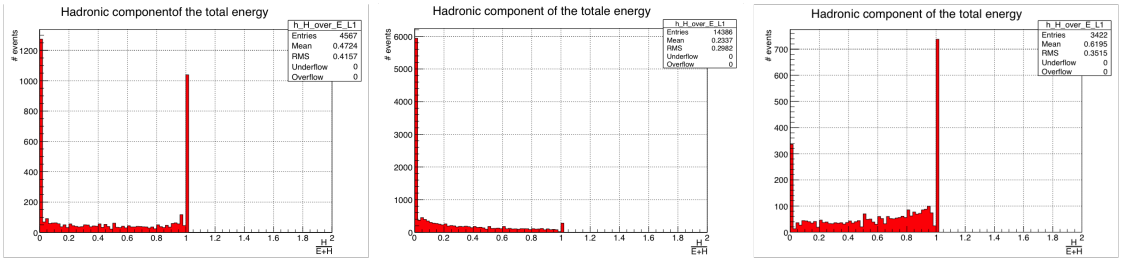


Figure 5.8: *Fraction of the τ energy deposited in HCAL respect the total τ energy for 1-prong (Left), 1-prong + π^0's (Center) and 3-prongs (Right) obtained from 2012 data.*

As can be seen, there is an average dependence of the hadronic energy fraction with the τ decay mode. In the majority of 1-prong + π^0's decaying τ's, the energy is mostly deposited in the ECAL as expected because of the presence of the γ coming from the π^0 disintegration. For the majority of the 3-prongs decaying τ's, given the presence of 3 charged hadrons, the energy is mostly collected by HCAL. Nevertheless, a non negligible fraction of the τ's, independently on their decay mode, deposit their energy in both calorimeters. Obviously, the energy recollection should be carried out in both calorimeters in order to correctly estimate the total τ visible energy. This is one of the modification component to the e/γ algorithm. An other important modification is related to the different shapes of the τ clusters compared to clusters from electromagnetic particles. The study of the single τ footprints revealed that some flexibility should be added to the clustering algorithm as to accommodate some energy deposit patterns that were not envisioned in the initial e/γ clustering. Indeed, in the e/γ algorithm, trigger towers only connected to the others by a corner are not allowed. Two examples of such *diagonal* topologies are presented in Fig.5.9.

From the study described above, the minimal set of modification to the e/γ clustering algorithm in order to trigger efficiently on the τ leptons are summarized in the following items:

- The energy of the cluster is estimated summing the energies deposited both in ECAL and in HCAL (in the e/γ case it was just ECAL);
- Removing of the H/E veto;
- Additional cluster shapes with diagonal contributions are allowed. They are shown in Fig.5.11;

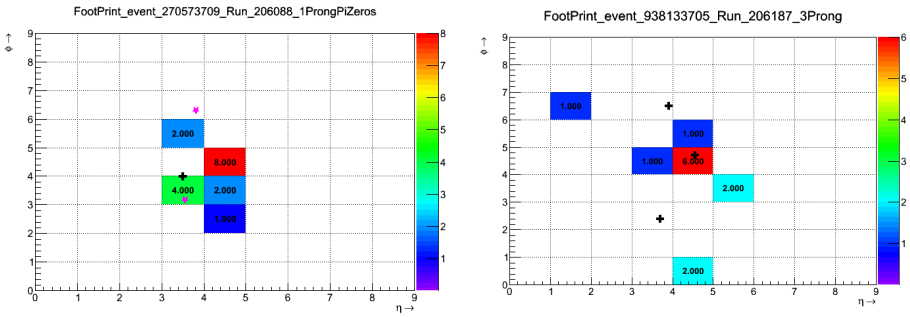


Figure 5.9: *Example of two tau footprints where diagonal trigger towers are present. The black cross indicate the position of the charged hadrons and the pink stars indicate the position of the photons at the ECAL surface. The energy inside each trigger tower is expressed in ADC counts*

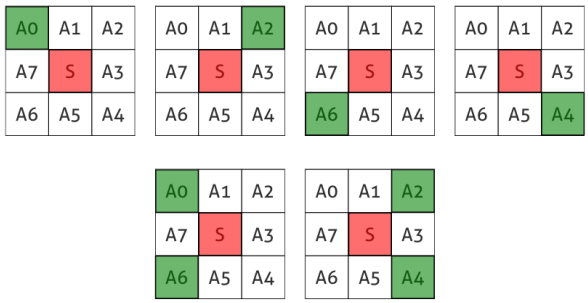


Figure 5.10: *Cluster shapes added to the already existing list of shapes of the e/γ clustering algorithm.*

5.5.3 Cluster merging algorithm

The strategy followed so far consisted in modifying the e/γ clustering for the τ leptons. This approach has however an intrinsic limitation. In 25% (resp. 10%) of the 3-prong (resp. 1-prong+π^0s) decays, a trigger tower with a substantial amount of energy is present outside the 3×3 cluster region and is found to originate from the τ. This is like a *secondary seed* that could originate a distinct cluster *a la* e/γ while it is in reality a *secondary cluster* created by the same τ. The spread of the energy deposit causing distinct reconstructed clusters is caused by the opening angle between the particles at the vertex, sometimes amplified by the magnetic

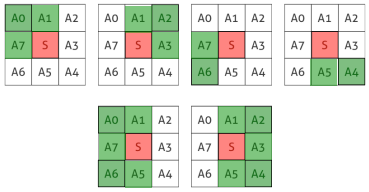


Figure 5.11: *Example of some cluster shapes already existing in the e/γ clustering algorithm and adopted also in the τ clustering.*

field. The distribution of the number of reconstructed clusters in a $\Delta R < 0.5$ cone around the offline reconstructed τ as function of the true τ transverse momentum is presented in Fig.5.12. One can conclude that for a $p_{\mathrm{T}}(\tau) > 20\,\mathrm{GeV}$ there is still a non negligible number of τ's with

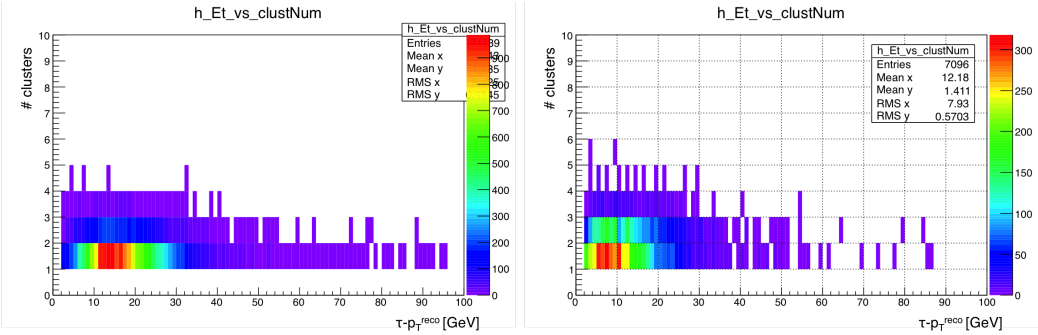


Figure 5.12: *Number of clusters per reconstructed τ as function of the true τ reconstructed transverse momentum for the 1-prong + π^0s (Left) and 3-prongs (Right) τ reconstructed decay modes obtained for data enriched in $Z \to \tau\tau$ events.*

more than 2 reconstructed clusters corresponding to one HPS-τ. In addition, as expected, the phenomenon is more pronounced in the 3-prongs case than in 1-prong + π^0's due to the fact that the charged hadrons with opposite sign bend in different directions under the action of the magnetic field increasing the probability to give rise to 2 or more distinct cluster *a la* e/γ. It is important to define an algorithm that is able to take care of two different clusters belonging to the same τ in order to recover the energy of the τ's that present an important spread of their final decay products at the ECAL entrance. One of the innovation of the τ clustering algorithm with respect to the e/γ one is its capability to recognize secondary clusters associated to the same τ lepton and merge them into a unique bigger final cluster. This *merging* procedure is run once all the clusters have been reconstructed in a region.

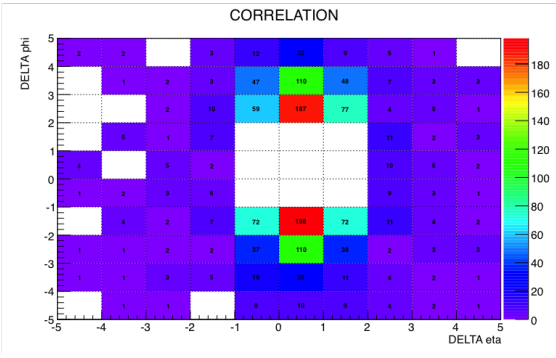


Figure 5.13: $\Delta\eta - \Delta\Phi$ *distribution between the seed of the secondary and the main one for a given reconstructed τ obtained for data enriched in $Z \to \tau\tau$ events. The 3×3 white region in the middle of the plot represent the dimension of the basic cluster: its size is the same as the e/γ one.*

A sample of τ leptons giving rise to two clusters has been constituted in order to define the merging criteria. For these events, the position of the seed of the less energetic with respect to the seed of the most energetic has been plotted. The result is shown in Fig. 5.13 where it appears clearly that the clusters whose seed is the same eta-band but shifted in phi with respect to the primary cluster are worth considering. More precisely, if the main cluster seed is $(i\eta_C, i\phi_C)$, the good candidates have their seed located at $i\eta_C - 1, i\phi_C \pm 2$, $i\eta_C, i\phi_C \pm 2$, $i\eta_C + 1, i\phi_C \pm 2$, $i\eta_C - 1, i\phi_C \pm 3$. The cluster merging procedure is sketched in Fig.5.14. Basically, if more than one cluster is found with a seed in the region allowing the merging, the secondary cluster is merged with the most energetic one giving a unique final cluster that will be identified as the L1-τ candidate.
In practice, the simple merging approach described is not straightforward hardware-wise. Indeed, due to the timing constraints, the clustering and the merging have to run almost in parallel. The phase-2 trigger system start to read the output from 2 ϕ slice at a time, starting from the center of the barrel and proceeding toward the endcaps. So the output of 144 trigger towers is read simultaneously and basic 3×3 clusters are dynamically built (more information in Ref. [124]). Once the clusters are built in a given ϕ slice at a fixed η, the algorithm search for each cluster the presence of a correspondent secondary cluster. An ordering operator between clusters (C_1 and C_2) has to be defined. It is defined as follows:

- $C_1 > C_2$ if $E_{\mathrm{T}}(C_1) > E_{\mathrm{T}}(C_2)$:
  - if $E_{\mathrm{T}}(C_1) = E_{\mathrm{T}}(C_2)$, $C_1 > C_2$ if $|\eta|(C_1) < |\eta|(C_2)$ (priority to more central clusters):
    - if $|\eta|(C_1) = |\eta|(C_2)$ $C_1 > C_2$ if $\phi(C_1) > \phi(C_2)$ (arbitrary condition):

The energy of the merged cluster is the sum of the energies of the primary and secondary clusters. As far as the position is concerned, temporarily the following formula has been used for the following studies:

$$\begin{cases} \eta_{L1-\tau} = \dfrac{\eta_{C_1} \cdot E_{C_1} + \eta_{C_2} \cdot E_{C_2}}{E_{C_1} + E_{C_2}} \\ \phi_{L1-\tau} = \dfrac{\phi_{C_1} \cdot E_{C_1} + \phi_{C_2} \cdot E_{C_2}}{E_{C_1} + E_{C_2}} \end{cases}$$

Because of the operations involved, in particular, the division, these formulae are not suitable in the hardware. It has been shown that the performance are marginally degraded when the position of the most energetic cluster is taken as the position of the merged cluster.

In Fig. 5.15, the energy response and the position resolution in a sample of 3-prong τ's are shown. The improvement brought by the merging can clearly be seen. As expected the energy response is smaller than 1, hence the need for an energy calibration.

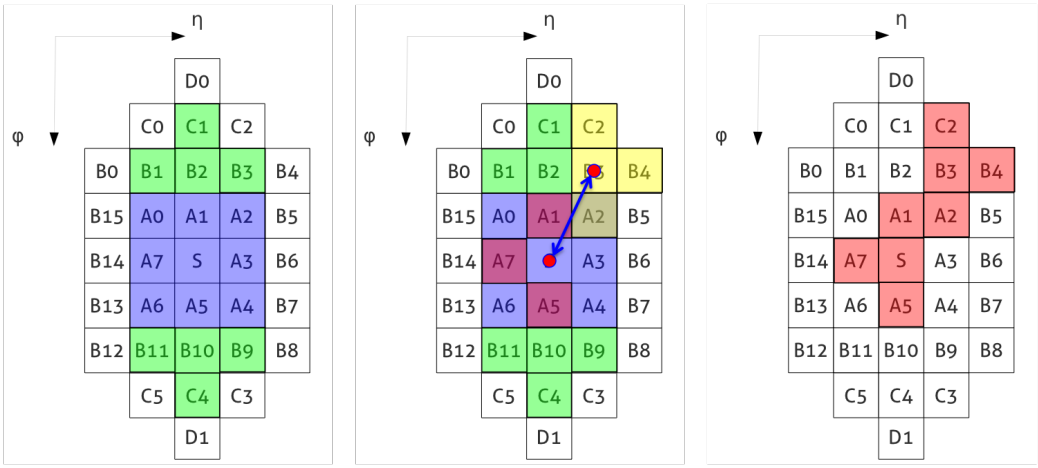


Figure 5.14: *Scheme of the merging algorithm, from the 2 separate clusters up to the final L1-τ candidate. The initial cluster is shown on the left. The blue area shows the protocluster surface, while the green area illustrates the zone where the search for a possible secondary seed is carried out. In the middle the two clusters found are shown: the main one seeded by the S TT, in red, and the secondary cluster, seeded by the B3 TT, in yellow. At this stage the trimming procedure has been already carried out The resulting merged cluster is shown on the right in red*

5.6 Calibration

As was seen in Fig.5.8 the τ decay products, mainly $\pi^{\pm}$ and γ coming from π^0, deposit energy both in ECAL and in HCAL calorimeters. Therefore, a calibration procedure based on a simple multiplicative factor would be suboptimal since the two calorimeters have different calibration. Moreover, the energy response depends also on the position of the cluster (η). The proposed technique to calibrate the τ cluster relies on a complex procedure already tested successfully to calibrate the charged hadrons in the Particle-Flow algorithm [96].

5.6.1 Energy-dependent calibration

In CMS, the ECAL is calibrated for photons, and the HCAL with 50 GeV pions not interacting in the ECAL. Because the HCAL response to hadrons is non-linear and because the ECAL response to hadrons is different from the response to photons, the ECAL and HCAL cluster energies need to be substantially recalibrated to get an estimate of the true hadronic energy deposits in ECAL and HCAL. The calorimetric transverse energy related to the τ cluster is assumed to have the following form:

$$E_{T_{calib}} = a\,(E_{\mathrm{T}}, \eta) \cdot E_{T_{ECAL}} + b\,(E_{\mathrm{T}}, \eta) \cdot E_{T_{HCAL}} + o_{eh}(\eta) \tag{5.3}$$

$$E_{T_{calib-had}} = c\,(E_{\mathrm{T}}, \eta) \cdot E_{T_{HCAL}} + o_{h}(\eta) \tag{5.4}$$

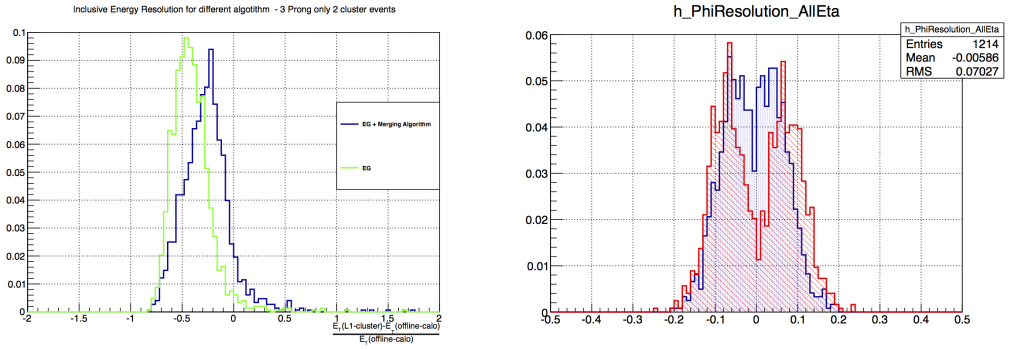


Figure 5.15: *Improvement of the energy (Left) and position (Right) resolution using the merging algorithm in the 3-prongs τ's giving rise at least to 2 clusters. The distributions were drawn both for the algorithm with merging and for the algorithm that use only a modified version of the e/γ clustering*

	barrel	endcaps
o_{eh}	1.5 GeV	1.5 GeV
o_h	0.5 GeV	0.0 GeV

Table 5.2: *Offset values that minimize the energy relative resolution different η region for different type of cluster*

where $E_{T_{ECAL}}$ and $E_{T_{HCAL}}$ are the transverse energies measured in ECAL and HCAL respectively, η the pseudo-rapidity of the cluster, and E_{T} an estimate of the true transverse energy, chosen to be either the total calorimetric transverse energy of the τ. For a given of value of the offset ($o_{eh}(\eta)$,$o_h(\eta)$) the coefficients a, b and c are obtained by minimizing, in each bin of E_{T}, the following χ^2:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(E^i_{T_{calib}} - E^i_{\mathrm{T}}\right)^2}{\sigma_i^2\left(E_{T_{calib}}\right)} \tag{5.5}$$

where E^i_{T} and σ_i are the true transverse energy and the expected calorimetric transverse energy resolution of the i^{th} single cluster, and where the sum extends over all events, separately in the barrel and in the endcaps region, and for clusters depositing energy either solely in the HCAL (Eq. 5.4), or in ECAL and HCAL (Eq. 5.3). The transverse energy resolution σ_i is determined iteratively as the Gaussian sigma of the $\left(E^i_{T_{calib}} - E^i_{\mathrm{T}}\right)$ distributions in each bin of true transverse energy. The offsets $o_{eh}(\eta)$ and $o_h(\eta)$ are chosen to minimize the transverse energy resolution for each choice of a,b (or c) coefficients. This coefficient found are reported in the Tab. 5.6.1 and each of them can be understood as a constant correction for the thresholds applied to the calorimetric cell energies in the clustering algorithm.

Since it is impossible to distinguish different τ decay modes at Level-1 stage, the calibration has been performed over a sample of 1-prong + π^0's events. Subsequently the calibration factors

obtained are used to calibrate a τ cluster independently on the τ decay mode. A Monte-Carlo sample of $H \to \tau\tau$ events produced through gluon fusion has been used in order to derive the calibration coefficients. The τ clusters that have been reconstructed by the new trigger algorithm have been selected and matched to the MC-truth τ using a ΔR criteria with a radius fixed at 0.5 and are required to have a $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.2$. In Fig. 5.16 the distribution of the transverse energy resolution as function of the offset values are shown, both for barrel and endcaps and for clusters that deposit energy only in HCAL or in ECAL+HCAL:

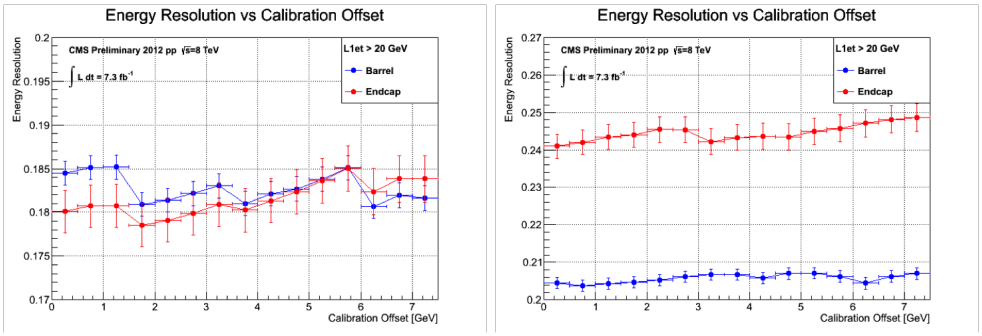


Figure 5.16: *Transverse energy relative resolution as function of different values for the offset in different η region, Barrel (Blue) and Endcap (Red), for cluster depositing in ECAL+HCAL (Left) and in HCAL only (Right)*

In Fig.5.17 the calibration coefficients corresponding to the choice of offset that minimize the cluster transverse energy resolution for the different region of the detector and for clusters that deposit energy only in HCAL or in ECAL+HCAL are shown.

The coefficients present values smaller than 1 in the low transverse energy region ($E_{\mathrm{T}} < 20\,\mathrm{GeV}$) because of the threshold effect on the cluster $E_{\mathrm{T}} > 20\,\mathrm{GeV}$. The correction factors are applied to cluster with a corrected E_{T} greater than 20 GeV, but the typical threshold used to trigger a τ lepton is around 30 GeV so the possible bias are minimized. As soon as the τ E_{T} is greater than 20 GeV the coefficients a and b present a peak in their values that can varies from 1.4 up to 2.0 depending on the cluster η that smoothly falls down and stabilize around 1.0 as $E_{\mathrm{T}}(\tau)$ increase. This behaviour of the coefficients could be explained considering that the energy spread is more important for low energy τ's and not all its energy can be collected. Moreover when a more energetic τ is considered, its decay products are almost collinear and, as consequence, the clustering algorithm is able to collect the almost totality of its energy and it is possible to explain in this way a calibration coefficient closer to 1.0. The c coefficient presents a shape in its dependence with respect to the $E_{\mathrm{T}}(\tau)$ that is similar to one showed by the a and b ones, but in this case all the values for the calibration coefficient are smaller than 1.

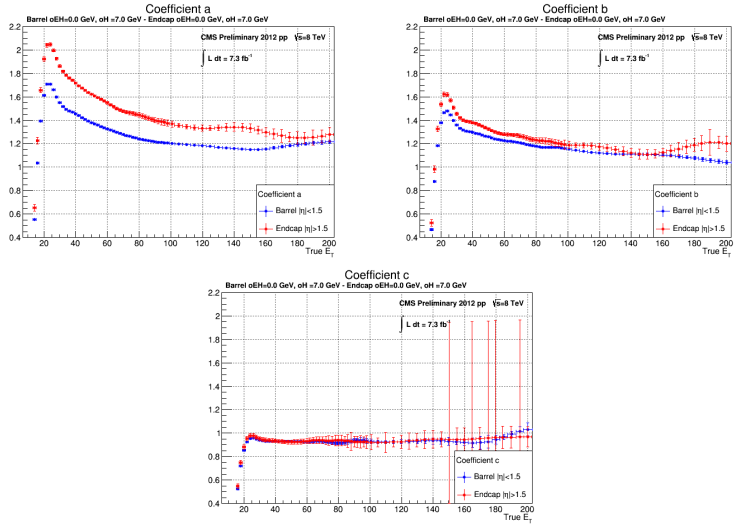


Figure 5.17: *The a (Left), b (Middle) and c (Right) coefficient obtained for the value of the offset that minimize the transverse energy resolution*

5.6.2 Calibration as a function of η

Once the energy dependent calibration has been carried out, an η-dependent calibration is then performed on top of it in order to bring the response flat in η. The response is plotted in different η bin. The E_{T} of the cluster is rescaled by a factor that depends on its η position. In Fig.5.18 the E_{T} response versus η of the cluster (expressed in trigger tower number) before and after the η-dependent correction are shown. It is possible to summarize the performances

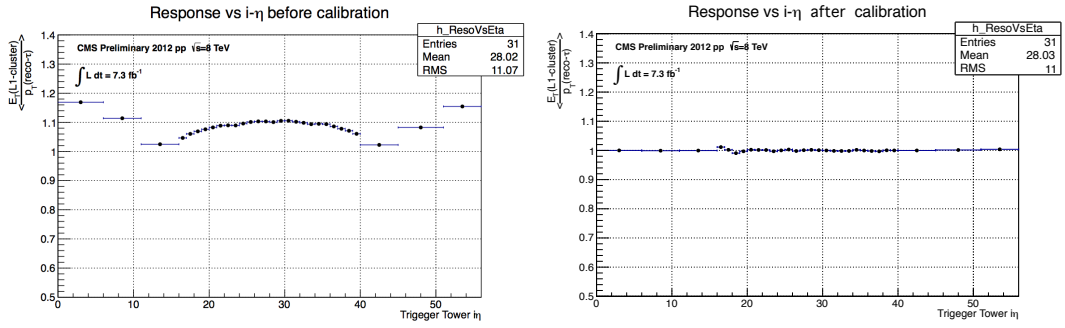


Figure 5.18: *Mean cluster E_{T} response as function of cluster η before the correction (Left) and after the correction (Right)*

of the cluster calibration procedure and of the achieved energy resolution deriving the τ trigger efficiency as a function of the offline τ calorimetric E_{T} (turn on curves). The turn on curves for clusters reconstructed in the barrel and in the endcaps separately for a L1 threshold of 30 GeV are illustrated in Fig 5.19. It is worth noting that an efficiency at plateau of $\sim 100\%$ is reached

and the turn on point (50% of efficiency) is very close to 30 GeV. This means that the calibration procedure restores properly the trigger response ~ 1.0 and is also almost η-independent. The performance of the τ trigger I have developed have been obtained on data and are studied in more details the next Section. The turn-on curves, comparing the trigger efficiency as function of the offline reconstructed τ transverse energy are obtained relaxing the cut on the $L1-\tau$ E_T down to 10 GeV.

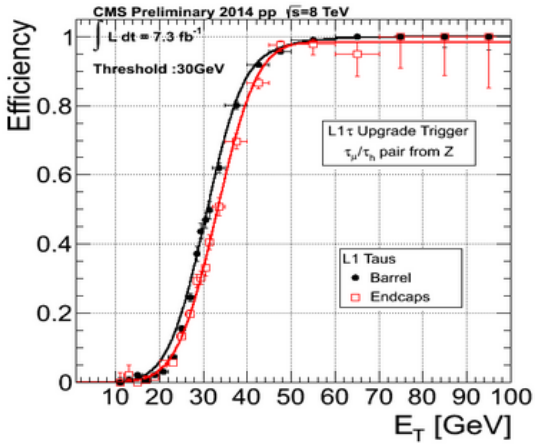


Figure 5.19: *Level-1 τ trigger efficiency as a function of the offline τ calorimetric E_{T} obtained in the barrel (black curve) and in the endcaps (red curve) regions.*

5.7 Performance and comparison with Run 1

In this Section a systematic comparison between the performance of the new stage-2 L1-τ trigger algorithm and those obtained with the Run 1 algorithm is performed. The tag and probe technique described in Section 5.3 has been applied in the Run D 2012 data. The distributions of the energy response and of the resolution are particularly important as well as the distributions of the angular resolution. Finally, in order to estimate the efficiency of the algorithms in triggering the τ lepton, the turn on curves were obtained for a fixed working point. For the new stage-2 L1-τ algorithm multiple turn on curves corresponding to different working point and different detector region are also presented.

In a first step, the performance in terms of energy response and resolution are assessed in Fig. 5.20 where the ratio of the L1 reconstructed p_{T} divided by the offline p_{T} are plotted. The Run 1 energy has been rescaled by a factor ~ 0.5. Indeed, in the Run 1 hardware, the τ candidates undergo the calibration of the jets, which is not adequate. As can be seen, the resolution for the stage-2 τ's is similar to the Run 1 while it is collecting energy in a surface ten times smaller. A smaller area used to collect energy directly translates directly into a smaller contamination from the pile-up events.

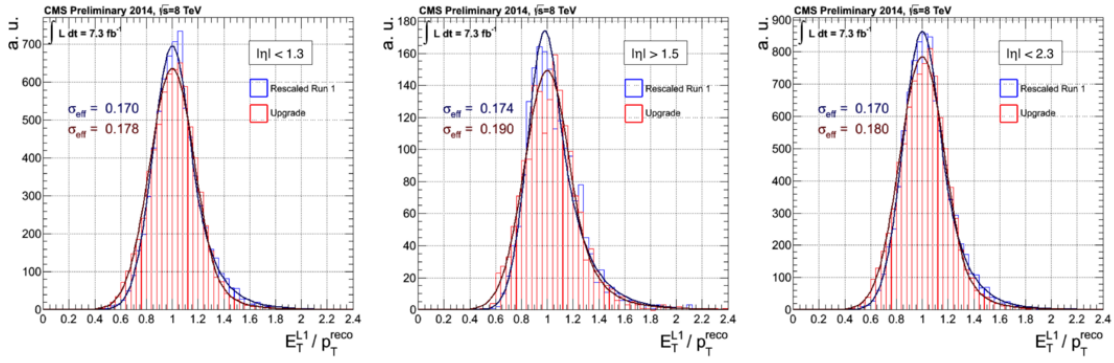


Figure 5.20: *Level-1 τ trigger energy response for τ's in the full calorimeters acceptance (Left), in the barrel (Center) and in the endcaps (Right). The response of the stage-2 upgrade trigger (red) is compared with that of the Run 1 algorithm (blue). The energy measurement of the latter has been rescaled to match the former. The stage-2 level-1 trigger upgrade τ reconstruction algorithm performs a dynamic clustering at the trigger tower level. The energy resolutions of the two algorithms are similar even though the upgrade algorithm is collecting the energy in a much smaller region*

The response and the resolutions are further compared in Fig. 5.21 and 5.22. The response and the relative distributions respectively defined as the mean and (relative) RMS of the distributions of Fig. 5.20 in given E_{T} bins. The resolution of the stage-2 algorithm is similar to the Run while using a much smaller surface of the calorimeter.

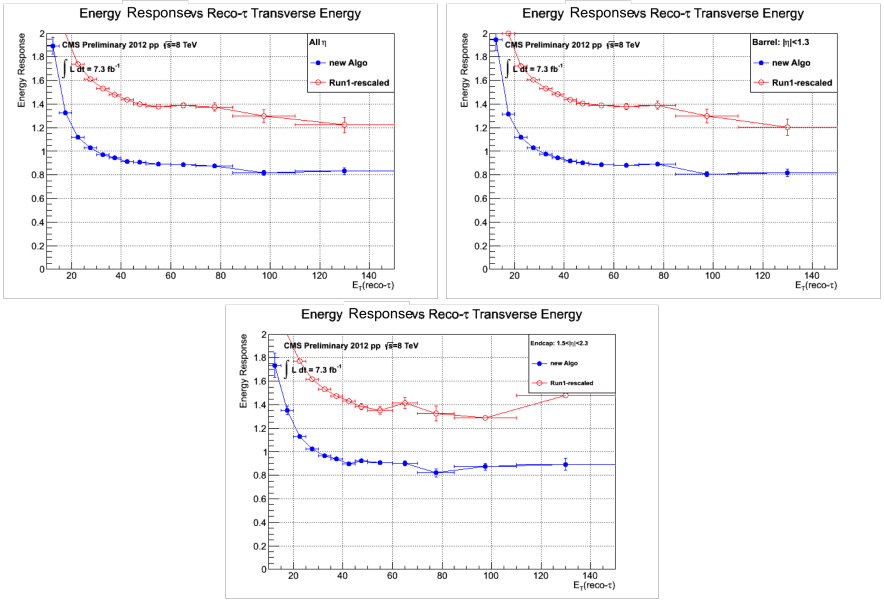


Figure 5.21: *Comparison between the stage-2 and Run 1 τ trigger algorithm: energy response as function of the reconstructed τ E_{T} considering clusters reconstructed in $|\eta| < 1.5$ (Up-Left), in $|\eta| > 1.5$ (Up-Right) and in $|\eta| < 2.2$ (Center).*

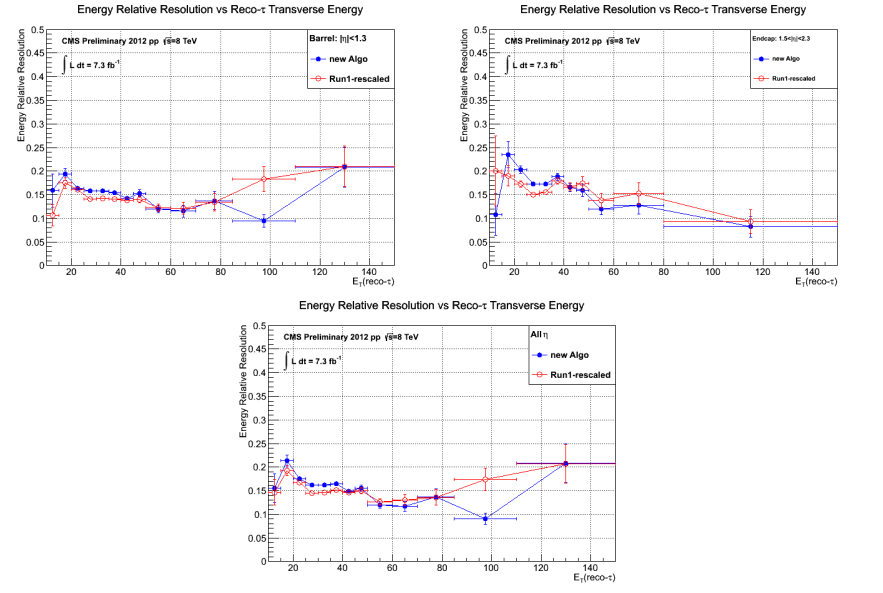


Figure 5.22: *Comparison between the stage-2 and Run 1 τ trigger algorithm: energy resolution as function of the reconstructed τ E_{T} considering clusters reconstructed in $|\eta| < 1.5$ (Up-Left), in $|\eta| > 1.5$ (Up-Right) and in $|\eta| < 2.2$ (Center).*

The position resolution in η and ϕ is then studied in Fig. 5.23 and Fig. 5.24 plotting the

difference between the η/ϕ position of the cluster and the η/ϕ of the reconstructed τ. The latter is evaluated at the vertex while the former is at the calorimeter surface, which is causing the ϕ width to be artificially enlarged as the effect of the magnetic field is not taken into account. The impact of the use of the single-tower granularity in the new hardware is very well visible and the position resolution is very much improved. The possibility to evaluate the invariant mass of the objects directly at L1 will benefit of an increased position resolution.

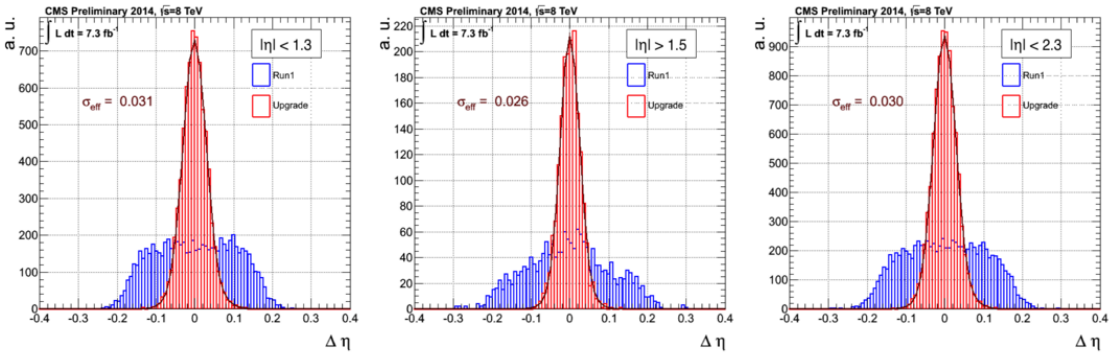


Figure 5.23: *Level-1 τ trigger pseudorapidity resolution for τ's in the full calorimeters acceptance (Left), in the barrel (Center) and in the endcaps (Right). The resolutions of the stage-2 upgrade trigger (red) is compared with those of the Run 1 algorithm (blue). The improvement in the position resolution with the upgrade system directly results from the access to the trigger tower granularity allowed by the new hardware.*

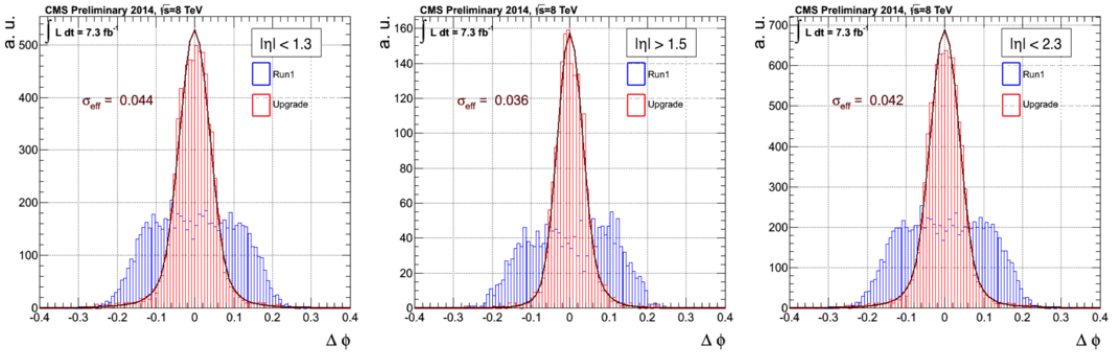


Figure 5.24: *Level-1 τ trigger phi resolution for τ's in the full calorimeters acceptance (Left), in the barrel (Center) and in the endcaps (Right). The resolutions of the stage-2 upgrade trigger (red) is compared with those of the Run 1 algorithm (blue). The improvement in the position resolution with the upgrade system directly results from the access to the trigger tower granularity allowed by the new hardware.*

The performance in energy response and resolutions directly translate into the sharpness of the trigger turn-on curves. In Fig. 5.25 the turn-on curves for Run 1 and in the upgrade algorithms are compared. The Run 1 response is still rescaled and it should be mentioned that the Run 1 algorithms features shape vetoes and isolation criteria. As can be seen, the efficiency of the Run 1 algorithm saturates at $\sim 70\%$ in the barrel and $\sim 40\%$ in the endcaps while the stage-2 algorithm reaches 100%, without isolation. The performance of the isolation will be assessed further on.

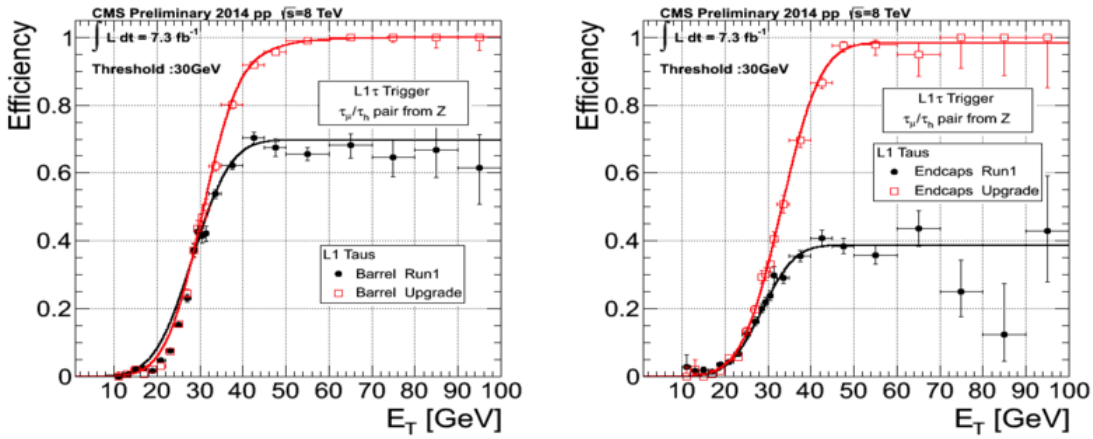


Figure 5.25: *Level-1 τ trigger efficiency as a function of the offline τ calorimetric energy for τ's in the barrel [$|\eta| < 1.5$] (Left) and in the endcaps [$|\eta| > 1.5$] (Right), for a level-1 transverse energy threshold of 30 GeV. The performance of the stage-2 upgrade trigger (red) is compared with those of the Run 1 algorithm (black). The energy measurement of the latter has been rescaled to match that of the former.*

Multiple turn-on curves for different thresholds are then presented in Fig. 5.26, Fig. 5.27 and Fig. 5.28.

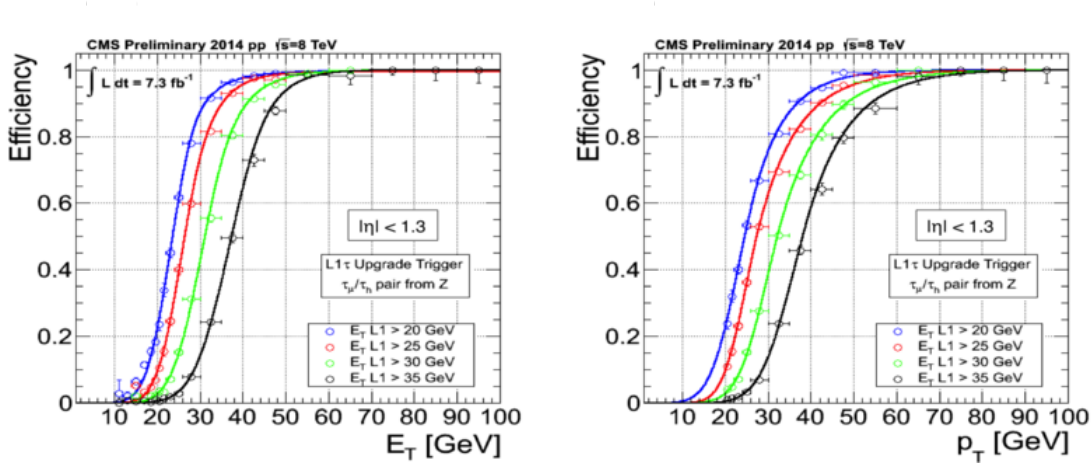


Figure 5.26: *Level-1 τ trigger efficiency as a function of the offline τ calorimetric energy, E_{T}, (Left) and as a function of the offline τ transverse momentum, p_{T}, (Right) for τ's in the barrel for different level-1 transverse energy thresholds:* 20 *GeV*, 25 *GeV*, 30 *GeV and* 35 *GeV*

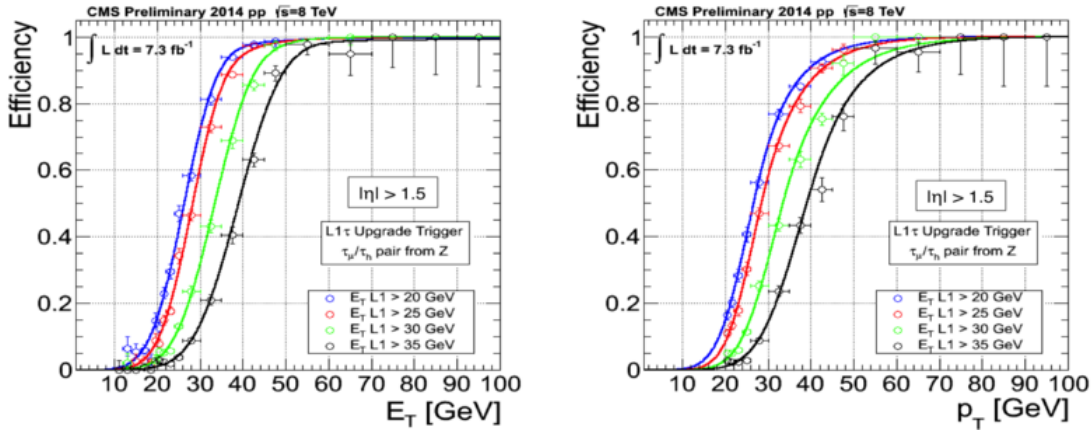


Figure 5.27: *Level-1 τ trigger efficiency as a function of the offline τ calorimetric energy, E_{T}, (Left) and as a function of the offline τ transverse momentum, p_{T}, (Right) for τ's in the endcaps for different level-1 transverse energy thresholds:* 20 *GeV*, 25 *GeV*, 30 *GeV and* 35 *GeV*

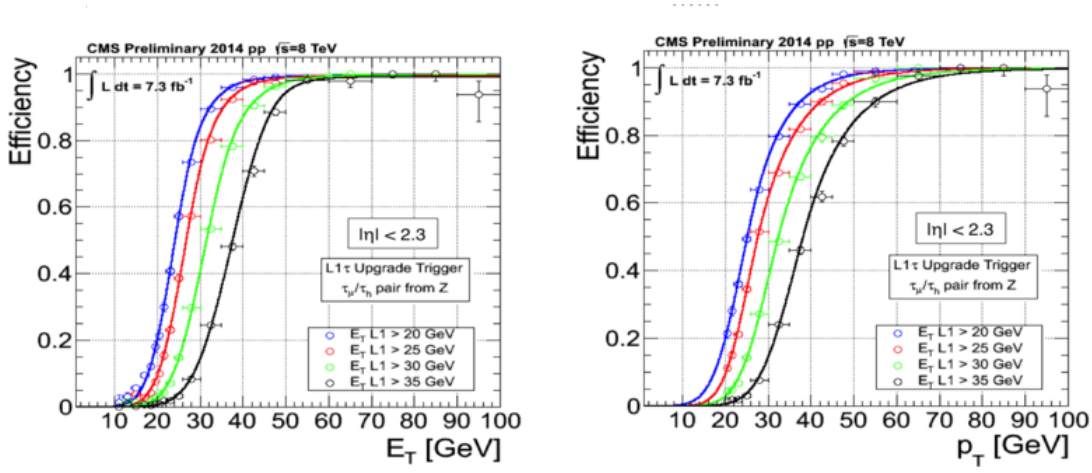


Figure 5.28: *Level-1 τ trigger efficiency as a function of the offline τ calorimetric energy, E_{T}, (Left) and as a function of the offline τ transverse momentum, p_{T}, (Right) for τ's in the full calorimeter acceptance for different level-1 transverse energy thresholds:* 20 *GeV,* 25 *GeV,* 30 *GeV and* 35 *GeV*

As already mentioned in the introduction of this Chapter, a very important figure of merit for a trigger algorithm is the control of the trigger rate. For both stage-2 and Run 1 algorithms, the trigger rate has been evaluated using 8 TeV minimum-bias samples with Level-1 μ with $p_{\mathrm{T}} > 12$ GeV from the Run D data taking period. The trigger rate has been estimated as the relative fraction of triggered events with at least 1 L1-τ candidate. In order to have a first estimation of the signal efficiency versus background rejection behaviour between the two algorithm, a ROC (Receiver Operating Characteristic) curve has been obtained and it is shown in Fig. 5.29. The signal efficiency on $Z \rightarrow \tau\tau$ data sample is estimated applying a threshold on the E_T of the L1-τ candidate of 10 GeV. The stage-2 algorithm is always more efficient that the Run 1 algorithm and has a higher background rejection. This comparison has been carried out once that a preliminary version of the isolation, detailed in the next Section, has been set up. The gain on the rate results from the fact that the stage-2 algorithm is collecting the energy in a region that is much narrower than the Run 1. It is therefore more difficult for a jet faking a τ to reach the threshold.

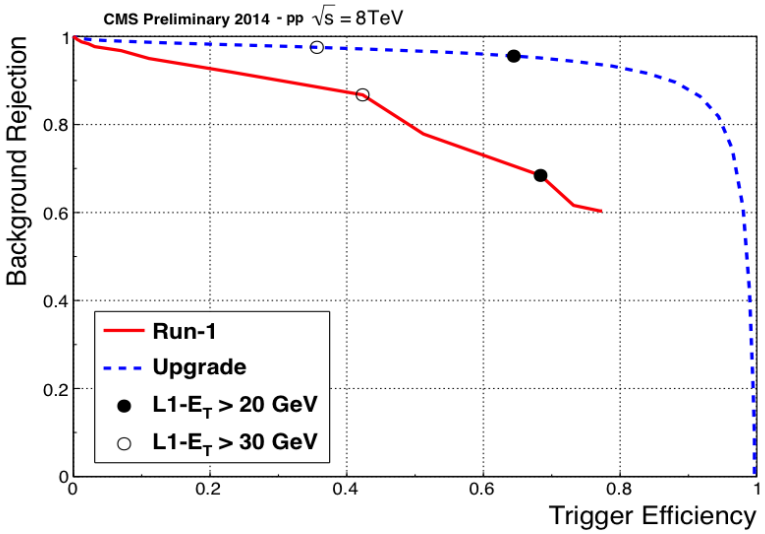


Figure 5.29: *Level-1 τ background rejection versus Level-1 tau signal efficiency evaluated for different values of the Level-1 transverse energy thresholds for τ's in the [-2.3, 2.3] pseudorapity region (barrel+endcaps). The stage-2 upgrade algorithm (dashed-blue) is compared to the Run 1 algorithm (red) and two particular working points are shown in the curves for both algorithm: Level-1 transverse energy above 20 GeV (black dots) and above 30 GeV (white circle). As visible in the turn-on plots above, the Run 1 algorithm efficiency saturates at 70% efficiency.*

In the ROC curve showed in Fig. 5.29 a preliminary isolation procedure (see next Section) is applied. It is possible to conclude that the stage-2 L1-τ algorithm, if adopted in a cross trigger path, e.g. with L1-μ, allows keeping the same *SingleMu* trigger threshold used during the Run 1, paying a very small price in term of efficiency reduction ($\sim 5\%$). In addition, the corresponding threshold on the τ_h leg turns out to be quite smaller than 20 GeV, allowing, in principle, to recover low p_{T} τ's. However, this is just one the ultimate goals. Indeed, one of the objective of the development of a new L1-τ trigger would be the possibility to trigger on the single τ object (*SingleTau* trigger) and on the di-τ_h pair, with an acceptable efficiency. The algorithm shows also that, with fine tuning of the isolation procedure, it is possible to reach impressive performance in the di-τ triggerring in term of signal efficiency versus background rejection, lowering the trigger threshold down to 29 GeV. In the next Section, a description of the dynamic isolation procedure I developed that lead to the plot illustrated in Fig. 5.29 and 5.30, is described. In addition, the main ideas that brought to a refinement of the isolation procedure, taking into account the pile-up effects, and obtained after that I left the trigger studies, is also given. This procedure, allows to even increase the performance of the new stage-2 L1-τ algorithm.

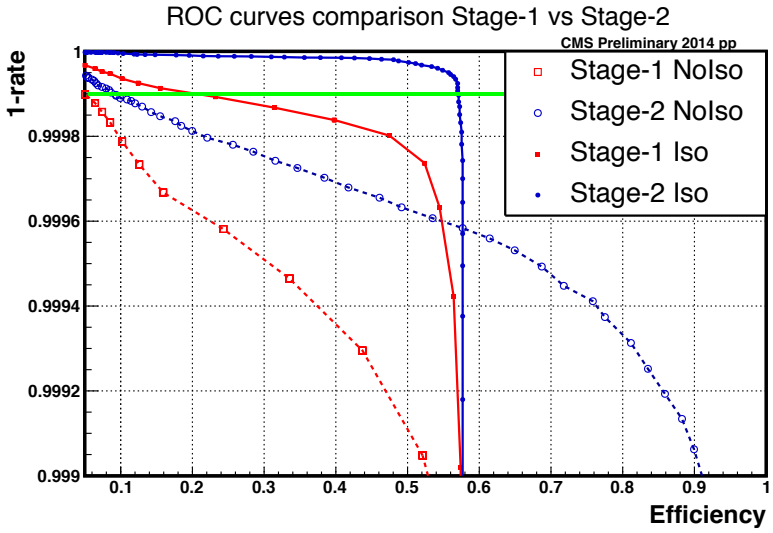


Figure 5.30: *Level-1 di-τ background rejection versus Level-1 di-τ signal efficiency evaluated for different values of the Level-1 transverse energy thresholds for τ's in the [-2.3, 2.3] pseudorapity region (barrel+endcaps). The stage-2 upgrade algorithm (dashed-blue) is compared to the stage-1 algorithm (red) and the working points that guarantee a trigger rate not-exceeding the 3 kHz bandwidth are shown on the curves for both the algorithms. In particular, to notice the low threshold of $\sim 29\,GeV$ needed to reach the goal with the new stage-2 L1-τ, allowing a di-τ trigger efficiency of 57%. The dashed lines represents the performance of both the algorithm obtained without any isolation requirement.*

5.8 Isolation studies

The isolation energy can be roughly defined as the energy deposited in the calorimeter cells around those that have been associated to a τ lepton. As processes that can fake a τ are generally characterized by a large value of isolation energy, while, on the contrary, real τ's are naturally isolated, this variable can be exploited to efficiently reduce the trigger acquisition rate, in order to have a better control of background and to improve its rejection. In this Section, the procedure used for the determination of the isolation energy is explained, and the development of the algorithms for background rejection are described. The final performance of the Level-1 Stage-2 τ algorithm are shown. In the end of the chapter, a summary of a more advanced isolation technique, developed after my work on the L1-τ algorithm, completes the development of the new τ trigger algorithm for the CMS phase-1 stage-2 upgrade.

Simulated events of $H \rightarrow \tau\tau$, both via gluon fusion and vector boson fusion, have been used. Unless specified, the term signal and background will refer to events coming from the samples indicated in table 5.3. For signal samples, the corresponding GEN-SIM-RAW sample is also used to retrieve the trigger information. The L1 τ are required to be geometrically matched ($\Delta R < 0.5$) to an offline τ that passes the usual τ analysis selections: $p_T > 20\,\text{GeV}$, $|\eta| < 2.1$, offline identification and finally, the offline τ must be geometrically matched ($\Delta R < 0.1$) to a generator level hadronic τ (built from all the visible decay products of the τ lepton). In particular, the last requirement ensures that fakes are removed from the offline τ collection, so that the isolation requirement usually applied in the analysis can be dropped; in this way, a larger part of the isolation energy spectra of the L1 candidate can be explored.

Determination of τ isolation energy

The τ isolation energy E_t^{iso} is defined as the difference of the energy deposited in a region of 5×9 in the $(i\eta, i\phi)$ plane around the τ candidate ($E_T^{5\text{x}9}$) and the energy deposited in a region of 2×5 or 3×5 dynamically assigned (see Fig. 5.31 and Fig. 5.32) around the trigger tower that contains the center of the merged cluster (L1-τ region). The center of the isolation window corresponds to the position of the seed of the cluster associated to the τ. The position and the width of the L1-τ region is assigned dynamically taking into account the relative position of the two merged clusters and the shapes information of the cluster before the merging. All energies are expressed in hardware values (ADC counts). The determination of isolation energy is sketched in Fig. 5.31 and Fig. 5.32. A comparison of its distribution for signal and background events is shown in figure 5.33.

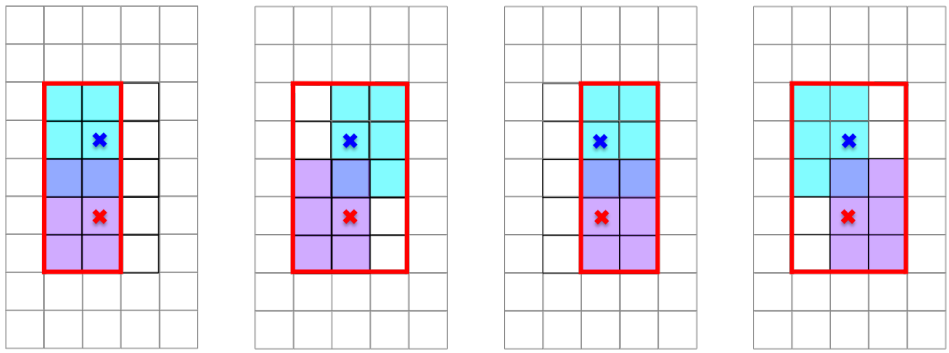


Figure 5.31: *Determination of the isolation energy for two merged clusters which seeds have the same i-η value*

Figure 5.32: *Determination of the isolation energy for two merged clusters which seeds are shifted by 1 TT in the η direction*

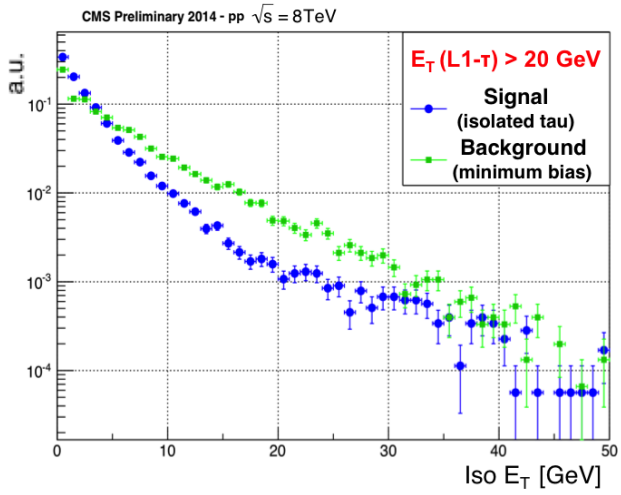


Figure 5.33: *Level-1 transverse energy distribution in the isolation region for the stage-2 upgrade τ trigger for signal (blue dots) and background (green dots) events. The isolation is computed by summing the energy in a 5×9 region in the (i-η; i-ϕ) plane, after subtracting the energy assigned to the Level-1 stage-2 tau candidate. The distributions for both signal and background are normalised to unity.*

The rejection of background is performed by applying a threshold on the isolation energy value. In order to choose such a threshold, in a first approach, the isolation efficiency on the background is shown as a function of the signal isolation efficiency for different values of the isolation threshold. The corresponding plot is shown in Fig. 5.34.

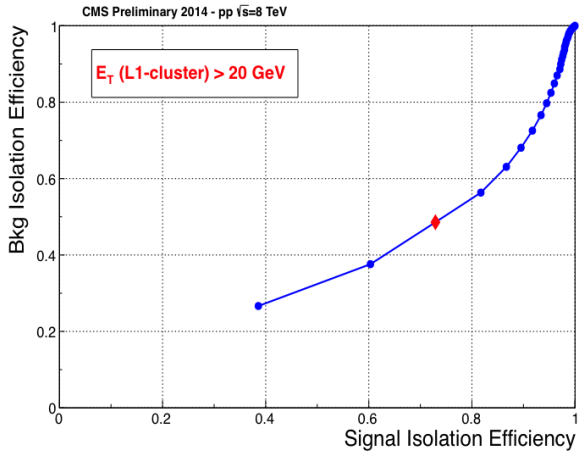


Figure 5.34: *The background isolation per-leg efficiency is shown as a function of the signal isolation efficiency for different values of cuts over the transverse energy in the isolation region for Level-1 stage-2 upgrade τ candidate with $E_T > 20$ GeV. The red marker shows the working point for a cut on the transverse energy in the isolation region less than 1 GeV: for this working point, the Run 1 algorithm and the upgrade stage-2 one present the same efficiency.*

The τ candidates have been considered isolated if the isolation energy was smaller than 1 GeV: such a value for the isolation threshold allows the two algorithms to have almost the same signal efficiency. This efficiency working point is also used to derive the relative trigger rate reduction as function of the L1-τ candidate threshold (see Fig. 5.35). In order to have a fair comparison between the two algorithms, taking into account the different calibrations and the different efficiencies, a correspondence between the L1 threshold and the offline p_{T} yielding a 50% L1 trigger efficiency is obtained. This correspondence is established producing a series of turn-on curves as function of the offline p_{T} made with various L1 thresholds, and considering the correspondent value of the offline p_{T} found at the *turn on* point.

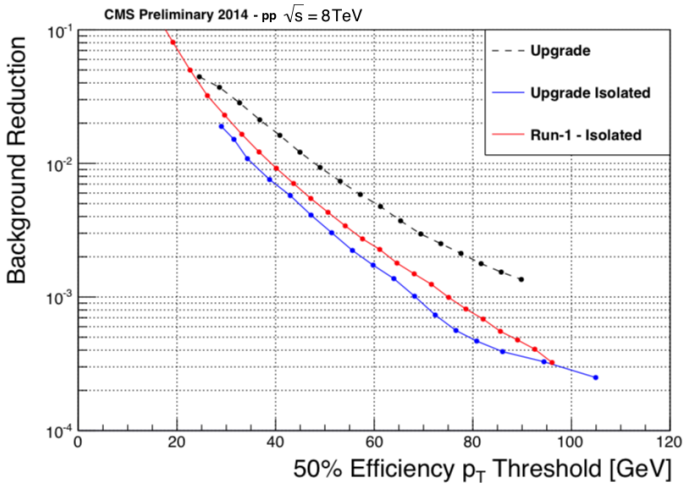


Figure 5.35: *Level-1 τ background reduction for L1 Et thresholds above 20 GeV. The background reduction obtained with the stage-2 upgrade non-isolated trigger (black-dashed) are compared with those obtained with the Run 1 algorithm (red) and with the stage-2 upgrade including the isolation (blue). The isolation threshold used ($\leq$ 1 GeV) is the one that allows an efficiency for the stage-2 upgrade algorithm comparable to the one of the Run 1 algorithm. From a series of turn-on curves as function of the offline p_{T} made with various L1 thresholds, a correspondence between the L1 threshold and the offline p_{T} yielding a 50% L1 trigger efficiency is obtained. The background reduction for various L1 E_{T} thresholds above 20 GeV is represented on the vertical axis. The 50% efficiency offline p_{T} for the chosen L1 E_{T} threshold is shown on the x-axis.*

It is possible to deduce from Fig. 5.35 that, using a very basic isolation technique, the new stage-2 L1-τ algorithm reaches a further 30% background reduction with respect to the one achieved by the Run 1 algorithm. The isolation implied, indeed, consists of a constant cut applied on the energy found in the isolation region and it doesn't take the effects of the pile-up into account. These results, together to the ones illustrated in the plots in Fig. 5.29 and 5.30 show that the new stage-2 L1-τ can be successfully used in a cross trigger or in a di-τ one. At the same times it is possible to conclude that an inclusive SingleTau trigger is difficult to sustain, it should have a threshold of O(100 GeV) in order to have a rate of a few 10's of kHz.

Further improvements are anyway possible on the isolation side, and they are described in the following.

5.8.1 Improvement of the isolation

Improvements of the isolation technique, carried out after I moved to the analysis, made the performance of the new stage-2 L1-τ trigger algorithm even more remarkable. Modification have been made in the determination of the isolation energy in order to made more easily its implementation at a hardware stage. Moreover, correlations between the isolation energy and the pile-up, as well as the pseudorapidity and the energy of the τ candidate, were exploited to achieve the best performance in terms of rate reduction, maintaining the signal efficiency as high as possible.

Determination of the τ isolation energy

The τ isolation energy E_t^{iso} is defined as the difference of the energy deposited in a region of 5×9 in the $(i\eta, i\phi)$ plane around the τ candidate ($E_t^{5\text{x}9}$) and the τ energy itself (E_t^{τ}):

$$E_t^{iso} = E_t^{5\text{x}9} - E_t^{\tau} \tag{5.6}$$

In the case of two merged clusters forming a τ candidate, the position of the main cluster is used; it has been explicitly checked that this choice does not affect the isolation energy distribution, and there is no need to move the isolation window center according to the relative position of the merged clusters. The determination of the isolation energy is sketched in figure 5.36. A comparison of its distribution for signal and background is shown in figure 5.37.

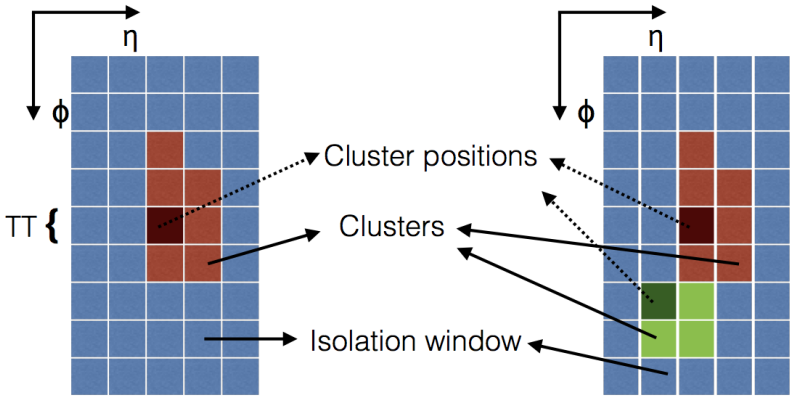


Figure 5.36: *Determination of the isolation energy for non merged (left) and merged (right) clusters.*

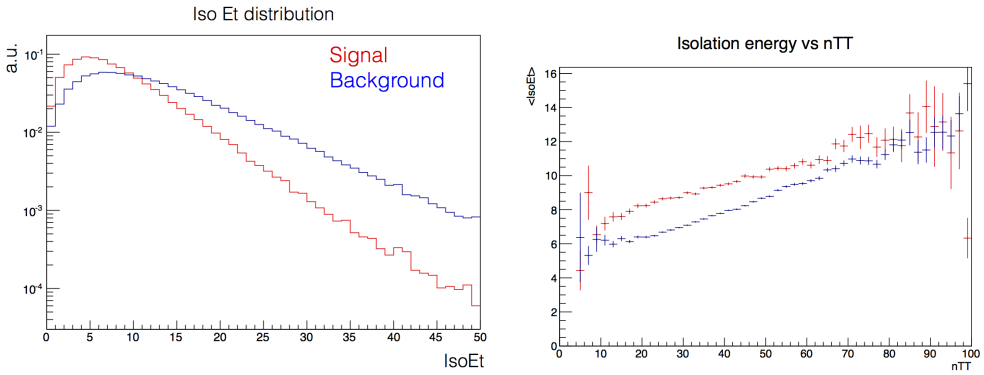


Figure 5.37: *Distribution of isolation energy for signal (red) and background (blue) events (Left). Distribution of isolation energy with the number of active trigger tower, i.e. the pile up activity.*

The isolation energy clearly depends on the amount of pile-up (PU) in the event, as it can be seen in Fig. 5.37 (Right). For the same reason, a dependence from the η position of the cluster is also present, as the pile-up energy density varies as function of η. Finally, E_T^{iso} also depends on the value of the τ energy itself. This is due to the clustering algorithm not being able to always collect the full energy deposited in the calorimeter; some towers where a small amount of energy has been deposited are discarded by the clustering algorithm, and by construction this leaked energy is summed up in the computation of $E_T^{5\text{x}9}$. The isolation energy is therefore a function of three variables:

$$E_T^{iso} = E_T^{iso}(\text{PU}, \eta, E_T^{\tau}). \tag{5.7}$$

As explained in Section 5.8.2, the PU variable is estimated from the number of central active trigger tower, nTT. Therefore, expressing the previous formula as a function of trigger quantities, we have:

$$E_T^{iso} = E_T^{iso}(nTT, i\eta, E_T^{\tau}) \tag{5.8}$$

The rejection of the background is performed by applying a threshold on the isolation energy value.

5.8.2 Pile-up estimator

The PU estimation is performed in a common way to the e/γ algorithm: PU is evaluated by counting the number of TT with $|i\eta| \leq 4$ (the eight central $|i\eta|$ *rings*) that have a transverse energy $E_T > 0$. The choice of the $i\eta$ range for this estimation is a trade-off between the precision in PU estimation and hardware resources constraints. A sketch of this procedure is shown in figure 5.38.

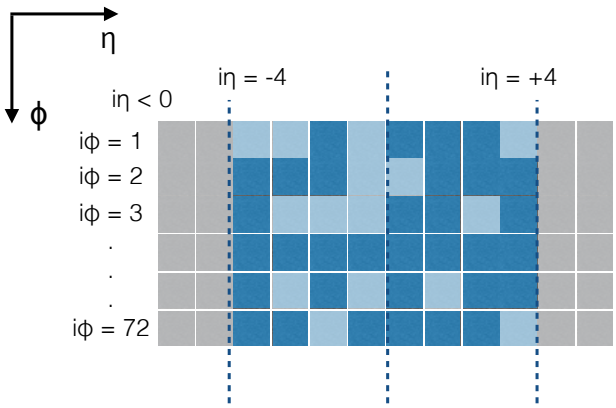


Figure 5.38: *Schematic representation of the determination of the PU activity. Depending by the $i\eta$ of the reconstructed cluster, a window of width $\pm|\eta| \leq 4$ TT is opened (azure cells) and the number of active TT (dark blue) inside this region is counted (nTT variable).*

5.8.3 Energy threshold determination

A energy threshold that corresponds to a flat efficiency with respect to nTT, $i\eta$ and E_T^τ is computed. A binned approach is used: for each 3-dimensional $(nTT, i\eta, E_T^\tau)$ bin the E_T^{iso} threshold value is computed as the one corresponding to a fixed signal efficiency in the bin. Different choices of the signal efficiency correspond to the determination of different isolation algorithm working point (WP).

5.8.4 New L1-τ trigger performance with isolation

The following results are shown as an example for three efficiency WP (70%, 80%, 90%) and compared to the stage-1 algorithm results, i.e. with a modified Run 1 hardware; this setup will be used until the end of 2015. One important difference between the stage-1 and the Run 1 algorithm is the removal of the shape vetoes that were limiting the efficiency. Performance in the non-isolated case is also shown. The rate reduction as a function of the L1 E_T^τ threshold is shown in figure 5.39. The target reduction is 10^{-4} corresponding to about 2.8 kHz at L1 output. The corresponding energy threshold is about 73 GeV for the non-isolation case and respectively 42, 35 and 30 GeV for the 90%, 80% and 70% WPs. The threshold for the same rate reduction is 45 GeV for the stage-1 algorithm.

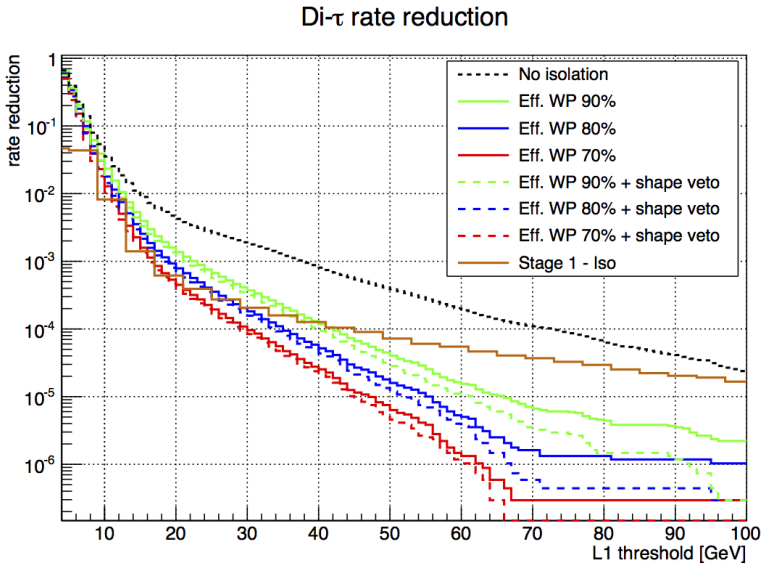


Figure 5.39: *Relative rate reduction as a function of L1 E_T^τ. A 10^{-4} relative reduction corresponds to about 2.8 kHz*

The turn-on curves corresponding to these thresholds are shown in figure 5.40 separately for barrel and endcaps. The turn-on curves for the same E_T^τ of 30 GeV are shown in figure 5.41. The performance of the stage-2 is very satisfying. A sharp turn-on is obtained, and the behaviour with respect to the isolation, which is a major handle to control the rate, is very clean. Indeed, as can be seen in Fig. 5.40, tightening the isolation cut impacts the plateau efficiency but does not affect the turn-on part. In Fig. 5.41 the Run 1 turn-on curves superimposed to several stage-2 turn-on curves are shown. The envelope of the stage-2 turn-on curves clearly covers those of stage-1, which points into the direction of implementing dynamical isolation working points depending on the energy, i.e., relaxing the isolation criteria at high p_{T}. This is rather easy to achieve through the use of Look Up Tables in the hardware.

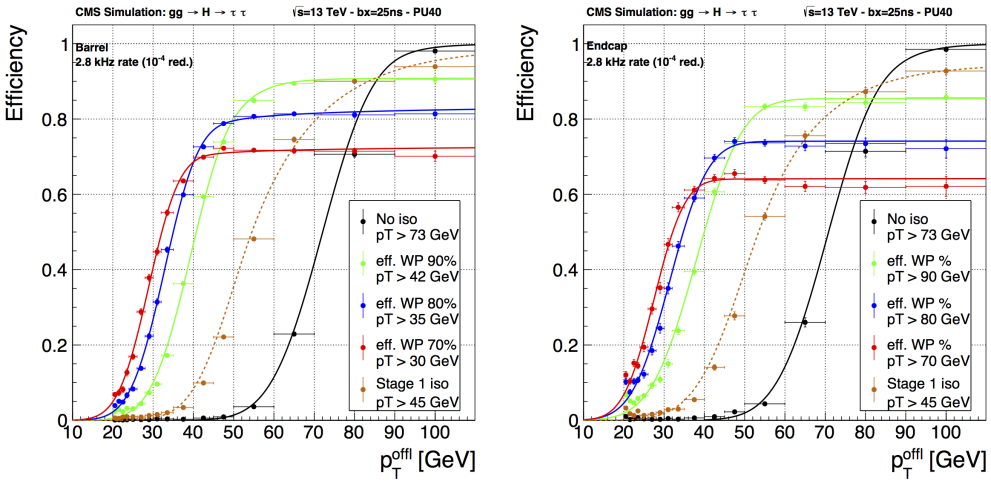


Figure 5.40: *Turn-on curves for the E_T^τ threshold corresponding to a 10^{-4} relative rate reduction*

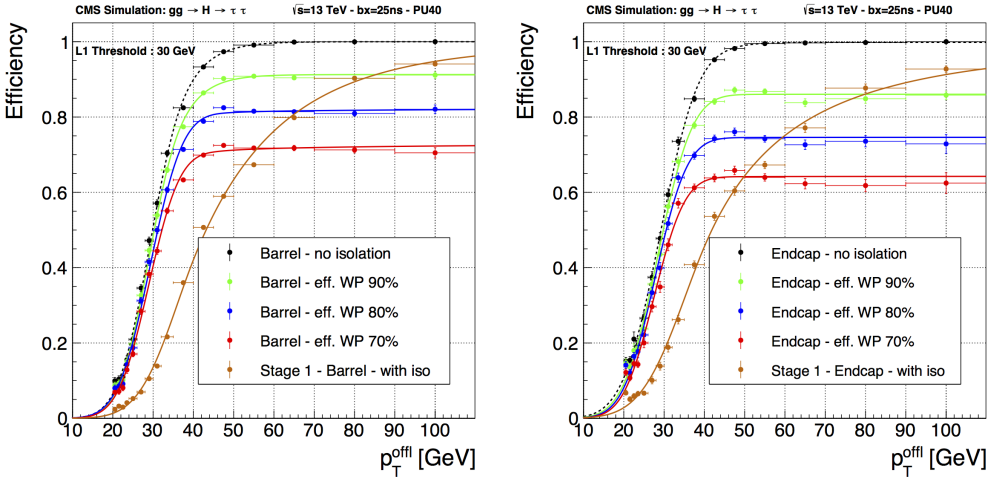


Figure 5.41: *Turn-on curves for the same E_T^τ threshold of 30 GeV*

5.9 Plans for Commissioning

The LHC Run 2 started during the Spring 2015. During the first period of the data taking, an upgraded version of the trigger architecture used during the Run 1 is responsible for the data acquisition. In parallel with the data taking, the new stage-2 upgraded trigger architecture is already running in shadow mode, in order to perform the commissioning of both the hardware systems and the new algorithms developed for the phase-1 CMS trigger upgrade: within them,

the new L1-τ trigger algorithm described in this Chapter. CMS will switch to the stage-2 trigger system and use it in production mode during LHCs annual technical stop at the beginning of 2016.

5.10 Conclusions

To guarantee a successful physics program in Run 2, the CMS Trigger and Data acquisition system has been consolidated. A key role to reach this goal is played by the L1 calorimeter Trigger hardware and architecture. The CMS L1 trigger system could thus benefit from the recent μTCA technology, allowing sophisticated algorithms to be deployed in order to identify different physical objects with the highest possible efficiency. The new L1 system is designed to provide handles to keep the trigger rate under control in the harsh LHC Run 2 environment. In this Chapter, a new algorithm completely dedicated to the identification of the τ lepton has been presented. This new algorithm relies on a finer granularity obtainable with the new μTCA technology that allows the individual calorimeter trigger tower readouts to be exploited. The core of the algorithm presented is based on a dynamical clustering algorithm shared together with the new L1 e/γ trigger. Substantial modifications with respect to the e/γ algorithm are introduced in order to take the different nature of the τ lepton and its different interaction with the CMS subdetectors into account. The cluster calibration procedure is also innovative, emulating the one used in the CMS offline reconstruction algorithms, e.g. the particle-flow. The comparison between the new τ trigger and the Run 1 algorithm has been studied. From this comparison emerges the superiority of the new τ trigger in terms of angular resolution allowing similar energy resolution despite exploiting less trigger towers. Moreover, the performance in terms of turn-on curves (efficiency) are sensitively increased with respect to the Run 1 algorithm, allowing a 100% plateau efficiency (without isolation). Finally, a deep study on the isolation at L1 has been presented, along with the handles put in place to mitigate the pile-up. The results are impressive in terms of rate reduction and efficiency. Moreover, the isolation algorithm is flexible enough to guarantee the possibility to choose between different working points. The algorithm as it is presented in this Chapter has already been implemented inside the official CMS software package. The impact of the new trigger on the main physics analysis is still being assessed as the 2016 L1 trigger menus are not designed yet, but as an example, the thresholds in di-τ triggers would become $\sim 29 - 30\,\text{GeV}$ instead of 46 GeV in the Run 1.

Chapter 6

Matrix Element Method approach in the $H \to \tau\tau \to \mu\tau_h$ channel

The LHC Run 2 has begun in spring 2015 delivering pp collisions at 13 TeV centre-of-mass energy. The instantaneous luminosity will reach $2 \times 10^{34} cm^{-2} s^{-1}$ and, at the end of Run 2, the total integrated luminosity is planned to reach $\sim 300 fb^{-1}$. A new era, characterized by precision measurements of the Higgs sector, has just began. Among the priority of the LHC physics program for Run 2: measurements of the coupling constants between the Higgs boson and the SM particles, refined measures of the Higgs boson mass and width as well as the precise characterization of the tensor structure of the trilinear coupling [68][69]. Deviations of the coupling constants values with respect to those predicted by the SM could represent hints for new physics. In addition, exclusive researches will be performed in order to prove the main Higgs boson production modes, measuring in a precise way the various Higgs boson production cross sections. In this context, the $H \to \tau\tau$ represents an excellent testbed. In particular, the VBF production mode will play, in the future, a leading role thanks to the distinctive topology that characterize such events. Hence the idea to develop multivariate strategies, like MVAs or matrix element method, to improve the efficiency and purity of the selections focussing on VBF topologies. The ME method relies on the computation of an event-by-event weight expressing the probability that a set of observables are results from a particular process. Such a method allows the full information of the event topology to be exploited thus improving the sensitivity of the $H \to \tau\tau$ analysis. Besides, the matrix element method has a direct connection with the theory and alternate models can be easily tested. It is also likely that the matrix method in $H \to \tau\tau$ could be exploited to test different theoretical hypothesis on the nature of the Higgs boson. Indeed, the various parameters of the model, like the spin, C and P of the Higgs boson, could be simultaneously tested. From these reasons, the idea to develop a matrix element approach in the $H \to \tau\tau$ search arose to light and has been preferred over other methods like boosted decision trees (BDTs). During my thesis, as described in this chapter, the pioneering work to

apply a matrix element method to the $H \to \tau\tau$ analysis involves only the most sensitive decay channel: $H \to \tau\tau \to \mu\tau_h$ where the Higgs boson is produced through the VBF mechanism. The evaluation of the matrix elements is performed at the leading order (LO) in the perturbation theory. Already at this stage, the implementation of such calculations requires hand calculations as well as powerful computing infrastructures.

6.1 The $H \to \tau\tau$ analysis strategy

From the experimental point of view, the $H \to \tau\tau$ analysis is very challenging because of the complexity of the final state topology and because of the presence of many SM reducible and irreducible SM backgrounds giving the same final state. Besides, the different τ decay modes and the presence of neutrinos in the final state represent a challenge to correctly reconstruct the energy and momentum of the τ leptons. The CMS Collaboration developed different strategies, based on the PF algorithms, to reconstruct the τ leptons efficiently and reduce as much as possible the contamination from other physics objects, such as hadronic jets, electrons and muons (see Section 3.4).

In addition to the large amount of the SM backgrounds, a further challenge in the $H \to \tau\tau$ analysis originates from the resonant nature of its main irreducible background: the Drell-Yan $Z/\gamma* \to \tau\tau$ process. Given the impossibility to measure directly the di-τ invariant mass because of the presence of the neutrinos, the visible invariant mass has been firstly used to separate the signal from the background. However, the presence of neutrinos in the decay chain as well as the resolution on the tau momentum in the 1-prong+π^0s channel, result in a broad distribution of the visible di-τ invariant mass. The background and signal peaks are mixed up because of the energy resolution: hence the challenge of the analysis.

In a second step, a special algorithm capable to estimate the di-τ invariant mass using a likelihood minimization approach (described in Section 3.4) has been introduced in the analysis (SVfit mass). A search for the $H \to \tau\tau$ in many decay channels (ee, $\mu\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$) has been performed by CMS exploiting a complex categorization of the events based on the number of the jets in the event, the p_T of the τ_h and the boost in the transverse plane of the Higgs boson candidate. The SVfit mass resolution varies from a decay mode to another and from category to category in a range between 10% and 20%.

The signal extraction is obtained through a global maximum-likelihood fit to the SVfit distribution (more details can be found in Chapter 7) in all channels (except ee and $\mu\mu$ where a multivariate analysis strategy is employed). The systematic uncertainties are treated as nuisance parameters that can vary in the fit according to their probability density functions. Using this analysis approach, the CMS Collaboration found a 3.4 σ evidence for the $H \to \tau\tau$ decay

[71]. The next goal during the LHC Run 2 will be to move towards a standalone discovery for $H \to \tau\tau$, already obtained through the combined results of ATLAS and CMS [66], revealing the nature of the Higgs boson and the fermion mass generation.

What is reported in the following is the first implementation of the Matrix Element method in the $H \to \tau\tau$ channel. The most sensitive category and decay channel, the VBF $H \to \tau\tau \to \mu\tau_h$ is considered. The Matrix Element Method, as it is developed in the present Chapter, is straightforwardly extensible to the VBF $H \to \tau\tau \to e\tau_h$ channel and, with some light changes in the integration procedure, also to the VBF $H \to \tau\tau \to \tau_h\tau_h$. The signal extraction is finally performed in the same way as for the SVfit mass: a binned likelihood is built after that a global maximum-likelihood fit has been performed on the matrix element weight distribution.

6.2 Introduction to the Matrix Element method

A new approach using the Matrix Element (ME) method is studied in the context of the search for the Standard Model Higgs boson produced through Vector Boson Fusion (VBF) and decaying into a τ lepton pair in the CMS experiment. Historically, the most spectacular implementation of this method was performed by the D0 Collaboration [125] [126], in the context of the measurement of the top quark mass, even if the oldest application in the high energy physics I could find is in the context of the search for the process $e^+e^- \to Z \to 4\ell$ in the CELLO experiment [127]. In CMS, the ME method was successfully adopted in the search for the Higgs boson produced in association with two top quarks and decaying into $b\bar{b}$ ($t\bar{t}H(b\bar{b})$ [128]. The ME method in the $H \to \tau\tau$ channel implies a complicated multi-dimensional integration of the LO Feynman amplitude both for the signal process $H \to \tau^+\tau^-$ and for the main backgrounds, the irreducible $Z \to \tau^+\tau^-$, in order to obtain an optimal separation between the signal and background hypotheses. The integral is evaluated numerically on an event-by-event basis, using the VEGAS [116] adaptive numerical integration technology. A brief description of the VEGAS integration program can be found in Section 6.3. The hard scattering amplitudes are also evaluated numerically, using Madgraph. The transfer functions model the decays of the stable and unstable particles produced by the main processes (i.e. $H \to \tau^+\tau^-$ or $Z \to \tau^+\tau^-$). They also model the detector resolution, in order to relate the set of measured quantities to the set of physical observables involved in the ME integral. The *transfer functions* taking care of the different τ decays are evaluated analytically, and it is one of the challenges of the present analysis. The ME approach offers several important advantages with respect to all other data analysis schemes commonly used in high energy physics [129]:

- The ME method is universal can be applied to a wide variety of processes for which theoretical models exist.
- The theoretical assumption on the model under study (matrix element, the mass and the

spin of the resonance, the parton distribution function, etc.) are tested in the analysis in the most efficient way: a direct link is established between theory and event reconstruction

- There are no restriction to the assumption on the detector resolution functions (incorporated in the transfer functions): it is possible to use very detailed detector models.
- According to the Neyman-Pearson lemma [130], the likelihood ratio, built with the results of the two integrals under the signal and background hypotheses, is the most powerful discriminant. Therefore, in theory, the ME method achieves the best signal to background separation
- Contrary to other multivariate analysis such MVA, the ME method does not need any training and it can be applied to sample with a limited statistic.

The ME method present however some limitation:

- The ME computation, as pointed out in Section 6.3, is heavily CPU time consuming
- In addition in a context of a model-independent analysis the ME is not longer exploitable because no model is available a priori (if effective Lagrangian approach are not considered)

In this chapter a brief description of the global form of the integrand is provided, the various elements are scrutinized starting with the scattering amplitude (Section 6.5) and the derivation and validation of the transfer functions (Section 6.6). Then, the need of a change of variable is explained in Section 6.7 and the corresponding Jacobian term derived. The final part of this chapter is dedicated to the validation of the algorithm used to finally reconstruct the 4-vector of the leptonic and hadronic τs (Section 6.8).

6.3 VEGAS: An Adaptative Multi-dimensional Integration Program

Computing the ME weight requires to evaluate a complex multidimensional integral. For practical purpose, this task requires a huge amount of computing time if performed with traditional Monte Carlo integration method. The computing time issue represents a big challenge in any analysis designed to exploits the ME method. In particular, this drawback becomes more important in complex analyses like the search for the $H \to \tau\tau \to \mu\tau_h$ because of samples for which the ME weight must be evaluated (in the case of the $H \to \tau\tau \to \mu\tau_h$ analysis this means $\sim O(1Mevents)$ of events processed). For this reason, an adaptive multi-dimensional integration method, VEGAS [116][117], has been chosen in order to speed up the integral convergence and allows the computation of the weight for many processes, as required by the physics analysis (see Chapter 7). Further improvements in the calculation speed result from the work on developments carried out on the analysis software itself that processes multiple events in parallel using

the Message Passing Interface (MPI) technology [131]. Performing the analysis described in this Chapter and in Chapter 7 was made possible thanks to the impressive performance reached using a MPI technology in term of computing time. The maximum performance in that case are achieved when just a single process is assigned to each CPU (or core in a multi-core machine). In the case of the following analysis a single process is represented by an LHC event (real or simulated).

6.3.1 The VEGAS algorithm in a nutshell

Given an integral of the form:

$$I = \int_{\Omega} f(\mathbf{x}) d^n\mathbf{x} \tag{6.1}$$

where Ω defines the integration volume, VEGAS performs an estimation of I by computing $f(\mathbf{x})$ at N random points (called also VEGAS shots) $\mathbf{x}_i$ and making a first estimation:

$$I \simeq S = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{\rho(\mathbf{x}_i)}, \tag{6.2}$$

where $\rho(\mathbf{x}_i)$ represent the density with which the random points are chosen in Ω. In addition, also the estimation of the uncertainty associated to S is evaluated:

$$\sigma^2 = \frac{1}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{f(\mathbf{x}_i)}{\rho(\mathbf{x}_i)}\right)^2 - S^2\right]. \tag{6.3}$$

In the same way VEGAS performs m estimates of S: $\{S_j\}_{j=0}^{m}$ and its uncertainty $\{\sigma_j\}_{j=0}^{m}$ providing a cumulative estimate $\bar{S}$:

$$I \simeq \bar{S} = \bar{\sigma}^2 \sum_{j=1}^{m}\frac{S_j}{\sigma_j^2}, \tag{6.4}$$

where $\frac{1}{\bar{\sigma}^2} = \sum_{j=1}^{m}\frac{1}{\sigma_j^2}$. The χ^2/dof is obtained as:

$$\chi^2/dof = \frac{1}{m-1}\sum_{j=1}^{m}\frac{\left(S_j - \bar{S}\right)^2}{\sigma_j^2} \tag{6.5}$$

and represents a test that quantifies how the different estimation S_j are consistent one with the other. If the algorithm works properly the expected χ^2/dof should be close to 1 [117]. In the simplest form of Monte Carlo integration, the random points are uniformly distributed (with $\rho(\mathbf{x}) = const$). In VEGAS, the density $\rho(\mathbf{x})$ is chosen such to minimize the σ_j^2 at each iteration j. In the first iteration the random number are uniformly distributed and the information gained about $f(\mathbf{x})$ is used to define a new density which reduces the σ_j^2 in the next iteration, using histograms of the densities projected on each variable direction. This procedure is repeated until the last iteration m. Theoretically, σ_j^2 is minimized when:

$$\rho(\mathbf{x}) = \frac{|f(\mathbf{x})|}{\int_{\Omega}|f(\mathbf{x})|d^n\mathbf{x}}. \tag{6.6}$$

This means that the VEGAS shots are concentrated where the integrand is larger in magnitude. In VEGAS this condition is approximated subdividing the integration volume in hypercubes and filling each hypercube with the same number (in average) of shots. From iteration to iteration, the sampling grid of the hypercube on each axis is adapted to concentrate hypercubes in the regions where $|f(\mathbf{x})|$ is larger. Eventually, VEGAS is able to concentrate the random shots in an adaptive way with respect to the integrand and the corresponding uncertainty σ decreases as $\sim \frac{1}{\sqrt{N \cdot m}}$ [117].

6.3.2 Performance

The VEGAS parameters have been optimized in order to perform successfully the integration of the ME weight for the $H \to \tau\tau \to \mu\tau_h$ analysis. In particular, as explained in more details in the following of this Chapter, two matrix element hypotheses have been tested. For the signal, the matrix element related to the VBF Higgs boson production has been considered while the matrix element for the Drell-Yan (DY) processes has been used to test the background hypothesis. The evaluation of the integration performance were obtained in term of χ^2/dof distribution as well as in term of relative uncertainty on the integral value. Finally the distribution of the computing time needed to perform the full integration procedure is shown. In our studies, the value of m (number of iteration) is always fixed at 5. Thus, one of the most important parameter to optimize is the number of VEGAS shots per event N needed to evaluate the ME integral with an acceptable χ^2/dof (distributed around 1) without increase too much the computing time per event. The distribution of the χ^2/dof obtained testing the VBF matrix element on simulated VBF $H \to \tau\tau$ events is shown in Fig. 6.1. It is possible to see how the tails of the χ^2/dof distribution progressively reduce as N =1000, 10000, 50000 respectively. Since not much improvements are noticed in the reduction of the tails in the χ^2/dof distribution passing from N=10000 up to N=50000, it is decided to fix N=10000.

χ^2/dof threshold	fraction of events [%]
$\chi^2/dof < 10$	99.76
$\chi^2/dof < 5$	99.84
$\chi^2/dof < 2.5$	94.8
$\chi^2/dof < 2.0$	89.36
$\chi^2/dof < 1.5$	77.14

Table 6.1: *Percentage of events with a χ^2/dof smaller than a fixed threshold for N=10000.*

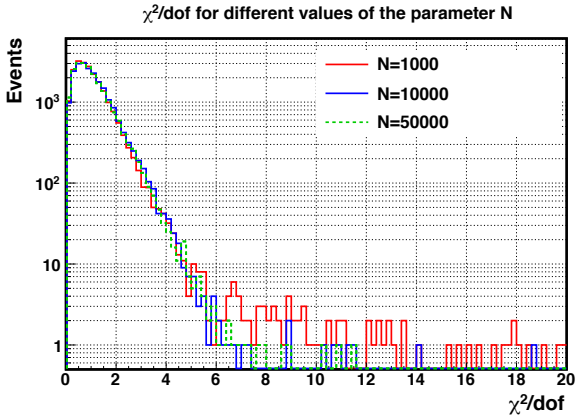


Figure 6.1: *Distribution of χ^2/dof for different values of N(=1000, 10000, 50000).*

Fixing N=10000, in Tab. 6.1 are shown the percentage of events with a χ^2/dof smaller than a fixed threshold.

With $m = 5$ and $N = 10000$ a scatter plot of the relative uncertainty on the integral value as function of the χ^2/dof has been obtained both testing the VBF and DY matrix element and it is shown in Fig. 6.2. It can be seen how the majority of the events present a χ^2/dof distributed around 1 with a relative uncertainty on the integral values that is between 0.5% and 5%.

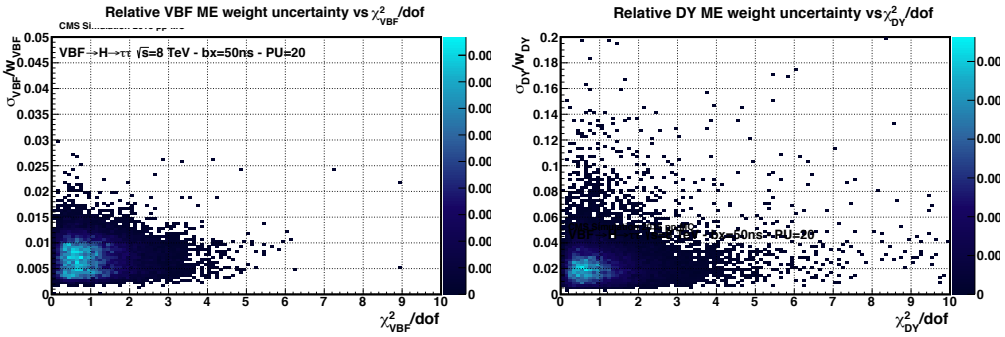


Figure 6.2: *Scatter plot of the relative uncertainty on the integral value $\bar{\sigma}/\bar{S}$ (for $m = 5$ and $N = 10000$) as function of the χ^2/dof obtained testing the signal (VBF $H \to \tau\tau$) matrix element (Left) and the Drell-Yan one (Right) on VBF $H \to \tau\tau$ simulated events.*

Finally, the distribution of the computing time per event is illustrated in Fig. 6.3. Given to the different complexity between VBF and DY matrix element, as explained more in detail in the next Section of this Chapter, the time needed to perform a full integration is quite different: in average $\sim 33\,s$ for the VBF ME and $\sim 53\,s$ for the DY ME.

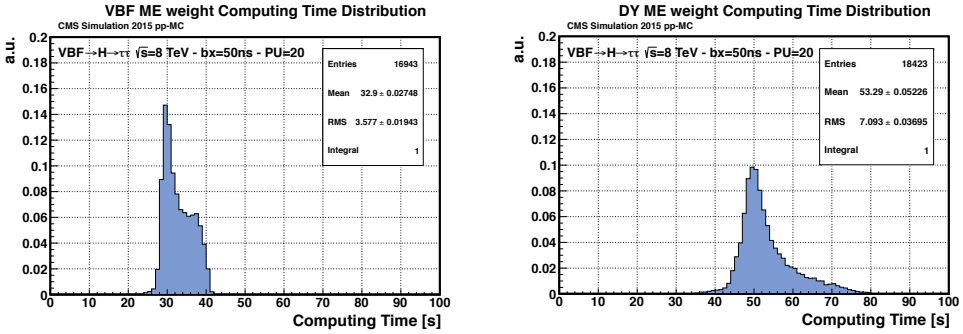


Figure 6.3: *Computing time per event (for $m = 5$ and $N = 10000$) distribution obtained testing the signal (VBF $H \to \tau\tau$) matrix element (Left) and the Drell-Yan one (right) on VBF $H \to \tau\tau$ simulated events.*

6.4 Computation of the Matrix Element weight

The ME method relies on the computation of a weight $w_i(\mathbf{y})$, which represents the probability that an event, characterized by a set of observables $\mathbf{y}$, is compatible with a process Ω. The process Ω would be for instance the signal from VBF-production of a Higgs decaying to τ

leptons or the Drell-Yan background. The weight w_Ω is given by the expression:

$$w_\Omega(\mathbf{y}) = \frac{1}{\sigma_\Omega} \sum_p \int d\mathbf{x} dx_a dx_b \frac{f(x_a, Q) f(x_b, Q)}{x_a x_b s} \delta^2(x_a P_a + x_b P_b - \sum_{p_k} p_k) |\mathcal{M}_\Omega(\mathbf{x})|^2 W(\mathbf{y}||\mathbf{x}), \tag{6.7}$$

where s represents the energy in the center-of-mass of the pp collision, the factor σ_Ω is fixed by the condition $\int d\mathbf{y} w_\Omega(\mathbf{y}) = 1$ and corresponds to the product of the cross-section of the process Ω times the detector acceptance and the efficiency of the analysis. Because of that normalization condition, all the overall multiplicative factors in the integrand will be omitted in the following developments, since they can be absorbed into σ_Ω. The summation is over the possible permutations p. In the case of the VBF signal, the two final jets being in principle indistinguishable, two permutations would have to be considered. Those permutations can also be taken into account in the computation of the matrix element squared $|\mathcal{M}_\Omega(\mathbf{x})|^2$, if we assign the quark q to the jet j and the quark q' to the jet j' and sum all the matrix elements over the possible permutations of the quarks. This is the approach adopted in this analysis. The $f(x, Q)$ term is the probability density as taken from the parton distribution functions (PDF) and $W(\mathbf{y}||\mathbf{x})$ is the product of all the transfer function.

The integration is performed over the Bjorken-fractions of the initial-state quarks x_a and x_b and the phase-space spanned by the final states particles. For VBF $H \to \tau\tau$, $dx_a dx_b d\mathbf{x} = dx_a dx_b d\mathbf{x}_q d\mathbf{x}_{q'} d\mathbf{x}_\tau d\mathbf{x}_{\bar{\tau}}$, with

$$d\mathbf{x}_p = \frac{d^3\vec{p}}{(2\pi)^3 2E_p}$$

The matrix element computed with MadGraph considers only the leading order (LO) diagrams and thus a perfect balance in the transverse plane is expected. This is not the case in data, where ISR, FSR and particles outside the detector acceptance or possible misreconstruction cause an imbalance in the transverse plane. The 4-dimensional $\delta^2(x_a P_a + x_b P_b - \sum_{p_k} p_k)$ function (expressing the total energy-momentum conservation) is thus replaced by the product of a 2-dimensional δ, allowing the conservation of the energy and enforcing the longitudinal component of the momentum, and the transfer function on the E_T^{miss}.

$$\delta^4 \to \delta^2 \left(E^{tot'}, P_z^{tot} \right) \cdot \mathcal{R}(\hat{\vec{p}}_T | \vec{P}_T^{tot}) \tag{6.8}$$

where the $\mathcal{R}(\hat{\vec{p}}_T | \vec{P}_T^{tot})$ term plays the role of a transfer function on the MET, and will be described in Section 6.6.4.

Therefore, we have to deal with a 14-dimensional integral. However, it is possible to reduce its dimensionality, taking into account some kinematic constraints and the reconstruction performance of the detector. Those dimensionality reductions will be described in the following sections and enable ultimately to compute a 4(5)-dimensional integral.

In this Chapter, we focus on the case where one of the τ decays into a muon or an electron (leptonic τ), while the other τ decays into hadrons (hadronic τ). In order to avoid double subscripts and unnecessarily heavy notations, the hadronic τ is assumed to be the positively

charged τ lepton ($\bar{\tau}$ denotes its four-vector), while the leptonic τ is assumed to be the negatively charged τ lepton (τ denotes its four-vector). The treatment of the cases where the two decay modes are switched is of course the same, up to polarization effects which are not taken into account in the present version of the analysis.

6.5 Scattering amplitude

The amplitude squared $|\mathcal{M}_\Omega(\mathbf{x})|^2$ accounts for the VBF-production of the Higgs and its decay into τ leptons. Its computation is done using an implementation based on C code generated with MadGraph. By default, it is averaged over spin states.

$$|\mathcal{M}_\Omega(\mathbf{x})|^2 = \bar{\sum_{r,r',s,s'}} |\mathcal{M}(q(r)q'(r') \to q(s)q'(s')H(\tau\bar{\tau})|^2$$

It is a function of x_a, x_b, q, q', τ and $\bar{\tau}$. The matrix element computed with MadGraph considers only the leading order (LO) diagrams and thus a perfect balance in the transverse plane is expected. In order to determine the 4-momentum of the incoming partons, the following procedure is applied. First, the total four-vector $P = q + q' + \tau + \bar{\tau}$, taking all the final-state particles into account, is computed. Then a transverse boost $-\vec{P}_T$ is applied to all the four-vectors. In such a way, the final particles four vectors are translated in a frame where the total transverse momentum is zero. The fractions of the proton energy carried by the partons a and b, labelled as $x_{a,b}$, can then be determined via

$$x_{a,b} = \frac{P^0 \pm |P^3|}{\sqrt{s}}$$

thanks to the two-dimensional δ-function, which ensures then the conservation of energy and longitudinal momentum between initial and final states. The incoming partons are assumed to have no transverse momentum. The matrix element is then evaluated using the four-vectors boosted in the transverse plane by the vector $\vec{P}_T$.

6.5.1 VBF matrix element

In the case of the VBF-production, the process can be generically represented as $q_1q_2 \to q_3q_4H(H \to \tau^-\tau^+)$. The full matrix element can therefore be written as

$$f(x_a,Q)f(x_b,Q)|\mathcal{M}(\mathbf{x})|^2 = \sum_{q_1q_2q_3q_4} f_{q_1}(x_1,Q)f_{q_2}(x_2,Q)|\mathcal{M}(q_1q_2 \to q_3q_4H, H \to \tau^-\tau^+)|^2(\mathbf{x})$$

where the q_i run over the possible quark and anti-quark flavors. Assuming the four-flavour scheme for the initial quarks (u, d, s, c), this leads to 48 processes and 168 matrix elements to evaluate, including all the possible permutations. Since the couplings of the Z boson to up-type and down-type quarks are different, as well as the coupling of the Z and of the W bosons with the Higgs boson, some differences are expected between the different processes.

To evaluate the different matrix elements involved, a leading-order Monte Carlo sample has been generated at parton level with MadGraph for the process $pp \to jjH(H \to \tau^-\tau^+)$, where p and j stand for g, u, d, c, s and the associated anti-quarks. To model the response of the VBF matrix elements to background events, a Monte Carlo sample has been generated with MadGraph for the process $pp \to jj\tau^-\tau^+$, where p and j stand for g, u, d, c, s and the associated anti-quarks. It does not include pure electroweak processes (like the VBF production of a Z boson) since they are suppressed by a factor $\alpha^2_{EW}/\alpha^2_{QCD}$ with respect to the other processes. In that case, the τ pair is produced from a Z boson or an off-shell Z boson or photon γ^*. The four-vectors of the particles at generator level are then used as inputs for the computation of the matrix elements.

In what follows, the processes will be divided into three categories: the $uu \to uuH$-like, the $ud \to udH$-like and the other processes. The $uu \to uuH$-like processes are the processes which involve a single quark flavor and are associated to two Feynman diagrams, as visible in Fig. 6.4. The $ud \to udH$-like processes are the processes for which the initial partons form a weak isospin-doublet and are also associated to two Feynman diagrams, one with W and one with Z boson. The other processes are associated to a single Feynman diagram, except for the processes like $u\bar{u} \to d\bar{d}H$ which have contributions from a VBF and from a VH diagram. The flavor repartition of the initial partons associated to the process which do not interfere one with each other is presented in Table 6.2.

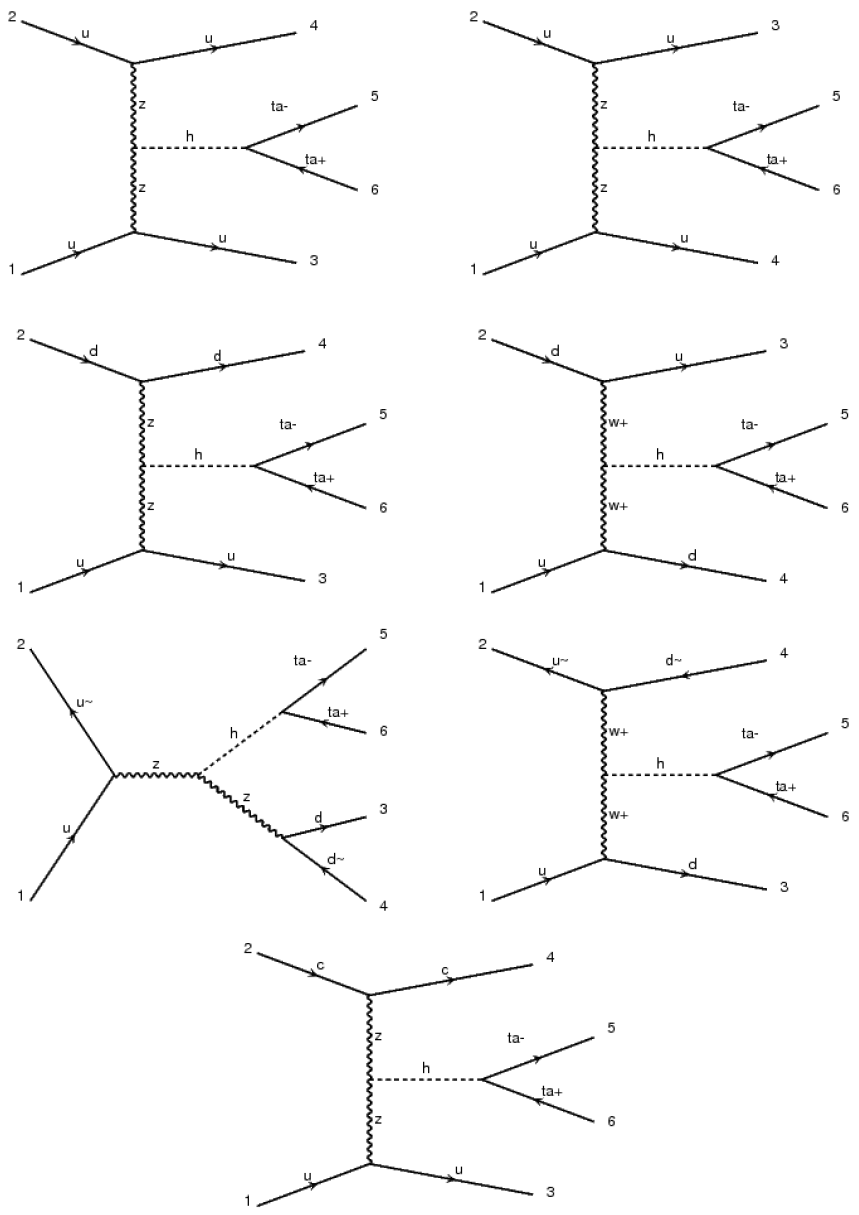


Figure 6.4: *Feynman diagrams of the processes* $uu \to uuH(H \to \tau^-\tau^+)$ *(*1^{st} *line),* $ud \to udH, H \to \tau^-\tau^+$ *(*2^{nd} *line),* $u\bar{u} \to d\bar{d}H, H \to \tau^-\tau^+$ *(*3^{rd} *line) and* $uc \to ucH, H \to \tau^-\tau^+$ *(*4^{th} *line).*

For the computation of the matrix element, the following labelling is used:

- p_1 is the four-vector of the incoming quark with $\eta > 0$
- p_2 is the four-vector of the incoming quark with $\eta < 0$
- p_3 is the four-vector of the outgoing quark with the highest η

Category of process	**Process**	**Fraction of the VBF events**
$uu \to uuH$-like	$uu \to uuH$	5.4%
	$dd \to ddH$	3.2%
	$qq \to qqH$, $q \neq u, d$	0.7%
	Total	**9.3%**
$ud \to udH$-like	$ud \to udH$	35.7%
	$cs \to csH + \bar{u}\bar{d} \to \bar{u}\bar{d}H$ $+\bar{c}\bar{s} \to \bar{c}\bar{s}H$	2.8%
	Total	**38.5%**
Other processes	$u\bar{u} \to d\bar{d}H$	8.5%
	$d\bar{d} \to u\bar{u}H$	5.6%
	$us \to dcH$	5.4%
	$u\bar{c} \to d\bar{s}H$	4.3%
	others	28.5%
	Total	**52.3%**

Table 6.2: *Repartition of the VBF events between the different classes of processes. For the other processes, only the ones representing more than 3% of all the VBF production are detailed.*

- p_4 is the four-vector of the outgoing quark with the lowest η

The matrix elements evaluation code has been generated with MadGraph such that the four-vector 1 and 3 (respectively 2 and 4) given as inputs are connected by a fermion line in one of the diagram involved in the process, as seen for example in the diagram $uc \to ucH$ in Fig. 6.4. The permutation where the p_i's are given as inputs in their natural order is referred as a "good" association. The permutation defined by switching p_3 and p_4, is referred to as a "bad" association. The distributions of each individual process has then been checked to identify the possible simplifications in the computation of the full matrix element (see Eq. 6.10). The comparisons for some $uu \to uuH$-like processes are for instance presented in Fig. 6.5. It can be seen that, as expected, the matrix elements for the processes $uu \to uuH$ and $cc \to ccH$ are equal and that they are very close to the ones for the processes $dd \to ddH$ and $ss \to ssH$[1]. Those comparisons enabled to identify in the end five representative VBF matrix elements. The distributions associated to those matrix elements are visible in Fig. 6.6.

[1]The pdfs weights are not included

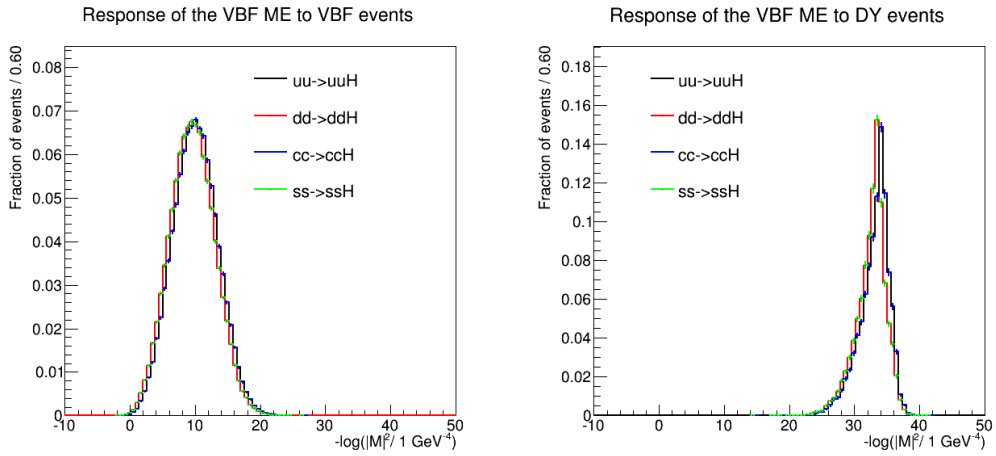


Figure 6.5: *Distributions of the matrix element squared for $uu \to uuH$-like processes in the VBF (Left) and the Drell-Yan sample (Right).*

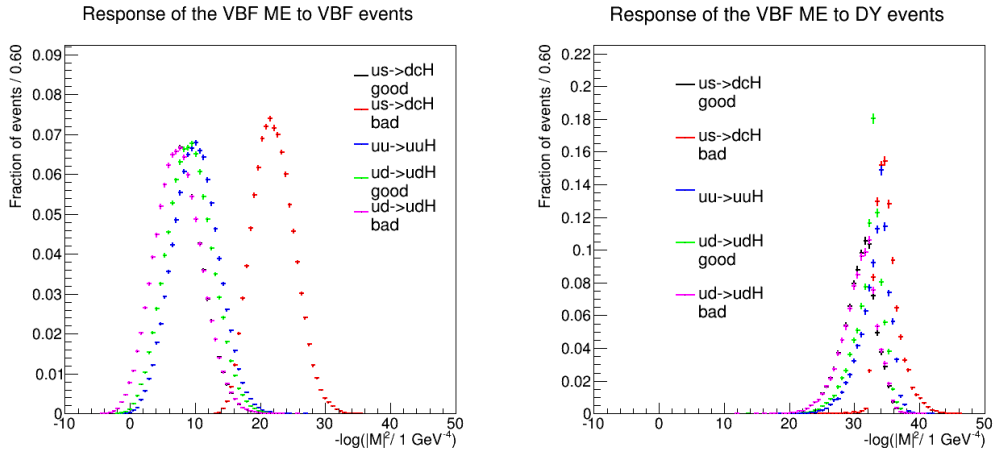


Figure 6.6: *Distributions of the matrix element squared for some representative processes, in the VBF (Left) and the Drell-Yan sample (Right). The distributions for the processes $us \to dcH$ good and $ud \to udH$ bad are almost superimposed.*

As expected, for the process $us \to dcH$ the good association yields a higher matrix element than the bad association. Indeed, the bad association corresponds to the configuration where the partons would go backwards (with respect to their initial direction) after the interaction. This is not the case for the process $ud \to udH$ because two Feynman diagrams are involved as seen in Fig. 6.4. Therefore the configuration where the final u quark goes in the same direction as the initial u quark gets a high matrix element because of the VBF diagram with Z bosons,

while the configuration where the final u quark goes in the same direction as the initial d quark gets a high matrix element because of the VBF diagram with W bosons.

To evaluate the full matrix element for the VBF production mode, all the contributions from the different processes have in principle to be taken into account, weighted by the associated PDF's.

$$\begin{aligned} f(x_1)f(x_2)|\mathcal{M}_{VBF}(\mathbf{x})|^2 &= \sum_{q_1,q_2,q_3,q_4} f_{q_1}(x_1)f_{q_2}(x_2)|\mathcal{M}(q_1q_2 \to q_3q_4H)|^2 \\ &= \sum_{q} f_q(x_1)f_q(x_2)|\mathcal{M}(qq \to qqH)|^2 \\ &+ \sum_{q_1 \neq q_2} f_{q_1}(x_1)f_{q_2}(x_2)|\mathcal{M}(q_1q_2 \to q_3q_4H)|^2 \\ &= \sum_{uu \to uuH-\text{like processes}} f_q(x_1)f_q(x_2)|\mathcal{M}(qq \to qqH)|^2 \\ &+ \sum_{ud \to udH-\text{like processes}} f_{q_1}(x_1)f_{q_2}(x_2)|\mathcal{M}(q_1q_2 \to q_3q_4H)|^2 \\ &+ \sum_{\text{other processes}} f_{q_1}(x_1)f_{q_2}(x_2)|\mathcal{M}(q_1q_2 \to q_3q_4H)|^2 \end{aligned} \tag{6.9}$$

Note that for $uu \to uuH$-like processes, a single permutation of the four-vectors has to be tested since there is a unique flavour participating to the interaction, whereas four permutations have to be taken into account for the $ud \to udH$-like and the other processes. In the computation of the full matrix element, this version is referred as the "full" matrix element. The computation is simplified as much as possible, taking into account the processes which have exactly the same matrix element (like $dd \to ddH$ and $ss \to ssH$ for instance).

Based on the previous observations, a simplification of the full matrix element has been tested and compared to the full version. In the simplified version, for each category ($uu \to uuH$-like, $ud \to udH$-like and other processes), the matrix element of all the processes is assumed to be equal to the matrix element of the representative process. For $uu \to uuH$-like processes, the matrix element is assumed to be $|\mathcal{M}(uu \to uuH)|^2$. For $ud \to udH$-like processes, the matrix element is assumed to be $|\mathcal{M}(ud \to udH)|^2$(good) for good associations and $|\mathcal{M}(ud \to udH)|^2$(bad) for bad associations. For the other processes, the matrix element is assumed to be $|\mathcal{M}(us \to dcH)|^2$(good) for good associations and $|\mathcal{M}(us \to dcH)|^2$(bad) for bad associations. The distributions of the full and of the simplified matrix element is presented in Fig. 6.7. As expected, they are very similar. However, since the full matrix element requires the computation of a lot of matrix elements, the simplified version turns out to be 10 times faster. For this reason, the simplified version is adopted in the rest of this thesis.

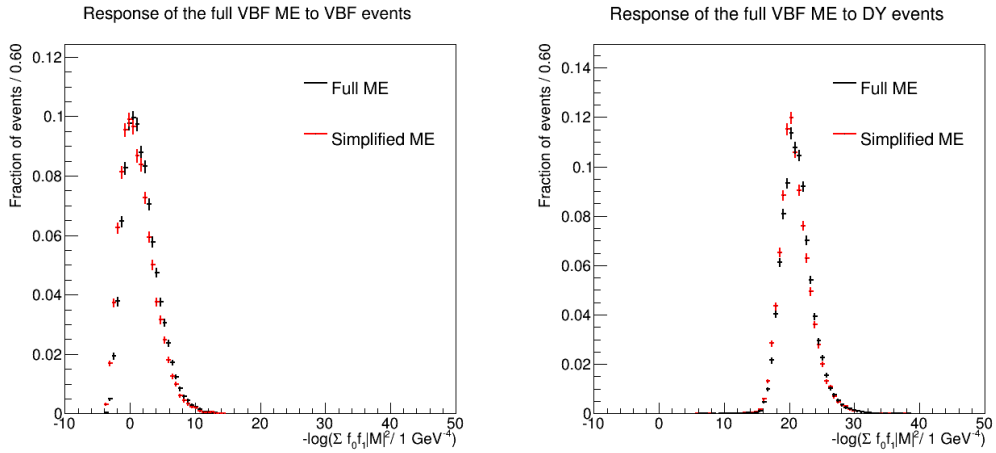


Figure 6.7: *Distributions of the matrix element squared with the full evaluation and the simplified one, in the VBF (Left) and the Drell-Yan sample (Right).*

6.5.2 Drell-Yan + 2 partons matrix element

The main irreducible background to the VBF $H \to \tau\tau$ search is represented by the Drell-Yan production of two τ leptons in association with two additional jets. It can be generically represented as $p_1p_2 \to p_3p_4Z/\gamma^*(Z/\gamma^* \to \tau^-\tau^+)$. The full matrix element can therefore be written as:

$$f(x_a,Q)f(x_b,Q)|\mathcal{M}(\mathbf{x})|^2 = \sum_{p_1p_2p_3p_4} f_{p_1}(x_1,Q)f_{p_2}(x_2,Q)|\mathcal{M}(p_1p_2 \to p_3p_4Z/\gamma^*(Z/\gamma^* \to \tau^-\tau^+))|^2(\mathbf{x}),$$

where the p_i's run over the possible quark and gluon flavors. Assuming no CKM mixing and nine possible flavors for the initial partons (g, u, d, s, c and the associated anti-quarks) and neglecting the pure electroweak processes (like the VBF production of a Z boson), this leads to 64 processes and 736 matrix elements to evaluate, including all the possible permutations. This number can be reduced, like for the VBF matrix element computation. To evaluate the different matrix elements involved, the same Monte Carlo samples as for the VBF matrix elements have been used.

In the following sections, the processes will be divided into seven categories: the $gg \to q\bar{q}$, the $q\bar{q} \to gg$, the $qg \to qg$, the $qq \to qq$, the $qq' \to qq'$, the $q\bar{q} \to q\bar{q}$ and the $q\bar{q} \to q'\bar{q}'$ processes. The number of processes in each category and the number of Feynman diagrams per process is indicated in Table 6.3. The repartition of the different categories in the Drell-Yan sample is presented in Table 6.4.

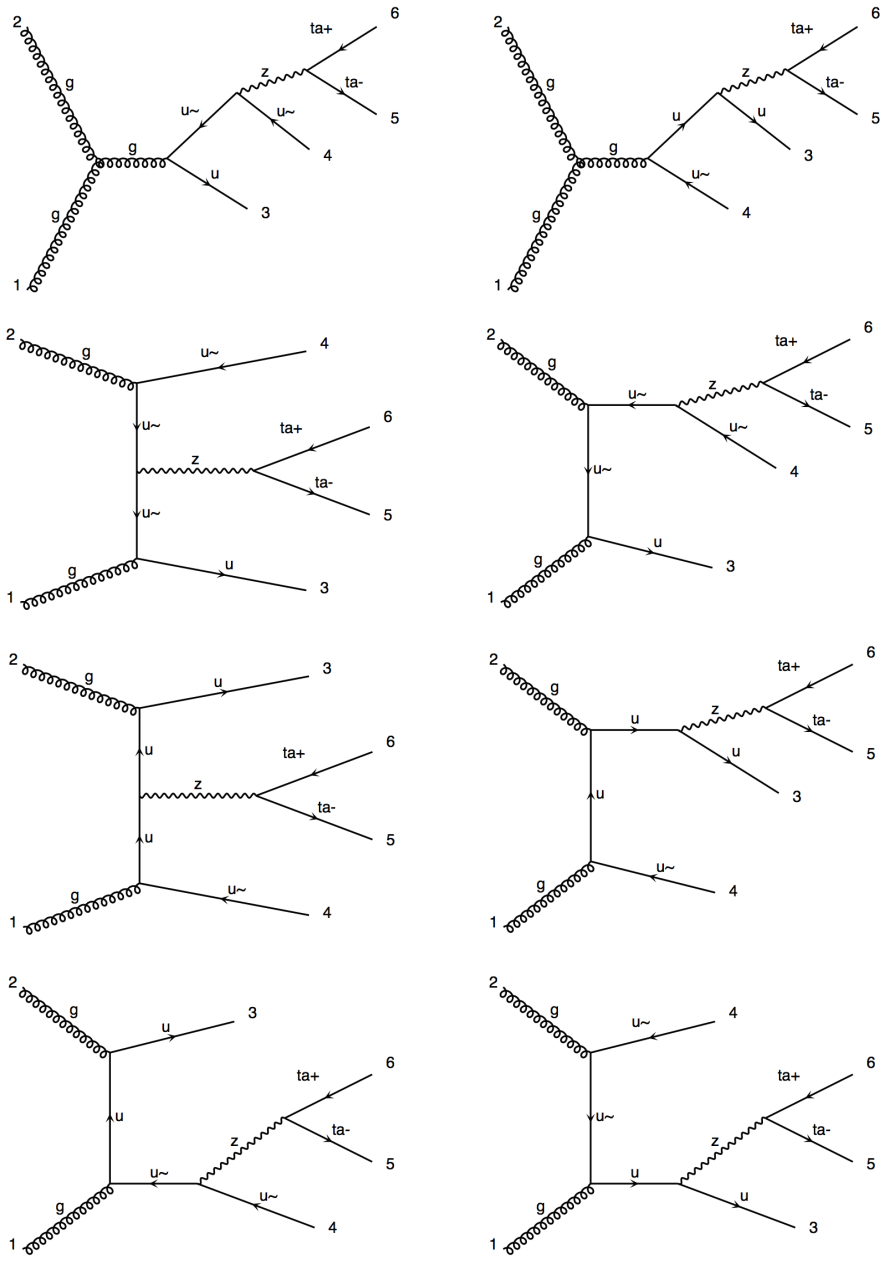


Figure 6.8: *Feynman diagrams of the process* $gg \to u\bar{u}Z(Z \to \tau^-\tau^+)$. *The diagrams including a* γ^* *are not represented there.*

By analogy with the evaluation of the VBF matrix element, the different possible permutations of the incoming and outgoing four-vectors are referred to as "good" and "bad" associations. Based on a similar analysis as for the VBF matrix element, the different matrix elements for the Drell-Yan + 2 partons process can be gathered into 11 categories:

Category of process	**Number of processes**	**Number of Feynman diagrams per process**
$gg \to q\bar{q}$	4	16
$q\bar{q} \to gg$	4	16
$qg \to qg$	8	16
$qq \to qq$	8	16
$qq' \to qq'$	24	8
$q\bar{q} \to q\bar{q}$	4	16
$q\bar{q} \to q'\bar{q}'$	12	8

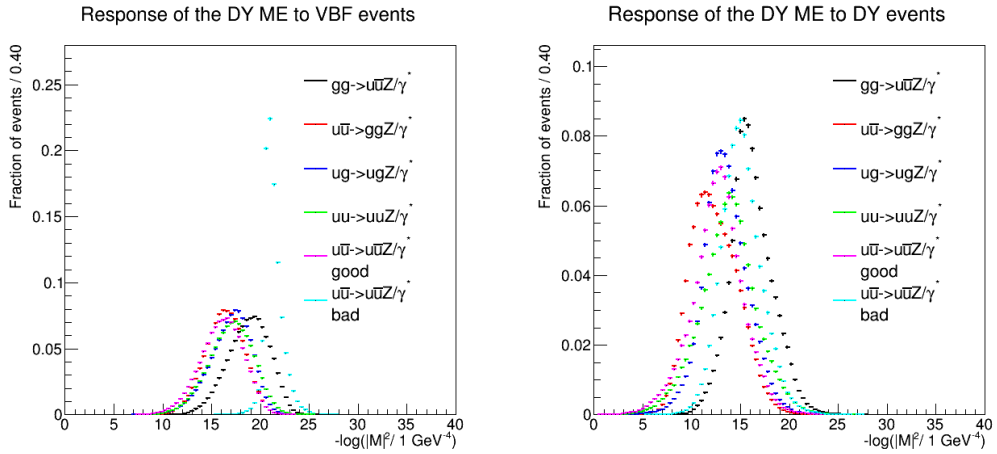
Table 6.3: *Repartition of the processes between the different categories and number of Feynman diagrams per process.*

Category of process	**Process**	**Fraction of the Drell-Yan events**
$gg \to q\bar{q}$	**Total**	**7.1%**
$q\bar{q} \to gg$	$u\bar{u} \to ggZ/\gamma^*$	3.3%
	others	3.7%
	Total	**7.0%**
$qg \to qg$	$ug \to ugZ/\gamma^*$	22.6%
	$dg \to dgZ/\gamma^*$	14.6%
	$cg \to cgZ/\gamma^*$	4.1%
	$sg \to sgZ/\gamma^*$	4.3%
	$\bar{u}g \to \bar{u}gZ/\gamma^*$	7.7%
	$\bar{d}g \to \bar{d}gZ/\gamma^*$	7.4%
	$\bar{c}g \to \bar{c}gZ/\gamma^*$	4.2%
	$\bar{s}g \to \bar{s}gZ/\gamma^*$	4.3%
	Total	**69.2%**
$qq \to qq$	**Total**	**2.5%**
$qq' \to qq'$	**Total**	**10.8%**
$q\bar{q} \to q\bar{q}$	**Total**	**2.5%**
$q\bar{q}' \to q\bar{q}'$	**Total**	**0.6%**

Table 6.4: *Repartition of the Drell-Yan events between the different categories of process. Only the individual processes representing more than 3% of all the Drell-Yan events are detailed.*

- the $gg \to q\bar{q}$ processes
- the $q\bar{q} \to gg$ processes
- the $qg \to qg$ processes
- the $qq \to qq$ processes
- the $qq' \to qq'$ processes for good associations
- the $uc \to uc$-like processes for bad associations, which gather the bad associations for the processes $uc \to ucZ/\gamma^*$, $\bar{u}\bar{c} \to \bar{u}\bar{c}Z/\gamma^*$, $ds \to dsZ/\gamma^*$ and $\bar{d}\bar{s} \to \bar{d}\bar{s}Z/\gamma^*$ (and all the ones obtained switching $u \leftrightarrow c$ and $d \leftrightarrow s$)
- the $ud \to ud$-like processes for bad associations, which gather all the other bad associations for the $qq' \to qq'$ processes
- the $q\bar{q} \to q\bar{q}$ processes for good associations
- the $q\bar{q} \to q\bar{q}$ processes for bad associations
- the $u\bar{u} \to c\bar{c}$-like processes for good associations, which gather the good associations for the processes $u\bar{u} \to c\bar{c}Z/\gamma^*$ and $d\bar{d} \to s\bar{s}Z/\gamma^*$ (and all the ones obtained switching $u \leftrightarrow c$ and $d \leftrightarrow s$)
- the $u\bar{u} \to d\bar{d}$-like processes, which gather the good and bad associations for the processes $u\bar{u} \to d\bar{d}Z/\gamma^*$ and $d\bar{d} \to u\bar{u}Z/\gamma^*$ and the bad associations for $u\bar{u} \to c\bar{c}Z/\gamma^*$ and $d\bar{d} \to s\bar{s}Z/\gamma^*$ (and all the ones obtained switching $u \leftrightarrow c$ and $d \leftrightarrow s$)

Inside each category, the most represented process (based on the PDFs) has been chosen as the representative process of the category. The distributions associated to those 11 representative processes are presented in Fig. 6.9.



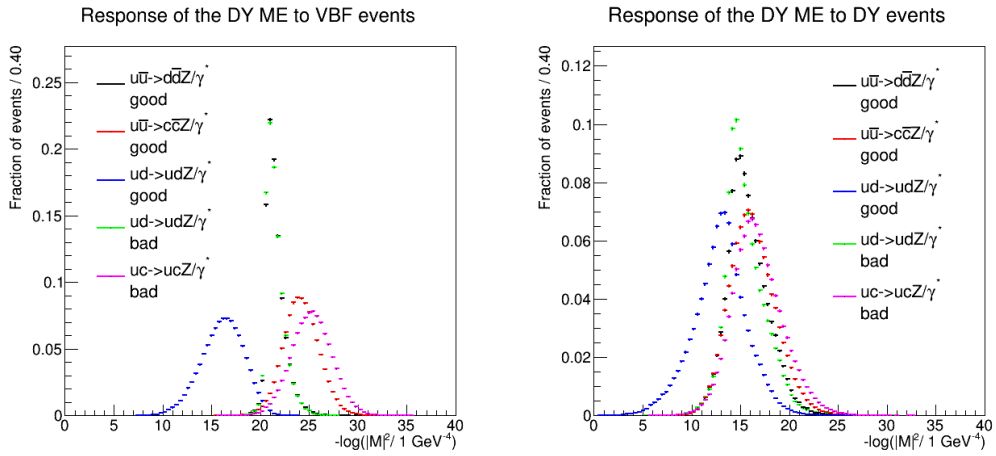


Figure 6.9: *Distributions of the matrix element squared for the 11 representative processes, in the VBF (Left) and the Drell-Yan sample (Right).*

A further simplification of the DY matrix element has been tested and compared to the full complete evaluation: for each of the 11 categories, the matrix element of all the processes is assumed to be equal to the matrix element of the representative process. As expected, the distributions are very similar. However, the simplified version results to be 7 times faster than the complete version since, the complete matrix element requires the evaluation of 92 matrix elements.

6.6 Transfer functions

The transfer function $W(\mathbf{y}||\mathbf{x})$ relates the set of four-momenta $\mathbf{x}$ with the set of observables $\mathbf{y}$. In the case of the VBF-production of a Higgs boson decaying to τ leptons, those observables are:

- the energy $\hat{E}_j$[2] and the direction $\hat{\vec{e}}_j$ of the two jets
- the three-momenta of the charged leptons $\hat{\vec{\ell}}$ from leptonic τ decays
- the three-momenta of the visible decay products $\hat{\vec{\pi}}$ from hadronic τ decays (either into 1 prong, 1 prong + π^0's or 3 prongs)
- the missing transverse energy $\hat{\vec{E}}_T^{miss}$

Given the performance of the reconstruction, based on the Particle Flow algorithm [96], and the predominant role of the tracking and the electromagnetic calorimeter in the objects involved, the following assumptions can be made:

[2]All the quantities like $\hat{x}$ are measured quantities.

- The direction of the quarks is assumed to be identical to that of the jets. In practice, the angular resolution for the jets reconstructed with the Particle Flow algorithm is better than 0.03 rad for $p_T > 20\,GeV$ [96].
- The momentum of the leptons is perfectly measured, both in direction and in magnitude. This assumption could be expressed mathematically by a Dirac delta of the form $\delta(\hat{\vec{\ell}} - \vec{\ell})$. In addition the lepton mass is neglected.
- The momentum of the visible decay products of hadronic τ's is perfectly measured, both in direction and in magnitude. This assumption could be expressed mathematically by a Dirac delta of the form $\delta(\hat{\vec{\pi}} - \vec{\pi})$

Given this assumption, the transfer function can be written as

$$W(\mathbf{y}||\mathbf{x}) = \delta(\hat{\vec{\ell}} - \vec{\ell}) \cdot \delta(\hat{\vec{\pi}} - \vec{\pi}) \cdot \prod_q \delta(\hat{\vec{e}}_j - \vec{e_q}) T_q(\hat{E}_j|E_q) \cdot T_l(\hat{\ell}|\tau_l) \cdot T_h(\hat{\pi}|\tau_h) \cdot T_{E_T}(\hat{\vec{\rho}}_T|\vec{P}_T),$$

where $T_l(\hat{\ell}|\tau_l)$ is the probability density to measure the momentum the momentum of the charged lepton $\hat{\ell}$ given the true momentum of the τ_l. The $T_h(\hat{\pi}|\tau_h)$ transfer function relates the momentum of the τ_h with the momentum of its visible decay products $\hat{\pi}$. The transfer function $T_q(\hat{E}_j|E_q)$ relates the energy of the reconstructed jet with the energy of the associated quark. The transfer function T_{E_T} compares the expected $\not{E}_T$ to the measured $\not{E}_T$ associated with the point in phase space considered for the integration. Each piece will be developed in the following sections.

The transfer functions, described in this Chapter, have been derived and validated using Monte Carlo simulation. In particular, the transfer functions for the jets are modelled comparing the four-vector of each reconstructed jet to the one of the corresponding quark at generator level. For the τ transfer function, the situation is more complicated because of the presence of real neutrinos in the decay chain and the different decay channels. In addition, a proper validation of their normalizations was needed. Finally, the transfer function for the missing transverse energy is checked. This transfer function, as it is described further in this section, plays a central role in the determination of the full matrix element score.

6.6.1 Jets

The transfer function for the jets is supposed to account for the experimental limits coming from the jet reconstruction procedure in terms of energy response and resolution. The conditional probability density that the observed jet energy ($\hat{E}_j$) corresponds to the real parton energy (E_q) is modelled via a Gaussian function that is parametrized as a function of the true parton energy:

$$T_q(\hat{E}|E,\eta) = \frac{1}{\sigma(E_q)\sqrt{2\pi}} e^{-\frac{\left(E_j - E_q\right)^2}{2\sigma^2(E_q)}}. \tag{6.10}$$

For this study, jets reconstructed with the Particle Flow algorithm were used and the proper jet corrections (described in the Chaper 3.2) are applied: the *L1* correction that takes care of the PU mitigation, the *L2* correction that aims to recover a relative jet response that is η independent and a *L3* correction whose purpose is to re-establish an absolute jet energy response close to ~ 1.0. The response and the resolution of the reconstructed jets have been studied as function of the true parton energy and are shown in Fig. 6.10, Fig. 6.11 and Fig. 6.12. Three different parametrizations of the jet transfer function have been derived for three η bins: one bin covering the barrel region ($|\eta| < 1.5$) and two bins covering the endcap region (up to $|\eta| = 4.5$). Above $|\eta| = 4.5$ no Jet Energy Correction (JEC) are available. The endcap region is thus subdivided in two η-zone: the one extending within $1.5 < |\eta| < 3.0$ and the other covering up to $|\eta| = 4.5$. In addition, a minimal transverse momentum of $15\,GeV$ has been required for the reconstructed jets. In the purpose of this analysis, the jet transfer function parametrization is considered independent on the parton's flavour.

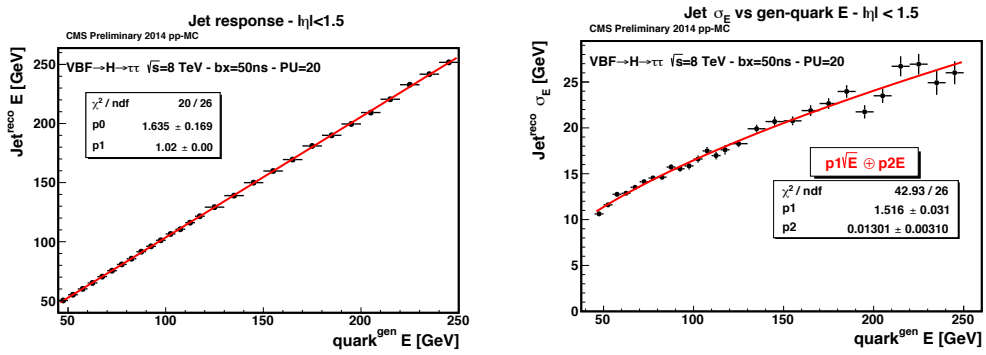


Figure 6.10: *Reconstructed jet energy as function of the generator parton energy (Left) and reconstructed jet energy resolution as function of the generator parton energy for reconstructed jet* $|\eta| < 1.5$ *(Left).*

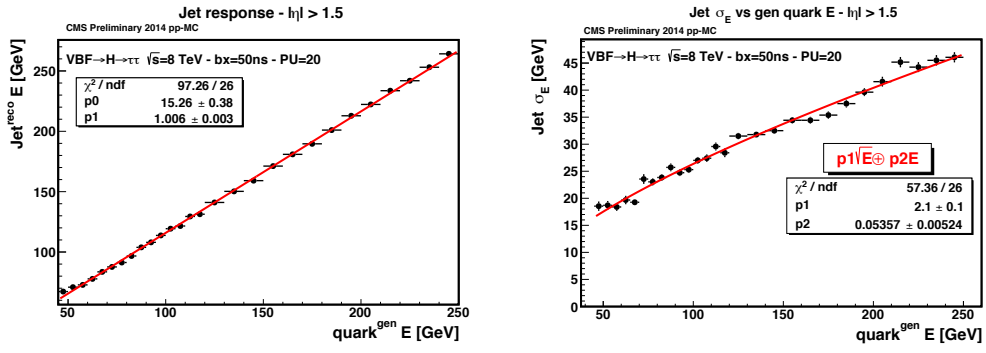


Figure 6.11: *Reconstructed jet energy as function of the generator parton energy (Left) and reconstructed jet energy resolution as function of the generator parton energy for reconstructed jet* $1.5 < |\eta| < 3.0$ *(Left).*

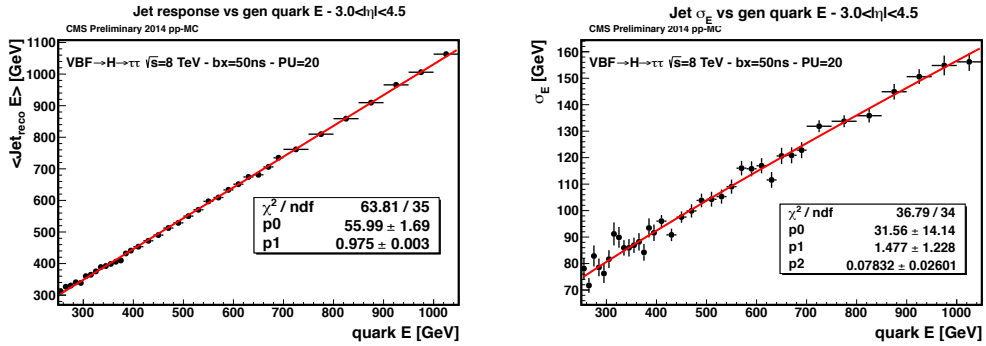


Figure 6.12: *Reconstructed jet energy as function of the generator parton energy (Left) and reconstructed jet energy resolution as function of the generator parton energy for reconstructed jet* $|\eta| > 3.0$ *(Left).*

The procedure adopted to derive the parametrization of the jet transfer function consists in obtaining, for different true parton energy bins, the corresponding reconstructed jet energy distribution that is then fitted with a Gaussian function in order to determine the associated mean value and sigma. A $\Delta R < 0.5$ matching criteria between the reconstructed jet and the parton at generator level has been required. An example of such a fit is shown in Fig.6.13, where the values of parton energy are considered in the range $[200, 210]$ GeV.

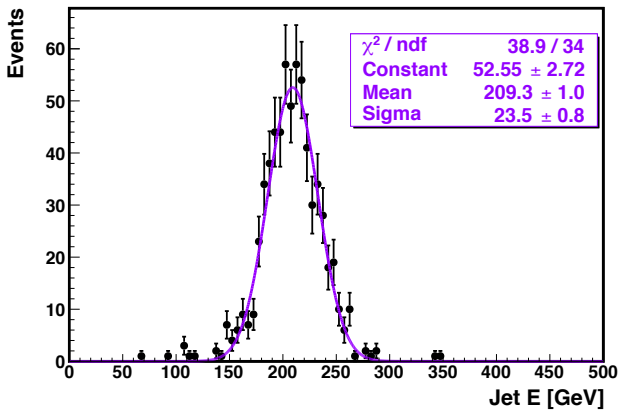


Figure 6.13: *Distribution of the reconstructed Jet energy for a bin in true parton energy* $[200, 210]\, GeV$. *The gaussian fit is also shown.*

The response and the resolution are fitted respectively with a linear function of the form $< E_j >= m(\eta) \cdot E_q + q(\eta)$ and with a more involved function $\sigma(E, \eta)) = p_1(\eta)\sqrt{E} \oplus p_2(\eta)E$. The values of the fit parameters are reported in Table 6.5.

The normalization term of the jet transfer function is such that the integral of the jet transfer function over the observable quantity (the reconstructed jet energy) is equal to unity. Given the Gaussian shape of Eq. 6.10, the normalization term is equal to $\frac{1}{\sigma(E_q)\sqrt{2\pi}}$.

η	m	q	p_0	p_1	p_2
$\|\eta\| < 1.5$	1.02 ± 0.009	1.6 ± 0.2	-	1.52 ± 0.03	0.013 ± 0.003
$1.5 < \|\eta\| < 2.5$	0.981 ± 0.003	17.4 ± 0.4	-	2.4 ± 0.1	0.020 ± 0.005
$\eta > 2.5$	0.975 ± 0.003	56.0 ± 1.7	31.6 ± 14.1	1.5 ± 1.23	0.08 ± 0.026

Table 6.5: *Table summarizing the various coefficients that parametrize the jet transfer function in the three eta bins considered.*

6.6.2 Hadronic τ

The variety of the τ hadronic decay modes translates into a complicated signature for the final state (the stable particles detectable give rise to a low-multiplicity jet). In order to validate the τ transfer function, a Monte Carlo sample is generated with τ pairs coming from Higgs boson decays. Each τ at generator level is associated with its daughters. The association procedure is checked by plotting the distribution of the invariant mass of the final hadronic system (Fig. 6.14), also called τ visible mass. For the 1-prong decay mode the τ visible mass is that of the pion and appears as peak in Fig. 6.14 at $139.6\,MeV/c^2$. For the 3-prongs decay mode the invariant mass of the hadronic system corresponds to the mass of the intermediate $a1(1260)$ resonance. As a matter of fact, because of the intrinsic decay width ($\sim 250\,MeV/c^2$) the τ visible mass distribution is spread out. For what regards the 1-prong + π^0s decays, the invariant mass of the final hadronic state has a more complex shape because of the coexistence of the two $a1(1260)$ and $\rho(770)$ intermediate resonances.

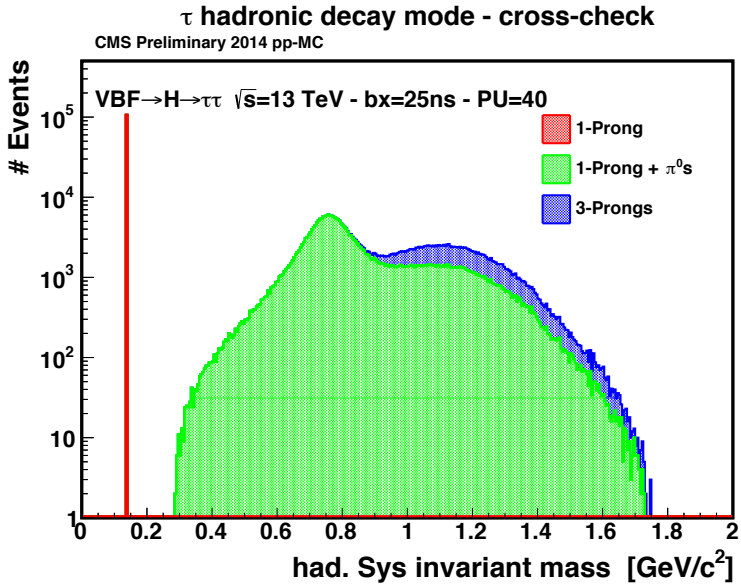


Figure 6.14: *Invariant mass distribution of the final hadronic system for all the τ decay mode analyzed.*

1-prong τ_h decay

In that case, the energy transfer function is given by:

$$T_{h1}(E_\pi|E_\tau) \propto \frac{d\Gamma}{dE_\pi} = \frac{d|\vec{\pi}|}{dE_\pi}\frac{d\Gamma}{d|\vec{\pi}|} \propto \frac{E_\pi}{|\vec{\pi}|}\int \frac{1}{E_\tau}\frac{|\vec{\pi}|^2 d\Omega_\pi}{(2\pi)^3 2E_\pi}\frac{d^3\vec{\nu}}{(2\pi)^3 2E_\nu}\delta^4(\tau-\pi-\nu)|\mathcal{M}_\Gamma(\tau\to\nu\pi)|^2,$$

with: $|\mathcal{M}_\Gamma(\tau\to\nu\pi)|^2$ the matrix-element corresponding to the 1-prong decay. The spin-averaged matrix element is given by

$$\frac{1}{2}\bar{\Sigma}|\mathcal{M}(\tau\to\pi\nu)|^2 = G_F^2|V_{ud}|^2 m_\tau^4 f_\pi^2\left(1-\frac{m_\pi^2}{m_\tau^2}\right).$$

As expected, it is independent on the four-vectors. Therefore it is just an overall multiplicative term and can be absorbed in the σ_Ω of Eq. 6.7. From the energy-momentum conservation, we then deduce that

$$\boxed{\cos\theta_{\tau\pi} = \frac{2E_\tau E_\pi - (m_\tau^2+m_\pi^2)}{2|\vec{\tau}||\vec{\pi}|}} \tag{6.11}$$

After the integration, one finally gets,

$$T_{h1}(E_\pi|E_\tau) \propto \frac{1}{|\vec{\tau}|E_\tau}.$$

Requiring in addition the normalization condition:

$$\int_{(m_\pi^2/m_\tau^2)E_\tau}^{E_\tau} dE_\pi T_{h1}(E_\pi|E_\tau) = 1$$

leads to

$$\boxed{T_{h1}(E_\pi|E_\tau) = \frac{1}{E_\tau\left(1-\frac{m_\pi^2}{m_\tau^2}\right)}}. \tag{6.12}$$

Now, assuming that we perfectly measure the momentum of the pion, both in magnitude and in direction, the overall transfer function for the hadronic τ is given by

$$T_{h1}(E_\pi|E_\tau)\delta(\hat{\vec{\pi}}-\vec{\pi})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \delta(\hat{\vec{\pi}}-\vec{\pi})\frac{|\vec{\tau}|^2}{E_\tau^2\left(1-\frac{m_\pi^2}{m_\tau^2}\right)}d|\vec{\tau}|d\cos\theta_{\tau\pi}d\phi_{\tau\pi}$$

We have to take the constraint on $\cos\theta_{\tau\pi}$ given by Eq. (6.11) into account. The measurement of the momentum of the pion allows removing the integration over $\cos\theta_{\tau\pi}$, leading to

$$\boxed{T_{h1}(\pi|\tau)\delta(\hat{\vec{\pi}}-\vec{\pi})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \frac{|\vec{\tau}|^2}{E_\tau^2\left(1-\frac{m_\pi^2}{m_\tau^2}\right)}d|\vec{\tau}|d\phi_{\tau\pi} \equiv \tilde{T}_{h1}(\hat{\pi}|\tau)d|\vec{\tau}|d\phi_{\tau\pi}}} \tag{6.13}$$

Moreover, for a physical solution, we must have $-1 \leq \cos\theta_{\tau\pi} \leq 1$. This constraint leads to:

$$\boxed{|s^-| \leq |\vec{\tau}| \leq s^+} \tag{6.14}$$

with

$$\boxed{s^{\pm} = \frac{(m_\tau^2 + m_\pi^2)|\vec{\pi}| \pm E_\pi(m_\tau^2 - m_\pi^2)}{2m_\pi^2}}. \tag{6.15}$$

In the ultra-relativistic regime Eq. (6.14) is equivalent to the natural constraint

$$\boxed{E_\pi \leq E_\tau \leq \frac{m_\tau^2}{m_\pi^2} E_\pi} \tag{6.16}$$

1-prong + π^0 and 3-prongs τ_h decays

The case of a hadronic τ decaying into 1-prong + π^0(s) and 3-prongs is in principle not as straightforward. Indeed, since it involves hadronic resonances, a simple analytic computation is not possible as was the case for the decay into 1-prong. The constraints given by Eq. 6.11 and 6.14 still remain valid.

Angular part

The two-body decay allows the cosine of the angle between the momentum vector of the τ and the momentum vector of the final hadronic system to be unambiguously expressed as a function of the energy of the tau and of the pion:

$$\boxed{\cos\theta_{\tau\pi} = \frac{2E_\tau \hat{E_\pi} - (m_\tau^2 + m_\pi^2)}{2|\vec{\tau}||\hat{\vec{\pi}}|}} \tag{6.17}$$

This quantity plays a crucial role in the reconstruction of the hadronic τ four-vector, as explained in Section 6.6 and 6.7. For this reason it is important to cross-check that, at generator level, the Eq. 6.17 is correct, and, in addition, that this expression is still valid for the hadronic decay modes that involve intermediate resonances. This quantity has been evaluated directly as the scalar product of the τ and hadronic system directions and superimposed to the one evaluated through the Eq. 6.17. The distribution of $\cos\theta_{\tau\pi}$ (where π stands for a generic hadronic system among those taken into consideration in this analysis) are illustrated in Fig.6.15.

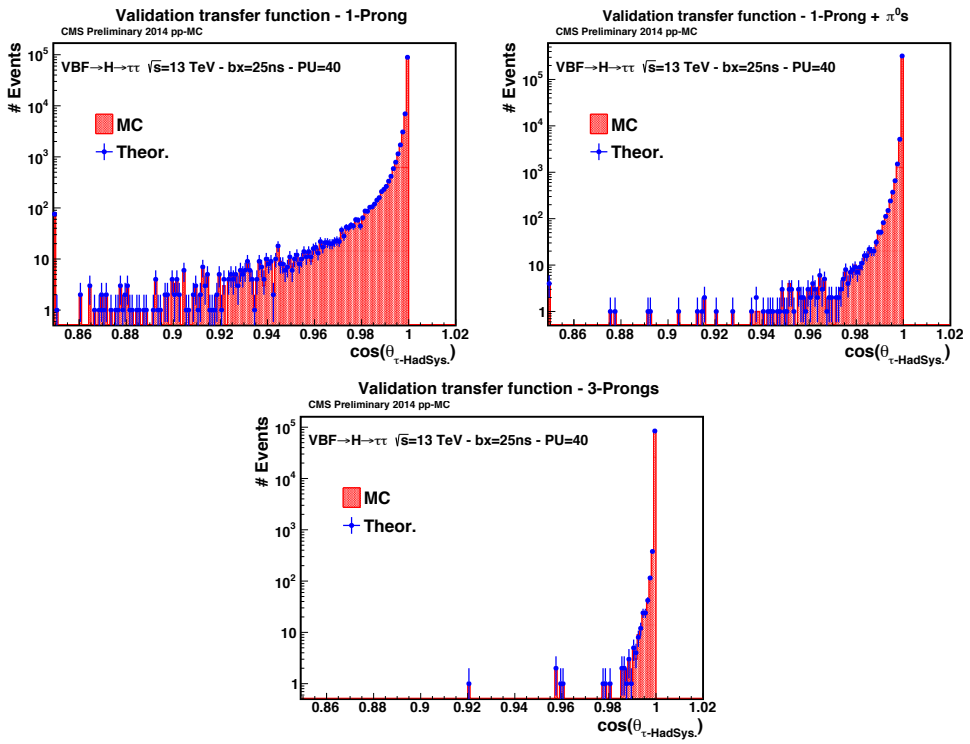


Figure 6.15: *Distribution for the cosinus of the angle between the τ and the hadronic system evaluated from the true information available at generator level (red histograms) and computed with the Eq. 6.17. The validation has been performed for the 1-prong, 1-prong+π^0(s) and 3-prong τ decays.*

Moreover, to better investigate the precision of the theoretical formulae, the scatter plot between $\cos\theta_{\tau\pi}$ evaluated directly from the simulation and the one coming from Eq. 6.17 are shown in Fig.6.16.

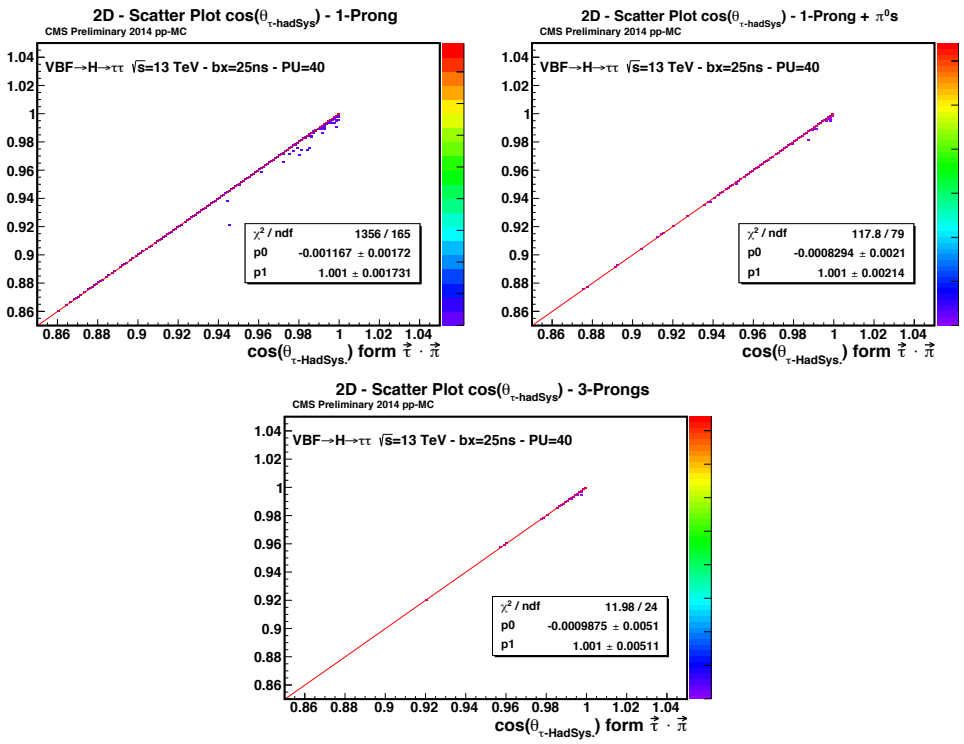


Figure 6.16: *On the y-axis is shown the cosine between the τ and its corresponding hadronic system as computed from Eq. 6.17, while on the x-axis is shown the value of the same cosine evaluated from the MC truth.*

As it can observed, the agreement between the $\cos\theta_{\tau\pi}$ (where π evaluated directly from the MC and those obtained through Eq. 6.17) is very satisfying. Only a small number of events do not lie on the diagonal. This is due to the very high numerical precision needed to compute the cosine from a difference between two quantities, the energy and the momentum, that are very close to each other when the particle is in an ultra relativistic regime. Also, the good agreement obtained using 1-prong+π^0s and 3-prongs τs is a confirmation that the 2-body decay approximation works well in describing the decay angles of the hadronic τ final products.

Energy transfer function

Unlike for the 1-prong case, the invariant mass of the visible decay products cannot be assumed to be fixed but has to be defined as $m_\pi^2 = E_\pi^2 - |\vec{\pi}|^2$.

Still, it is reasonable to assume that, for a fixed value of m_π^2, the matrix element $|\mathcal{M}_\Gamma(\tau \to \nu\pi)|^2$ does not depend on any other kinematic quantity. Therefore, we expect a flat distribution for the variable $z = E_\pi/E_\tau$ for z between $\frac{m_\pi^2}{m_\tau^2}$ and 1 (from Eq. 6.12). This is indeed what is observed in Fig. 6.19 and in Fig. 6.20. Based on that, the transfer function for hadronic τ's decaying into 3 particles is assumed to be the same as 6.12, which leads again to the following

transfer function:

$$\boxed{T_{h3}(\pi|\tau)\delta(\hat{\vec{\pi}} - \vec{\pi})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \frac{|\vec{\tau}|^2}{E_\tau^2} d|\vec{\tau}|d\phi_{\tau\pi} \equiv \tilde{T}_{h3}(\hat{\pi}|\tau)d|\vec{\tau}|d\phi_{\tau\pi}}. \tag{6.18}$$

As part of the validation of the hadronic transfer function, it is worth checking the boundaries of the integral over $\vec{\tau}$. The inequality to be satisfied is the following (also reported in Eq. 6.14):

$$\boxed{|\hat{t}^-| \leq |\vec{\tau}| \leq \hat{t}^+} \tag{6.19}$$

with

$$\boxed{\hat{t}^\pm = \frac{(m_\tau^2 + m_\pi^2)|\hat{\vec{\pi}}| \pm \hat{E}_\pi(m_\tau^2 - m_\pi^2)}{2m_\pi^2}} \tag{6.20}$$

In order to verify such quantities, scatter plots of $\frac{|t^-|}{\vec{\tau}}$ and $\frac{\vec{\tau}}{t^+}$ have been obtained and shown in Fig. 6.17. As expected, the points are all located within the unit square.

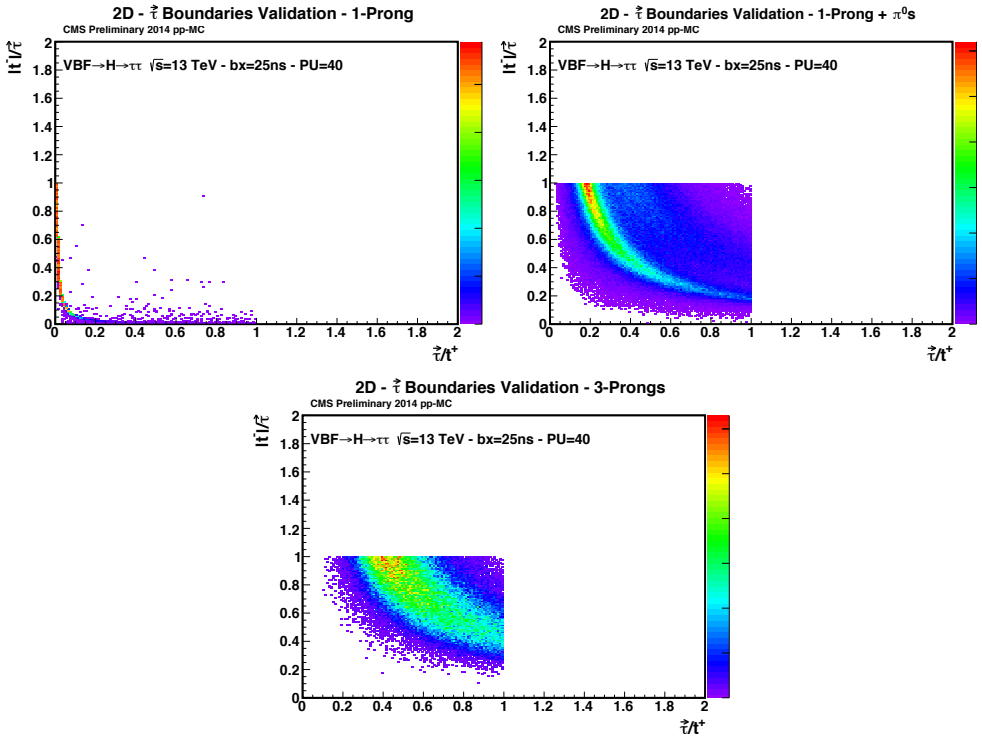


Figure 6.17: *Validation of the integral boundaries for the τ hadronic decays.*

For what regards the energy transfer function for an hadronic τ, the situation is more complex. Indeed, the 2-body decay approximation for 1-prong + π^0s and 3-prongs cases, as already shown, doesn't have an impact on the determination of the angle between the τ and its

decay products since the decay products are in general heavily boosted. Thus, Eq. 6.17 could be adopted for all the hadronic τ decay modes.

The distributions of $z = E_{vis}/E_\tau$, the daughters energy ($E_{vis} < E_\tau$ always because of neutrinos) over the energy of the τ, are affected by the intermediate resonance mass and width. The plot in Fig. 6.18 shows the fractional energy carried by the τ visible daughter(s) normalized to the mother τ energy for different decay modes at the generator level. The theoretical derivation of such distributions is, in principle, feasible [75], but not trivial. The solution adopted in the present analysis is to model the fraction of visible energy over the τ energy (called z) as a function of the visible mass, independently on the decay channel. This approach is justified by the plots visible in Fig. 6.19 (respectively 6.20) for the 1-prong + π^0(s) (respectively 3-prongs) where it can be seen that the transfer function can be modelled by a Heaviside distribution, the turn-on point of that function being m^2_{vis}/m^2_τ.

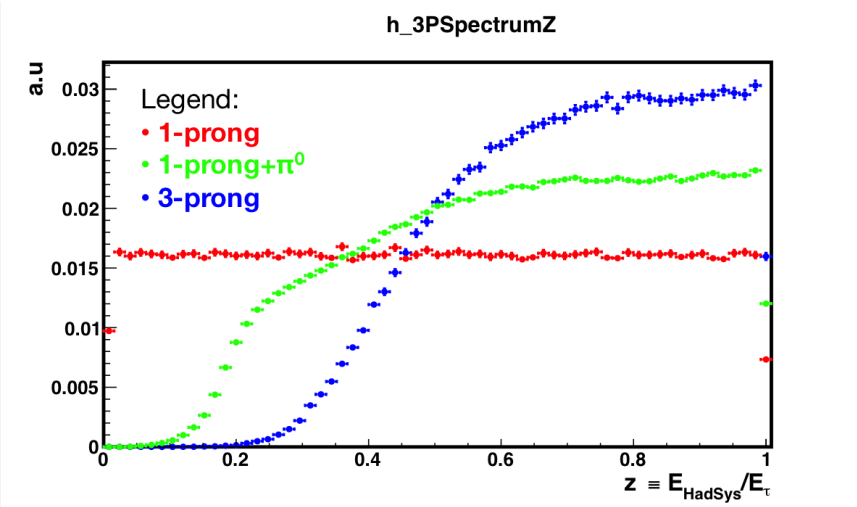


Figure 6.18: *$z = E_{vis}/E_\tau$ distribution for the different hadronic τ decay modes.*

Thus, the normalization factor is calculated in a straightforward way requiring the integration over the z distribution to be equal to 1. In this way, it is possible to write the energy transfer function for an hadronic τ independently on its decay mode:

$$T_{h1}(E_\pi|E_\tau) = \frac{1}{1 - \frac{m^2_{vis}}{m^2_\tau}} \cdot \Theta(\frac{m^2_{vis}}{m^2_\tau}) \tag{6.21}$$

In order to validate the analytical expression in Eq. 6.21, the z distribution obtained from Monte Carlo is compared directly to the normalized function for different bins in visible τ mass as illustrated in Fig. 6.19 and 6.20. A good agreement between the function described in 6.21 and the simulation is found. The small disagreements appearing around the turn on point of the Heaviside function with respect to what is obtained from the simulation is due to binning

effects. Indeed in order to have enough statistic, finite bins in the visible τ mass have been considered.

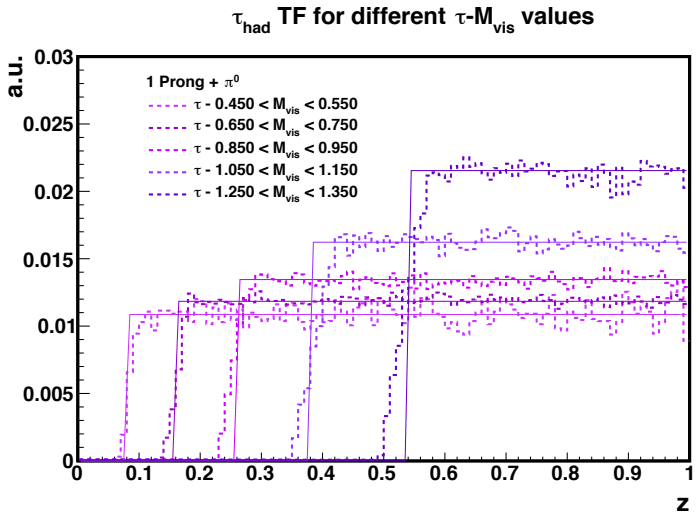


Figure 6.19: *$z = E_{vis}/E_\tau$ distribution and corresponding transfer function (continuous line) for the τ decaying into 1-prong + π^0 in different visible τ mass bins.*

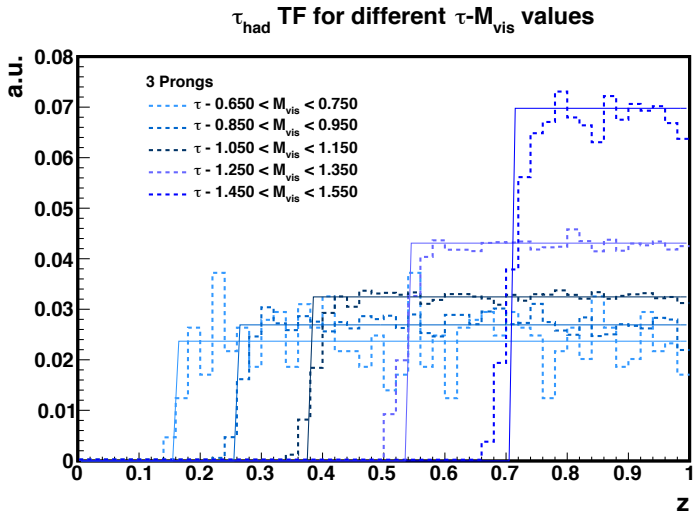


Figure 6.20: *$z = E_{vis}/E_\tau$ distribution and corresponding transfer function (continuous line) for the τ decaying into 3-prongs in different visible τ mass bins.*

To definitely validate the hadronic τ transfer function as reported in Eq. 6.21, a convolution between the $T_{h1}(E_\pi|E_\tau)$ and the visible τ mass spectrum has been performed in order to recover the inclusive z distribution shown in Fig. 6.18. The results of this check are illustrated separately

for 1-prong + π^0 and 3-prong decay modes in Fig. 6.21 and in Fig. 6.22, respectively. The inclusive z distribution obtained through the convolution of $T_{h1}(E_\pi|E_\tau)$ with the visible τ mass spectrum is compared directly to the one extracted from the Monte Carlo. It is possible to conclude from this comparison the agreement is almost perfect, and thus the adoption of Eq. 6.21 to model the τ hadronic tranfer function is justified.

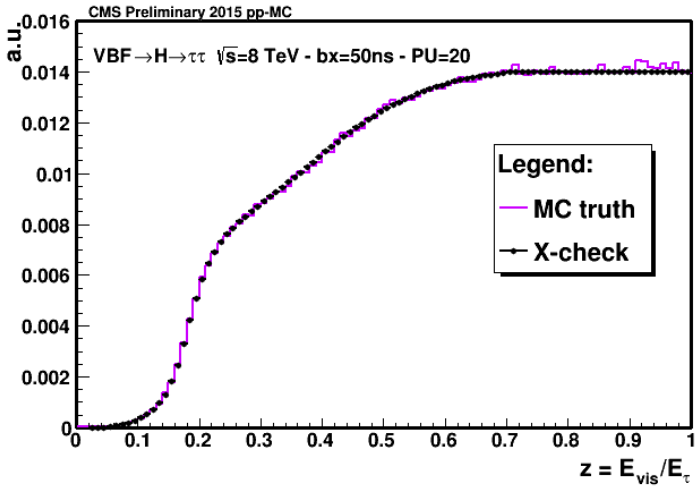


Figure 6.21: *Comparison between the $z = E_{vis}/E_\tau$ inclusive distribution obtained through the convolution of the τ hadronic transfer function with the τ visible mass spectrum (black dots) and the MC truth (violet histogram) for the τ decaying into 1-prong + π^0.*

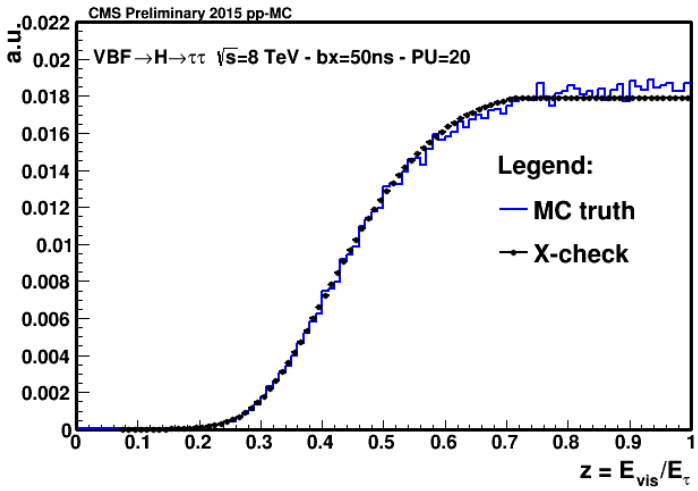


Figure 6.22: *Comparison between the $z = E_{vis}/E_\tau$ inclusive distribution obtained through the convolution of the τ hadronic transfer function with the τ visible mass spectrum (black dots) and the MC truth (blue histogram) for the τ decaying into 3-prong.*

6.6.3 Leptonic τ

In the case of the leptonic decays of the τ, the leptonic τ transfer function is given by

$$T_l(\ell|\tau) \propto \frac{d^3\Gamma}{d^3\vec{\ell}} \propto \frac{1}{E_\tau}\int \frac{1}{(2\pi)^3 2E_\ell}\frac{d^3\vec{\nu}}{(2\pi)^3 2E_\nu}\frac{d^3\vec{\bar{\nu}}}{(2\pi)^3 2E_{\bar{\nu}}}\delta^4(\tau-\ell-\nu-\bar{\nu})|\mathcal{M}_\Gamma(\tau\to\ell\nu\bar{\nu})|^2$$

with $|\mathcal{M}_\Gamma(\tau \to \ell\nu\bar{\nu})|^2$ the matrix-element corresponding to the leptonic decay. The spin-averaged matrix element is given by

$$\frac{1}{2}\bar{\Sigma}|\mathcal{M}(\tau\to\ell\nu\bar{\nu})|^2 = 64G_F^2(\nu.\ell)(\bar{\nu}.\tau).$$

One gets then:

$$T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell}\int \frac{d^3\vec{\nu}}{2E_\nu}\frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}}(\nu.\ell)(\bar{\nu}.\tau)\delta^4(\tau-\ell-\nu-\bar{\nu})$$

A quite long calculation then leads to:

$$\boxed{T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell}[(\ell.\tau)(m_\tau^2+m_\ell^2-2l.\tau)+2(l.\tau-m_\ell^2)(m_\tau^2-\ell.\tau)]} \tag{6.22}$$

Moreover, the requirement that E_ν and $E_{\bar{\nu}}$ are positive leads to the following constraints:

$$\boxed{\cos\theta_{\tau\ell} \geq \frac{2E_\tau E_\ell - m_\tau^2 - m_\ell^2}{2|\vec{\tau}||\vec{\ell}|} \equiv \alpha}. \tag{6.23}$$

This constraint has to be taken into account in the integration. Indeed, since we must have $\cos\theta_{\tau\ell} \leq 1$, this puts a constraint over the integration range for $|\vec{\tau}|$. The same kind of computation as for the hadronic case leads to the constraint: (with no definite value for t^-, contrary to the hadronic case).

$$\boxed{t^- \leq |\vec{\tau}| \leq t^+}, \tag{6.24}$$

and

$$\boxed{t^{\pm} = \frac{(m_\tau^2 + m_\ell^2)|\vec{\ell}| \pm E_\ell(m_\tau^2 - m_\ell^2)}{2m_\ell^2}}. \tag{6.25}$$

Again, in the ultra-relativistic regime, Eq. (6.24) is equivalent to the constraint:

$$\boxed{E_\ell \leq E_\tau \leq \frac{m_\tau^2}{m_\ell^2} E_\ell}. \tag{6.26}$$

The leptonic τ transfer function is directly connected to the differential decay width describing the τ disintegration:

$$\frac{d^3\Gamma}{d^3\vec{\ell}} \propto \frac{1}{E_\tau \hat{E}_l} \left(Q^2(\ell \cdot \tau) + 2(Q \cdot \ell)(Q \cdot \tau)\right) \tag{6.27}$$

Where $Q = \ell - \tau$. It is possible, therefore, to validate the leptonic transfer function equation verifying the corresponding differential decay width expression. The decay rate for the process $\tau \to \ell\nu\bar{\nu}$ is function of two variables: the energy of the lepton in the final state and the cosine of the angle between the lepton momentum and the τ momentum one. The transfer function cannot however be validated under the form presented in Eq. 6.27. Indeed, a very high precision would be needed to take the distribution in $\cos\theta_{\tau\ell}$ into account: the cosine is very close to 1 and a very high precision on the energy, momentum and mass estimation is required in order to get the right value for $\cos\theta_{\tau\ell}$. Therefore the strategy adopted to validate the leptonic transfer function consists in:

1. Validating the bidimensional distribution of Eq. 6.27 in the τ center-of-mass frame (angles are less boosted);

2. Validating the $d\Gamma/dz$ distribution in the laboratory frame, where $z = E_\ell/E_\tau$. To do so, some calculations are needed to evaluate the theoretical formula to be compared to the z distribution obtained from the MC. This calculation has been carried out in the context of the collinear approximation and it is shown in the following.

As far as approach 1) is concerned, Eq. 6.27 in the τ center-of-mass frame reduces to:

$$\frac{d\Gamma}{d^3\ell} \propto \left[\left(m_\tau^2 + m_\ell^2 - 2m_\tau \hat{E}_\ell\right) + 2m_\tau \left(m_\tau - \hat{E}_\ell\right)\left(m_{tau}\hat{E}_\ell - m_\ell^\ell\right)\right],$$

$$\frac{d\Gamma}{d^3\ell} \propto \vec{\tau} E_\tau \left[3m_\tau^2 - 4m_\tau \hat{E}_\ell\right]$$

It is thanks to the fact that the τ leptonic transfer function is proportional to the $\tau \to \mu\nu(s)$ differential decay width that it is possible to easily check the behaviour of the 2D transfer function at least in the τ rest frame. As can be seen, it does not depend on $cos\theta_{\tau\ell}$ and therefore the correspondent angular part is expected to be flat. Considering the 2D distribution in Fig. 6.23, where a comparison has been made between the $d\Gamma/(dEd\cos\theta)$ evaluated in the τ rest frame obtained by the theoretical equation (Fig. 6.23-Left) and the one obtained from the simulation (Fig. 6.23-Right), it is clear that the angular dependence is flat, as expected. Moreover there is a very good agreement between theoretical prediction and simulation, as can be seen in Fig. 6.23. The energy-profile of the distribution follows, in both the case, the expected behaviour for a 3-body decay involving 2 neutrinos in the mother particle rest frame.

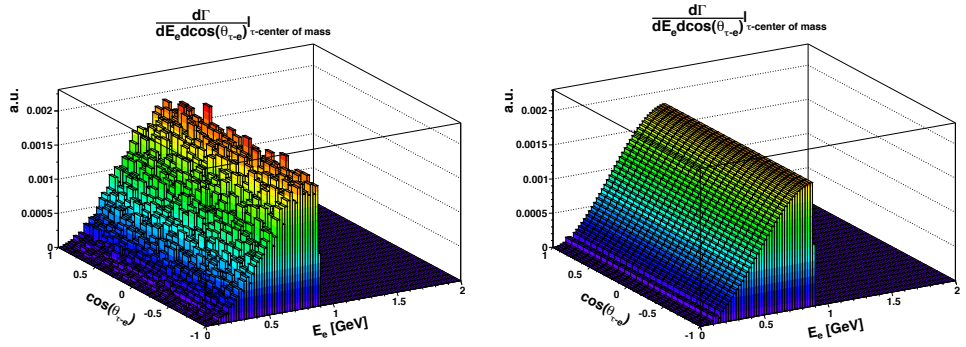


Figure 6.23: *2D distribution for the leptonic decay width obtained using MC events (Left) and as it is modelled by the theoretical formula of Eq. 6.27.*

Practically, even though the transfer functions for the τ decay could be evaluated in the τ rest frame, it is more convenient to evaluate it in the laboratory frame, following the approach 2).

Collinear approximation

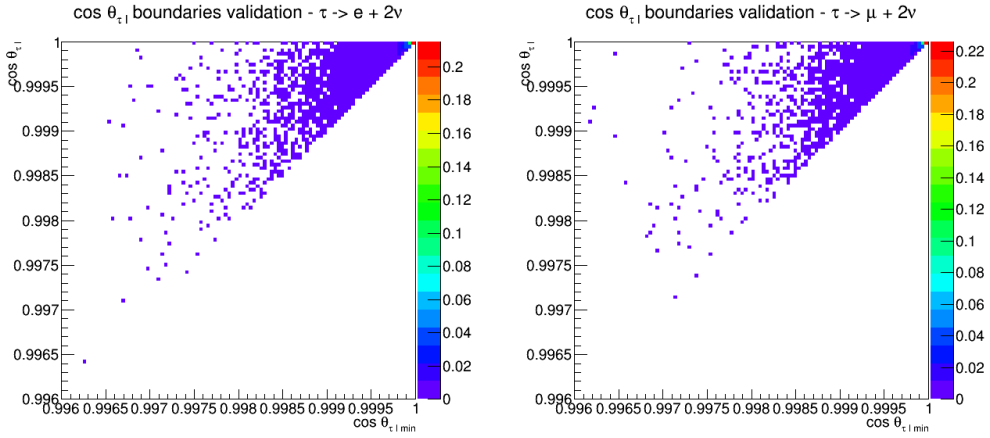


Figure 6.24: *Angular distributions for the leptonic τ decay in the electron (Left) and muon (Right) decay channels. The distributions are obtained from generator level in a VBF sample.* $\cos\theta_{\tau l\,min}$ *is defined by Eq. (6.23).*

As can be seen in Fig. 6.24, the direction of the charged lepton coming from the leptonic τ decay is extremely close to the direction of the τ. Therefore the angular transfer function could be replaced by a δ function, which enforces that the direction of the τ lepton is directly given by the measured direction of the charged lepton.

$$\boxed{T_{l\theta}(l|\tau) = \delta(\vec{e}_\tau - \hat{\vec{e}}_l)} \tag{6.28}$$

This offers the advantage to remove two integration variables and the computation time required for the MEM weight can then be considerably reduced. The only remaining part in the lepton transfer function is therefore the energy transfer function.

For $m_l = 0$, we have

$$\frac{d\Gamma}{dl^3}(\tau \to l\nu\bar{\nu}) \propto \left(1 - \frac{|\vec{\tau}|}{E_\tau}\cos\theta_{\tau l}\right)\left(3m_\tau^2 - 4E_l E_\tau\left(1 - \frac{|\vec{\tau}|}{E_\tau}\cos\theta_{\tau l}\right)\right)$$

$$= 3m_\tau^2 - 4E_l E_\tau + \left(8E_l|\vec{\tau}| - 3m_\tau^2\frac{|\vec{\tau}|}{E_\tau}\right)\cos\theta_{\tau l} - 4\frac{E_l}{E_\tau}|\vec{\tau}|^2\cos^2\theta_{\tau l}$$

We have:

$$\alpha \equiv \frac{2E_\tau E_l - m_\tau^2}{2|\vec{\tau}|E_l} \leq \cos\theta_{\tau l} \leq 1$$

Therefore,

$$\frac{d\Gamma}{dE_l} \propto \int_{\alpha}^{1} d(\cos\theta_{\tau l}) E_l^2 \frac{d^3\Gamma}{dl^3}$$

$$= (1-\alpha)E_l^2 \left(3m_\tau^2 - 4E_l E_\tau + \frac{1}{2}\left(8E_l|\vec{\tau}| - 3m_\tau^2\frac{|\vec{\tau}|}{E_\tau}\right)(1+\alpha)\right.$$

$$\left. -\frac{4}{3}\frac{E_l}{E_\tau}|\vec{\tau}|^2(1+\alpha+\alpha^2)\right)$$

$$= \frac{m_\tau^2 - 2E_l(E_\tau - |\vec{\tau}|)}{2E_l|\vec{\tau}|}E_l^2\left(3m_\tau^2 - 4E_lE_\tau + 4E_l|\vec{\tau}| - \frac{3}{2}m_\tau^2\frac{|\vec{\tau}|}{E_\tau} + 4E_\tau E_l - 2m_\tau^2 - \frac{3}{2}m_\tau^2 + \frac{3}{4}\frac{m_\tau^4}{E_\tau E_l}\right.$$

$$\left. -\frac{4}{3}E_lE_\tau + \frac{4}{3}\frac{E_l}{E_\tau}m_\tau^2 - \frac{4}{3}|\vec{\tau}|E_l + \frac{2}{3}m_\tau^2\frac{|\vec{\tau}|}{E_\tau} - \frac{4}{3}E_\tau E_l + \frac{4}{3}m_\tau^2 - \frac{1}{3}\frac{m_\tau^4}{E_\tau E_l}\right)$$

$$= \frac{m_\tau^2 - 2E_l(E_\tau - |\vec{\tau}|)}{2E_l|\vec{\tau}|}E_l^2\left(\frac{5}{6}m_\tau^2 - \frac{8}{3}E_\tau E_l + \frac{8}{3}E_l|\vec{\tau}| - \frac{5}{6}m_\tau^2\frac{|\vec{\tau}|}{E_\tau} + \frac{5}{12}\frac{m_\tau^4}{E_\tau E_l} + \frac{4}{3}\frac{E_l}{E_\tau}m_\tau^2\right)$$

$$= \frac{m_\tau^2 - 2E_l(E_\tau - |\vec{\tau}|)}{2E_\tau|\vec{\tau}|}\left(\frac{5}{6}m_\tau^2E_l(E_\tau - |\vec{\tau}|) - \frac{8}{3}E_l^2E_\tau(E_\tau - |\vec{\tau}|) + \frac{5}{12}m_\tau^4 + \frac{4}{3}m_\tau^2E_l^2\right).$$

For $m_\tau \ll E_\tau$:

$$E_\tau - |\vec{\tau}| = E_\tau - E_\tau\left(1 - \frac{m_\tau^2}{E_\tau^2}\right)^{1/2} = E_\tau - E_\tau\left(1 - \frac{1}{2}\frac{m_\tau^2}{E_\tau^2} - \frac{1}{8}\frac{m_\tau^4}{E_\tau^4} + o\left(\frac{m_\tau^4}{E_\tau^4}\right)\right)$$

$$= \frac{1}{2}\frac{m_\tau^2}{E_\tau} + \frac{1}{8}\frac{m_\tau^4}{E_\tau^3} + o\left(m_\tau^4\right).$$

We have then:

$$\frac{d\Gamma}{dE_l} \propto \frac{m_\tau^2}{2E_\tau^2}\left(1 - \frac{E_l}{E_\tau}\right)\left(\frac{5}{12}m_\tau^4\frac{E_l}{E_\tau} - \frac{4}{3}m_\tau^2E_l^2 - \frac{1}{3}m_\tau^4\frac{E_l^2}{E_\tau^2} + \frac{5}{12}m_\tau^4 + \frac{4}{3}m_\tau^2E_l^2 + o(m_\tau^4)\right)$$

$$\boxed{T_{lE}(l|\tau) \propto \frac{1}{E_\tau^2}(1-z)\left(5 + 5z - 4z^2\right)}, \tag{6.29}$$

with $z = E_l/E_\tau$. To normalize it we write it as:

$$T_{lE}(l|\tau) = \frac{\alpha(\tau)}{E_\tau^2}(1-z)\left(5 + 5z - 4z^2\right).$$

The normalization is then imposed from:

$$\int_0^{E_\tau} dE_l T_{lE}(l|\tau) = \frac{\alpha(\tau)}{E_\tau} \int_0^1 dz\,(1-z)\left(5+5z-4z^2\right) = 3\frac{\alpha(\tau)}{E_\tau}.$$

Therefore $\alpha(\tau) = E_\tau/3$ and we have:

$$\boxed{T_{lE}(l|\tau) = \frac{1}{3E_\tau}\,(1-z)\left(5+5z-4z^2\right)} \tag{6.30}$$

To summarize, in the collinear approximation, the overall transfer function for the leptonic τ allows to discard the two angular variables, which only leaves the integration over the τ momentum, resulting in:

$$\boxed{T_{lE}(\ell|\tau)\delta(\vec{e}_\tau - \hat{\vec{e}}_\ell)\delta(\hat{\vec{\ell}} - \vec{\ell})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \frac{1}{3}\,(1-z)\left(5+5z-4z^2\right)\frac{|\vec{\tau}|^2}{E_\tau^2}d|\vec{\tau}| \equiv \tilde{T}_l(\ell|\tau)d|\vec{\tau}|} \tag{6.31}$$

Even if all the information related to the shape of the τ leptonic transfer function are included in the Eq. 6.22, it is worth to perform the calculation in order to obtain Eq. 6.30. Indeed, Eq, 6.22 is written in a covariant form, coming directly from the application of the Feynman diagrams rules, in other words it represents all our knowledge regarding the τ leptonic decays in the context of the SM condensed in one equation. The equation, as it is written in its covariant form, it is not trivial to be validated. For this reason it was so important getting Eq. 6.30 factorizing the angular and energetic part in order to validate it explicitly and get rid of 2 integration variables. The first validation step is to prove that the expression found theoretically, Eq. 6.30, reproduces the fractional energy distribution for leptonic τ from the MC sample. In Fig. 6.25 the plots obtained for the z distribution ($z = E_\ell/E_\tau$) when $\ell = \mu$ and $\ell = e$ are shown together with the theoretical shape. In this case, only a check of the z dependence is performed, thus all the distributions are normalized to unity.

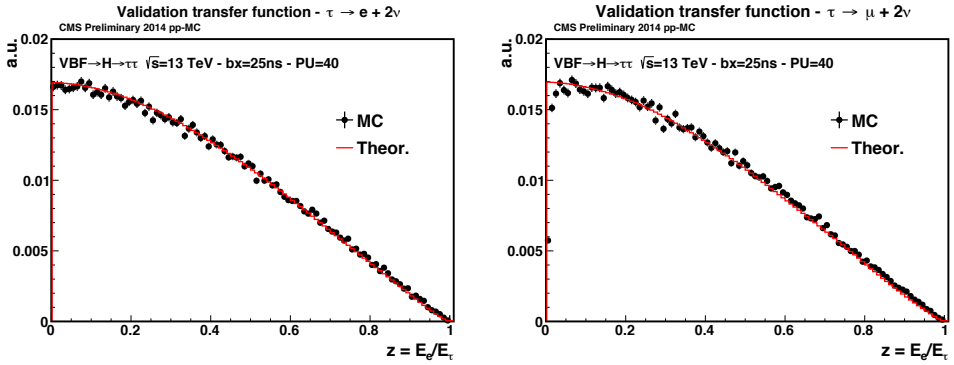


Figure 6.25: *Validation of the energy transfer function for the leptonic τ. The z distribution obtained from Eq 6.31 (black dots) is superimposed to the one obtained from the simulation.*

An almost perfect agreement is found. A small mismatch between the theoretical formulation and the simulation is present in the low z regime. The reason lies in the fact that, in the theoretical computation, the assumption of $m_\ell = 0$ has been adopted. This explains why the difference is more visible in the muonic decay mode. The result obtained is in agreement with the one reported in [75]. The normalization coefficient of Eq. 6.31 can be computed and it turns out to be simply $1/3E_\tau$.

To validate the transfer function including its normalization factor, the muon p_T spectrum for small bins of τ p_T is drawn. Then, the transfer function is plotted on top of it, taking into account the bin-width in the normalization. Figures 6.26, 6.27 and 6.28 show a good agreement between the real and predicted spectrum, both in terms of shape and normalization.

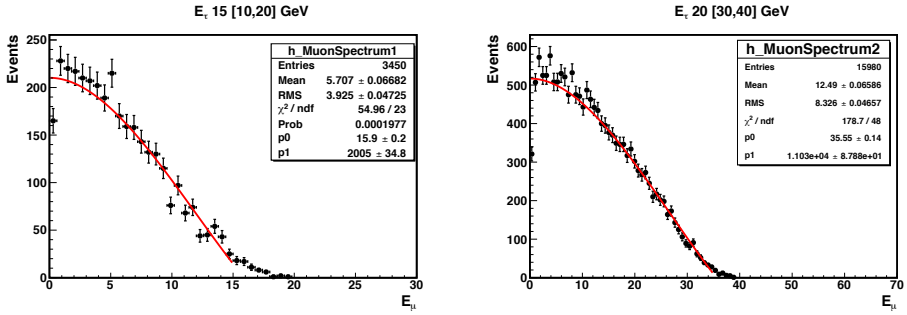


Figure 6.26: *Validation of the energy transfer function for the leptonic τ. The muon spectrum obtained from Eq 6.31 (black dots) is superimposed to the one obtained from the simulation. The μ energy is plotted in different E_τ bins: 15 GeV (Left) and 35 (Right).*

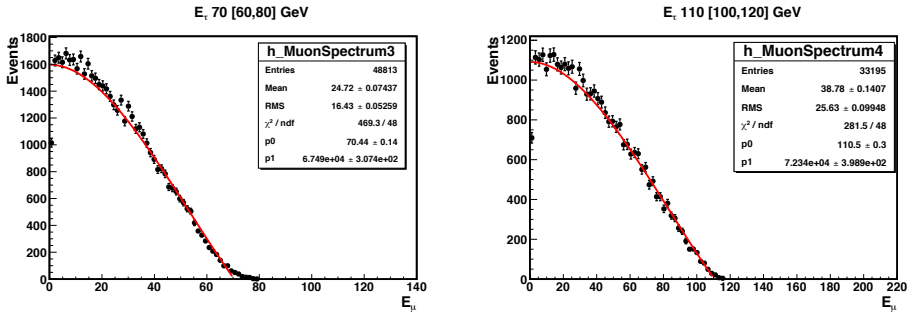


Figure 6.27: *Validation of the energy transfer function for the leptonic τ. The muon spectrum obtained from Eq 6.31 (black dots) is superimposed to the one obtained from the simulation. The μ energy is plotted in different E_τ bins: 70 GeV (Left) and 110 (Right).*

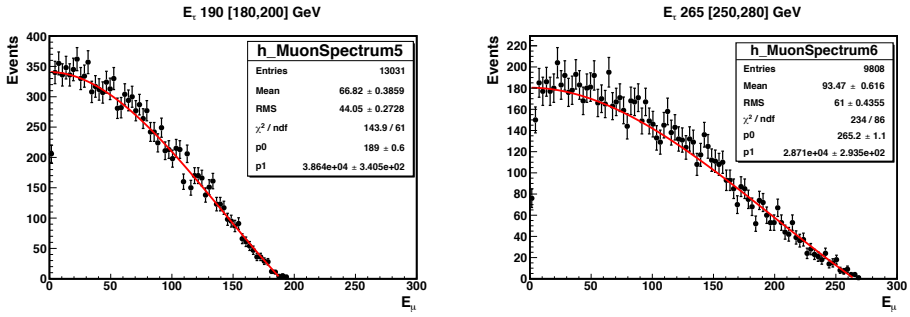


Figure 6.28: *Validation of the energy transfer function for the leptonic τ. The muon spectrum obtained from Eq 6.31 (black dots) is superimposed to the one obtained from the simulation. The μ energy is plotted in different E_τ bins: 190 GeV (Left) and 265 (Right).*

6.6.4 $\vec{E}_T^{miss}$ transfer function

So far, the TF that have been described had a direct correspondence with one of the integration variable: energy of a quark or of a τ. In that respect, the E_T^{miss} transfer function is peculiar as it does not have a direct corresponding integration variable. As a matter of fact, its origin lies in the δ^4 energy-momentum conservation, reduced in this case to the transverse momentum conservation. The $\vec{E}_T^{miss}$ transfer function, nevertheless, plays a crucial role in the integrand as there are several neutrinos in the final state resulting from the τ leptons decays and it quantifies the compatibility between the predicted E_T^{miss} and the measured one. The recoil transfer function is parametrized in CMS by a two-dimensional Gaussian:

$$\boxed{T_{E_T}(\hat{\vec{\rho}}_T|\vec{P}_T) = \frac{1}{2\pi\sqrt{|V|}}\exp(-\frac{1}{2}(\hat{\vec{\rho}}_T - \vec{P}_T)^T V^{-1}(\hat{\vec{\rho}}_T - \vec{P}_T))} \tag{6.32}$$

where $\vec{\rho}_T$ is defined as $\vec{\rho}_T = \vec{E}_T^{miss} + \sum_{l,\pi}\vec{p}_T + \sum_{jets}\vec{p}_T$. It corresponds to the opposite of the recoil, i.e. the transverse component of the vectorial sum of all the particles not clustered in jets or

identified as leptons, but that are, of course, included in the $\vec{E}_T^{miss}$ computation. In other words, it also corresponds to the overall boost of the jet+Higgs boson system. The ρ_T is then compared with the simulated transverse boost of the jets+Higgs boson system. The $\vec{P}_T = \sum_{\tau} \vec{p}_T + \sum_{q} \vec{p}_T$ vector is the overall boost associated to the point in phase space considered for the integration. If the jets and the visible part of the τs were measured perfectly, the $(\rho_T - P_T)$ vector would correspond to the difference between the measured $\vec{E}_T^{miss}$ and the transverse component of the momentum of all the neutrinos. Finally, V is the $\vec{E}_T^{miss}$ covariance matrix. V is a 2×2 matrix and when it is diagonal it takes the form:

$$\begin{pmatrix} \sigma^2_{E_T} & 0 \\ 0 & E_T^2\sigma^2_{\phi} \end{pmatrix}$$

where σ_{E_T} is the resolution on the magnitude of the E_T^{miss} and σ^2_{ϕ} is the resolution of the E_T^{miss} along the ϕ direction. A non-diagonal form (x-y) representation is used in the present case.

The recoil transfer function is of capital importance in the full integral evaluation procedure. It is up to this TF to regularise very boosted configurations. Indeed, in the numerical integration process, extremely boosted configurations can be picked-up. Due to the change of frame described earlier, the total boost of the jets+Higgs boson system does not intervene in the ME computation and such boosted configurations can therefore have large ME contribution. Such boosted topologies however also result in large predicted E_T^{miss} that, in most of the cases, are not compatible with the measured $\vec{E}_T^{miss}$ vector.

The validation of the $\vec{E}_T^{miss}$ transfer function thus proceeded in several step:

- cross-check the different parts of the expressions for $\hat{\vec{\rho}}_T$ and $\vec{P}_T^{tot}$ vectors
- characterize the behaviour and the response of the transfer function

The validation of the reconstructed $\vec{E}_T^{miss}$ vector is carried out in the VBF $H \to \tau\tau$ Monte Carlo generated events at $\sqrt{s} = 8\,TeV$. In a first step, it is checked that the $\vec{E}_T^{miss}$ reconstructed in these events behaves as expected. In these studies the particle flow $\vec{E}_T^{miss}$ has been used, corrected for the *Type-I* correction. This correction is a propagation of the JEC to the missing transverse energy. The *Type-I* correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied. As already done for the Jet transfer function study, the corrections applied to the Jets are the full *L1L2L3* chain.

$\vec{E}_T^{miss}$ response and resolution

A convenient method to study the performance of the $\vec{E}_T^{miss}$ consists in decomposing the $\vec{E}_T^{miss}$ vector on two orthogonal directions, for example the x and y directions. It is however even better to decompose the MET on the axis parallel to the Higgs boson p_T vector and the corresponding

perpendicular axis. Indeed, the response of the detector to the recoil against the boson could differ from the response to other particles, e.g. from jets. The decomposition of the recoMET-GenMet on the two axes is presented in Fig. 6.29 where it can be seen that the response is indeed slightly different. On the parallel axis, the mean value is $1\,GeV$ and the resolution is slightly worse than on the perpendicular direction. The results obtained are shown in Fig.6.29 and in Fig. 6.30.

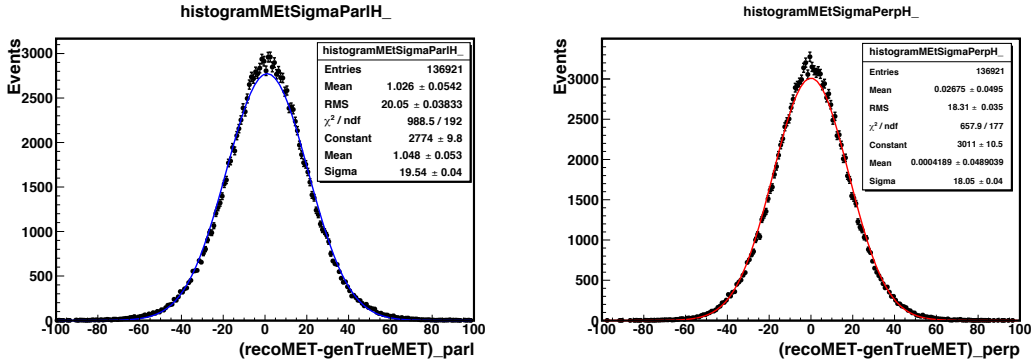


Figure 6.29: $\vec{E}_T^{miss}$ *response plots. The difference (recoMET - genTrueMET) is projected in the parallel (Left) and perpendicular (Right) directions with respect to the Higgs* $\vec{p}_T$.

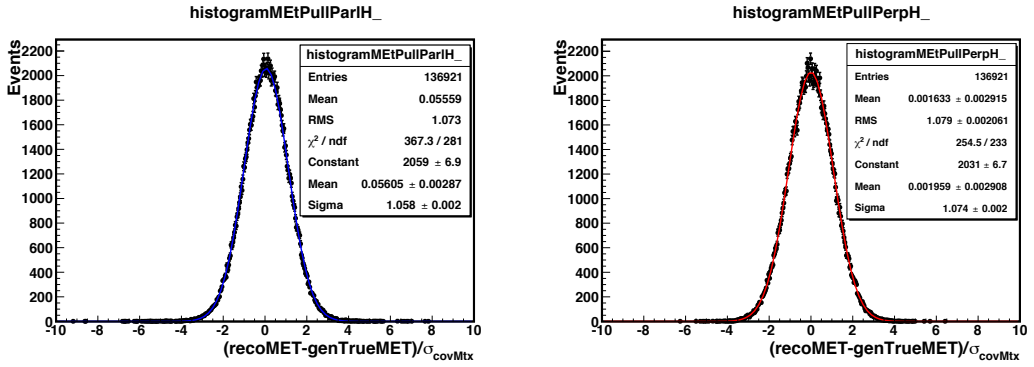


Figure 6.30: $\vec{E}_T^{miss}$ *pull distribution plots. The difference (recoMET - genTrueMET) is divided by the* σ_{covMtx} *and projected in the parallel (Left) and perpendicular (Right) direction with respect to the Higgs* $\vec{p}_T$.

The V covariance matrix can be used to compute the expected σ for the two considered directions. The pull distributions thus obtained are presented in Fig. 6.30 together with the Gaussian fits and their parameters. The results are satisfying and match the expectations.

Reconstruction of the neutrinos system

Now that the *reconstructed* E_T^{miss} has been checked, the *predicted* E_T^{miss} counterpart should be studied as well. As alluded to earlier, for each shot within the numerical integration, the neutrino momentum is deduced making the difference between $\sum_{vegas} \vec{\tau}$ and $(\vec{\tau}_{vis} + \vec{\mu}_{reco})$ (from here after called *reconstructed neutrinos*). To evaluate whether this indirect neutrino reconstruction is valid or not, the following check has been made in the same simulation as before. For this check, since we are obviously not in the context of a numerical integration: $\sum_{vegas} \tau$ correspond to the true τ momenta taken at generator level; the distributions comparing the magnitude of computed neutrino momentum and the true neutrino momentum is shown in Fig. 6.31. A satisfying agreement can be observed. The small differences in the shape are attributed to the imperfect energy scale of the reconstructed visible tau energy.

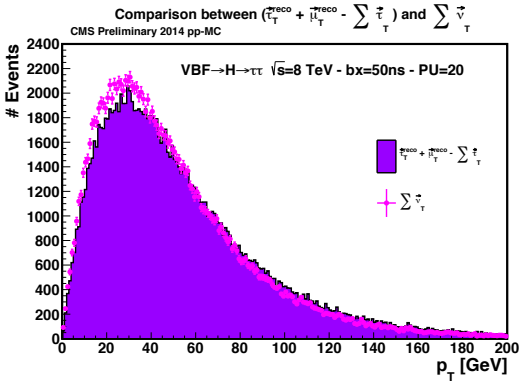


Figure 6.31: *Inclusive distribution comparison between the transverse component of reconstructed neutrino system with respect to the MC-truth one.*

$\vec{E}_T^{miss}$ transfer function validation

Finally the transfer function shape validation proceeds through a scatter plot showing the weight evaluated by the expression $\exp(-\frac{1}{2}(\hat{\vec{\rho}}_T - \vec{P}_T)^T V^{-1}(\hat{\vec{\rho}}_T - \vec{P}_T))$ as function of the vectorial difference $(\hat{\vec{\rho}}_T - \vec{P}_T^{tot})$ projected in the x and y direction. Obviously, from the formula, the maximum value of the weight corresponds to $(\hat{\vec{\rho}}_T - \vec{P}_T^{tot}) = \vec{0}$. This means that when the global boost chosen by Vegas corresponds to the measured one, the transfer function returns the higher possible response, as can be seen in Fig. 6.32 that is simply showing a 2D-Gaussian function.

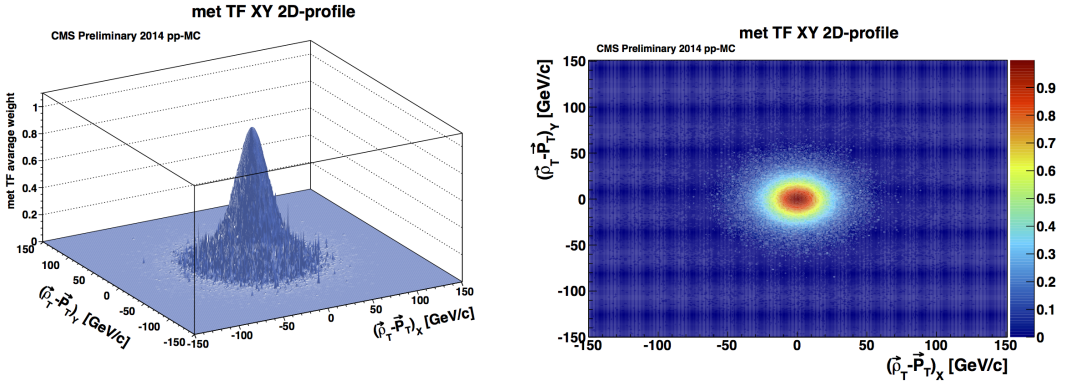


Figure 6.32: *Shape of the missing transverse energy transfer function with the X, Y values of the vectorial difference between* $\hat{\vec{p}}_T - \vec{P}_T^{tot}$. *A 3D plot is shown (Left) with its projection of the z-axis (Right).*

In order to directly determine the response of the $\vec{E}_T^{miss}$ transfer function to the reconstructed τ vectors, a simulation has been performed scaling both the τ vectors magnitude. In this way it is possible to produce a 3-dimensional plot between the 2 independent scale factors applied to the τ vectors and the corresponding transfer function value. What emerges from Fig.6.33 is that the maximum of the transfer function is positioned in correspondence of the true τ vector magnitude, as expected. In addition, a tau momentum 20% smaller than in reality results in about 30% lower TF.

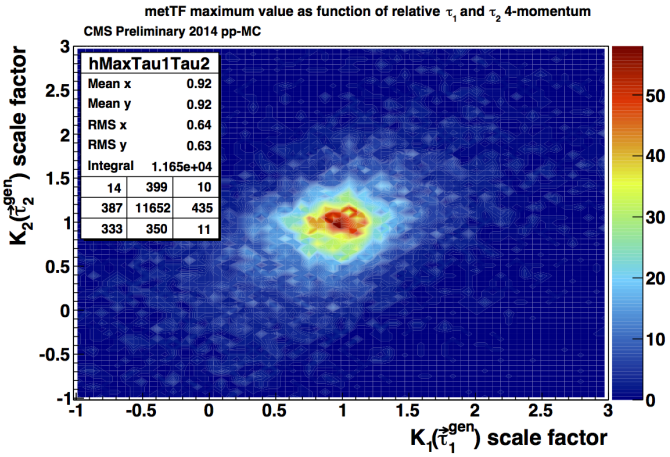


Figure 6.33: $\vec{E}_T^{miss}$ *TF value as function of two scale factor applied on the magnitudes of the true* τ *vectors momenta.*

6.7 Integration over $m^2_{\tau\bar{\tau}}$

Given the narrow width of the Higgs boson, the numerical Monte Carlo integration would be very much, if not completely, inefficient. Indeed, if the momenta of the two τ leptons were to be

shot randomly, the probability that their invariant mass corresponds to the Higgs boson mass within its width would be infinitesimal. As a result, a change of variable is introduced in the following section to deal with the resonant spectrum in the di-τ invariant mass $m_{\tau\bar{\tau}}$, constrained by:

$$\boxed{m^2_{\tau\bar{\tau}} = (\tau + \bar{\tau})^2 = 2m^2_\tau + 2(E_\tau E_{\bar{\tau}} - |\vec{\tau}||\vec{\bar{\tau}}|\cos\theta_{\tau\bar{\tau}})} \tag{6.33}$$

where $\theta_{\tau\bar{\tau}}$ is the angle between between the two τ lepton momenta. This change of variable should account for the constraints associated to the transfer functions of the τ leptons. As a remainder, we assume here that the τ decays leptonically and the $\bar{\tau}$ decays hadronically. The $\tau_h\bar{\tau}_l$ can of course be treated in the same way by exchanging the momenta of the two particles. We treat all the hadronic decay modes, therefore $\tilde{T}_h(\pi|\bar{\tau})$ can be $\tilde{T}_{h1}(\pi|\bar{\tau})$, $\tilde{T}_{h1+\pi^0 s}(\pi|\bar{\tau})$ or $\tilde{T}_{h3}(\pi|\bar{\tau})$ where π represents the sum of all the visible decay products from the hadronic $\bar{\tau}$. Moreover, we assume that the τ and the charged lepton ℓ are collinear. At this stage, the integration variables are therefore $d|\vec{\tau}|d|\vec{\bar{\tau}}|d\phi_{\bar{\tau}\pi}$ and a possible change of variables would be $d\phi_{\bar{\tau}\pi} \to dm^2_{\tau\bar{\tau}}$. We have to evaluate:

$$\tilde{T}_l(\ell|\tau)\tilde{T}_h(\pi|\bar{\tau})d|\vec{\tau}|d|\vec{\bar{\tau}}|d\phi_{\bar{\tau}\pi}$$
$$= \tilde{T}_l(\ell|\tau)\tilde{T}_h(\pi|\bar{\tau})d|\vec{\tau}|d|\vec{\bar{\tau}}|d\phi_{\bar{\tau}\pi}d\cos\theta_{\bar{\tau}\pi}\delta\left(\cos\theta_{\bar{\tau}\pi} - \frac{2E_{\bar{\tau}}E_\pi - (m^2_\tau + m^2_\pi)}{2|\vec{\bar{\tau}}||\vec{\pi}|}\right)$$

and we have:

$$t^- \leq |\vec{\tau} \leq |t^+ \text{ and } |s^-| \leq |\vec{\bar{\tau}}| \leq s^+,$$

with:

$$t^\pm = \frac{(m^2_\tau + m^2_\ell)|\vec{\ell}| \pm E_\ell(m^2_\tau - m^2_\ell)}{2m^2_\ell}$$
$$s^\pm = \frac{(m^2_\tau + m^2_\pi)|\vec{\pi}| \pm E_\pi(m^2_\tau - m^2_\pi)}{2m^2_\pi}.$$

Those constraints come from the condition $-1 \leq \cos\theta_{\bar{\tau}\pi} \leq 1$ and $-1 \leq \cos\theta_{\tau\ell} \leq 1$. To take the simultaneous constraints on $\cos\theta_{\tau\bar{\tau}}$ from Eq. (6.23) and on $\cos\theta_{\bar{\tau}\pi}$ into account it is convenient to introduce the change of variables $d\phi_{\bar{\tau}\pi} \to d(\cos\theta_{\tau\bar{\tau}}) \to dm^2_{\tau\bar{\tau}}$. For that, we have to determine $\vec{\bar{\tau}}$ as a function of $(|\vec{\bar{\tau}}|, \cos\theta_{\bar{\tau}\pi}, \cos\theta_{\tau\bar{\tau}})$. Let's introduce a Cartesian system of coordinates $(\vec{e}_\pi, \vec{e}_x, \vec{e}_y)$, with $\vec{e}_\pi$ the direction of π, $\vec{e}_x = \frac{1}{\sin\theta_{\tau\pi}}(\vec{e}_\tau - \cos\theta_{\tau\pi}\vec{e}_\pi)$ and $\vec{e}_y = \vec{e}_\pi \wedge \vec{e}_x$. In that frame represented in Fig. 6.34, $\vec{\bar{\tau}}$ is given by:

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\cos\theta_{\bar{\tau}\pi}\vec{e}_\pi + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\vec{e}_x + \sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\vec{e}_y)$$

.

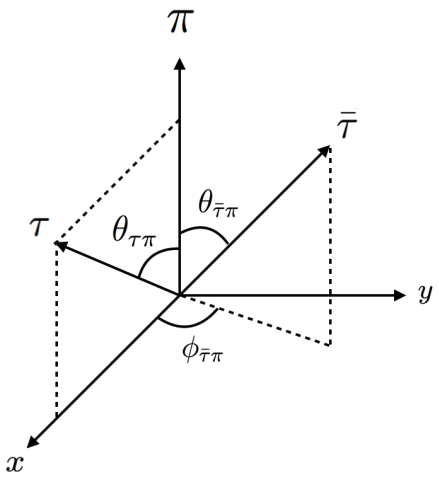


Figure 6.34: *Cartesian coordinate system used to describe a $\tau_l\bar{\tau}_h$ decay.*

One gets then:

$$\cos\theta_{\tau\bar{\tau}} = \cos\theta_{\bar{\tau}\pi}\cos\theta_{\tau\pi} + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi}.$$

Therefore:

$$d\phi_{\bar{\tau}\pi} = \frac{1}{\left|\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}}\right|} d(\cos\theta_{\tau\bar{\tau}}),$$

with:

$$\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} = -\sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi}.$$

Let's now introduce another system of coordinates (non-Cartesian) $(\vec{e}_\pi, \vec{e}_\tau, \vec{e}_y)$. We can then write $\vec{\bar{\tau}}$ in that frame as:

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\alpha\vec{e}_\pi + \beta\vec{e}_\tau + \gamma\vec{e}_y).$$

With that parametrization, $\gamma = \vec{e}_{\bar{\tau}}.\vec{e}_y = \sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}$, one then gets:

$$1 = \alpha^2 + \beta^2 + \gamma^2 + 2\alpha\beta\cos\theta_{\tau\pi},$$

$$\cos\theta_{\bar{\tau}\pi} = \alpha + \beta\cos\theta_{\tau\pi} \ , \ \cos\theta_{\tau\bar{\tau}} = \alpha\cos\theta_{\tau\pi} + \beta.$$

Therefore:

$$\alpha = \frac{\cos\theta_{\bar{\tau}\pi} - \cos\theta_{\tau\pi}\cos\theta_{\tau\bar{\tau}}}{\sin^2\theta_{\tau\pi}} \ , \ \beta = \frac{\cos\theta_{\tau\bar{\tau}} - \cos\theta_{\tau\pi}\cos\theta_{\bar{\tau}\pi}}{\sin^2\theta_{\tau\pi}},$$
$$\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi}.$$

This system has a solution if and only if $\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi} \geq 0$, with α and

β defined by the previous equations. However, this is not always guaranteed. Indeed, it has been searched for a possible three-momentum $\vec{\bar{\tau}}$, which fulfills at the same time constraints on $\cos\theta_{\tau\bar{\tau}}$ and $\cos\theta_{\bar{\tau}\pi}$ (which depend on $|\vec{\bar{\tau}}|$). In geometrical terms, this is equivalent to look for the intersections of two cones which axes are given by $\vec{e}_\tau$ and $\vec{e}_\pi$ and which opening angles are $\theta_{\tau\bar{\tau}}$ and $\theta_{\bar{\tau}\pi}$. Now, the axes being fixed, it is possible that the opening angles are too small for the cones to intersect. This would be the case for instance if the momentum $|\vec{\bar{\tau}}|$ shot is too high, because in that case $\bar{\tau}$ should be at the same time almost collinear with τ and π.

If the constraint $\gamma^2 \geq 0$ is fulfilled, it is possible to define

$$\gamma = \pm\sqrt{1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi}}$$

The so-called "boost" associated to the outgoing particles, defined by $\vec{P}_T = \vec{\tau}_T + \vec{\bar{\tau}}_T + \vec{q}_T + \vec{q'}_T$, is then computed for the two signs. The sign of γ is consequently chosen as the one leading to the smallest boost in magnitude. At leading order, all the outgoing particles are indeed expected to be perfectly balanced in the transverse plane, having thus $P_T = 0$. This procedure to determine the sign of γ has been proven to be very efficient at generator level (see Fig. 6.35).

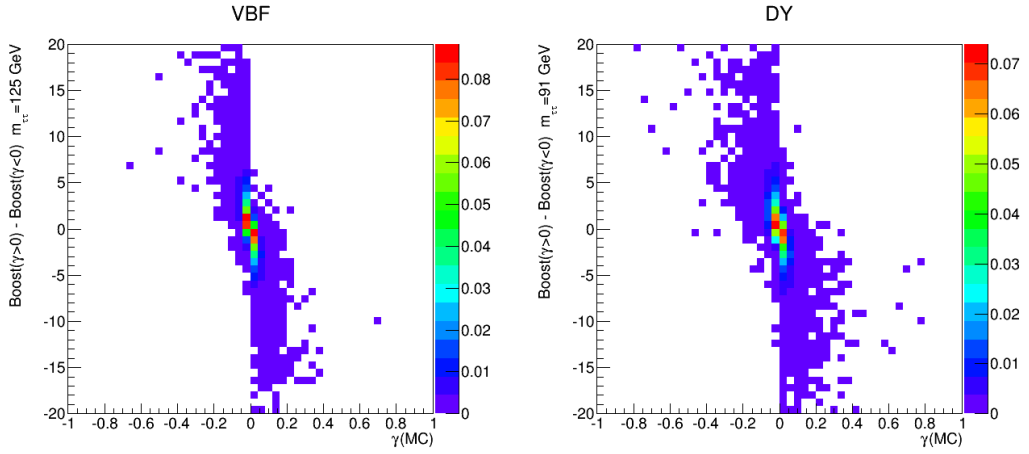


Figure 6.35: *Distribution of $\gamma(MC) = \vec{e}_{\bar{\tau}}.\vec{e}_y/$ and $P_T(\gamma > 0) - P_T(\gamma < 0)$ in leading-order Monte-Carlo samples for VBF (Left) and DY+2 jets (Right). All the quantities have been evaluated using the kinematic reconstruction detailed previously, fixing the integration variables at their generator-level values. A clear correlation can be observed between the sign of γ and the sign of $P_T(\gamma > 0) - P_T(\gamma < 0)$.*

The constraint $\gamma^2 \geq 0$ could in principle be translated on the integration range for one of the integration variable but it may not be possible to do it analytically. Since the region $\gamma^2 \geq 0$ has a non-zero measure in the phase-space, the following strategy has been first tested: if the choice of the point in the phase space is such that the inequality is not fulfilled then this point has a null contribution to the integral. However this solution turns out to be relatively inefficient, since when picking random values of $|\vec{\tau}|$ and $|\vec{\bar{\tau}}|$, most of the time the constraints on the angles

are such that there is no intersection of the two cones of direction $\vec{e}_\tau$ and $\vec{e}_\pi$ with opening angles $\theta_{\tau\bar{\tau}}$ and $\theta_{\bar{\tau}\pi}$. Only for a very small region in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space this constraint is fulfilled, as shown in Fig. 6.36.

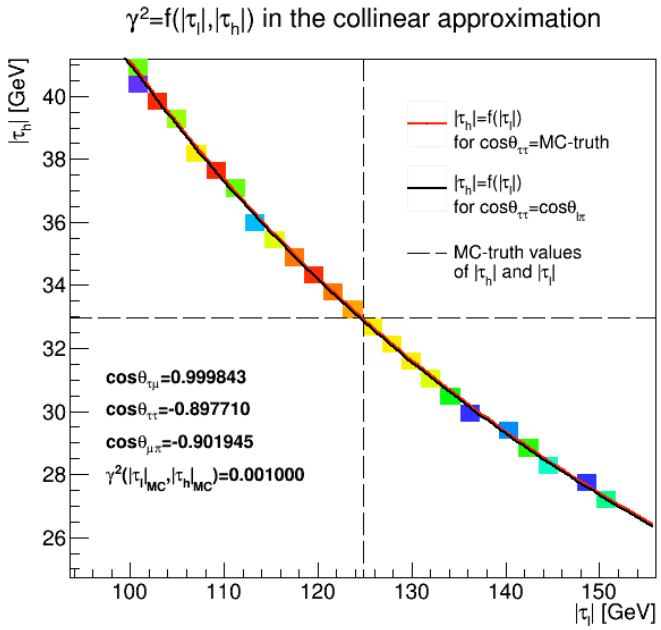


Figure 6.36: *Region where $\gamma^2 > 0$ (colored region) for a single VBF event in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space. The red (black) line corresponds to the function $|\vec{\tau}_h| = f(|\vec{\tau}_l|)$ determined from Eq. (6.33) using $m_{\tau\bar{\tau}}$ =125 GeV and the true value of $\cos\theta_{\tau\bar{\tau}}$ (using $\cos\theta_{\tau\bar{\tau}} = \cos\theta_{\ell\pi}$).*

Therefore, rather than a change of variables $d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \to d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}dm^2_{\tau\bar{\tau}}$, the change of variables $d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \to d\cos\theta_{\bar{\tau}\pi}d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}}$ is introduced. It has the following advantage: since the direction of the τ leptons is close to the direction of their visible decay products, $\cos\theta_{\ell\pi}$ gives a good estimator of $\cos\theta_{\tau\bar{\tau}}$. In that case, it is possible to integrate $\cos\theta_{\tau\bar{\tau}}$ in a small window around $\cos\theta_{\ell\pi}$ such that $\gamma^2 \geq 0$ in most of the integration domain.

To take all of the kinematically allowed configurations into account, the integration domain should cover at least the whole region of phase-space for which $\gamma^2 \geq 0$. Although not solvable analytically, this constraint could be tested prior to the integration process for different values of the integration variables. We can therefore determine in advance what will be the region of the phase space that will contribute to the integral and reduce the integration range consequently. The advantages of the new change of variables is that this region is now more extended in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$ space than previously in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space and we know its approximate location around $\cos\theta_{\tau\bar{\tau}} \approx \cos\theta_{\ell\pi}$, as visible in Fig. 6.37.

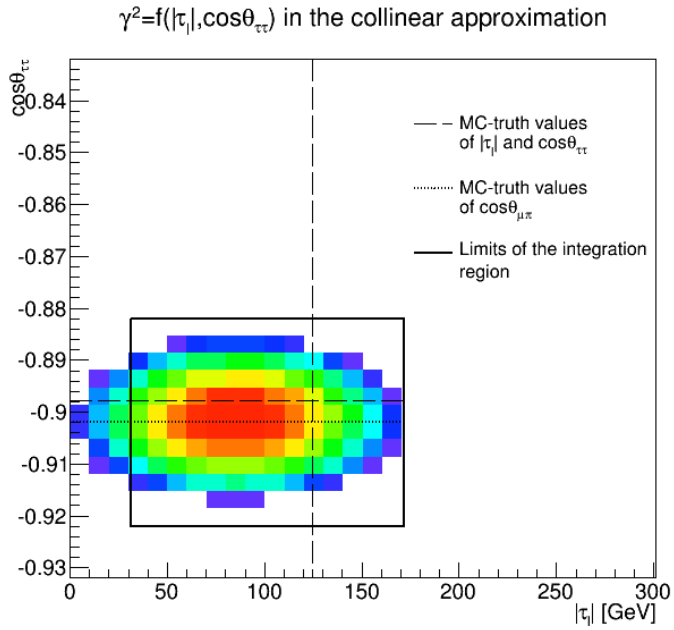


Figure 6.37: *Region where $\gamma^2 > 0$ (colored region) for a single VBF event in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$ space. The black rectangle corresponds to the limits of the integration region, which is determined prior to the integration. The lower limit for $|\vec{\tau}|$ takes into account t^- from Eq. (6.24), which explains why the rectangle does not cover the whole region $\gamma^2 > 0$.*

The change of variables $d|\vec{\tau}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \to d\cos\theta_{\bar{\tau}\pi}d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}}$ introduces the following Jacobian term:

$$d|\vec{\tau}|d\phi_{\bar{\tau}\pi} = \frac{1}{\begin{vmatrix} \frac{\partial m^2_{\tau\bar{\tau}}}{\partial|\vec{\tau}|} & \frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial|\vec{\tau}|} \\ \frac{\partial m^2_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} & \frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} \end{vmatrix}} d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}}$$

with:

$$\cos\theta_{\tau\bar{\tau}}\left(|\vec{\tau}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi}\right) = \cos\theta_{\bar{\tau}\pi}\cos\theta_{\tau\pi} + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi}$$

$$m^2_{\tau\bar{\tau}}\left(|\vec{\tau}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi}\right) = 2m^2_\tau + 2\left(E_\tau E_{\bar{\tau}} - |\vec{\tau}||\vec{\tau}|\cos\theta_{\tau\bar{\tau}}(|\vec{\tau}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi})\right).$$

One gets then:

$$\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial|\vec{\tau}|} = 0,$$

$$\frac{\partial m^2_{\tau\bar{\tau}}}{\partial|\vec{\tau}|} = 2\left(E_\tau\frac{|\vec{\tau}|}{E_{\bar{\tau}}} - |\vec{\tau}|\cos\theta_{\tau\bar{\tau}}\right),$$

$$\frac{\partial \cos\theta_{\tau\bar{\tau}}}{\partial \phi_{\bar{\tau}\pi}} = -\sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi} = -\gamma\sin\theta_{\tau\pi}.$$

We now have to determine $|\vec{\tilde{\tau}}|$ as a function of $|\vec{\tau}|$, $\cos\theta_{\tau\bar{\tau}}$ and $m^2_{\tau\bar{\tau}}$ from Eq. (6.33). Defining $M^2 = \frac{1}{2}m^2_{\tau\bar{\tau}} - m^2_\tau$, we get

$$\boxed{|\vec{\tilde{\tau}}| = \frac{M^2|\vec{\tau}|\cos\theta_{\tau\bar{\tau}} + E_\tau\sqrt{(M^2)^2 - m^2_\tau\left(m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}\right)}}{m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}}} \tag{6.34}$$

with

$$\boxed{|\vec{\tau}| \leq \frac{m_{\tau\bar{\tau}}}{|\sin\theta_{\tau\bar{\tau}}|}\sqrt{\frac{m^2_{\tau\bar{\tau}}}{4m^2_\tau} - 1} \equiv u^+} \tag{6.35}$$

All the elements necessary to the evaluation of the integrand have therefore been determined and the weight can then be computed using the numerical integration suite.

$$w_i(\mathbf{y}) \propto \int \sum_p \frac{f(x_a,Q)f(x_b,Q)}{x_a x_b s}|\mathcal{M}_\Omega(\mathbf{x})|^2 \cdot T_q(\hat{E}_j|E_q,\eta_q)E_q \cdot T_q(\hat{E}_{j'}|E_{q'},\eta_{q'})E_{q'}$$

$$T_{E_T}(\hat{\vec{p}}_T|\vec{P}_T)\cdot\tilde{T}_l(\ell|\tau)\cdot\tilde{T}_h(\pi|\bar{\tau})\frac{1}{\left|E_\tau\frac{|\vec{\tilde{\tau}}|}{E_{\bar{\tau}}} - |\vec{\tau}|\cos\theta_{\tau\bar{\tau}}\right|}\frac{1}{|\gamma\sin\theta_{\tau\pi}|}d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}}dE_q dE_{q'}$$

The integration over $m^2_{\tau\bar{\tau}}$ is removed by fixing its value. By default, it is chosen as the invariant mass squared of the resonance involved in the process Ω. Yet, it is possible that this value of $m_{\tau\bar{\tau}}$ does not lead to a region $\gamma^2 \geq 0$ compatible with the constraint on t^-. This would correspond to cases where this region lies to the left of t^- in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$-plane and would indicate that the visible decay products of the di-τ pair are not compatible with that value of $m_{\tau\bar{\tau}}$. This happens typically for some VBF events, when testing the Drell-Yan background hypothesis with $m_{\tau\bar{\tau}} = 91\,GeV$. This happens for some background events as well. In order to define a non-zero value of both the signal and background weights for every event, the solution adopted consists in increasing the integration interval in $m_{\tau\tau}$ in the integration over the background matrix element. The boundaries are, in that case, constrained by the visible di-τ invariant mass and an upper limit of $180\,GeV$. The integration over the signal matrix element, instead, still concentrates in a very narrow region around the Higgs mass value. The choise to adopt this solution was driven by the fact that events coming from DY processes present a $m_{\tau\tau}$ distribution that, because of its tails, is spread out in a wide range. On the contrary, events coming from the decay of the Higgs boson show a $m_{\tau\tau}$ distribution peaked in a very narrow region around the Higgs boson mass. The distribution of $m_{\tau\tau}$ for DY and VBF events are illustrated in Fig. 6.38

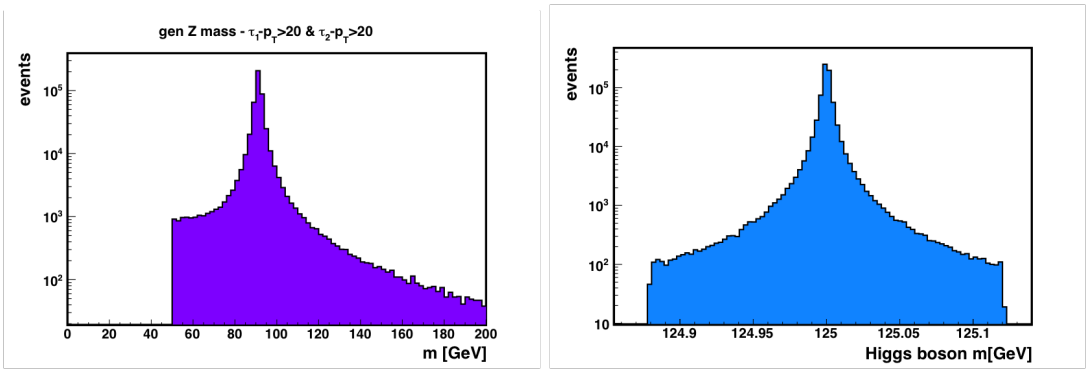


Figure 6.38: *$m_{\tau\tau}$ distribution for DY events (Left) and for VBF events (Right).*

We are then left with a five-dimensional integration over $d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}}dE_q dE_{q'}$. The integration boundaries on $d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}}$ are fixed according to the procedure described in the previous sections, while the integration boundaries on the energy of the quarks are determined as the 95% confidence interval associated to the transfer function of the jets.

6.8 Validation of the τ four-vectors kinematic reconstruction

One of the main difficulties in the implementation of the ME method in the $H \to \tau\tau$ analysis lies in the kinematic reconstruction of the τ four-vectors from the physical quantities chosen to parametrize the phase-space. The reconstruction procedure of the τ four-vectors is described in details in the Section 6.7 of this chapter. In order to check that all the τ four-vector reconstruction procedure was implemented correctly, some physical quantities are plotted superimposed to the Monte-Carlo truth. No integration is carried out to estimate the τ four-vector reconstructed by the ME algorithm but the input variable are fixed to the Monte Carlo truth as well. So what is shown in Fig. 6.39, 6.41 and 6.43 is a consistency check that demonstrates that the calculations put in place to reconstruct the τ four-vectors are correct. Figures 6.40, 6.42 and 6.44 further illustrate the agreement achieved by the kinematic reconstruction in term of scatter plots between the Monte Carlo simulation and the kinematic reconstruction for different physical quantities. A linear fit has been also performed in order to quantify the degree of the correlation. Incidentally, these plots also demonstrate the validity of the collinear approximation.

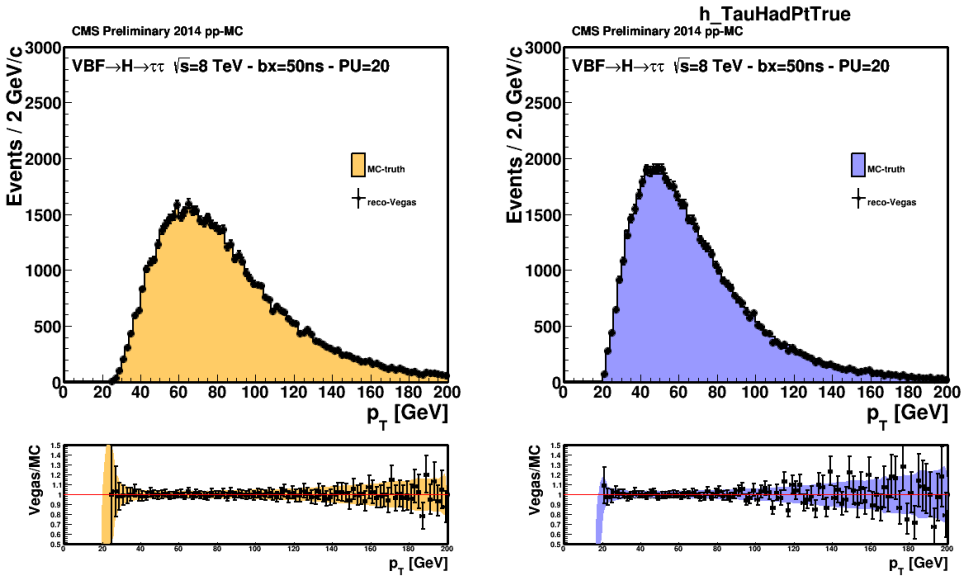


Figure 6.39: *p_T distribution of the leptonic (Left) and hadronic (Right) τ as it is reconstructed by the ME algorithm (black dots) and superimposed to the MC truth (coloured histograms).*

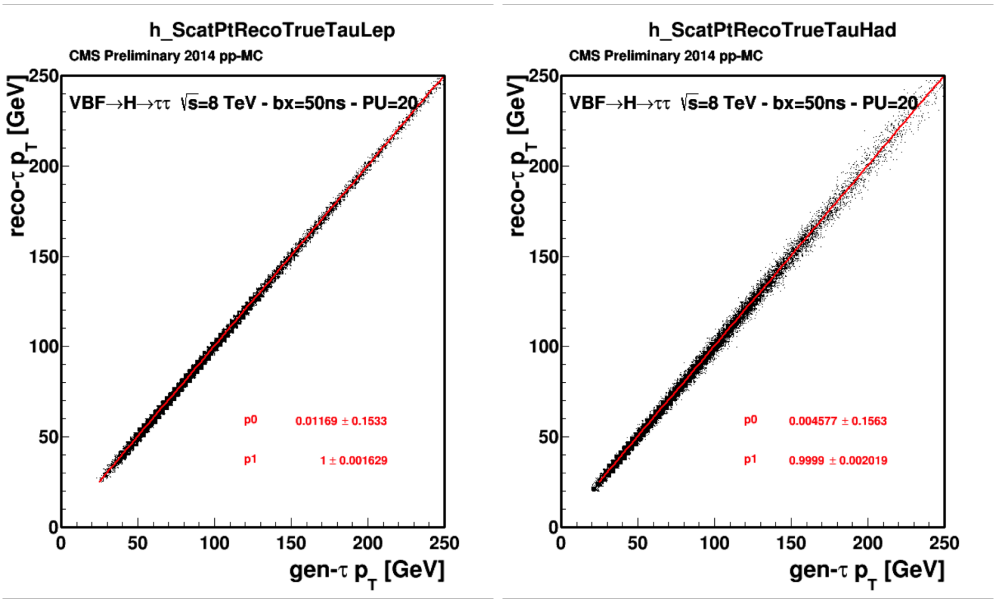


Figure 6.40: *Scatter plot between the p_T reconstructed by the ME algorithm and the MC truth for the leptonic (Left) and hadronic (Right) τ (black dots). A linear fit has been performed and superimposed to the plot (red line) together with the fit parameters.*

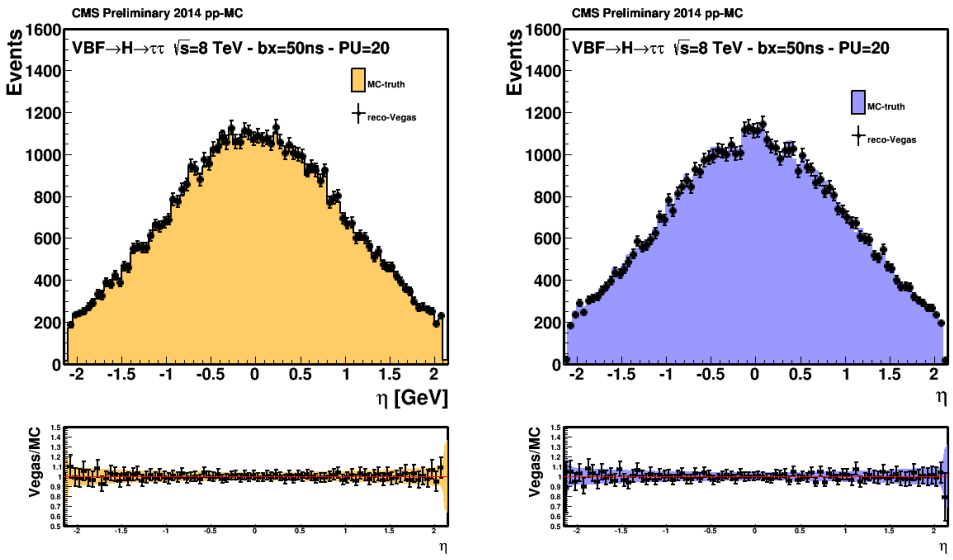


Figure 6.41: *η distribution of the leptonic (Left) and hadronic (Right) τ as it is reconstructed by the ME algorithm (black dots) and superimposed to the MC truth (coloured histograms).*

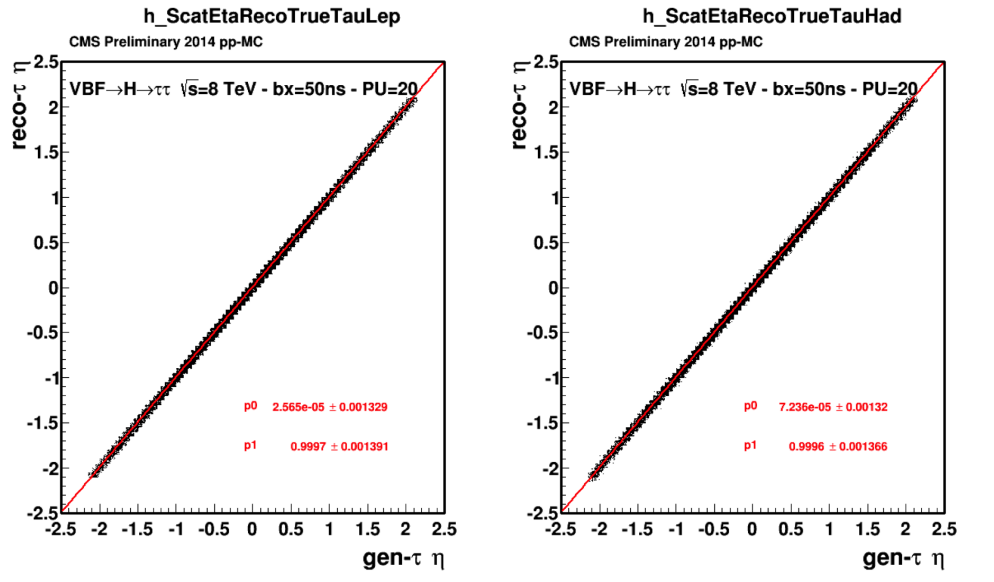


Figure 6.42: *Scatter plot between the η reconstructed by the ME algorithm and the MC truth for the leptonic (Left) and hadronic (Right) τ (black dots). A linear fit has been performed and superimposed to the plot (red line) together with the fit parameters.*

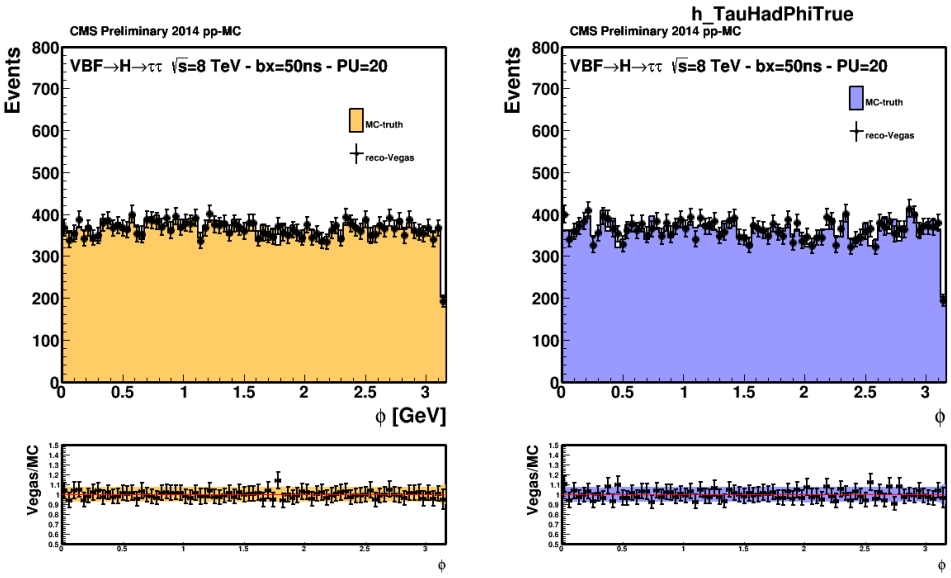


Figure 6.43: *ϕ distribution of the leptonic (Left) and hadronic (Right) τ as it is reconstructed by the ME algorithm (black dots) and superimposed to the MC truth (coloured histograms).*

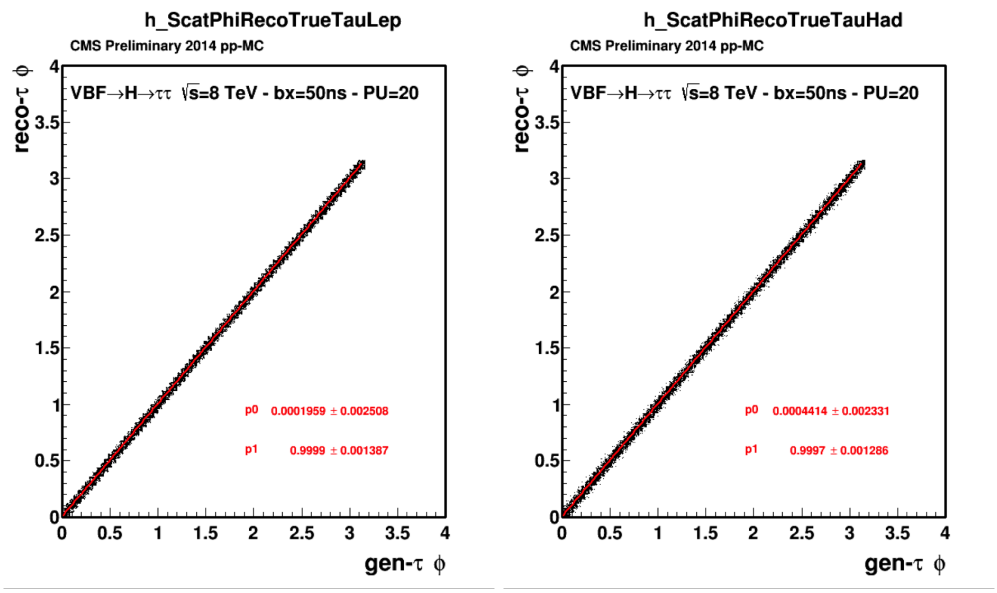


Figure 6.44: *Scatter plot between the ϕ reconstructed by the ME algorithm and the MC truth for the leptonic (Left) and hadronic (Right) τ (black dots). A linear fit has been performed and superimposed to the plot (red line) together with the fit parameters.*

6.8.1 Conclusion

Even if the concept of ME starts to be well known in particle physics, its actual implementation can be tricky. Indeed, the kinematic constraints have to be put by hand, some changes of variables can be needed, and the transfer functions should be carefully validated. In addition, the work on the ME itself is quite specific. In particular, the calculations to perform in order to reconstruct the four-vectors of the particles in the final state in function of the integration variables depend on the specific process considered. The difficulty of this task varies from analysis to analysis and strongly depends on the phase-space of the process considered. However, the principle of MEM is universal and it can be applied to a large variety of analysis for which a precise theoretical model is known a priori. In addition, a strong point in favour of the MEM compared to other multivariate analysis approach, is that MEM does not require any training, and thus does not suffer from lack of statistic in rare processes or due to limitation in the number of simulated events used to model the signal or backgrounds contributions in the analyses.

Chapter 7

Matrix Element Method in the $H \to \tau\tau \to \mu\tau_h$ channel

This Chapter is dedicated to the application of the Matrix Element Method described in Chapter 6 to the search for the SM Higgs boson decaying into a τ-lepton pair in the $\mu\tau_h$ final state. This is the most sensitive channel thanks to the efficient muon reconstruction and identification and high branching ratio of the hadronic τ decay (see Section 3.1.5 and 3.4). Since this work illustrates the development of the MEM in the context of the $H \to \tau\tau$ analysis for the very first time, more details are provided on the intrinsic performance of the method itself, like its discrimination power, tested on VBF $H \to \tau\tau$ and Drell-Yan simulated events. A comparison with the already existing SVfit mass approach (see Section 7.7.2) is also given, showing the superiority of the MEM in reducing the background contamination for the same signal efficiency. Finally, the full analysis results for the $H \to \tau\tau \to \mu\tau_h$ channel are shown as they are obtained both through the SVfit and MEM method, followed by a discussion on their interpretation. As a foreword, the analysis strategy adopted by the CMS Collaboration in the search for the SM Higgs boson decaying into a τ pair is depicted in details. The search for the SM $H \to \tau\tau$ has been performed in CMS covering all the six possible τ decay mode combinations: ee, $\mu\mu$, $e\mu$, $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$. In this Chapter the pp collisions data recorded by CMS during the 2012 at $\sqrt{s} = 8\,TeV$ corresponding to an integrated luminosity of $19.7\,fb^{-1}$ for the $\mu\tau_h$ channel are used. To exploit the different Higgs boson production mechanism, an exclusive event categorization based on the number of reconstructed jets in the event, on the p_T of the hadronic τ and on the p_T of the di-τ system has been carried out in the *standard* analysis. This categorization is preserved also in the case of MEM. Indeed, only the VBF categories exploit the MEM while all the others are used to constrain the backgrounds yields and shapes. A description of the events selection applied in the SM $H \to \tau\tau$ analysis is given in Section 7.4 together with a detailed definition of the categorization. The major backgrounds and their estimation are also described in Section 7.5. The identification and the isolation criteria applied to the τs are detailed. Even

if this topic is already treated in a more general context in Chapter 3.4, it represents a powerful handle to increase the purity of the selected τ's and to reduce the fake rate coming from hadronic jets, while preserving a reasonable selection efficiency.

7.1 The SM $H \to \tau\tau$ analysis in a nutshell

In this section, the main aspects of the published CMS $H \to \tau\tau$ [71] legacy analysis are presented. A 3.4 sigma evidence is obtained. This analysis is used as benchmark to evaluate the improvements brought by the application of the matrix element method to the $H \to \tau\tau \to \mu\tau$ channel.

7.2 Samples and data/MC corrections

As explained in length in the previous chapters, the L1 selection represents the first step of any analysis. It defines the physics content present in each event and the phase space available for the physics analysis. A more detailed description of the CMS trigger used during the LHC Run 1 is illustrated in Section 2. In the context of the SM $H \to \tau\tau$ analysis, different trigger paths are used depending on the final state considered. In addition, the threshold imposed to each triggered object has continuously evolved to match the LHC performance. The events that pass the L1 and high level trigger are processed offline. At the analysis level, the events are required to have fired a specific HLT path, specific to the final state considered. The PF reconstruction (see 3.1) is performed on the selected events. The jets seed the HPS algorithm that is able to identify jets coming from τ-lepton and determine the τ decay mode in according to the Tab. 3.1. The HLT bandwidth dedicated to the $H \to \tau\tau$ search in the semileptonic channels is $\sim 41\,\mathrm{Hz}$, representing $\sim 41\%$ of the bandwidth dedicated to the Higgs searches and about $\sim 14\%$ of the total bandwidth available at HLT level for CMS. At the analysis level, a muon with with a $p_T > 17(18)\,GeV$ and $|\eta| < 2.1$ and a PF τ with $p_T > 20\,GeV$ are required.

Dataset: The full reconstruction of the events that passed the HLT has been performed taking into account the latest alignment and calibration of each CMS subdetectors. The data thus processed at this stage constitutes the dataset available for the analysis. The SM $H \to \tau\tau$ search has been performed over a total integrated luminosity of $19.7\,fb^{-1}$, collected from pp collisions at $\sqrt{s} = 8\,TeV$.

Simulated datasets: The simulated events are produced using different Monte Carlo generators, and they are used to evaluate the yields of the signal and of many of the SM backgrounds. The signal samples are produced for each of the main Higgs boson production mechanism. For the gluon-gluon fusion and the VBF, the POWHEG [132][133][134][135] Monte Carlo generator is used, while for the Higgs production in association with a vector boson, the PYTHIA [136]

generator is used. In addition, for each Higgs boson production mechanism, different simulated samples are generated for each value of the Higgs boson mass in a range between $110\,GeV$ and $145\,GeV$, in $5\,GeV$ steps. The Higgs boson production cross sections and branching ratios, with their relative uncertainties, are calculated in Refs. [137][138][139]. As far as the background simulated samples are concerned, the $Z \to \ell\ell, (\ell = e, \mu, \tau)$, $W + jets$, $t\bar{t}$ and di-boson processes are simulated with MadGraph [140]. Single-top events are simulated using POWHEG. Exclusive samples for $Z \to \ell\ell$ and $W + jets$, binned in jet multiplicities, are produced in order to increase the statistic in regions characterized by a high signal purity. All the simulated processes are normalized to their NNLO production cross section, except the di-boson and single-top processes that are normalized to their NLO cross sections. Madgraph and POWHEG are interfaced with PYTHIA in order to simulate the parton shower and the fragmentation. To simulate the effect of the pile-up, pre-simulated min-bias events generated with PYTHIA are used. Randomly chosen min-bias events are superimposed to the hard interaction before being processed through the standard digitization and reconstruction chain. The τ-lepton decays are modelled with the TAUOLA [141] package, while the tau polarization effect are simulated with the TAUSPINER [142] framework.
Each event in each MC sample is thus reweighted by a factor w_{MC}:

$$w_{MC} = \frac{\sigma \cdot \mathcal{L} \cdot \varepsilon}{N_{proc}}, \tag{7.1}$$

where σ is the corresponding process production cross section, $\mathcal{L}$ is the total integrated luminosity recorded, ε is the efficiency of the possible event selection inserted in the analysis chain (described in Section 7.4) and N_{proc} is the number of the MC generated events.

Embedded samples: The Drell-Yan $Z/\gamma^* \to \tau\tau$ process constitutes the main irreducible background in the SM $H \to \tau\tau$ analysis. It is mandatory to simulate it in the most precise way, since small variations of its shape can affect the event yield in the signal region and the data/Monte-Carlo agreement. The hadronic τ decays are very well described by generators like TAUOLA. To reduce the uncertainties related to the DY events with high jet multiplicities, a data-driven method has been developed by the CMS collaboration to simulate $Z \to \tau\tau$ events. This method is called "embedding" and consists in selecting $Z \to \mu\mu$ events from data and create generated events with two τ's whose four-vectors correspond to the four-momenta of the muons. The decay of the τ lepton is then simulated by TAUOLA. The events are then processed through the full simulation chain of CMS before being embedded into the original $Z- > \mu\mu$ events, after removal of the muons. This approach presents the advantage that the pile-up content, the jet content and kinematics as well as the missing transverse energy are taken directly from data.

7.2.1 Analysis objects

Given the various τ decay modes, different particles in the final state are expected. Depending on the channel analysed (ee, $\mu\mu$, $e\mu$, $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$) electrons, muons, τ_h are reconstructed as

explained in more details in the Chapter 3. The jet and $\vec{\not{E}}_T$ reconstruction used in the $H \to \tau\tau$ analysis have been described also in Chapter 3.

7.2.2 Corrections to the simulation

Even if a lot of effort has been invested in developing the most accurate simulation, its prediction present some limitations. In particular, it is possible to isolate two main source of possible differences with respect to a real pp collisions. The first one originates from the theoretical evaluation of the production differential cross sections. Indeed, all the cross sections are evaluated at given order in perturbation theory. This effect could be negligible in some cases, when corrections coming from higher orders in the perturbation expansion are smaller. However, in some other cases, like in the p_T distribution of the Higgs boson events produced through gluon-gluon fusion, this effect is more visible. The second source of possible disagreement between data and simulation arises from the fact that it is almost impossible to predict the real amount of pile-up interactions as well the precise calibration and alignment condition of each CMS subdetector prior to the data-taking. Indeed, the number of pile-up interactions directly depends on the machine conditions, and it is far to be a constant during the whole data-taking period. The MC samples are produced well before the data taking, because of the huge computing time required. In these conditions, the calibration of each of the CMS subdetector can change with time. In the following the most important data/MC corrections applied to the simulation are listed along what is described in [71].

Pile-up. It is observed that the distribution of the number of the pile-up interactions is different in data and in simulation. Figure 7.1 shows the number of reconstructed primary vertices per event in data and in the simulation.

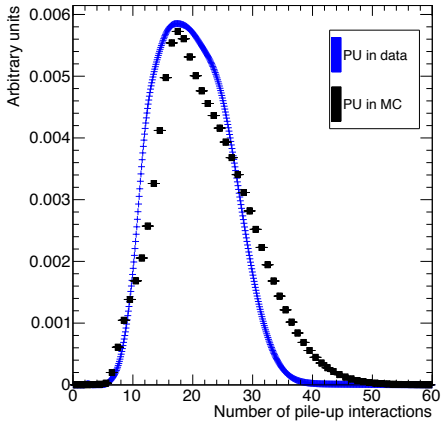


Figure 7.1: *Distribution of the reconstructed primary vertices for data (black dots) and simulation (blue line) before the pile-up reweighting.*

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 $P(N_{PU}^{data})$ and in the simulation$P(N_{PU}^{MC})$ (pile-up reweighting):

$$w_{PU} = \frac{P(N_{PU}^{data})}{P(N_{PU}^{MC})} \tag{7.2}$$

Trigger efficiency. The trigger efficiency is computed as function of the p_T of the reconstructed object (turn-on curves). The turn-on curves for each object (e, μ, τ_h) are obtained both with simulated events and with data using a tag-and-probe technique. Some differences (always smaller the 5%) are observed between data and simulation. Examples of the turn on mentioned are provided for the μ and τ_h, shown in Fig. 7.2 and in Fig. 7.3, respectively.

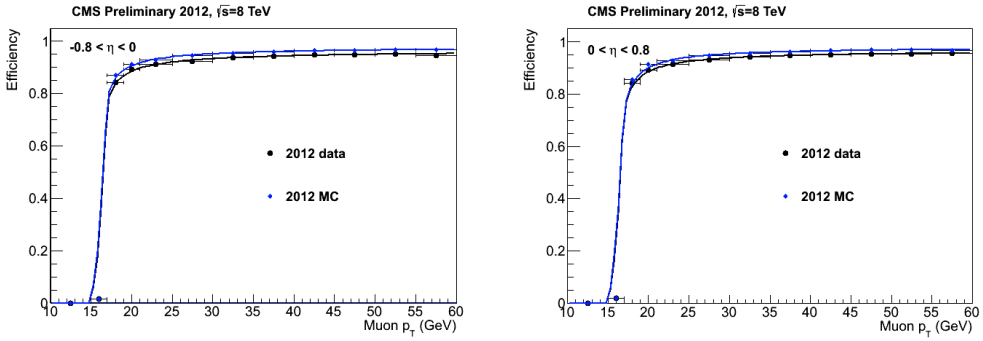


Figure 7.2: *Turn on curves measured in data and in simulation for the muon trigger as a function of the reconstructed muon p_T in the $|\eta| < 0.8$ region.*

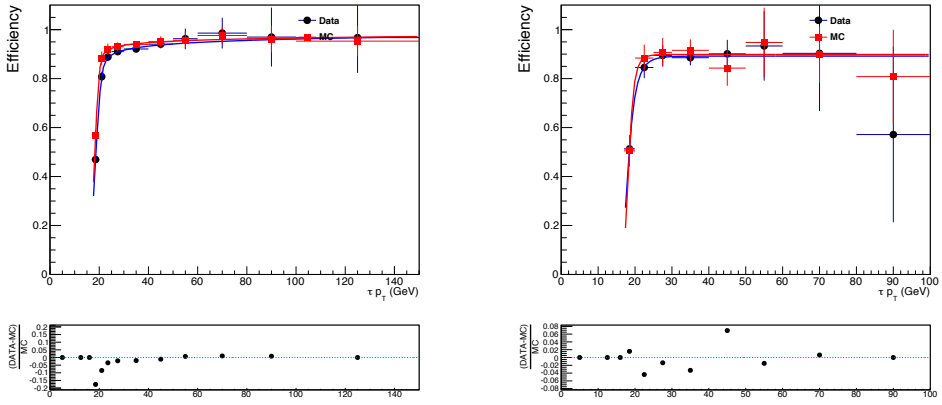


Figure 7.3: *Turn on curves measured in data and in simulation for the τ_h trigger as a function of the reconstructed τ_h p_T in the barrel (Left) and endcap (Right) region.*

In the $H \to \tau\tau$ analysis, a cross trigger is always applied. It means that at the trigger level the presence of two objects is required, depending on the particular final state considered. In a good approximation, the efficiency to trigger these two objects can be considered as the product of the efficiency to trigger each one of the single object. As a result, a correction to the simulation is applied, and in particular, for the semileptonic channels it takes the form:

$$w_{trig} = \frac{\varepsilon^{data}}{\varepsilon^{MC}} = \frac{\varepsilon^{data}(p_T^{\ell})\varepsilon^{data}(p_T^{\tau_h})}{\varepsilon^{MC}(p_T^{\ell})\varepsilon^{MC}(p_T^{\tau_h})} \tag{7.3}$$

The p_T dependence of the turn-on curves is modelled with a function proportional to the integral of a double Crystal Ball function.

The τ_h decay mode reweighting. The τ_h is reconstructed with the HPS algorithm (see Section 3.4). It turns out that differences in the fraction of the events reconstructed in each particular decay mode are observed between the data and the simulation. In order to correct the MC sample, six correction factors (one per decay mode and per barrel/endcap $|\eta|$ region) have been derived. It is important to stress the fact that the corrections are derived in a way that guarantees the conservation of the total number of reconstructed τ_h independently on the decay mode.

Higgs boson p_T reweighting in the gluon-gluon fusion process. The gluon-gluon fusion, through a loop of heavy quark (mainly the top quark), is the dominant production mechanism of the SM Higgs boson at the LHC. The differential Higgs boson production cross section versus τ p_T, $\frac{d\sigma_H}{dp_T}(q_T, m_H, s)$ is a function of q_T, the transverse momentum of the Higgs boson, m_H is the Higgs boson mass and and s, the energy squared in the pp collision centre-of-mass. The quantity $\frac{d\sigma_H}{dp_T}(q_T, m_H, s)$ represents the spectrum in transverse momentum of the Higgs boson produced through gluon-gluon fusion. In the simulated signal sample used in the $H \to \tau\tau$ analysis, the $\frac{d\sigma_H}{dp_T}(q_T, m_H, s)$ is evaluated at a fixed order in perturbation theory at next-to-leading order (NLO). In general, this approach is justified when the magnitude of $q_T \sim m_H$ since, in this case, the perturbative series is controlled by α_S^n, with $\alpha_S(m_H) < 1$ and $n \geq 1$ [143][144]. However, in the small q_T region, the perturbative expansion is spoilt by the presence of logarithmic terms of the form $\propto \alpha_S^n log^m(m_H^2/q_T^2)$ with $1 \leq m \leq 2n-1$; due to multiple soft-gluon emission to all orders in QCD perturbation theory. These terms, indeed, diverge as $q_T \to 0$. In order to obtain a reliable prediction for $\frac{d\sigma_H}{dp_T}(q_T, m_H, s)$, a resummation to all orders is needed [143][144]. The q_T spectra evaluated at NLO, compared to the one obtained at next-to-next-to-leading logarithmic (NNLL) resummation combined with the NLO fixed-order calculation are illustrated in Fig. 7.4. It implies that a reweighting of the Higgs p_T evaluated at NLO is needed to match the distribution at NNLL+NLO. The importance of this reweighting is due to the fact that, at the analysis level, the signal yield depends on the Higgs boson p_T because of the cuts on the transverse momenta.

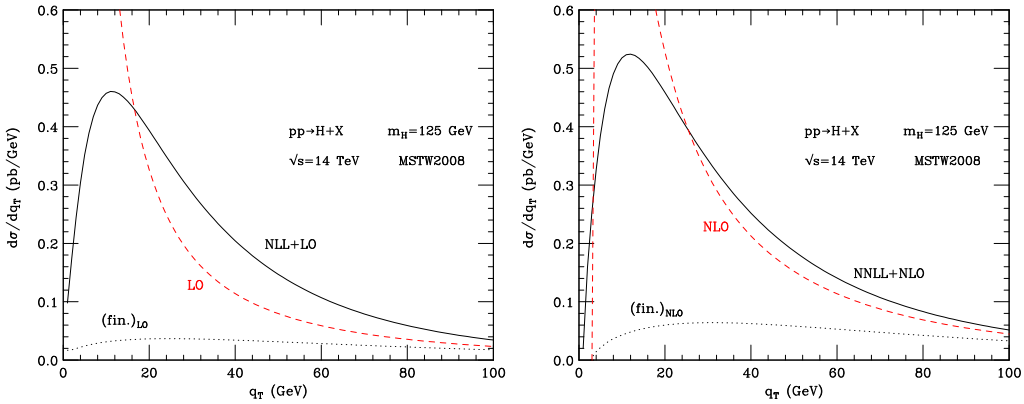


Figure 7.4: *Higgs boson gluon-gluon fusion differential cross section as a function of the Higgs boson transverse momentum evaluated from the program HQT2 [143] at* $\sqrt{s} = 14\,TeV$ *and* $m_H = 125\,GeV$*. LO fixed-order computation (red-dashed) and NLL (next-to-leading logarithmic resummation)+LO computation (solid black) (Left). NLO fixed-order computation (red-dashed) and NNLL (next-to-next-to-leading logarithmic resummation)+LO computation (solid black) (Right). From Ref. [143]*

Other corrections Other corrections are applied to MC sample: τ_h energy scale (see also 3.4.2), jet energy scale (see also 3.2), b-tag efficiency and mis-tagging rate, lepton identification and isolation efficiencies, $\vec{\not{E}}_T$ recoil corrections (see Sec. 3.3) and τ_h p_T reweighting in the W+jets events. A detailed treatment of these corrections can be found in Ref. [145].

7.3 Statistical interpretation of the results

In this Section. the statistical procedure used to interpret the results in the SM $H \to \tau\tau$ search is detailed. The signal extraction is performed through a binned maximum likelihood fit to the SVfit mass distribution (or on the matrix element likelihood ratio, as explained in subsection 7.7.2). The fit is performed simultaneously in all the categories and in all the channels analysed.

Likelihood model: The fit is performed both under the signal + background hypothesis (H_1) and under the background only hypothesis (H_0). In what follows, the attention is given to the case where SVfit is used as the fitted variable, but the same concepts are still valid also in the case the MEM likelihood ratio distribution is used. The statistical approach adopted by the LHC experiments is called *modified CLs method* [146]. It is based on a profile likelihood ratio as it is recommended by the LHC Higgs Combination Group [147] [148]. The results of an observation is an array of data yields in each bin of all the mass distribution entering in the combined fit. The expected number of events in the i^{th} bin is:

$$\nu_i(\mu, \vec{\theta}) = \mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta}), \tag{7.4}$$

where s_i and b_i represents the signal and background yields, respectively. These yields depend on a set of nuisance parameters: $\vec{\theta}$. Typically, nuisance parameters are associated to systematic uncertainties. The signal strength modifier, μ, is a free parameter of the fit. It is free to rescale the signal event yield by a factor common to all the Higgs boson production mechanisms. The H_1 hypothesis correspond to a value of $\mu = 1$, while H_0 corresponds to $\mu = 0$. Thus, the probability to observe n events under the assumption $\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})$ is given by the product of the probabilities to observe n_i events in each bin. In each bin, such a probability follows a Poissonian distribution. It is possible to write the likelihood function:

$$\mathcal{L}\left(n|\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})\right) = \prod_i \frac{\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})}{n_i!} e^{-[\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})]} \prod_j \rho\left(\theta_j|\hat{\theta}_j\right). \tag{7.5}$$

In Eq. 7.5, the values θ_j represent the parameters assigned to each systematic. The term $\rho\left(\theta_j|\hat{\theta}_j\right)$ is a probability density function. It represents the probability that the true value of the nuisance parameter j^{th} is equal to θ_j given the best estimate $\hat{\theta}_j$. The best estimate is often obtained by auxiliary measurements. The nuisance parameters $\vec{\theta}$ are thus included in the fit as constrained parameters, and their values are estimated by the fit simultaneously with μ. Their possible values are however constrained by the measured uncertainties on the systematics. So. *de facto*, each θ_j can vary within a confidence interval. The nuisance parameters can be grouped into two main classes: the yield and the shape uncertainties.

Yield uncertainties: These uncertainties affect the event yield estimation of any process considered in the analysis (signal or backgrounds) and, depending on the origin of the uncertainties, the $\rho(\theta)$s of the corresponding nuisance parameters are usually modelled with two different types of probability density functions (a log-normal or Gamma distribution). The uncertainties related to the determination of the various normalizations are modelled with a log-normal distribution of the type:

$$\rho(\theta) = \frac{1}{\sqrt{2}\ln\kappa} \exp - \frac{\ln\left(\theta/\hat{\theta}\right)^2}{2(\ln\kappa)^2} \frac{1}{\theta}, \tag{7.6}$$

where $\kappa = 1 + \varepsilon$ and ε represents the relative scale of the uncertainties. Some examples of this uncertainties are the cross section uncertainties, the identification and isolation efficiencies, the extrapolation factors used to estimate the backgrounds contributions, etc. The treatment of the nuisance parameters is complicated by the fact that many processes, channels or categories can be affected by the same nuisance parameter. In that case, the uncertainties associated to that nuisance parameter are considered to be fully correlated or anti-correlated.

The second type of yield uncertainty comes from auxiliary measurements; typically, the events counted in a side-band region used to constrain the event yield in the signal region and it has a statistical origin. For this reason, they are modelled by a Gamma distribution:

$$\rho(n) = \frac{1}{\alpha} \frac{(n\alpha)^N}{N!} \exp\left(-n/\alpha\right), \tag{7.7}$$

where N is the number of events in the control region used to extrapolate the number of events in the signal region $n = \alpha N$. It is possible that the parameter α is affected also by the first type of yield uncertainties described above.

Shape uncertainties: The nuisance parameters that are sensitive to the scale of a variable correlated with SVfit (or to the variable used to be fitted) lead to shape template uncertainty. An example of such systematic is provided by the τ energy scale (see also 3.4.2). 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 in order to interpolate between the nominal template ($\lambda = 0$) and the two deviated templates. The effect of all the shape nuisances is additive.

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:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \tag{7.8}$$

with $\hat{\mu}$ and $\hat{\vec{\theta}}$ are values maximizing $\mathcal{L}(\mu, \vec{\theta}_\mu)$, μ the tested signal strength, and $\hat{\vec{\theta}}_\mu$ the set of nuisance parameter maximizing $\mathcal{L}(\mu, \vec{\theta}_\mu)$ for each fixed value of μ. In addition, the constraint $0 \leq \hat{\mu} \leq \mu$ is enforced in order to ensure the physical meaning of the results. The test statistic q_μ is then defined as:

$$q_\mu = -2 \ln \lambda(\mu) = \begin{cases} -2 \ln \frac{\mathcal{L}(\mu, \vec{\theta}_\mu)}{\mathcal{L}(0, \hat{\vec{\theta}})} & \text{if } \hat{\mu} < 0 \\ -2 \ln \frac{\mathcal{L}(\mu, \vec{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 0 \text{ if } \mu < \hat{\mu} \end{cases}$$

By construction, $q_\mu \geq 0$. It is a measure of the incompatibility between data and the $(\mu \cdot s + b)$ hypothesis. The higher it is, the less data is compatible with the hypothesis on μ.

Limit setting procedure

In order to establish confidence levels on μ, the probability density functions $f(q_\mu | \mu \cdot s + b)$ as a function of μ, or sampling distribution of q_μ is used. It is computed for the H_0 and H_1 hypotheses 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 under the signal plus background hypothesis (p_μ) and under the background only hypothesis ($1 - p_b$) given the actual observed value of the test statistic q_μ^{obs}:

$$p_\mu = \int_{q_\mu^{obs}}^{\inf} f(q_\mu|\mu) dq_\mu \tag{7.9}$$

$$1 - p_b = \int_{q_\mu^{obs}}^{\inf} f(q_\mu|0) dq_\mu \tag{7.10}$$

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} becomes very small due to a negative background fluctuation, which could lead to an exclusion of a value at 95% C.L. not originated by the absence of signal. The alternative approach used is then the CL_s method, defined as:

$$CL_s = \frac{p_\mu}{1 - p_b} = \frac{CL_{s+b}}{CL_b} \tag{7.11}$$

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 under 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.

Significance and p-value

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:

$$q_0 = \begin{cases} -2 \ln \frac{\mathcal{L}(0,\vec{\theta}_0)}{\mathcal{L}(\hat{\mu},\vec{\theta})} & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{if } \hat{\mu} < 0 \end{cases}$$

In the case where $\hat{\mu} < 0$, the test statistic is zero as this case happens when the background fluctuates negatively, and 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:

$$p_0 = \int_{q_0^{obs}}^{\inf} f(q_0|\mu = 0) dq_0, \tag{7.12}$$

where q_0^{obs} is the observed value of q_0 in data and $f(q_0|\mu = 0)dq_0$ the pdf of q_0 given the background only hypothesis. The p_0 is the probability that the excess is due to a fluctuation of

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 7.1: *Correspondance between the significance and the p-value.*

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 ($\phi^{-1} = \sqrt{2}\text{erf}(2x-1)$ with $x \in [0,1]$). 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 Tab. 7.1.

7.4 Events selection

7.4.1 Baseline event selection

The baseline di-τ selection selects one signal candidate per event. It ensures that all the selected events pass quality criteria and suppresses the reducible background. The objects used to apply the pre-selection are those described in Section 7.2.1. Some of the inclusive selections are common to all channels considered in the $H \to \tau\tau$ analysis while some other selections depend on the particular final state considered. As in this Chapter, I focus on the semileptonic $\mu\tau_h$ channel, I will describe only the selection related to this channel. The events containing a reconstructed $\mu\tau_h$ pair are selected requiring the following criteria:

Inclusive selections common to all decay channels:

- The primary vertex has to pass the quality criteria described in Section 3.1.2;
- The lepton and the τ_h are required to be of opposite charge;
- If several pairs pass the previous criteria, the one with the highest scalar sum of the muon and τ_h momenta, $p_T(\ell) + p_T(\tau_h)$, is chosen as the signal candidate;
- In order to reduce the W+jets background, the transverse mass of the lepton (e or μ) and

the missing transverse energy, defined as:

$$m_T(\ell, \vec{\not{E}}_T) = \sqrt{2 p_T^{\ell} \not{E}_T (1 - \cos\Delta\phi)} \tag{7.13}$$

is required to be smaller than $30\,GeV$. For simulated events, the MVA-based recoil corrected missing transverse energy is applied (see Section 3.3 for more details);

- Additionally to the second electron or muon veto, another lepton veto is applied in both channels requiring the selected events to contain no additional loosely identified τ's, electrons or muons.

- To suppress the $t\bar{t}$ background a veto on b-tagged jets with a $p_T > 20\,GeV$ is applied.

Inclusive selections specific to the $\mu\tau_h$ channel:

- The event has to be triggered by any of the $\mu\tau_h$ trigger paths [71];

- One muon passing the tight working point identification criteria (see Section 3.1.5), matching the HLT muon filter object of the corresponding trigger path by $\Delta R = 0.5$, is required. It should fulfill $p_T(\mu) > 20\,GeV$ and $|\eta(\mu)| < 2.1$, and pass the isolation cut $I_{rel} < 0.1$ (see Section 3.1.5);

- One hadronically decaying τ lepton passing the $I_\tau < 1.5\,\text{GeV}$ of the isolation criteria described in 3.4.2, and matching the HLT τ filter object of the corresponding trigger path by $\Delta R = 0.5$ is required. It should be such that $p_T(\tau) > 30\,GeV$ and $|\eta(\tau)| < 2.1$. 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 $e \to \tau_h$ and $\mu \to \tau_h$ fake rates, respectively;

- 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 to further suppress the $Z/\gamma^* \to \mu\mu$ background.

7.4.2 Categorization

In order to increase the sensitivity of the $H \to \tau\tau$ analysis, the events are split into mutually exclusive categories. The gain is maximized when the events are sorted in categories with very different signal over background ratio since:

$$\frac{S}{\sqrt{B}} < \sqrt{\left(\frac{S_1}{\sqrt{B_1}}\right)^2 + \left(\frac{S_2}{\sqrt{B_2}}\right)^2 + ... + \left(\frac{S_n}{\sqrt{B_n}}\right)^2} \tag{7.14}$$

where $S = S_1 + S_2 + ... + S_n$ and $B = B_1 + B_2 + ... + B_n$. The categories are defined depending on the number of the energetic jets reconstructed in the event, on the p_T of the hadronic τ and

on the p_T of the reconstructed di-τ system. In Fig. 7.5, a sketch summarizing the categories defined for the $\mu\tau_h$ channel is shown. In this Section, a detailed description of the categories involved in the $\mu\tau_h$ channel is provided. More details concerning the other channels considered in the $H \to \tau\tau$ analysis can be found in Ref. [71].

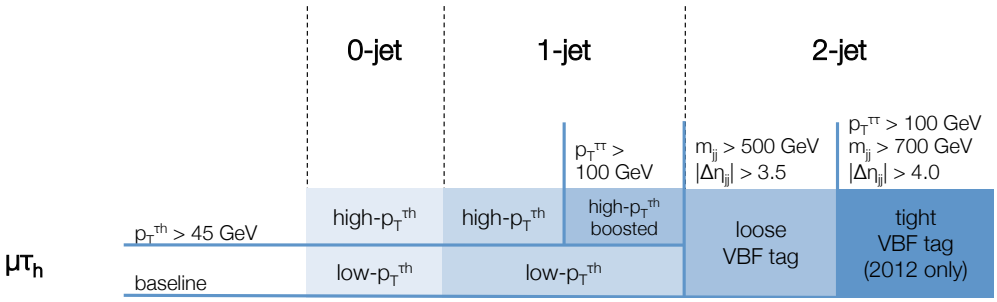


Figure 7.5: *Categories defined in the $H \to \tau\tau \to \mu\tau_h$ search.*

VBF categories

The VBF categories are aimed to exploit the Higgs boson produced through the vector boson fusion mechanism. This production mode has a well defined topology, as described in 1.5.1. In the events belonging to the VBF categories, at least two energetic jets, i.e. with a $p_T > 30\,GeV$, are required in the event. In addition, two exclusive subcategories are defined depending on the requirements on the di-jet invariant mass (m_{jj}) and on the difference of the jet pseudorapidity $\Delta\eta_{jj}$ between the two leading jets in the events. The VBF-Tight category is thus defined requiring $m_{jj} > 700\,GeV$ and $|\Delta\eta_{jj}| > 4.0$ with an additional requirement on $p_T^{\tau\tau} > 100\,GeV$ for the di-τ system and a veto on additional jets: no jet with a p_T larger than $20\,GeV$ should be present in the region delimited by the two highest-p_T jets. The VBF-Loose category is characterized by looser requirements: $m_{jj} > 400\,GeV$ and $|\Delta\eta_{jj}| > 3.5$; the events already entered in the VBF-Tight category are not considered in the VBF-Loose one. All the events that do not enter in any of the VBF categories are classified into the 1-Jet and 0-Jet categories.

1-Jet categories

The 1-Jet categories are designed to exploit the Higgs boson production through the gluon-gluon fusion mechanism and recover the events produced through the VBF mechanism having one of the two energetic jets outside the detector acceptance. As the Higgs boson is produced with a transverse boost ($\sim 80\,\text{GeV}$), there is often a recoiling jet. Therefore, at least one energetic jet ($p_T > 30\,GeV$) is required to be reconstructed in the event. Three different subcategories are defined depending on the p_T of the hadronic τ: in the 1-Jet-Low category $30\,GeV < p_T(\tau_h) <$

$45\,GeV$ is required, while the 1-Jet-High category is defined by $p_T(\tau_h) > 45\,GeV$. Finally, in the 1-Jet-High-Boosted category it is required in addition that $p_T^{\tau\tau} > 100\,GeV$ for the di-τ system.

0-Jet categories

The 0-Jet categories include the majority of events and are mainly used to constrain the background shapes and normalizations in the final fit procedure (see Section 7.7.3) and to estimate the energy scales. There are two 0-Jet subcategories: the 0-Jet-Low category that includes all the events where no energetic jet has been reconstructed and $30\,GeV < p_T(\tau_h) < 45\,GeV$, while the 0-Jet-High category is defined as collection of the events with no energetic jet reconstructed in the event despite a hadronic τ with $p_T > 45\,GeV$.

7.5 Background estimation

The main background in the $H \to \tau\tau$ analysis are estimated using data-driven techniques. The major background sources for the $\mu\tau_h$ channel come from $Z/\gamma^* \to \tau\tau$ Drell-Yan production QCD multi-jets events, W+jet production, di-boson production (WW, WZ and ZZ), single-t and $t\bar{t}$ production. In addition, events coming from Drell-Yan production of $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to ee$ contribute to the main background in $H \to \tau\tau \to e\tau_h/ee/\mu\mu$ channels. The background contributions are, as much as possible, evaluated through data-driven techniques. In the following, I concentrate on the description of the main background sources and evaluations in the $\mu\tau_h$ channel. More details about the estimation of all the others backgrounds that dominate in other channels but the $\mu\tau_h$ can be find in Ref. [71].

7.5.1 Irreducible $Z/\gamma^* \to \tau\tau$

The $Z/\gamma^* \to \tau\tau$ Drell-Yan (DY) production represents the main irreducible background to the SM $H \to \tau\tau$ analysis, because of the resonant nature of the process and because it gives rise to genuine τ-leptons in the final state. For these reasons, it is of capital importance to estimate the shape and the normalization of this background in the most accurate way.

Shape

The mass shape of events coming from DY $Z/\gamma^* \to \tau\tau$ is estimated from the embedded technique described in Section. 7.2. The baseline selection listed in Section 7.4 and the full set of the data to simulation corrections described in Section 7.2 are applied to the DY $Z/\gamma^* \to \tau\tau$ embedded events. The normalization of this background has been obtained using DY $Z/\gamma^* \to \tau\tau$ simulated events.

Normalization

in order to extract the proper normalization factors in each of the categories considered, an extrapolation factor $\varepsilon^{Emb,cat}_{Z\to\tau\tau}$ from the baseline to the category selection is applied. This extrapolation factor has been estimated thanks to relationship:

$$\varepsilon^{Emb,cat}_{Z\to\tau\tau} = \frac{N^{Emb,cat}_{Z\to\tau\tau}}{N^{Emb,inc}_{Z\to\tau\tau}}. \tag{7.15}$$

The quantities $N^{Emb,inc}_{Z\to\tau\tau}$ and $N^{Emb,cat}_{Z\to\tau\tau}$ represent, respectively, the number of events in the embedded sample that pass the baseline selection (without the m_T cut) and the number of events that satisfy in addition the specific category selection. Thus,the normalization factor in a given category is given by:

$$N^{cat}_{Z\to\tau\tau} = N^{MC,inc}_{Z\to\tau\tau} \times \varepsilon^{Emb,cat}_{Z\to\tau\tau}, \tag{7.16}$$

where $N^{MC,inc}_{Z\to\tau\tau}$ is the normalization taken from Monte-Carlo simulation with the baseline selection (without the m_T cut).

7.5.2 Irreducible W+jets

The $W + jets$ background is a major background in the semileptonic channels. Indeed, a real lepton comes from the W decay while a jet can fake the hadronic τ.

Shape

The invariant mass shape is modelled using Monte Carlo simulation in each category. In the particular case of the VBF categories, some selection criteria have been relaxed in order to obtain smoother distributions [71].

Normalization

The normalization of the $W+jets$ background has been carried out using data-driven techniques. A data control region enriched in $W + jets$ events has been defined requiring $m_T(\ell, \vec{\not{E}}_T) > 70\,GeV$. In particular, for the VBF category, a cut $m_T(\ell, \vec{\not{E}}_T) < 120\,GeV$ has been also applied in order to reduce the contamination in the $W + jets$ enriched region by non-$W + jets$ events populating the tails of the m_T distribution. The other source of background are estimated through Monte-Carlo simulation and then subtracted:

$$N_{W+jets} = r_W \times \left[N^{Data} - N^{MC}_{Z\to\tau\tau} - N^{MC}_{Z\to\ell\ell} - N^{MC}_{t\bar{t}} - N^{MC}_{VV}\right]^{m_T>70} \tag{7.17}$$

where the extrapolation factor is estimated by the ratio between the number of simulated events found in a $m_T < 30\,GeV$ region (same as the signal) and in $m_T > 70\,GeV$ sideband region:

$$r_W = \frac{N^{MC,m_T<30}_{W+jets}}{N^{MC,m_T>70}_{W+jets}}. \tag{7.18}$$

In Fig. 7.6 the distribution of the transverse mass $m_T(\ell, \vec{E}_T)$ for each background and signal processes can be observed.

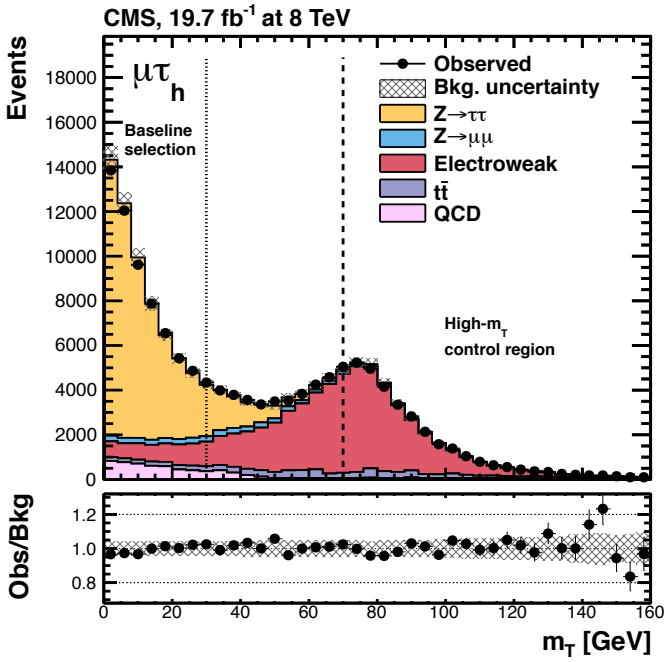


Figure 7.6: *Transverse mass distribution (m_T) for all background and signal contribution. The two vertical lines represent the limits of the signal enriched and W + jets enriched regions.*

7.5.3 The QCD multi-jets events

The QCD multi-jet represents another major background because of its huge production cross section at the LHC ($\sim 10\,mb$). In these events, the reconstructed lepton is originated from an heavy flavour decay or a misidentified jet, while the τ_h comes from a misidentified jet. The contribution of this background is estimated using data-driven techniques by counting the events present in a control region enriched in QCD multi-jet events.

Shape

The reconstructed di-τ shape is obtained from events that contain a muon and a hadronic τ's with the same charge (SS events). In addition, since the QCD processes are often accompanied by soft radiation, the lepton is required to be anti-isolated. The other background processes are estimated through Monte-Carlo simulation and then subtracted. A special treatment is deserved to the VBF categories. Indeed, due to the lack of statistic in these categories, in order to obtain a

smooth template for the QCD multi-jet shape, the selection criteria have been relaxed, requiring $p_T(jet) > 20\,GeV$, $m_{jj} > 200\,GeV$ and $|\Delta\eta_{jj}| > 2.0$ in addition with a central jet veto.

Normalization

The normalization factor for this background is obtained from data, considering SS events in a QCD-enriched control region, after the subtraction of all the others background estimated using Monte-Carlo simulation (except for the $W+jets$ that is estimated with data-driven techniques):

$$N_{QCD} = r_{OS/SS} \times \left[N^{Data} - N^{MC}_{Z\to\tau\tau} - N^{MC}_{Z\to\ell\ell} - N^{MC}_{t\bar{t}} - N^{MC}_{VV} - N^{Data-driven}_{W+jets} \right]^{SS} \quad (7.19)$$

The extrapolation factor $r_{OS/SS}$ is obtained as the ratio of OS over SS anti-isolated events (reversed lepton isolation + relaxed τ isolation) in a QCD-enriched control region:

$$r_{OS/SS} = \frac{N^{Data,OS,AntiIso}_{QCD}}{N^{Data,SS,AntiIso}_{QCD}} = 1.06. \quad (7.20)$$

7.5.4 $t\bar{t}$ production

The t-quark decays into a b-quark and a vector boson W. Subsequently, the vector boson can, in turn, decay ($W \to \ell\nu_\ell$) with a lepton in the final state. Therefore, depending on the lepton flavour coming from the W decay, the $t\bar{t}$ production can give rise to a true $\mu\tau$ pair or, one of the jets originating by the b-quark can fake an hadronic τ while the lepton coming from the other W is not-reconstructed or outside the detector acceptance. This source of background is modelled by Monte-Carlo simulations. In the VBF categories, the shape is taken from events with VBF relaxed selections. The $t\bar{t}$ contribution is normalized to its cross section evaluated at NNLO. An efficient way to reduce this background consists in applying a b-tagged jet veto in all the categories.

7.5.5 Di-boson and single-t

This source of background, in the semileptonic channels and, in particular, in the $\mu\tau$ one, is very small. This background is modelled with Monte Carlo simulations and its invariant mass shape in the VBF categories is estimated from data relaxing the VBF selections.

7.6 Systematic uncertainties

The simultaneous adjustment of the shape and normalization of the simulated signal and background $m_{\tau\tau}$ distributions in all the categories and channels allows the estimation of the compatibility between the data and the H_0 or H_1 hypotheses (see Section 7.3). The systematic uncertainties detailed in this Section are used to define the probability density functions $\rho(\hat{\theta}|\theta)$

of the associated nuisance parameters considered in the maximum likelihood fit. It is possible to classify the systematic uncertainties into two main classes: the uncertainties coming from the theory, affecting, in particular, the estimation of the signal production cross sections and those related to the experiment. In this section the main systematic uncertainties considered in the $H \to \tau\tau$ analysis are listed, and their values are summarized in Tab 7.2. It is beyond the scope of this Chapter to provide a exhaustive dissertation of all the systematics present in such a sophisticated analysis. For more detailed information it is possible to make reference to [119], [145] and [71].

7.6.1 Theoretical uncertainties

The theoretical systematics are related to the uncertainties on the theoretical calculation of the signal production cross section. They have several origins: uncertainties on the parton distribution function (PDF), strong coupling (α_S), the renormalization and factorization scales. In addition, the uncertainties on the simulation of the parton showering, of the hadronization and of the underlying events are included in this class of systematics. In particular, the normalization uncertainty on the signal yield in the VBF categories is estimated to be $\sim$ 4% [137]. the systematics on the computation of the gluon-gluon fusion cross section ranges from 10% up to 40% depending on the category considered. Uncertainties on the parton distribution function are estimated to be around 8%.

7.6.2 Experimental uncertainties

The experimental systematics are further subdivided into two subclasses depending on whether they impact the shape of the $m_{\tau\tau}$ distribution of the simulated processes or their normalization.

Shape uncertainties

In this subclass are included the systematics on the τ_h energy scale (3%) (see Section 3.4), the electron energy scale (1%), the jet energy scale (strongly dependent on the η and p_{T} of the jets, but ranging between 3% and 20% over among the different categories considered in the analysis). Finally, the $\not\!\!E_T$ scale (2%-8% depending on the category considered) is included in this subclass of systematics uncertainties.

Normalization uncertainties

The experimental uncertainties on the normalization of the main irreducible background, the DY $Z/\gamma^* \to \tau\tau$, after the baseline selection are estimated to be about 3%. In addition, the extrapolation factors used to estimate the DY $Z/\gamma^* \to \tau\tau$ contribution in the different categories are estimated with a systematic that is dominated by the statistical uncertainties and range from 2% up to 14% depending on the category considered. The evaluation of the normalization for

the $W + jets$ background has been derived directly from data, considering the high-m_T sideband region. In addition, an uncertainty on the extrapolation factor is also present. Practically, the main contribution in this systematic uncertainty is the limited statistics of the sample used to derive the extrapolation factors.
The $W + jets$ normalization overall systematics ranges between 10% and 25%. As far as the QCD multi-jet background is concerned, there are two sources of systematics: the first is related to the limited statistics of the QCD-enriched control sample and the second one arises from the errors on the estimation of the extrapolation factors. For the QCD multi-jet normalization the total uncertainties range from 6% up to 35%. The other minor systematic uncertainties affecting the estimation of the other background normalizations are listed in Tab. 7.2.
In the following of this chapter, the $H \to \tau\tau \to \mu\tau_h$ is performed using two different analysis approaches: the standard one, that aims to reproduce the results already obtained in the CMS published analysis in the $\mu\tau_h$ channel [71], and the upgraded one, in which the MEM developed in Chapter 6 is used in the VBF categories. The baseline selection and categorization, the backgrounds estimation and the systematic uncertainties are considered the same in the two analysis approaches.

7.7 Performance and results of the MEM in the VBF $H \to \tau\tau \to \mu\tau_h$ search

The MEM approach consists in building an event-by-event discriminant that is proportional to the probability that a particular observed event has been produced by a given process for which a theoretical model is known *a priori*. In the context of the VBF $H \to \tau\tau \to \mu\tau_h$ search, as detailed in Chapter 6, the theoretical model assumed is the SM, and the events are tested against two particular hypotheses: the signal one (VBF $H \to \tau\tau$) and the background one (Drell-Yan (DY) $Z/\gamma^* \to \tau\tau$). This choice is justified by the fact that in the $\mu\tau_h$ channel the event rate expected for the Drell-Yan process dominates the rate expected for all the other SM backgrounds. Thus, for each event, the integrations over the phase-space of the final state particles, the Bjorken fraction of the incoming partons and over the matrix element squared of the tested processes, have been performed for both the signal and background hypotheses. The notation for the weights w_i, with $i = VBF, DY$, represents the final values obtained, for each event, from the full integration when the matrix element for the VBF or DY processes are respectively tested. As explained in Section 7.7.1 in more details, the final discrimination variable results in a simple function of w_{VBF} and w_{DY}. The weight w_i have first been obtained for the VBF (blue) and DY (red) simulated events and are illustrated in Fig.7.7, 7.8 and 7.9 for the different reconstructed τ decay modes, while the inclusive distributions are shown in Fig. 7.10. On the left plots, the matrix element tested is the VBF one, while on the right plots, the DY matrix element is considered. Since the w_i can take values that differ by several order of

Incertitude	0-Jet	1-Jet	VBF
e Id + trigger	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$
μ Id + trigger	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$
τ_h Id + trigger	$\pm 8\%$	$\pm 8\%$	$\pm 8\%$
τ_h energy scale	$\pm 3\%$	$\pm 3\%$	$\pm 3\%$
e energy scale	$\pm 1\%$	$\pm 1\%$	$\pm 1\%$
$jets$ energy scale	$\mp 3 - 15\%$	$\pm 1 - 6\%$	$\pm 5 - 20\%$
$\not\!\!E_T$ energy scale	$\mp 2 - 7\%$	$\pm 2 - 7\%$	$\pm 5 - 8\%$
$\mathcal{L}_{int}$	$\pm 4.2\%$	$\pm 4.2\%$	$\pm 4.2\%$
b-jets efficiency	$\mp 2\%$	$\mp 2 - 3\%$	$\mp 3\%$
b-jets fake	$\mp 2\%$	$\mp 2\%$	$\mp 2 - 3\%$
DY $Z/\gamma^* \to \tau\tau$ norm.	$\pm 3\%$	$\pm 3\%$	$\pm 3\%$
$\varepsilon_{Z\to\tau\tau}^{Emb,cat}$	$\pm 0 - 5\%$	$\pm 3 - 5\%$	$\pm 10 - 13\%$
$W + jets$ norm.	$\pm 20 - 27\%$	$\pm 10 - 33\%$	$\pm 12 - 30\%$
QCD norm.	$\pm 6 - 32\%$	$\pm 9 - 30\%$	$\pm 19 - 35\%$
$t\bar{t}$ norm.	$\pm 10\%$	$\pm 10\%$	$\pm 12 - 33\%$
di-boson norm.	$\pm 15 - 30\%$	$\pm 15 - 30\%$	$\pm 15 - 100\%$
$ZL\mu \to \tau_h$ norm.	$\pm 30\%$	$\pm 30\%$	$\pm 30\%$
$ZLe \to \tau_h$ norm.	$\pm 20\%$	$\pm 36\%$	$\pm 22\%$
$ZJjet \to \tau_h$ norm.	$\pm 20\%$	$\pm 20\%$	$\pm 40\%$
PDF	-	$\pm 2 - 8\%$	$\pm 2 - 8\%$
$\mu_r/\mu_f(gg \to H)$	-	$\pm 10\%$	$\pm 30\%$
$\mu_r/\mu_f(qq \to H)$	-	$\pm 4\%$	$\pm 4\%$
$\mu_r/\mu_f(qq \to VH)$	-	$\pm 4\%$	$\pm 4\%$
UE + PS	-	$\pm 4\%$	$\pm 4\%$

Table 7.2: *Theoretical and experimental systematics uncertainties. The symbol $\mp$ means that the systematic considered is anti-correlated with respect to the other categories.*

magnitude, the quantity considered in the plots is $-log(w_i)$.

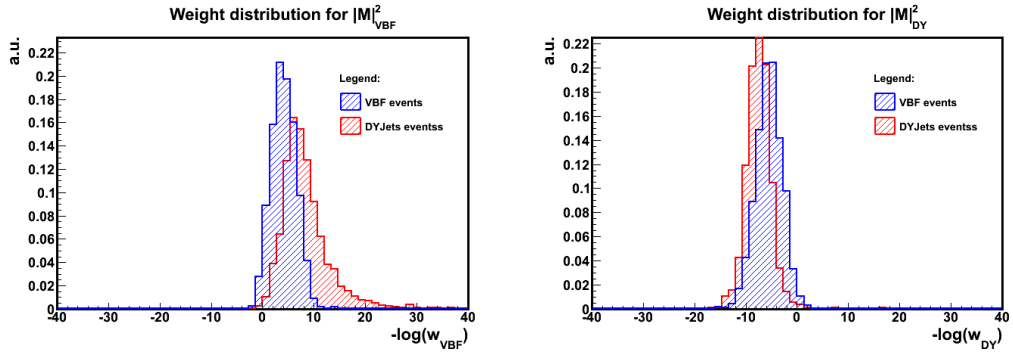


Figure 7.7: *Distribution of the $-log(w_{VBF})$ (Left) and $-log(w_{DY})$ (Right) obtained for the VBF $H \to \tau\tau$ (blue histograms) and for the DY $Z/\gamma^* \to \tau\tau$ Monte-Carlo simulated events. Only events with the τ_h decaying 1-prong are considered.*

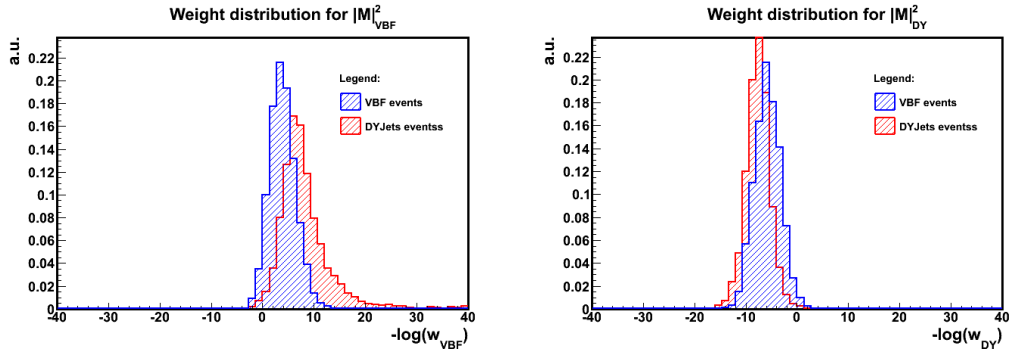


Figure 7.8: *Distribution of the $-log(w_{VBF})$ (Left) and $-log(w_{DY})$ (Right) obtained for the VBF $H \to \tau\tau$ (blue histograms) and for the DY $Z/\gamma^* \to \tau\tau$ Monte-Carlo simulated events. Only events with the τ_h decaying 1-prong+π^0s are considered.*

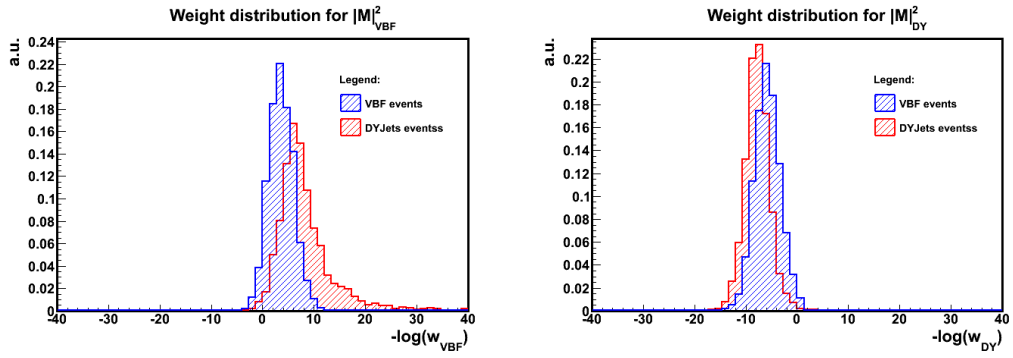


Figure 7.9: *Distribution of the $-log(w_{VBF})$ (Left) and $-log(w_{DY})$ (Right) obtained for the VBF $H \to \tau\tau$ (blue histograms) and for the DY $Z/\gamma^* \to \tau\tau$ Monte-Carlo simulated events. Only events with the τ_h decaying 3-prongs are considered.*

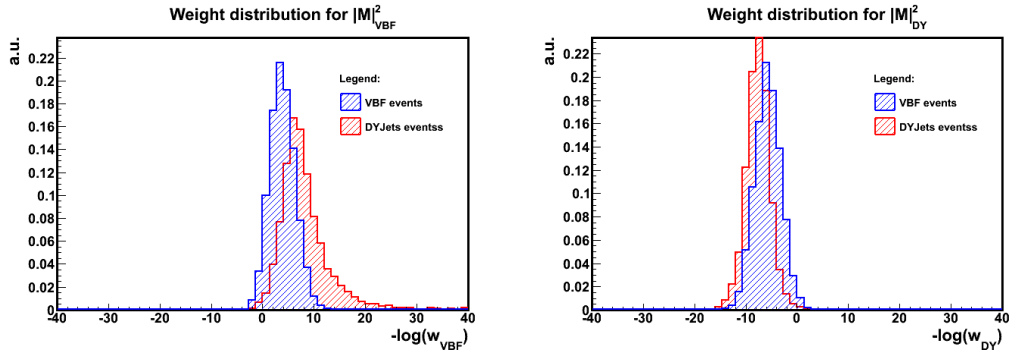


Figure 7.10: *Distribution of the $-log(w_{VBF})$ (Left) and $-log(w_{DY})$ (Right) obtained for the VBF $H \to \tau\tau$ (blue histograms) and for the DY $Z/\gamma^* \to \tau\tau$ Monte-Carlo simulated events. All the hadronic τ decay modes are considered.*

It is worth pointing out that a discrimination between the signal and background events is already clear at this stage of the analysis. In particular, it is interesting to note the different orders of magnitude between the weights evaluated under the two different hypotheses. For example, the VBF events, under the VBF matrix element hypothesis, show a $-log(w_{VBF})$ distribution that peaks at a value that is ~ 2 orders of magnitude greater than the peak value of the same $-log(w_{VBF})$ distribution for DY events (see Fig. 7.7, 7.8, 7.9 and 7.9, left plots). In addition, the distribution of $-log(w_{VBF})$ for DY events presents a longer tail in the high $-log(w_{VBF})$ region with respect to the $-log(w_{VBF})$ distribution for the VBF events. This can be interpreted as the fact that topologies very far from the VBF one (central jets, lower p_T of the leptons in the final states, small di-jet invariant mass, etc.) take a penalty when they are required to be compatible with events produced by the VBF process. Analogous considerations hold for

the $-log(w_{DY})$ distributions (see Fig. 7.7, 7.8, 7.9 and 7.9, right plots). The only difference in that case is that the "order" of the distribution with respect to the x-axis is inverted: the $-log(w_{DY})$ distribution for DY events is shifted towards higher values with respect the one for VBF events. This can be understood since in that case the events are required to be compatible with events produced by the DY process, so VBF events take a lower $-log(w_{DY})$ score. A final remark regarding plots in Figs. 7.7, 7.8, 7.9 and 7.9: it is possible to notice that the absolute mean value of the $-log(w_{VBF})$ distribution for VBF events is ~ 4 order of magnitude lower with respect to the $-log(w_{DY})$ distribution for DY events. Also this feature has a physical explanation and it is related to the ratio between the VBF $H \to \tau\tau$ production cross section and the one of DY $Z/\gamma^* \to \tau\tau$.

7.7.1 Likelihood ratio

As mentioned above, starting from the w_i obtained for the two signal and background hypotheses, it is possible to build a variable that is proportional to the probability that a particular event is generated by one of the two processes. This variable is the likelihood ratio and it is chosen to be defined (among its possible definitions) as:

$$\boxed{\mathcal{L} = \frac{w_{VBF}}{w_{VBF} + k \cdot w_{DY}}} \tag{7.21}$$

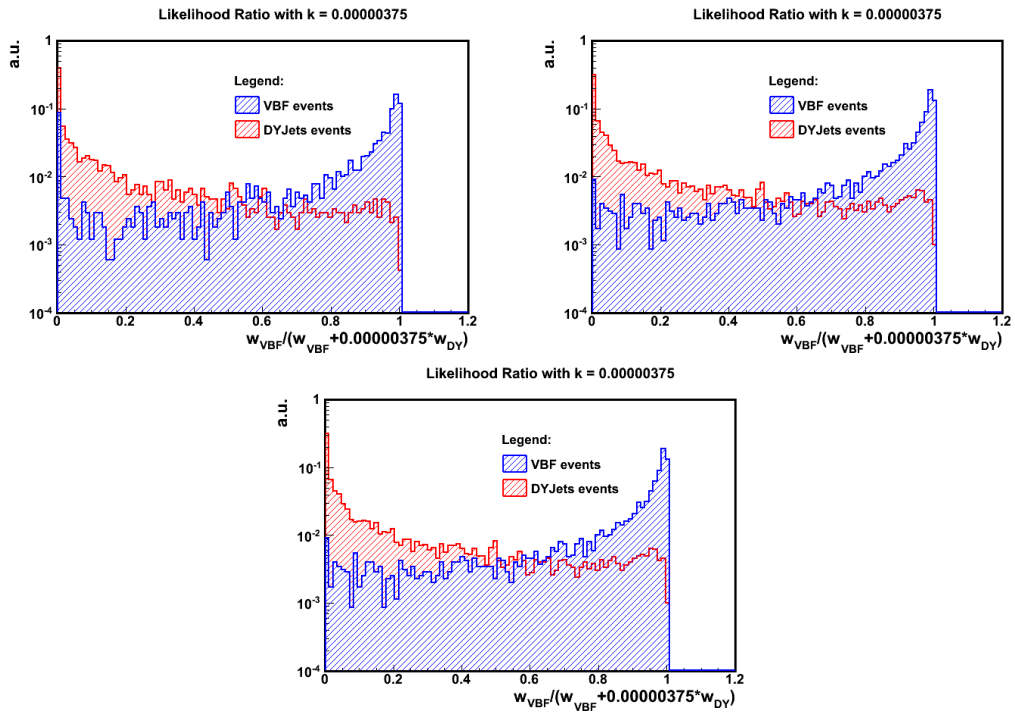


Figure 7.11: *Likelihood ratio distribution for VBF $H \to \tau\tau$ simulated events (blue) and DY $Z/\gamma^* \to \tau\tau$ simulated events (red) for τ's decays 1-prong (Up-Left), 1-prong+π^0's (Up-Right) and 3-prongs (Down-Center).*

This variable combines the information coming from the two distinct integrations discussed in the previous Section. It gives information not only about how a particular event is *VBF-like* but also how it is *not-DY-like*. The distributions for the likelihood ratio have been obtained for the different τ decay modes, as shown in Fig. 7.11. The inclusive likelihood ratio distribution is illustrated in Fig. 7.12. The factor k that appears in Eq. 7.21, should in principle be equal to the ratio between the VBF $H \to \tau\tau$ and DY $Z/\gamma^* \to \tau\tau$ production cross sections, $\frac{\sigma_{VBF}}{\sigma_{DY}}$, evaluated in the analysis acceptance. Since, it has been observed that the value of the constant k, with $k \neq 0$, does not change the signal to background discrimination, the value of k is chosen such that the likelihood ratio distribution for signal events and the one for background events cross each other at a value of $\mathcal{L} \sim 0.5$.

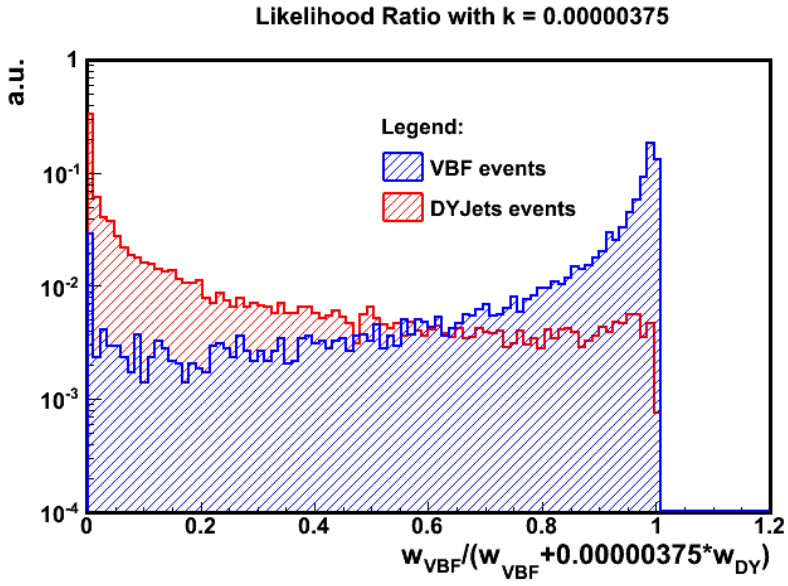


Figure 7.12: *Likelihood ratio distribution for VBF $H \to \tau\tau$ simulated events (blue) and DY $Z/\gamma^* \to \tau\tau$ simulated events (red) for τ's decaying 1-prong, 1-prong+π^0's and 3-prongs.*

The distributions in Fig. 7.12 show an impressive discrimination between VBF and DY events.

7.7.2 Comparison with previous methods (svFit)

The likelihood ratio defined in Eq. 7.21 and illustrated of Fig. 7.12 is able to discriminate events with a VBF topology from events that are produced through the DY process. In order to quantify the discrimination power of Eq. 7.21 with respect to the one achieved using the SVfit mass, the Receiver Operating Characteristic (ROC) curves have been studied. The ROC curves are obtained plotting the background reduction, i.e. the complementary probability to select a DY event, versus the signal efficiency, i.e. the probability to select a signal event. The discrimination of the MEM likelihood ratio is expected to be greater with respect to the one achieved by SVfit, since in the case of MEM, the whole event topology, included the information on the jets, is exploited. In order to carry out a fair comparison between MEM and SVfit, the events are required to pass the same baseline selections, as described in Section 7.4, with additional requirements in order to concentrate on the events that present a *VBF-like* topology. Indeed, the MEM, as developed in Chapter 6, is supposed to discriminate signal of a VBF produced Higgs boson decaying into a τ pair from the SM backgrounds (in particular from DY events), but it is not thought to exploit the other Higgs boson production mechanisms. For this reason, two further selections have been considered on top of the baseline one:

- **VBF very loose category:** two jets with a $p_T > 30\,GeV$ are required with $\Delta\eta_{jj} > 1.5$ and a di-jet invariant mass $m_{jj} > 150\,GeV$;
- **VBF loose category:** This category is defined to exactly match the VBF categories defined in the *standard* $H \to \tau\tau$ analysis as detailed in Section 7.5 (two jets with a $p_T > 30\,GeV$ are required with a $\Delta\eta_{jj} > 3.5$ and a di-jet invariant mass $m_{jj} > 500\,GeV$).

The signal and background efficiencies have been evaluated as the fraction of events passing a cut on the MEM likelihood ratio (N_{MEM}) or on the SVfit (N_{SVfit}) with respect to the total number of events considered (N_{TOT}):

$$\varepsilon = \frac{N_{MEM(SVfit)}}{N_{TOT}} \tag{7.22}$$

Given the sharpness of the MEM $\mathcal{L}$ distributions around 0 and 1, the ROC curves have been obtained using an unbinned procedure in order to avoid undesired binning effects. For what regards the efficiency estimation with the MEM $\mathcal{L}$, no further requirements, such as a quality criteria on the χ^2/dof of the w_{VBF} or w_{Dy} integrals, are required. The events for which one of the two integral computations had not converged are considered as lost, and thus, they are included in the denominator of the signal efficiency computation (Eq. 7.22) and considered as a source of inefficiency. The events that present both $w_{VBF} = 0$ and $w_{DY} = 0$, have been assigned a likelihood ratio value of zero and are considered as lost. On the contrary, the events that present only one of the two $w_i = 0$ are kept. The number of the events in the denominator of Eq. 7.22, N_{TOT}, is exactly the same for the two ROC curves. The SVfit ROC curves are obtained varying the threshold of the di-τ invariant mass (SVfit mass). The ROC curves obtained for the VBF very loose selections are illustrated in Fig. 7.13, while the ones obtained for the VBF loose selection are shown in Fig. 7.14.

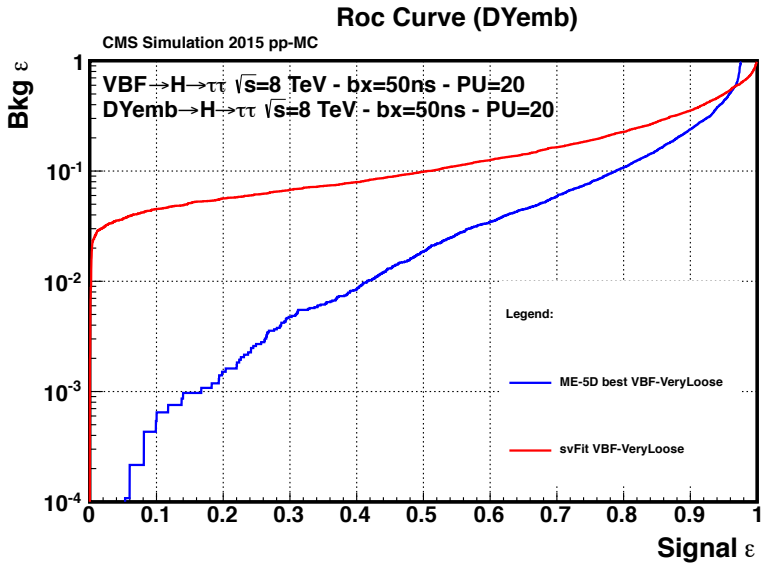


Figure 7.13: *ROC curves obtained cutting on the SVfit variable (red) and on the MEM $\mathcal{L}$ one for the VBF very loose selection criteria.*

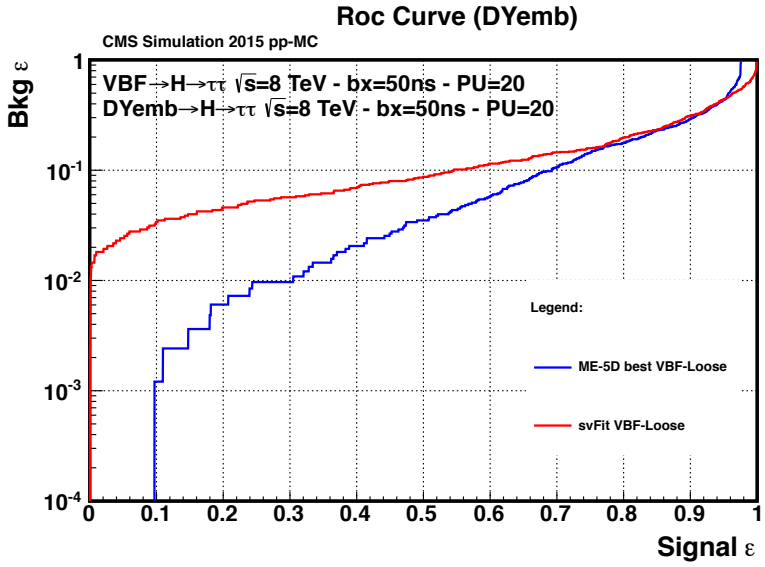


Figure 7.14: *ROC curves obtained cutting on SVfit (red) and on the MEM $\mathcal{L}$ for the VBF loose selection criteria.*

Applying both the VBF selections, it is possible to notice a small signal inefficiency ($\lesssim 5\%$)in the very high signal efficiency region for the MEM $\mathcal{L}$ ROC curves. After an investigation on

these events, it emerged that all of them present a null value for both the w_{VBF} and w_{DY} weights. However, the region between 90% and 100% is not a region of interest in the $H \to \tau\tau$ analysis since the DY background is way too high. It is possible to infer, instead, how the background rate drastically decreases in almost all the regions of the signal efficiency using the MEM $\mathcal{L}$ discriminator, compared to the background rate reduction observed using SVfit. This behaviour is even more pronounced when only the VBF very loose criteria is applied. This is because the MEM approach is able to exploit the full event information and recognize if an event is compatible with a VBF topology, while the SVfit method exploits only the information regarding the τ decay products and the observed missing transverse energy and needs to rely on event categories to target the VBF process. In the region of interest, where the signal efficiency is $\sim 40\%$ for the VBF loose selection, the MEM achieves a background rejection greater than four times the one reached by the SVfit analysis (see Fig. 7.14).

Signal extraction

The signal extraction in the published $H \to \tau\tau$ analysis [71] is performed through a binned maximum likelihood fit on the SVfit mass distribution in all the categories and in all the channels analysed, and the same approach is followed here. In general, once a discriminating variable, such as SVfit or the MEM $\mathcal{L}$, is identified, it is possible to perform the signal extraction in different ways. Indeed, it is possible to select and count the events that pass an optimized cut on the discriminating variable or, as it is done in the $H \to \tau\tau$ analysis, to fit the shape of its distribution. In order to carry out the signal extraction using the MEM $\mathcal{L}$ variable, in a first approach, it has been chosen to keep the same signal extraction as the one used in existing $H \to \tau\tau$ analysis, i.e. to fit the MEM $\mathcal{L}$ distribution. In addition, in order to make a fair comparison of the results obtained using the two methods, the same event categorization, detailed in Section 7.5, has been kept. This choice is, most probably, non optimal for the MEM approach. Indeed, the categorization adopted in the $H \to \tau\tau$ analysis makes use of some variables, like the number of the energetic jets in the event, their invariant mass and pseudorapidity, that are naturally exploited in the MEM computation of the likelihood ratio, and so, there is, in principle, no need to cut the events with a $\Delta\eta_{jj} < 3.5$ or a $m_{jj} < 500\,GeV$; it should be, indeed, up to the MEM $\mathcal{L}$ to assign to them a small probability to be originated by a VBF process. In addition, given the limited statistic available in the VBF categories, as they are defined in Section 7.5, it is not possible to perform the maximum likelihood fit to the VBF categories only. In that case, the backgrounds yields and shapes are poorly constrained and the fit shows difficulties to converge. For this reason, an *hybrid* approach has been chosen, and it has been decided to use the 0-jet and 1-jet categories to constraint the background. In these categories, the fit is still performed on the SVfit distributions. Instead, for the VBF categories, the fit is performed on the MEM $\mathcal{L}$ distribution. The results are then given in terms of combined 95% CL exclusion limits and significance for all the categories in the $\mu\tau_h$ analysis. It is possible, *a posteriori*, to extract information about the improvement achieved by using the MEM $\mathcal{L}$ in the VBF only

categories, as it is described in Section 7.7.4. Before performing the maximum likelihood fit, it is important to check, in each category, the agreement data/Monte-Carlo, in order to check if the backgrounds have been modelled properly. The so-called *prefit* plots are shown in Fig. 7.15, 7.16, 7.17, 7.18, 7.19, 7.20 and 7.21, both for the *standard* SVfit and for the upgraded SVfit+MEM analyses, while the estimation of the event yields for each background and signal process as well as for the data are collected in Tab. 7.3 and 7.4. In principle, the prefit plots obtained with the two analysis strategies should be identical. However, since the 0-jet and 1-jets categories are defined required an anti-VBF cut, a possible migration of events from the VBF categories that present problem with the integral convergence in the w_i estimation can occur.

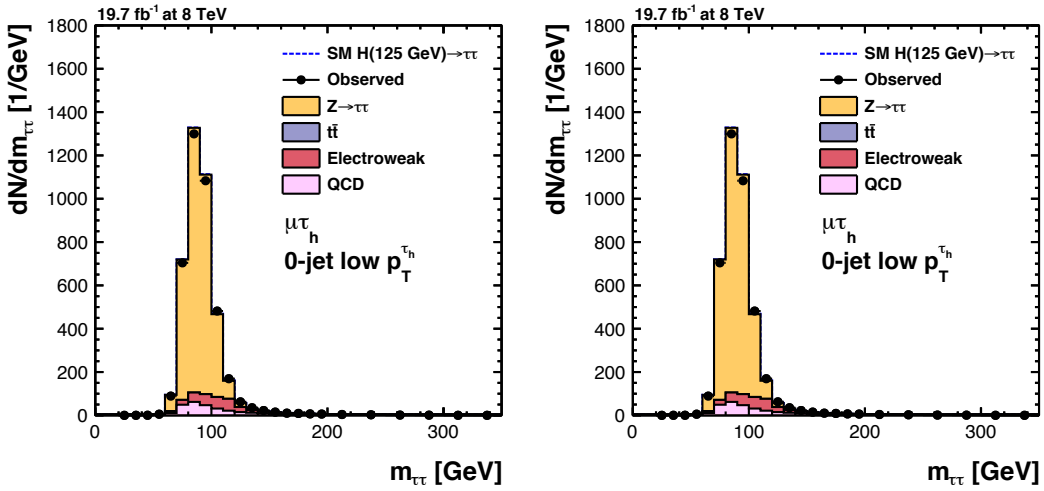


Figure 7.15: *Prefit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 0-jet-Low category.*

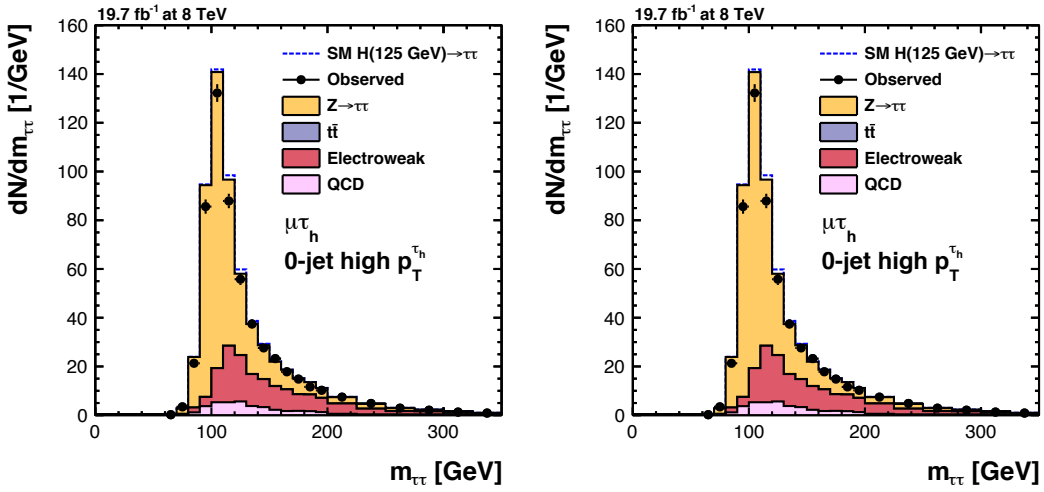


Figure 7.16: *Prefit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 0-jet-High category.*

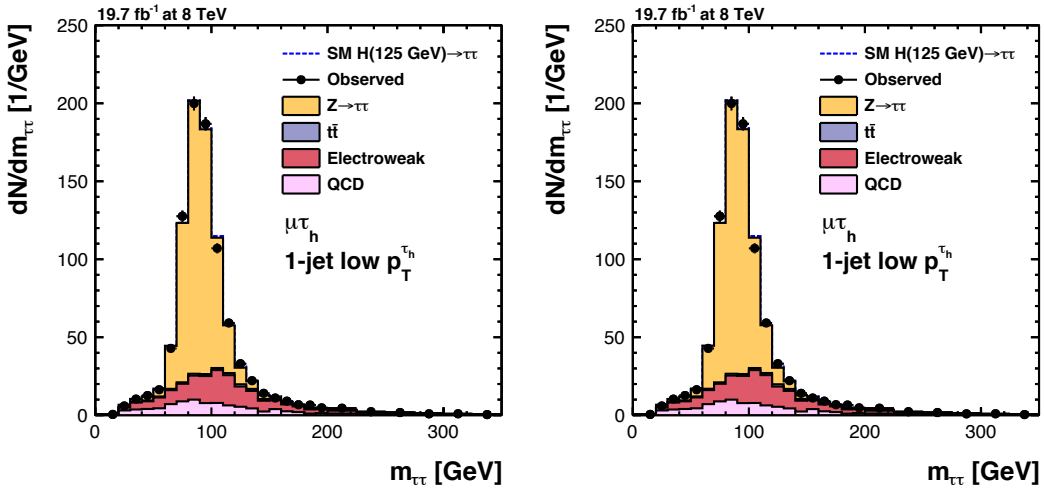


Figure 7.17: *Prefit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-Low category.*

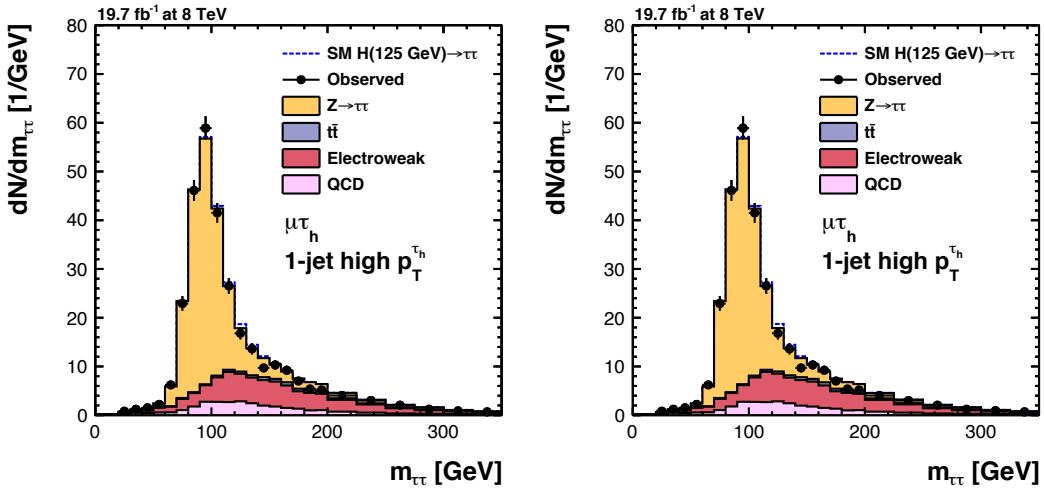


Figure 7.18: *Prefit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-High category.*

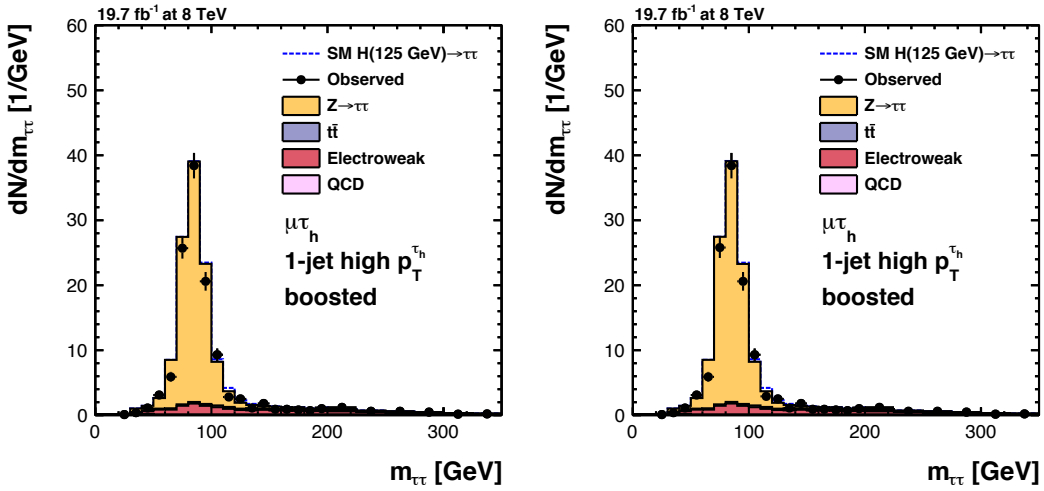


Figure 7.19: *Prefit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-High-Boosted category.*

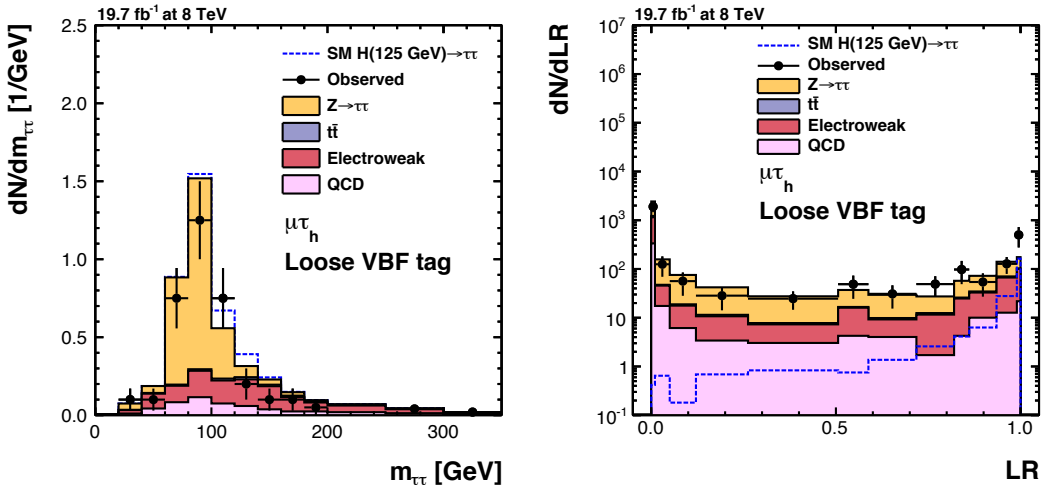


Figure 7.20: *Prefit plots of the SVfit invariant mass distribution for the SVfit analysis (Left) and of the MEM $\mathcal{L}$ for the SVfit+MEM (Right) analysis for the VBF-Loose category.*

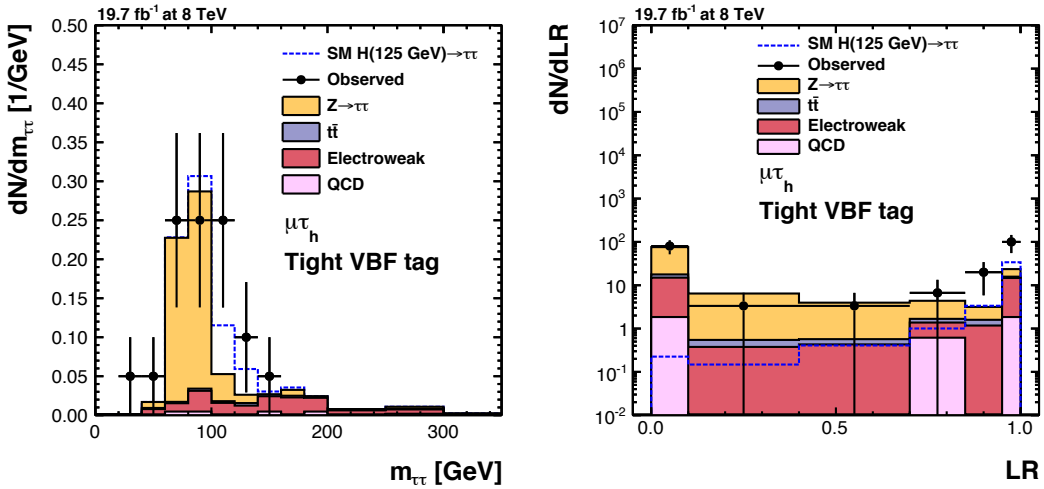


Figure 7.21: *Prefit plots of the SVfit invariant mass distribution for the SVfit analysis (Left) and of the MEM $\mathcal{L}$ for the SVfit+MEM (Right) analysis for the VBF-Tight category.*

Overall, a very satisfactory agreement is achieved in all prefit plots. In particular, the good agreement found in the 0-jet categories, where the contribution of the Higgs boson signal is negligible compared to the one of the SM backgrounds, means that the various background estimations, carried out as described in Section 7.5, model the real background shapes and yields observed in data in a proper way.

	Categories - SVfit analysis						
process	0-Jet-Low	0-Jet-High	1-Jet-Low	1-Jet-High	1-Jet-High	VBF-Loose	VBF-Tight
ggH	87.2803	69.6778	38.1375	31.5425	12.2295	1.12533	0.321448
qqH	0.805452	0.698748	6.36694	4.5556	3.11194	3.57303	2.13364
WH	0.140404	0.219677	0.998601	0.808492	0.713038	0.0118255	0
ZH	0.0773951	0.121092	0.550459	0.445665	0.393048	0.00651859	0
ZTT	34552.6	4195.75	6100.23	2059.96	1086.97	49.5093	10.6038
QCD	3043.98	417.87	964.131	337.123	17.3306	11.435	0.468394
W	2166.99	1191.65	1452.74	622.552	137.501	22.2504	3.08037
ZJ	78.0996	33.5201	129.008	44.9581	9.15696	0.306139	0
ZL	709.814	237.303	103.195	56.6016	4.5143	0.614786	0.01853
TT	3.98023	6.00157	101.561	62.8972	35.9772	2.06044	0.513359
VV	56.413	57.4561	105.789	52.5401	45.6388	0.874773	0.239408
observed	40213	5772	9023	3146	1252	71	20

Table 7.3: *Event yields for all the simulated background and signal processes and for the data obtained in the SVfit analysis.*

	Categories - SVfit + MEM analysis						
process	0-Jet-Low	0-Jet-High	1-Jet-Low	1-Jet-High	1-Jet-High	VBF-Loose	VBF-Tight
ggH	87.3037	69.6778	38.1375	31.5897	12.3017	1.10311	0.302962
qqH	0.805452	0.698748	6.40756	4.58149	3.17559	3.50161	2.0681
WH	0.138838	0.219677	0.998674	0.808752	0.717454	0.0118255	0
ZH	0.0765318	0.121092	0.550499	0.445808	0.395482	0.00651859	0
ZTT	34552.8	4195.47	6101.11	2060.22	1087.34	43.9891	9.54933
QCD	3044.07	417.87	944.104	329.231	17.5208	8.58809	0.369191
W	2166.98	1191.65	1412.61	607.379	126.965	19.5029	2.17005
ZJ	78.0996	33.5201	129.008	44.9581	9.15696	0.306139	0
ZL	709.814	237.303	103.195	56.6016	4.5143	0.614786	0.01853
TT	3.98023	6.00157	172.24	86.1704	54.2543	1.7033	0.513359
VV	56.413	57.4561	105.871	52.707	45.7259	0.716217	0.239408
observed	40213	5772	9023	3147	1254	71	18

Table 7.4: *Event yields for all the simulated backgrounds and signal processes and for the data obtained in the SVfit + MEM analysis.*

7.7.3 Results and conclusion

In the maximum likelihood fit, the nuisance parameters are free to move around their estimated best value in a range determined by the experimental uncertainties, independently in each category. The global fit over all the categories is designed to find the values for the nuisance parameters that provide the best agreement with the data. For this reason, the background shapes and yields are allowed to be slightly modified while the fit is performed. The data/Monte-Carlo agreement after the fit is visible in the *postfit* plots illustrated in Fig. 7.22, 7.23, 7.24, 7.25, 7.26, 7.27 and 7.28.

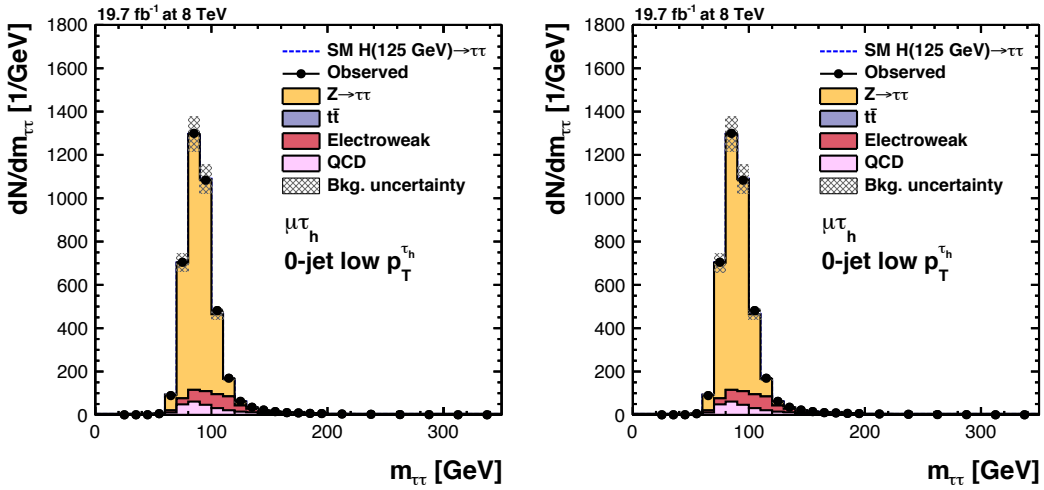


Figure 7.22: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 0-jet-Low category.*

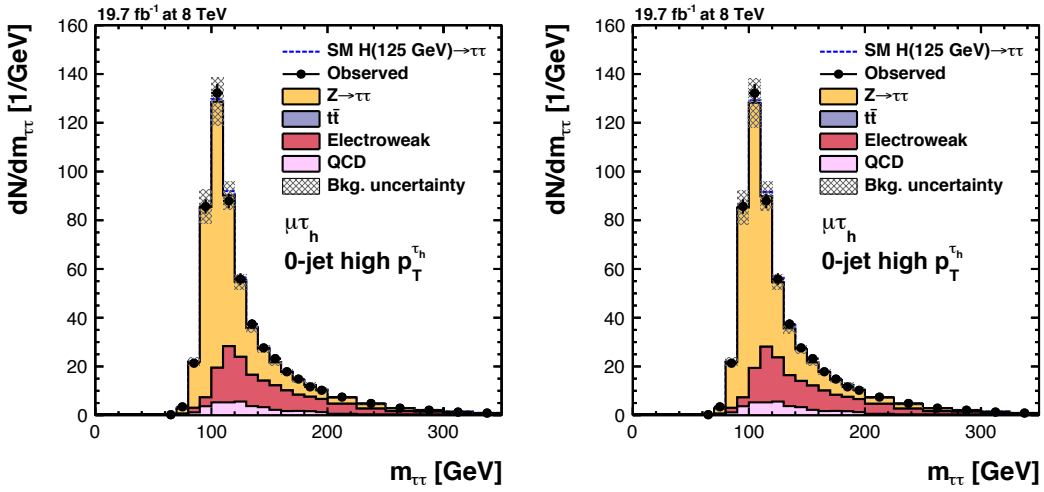


Figure 7.23: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 0-jet-High category.*

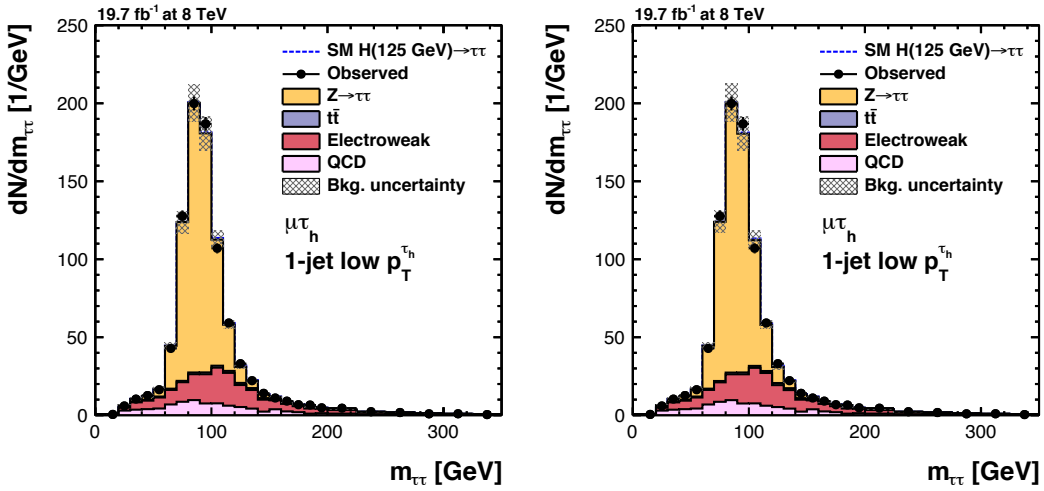


Figure 7.24: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-Low category.*

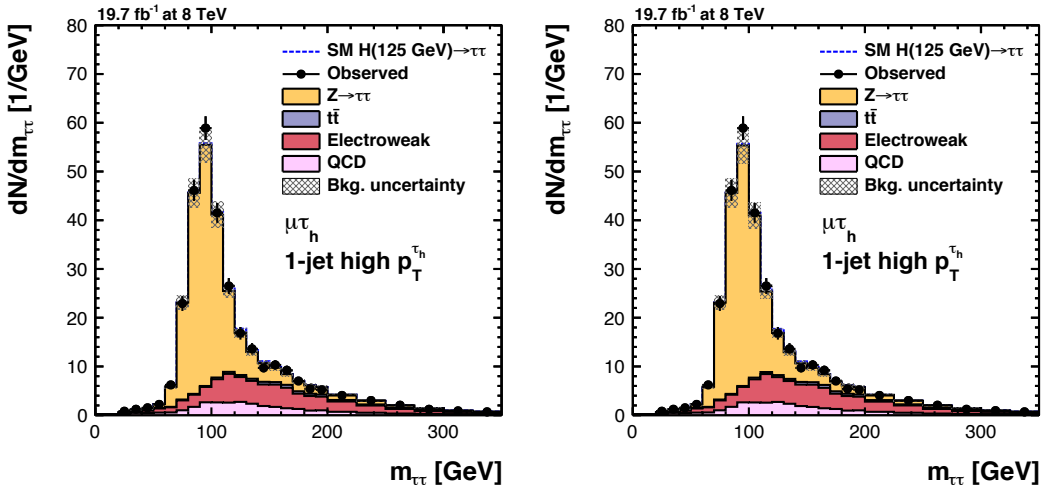


Figure 7.25: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-High category.*

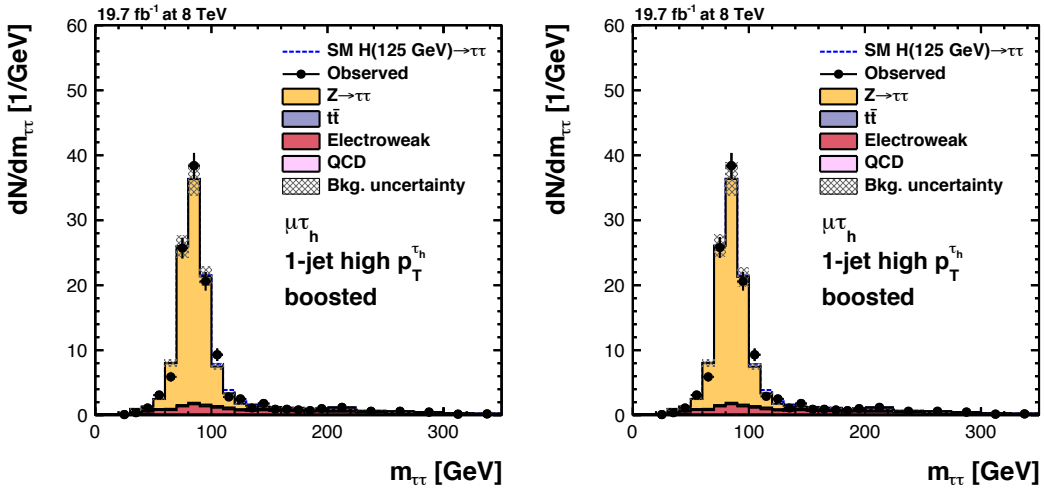


Figure 7.26: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and SVfit+MEM (Right) analysis for the 1-jet-High-Boosted category.*

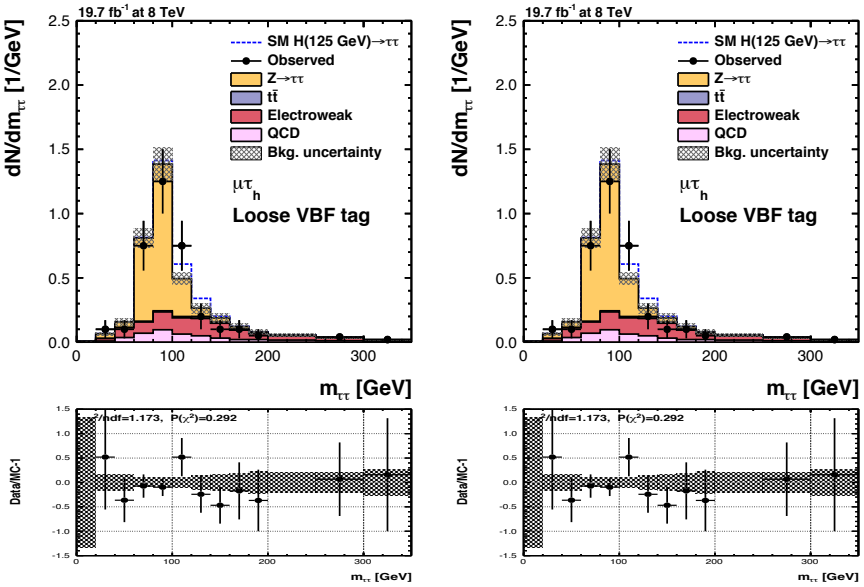


Figure 7.27: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and the Likelihood ratio distribution for the SVfit+MEM (Right) analysis for the VBF-Loose category. Below the invariant mass distribution the agreement between data and MC simulation is illustrated.*

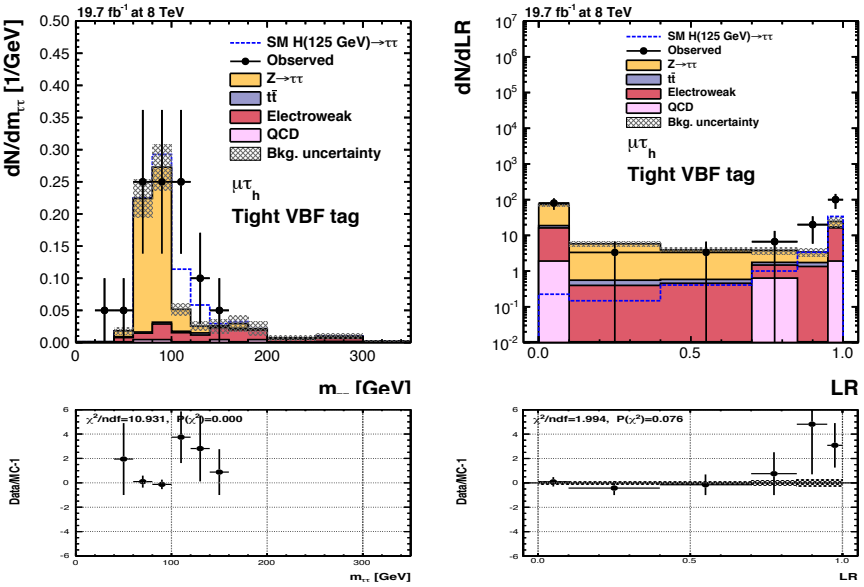


Figure 7.28: *Postfit plots of the SVfit invariant mass distribution for the SVfit (Left) and the Likelihood ratio distribution for the SVfit+MEM (Right) analysis for the VBF-Tight category. Below the invariant mass distribution the agreement between data and MC simulation is illustrated.*

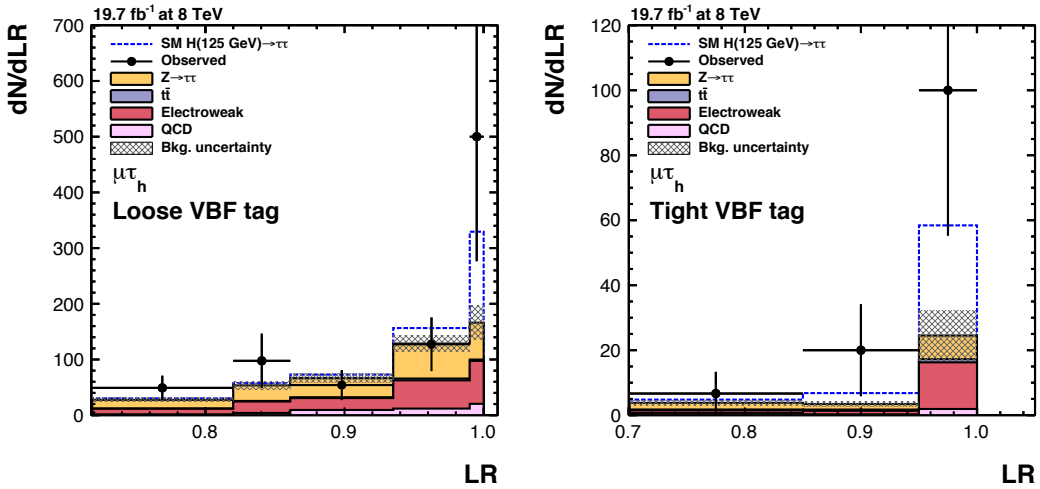
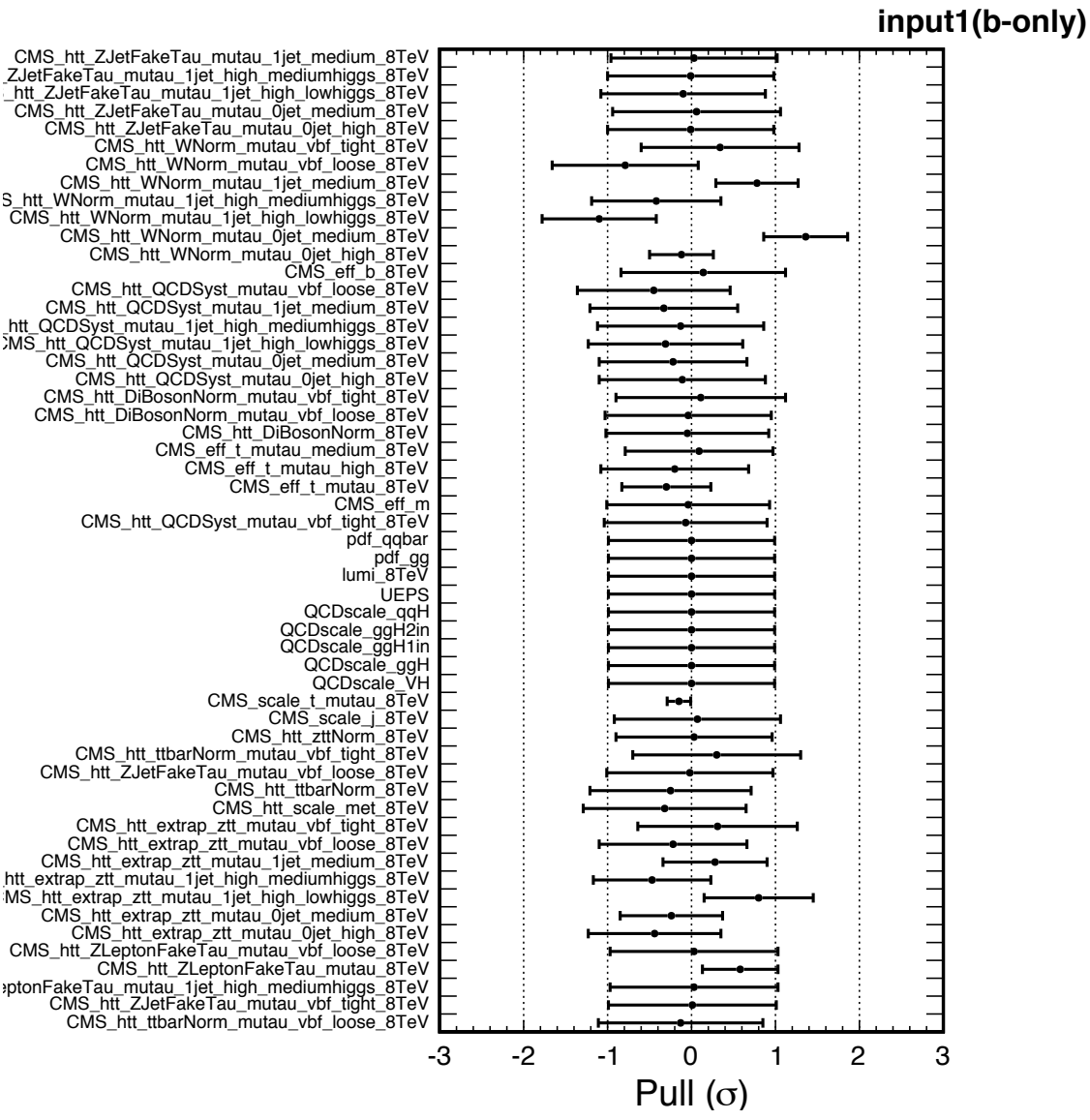


Figure 7.29: *Postfit plots of the Likelihood ratio distribution zoomed in the high purity signal region for the SVfit+MEM analysis. The distribution are illustrated for the VBF-Loose (Left) and the VBF-Tight (Right) categories.*

Figure 7.29 illustrates the bin content in the high purity signal region of the distribution shown in Fig. 7.28 (Right) and Fig. 7.27 (Left). The binning adopted in the histograms related to the Likelihood ratio distributions has been chosen in order to provide an almost equal number of entries in each bin. An excess of events is present in the last bin of the Likelihood ratio distribution in the VBF-Loose category can be noticed. This excess is even more remarkable in the VBF-Tight category. Due to *lucky* statistical fluctuation, the number of observed events, both in VBF-Loose and VBF-Tight categories, exceeds the number of events expected in the hypothesis of the signal (SM H $\to \tau\tau$) plus the SM backgrounds. As it is shown in Tab. 7.8, the signal strength modifier (μ) is expected to be greater than 1. In Tab. 7.5, the event content of the last and second-to-last bins are shown for each considered component of the SM background, for the total amount of the SM background with its uncertainties, for the expected SM H $\to \tau\tau$ signal and for the observed data, both for the VBF-Loose and VBF-Tight categories. The pulls distributions after the fit have been compared in the two analysis approaches and shown in Fig. 7.30 and 7.31. As it is possible o notice, the pulls values associated to the most important nuisance parameter are in agreement among the two analysis strategies.

	VBF-Loose		VBF-Tight	
process	2^{nd}-to-last bin	last bin	2^{nd}-to-last bin	last bin
DY	3.45	0.66	0.17	0.36
QCD	0.65	0.20	0.00	0.10
EWK	3.43	0.77	0.14	0.72
$t\bar{t}$	0.19	0.02	0.04	0.05
Total Background	7.07 ± 0.83	1.66 ± 0.30	0.34 ± 0.09	1.22 ± 0.39
SM ($H \to \tau\tau$) Signal	1.53	1.63	0.34	1.70
Observed	7	5	2	5

Table 7.5: *Postfit event yields in the two last bin in the high purity signal region in the Likelihood ratio distribution for all the simulated backgrounds and signal processes and for the data obtained in the SVfit + MEM analysis for both the VBF categories.*



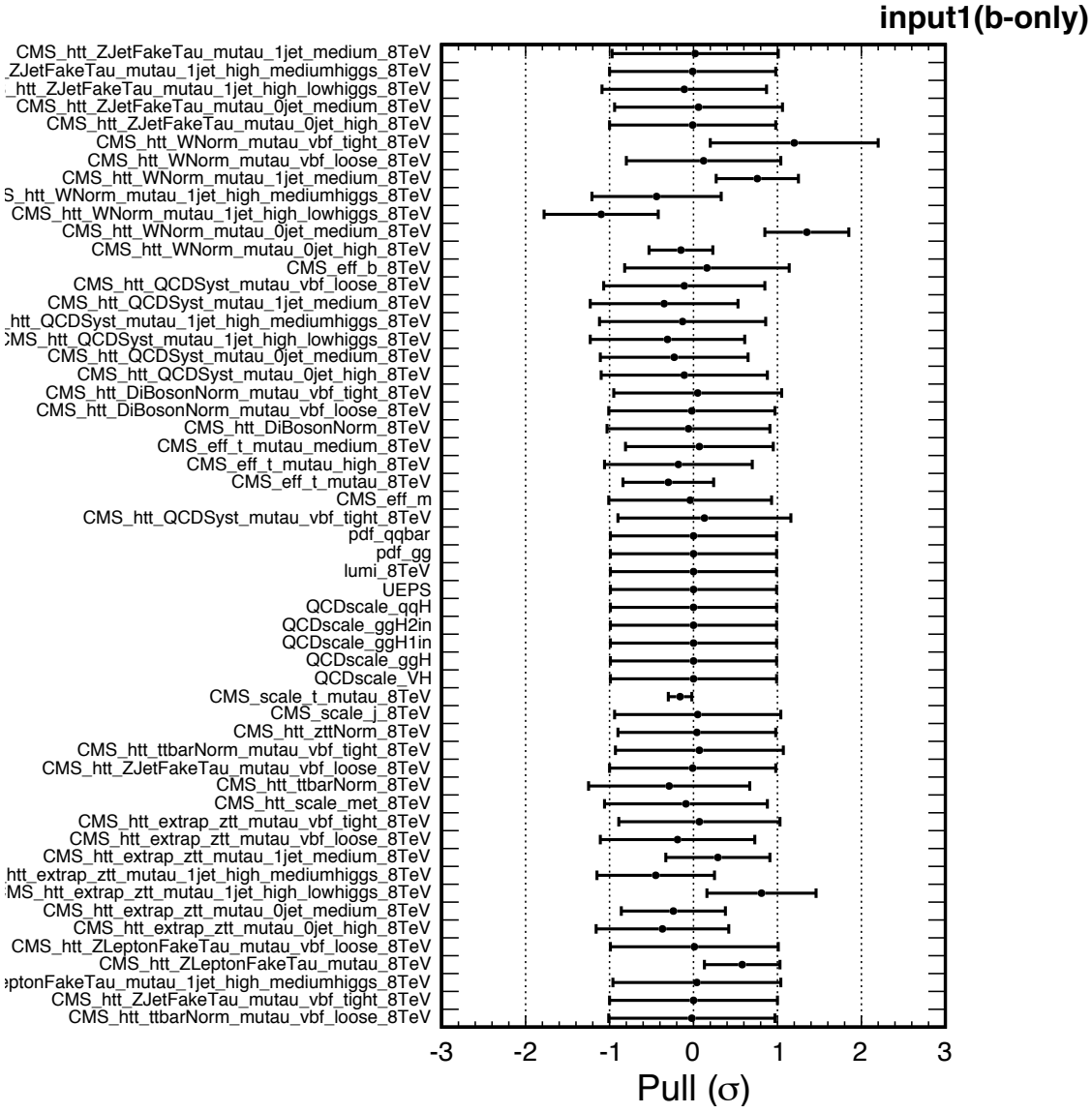


Figure 7.31: *Pulls on the nuisance parameters for the $\mu\tau_h$ channel for the SVfit + MEM analysis.*

Limits and significance

The results of the fit are finally given in terms of the expected 95% CL exclusion limits, significances and p-values for different values of the Higgs boson mass m_H hypotheses in a range from $110\,GeV$ to $145\,GeV$. Since the Higgs boson has been already discovered and its mass has been measured with a small uncertainty (a fraction of percent), the whole $H \to \tau\tau$ analysis has been optimized to search for a SM model Higgs boson with $m_H = 125\,GeV$. For this reason, one

expects that the analysis sensitivity is optimal around of $m_H = 125\,GeV$. The limits have been calculated for both the SVfit and SVfit+MEM analysis and are illustrated in Fig. 7.32. The exclusion limits set an upper value on the possibility to exclude a signal coming from an Higgs boson produced with a cross section lower than or equal to σ/σ_{SM}. The lower the exclusion limit is, the better the analysis sensitivity. The values of the 95% CL exclusion limits for each m_H are summarized in the Tab. 7.6 for the SVfit analysis and in Tab. 7.7 for the SVfit+MEM one.

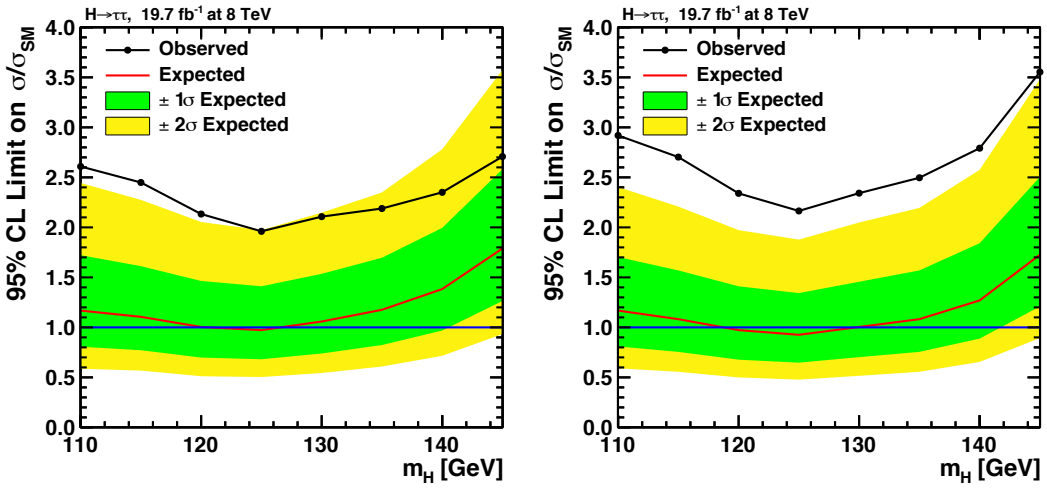


Figure 7.32: *Expected and observed 95% CL exclusion limits on the σ/σ_{SM} value for different values of the Higgs boson mass m_H for the $\mu\tau_h$ channel. The σ_{SM} represents the Higgs boson production cross section. The limits are obtained for both the SVfit (Left) and SVfit + MEM (Right) analyses approaches. The red curves represents the mean value of the expected limit, while the yellow bands represent the 65% uncertainties region and the green bands the 95% uncertainties region.*

From Fig. 7.32 it is already possible to appreciate the improvement on the value of the expected 95% CL exclusion limits brought by the introduction of the MEM approach in the $H \to \tau\tau \to \mu\tau_h$ search. This improvement is quantified and reported in the summary table 7.9. The observed limits in Fig. 7.32 are in agreement with the presence of an excess of events.

The concept of statistical significance is formally introduced in the context of hypothesis testing [149][150]. In the case of the $H \to \tau\tau$ analysis, these hypotheses are:

- H_0: the observed distributions in data is generated by the SM background processes only;
- H_1: the observed distributions in data is generated by the SM background processes plus the SM Higgs boson;

The statistical significance, as mentioned in Sec. 7.3, is strictly related to the p-value that quantifies the probability that the observed deviation from the H_0 hypothesis is due to a statisti-

95% C.L. Exclusion Limits - SVfit						
m_{H}	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Obs.
110 GeV	$5.86 \cdot 10^{-1}$	$8.07 \cdot 10^{-1}$	1.17	1.72	2.44	2.61
115 GeV	$5.68 \cdot 10^{-1}$	$7.72 \cdot 10^{-1}$	1.11	1.61	2.28	2.45
120 GeV	$5.12 \cdot 10^{-1}$	$6.98 \cdot 10^{-1}$	1.00	1.46	2.05	2.13
125 GeV	$5.03 \cdot 10^{-1}$	$6.81 \cdot 10^{-1}$	0.97	1.41	1.97	1.96
130 GeV	$5.44 \cdot 10^{-1}$	$7.39 \cdot 10^{-1}$	1.06	1.54	2.15	2.11
135 GeV	$6.09 \cdot 10^{-1}$	$8.23 \cdot 10^{-1}$	1.18	1.70	2.35	2.19
140 GeV	$7.16 \cdot 10^{-1}$	$9.68 \cdot 10^{-1}$	1.38	1.99	2.78	2.35
145 GeV	$9.33 \cdot 10^{-1}$	1.26	1.79	2.58	3.57	2.71

Table 7.6: *Values of the expected and observed 95% CL exclusion limits on σ/σ_{SM} for different values of the Higgs boson mass m_H. The σ represents the Higgs boson production cross section obtained for the SVfit analysis approach. The values of the 65% (1σ) and 95% (2σ) lower and upper limit are listed.*

95% C.L. Exclusion Limits - SVfit+MEM						
m_{H}	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Obs.
110 GeV	$5.91 \cdot 10^{-1}$	$8.10 \cdot 10^{-1}$	1.17	1.70	2.40	2.92
115 GeV	$5.56 \cdot 10^{-1}$	$7.55 \cdot 10^{-1}$	1.08	1.57	2.21	2.70
120 GeV	$5.00 \cdot 10^{-1}$	$6.77 \cdot 10^{-1}$	0.97	1.41	1.97	2.34
125 GeV	$4.77 \cdot 10^{-1}$	$6.47 \cdot 10^{-1}$	0.93	1.34	1.88	2.16
130 GeV	$5.16 \cdot 10^{-1}$	$7.03 \cdot 10^{-1}$	1.00	1.46	2.05	2.34
135 GeV	$5.56 \cdot 10^{-1}$	$7.55 \cdot 10^{-1}$	1.08	1.57	2.19	2.50
140 GeV	$6.55 \cdot 10^{-1}$	$8.88 \cdot 10^{-1}$	1.27	1.84	2.57	2.79
145 GeV	$8.94 \cdot 10^{-1}$	1.21	1.73	2.50	3.48	3.55

Table 7.7: *Values of the expected and observed 95% CL exclusion limits on σ/σ_{SM} for different values of the Higgs boson mass m_H. The σ represents the Higgs boson production cross section obtained for the SVfit + MEM analysis approach. The values of the 65% (1σ) and 95% (2σ) lower and upper limit are listed.*

cal fluctuation in the observed data under the background only hypothesis. The significance and p-value quantify the presence of an excess and their distribution are obtained from the combined fit both for the SVfit and SVfit+MEM analysis as illustrated in Fig. 7.33 and 7.34 respectively. A value of significance equal to $S = 2.14$ at $m_H = 125\,GeV$ is expected in the SVfit analysis while a value of $S = 2.31$ at $m_H = 125\,GeV$ is expected in the SVfit+MEM analysis.

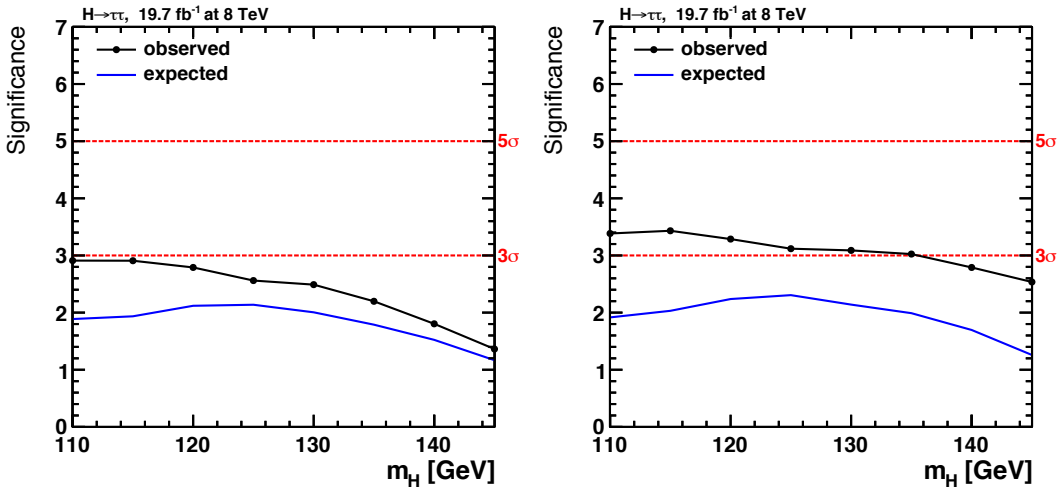


Figure 7.33: *Expected significance computed for different values of the Higgs boson mass m_H. The significance is obtained for both the SVfit (Left) and SVfit + MEM (Right) analyses approaches.*

The observed significance, in both the analysis, is higher than the expected one (see Fig. 7.27, 7.28 and 7.29), because the event excess is slightly above the expectation for the SM H $\to \tau\tau$ plus SM background hypothesis. For the SVfit + MEM case, the observed significance exceed the 3σ (3.12σ at $m_H = 125\,\text{GeV}$) leading to an evidence of the decay process H $\to \tau\tau$ in the sole $\mu\tau_h$ channel. Another way to quantify the observed excess is through the p-value distribution shown in Fig. 7.34.

	SVfit	SVfit + MEM
Observed μ at $m_H = 125\,\text{GeV}$	$1.08^{+47\%}_{-43\%}$	$1.29^{+40\%}_{-36\%}$

Table 7.8: *Signal strength modifier (μ) obtained for the SVfit and SVfit + MEM analysis for a value of $m_H = 125$ GeV.*

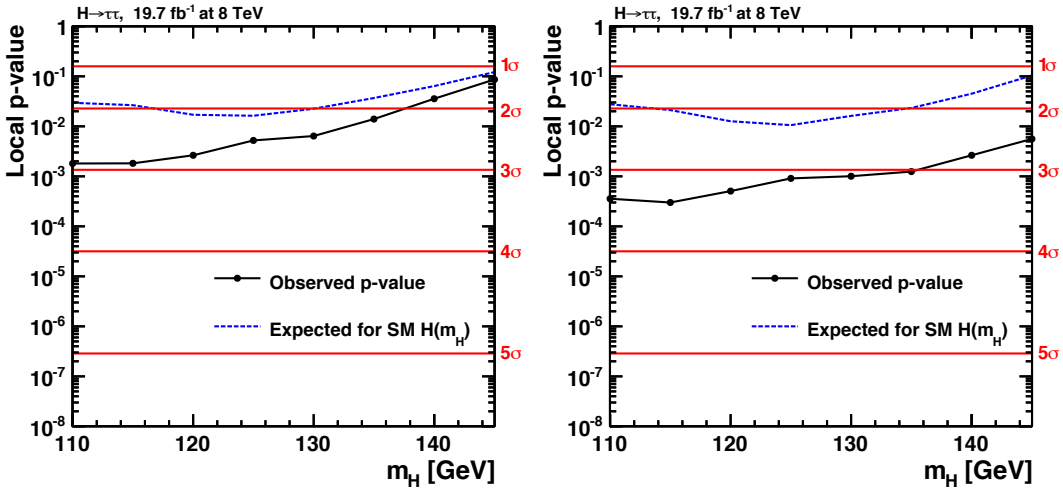


Figure 7.34: *Expected p-value computed for different values of the Higgs boson mass m_H. The p-value is obtained for both the SVfit (Left) and SVfit + MEM (Right) analysis approaches.*

As can be concluded from the exclusion limits, significance and p-value results obtained for the two different analysis approaches, the MEM brings a substantial improvements to the analysis sensitivity. The improvement brought by the MEM approach in the VBF $H \to \tau\tau \to \mu\tau_h$ channel in the full category combination is summarized in Tab. 7.10. Instead, in Tab. 7.8, the signal strength modifier has been listed both for the SVfit and the SVfit + MEM analysis, evaluated at a value of $m_H = 125\,\text{GeV}$.

Both the signal strength modifiers are compatible with 1.0, i.e., with the SM H $\to \tau\tau$ hypothesis.

Signal injection test

In order to check the consistency of the results shown in the previous section, and search for eventual bias in the limit extraction procedure, a test has been made replacing the data by the sum of the SM background and Higgs boson signal simulated through the Monte-Carlo. This procedure, called also *signal injection test*, aims at obtaining a distribution for the signal strength modifier μ as function of the different values of the Higgs boson mass from the global fit in all the categories (see Section 7.3). In this test, the mass of the injected Higgs boson signal

$H \to \tau\tau \to \mu\tau_h$ - all categories - $m_H = 125\,GeV$			
	SVfit	**SVfit+MEM**	**SVfit+MEM improvement**
95% CL excl. lim.	**0.973**	**0.926**	**4.8%**
Significance	**2.14**	**2.58**	**20.6%**

Table 7.9: *Summary of the expected 95% CL exclusion limits and significance for $m_H = 125\,GeV$ obtained for the different analysis approaches and the relative improvement brought by the SVfit + MEM analysis with respect to the SVfit one.*

is $m_H = 125\,GeV$. The distribution of the μ parameter is expected to be compatible with 1. Indeed, the injected signal is the SM higgs boson with $m_H = 125$. Indeed, it is like an artificial data distribution made by the background plus signal events is tested against the background plus signal hypothesis (H_1). The distribution of the μ values are compared through the plots shown in Fig. 7.35 for the two different analysis approaches.

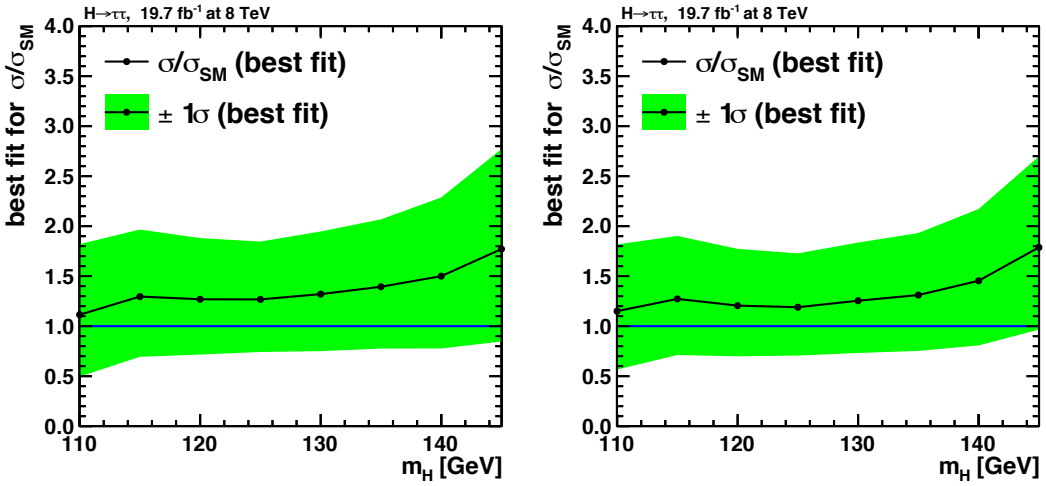


Figure 7.35: *Observed signal strength μ obtained after the signal injection for SVfit (Left) and SVfit + MEM (Right) analysis approached.*

As it is possible to deduce from Fig. 7.35, the values for the μ in the signal injection test are both compatible with 1 and the error band is reduced in the SVfit + MEM case.

7.7.4 Sensitivity estimate in VBF categories

As already mentioned in Section 7.7.2, it is possible to estimate the sensitivity of the analysis in the sole VBF categories, as well as the improvements brought by the introduction of the MEM approach in these two last categories. If the VBF-Loose and VBF-Tight categories do

not have enough statistics to properly constrain the backgrounds yields and shapes (indeed, the observed number of events in the two categories is $\sim$ 70 and $\sim$ 20, respectively) the other 5 categories (0-Jets-Low/High and 1-Jet-Low/High/HighBoosted) have enough statistics to do it. In addition, in these five categories, the SVfit mass distribution is used as variable to be fitted for both analysis approaches. Hence, it is possible to evaluate the absolute improvement on the expected 95% CL exclusion limits and the significance brought by the VBF categories for the two analysis approaches. Then, from these, it is possible to quantify the relative gain of the MEM approach with respect to the SVfit one from the ratio of the absolute difference between the limits obtained with all the categories and those obtained from the high statistics categories only. The same concepts apply in the evaluation of the expected significance improvements. The expected 95% CL exclusion limits and significance are shown for the SVfit analysis in Fig. 7.36.

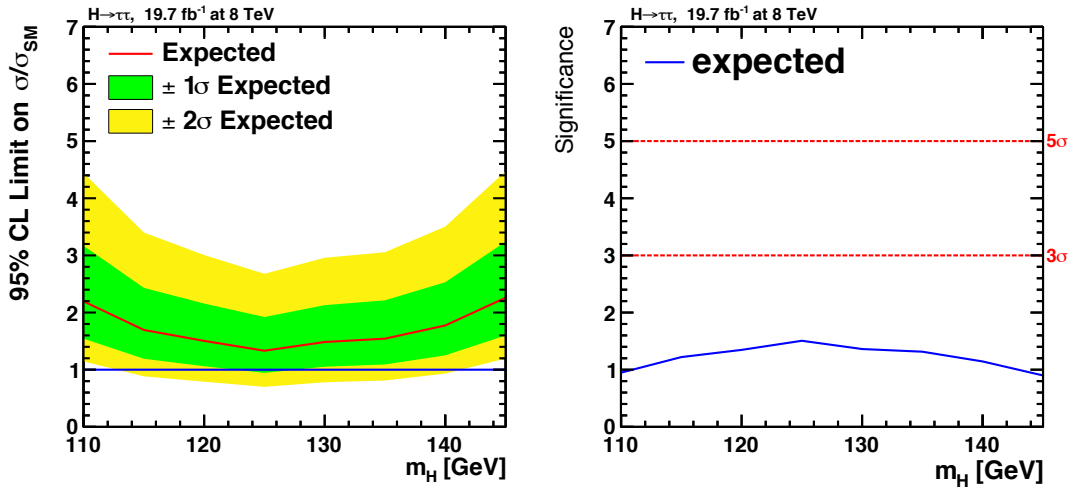


Figure 7.36: *Expected 95% CL exclusion limits (Left) and significance (Right) as function of the Higgs boson mass hypothesis $m_H = 125\,GeV$ obtained from the simultaneous fit to the 0-Jets-Low/High and 1-Jet-Low/High/HighBoosted categories only.*

Comparing the exclusion limits and significance shown in Fig. 7.36 and those illustrated in Fig 7.32 and 7.33, it is possible to summarize the results in the following table 7.10:

$\mathbf{H \to \tau\tau \to \mu\tau_h}$ **- all categories -** $\mathbf{m_H = 125\,GeV}$			
	SVfit	**SVfit+MEM**	**SVfit+MEM improvement**
Significance	**2.14**	**2.31**	**7.9%**
$\mathbf{H \to \tau\tau\mu\tau_h}$ **- 0-Jets + 1-Jet categories -** $\mathbf{m_H = 125\,GeV}$			
Significance	**1.51**	**1.51**	**-**
$\mathbf{H \to \tau\tau\mu\tau_h}$ **- VBF-Loose + VBF-Tight -** $\mathbf{m_H = 125\,GeV}$			
Significance	$\mathbf{\Delta = 0.63}$	$\mathbf{\Delta = 0.80}$	**27.0%**

Table 7.10: *Summary of the expected 95% CL exclusion limits and significance for $m_H = 125\,GeV$ obtained for the different analysis approaches and the relative improvement brought by the SVfit + MEM analysis with respect to the SVfit one. The values Δ indicates the absolute difference between the limit (significance) obtained using only the 0-Jet and 1-Jet categories and the limit (significance) obtained with the inclusion of all the seven categories.*

The improvements brought by using the MEM appear even more impressive if it is considered that the standard analysis (SVfit) would need 45% more integrated luminosity to reach the same performance as the SVFit+MEM analysis.

7.8 Conclusions and perspective

The Matrix Element Method (MEM) has been developed in the context of the VBF $H \to \tau\tau \to \mu\tau_h$ analysis and detailed in the Chapter 6. Making use of the full event kinematic information, a large separation power between the VBF $H \to \tau\tau$ signal and the main irreducible background (DY $Z/\gamma^* \to \tau\tau$) has been found observing the MEM likelihood ratio distribution. The MEM method itself has been compared to the SVfit algorithm, applied so far in the CMS $H \to \tau\tau$ analysis, and it shows a superiority of a factor 4-10 (depending on the signal efficiency working point) in the background rejection. The MEM has been applied for the first time in the $H \to \tau\tau$ analysis, in the two VBF categories. In the high statistics categories used to constrain the background yields, given the small presence of VBF events, the SVfit variable is still used since the MEM has been developed appositely to discriminate signal events produced through VBF mechanism. The results have been obtained through a maximum likelihood fit to all the analysis categories, and show a global overall improvement in the analysis expected significance of $\sim 8\%$ for a $m_H = 125\,GeV$, while, in the sole VBF categories, it reaches $\sim 30\%$. Given the results obtained in this first application of the MEM to the $H \to \tau\tau$ analysis, new developments are already taken in consideration to further improve the performance of the method. The MEM as it is applied in the $\mu\tau_h$ channel can be easily improved to treat also the $e\tau_h$ decay mode. In addition, with some calculation on the integral dimensionality reduction, it is possible to apply easily the method to ee, $\mu\mu$ and $\tau_h\tau_h$ channels. Dedicated VBF categories defined with much

looser cuts then one used in the SVfit analysis can be defined in order to optimally exploits the MEM capability to recognize events produced through the VBF mechanism. In addition, development on the method itself are envisaged, like the possibility to introduce also the matrix element for the $W + jets$ background. It implies introducing new transfer function for the jets faking τ's. Finally, in a larger prospective, an interesting development is represented by the possibility to exploit the τ polarization in the MEM. Such improvements would provide additional handle to reduce the $Z/\gamma^* \to \tau\tau$ background, for instance in the case of a ME for gluon-gluon fusion that could not exploit the jet topologies. One could foresee an application in the context of CP measurements as well. To conclude, during the LHC Run 2 the analyses will certainly concentrate more on the exclusive Higgs boson production mechanisms, and in this context, the MEM approach in the VBF $H \to \tau\tau$ will play a leading role, improving the sensitivity of the channel that already sets the tightest constraints on the VBF process as it clearly illustrated in Fig. 7.37.

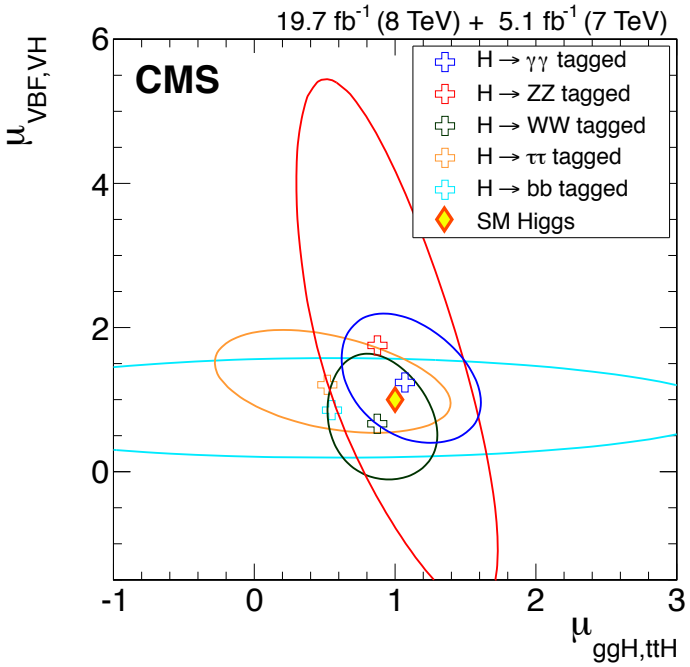


Figure 7.37: *Constraints on the signal strength (μ) for the VBF-VH and ggF-$t\bar{t}$ Higgs boson production modes obtained through the analysis of the 5 main Higgs boson decay channels. The yellow cross indicate the value for μ's for the SM Higgs bosos [151].*

Conclusion

This thesis work has been carried out in a very exiting period in the history of particle physics. The Higgs boson had been searched for for several decades, and was finally discovered on the 4th July 2012 with a mass close to the $125\,GeV/c^2$. In this context, I started my work as a member of the CMS collaboration a few months after the elusive particle was discovered through its bosonic decay channels. The coupling between the Higgs boson and the vector bosons of the Standard Model, proved the electroweak spontaneous symmetry breaking mechanism, hypothesized in 1964 by the independent works of Robert Brout, François Englert, Peter Higgs, Carl Richard Hagen, Gerald Guralnik et Thomas Kibble. This discovery was made possible thanks to the most advanced technologies developed in the field of accelerator and particle detector physics, that lead to the design, construction and operation of the Large Hadron Collider and its experiments, in particular ATLAS and CMS. This remarkable discovery led to the awarding of the Nobel Prize to Peter Higgs and François Englert, in 2013.

However, at that time, there was no evidence for the coupling between the Higgs boson and the leptons. The mechanism allowing the fermions to acquire mass was not proven. In this context, given the mass of the discovered Higgs boson, the best way to probe the Yukawa couplings relied on the search for the Higgs boson decaying into a τ-lepton pair. This search was already started when I begun my thesis, and the main goal at that time was to increase the sensitivity of the analysis in order to maximally exploit the proton-proton collision data provided by the LHC. Given the variety of the τ lepton decay modes and, in particular, the major challenge in reconstructing the τ hadronic decays, it turn out that $H \to \tau\tau \to \ell\tau_h$ represent the most sensitive decay channels. In these channels, one τ decays into a lepton, either an electron or a muon, efficiently identified and reconstructed by CMS with an optimal resolution, and the other τ decays into hadron, increasing the analysis acceptance thanks to the large branching fraction charactering this decay modes. A non-negligible amount of the τ's energy is carried by the undetectable neutrinos. The LHC environment, with a mean value of 21 hard proton-proton interaction and a bunch crossing of 50 ns, together with the hardware recording limitation of the online trigger system (L1), lead to impose high transverse momentum thresholds to the online

lepton selections, leading to a loss of $\sim 57\%$ of the possible signal events. Thus the idea to conceive a new trigger capable to select leptons with relaxed transverse momentum selections in order to recover a fraction of the events that, otherwise, would be lost. To keep the trigger rate under control, the new trigger was designed to make use of the missing transverse energy to further filter the events. I gave my personal contribution in the evaluation of the performance of such a trigger and, in particular, in the aspects related to the characterization of the missing transverse energy computed at L1. After the application of the selection based on the missing transverse energy threshold, the $\sim 47\%$ of the predicted $H \to \tau\tau$ events have been recovered for the $\mu\tau_h$ channel. Unfortunately, this trigger was introduced late in the data taking and it lead to an improvement of $\sim 2\%$ on the final exclusion limits for the search of the SM Higgs boson decaying into a τ-lepton pair.

However, the overall improvement given by the introduction of the soft-lepton trigger in the $H \to \tau\tau$ analysis was diluted into the inclusive analysis and was judged small ($\sim 2\%$). In addition, in the perspective of the LHC Run 2, started in Spring 2015, there were other challenges that limited the applicability of the approach consisting in controlling the rate trough the MET. Indeed, during the Run 2, the instantaneous luminosity of the machine will constantly increase up to $2 \times 10^{34} cm^{-2} s^{-1}$ and the energy in the center-of-mass of the proton-proton collisions has been already increased up to $13\,TeV$. In this context, for a given trigger algorithm and a given threshold, the trigger rate will be increased by a factor 6. In such extreme environment, lowering the threshold on the lepton selection can translate into the impossibility to maintain a sustainable event rate. For these reasons, during the first LHC long shutdown, it has been decided to face the problem from another perspective in order to preserve and even improve the physics program based on τ leptons acting at the trigger level. Indeed, the CMS data and acquisition system has been upgraded. The new trigger architecture was designed adopting the recent μ−TCA technology allowing the granularity of each trigger tower to be exploited at L1. My personal contribution in this project consisted in the design of a completely new τ trigger algorithm for the CMS stage 2 upgrade. This algorithm makes use of a dynamical clustering technique, developed for the upgraded e/γ algorithm. The main challenge in the design of such τ trigger came from the variety of the τ hadronic decay modes and the different particles present in the final state. In order to design the best possible algorithm able to account for the different cluster shapes originating from the variety of the τ decays, a deep study on the τ energy depositions in the CMS calorimeters has been carried out on the 2012 data. In addition, a dedicated energy calibration of the τ candidate, that treats separately the electromagnetic and hadronic part of the clusters energy, has been carried out. The isolation procedure has been also conceived in a dynamical way, as the clustering procedure. The possibility to perform an online pile up subtraction has been exploited in order to further control the trigger rate. The performance of the algorithm emulated in the 2012 data in terms of energy and angular resolutions are impressive. Indeed, the new algorithm provides the same energy resolution as

the previous ones but with the usage of the $\sim 10\%$ of the trigger towers. For what regards the angular resolution, an improvement of a factor 4 has been achieved thanks to the finer granularity accessible in the new system. The trigger threshold adopted on the di-τ trigger, used in the $H \to \tau\tau \to \tau_h\tau_h$ analysis, can be lowered down to $29\,GeV$, compared to $45\,GeV$ adopted with the previous system, increasing the acceptance of the analysis by $\sim 25\%$. One of the strong points of the new trigger algorithm is its flexibility. Indeed, its large dynamical range in term of signal efficiency for a sustainable trigger rate, allows for the usage of multiple working points suited for different possible analysis.

In the last period of my thesis, finally, I concentrated my studies on the possibility to further increase the sensitivity in the $H \to \tau\tau$ search acting at the final step of the analysis: the signal extraction procedure. In this thesis the search for the Higgs boson decaying into τ lepton pairs with the Matrix Element Method is presented. This work represents the first application of the Matrix Element Method in the $H \to \tau\tau$ analysis. I first focused on the development of the Matrix Element Method targeting the characterization of the Higgs boson events produced through the vector boson fusion mechanism, with the Higgs boson decaying into the di-τ semileptonic channels. It is, indeed, the most sensitive channels in the $H \to \tau\tau$ analysis. Moreover, the topology of the VBF production mechanism is well suited to multivariate analysis technique such as the Matrix Element Method, thanks to the relevant informations brought, not only by the di-τ system, but also by the kinematic of the high energy jets in the event. The main challenge in the $H \to \tau\tau$ analysis is to distinguish the Higgs boson resonant peak in the di-τ invariant mass distribution from the one generated by events arising from DY $Z/\gamma^* \to \tau\tau$ events, representing the main irreducible background of the analysis. Thus, the Matrix Element Method is used to test the hypothesis that an event has been generated by a Higgs boson or by a DY process, assigning a probability value to each event considered. The results obtained for the $\mu\tau_h$ channels have been compared to those obtained in the CMS analysis. In particular, a direct comparison in term of signal to background separation power obtained through the Matrix Element Method and the di-τ invariant mass reconstruction method used so far in CMS, the SVfit algorithm, has been carried out using Monte Carlo simulated events. For a given signal efficiency, the background rejection using the Matrix Element Method can be improved by an order of magnitude. The application of the Matrix Element Method increased the inclusive analysis sensitivity in the $\mu\tau_h$ channel of $\sim 8\%$ in term of expected significance and, considering the sole VBF categories, this improvement reaches $\sim 27\%$. The Matrix Element Method can be further extended in order to consider other τ decay modes, in particular the $e\tau_h$, $\tau_h\tau_h$, ee and $\mu\mu$ final state and to target other production mechanism, e.g. the gluon-gluon fusion process. Moreover, other matrix element related to other background processes other then the Drell-Yan can be considered.

Recently, the combination of the results obtained by the ATLAS and CMS Collaborations in

the search for the SM Higgs boson decaying into a τ-lepton pair, lead to an observation of such a process with a significance greater than 5σ. However, no standalone observation, by none of the two collaborations, has been claimed yet. In this context, the Matrix Element Method can play a leading role in CMS and, looking even farther, it can contribute during the LHC Run 2 to a precise measurement either of the Yukawa coupling between the Higgs boson and the τ lepton and of a characterization of the exclusive Higgs boson production mechanisms.

Bibliography

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

[2] S.L. Glashow. Partial Symmetries of Weak Interactions. *Nucl.Phys.*, 22:579–588, 1961.

[3] F. J. Hasert et al. Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment. *Phys. Lett.*, B46:138–140, 1973.

[4] G. Arnison et Al. Further evidence for charged intermediate vector bosons at the sps collider. *Physics Letters B*, 129(34):273 – 282, 1983.

[5] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group. Precision electroweak measurements on the Zresonance. *Phys.Rept.*, 427:257–454, 2006. 302 pages, v2: minor corrections and updates of references. Accepted for publication by Physics Reports, v3: further small corrections and journal version Report-no: CERN-PH-EP/2005-041, SLAC-R-774.

[6] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[7] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[8] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[10] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

[11] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967.

[12] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[13] Georges 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.*, B716:1–29, 2012.

[14] Amos Breskin and Rdiger Voss. *The CERN Large Hadron Collider: Accelerator and Experiments*. CERN, Geneva, 2009.

[15] CMS Collaboration. *The CMS Physics Technical Design Report* - volume 2 - physics performance. *J. Phys. G: Nucl. Part. Phys. 34 CMS TDR 8.2*, page 995, 2007.

[16] G. Weiglein et al. Physics interplay of the LHC and the ILC. *Phys.Rept.*, 426:47–358, 2006.

[17] Yu. A. Golfand and E. P. Likhtman. Extension of the Algebra of Poincare Group Generators and Violation of p Invariance. *JETP Lett.*, 13:323–326, 1971. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[18] J. Wess and B. Zumino. Supergauge transformations in four dimensions. *Nuclear Physics B*, 70(1):39 – 50, 1974.

[19] S. Ferrara and B. Zumino. Supergauge Invariant Yang-Mills Theories. *Nucl. Phys.*, B79:413, 1974.

[20] R. Barbieri, S. Ferrara, and C.A. Savoy. Gauge models with spontaneously broken local supersymmetry. *Physics Letters B*, 119(4):343 – 347, 1982.

[21] Savas Dimopoulos and Greg L. Landsberg. Black holes at the LHC. *Phys. Rev. Lett.*, 87:161602, 2001.

[22] Steven B. Giddings and Scott D. Thomas. High-energy colliders as black hole factories: The End of short distance physics. *Phys. Rev.*, D65:056010, 2002.

[23] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Gttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[24] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.

[25] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214 – 215, 1964.

[26] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365 – 368, 1973.

[27] J. Goldstone. Field theories with superconductor solutions. *Il Nuovo Cimento (1955-1965)*, 19:154–164, 1961. 10.1007/BF02812722.

[28] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, Aug 1962.

[29] Alessandro Strumia and Francesco Vissani. Neutrino masses and mixings and... 2006.

[30] Makoto Kobayashi and Toshihide Maskawa. *cp*-violation in the renormalizable theory of weak interaction. *Progress of Theoretical Physics*, 49(2):652–657, 1973.

[31] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008.

[32] Urs M. Heller, Markus Klomfass, Herbert Neuberger, and Pavlos M. Vranas. Numerical analysis of the Higgs mass triviality bound. *Nucl. Phys.*, B405:555–573, 1993.

[33] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[34] Vardan Khachatryan et al. Constraints on the higgs boson width from off-shell production and decay to z-boson pairs. *Physics Letters B*, 736:64 – 85, 2014.

[35] Georges Aad et al. Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector. *Eur. Phys. J.*, C75(7):335, 2015.

[36] *https* : *//twiki.cern.ch/twiki/bin/view/lhcphysics/crosssections*.

[37] R. Barate et al. Search for the standard model higgs boson at lep. *Phys.Lett.*, B565:61–75, 2003.

[38] S. Dittmaier and M. Schumacher. The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate. *Prog. Part. Nucl. Phys.*, 70:1–54, 2013.

[39] Ben Kilminster. Higgs searches at the tevatron. *PoS*, ICHEP2010:544, 2010.

[40] The T.E.V.N.P.H. Working Group. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data. In *Proceedings, 35th International Conference on High energy physics (ICHEP 2010)*, 2010.

[41] M. Spira, A. Djouadi, D. Graudenz, and P.M. Zerwas. Higgs boson production at the LHC. *Nucl.Phys.*, B453:17–82, 1995.

[42] A. Djouadi, M. Spira, and P. M. Zerwas. Production of Higgs bosons in proton colliders: QCD corrections. *Phys. Lett.*, B264:440–446, 1991.

[43] Robert V. Harlander. Virtual corrections to $gg \to$ H to two loops in the heavy top limit. *Phys. Lett.*, B492:74–80, 2000.

[44] S. Dawson. Radiative corrections to Higgs boson production. *Nucl. Phys.*, B359:283–300, 1991.

[45] Stefano Catani, Daniel de Florian, Massimiliano Grazzini, and Paolo Nason. Soft gluon resummation for Higgs boson production at hadron colliders. *JHEP*, 07:028, 2003.

[46] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati. NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders. *Phys. Lett.*, B670:12–17, 2008.

[47] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini. Two loop light fermion contribution to Higgs production and decays. *Phys. Lett.*, B595:432–441, 2004.

[48] Julien Baglio and Abdelhak Djouadi. Higgs production at the lHC. *JHEP*, 03:055, 2011.

[49] R.N. Cahn and Sally Dawson. Production of very massive higgs bosons. *Physics Letters B*, 136(3):196 – 200, 1984.

[50] T. Han, G. Valencia, and S. Willenbrock. Structure-function approach to vector-boson scattering in *pp* collisions. *Phys. Rev. Lett.*, 69:3274–3277, Dec 1992.

[51] Matteo Cacciari, Frdric A. Dreyer, Alexander Karlberg, Gavin P. Salam, and Giulia Zanderighi. Fully differential VBF Higgs production at NNLO. 2015.

[52] Evidence for a new state in the search for the standard model higgs boson in the h to zz to 4 leptons channel in pp collisions at sqrt(s) = 7 and 8 tev. 2012.

[53] Higgs to ZZ to 2l 2nu. Technical Report CMS-PAS-HIG-12-023, CERN, Geneva, 2012.

[54] Search for the standard model higgs boson in the decay channel h to zz(*) to $q\bar{q}\ l^-l^+$ at cms. 2011.

[55] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[56] Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at $\sqrt{s} = 7$ and 8 TeV. 2013.

[57] Evidence for a new state decaying into two photons in the search for the standard model Higgs boson in pp collisions. Technical Report CMS-PAS-HIG-12-015, CERN, Geneva, 2012.

[58] Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel. (CMS-PAS-HIG-13-001), 2013.

[59] Vardan Khachatryan et al. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *Eur. Phys. J.*, C74(10):3076, 2014.

[60] Evidence for a particle decaying to W+W- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at the LHC. Technical Report CMS-PAS-HIG-13-003, CERN, Geneva, 2013.

[61] Search for the Standard Model Higgs boson in the H to WW to lvjj decay channel. Technical Report CMS-PAS-HIG-12-021, CERN, Geneva, 2012.

[62] Search for the Standard Model Higgs boson in the H to WW to lvjj decay channel. Technical Report CMS-PAS-HIG-12-003, CERN, Geneva, 2012.

[63] Search for WH to 3 leptons. Technical Report CMS-PAS-HIG-11-034, CERN, Geneva, 2012.

[64] VH with H→WW→ $\ell\nu\ell\nu$ and V→jj. Technical Report CMS-PAS-HIG-12-014, CERN, Geneva, 2012.

[65] Serguei Chatrchyan et al. Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks. *Phys. Rev.*, D89(1):012003, 2014.

[66] Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at sqrts = 7 and 8 TeV. Technical Report CMS-PAS-HIG-15-002, CERN, Geneva, 2015.

[67] Vardan Khachatryan et al. Search for a standard model-like Higgs boson in the $^+$ and e^+e decay channels at the LHC. *Phys. Lett.*, B744:184–207, 2015.

[68] Tilman Plehn, David Rainwater, and Dieter Zeppenfeld. Determining the structure of higgs couplings at the cern large hadron collider. *Phys. Rev. Lett.*, 88:051801, Jan 2002.

[69] A. Djouadi, R. M. Godbole, B. Mellado, and K. Mohan. Probing the spin-parity of the Higgs boson via jet kinematics in vector boson fusion. *Phys. Lett.*, B723:307–313, 2013.

[70] D. Rainwater, D. Zeppenfeld, and K. Hagiwara. Searching for $h \to \tau\tau$ in weak boson fusion at the cern lhc. *Phys. Rev. D*, 59:014037, Dec 1998.

[71] Serguei Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons. *JHEP*, 1405:104, 2014.

[72] J. D. Bjorken. Rapidity gaps and jets as a new-physics signature in very-high-energy hadron-hadron collisions. *Phys. Rev. D*, 47:101–113, Jan 1993.

[73] J. D. Bjorken. *Int. J. Mod. Phys. A*, 7:4189, 1992.

[74] C. Amsler. Review of particle physics. *Physics Letters B*, 667(15):1 – 6, 2008. Review of Particle Physics.

[75] B.K. Bullock, K. Hagiwara, and A.D. Martin. Tau polarization and its correlations as a probe of new physics. *Nuclear Physics B*, 395(3):499 – 533, 1993.

[76] Search for WH in Final States with Electrons, Muons, Taus. Technical Report CMS-PAS-HIG-12-006, CERN, Geneva, 2012.

[77] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[78] Martina Malberti. *"Prospects for the precision determination of the W boson mass with the CMS detector at the LHC". [CMS AN -2009/098].* PhD thesis, A.A 2006-2007.

[79] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[80] Petra Merkel. Cms tracker performance. Technical Report CMS-CR-2012-102. CERN-CMS-CR-2012-102, CERN, Geneva, May 2012.

[81] CMS Collaboration. ECAL-TDR CERN/LHCC 97-33 CMS TDR 4. *Nature*, 15 December 1997.

[82] Vardan Khachatryan et al. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. 2015.

[83] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. Submitted to *JINST*, 2015.

[84] M. Anfreville, D. Bailleux, J.P. Bard, A. Bornheim, C. Bouchand, E. Bougamont, M. Boyer, R. Chipaux, V. Daponte-Puill, M. Dejardin, J.L. Faure, P. Gras, P. Jarry, C. Jeanney, A. Joudon, J.P. Pansart, Y. Penichot, J. Rander, J. Rolquin, J.M. Reymond, J. Tartas, P. Venault, P. Verrecchia, L. Zhang, K. Zhu, and R.Y. Zhu. Laser monitoring system for the {CMS} lead tungstate crystal calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 594(2):292 – 320, 2008.

[85] *https* : *//twiki.cern.ch/twiki/pub//cmspublic/ecaldpgresultscmsdp*2012027/.

[86] P Adzic. Energy resolution of the barrel of the cms electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004, 2007.

[87] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013.

[88] CMS Collaboration. Performance of the cms hadron calorimeter with cosmic ray muons and lhc beam data. *Journal of Instrumentation*, 5(03):T03012, 2010.

[89] S. Abdullin at al. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *European Physical Journal C*, 55:159–171, 2008.

[90] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[91] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. 2000.

[92] Nicol Cartiglia. Measurement of the proton-proton total, elastic, inelastic and diffractive cross sections at 2, 7, 8 and 57 TeV. 2013.

[93] *https* : *//twiki.cern.ch/twiki/bin/view/cmspublic/physicsresultsjme*2012*jec*.

[94] H. Videau. Energy flow or Particle flow - The technique of "energy flow" for pedestrians. In J.-C. Videau, H.; Brient, editor, *International Conference on Linear Colliders - LCWS04*, pages 105–120, Paris, France, April 2004. Ecole Polytechnique Palaiseau.

[95] Performance of the aleph detector at lep. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 360(3):481 – 506, 1995.

[96] CMS Collaboration. Particle–flow event reconstruction in CMS and performance for jets, taus, and MET. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[97] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *J. Instrum.*, 9(arXiv:1405.6569. CERN-PH-EP-2014-070. CMS-TRK-11-001):P10009. 80 p, May 2014. Comments: Replaced with published version. Added journal reference and DOI.

[98] CMS Collaboration. Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector. CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[99] *https* : *//twiki.cern.ch/twiki/bin/view/cmspublic/lumipublicresults*.

[100] Pierre Billoir. Progressive track recognition with a kalman-like fitting procedure. *Computer Physics Communications*, 57(13):390 – 394, 1989.

[101] P. Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the kalman filtering method. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 294(12):219 – 228, 1990.

[102] Rainer Mankel. A concurrent track evolution algorithm for pattern recognition in the hera-b main tracking system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(2):169 – 184, 1997.

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

[104] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

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

[106] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.

[107] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011.

[108] https://twiki.cern.ch/twiki/bin/viewauth/cms/newjecplots13.

[109] S. Chatrchyan. Missing transverse energy performance of the CMS detector. *J. Instrum.*, 6(arXiv:1106.5048. CMS-JME-10-009. CERN-PH-EP-2011-051):P09001. 56 p, Jun 2011.

[110] The CMS collaboration. Performance of the cms missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ tev. *Journal of Instrumentation*, 10(02):P02006, 2015.

[111] Study of tau reconstruction algorithms using pp collisions data collected at sqrt(s) = 7 TeV. Technical Report CMS-PAS-PFT-10-004, 2010.

[112] Tau identification in CMS. *CMS-PAS-TAU-11-001*, 2011.

[113] "CMS Collaboration The Tau Physics Object Group". "performance of tau reconstruction and identification in pp collisions at $\sqrt{s} = 8\,tev$". *"CMS Analysis Note"*, "AN-14-008", "2014".

[114] D L Rainwater, Dieter Zeppenfeld, and K Hagiwara. Searching for $H \to \tau\tau$ in weak boson fusion at the LHC. *Phys. Rev. D*, 59(hep-ph/9808468. MAD-PH-1057. 1):14037. 32 p, Aug 1998.

[115] E. K. Friis J. Conway and C. Veelken. Likelihood based mass reconstruction in events containing tau leptons. *CMS Note*, 2011.

[116] G Peter Lepage. A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics*, 27(2):192 – 203, 1978.

[117] G Peter Lepage. Vegas-an adaptive multi-dimensional integration program. Technical report, 1980.

[118] S. Jadach, Z. Was, R. Decker, and Johann H. Kuhn. The tau decay library TAUOLA: Version 2.4. *Comput. Phys. Commun.*, 76:361–380, 1993.

[119] Nadir Daci. *Electron selection and search for the Higgs boson decaying into tau leptons pairs with the CMS detector at the LHC.* Theses, Université Paris Sud - Paris XI, October 2013.

[120] "CMS Collaboration". "investigating the performance of the level 1 energy sum triggers, et, ht, etmissmiss and htmiss, at the cms experiment". *"CMS Analysis Note"*, "AN-2010-263", "2010".

[121] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, CERN, Geneva, Jun 2013. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[122] P. Klabbers et al. CMS level-1 upgrade calorimeter trigger prototype development. *JINST*, 8:C02013, 2013.

[123] P. Klabbers et al. CMS calorimeter trigger Phase I upgrade. *JINST*, 7:C01046, 2012.

[124] Jean-baptiste Sauvan. Performance and upgrade of the CMS electromagnetic calorimeter trigger for Run II. Technical Report CMS-CR-2014-068, CERN, Geneva, May 2014.

[125] J C Estrada-Vigil and T Ferbel. *Maximal Use Of Kinematic Information For The Extraction Of The Mass Of The Top Quark In Single-lepton Tt Bar Events At Do.* PhD thesis, Rochester Univ., Rochester, NY, 2001.

[126] V. M. Abazov et al. A precision measurement of the mass of the top quark. *Nature*, 429:638–642, 2004.

[127] Francois Le Diberder. *Study of the annihilation reaction e+ e- into four leptons.* PhD thesis, Orsay, LAL, 1988.

[128] Search for ttH production using the Matrix Element Method. Technical Report CMS-PAS-HIG-14-010, CERN, Geneva, 2014.

[129] Igor Volobouev. Matrix Element Method in HEP: Transfer Functions, Efficiencies, and Likelihood Normalization. 2011.

[130] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 231(694-706):289–337, 1933.

[131] Edgar Gabriel, Graham E. Fagg, George Bosilca, Thara Angskun, Jack J. Dongarra, Jeffrey M. Squyres, Vishal Sahay, Prabhanjan Kambadur, Brian Barrett, Andrew Lumsdaine, Ralph H. Castain, David J. Daniel, Richard L. Graham, and Timothy S. Woodall. Open MPI: Goals, concept, and design of a next generation MPI implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary, September 2004.

[132] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *JHEP*, 0711:070, 2007.

[133] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *JHEP*, 1006:043, 2010.

[134] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. Nlo higgs boson production via gluon fusion matched with shower in powheg. *JHEP*, 0904:002, 2009.

[135] Paolo Nason and Carlo Oleari. Nlo higgs boson production via vector-boson fusion matched with shower in powheg. *JHEP*, 1002:037, 2010.

[136] Torbjorn Sjostrand, Leif Lonnblad, Stephen Mrenna, and Peter Z. Skands. Pythia 6.3 physics and manual. 2003.

[137] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. 2011.

[138] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. 2012.

[139] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013.

[140] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011.

[141] Z. Was. Tauola the library for tau lepton decay, and kkmc / koralb / koralz /... status report. *Nucl.Phys.Proc.Suppl.*, 98:96–102, 2001.

[142] Czyczula, Z., Przedzinski, T., and Was, Z. Tauspinner program for studies on spin effect in tau production at the lhc. *Eur. Phys. J. C*, 72(4):1988, 2012.

[143] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC. *JHEP*, 11:064, 2011.

[144] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini. Higgs boson production at the LHC: transverse momentum resummation effects in the H $\to 2\gamma$, H $\to WW \to l\nu l\nu$ and H $\to ZZ \to 4l$ decay modes. *JHEP*, 06:132, 2012.

[145] Ivo Nicolas Naranjo Fong. *Tau lepton reconstruction and search for Higgs bosons decaying to tau pairs in the CMS experiment at the LHC.* Theses, Ecole Polytechnique, October 2014.

[146] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys.*, G28:2693–2704, 2002. [,11(2002)].

[147] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[148] https://twiki.cern.ch/twiki/bin/viewauth/cms/swguidehiggsanalysiscombinedlimit. *CMS Higgs PAG.*

[149] "http://pdg.lbl.gov/".

[150] Pekka K. Sinervo. Signal significance in particle physics. In *Advanced statistical techniques in particle physics. Proceedings, Conference, Durham, UK, March 18-22, 2002*, pages 64–76, 2002.

[151] V. Khachatryan. Precise determination of the mass of the higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *The European Physical Journal C*, 75(5), 2015.