université PARIS-SACLAY

NNT : 2017SACLX059

Thèse de doctorat

Search for Higgs boson pair production in the $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel with the CMS detector at the LHC

Recherche de la production de paires de bosons de Higgs dans le canal de désintégration $\mathrm{b\bar{b}}\tau^+\tau^-$ avec le détecteur CMS auprès du LHC

Thèse de doctorat de l'Université de Paris-Saclay
préparée à l'École polytechnique

École doctorale n° 576
Particules, Hadrons, Énergie, Noyau, Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)
Spécialité de doctorat : Physique des particules

Thèse présentée et soutenue à l'École polytechnique, le 5 octobre 2017, par

Luca Cadamuro

Composition du Jury :

Tejinder Singh Virdee	IC, London	Président du Jury
Aleandro Nisati	INFN, Roma	Rapporteur
Dirk Zerwas	LAL, Orsay	Rapporteur
Anne-Catherine Le Bihan	IPHC, Strasbourg	Examinatrice
Christoph Englert	University of Glasgow	Examinateur
Yves Sirois	LLR, Palaiseau	Directeur de thèse
Roberto Salerno	LLR, Palaiseau	Co-directeur de thèse

Abstract

This thesis describes a search for Higgs boson pair (HH) production using proton-proton collision data collected at $\sqrt{s} = 13\,\text{TeV}$ with the CMS experiment at the CERN LHC. Events with one Higgs boson decaying into two b quarks and the other decaying into two τ leptons ($\text{HH} \to \text{b}\bar{\text{b}}\tau^+\tau^-$) are explored to investigate both resonant and nonresonant production mechanisms. HH production gives access to the Higgs boson trilinear self-coupling and is sensitive to the presence of physics beyond the standard model.

A considerable effort has been devoted to the development of an algorithm for the reconstruction of τ leptons decays to hadrons (τ_h) and a neutrino for the Level-1 calorimeter trigger of the experiment, that has been upgraded to face the increase in the centre-of-mass energy and instantaneous luminosity conditions expected for the LHC Run II operations. The algorithm implements a sophisticated dynamic energy clustering technique and dedicated background rejection criteria. Its structure, optimisation and implementation, its commissioning for the LHC restart at 13 TeV, and the measurement of its performance are presented. The algorithm is an essential element in the search for HH production.

The investigation of the $\text{HH} \to \text{b}\bar{\text{b}}\tau^+\tau^-$ process explores the three decay modes of the $\tau^+\tau^-$ system with one or two τ_h in the final state. A dedicated event selection and categorisation is developed and optimised to enhance the sensitivity, and multivariate techniques are applied for the first time to these final states to separate the signal from the background. Results are derived using an integrated luminosity of 35.9 fb^{-1}. They are found to be consistent, within uncertainties, with the standard model background predictions. Upper limits are set on resonant and nonresonant HH production and constrain the parameter space of the minimal supersymmetric standard model and anomalous Higgs boson couplings. The observed and expected upper limits are about 30 and 25 times the standard model prediction respectively, corresponding to one of the most stringent limits set so far at the LHC.

Finally, prospects for future measurements of HH production at the LHC are evaluated by extrapolating the current results to an integrated luminosity of 3000 fb^{-1} under different detector and analysis performance scenarios.

Résumé

L'exploration du secteur scalaire du modèle standard de la physique de particules (SM) est l'un des objectifs principaux des recherches poursuivies auprès du grand collisionneur de hadrons (LHC) du CERN à Genève. Avec la découverte du boson de Higgs en 2012 par les expériences ATLAS et CMS, les efforts de la communauté scientifique se focalisent maintenant sur une meilleure caractérisation de cette particule. Le travail présenté dans cette thèse s'inscrit dans ce context en étudiant la production de paires de bosons de Higgs (HH) dans le canal de désintégration avec une paire de quarks b et une paire de leptons tau. Ce canal bénéficie d'un grand rapport de branchement (7.3%) et d'une faible contamination par les bruits de fond du SM, constitués principalement de paires de quark top ($\mathrm{t\bar{t}}$), de production Drell-Yan de leptons τ ($\mathrm{Z}/\gamma^* \rightarrow \tau^+\tau^-$), et de la production multiple de gerbes hadroniques. Plusieurs résultats ont été publiés au cours de cette thèse, en utilisant des échantillons de données de 2.7, 12.9, et 35.9 fb^{-1} respectivement. La discussion se focalise sur ces derniers résultats, où une meilleure sensibilité est obtenue grâce à la grande luminosité intégrée et à l'optimisation des techniques d'analyse.

Dans le contexte du SM, la mesure de la section efficace de production de HH permet d'extraire la valeur de l'auto-couplage du boson de Higgs, λ_{HHH}. Ce couplage est directement lié à la forme du potentiel scalaire et précisément prédit par la théorie à partir de la masse du boson de Higgs et de la valeur moyenne sur le vide du champ scalaire. Sa détermination représente donc une vérification de la validité et de la cohérence du SM et donne accès à une propriété du potential scalaire inexplorée jusqu'à présent. Auprès du LHC, le mode dominant de production de HH est la fusion de gluons où une faible section efficace, environ 33.5 fb à 13 TeV, résulte d'une interférence négative entre les deux principaux diagrammes de production. Cependant, la production de HH est sensible à la présence de contributions provenant d'une physique au-delà du SM (BSM), qui peuvent significativement modifier sa section efficace et ses propriétés cinématiques. Ces effets peuvent être observés comme une production résonante ou non-résonante de paires de bosons de Higgs. L'existence de la première est prédite dans plusieurs modèles qui étendent le secteur scalaire du SM ou qui présentent un nombre additionnel de dimensions spatiales. La deuxième est possible en présence de couplages anormaux du boson de Higgs, qui sont décrits avec une théorie efficace à l'échelle du TeV.

Le contexte expérimental de cette thèse est celui du redémarrage du LHC en 2015 pour son "Run II", une nouvelle phase de collisions proton-proton à une énergie dans le centre de masse de 13 TeV et une luminosité qui pourra atteindre $2-2.2\times10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, le double de la valeur nominale initialement prévue pour la machine. Pour assurer le succès du programme de physique de l'expérience CMS, le premier niveau (L1) de son système de déclenchement a été profondément amélioré. La mise à jour de son architecture et l'installation d'une nouvelle électronique avec de grandes capacités de calcul offre la possibilité de développer des algorithmes innovants pour l'identification des désintégrations des leptons τ en hadrons (τ_{h}) et un neutrino. Ces algorithmes permettent d'utiliser de

façon optimale l'information des calorimètres du détecteur CMS et améliorent significativement l'efficacité de reconstruction des τ_{h} et la réduction du bruit de fond, assurant une réduction du taux des déclenchements et un fonctionnement optimal dans les conditions des collisions du Run II.

L'algorithme τ du déclenchement au L1 réunit de façon dynamique les dépôts d'énergie significatifs dans des agrégats qui peuvent être regroupés pour mieux reconstruire les désintégrations avec plusieurs particules dans l'état final. Une calibration est appliquée à l'énergie de l'agrégat et sa forme, ainsi que l'énergie des dépôts environnants, sont utilisées pour rejeter les sources de bruit de fond. Le travail présenté dans cette thèse porte sur le développement et l'optimisation de l'algorithme, ainsi que sa mise en oeuvre dans l'électronique du système de déclenchement L1 et la vérification de son fonctionnement avec les données enregistrées dans des prises de données en parallèle effectuées en 2015. La performance est enfin mesurée en utilisant des évènements $\mathrm{Z} \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ enregistrés en 2016, période durant laquelle le système de déclenchement a été mis en service avec succès.

La recherche de la production de $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^{+}\tau^{-}$ bénéficie largement de ces améliorations du système de déclenchement. Elle utilise en fait les trois modes de désintégration où l'un des deux leptons τ se désintègre en τ_{h} et un neutrino et l'autre en muon, électron, ou encore τ_{h}, avec les neutrinos associés. Ces états finaux sont indiqués comme $\tau_{\mu}\tau_{\mathrm{h}}$, $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$, et $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. Les évènements sont sélectionnés selon la présence d'une paire de leptons et de jets compatibles avec des désintégrations $\mathrm{H} \to \tau^{+}\tau^{-}$ et $\mathrm{H} \to \mathrm{b\bar{b}}$. Ils sont divisés par état final de la paire $\tau^{+}\tau^{-}$, et catégorisés selon les propriétés de la paire $\mathrm{b\bar{b}}$. En particulier, une catégorie est définie pour les évènements où la paire de jets a un grand "boost" de Lorentz, ce qui améliore la reconstruction d'objets partiellement superposés et augmente la sensibilité pour des résonances d'une masse d'environ 700 GeV ou au-delà. Des méthodes multivariées sont enfin mises en place pour rejeter le bruit de fond $\mathrm{t\bar{t}}$ dans les état finaux $\tau_{\mu}\tau_{\mathrm{h}}$ et $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$. Les variables ont été choisies en fonction des propriétés géométriques des évènements du signal et du bruit de fond et elles sont combinées à l'aide d'arbres de décision entraînés sur des évènements simulés.

La modélisation des processus de physique est effectuée à l'aide d'une simulation Monte Carlo (MC). Des échantillons d'évènements du signal sont générés pour différentes hypothèses de production résonante et non résonante. Dans un premier cas, la valeur de la masse de la résonance $\mathrm{m_X}$ est comprise entre 250 et 900 GeV pour les hypothèses de spin 0 et 2. Dans un deuxième cas, le signal est paramétré en fonction des couplages anormaux du boson de Higgs grâce à une technique de repondération des évènements simulés. La production multiple de gerbes hadroniques est évaluée avec les données dans des régions où la contribution du signal est absente. Le bruit de fond de production Drell-Yan est estimé avec la simulation et corrigé avec des évènements $\mathrm{Z} \to \mu^{+}\mu^{-}$. Les autres bruits de fond sont entièrement estimés avec la simulation MC et des corrections sont appliquées pour améliorer la description des données. Les différences résiduelles sont prises en compte comme des erreurs systématiques dans l'analyse statistique.

La présence d'un signal est recherchée avec des observables différents dans le cas de production résonante et non-résonante et qui ont été développés pour optimiser la sensibilité de la recherche. Ceux-ci correspondent à une reconstruction cinématique de la masse invariante à quatre corps du système $\mathrm{b\bar{b}}\tau^{+}\tau^{-}$ et à une généralisation de la masse transverse. Aucun excès significatif n'est observé et les résultats sont en accord, compte tenu des incertitudes, avec les prédictions du SM. Des limites supérieures à un niveau

de confiance de 95% sont évaluées. Pour la production d'une résonance X, elles varient entre environ 500 et 5 fb pour $\sigma(gg \to X) \times (X \to HH \to b\bar{b}\tau^+\tau^-)$. Elles sont interprétées dans le contexte des scénarios du modèle supersymétrique minimal et permettent d'exclure une région de son espace des paramètres. Pour la production non-résonante, des limites supérieures sont évaluées pour différentes combinaisons de couplage du boson de Higgs, et des régions dans l'espace des couplages anormaux avec le quark top et λ_{HHH} sont exclues. Pour le signal prévu dans le SM, la limite supérieure observée et attendue correspond respectivement à environ 30 et 25 fois la prédiction théorique. Ceci représente l'un des résultats les plus sensibles jamais atteints au LHC et améliore significativement les limites précédentes.

Les perspectives pour l'observation de la production de HH dans les futures opérations du LHC et dans sa phase à haute luminosité (HL-LHC) sont enfin évaluées avec une extrapolation des résultats obtenus. Différents scénarios sont définis selon la performance attendue du détecteur et des techniques d'analyse dans les conditions futures du LHC et HL-LHC. Les limitations majeures à la sensibilité obtenues à présent sont discutées et des améliorations possibles sont prises en compte. Les résultats montrent que, dans le canal $HH \to b\bar{b}\tau^+\tau^-$, une sensibilité à la production HH du SM entre 0.6 et 1.4σ est attendue. Une observation nécessitera d'explorer et de combiner plusieurs états finaux. L'étude de la production de paires de bosons de Higgs, permettant à la fois de vérifier le SM et d'étudier ses possibles extensions, représente un important défi qui sera à la frontière des recherches futures auprès du LHC.

Acknowledgments

As this thesis comes to its conclusion, I wish to thank all the people that have accompanied me during these three years that marked an incredibly enriching period of my life.

I would like to thank the president of the Jury, Jim, the referees, Aleandro and Dirk, and the examiners, Anne-Catherine and Christoph, for accepting to review the manuscript and for the stimulating discussion and comments that followed.

My deepest gratitude is for my supervisors, Yves and Roberto, whose endless and thorough guidance has been essential to shape this work.
Without Yves, this thesis would not even have started, and I thank him for his support since the very beginning of the PhD. His vision of physics and his enthusiasm have been an extraordinary motivation to carry on this project, and he has been a point of reference throughout all its development.
And I am indebted to Roberto for all the time that he dedicated to me, following my work day-by-day. He formed me as a physicists, and I learned how to prepare, develop, and present an original scientific work. A lot of this thesis, in its form and its results, owes to the discussions that we had throughout these three years.

Thanks to all the members of the Laboratoire Leprince-Ringuet that have accompanied me in the adventure of the PhD. Thanks to the students, the postdocs, and the permanent researchers for the knowledge that they shared with me in all the interesting discussions at work, in the *cafétéria* or at the canteen. Thanks also to the IT and administrative teams for their prompt help that allowed this thesis to run smoothly.

This work is however only a small contribution to the huge program of exploration of particle physics performed at CMS. It is a privilege and an honour for me to be part of such an important project. I am in particular grateful to the members of the L1 trigger team from the LLR and UK groups. And I would like to thank the colleagues from LLR and the Italian institutes of Milano, Bari, and Pisa for the path that we walked together in the exploration of Higgs boson pair production.

A special thanks goes to Alex, for introducing me to the world of the trigger and of the "action" of the data taking. Working and teaching together has been pleasant, intense, and always interesting, and I am grateful for all the nice time that we spent together.
I also wish to thank Giacomo for all the work that we did side-by-side in the development of the analysis. I was happy to find not only a very competent colleague, but also a friend with whom I shared a lot of cheerful moments outside the walls of the laboratory.

And it is outside of the laboratory that I have found an incredible support from many people that I have met throughout these three years.

I am grateful to Louise for her kindness and generosity that helped me feel at home in Paris.
And a huge *merci* goes to Muriel, for her unconditioned generosity and hospitality in the occasion of my stay at CERN.

Thanks to my close friends in Italy, that never made me feel the distance from home, and to the extraordinary friends that I met in Paris that made me find a new home there.

We have spent incredible moments together, and they made me reconsider how many people can sit at the same table for a dinner, or fit in an apartment for a party.

I would like to thank my parents, because they have always been by my side although we were in two different countries. The conclusion of the PhD is only the last step of a long road where I could always count on their endless support, and I will never be capable of expressing enough gratitude for this.

And finally, I wish to thank Bea. She has been next to me all the time, sustaining me in the difficult moments and bearing all the difficulties with me. Without her strength and her constant encouragement this thesis wouldn't have been possible. Thanks for what we have shared in these years, for what we have experienced and lived, and for bringing joy in every moment that we spent together.

Contents

Introduction

With the ambitious goal of explaining the laws that regulate our Universe, particle physics has shaken our understanding of its very fundamental elements: space, time, and matter. After the first exploration of the subatomic scales at the end of the 19^{th} century, our knowledge of the constituents of the matter and their relation to the frame of space and time in which their interactions take place is now well supported by the two main theoretical pillars of particle physics, quantum mechanics and special relativity. Together with the general relativity, they provide a mathematical description that is capable of describing the behaviour of our Universe at spatial scales that extend over more than 40 orders of magnitude, and they can elucidate its past history, and predict its destiny. This is a remarkable result with no equal in any other experimental science that had, and is continuing to produce, a profound impact on our society and in our perception and understanding of the Universe.

The mathematical theory that supports this description is the standard model of particle physics (SM), a quantum field theory formalized between the 1960s and 1970s. It relies on the mathematically elegant concept of local gauge invariance under symmetry groups of its Lagrangian $\mathcal{L}$, the function from which the equations of motion of a physical system can be derived. This is at the origin of the unification of electromagnetic and weak forces, predicts the existence of photon, $\mathrm{W}^{\pm}$ and Z bosons, and determines their mutual interactions and those with the matter. The same principle is used to explain the existence of the gluon as the mediator of the strong force and its interactions with the quarks.

The SM has encountered a tremendous success because of its high predictive power and the large number of experimental confirmations to which it has been subjected. After the discovery of the $\mathrm{W}^{\pm}$ and Z bosons at the CERN super proton synchrotron (SPS) and the observation of the top quark at the FNAL Tevatron, extensive verifications of its prediction have been performed, most notably at the CERN large electron-positron collider (LEP). The SM has been capable of correctly describing all the measurements performed until now at the GeV and TeV energy scales.

However, this picture alone cannot account for the experimental observation of massive fermions and W and Z bosons, and predicts unphysical properties for vector boson scattering processes. The cornerstone for the completion of the SM is found in the introduction of the Brout-Englert-Higgs (BEH) mechanism, as called from the name of the physicists that proposed it in 1964. The BEH mechanism postulates the existence of a doublet of complex scalar fields that causes a spontaneous electroweak symmetry breaking (EWSB), thus generating the masses of the vector bosons, providing naturally mass terms for the fermion masses via a Yukawa interaction, and ensuring the unitarity of the theory. The direct consequence of the BEH mechanism is the presence of a scalar boson, called the Higgs boson (H), with a mass m_{H} that is a free parameter of the model. Following the incorporation of the BEH mechanism in the electroweak model by Weinberg and Salam in

1967 and the proof by 't Hooft and Veltman in 1972 that the theory was renormalizable, the search for the Higgs boson became the main goal of the researches performed at high energy colliders. However, even if the value of m_{H} could be constrained with theoretical arguments and direct searches at the LEP and Tevatron colliders, its existence could not be directly proved for almost half a century.

The Large Hadron Collider (LHC) was built to give an answer to this question by elucidating the properties of the EWSB mechanism and exploring the physics at the TeV energy scale. It is designed to collide protons at a centre-of-mass energy up to 14 TeV, in four interactions points instrumented with an equal number of detectors. The physics programme of the two general-purpose detectors, A Toroidal LHC Apparatus (ATLAS) and the Compact Muon Solenoid (CMS), is complemented by the measurements performed with the LHC beauty (LHCb) and A Large Ion Collider Experiment (ALICE) detectors, opening new horizons for the study of particle physics. The realization of the LHC has been a titanic endeavour that involved thousands of physicists and engineers from the entire world to design, build, commission, and run what is probably the most complex machine ever realized. After a 20-years long, eventful phase of construction and commissioning, after the first operations in 2009 and the first hints in the pp collision data in 2009, the LHC made it possible to answer the dilemma: in July 2012, the ATLAS and CMS collaborations announced the discovery of a new boson with a mass close to 125 GeV. The data collected between 2010 and 2012, in the machine operations denoted as the Run I, were crucial to characterize the Higgs boson.

The Higgs boson is something unique in the SM. It is the only known elementary boson with a spin of 0, a scalar, and whose existence does not arise from a local gauge invariance mechanism. It breaks the degeneracy between the three families of fermions by conferring them their mass with couplings of different strengths, and its observation consequently proves the purely quantum-relativistic nature of the mass. The discovery of the Higgs boson marked a milestone in the history of physics, not only because it completes the SM with its last missing piece, but because it opens up the study of a new sector of the theory, the scalar sector. It is now of utmost importance to precisely characterize the properties and couplings of the Higgs boson. The measurements performed with the aforementioned LHC Run I data showed a good agreement with the predictions of the SM. The exploration of the scalar sector of the SM now requires a more precise determination of its couplings and properties, including the measurement of rare production and decay modes. In this context, the Higgs boson self-interactions are of particular interest because they provide invaluable information to reconstruct the shape of the scalar potential itself. As little is currently known experimentally about these interactions, their measurement represents one of the main elements for a complete characterization of the scalar sector of the SM.

The observation of the Higgs boson completes the theoretical description of the SM, making it a theory a priory valid up to the Planck scale. However, it raises further questions deeply related to its scalar sector. First, the SM does not provide a mechanism that determines its arbitrary parameters. In particular, it does not explain why three families of fermions exist, what is at the origin of their couplings to the Higgs boson, and why their values range over several order of magnitude. Being responsible for the breaking of the degeneracy between the fermion families, the scalar sector of the SM has a direct role in this context. Second, the mass of the Higgs boson is not protected by any fundamental symmetry of the theory, making it largely sensitive to divergent radiative corrections.

These corrections need to be finely tuned to account for the observed Higgs boson mass of about 125 GeV, raising an issue about the naturalness of the theory. Finally, this specific mass value implies an instability of the vacuum, as the evolution of the Higgs self-coupling (and consequently of the shape of the scalar potential) with the energy implies that at higher scales this can become negative. This results in a metastable conditions of our Universe, that could collapse into a different vacuum state.

These theoretical considerations should be regarded together with the limitations of the SM in explaining the experimental observations at cosmological scales. The SM does not provide a mechanism that is responsible for the observed matter-antimatter asymmetry in the Universe, nor it accounts for the existence of a stable particle that is compatible with the dark matter, of which we observe the gravitational effects. Although being extremely successful at describing the phenomenology of colliders experiment, the SM seems incomplete.

When considered together, these open questions seem to indicate that the SM is only a part of a more extended physics model. The existence of such physics beyond the SM (BSM) is a conundrum which currently focuses the efforts of the theoretical and experimental communities. Many different models have been devised to solve these problems and provide an explanation to the arbitrariness and unnaturalness of many aspects of the SM. Being something intimately different from all the other particles of the SM, the Higgs boson provides a preferential way to search for these SM extensions. The nature of BSM physics is still an open question, as no clear signs of its presence have been found in Run I data at the LHC.

With the restart of the LHC in 2015 for its Run II these questions might find an answer. The increase in the centre-of-mass energy of pp collisions from 8 to 13 TeV and the higher instantaneous luminosity allow for a more precise test of the Higgs boson properties and the exploration of its rare decay modes, where deviations from the SM might be found. At the same time, it opens the way to direct searches for experimental signatures of BSM physics at the TeV scale.

The work presented in this thesis is situated in this context and aims at the exploration of the EWSB properties and of extensions of the SM via the study of Higgs boson pair (HH) production. This process allows for probing the very fundamental nature of the Higgs scalar field, as it directly involves the trilinear self-coupling of the Higgs boson which, as mentioned above, depends on the shape of the scalar potential itself. The observation of this process thus represents a crucial test of the validity of the SM. Its measurement at the LHC is particularly challenging because of its small cross section. However, the direct relation to the scalar potential makes HH production very sensitive to the presence of BSM contributions, that could manifest either directly as new states decaying to a HH pair (resonant production), or as contributions in the quantum loops that would modify its cross section and kinematic properties (nonresonant production).

The search for HH production discussed in this thesis explores the decay channel where one Higgs boson decays to a b quark pair and the other to a τ lepton pair: the relatively large branching fraction of 7.3% and the relatively small contamination from SM backgrounds make it one of the most sensitive decay channels to study HH production at the LHC. The experimental challenge is represented by the identification of the decays of τ leptons to hadrons (τ_{h}) and a neutrino, accounting for about 65% of the total decays. This is particularly difficult in the dense environment of the LHC, where proton bunches collide at the centre of the CMS detector every 25 ns and up to 40 simultaneous interactions on

average take place at each crossing. The sensitivity thus crucially depends on the efficiency of the identification of these decays and on the rejection of the abundant jet background. To achieve this goal, the τ_{h} identification performance must be ensured at the very first event selection step that is performed by the Level-1 (L1) trigger, the hardware system that filters collision events and selects those to be recorded for subsequent analysis.

I had the chance to start my thesis work at the beginning of fall 2014, when the end of the two-years long LHC shutdown was approaching and the machine was being prepared for the restart. It was a moment of great excitement for the forthcoming collisions at an unprecedented centre-of-mass energy. The CMS experiment was getting ready for Run II by completing an important upgrade program of its L1 trigger system. This provided the possibility to develop and implement innovative approaches to τ_{h} identification at hardware level. I had thus the opportunity to work on the development of the τ_{h} identification algorithm, taking part to the ongoing project at the LLR laboratory. My work consisted in the completion of the existing algorithm structure, and in the development of criteria to reject the jet background. This resulted in the introduction of an isolation discriminant and in the study of a criterion based on the topology of the energy deposit, essential ingredients to reduce the trigger rate and set low energy thresholds on the reconstructed objects to maximize the acceptance to physics signals.

The completion of the τ algorithm structure allowed me to determine its performance using simulated collision events and to quantify the improvements with respect to the Run I trigger system. However, the algorithm had to be translated from the "offline" world in which it had been developed to the "online" world of data taking, where it had to be implemented in the electronic boards of the L1 trigger system. My activity then focused on the optimization and streamlining of the algorithm, fulfilling the hardware specifications with minimal impact in the performance. With the algorithm implemented, I contributed to commission the L1 trigger system and to verify its performance using data taken in parallel runs in 2015, ensuring the readiness of the upgrade for 2016 collisions. The measurement of the trigger performance with Z boson decay events in 2016 data completed the picture and allowed me to verify that the original performance, established with the initial design more than one year before, had been met in the data taking.

The search for HH production in the $\mathrm{b\bar{b}}\tau^{+}\tau^{-}$ decay channel proceeded in parallel, and its sensitivity largely benefited of the trigger upgrade. My work consisted in the complete design and preparation of the analysis strategy, from the development of specific data analysis tools to the definition of the analysis methods. This comprised the determination of the object and event selections, of the methods for background estimation and the study of systematic uncertainties, and the choice of sensitive observables to search for the presence of a signal. The three $\tau^{+}\tau^{-}$ final states involving at least one τ_{h} candidate have been explored and, for the first time in the $\mathrm{b\bar{b}}\tau^{+}\tau^{-}$ decay channel, multivariate methods have been developed to reject the background contribution.

The analysed dataset evolved rapidly and, moving in the unknown territory of 13 TeV collisions, data needed to be promptly explored to search for possible presence of BSM physics, as well as to assess the sensitivity of the search. Results have thus been published as Physics Analysis Summaries (PASes) for a dataset of 2.3, 12.9, and 35.9 fb^{-1}. The analysis strategy has progressively been refined to better exploit the increasingly large amount of data. I could experience the joys and sorrows of scientific writing in a large collaboration as the editor of some of these PASes and of the paper summarizing the results on the full 2016 dataset. This paper has been submitted to Physics Letters B for

peer-review as this thesis is being completed. I had the pleasure to present these results on behalf of the ATLAS and CMS Collaborations at the 52$^{\text{nd}}$ Rencontres de Moriond international conference in 2017.

The results described in this thesis currently represent one of the most sensitive results on HH cross section in both its resonant and nonresonant production modes, and significantly improve over the previous limits from the LHC Run I. They are part of a vast effort of the high energy physics community in the study of the EWSB and searches for BSM physics. In this exciting exploration, occasionally marked by fluctuations subsequently disproved by the analysis of larger datasets, no evidence for BSM physics has been found thus far, with no exception for the HH search in the $\mathrm{b\bar{b}}\tau^+\tau^-$ final state. These results however constrain the parameter space of BSM physics models, as well as anomalous couplings of the Higgs boson.

Chapters 1 and 2 introduce the reader to the theoretical and experimental contexts, discuss the motivations of a search for HH production, and present the experimental apparatus that is used to perform the search. The subsequent chapters describe my personal contribution to this subject and summarize my activity during three years of PhD. In Chapter 3 the structure of the upgrade τ algorithm, its development, commissioning and performance measured with data are discussed. Chapter 4 introduces the experimental challenges of the $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel and the techniques developed to select the collision events and optimize the sensitivity of the search. It is followed by Chapter 5, where the methods developed for the modelling of the signal and background processes, and the corresponding uncertainties, are discussed. The results and their interpretation are presented in Chapter 6. Given the importance of the HH production and measurement of the Higgs boson trilinear coupling in the context of the SM, a projection of the sensitivity for future LHC operations under different detector performance scenarios are presented in Chapter 7.

Chapter 1

Higgs boson pair production

The standard model of particle physics (SM) is a renormalizable quantum field theory that describes the phenomena at the subnuclear scales. It provides a unified description of the strong, weak, and electromagnetic forces, and incorporates a minimal scalar sector that is at the origin of the spontaneous breaking of the electroweak symmetry and of the masses of the fermions. The SM is well corroborated by experimental observations at collider experiments, and received further confirmation with the recent discovery of the Higgs boson. Despite the excellent agreement with direct experimental tests performed so far, some observations from the subnuclear to the astrophysical scales, as well as theoretical considerations, suggest that it is incomplete and that a broader theory exists beyond its current formulation. This physics beyond the SM (BSM) is possibly connected to the scalar sector, and the Higgs boson discovery opens new ways to its exploration. Being intimately related to the nature of the scalar sector, the production of Higgs boson pairs (HH) at the LHC can give invaluable information in this context. It allows for determining the Higgs boson self-interaction, and provides a fertile ground to search for the signs of BSM physics.

This chapter discusses the importance of the study of HH production in the context of both SM and BSM physics. After introducing the SM gauge structure and its scalar sector, with a focus on the Higgs boson properties and couplings, HH production in the SM is reviewed. This process is subsequently discussed in the context of BSM models, that can result in resonant or nonresonant HH signatures. Finally, the phenomenology in collider experiments and results previously obtained at the LHC are discussed.

1.1 The standard model of particle physics

The gauge sector and the scalar sector are the foundations of the SM. They are closely interconnected but rely on different theoretical concepts of gauge invariance and spontaneous symmetry breaking. They are discussed separately in what follows, with a focus on the properties of the Higgs boson and a summary of experimental measurements.

1.1.1 Fields and gauge structure of the standard model

The mathematical formulation of the SM is based on the local gauge invariance of its Lagrangian under the gauge group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{SU}(1)_Y$ to explain the strong, weak, and electromagnetic interactions. In particular, the $\mathrm{SU}(3)_C$ invariance results in the existence of "gluons" (g) as the mediators of the strong force, which is described by the quantum chromodynamics (QCD). The $\mathrm{SU}(2)_L \times \mathrm{SU}(1)_Y$ symmetry explains jointly

the weak and electromagnetic forces, mediated by the $W^{\pm}$ and Z bosons and the photon (γ), respectively. The SM formulation does not encompass the gravitational interaction, which is negligible at the subnuclear scales.

Matter is described in the SM by fermion fields of spin $\frac{1}{2}$, which interactions are mediated by spin-1 boson fields. Experimental observations show that twelve physical fermion fields exist, six "quark" fields and six "lepton" fields. They are organized in three families, made up of two quarks of electric charge $+\frac{2}{3}$ and $-\frac{1}{3}$ and two leptons of electric charge -1 and 0. Fermions in one family and their counterparts in the others have identical properties and only differ in their mass. This is related to their coupling strength to the scalar field, that is described in the next section. To each fermion corresponds an antiparticle with identical properties but opposite quantum numbers.

Quarks

Quarks are subject to all the three forces and, in particular, are the only fermions to possess a "colour" charge to which QCD owes its name. The first family of quarks is composed of the up (u) and down (d) quarks, with a mass of a few MeV. The former has a positive electric charge of $+\frac{2}{3}$ while the latter has a negative electric charge of $-\frac{1}{3}$. Being the lightest quarks, they are stable and compose the ordinary matter. Their counterparts in the second family are the charm (c) and the strange (s) quarks, of masses of about 1.28 GeV and 95 MeV respectively. Finally, the third family is composed of the top (t) and bottom (b) quarks, which masses are about 173 and 4.2 GeV respectively.

Because of the QCD colour confinement properties, quarks do not exist as free states but can be experimentally observed only as bound states. Collider experiments thus detect "mesons", that are composed by a quark-antiquark pair, or "baryons", composed by three quarks. The proton and neutron composing the ordinary matter are stable examples of the latter. Mesons and baryons are collectively denoted as hadrons. The creation of hadrons from a single quark produced in a collider experiment is a complex process that takes the name of "hadronization". As its timescales, related to the QCD energy scales, are of the order of 10^{-24}s, hard scatter and hadronization phenomena can be treated separately thanks to a factorization of their effects. The top quark represents an exception in this sense, as its lifetime is so short ($\approx 0.5 \times 10^{-24}$ s) that it decays before bound states can be formed.

Quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, as quark mass eigenstates do not correspond to the weak interaction eigenstates. Their mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix.

Leptons

Leptons have no colour charge and are subject only to the electromagnetic and weak forces. The charged leptons of the three families are respectively denoted as the electron (e), muon (μ) and tau lepton (τ). The electron is stable, being the lightest lepton with a mass of 511 keV. The muon has a mass of 105.7 MeV and a lifetime of 2.2 µs, that is sufficiently long to consider it as a stable particle at an LHC experiment given the detector size and the typical muon momentum. Finally, the tau lepton has a mass of 1.8 GeV and a lifetime of 2.9×10^{-13} s, that is instead short enough for observing it only through its decay products. Most notably, the tau is the only lepton that has a sufficient mass to decay to semi-leptonically.

To each lepton corresponds a neutrino, respectively denoted as ν_{e}, ν_μ, and ν_τ. Being electrically neutral, neutrinos interact with the matter only via the weak force and consequently they are not directly detectable at collider experiments. Little is known about their masses, but the observation of their flavour oscillations prove that they are not zero. The mixing of weak and mass eigenstates is represented by the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix.

Strong interaction

QCD is built on the local gauge invariance under the $\mathrm{SU}(3)_C$ group, the subscript denoting the relation to the colour charge arising from this symmetry. The free-field Dirac Lagrangian density of a massless spin-$\frac{1}{2}$ fermion (the quark fields in this case) is:

$$\mathcal{L} = \bar{\psi}(x)\left(i\gamma^\mu\partial_\mu\right)\psi(x) \tag{1.1}$$

where ψ is the fermion field at the space-time coordinate x and γ^μ are the Dirac matrices. The notation $\not{\partial} \equiv \gamma^\mu\partial_\mu$ is also used in some cases. The following discussion on QCD is valid in presence of a $m\bar{\psi}\psi$ mass term. The reason to consider massless fermions is explained in the context of the electroweak interactions described in the next section. The fermion field transforms in the following way under the $\mathrm{SU}(3)_C$ group:

$$\psi(x) \to e^{ig\frac{\lambda^a}{2}\theta_a(x)}\psi(x) \tag{1.2}$$

where $\frac{\lambda^a}{2}$ are the eight Gell-Mann matrices that generate the group. An important remark is that the derivatives $\partial_\mu\psi(x)$ do not transform in the same way. Hence, the only way for the Lagrangian density (1.1) to be invariant under the transformation (1.2) is to redefine the derivative ∂_μ as covariant derivatives:

$$D_\mu = \partial_\mu - igA^a_\mu(x)\frac{\lambda^a}{2} \tag{1.3}$$

where the gauge vector fields $A^a_\mu(x)$ correspond to the eight gluons that mediate the strong force. To satisfy the local gauge invariance of the Lagrangian, the gluon fields must transform as:

$$A^a_\mu \to A^a_\mu + \partial_\mu\theta^a + gf^{abc}A^c_\mu\theta^c \tag{1.4}$$

The f^{abc} symbols denote the structure constants of the group that are defined from the commutation rules $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = if^{abc}\frac{\lambda^c}{2}$. The introduction of the vector fields ensure that the covariant derivative transforms in the same way as the fermion fields and that, consequently, the Lagrangian density is invariant under the local gauge transformation. The Lagrangian density can be completed with a kinetic term for the gluon fields in the form:

$$-\frac{1}{4}F^{\mu\nu}_a F^a_{\mu\nu} \tag{1.5}$$

where:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \tag{1.6}$$

The complete QCD Lagrangian density thus reads:

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu}\right)\psi - g\bar{\psi}(x)\gamma^{\mu}\frac{\lambda_a}{2}\psi A_{\mu}^{a} - \frac{1}{4}F_{a}^{\mu\nu}F_{\mu\nu}^{a} \tag{1.7}$$

with a summation over all quark fields being implied. The first term is the same as in the original Lagrangian density and represents the free-field propagation of the quark. The second one stems from the introduction of the covariant derivative and represents the interaction of the quark with the vector field A_{μ}. The strength of the interaction is parametrized by the constant g, usually redefined as the strong coupling constant $\alpha_{\mathrm{s}} = g^2/4\pi$. The third term has been introduced as the kinetic term of the vector field. The generators of the $\mathrm{SU}(3)_C$ group do not commute and therefore the structure constants f^{abc} are not all zero. As a consequence, the $gf^{abc}A_{\mu}^{b}A_{\nu}^{c}$ terms in Eq. (1.6), when inserted in the kinetic term of Eq. (1.7), result in cubic and quartic self-interactions of the gluon fields. Such interactions between the force mediators are a general property of non-abelian gauge theories.

Requiring the local gauge invariance led to the introduction of gauge bosons (the gluons) and to the description of their interactions with the fermion fields (the quarks). Choosing the $\mathrm{SU}(3)_C$ group implies the presence of eight generators, the eight gluons, that are mathematically described by the adjoint representation of the group (**8**), and differ by the colour charge that they carry. Quarks and anti-quark are instead described in the simplest non-trivial representations of $\mathrm{SU}(3)_C$, **3** and $\bar{\mathbf{3}}$, which explains the three colour quantum numbers of quarks. It is important to remark that explicit mass terms in the form $A_{\mu}^{a}A_{a}^{\mu}$ for the gauge bosons in Eq. (1.7) would break the gauge invariance of the Lagrangian.

Electroweak interaction

Electroweak interactions are explained in the SM with the same local gauge invariance mechanism as strong interactions by imposing a symmetry under the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group. Experimental observations show that parity is violated by weak interactions, which is accounted for in the theoretical description by assigning different interactions to fermions of opposite chiralities. The left and right chiral components of a field are defined from the $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix. This is used to define left and right chirality projection operators as $\frac{1-\gamma^5}{2}$ and $\frac{1+\gamma^5}{2}$, respectively. In the limit of a massless particle, the chirality corresponds to the helicity, that is defined as the normalized projection of the spin vector onto the spatial momentum vector.

The $\mathrm{SU}(2)_L$ gauge group is a non-abelian group to which the weak isospin quantum number (I_3) is associated. The gauge invariance under this group results in the presence of 3 gauge fields W_{μ}^{i} ($i = 1, 2, 3$). Fermion fields of left chirality are represented by $\mathrm{SU}(2)_L$ doublets while fermions of right chirality are $\mathrm{SU}(2)_L$ singlets and do not interact with the W_{μ}^{i} fields.

The $\mathrm{U}(1)_Y$ gauge group is abelian and is associated to the weak hypercharge Y. A single gauge field, denoted as B_{μ}, results from the $\mathrm{U}(1)_Y$ local gauge invariance, and interacts with both ψ_L and ψ_R. The $\mathrm{U}(1)_Y$ group can be seen in close analogy with the electromagnetic $\mathrm{U}(1)_{\mathrm{em}}$ group associated to the electric charge Q, but is distinct from it.

The Gell-Mann–Nishijima formula determines the relation with the electric charge:

$$Q = I_3 + \frac{Y}{2} \tag{1.8}$$

Fields can therefore be represented as one doublet and two singlets:

$$\begin{aligned} \Psi_L &\equiv \frac{1-\gamma^5}{2}\begin{pmatrix}\psi\\ \psi'\end{pmatrix} = \begin{pmatrix}\psi_L\\ \psi'_L\end{pmatrix} \\ \psi_R &\equiv \frac{1+\gamma^5}{2}\psi \\ \psi'_R &\equiv \frac{1+\gamma^5}{2}\psi' \end{aligned} \tag{1.9}$$

The fields ψ and ψ' represent either the neutrino and charged lepton fields or the up– and down–type quark fields. The two sectors are however separate and neither the strong nor the electroweak interactions can transform quark fields into lepton fields or vice-versa: a direct lepton-quark coupling is not predicted in the SM.

Under this notation, the Lagrangian density can be written as composed of three parts:

$$\mathcal{L} = i\bar{\Psi}_L \not{D} \Psi_L + i\bar{\psi}_R \not{D} \psi_R + i\bar{\psi}'_R \not{D} \psi'_R \tag{1.10}$$

where the covariant derivative, that is introduced to ensure the gauge invariance, is defined as:

$$D_\mu = \partial_\mu - igW^i_\mu T_i - ig'\frac{Y_\psi}{2}B^\mu \tag{1.11}$$

with $T_i = \frac{\sigma_i}{2}$ (the Pauli matrices, generators of the $\mathrm{SU}(2)_L$ group) for the Ψ_L field and 0 for ψ_R and ψ'_R. The coupling constants g and g' define the strength of the interactions. More interestingly, the Lagrangian density (1.10) can be rewritten as:

$$\mathcal{L} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{CC}} + \mathcal{L}_{\text{NC}} \tag{1.12}$$

with the three parts defined respectively as:

$$\begin{aligned} \mathcal{L}_{\text{kin}} &= i\bar{\Psi}_L \not{\partial} \Psi_L + i\bar{\psi}_R \not{\partial} \psi_R + i\bar{\psi}'_R \not{\partial} \psi'_R \\ \mathcal{L}_{\text{CC}} &= gW^1_\mu \bar{\Psi}_L \gamma^\mu \frac{\sigma_1}{2} \Psi_L + gW^2_\mu \bar{\Psi}_L \gamma^\mu \frac{\sigma_2}{2} \Psi_L \\ &= \frac{g}{\sqrt{2}} W^+_\mu \bar{\Psi}_L \gamma^\mu \sigma^+ \Psi_L + \frac{g}{\sqrt{2}} W^-_\mu \bar{\Psi}_L \gamma^\mu \sigma^- \Psi_L \\ &= \frac{g}{\sqrt{2}} W^+ \bar{\psi}_L \gamma^\mu \psi'_L + \frac{g}{\sqrt{2}} W^- \bar{\psi}'_L \gamma^\mu \psi_L \\ \mathcal{L}_{\text{NC}} &= \frac{g}{\sqrt{2}} W^3_\mu \left[\bar{\psi}_L \gamma^\mu \psi_L - \bar{\psi}'_L \gamma^\mu \psi'_L\right] \\ &\quad + \frac{g'}{\sqrt{2}} B_\mu \left[Y_{\Psi_L}(\bar{\psi}_L \gamma^\mu \psi_L + \bar{\psi}'_L \gamma^\mu \psi'_L) + Y_{\psi_R} \bar{\psi}_R \gamma^\mu \psi_R + Y_{\psi'_R} \bar{\psi}'_R \gamma^\mu \psi'_R\right] \end{aligned} \tag{1.13}$$

and

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \\ \sigma_\mu^\pm &= \frac{1}{2}(\sigma^1 \pm i\sigma^2) \end{aligned} \tag{1.14}$$

There is therefore a charged current interaction that couples the fields ψ_L and ψ'_L and is mediated by the $W^\pm$ fields, that correspond to the charged weak bosons interacting with fermions. A neutral current interaction also exists, although neither the W_μ^3 nor the B_μ fields can be interpreted as the photon field since they couple to neutral fields. However, we can express them in terms of the physical Z_μ field (the neutral Z boson field) and the A_μ field (the photon field) through a linear superposition parametrized with the Weinberg angle θ_W:

$$\begin{aligned} B_\mu &= A_\mu \cos\theta_W - Z_\mu \sin\theta_W \\ W_\mu^3 &= A_\mu \sin\theta_W + Z_\mu \cos\theta_W \end{aligned} \tag{1.15}$$

Upon substituting this expression in Eq. (1.13), two neutral current interactions appear, the one with the A_μ field being determined by a coupling of strength

$$g \sin\theta_W I_3 + g' \cos\theta_W \frac{Y}{2} \tag{1.16}$$

The unification of the weak and electromagnetic forces is completed if we require this expression to be equal to the coupling constant of the photon field eQ. We can arbitrarily set $Y_{\Psi_L} = -1$ as the hypercharge only appears multiplied by g', and solve the equation by using $Q = 0$ for the neutrino fields and $Q = -1$ for the lepton fields. Upon substitution, Eq. (1.8) is obtained together with the following relation between the coupling constants:

$$g \sin\theta_W = g' \cos\theta_W = e \tag{1.17}$$

The Lagrangian density of Eq. (1.12) only needs to be completed with the kinetic terms of the gauge fields. The full electroweak Lagrangian can be expressed in a compact form as:

$$\mathcal{L}_{\text{EWK}} = i\bar{\Psi}_L \not{D} \Psi_L + i\bar{\psi}_R \not{D} \psi_R + i\bar{\psi}'_R \not{D} \psi'_R - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} W_i^{\mu\nu} W_{\mu\nu}^i \tag{1.18}$$

The field strength tensors are:

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\varepsilon^{abc} W_\mu^b W_\nu^c \end{aligned} \tag{1.19}$$

Equation (1.18) contains the free fermion Dirac Lagrangian and the charged and neutral current interactions of the fermions previously discussed. The kinetic terms of the W_μ^i field, once developed, show a rich structure of self-interactions of the gauge bosons. Trilinear (ZWW, γWW) and quadrilinear (ZZWW, $\gamma\gamma$WW, γZWW, WWWW) interactions are predicted by the theory. As already observed for the strong interactions, explicit mass terms of the gauge fields would break the gauge invariance. Direct fermion mass terms are also not allowed, because they are not invariant under the gauge transformation: the

left and right chiralities of the fields transform differently under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ while the mass can be decomposed as $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$.

Summary of particle properties and interactions

The fermion fields are summarized under their $\mathrm{SU}(2)_L$ representation in Table 1.1. In the SM, quarks and leptons have the same structure under the $\mathrm{SU}(2)_L$ group. Left and right chirality fields are respectively a double and a singlet of the $\mathrm{SU}(2)_L$ group and, consequently, only the former have a charged weak interaction, that is mediated by the $\mathrm{W}^{\pm}$ bosons. Neutral weak interactions are mediated by the Z bosons, that interact with both chiral components, albeit with a different strength, thanks to the mixing of the gauge fields via the Weinberg angle θ_W. The electromagnetic force, mediated by the photon γ, is not sensitive to the chirality of the fermion fields and its interaction depends on the charge Q, that is related to the hypercharge Y and the weak isospin I_3. The table shows the different lepton fields according to their flavours and to their chirality. Quark fields exist in three additional types according to their colour charge, as they represent a triplet of the $\mathrm{SU}(3)_C$ group. In contrast, leptons have no colour charge and thus do not interact with the strong force. Interactions can change the quantum numbers of the fields through the charge carried by the mediators. Charged weak interactions change the weak isospin (and consequently the electric charge) and strong interactions change the colour charge of quarks.

Type	1^{st} gen.	2^{nd} gen.	3^{st} gen.	I_3	Y	Q	$\mathrm{SU}(3)_C$
Quarks	$\begin{pmatrix}\mathrm{u}_L\\ \mathrm{d}_L\end{pmatrix}$	$\begin{pmatrix}\mathrm{c}_L\\ \mathrm{s}_L\end{pmatrix}$	$\begin{pmatrix}\mathrm{t}_L\\ \mathrm{b}_L\end{pmatrix}$	$\begin{pmatrix}1/2\\ -1/2\end{pmatrix}$	$1/3$	$\begin{pmatrix}2/3\\ -1/3\end{pmatrix}$	triplet
	u_R	c_R	t_R	0	$4/3$	$2/3$	
	d_R	s_R	b_R	0	$-2/3$	$-1/3$	
Leptons	$\begin{pmatrix}\nu_{\mathrm{e},L}\\ \mathrm{e}_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\mu,L}\\ \mu_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\tau,L}\\ \tau_L\end{pmatrix}$	$\begin{pmatrix}1/2\\ -1/2\end{pmatrix}$	-1	$\begin{pmatrix}0\\ -1\end{pmatrix}$	singlet
	e_R	μ_R	τ_R	0	-2	1	
	$\nu_{\mathrm{e},R}$	$\nu_{\mu,R}$	$\nu_{\tau,R}$	0	0	0	

Table 1.1 – Fermion fields under their $\mathrm{SU}(2)_L$ representation. The $_L$ and $_R$ subscripts denote respectively the left and right chiralities. Quarks and lepton fields are separately listed. They differ for the weak hypercharge Y (and consequently by the electric charge $Q = I_3 + Y/2$) quantum numbers, as well as for their colour charge under the $\mathrm{SU}(3)_C$ group.

This formulation of the SM is of extraordinary beauty and elegance. Matter fields are completely described in terms of quantum numbers, and their interactions follow from the application of a symmetry principle to the Lagrangian density. Differences in the representation of the fields under a specific group completely determine the phenomenology that we observe experimentally. The electromagnetic and weak force are jointly described, and all the relevant forces at the subnuclear scale are explained from a common symmetry principle. However, the theory requires both the fermions and the gauge bosons to be massless, as any explicit mass term would violate the gauge invariance itself. This is in

clear contrast with the experimental observation of massive weak bosons and fermions. A simple addition "a posteriori" of the mass terms produces a theory that is not renormalizable, and results in unhphysical predictions for scattering of longitudinally polarized vector bosons. The solution needed to ensure the unitarity of the theory and to explain bosons and fermions masses is provided by the Brout-Englert-Higgs mechanism, a natural way of breaking the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry to $\mathrm{U}(1)_{\mathrm{em}}$ without explicitly violating the local gauge invariance.

1.1.2 The Brout-Englert-Higgs mechanism

The Brout-Engler-Higgs (BEH) mechanism was proposed in 1964 independently by physicists Englert and Brout [1], Higgs [2], and also by Guralnik, Hagen, and Kibble [3] as a solution to generate the gauge boson masses and explain the fermion masses. The mechanism is based on the concept of spontaneous symmetry breaking, a phenomenon that is often observed in Nature whenever individual ground states of a system do not satisfy the symmetries of the system itself.

A familiar example is a pencil vertically placed on a table, for which the gravitational force is symmetric for rotations around the vertical axis, but the ground state corresponds to a pencil laying horizontally on the table and pointing to a specific direction: the ground state does not satisfy the symmetry of the force acting on the system. The symmetry is said to be "spontaneously broken" because the ground state is not invariant under it, but the ground state chosen is only one out of the continuum of the possible ground states. As these are interrelated by the original symmetry of the system, the latter is "hidden" among these multiple choices.

In the BEH mechanism, spontaneous symmetry breaking is realized through the introduction of a complex scalar doublet of fields:

$$\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} \tag{1.20}$$

An important remark is that the field must be scalar to satisfy space isotropy, otherwise the expectation value on the vacuum would be frame-dependent. Moreover, the expectation value on the vacuum must be constant to satisfy space homogeneity. The field has an hypercharge $Y_\Phi = 1$ and thus its covariant derivative is:

$$D_\mu = \partial_\mu - igW_\mu^i \frac{\sigma_i}{2} - \frac{1}{2} ig' B_\mu \tag{1.21}$$

The BEH lagrangian can consequently be written as:

$$\mathcal{L}_{\mathrm{BEH}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \tag{1.22}$$

where the potential $V(\Phi^\dagger \Phi)$ is defined as:

$$V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \qquad \text{with } \mu^2, \lambda > 0 \tag{1.23}$$

All the doublets that satisfy the condition:

$$|\Phi^2| = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.24}$$

are minima of these potential, and are connected through gauge transformations that change the phase of the field Φ but not its modulus. The quantity v is called the vacuum expectation value (VEV) of the scalar potential. Once a specific ground state is chosen, the symmetry is explicitly broken but the Lagrangian is still gauge invariant with all the important consequences for the existence of gauge interactions.

If the symmetry is spontaneously broken to the ground state that is parallel to the ϕ^0 component of the doublet, it can be shown that this specific ground state is still invariant under the $U(1)_{\text{em}}$ symmetry group. As a consequence, the field expansion around this minimum is written as:

$$\Phi(x) = \frac{1}{\sqrt{2}} \exp\left[\frac{i\sigma_i\theta^i(x)}{v}\right] \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.25}$$

This corresponds to the presence of a scalar real massive field H and of three massless fields θ^i. The latter are expected as consequence of the Goldstone theorem [4], that states that the spontaneous breaking of a continuous symmetry generates as many massless bosons (the Goldstone boson) as broken generators of the symmetry. However, such massless bosons are not observed in Nature. They can be removed with an $\text{SU}(2)_L$ transformation that consists in the choice of a specific gauge called "unitary gauge":

$$\Phi(x) \to \Phi'(x) = \exp\left[\frac{-i\sigma_i\theta^i(x)}{v}\right] \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.26}$$

After this transformation, only the real scalar field $H(x)$ remains and its quanta correspond to a new physical massive particle, the Higgs boson (H).

Upon substitution of Eq. (1.21) and of the expression of $\Phi(x)$ in the unitary gauge in Eq. (1.22), the BEH Lagrangian reads:

$$\begin{aligned} \mathcal{L}_{\text{BEH}} = & \frac{1}{2}\partial^\mu H \partial_\mu H - \frac{1}{2}\left(2\lambda v^2\right) H^2 \\ & + \left[\left(\frac{gv}{2}\right)^2 W^{\mu+} W^-_\mu + \frac{1}{2}\frac{(g^2 + g'^2)v^2}{4} Z^\mu Z_\mu\right]\left(1 + \frac{H}{v}\right)^2 \\ & + \lambda v H^3 + \frac{\lambda}{4} H^4 - \frac{\lambda}{4} v^4 \end{aligned} \tag{1.27}$$

The first line represents the evolution of the scalar Higgs field, that has a mass $m^2_{\text{H}} = 2\lambda v^2 = 2\mu^2$. It is a free parameter of the theory, directly related to the parameter μ of the scalar potential. The second line represents the mass terms of the weak bosons (those that multiply the constant term), of mass:

$$\begin{aligned} m^2_{\text{W}} &= \frac{g^2 v^2}{4} \\ m^2_{\text{Z}} &= \frac{(g^2 + g'^2)v^2}{4} = \frac{m^2_{\text{W}}}{\cos^2\theta_W} \end{aligned} \tag{1.28}$$

It can be observed that the Goldstone bosons, removed with the transformation of Eq. (1.26), are absorbed as additional degrees of freedom of the $\text{W}^\pm$ and Z bosons, corresponding to their longitudinal polarizations: the mechanism confers mass to the weak bosons. The

second line of Eq. (1.27) also describes the interactions of the weak bosons with the Higgs field. There are a HWW and a HZZ interactions from the $2H/v$ term and a HHWW and a HHZZ interaction from the H^2/v^2 term. The third line shows that cubic and quartic self-interactions of the Higgs boson are predicted. The BEH potential can be rewritten in terms of a trilinear and a quadrilinear coupling as:

$$V(H) = \frac{1}{2}m_{\mathrm{H}}^2 H^2 + \lambda_{\mathrm{HHH}} v H^3 + \frac{1}{4}\lambda_{\mathrm{HHHH}} H^4 - \frac{\lambda}{4}v^4 \tag{1.29}$$

with the two couplings constants defined as[1]:

$$\lambda_{\mathrm{HHH}} = \lambda_{\mathrm{HHHH}} = \lambda = \frac{m_{\mathrm{H}}^2}{2v^2} \tag{1.30}$$

An important remark is that both Higgs boson self-couplings are directly related to the parameters of the scalar potential and are entirely determined from the Higgs boson mass and the VEV. Their measurement thus represents a test of the validity and coherence of the SM. In a wider perspective, the Higgs boson self-couplings have no equal in the SM: in contrast to the weak boson self-interactions, that have a gauge nature, the Higgs boson self-interactions are purely related to the scalar sector of the theory. They are responsible for the mass of the Higgs boson itself, as shown in the mass term of the Lagrangian (1.27). Their experimental determination is thus crucial to reconstruct the Higgs boson potential and explore the nature of the EWSB.

Finally, there is a constant term in the Lagrangian density of Eq. (1.27). While this is irrelevant in the SM, it contributes to the vacuum energy, which is related to the cosmological constant that determines the curvature of the Universe. The value of this constant predicted in the SM is not compatible with astronomical observations. This is a puzzle that requires either a proper quantum theory of gravity with additional interactions or a mechanism to reduce the Higgs field vacuum energy density.

There are at this point two free parameters of the BEH mechanism: the VEV v and the Higgs boson mass m_{H}. The first corresponds to the energy scale of the electroweak symmetry breaking and can be computed from the Fermi constant G_F that is precisely determined from the muon lifetime:

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g}{2\sqrt{2}}\right)^2 \frac{1}{m_{\mathrm{W}}^2} \Longrightarrow v = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246\,\mathrm{GeV} \tag{1.31}$$

Fermions have been assumed to massless until now. Mass terms are generated by the Higgs field itself through a Yukawa interaction that couples the left and right chiral fields. Denoting as ψ and ψ' the up- ($I_3 = 1/2$) and down-type ($I_3 = -1/2$) fermions, the Yukawa Lagrangian density is:

$$\mathcal{L}_{\mathrm{Yukawa}} = -y_{f'}\left(\bar{\Psi}_L \Phi \psi'_R + \bar{\psi}'_R \Phi^\dagger \Psi_L\right) - y_f\left(\bar{\Psi}_L \tilde{\Phi} \psi_R + \bar{\psi}_R \tilde{\Phi}^\dagger \Psi_L\right) \tag{1.32}$$

with

$$\tilde{\Phi} = i\sigma_2 \Phi^* = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix} \quad \xRightarrow{EWSB} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \tag{1.33}$$

[1] In the literature, the cubic Higgs interaction term is sometimes written as $(\lambda_{\mathrm{HHH}}/6)H^3$, so that the trilinear coupling is defined as $\lambda_{\mathrm{HHH}} = 3m_{\mathrm{H}}^2/v$.

Although not discussed here for simplicity, the Lagrangian of Eq. (1.32) can be generalized to include fermion mixing through combinations of the mass eigenstates of the fields Ψ_L, ψ_R and ψ'_R. The Yukawa Lagrangian density is Lorentz and gauge invariant and renormalizable, so that it can be included in the SM Lagrangian density. After EWSB, Eq. (1.32) is expressed as:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_f m_f(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)\left(1 + \frac{H}{v}\right) \tag{1.34}$$

where the sum runs on both up– and down–type fermions and the mass terms are:

$$m_{f^{(\prime)}} = y_{f^{(\prime)}}\frac{v}{\sqrt{2}} \tag{1.35}$$

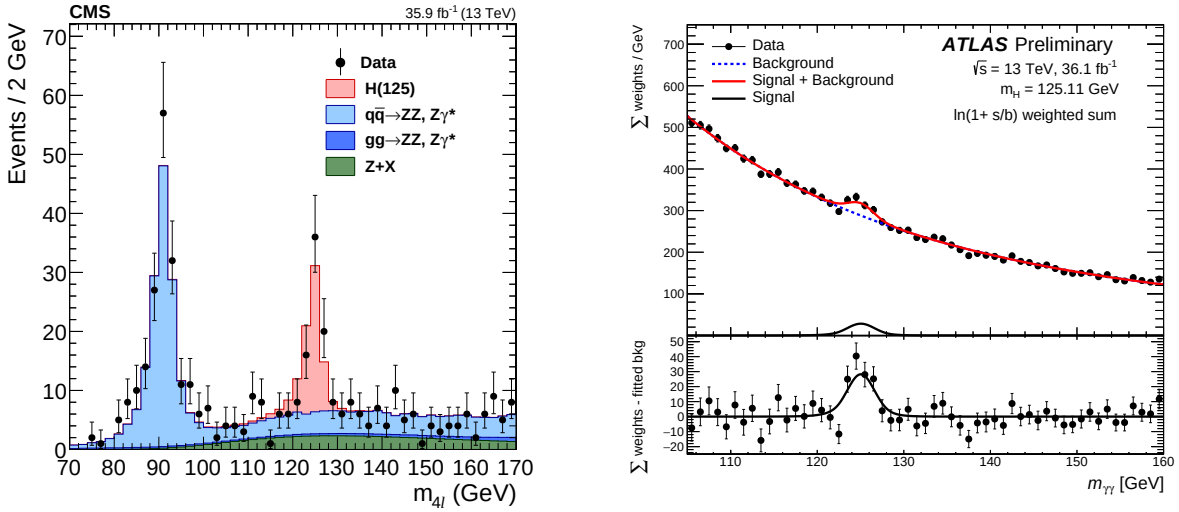
Fermion masses are thus explained in the SM as the interaction of the fermion fields with the Higgs field, which changes the chirality of the fermions. The strengths of the interactions are directly related to the fermion masses, and are free parameters of the theory. An important remark is that the SM does not explain the origin of these couplings and, consequently, the hierarchy of the three fermion families.

In conclusion, the BEH mechanism solves the aforementioned problems of the electroweak theory of the SM. Upon breaking the electroweak symmetry, the scalar field generates Goldstone bosons that are absorbed as degree of freedoms of the vector boson fields, which become massive. The Higgs boson contributions to the quantum loops in the scattering of longitudinally polarized vector bosons regularizes the process and ensure its unitarity at the TeV scale and beyond. Finally, the Higgs boson couples the left and right chiral components of the fermion fields in a Yukawa interaction, determining the fermion masses with a purely quantum-relativistic mechanism. Finally, the theory obtained by incorporating the BEH mechanism in the electroweak theory is renormalizable, as demonstrated by 't Hooft and Veltman [5].

1.1.3 The Higgs boson: phenomenology and experimental status

Experimental confirmation of the BEH mechanism came in July 2012 with the discovery of a new scalar boson of a mass of approximately 125 GeV announced by the ATLAS and CMS Collaborations [6, 7, 8]. The discovery was performed in the data collected at $\sqrt{s} = 8$ TeV (Run I) and lead by the high resolution $\text{H} \to \text{ZZ}^* \to \ell^+\ell^-\ell'^+\ell'^-$ ($\ell, \ell' = \text{e}, \mu$) and $\text{H} \to \gamma\gamma$ decay channels. The existence of this scalar particle is now firmly established and further confirmed with the data collected at $\sqrt{s} = 13$ TeV (Run II), as shown in Figure 1.1.

With the observation of m_{H} and the measurement of its mass, the last important parameter of the SM has been determined. However, the observation of this new particle only represents the first step in the exploration of the EWSB. The efforts of the experimental community are thus focusing on the characterization of this boson. The measurements of its couplings via exclusive production modes and decay channels, of its spin-parity, and of its differential production cross section, need to be thoroughly investigated to verify that they correspond precisely to the SM predictions.



(a) $\mathrm{H} \to \mathrm{ZZ}^* \to \ell^+\ell^-\ell'^+\ell'^-$ channel (CMS)

(b) $\mathrm{H} \to \gamma\gamma$ channel (ATLAS)

Figure 1.1 – Four-lepton invariant mass (a) and photon pair invariant mass (b) distributions in events collected at $\sqrt{s} = 13\,\mathrm{TeV}$. The two figures are taken from Refs. [9] and [10].

Several mechanisms contribute to the production of a Higgs boson at a pp collider. The dominant one, with a cross section of about 49 pb, is the gluon fusion (ggF) production, that proceeds through a heavy quark loop. The second most frequent mechanisms, about 10 times rarer than ggF, is vector boson fusion (VBF), where the Higgs boson is produced in association with a jet pair of large invariant mass. The third main mechanism is the production in association with a single vector boson (VH, $\mathrm{V} = \mathrm{W}^{\pm}, \mathrm{Z}$). Studying the rare VH and VBF production mechanisms allows for probing the Higgs boson coupling to vector bosons. Finally, Higgs bosons can be produced in association with a pair of top quark ($\mathrm{t\bar{t}H}$) or a single top quark (tH). The $\mathrm{t\bar{t}H}$ mechanism is of particular interest as it allows for the direct determination of the magnitude of the top Yukawa coupling y_t, in contrast with the indirect determination from ggF. The tH mechanism, about a factor of 10 rarer than the previous, allows for the determination of the sign of y_t. The cross sections of these production modes as a function of $\sqrt{s}$ are summarized in Figure 1.2a.

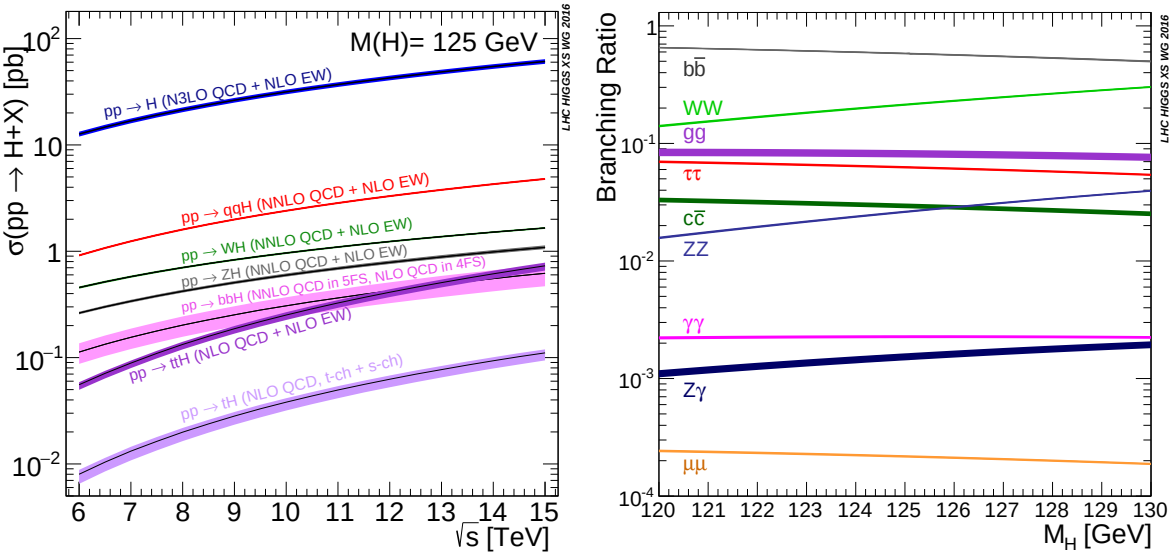
The Run I Higgs boson discovery was performed inclusively for all the production mechanisms. The combination of the high-resolution channels between the ATLAS and CMS experiments resulted in a precise determination of m_H of:

$$m_\mathrm{H} = 125.09 \pm 0.21(\mathrm{stat.}) \pm 0.11(\mathrm{syst.})\,\mathrm{GeV} \tag{1.36}$$

Exclusive production modes are now being probed and, in particular, there is increasing evidence for $\mathrm{t\bar{t}H}$ production [11, 12].

Higgs boson decays branching fractions are shown in Figure 1.2b as a function of m_H, and are summarized in Table 1.2 for a Higgs boson of mass $m_\mathrm{H} = 125.09\,\mathrm{GeV}$. The aforementioned $\mathrm{H} \to \mathrm{ZZ}^*$ and $\mathrm{H} \to \gamma\gamma$ decay modes are amongst the rarest but are experimentally advantageous because of the high signal-to-background (S/B) ratio and the excellent invariant mass resolution. The mass $m_\mathrm{H} \approx 125\,\mathrm{GeV}$ turns out to be one of

the most difficult values to reach experimentally, as all masses below about 120 GeV and above 130 GeV had been excluded with 2011 data [13]. However, after the Higgs boson discovery, this specific mass value provides several final states that can be simultaneously probed. Decays of the Higgs boson to $W^{\pm}W^{\mp *}$ or $Z\gamma$, as well as to fermions in the $H \to \tau^+\tau^-$, $H \to b\bar{b}$, $H \to \mu^+\mu^-$, and $H \to c\bar{c}$ decay channels, can thus be eventually probed at the LHC. The latter is challenging because of the low S/B ratio and the tiny branching fractions, but of particular interest as it represents the only direct way to probe Higgs boson decays to up-type fermions.



(a) Production cross section

(b) Decay branching fraction

Figure 1.2 – (a) Higgs boson production cross section as a function of $\sqrt{s}$ for different production mechanisms. (b) Branching fractions of the decay of a Higgs boson as a function of m_H.

Decay mode	$\mathcal{B}$ [%]
$H \to b\bar{b}$	$58.09^{+0.72}_{-0.73}$
$H \to W^{\pm}W^{\mp *}$	$21.52^{+0.33}_{-0.33}$
$H \to gg$	$8.18^{+0.42}_{-0.42}$
$H \to \tau^+\tau^-$	$6.27^{+0.10}_{-0.10}$
$H \to c\bar{c}$	$2.88^{+0.16}_{-0.06}$
$H \to ZZ^*$	$2.641^{+0.040}_{-0.040}$
$H \to \gamma\gamma$	$0.2270^{+0.0047}_{-0.0047}$
$H \to Z\gamma$	$0.1541^{+0.0090}_{-0.0090}$
$H \to \mu^+\mu^-$	$0.02171^{+0.00036}_{-0.00037}$

Table 1.2 – Branching fractions of the main Higgs boson decay modes for a SM Higgs boson of mass $m_H = 125.09$ GeV. Theoretical uncertainties combine the uncertainties on the Higgs boson partial width, on the value of α_s, and on the quark masses.

The Higgs boson was also shown to have a spin-parity $J^P = 0^+$ [14], and the combination of several decay channels indicated that its decay rates and coupling strengths are compatible with the SM expectation [15]. The combined measurement performed by the ATLAS and CMS experiments [16] further confirmed the agreement with the SM predictions. The exploration of the Higgs boson properties with Run II data continues to reveal important information about this particle. The existence of decays to fermions has been established in the $\tau^-\tau^+$ decay mode by the CMS experiment [17], complementing the previous Run I observation from the combination of the two experiments [16]. Further information on the couplings to fermions is also coming from the evidence for the decay of the Higgs boson to $\mathrm{b\bar{b}}$ [18, 19], as well as from the stringent upper limits on the decay branching fraction to muon pairs [20] that amount to 2.9 times the SM prediction.

The investigation of the Higgs boson properties is showing good agreement with the SM predictions. Although constraints on its couplings partially depend on assumptions on BSM contributions and still leave some space for possible deviations [16], the scalar boson discovered five years ago is today precisely known and is compatible within the uncertainties with a SM Higgs boson. Most importantly, it has been experimentally observed that this particle breaks the degeneracy between the three fermion families by coupling proportionally to their mass. As summarized in Figure 1.3, couplings of the Higgs boson are probed over about three orders of magnitude and the dependence of their strength on the boson and fermion masses is established. However, one important element is missing in this figure: the Higgs boson itself. With its mass now known with precision, the value of its self-coupling can be computed from Eq. (1.30) to be $\lambda_{\mathrm{HHH}} \approx 0.13$, completely determined in the SM from m_{H} and v. Experimentally measuring λ_{HHH} would allow to verify if this coupling really fits in the global scheme illustrated in Figure 1.3 at an ordinate of $\sqrt{2\lambda_{\mathrm{HHH}}} = m_{\mathrm{H}}/v \approx 0.51$, providing a test of the validity of the SM. As this coupling is responsible for the Higgs boson mass itself, it is related to the very fundamental properties of the EWSB and of the BEH mechanism.

The λ_{HHH} coupling can be directly probed in Higgs boson pair (HH) production. Similarly, the measurement of the quadrilinear coupling λ_{HHHH}, a further probe of the BEH potential, requires the study of triple Higgs final state. The production of the latter is however extremely rare in the SM, with a cross section of about 80 ab at $\sqrt{s} = 14\,\mathrm{TeV}$ [21], out of the experimental reach of the LHC. In contrast, HH production, although challenging, can be experimentally probed at the LHC.

It has also been recently suggested that λ_{HHH} could be determined from precision measurements at the LHC. Its effects could be observed either in electroweak precision observables [22], or from precision measurements of single Higgs boson production, where the radiative corrections due to the trilinear coupling can be sizeable [23]. In the latter case, significant information can be extracted from both the total cross section and the differential Higgs boson p_{T} distribution [24, 25]. However, as λ_{HHH} is completely determined from m_{H} and the vacuum expectation value v, these indirect constraints need to assume a variation of the trilinear coupling which often depends on the theoretical assumptions through which this is realized. Moreover, some authors point out that the constraint from a global fit of single Higgs boson measurements are subject to degeneracies that can be only solved by including information from HH production [26]. The direct determination of λ_{HHH} from HH production is thus an essential step in the understanding of the BEH mechanism and, for this reasons, it represents one of the main goals of the LHC physics programme.

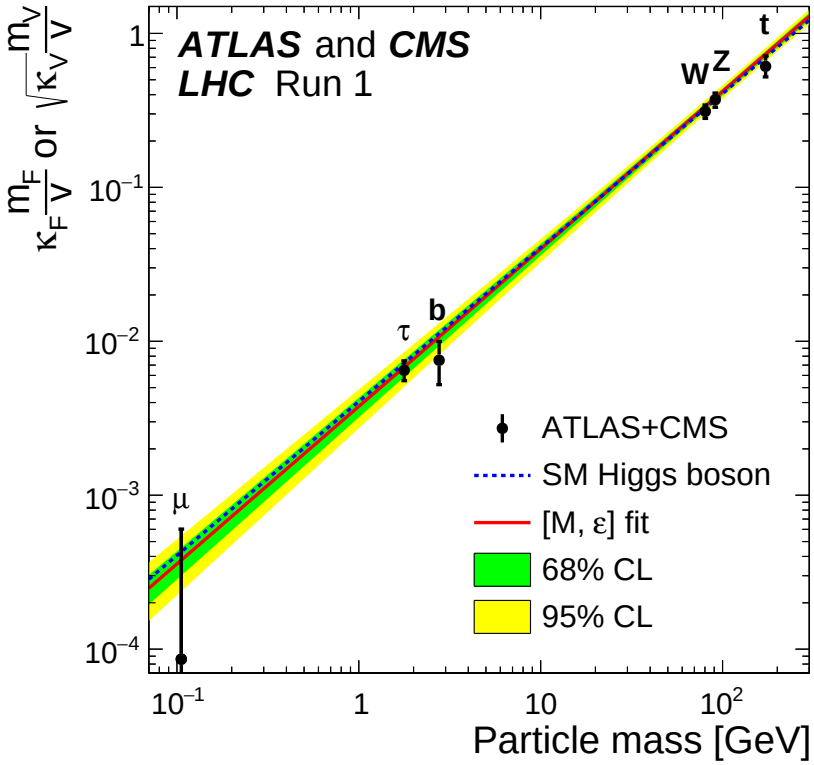
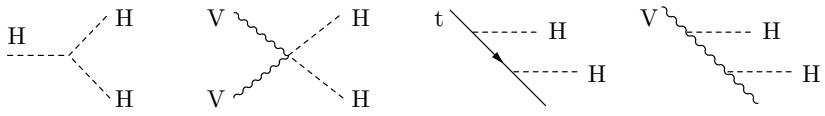


Figure 1.3 – Normalized Higgs boson coupling constant as a function of the boson or fermion mass. The figure is taken from Ref. [16]

1.2 Higgs boson pair production in the SM

It is known since a long time that λ_{HHH} can be extracted from the measurement of the Higgs boson pair production cross section. The role of the trilinear coupling in this context was highlighted back in 1988, well before the Higgs boson discovery, when the first computation of the cross section was performed [27].

However, the λ_{HHH} coupling represents only one of the possible interactions that contribute to HH production. In general, a pair of on-shell Higgs bosons can be produced in the final state of a collision through any of the following diagrams:

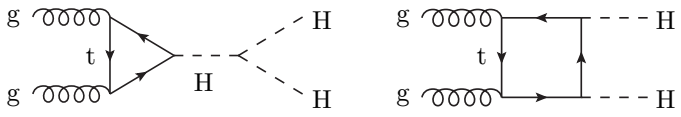


The production mechanisms that are detailed in this section are characterized by different combinations of these interactions, in which the λ_{HHH} contribution must be disentangled from other effects.

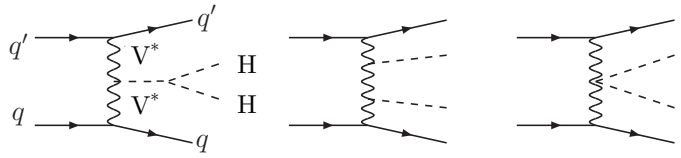
1.2.1 Production mechanisms

At the LHC, Higgs boson pairs can be produced through the five main mechanisms [28, 29] that are listed below in decreasing order of their cross section. Some representative Feynman diagrams illustrate the Higgs boson couplings involved.

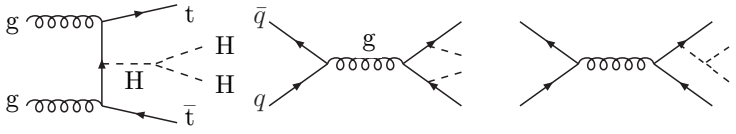
- **gluon fusion production** $gg \to HH$. It involves either the production of a Higgs boson pair through the trilinear Higgs boson self-coupling, or the radiation of two on-shell Higgs bosons from a heavy quark loop. The cross section consequently depends on λ_{HHH} and on the top quark Yukawa couplings y_t. The contribution from b quarks is smaller than 1% at leading order and can be neglected given the current accuracy of the theoretical computations and the experimental sensitivity.



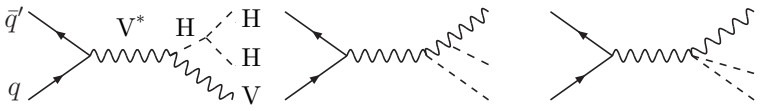
- **vector boson fusion (VBF) production** $qq' \to jjHH$. In addition to the Higgs boson trilinear coupling, this process also depends on the quadrilinear coupling of a Higgs boson pair to a vector boson pair as well as on the single Higgs boson coupling to vector bosons. Despite its cross section being one order of magnitude smaller than the gluon fusion one, the two final state jets provide a clean signature that can be used to discriminate signal events from background.



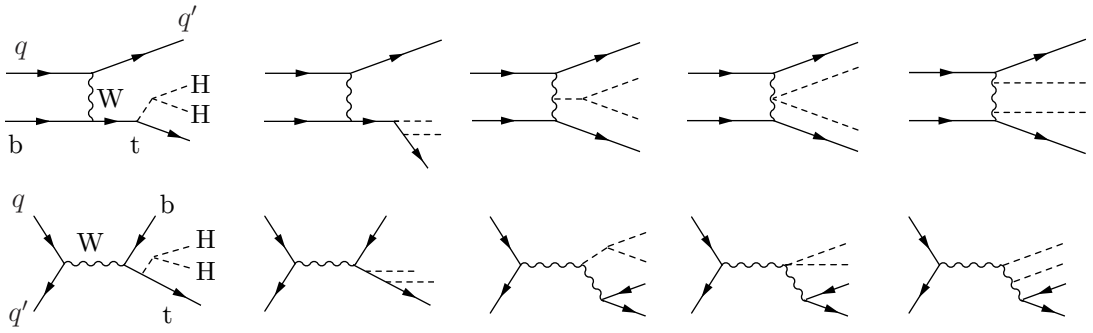
- **top quark pair associated production** $qq'/gg \to t\bar{t}HH$. It is a HH variant of the single Higgs boson pair production in association to a top quark pair ($t\bar{t}H$), where either two Higgs bosons are radiated from the top quarks or are produced from the Higgs boson self-coupling. Its cross section exceeds the one from VBF HH production at high transverse momenta of the HH pair and for high centre-of-mass energies.



- **Vector boson associated production** $qq' \to VHH$ with $V = W^{\pm}, Z$. It involves the same Higgs boson couplings as VBF production, but an on-shell vector boson is present in the final state.



- **single top quark associated production** $qq' \to \mathrm{t}j\mathrm{HH}$. It can proceed through either the t- or s-channel, that are respectively illustrated in the top and bottom row of the diagrams below. The t-channel diagrams are illustrated for simplicity in the so-called 5F scheme [30]. It is the only process that is sensitive at the same time to the HH couplings to vector bosons and to top quarks and to their relative phase. However, its cross section is so small that it can hardly be investigated at the LHC, but could be studied in a future higher energy collider.



The cross sections of these production mechanisms at different centre-of-mass energies are summarized in Table 1.3. The cross section for gluon fusion is computed at the next-to-next-to-leading order (NNLO) of the theoretical perturbative QCD calculation, including next-to-next-to-leading-logarithm (NNLL) corrections and finite top quark mass effects at next-to-leading order (NLO). The theoretical uncertainties include uncertainties in the QCD factorization and renormalization scales, α_{S}, parton distribution functions (PDF), and unknown effects from the finite top quark mass at NNLO. The cross sections for VHH are computed at the NNLO and those of the other processes at the NLO of the perturbative QCD calculation.

A graphical comparison of the cross sections as a function of the centre-of-mass energy is shown in Figure 1.4. HH production is in general very rare at the LHC. As a consequence, experimental searches, including the one presented in this thesis, focus on the dominant gluon fusion production, as other production mechanisms seem presently out of reach. Nevertheless, there is an interest of going beyond gluon fusion: VBF HH production could for example provide additional handles for the measurement of λ_{HHH} and give access to the VVHH interaction that is currently unexplored. While this surely a possibility for future HH searches, it is not investigated further here, and the symbol σ_{HH}, whenever not ambiguous, will thus be used in the following to denote the gluon fusion HH production cross section.

An important property of the gluon fusion production channel should be highlighted at this point. The two production diagrams discussed before have amplitudes that are about the same order of magnitude, but interfere destructively. Combined with the restricted phase space of production of two Higgs bosons, this results in the small cross section discussed above. However, the destructive interference makes HH production extremely sensitive to physics beyond the SM (BSM). BSM physics contributions might alter the destructive interference and produce large modifications that can be probed with the current LHC data. We thus move from the perspective of HH production as a test of the SM, to the one of HH production as a probe of BSM physics, which is the topic of the next section.

Production mode	σ[fb]
	$\sqrt{s} = 8\,\text{TeV}$
Gluon fusion	$10.15^{+4.1\%}_{-5.7\%}$ (scale) $\pm 3.1\%$(PDF) $\pm 2.6\%(\alpha_s) \pm 5.0\%$(top)
VBF	$0.459^{+3.2\%}_{-3.6\%}$ (scale) $\pm 2.6\%$(PDF + α_s)
$t\bar{t}HH$	$0.174^{+2.8\%}_{-10.6\%}$(scale) $\pm 3.9\%$(PDF + α_s)
W^+HH	$0.145^{+0.43\%}_{-0.52\%}$(scale) $\pm 2.8\%$(PDF + α_s)
W^-HH	$0.0677^{+1.0\%}_{-1.2\%}$ (scale) $\pm 3.7\%$(PDF + α_s)
ZHH	$0.143^{+2.7\%}_{-2.3\%}$ (scale) $\pm 2.6\%$(PDF + α_s)
$tjHH$	$0.00540^{+5.4\%}_{-3.1\%}$ (scale) $\pm 5.6\%$(PDF + α_s)
	$\sqrt{s} = 13\,\text{TeV}$
Gluon fusion	$33.49^{+4.3\%}_{-6.0\%}$ (scale) $\pm 2.1\%$(PDF) $\pm 2.3\%(\alpha_s) \pm 5.0\%$(top)
VBF	$1.62^{+2.3\%}_{-2.7\%}$ (scale) $\pm 2.3\%$(PDF + α_s)
$t\bar{t}HH$	$0.772^{+1.7\%}_{-4.5\%}$ (scale) $\pm 3.2\%$(PDF + α_s)
W^+HH	$0.329^{+0.32\%}_{-0.41\%}$(scale) $\pm 2.2\%$(PDF + α_s)
W^-HH	$0.173^{+1.2\%}_{-1.3\%}$ (scale) $\pm 2.8\%$(PDF + α_s)
ZHH	$0.362^{+3.4\%}_{-2.6\%}$ (scale) $\pm 1.9\%$(PDF + α_s)
$tjHH$	$0.0281^{+5.2\%}_{-3.2\%}$ (scale) $\pm 4.5\%$(PDF + α_s)
	$\sqrt{s} = 14\,\text{TeV}$
Gluon fusion	$39.59^{+4.4\%}_{-6.0\%}$ (scale) $\pm 2.1\%$(PDF) $\pm 2.2\%(\alpha_s) \pm 5.0\%$(top)
VBF	$1.95^{+1.8\%}_{-2.3\%}$ (scale) $\pm 2.4\%$(PDF + α_s)
$t\bar{t}HH$	$0.949^{+1.8\%}_{-4.8\%}$ (scale) $\pm 3.2\%$(PDF + α_s)
W^+HH	$0.368^{+0.33\%}_{-0.39\%}$(scale) $\pm 2.1\%$(PDF + α_s)
W^-HH	$0.197^{+1.2\%}_{-1.3\%}$ (scale) $\pm 2.7\%$(PDF + α_s)
ZHH	$0.414^{+3.5\%}_{-2.7\%}$ (scale) $\pm 1.8\%$(PDF + α_s)
$tjHH$	$0.0364^{+3.7\%}_{-1.3\%}$ (scale) $\pm 4.7\%$(PDF + α_s)

Table 1.3 – Cross section for different HH production modes assuming a Higgs boson mass of 125.09 GeV. The gluon fusion cross section is computed at NNLO of perturbative QCD calculation, with NNLL corrections and finite top quark mass effects at NLO. The cross section of the VHH, $V = W^{\pm}, Z$ are computed at NNLO QCD and those of the other processes at NLO QCD. The values are taken from Ref. [21].

1.3 Beyond the SM

Theoretical considerations and experimental results indicate that the SM is incomplete. Once compared to astrophysical observations and combined with cosmological models, the SM cannot provide a suitable mechanism that is responsible for the matter-antimatter asymmetry in the Universe, nor it predicts the existence of a particle species that is compatible with the dark matter, which is observed from its gravitational effects. The SM is also unsatisfactory under three main theoretical aspects deeply related to Higgs boson physics. First, it cannot currently provide an explanation for the existence of three families of fermions, identical under all aspects but for their couplings with the Higgs boson, that span over several orders of magnitude. Secondly, the mass of the Higgs boson

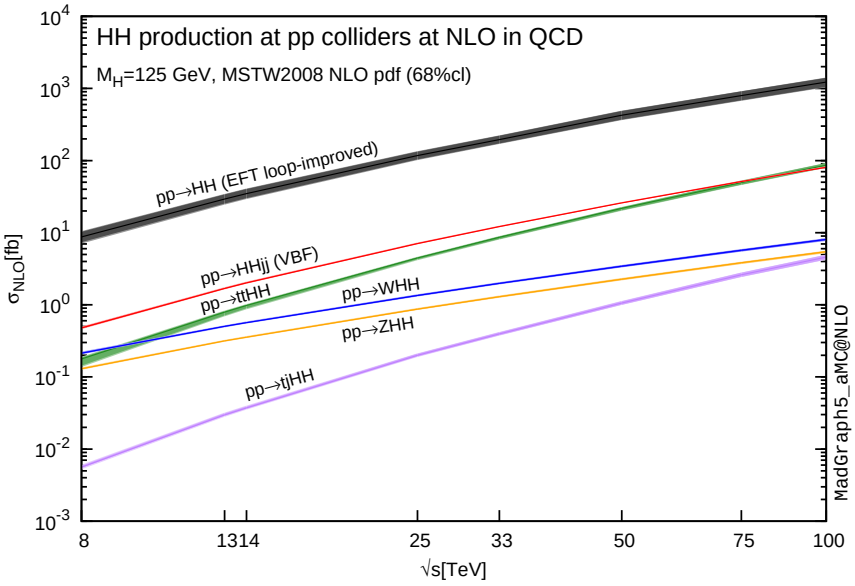


Figure 1.4 – Total cross section for HH production in pp collisions for the production modes described in the text. The cross sections are computed at the NLO accuracy and the bands shown the linear combination of the theoretical errors on the scale and PDF uncertainties. The figure is taken from Ref [29].

is not protected by any fundamental symmetry of the SM and it is subject to quadratically divergent radiative corrections, that call for a mechanism to stabilize them to avoid an unnatural fine tuning. Finally, the validity of the theory up to the Planck scale can only be ensured if the scalar potential is bounded from below, which guarantees the stability of the vacuum. From the values of m_{H} and m_{t} presently measured, a metastability condition of the scalar potential appears to be favoured [31], challenging the long term existence of the electroweak vacuum. The stability of the Higgs potential at higher energy scales is also related to its possible role in the inflation of the primordial Universe [32, 33]. Both the vacuum stability and the role of the Higgs as an inflaton depend on the shape of the scalar potential, that is determined from the running value of the BEH potential parameter λ. This parameter is deeply connected to the Higgs boson self-couplings which are one of the main topics investigated in this thesis.

It is natural in this context to think that the SM is only the manifestation of a more extended theory beyond it, that exists below the Planck scale and regulates the aforementioned problems of the SM. The presence of BSM physics could provide a solution to these problems by changing profoundly the structure of the SM while preserving its incredible success at describing the phenomenology of collider experiments until now.

In this context, HH production is both a probe for BSM physics, and a balance to discriminate between possible alternatives. If the scale of BSM physics is at the LHC reach, new states can be directly produced and subsequently decay to a HH pair. The experimental signature of this resonant production mechanism is an enhancement of σ_{HH} at a specific value of m_{HH}, corresponding to mass of the resonance. If instead the scale of BSM physics is significantly higher than the LHC centre-of-mass collision energy, its effects could still be observed as a nonresonant enhancement of the production cross section, due to either new particles in the quantum loops or to anomalous Higgs boson couplings.

1.3.1 Resonant BSM HH production

Final states with a Higgs boson pair can be used to generically probe BSM physics models predicting the existence of a new resonance X of mass $m_X > 2m_H$ that couples significantly to the Higgs boson. The presence of a common and clear signature suggests the possibility of a model-independent approach, that is adopted for the search described in this thesis. The results subsequent reinterpretation of the results provides a general and efficient way to explore a broad class of BSM physics models.

The profound relation between HH production and the properties of the BEH mechanism suggest that resonant HH signatures can appear in models with an extended scalar sector, or in models with warped extra dimensions that alter the relation between the Higgs field and the matter fields. Some of this BSM physics scenarios are discussed here. However, this section is not meant to be an extensive summary of all the possible models predicting a resonant HH production. The discussion instead focuses on a few representative examples to show that, despite their different theoretical assumptions, they can simultaneously be probed in HH production. An important remark is that, depending on the model, the mass of the resonance can range from the kinematic threshold of 250 GeV up to several TeV. From the experimental point of view, this requires the development of complementary analysis methods and the analysis of several final states to ensure a high acceptance over the entire mass range.

Higgs singlet model

The most simple extension of the scalar sector is the Higgs singlet model [34, 35, 36], that postulates the existence of an additional Higgs singlet S together with the Higgs doublet Φ:

$$\begin{aligned} \Phi &= \begin{pmatrix} \phi_+ \\ \tilde{\phi}_0 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \frac{\phi_0 + v}{\sqrt{2}} \end{pmatrix} \\ S &= \frac{s + \langle S \rangle}{\sqrt{2}} \end{aligned} \tag{1.37}$$

The corresponding potential is described by 5 parameters:

$$V = -m^2 \Phi^\dagger \Phi - \mu^2 S^2 + \lambda_1 (\Phi^\dagger \Phi)^2 + \lambda_2 S^4 + \lambda_3 \Phi^\dagger \Phi S^2 \tag{1.38}$$

where a Z_2 symmetry $\Phi \to \Phi$ and $S \to -S$ is assumed. After EWSB, both $\tilde{\phi}_0$ and S acquire a VEV, which in analogy of the BEH mechanism results in the existence of two physical fields, corresponding to a heavy and a light scalar. These are conventionally denoted as h and H, where the lightest scalar is interpreted as the Higgs boson H. The usage of italic and roman symbols is intended here to partially remove the conflict between the SM and this BSM model notation. In general, it is assumed here that $m_H > m_h \approx m_H = 125\,\text{GeV}$. The (H, h) notation is used in this context whenever not ambiguous.

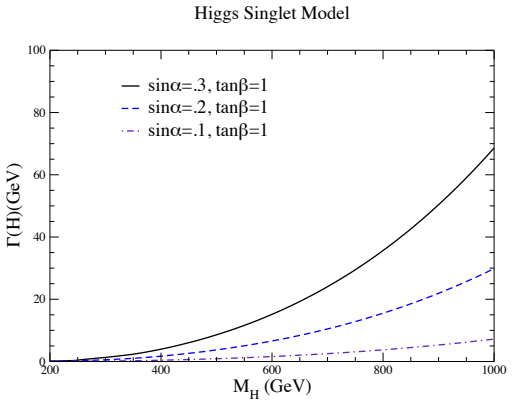
The two physical fields are mixtures of the original fields:

$$\begin{aligned} h &= \cos\alpha\, \phi_0 - \sin\alpha\, s \\ H &= \sin\alpha\, \phi_0 + \cos\alpha\, s \end{aligned} \tag{1.39}$$

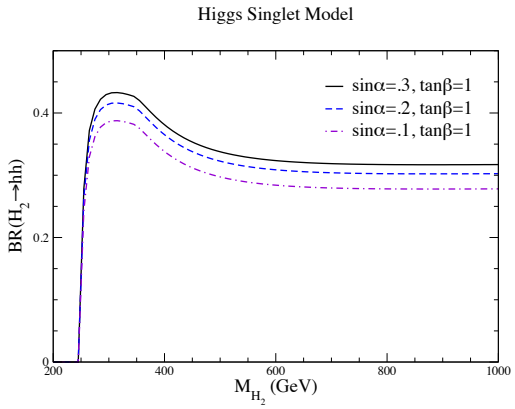
A modification of the Higgs boson trilinear coupling and the presence of a Hhh coupling are predicted, resulting in the LO relations:

$$\begin{aligned} \lambda_{hhh} &= \frac{-3m_h^2}{v}\left(\cos^3\alpha - \tan\beta\sin^3\alpha\right) \\ \lambda_{Hhh} &= \frac{-m_h^2}{v}\sin(2\alpha)(\cos\alpha + \sin\alpha\tan\beta)\left(1 + \frac{m_H^2}{2m_h^2}\right) \end{aligned} \tag{1.40}$$

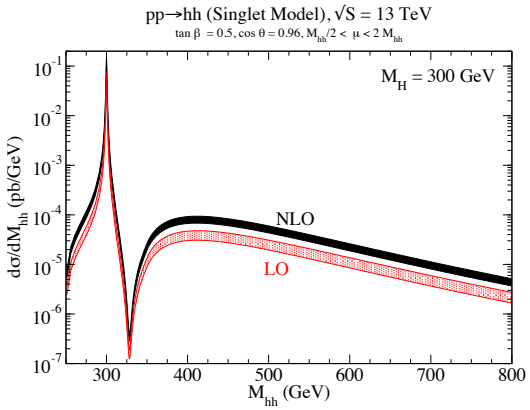
where $\tan\beta = v/\langle s\rangle$ is the ratio of the VEV of the two fields. A constraint $|\cos\alpha| > 0.94$ follows from single Higgs boson measurements [37] and combined constraints on $\cos\alpha$ and m_H come from W boson mass measurements [38, 39]. The width of the H resonance and its decay branching fractions to a hh pair are shown in Figures 1.5a and 1.5b. The latter are sizeable for resonance masses up to the TeV scale. Therefore, a clear experimental signature of the singlet Higgs model is a resonant enhancement of the m_{hh} differential cross section, as illustrated in Figure 1.5c. The finite width of the resonance H also results in interference effects with nonresonant production that could be used to further characterize a data excess.



(a) H width



(b) $H \to hh$ branching fraction



(c) Differential cross section in m_{hh}

Figure 1.5 – (a) intrinsic width, (b) hh decay branching fraction of the heavier scalar H, and (c) expected m_{hh} differential distribution in the Higgs singlet model for a few representative values of the parameters. The figures are taken from Refs. [21, 39].

2HDM and MSSM

Two-Higgs-doublet models (2HDM) postulate the existence of an additional Higgs doublet field. Such extension of the scalar sector is of particular interest since the presence of two Higgs doublets H_u and H_d is required in low-energy supersymmetric (SUSY) scenarios to break the electroweak symmetry. In this sense, 2HDM models describe the Higgs phenomenology of SUSY theories and represent the contact point between the two sectors.

The extension of the scalar potential with the addition of a Higgs doublet can be realized in two main ways by requiring that all quark fields couple with the same doublet (type I 2HDM) or that right-handed quarks of charge $+2/3$ and $-1/3$ couple to two distinct doublets (type II 2HDM). The most general 2HDM form have a very rich vacuum structure, with a corresponding variety of phenomenological consequences that are extensively described in Ref. [40]. For the scope of this work, it is possible to focus on the minimal supersymmetric extension of the SM (MSSM), which Higgs sector is represented by the type II 2HDM.

In the MSSM, the addition of a complex Higgs doublet implies the existence of five new particles: two CP-even (i.e scalars) neutral h and H, one CP-odd (i.e pseudoscalar) A, and two charged scalars H^+ and H^-. As for the singlet Higgs model, it is assumed that $m_H > m_h$ and one of the two particles (usually the lightest one) is interpreted as the scalar boson observed at the LHC. At the tree level, the model can be completely described in terms of two parameters, that are usually chosen as the mass of the pseudoscalar m_A and the ratio of the VEV of the two fields $\tan\beta = v_u/v_d$. The mass of the charged scalar correspond to $m^2_{H^\pm} = m^2_A + m^2_{\mathrm{W}}$, while the mass of the neutral scalars are give from the mass matrix:

$$\mathcal{M}^2_{\text{tree}} = m^2_{\mathrm{Z}} \begin{pmatrix} \cos^2\beta & -\sin\beta\cos\beta \\ -\sin\beta\cos\beta & \sin^2\beta \end{pmatrix} + m^2_A \begin{pmatrix} \sin^2\beta & -\sin\beta\cos\beta \\ -\sin\beta\cos\beta & \cos^2\beta \end{pmatrix} \tag{1.41}$$

that once diagonalized by a rotation of an angle α results in the masses of the two scalars:

$$m^2_{h,H} = \frac{1}{2}\left(m^2_A + m^2_{\mathrm{Z}} \mp \sqrt{(m^2_A + m^2_{\mathrm{Z}})^2 - 4m^2_A m^2_{\mathrm{Z}}\cos^2 2\beta}\right) \tag{1.42}$$

The introduction of an additional doublet produces deviations of the couplings of the Higgs bosons and, more importantly for this context, implies the presence of a Hhh coupling that produces a resonant $H \to hh$ signature.

The Run I data collected at the LHC strongly constrain the MSSM parameter space. In particular, the non-observation of SUSY particles and of additional charged or scalar particles, and the measured Higgs boson mass and its couplings, disfavour high $\tan\beta$ values. This motivates the exploration of the low $\tan\beta$ regions where $H \to hh$ production is crucial to conquer more territory in the MSSM parameter space.

This exploration requires to take into account radiative corrections from top quarks and their supersymmetric partners (stops). These introduce a dependence on the underlying SUSY models and make significantly more complex the tree-level parametrization. In particular, the mass matrix of the neutral scalars is modified as:

$$\mathcal{M}^2 = \mathcal{M}^2_{\text{tree}} + \begin{pmatrix} \Delta\mathcal{M}^2_{11} & \Delta\mathcal{M}^2_{12} \\ \Delta\mathcal{M}^2_{12} & \Delta\mathcal{M}^2_{22} \end{pmatrix} \tag{1.43}$$

The tree level parametrization in terms of m_A and $\tan\beta$ can be recovered by taking into

account the experimental observation of a scalar boson, corresponding to an additional input $m_h \approx 125\,\text{GeV}$ to the model. Two alternative approaches exist to include this information in a MSSM scenario [41].

The first one, denoted as the hMSSM [42, 43], combines the input m_h value with the approximation $\Delta\mathcal{M}^2_{11} = \Delta\mathcal{M}^2_{12} = 0$ and the assumption that SUSY particles are heavy enough to have negligible effects on the Higgs sector apart from the $\Delta\mathcal{M}^2_{22}$ term itself. Under these assumptions, the hMSSM represents effectively a "model-independent" parametrization of the MSSM that can be expressed in terms of the usual m_A and $\tan\beta$ parameters.

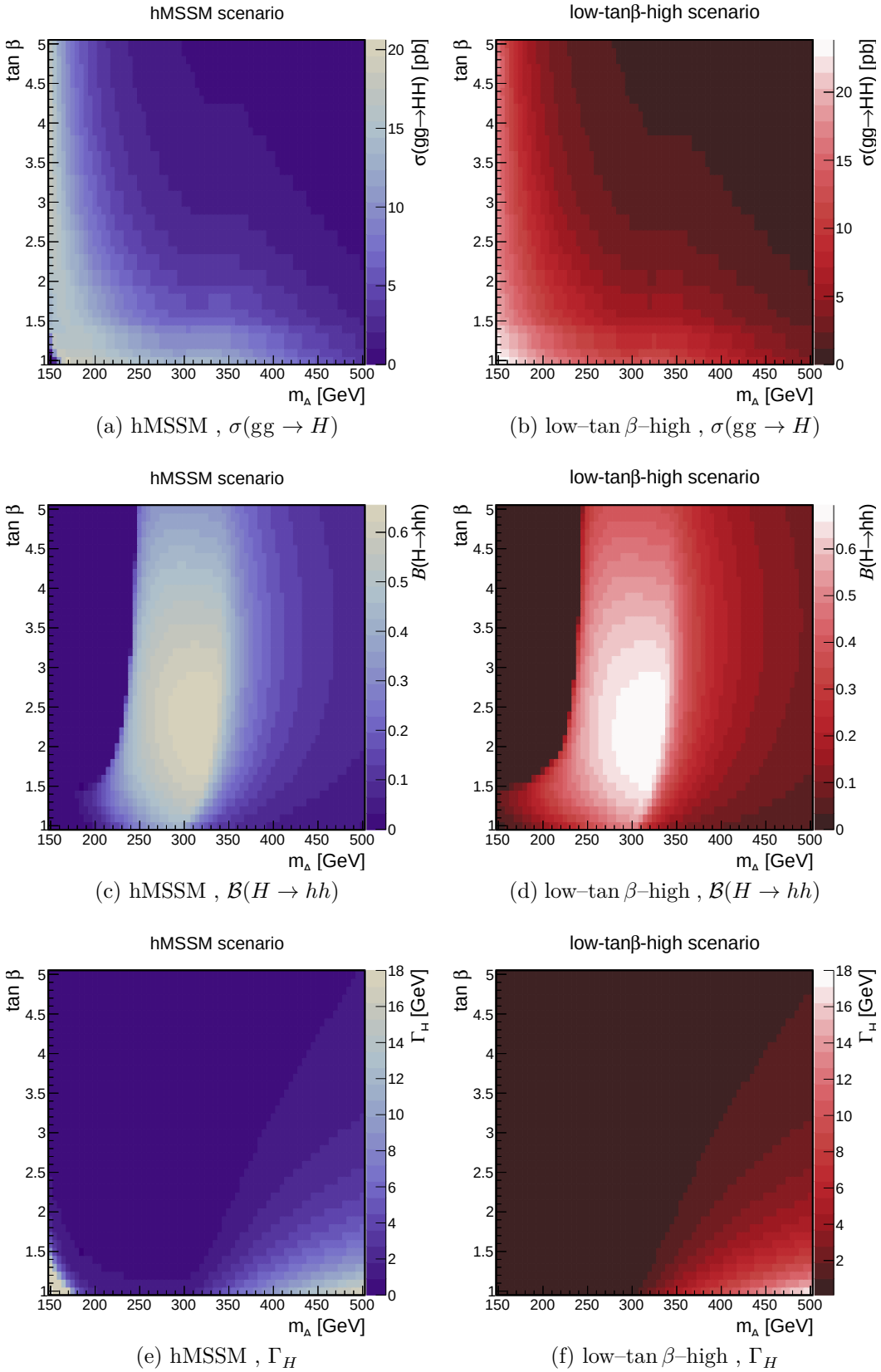
A second approach, denoted as the "low–$\tan\beta$–high" scenario [44], proceeds in the opposite way and scans the possible combinations of SUSY parameters to find those that are compatible with the measured Higgs boson mass $m_h \approx 125\,\text{GeV}$. The scan is performed under the assumptions that, as the name of the scenario suggests, the values of $\tan\beta$ are low and supersymmetric particles have a masses that are high enough to have negligible impact on the Higgs boson decays.

The predictions of the two models for the production cross section $\sigma(\text{gg} \to H)$ and the decay branching fraction $\mathcal{B}(H \to hh)$ are shown in Figure 1.6. Although numerical difference exist, both models show similar properties, with a cross section that decreases as m_A and $\tan\beta$ increase. The decay branching fraction to hh has a maximum for $m_A \approx 300\,\text{GeV}$ and $\tan\beta \approx 2$, where searches for Higgs boson pair production are expected to be the most sensitive. The upper left region is kinematically forbidden because it corresponds to $m_H < 250\,\text{GeV}$, while for $m_H > 2m_\text{t}$ the decay to top quark pair becomes kinematically allowed and the branching fraction of the decay to hh decreases quickly. The effects of the mass degeneracy $m_H \approx m_A$ for large $\tan\beta$ values that is typical of MSSM models can be clearly observed in these figures. The intrinsic width of the H resonance becomes relevant as the $\text{t}\bar{\text{t}}$ decay channel opens up. However, this effect is negligible with respect to the experimental resolution on m_H in the search performed in the $hh \to \text{b}\bar{\text{b}}\tau^+\tau^-$ decay channel that is presented in this thesis, allowing for a MSSM interpretation of results derived under a narrow resonance hypothesis.

A summary of the excluded regions of the hMSSM plane from Run I searches is shown in Figure 1.7. As anticipated, hh final states are crucial to cover the low $\tan\beta$, low m_A region. The combined reduction of the cross section and of the decay branching fractions make the low $\tan\beta$, high m_A regions experimentally hard to probe.

Warped extra dimensions

Proposals for a spacetime with more than three spatial dimensions have since long time been formulated as an attempt to unify the forces of Nature. Since the first idea by Kaluza and Klein back in the 1920's, extra dimensions have become an important part of quantum gravity theories and found a key role in string theories. In particular, as explained by Arkani-Hamed, Dimopoulos and Dvali in 1998 [46], the weakness of the gravitational force can be explained by its propagation through these additional dimensions. Of particular interest for the collider phenomenology [47, 48] is the mechanism proposed by Randall and Sundrum (RS) [49] where the extra dimensions are compactified ("warped extra dimensions") between two points of the space referred to as "branes". One brane corresponds to the Planck scale (M_Pl) and the other to the electroweak scale, and the region separating them, the "bulk", is controlled by an exponential metric. The



(a) hMSSM , $\sigma(\mathrm{gg} \to H)$ (b) low–tan β–high , $\sigma(\mathrm{gg} \to H)$

(c) hMSSM , $\mathcal{B}(H \to hh)$ (d) low–tan β–high , $\mathcal{B}(H \to hh)$

(e) hMSSM , Γ_H (f) low–tan β–high , Γ_H

Figure 1.6 – Cross section for H production (upper panels), decay branching fraction to hh (central panels), and width of the H resonance (bottom panels) for the hMSSM (left panels, blue) and the low–tan β–high (right panels, red) scenarios. The values are shown as a function of the m_A and tan β parameters of the model.

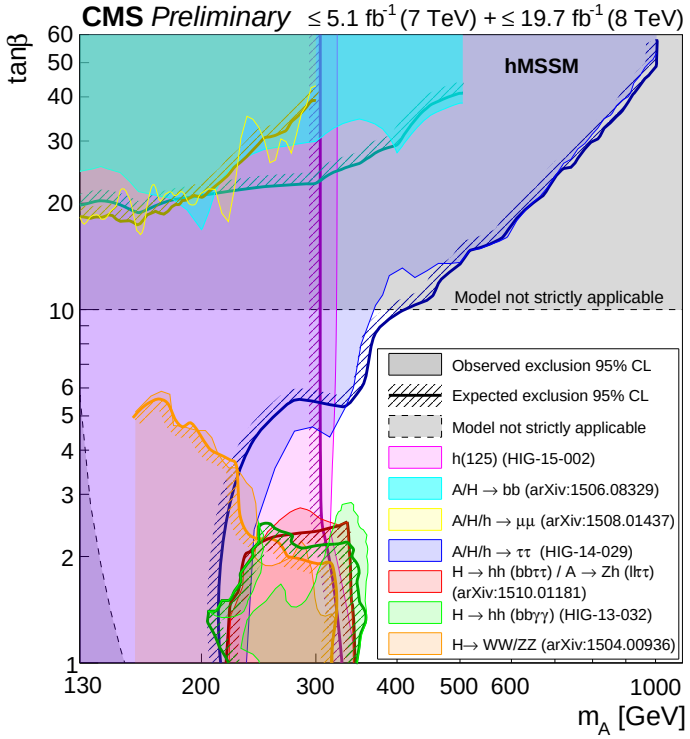


Figure 1.7 – Summary of the $(m_A, \tan\beta)$ regions excluded at the 95% confidence level in Run I searches performed by the CMS experiment. The figure is taken from Ref. [45].

gap between these two fundamental scales of the Nature is parametrized by a warp factor k.

The consequence of these models is the existence of new particles of spin 2 ("graviton", G) and of spin 0 ("radion", R) than can decay to a pair of Higgs bosons. The former is the mediator of the gravitational force itself, while the latter is required to stabilize the size of the extra dimension l. The phenomenology varies whether only the gravitational fields or the SM fields as well are allowed to propagate in the bulk. In the former case, denoted as RS1 model, couplings of the graviton to matter fields are determined by $\tilde{k} = k/\overline{M}_{\mathrm{Pl}}$, with $\overline{M}_{\mathrm{Pl}} = M_{\mathrm{Pl}}/\sqrt{8\pi}$. In the latter case, the bulk RS model, the couplings depend on the localization of the SM fields in the bulk. This class of models is theoretically interesting because it allows for a Higgs sector at the TeV scale while, at the same time, allowing for high-energy unification of gauge couplings and providing a natural hierarchy of masses. In particular, light quarks can be localized close to the Planck brane while top quarks are close to the TeV brane (elementary top hypothesis), providing a large top quark mass. The properties of the radion resonance are similar in the two models, and usually parametrized in terms of $\tilde{k}$ and of $\Lambda_{\mathrm{R}} = \sqrt{6}\exp(-kl)\overline{M}_{\mathrm{Pl}}$, the latter usually interpreted as the ultraviolet cutoff scale of the theory. The graviton production cross section is proportional to $\tilde{k}^2$ and the radion cross section is proportional to $1/\Lambda_{\mathrm{R}}^2$ [50]. A mixing between the radion and the Higgs boson is in principle possible, although this possibility is not explored here. The intrinsic total width of the radion is inversely proportional to Λ_{R}^2, and is typically experimentally negligible in low resolution channels such as $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^+\tau^-$ for radion masses up to the TeV scale. Most importantly, the decay channel $\mathrm{R} \to \mathrm{HH}$ has a large branching fraction for $m_{\mathrm{R}} > 250\,\mathrm{GeV}$, with a value of $\mathcal{B} \approx 24\%$ that is constant as a

function of m_{R} and has little dependence on the model parameters [51]. Together with the $\mathrm{R} \to \mathrm{W^+W^-}$ and $\mathrm{R} \to \mathrm{ZZ}$, that have similar branching fraction, $\mathrm{R} \to \mathrm{HH}$ is one of the dominant decay modes resulting in a large sensitivity from HH searches. Similarly, the decay branching fraction of a graviton to a HH pair is approximately constant as a function of m_{G} for $m_{\mathrm{G}} \gtrsim 500\,\mathrm{GeV}$ and can be as large as 10% depending on the model considered and of its parameters. Searches at the LHC must therefore be able to investigate both spin hypotheses in a broad mass range.

1.3.2 Nonresonant BSM HH production

BSM physics signatures in HH final states can also be probed in nonresonant HH production. If the scale of BSM physics is sufficiently high that resonances cannot be directly produced, its effects can be observed through the contributions in the quantum loops responsible for HH production. Experimental signatures are in this case an enhancement of the HH production rate and a sizeable modification of the Higgs boson pair kinematic properties. In the following, the HH phenomenology in presence of an anomalous trilinear Higgs boson coupling is first discussed, and serves as in introduction to a more general approach based on an effective SM field theory.

Anomalous λ_{HHH} coupling

The value of λ_{HHH} is completely determined in the SM once the values of v and m_{H} are known. However, several BSM models predict a modification of the trilinear Higgs boson self-coupling, modifying the properties of HH production. This can provide the first hints at the LHC of the presence of BSM physics, and serve as an important criterion to discriminate between alternative models [52]. In this context, a parametric approach is adopted and consist in considering the λ_{HHH} value in Eq. (1.29) as a free parameter. Deviations from the SM prediction are quantified with the ratio $k_\lambda = \lambda_{\mathrm{HHH}}/\lambda_{\mathrm{HHH}}^{\mathrm{SM}}$. This coupling rescaling approach is usually referred to as κ–framework. It is adopted in the aforementioned study of λ_{HHH} constraints via single Higgs measurements [23], and an interpretation of the experimental results in terms of k_λ opens up the possibility to set combined constraints on the value of the trilinear Higgs boson self-coupling.

The λ_{HHH} coupling is largely underconstrained by experimental data. Theoretical arguments based on the unitarity of the theory with a modified coupling can be used to constrain its value, but crucially depend on the assumptions under which the SM potential is deformed. Bounds of $|k_\lambda| \lesssim 6$ [53, 26], $|k_\lambda| \lesssim 8\pi/3$ [54, 55], or $|k_\lambda| \lesssim 20$ [56] are set depending on the underlying assumption. In general, large absolute values of k_λ are allowed.

As is illustrated in Figure 1.8, the modification of the value of λ_{HHH} has a profound impact on the HH production cross section. The interference between the different diagrams contributing to HH production results in a minimum of the cross section that is not centred at $k_\lambda = 0$. The dependence of the gluon fusion cross section on k_λ is quadratic. Taking the more general case, further discussed in the next section, where also the Higgs boson Yukawa coupling to the top quark can vary as $k_{\mathrm{t}} = y_{\mathrm{t}}/y_{\mathrm{t}}^{\mathrm{SM}}$, the cross section is parametrized by the following function [57]:

$$\frac{\sigma_{\mathrm{HH}}}{\sigma_{\mathrm{HH}}^{\mathrm{SM}}}(13\,\mathrm{TeV}) = 0.28\,k_\lambda^2 k_{\mathrm{t}}^2 - 1.37\,k_\lambda k_{\mathrm{t}}^3 + 2.09\,k_{\mathrm{t}}^3 \tag{1.44}$$

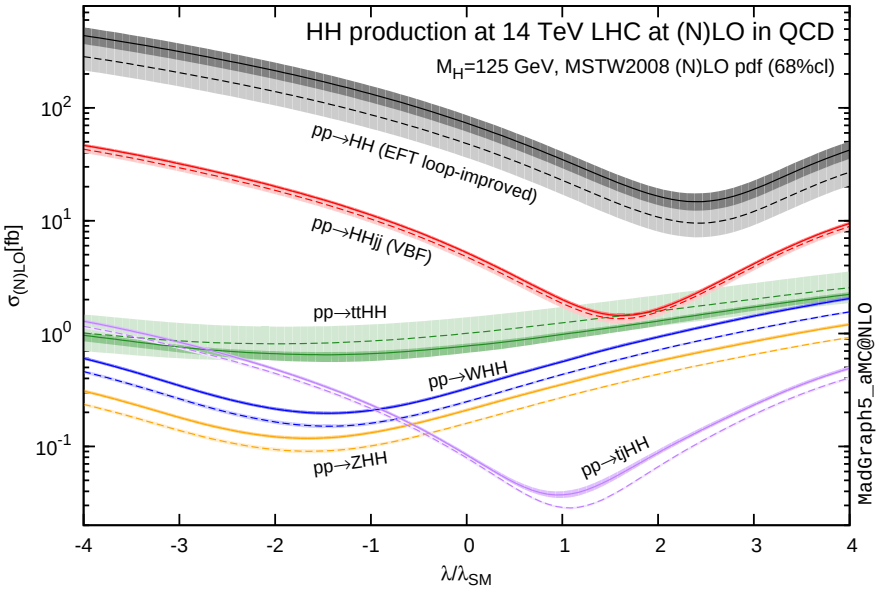


Figure 1.8 – HH production cross section as a function of the coupling modifier $k_\lambda = \lambda_{\mathrm{HHH}}/\lambda_{\mathrm{HHH}}^{\mathrm{SM}}$ for several production mechanisms. The dashed and solid lines denote respectively the LO and NLO predictions and the bands indicate the PDF and scale uncertainties added linearly. The figure is taken from Ref. [29].

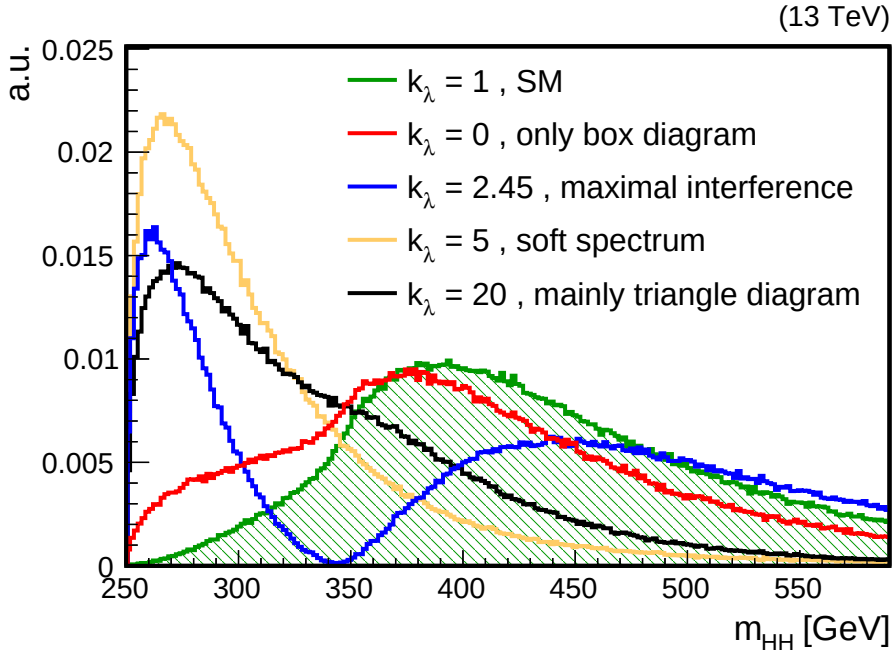


Figure 1.9 – Comparison of the m_{HH} distribution for different values of k_λ.

that has a minimum for $k_\lambda/k_{\mathrm{t}} = 2.45$. Gluon fusion is the dominant production mode also in presence of anomalous couplings and is consequently the only one investigated in the following.

Variations of the k_λ value not only affect the cross section but also the HH pair kinematics, as illustrated in Figure 1.9 for the m_{HH} distribution under a few representative values of k_λ. The SM case, corresponding to the green curve, shows a broad peak about

$m_{\mathrm{HH}} = 400\,\mathrm{GeV}$. Its shape is the result of the interference of the two "triangle" and "box" diagrams described in Section 1.2.1. The former, that involves λ_{HHH}, has an important role to populate the low m_{HH} region (black curve, $k_\lambda = 20$) while the latter significantly contributes to the high m_{HH} tail (red curve, $k_\lambda = 0$). The interference effects strongly influence the m_{HH} distribution as λ_{HHH} is modified. A soft m_{HH} spectrum is observed for $k_\lambda \approx 5$ (yellow curve), while in correspondence of the maximal interference for $k_\lambda = 2.45$ (blue curve) a characteristic double peak structure is observed. These effect have important consequences for the experimental searches, that are sensitive to anomalous λ_{HHH} couplings through both the total HH production cross section and the kinematic distribution of HH events.

Effective field theory

In the previous section λ_{HHH} has been treated as a free parameter and allowed to vary from the SM prediction. This has the advantage to cover multiple BSM scenario from a simple parametrization of the induced coupling modifications at the TeV scale. Results can be subsequently reinterpreted in a specific model through a comparison for the predicted λ_{HHH} deviations. A generalization of this approach with a systematic method is provided by the effective field theory (EFT). If the scale of BSM physics is assumed to be beyond the direct reach of the LHC, we can approximate its effects through an addition of higher order operators to the $d \leq 4$ SM Lagrangian. These additional operators are suppressed by powers of a scale Λ. From a bottom-up perspective, Λ can be interpreted as the scale up to which only SM fields propagate, while from a top-down perspective it is the energy scale of the BSM physics itself. The theory thus obtained is not renormalizable, but this does not constitute a problem in this context as an EFT only represents the lower energy manifestation of a more extended (and renormalizable) theory at higher scales. Considering a universal flavour structure and no CP violation, there is only one dimension–5 operator that has the effect of introducing neutrino masses $m_\nu \propto v^2/\Lambda^2$. It can be neglected in this context, so that dimension–6 operators are relevant and the EFT Lagrangian can be written as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^6 + \cdots \tag{1.45}$$

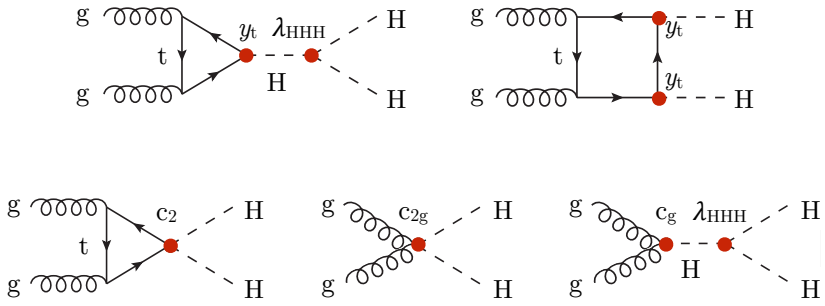
and the BSM physics is fully parametrized in terms of the Wilson coefficients c_i. Once the EFT defined, any UV-complete BSM model can be matched to it, i.e. reduced to its lower scale manifestation to derive an expression of the c_i coefficients in terms of the fundamental model parameters. From an experimental point of view, Eq. (1.45) provides a generic parametrization to investigate several BSM signatures with a model-independent approach.

In the context of HH production, a relevant EFT can be constructed as detailed in Ref. [58]. Following the procedure in Ref. [21], the EFT Lagrangian can be rewritten in terms of effective Higgs boson couplings to provide a simple physics interpretation of the effects of dimension–6 operators. The relevant terms of the Lagrangian for HH processes

initiated by gluon-fusion are given by:

$$\begin{aligned}\mathcal{L}^{\mathrm{HH}} = &\frac{1}{2}\partial_\mu H \partial^\mu H - \frac{m_{\mathrm{H}}^2}{2} H^2 - k_\lambda \lambda^{\mathrm{SM}} v H^3 \\ &- \frac{m_{\mathrm{t}}}{v}\left(v + k_{\mathrm{t}} H + \frac{\mathrm{c}_2}{v} HH\right)(\bar{t}_L t_R + \mathrm{h.c.}) \\ &+ \frac{\alpha_s}{12\pi v}\left(\mathrm{c_g} H - \frac{\mathrm{c}_{2\mathrm{g}}}{2v} HH\right) G^A_{\mu\nu} G^{A,\mu\nu}\end{aligned} \tag{1.46}$$

The physical interpretation of this Lagrangian is the presence of anomalous λ_{HHH} and y_{t} couplings and of three BSM contact interactions representing ttHH (c_2), ggHH ($\mathrm{c}_{2\mathrm{g}}$), and ggH ($\mathrm{c_g}$) vertices. In a linear realization of the EWSB, the relation $\mathrm{c}_2 = -\mathrm{c_g}$ holds. The box and triangle diagrams involved in gluon fusion can thus be modified with respect to their SM value and three new diagrams are predicted at the same perturbative order. The diagrams involved at LO in HH production via gluon fusion are shown below, where the BSM couplings are highlighted in red.



The cross section for HH production, normalized to the SM prediction, under the combination of these five diagrams can be generically parametrized from the square of their amplitude sum as:

$$\begin{aligned}R_{\mathrm{HH}} = \frac{\sigma_{\mathrm{HH}}}{\sigma_{\mathrm{HH}}^{\mathrm{SM}}} \overset{\text{LO}}{=} &A_1 k_{\mathrm{t}}^4 + A_2 \mathrm{c}_2^2 + \left(A_3 k_{\mathrm{t}}^2 + A_4 \mathrm{c_g}^2\right) k_\lambda^2 + A_5 \mathrm{c}_{2\mathrm{g}}^2 \\ &+ \left(A_6 \mathrm{c}_2 + A_7 k_\lambda k_{\mathrm{t}}\right) k_{\mathrm{t}}^2 + \left(A_8 k_{\mathrm{t}} k_\lambda + A_9 \mathrm{c_g} k_\lambda\right) \mathrm{c}_2 \\ &+ A_{10} \mathrm{c}_2 \mathrm{c}_{2\mathrm{g}} + \left(A_{11} \mathrm{c_g} k_\lambda + A_{12} \mathrm{c}_{2\mathrm{g}}\right) k_{\mathrm{t}}^2 \\ &+ \left(A_{13} k_\lambda \mathrm{c_g} + A_{14} \mathrm{c}_{2\mathrm{g}}\right) k_{\mathrm{t}} k_\lambda + A_{15} \mathrm{c_g} \mathrm{c}_{2\mathrm{g}} k_\lambda\end{aligned} \tag{1.47}$$

that in the limit $\mathrm{c}_2 = \mathrm{c_g} = \mathrm{c}_{2\mathrm{g}} = 0$ reduces to the $k_\lambda, k_{\mathrm{t}}$ parametrization of Eq. (1.44). The coefficients A_i are determined from a simultaneous fit of the cross section obtained from a LO simulation [57]. The total gg $\to$ HH cross section is computed as the product of R_{HH} with the HH cross section value computed at NNLO+NNLL with finite top quark mass effects at NLO reported in Table 1.3. Uncertainties on the A_i coefficients from PDF and α_{s} are found to be below the 1% level and are covered, together with missing orders of the BSM cross section, by the SM cross section uncertainty.

Anomalous Higgs boson couplings can not only largely enhance the cross section for HH production, but modify significantly the kinematic properties of HH events. The modelling of the HH kinematics is done by generalizing the parametrization of Eq. (1.47) and is discussed in detail in Section 5.2 of Chapter 5.

The separate contribution of each diagram is illustrated in Figure 1.10. It should be noted that the contribution from the triangle diagram cannot be isolated by setting to zero the other couplings, as its amplitude squared depends quadratically on y_t. However, as already illustrated in Figure 1.9, it mostly contributes to the low m_{HH} region. The diagram involving the λ_{HHH} and c_g couplings contributes as well to the low m_{HH} region while those diagrams involving c_2 and c_{2g} have significant impact to the high m_{HH} region, the latter extending significantly beyond 1 TeV. As already observed in the simple case discussed in the previous section, these five contributions have a non trivial interference that can produce a large variety of HH signal topologies.

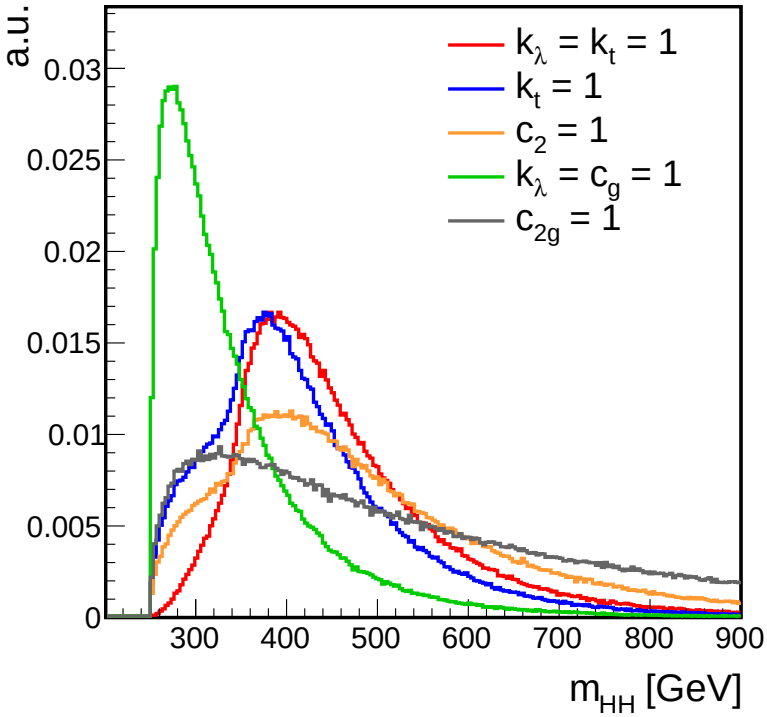


Figure 1.10 – Comparison of the m_{HH} distributions for different combinations of the BSM couplings. All the couplings not explicitly indicated in the legend are set to zero.

Exploring all the possible combinations of the five couplings is clearly not feasible for an experimental search in terms of complexity of the combinations and computing time. An approach discussed in Ref. [59] consists in defining "shape benchmarks", combinations of the five EFT parameters which topologies are representative for large regions of the five-dimensional parameter space. The shape benchmarks are defined by scanning a sample of 1507 points generated in a five-dimensional grid and by regrouping those with similar kinematic properties. The latter are completely described at LO by two parameters that are taken as m_{HH} and and the absolute value of the cosine of the polar angle of one Higgs boson with respect to the beam axis, $|\cos\theta^*|$, as discussed in more detail in Section 5.2 of Chapter 5. The similarity between two shapes is quantified through a metric defined from a binned likelihood ratio test statistics. Twelve shape benchmarks are defined with this procedure, and their corresponding shapes are shown in Figure 1.11. The corresponding

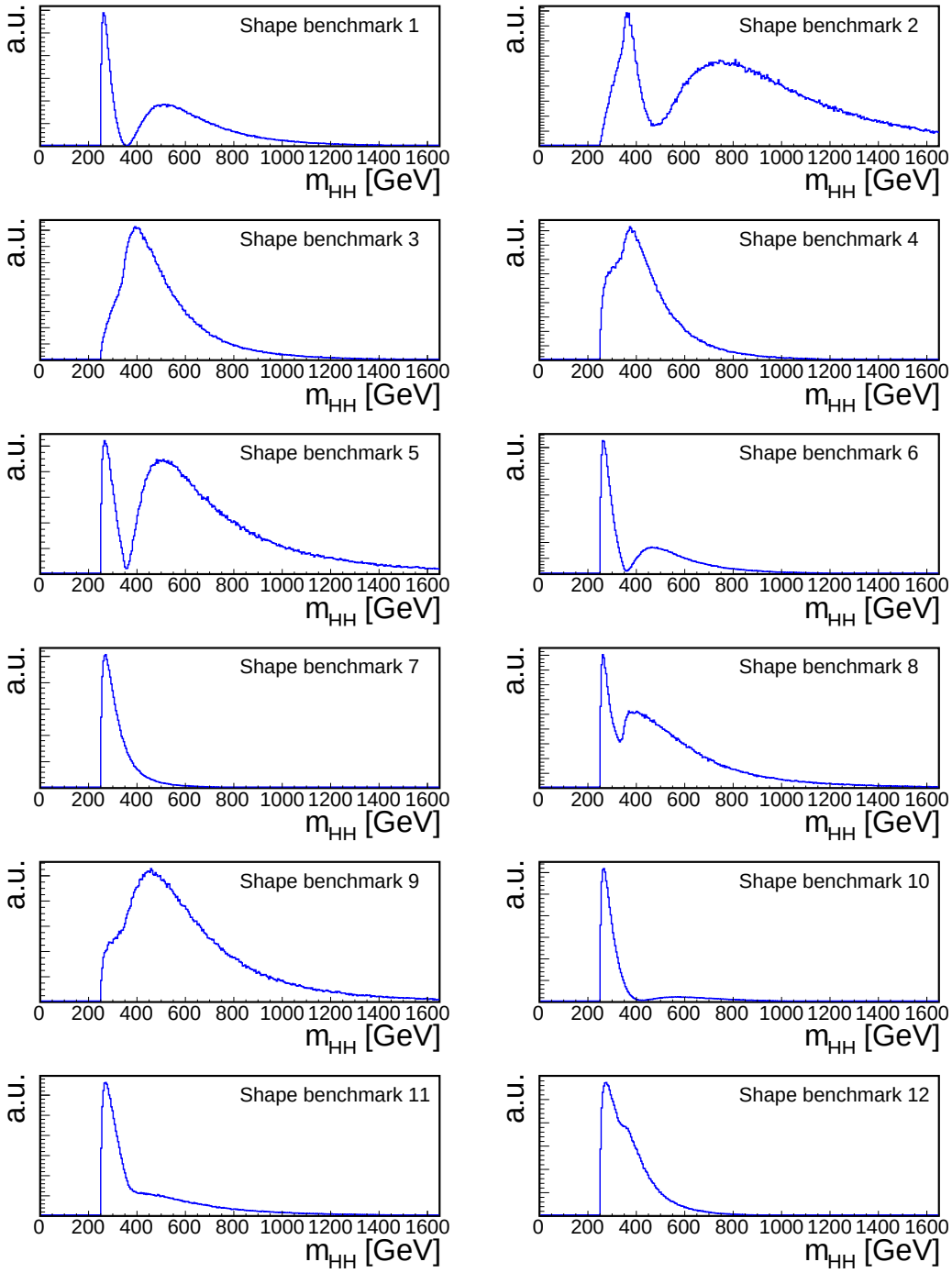


Figure 1.11 – Distribution of the Higgs boson pair invariant mass m_{HH} for the twelve shape benchmarks. Each distribution is arbitrarily normalized to a unitary area.

combination of the five couplings is reported in Table 1.4. These twelve representative shapes largely differ from the m_{HH} range that they cover and from their single or double peak structure. Some of these benchmarks, such as the number 2, have a high m_{HH} tail that extends well beyond 1 TeV, for others such as number 7 the majority of the events is at low m_{HH}.

Benchmark nr.	k_λ	k_t	c_2	c_g	$\mathrm{c}_{2\mathrm{g}}$
1	7.5	1.0	-1.0	0.0	0.0
2	1.0	1.0	0.5	-0.8	0.6
3	1.0	1.0	-1.5	0.0	-0.8
4	-3.5	1.5	-3.0	0.0	0.0
5	1.0	1.0	0.0	0.8	-1.0
6	2.4	1.0	0.0	0.2	-0.2
7	5.0	1.0	0.0	0.2	-0.2
8	15.0	1.0	0.0	-1.0	1.0
9	1.0	1.0	1.0	-0.6	0.6
10	10.0	1.5	-1.0	0.0	0.0
11	2.4	1.0	0.0	1.0	-1.0
12	15.0	1.0	1.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

Table 1.4 – Values of the effective Lagrangian couplings that define the twelve shape benchmarks.

The EFT approach and the resulting couplings in an effective Lagrangian are a useful, model-independent parametrization of BSM models. Many examples of such ultraviolet complete models that can result in anomalous nonresonant HH production are found in literature. Some examples, partially discussed in the context of resonant HH production, are multiplet extensions of the scalar sector [60, 61] and 2HDM models [62]. Other examples are constituted by vector-like quarks [63] and vector-like leptons [64], as well as from composite Higgs models [65, 66, 67, 68] where the Higgs boson is described as a bound state of an higher energy theory, in analogy to the pion in QCD. In all these models, an effective description at lower energies can realized in the EFT approach. Once the complete model is matched to the EFT, anomalous couplings arise and are directly related to the fundamental parameters of the original model. In particular, in the case of a LO matching in the models mentioned above, c_2 interactions usually arise and are accompanied by k_λ and k_t values different from one. These examples show the effectiveness of the EFT approach in parametrizing a large variety of BSM physics models, even in presence of profoundly different theoretical motivations behind the specific models.

1.4 Searching for Higgs boson pair production at the LHC

Final states with a pair of Higgs boson are phenomenologically very rich and can be explored in several decay channels. As discussed above, there is a large variety of BSM models that can manifest either in resonant or nonresonant HH production. Depending

on the specific model, Higgs bosons can have a low transverse momentum or, inversely, be highly boosted. Searches at the LHC thus need to explore several HH decay channels and to make use of complementary analysis techniques to be sensitive to this large variety of signals.

1.4.1 HH decay channels

Measuring the production of Higgs boson pairs at the LHC requires to reconstruct their decay products in the detector and to discriminate them from the large background. The choice of the decay channel of the HH system is crucial in this sense and determines a different trade-off between the branching fraction and the background contamination. Higgs boson pair production, at least in the context of the SM, is characterized by tiny cross sections, so that decay channels with a sizeable branching fraction are preferred. Referring to the single Higgs boson branching fractions of Table 1.2, this consists in requiring that least one Higgs boson decays to a $b\bar{b}$ or a $W^{\pm}W^{\mp *}$ pair. In the following, HH production and subsequent decays are considered as independent processes and, in particular, SM branching fractions are assumed. Although this might not be the case for some BSM scenarios, good agreement has been observed thus far between the measured Higgs boson coupling strengths and the SM predictions. Anomalous Higgs boson couplings to fermions and gauge bosons can be probed in single Higgs measurements thanks to the larger cross section and are of limited interest for HH searches. The decay branching fractions for some selected HH final states are shown in Figure 1.12 and those currently

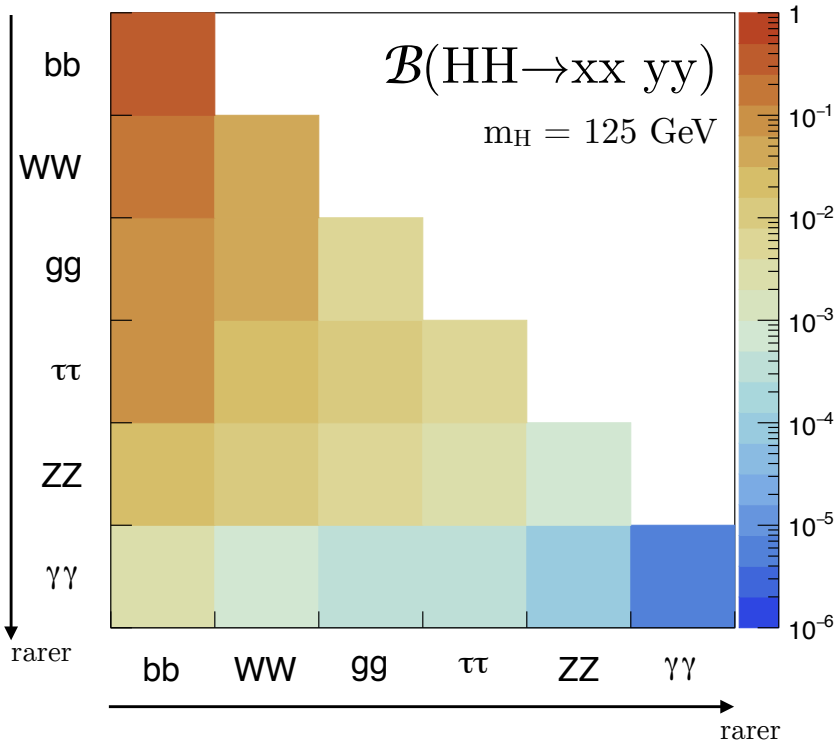


Figure 1.12 – Branching fractions for the decay of a HH pair to a selected group of final states. The decays of the two Higgs bosons are indicated in the two axes of the figure.

investigated at the LHC are summarized in Table 1.5 together with the expected number of events from SM HH production for different integrated luminosities and centre-of-mass energies.

Final state	$b\bar{b}b\bar{b}$	$b\bar{b}VV$	($\to b\bar{b}\ell^+\nu_\ell\ell^-\bar{\nu}_\ell$)	$b\bar{b}\tau^-\tau^+$	$b\bar{b}\gamma\gamma$	$W^\pm W^{\mp *}\gamma\gamma$	($\to qq'\ell\nu_\ell\gamma\gamma$)
Branching fraction	33.6%	27.9%	2.7%	7.3%	0.26%	0.098%	0.028%
N, 13 TeV, 35.9 fb^{-1}	404	335	33	88	3.1	1.2	0.33
N, 13 TeV, 300 fb^{-1}	3376	2803	271	733	26	9.8	2.8
N, 14 TeV, 3000 fb^{-1}	39907	33137	3207	8670	309	116	33

Table 1.5 – Branching fractions of the HH final states explored by the ATLAS and CMS Collaborations. In the table $\ell = e, \mu$ and $V = W^\pm, Z$. The corresponding number of events N expected for SM HH production for different centre-of-mass energies and integrated luminosities is also given.

The sensitivity to HH production at the LHC is driven by four main channels:

- $HH \to b\bar{b}b\bar{b}$ is characterized by the highest branching fraction but is affected by a copious multijet background. It has consequently little sensitivity to low m_{HH} values but can profit from the large signal yields to probe regions up to $m_{HH} \approx 3$ TeV.
- $HH \to b\bar{b}VV$ ($V = W^\pm, Z$) profits from a sizeable branching fraction and a reduced background contamination. Searches typically focus on the $b\bar{b}W^+W^-$ decay where the two W bosons decay leptonically, which reduces the branching fraction by about a factor of 10. This channel suffers from a large contamination from $t\bar{t} \to b\bar{b}W^+W^-$ irreducible background.
- $HH \to b\bar{b}\tau^-\tau^+$ represents an optimal compromise between the branching fraction and the background contamination. Contributions from the irreducible $t\bar{t}$ background are suppressed with respect to the $b\bar{b}W^+W^-$ because of the branching fraction $\mathcal{B}(W \to \tau\nu_\tau) \approx 11\%$.
- $HH \to b\bar{b}\gamma\gamma$ is a very pure final state but suffers from a small branching fraction. The clean signature of the photon pairs results in a high signal selection efficiency and provides a powerful tool to separate signal events from the background through the use of the photon pair invariant mass.

To simplify the notation, the indication of the quark and lepton charge is omitted in what follows.

Many other final states can be studied at the LHC to improve the sensitivity of experiments to HH production. Those listed above represent nevertheless the decay channels that are expected to be the most sensitive and which combination can ensure the largest coverage of the possible HH topologies. Experimental challenges are very different depending on the final state considered. The exploration of HH production in its bbbb final state crucially relies on the capability to identify jets from b quarks and to reject instrumental background from the misidentification of gluon or light flavour quark jets.

Inversely, final states such as bbVV and bb$\gamma\gamma$ are mostly affected from irreducible backgrounds, that can be statistically suppressed only by exploiting the kinematic properties of the selected events.

The bb$\tau\tau$ final state, that is the topic of this thesis, represents an intermediate and particularly interesting situation. As the tau lepton is unstable and can decay to either leptons or hadrons in associations to neutrinos, the searches must exploit several final states. Neutrinos from τ decays do not allow for a complete reconstruction of the event, and final states where the tau leptons decay to hadrons and neutrinos must be distinguished from instrumental backgrounds caused by the misidentification of a quark or gluon jet. At the same time, irreducible background contamination also affects this decay channel and calls for the usage of the event kinematic properties to reduce it. For these reasons, the bb$\tau\tau$ decay channel is probably one of the most challenging at the LHC, but the effort is rewarded by one of the best sensitivities over several resonant and nonresonant HH signal hypotheses, as pointed out in several phenomenological studies [28, 69, 70]. The properties of the bb$\tau\tau$ decay channel and the major background contributions are further discussed in Section 4.1 of Chapter 4 and throughout the rest of this thesis.

1.4.2 Previous searches for Higgs boson pair production

Searches for Higgs boson pair production at the LHC in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ (Run I) have been performed by both the ATLAS and CMS Collaborations.

The ATLAS Collaboration explored the bbbb, bb$\tau\tau$, bb$\gamma\gamma$, and WW$^*\gamma\gamma$ final states, and evaluated their combined sensitivity [71]. A modest local excess of about 2.4 standard deviations above the background expectation was observed in the bb$\gamma\gamma$ decay channel [72] but it has not been confirmed in more recent results in the same final state [73]. Consequently, the combination of the results do not show significant deviations from the SM expectation and 95% confidence level upper limits are set on both resonant and nonresonant HH production. The observed and expected upper limits corresponds to 70 and 48 times the SM prediction, respectively.

The CMS Collaboration explored and combined the bbbb, bb$\tau\tau$, and bb$\gamma\gamma$ final states [74]. Results are found to be in agreement with the SM predictions and are used to set upper limits on both the resonant and nonresonant production mechanisms. An observed upper limit of 43 times the SM HH cross section is set, for an expected upper limit of 47 times the SM prediction. Anomalous Higgs boson couplings were explored in the bb$\gamma\gamma$ final state [75], and values of the anomalous trilinear Higgs boson self-couplings $k_\lambda < -17.5$ and $k_\lambda > 22.5$ were excluded at the 95% confidence level. Resonant HH production is also probed in multilepton final states [76] and in dedicated searches for high mass resonances in the bbbb [77] and bb$\tau\tau$ [78] decay channels that make use of specific reconstruction techniques for highly boosted objects.

A comparison of the 95% confidence level upper limits on resonant HH production as a function of the resonance mass $\mathrm{m_X}$ is shown in Figure 1.13. An important remark is the complementarity of the different final states as a function of the mass value considered. The exploration of several final states is therefore necessary to probe in the most effective way BSM physics at the LHC in HH processes. This is also valid for nonresonant production, where the SM represents a particularly interesting case. The m_{HH} distribution in the SM has a broad peak about 400 GeV, a value where several final states achieve similar sensitivities, as also seen in the combination of Run I results [71, 74]. Probing many HH

signatures at the LHC is thus necessary to measure HH production and investigate the nature of EWSB.

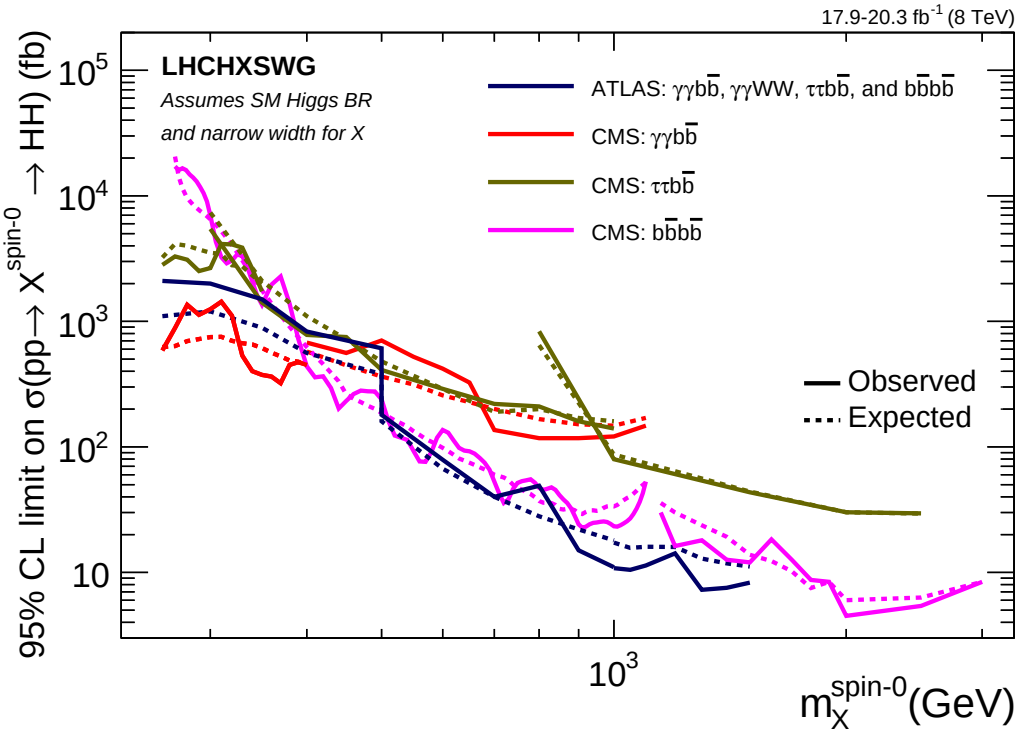


Figure 1.13 – 95% confidence level upper limits on the production cross section of resonant HH production for different final states explored by the ATLAS and CMS Collaborations at $\sqrt{s} = 8\,\text{TeV}$. The figure is taken from Ref. [21].

Chapter 2

Experimental apparatus

The European Organization for Nuclear Research (CERN) is today at the forefront of particle physics research and technology development, of international collaboration, and of education. It catalyses the scientific activity in the high energy physics domain, with about 10 000 people from more than 800 institutes and universities of 76 different countries in the world using its facilities. The CERN laboratories host the Large Hadron Collider (LHC), designed to accelerate protons inside its 26.7 km long tunnel to a centre-of-mass energy of 14 TeV. The LHC is the largest and most powerful particle accelerator ever built and represents today the frontier of the research in high energy physics.

The LHC collides the particle beams in four interaction points, instrumented with an equal number of detectors. In one of these four points is installed the Compact Muon Solenoid (CMS) experiment, designed to explore the physics at the TeV scale. The CMS Collaboration involves the participation of more than 3500 scientists from 47 different countries. It is in this unique experimental and social context that the work presented in this thesis has been developed.

This chapter introduces the properties and operations of the LHC and the structure of the CMS detector used to collect the data analysed in this thesis. The algorithms to reconstruct the particles produced in the collisions from the raw detector data are presented. Finally, the trigger system of the CMS experiment is introduced and its recent trigger system upgrade is detailed.

2.1 The Large Hadron Collider

The LHC is designed to collide protons at a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ with an instantaneous luminosity $\mathcal{L} = 10^{34}\text{cm}^{-2}\,\text{s}^{-1}$, as well as lead ions at a centre-of-mass energy of 2.76 GeV per nucleon and $\mathcal{L} = 10^{27}\text{cm}^{-2}\,\text{s}^{-1}$ [79, 80]. The realization of the LHC constituted a two decade-long international endeavour. Its first proposal dates back to 1984 with the first official recognition of the project, subsequently approved in 1994 and inaugurated in 2008.

2.1.1 Design and specifications

The LHC is installed in a 26.7 km long tunnel built between 1984 and 1989 to host the CERN Large Electron Positron (LEP) collider. The tunnel is located in the region nearby Geneva and extends across the French and the Swiss borders. In the LHC, two separate, counter-rotating particle beam lines are kept in orbit in two magnetic channels thanks to

the field generated by superconduting niobium-titanium coils. The particles are steered by a magnetic field of 8.3 T generated by a current of about 11 kA in the 1232 dipole magnets, each measuring 15 metres of length and 35 tonnes of weight. The stability of the beam dynamics is ensured by 392 quadrupoles magnets measuring 5 to 7 metres of length, that focus the particles and keep them in a narrow beam. Special quadrupoles are installed in front of the collision points to squeeze the beams and increase the proton density in the collisions. Superconducting magnets are cooled with superfluid helium-4 and kept to a working temperature of 1.9 K.

The LHC is the last element of an injection chain composed of several smaller particle accelerators, schematically represented in Figure 2.1. Hydrogen atoms are stripped of their electrons in a duo-plasmatron source and are accelerated to an energy of 50 MeV in the Linear Accelerator (LINAC2), which feeds the Proton Synchrotron Booster (PSB) where protons are accelerated to 1.4 GeV. The beam is then injected into the Proton Synchrotron (PS) for a further acceleration to 25 GeV, and subsequently into the Super Proton Synchrotron (SPS) where protons reach an energy of 450 GeV. The proton beams are finally transferred to the two LHC beam pipes, where the beams are accelerated and shaped into proton bunches thanks to radio-frequency cavities operated at 400 MHz. Once the proton reach the nominal energy and the beam dynamics is stabilized, protons are brought to collide in four points along the LHC ring.

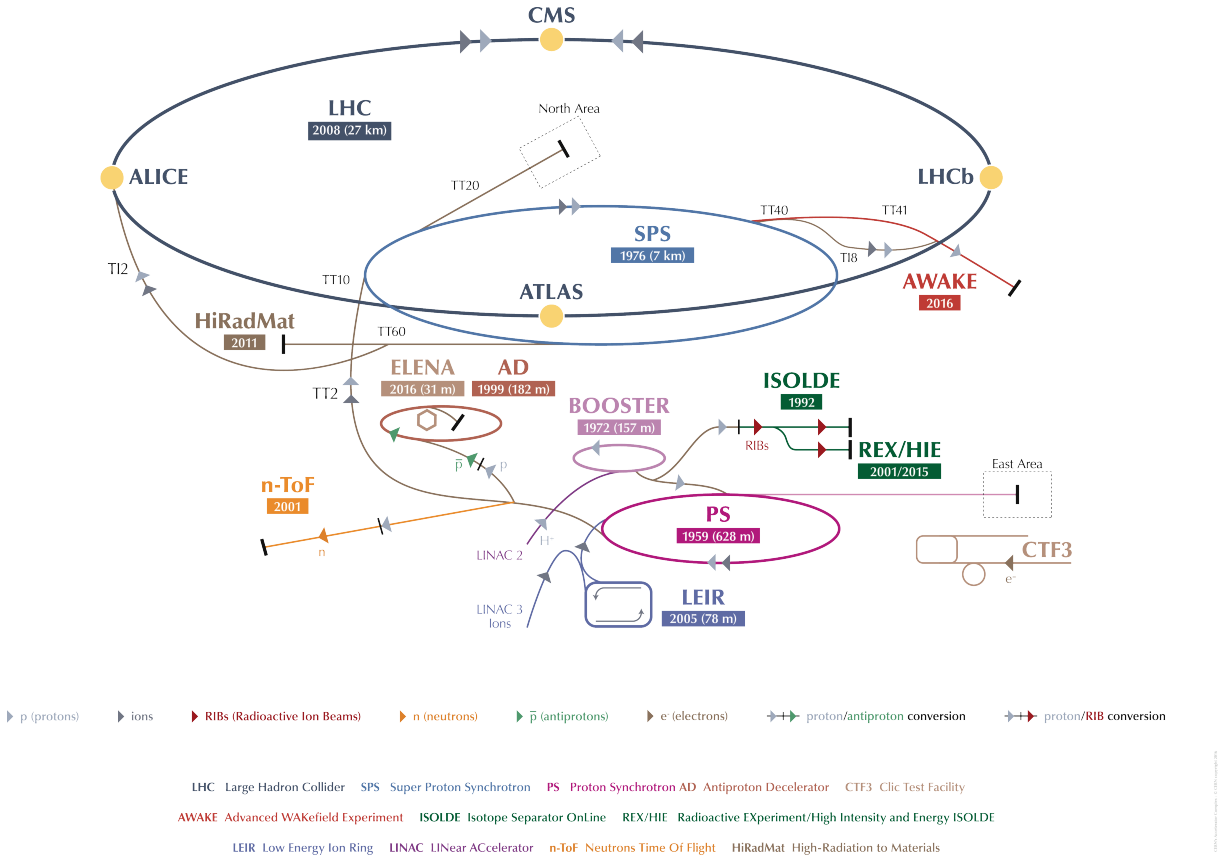


Figure 2.1 – Representation of the CERN accelerator complex [81].

An important parameter of the LHC machine is the instantaneous luminosity $\mathcal{L}$ of the collisions, that depends on the beam properties as [82]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \varepsilon_n \beta^*} F \tag{2.1}$$

N_b is the number of particles in each of the n_b bunches per beam, that revolve in the tunnel with a frequency f_{rev}. The symbol γ_r denotes the relativistic factor. The shape and focus of the beam are described by its transverse emittance ε_n and its beta function β^*, or focal length, at the collision point. The factor F accounts for the geometric reduction of the instantaneous luminosity, and depends on the beam crossing angle θ_c and on the transverse and longitudinal r.m.s. bunch sizes σ_{xy} and σ_z at the interaction point (i.p.) as:

$$F = \left(1 + \frac{\theta_c \sigma_z}{2\sigma_{xy}}\right)^{-\frac{1}{2}} \tag{2.2}$$

The nominal values of the LHC machine parameters are summarized in Table 2.1.

$\sqrt{s}$	centre-of-mass energy	14 TeV
Δt_b	bunch spacing	25 ns
N_b	particles per bunch	1.15×10^{11}
n_b	bunches per beam	2808
f_{rev}	revolution frequency	11.2 kHz
ε_n	transverse beam emittance	3.75 µm
β^*	beta function	0.55 m
θ_c	crossing angle at i.p.	285 µrad
σ_{xy}	transverse r.m.s. bunch size at i.p.	16.7 µm
σ_z	longitudinal r.m.s. bunch size	7.55 cm

Table 2.1 – Nominal parameters of the LHC machine in pp collisions.

The integrated luminosity $L = \int \mathcal{L}\,\mathrm{dt}$ is a measure of the total amount of collisions produced. LHC downtimes must be taken into account upon performing the integration, and correspond to the time needed for maintenance, filling, and ramping of the magnetic field. In 2016, the LHC achieved a record-breaking performance of colliding protons for about 70% of the time dedicated to operations; machine commissioning and winter shutdown are clearly not taken into account. The luminosity constitutes the coefficient of proportionality between the number of events N produced for a specific process and its cross section σ:

$$N = L \times \sigma \tag{2.3}$$

The four collision points of the LHC are instrumented with particle detectors installed in underground caverns. "A Toroidal LHC ApparatuS" (ATLAS) and the "Compact Muon Solenoid" (CMS) experiments are installed in the diametrically opposite Points 1 and 5 of the LHC, where the highest instantaneous luminosity of collision is produced. They are designed as hermetic, multi-purpose detectors that surround the interaction point and measure the debris of proton and ion collisions. The "LHC beauty" (LHCb), located at Point 8 is a forward, one-arm spectrometer devoted to the study of CP-violation in B

hadrons. "A Large Ion Collider Experiment" (ALICE) is installed in Point 2 and is built to study heavy ion collisions and quark-gluon plasmas.

The LHC also hosts three smaller size experiments. The "LHC forward" (LHCf) and the "TOTal Elastic and diffractive cross section Measurement" (TOTEM) experiments, located a hundred of meters away on either side from the ATLAS and CMS interaction points, are dedicated to pp interaction cross section measurements and forward diffractive physics. LHCb shares its cavern with the "Monopole and Exotics Detector at the LHC" (MOEDAL) experiment, dedicated to the search for magnetic monopoles.

2.1.2 Operations

After more than a decade of construction and installation, the first proton beam circulated in the LHC on September 10$^{\text{th}}$, 2008. An incident occurred a week later due to a faulty electrical connection between two magnets, causing the release of helium into the tunnel and mechanical damage. Repair works were promptly achieved and the LHC was back to operations in November 2009. After machine commissioning and collisions at lower energy, the first, high energy collisions took place on March 30$^{\text{th}}$, 2010. This moment marked the beginning of the so-called Run I, the fruitful data taking era that lasted until 2012. It was decided not to operate the LHC at its design parameters, and pp collisions took place at a centre of mass energy of 7 TeV, soon increased to 8 TeV, with a bunch spacing of 50 ns. About 45 pb^{-1} and 6 fb^{-1} were collected by the CMS experiment at $\sqrt{s} = 7$ TeV in 2010 and 2011, respectively, and a larger datasets of 23 fb^{-1} was recorded at 8 TeV in 2012. These data allowed for the discovery of the Higgs boson and for a precise measurement of its properties.

The LHC operations halted in 2012 for a two-year long shutdown (LS1). Important renovation and consolidation works were performed to push the LHC performance towards the design parameters. In particular, the magnets were trained to withstand higher currents for an increase of the energy per beam to 6.5 TeV. LS1 represented as well the opportunity for the experiments to complete a series of important detector upgrades to cope with the harsher collision conditions. This included the replacement of a part of the CMS trigger electronics in the context of the L1 trigger upgrade discussed in Section 2.4.3.

LHC operations restarted in 2015 at a centre-of-mass energy of 13 TeV. With $\sqrt{s}$ almost doubled with respect to the LHC Run II, the experiments can extend their exploration of the "terra incognita" of physics at the TeV scale. After a short collision phase with 50 ns bunch spacing, the nominal spacing of 25 ns was reached. Operations in 2015 were focused onto the commissioning of the LHC at the new energy, and the instantaneous luminosity was not increased beyond $5 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$. The CMS Collaboration experienced an eventful 2015 data taking with the recommissioning of the detector and the ongoing upgrade of the trigger system. The experiment suffered of problems with the cryogenic circuit of its magnet, that had to be switched off for a part of the operations. An integrated luminosity of 2.9 fb^{-1} was collected with a 3.8 T magnetic field.

In 2016, the LHC was ready to deliver a large dataset to the experiments. The instantaneous luminosity rose up to $1.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$, beyond the original machine design specifications. An integrated luminosity of 12.9 fb^{-1} was collected by July 2016 and an unprecedented 35.9 fb^{-1} dataset was recorded over the full year. These numbers refer to the integrated luminosity collected with the CMS detector and validated for a use in physics analyses, and include dead times of the experiment. Data collected in 2016 are

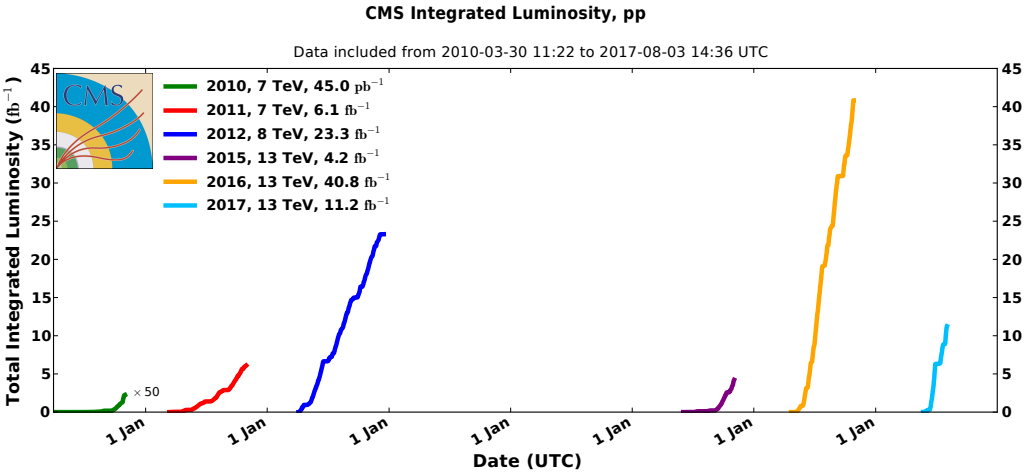
used to derive the results presented in this thesis. The 2016 LHC operations can be regrouped into several periods, labelled with a letter from A to H. Period A was devoted to the development and commissioning of the machine for the 2016 restart, while periods from B to H were used for the physics. As shown in Table 2.2, 2016 pp operations lasted from the end of April to the end of October, when the LHC underwent one month of pPb collisions, and the instantaneous luminosity was progressively increased. This can also be seen from the summary of the LHC performance in terms of peak instantaneous and integrated luminosities shown in Figure 2.2, where the machine record-breaking performance in 2016 can be observed.

2016 period	Time	LHC fills	Peak $\mathcal{L}$ $\times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$	L delivered fb^{-1}
B	28 Apr–21 Jun	4879–5030	8.4	6.1
C	24 Jun–4 Jul	5038–5071	10.0	3.2
D	4 Jul–15 Jul	5072–5095	10.4	4.6
E	15 Jul–25 Jul	5096–5117	12.2	4.6
F	29 Jul–14 Aug	5134–5198	12.6	3.4
G	14 Aug–16 Sep	5199–5303	13.3	8.5
H	16 Sep–28 Oct	5304–5471	15.3	10.0

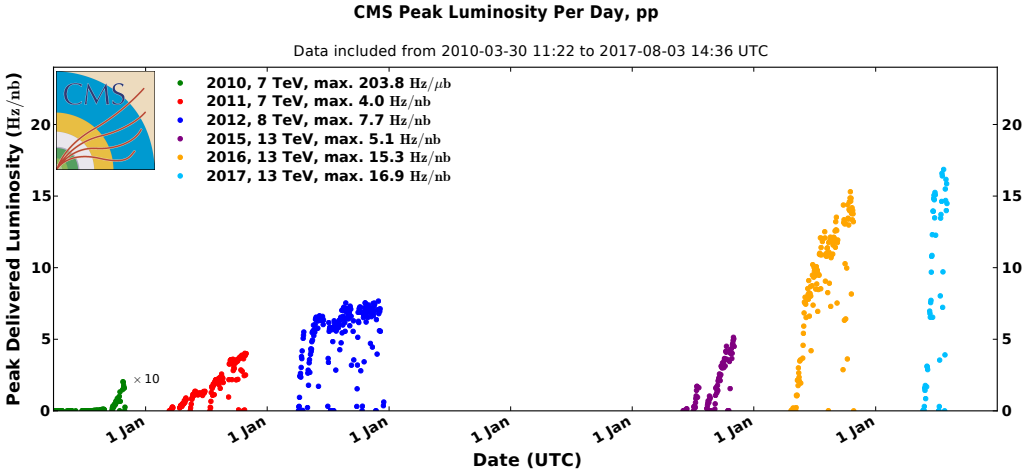
Table 2.2 – Summary of the 2016 LHC operations for physics. For each period, denoted with a letter from B to H, the corresponding time and LHC fill ranges, the peak instantaneous luminosity $\mathcal{L}$, and the total integrated luminosity L delivered to CMS are reported.

LHC operations are continuing at full swing in 2017 as this thesis is being completed, and will proceed until the end of 2018. By that date, that will mark the end of the Run II, the experiments are expected to collect an integrated luminosity ranging between 100 and 150 fb^{-1}. LHC operations will then halt in 2019 for a second long shutdown (LS2) devoted to upgrades of the machine injectors in view of the future high luminosity phase. As in LS1, this will be as well the occasion for an upgrade of the detectors. The LHC is planned to restart in 2021 for its Run III, three years of operations at $\sqrt{s} = 14\,\mathrm{TeV}$ at a peak luminosity twice the original machine design. Within the three years of the Run III, experiments are foreseen to record an integrated luminosity of about 300 fb^{-1}.

The third long shutdown (LS3) starting in 2024 will conclude the Phase I of LHC operations that started back in 2008. During 30 months the LHC will undergo profound changes [83]. New Nb_3Sn superconducting quadrupole magnets, capable of generating a field up to 12 Tesla, will be installed at the ATLAS and CMS interaction points to focus the beams. Compact superconducting cavities ("crab cavities") will be used to precisely rotate the proton bunches before the collision, reduce the crossing angle and enhancing the factor F in Eq. (2.1). The luminosity in the interaction points will be levelled to ensure uniform conditions and long machine fills. These upgrades will increase the instantaneous luminosity of a factor of five with respect to the original design specifications and the integrated luminosity by a factor of ten. The high-luminosity LHC (HL-LHC), or Phase II, will represent an unprecedented way to study very rare phenomena at the LHC. The machine is expected to deliver, during a decade of operations, an integrated luminosity of about 3000 fb^{-1}. The unprecedented conditions of the collisions and, in particular, an



(a) Integrated luminosity



(b) Peak instantaneous luminosity

Figure 2.2 – Total integrated luminosity (a) and peak instantaneous luminosity (b) of the LHC as a function of the year, as measured by the CMS experiment.

average number of simultaneous interactions in one bunch crossing of $\langle\mu\rangle = 140$, have already initiated an important upgrade programme of the experiments.

A summary of past operations and the baseline future schedule for LHC and HL-LHC is shown in Figure 2.3.

2.2 The CMS experiment

The CMS detector has been built to explore the physics at the TeV scale in many different signatures and final states. It has been consequently designed as a multi-purpose detector, that hermetically surrounds the interaction point in the underground caver of Point 5 in Cessy, and is instrumented with several subsystems developed for the identification and measurement of different types of particles. The detector has a cylindrical structure with a diameter of 15 m and a length of 21.5 m, and an overall weight of about 12 500 t.

Collisions take place in the centre of the CMS experiment every 25 ns, implying that new waves of particles leave the interaction point before those produced in the previous

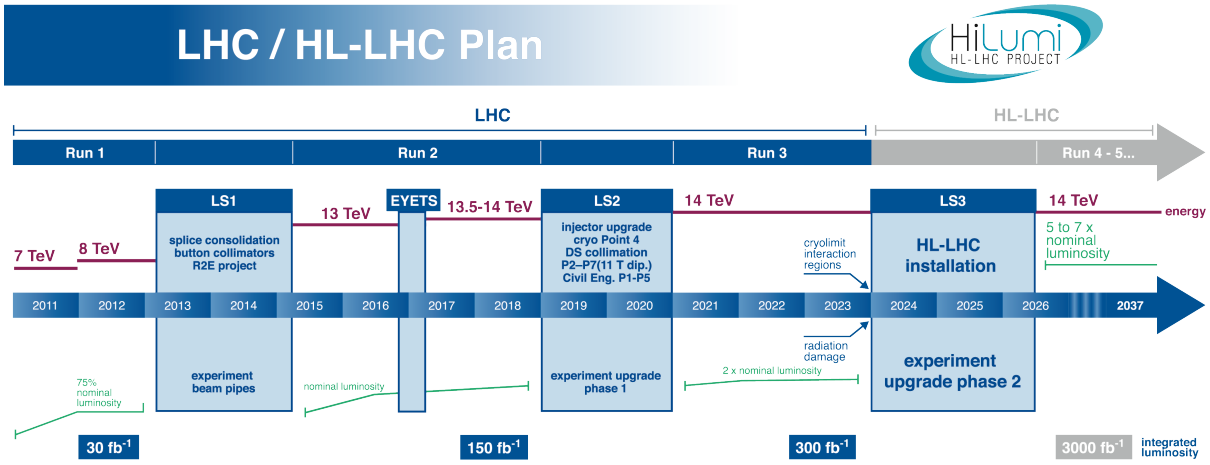


Figure 2.3 – Baseline schedule of LHC and HL-LHC operations. The figure is taken from Ref. [84].

bunch crossing have even escaped the external surface of the detector. In addition, multiple proton interactions can take place within each bunch crossing. These two effects are globally denoted as out-of-time and in-time "pileup" and overlap to the signal of interest represented by the hard-scatter interaction. These challenging conditions call for a detector design that is highly granular, fast in its response, and resistant to the radiation. At the same time, it must be capable of precisely measuring the energy and the momentum of the final state particles and to identify them. The design of the detector, detailed in Section 2.2.2, has been conceived to fulfil these requirements.

2.2.1 Coordinate system

A right-handed Cartesian coordinate system is used to describe the detector and the collision products. It is defined with its centre in the nominal interaction point, the x axis pointing to the centre of the LHC ring, the y axis pointing upwards, and the z axis pointing in the anticlockwise proton beam direction.

Given the cylindrical structure of the detector, a polar system is also used. The azimuthal angle φ is defined in the (x, y) or "transverse" plane as the angle formed with respect to the positive x axis, and the radial coordinate in this plane is denoted as r. The polar angle θ is defined in the (r, z) plane as the angle formed with the z axis and usually converted into the pseudorapidity $\eta = \ln \tan(\theta/2)$. The spatial separation of two particles can be expressed in terms of their angular distance as $(\Delta \mathrm{R})^2 = (\Delta\eta)^2 + (\Delta\varphi)^2$.

The projection of the momentum of a particle onto the transverse plane is referred to as the "transverse momentum" or p_{T}, and has the advantage to be independent on the Lorentz boost resulting from the initial momentum of the interacting partons along the z axis.

2.2.2 Detector structure

The CMS detector [85] is constituted of a central section, or "barrel", and two forward regions, or "endcaps", as it can be observed in the schematic representation of Figure 2.4.

The specific boundaries in η between the two regions depend on the subsystem considered.

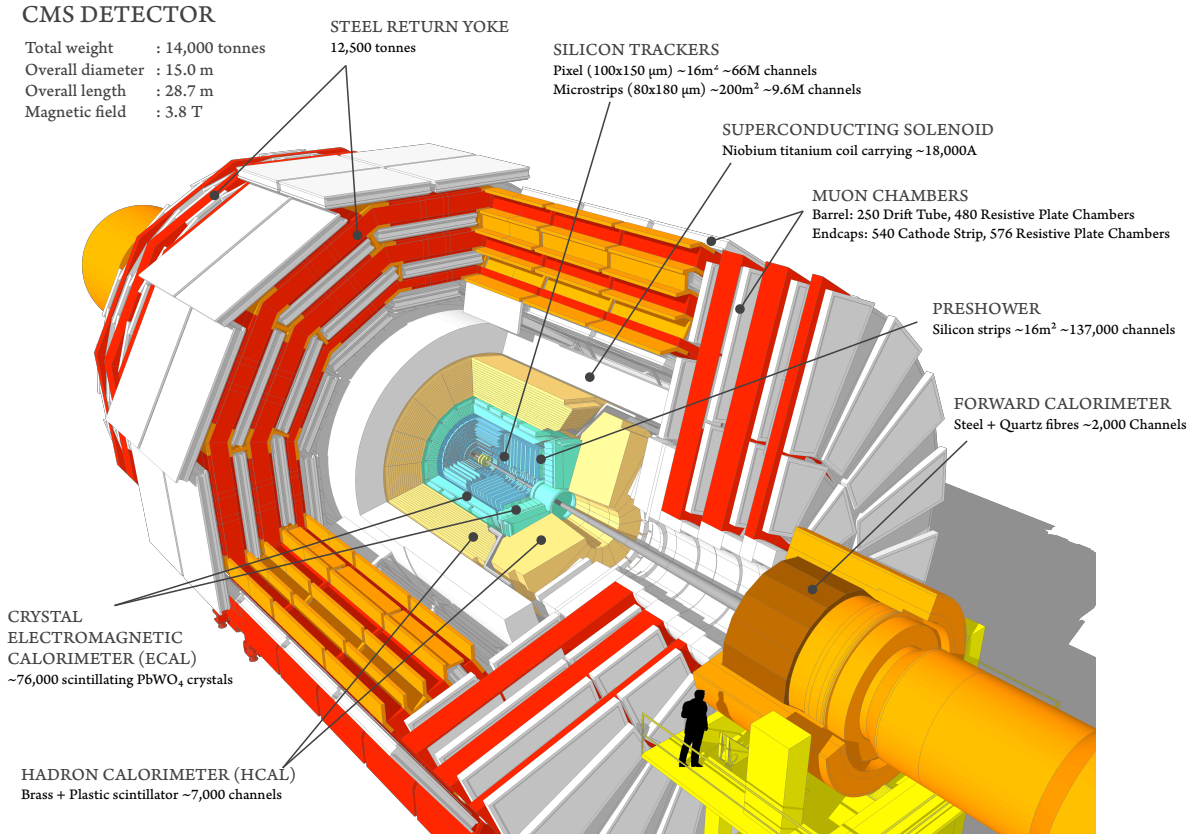


Figure 2.4 – Perspective view of the CMS detector layout.

CMS is instrumented with multiple, concentric layers of detectors to identify and measure the particles produced in the collisions. The interaction point is surrounded by pixel tracker and strip tracker detectors to precisely measure the positions of the interaction points (or "collision vertices") and the trajectory and momentum of the charged particles. The electromagnetic and a hadron calorimeters are located around the tracking systems and designed to absorb electrons, photons, and hadrons within their volume to measure the energy deposited. Muons can traverse the calorimeters and are measured in muon tracking systems located in the outermost part of the detector.

The core of the experiment is a niobium-titanium superconducting solenoid of 6 m of diameter. It is operated at a temperature of 4.5 K and generates a 3.8 T magnetic field along the z axis. This strong magnetic field is used to bend the charged particles and measure their transverse momentum with the tracking subdetectors. The tracker and calorimeters systems are located inside the solenoid, which poses tight constraints of their size and, in the case of the calorimeters, requires high density materials to contain the incoming particles and their secondary interaction products. The return field of the magnet has an intensity of about 2 T and is used to measure the transverse momentum with the muon detectors located inside in the iron structure that surrounds the solenoid. This causes the muons trajectories to be bent in opposite directions in the inner tracker and muon systems, a characteristic feature to which the CMS experiment owes its logo.

The information from the individual subdetectors are often redundant and can be combined to improve the reconstruction of final state objects, as it is discussed in Section 2.3.

Inner tracking systems

The inner tracking detectors are located directly around the interaction point. Their are constituted by a volume of 5.6 m of length and 2.4 m of diameter that is instrumented with silicon sensors that are sensitive to the passage of charged particles.

Thanks to the uniform magnetic field within the tracking detector volume, the informations on the position of charged particles within each silicon detector, or "hits", are combined to measure of the momentum and charge of these particles. The spatial measurement provided by the tracking system also allows for the determination of the hard scatter interaction point ("primary vertex") and its discrimination against additional interactions from pileup in the event. It also allows for the reconstruction of in-flight decays such as those of B hadrons or τ leptons ("secondary vertex").

To fulfil a requirement of precise spatial measurement while being exposed to a large flux of particles, the tracking detector is finely segmented and equipped with fast readout on-board electronics. Moreover, to minimize the impact of the tracking measurement on the passage of charged particles, its design is optimized to use a minimum amount of material. The silicon detector technology deployed in the CMS tracking system addresses these needs by providing a large surface of thin, finely segmented, active detectors.

The detector occupancy rapidly decreases with the radial distance r as the particle flux with a r^{-2} dependence. Higher spatial precision is also required close to the beam pipe for the identification of the interaction vertices. As a result, two silicon detector technologies are used in the CMS tracking system, as illustrated in Figure 2.5.

The innermost region is constituted by the pixel tracking detector, disposed in three cylindrical layers in the barrel and two disks in the endcap, at a distance from the interaction point that varies between 4.4 and 10.2 cm. Each pixel measures about $100 \times 150\,\mu\text{m}$,

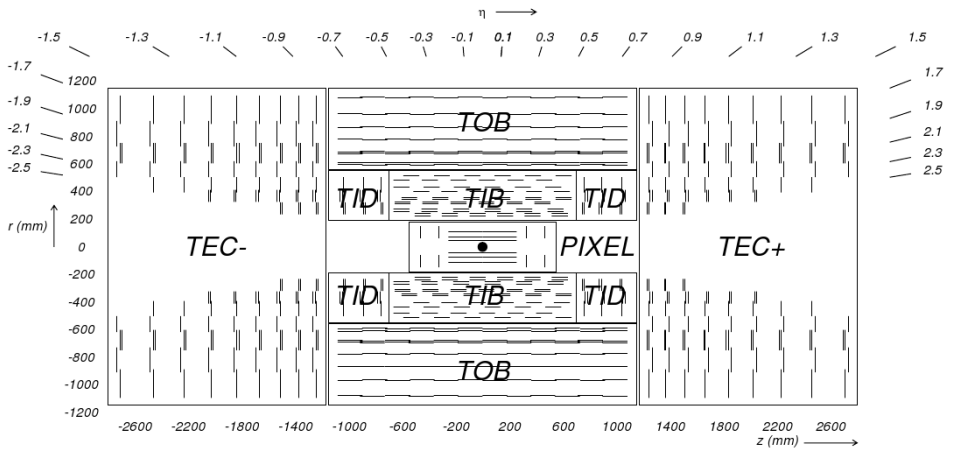


Figure 2.5 – Longitudinal view of the CMS inner tracking system layout. The pixel detector is located directly in front of the interaction point. The strip tracker detector is composed of the tracker innner barrel (TIB) and tracker inner disks (TID), surrounded by the tracker outer barrel (TOB) and the tracker endcaps (TEC).

resulting in about 66 million of pixels covering an area of $1\,\mathrm{m}^2$. The spatial resolution of each pixel is of $10\,\mu\mathrm{m}$ in the (r, φ) plane and of $20\,\mu\mathrm{m}$ along the z direction.

The outermost region uses strip silicon sensors with two different strip pitches depending on the distance from the interaction point. The barrel tracker region is constituted by 4+6 active layers and the endcap region by 3+9 endcap disks, that extend the tracking measurement up to a pseudorapidity $|\eta| = 2.5$. The resolution on the single point ranges from 20 to $50\,\mu\mathrm{m}$ in the radial direction and from 200 to $500\,\mu\mathrm{m}$ in the longitudinal one, depending on the value of r. The entire strip tracker detector is composed of about 9.6 millions of silicon strips.

The strip tracker and the pixel detectors are operated at a temperature of about $-15\,^{\circ}\mathrm{C}$ and $-20\,^{\circ}\mathrm{C}$, respectively. This is necessary to minimize the damage caused by ionizing radiation to silicon detectors, and requires the presence of an efficient cooling system to absorb the heath produced by the on-board electronics.

The pixel detector has been recently upgraded in view of the 2017 data taking [86]. The upgraded pixel detector features four layers in the barrel and three disks in the endcaps, providing an additional measurement point in both regions. The innermost disk has been installed closer to the nominal interaction point, at a distance of 3 cm for the barrel detector. The material budget has also been largely reduced, with the barrel and endcap detectors weighing 40% and 80% of the current detectors, respectively. The new layout of the upgraded pixel detector is illustrated in Figure 2.6.

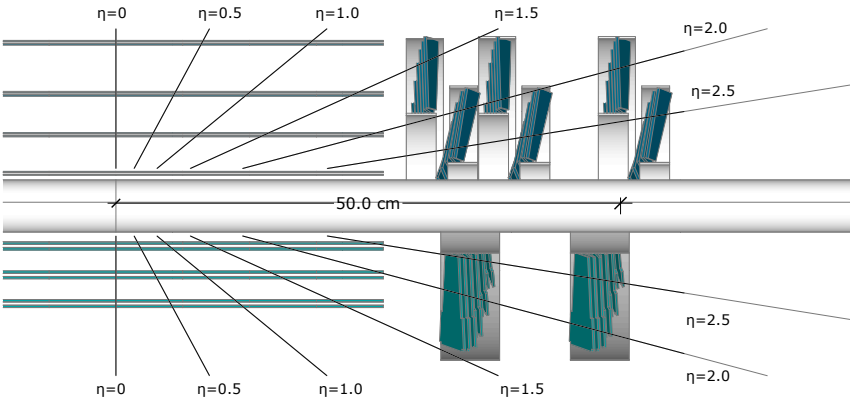


Figure 2.6 – Longitudinal view of the current pixel detector (top, blue) and of the upgraded detector (bottom, green) installed for the 2017 data taking.

Electromagnetic calorimeter

The electromagnetic calorimeter, or ECAL, is designed to measure the energy of incident electrons and photons. The CMS ECAL is a homogeneous and highly granular calorimeter constituted of lead tungstate crystals ($PbWO_4$). The energy measurement is based on the conversion of the incident electron or photon to an electromagnetic shower, that interacts with the crystal material producing scintillation light. The crystals are at the same time the dense interacting material and the active scintillating medium, resulting in an excellent energy resolution.

The choice of $PbWO_4$ is motivated by its high density ($8.28\,g/cm^3$), small radiation length ($X_0 = 0.89\,cm$) and short Molière radius ($R = 2.2\,cm$). These parameters ensure an excellent containment of the electromagnetic shower within the crystals, which have a length of approximately $25\,X_0$. The lead tungstate is radiation hard and about 80% of its scintillation light is produced within 25 ns, making it ideal for the high instantaneous luminosity collisions of the LHC and the proton bunch spacing. The disadvantage of this material is the relatively low light yield, corresponding to about 30 photons per MeV of deposited energy, which calls for the usage of photodetectors with internal amplification, as detailed below.

The barrel part of the ECAL is constituted by 61 200 crystals with a frontal transverse section of $22 \times 22\,mm^2$ and a length of 23 cm, and ensures the coverage of the region $|\eta| < 1.479$. The two endcaps are each made of 7324 crystals of a frontal transverse section of $28.62 \times 28.62\,mm^2$ and a length of 22 cm, and extend the coverage up to $|\eta| < 3.0$. The layout of the crystals in the ECAL is illustrated in Figure 2.7. Crystals in the barrel are organized in 36 "supermodules", each covering half a barrel region and an angle of 20° in φ, and made of four "modules" where single crystals are mounted in a mechanic support. Crystals in each endcap are disposed in two semi-circular "dees". In both the barrel and the endcaps, crystals are mounted in a quasi-projective geometry, with their axes being tilted of 3° with respect to the direction that points to the nominal interaction region. This ensures that no particle escapes the ECAL active volumes from the interstices between the crystals.

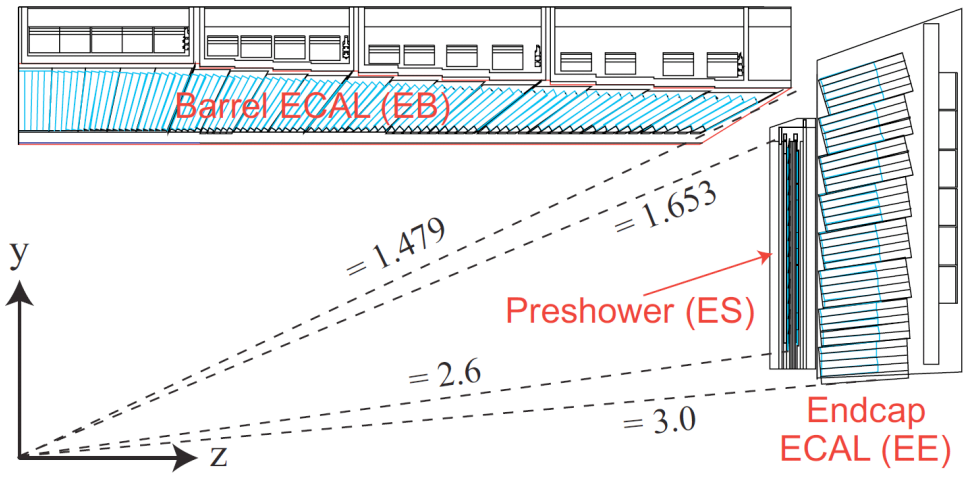


Figure 2.7 – Longitudinal view of the ECAL layout, representing one quarter of the detector. The barrel and endcaps sections and the preshower detector are shown.

The crystal scintillation light is read out by detectors that are designed to work in the high magnetic field to which they are exposed and to be resistant to the radiation. The barrel part of the ECAL is instrumented with silicon avalanche photodiodes (APDs) while vacuum phototriodes are deployed in the endcaps. The signals from these photodetectors are amplified and shaped by the front-end electronics and sampled at a frequency of 40 MHz with a 12-bit analog-to-digital converter. The ECAL is operated at a temperature of 18 °C. To control the changes in the light yield of the crystals (−2%/°C) and of the photodetectors (−2.3%/°C) on the temperature, the latter is stabilized within 0.05 °C in EB and 0.1 °C in EE by a cooling system.

A sampling preshower (ES) is installed in front of the two endcaps to improve the discrimination of single photons from $\pi^0 \to \gamma\gamma$ decays. It is constituted by two layers of lead absorber in which the electromagnetic shower is initiated, followed by a layer of 2 mm-wide silicon strips to measure the deposited energy and the transverse profile of the shower shape.

The large doses of radiation to which the crystals are exposed causes a change in their transparency, that is naturally recovered at the operating ECAL temperature. As a consequence, the ECAL undergoes cycles of transparency reduction and recovery that correspond to the collisions and refill operations of the LHC. This effect is monitored and corrected with the injection of a 440 nm laser light in each crystal to derive time-dependent correction factors that are applied to the response.

Tracking and calorimeter detectors in CMS provide complementary measurements. The former can identify only charged particles, and the precision of its momentum measurement increases as p_{T} decreases, because of the larger curvature of the trajectory. Inversely, the latter can measure both charged and neutral particles with a resolution that increases with the particle energy itself, because of the reduced impact of two of the three main effects that determine resolution of a generic calorimeter. The first one is a stochastic term that depends on the number n of scintillation photons (or elementary information carriers in general) produced in the interaction as $\sqrt{n}$, where n is in turn proportional to the incident particle energy E. A second term accounts for the noise in the detector and does not depend on E. Finally, a third term is related to detector inhomogeneities, resulting in an error that amounts to a constant fraction of E. The combined effect of these three factors results in a dependence of the calorimetric energy resolution σ on the particle of energy E as:

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

where S, N and C denote the stochastic, noise, and constant terms, respectively. In ECAL test beam studies performed with incident electrons [87], the values $S = 2.8\%$, $N = 12\%$, and $C = 0.3\%$ were obtained. The ECAL response is calibrated to determine both the absolute energy scale and the channel-to-channel intercalibration, to compensate in particular for the intrinsic crystal light yield variations ($\approx 15\%$) and the spread in the EE phototriodes ($\approx 25\%$). The initial calibration derived from laboratory studies and cosmic rays exposures of crystals is now complemented with in-situ measurements based on collision events. As a result, the energy resolution for 45 GeV electrons is of about 2% in the barrel and 2–5% elsewhere, and increases to about 1.5% for electrons in the central part of the detector with little energy radiated by bremsstrahlung. An extensive description of the ECAL performance can be found in [88].

Hadronic calorimeter

Hadrons typically traverse the ECAL volume without being stopped. The hadronic calorimeter, or HCAL, is designed to absorb them within its volume and measure their energy. Compared to electron and photon interactions in the ECAL, hadron energies are intrinsically more difficult to measure from hadron showers induced in the HCAL. Nuclear and hadronic interaction result in non-Poissonian effects in the shower development, where many undetectable particles can also be produced. The presence of π^0 decaying to

photon pairs also results in an electromagnetic component of the shower with a different response from the HCAL itself. All these effects limit the intrinsic resolution on hadron energies, which can be improved offline with the usage of the particle flow reconstruction techniques detailed in Section 2.3. Despite these limitations, the HCAL is an essential element in the reconstruction of final states containing hadron jets or non-interacting particles such as neutrinos, where the calorimeter hermeticity and geometrical coverage allows for the computation of the imbalance in the transverse momentum sum of the event.

The barrel (HB) and endcap (HE) sections of the HCAL instrument respectively the regions $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$. Both the HB and HE are sampling calorimeters composed of a brass absorber and of active plastic scintillating tiles. The scintillation light is collected by wavelength shifter fibres embedded in the tiles and read out by hybrid photodiodes (HPDs). Each readout cell is formed by the addition of a "tower" of scintillating tiles in a spatially localized region, and has a transverse $\Delta\eta \times \Delta\varphi$ dimension of about 0.087×0.087 in the HB and of about 0.17×0.17 in the HE. As the HCAL is located between the ECAL and the internal surface of the solenoid, the limited space does not allow for a full containment of the secondary interaction shower. The detector is complemented by a outer hadron calorimeter (HO) located outside the solenoid, which extends the total interaction depth to about $11\,\lambda_0$, where the constant indicates the average interaction length of hadrons in the calorimeter. The energy measurement in the forward region is complemented by the forward hadronic calorimeter (HF), that is located 11.2 m away from the interaction point and measures hadron interactions up to $|\eta| = 5.2$. Because of the higher radiation levels in the forward region, the HF is composed of steel absorbers and quartz fibres that produce light by Cherenkov effect, which is measured by photomultiplier tubes (PMTs). Fibres of two different lengths are installed to estimate the electromagnetic and hadronic components of the shower. The global layout of the HCAL is illustrated in Figure 2.8.

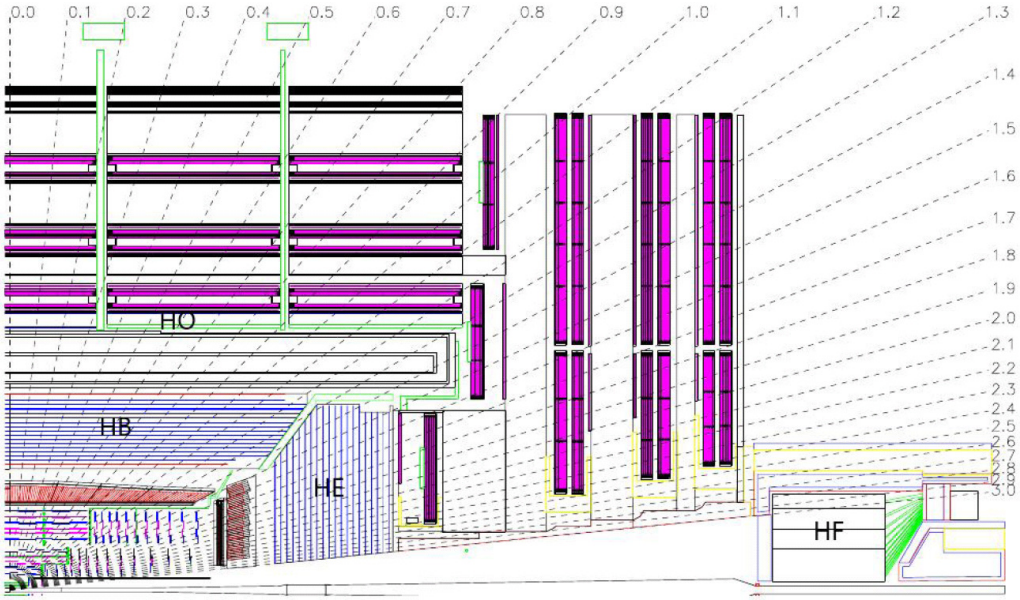


Figure 2.8 – Longitudinal view of the HCAL layout, where are visible the barrel (HB) and endcaps (HE) detectors located inside the solenoid, the outer detector (HO) outside the solenoid, and the forward calorimeter (HF).

The overall HCAL performance is dominated by the imperfect containment of the hadronic shower, which results in a resolution sampling term of about 110% and a constant term of 9%, as measured in pion test beams [89].

Muon detectors

Because of their typical energy, muons produced in collisions at the LHC have minimal energy loss rates. As a consequence, they traverse the ECAL, the HCAL, and the solenoid volumes without being stopped and are identified and measured in the muon detectors located in the outermost part of CMS. The muon momentum is measured using the return field of the solenoid inside the iron structure in which the muon detectors are embedded, and complements the measurement from the inner tracker previously discussed. CMS is instrumented with three types of muon detectors, chosen accordingly to the expected background rates and uniformity of the magnetic field, as it is illustrated in Figure 2.9.

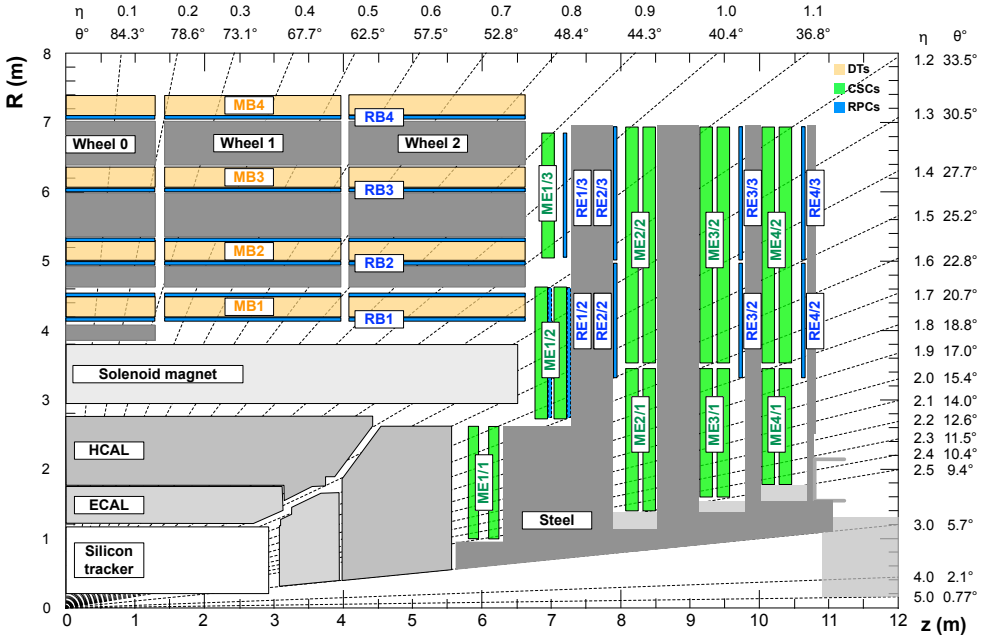


Figure 2.9 – Layout of the muon systems of the CMS experiment. A quarter of the detector in its longitudinal view is shown. Orange, green, and blue regions denote respectively the DTs, CSCs, and RPCs subsystems.

In the barrel region, CMS deploys 250 drift tubes (DTs) uniformly distributed in the five barrel sections, or "wheels". Each wheel hosts four concentric rings of DT stations, organized in twelve contiguous sectors. The basic element of the DT detector is a rectangular cell of transversal size $4.2 \times 1.3\,\mathrm{cm}^2$, containing an anode wire and filled with a Ar/CO_2 gas mixture. Electrodes on the top and bottom of the cell ensure a constant field and a uniform drift velocity of about $55\,\mu\mathrm{m/s}$, while cathodes are placed on the sides of the cell. DT cells are organized in three groups of four elements (three "super-layers", SLs) that together compose a DT chamber. Muon traversing each group ionize the gas, and their position and angle are measured from the time needed by the electrons to drift toward the anode wires. The middle SL is oriented to measure the coordinate position in the z direction while the other two SLs provide a measurement in the (r, φ) plane. Each

DT cell has a spatial resolution of about 200 µm, resulting in a resolution of $80 - 120$ µm for the global chamber measurement [90].

Cathode strip chambers detectors (CSCs) are used to instrument the endcap regions of CMS ($0.9 < |\eta| < 2.1$). The different choice of detector technology is imposed by the higher background rates and the stronger magnetic field. CSCs are detectors designed in a trapezoidal shape and composed of six layers of anode wires interposed between seven segmented cathode plates disposed in the perpendicular direction. CSCs contain a $Ar/CO_2/CF_4$ gas mixture, which is ionized upon the passage of a muon. The resulting signals induced on the wires and on the strips are interpolated and provide a position measurement in the (r, φ) plane (anode wires) and along the z direction (strips). The CSC is a fast detector, capable of identifying the bunch crossing of a pp collision, and achieves a spatial resolution of $40 - 150$ µm [90].

Resistive plate chambers (RPCs) are installed in both the barrel and endcaps and cover the region $|\eta| < 1.6$. RPCs are formed by two gaps consisting of two resistive Bakelite layers of 2 mm thickness separated by a 2 mm volume filled with a $C_2H_2F_4/i - C_4H_{10}/SF_6$ gas mixture. The detector is operated in avalanche mode and, when traversed by a muon, an avalanche is generated by the high electric field inside the gas volume and is read out by strips located on the outer surface of the gap. Although RPCs have a modest spatial resolution of $0.8 - 1.2$ cm [90], they have excellent timing properties with a resolution of the order of the ns, allowing for the determination of the pp bunch crossing.

2.3 Physics object identification and reconstruction

The raw detector information is combined and used to reconstruct "physics objects", that constitute the input of all the data analyses. A global event reconstruction is performed to identify a few elementary objects: charged and neutral hadrons, electrons, photons, and muons. These are subsequently combined to reconstruct more complex objects such as semileptonic τ decays, jets, and missing transverse momentum.

2.3.1 Global event reconstruction

The high granularity of the tracker and ECAL detectors, the strong magnetic field to separate neutral and charged particles, the hermeticity of the HCAL and HF, and the excellent muon system resolution make the CMS detector ideally suited to identify and measure the individual final state particles. The particle flow (PF) algorithm [91] is designed to exploit the redundant measurements from the CMS subsystems and reconstruct physics objects from raw detector data, performing a global event reconstruction. The philosophy of this approach is illustrated in Figure 2.10, where the typical signatures of different particles in the CMS detector are compared. The trajectory of charged particles, or tracks, are reconstructed from the hits in the tracker systems, and matched to deposits in the ECAL only (electrons) or in the HCAL as well (charged hadrons). The absence of a track in front of a calorimetric deposit reveals the passage of a photon or a neutral hadrons. Finally, the presence of a track in the muon systems identifies the interaction of a muon. The combination of these redundant measurements has a positive impact on all physics objects, and the largest effects are observed in jet, tau leptons, and missing transverse momentum, three key elements of the $HH \to b\bar{b}\tau^+\tau^-$ search.

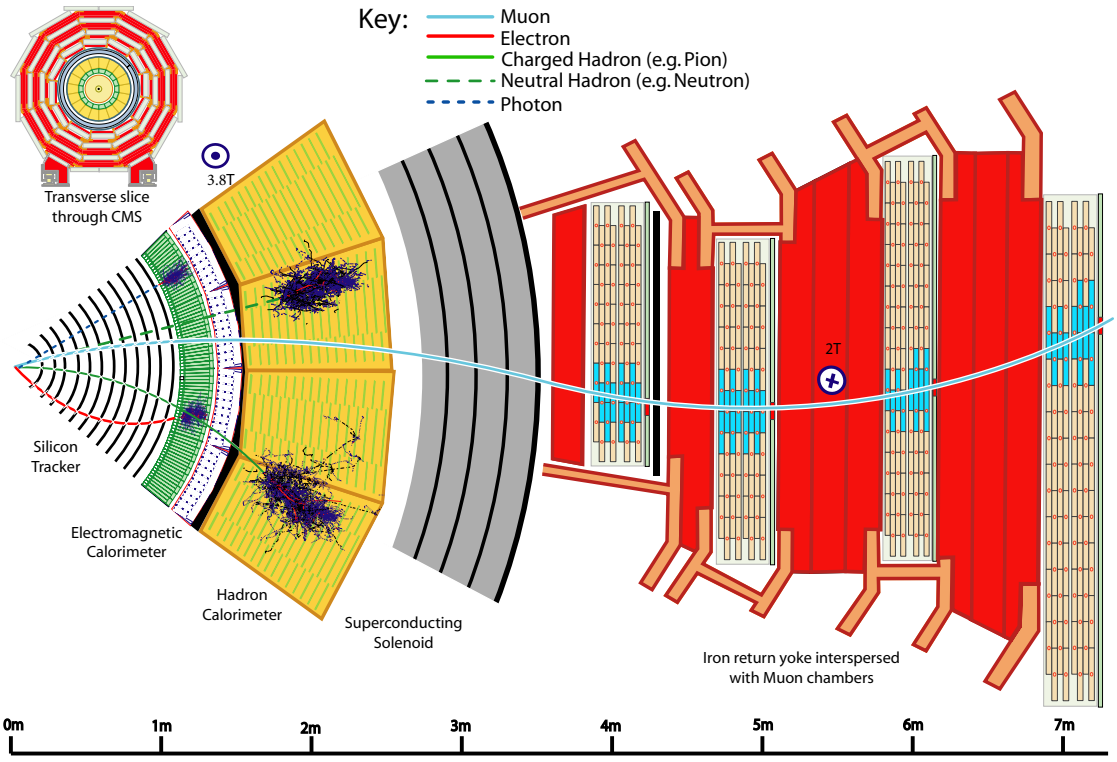


Figure 2.10 – Illustration of the experimental signature of final state particles in the CMS detector.

This simplified picture is complicated by earlier interactions of particles within the tracker, representing up to 2 radiation lengths of material (at $|\eta| \approx 1.5$) in front of the calorimeters. This result in a probability of photon conversion to e^-e^+ or of bremsstrahlung emission from electron of about 85%, and a probability of hadron nuclear interaction of about 20%. Similarly, muon can undergo multiple scattering before reaching the muon detectors, with a subsequent degradation of the momentum resolution. To overcome these problems, advanced specific algorithms have been developed to reconstruct the key elements of the PF algorithm, namely tracks and energy clusters.

Tracks must be reconstructed with an efficiency as close to 100% as possible. This is especially important when they originate from charged hadrons in a jet, as the complementary measurement based solely on the calorimeters is not fully efficient and suffers from a direction bias and energy degradation. At the same time, a low rate of erroneously reconstructed tracks from random hit association must be achieved to ensure a good description of the event. An iterative tracking procedure [92] fulfils these requirements by initially applying strict quality criteria on track seeding and reconstruction, removing hits unambiguously assigned, and progressively loosening the quality criteria to increase the efficiency.

The energy deposits in the CMS calorimeters are grouped together with the PF clustering algorithm. The clustering algorithm is operated separately in the preshower, ECAL, and HCAL subdetectors. It identifies local maxima of energy, or "seeds", regroups the neighbouring energy deposits that satisfy topological and energy criteria, and individuates in the deposits thus selected the PF cluster itself. This procedure has a close analogy with the dynamic clustering developed for the L1 calorimeter trigger upgrade discussed

in Section 3.4.1 of Chapter 3, which was inspired from the PF approach and adapted to the hardware trigger constraints.

The individual PF elements can be associated, or "linked", to create "PF blocks". The PF identification proceeds by analysing the structure and properties of these blocks to identify the final physics objects. In a simplified description, charged hadrons are identified from the combination of a track and of a cluster, while muon tracks are reconstructed independently as detailed in Section 2.3.2 and are not considered as charged hadrons. Electron reconstruction relies on a dedicated tracking procedure to account for bremsstrahlung radiation and combines several PF clusters to recover the energy of radiated photons, as summarized in Section 2.3.3. Once all tracks have been assigned to a candidate, the remaining clusters are considered as photons in case of ECAL deposits and as neutral hadrons in case of HCAL deposits. After the association of all the PF elements, the redundant information from the CMS subsystems is combined to estimate their momentum. The total calorimetric energy is the linear sum of the calibrated ECAL and HCAL energies. In case this is not found in agreement, within the expected uncertainties, with the associated track momentum, an overlap between a charged and neutral candidate is assumed and the energy of the latter is estimated as the difference of the two measurements. A more detailed description of the linking and reconstruction of PF candidates can be found in Ref. [91].

The output of the PF algorithms is a list of PF candidates: charged and neutral hadrons, electrons, photons, and muons. These are combined to reconstruct other objects such as jets, tau leptons decaying semileptonically (τ_{h}), and missing transverse momentum ($\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$).

2.3.2 Muon reconstruction

Muons leave a very clean signature in the CMS detector thanks to their interactions in the muon spectrometers. As a consequence, muon tracks are reconstructed with dedicated algorithms that are independent from the iterative PF tracking discussed above, and are based on a Kalman filter method that accounts for the muon energy loss in the detector materials. Three muon reconstruction algorithms are defined and exploit differently the subdetectors information [93]:

- **Standalone muons** rely solely on the information from the muon subdetectors. Patterns in the DT, CSC, and RPC detectors are combined and fitted into a standalone muon track.

- **Tracker muons** are reconstructed from hits in the inner tracking detectors. Each track with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and a total momentum larger that 2.5 GeV is extrapolated to the muon subsystems, where the presence of at least one muon segment at a compatible position is required.

- **Global muons** combine the information from both detectors, by propagating inner tracks and standalone muon tracks to a common surface and verifying their compatibility. The combined collection of hits is fitted together to a global muon track. Because of the size of muon spectrometers, their inclusion in a global fit improves the p_{T} resolution for muons of $p_{\mathrm{T}} \gtrsim 200\,\mathrm{GeV}$.

Owing to the high reconstruction efficiency in both tracker and muon systems, about 99% of muons are reconstructed either as tracker and global muons, and those candidates that share the same inner tracks are merged into a single object. As standalone-only muons have a worse momentum resolution and show a high contamination from cosmic rays background, they are seldom used for physics analyses.

Muon charge and momentum assignments are computed solely from the tracker measurement for muons of $p_{\mathrm{T}} < 200\,\mathrm{GeV}$, as multiple scattering effects degrade the measurement of the muon detectors. The global track curvature is instead used for muon with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, if the charge-to-momentum ratio agrees within two standard deviations from the tracker only measurement. The muon transverse momentum resolution thus achieved ranges between 1 to 6%, depending on the η coordinate, for muons with $p_{\mathrm{T}} < 100\,\mathrm{GeV}$, and is better than 10% for central muons of $p_{\mathrm{T}} = 1\,\mathrm{TeV}$.

2.3.3 Electron reconstruction

Electron reconstruction is complicated by their interaction in the inner tracking material before they reach the ECAL. Tracker algorithms must take into account the non-Gaussian energy loss, and clustering algorithms must collect the bremsstrahlung photon energy deposits that can be located away from the electron interaction point in the ECAL. The electron reconstruction algorithm [94] addresses these two effects with a dedicated tracking and an advanced energy clustering.

The latter are built by regrouping PF ECAL clusters in "superclusters". This procedure identifies a seed cluster and gathers together the energy deposits associated to bremsstrahlung photons. Preshower energy clusters in the endcaps are also taken into account in the procedure. The supercluster aggregation depends on the cluster E_{T} and exploits the correlation between their η and φ positions, preferring clusters that are spread along the φ direction because of the magnetic field.

Tracks are reconstructed with a Gaussian sum filter (GSF) method. In contrast to the Kalman filter, the GSF method accounts for the large bremsstrahlung energy emission by approximating the radiated energy loss probability with a sum of Gaussian distributions. The GSF tracking is initiated, or "seeded", by to complementary algorithms. An ECAL-seeding procedure makes an estimate for the track position starting from the PF superclusters, while a tracker-seeding relies on the general charged particle iterative tracks and looks for a correspondence with a PF supercluster.

GSF tracks and PF superclusters are associated into an electron candidate if they satisfy some loose requirements on their qualities and matching. They are subsequently used to estimate the electron charge and its momentum, the latter being computed from a combination of GSF track curvature and supercluster total energy. To improve the resolution, electrons are classified in five categories depending on their quality and bremsstrahlung properties. The momentum resolution for electrons produced in Z boson decays ranges between 1.7 and 4.5% depending on the electron category and position in the detector.

2.3.4 Tau reconstruction

With a mean lifetime of about $2.9 \times 10^{-13}\,\mathrm{s}$, tau leptons decay within a few millimetres from their production point for the typical Lorentz boosts at the LHC. While fully leptonic decays to a electron or a muon in association with neutrinos are reconstructed from the

respective object algorithms detailed above, semileptonic decays to hadrons (τ_{h}) and a neutrino result in small and collimated hadron jet that requires a specific reconstruction algorithm. The decay can occur through an intermediate $\rho(770)$ or $\mathrm{a}_1(1260)$ resonance, and result in different multiplicities of charged and neutral hadrons, usually pions, as summarized in Table 2.3. Decays containing one and three charged hadrons are generally referred to as one- and three-prong decays, respectively.

Decay mode	Meson resonance	$\mathcal{B}$ (%)
$\tau^- \to \mathrm{e}^-\overline{\nu}_{\mathrm{e}}\nu_\tau$		17.8
$\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$		17.4
$\tau^- \to \mathrm{h}^-\nu_\tau$		11.5
$\tau^- \to \mathrm{h}^-\pi^0\nu_\tau$	$\rho(770)$	25.9
$\tau^- \to \mathrm{h}^-\pi^0\pi^0\nu_\tau$	$\mathrm{a}_1(1260)$	9.5
$\tau^- \to \mathrm{h}^-\mathrm{h}^+\mathrm{h}^-$	$\mathrm{a}_1(1260)$	9.8
$\tau^- \to \mathrm{h}^-\mathrm{h}^+\mathrm{h}^-\pi^0$		4.8
Other modes with hadrons		3.3
Total with hadrons		64.8

Table 2.3 – Branching fraction of the decays of a τ lepton. The symbol h generically refers to a charged pion or kaon. For the decay occurring through an intermediate meson resonance, the corresponding name is indicated.

The τ_{h} reconstruction algorithm should be able to determine the τ decay mode, identify PF candidates associated to both charged hadrons and photons from $\pi^0 \to \gamma\gamma$ decays, and regroup them together to estimate the τ_{h} kinematic properties. The hadrons plus strips (HPS) algorithm [95, 96, 97] is designed to perform this tasks using the PF candidates previously reconstructed. The HPS reconstruction is initiated by PF jets that are formed as detailed in Section 2.3.5. The algorithm analyses the PF candidates composing each jet to verify their compatibility with a τ_{h} object. The contribution from neutral pions in $\pi^0 \to \gamma\gamma$ can appear either directly as photon PF candidates, or as electron candidates clustered inside the jet, because of the large $\gamma \to \mathrm{e}^-\mathrm{e}^+$ conversion probability. Photon and electron PF candidates of $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ are thus clustered into "strips" with and iterative procedure. Electron and photon candidates within a clustering region around the strip are added to the strip itself, which position is recomputed as a p_{T}-weighted average. The strip creation ends when no candidates are found within the clustering region. A dynamic strip reconstruction, introduced in the Run II, defines the $\Delta\eta$ and $\Delta\varphi$ clustering window sizes as functions of the strip p_{T} itself, to ensure an optimal collection of the energy and minimize the impact of background.

The strips and the charged hadrons in the jet are combined together to reconstruct any of these decay modes:

- $\mathrm{h}^\pm$, single charged hadron with no strips
- $\mathrm{h}^\pm\pi^0$, single charged hadron with one strip
- $\mathrm{h}^\pm\pi^0\pi^0$, single charged hadron with two strips
- $\mathrm{h}^\pm\mathrm{h}^\mp\mathrm{h}^\pm$, three charged hadrons

The three-prong decay in association to a π^0 is not considered as a valid topology because of its small branching fraction and large contamination from quark or gluon jets. Quality and invariant mass criteria are applied onto all the valid decay mode hypotheses to verify their compatibility with a τ decay. The selections applied depend on the decay mode considered and on the e/γ candidates clustered into the strip. The specific selections used in Run II are detailed in Ref. [97]. In general, it is verified that tracks originate from the same vertex for multi-prong decays, that the combined invariant mass is compatible with the one of the intermediate meson resonance, and that the total electric charge is ± 1. Decay mode hypotheses are also discarded if they include additional charged hadrons or strips outside of a signal cone centred on the τ_{h} momentum axis and with and aperture of $\Delta R = 3\,\mathrm{GeV}/p_{\mathrm{T}}$ (bounded to $0.05 < \Delta R < 0.1$). The $\mathrm{h}^{\pm}\pi^0$ and $\mathrm{h}^{\pm}\pi^0\pi^0$ decays are analysed together and commonly referred to as $\mathrm{h}^{\pm}\pi^0$s. In case multiple decay mode hypotheses are satisfied, the one with the largest p_{T} is retained, resulting in an unique association of a τ_{h} candidate to a jet.

The τ_{h} reconstruction discussed here is complemented by isolation and identification methods to reject the large quark and gluon jet background and erroneous reconstruction of muons and electrons. These methods, together with the optimization of their selection in the context of the $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^+\tau^-$ search, are detailed in Section 4.3.3 of Chapter 4.

2.3.5 Jet reconstruction

As quarks and gluons undergo an hadronization process, the estimation of their initial momentum requires the recollection and measurement of the hadronization products. Jets are thus reconstructed by clustering the PF candidates with the anti-k_{T} algorithm [98, 99]. The algorithm iteratively combines PF candidates that are close to each other according to a metric, that is defined to produce jets of an approximate conic shape clustered around the hardest particles in the event. The size of the jet cone is determined by the distance parameter R at which the algorithm is operated. Both the values $\mathrm{R} = 0.4$ and $\mathrm{R} = 0.8$ are used for the search presented in this thesis. The anti-k_{T} algorithm is resilient against infrared and collinear effects, i.e. it is not affected by soft radiation or collinear parton splitting.

The jet four momentum is computed as the vector sum of the clustered PF candidates four momenta, and a set of corrections are applied to calibrate the jet response using the information of generated particles in a simulation. These corrections of the jet energy scale take into account the contribution from pileup in the event, nonlinearities in the detector response to hadrons, and residual differences between the data and the simulation used for the method. They are validated using dijet, multijet, γ+jets and leptonic Z+jets events [100, 101, 102]. Typical jets resolutions achieved are of about 15–20% for at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV.

2.3.6 Missing transverse momentum reconstruction

The presence of undetected final state particles such as neutrinos can be indirectly inferred from the imbalance of the total transverse momentum vector sum. The negative projection of this vector onto the transverse plane is denoted as missing transverse momentum ($\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$), and is an important quantity in the selection and reconstruction of $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^+\tau^-$ events.

The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector is reconstructed with the PF algorithm [103] as the negative of the vectorial sum of the transverse momenta of the PF candidates reconstructed in the event. As inefficiencies of the tracking algorithm, minimal thresholds in the calorimeter energy estimation, and nonlinearities of the energy response of the calorimeters for hadronic particles can introduce a bias in the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ determination, a correction is applied by propagating to the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ sum the jet energy corrections introduced in Section 2.3.5. In particular, the corrected $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector is estimated as:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,corr}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T}}) \tag{2.5}$$

i.e. taking into account the difference between the initial jet $\vec{p}_{\mathrm{T}}$ and its corrected value $\vec{p}_{\mathrm{T}}^{\,\mathrm{corr}}$. If a muon is found within the jet cone, its four-momentum is subtracted from the jet momentum when computing the correction and then added back into the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ sum.

2.4 Trigger system

Proton-proton collisions occur in the centre of the CMS detector every 25 ns, and generate a huge amount of information in the detector, corresponding to about 70 terabytes of data every second. No technology exists nowadays to read out, store and analyse such volumes of data. However, the large majority of the collisions result in low-energy proton-proton interactions that are not interesting to pursue the physics programme of CMS. As it can be observed in Figure 2.11, even the most frequent processes studied at the LHC have a cross section of the order of 10^5 pb, that is 5 orders of magnitude smaller than the total proton-proton interaction cross section $\sigma_{\mathrm{pp}} \approx 10^{11}$ pb [104].

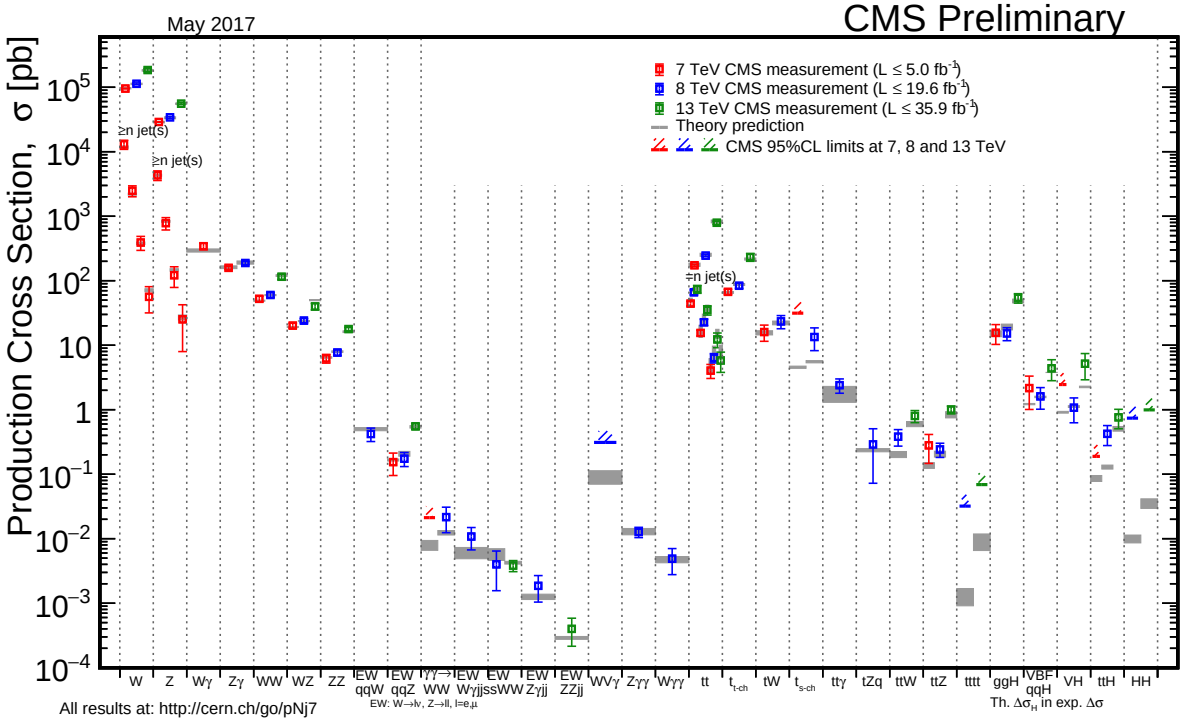


Figure 2.11 – Summary of the cross section measurements of SM processes at CMS. Values are to be compared to the total pp interaction cross section of about 10^{11} pb.

The role of the trigger system of the CMS experiment is to identify and select the interesting collision events, thus reducing the acquisition rate by a factor of about 10^5. The trigger is at the interface between the "online" data taking and the "offline" data analysis, and must at the same satisfy the technical constraints of the former and ensure a high efficiency for the latter.

2.4.1 Structure of the trigger system

The CMS experiment adopted a two-tiered approach in the development of its trigger system. The Level-1 (L1) trigger is composed of custom hardware that processes the information from calorimeters and muon systems only with reduced granularity. It has a fixed latency (i.e. the time available for data processing) of 3.8 µs, in which the event accept decision is made and the event rate is reduced down to about 100 kHz. Following this first selection, the high-level trigger (HLT) can access the complete detector information at the full granularity to perform an event reconstruction that is similar to the one performed offline. The HLT is implemented in a farm of commercial processors, where sophisticated algorithms running on its 22 000 CPU cores produce a decision in an average time of about 220 µs and further reduce the trigger rate below 1 kHz. The events thus selected are recorded on the tapes of the CERN Tier-0 and become available for subsequent offline analysis.

The very different technologies used in the L1 and HLT systems are due to the distinct constraints on the processing time and on the volume of data to be analysed. Despite these differences, both systems implement flexible and configurable algorithms that identify and reconstruct physics objects and combine their properties to perform the event accept decision. In 2016, the L1 system implemented a "menu" of about 300 algorithms or "seeds", that has been extended to about 500 in view of 2017 collisions. Any seed is assigned an adjustable factor f, or "prescale", that reduces the trigger rate of $1/f$ by retaining only one accept decision every f occurrences. The set of prescale values is changed during the data taking runs as a function of the instantaneous luminosity $\mathcal{L}$ to maintain a constant trigger rate when $\mathcal{L}$ is reduced and consequently maximize the signal acceptance. Similarly, the HLT implements about 500 "paths", i.e. sequences of algorithmic operations that include the object reconstruction, identification and selection. Events that satisfy all the requirements of one trigger paths are directed to a corresponding data stream for their storage on tape. These streams include events for physics analyses and detector calibration, alignment and monitoring, and differ by the amount of detector information stored. A set of prescale factors is also associated to a menu of HLT paths, and evolves with $\mathcal{L}$.

Object reconstruction in the L1 trigger is performed separately using the inputs from the calorimeter and the muon subdetectors. The former are organised into trigger towers (TT), calorimeter readout units that are combined into objects representing jets, electrons, photons, and τ_{h}, and used to compute energy sums. As no information from the tracking detector is available, electron and photons result in a similar experimental signature and are both reconstructed as an e/γ object. Similarly, hits in the DT, CSC, and RPC subdetectors are combined to reconstruct muon tracks. Both the L1 calorimeter and muon trigger systems have been upgraded for the LHC Run II to improve their performance under the high instantaneous luminosity and pileup conditions expected. After a short

introduction to the Run I trigger system in Section 2.4.2, the upgraded L1 system is described in detail in Section 2.4.3.

The HLT implements an online object reconstruction and selection that is a streamlined version of the offline reconstruction algorithms. Some general principles are adopted in the HLT algorithm development and optimization. HLT object reconstruction is usually performed only locally around the L1 seed objects, reducing the time needed to read the raw detector information and to process it. Selections on variables that discriminate the signal of interest from the background are applied as early as possible to optimize the processing time, and priority is given to the least time-consuming algorithmic steps. As a consequence, calorimeters and muon systems information at full granularity is typically used earlier than tracker information. With these optimizations, HLT reconstruction in Run II followed the PF approach of reconstructing PF candidates with simplified clustering and track reconstruction algorithms. Jets are formed by clustering together these PF candidates with the anti-k_{T} algorithm. The presence of secondary displaced vertices inside the jet is used to determine whether the jet is compatible with the hadronization of a b quark. Muons are initially built from patterns of CSC and DT segments, subsequently combined to inner tracks locally reconstructed and globally fitted into a muon track. Isolation criteria based on tracks around the muon candidates and calorimetric information are used to reduce the trigger rate. Electron reconstruction closely follows the offline algorithm detailed in Section 2.3.3 and makes use of ECAL superclusters locally reconstructed around L1 e/γ seed and matched to inner tracks reconstructed with a GSF tracking algorithm adapted to HLT timing constraints. Pileup-resilient isolation criteria, based on the reconstructed PF candidates, can be applied to reduce the trigger rate. The reconstruction of τ_{h} objects at HLT is also similar to the HPS algorithm detailed in Section 2.3.4. The HLT algorithm considers up to 3 charged PF candidates clustered inside the PF jet and builds e/γ strips. The combination of the selected charged candidates and strips forms the τ_{h} HLT candidate. Timing constraints do not presently allow for evaluating the τ decay mode from all the possible combinations of charged tracks and strips as done in the HPS algorithm. Consequently, HLT τ_{h} reconstruction has a larger efficiency with respect to the offline algorithm but a background contamination of about one order of magnitude larger.

2.4.2 Run I L1 trigger system

The trigger system used for the Run I was designed to separately analyse the information from the calorimeter subdetectors (calorimeter trigger) and muon subdetectors (muon trigger), before combining it inside a global trigger (GT), as schematically represented in Figure 2.12. The calorimeter trigger followed a two-layered design: limited portions of the detector, corresponding to regions of 4×4 TTs, were processed in parallel by the boards of the regional calorimeter trigger (RCT), and subsequently combined inside the global calorimeter trigger (GCT). Each RCT had a partial view of the detector information, while the GCT could access the totality of this information but at a reduced granularity, limited by the size of the RCT itself. Similarly, hits in the DT, RPC, and CSC subdetectors were independently regrouped into track segments and combined inside a global muon trigger (GMT), where the muon candidates were identified and their p_{T} estimated. The GCT and GMT outputs were transmitted to the GT, where the decision whether accepting or rejecting the event was taken.

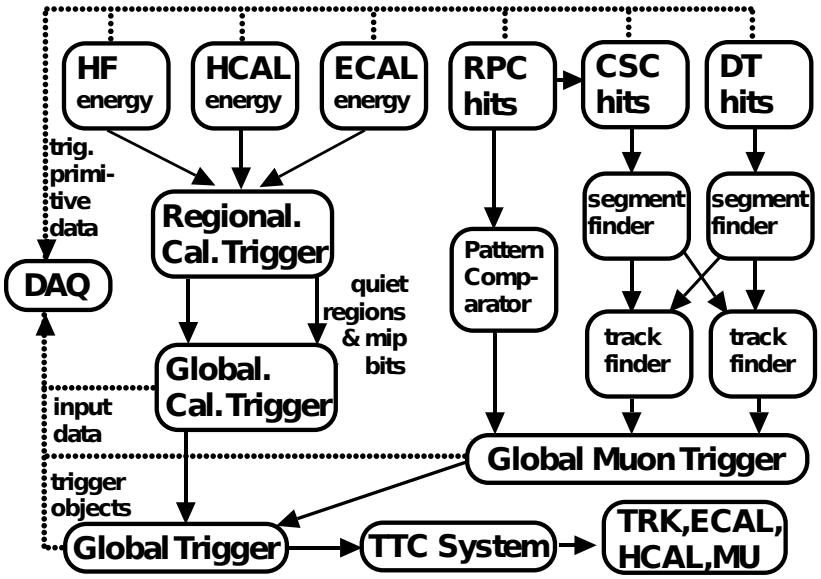


Figure 2.12 – Layout of the Run I trigger system. Information from limited portions of the calorimeter subdetectors (regions of 4×4 TTs of size) is processed in the regional calorimeter trigger (RCT) and subsequently combined in the global calorimeter trigger (GCT). Inputs from the three muon systems are separately processed and combined in the global muon trigger (GMT). The outputs of these systems are collected by the global trigger (GT) to perform the final event accept decision. The figure is taken from Ref. [105].

2.4.3 Upgrade of the L1 trigger system

After the conclusion of the long shutdown in 2015, the LHC machine increased the centre-of-mass energy of pp collisions from 8 to 13 TeV and halved the proton bunch spacing from 50 to 25 ns. These factors, combined with changes in the parameters of the machine, resulted in an increase of the instantaneous luminosity ($\mathcal{L}$) by about a factor 4 with respect to the Run I. During the 2016 data taking, values of $\mathcal{L}$ up to $1.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ were obtained in pp collisions, and this value is expected to further increase up to 2 or $2.2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ in 2017. The L1 trigger system deployed during the Run I was unable to withstand such harsh luminosity conditions without a significant increase in the trigger threshold, that would have largely reduced the acceptance to many processes with a detrimental effect for the physics programme of the CMS experiment. A major upgrade of the entire system [106, 107] has therefore been realized to maintain and improve the performance under high instantaneous luminosity and pileup conditions.

As schematically represented in Figure 2.13, the upgrade system architecture retains the subdivision in three main parts as implemented in the Run I trigger. Information from the ECAL, HCAL, and HF subdetectors is processed in the calorimeter trigger, while information from the DT, RPC, and CSC subdetectors is processed in the muon trigger. The output of these subsystems is collected by the micro global trigger (μGT), that combines it to perform the event accept or reject decision.

The upgrade benefits of the introduction of advanced mezzanine cards (AMC) technology, electronics boards that mount powerful field-programmable gate arrays (FPGAs)

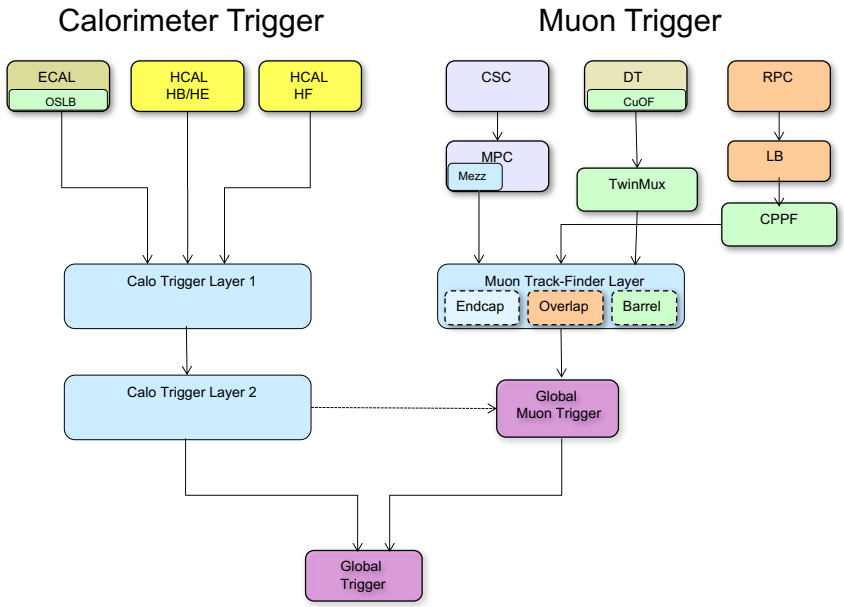


Figure 2.13 – Layout of the upgraded L1 system. Information from the calorimeter and muon detectors is separately processed and combined inside the global trigger where the event accept decision in performed. The figure is adapted from Ref. [106].

and fit into the μTCA telecommunication standard. The FPGAs are electronic circuits which functionality can be configured using a hardware description language (HDL). This guarantees a very high flexibility in the design of sophisticated algorithms, that can be tailored for the luminosity conditions of the LHC Run II and the response of the CMS detector. Owing to the common communication standard, the three subsystems use similar hardware, which ensures the flexibility and scalability of the trigger system. The communication between these boards is ensured by optical serial links with a bandwidth of 10 Gb/s, that replaced the copper parallel links limited to 1.2 Gb/s used in the Run I system to maximize data throughput. The upgrade consisted in a complete replacement of the hardware and, consequently, in the development and commissioning of new simulation software, monitoring and configuration systems, databases, timing and data acquisition interfaces.

Calorimeter trigger

The experience with the Run I trigger showed that improved granularity of the input and global view of the detector were key factors to improve the jet, τ_{h}, and e/γ identification and reconstruction. Consequently, the upgraded calorimeter system [108] has been designed to access the whole detector information at the full TT granularity instead of the 4×4 TT granularity of the RCT regions. This requires the transmission of the totality of the TTs corresponding to a specific bunch crossing to a single electronic board, an amount of data much larger than the one individually treated by each RCT board of the Run I algorithm. In the upgraded calorimeter trigger, this is made possible with the usage of a time-multiplexed trigger (TMT) architecture built into a two-layered system, schematically illustrated in Figure 2.14.

Inputs from the calorimeter subdetectors are first treated in parallel by the 18 boards of the Layer-1 system, that perform pre-processing operations such as the computation

and calibration of the total TT energy from the ECAL and HCAL energies, and the timing organization of data. The output is then distributed to one out of nine processing nodes of the Layer-2, where the identification and reconstruction algorithms are implemented; an additional redundant node is also available to redirect the data in case of failure of any other node. The output of the Layer-2 is collected by a demultiplexer node, that reorganizes the reconstructed objects, converts their energy and position coordinates to a specific format, and transmits them to the μGT. The usage of nine boards in the Layer-2 of the calorimeter trigger, each processing the information from consecutive events, introduces an additional latency of $9 \times 25\,\text{ns}$ with respect to a non-multiplexed system, and provides flexibility for the development of sophisticated algorithms.

In this approach, data distribution and algorithm implementation are physically separated in two processing layers, that can use different hardware and be optimized independently. The Layer-1 is instrumented with 18 calorimeter trigger processor cards (CTP7) [109], AMCs that embed a Xilinx Virtex-7 FPGA and are optimized for the data sharing within the crate in which they are mounted. The Layer-2 is constituted by 10 master processor cards (MP7) [110], also embedding a Xilinx Virtex-7 FPGA and designed as a generic stream-processing engine to provide the best flexibility in the development and implementation of the algorithms. The interface with the CMS data acquisition and the synchronization of the system is ensured by an AMC13 card [111, 112]. Every CTP7 provides its output on four optical links, and each MP7 board consequently receives the totality of the detector information on 72 input links through the connections realized in a custom patch panel. Because the Layer-1 adopts a parallel approach, where each CTP7 processes the information from a 20° region of the detector, data must be reorganized serially before being sent to the Layer-2, introducing an important latency in the process. This is compensated by an MP7 clock frequency of 240 MHz, six times faster than the LHC collision frequency, that allows for the transmission of the totality of the data within less than seven bunch crossings. To further reduce the latency, algorithms are designed to start the processing as soon as a minimal amount of data is received.

The change of paradigm with respect to the Run I calorimeter trigger offers unprecedented possibilities in the development of efficient identification algorithms at the hardware level. This will be illustrated in this thesis in the context of τ_h identification, to which Chapter 3 is entirely devoted, and where the largest improvements with respect to the Run I performance have been achieved. Nevertheless, the reconstruction of all the objects benefits of the upgraded calorimeter trigger [113]. The access to the full TT granularity improves the spatial resolution by more than a factor of 4 and opens up the possibility to precisely identify and cluster the relevant energy deposits, improving the trigger resilience to pileup. The improved granularity allows for computing the invariant mass or the spatial correlation of object combinations, allowing the selection of specific signal topologies. This ultimately brings HLT identification criteria to the hardware level of the L1 trigger.

Muon trigger and global trigger

The upgrade of the muon trigger changes the muon track reconstruction approach used at Run I and, instead of combining the three subdetectors in a later stage, exploits the redundancy of the systems at an earlier stage. The DT, RPC, and CSC track finders are thus replaced by a barrel (BMTF), overlap (OMTF), and endcap (EMTF) track finder systems. The BMTF covers the region $|\eta| < 0.83$ and uses the information from DT and

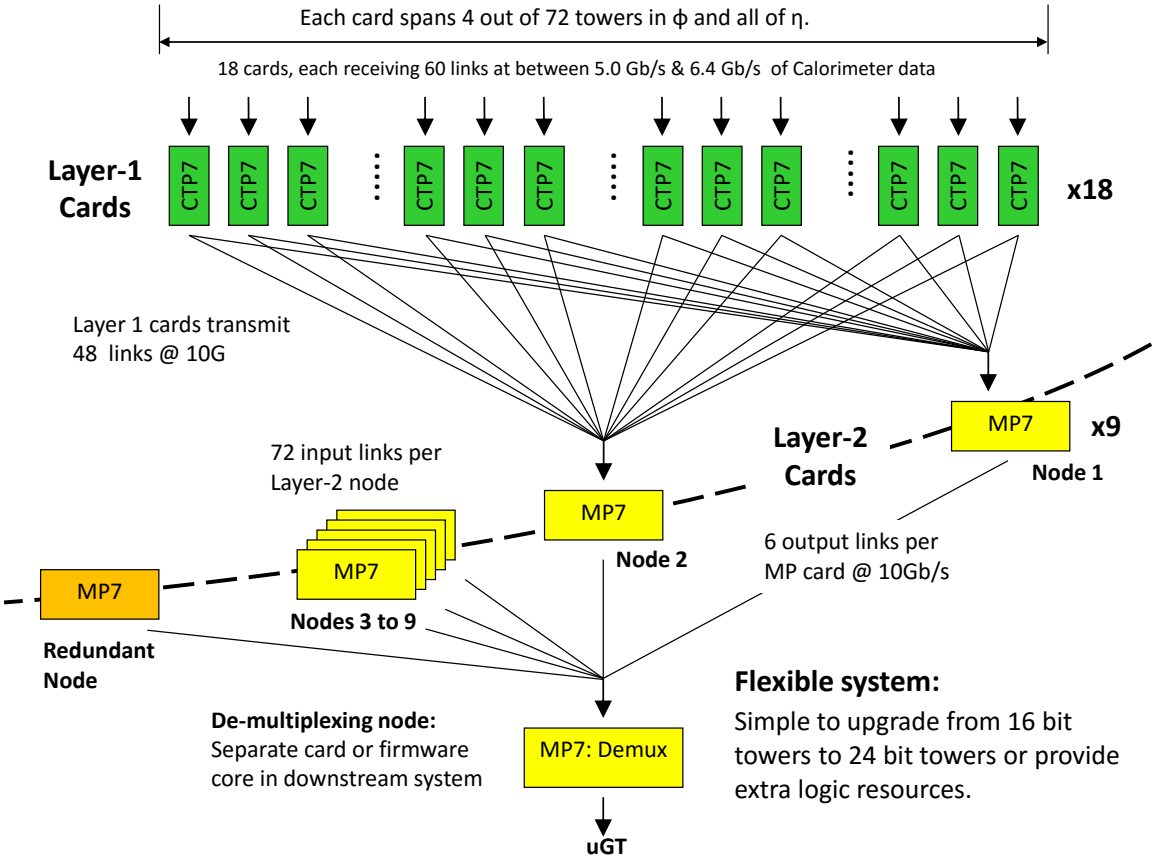


Figure 2.14 – Layout of the upgraded calorimeter trigger system. Inputs from the calorimeter subdetectors are pre-processed in the Layer-1 and distributed to one out of the nine processing nodes of the Layer-2, with an additional redundant node being available for data redirection. The output is then collected by a demultiplexer node and sent to the μGT. The figure is taken from Ref. [106].

RPC subdetectors, that are combined into "super-primitives" in the TwinMUX system: their redundancy improves the precision in the determination of the muon hit position, and the bending angle information available from the DT is used in a track finding algorithm based on "road search" extrapolation. The number of the hits and the quality of the extrapolation procedure determine a muon quality criterion that is used to control the trigger rate. The OMTF covers the intermediate region $0.83 < |\eta| < 1.24$ while the EMTF processes the forward region $|\eta| > 1.24$. Because of the very short latency of about 750 ns available for these systems, fast pattern-recognition algorithms are implemented and convert patterns of hits into p_{T} assignments. The number and topologies of the hits are also used to determine muon quality criteria. Once again, the redundancy of CSC and RPC information is exploited with the usage of a concentration, pre-processing and fan-out (CPPF) card. The outputs of the BMTF, OMTF, and EMTF systems is collected by the global muon trigger (μGMT), that ranks the muon candidates by p_{T} and quality, removes duplicates reconstructed at the boundaries of the system, and transmits the output to the μGT. The very different features of the algorithms implemented in the three muon trigger systems result in the usage of different hardware. In the BMTF (where large computing power is needed) and in the TwinMUX, the MP7 cards are used,

showing the flexibility of this hardware and the high level of standardization achieved in the upgraded L1 trigger system. In the OMTF and EMTF, where large memories are needed to store the associative patterns, modular track finder (MTF7) cards [114] are used.

The final decision whether accepting or rejecting the event is taken in the μGT system, and is based on the properties of the objects reconstructed in the calorimeter and muon trigger systems. The μGT is instrumented with MP7 boards that, owing to their large computing power, make it possible to compute multi-object correlation and global event quantities to enhance the signal acceptance to specific processes and reduce the event acquisition rate. In addition to counting the number of the objects above a specific p_{T} threshold as done in the Run I GT, the μGT can thus compute the invariant mass of pairs of reconstructed candidates, their spatial separation, and their spatial momentum sum. These quantities are used to identify the decays of resonances or topologies associated to specific production mechanisms such as the vector boson fusion. These algorithms were implemented in 2016 as a menu of about 300 seeds, running on four MP7 boards. Two additional boards have been added in view of of the 2017 data taking, allowing the implementation of a menu containing about 500 seeds.

Timeline of the upgrade

To ensure a high reliability of the system and minimize the risk of failures during the data taking, the CMS Collaboration adopted a strategy consisting in a progressive upgrade of the system that has been realized between 2015 and 2016.

The Run I GT was upgraded to the μTCA-based μGT since the beginning of the 2015 data taking, although with a reduced number of MP7 boards that has been extended subsequently.

The hardware of the muon trigger system was updated for 2015 collisions where the BMTF, OMTF, and EMTF systems were used in the data taking with reduced input information. The combination of separate muon system information was commissioned during 2015 and 2016 collisions. The full redundancy of the subdetector was thus not available for the results presented in this thesis, but the TwinMUX and CPPF systems are now operational for 2017 data taking.

An improved version of the Run I calorimeter trigger, denoted as the "stage-1", was deployed for the 2015 data taking [115]. The stage-1 was designed as an interim system, and maintained the limitations of the Run I algorithm in terms of regional view of the detector. Its identification algorithms implemented pileup estimators and other ameliorations, allowing it to cope with the luminosity conditions of 2015 data taking that did not exceed an instantaneous luminosity of $5.2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In parallel, the outputs from the ECAL, HCAL, and HF subdetectors were duplicated and used to commission the "stage-2", the full upgrade of the L1 calorimeter trigger that is simply referred to as the "upgraded system" in the rest of this document. The upgraded system was installed and included in the CMS data acquisition system in 2015, although not used to perform the event accept decision. This made it possible to measure its performance in data and to commission it without affecting the data taking operations, as it is described in Section 3.7 of Chapter 3 in the context of the τ algorithm. Once the commissioning completed, the upgraded system was deployed for the 2016 data taking. It served as the CMS L1 calorimeter trigger for the entire data taking and used to collect the data on which are based the results presented in this thesis.

Chapter 3

The L1 τ trigger

Tau leptons are the heaviest leptons known. In the SM, their large mass arises from the sizeable Yukawa coupling to the Higgs boson, which makes them an essential tool in the exploration of the electroweak symmetry breaking and, more generally, of BSM physics. Decays of the Higgs boson to τ leptons represent the most sensitive final state to study its direct coupling to fermions, and many extensions of the scalar sector of the SM can only be probed in final states containing τ leptons. A few examples are heavy scalar resonances decaying to a τ lepton pair and charged Higgs bosons decays $H^+ \to \tau^+\nu_\tau$ predicted in the MSSM, as discussed in Chapter 1, or searches from lepton flavour violation using the $H \to \mu^\pm\tau^\mp$ process. In the context of HH searches, the $HH \to b\bar{b}\tau^+\tau^-$ decay channel is one of the most sensitive to both SM and many BSM production modes as it is discussed in this thesis.

The sensitivity of analyses involving τ leptons crucially depends on the capability to identify and reconstruct them. This starts with the very first selection that is performed by the L1 trigger. Decays of the τ lepton to a muon or an electron and neutrinos can be reconstructed at L1 using the clean signatures of these particles in the detector. In contrast, decays of a τ lepton to final state hadrons (τ_h) and a neutrino are extremely challenging to identify given their similarity to hadron jets initiated by quarks or gluons. Further constraints are induced from hardware limitations in the L1 trigger, requiring a high optimization of the sequence of algorithm operations to be performed.

In the Run I, a simple approach to τ_h reconstruction based on a regional jet identification has been adopted because of the L1 calorimeter trigger structure and computing power of its boards, and resulted in a limited τ_h identification efficiency. The upgrade of the L1 system gives the possibility to develop, for the first time at a hadron collider, an efficient and dedicated τ finder algorithm at the hardware trigger level. I had the chance to start my PhD work during the finalization of the upgrade project and to contribute to all the phases of the algorithm development, starting from the design of the algorithm itself, moving subsequently to its optimization to improve the performance and adapt it to the hardware constraints, up to the commissioning phase, and finally to the measurement of its performance with the data collected by the CMS experiment. This has been a very interesting activity, that is still evolving, where the most rewarding part has been the successful performance to select events during all the 2016 CMS data taking phase.

This chapter describes the τ algorithm developed for the L1 upgrade and its performance. The discussion is introduced by a general overview of the experimental challenges of τ_h identification at the hardware level, followed by a short description of the algorithm used in the Run I and of its limitations. The τ algorithm developed for the L1 trigger upgrade was progressively optimized to fulfil performance and hardware constraints, and

is further being tuned to 2017 LHC conditions. The description that is reported here focuses on its version optimized for the 2016 data taking. A summary of the changes and optimizations occurred during the development is given in the context of the algorithm integration in the L1 electronics. Finally, the commissioning of the algorithm with 2015 data is presented, and its performance measured with the data collected in 2016.

3.1 Experimental challenges of a Level-1 τ trigger

The L1 τ algorithm represents the first level of the CMS event selection of final states with τ leptons decaying semileptonically. As such, it must satisfy demanding requirements of high signal efficiency (to maximise the sensitivity of the physics analyses) and large background rejection (to maintain the trigger rate under control). In addition, tight technical constraints are imposed by the electronics and the structure of the trigger system.

The experimental signature of a τ_{h} candidate is a narrow and collimated jet. This characterizes it against jets initiated by a quark or a gluon, that are typically broader and with higher particle multiplicity. As discussed in Section 2.4 of Chapter 2, the L1 τ algorithm has no access to the tracker detector information. The first challenge is thus represented by the limited detector information. It requires the development of a reconstruction algorithm that relies solely on the calorimeters inputs. Localized high energy deposits hint at the presence of a τ_{h} interaction, while a larger energy spread is usually associated to a jet background. A τ algorithm must be able to efficiently identify and collect all the relevant energy deposits and use them to estimate the τ_{h} energy.

The second challenge is represented by the background rejection. As most of the proton bunch crossings produce deposits of energy in the calorimeters, an efficient reduction must be achieved to maintain the trigger rate at a level that can be managed by the CMS data acquisition system. While part of the background can be suppressed with the application of a minimal threshold on the τ_{h} energy, the development of complementary rejection criteria is mandatory to maintain these thresholds sufficiently low and consequently ensure a high efficiency to signal processes. This is particularly demanding under the pileup (PU) and instantaneous luminosity conditions of the LHC Run II, possibly as high as $2.2 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ with an average PU of about 40. These harsh collision conditions complicate the background rejection because of the increase in the energy and number of jets, and require a τ algorithm that is resilient against the PU.

A third challenge is represented by the technical limitations of the L1 trigger itself. Being implemented in an electronics system that is fully synchronous with the LHC, the entire L1 trigger has a fixed time to perform a decision ("latency") of 3.8 µs. This time must account for the data transfer and processing, and consequently only about 1 µs of latency is available for the algorithms. The electronic boards have also a finite amount of logic gates and memory available for the algorithm implementation ("hardware resources"). The combination of these effects limits the number and complexity of the operations that can be performed and requires a high optimization of the τ algorithm.

The upgrade of the hardware system of the L1 upgrade trigger discussed in Section 2.4.3 of Chapter 2 offers unprecedented possibilities to face these challenges. Thanks to the large computation power of the FPGA implemented in the electronic boards and to the additional latency provided by the time multiplexed architecture, sophisticated algorithms can be developed for the identification of τ_{h} interactions in the calorimeters.

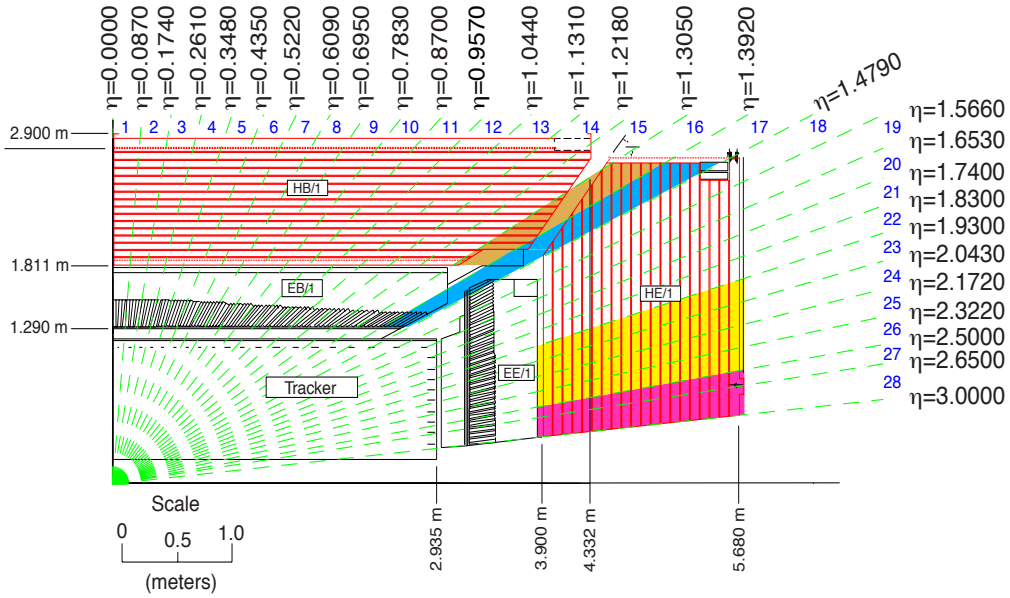
These include a dynamic clustering of the relevant energy deposits, an isolation criterion, and background rejection based on the topology of the energy deposits.

3.2 Inputs to calorimeter trigger algorithms

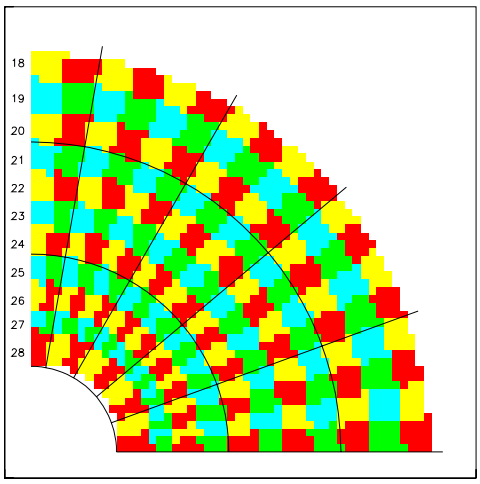
As introduced in Section 2.4.1 of Chapter 2, the calorimeter trigger algorithms receive information from the ECAL, HCAL, and HF subdetectors, and use them to reconstruct e/γ, τ_{h}, and jet candidates, and to compute energy sums. The information is transmitted from the calorimeters to the trigger system in the form of "trigger primitives", that are digital quantities corresponding to the 40 MHz digital samplings of the detector pulses. As these pulses span over multiple bunch crossings, the trigger primitives generator electronics assign every trigger primitive to a specific bunch crossing using a digital filtering combined with peak finding techniques. The projection onto the transverse plane of the momentum vector originating in the detector centre and pointing to the calorimeter cells is computed for every trigger primitive from the sampled energy and is denoted as E_{T}.

Trigger primitives are organized in trigger towers (TT), each encoding the sum of the ECAL and HCAL energies deposited in a localized detector region. In the barrel part of the CMS detector, each TT has an extension of 0.087×0.087 in the η and φ directions, corresponding to 5×5 ECAL crystals and one HCAL readout unit. Each half-barrel is therefore divided in 17 towers in η and 72 towers in φ, identified in a Cartesian coordinate system by a pair of indexes $(i\eta, i\varphi)$, as schematically represented in Figure 3.1a. In this Cartesian notation used in the following, the $i\varphi$ position is represented by a number between 1 and 72, and the $i\eta$ position is a positive or negative index which absolute value increases with the absolute value of η, while 0 is not considered as a valid position. The towers with $|i\eta| \leq 17$ thus correspond to the barrel. A more complex definition of the TT exists in the endcap regions of the detector because of the different geometry of the ECAL crystals, as represented in Figure 3.1b. The size and number of ECAL crystals of each TT consequently increase with η, matching the corresponding readout unity of the HCAL. Each endcap region contains 11 towers in η and 72 towers in φ identified by $(i\eta, i\varphi)$ coordinates. In the TT geometry, the endcaps are composed by the towers $18 \leq |i\eta| \leq 28$. Inputs from the HF subdetector are organized in a coarser granularity, with 4 towers in η and 18 in φ, covering the region $29 \leq |i\eta| \leq 32$. Further details on the complete TT geometry can be found in Ref. [105]. The value of E_{T} is encoded in a digital representation using a linear scale with a 0.5 GeV unit. This is therefore the minimal non-null energy value that is represented in a TT and the granularity to which the energy can be expressed. Any TT with an energy $E_{\mathrm{T}} \geq 0.5$ GeV is referred to as an "active tower".

Each TT is represented in a digital form by 16 bits that encode the total transverse energy (ECAL and HCAL transverse energy sum) on 9 bits, the ratio of the ECAL and HCAL energies on 5 bits and HCAL and ECAL quality flags on the two remaining 2 bits. In the time-multiplexed architecture of the L1 trigger discussed in Section 2.4.3 of Chapter 2, any MP7 board in the Layer-2, where the algorithms are implemented, receives from the Layer-1 the complete set of TT from a specific bunch crossing. Any of the 72 input links to the MP7 boards transmits 32 bits for every card clock cycle, corresponding to a pair of TT at the same $i\varphi$ and opposite $i\eta$ positions, i.e. a pair $(i\eta, i\varphi)$ and $(-i\eta, i\varphi)$. Consequently, each MP7 board receives simultaneously two "rings" of TT at every clock cycle, starting from the most central ring up to the most peripheral ones.



(a) r–z view



(b) x–y view , endcap

Figure 3.1 – (a) Layout of the TT boundaries in the r–z plane. Each TT regroups inputs from both the ECAL and HCAL subdetectors. (b) Layout of the TT definition in the endcap regions of the detector. Each square denotes an ECAL crystal, and regions with the same colors represent one TT. Both figures are taken from Ref. [105].

3.3 The Run I τ algorithm

Following the organization of the TT inputs in RCT regions, i.e. groups of 4×4 TT, as discussed in Section 2.4.2 of Chapter 2, the τ algorithm implemented in the Run I trigger system [116] adopted a regional approach. L1 jets are first identified and reconstructed, and subsequently investigated using isolation and τ_{h} identification criteria. The jet reconstruction is based on the RCT calorimeter regions, combined into 3×3 groups to define the jet active area, as schematically represented in Figure 3.2. This area corresponds approximately to the area of a jet reconstructed offline with the anti-k_{T} jet algorithm, operated in Run I with a parameter of $\mathrm{R} = 0.5$. The central calorimeter region is defined

as the one with the highest E_{T} sum, and the jet energy is computed as the E_{T} sum over all the TT included in the nine calorimeter regions. An isolation criterion is defined by requiring that at least seven out of the eight non-central trigger regions contain an energy deposit $E_{\mathrm{T}} < 2\,\mathrm{GeV}$, and is complemented by an identification criterion that requires the energy deposits in each calorimeter region not to be spread over more than 2×2 TT. The energy deposit topologies that are accepted by this identification requirement are illustrated in Figure 3.2.

Although providing a simple and robust method for τ_{h} identification, this approach has an efficiency that is limited to about 60%, and larger inefficiencies are observed at high p_{T} because of the definition of the isolation criterion. For this reasons the L1 τ trigger needed to be complemented by jet triggers, which in turn required high p_{T} trigger thresholds: the double τ_{h} L1 trigger seeding the HLT path used in Run I searches [117, 118] required the presence of a L1 τ_{h} candidate pair of $p_{\mathrm{T}} > 44\,\mathrm{GeV}$, or of a pair of central jets of $p_{\mathrm{T}} > 64\,\mathrm{GeV}$.

While this approach to τ_{h} identification at L1 trigger level has guaranteed a sufficient identification efficiency for Run I, it would not have been sustainable for the Run II data taking, where both the instantaneous luminosity and the number of interactions per proton bunch crossing increased by more than a factor of 2 and are expected to be even larger in the upcoming data taking phase of 2017. As a consequence, a profound change in the approach to τ_{h} identification has been adopted for the L1 trigger upgrade.

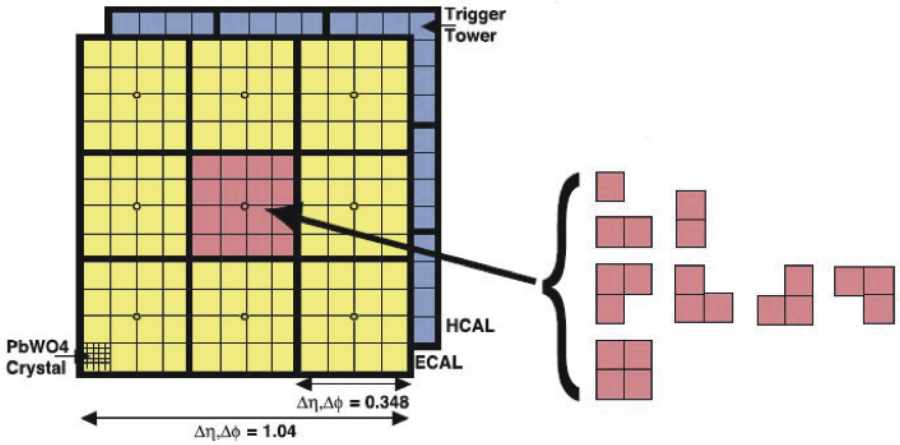


Figure 3.2 – Schematic representation of the calorimeter regions considered in the τ algorithm for Run I. The jet and τ_{h} reconstruction region is composed of 3×3 calorimeter regions, each composed of 4×4 trigger towers. The central calorimeter region, highlighted in red, is used to reconstruct the τ_{h} candidate, and the surrounding 8 calorimeter regions are used to compute isolation criteria. A shape criterion is also computed as shown on the right. If any of the 9 trigger regions fails the shape criterion the τ_{h} is rejected. The figure is taken from Ref. [116].

3.4 The τ trigger algorithm for the CMS L1 trigger upgrade

The τ algorithm developed for the L1 upgrade is composed of five steps, designed to collect the relevant energy deposits, estimate the τ_{h} energy, and reject the background:

1. clustering of TTs to identify the relevant localized energy deposits
2. merging of clusters into a single L1 τ_{h} candidate
3. energy calibration to improve the τ_{h} scale and resolution
4. isolation criterion to reject quark and gluon jet background
5. veto of background-like clusters based on the cluster shape

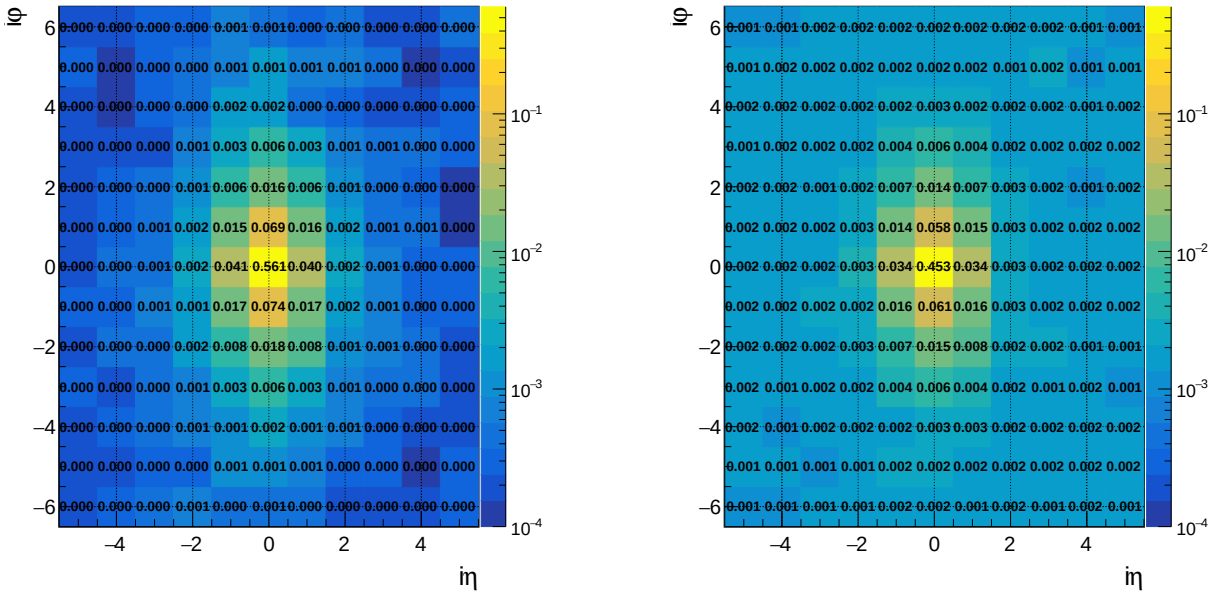
In contrast to the regional approach of the Run I algorithm, the TMT architecture discussed in Section 2.4.3 of Chapter 2 distributes the complete set of TT of an event to a single MP7 board where the τ algorithm is implemented. As detailed previously in Section 3.2, two "rings" of 72 TT are received simultaneously as input at every clock cycle, starting from the two most central ones at $i\eta = \pm 1$ and progressively moving to the edges of the detector. In order to minimize the latency, the algorithm is designed to start as soon as a minimal amount of information has been received, resulting in a fully pipelined structure.

The algorithm is implemented in the MP7 cards of the Layer-2 of the trigger system. Algorithms, input and output links control, monitoring and configurations registers are globally denoted as the board "firmware". The firmware is developed in VHDL (Very high speed integrated circuits Hardware Description Language), a language that allows to configure the internal connections of the FPGA embedded in the MP7 cards to perform a specific set of instructions. The development of the algorithm and the study of its performance, as well as its simulation in MC samples, is performed using an "emulator", i.e. a C++ code that replicates the behaviour of the VHDL implementation, and has been used to derive the majority of the results shown in this section.

3.4.1 Clustering

This first step of the τ algorithm aims at identifying the energy deposits in the calorimeters due to the τ_{h} interaction, including the energy emitted by bremsstrahlung radiation and from $\pi^0 \to \gamma\gamma$ decays, while minimizing the impact of smaller and more diffuse energy contributions due to PU. The spatial distribution of the energy deposited in the calorimeters by a τ_{h} interaction is represented in Figure 3.3 for simulated $\mathrm{Z} \to \tau\tau$ events and in presence or absence of the pileup contribution. About 55% of the energy is deposited in a single TT and about 85% is found in the neighbour TTs (within a distance of one unity in $i\eta$ or $i\varphi$), while the presence of pileup in the event produces a low-energy and diffuse contribution. The clustering step of the τ algorithm has consequently been designed to efficiently collect these specific configurations of energy deposits in the calorimeters by selecting small groups of TTs to minimize the impact of pileup.

The creation of a L1 τ candidate is initiated, or "seeded", by local maxima of energy in a region that extends over 3 TTs along the η direction and over 9 TTs along the φ one.



(a) Without pileup (b) With pileup

Figure 3.3 – Average repartition of the energy in simulated $Z \to \tau\tau$ event in absence (a) and presence (b) of pileup contribution in the event. The figure is centred around the local maximum of energy and is realized by averaging the energy repartition patterns observed in a sample of about two thousand τ_h candidates in the barrel part of the detector.

Because of the presence of charged pions in a τ_h candidate, energy deposits are expected in both the ECAL and HCAL subdetectors, and the sum of the two energies is used. Ambiguities in the presence of equal energy TTs inside the same region are resolved by applying asymmetrically a strict ($<$) or loose ($\leq$) energy comparison depending on the relative position of the towers. Valid seed towers are required to have a transverse energy $E_T \geq 4\,\text{GeV}$, and to be in the barrel or endcap regions of the detector but not in the forward calorimeters, corresponding to a position $|i\eta| \leq 28$.

The eight TTs within a distance of one unit in $i\eta$ or $i\varphi$ from the seed are regrouped in a "proto-cluster" if they have a transverse energy $E_T \geq 2\,\text{GeV}$. The two TT with the same $i\eta$ position as the seed one and $\Delta(i\varphi) = \pm 2$ are also added to the proto-cluster if they satisfy the same energy requirement and if, in addition, the TT that separates them from the seed is also included. The extensions along the φ direction are needed to recover the bremsstrahlung radiation and, in the case of multi-particle τ decays, part of the energy that is spread in the φ direction because of the separation of final state hadrons due to the magnetic field. In total, up to 11 TTs can be regrouped inside a proto-cluster, as schematically represented in Figure 3.4.

The proto-cluster can extend over a maximum of 3 TTs in the $i\eta$ direction, but the energy deposition patterns from a τ_h interaction are typically more compact. This feature cannot be observed from Figure 3.3 because the average performed on multiple candidates results in an artificial spread of the distribution. Using the notation of Figure 3.4, the groups of TT in $(2, 3, 4)$ and $(2', 3', 4')$ can be denoted as the right and left side of the proto-cluster, respectively. It is observed in simulated $Z \to \tau\tau$ events that about 70% of

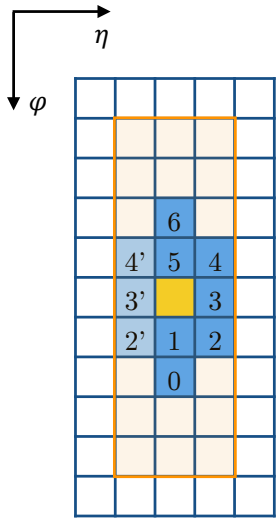


Figure 3.4 – Schematic representation of the area considered in the clustering step of the τ algorithm. Every square denotes a TT, the yellow one is the seed TT and the blue ones are those that can be grouped inside a proto-cluster. The orange shading represents the area in which the seed TT is required to be a local maximum. Both the sides labelled with primed and non-primed indices are considered for the creation of the proto-cluster, but the lowest energy side is subsequently removed, resulting in the final cluster as detailed in the text.

the proto-clusters do not have any active tower in either their left or right region, with about 25% of these clusters having no active tower in both regions. For the resulting 30% of the proto-clusters, the energy deposit is highly asymmetric, and the lowest energy one contributes only to a small fraction of the total energy collected (less than 10% for 90% of the reconstructed clusters), as it is shown in Figure 3.5. Therefore, the proto-cluster side with the lowest E_{T} sum is removed, resulting in the final cluster. In the following, this procedure is referred to as "lateral trimming". A bit is set to 1 or 0 whether the trimming occurs on the left or right side of the proto-cluster. This quantity is part of the numerical representation of the cluster in the firmware implementation, and is used later in the τ algorithm.

The position assigned to the cluster corresponds to the center of the seed TT. Depending on the cluster shape, an additional offset of 1/4 of the TT size in the η and φ directions can be assigned to improve the precision. Such offset is determined according to asymmetries in the energy pattern depositions inside the cluster. For the φ direction, this is done by comparing the energy deposited in the upper and lower parts of the cluster, i.e. in the selected towers $(0, 1, 2/2')$ and $(4/4', 5, 6)$ in the notation from Figure 3.4, and shifting the φ position towards the highest energy part. Similarly, for the η direction, a positive or negative offset is assigned depending on the presence of active towers in its right or left regions, according to the lateral trimming bit. The nine positions within the TT surface that can result from the assignment of the φ and η offsets are summarized in Figure 3.6.

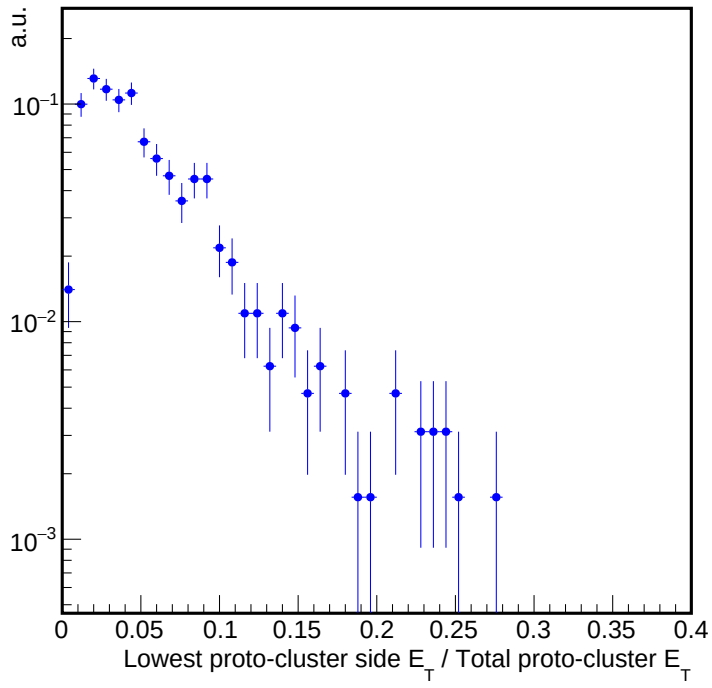


Figure 3.5 – Distribution of the lowest-E_{T} proto-cluster side and of the complete proto-cluster energies ratio. Only proto-clusters with at least one active towers in both their right and left regions are considered, corresponding to about 30% of the total number of proto-clusters formed in simulated $\mathrm{Z} \to \tau\tau$ events.

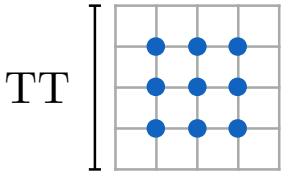
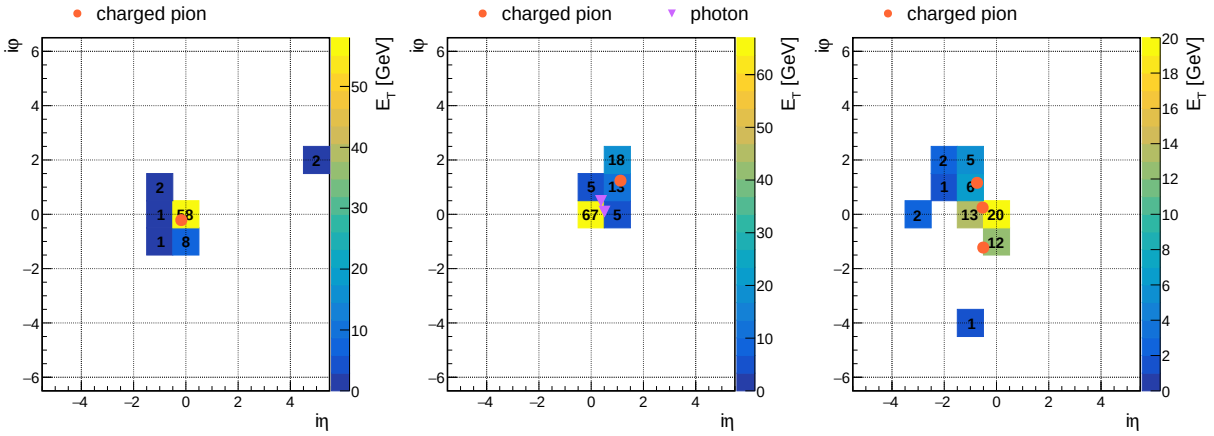


Figure 3.6 – Schematic representation of the cluster fine grain position assignment. Depending on the number and position of the TTs grouped, the cluster position is assigned an offset of one fourth the size of the seed TT, resulting in one out of the nine possible positions indicated with the blue dots.

3.4.2 Merging

The clustering procedure described in the previous section collects efficiently the energy from the interaction of a hadron issued from a τ decay in the calorimeter. However, decays containing multiple hadrons such as $\tau \to \pi^{\pm}\pi^{0}\nu_{\tau}$ or $\tau \to \pi^{\pm}\pi^{\pm}\pi^{\mp}\nu_{\tau}$ are significantly spread over the detector due to the magnetic field, and their energy is only partly collected into a single cluster. The $\Delta(i\varphi) = \pm 2$ extensions are not wide enough to collect the totality of the energy. The usage of a larger cluster area would make the algorithm more sensitive to PU contribution, but even more importantly introduces a real challenge for the firmware implementation, because of hardware constraints that require the clustering phase to be shared with the e/γ algorithm. Some examples of typical energy deposits are shown in Figure 3.7, where the spread along the φ direction is clearly visible for decays



(a) 1-prong (b) 1-prong + π^0 (c) 3-prong

Figure 3.7 – Examples of energy deposition patterns in the calorimeter subdetectors in case of a 1-prong (a), 1-prong + π^0 (b), or 3-prong (c) decay of a τ lepton. The color scale and the numbers in the figures denote the energy deposited in the corresponding TT in GeV units. The increase of the spatial spread of the energy deposits for decay modes with more than one particle in the final state is clearly visible. The cluster merging procedure is put in place to recover the energy not collected within a single cluster.

with multiple hadrons or with photons in the final state.

In order to recover this fraction of the energy, nearby clusters can be merged into a single L1 τ candidate. The "main" cluster is the one seeded by the TT with the largest energy in a window extending 3×9 TT in the η and φ directions. Its coordinates $i\eta^{\text{main}}$ and $i\varphi^{\text{main}}$ determine the coordinates of the L1 τ_{h} candidate. A "secondary" cluster is associated to the main one if the corresponding seed is found within one of the eight positions that are represented in Figure 3.8. Valid secondary clusters seeds must be local energy maxima with respect to the neighbouring TTs (i.e. within a 3×3 TT region centred on them). Secondary clusters are constructed in the same way as the main one as detailed in Section 3.4.1, but they do not undergo the lateral trimming. Following the numbering convention in Figure 3.8, clusters that are seeded in positions number 2 or 5 are not considered if the seed TT is already included in the main cluster. This pattern is chosen according to the typical spread of the energy deposit in τ lepton decays, that occurs in the φ direction because of the effect of the magnetic field. In case multiple secondary cluster candidates are found, the one with the largest energy is selected. Equal energy secondary cluster candidates on the same φ side with respect to the main cluster can only occur in sites 4 and 6 if $\varphi < \varphi^{\text{main}}$ and in sites 1 and 3 if $\varphi > \varphi^{\text{main}}$ because of the local 3×3 energy maximum requirement on their seed. If this case occurs, ambiguities are resolved by selecting the site with the lowest value of $|\eta|$. In case secondary clusters with the same energy are found on the opposite φ sides with respect to the main one, the cluster with the smallest $i\varphi$ is arbitrarily chosen.

Latency constraints require that both the main and the secondary clusters are constructed at the same time, therefore some TTs can be simultaneously associated to two different clusters, resulting in a double counting of their energy. An overlap removal procedure is put in place to solve this problem by removing these TT from the main cluster.

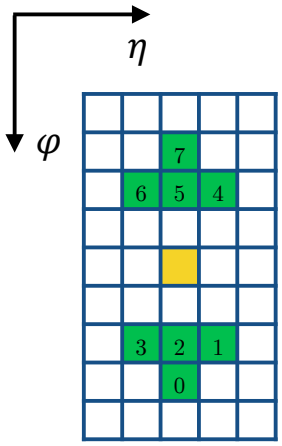


Figure 3.8 – Schematic representation of the area used for cluster merging. Every square represents a TT, the yellow one is the seed TT and the green one are the TTs considered as seeds for the secondary cluster. The numbering scheme of the secondary cluster seeds adopted in the firmware implementation is reported in the figure.

The towers that are associated to multiple clusters can be computed a priori knowing the secondary cluster position, the trimming direction of the main cluster, and whether the main and secondary clusters contain the TT at $\Delta(i\varphi) = \pm 2$ from the seed tower. The positions of the overlapping TTs to be rejected are encoded into 7 bits (corresponding to the 7 position of a TT inside a cluster) that are associated to the 256 possible combinations of inputs using a look-up table (LUT).

The fraction of merged clusters is about 15% for τ_{h} candidates for a typical signal sample. This value is however largely dependent on the p_{T} of the τ_{h} and on its reconstructed decay mode, as shown in Figure 3.9. In particular, up to 35% of low p_{T}, three-prongs decays are identified as merged cluster, and the fraction decreases at higher p_{T} values because of the smaller bending of the charged hadrons due to the magnetic field.

3.4.3 Calibration

After the creation and merging of the TT clusters, a direct estimation of the τ_{h} candidate energy ($E_{\mathrm{T}}^{\mathrm{raw}}$) is computed as the sum of the clustered TTs energies. A calibration procedure is subsequently performed to correct this value and improve the resolution with respect to the energy of the τ_{h} candidate reconstructed offline. Although the energy of the single TTs is already calibrated in the Layer-1 of the trigger system, this second calibration procedure is motivated by the presence of residual energy losses due to the clustering, especially in the endcap regions of the detector, characterized by a more complex TT geometry. Non-linearities in the calorimeter response and differences between the ECAL and HCAL response can also introduce additional effects that deteriorate the resolution. Finally, the merging procedure can introduce systematic differences between merged and unmerged clusters that must be taken into account and corrected.

The calibration procedure implemented consists in correcting $E_{\mathrm{T}}^{\mathrm{raw}}$ by a multiplicative factor c to compute the calibrated energy E_{T}. The factor c is a function of $E_{\mathrm{T}}^{\mathrm{raw}}$ itself and of the $i\eta$ coordinate, of the merging status of the cluster (i^{merged}), and of the presence of an energy deposit $E_{\mathrm{T}} \geq 0.5\,\mathrm{GeV}$ in the ECAL subdetector corresponding to the seed TT

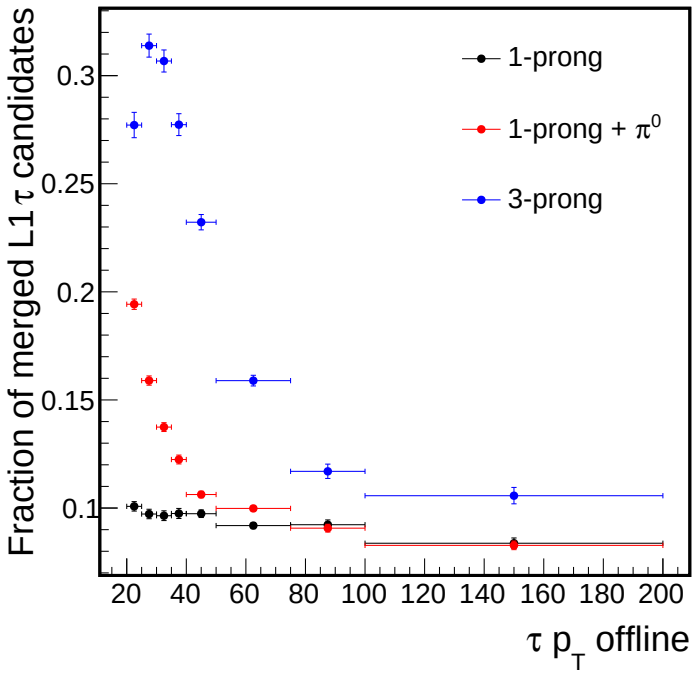


Figure 3.9 – Fraction of L1 τ_{h} candidates reconstructed as the merging of two clusters as a function of the transverse momentum of the τ_{h} candidates reconstructed offline. The black red, and blue, curves denote respectively the 1-prong, 1-prong+π^0, and 3-prong reconstructed decays of the τ lepton. The value of each bin is computed independently from the others from the number of τ leptons reconstructed in the corresponding decay mode and p_{T} range.

of the main cluster (i^{EM}):

$$E_{\mathrm{T}} = c(E_{\mathrm{T}}^{\mathrm{raw}}, i\eta, i^{\mathrm{merged}}, i^{\mathrm{EM}}) \times E_{\mathrm{T}}^{\mathrm{raw}} \tag{3.1}$$

The inputs i^{merged} and i^{EM} are represented on a single bit while $i\eta$ and $E_{\mathrm{T}}^{\mathrm{raw}}$ are expressed on 5 and 13 bits respectively. Because hardware resources constraints limit the total number of input addresses of a LUT (or "LUT size"), the $i\eta$ and $E_{\mathrm{T}}^{\mathrm{raw}}$ values are compressed, i.e. converted with a non linear transformation to another value that is expressed on 4 and 5 bits respectively. This transformation is further encoded in a separate LUT, and for the $E_{\mathrm{T}}^{\mathrm{raw}}$ compression values of $E_{\mathrm{T}}^{\mathrm{raw}} \geq 256$ (in hardware units) are saturated to $E_{\mathrm{T}}^{\mathrm{raw}} = 255 = 2^8 - 1$ to further reduce the LUT size. As the typical trigger thresholds for τ_{h} candidate range between 30 and 40 GeV, a precise calibration of high energy candidates is not crucial and this solution is adopted to improve the performance of the calibration method in the region of interest.

The factor c is encoded on 10 bit and represents a decimal value in a linear scale between 0 and 2, while $E_{\mathrm{T}}^{\mathrm{raw}}$ is encoded on 13 bit and expressed in hardware units. The calibration consists in the computation of the product $cE_{\mathrm{T}}^{\mathrm{raw}}$, that needs 23 bits to be represented. The result is truncated to the 14 highest significant bits to perform the division by 2^{10}, taking into account the scale on which c is expressed. Finally, the result is saturated to 12 bits, i.e. E_{T} is set to $2^{12} - 1$ if it exceeds such value.

The values of the c factors are computed in simulated events in exclusive intervals of the calibration input variables. For any of these intervals, the ratio of the uncalibrated τ_{h} candidate energy $E_{\mathrm{T}}^{\mathrm{raw}}$ to the τ_{h} momentum reconstructed offline $p_{\mathrm{T}}^{\mathrm{offline}}$ is considered and the value of c is computed as the inverse of the average of this distribution. The effect of the calibration on the energy of the L1 τ_{h} candidates is shown in Figure 3.10. The calibration constants clearly depend on the TT calibration performed in the Layer-1 of the trigger system. Although this was changed a few times during the preparation of the 2016 data taking and is being updated for the upcoming 2017 data taking, the τ_{h} calibration procedure described before is able to compensate for any difference in the inputs, providing a robust determination of the τ_{h} energy.

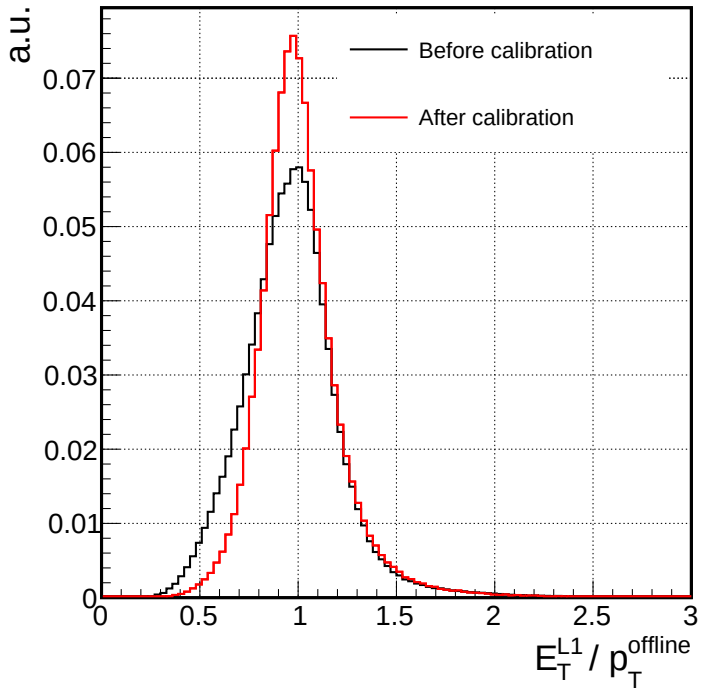


Figure 3.10 – Comparison of the energy response of the L1 τ algorithm before (black) and after (red) the calibration.

3.4.4 Isolation

An efficient τ_{h} identification and a precise estimate of its energy are achieved by the previous steps of the algorithm. However, background from quark and gluon jets is also reconstructed and needs to be identified and rejected to reduce the trigger rate. As τ_{h} energy deposits are typically narrower than background ones, the application of an isolation criterion, that exploits the presence of calorimeter activity around the selected candidate, provides a good way to identify and reject the background.

The isolation energy $E_{\mathrm{T}}^{\mathrm{iso}}$ is computed as the difference between the energy sum of the TTs within a region extending 6×9 in $i\eta$ and $i\varphi$ around the τ_{h} candidate, and the uncalibrated energy $E_{\mathrm{T}}^{\mathrm{raw}}$ of the L1 candidate itself, i.e.

$$E_{\mathrm{T}}^{\mathrm{iso}} = E_{\mathrm{T}}^{6\times9} - E_{\mathrm{T}}^{\mathrm{raw}} \tag{3.2}$$

The isolation region considered is schematically represented in Figure 3.11. The uncalibrated $E_{\mathrm{T}}^{\mathrm{raw}}$ energy is used in the formula (3.2) to ensure that homogeneous quantities are subtracted and that $E_{\mathrm{T}}^{\mathrm{iso}}$ is not biased by the application of the calibration constants, and also because of latency constraints in the algorithm implementation: the calibrated energy is computed in parallel to the isolation sum, and there is no access to its value in the isolation criterion computation. The 6×9 region is centred around the τ_{h} main cluster depending on its original trimming direction: if the left part of the cluster was removed, the region extends by one additional $i\eta$ unit towards the right side of the cluster and vice versa, as represented in Figure 3.11. The specific choice of the isolation region size is a trade-off between an optimal estimation of $E_{\mathrm{T}}^{\mathrm{iso}}$ and hardware resources constraints, and allows for the sharing of the energy sum computation between the τ, e/γ, and jet algorithms as it will be discussed further in Section 3.6.

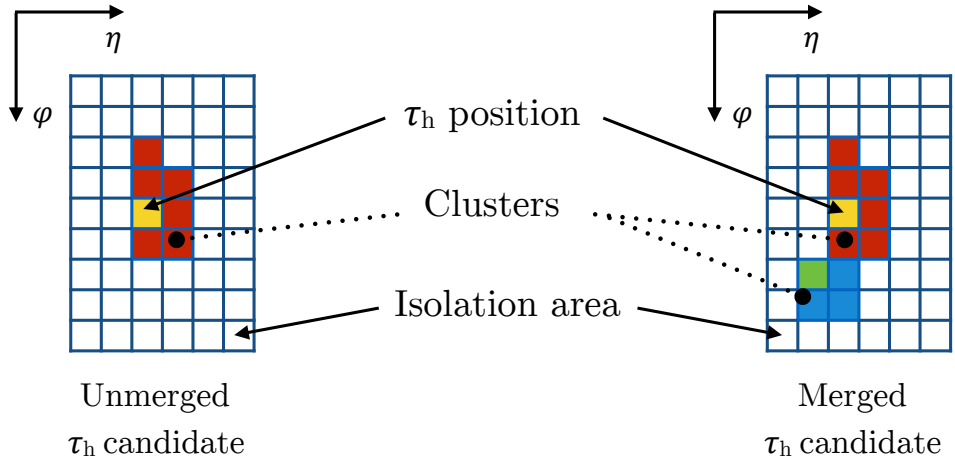


Figure 3.11 – Schematic representation of the isolation region considered for unmerged (left) and merged (right) L1 τ_{h} candidates.

A comparison of the signal and background separation that is achieved with the $E_{\mathrm{T}}^{\mathrm{iso}}$ variable is shown in Figure 3.12. The signal in the figure corresponds to simulated events of Higgs boson production via vector boson fusion and subsequent decay to a τ lepton pair. Background events are data collected with the CMS detector in high PU runs in 2016. These data are selected using a set of triggers that are synchronized on a specific proton bunch crossing, thus providing an unbiased sample of the calorimeter activity in proton-proton collisions (also referred to as "zero bias events"). The distribution of the PU is compatible between the simulation and these data to ensure a fair comparison of the $E_{\mathrm{T}}^{\mathrm{iso}}$ distribution.

The L1 τ_{h} candidate isolation is computed by comparing the isolation energy to a threshold value ξ, i.e. the candidate is isolated if $E_{\mathrm{T}}^{\mathrm{iso}} < \xi$. A comparison of the signal efficiency and background rejection that is achieved by varying the value of ξ is shown in Figure 3.13 for three separate intervals of the L1 τ_{h} candidates E_{T}. In the figure, and in the following ones, the background rejection is defined as $1-\varepsilon$, where ε is the background selection efficiency. A rejection of the background between 60% and 70% for a signal efficiency of about 80% can typically be achieved.

The definition of the isolation threshold ξ must ensure a uniform performance over the whole detector, E_{T} range considered, and range of number of PU interactions expected for the collisions: the value of ξ is consequently a function of the L1 τ_{h} candidate $i\eta$

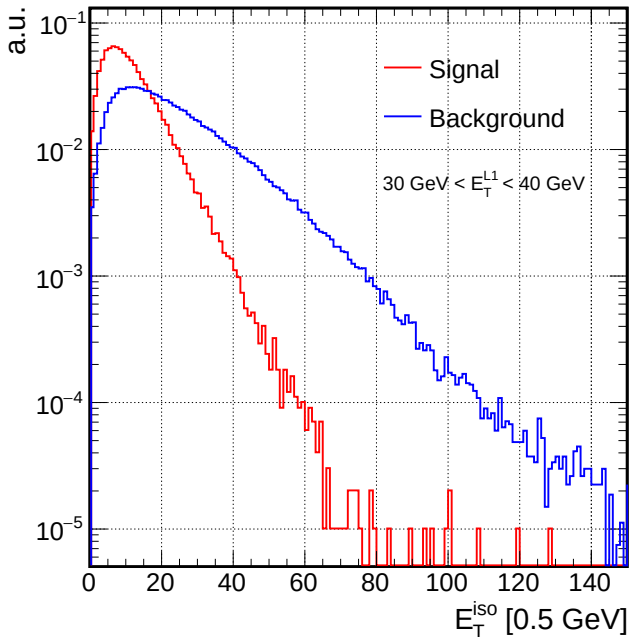


Figure 3.12 – Distribution of the isolation energy $E_{\mathrm{T}}^{\mathrm{iso}}$ for τ_{h} candidates (red) and jet background candidates (blue). For both distributions the L1 τ_{h} candidates are required to have an energy between 30 and 40 GeV. The signal corresponds to simulated events with Higgs boson produced via vector boson fusion and decaying to a τ lepton pair. The L1 τ_{h} candidates are required to be matched to a genuine τ_{h} in the simulation. Background is constituted of zero bias events collected in 2016 high PU special runs. The simulated PU conditions of the signal are the same as those measured for the background.

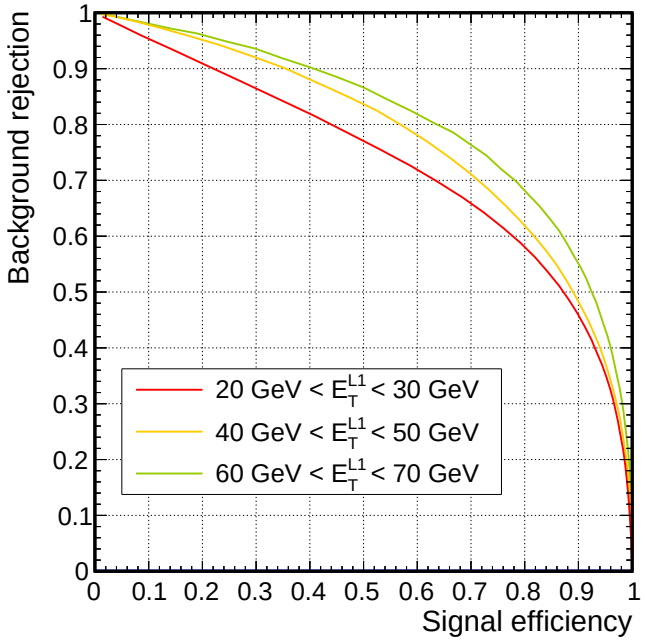
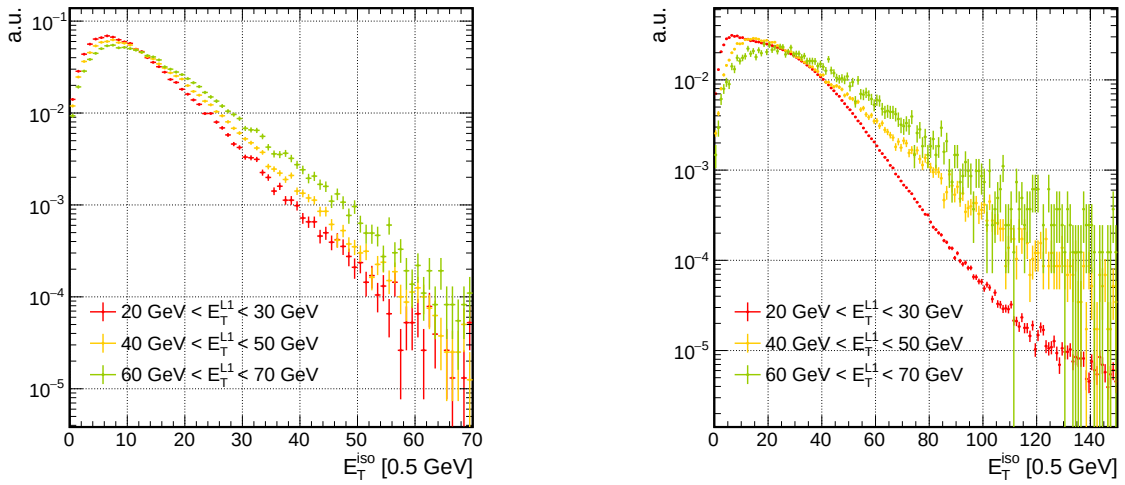


Figure 3.13 – Signal efficiency (simulated candidates from H $\rightarrow \tau\tau$ simulated events) versus background rejection (τ_{h} candidates from zero bias 2016 data) achieved by varying the upper threshold on $E_{\mathrm{T}}^{\mathrm{iso}}$.

position and energy, and of the number of PU interactions in the event. The dependence of $E_{\mathrm{T}}^{\mathrm{iso}}$ on $i\eta$ mainly arises from the different TT geometry in the barrel and endcap region of the detector, while the dependence on the E_{T} of the L1 τ_{h} arises from the imperfect collection of the energy deposited in the calorimeter by the clustering procedure. The distribution of $E_{\mathrm{T}}^{\mathrm{iso}}$ for different intervals of the L1 τ_{h} E_{T} is compared in Figure 3.14 for signal and background events, and in both cases a dependence of the distribution on E_{T} is observed. The variation is more evident for background candidates as the energy deposits are typically more spread and not fully grouped into a L1 cluster, but the same effect is visible also in the case of signal candidates and must be taken into account.

The number of PU interactions in the event is estimated by counting the number of active TTs (N_{TT}) in the event in the central region of the detector $|i\eta| \leq 4$, as schematically represented in Figure 3.15. Because eight "rings" in the φ direction are considered and each one contains 72 TTs, this number ranges from 0 to 576, making it little sensitive to the presence of a L1 τ_{h} candidate in the N_{TT} counting region given its very localized interaction in the detector. The N_{TT} variable thus defined provides a fast, robust, and reliable estimation of the PU in the event. This can be observed from its correlation with the number of primary vertices N_{vtx} in the event, which is itself proportional to the number of PU interactions. The dependence of N_{TT} on N_{vtx} is shown in Figure 3.16. The choice of the $i\eta$ region for the N_{TT} computation is again a trade-off between the statistical precision and the latency of the algorithm.



(a) Signal (b) Background

Figure 3.14 – Distribution of the isolation energy $E_{\mathrm{T}}^{\mathrm{iso}}$ in different intervals of τ_{h} candidates E_{T} for signal (a) and background (b). Signal candidates are obtained from simulated $\mathrm{H} \to \tau\tau$ events and background candidates are obtained from zero bias data in high PU 2016 collisions. All the distribution are normalized to a unitary area. Note that the horizontal scale of the two plots is different.

In conclusion, the threshold ξ depends on the L1 τ_{h} candidate energy and $i\eta$ position and on the estimator of the PU in the event, i.e.

$$\xi \equiv \xi(E_{\mathrm{T}}^{\mathrm{raw}}, i\eta, N_{\mathrm{TT}}) \tag{3.3}$$

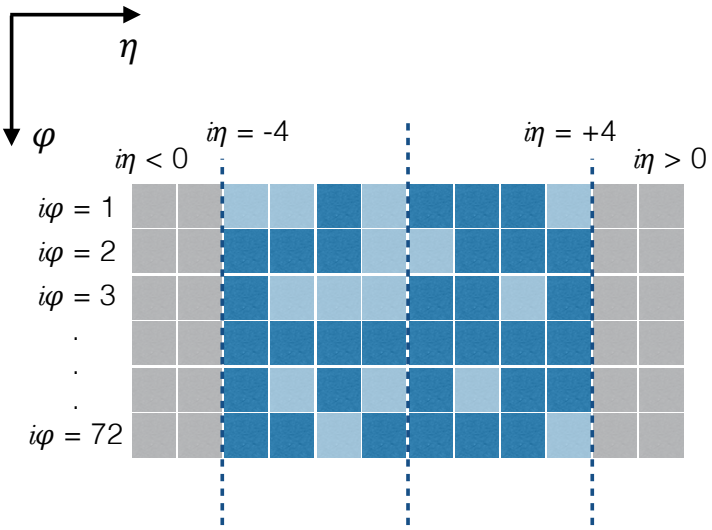


Figure 3.15 – Schematic representation of the PU estimation using the N_{TT} variable. Only the TTs in the region $|i\eta| \leq 4$ are considered and are highlighted with the color in the scheme. N_{TT} corresponds to the number of TTs in this region with $E_{\mathrm{T}} \geq 0.5\,\mathrm{GeV}$ that are represented in dark blue.

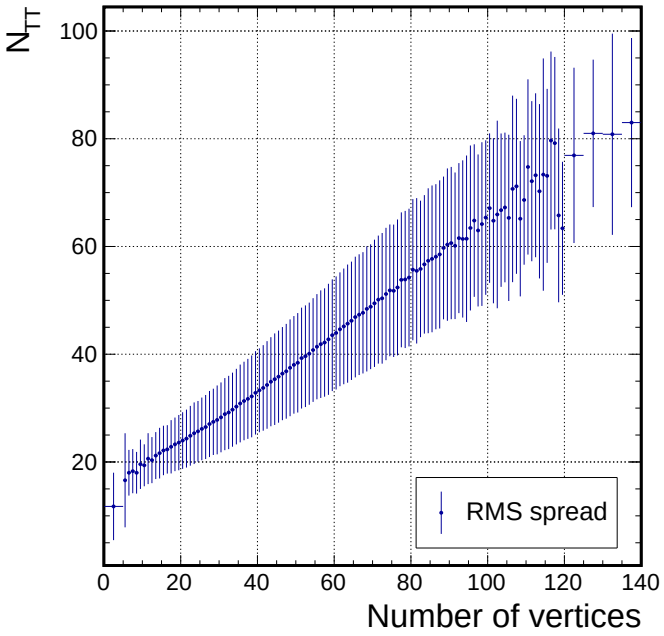


Figure 3.16 – Average of the N_{TT} distribution as a function of the number of reconstructed primary vertices. The vertical error bars denote the RMS spread of the distribution in each interval of the number or primary vertices considered. The high correlation between the two variables shows the effectiveness of N_{TT} as an event PU estimator.

The values of the ξ thresholds are encoded in a LUT as a function of its inputs, where the values of $E_{\mathrm{T}}^{\mathrm{raw}}$ and $i\eta$ are compressed with the same scheme used in the calibration. Similarly, the value of N_{TT}, expressed on 10 bits, is compressed with a dedicated LUT to 5 bits.

The isolation thresholds are derived in separate intervals of $E_{\mathrm{T}}^{\mathrm{raw}}$, $i\eta$, and N_{TT} using simulated events of gluon-gluon and vector boson fusion Higgs production and Drell-Yan production of τ lepton pairs. In any of these intervals, the value of ξ is chosen according to a target signal efficiency, that is constant as a function of $i\eta$ and N_{TT} but can be varied as a function of $E_{\mathrm{T}}^{\mathrm{raw}}$. A uniform efficiency over the entire p_{T} range has been used to assess the performance of the τ algorithm, as it is discussed futher in Section 3.5. However, given the steeply falling E_{T} spectrum of background events, as it can be observed in Figure 3.17 for zero bias events, larger background rejection can be achieved with a tighter isolation criterion at low E_{T} values, that is progressively relaxed until reaching a 100% efficiency. This approach was adopted for the 2016 data taking, where the signal efficiency ε is defined as:

$$\varepsilon = \begin{cases} \varepsilon_0 & \text{if } E_{\mathrm{T}}^{\mathrm{raw}} < A \\ \varepsilon_0 + \dfrac{1-\varepsilon_0}{B-A}\left(E_{\mathrm{T}}^{\mathrm{raw}} - A\right) & \text{if } E_{\mathrm{T}}^{\mathrm{raw}} \geq A \text{ and } E_{\mathrm{T}}^{\mathrm{raw}} < B \\ 1 & \text{if } E_{\mathrm{T}}^{\mathrm{raw}} \geq B \end{cases} \tag{3.4}$$

corresponding to a constant efficiency ε_0 below an energy threshold A, that is linearly increased up to 100% in correspondence of a second energy threshold B. The values of ε_0, A and B are optimized according to the expected PU and luminosity conditions. For most of the 2016 data taking (corresponding to the periods from B to G, as discussed in Section 2.1.2 of Chapter 2), where an instantaneous luminosity of $\mathcal{L} = 1.1 - 1.3 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ was expected from the LHC, ε_0, A and B have been set to 80%, 20 GeV and 50 GeV respectively. For the last part of the data taking (period H), where $\mathcal{L}$ was increased to $1.5 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the isolation criterion has been tightened by setting these parameters to 70%, 25 GeV and 70 GeV respectively.

3.4.5 Shape veto

The TT granularity of the trigger system, combined with the dynamic clustering technique, provides an additional information that can complement the isolation criterion and be used to reject background events: the shape of the L1 τ_{h} cluster. Larger TT aggregates are expected in the case of quark and gluon jets, while more compact ones are produced in τ_{h} decays given the more localized energy deposition. While broad jets typically lead to a significant energy deposition in the isolation region, the energy of smaller jets can be mostly collected into one cluster or two merged clusters and hence satisfy the isolation requirement, but result in a larger size of the cluster or in a less compact shape. The shape veto consists in the identification and rejection of such cluster topologies. In the following, only the shape of the main cluster is studied; the usage of the combined information of the shape of the main and secondary clusters for L1 τ_{h} candidates that undergo cluster merging represents a real challenge from the point of view of hardware resources and has not been explored at the moment.

A τ_{h} cluster, as described in Section 3.4.1, can regroup up to 7 TTs around the seed one, which position was schematically represented in Figure 3.4. Following the numbering

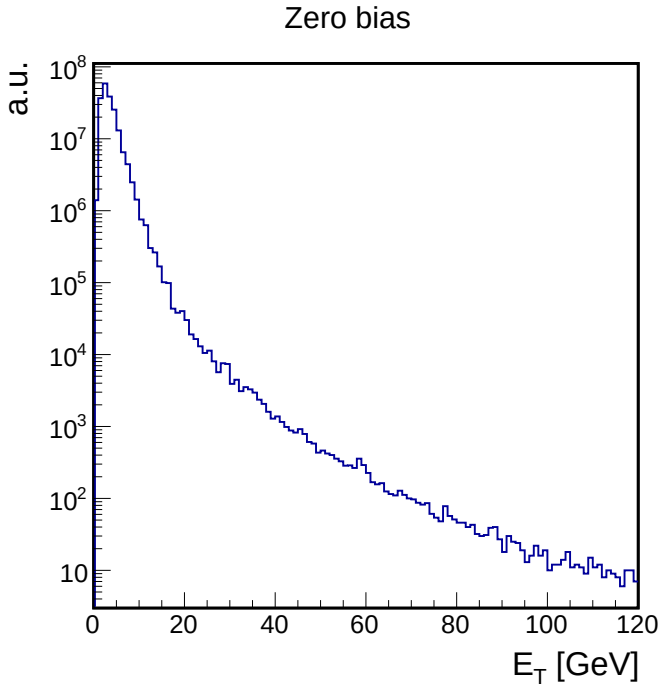


Figure 3.17 – Energy distribution of τ_h candidates in events collected with a zero bias trigger in high PU pp runs in 2016.

convention introduced in the scheme, the shape of the cluster is represented in a binary form as $S = b_6b_5b_4b_3b_2b_1b_0$, where b_i is 1 if the TT is included in the cluster and 0 otherwise. Because of the cluster lateral trimming, the same representation can be used for two shapes that differ only by a reflection with respect to the $i\varphi$ direction (i.e. reflection with respect to a vertical axis in the scheme): in the binary representation of S, this is achieved through the exchange of the indices $(2, 3, 4) \leftrightarrow (2', 3', 4')$. This representation gives $2^7 = 128$ different possible values or S, but not all of these can be produced by the algorithm. In particular, configurations with $(b_0 = 1 \wedge b_1 = 0)$ or $(b_6 = 1 \wedge b_5 = 0)$, corresponding to 56 different shapes, cannot be formed by construction, reducing the number of possible cluster shapes to 72. Finally, the symmetry of the shape in the $i\varphi$ direction (i.e. reflection with respect to an horizontal axis in the scheme) can also be exploited to further reduce the possible configurations of S to be studied. This symmetry is represented by the exchange $(b_6b_5b_4b_3b_2b_1b_0) \leftrightarrow (b_0b_1b_2b_3b_4b_5b_6)$, allowing to reduce further the number of independent shapes by 30 units. Accounting for the reflections with respect to both the $i\eta$ and $i\varphi$ directions, 42 independent shapes are identified. They constitute the variable that is used to separate background from signal events. The determination of this set of independent shapes is important to optimize the development of the shape veto and to avoid the introduction of a bias in the treatment of symmetric shapes, for which no physical difference in the production is expected. The 42 shapes are shown in Figure 3.18 together with their binary representation.

An example of the distribution of these shapes is shown in Figure 3.19 for the three decay modes of the τ lepton separately and for background events. As expected, large, non-compact clusters are more frequent in the case of background events, such as the shapes 99, 103, 111, and 127. The shape of the cluster also proves to be dependent on the τ_h decay mode; although not enough sensitive when considered alone to determine the

Figure 3.18 – Overview of the 42 symmetric shapes of a L1 τ_{h} cluster. Each square denotes a TT, and the yellow one denotes the seed of the cluster. The number below each shape represents its binary representation in the format discussed in the text and its conversion in decimal format is reported below between brackets for an easier comparison with the other figures.

decay mode, it shows an important correlation with it and, combined with other observables such as the fraction of the energy deposited in the ECAL and HCAL calorimeters and the merging information, could provide the capability to identify the τ lepton decay mode at the calorimeter hardware level.

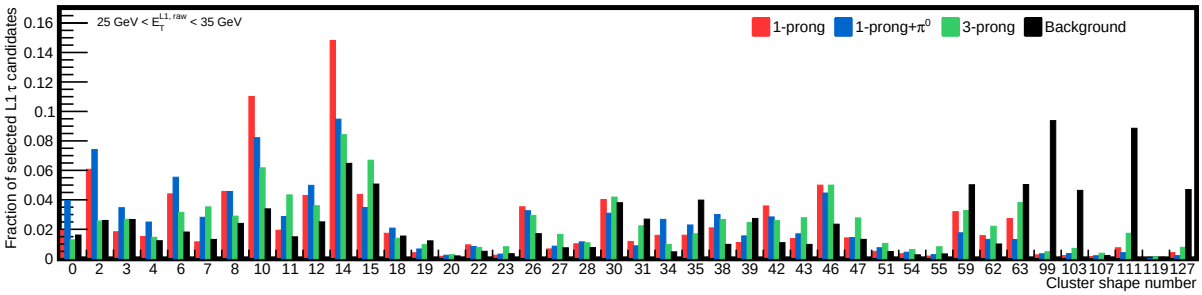


Figure 3.19 – Distribution of the selected shapes for the three reconstructed decay modes of the τ lepton (red, blue and green) and for background events (black). The statistical error in the plot is negligible for most cases and not shown for better readability.

The distribution of the shapes is however correlated with the isolation and p_{T} of the L1 τ_{h} candidates. For this reason, a shape-based discriminant is derived for isolated τ_{h} candidates in separate intervals of their $E_{\mathrm{T}}^{\mathrm{raw}}$ value. The selected cluster shapes are ordered according to their S/B ratio, where S and B denote the frequency of a specific shape for the selected signal and background events, respectively. In this way, cluster shapes with a value of S/B smaller than a specific threshold are considered as background-like and the corresponding L1 τ_{h} candidates rejected. The background rejection as a function of the signal efficiency in three separate intervals of $E_{\mathrm{T}}^{\mathrm{raw}}$ is illustrated in Figure 3.20. It can be observed that about 25% of the background can be rejected for a signal efficiency of 95%.

These considerations set the basis for a shape veto method to reduce the background contribution. As for the isolation criterion, its efficiency can be kept constant or modified to recover a 100% efficiency at high p_{T}; in the method detailed below, the shape veto is progressively relaxed as a function of the $E_{\mathrm{T}}^{\mathrm{raw}}$ value of the L1 τ_{h} candidate. Also, in this first study to assess the potentialities of the shape veto discriminant, no dependence on the cluster $i\eta$ position in considered, nor on its i^{merged} status, but their inclusion could significantly improve the performance of the method although requiring some optimization for the implementation in the firmware.

The reduction of a double τ_{h} trigger rate as a function of the E_{T} threshold applied on both L1 τ_{h} candidates is illustrated in Figure 3.21a. For typical trigger thresholds between 25 and 35 GeV, a reduction of the rate between 10 and 20% can be achieved, a factor that can be crucial to fit the rate of τ_{h} triggers in the values allowed by the CMS data acquisition system. Inversely, for the same trigger rate, the application of the shape veto makes it possible to reduce the trigger threshold by about 2 GeV. As illustrated in Figure 3.21b, the acceptance to signal processes is increased in the experimentally challenging low E_{T} region and the overall efficiency improved. The shape veto efficiency on the signal is not fully compensated by the lower trigger threshold only for offline τ_{h} candidates of $p_{\mathrm{T}} \approx 40$ GeV, where a small reduction of the acceptance of about 2% is observed. This effect is due to the relaxation of the shape veto only for candidates with $E_{\mathrm{T}}^{\mathrm{raw}} > 50$ GeV, but such value can be tuned and take part to a combined optimization of

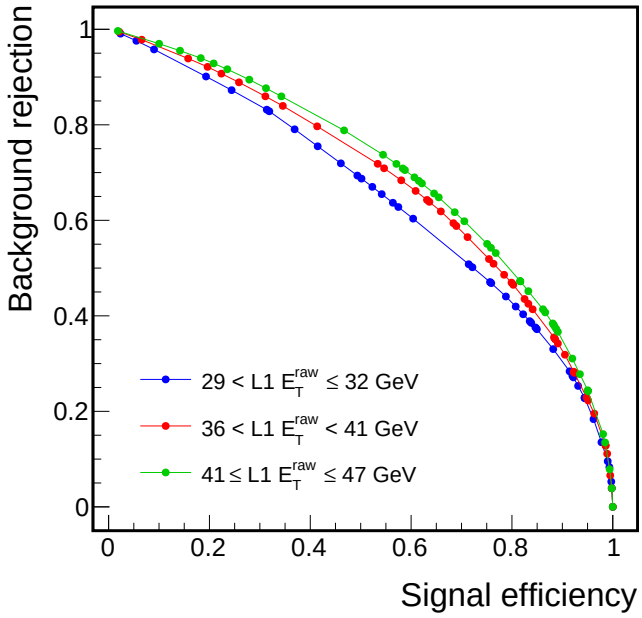
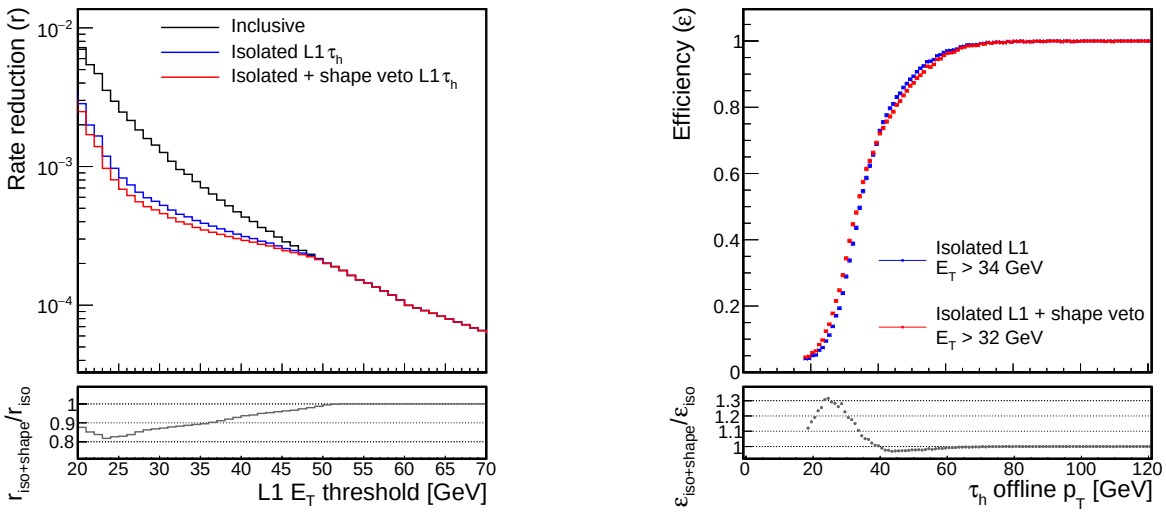


Figure 3.20 – Signal efficiency versus background rejection achieved with the usage of a cluster shape discriminant for three separate intervals of L1 $E_{\mathrm{T}}^{\mathrm{raw}}$.



(a) Rate reduction

(b) Efficiency

Figure 3.21 – (a): double τ_{h} trigger rate reduction as a function of the E_{T} threshold applied on both L1 τ_{h} candidates. The reduction is computed from an inclusive zero bias data sample as the fraction of events containing a two L1 τ_{h} candidates that satisfy the E_{T} selection (black) together with the isolation (blue) and the shape veto (red) criteria. (b): τ_{h} efficiency as a function of the offline τ_{h} reconstructed transverse momentum with the application of the isolation criterion (blue) and of both the isolation and shape veto criteria (red). The E_{T} thresholds applied are 34 and 32 GeV respectively and correspond to the same trigger rate.

both the isolation and shape veto criteria to improve the signal acceptance over the full p_{T} range.

This shape veto criterion requires a considerable amount of hardware resources because of the cluster representation and the need of a dedicated LUT implementing it, making it necessary to optimize its implementation to fulfil hardware requirements. For this reason it was not deployed for 2016 collisions where a stable and commissioned version of the firmware was needed to be rapidly produced. The shape veto discriminant has been included in the Layer-2 MP7 firmware for 2017 collisions by combining it with the cluster double counting LUT described in Section 3.4.2 and can now be optimized and used in the data taking.

3.4.6 Sorting and data transmission to μGT

Requiring the seed TT to have $|i\eta| \leq 28$ and to be a local maximum in a region of 3×9 in $i\eta$ and $i\varphi$ respectively implies that up to 144 τ_{h} candidates can be identified (one every 3 TT out of the $28 \times 2 = 56$ positions available in the η direction and one every 9 TT out of the 72 positions available in the φ direction). As this very large number of candidates cannot be transmitted to the μGT (and only a few τ_{h} candidates at most per event are expected to be relevant for the trigger decision) the 6 highest energy candidates in the positive and negative η regions of the detector are selected. The two detector regions are treated separately to minimize the latency required for the sorting operation: given the pipelined structure of the algorithm, the sorting operation in each half-detector can start as soon as the very first candidates in the central part of the detector are produced. A standard bitonic sorting algorithm, particularly suited for this operation given its parallel structure, is used. It should be noted that the sorting is performed simultaneously on all the τ_{h} candidates, irrespectively of their isolation status. It has been verified that this does not limit the trigger efficiency even in processes with a high hadronic activity such at $\mathrm{t\bar{t}}$ production. The 12 highest E_{T} candidates thus selected are transmitted to the μGT that will used them to perform the final event accept decision.

3.5 τ trigger performance on simulated events and comparison with Run I

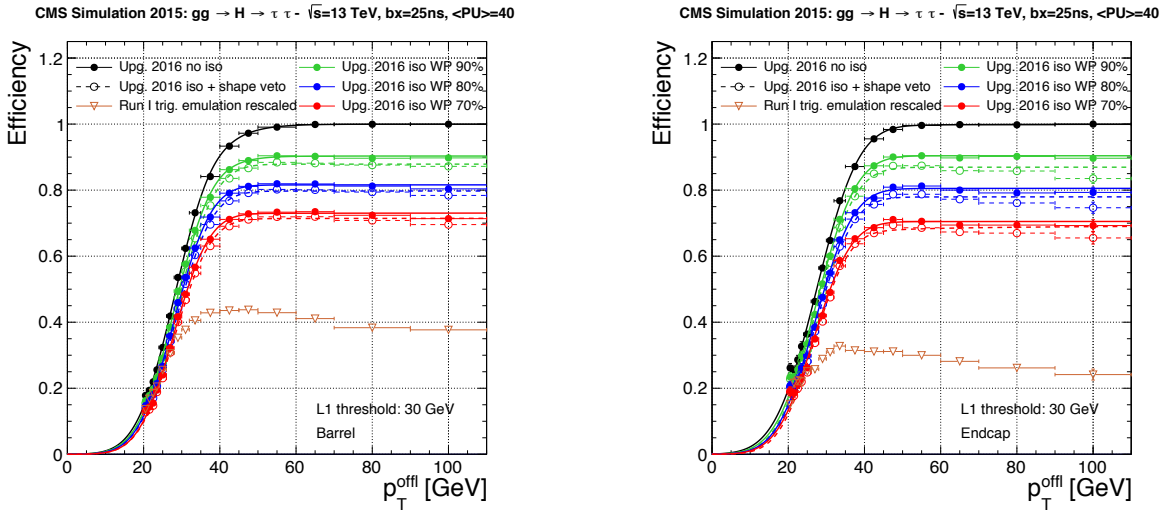
The performance of the upgrade τ algorithm has been compared to the one of Run I τ algorithm on simulated MC events. The purpose of this comparison, that was performed before the firmware integration of the calorimeter trigger algorithms described in Section 3.6, was to establish a reference in view of the subsequent changes that would have been required to fulfil implementation constraints, as well as to assess the physics potential of the new algorithm.

For these reasons, some differences exist between the algorithm version used in this comparison and the one described in the previous section. Their detailed description is summarized in [119, 120], and the major change consists in a different calibration procedure where the separate sums of the energy deposited in the ECAL and HCAL subdetectors are combined linearly. Minor differences concern the size of the isolation window (5×9) and the way secondary clusters are formed and selected. However, after the optimization of the algorithm, these changes had a small or negligible impact on the

performance, and the results shown below represent well the improvements in the upgrade algorithm with respect to the Run I.

The comparison has been performed using a simulated sample of $gg \to H \to \tau\tau$ for the signal and of zero bias events for the background. The expected Run II collisions conditions were simulated, namely a bunch spacing of 25 ns and an average number of PU interactions of 40.

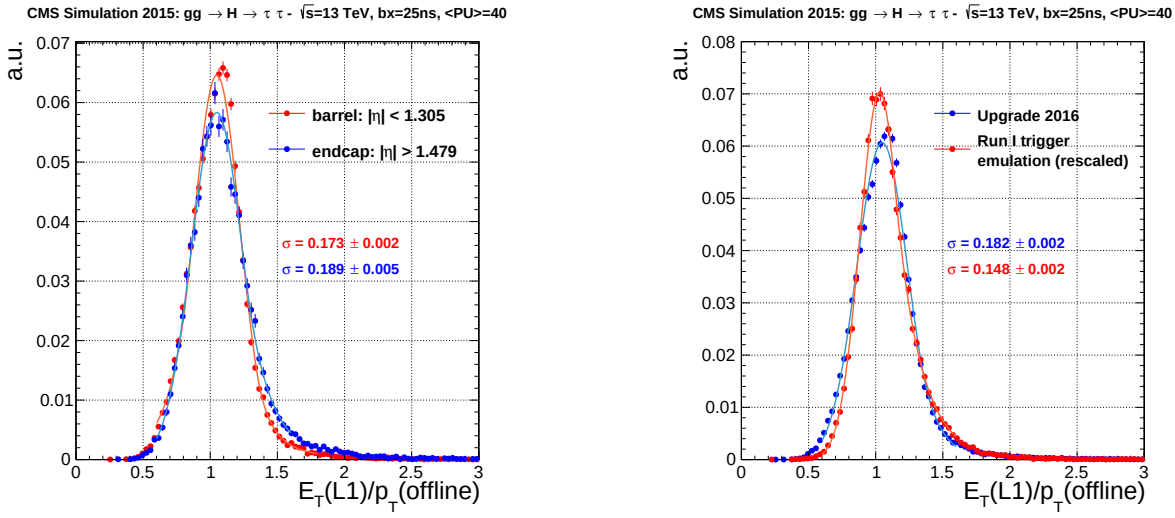
The comparison of the trigger efficiency as a function of the p_T of the τ_h reconstructed offline, computed as detailed later in Section 3.8.2, is shown in Figure 3.22 for the barrel and endcaps regions separately. This efficiency curve is usually referred to as "turn-on" because of its shape, where a plateau is present and is preceded by a rise which steepness is mainly determined by the response of the detector and the resolution achieved by the algorithm in collecting the deposited energy. While the Run I algorithm efficiency is limited to about 40% and decreases at high p_T, the upgrade algorithm shows a sharp rise, a 100% efficiency plateau in the absence of isolation, and stability of the performance after the application of the isolation criterion. The application of the cluster shape veto has little impact on the efficiency as expected. It should be noted that neither the isolation nor the shape veto criteria are relaxed at high p_T in this case.



(a) Barrel (b) Endcaps

Figure 3.22 – Trigger efficiency as a function of the p_T of the offline reconstructed τ_h. The barrel (a) and endcaps (b) regions are shown separately. The green, blue, and red curves denote the application of the isolation requirement with a signal efficiency of 90%, 80%, and 70% respectively, while in the black curve no isolation requirement is applied. The dashed curve denote the efficiency after the application of the shape veto. The efficiency of the Run I algorithm is shown in orange.

The energy response of the algorithms is shown in Figure 3.23. Similar performance is achieved in the barrel and endcap regions of the detector by the upgrade algorithm. Its response is very close to the one achieved by the Run I algorithm, but uses only about 10% of the TT to reconstruct the τ_h candidate. This highlights the excellent performance of the dynamic TT clustering, that significantly reduces the background contribution. The value of E_T in the the Run I algorithm has been scaled by a factor $k = 0.628$ for a better visual comparison of the two curves. This wrong global energy scale results from



(a) Upgrade, barrel vs endcap

(b) Run I vs upgrade algorithms

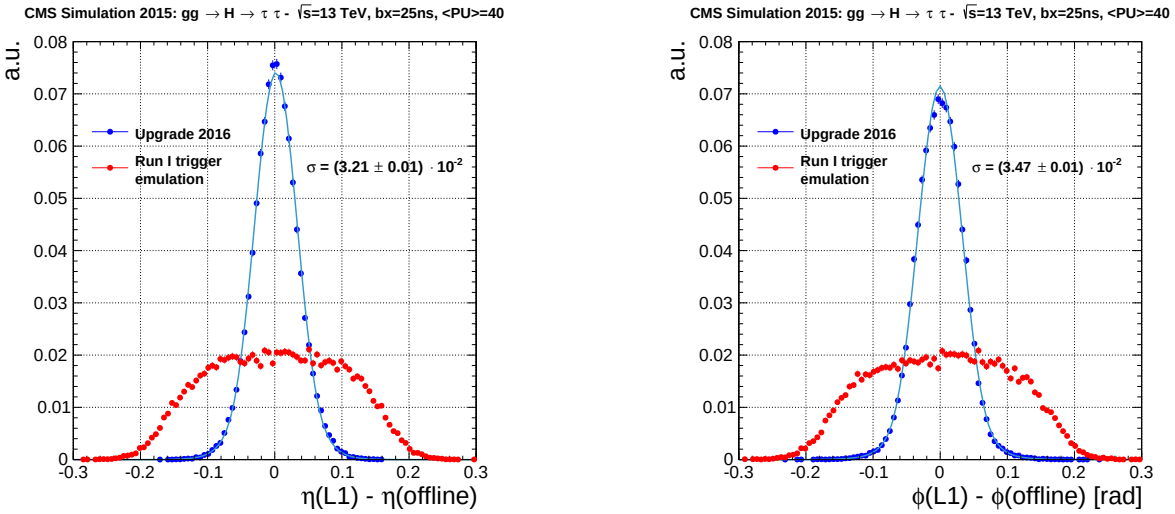
Figure 3.23 – (a) Comparison of the energy response of the upgrade τ algorithm in the barrel (red) and endcaps (blue) regions of the detector. (b) Comparison of the energy response for the Run I (red) and upgrade (blue) algorithms.

the application of the jet calibration, as L1 τ_{h} candidates were reconstructed from them and no dedicated τ_{h} calibration was put in place.

The η and φ position resolution of the Run I and upgrade τ algorithms are compared in Figure 3.24. The effect of the improved granularity of the latter is clearly visible, as the position resolution improves from the size of the calorimeter region to the size of the single TT.

The stability of the efficiency as a function of the number of PU interactions in the event is summarized in Figure 3.25. Variations within about 5% over the entire PU range are observed after the application of both the isolation and the shape veto criteria, making the algorithm suitable for the usage in the harsh running conditions foreseen for the LHC Run II.

Finally, the trigger rate as a function of the threshold applied on the L1 τ_{h} candidates E_{T} is shown in Figure 3.26a for a double τ_{h} trigger, defined by requiring the presence of two τ_{h} candidates in the event with a value of E_{T} larger than the threshold indicated on the horizontal axis and $|\eta| < 2.1$. As done for the energy response comparison, the E_{T} value for the Run I has been rescaled to account for its wrong energy scale. The horizontal dashed line denotes a rate of 3 kHz, that was considered as a typical target rate for such trigger. It can be observed that thresholds as low as 29 GeV can be achieved, and that a threshold of 40 GeV is sufficient to keep the rate under control even in presence of a loose isolation requirement. Although the rate achieved by the Run I algorithm may look smaller beyond a threshold of 30 GeV, its performance suffers from the very low signal efficiency. A comparison of these two quantities is shown in Figure 3.26b, where it can be observed that for the same background rejection (i.e. the same trigger rate), significantly larger signal efficiencies are achieved by the upgrade τ algorithm. Background rejection values of 0.995 or larger are illustrated in the figure for a better comparison of the Run I and upgrade algorithms, but the region experimentally accessible with a sustainable trigger rate corresponds to rejection values larger than 0.999. Rejection



(a) η (b) φ

Figure 3.24 – Resolution in the η (a) and φ (b) positions of the Run I (red) and upgrade (blue) τ algorithms.

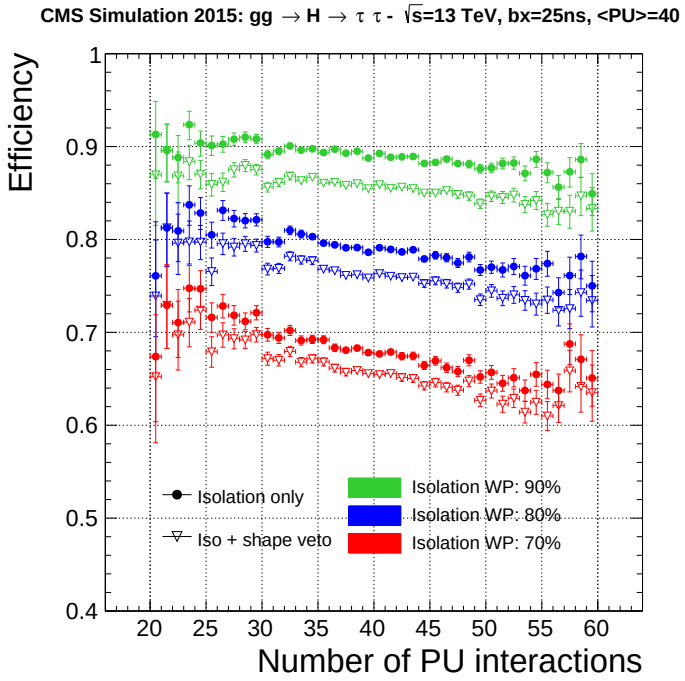
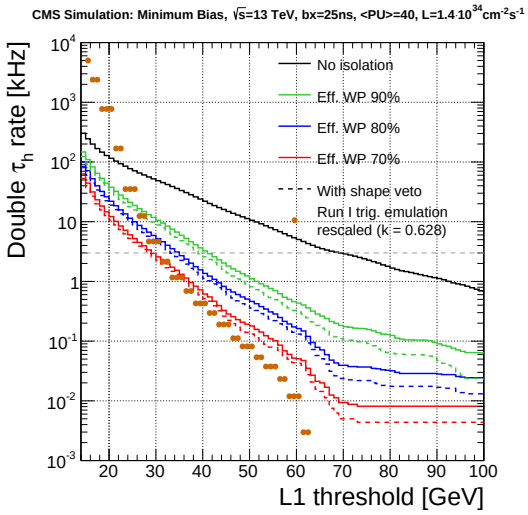


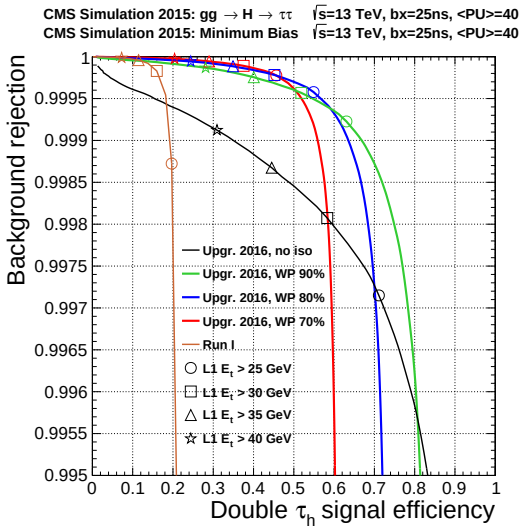
Figure 3.25 – τ_{h} efficiency reconstruction as a function the number of PU interactions in the event after the application of isolation requirements corresponding to a signal efficiency of 90% (green), 80% (blue) and 70% (green). Open markers denote the additional application of the cluster shape veto. A reduction of the efficiency of about 5% or smaller across the whole range of number of PU interactions is achieved.

values below 0.995 correspond to low E_{T} trigger threshold and, consequently, a vertical asymptotic behaviour is observed, as the signal efficiency is completely determined by the isolation working point.

These results demonstrate the improvements in τ_{h} identification that can be achieved



(a) Rate



(b) Signal efficiency vs bakground rejection

Figure 3.26 – (a) Double τ_{h} trigger rate as a function of the E_{T} threshold applied. Both τ_{h} candidates must have $|\eta| < 2.1$. (b) Signal efficiency versus background rejection for the double τ_{h} trigger. In both figures, the green, blue, and red curves denote the application of the isolation criterion with 90%, 80%, and 70% efficiency respectively, while no isolation criterion is applied for the black curve. The Run I algorithm performance is shown in orange.

with the trigger upgrade, showing that dedicate, highly efficient τ algorithm can be deployed at the hardware trigger level.

3.6 τ trigger firmware integration

The first developments and performance evaluation of the τ and e/γ, jet and energy sum algorithms were performed independently for each object. However, the design of the upgrade trigger system required all of them to be implemented into a single MP7 board. This represented a tremendous challenge to the firmware development for all the algorithms, that were designed to perform sophisticated operations and needed consequently to be simplified without loss in their performance. For the τ algorithm, a detailed description of the changes occurred with respect to previous algorithm version is documented in Refs. [121, 119]. Here, it is sufficient and important to underline the role that these constraints played in forging the algorithm as presented above.

The usage of hardware resources can be related to the volume and complexity of the data processed: more information need to be transmitted in the internal links of the FPGA, occupying more logic units. Instead, the latency of the algorithm depends on the amount of sequential operations that need to be performed. The optimization of the resources usage goes through a reduction in the quantity of information that is processed and in an efficient sharing of the same logic operations between the different algorithms. The minimization of the latency is instead achieved by performing these operations in parallel, as soon as a minimal amount of data has been received, without the need to wait for other operations to be concluded.

Firmware	Resources used	Algorithm fraction
τ	7.8%	19%
e/γ	3.1%	7%
Jets and sums	9.5%	23%
Common e/γ and τ	13.1%	31%
Common all algorithms	8.5%	20%
Total algorithms	42%	-
Core	23.6%	-

Table 3.1 – Usage of the hardware resources in the firmware implementation.

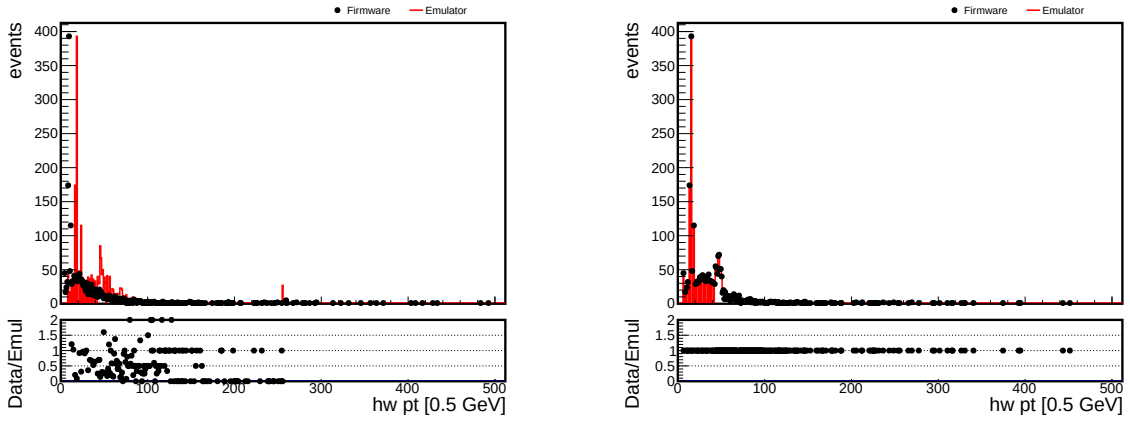
With these two general guidelines in mind, it can be observed how the very first operation of the τ algorithm, the clustering, has been shared with the e/γ and jet algorithms. The construction of a jet needs to be initiated by a local maximum of energy in a 9×9 region in $i\eta$ and $i\varphi$. This region is constructed in successive steps starting from 3×1 regions, that are subsequently combined into 3×3, 3×9, and 9×9 regions. These intermediate products are therefore reused for the determination of the main cluster (3×9) and secondary cluster (3×3) seeds, as well as for the determination of the energy deposited in the isolation region (6×9).

The calibration procedure also needed to be modified, as it was initially designed to combine the ECAL and HCAL energy sums independently, with different coefficients depending on the presence of energy deposits in the ECAL subdetector. This operation, too expensive in terms of resources, has been replaced by the calibration described in Section 3.4.3, where the i^{EM} flag (the presence of an energy deposit in the electromagnetic calorimeter) ensures that a similar performance is achieved. Similarly, given the large amount of resources needed to compute energy sums, the secondary clusters do not undergo the lateral trimming that is performed on the main one.

These changes, and many more that have not been detailed here, were carefully studied and optimized to have little or no impact to the performance of the algorithm, and made it possible to successfully implement it in the MP7 firmware. The amount of resources used in the firmware implementation is detailed in Table 3.1, where it can be observed that 42% of the resources are devoted to the algorithms and 23.6% to the core firmware (input and output links control, configuration and monitoring). The majority of the resources devoted to the algorithms are shared among them. These numbers are related to the amount of logic unities used by the implementation, and it should be noted that 70% represents approximately the maximum fraction that can be used: higher occupancies prevent a signal to propagate correctly into the FPGA (generating timing errors, i.e. desynchronization of the signals inside the FPGA) and do not make the algorithm function properly.

3.7 Commissioning with 2015 data

The firmware implementation needs to be validated to ensure that no differences exists with respect to the algorithm emulation in C++, and that the expected performance is achieved. The commissioning is performed in several phases, that begin with the test



(a) Before (b) After

Figure 3.27 – Comparison of the firmware and emulator before (a) and after (b) its commissioning using input patterns.

of the system in a controlled environment at the laboratory, where a single MP7 card is controlled and detailed monitoring and debug are possible, up to the final installation and integration in the racks of the CMS service cavern at the LHC.

Simulated data ("patterns") are injected in the MP7 card to record the output, as if data were received from the Layer-1 of the trigger system. These input patterns are also converted and processed through the C++ emulator to verify the consistency of the firmware implementation. In this way, any difference in the output between the emulator and the firmware can be identified and fixed. In parallel, a test bench, consisting of an emulation of the VHDL code performed on a calculator, is used to get access to intermediate quantities of the algorithm and not only to the MP7 output, allowing for more flexibility in the study of potential discrepancies. These steps have been performed on MP7 cards installed at LLR (France) and CERN (Switzerland), and in a test bench at Imperial College (UK), and allowed to identify and solve discrepancies due to the choice of the secondary cluster and removal of overlapping towers in the merging phase, isolation and calibration. The resolution of the discrepancies has been an intensive task, that required a detailed analysis of the algorithm in all its elementary operations. The firmware structure and its internal representation of the τ_{h} objects were thoroughly studied, comparing them with the C++ emulator where a different representation and algorithm logic are used. This challenging but important work resulted in the complete correspondence of the firmware and emulator algorithm descriptions. As an example, a comparison performed on the same input pattern of 1000 $\mathrm{H} \rightarrow \tau\tau$ events before and after the commissioning is presented on Figure 3.27 for the E_{T} distribution. The 100% agreement achieved after the commissioning can be observed, and the same agreement has been verified on the position and isolation of L1 τ_{h} candidates. It should be noted that modifications were also introduced in the C++ emulator and not only in the firmware, for example concerning the energy saturation of the candidates, as it can be observed in the figure.

This work proceeded in parallel with the installation of the MP7 boards in the service cavern of the CMS experiment. The links between the calorimeter trigger primitives generators, the Layer-1, the Layer-2, and the μGT were tested and, once this operation completed, the synchronization between these different systems was verified and adjusted.

The software of the data acquisition (DAQ) system of the experiment were also updated to ensure that the upgrade trigger information were correctly recorded and stored as a part of the event information. The readiness of the upgrade trigger system installation was achieved in fall 2015, when parallel runs of the upgrade system during the 2015 CMS data taking could be performed. These parallel runs consisted in the introduction of the upgrade system in the CMS experiment DAQ system. The output of the upgrade trigger system was not used to perform the final event accept or reject decision in the global trigger, but was stored as a part of the event information. The data collected in this way allowed us to properly configure the trigger system and its correct alignment and synchronization with the rest of the trigger electronics, and to measure its performance in a real collision environment. These operations ensured the readiness of the L1 trigger system for the 2016 data taking. The upgraded system was deployed as the baseline CMS trigger during the technical stop in winter 2016, and its configuration and synchronization once again verified during cosmic runs (i.e. in absence of beams circulating in the LHC) as well as "splashes", events where single bunches of protons are injected in the LHC and collide on the closed beam collimators 150 m upstream from the interaction point, producing a large multiplicity of particles that are measured in the CMS detector. At that point, the upgraded trigger system was ready for the 2016 data taking.

3.8 Deployment in 2016 data taking and performance

The successful commissioning of the τ algorithm was followed by its deployment for the 2016 data taking. L1 trigger seeds involving τ_{h} objects were defined to target several physics signal processes, including $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^{+}\tau^{-}$ events searched for in this thesis. The data collected in 2016 were used to measure the performance of the algorithm.

3.8.1 Main L1 τ seeds

The reconstructed L1 τ_{h}, e/γ, and jet objects, and TT energy sums computed are transmitted from the Layer-2 of the calorimeter trigger to the micro global trigger (μGT), that also receives muon candidates from the L1 muon trigger. The μGT combines them into "L1 seeds", i.e. conditions on the L1 candidates properties that are used to perform the final decision to accept or reject the event and that initiate the subsequent event reconstruction at the HLT. In case the trigger rate of a L1 seed is too high to be withstood by the data acquisition system of the experiment, this seed can be prescaled by a factor $f > 1$, meaning that its positive decision will be taken into account only one time every f occurrences.

The τ_{h} candidates reconstructed by the upgrade τ algorithm played a key role in the physics programme of the LHC in 2016 and were used in a variety of L1 seeds that can be classified in three main categories:

- Single object triggers: the μGT requires the presence of a single τ_{h} candidate satisfying certain position and momentum criteria
- Multiple object triggers: the μGT requires the presence of multiple τ_{h} candidates satisfying certain position and momentum criteria

- Cross triggers: the μGT requires the simultaneous presence of τ_{h} candidates and other objects such as a μ or an e/γ candidate

Single object triggers have a wide acceptance and can be used for different signals of interest, and their efficiency can be easily measured as no correlation with other objects needs to be taken into account; this makes them ideal candidates for precision physics such as the study of top quark properties, where events are typically selected with single electron or muon triggers. They result however in a high trigger rate and require high thresholds or prescale factors to be used in high luminosity conditions. Multiple object and cross triggers target instead specific signal topologies, and are used to improve the acceptance while maintaining a low trigger rate. For trigger objects with a high background contamination such as τ_{h}, these are the most efficient L1 seeds and can be used to select event where a τ pair is produced and decays to two τ_{h} jets and neutrinos (double τ_{h} trigger) or to a τ_{h} and an electron or muon with neutrinos (τ_{h} plus e or μ cross trigger).

The main L1 trigger seeds used for the 2016 data taking involving τ_{h} candidates are summarized in Table 3.2. The trigger rate is kept under control with the application of a $|\eta| < 2.1$ restriction on the τ_{h} position (higher levels of noise and background are typically observed in the forward regions) and by the application of the τ_{h} isolation criterion. Consequently, low trigger thresholds can be maintained, and ensure a high efficiency in the central part of the detector, where most of the target physics signals events are produced. The trigger rate reported in the table is measured in the data taking run 282092 (that took place from 02/10/2016, 13:13 to 03/10/2016, 04:05), that represents well the luminosity conditions experienced in the 2016 data taking. The trigger rate decreases with time because of the reduction of the proton beam intensities, as it can be observed in Figure 3.28 for the double τ_{h} trigger. The values quoted in the table correspond to the maximum rate at the beginning of the run, where a luminosity of $1.3 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ was measured. For comparison, the luminosity at the end of the run is $5.2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The trigger paths listed in Table 3.2 were used to collect a large fraction of the data exploited in physics analyses with τ leptons in the final state. In the case of the $\mathrm{HH} \to \mathrm{b\bar{b}}\tau^{+}\tau^{-}$ search presented in this thesis, the double τ_{h} trigger was used to collect the data analysed in the final state where both τ leptons decay to hadrons and a neutrino, that is the most sensitive to HH production as it will be detailed in Chapter 6. This search therefore largely benefits of the improvements in the τ_{h} identification at L1 that are achieved with the trigger upgrade.

3.8.2 Measurement of the performance with 2016 data

The τ trigger algorithm was deployed for the 2016 data taking and performed successfully during this period. The data collected in 2016 were used to measure the performance of the algorithm, an essential operation to verify that the expected performance was achieved and to monitor the proper functioning of the algorithm during the CMS operations. The results presented here are derived using the full dataset collected in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1}. The performance is measured with τ_{h} candidates from a sample of $\mathrm{Z} \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ events selected with a tag-and-probe technique.

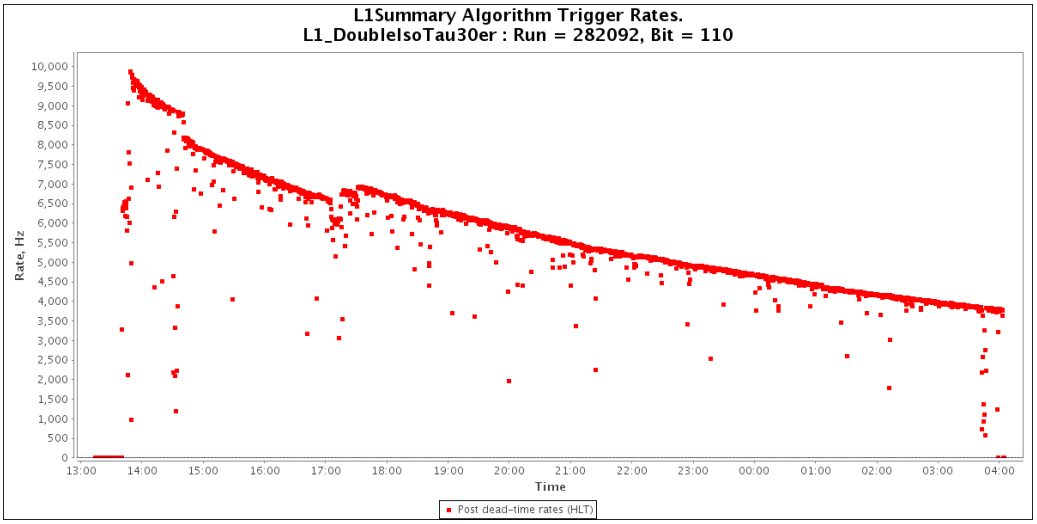


Figure 3.28 – Double τ_{h} trigger rate as a function of the time. The decrease in the trigger rate is due to the reduction of the proton beam intensities. Discontinuities in the curve are due to adjustments of the LHC machine parameters that are regularly performed.

L1 seed	Max. rate	Physics target
One τ_{h}, $p_{\mathrm{T}} > 120\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$	8.3 kHz	boosted $\mathrm{H} \to \tau\tau$
Two isolated τ_{h}, $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$	9.9 kHz	$\mathrm{H} \to \tau\tau \to \tau_{\mathrm{h}}\nu_\tau\tau_{\mathrm{h}}\nu_\tau$ including $\mathrm{HH} \to \mathrm{bb}\tau\tau$
One τ_{h}, $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$ One μ, $p_{\mathrm{T}} > 18\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$	3.3 kHz	$\mathrm{H} \to \tau\tau \to \tau_{\mathrm{h}}\nu_\tau\mu\nu_\tau\nu_\mu$ including $\mathrm{HH} \to \mathrm{bb}\tau\tau$
One τ_{h}, $p_{\mathrm{T}} > 26\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$ One isolated e, $p_{\mathrm{T}} > 22\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$ $\Delta\eta(\mathrm{e}, \tau_{\mathrm{h}}) > 0.2$	2.1 kHz	$\mathrm{H} \to \tau\tau \to \tau_{\mathrm{h}}\nu_\tau\mathrm{e}\nu_\tau\nu_{\mathrm{e}}$ including $\mathrm{HH} \to \mathrm{bb}\tau\tau$

Table 3.2 – Main L1 trigger seeds used for physics in 2016 data taking. The definition of the seed, the maximum trigger rate in the reference run, and the target physics signals are reported. Double τ_{h} trigger and cross-triggers target generically $\mathrm{H} \to \tau\tau$ decays and include the $\mathrm{HH}\mathrm{bb}\tau\tau$ signal investigated in this thesis. The rate values are taken from run 282092, and the maximum trigger rate is observed at the beginning of the run for an instantaneous luminosity of $1.3 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Tag-and-probe event selection and background subtraction

The $\mathrm{Z} \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_{\mathrm{h}}\nu_\tau$ process is particularly suited for the trigger efficiency measurement because of the clean signature provided by the muon, that is exploited for event trigger and selection, as well as for the kinematic properties of the $\mathrm{Z} \to \tau\tau$ decays that are used to identify the τ_{h} candidates. Events are selected with a tag-and-probe technique, requiring the presence of a muon (tag) satisfying identification, isolation and trigger requirements, in association with a τ_{h} candidate reconstructed offline (probe), with a kinematic that is compatible with the decay of a Z boson. As no trigger requirement is applied directly on the τ_{h} candidate, this selection provides an unbiased data sample

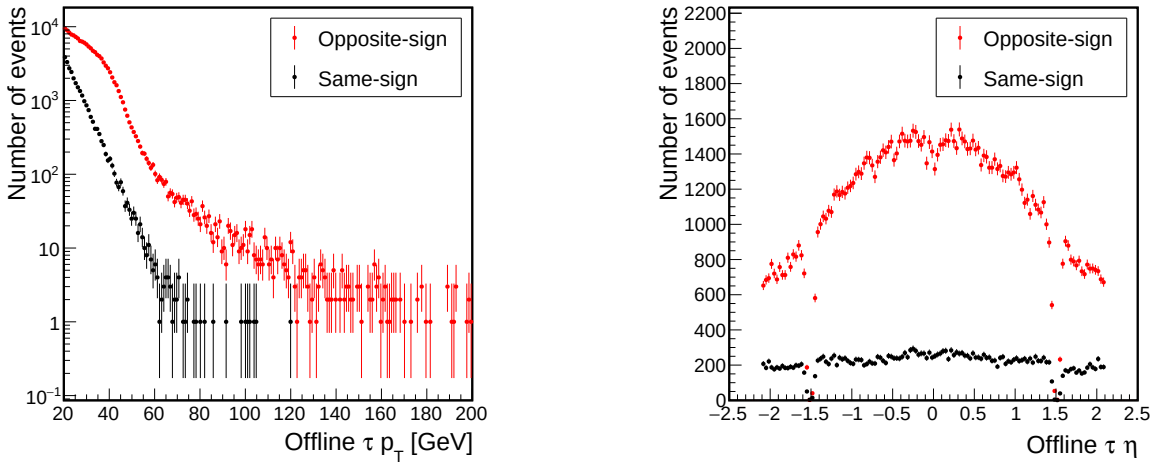
that is used to measure the L1 τ_{h} reconstruction efficiency.

Events are recorded with a single muon HLT trigger that requires the presence of an isolated muon of $p_{\mathrm{T}} > 22\,\mathrm{GeV}$ and $|\eta| < 2.1$. The HLT trigger muon object must correspond to a muon reconstructed offline with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$, $|\eta| < 2.1$, that satisfies loose identification criteria and the relative isolation criterion $\mathcal{I}_{\mu}^{\mathrm{rel}} < 0.15$; the definition of these criteria is detailed in Chapter 4. The specific values used for this measurement are chosen to ensure the quality of the reconstructed muon candidate with a reduced presence of background events, and the properties of the muon candidate do not enter in the computation of the trigger efficiency. These events must also contain a τ_{h} candidate reconstructed offline with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.1$ that satisfies the identification criterion, the tight working point of the multivariate isolation discriminant, the tight working point of the anti-muon discriminant, and the loose working point of the anti-electron discriminant, as described in Chapter 4. The invariant mass $m_{\mu\tau_{\mathrm{h}}}$ of the selected τ_{h} and μ candidates is required to be in the interval $40 < m_{\mu\tau_{\mathrm{h}}} < 80\,\mathrm{GeV}$, corresponding to the position of the peak in the distribution of this variable in $\mathrm{Z} \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ events. In case multiple τ_{h} candidates satisfy the previous requirements, the most isolated τ_{h} candidate is chosen. The spurious contribution from $\mathrm{Z} \to \mu\mu$ events is reduced by rejecting events that contain more than one muon with $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.4$, and $\mathcal{I}_{\mu}^{\mathrm{rel}} < 0.3$, while the contribution arising from $\mathrm{t\bar{t}}$ background is suppressed by rejecting the events with at least one jet with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$ that satisfies the medium working point of the algorithm that identifies jets produced from b quark hadronization (CSV b tagging algorithm). Finally, the contribution from W boson production in association with jets is reduced with the requirement on the transverse mass of the muon $m_{\mathrm{T}}^{\mu} < 30$ GeV, where the m_{T} variable is defined as:

$$m_{\mathrm{T}}^{\mu} = \sqrt{\left(p_{\mathrm{T}}^{\mu} + p_{\mathrm{T}}^{\mathrm{miss}}\right)^2 - \left(\vec{p}_{\mathrm{T}}^{\,\mu} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right)^2} \tag{3.5}$$

In this formula, $\vec{p}_{\mathrm{T}}^{\,\mu}$ denotes the transverse momentm vector of the muon and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ denotes the imbalance in the sum of the energy of the reconstructed particle-flow objects in the event, as detailed in Section 4.3.4 of Chapter 4.

These requirements have been optimized to maximize the purity of $\mathrm{Z} \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ signal in the event selection. However, a residual fraction of about 20% of the selected events is due to background processes where a quark or gluon jet is misidentified as a τ_{h} candidate, and originates mainly from QCD multijet and W boson production in association with jets. This background contribution is mostly relevant for τ_{h} with low p_{T} given the kinematic properties of such processes. The contribution of this background is estimated from the data using the relative sign of the electric charge of the selected μ and τ_{h} candidates. In signal events, the two candidates are expected to have opposite electric charge, while in background events same sign and opposite sign pairs are almost equally probable. Therefore, the number of background events is estimated as the number of events with a same sign (μ, τ_{h}) pair (SS) and is subtracted from the signal region, defined by the presence of an opposite sign (μ, τ_{h}) pair (OS). The distribution of events in these two regions is shown in Figure 3.29, where the larger relative importance of the background at low p_{T} values and in the endcap regions of the detector can be observed.



(a) p_{T} distribution

(b) η distribution

Figure 3.29 – Distribution of the p_{T} (a) and η (b) of the τ_{h} candidate reconstructed offline. The red and the black curve denotes the events with a muon and a τ_{h} pair of opposite and same electric charge respectively. The latter is used for the estimation of the background contribution. The regions with a low number of events at $|\eta| \approx 1.5$ correspond to the transition between the barrel and the endcap regions of the detector. Candidates in this region are rejected by the anti-electron discriminant, because the large amount of material increases the probability of erroneous τ_{h} identification.

Results

An offline τ_{h} candidate is considered as successfully reconstructed by the trigger algorithm if a L1 τ_{h} candidate is found within an angular distance $\Delta R < 0.5$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$. In case multiple L1 candidates are found, the one with the smallest value of ΔR is chosen, so that at most one L1 τ_{h} candidate is associated to an offline candidate. Additional requirements such as a minimal E_{T} value or the isolation criterion can also be applied on the L1 candidate in the study of the efficiency.

Upon evaluating the position resolution, the effect of the magnetic field that bends the trajectory of charged hadrons must be taken into account. As illustrated in Figure 3.30, τ_{h} objects impact the ECAL internal surface at a φ position that is systematically different from the one of emission at the production vertex. The bias depends on the p_{T} of the τ_{h} and on its electric charge. Consequently, the reconstructed τ_{h} positions at the entrance of the ECAL subdetector ($\eta_{\mathrm{calo}}^{\mathrm{offline}}$ and $\varphi_{\mathrm{calo}}^{\mathrm{offline}}$) are compared with the L1 positions to account for this effect. In case the τ_{h} is composed of multiple PF candidates, its $\eta_{\mathrm{calo}}^{\mathrm{offline}}$ and $\varphi_{\mathrm{calo}}^{\mathrm{offline}}$ coordinates are computed as the average of the $\varphi_{\mathrm{calo}}^{\mathrm{offline}}$ coordinates of its components, weighted by their energy. The position resolution measured in the data is shown in Figure 3.31. A full width at half maximum resolution of about 0.08 rad in η and 0.1 rad in φ is observed. The worse resolution on the φ coordinate is explained by the bending of the charged particles in the magnetic field, that causes a spread of the energy deposits in this direction and reduces the precision to which their energy barycentre can be estimated as the position of the highest energy TT. This can also be observed by separately comparing the position resolution for the three reconstructed decay modes of the selected τ leptons,

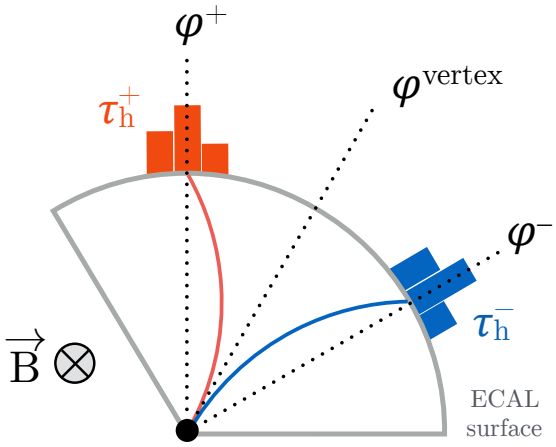


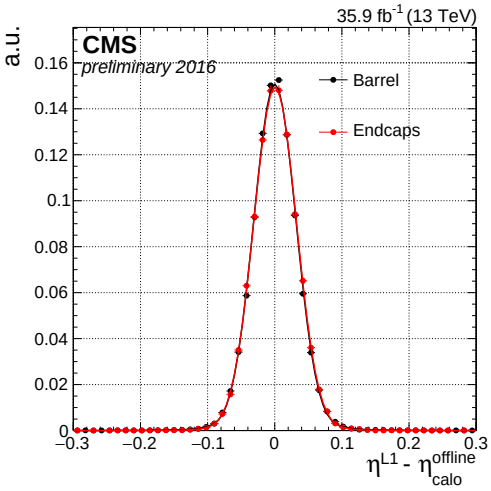
Figure 3.30 – Schematic illustration of the bias in the τ_{h} φ position between a calorimetric measurement and the offline reconstruction. The former corresponds to the position of the τ_{h} cluster at the ECAL internal surface, while the latter corresponds to the angle of emission at the production vertex ($\varphi^{\mathrm{vertex}}$). Without accounting for this effect, an opposite systematic shift in the position is observed for positive (φ^{+}) and negative (φ^{-}) τ_{h} candidates.

as shown in Figure 3.32. While the resolution on the η coordinate is similar for the three decay modes, the resolution in the φ position is worse for multi-particles τ decays, and in particular for 3-prong final states where the charged hadron spread due to the magnetic field is larger. Finally, the small negative bias of about 0.02 rad in the φ position that is observed is due to the numerical conversion of the hardware position $i\eta$ in the μGT coordinate representation to a physical φ value. It is therefore an effect external to the L1 trigger and introduced by the software that interprets the L1 trigger internal object representation (usually referred to as "unpacker"), that is being corrected for the 2017 data taking.

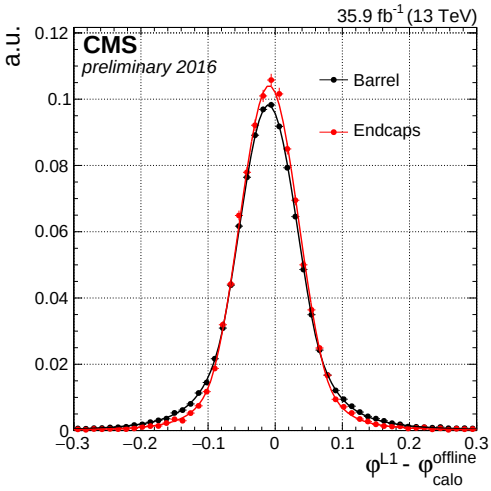
The energy response of the algorithm is shown in Figure 3.33a for the barrel and endcaps regions separately. In Figure 3.33b, the response measured inclusively in the barrel and endcap regions is compared to the one observed in a simulated sample of $\mathrm{Z} \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ events, selected with the same requirements as the data. The figure shows that the simulation reproduces well the data, and only a small difference is observed in the two cases. The evolution of the energy resolution as a function of the reconstructed p_{T} of the τ_{h} is shown in Figure 3.34. The improvement at higher energies is due to the variation of the calorimeter response as a function of the deposited energy, as discussed in Section 2.2.2 of Chapter 2. The stability of the energy resolution versus the τ_{h} η position in the detector is illustrated in Figure 3.35a, and the resilience against the pileup contribution is demonstrated in Figure 3.35b, where the number of reconstructed vertices is used as an estimator of the number of pileup collisions in the event. These results confirm that the dynamic TT clustering correctly individuates the energy deposits and rejects soft and diffuse contribution from pileup, and demonstrate that the target performance of the τ algorithm is met in Run II collisions.

The τ algorithm reconstruction efficiency ε is computed with the formula:

$$\varepsilon = \frac{N_{\mathrm{OS}}^{\mathrm{pass}} - N_{\mathrm{SS}}^{\mathrm{pass}}}{N_{\mathrm{OS}}^{\mathrm{tot}} - N_{\mathrm{SS}}^{\mathrm{tot}}} \tag{3.6}$$

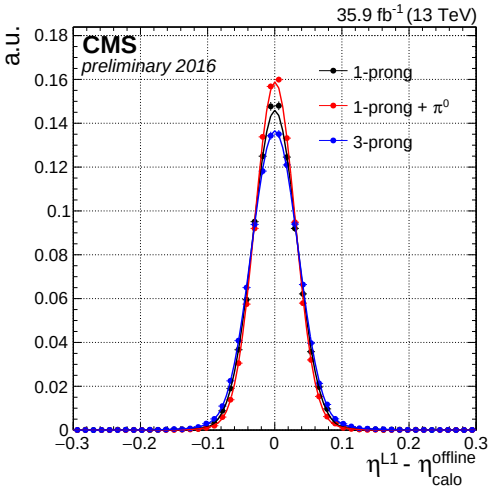


(a) η resolution

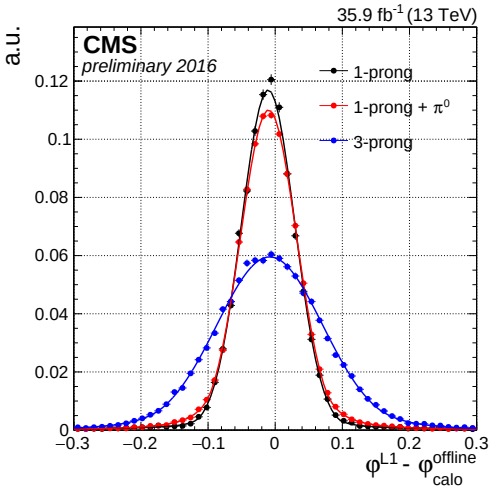


(b) φ resolution

Figure 3.31 – L1 τ algorithm resolution on the η (a) and φ (b) positions, computed inclusively for all the selected events. The resolution is shown separately for the barrel (black) and endcaps (red) detector regions. The offline τ_{h} candidate position is computed at the entrance of the ECAL subdetector to remove any bias from the magnetic field, as detailed in the text. Data are interpolated with a symmetric, two-sided Crystal Ball function.

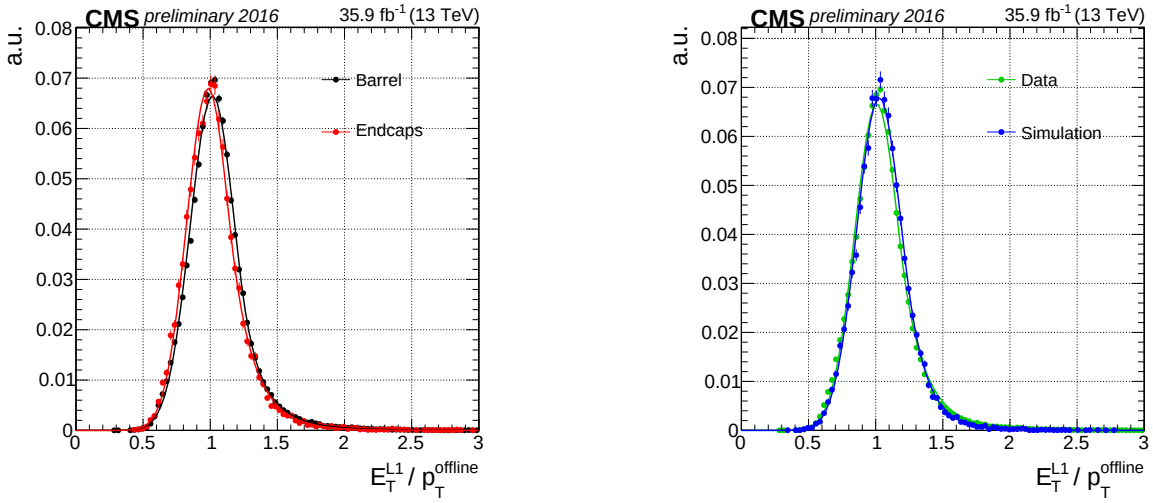


(a) η resolution



(b) φ resolution

Figure 3.32 – L1 τ algorithm resolution on the η (a) and φ (b) position, separately shown for 1-prong (black), 1-prong + π^0 (red) and 3-prong (blue) decays. The offline τ_{h} candidate position is computed at the entrance of the ECAL subdetector to remove any bias from the magnetic field, as detailed in the text. Resolution curves are computed inclusively for all the selected events, and data are interpolated with a symmetric, two-sided Crystal Ball function.



(a) Data: barrel and endcaps (b) Data and simulation

Figure 3.33 – (a) Energy response of the τ algorithm measured with the data collected in 2016. The barrel (black) and endcaps (red) regions are shown separately. Good uniformity of the response is observed in these two regions of the detector. (b) Comparison of the energy response in data (green) and in a sample of simulated $Z \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ events (blue) selected with the same requirements. The response in both figures is shown inclusively for all the selected τ_h candidates with $E_T > 20\,GeV$. Data are interpolated with a one-sided Crystal Ball function.

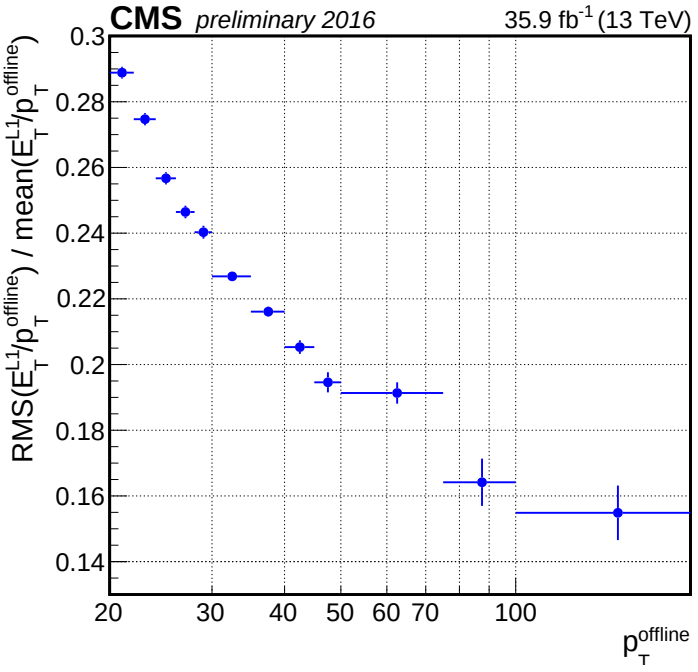


Figure 3.34 – Energy resolution as a function of the p_T of the τ_h reconstructed offline. The improvement of the resolution at high E_T is expected due to the improvement of the calorimeter response at higher energies.

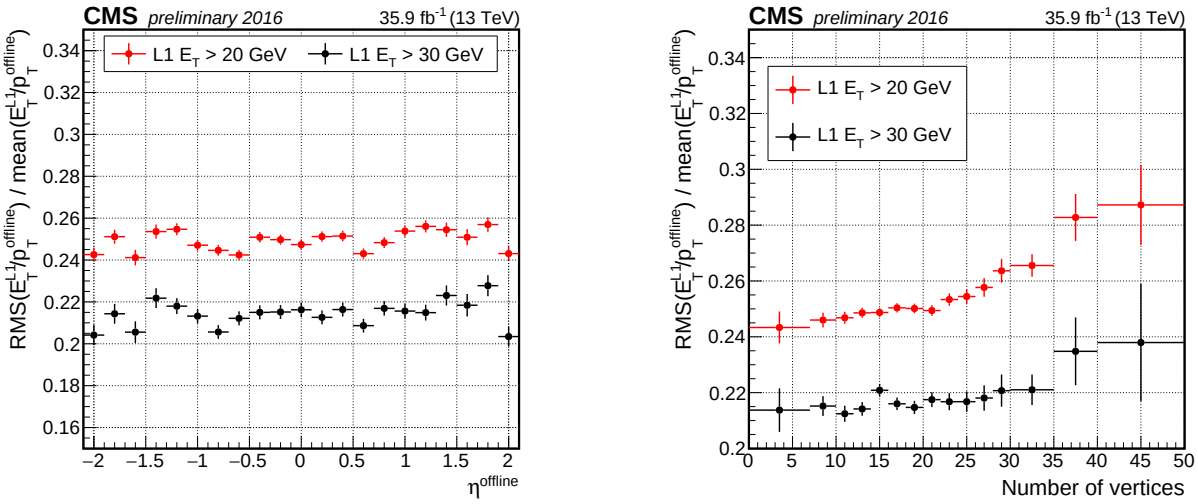
Here N_{OS} and N_{SS} are the number of candidates in the OS and SS regions respectively and the "tot" and "pass" superscripts denote the total number of events selected and those reconstructed by the L1 τ algorithm, respectively. These numbers can be computed in separate intervals of reconstructed offline quantities to study the dependence of the efficiency on such quantities.

Efficiency as a function of the position in the detector and of the number of reconstructed vertices in the event are shown in Figure 3.36. Stability at the level of 5% or better over the whole detector and the range of pileup interactions experienced in 2016 collisions are observed, both in absence and presence of the isolation criterion. Resilience against pileup effects is observed even when the isolation criterion is applied. The efficiencies are shown separately for the data taking periods corresponding to runs B to G and run H, because of the changes in the relaxation of the isolation criterion as a function of E_{T} detailed in Section 3.4.4.

The efficiency as a function of the offline τ_{h} p_{T} is shown in Figure 3.37 for different thresholds on the L1 E_{T} for all the τ_{h} reconstructed candidates and the isolated ones. The excellent response resolution results in a sharp rise of the efficiency, that reaches a flat plateau of 100% also in presence of the isolation criterion, as a consequence of the relaxation at high E_{T}. Uniformity between the efficiency in the barrel and endcap regions of the detector is illustrated in Figure 3.38a where very similar performance can be observed also with the application of the isolation criterion. It is also interesting to observe the different response of the algorithm to the three reconstructed τ decay modes, as shown in Figure 3.38b. The comparison of the efficiency for 1-prong and 1-prong+π^0 decays shows the better resolution in the latter case, because of the better reconstruction of photon energy in the ECAL subdetector. Differences in the scale and resolution for 3-prong decays are also observed and arise from the fact that no dedicated calibration can be performed on these specific topologies, because limited discrimination can be achieved at the calorimeter level in absence of tracking information. While this could partially be improved with the usage of additional inputs to the calibration such as the shape of the cluster, serious limitations might come from hardware constraints.

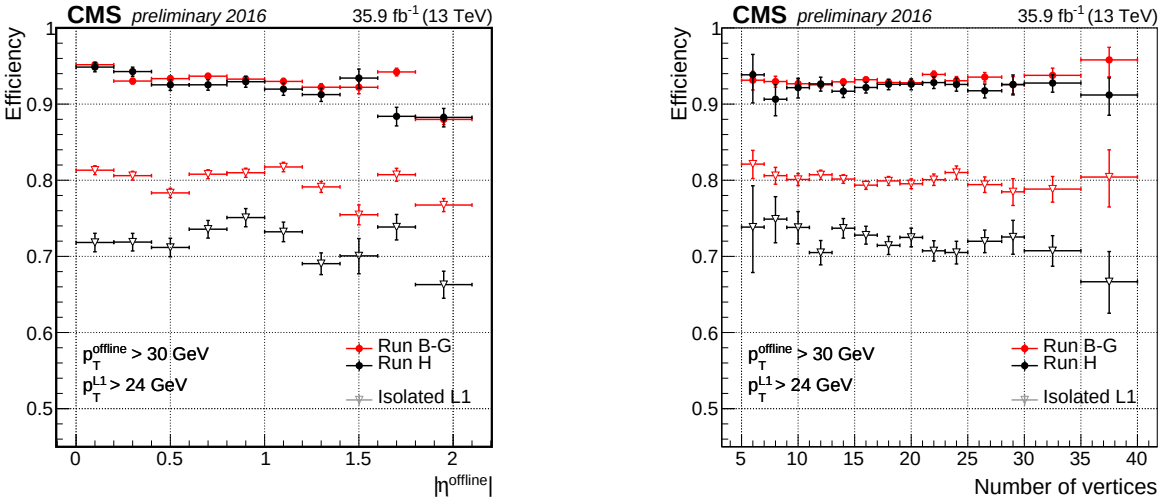
It should be noted that the isolation criterion shown in the previous figure only represents an average of the efficiency achieved in 2016. The separate efficiency in 2016 data taking periods from B to G and the period H are shown in Figure 3.39a. Owing to the tighter isolation criterion in the latter case, the same E_{T} threshold of 30 GeV could be maintained for the entire 2016 data taking. Finally, a comparison between the measured efficiency in data and the one expected from simulated $\mathrm{Z} \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_{\mathrm{h}}\nu_\tau$ events is shown in Figure 3.39b, where the good agreement can be observed. Small differences between the data and simulation efficiencies can be ascribed to the small differences in the energy response previously detailed, as well as to the differences in the simulation of collision conditions and particle interaction in the detector, that are usually corrected for at the data analysis level in terms of multiplicative factors in the object reconstruction efficiency. They do not arise from differences in the implementation of the algorithms in the firmware and its simulation, as excellent agreement has been observed by simulating the algorithm response and comparing it to the firmware output as described in Section 3.7.

These results summarize the excellent τ trigger performance achieved thanks to the upgrade of the L1 trigger system. High efficiency and uniormity of L1 τ_{h} identification are measured in data, showing that the algorithm design works as expected in real collisions and is well reproduced in the simulation.



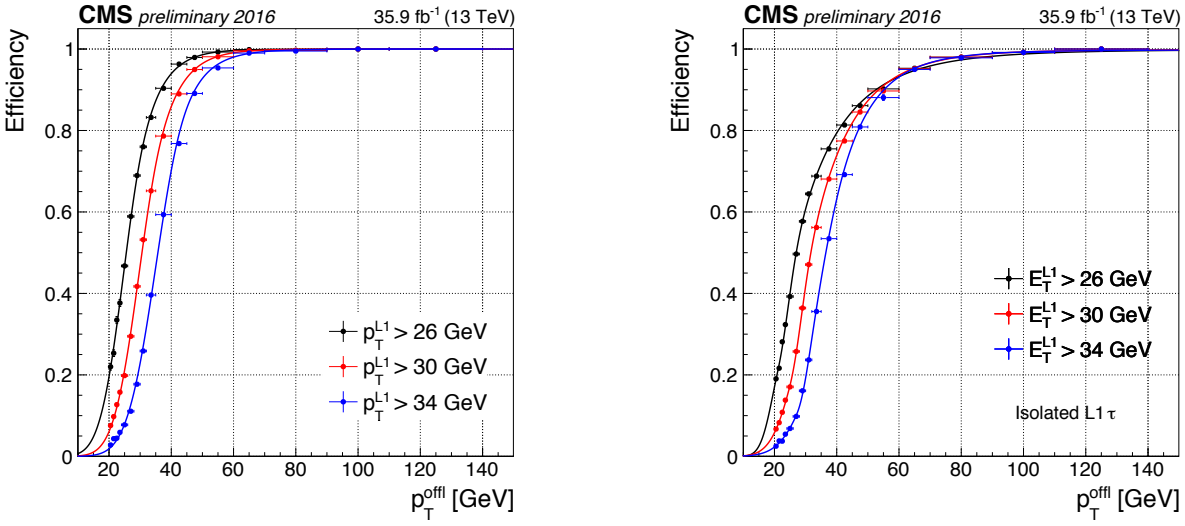
(a) η position (b) Number of vertices

Figure 3.35 – Energy resolution as a function of the η position (a) and of the number of vertices in the event (b). The red and black curve denote the events with and L1 τ_{h} E_{T} larger than 20 and 30 GeV, respectively. Good uniformity over the entire detector and resilience against pileup contributions are observed.



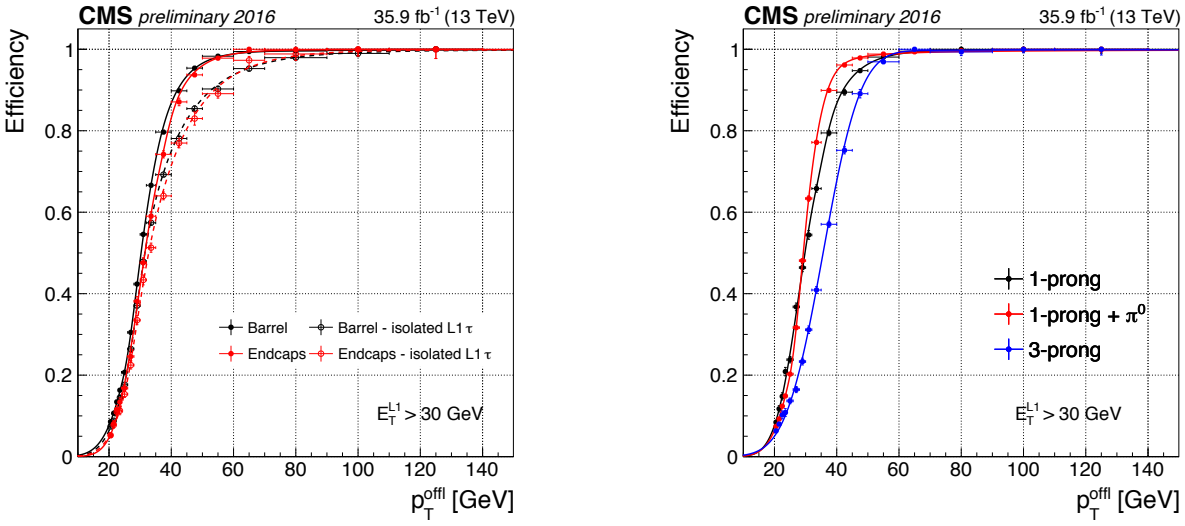
(a) η position (b) Number of vertices

Figure 3.36 – Reconstruction efficiency of τ_{h} candidates as a function of the η position (a) and of the number of vertices in the event (b). Full and open marker denote the absence and presence of the isolation requirement, respectively. The data taking periods corresponding to runs B to G (red) and H (black) are shown separately because of the changes occurred in the isolation criterion.



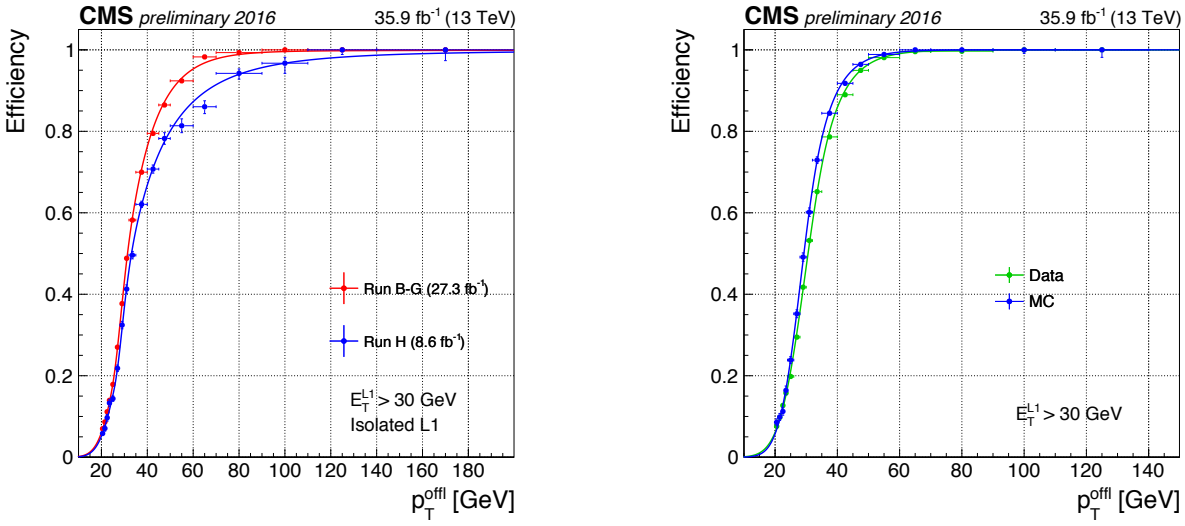
(a) (b)

Figure 3.37 – Efficiency as a function of the p_{T} of the reconstructed offline τ_{h} candidate for three different L1 E_{T} thresholds for all the reconstructed τ_{h} candidates (a) and the isolated ones (b). The L1 E_{T} threshold values correspond to those typically used for an isolated double τ_{h} trigger in 2016 data taking. Data in (a) are interpolated with the cumulative distribution of a Crystal Ball function. The same function, convolved with the linear relaxation function of the isolation, is used to interpolate the data in (b).



(a) (b)

Figure 3.38 – (a) Comparison of the efficiency for the barrel (black) and endcaps (red) regions of the detector, for all the reconstructed L1 τ_{h} candidates (solid line) and the isolated ones (dashed line). (b) Comparison of the inclusive efficiency for the three reconstructed τ lepton decay modes. The different steepness and position of the efficiency rise are related to the differences in the resolution and scale to the three decay modes, that cannot be fully distinguished at the L1 trigger because of the absence of tracking information. Data are interpolated with the cumulative distribution of a Crystal Ball function.



(a) (b)

Figure 3.39 – (a) Comparison of the efficiency for isolated L1 candidates in the data taking period corresponding to Run B to G (red), that took place from May to September 2016, and Run H (blue) that took place from September to October 2016. The isolation criterion in Run H has been tightened to cope with the increase in the instantaneous luminosity conditions of the LHC. (b) Comparison of the efficiency measured in data inclusively for the B–H periods (green) and in simulated MC events (blue) selected with the same criteria. The small difference between the measurement and the simulation can be ascribed to small discrepancies in the energy response, in turn due to the expected differences in the simulation of particle interaction and calorimeter response. Data in are interpolated with the cumulative distribution of a Crystal Ball function, convolved in (a) with the linear relaxation function of the isolation requirement.

Chapter 4

Event selection and categorization

The exploration of Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel requires the experimental capability to identify and reconstruct several different types of final state objects and to use them for the selection of signal-like events. This in turn requires the reconstruction of the $H \to \tau^+\tau^-$ and $H \to b\bar{b}$ decays and the usage of their properties to identify the specific signature of signal events and to reject background ones. These selections and techniques globally constitute the "analysis strategy", the backbone of the $HH \to b\bar{b}\tau^+\tau^-$ search to which this chapter is devoted.

These methods were developed to analyse pp data collected with the CMS detector at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016. It is important to recall that the amount of integrated luminosity analysed evolved quickly in these two years and that the analysis strategy was subsequently adapted to profit of the increasing amount of events. The LHC 2015 operations resulted in about 3.8 fb^{-1} of data recorded by the CMS experiment, but problems with the cryogenic system of the superconducting magnet affected part of the data taking operations and about 2.7 fb^{-1} of good quality data were recorded with the magnet on. These data were promptly explored and we documented the searches for HH production in the $b\bar{b}\tau^+\tau^-$ decay channel in two Physics Analyses Summaries (PASes), devoted to the nonresonant [122] and resonant [123] production modes, respectively. CMS-specific details are further described in the internal analysis notes (ANs) of Refs. [124, 125]. The 2015 dataset allowed us to probe HH production in the previously unexplored 13 TeV energy regime, and to verify the validity of the analysis strategy that we developed before the LHC restart. After this first phase, 2016 represented the first year of the LHC Run II mainly devoted to physics production. The International Conference on High Energy Physics held in summer 2016 (ICHEP 2016) represented an important moment to gather the result of the physics exploration at 13 TeV using the first part of this dataset, corresponding to an integrated luminosity of 12.9 fb^{-1}. The HH analysis strategy evolved in parallel with the increasing size of the dataset analysed and we introduced some important ameliorations. The results, documented in PASes [126, 127] with supporting ANs [128, 129], were presented at the conference. Finally, the totality of the data collected in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1}, was analysed between the end of 2016 and the beginning of 2017. The analysis strategy was further improved and optimized to enhance the sensitivity to both SM and BSM HH production. These results, documented in PAS [130] and supported by AN [131], were the highlight of the summary of HH searches that I presented on behalf of the ATLAS and CMS Collaborations at the Rencontres de Moriond 2017 conference [132]. I edited the corresponding paper [133] that was submitted to the Physics Letters B journal in time for the European Physical Society Conference on High Energy Physics held in 2017 (EPS-HEP 2017).

In this thesis I will focus on these latest result with the purpose of clarity and coherence in the description of the analysis strategy, but the reader should be aware that the earlier ones were crucial towards the understanding of the $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel and the optimization of the analysis strategy. Earlier results are summarized in the Appendix A of this thesis.

The first part of this chapter introduces the experimental challenges of the $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel. The discussion focuses on the experimental signature of the signal processes under study and on the main background sources. These considerations provide the bases for the development of the analysis strategy, that is covered in the rest of the chapter. The discussion is structured in four main parts, corresponding to trigger requirements, object preselections, event categorization, and definition of the signal regions. Each part relies on the previous one and improves the discrimination from background processes with an increasing complexity of the selections applied. Trigger requirements represent the first level of event selection and are necessary to store events offline for subsequent analysis. They are followed by preselections on final state objects, i.e. specific requirements on the quality of standard object reconstruction and identification that are optimized for the signal processes under study. Objects satisfying these quality criteria are combined into a $\mathrm{H} \to \tau^+\tau^-$ and a $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate, which properties are used to classify the events depending on their $\tau^+\tau^-$ final state and $\mathrm{b\bar{b}}$ topology. This event categorization makes it possible to improve the sensitivity to specific experimental signatures. Finally, the separation from the background in these event categories is enhanced with dedicated selections that exploit the invariant mass of the $\tau^+\tau^-$ and $\mathrm{b\bar{b}}$ pairs and their kinematic properties with a multivariate method. The combined efficiency of these four elements of the analysis strategy is summarized and discussed at the end of the chapter.

4.1 The $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel

The τ leptons are unstable and can undergo a decay to a lighter lepton or hadrons in association to neutrinos. As a consequence, the exploration of HH production in the $\mathrm{b\bar{b}}\tau^+\tau^-$ decay channel requires the study of several final states. In the following, the notation τ_ℓ ($\ell = \mathrm{e}, \mu$) is used to denote the decay of a τ lepton to lighter charged leptons and two neutrinos, i.e. the $\tau \to \ell\nu_\ell\nu_\tau$ decay, and the symbol τ_h previously introduced denotes the decay of a τ lepton to final state hadrons and a ν_τ neutrino. The indication of the quark and lepton charges, whenever not ambiguous, will also be omitted for simplicity of notation. The decay of a $\tau\tau$ pair can result in six final states, summarized in Table 4.1 together with their respective branching fractions. These values have been computed from the τ lepton decays branching fractions reported in Ref. [134].

Decay mode	$\mathcal{B}$
$\tau_\mu\tau_\mathrm{h}$	22.5%
$\tau_\mathrm{e}\tau_\mathrm{h}$	23.1%
$\tau_\mathrm{h}\tau_\mathrm{h}$	42.0%
$\tau_\mu\tau_\mathrm{e}$	6.2%
$\tau_\mu\tau_\mu$	3.0%
$\tau_\mathrm{e}\tau_\mathrm{e}$	3.2%

Table 4.1 – Branching fractions $\mathcal{B}$ of the six $\tau\tau$ decay final states.

The search that is described in this thesis explores the $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ final states, that together represent about 88% of the $\tau\tau$ decays. Little sensitivity is expected in the fully leptonic final states given their small branching fraction and the large contamination from Drell-Yan electron and muon pair production (for $\tau_\mu\tau_\mu$ and $\tau_e\tau_e$) and $t\bar{t}$ (for $\tau_\mu\tau_\mu$, $\tau_e\tau_e$, and $\tau_\mu\tau_e$) backgrounds, and they are therefore not considered.

The search aims at exploring both resonant and nonresonant HH production mechanisms: a high sensitivity to these different signal types is important to probe the nature of the electroweak symmetry breaking and of possible extensions of the SM as described in Section 1.3 of Chapter 1. Nonresonant HH production is considered both in the context of the SM and in the framework of an effective Lagrangian parametrization. The signal is characterized from the anomalous Higgs boson couplings predicted in this context, and its modelling is further discussed in Section 5.2 of Chapter 5. Resonant $gg \to X \to HH$ production is explored for mass resonances m_X ranging from 250 to 900 GeV under both the spin-0 and spin-2 hypotheses. The upper limit is chosen as the typical range for which a separate reconstruction of two τ decays is known to be efficient. For higher values of m_X, the τ decay products tend to overlap and dedicated reconstruction of such "boosted τ" topologies must be used, as it is documented in the Run I search for high mass resonant HH production [78].

The analysis techniques described in this chapter have been designed to be sensitive to a large variety of different signal processes, and the event selection strategy has been consequently structured in two main parts. The first one is a preselection that aims at correctly reconstructing and identifying the decay products of the Higgs bosons pair with the best acceptance for all the different signal hypotheses. It is followed by a second step where the sensitivity to specific signal topologies is maximized using dedicated selections and event categorization.

Trigger algorithms requiring the presence of e, μ, or τ_h candidates in the event are used to record collision events as further discussed in Section 4.2. In the preselection step, baseline criteria are applied on the reconstructed leptons to ensure their good identification and to reconstruct the $H \to \tau\tau$ decay. The presence of two jets in the event is subsequently required to reconstruct the $H \to bb$ decay. These preselections on the physics objects are detailed in Sections 4.3 and 4.4, together with the corrections applied to ensure their correct modelling in the Monte Carlo (MC) simulation.

Events are subsequently categorized in separate topologies depending on the spatial overlap of the jets from the selected $H \to bb$ candidates, improving the sensitivity to high m_X resonant production where substantial overlap is expected. A b tagging discriminant is used to select jets originating from b quark hadronization, and an event categorization based on the number of b-tagged jets is applied given the limited discriminant efficiency. The invariant mass of the jet and lepton pairs is exploited to further reject background events not compatible with the decay hypothesis of a Higgs boson pair decay. Finally, multivariate methods are developed in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states to further suppress the residual backgrounds.

Multiple sources of background affect the $bb\tau\tau$ search, and can be qualitatively classified as "irreducible" and "reducible" processes. Irreducible backgrounds result in a final state with the same particle content as the $bb\tau\tau$ signal, and have a large probability to be preselected. The main irreducible backgrounds are $t\bar{t} \to b\bar{b}W^\pm W^\mp \to b\bar{b}\ell^\pm\overset{(-)}{\nu}_\ell(\nu_\tau)\tau^\mp\nu_\tau$ ($\ell = e, \mu, \tau$) production and Drell-Yan production of a τ pair in association to a b quark pair, with minor contributions from the rarer ZZ and ZH processes. The $t\bar{t}$ contamination

is particularly important in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states, as both the prompt lepton production ($W \to \ell\nu_\ell$) and the production mediated from τ decays ($W \to \tau\nu_\tau \to \mu\nu_\mu\nu_\tau$) are possible, in contrast with the $\tau_h\tau_h$ final state. Reducible backgrounds instead arise from erroneous object identification, the most common sources being the misidentification of a jet initiated by a quark or gluon as a τ_h, or the erroneous identification of a light flavour jet as a b jet. Although the misidentification probabilities for these objects are usually small, the cross section of the background processes is many order of magnitude larger than the one expected for the signal, and they consequently result in a sizeable contribution. The Drell-Yan production of a τ pair in association to a light flavour quark pair, $t\bar{t}$ production with the decay of a W boson to quarks, and QCD multijet production are examples of such reducible processes. Additional minor reducible background are W boson production in association with jets, and single top quark production in association to a vector boson.

Different strategies must be put in place to reject these two classes of background. In the case of reducible backgrounds, tight object quality criteria are applied to reduce the misidentification rate, but at the price of a reduced signal efficiency. The best requirements are determined as a trade off between these two effects, and an example of this optimization procedure is discussed in Section 4.3.3 for the choice of the τ_h isolation criterion. Event categorization according to the quality of the reconstructed objects is another solution, that is adopted for the b tagging number categorization previously discussed. Irreducible background sources can be instead suppressed statistically only by exploiting their kinematic differences with respect to the signal processes. The invariant mass requirements and the multivariate methods, that are discussed in Section 4.6, have been developed for this purpose.

The estimation of the reducible and irreducible background processes is performed with a combination of MC simulation and data-driven methods. The comparison between these predictions and the data that are shown in this chapter are realized with the background estimation methods that are discussed in Chapter 5.

4.2 Trigger requirements

The data events recorded offline for this analysis are selected with a set of HLT triggers that require the presence of muon, electron, or τ_h candidates in the event depending on the $\tau\tau$ final state considered. In the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states, triggers requiring the presence of a single lepton in the events are used. In the $\tau_h\tau_h$ final state, a trigger path requiring the presence of two τ_h candidates is used. This trigger path is seeded by L1 τ_h candidates reconstructed with the trigger algorithm that has been described in detail in Chapter 3. The objects requirements corresponding to these trigger paths are reported in Table 4.2.

In the $\tau_\mu\tau_h$ final state, the logical "or" of two trigger paths with a different definition of the isolation is used. Muon reconstruction at HLT is initiated by the corresponding L1 μ candidates and combines the hits in the muon systems and, subsequently, those from the tracking subdetectors. A calorimetric isolation criterion is computed using the information from the ECAL and HCAL subdetectors, and a tracker isolation selection is based on the tracks reconstructed at HLT around the μ candidate. With the progressive increase of the instantaneous luminosity of the collisions, an additional requirement on the position of the μ of $|\eta| < 2.1$ was applied to reduce the trigger rate. This is accounted

for in the subsequent data analysis by applying the same muon position criterion in the preselections, as discussed later.

In the $\tau_e\tau_h$ final state the presence of an electron, reconstructed with an approach similar to the offline electron reconstruction, is required. An isolation criterion is computed as the scalar p_T sum of the PF clusters and tracks reconstructed at the HLT within a distance $\Delta R < 0.3$ around the electron candidate.

Finally, in the $\tau_h\tau_h$ final state two HLT τ_h candidates are required at trigger level. These are built from the PF candidates reconstructed within an angular distance $\Delta R = 0.8$ around PF jets. Up to three charged hadrons are combined with the π^0 candidates within a signal cone that has an aperture ΔR between 0.08 and 0.12 depending on the p_T of the object. This approach is similar to the offline HPS algorithm used for τ_h reconstruction, but timing constraints do not allow for the computation of all the PF candidates combinations and, consequently, for the reconstruction the τ decay mode, increasing the background contamination. The isolation is computed from the reconstructed tracks within a distance $\Delta R = 0.4$. In the data taking period ranging from B to G, tracks participating to the isolation sum were required to have at least 5 hits in the tracking subdetectors, of which 3 in the pixel system. For the data period H, further background rejection has been achieved by loosening the quality of the tracks down to 3 hits in the tracking system (with no constraint on pixel hits). An additional criterion on the scalar p_T sum of neutral candidates reconstructed at HLT in the isolation cone has also been applied to reduce the trigger rate (combined isolation).

$\tau_\mu\tau_h$ final state		
One μ , $p_T > 22\,\text{GeV}$, Calo. isolation One μ , $p_T > 22\,\text{GeV}$, Tracker isolation	run < 274954	2.8 fb^{-1}
One μ , $p_T > 22\,\text{GeV}$, $\lvert\eta\rvert < 2.1$, Calo. isolation One μ , $p_T > 22\,\text{GeV}$, $\lvert\eta\rvert < 2.1$, Tracker isolation	run ≥ 274954	33.1 fb^{-1}

$\tau_e\tau_h$ final state		
One e , $p_T > 25\,\text{GeV}$, $\lvert\eta\rvert < 2.1$, Tight isolation	All runs	35.9 fb^{-1}

$\tau_h\tau_h$ final state		
Two τ_h, $p_T > 35\,\text{GeV}$, $\lvert\eta\rvert < 2.1$, Medium iso.	Period B to G	27.3 fb^{-1}
Two τ_h, $p_T > 35\,\text{GeV}$, $\lvert\eta\rvert < 2.1$, Medium comb. iso.	Period H	8.6 fb^{-1}

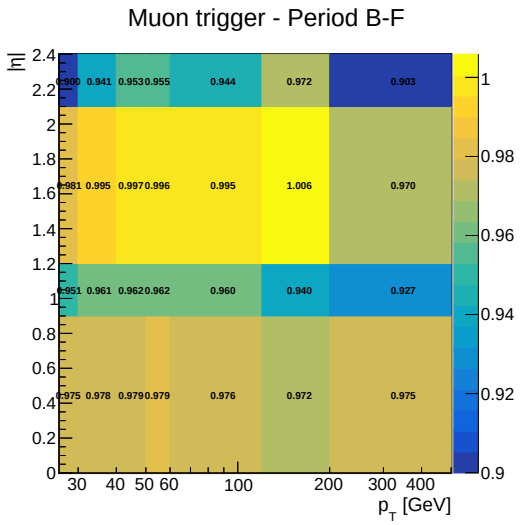
Table 4.2 – Trigger paths used in the $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ final states. The corresponding run interval or data taking periods and the integrated luminosity collected are also reported. The logical "or" of the two muon paths covering the same run periods is used. Details about the isolation criteria reported in the table are discussed in the text.

The same trigger selections are applied onto MC simulated events. The resulting trigger efficiencies are compared to those measured in data and systematic differences are corrected. For the single muon triggers used in the $\tau_\mu\tau_h$ final state, the efficiencies are measured using $Z \to \mu\mu$ events selected with a tag and probe technique, that consists in a selection of a sample of muon candidates that is unbiased with respect to trigger requirements and is consequently used to compute the efficiency of the latter. The selection method identifies one "tag" muon with tight trigger, reconstruction and identification criteria, and uses the kinematics of the $Z \to \mu\mu$ decay to identify a second "probe" muon without directly applying trigger criteria on it. The fraction of the probe muons that satisfy the trigger requirements is subsequently verified to measure the efficiency. This method is analogous to the tag and probe technique described in the context of the L1 τ_h trigger efficiency measurement that is described in Section 3.8.2 of Chapter 3. A correction factor for the muon trigger efficiency, or "scale factor" (SF), is derived as a function of the p_T and η values of the reconstructed muon. The efficiency of the single electron trigger is measured using $Z \to ee$ events selected with the same method and SFs are derived as a function of the electron p_T, separately for the barrel and endcaps regions. The efficiency of the $\tau_h\tau_h$ trigger is measured using $Z \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ events that are selected as described in Section 3.8.2 of Chapter 3. Part of my work has been devoted to the extension of the tag and probe procedure used for the L1 trigger to HLT efficiency measurement and to the derivation and cross check of the results. The SF for the $\tau_h\tau_h$ trigger are defined as a function of the τ_h candidate p_T, as well as of the decay mode of the lepton.

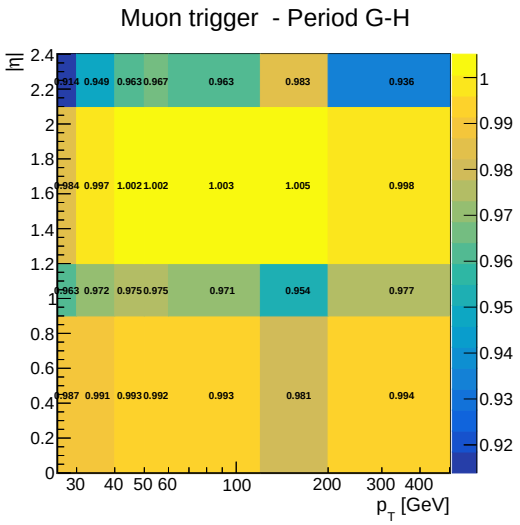
The values of these SFs are shown in Figure 4.1 for the three final states studied. They are applied onto the selected electron or muon candidate for the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final state, and to both selected τ_h candidates for the $\tau_h\tau_h$ final state. Muon SFs are derived separately for the B–F and G–H data taking periods, as different performance of the strip tracker detector was observed. An average SF, weighted on the relative luminosity corresponding to the two periods, is applied to the simulated events. Similarly, in the case of $\tau_h\tau_h$ trigger, two sets of SFs are derived separately for the data taking periods ranging from B to G and the period H, because of the changes occurred in the τ_h isolation both at L1 and HLT to cope with the higher instantaneous luminosity conditions. As only the isolation corresponding to the former period is simulated in the MC samples, the two sets of SFs are averaged and weighted on the integrated luminosity of the corresponding data taking periods. The SFs are derived separately for the three reconstructed τ lepton decay modes.

4.3 Preselection of $H \to \tau\tau$ objects

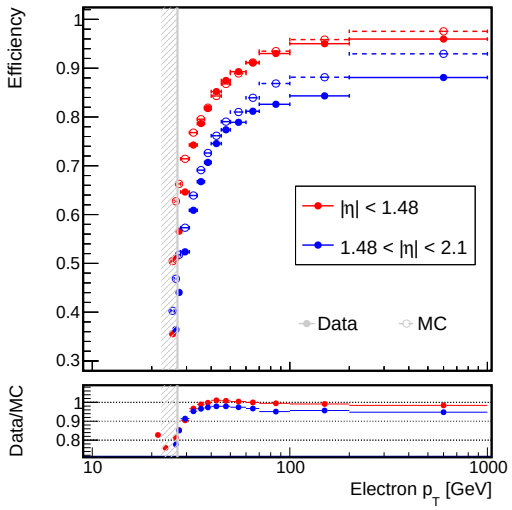
The exploration of the final states of the $H \to \tau\tau$ decay requires the reconstruction and selection of electrons, muons, and τ_h objects in the event, as well as of the imbalance in the transverse momentum due to neutrinos from τ decays. Quality criteria are thus applied on the objects reconstructed with the standard algorithms described in Section 2.3 of Chapter 2 and are optimized for this search. Data are used to verify that objects are well described in the MC simulation and to correct for any possible discrepancy. The specific choices on the final state objects produced in the $H \to \tau\tau$ decay and the corrections applied onto the MC simulation are described in this section.



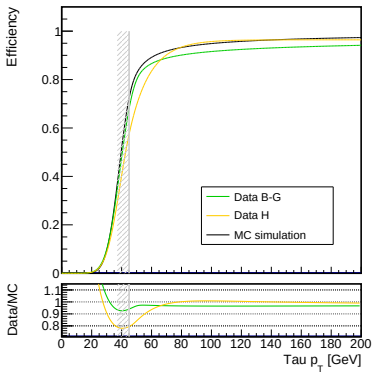
(a) Muon, period B–F



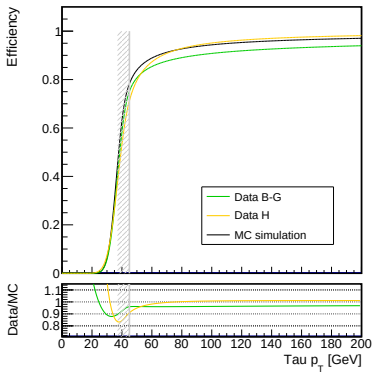
(b) Muon, period G–H



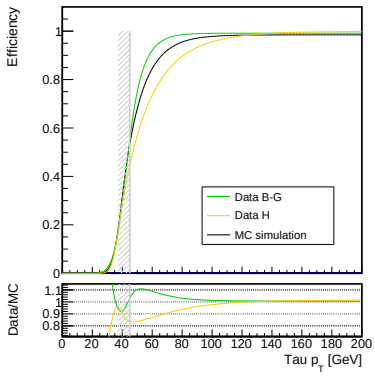
(c) Electron



(d) τ_h– 1-prong decays



(e) τ_h– 1-prong + π^0 decays



(f) τ_h– 3-prongs

Figure 4.1 – Trigger SFs for the single muon (a), (b), single electron (c) and τ_h (d), (e), (f) triggers. Muon SFs are computed for separate data taking periods, and τ_h SFs are computed separately for the three reconstructed decay modes of the τ lepton.

4.3.1 Electron selection

Electrons are reconstructed by combining ECAL and tracker informations using the algorithm described in Section 2.3.3 of Chapter 2. The reconstruction is initiated from clusters of energy deposits in the ECAL, which are then matched to tracks in the inner silicon tracker refitted by a Gaussian-Sum Filter (GSF).

Electron identification relies on a multivariate approach consisting of a boosted decision tree (BDT) classifier [135, 136], that combines observables sensitive to the amount of bremsstrahlung along the electron trajectory, the geometrical and momentum matching between the electron trajectory and associated clusters, shower-shape observables, and electron conversion variables. The BDT discriminant is trained on electrons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ in three separate pseudorapidity intervals, corresponding to two regions in the barrel and the detector endcaps. For each of these regions, selections on the discriminator value ("working point", or WP) are defined and correspond to different signal efficiencies and hadron jet misidentification probabilities.

An isolation criterion is applied to further suppress the contamination from the hadron jet background. The relative electron isolation $\mathcal{I}_{\mathrm{e}}^{\mathrm{rel}}$ is defined as the sum of the transverse momenta of PF candidates reconstructed within a distance $\Delta R < 0.3$ from the electron, normalized to its transverse momentum:

$$\mathcal{I}_{\mathrm{e}}^{\mathrm{rel}} = \frac{\sum p_{\mathrm{T}}^{\mathrm{charged}} + \max\left(0, \sum p_{\mathrm{T}}^{\mathrm{neutral}} + \sum p_{\mathrm{T}}^{\gamma} - \frac{1}{2}\sum p_{\mathrm{T}}^{\mathrm{PU}}\right)}{p_T} \tag{4.1}$$

where $\sum p_{\mathrm{T}}^{\mathrm{charged}}$, $\sum p_{\mathrm{T}}^{\mathrm{neutral}}$, and $\sum p_{\mathrm{T}}^{\gamma}$ are respectively the scalar sums of the transverse momenta of charged hadrons, neutral hadrons, and photons originating from the primary vertex, while $\sum p_{\mathrm{T}}^{\mathrm{PU}}$ is the sum of transverse momenta of charged hadrons not originating from the primary vertex.

Electron candidates considered in this search must satisfy the "tight" WP of the BDT discriminant, corresponding to a signal efficiency of about 80%. They must also have $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and $|\eta| < 2.1$, where the values are chosen as a function of the corresponding geometrical and momentum acceptance of the single electron HLT trigger requirement, and satisfy the isolation criterion $\mathcal{I}_{\mathrm{e}}^{\mathrm{rel}} < 0.1$. Finally, the associated GSF track must have a distance from the primary vertex of $\Delta_{xy} < 0.045\,\mathrm{cm}$ in the transverse plane and of $\Delta_z < 0.2\,\mathrm{cm}$ in the longitudinal direction.

A correction factor is applied to the MC simulation to account for differences with respect to data in the electron isolation and identification efficiencies. These are measured using $\mathrm{Z} \to \mathrm{ee}$ events selected with a tag and probe technique as done for the trigger SFs. The combined efficiencies and their ratio, used to correct the MC simulation, is shown in Figure 4.2. The magnitude of the correction in the region of interest of this search ranges between 5 and 15%, and the agreement observed between the data and the MC simulation after its application is illustrated in Figure 4.3 for events selected in the $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final state.

4.3.2 Muon selection

As detailed in Section 2.3.2 of Chapter 2, muons are reconstructed from the hits in the tracker subdetector (tracker muon algorithm) and in the muon chambers detectors (standalone muon algorithm), that can be combined together to improve the quality of the

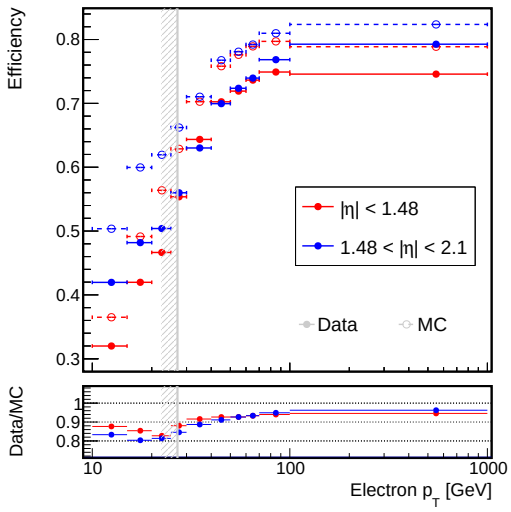
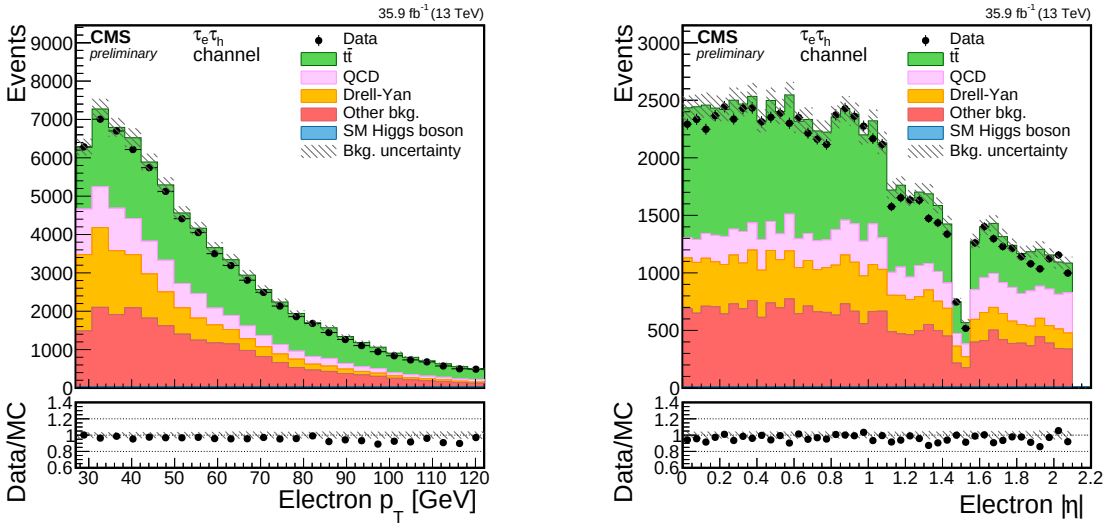


Figure 4.2 – Data and MC simulation efficiencies for the combination of identification and isolation criteria described in the text (top panel) and their ratio (bottom panel) used to correct the MC simulation. The red and blue markers denote the two separate pseudorapidity intervals used to derive the correction factors. The shaded vertical grey line denotes the p_{T} selection applied in this analysis.



(a) e p_{T} (b) e $|\eta|$

Figure 4.3 – Electron p_{T} (a) and $|\eta|$ (b) distributions for events selected in the $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final state.

reconstruction (global muon algorithm) [93]. A Kalman filter method accounts for energy losses from the interaction in the material, and the curvature of the resulting track is used to estimate the muon momentum. Muon identification relies on the quality of the reconstructed muon track and on the number of hits. Several working points are defined accordingly, and the "loose" and "tight" ones are applied in this search. The former requires that the muon satisfies the general PF identification criterion and that it is reconstructed either with the tracker or global algorithms. The latter restricts the selection to muons reconstructed with the global algorithm, and puts a stringent requirement on the

χ^2 of the associated track fit. In addition, requirements on the minimal number of hits in the muon chambers, strip tracker, and pixel detectors, as well as on the distance from the primary interaction vertex, are applied to suppress the erroneous identification of hadrons escaping the calorimeter volume, cosmic rays, and in-flight decays of other particles. The efficiency of the loose and tight working points is of about 99% and 96% respectively. As for electrons, a relative isolation criterion $\mathcal{I}_{\mu}^{\text{rel}}$ is defined with the formula (4.1) using the PF candidates within a distance $\Delta R < 0.4$ around the muon.

Muon candidates considered in this search must satisfy the tight isolation criterion and have $p_{\text{T}} > 23\,\text{GeV}$ and $|\eta| < 2.1$. The distance of the muon track from the primary vertex must be $\Delta_{xy} < 0.045\,\text{cm}$ in the transverse plane and $\Delta_z < 0.2\,\text{cm}$ in the longitudinal direction, and the candidate must satisfy the isolation criterion $\mathcal{I}_{\mu}^{\text{rel}} < 0.15$.

Correction factors are applied to the MC simulation to take into account differences with respect to data in the muon isolation and identification efficiencies, and are derived using $\text{Z} \to \mu\mu$ events selected with the tag and probe technique previously discussed. In the data taking periods ranging from B to F (discussed in Section 2.1.2 of Chapter 2) an inefficiency of the strip tracker was observed. The correction factors are consequently computed separately for two datasets, corresponding to the periods from B to F and for the periods G and H. The former are globally a few percent lower than the latter, as it can be observed from the values reported in Figure 4.4. The two sets of correction factors are combined as an average weighted on the relative luminosity of the two datasets considered. The agreement between the simulation and the observed data after the application of these correction factors for the selected muon p_{T} and η in $\tau_{\mu}\tau_{\text{h}}$ events is illustrated in Figure 4.5.

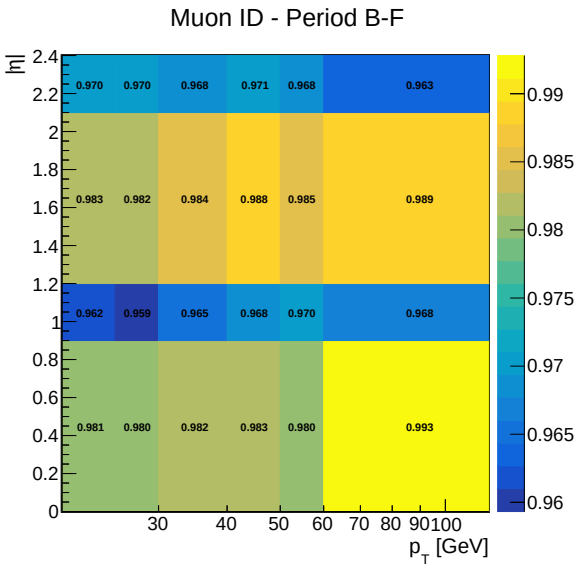
4.3.3 Tau lepton selection

Decays of τ into hadrons and a neutrino are reconstructed with the hadrons plus strips (HPS) identification algorithm that has been described in Section 2.3.4 of Chapter 2. It is recalled here that the τ_{h} identification algorithm analyses and combines the PF constituents of the reconstructed jets to identify the specific τ decay mode and reject jets issuing from a quark or gluon decay. A τ lepton decay often involves one or two π^0, that subsequently decay into a photon pair. Given the large probability for these photons to convert into an e^-e^+ pair within the detector material, the algorithm regroups photons and electrons constituents of the jets into "strips" using a dynamic clustering technique. Four signatures are considered, corresponding to a single charged hadron in association to zero, one, or two strips, and three charged hadrons with no strips associated.

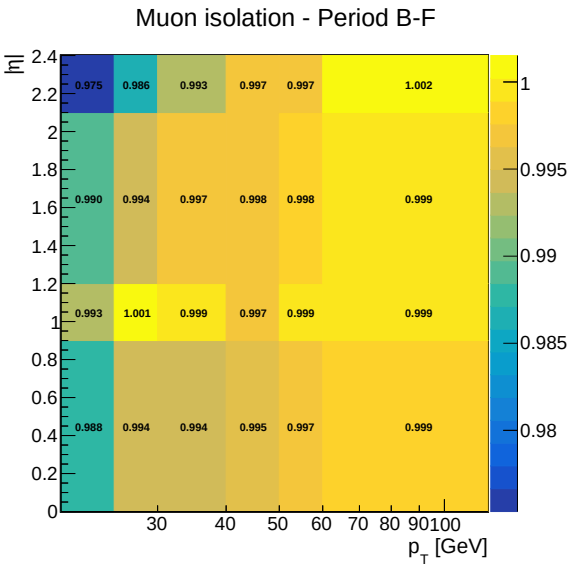
Quark and gluon jets constitute the dominant background for τ_{h} identification, and isolation criteria are used to reduce their contamination. These criteria are based on the PF candidates reconstructed inside an "isolation cone", i.e. within a distance $\Delta R < 0.5$ from the τ_{h} candidate itself. Three isolation discriminants, exploiting in a different way the properties of such PF candidates, have been studied.

The most straightforward approach consists in the computation of the scalar p_{T} sum of the charged hadron PF candidates with associated tracks originating within a distance of 0.2 cm from the τ_{h} production vertex ($\sum p_{\text{T}}^{\text{ch}}$). An isolated τ_{h} candidate is defined by requiring $\sum p_{\text{T}}^{\text{ch}}$ to be smaller than a fixed threshold.

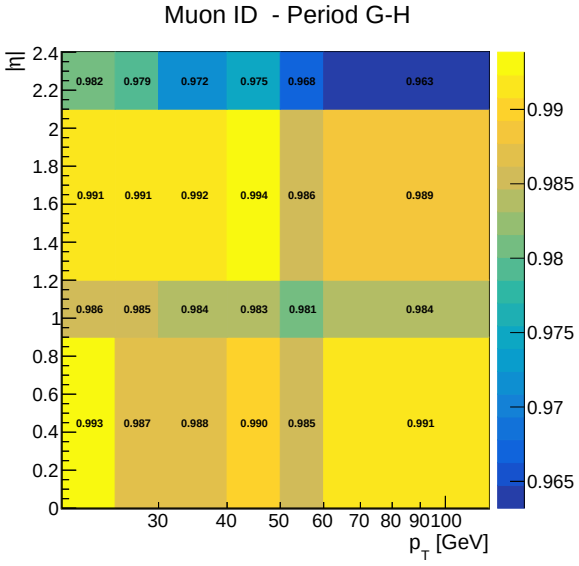
A second approach combines this variable with the sum of the transverse momenta of photons reconstructed in the isolation cone ($\sum p_{\text{T}}^{\gamma}$), and corrects its value for the pileup contribution in the event estimated from the charged candidates that do not originate



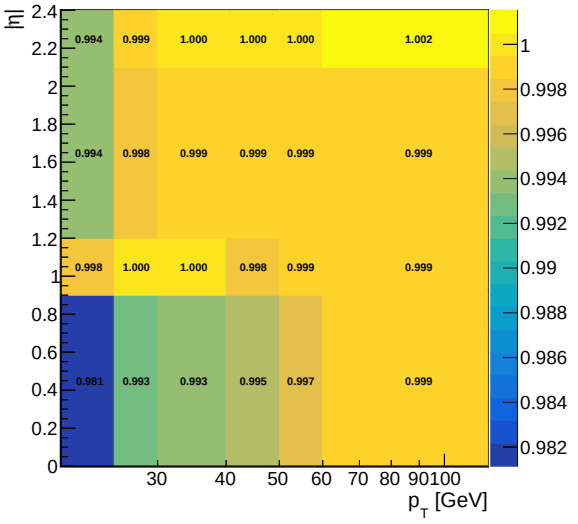
(a) Identification - periods B to F



(b) Isolation - periods B to F

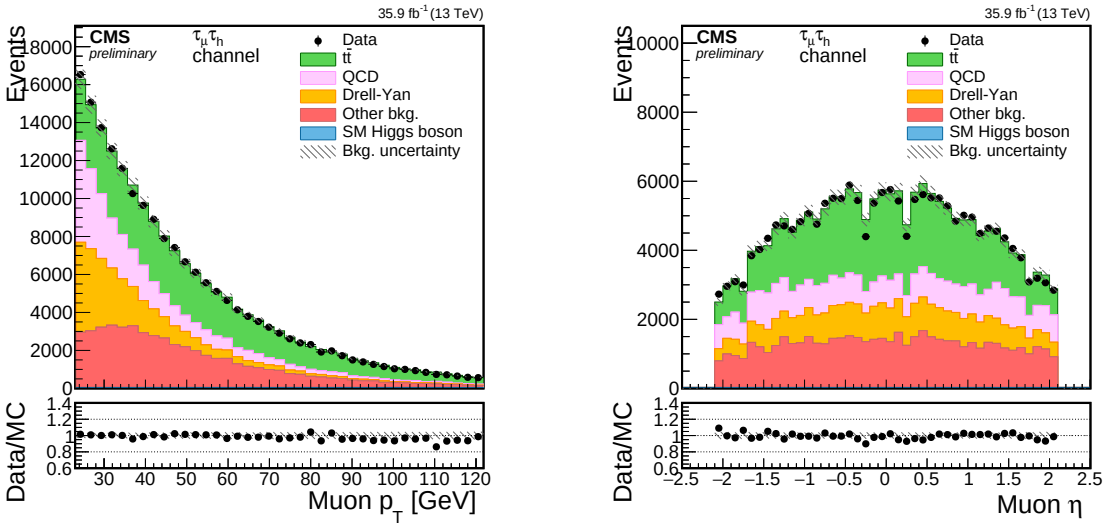


(c) Identification - periods G and H



(d) Isolation - periods G and H

Figure 4.4 – MC simulation correction factors for the muon identification (left) and isolation (right) efficiencies for the data taking period between run B and F (top) and G and H (bottom).



(a) μ p_{T} (b) μ η

Figure 4.5 – Muon p_{T} (a) and η (b) distributions for events selected in the $\tau_\mu\tau_{\mathrm{h}}$ final state.

from the τ_{h} production vertex. The isolation $\mathcal{I}_\tau$ is consequently defined as:

$$\mathcal{I}_\tau = \sum p_{\mathrm{T}}^{\mathrm{ch}}(\Delta_{\mathrm{z}} < 0.2\,\mathrm{cm}) + \max\left(\sum p_{\mathrm{T}}^{\gamma} - \Delta\beta \sum p_{\mathrm{T}}^{\mathrm{ch}}(\Delta_{\mathrm{z}} > 0.2\,\mathrm{cm}), 0\right) \tag{4.2}$$

where the $\Delta\beta$ factor is determined as a function of $\sum p_{\mathrm{T}}^{\gamma}$. Loose, medium, and tight working points are defined by requiring $\mathcal{I}_\tau$ to be smaller than 2.5, 1.5, and 0.8 GeV respectively. In addition, a selection on the p_{T} sum of the electron and photon candidates ($\sum p_{\mathrm{T}}^{\mathrm{e}/\gamma}$) that were clustered into the HPS algorithm strips but are not part of the τ_{h} candidate is applied to further reduce the contamination from jet background. This selection is defined as:

$$\sum p_{\mathrm{T}}^{\mathrm{e}/\gamma}(\Delta R > R_{\mathrm{sig}}) < 0.1 \cdot p_{\mathrm{T}} \tag{4.3}$$

where $0.05 < R_{\mathrm{sig}} = 3\,\mathrm{GeV}/p_{\mathrm{T}} < 0.1$ and p_{T} is the transverse momentum of the τ_{h} candidate. The combination of the selections on the quantities defined in formulae (4.2) and (4.3) defines a "cut-based" isolation discriminant.

The third and most advanced isolation discriminant improves the rejection of quark and gluon jets by combining the isolation variables previously described with τ lifetime related variables. A multivariate analysis (MVA) method consisting of a BDT is used. Its input variables include the isolation and energy sums previously defined in the formulae (4.2) and (4.3), the reconstructed τ_{h} decay mode, the value and significance of the impact parameter of its highest p_{T} track, the value and significance of the distance between the τ production and decay vertices, and the photon and electron candidates multiplicities and topology inside the τ_{h} signal cone. Further details on such variables are provided in [96, 97]. Selections on the BDT output are defined as a function of the p_{T} of the τ_{h} to ensure a constant signal efficiency.

The comparison of the performance of the three isolation criteria is summarized in Figure 4.6, where the signal corresponds to SM HH events and the multijet background is

estimated from the data, as detailed in Section 5.4 of Chapter 5. It is shown for selected $\tau_h\tau_h$ events that must contain two jets compatible with a H → bb decay, as described in Section 4.6. The signal efficiency ε is defined as the fraction of events where both τ_h satisfy the isolation criteria, while the background rejection is defined as $1 - \varepsilon$. The superior performance of the MVA method can be observed, with its medium WP appearing as the best compromise between background rejection and signal efficiency. The improvement in the sensitivity to SM and BSM HH production with the usage of this WP is as large as 20% in comparison to other WPs.

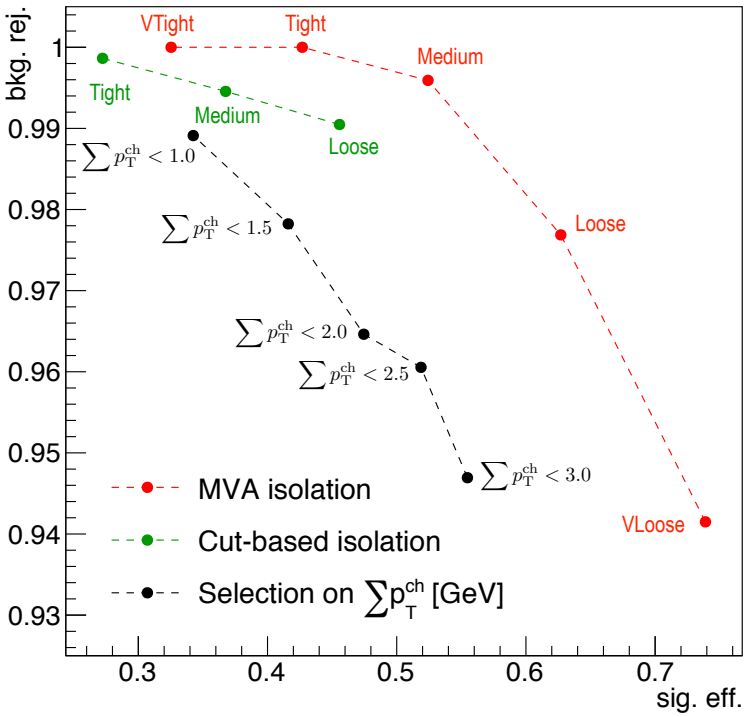
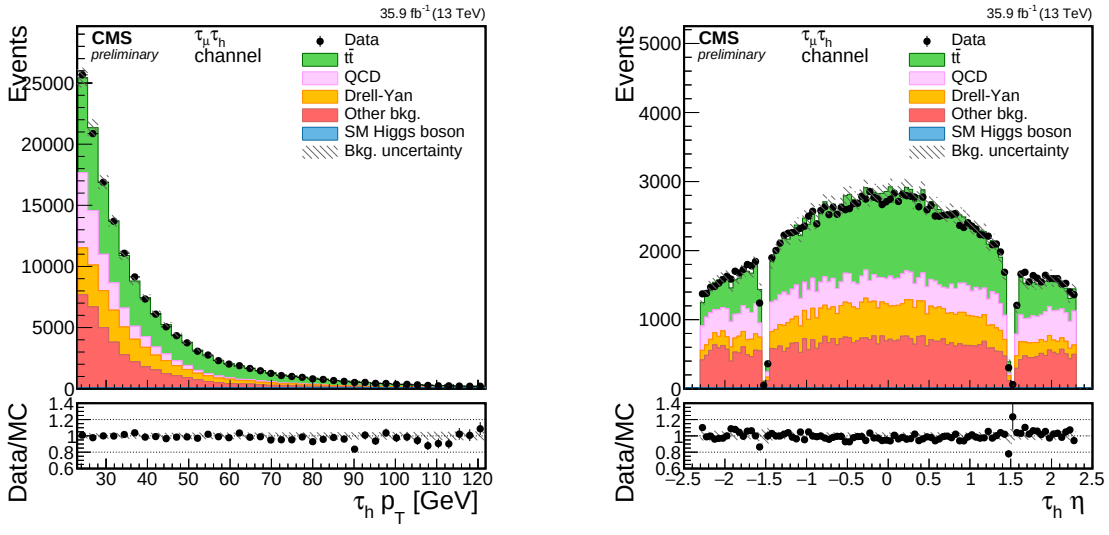


Figure 4.6 – Comparison of the efficiency on HH signal versus rejection of multijet background for the three isolation criteria described in the text.

Electron and muons constitute an additional background source in τ_h identification, and their contribution is suppressed with dedicated discriminants. The anti-electron discriminant consists of a BDT that is based on several variables related to the fraction of energy deposited in the ECAL and HCAL subdetectors, to the multiplicity, topology, and energy of the photons inside the τ_h candidate signal cone, and to the reconstructed tracks and their changes in curvature due to bremsstrahlung radiation. A detailed list of the input variables can be found in [96, 97]. Several working points are provided for the anti-electron discriminant, and two of them are used in this search, denoted as "very loose" and "tight" WPs. The former results in a signal and background efficiency of 90% and 5% respectively, while for the latter these values are 75% and 0.1–0.4% respectively. The anti-muon discriminant rejects τ_h candidates in case signals in the muon subdetectors are found and are aligned to the τ_h direction. Two working points of the anti-muon discriminant, differing for the quality requirements on these signals, are denoted as "loose" and "tight" and result in a probability of a muon to be misidentified as a τ_h of about 0.1–0.3% and 0.05–0.1%, respectively. The corresponding signal efficiencies are about 99% in both cases for the p_T range of interest of this search.

The τ_h candidates used in this search must satisfy the identification criterion and the medium WP of the MVA isolation discriminant. This WP has an efficiency of about 60% for genuine τ_h decays, approximately flat as a function of p_T, and a misidentification probability of quark and gluon jets that ranges from about 2% at 30 GeV to 0.1% at 100 GeV [97]. It should be noted that the corresponding signal efficiency in Figure 4.6 is higher than $(60\%)^2 \approx 35\%$, because it is computed for candidates that already satisfy isolation criteria at the HLT level. For the τ_h candidates considered in this search, the highest p_T track associated must be within a distance along the beam direction $\Delta_z < 0.2$ cm from the primary vertex of the event. In the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states, τ_h candidates must have $p_T > 20$ GeV and $|\eta| < 2.3$, while in the $\tau_h\tau_h$ final state the selected candidates must have $p_T > 45$ GeV and $|\eta| < 2.1$ because of the trigger requirements previously discussed. The very loose WP of the anti-electron discriminant is applied in the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ final states, while the tight one is required in $\tau_e\tau_h$ events. The different WP is chosen to better suppress the large $Z/\gamma^* \to ee$ background in the latter case. Similarly, the loose WP of the anti-muon discriminant is applied in $\tau_e\tau_h$ and $\tau_h\tau_h$ events while the tight WP is applied in the $\tau_\mu\tau_h$ final state.

The combined τ_h reconstruction, identification, and isolation efficiency has been measured using $Z \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ events selected with the tag and probe technique already described. Two independent measurements are performed using as sensitive variables the invariant mass of the μ and τ_h selected candidates and the number of signal and isolation tracks associated to the τ_h candidate. The efficiencies are found to be compatible with both methods between data and MC simulation within a 5% uncertainty in the measurement, and no further corrections are applied. The agreement between the data and the MC simulation can be observed in Figure 4.7 for the p_T and η distribution of τ_h candidates selected in $\tau_\mu\tau_h$ events. Candidates in the two regions around $|\eta| = 1.5$ are vetoed by the anti-electron discriminant because of the high probability of an electron to interact in the CMS detector material and be consequently misidentified as a τ_h.



(a) τ_h p_T (b) τ_h η

Figure 4.7 – τ_h p_T (a) and η (b) distributions for events selected in the $\tau_\mu\tau_h$ final state.

4.3.4 Missing transverse momentum

The imbalance of the transverse momentum sum of the reconstructed particles ($\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$) in $\mathrm{HH} \to \mathrm{bb}\tau\tau$ events mainly originates from neutrinos from the two τ leptons decays. A fraction is also due to neutrinos from the decays of the B hadrons produced in the hadronization of the two b quarks, but accounts only for a small part of the total magnitude of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, as it is illustrated in Figure 4.8 for several resonant HH signals. It can be observed that the relative contribution of neutrinos from B decays to the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event decreases as $\mathrm{m_X}$ becomes larger. While the momentum of neutrinos from τ decays is directly related to the τ lepton momentum and thus increases with $\mathrm{m_X}$, hadronization and fragmentation effects distribute the original b quark momentum over a multitude of final products, reducing the dependence of the B hadron momentum on the original b quark momentum. Although not directly used to select $\mathrm{bb}\tau\tau$ events, the magnitude and direction of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector are combined to other observables and provide an efficient way to reject the $\mathrm{t\bar{t}}$ background, as it is discussed further in Section 4.6.2.

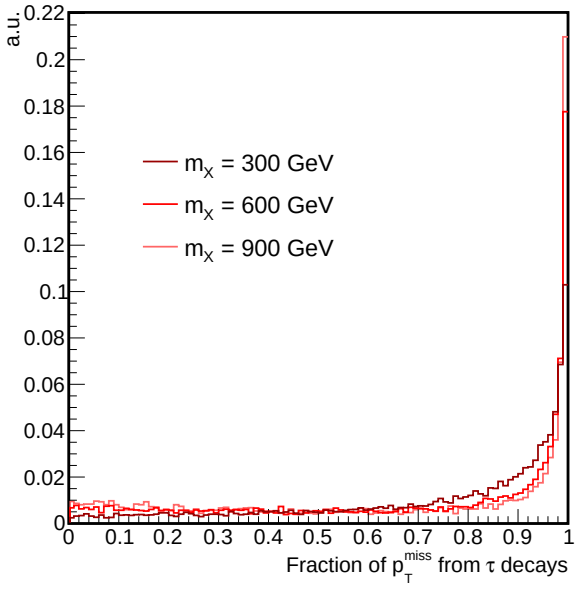
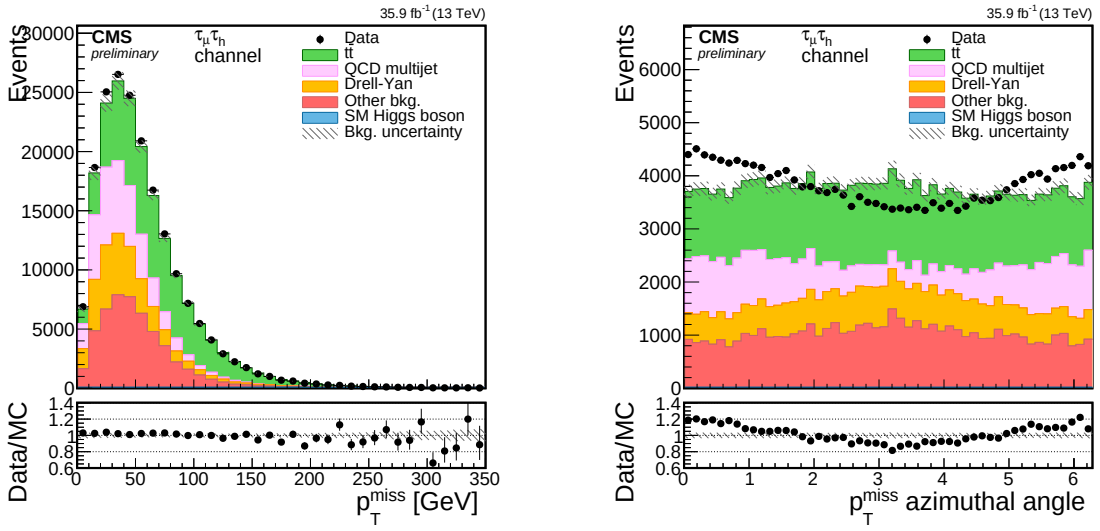


Figure 4.8 – Distribution of the fraction of missing transverse momentum due to neutrinos produced in τ decays with respect to the total transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event, for resonant HH production under the hypothesis of a resonance of mass $\mathrm{m_X}$ of 300, 600, and 900 GeV. The value of $p_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the vector sum of the neutrinos produced in the MC simulation.

As detailed in Section 2.3.6 of Chapter 2, the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector is reconstructed with the PF algorithm. The distributions of the magnitude and direction of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector are illustrated in Figure 4.9 using the events selected in the $\tau_\mu\tau_{\mathrm{h}}$ final state. Good agreement with the MC simulation is observed for the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ magnitude, while discrepancies can be noticed for the azimuthal angle distribution. This effect is explained from anisotropies and inefficiencies in the detector response, misalignments, inactive calorimeter cells, and displacements of the pp interaction region ("beam spot"), that are not fully taken into account in the simulation. The direction of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ with respect to other objects in the

event is instead well modelled and can be used to separate the $t\bar{t}$ background from the HH signal, as discussed in Section 4.6.2.



(a) Magnitude (b) Azimuthal angle

Figure 4.9 – Distribution of the $\vec{p}_T^{miss}$ vector magnitude (a) and azimuthal angle (b) in events selected in the $\tau_\mu\tau_h$ final state. The disagreement observed in (b) comes from anisotropies, misalignments, and inefficiencies of the detector not accounted for in the simulation. They do not constitute a concern for the analysis, as the direction of the $\vec{p}_T^{miss}$ vector with respect to other objects in the event is instead correctly modelled, as shown for example in Figure 4.23.

4.3.5 Other selections

The two reconstructed leptons in the $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ final states are required to be separated by a distance $\Delta R > 0.1$ to reject those cases where the same PF candidate is simultaneously used by two different object identification algorithms.

The two leptons forming the $H \to \tau\tau$ candidate are also required to have opposite electric charge. The probability of wrong electric charge assignment is about 0.2% for τ_h objects [97], and largely inferior for electrons and muons [137, 138], making this requirement highly efficient in the selection of signal events.

The reconstructed leptons must correspond to the HLT candidates on which the trigger decision is made. Because the trigger efficiencies are computed and measured separately for single objects, this match ensures that no bias is introduced when applying the trigger SFs to simulated events. The correspondence is ensured by requiring the selected muon and electron in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states to be within a distance $\Delta R = 0.5$ from the corresponding HLT objects that satisfy the trigger requirements. The same spatial matching requirement with HLT objects is applied on both τ_h candidates reconstructed in the $\tau_h\tau_h$ final state.

Events containing additional isolated electrons or muons are rejected to suppress the contribution from Drell-Yan production. As no additional leptons are expected in the case of the signal process, this requirement is highly efficient for signal events. Electrons considered for this veto must be identified with the loose WP of the multivariate isolation

discriminant, corresponding to an identification efficiency of about 90%. They must have $p_T > 10\,\text{GeV}$, $|\eta| < 2.5$, and $\mathcal{I}_e^{\text{rel}} < 0.3$. Muons must satisfy the loose WP of the muon identification criteria, and have $p_T > 10\,\text{GeV}$, $|\eta| < 2.4$, and $\mathcal{I}_\mu^{\text{rel}} < 0.3$. The event is rejected if at least one electron or muon besides the one associated to the $H \to \tau\tau$ candidate previously selected satisfies these criteria.

4.4 Preselection of H → bb objects

The two b quarks produced in the $H \to bb$ decay are experimentally observed as hadron jets in the detector. Their reconstruction and separation from jets produced by light flavour quarks and gluons is essential to identify signal events.

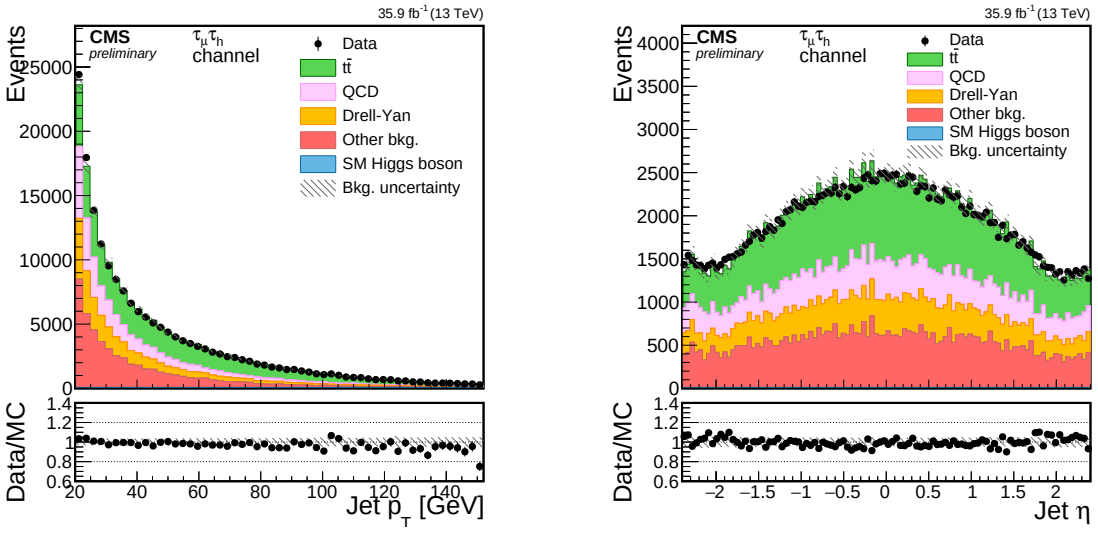
4.4.1 Jet selection

Jets are reconstructed from PF candidates using the anti-k_T algorithm operated with a distance parameter $R = 0.4$ (AK4 jets) and $R = 0.8$ (AK8 jets), as detailed in Section 2.3.5 of Chapter 2.

The AK4 jets considered in this search are required to have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$. The η requirement is due to the necessity to apply jet b tagging criteria, as described in the next section, that can be computed only in the regions where tracking information is available. Jets reconstructed within a distance $\Delta R = 0.5$ from the two selected $H \to \tau\tau$ signal leptons are not considered.

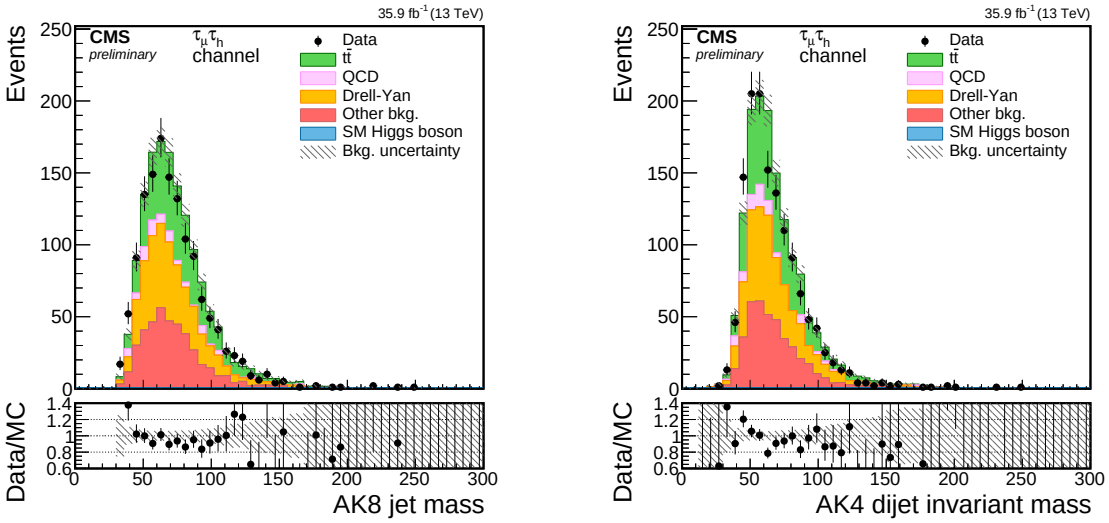
A particle flow jet identification criterion is applied to suppress jets poorly reconstructed or due to noise in the detector. Jet related observables used in this identification criterion correspond to the fraction of charged and neutral hadrons clustered within the jet, the charged hadron multiplicity, and the fraction of energy deposited in the ECAL subdetector by the neutral and charged hadron candidates [102]. The identification efficiency has been measured in data using di-jet and inclusive jet event samples and corresponds to about 99.7% for the $|\eta|$ region of interest of this analysis, while the background rejection is measured in events selected with a zero bias trigger and ranges between 99.8 and 100%. The good modelling of jet position and momentum in the simulation is illustrated in Figure 4.10 for events selected in the $\tau_\mu\tau_h$ final state.

The AK8 jet algorithm better reconstructs high Lorentz boost $H \to bb$ decays, where the two b quarks are typically close to each other and their hadronization products partially overlap. The "soft drop" jet grooming algorithm [139] is used to identify the contribution of the two b quarks inside the reconstructed AK8 jet and to mitigate the contribution from initial state radiation, pileup, and underlying event, that can be sizeable given the large radius of the jet. The algorithm recursively removes the soft, wide-angle radiation, and the resulting jet is iteratively decomposed into sub-jets [140]. This efficiently identifies the contributions in the initial AK8 arising from the two quarks in $H \to bb$ decays and improves the resolution on the jet invariant mass. The distribution of the AK8 jet mass thus computed is shown in Figure 4.11 for the $\tau_\mu\tau_h$ final state, where it is compared to the invariant mass of the two AK4 jets for the same events selected.



(a) Jet p_{T} (b) Jet η

Figure 4.10 – Jet p_{T} (a) and η (b) distributions for events selected in the $\tau_{\mu}\tau_{\mathrm{h}}$ final state.



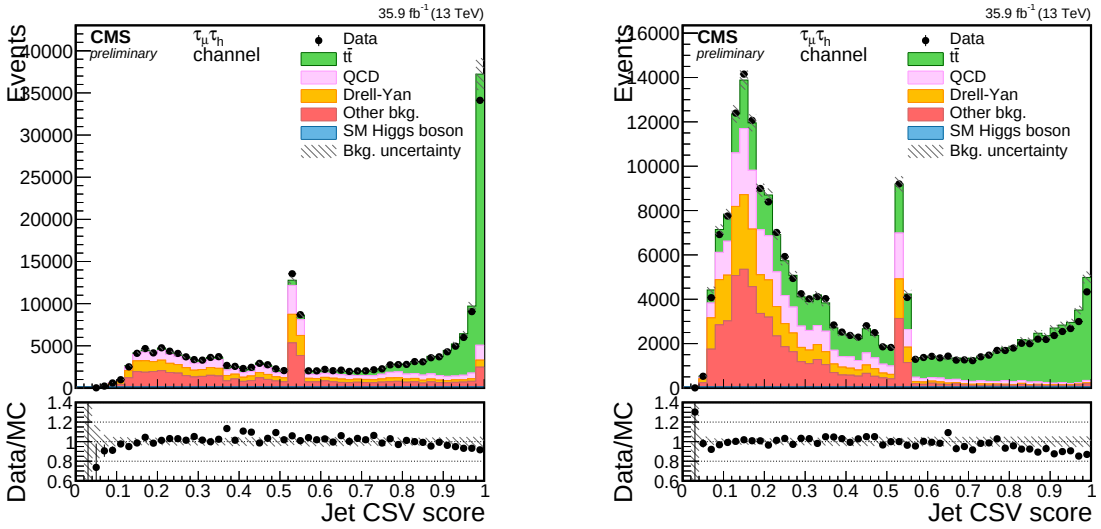
(a) AK8 jet mass (b) AK4 jet pair invariant mass

Figure 4.11 – (a) AK8 jet invariant mass and (b) invariant mass of the two AK4 jets selected in the same events.

4.4.2 Identification of b jets

Jets originating from b quarks are identified using the PF-based combined secondary vertex (CSV) algorithm [141]. This method exploits the long lifetime of hadrons containing b quarks, that usually can decay a few millimetres away from the primary interaction vertex, as well as their probability to undergo a decay that contains an electron or a muon. It combines the information from the reconstructed tracks and from their production vertex into a single discriminant using a multivariate technique. A jet is defined as b-tagged if the value of the CSV discriminant associated is larger than a fixed threshold, which choice determines a different trade-off between the efficiency of genuine b jet identification and

the misidentification probability for light flavour jets. In this search two thresholds, corresponding to a "loose" and "medium" WP, are used. The former corresponds to a b jet identification efficiency of about 80% and misidentification rate of 10%, while the latter achieves a signal identification efficiency of about 65% for a background misidentification rate of 1%. The medium WP corresponds to a selection CSV > 0.8484 and the loose one to CSV > 0.5426. The distribution of CSV observed in the data is shown in Figure 4.12.



(a) Highest CSV jet

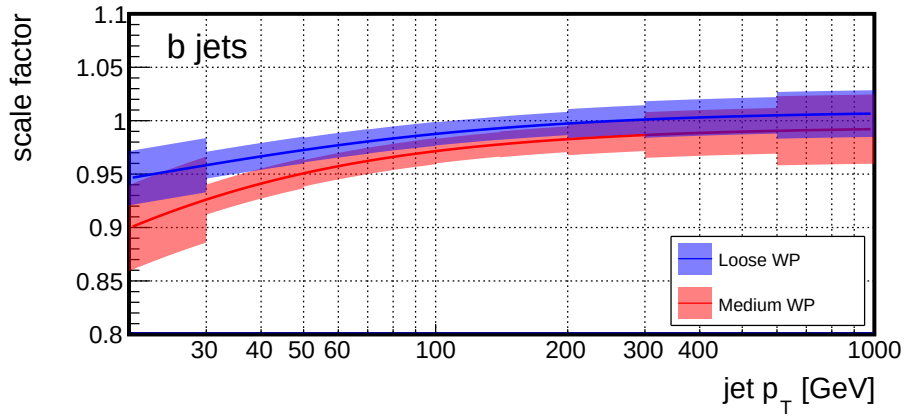
(b) Second-to-highest CSV jet

Figure 4.12 – Distribution of the CSV b tagging algorithm output for the two jets selected in $\tau_\mu \tau_h$ events.

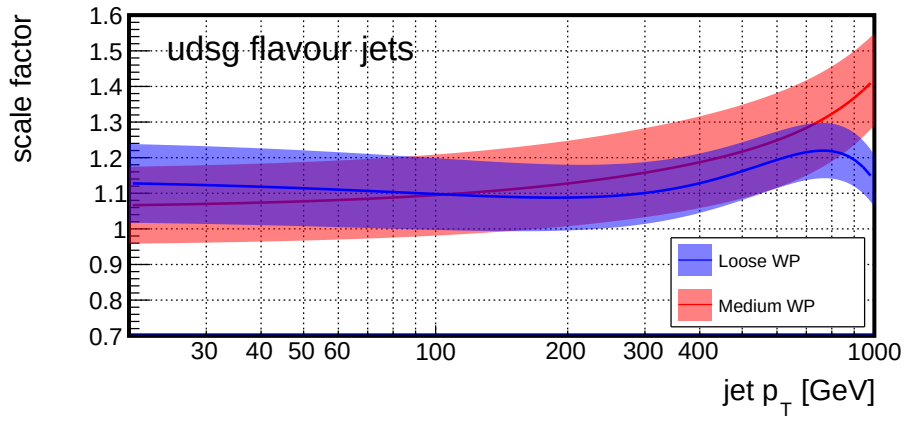
A disagreement up to 10% is observed at high CSV values between the data and its modelling in the MC simulation. To correct for this effect, the tagging efficiency for b and c jets is measured using $t\bar{t}$ and QCD events containing muons, while misidentification rates for light flavour jets are measured in inclusive QCD and Drell-Yan events. The ratio of the efficiencies observed in data and MC simulated events, denoted as "b tag SF", is derived in these data samples as a function of the jet p_T and η. For the MC simulation, the b tag SF is separately computed according to the underlying jet flavour, which is determined by re-clustering the PF candidates together with the generated B and charmed (C) hadrons after rescaling their four-momenta to a very small value (a factor of 10^{-18} is applied). In this way, B and C hadrons have negligible impact on the jet properties, and their are used to determine the flavour of the jet they are clustered into.

The values of the measured SF as a function of the jet p_T are shown in Figure 4.13. Efficiencies for genuine b jets are found to be usually overestimated in the MC simulation, while the misidentification probability of light flavour and gluon jets is typically underestimated.

A correction ω is consequently introduced in the simulated samples to account for this difference in the b tagging algorithm. The probability of a given simulated event, containing n jets divided into a "tagged" and an "untagged" subsets, to satisfy the b tagging criteria applied is defined as P^{MC}. It depends not only on the probability of genuine b jets to be identified by the algorithm, but also on the probability of light flavour and gluon jets to be erroneously identified. An identical event in measured data would satisfy the b tagging criteria with a probability P^{data}, and the value of ω is defined



(a) b jets



(b) udscg jets

Figure 4.13 – Measured b tag SFs as a function of the jet p_{T} for b jets (a) and udscg jets (b). The blue and red curves denote the loose and medium WP respectively. The shaded bands denote the systematic uncertainty in the measurement.

as:

$$\omega = \frac{P^{\mathrm{data}}}{P^{\mathrm{MC}}} \tag{4.4}$$

The two probabilities can be computed from the properties of all the n selected jets as

$$\begin{aligned} P^{\mathrm{data}} &= \prod_{i \in \text{tagged}} \mathrm{SF}_i \varepsilon_i \prod_{j \in \text{not-tagged}} (1 - \mathrm{SF}_j \varepsilon_j) \\ P^{\mathrm{MC}} &= \prod_{i \in \text{tagged}} \varepsilon_i \prod_{j \in \text{not-tagged}} (1 - \varepsilon_j) \end{aligned} \tag{4.5}$$

where ε_i denotes the b tagging efficiency as computed from MC simulated events. The value of ε_i is defined separately for b, c, and light jet flavours, is a function of the jet p_{T} and η, and is obtained from the events selected in the simulated $\mathrm{t\bar{t}}$ sample. The correction ω thus computed modifies the yield associated to a simulated event and corrects it for the different probability to satisfy the b tagging criteria applied with respect to the measured data.

4.5 Event categorization

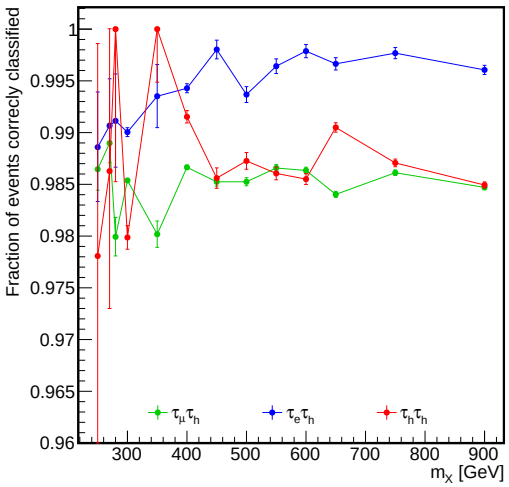
The preselected muon, electron, and τ_h objects are used to construct the $H \to \tau\tau$ candidate and the preselected jets are combined into the $H \to bb$ candidate. The events are classified in three $\tau\tau$ final states and in three categories of bb quality and topology, for a total of nine signal regions simultaneously explored.

4.5.1 $H \to \tau\tau$ final state assignment and selection

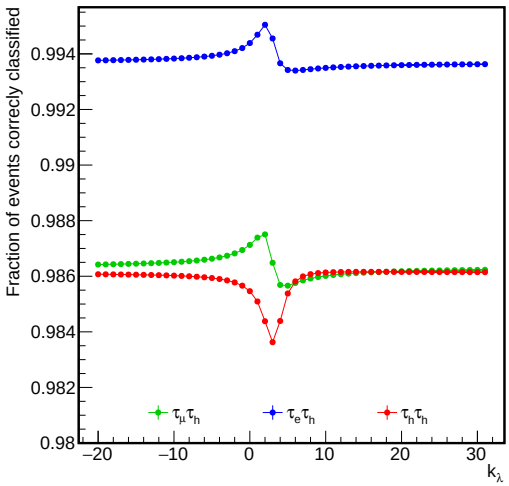
The $\tau\tau$ final state of an event is determined according to the number of electrons and muons that satisfy the preselections discussed in Sections 4.3.1 and 4.3.2. If at least a muon candidate is found, the event is assigned the $\tau_\mu\tau_h$ final state, otherwise it is assigned the $\tau_e\tau_h$ final state if it contains at least one electron candidate. Finally, if no electron nor muon are found, the event is assigned the $\tau_h\tau_h$ final state. Although events containing multiple electron and muon candidates are in principle considered, they are removed by the application of the additional lepton veto described in Section 4.3.5 and, as a result, at most one muon or electron candidate per event is selected. This ensures that the three final states are mutually exclusive and proves to be very effective in the correct determination of the $\tau\tau$ final state. The purity of the assignment is illustrated for resonant and nonresonant HH signals in Figure 4.14, where it can be observed that, after the preselections and the application of the additional lepton veto, the fraction of events correctly classified ranges between 98% and 99.8% depending on the $\tau\tau$ decay mode and signal kinematics. The dominant contamination in the $\tau_\mu\tau_h$ final state is represented by $\tau_\mu\tau_e$ events where the electron is outside the detector acceptance or not identified, and the event is consequently not rejected by the additional lepton veto. Such decays have a smaller impact on the $\tau_e\tau_h$ final states because of the higher reconstruction and identification efficiencies for muons. The dominant contamination in $\tau_h\tau_h$ final states is from $\tau_e\tau_h$ decays where the electron does not satisfy reconstruction quality and isolation criteria and is instead reconstructed as a τ_h. The fraction of generated $\tau\tau$ decays selected in the three final states is illustrated in Figure 4.14c in the case of the SM HH signal.

After the final state is assigned, the $H \to \tau\tau$ candidate pair is chosen as the one satisfying the specific final state selection criteria detailed in Section 4.3, but without applying the τ_h isolation nor the pair electric charge requirements, because both criteria are exploited at a later stage to define background-enriched regions for data-driven estimates. If no pair is found satisfying the selection criteria, the event is rejected. Inversely, in case multiple pairs are found, ambiguities are solved with a dedicated choice algorithm. Such situation is possible as, without applying an isolation criterion, little separation exists between genuine τ_h candidates and jets.

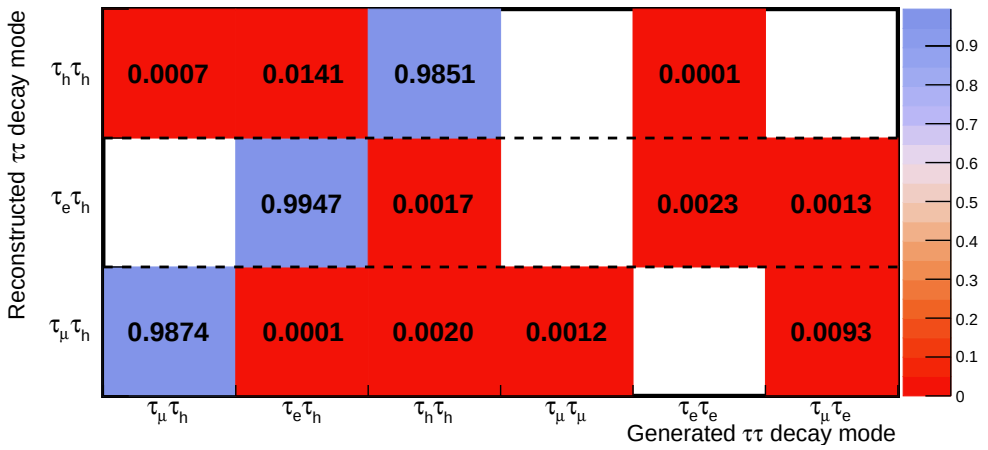
The choice is particularly delicate given the large number of jets typically observed in an event that can be reconstructed as a τ_h. This is illustrated for the $\tau_h\tau_h$ final state in Figure 4.15, where the solid lines denote all the selected events and the dashed lines those where both leptons in the selected $H \to \tau\tau$ pair correspond to genuine τ_h decays. In absence of the isolation criterion, as in Fig. 4.15a, many $H \to \tau\tau$ candidates that satisfy the preselections (without the application of the isolation and pair charge requirements) contain at least one misidentified quark or gluon jet: in the majority of these events, the genuine τ_h does not satisfy the p_T requirement and another jet in the event is selected to form the $\tau_h\tau_h$ pair. A subsequent application of the isolation criterion, shown in



(a) Resonant production



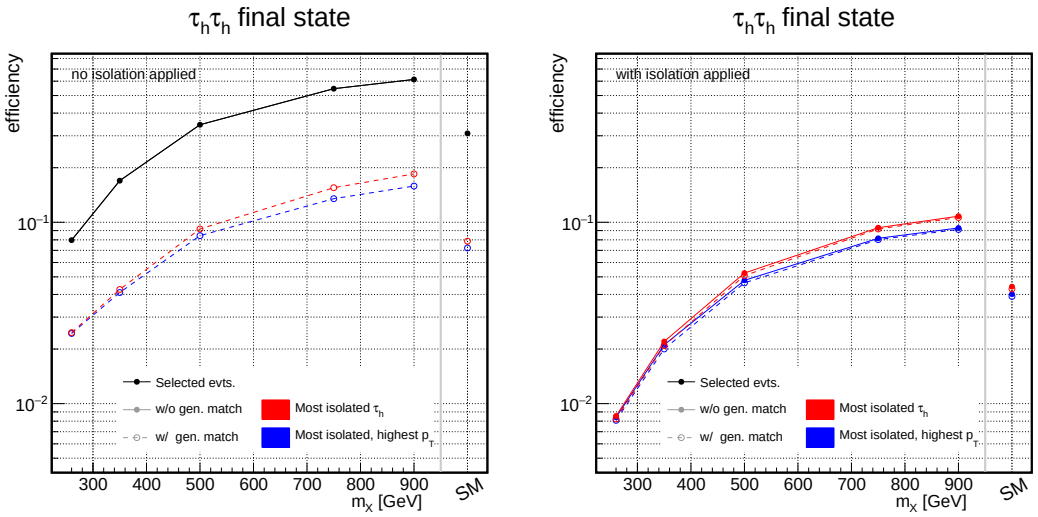
(b) Nonresonant production



(c) SM HH production

Figure 4.14 – (a), (b) Fraction of signal events where the final state assigned corresponds to the generated $\tau\tau$ decay mode. The green, blue, and red lines denote the $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ final states, respectively. The results are shown for resonant HH production as a function of the resonance mass m_X and for nonresonant production as a function of the anomalous trilinear coupling modifier k_λ. (c) Purity of the final state assignment per generated $\tau\tau$ decay mode in the case of SM HH production. The fractions are normalized to unity for each horizontal line of the plot.

Fig. 4.15b, rejects the majority of these misidentified candidates. Two possible choices of the H $\to \tau\tau$ candidate are illustrated in the figure. One corresponds to the choice of the two most isolated τ_{h} candidates (red line) while the other corresponds to the choice of the pair having with the most highest-p_{T} isolated candidate (i.e. all the possible $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ pairs are formed and the isolation of the highest p_{T} candidate only is compared amongst the pairs). The former results in a better efficiency in the identification of the H $\to \tau\tau$ candidate, confirming that the τ_{h} isolation is the most powerful criterion to separate signal from jet background.



(a) Before τ_{h} isolation

(b) After τ_{h} isolation

Figure 4.15 – Efficiency in the selection of the H $\to \tau\tau$ lepton candidates in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state under the application of two different algorithms for the candidates choice (red and blue lines) as detailed in the text. The solid and dashed lines denote respectively all the selected events, and those where the selected lepton candidates correspond to a genuine τ_{h} decay. The efficiencies are shown separately before (a) and after (b) the application of the isolation criterion on the τ_{h} candidates. Their values are indicated as a function of the resonance mass hypothesized, and for SM HH production.

Following these observations, ambiguities are solved by selecting the τ_{h} candidates with the largest output of the multivariate isolation discriminant. In rare cases of equality of this value for two or more candidates, the one with the highest p_{T} is selected. One candidate is retained in the $\tau_{\mu}\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states, while two candidates are selected with this criterion in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state.

4.5.2 H $\to$ bb selection and event categorization

The two jets composing the H $\to$ bb candidate are chosen as those satisfying the preselections detailed in Section 4.4. In case more than two jets are found, those with the largest output of the b tagging discriminant are chosen. This selection criterion has been compared to other alternatives, namely the selection of the two highest p_{T} jets and of the pair with the invariant mass closer to the expected H boson mass peak. The former was observed to be less efficient in the selection of the correct jet candidates, while the latter

introduced a bias in the background distribution, making it artificially similar to signal events with a consequent reduction of the sensitivity.

Events are categorized according to the topology of the two selected jets to improve the sensitivity over the entire mass range studied for resonant production. The separation of the two b quarks produced in the Higgs boson decay depends on the Lorentz boost of the Higgs boson ($\gamma = E/m_{\mathrm{H}}$) approximately as $\Delta R \approx 2/\gamma$ [140]. Three topologies can consequently be observed experimentally. For low values of γ, the ΔR separation of the two quarks is larger than 0.8 and the two jets are spatially distant and separately reconstructed by the AK4 algorithm. At high values of γ the separation becomes smaller than 0.4 and the AK4 algorithm is unable to reconstruct two distinct jets. An intermediate regime finally exists for $0.4 < \Delta R < 0.8$ where the two jets partially overlap and can therefore be both reconstructed individually by the AK4 algorithm or together as a single jet by the AK8 algorithm. The signal processes explored in the resonant search, extending from $\mathrm{m_X} = 250\,\mathrm{GeV}$ up to $\mathrm{m_X} = 900\,\mathrm{GeV}$, correspond to the low and intermediate boost regimes, with jet overlap starting to occur significantly for $\mathrm{m_X} \gtrsim 700\,\mathrm{GeV}$.

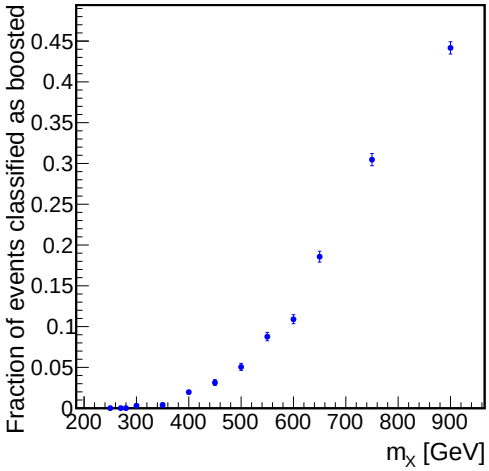
Signal events are consequently categorized into a "resolved" and a "boosted" category. Events in the boosted category must contain at least one AK8 jet of invariant mass $m_{\mathrm{AK8}} > 30\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 170\,\mathrm{GeV}$. The AK8 jet must have a substructure composed of at least two sub-jets, each spatially matched to the previously selected AK4 jets within an angular distance $\Delta R = 0.4$. Events failing such requirements are classified in the resolved category. The requirement on the AK8 jet invariant mass, estimated with the soft drop jet grooming algorithm, is applied to reject jets initiated by a single quark or gluon as well as those asymmetric decays where the grooming algorithm removes the decay particles of one of the two b quarks. The fraction of events that is categorized as "boosted" is shown in Figure 4.16 for resonant HH production as a function of $\mathrm{m_X}$ and for different nonresonant signal hypotheses. The fraction of events classified as boosted is sizeable for high mass resonances as well as for those nonresonant signals where large m_{HH} values are expected, such as for the shape benchmark 2. In the case of SM production, about 8% of the events are classified as boosted.

The usage of the AK8 jet reconstruction provides a clear separation from the $\mathrm{t\bar{t}}$ background, for which the two b jets are typically spatially distant. Only about 0.7% of the preselected $\mathrm{t\bar{t}}$ events contain an AK8 which mass, p_{T}, and substructure are compatible with the requirements applied in the boosted category.

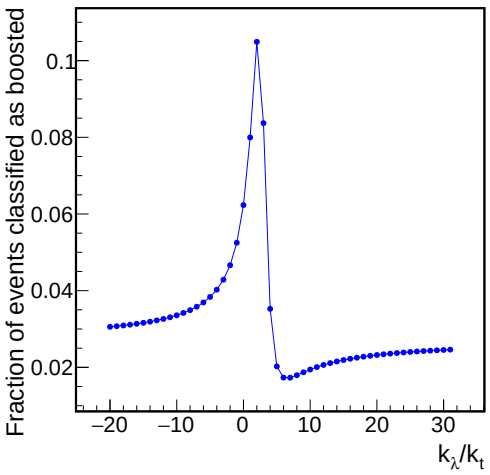
Additional b tagging criteria are applied to the events thus selected to reject background events containing light jets. The medium working point of the b tagging algorithm is used in the resolved category, and events are classified into a two b-tagged (2b) and one b-tagged (1b1j) jet event categories. The former represent a high purity category where the signal selection efficiency is about $0.65^2 \approx 40\%$, and is complemented by the latter that collects approximately $2 \cdot 0.65(1 - 0.65) \approx 45\%$ of the signal events. Given the smaller number of events expected in the boosted category, both selected jets are required to satisfy the loose working point of the b tag discriminant and no further classification is performed.

4.6 HH signal regions

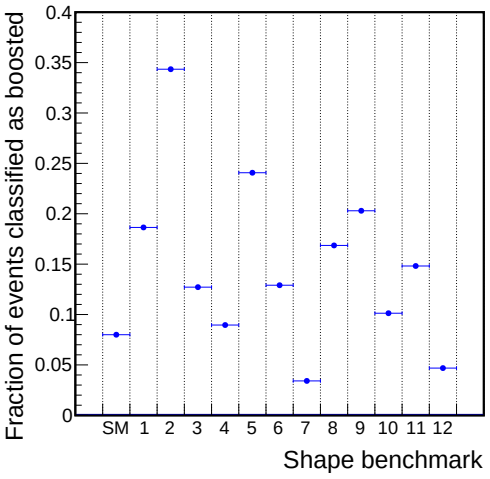
HH signal regions are defined to maximise the signal purity in the event categories defined above. A selection is performed on the invariant mass of the $\mathrm{H} \to \tau\tau$ and $\mathrm{H} \to \mathrm{bb}$



(a) Resonant HH signal



(b) Nonresonant k_λ/k_t



(c) Nonresonant shape benchmarks

Figure 4.16 – Fraction of events classified as boosted as a function of the resonance mass hypothesis for resonant HH production (a) and as a function of the ratio of the anomalous couplings k_λ/k_t (b) and for the shape benchmarks (c) for nonresonant production.

candidates, and their properties are combined with a BDT to reject the $t\bar{t}$ background contribution in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states of the resolved event categories.

4.6.1 Invariant mass selection

The invariant masses of the $\tau\tau$ ($m_{\tau\tau}$) and bb (m_{bb}) object pairs provide a clear signal signature. These variables are used to reduce the background contamination by retaining only events compatible with a HH $\to$ bb$\tau\tau$ decay.

The $m_{\tau\tau}$ value is reconstructed using the SVfit algorithm [142], a dynamical likelihood technique that combines the measured momenta of the visible τ decay products with the missing transverse momentum in the event. The kinematics of a $\tau \to \tau_h\nu_\tau$ decay is

described by six parameters, corresponding to the polar and azimuthal angle of the τ_{h} in the τ rest frame, the spatial boost vector of the τ lepton with respect to the laboratory rest frame, and the invariant mass of the τ decay products, corresponding to the τ_{h} candidate invariant mass. An additional parameter is needed for a $\tau \to \ell\nu_\ell\nu_\tau$ decay ($\ell = \mathrm{e}, \mu$) and it is usually chosen as the invariant mass of the two neutrinos. Only four of these parameters can be determined by measurements with the detector, and the two components of the missing transverse momentum can be used as further constraints in the hypothesis that $p_{\mathrm{T}}^{\mathrm{miss}}$ is mainly due to neutrinos from τ decays. As a result, there are two, three, or four unconstrained parameters, $\mathbf{q}$, in the decay of a $\tau\tau$ pair to $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $\tau_\ell\tau_{\mathrm{h}}$, and $\tau_\ell\tau_\ell$ respectively. The SVfit algorithm computes a conditional probability $P(m_{\tau\tau}|\mathbf{q},\mathbf{x})$ using the measured lepton momenta $\mathbf{x}$ and a τ kinematic decay model, and determines $m_{\tau\tau}$ by maximizing P under the marginalization on $\mathbf{q}$. The complete four-momentum p of the $\tau\tau$ pair can similarly be reconstructed by minimizing a function $P(p|\mathbf{q},\mathbf{x})$. The usage of the SVfit algorithm improves the resolution on $m_{\tau\tau}$ by about 30% over the visible invariant mass of the selected τ lepton decay products, as it can be observed in Figure 4.17. The shift of the Higgs boson mass peak position of about 7% below the 125 GeV value is due to the contribution to $p_{\mathrm{T}}^{\mathrm{miss}}$ of B mesons decays from b quarks hadronization. Although being typically small, as shown in Section 4.3.4, this contribution is not accounted for in the SVfit algorithm kinematic model. The good agreement between the MC simulation and the observed data in the $m_{\tau\tau}$ variable distribution is shown in Figure 4.18 for the three final states ($\tau_\mu\tau_{\mathrm{h}}$, $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$, $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$) and the three event categories (resolved 1b1j, resolved 2b, boosted) described before in this chapter.

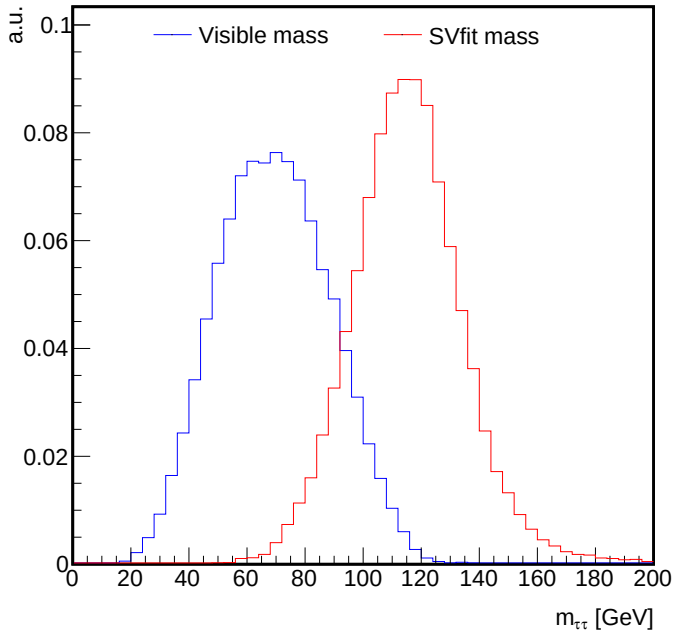


Figure 4.17 – Comparison of $m_{\tau\tau}$ determined as the visible invariant mass (blue) and using the SVfit algorithm (red) for the simulated SM HH $\to$ bb$\tau\tau$ events.

The m_{bb} value is computed as the AK4 jet pair invariant mass in the resolved categories and as the AK8 jet mass in the boosted category. Unmeasured energy of neutrinos from B meson decays causes the peak of the m_{bb} distribution to be shifted below the Higgs boson

mass of 125 GeV. The usage of regression techniques that exploit jet and event related observables to estimate such energy loss has been investigated. Despite an improvement of the m_{bb} resolution of about 15%, little impact on the sensitivity was observed given the changes induced in the background kinematic distributions, and therefore these techniques have not been retained.

The good modelling of m_{bb} in the three final states and in the three event categories is shown in Figure 4.19. The apparent deterioration of the m_{bb} resolution for signal events (solid lines in the figure) in the resolved 1b1j category with respect to the 2b one is due to the selection of a jet not issuing from a $\mathrm{H} \to \mathrm{bb}$ decay. In these cases, the genuine b jet is often outside the detector acceptance or below the p_{T} threshold applied.

For events classified in the resolved category, an invariant mass criterion is defined as:

$$\frac{\left(m_{\tau\tau} - 116\,\mathrm{GeV}\right)^2}{\left(35\,\mathrm{GeV}\right)^2} + \frac{\left(m_{\mathrm{bb}} - 111\,\mathrm{GeV}\right)^2}{\left(45\,\mathrm{GeV}\right)^2} < 1 \tag{4.6}$$

This selection corresponds to an ellipse in the ($m_{\tau\tau}$, m_{bb}) plane, centred about the reconstructed Higgs boson mass peak and whose semi-axes are optimized on the mass resolutions to give a signal efficiency of about 80% and a $\mathrm{t\bar{t}}$ background rejection around 85% in the 2b category. The distribution of SM HH signal and $\mathrm{t\bar{t}}$ background events in the $(m_{\tau\tau}, m_{\mathrm{bb}})$ plane for the three final states of the 2b resolved category is shown in Figure 4.20 together with the selected mass region. When compared with a "rectangular" selection, i.e. two independent requirements applied on $m_{\tau\tau}$ and m_{bb}, that gives the same signal efficiency, this optimized "elliptical" selection improves the background rejection by 5 to 10% depending on the process considered.

A different mass requirement is applied onto events in the boosted category as:

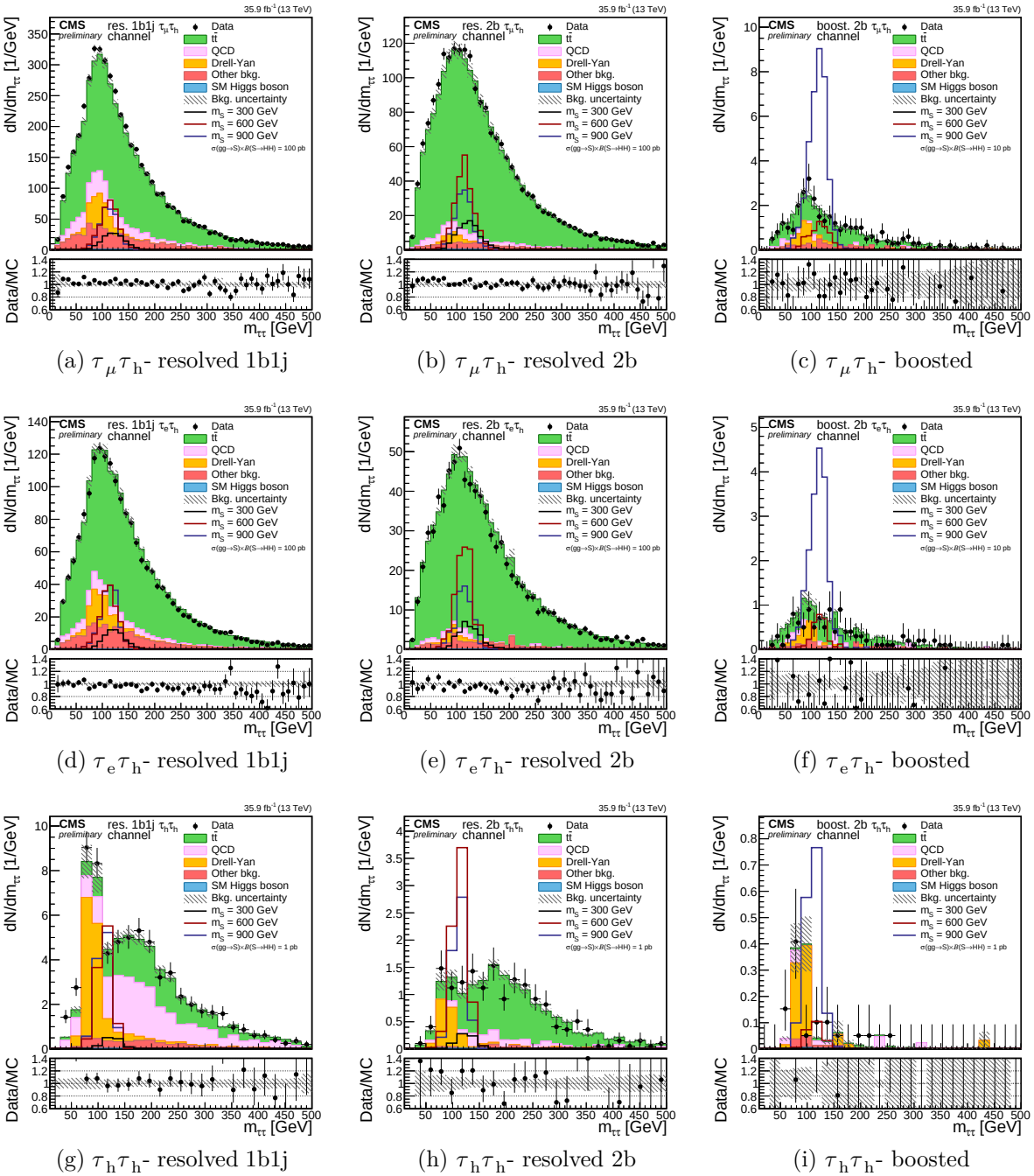
$$\begin{aligned} 80 < m_{\tau\tau} < 152\,\mathrm{GeV} \\ 90 < m_{\mathrm{bb}} < 160\,\mathrm{GeV} \end{aligned} \tag{4.7}$$

The values have been chosen as a function of the $m_{\tau\tau}$ and m_{bb} resolutions. The tighter selection on m_{bb} with respect to the resolved category stems from the better resolution achieved with the AK8 algorithm at high p_{T} regimes. The resulting efficiency on signal events is of about 85% for a total background reduction of 80%.

4.6.2 Multivariate method for $\mathrm{t\bar{t}}$ rejection

Different background contaminations are observed depending on the $\tau\tau$ final state and bb category considered. As illustrated in Figures 4.18 and 4.19, the two resolved categories of the $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states are similarly characterized by a large contribution from $\mathrm{t\bar{t}}$ production, while the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state has a larger contamination from multijet and Drell-Yan backgrounds, especially in the single b tag event category, and a total event yield almost two orders of magnitude smaller than the other two final states. The larger $\mathrm{t\bar{t}}$ contamination in the $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states is due to the direct production of a prompt muon or electron and the lower p_{T} thresholds applied. For all the boosted categories, the $\mathrm{t\bar{t}}$ production is suppressed and other backgrounds acquire a larger relative importance.

The above observations motivate the development of a dedicated technique to suppress the $\mathrm{t\bar{t}}$ production in resolved $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states. The choice made for this search

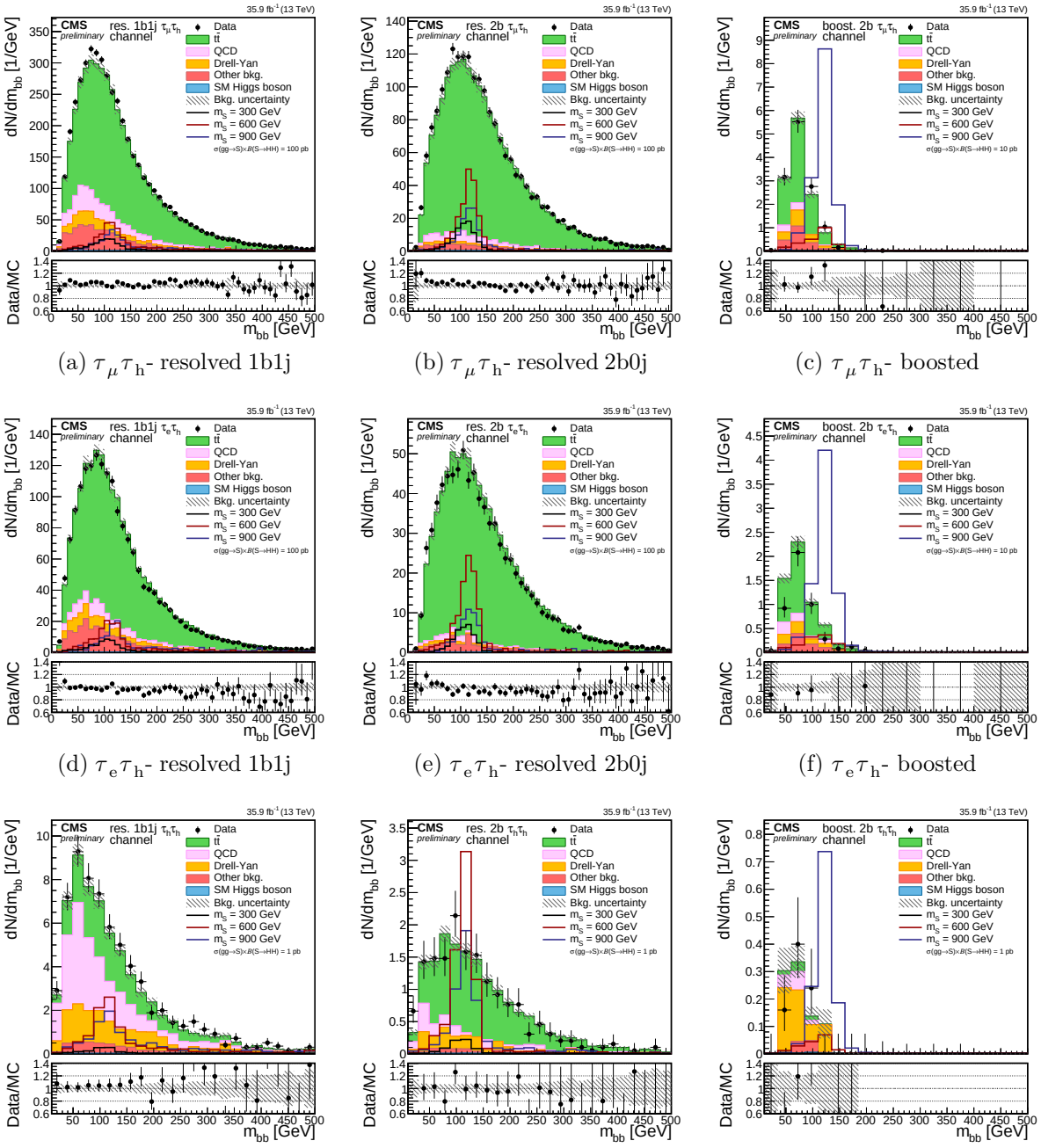


(a) $\tau_\mu\tau_h$- resolved 1b1j (b) $\tau_\mu\tau_h$- resolved 2b (c) $\tau_\mu\tau_h$- boosted

(d) $\tau_e\tau_h$- resolved 1b1j (e) $\tau_e\tau_h$- resolved 2b (f) $\tau_e\tau_h$- boosted

(g) $\tau_h\tau_h$- resolved 1b1j (h) $\tau_h\tau_h$- resolved 2b (i) $\tau_h\tau_h$- boosted

Figure 4.18 – Distribution of $m_{\tau\tau}$ computed with the SVfit algorithm in the $\tau_\mu\tau_h$ (first row), $\tau_e\tau_h$ (second row), and $\tau_h\tau_h$ (third row) final states for the resolved 1b1j (first column), resolved 2b0j (second column), and boosted (third column) event categories. The expectations for the a resonant spin-0 signal of mass m_S of 300, 600, and 900 GeV are also shown.

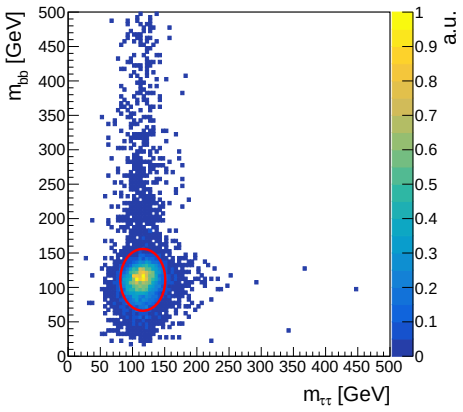


(a) $\tau_\mu\tau_h$- resolved 1b1j (b) $\tau_\mu\tau_h$- resolved 2b0j (c) $\tau_\mu\tau_h$- boosted

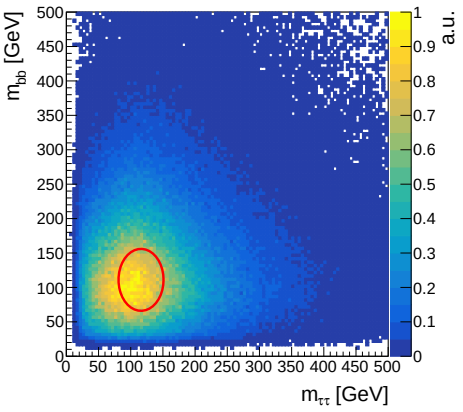
(d) $\tau_e\tau_h$- resolved 1b1j (e) $\tau_e\tau_h$- resolved 2b0j (f) $\tau_e\tau_h$- boosted

(g) $\tau_h\tau_h$- resolved 1b1j (h) $\tau_h\tau_h$- resolved 2b0j (i) $\tau_h\tau_h$- boosted

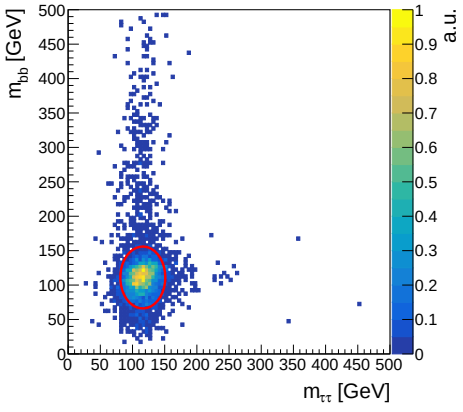
Figure 4.19 – Distribution of m_{bb} in the $\tau_\mu\tau_h$ (first row), $\tau_e\tau_h$ (second row), and $\tau_h\tau_h$ (third row) final states for the resolved 1b1j (first column), resolved 2b0j (second column) and boosted (third column) event categories. The expectations for the a resonant spin-0 signal of mass m_S of 300, 600, and 900 GeV are also shown.



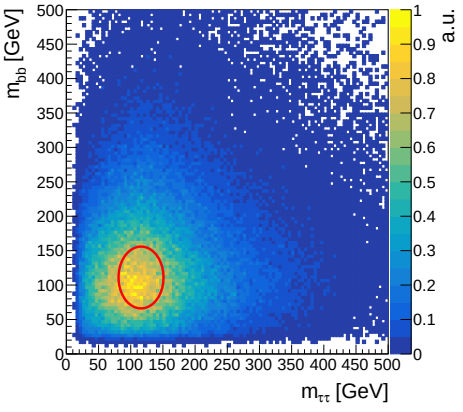
(a) SM HH signal, $\tau_\mu\tau_h$



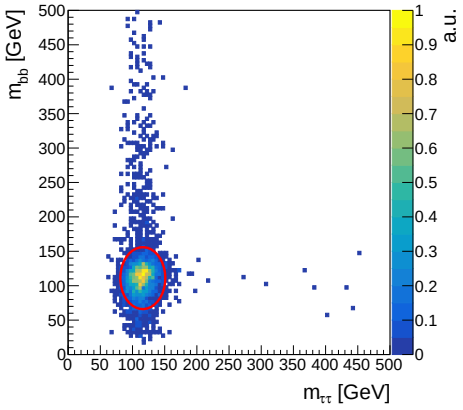
(b) $t\bar{t}$ background, $\tau_\mu\tau_h$



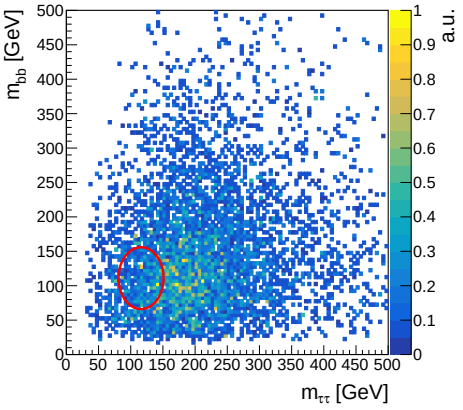
(c) SM HH signal, $\tau_e\tau_h$



(d) $t\bar{t}$ background, $\tau_e\tau_h$



(e) SM HH signal, $\tau_h\tau_h$



(f) $t\bar{t}$ background, $\tau_h\tau_h$

Figure 4.20 – 2D distributions of $m_{\tau\tau}, m_{bb}$ for the HH SM signal (left column) and for the $t\bar{t}$ background (right column) after the $\tau\tau$ and bb candidates selections and before the invariant mass requirements in the $\tau_\mu\tau_h$ (top row), $\tau_e\tau_h$ (central row), and $\tau_h\tau_h$ (bottom row) final states and resolved bb category. The red ellipse shows the region defined by the selection in formula (4.6).

is the development of a multivariate discriminant in the form of a BDT. The usage of a BDT method is ideally suited to combine multiple kinematic observables into a single discriminant and to fully exploit their correlated information.

The method must be designed to achieve a large background rejection with a signal efficiency that is at the same time high and uniform for the different signal processes studied. It should also be as little correlated as possible with the observables used to search for the presence of a signal that, as described in Section 6.2 of Chapter 6, are closely related to the invariant mass of the selected leptons and jets: a large correlation would reduce the separation achieved with these observables and, consequently, the sensitivity of the search.

To improve the sensitivity for the various signal processes studied, two separate trainings are performed for the low-mass (LM) and high-mass (HM) regions of the resonant search, defined by $\mathrm{m_X} \leq 350\,\mathrm{GeV}$ and $\mathrm{m_X} > 350\,\mathrm{GeV}$ respectively. The separation of the LM and HM regimes at $2m_\mathrm{t}$ has been determined as the value that better ensures the similarity of the distributions of the BDT input variables for different $\mathrm{m_X}$ hypotheses and, consequently, the higher uniformity of the BDT performance in these two regimes. The usage of two mass regimes is a compromise between the complexity and the performance of the search: while in principle the most efficient approach would be a dedicated training for each signal tested, this becomes soon very impractical and difficult to control and validate. For the nonresonant search, the LM training is used as it performs optimally even when compared to a dedicated training on nonresonant signals.

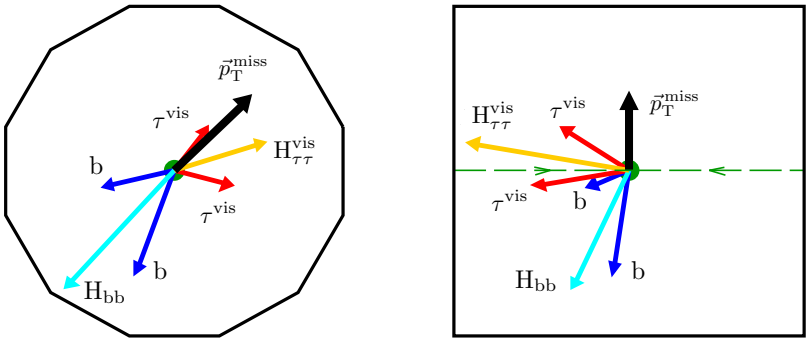
Input variables

The BDT input variables are chosen by considering the kinematic differences between signal and background events. Their choice is restricted to quantities related to the event topology, that have a moderate correlation with the discriminating observables used for signal extraction. For this reason, the transverse momenta of the selected lepton and jet candidates, although providing additional discrimination, are not considered. Furthermore, they largely depend on the signal hypothesis and are sensitive to higher-order effects that are not taken into account in the leading-order MC simulation.

A schematic illustration of a SM HH and of a $\mathrm{t\bar{t}}$ event is shown in Figure 4.21. It can be seen that the bb and the $\tau\tau$ systems are typically produced in opposite hemispheres of the detector (in the transverse plane) for the signal processes, as the two Higgs bosons are usually produced back-to-back, while this is not true for $\mathrm{t\bar{t}}$ events where the two top quarks are produced back-to-back and each top quark decays into a b and a lepton in association to neutrinos. The direction of the $\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$ vector in the transverse plane is also typically aligned to the H boson reconstructed from the $\tau\tau$ pair, as the missing momentum in the event mainly comes from neutrinos in τ decays, while it expected to be randomly distributed in the case of $\mathrm{t\bar{t}}$ production. For signal events, this results in a small angular separation $\Delta\varphi$ in the transverse plane between the $\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$ vector and the electron or muon momentum vector, $\vec{p}_\mathrm{T}^{\,\ell}$. Consequently, a variable with a large separation power is the transverse mass m_T of the lepton and τ_h candidates selected, that is defined as:

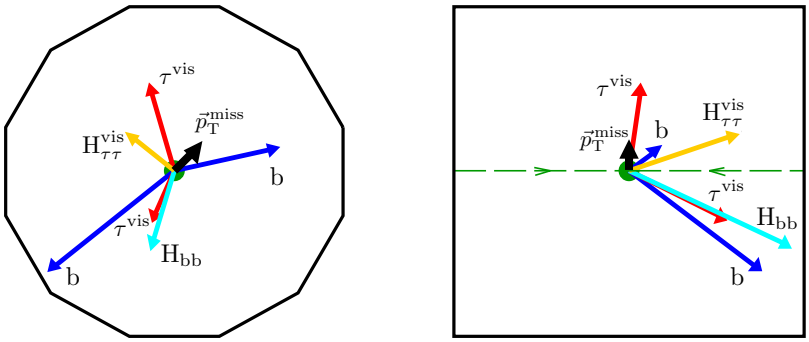
$$\begin{aligned} m_\mathrm{T}(\ell) &= \sqrt{\left(p_\mathrm{T}^{\mathrm{miss}} + p_\mathrm{T}^{\ell}\right)^2 - \left(\vec{p}_\mathrm{T}^{\,\mathrm{miss}} + \vec{p}_\mathrm{T}^{\,\ell}\right)^2} \\ &= \sqrt{2 p_\mathrm{T}^{\mathrm{miss}} p_\mathrm{T}^{\ell} \left(1 - \cos\Delta\varphi\right)} \end{aligned} \tag{4.8}$$

Signal events typically have small m_{T} values, while for $\mathrm{t\bar{t}}$ events its distribution peaks about the W boson mass because of the presence of a $\mathrm{W} \to \ell\nu_\ell$ decay.



(x, y) plane (z, y) plane

(a) HH event



(x, y) plane (z, y) plane

(b) $\mathrm{t\bar{t}}$ event

Figure 4.21 – Schematic representation of a typical HH (a) and $\mathrm{t\bar{t}}$ (b) event.

Following the above considerations, a final set of eight variables has been chosen as input to the LM and HM BDT trainings out of a larger set of inputs, that has been recursively pruned of the least discriminating and the most correlated variables. The input variables are listed and described in Table 4.3. The symbols $\mathrm{H}_{\tau\tau}$ and $\mathrm{H_{bb}}$ denote the four-momenta of the H boson candidates reconstructed from the $\tau\tau$ pair with the SVfit algorithm and from the selected jet pair, respectively, and ℓ denotes the selected electron or muon. Little correlation among the majority of the input variables is observed in both signal and background events, as shown in Figure 4.22.

Their distributions for the selected events are shown in Figures 4.23 and 4.24 for the $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states respectively. The MC simulation describes well the observed data. The expected distributions for nonresonant SM, $\mathrm{m_X} = 300\,\mathrm{GeV}$ (LM regime), and $\mathrm{m_X} = 600\,\mathrm{GeV}$ (HM regime) signals, normalized to an arbitrary cross section, are also shown for comparison.

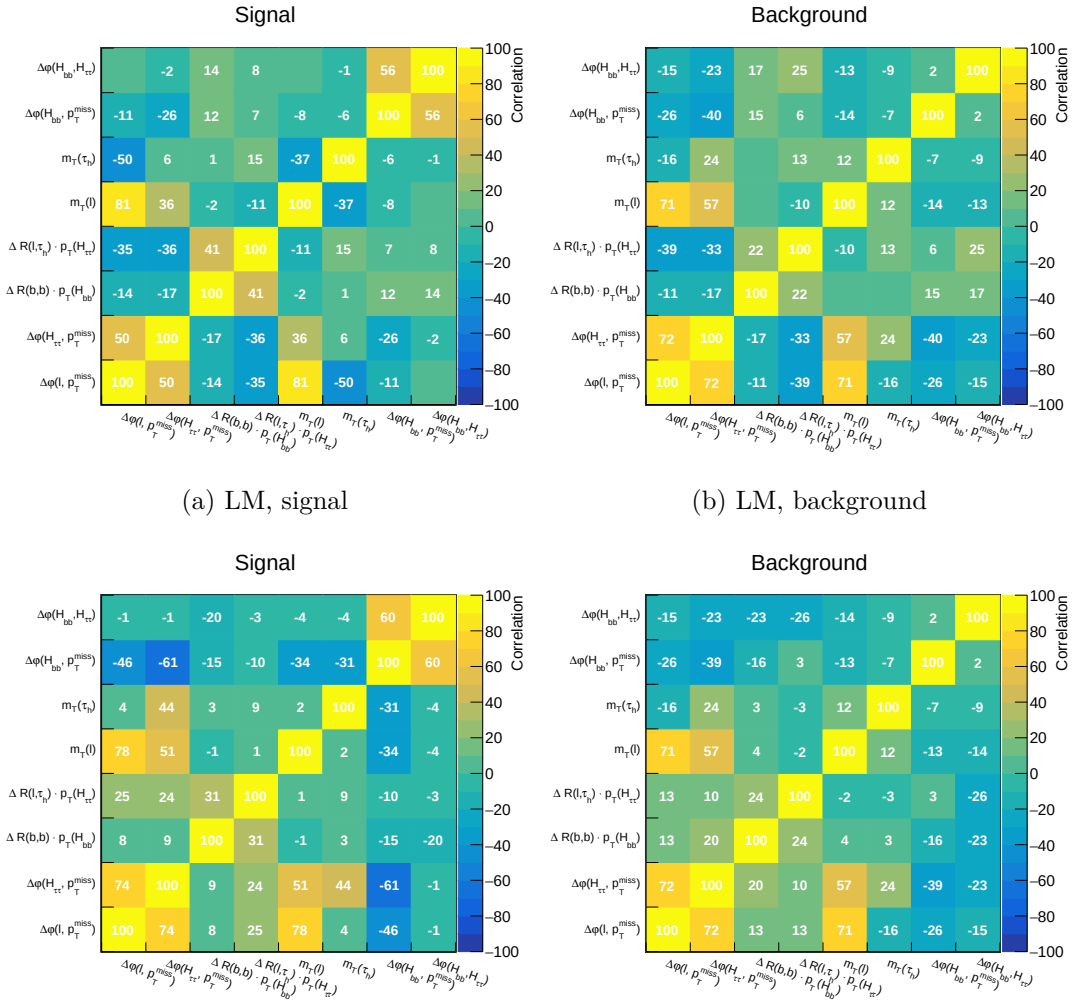
Training

The training procedure consists in the combination of the input variables into a unique value s, or "score", that separates background-like events (low s) from signal-like events

Variable	Description	LM	HM
$\Delta\varphi(\mathrm{H_{bb}}, \mathrm{H}_{\tau\tau})$	angular separation in the transverse plane between the two reconstructed Higgs bosons	✓	✓
$\Delta\varphi(\mathrm{H}_{\tau\tau}, p_{\mathrm{T}}^{\mathrm{miss}})$	angular separation in the transverse plane of the $\mathrm{H} \to \tau\tau$ candidate momentum and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vector	✓	✓
$\Delta\varphi(\mathrm{H_{bb}}, p_{\mathrm{T}}^{\mathrm{miss}})$	angular separation in the transverse plane of the $\mathrm{H} \to \mathrm{bb}$ candidate momentum and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vector	✓	✓
$\Delta\varphi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$	angular separation in the transverse plane of the reconstructed muon or electron and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vector	✓	✓
$m_{\mathrm{T}}(\ell)$	transverse mass of the selected muon or electron	✓	✓
$m_{\mathrm{T}}(\tau_{\mathrm{h}})$	transverse mass of the selected τ_{h}	✓	✓
$\Delta R(\mathrm{b}, \mathrm{b})$	separation of the two reconstructed jets	-	✓
$\Delta R(\ell, \tau_{\mathrm{h}})$	separation of the two reconstructed leptons	-	✓
$\Delta R(\mathrm{b}, \mathrm{b}) \cdot p_{\mathrm{T}}(\mathrm{H_{bb}})$	separation of the two reconstructed jets, corrected for the H boson p_{T}	✓	-
$\Delta R(\ell, \tau_{\mathrm{h}}) \cdot p_{\mathrm{T}}(\mathrm{H}_{\tau\tau})$	separation of the two reconstructed leptons, corrected for the H boson p_{T}	✓	-

Table 4.3 – Description of the input variables used for the low-mass (LM, $\mathrm{m_X} \leq 350\,\mathrm{GeV}$) and high-mass (HM, $\mathrm{m_X} > 350\,\mathrm{GeV}$) BDT trainings. The LM training is also used to search for nonresonant signals. $\mathrm{H}_{\tau\tau}$ and $\mathrm{H_{bb}}$ denote the four-momenta of the H boson candidates reconstructed from the $\tau\tau$ pair with the SVfit algorithm and from the selected jet pair, respectively, and ℓ denotes the selected electron or muon.

(high s). More specifically, the training consists in the creation of a "forest" of binary decision trees, weak classifiers that apply a sequence of selections on the input variables. Each selection consists of a threshold on the input variable that better separates signal and background events. All the variables in Table 4.3 are investigated, and the figure of merit to determine the optimal one and the corresponding threshold is the Gini index, defined as $G = p(1-p)$, with p indicating the purity, i.e. the fraction of signal events correctly classified. The number of successive selections applied by each binary tree, or "splitting level", is set to three in this training. Each individual tree alone provides little separation between signal and background: the "learning" phase consists in creating a large collection of tree in the forest (500 trees are used for this training), and making them aware of the events that were erroneously classified by the previous iteration. In practice, this is done by increasing the relative importance of events erroneously classified, which determines how the method learns from the events themselves. The rate of erroneous



(a) LM, signal

(b) LM, background

(c) HM, signal

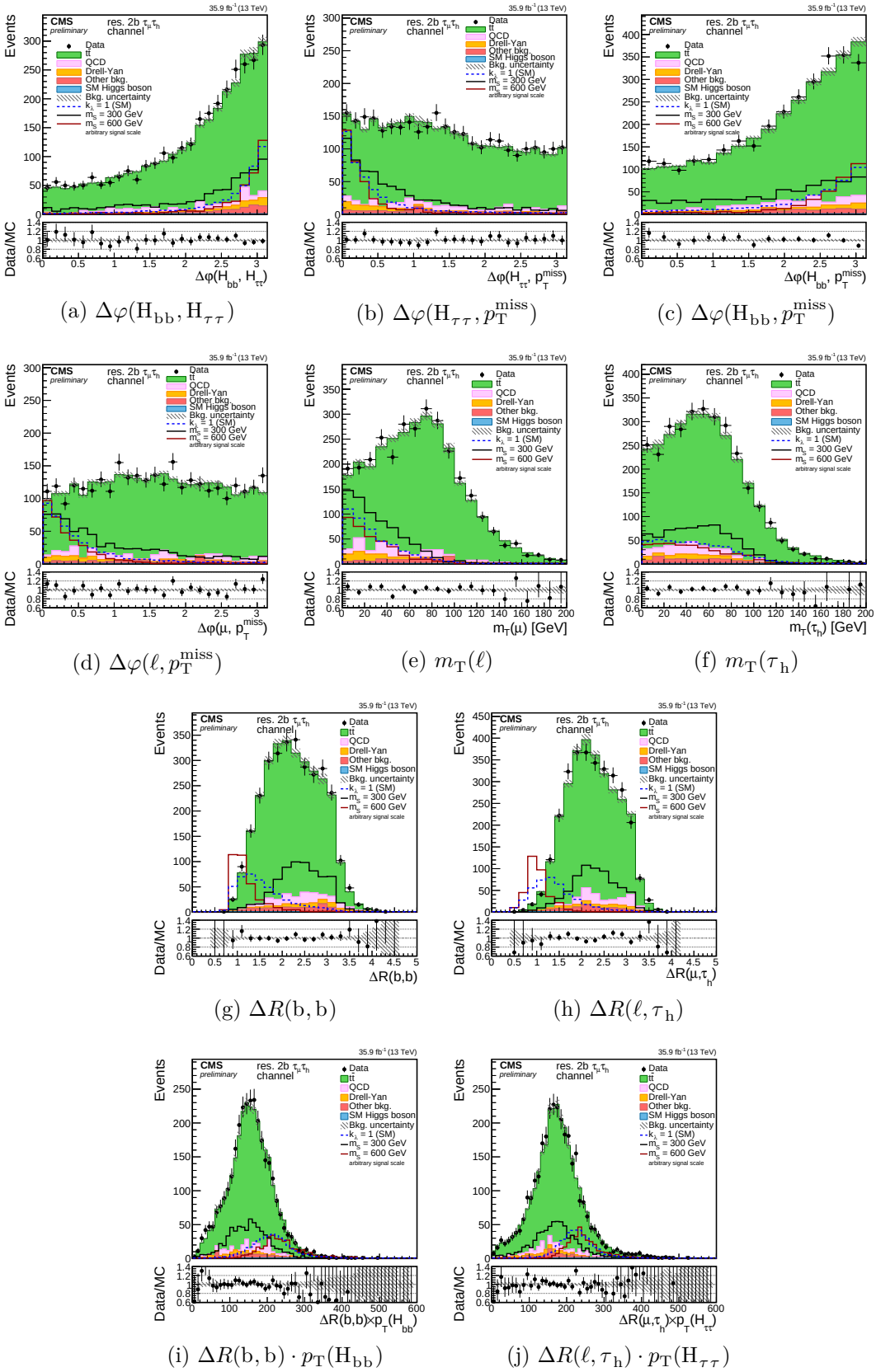
(d) HM, background

Figure 4.22 – Linear correlation coefficients of the BDT input variables for the LM training (top row) and HM training (bottom row) for signal (left column) and background (right column) training samples.

classification is estimated through a loss function that is minimized numerically with a gradient technique [135] using the TMVA software [136].

The importance of each variable x_i in the training, or "ranking", is determined from the number of times x_i is used for the splitting of a binary tree, and weighting each occurrence by the square of the gain achieved in the separation and the number of events classified. The ranking thus computed is shown in Table 4.4.

The output s of the method thus trained is computed as the weighted sum of all the binary trees in the forest. Some care is required to ensure that the method performs properly and that statistical fluctuations in the input events are not regarded as real differences, an effect known as "overtraining". To minimize such effect, the number of events available for the training has been maximized by combining events in the $\tau_\mu\tau_\mathrm{h}$ and $\tau_\mathrm{e}\tau_\mathrm{h}$ final states, after verifying that the corresponding distributions of the input variables are consistent, as seen from Figures 4.23 and 4.24. The presence of overtraining is checked by dividing the input signal and background datasets in two equal size "training"



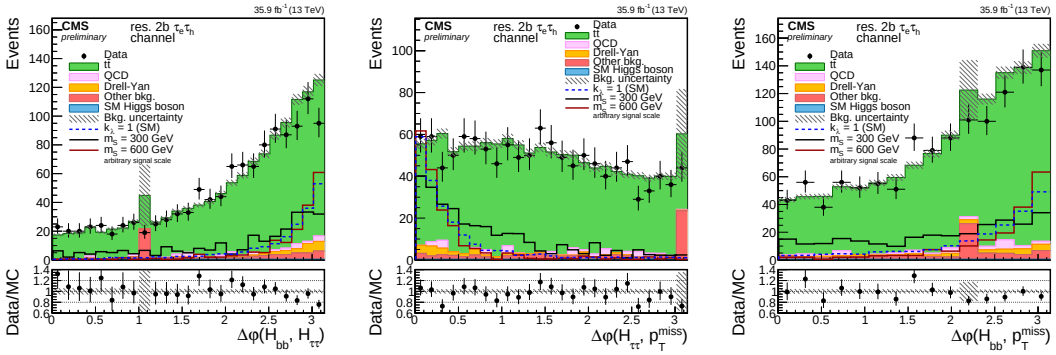
(a) $\Delta\varphi(\mathrm{H_{bb}}, \mathrm{H}_{\tau\tau})$ (b) $\Delta\varphi(\mathrm{H}_{\tau\tau}, p_\mathrm{T}^\mathrm{miss})$ (c) $\Delta\varphi(\mathrm{H_{bb}}, p_\mathrm{T}^\mathrm{miss})$

(d) $\Delta\varphi(\ell, p_\mathrm{T}^\mathrm{miss})$ (e) $m_\mathrm{T}(\ell)$ (f) $m_\mathrm{T}(\tau_\mathrm{h})$

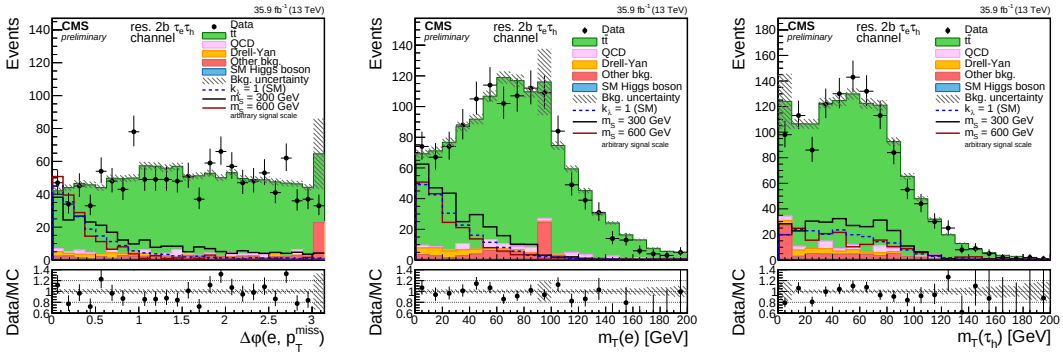
(g) $\Delta R(\mathrm{b},\mathrm{b})$ (h) $\Delta R(\ell, \tau_\mathrm{h})$

(i) $\Delta R(\mathrm{b},\mathrm{b}) \cdot p_\mathrm{T}(\mathrm{H_{bb}})$ (j) $\Delta R(\ell, \tau_\mathrm{h}) \cdot p_\mathrm{T}(\mathrm{H}_{\tau\tau})$

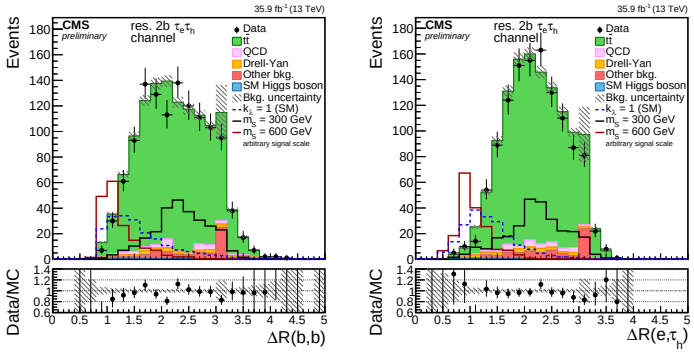
Figure 4.23 – Distribution of the BDT input variables before the application of the BDT discriminant selection for the $\tau_\mu\tau_\mathrm{h}$ final state. Points with error bars represent the data, shaded histograms represent the backgrounds and solid lines denote the signal expectations, normalized to an arbitrary cross section to be clearly visible in the plots.



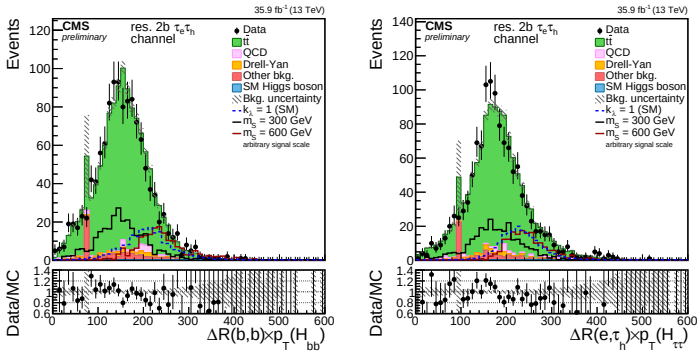
(a) $\Delta\varphi(\mathrm{H_{bb}}, \mathrm{H}_{\tau\tau})$ (b) $\Delta\varphi(\mathrm{H}_{\tau\tau}, p_{\mathrm{T}}^{\mathrm{miss}})$ (c) $\Delta\varphi(\mathrm{H_{bb}}, p_{\mathrm{T}}^{\mathrm{miss}})$



(d) $\Delta\varphi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ (e) $m_{\mathrm{T}}(\ell)$ (f) $m_{\mathrm{T}}(\tau_{\mathrm{h}})$



(g) $\Delta R(\mathrm{b}, \mathrm{b})$ (h) $\Delta R(\ell, \tau_{\mathrm{h}})$



(i) $\Delta R(\mathrm{b}, \mathrm{b}) \cdot p_{\mathrm{T}}(\mathrm{H_{bb}})$ (j) $\Delta R(\ell, \tau_{\mathrm{h}}) \cdot p_{\mathrm{T}}(\mathrm{H}_{\tau\tau})$

Figure 4.24 – Distribution of the BDT input variables before the application of the BDT discriminant selection for the $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final state. Points with error bars represent the data, shaded histograms represent the backgrounds and solid lines denote the signal expectations, normalized to an arbitrary cross section to be clearly visible in the plots.

LM BDT			HM BDT		
Rank	Variable	Importance	Rank	Variable	Importance
1	$m_{\mathrm{T}}(\ell)$	0.19	1	$\Delta R(\ell, \tau_{\mathrm{h}})$	0.29
2	$\Delta R(\ell, \tau_{\mathrm{h}}) \cdot p_{\mathrm{T}}(\mathrm{H}_{\tau\tau})$	0.15	2	$\Delta R(\mathrm{b}, \mathrm{b})$	0.19
3	$\Delta\varphi(\mathrm{H}_{\tau\tau}, p_{\mathrm{T}}^{\mathrm{miss}})$	0.14	3	$\Delta\varphi(\mathrm{H}_{\tau\tau}, p_{\mathrm{T}}^{\mathrm{miss}})$	0.14
4	$\Delta R(\mathrm{b}, \mathrm{b}) \cdot p_{\mathrm{T}}(\mathrm{H}_{\mathrm{bb}})$	0.12	4	$m_{\mathrm{T}}(\ell)$	0.13
5	$\Delta\varphi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$	0.11	5	$\Delta\varphi(\mathrm{H}_{\mathrm{bb}}, \mathrm{H}_{\tau\tau})$	0.07
6	$\Delta\varphi(\mathrm{H}_{\mathrm{bb}}, \mathrm{H}_{\tau\tau})$	0.11	6	$\Delta\varphi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$	0.07
7	$m_{\mathrm{T}}(\tau_{\mathrm{h}})$	0.10	7	$\Delta\varphi(\mathrm{H}_{\mathrm{bb}}, p_{\mathrm{T}}^{\mathrm{miss}})$	0.06
8	$\Delta\varphi(\mathrm{H}_{\mathrm{bb}}, p_{\mathrm{T}}^{\mathrm{miss}})$	0.09	8	$m_{\mathrm{T}}(\tau_{\mathrm{h}})$	0.05

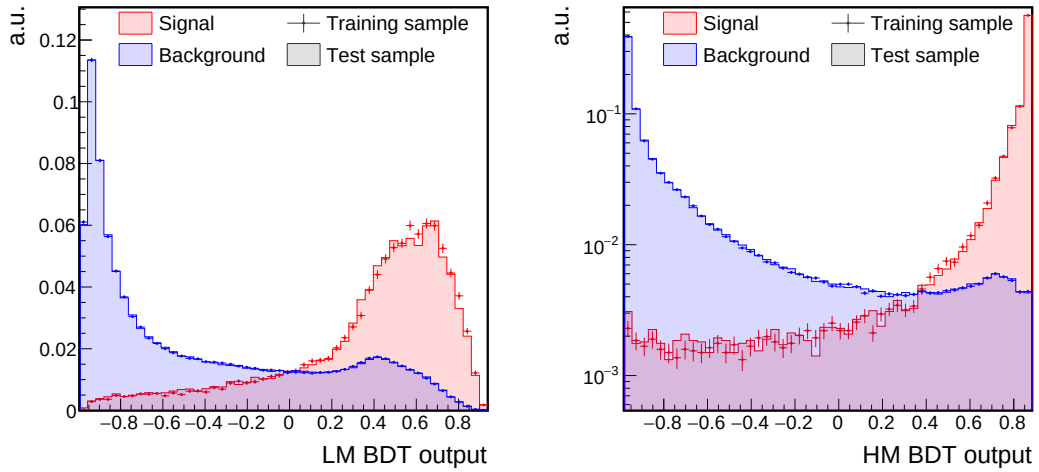
Table 4.4 – Ranking of the BDT input variables for the LM (left) and HM (right) trainings.

and "test" samples, the former being used to train the method and the latter to verify its output. During the training of each tree, only a subset of the events randomly chosen from the training sample is used to further reduce overtaining effect ("bagging" technique). A comparison of the LM and HM BDT outputs for the training and test samples is shown in Figure 4.25, where the agreement of the distributions is be observed.

All the signal samples in the LM and HM regimes are combined in the training to increase the number of events available and to cover the entire phase space of the signal processes in the two regions. This variety of inputs prevents the training to be fully optimized for a single $\mathrm{m_X}$ value with a suboptimal performance for other resonance masses. An arbitrary choice is made to normalize each sample in the training according to the radion cross section production times the branching fraction of its decay to HH [50], which decreases as a function of $\mathrm{m_X}$ and stabilizes the performance for different signal hypotheses. This has been compared to an alternative choice of normalizing all the signal processes to the same cross section, which resulted in a worse performance for low $\mathrm{m_X}$ samples without a significant gain at higher mass values. The validity of the training on the combined set of samples is verified by comparing the performance against a training on a specific $\mathrm{m_X}$ value. The comparison is illustrated in Figure 4.26 for $\mathrm{m_X} = 300\,\mathrm{GeV}$. For a given signal efficiency, the training performed exclusively on $\mathrm{m_X} = 300\,\mathrm{GeV}$ achieves a background rejection that is only a few percent better than the combined one. The BDT methods thus developed are reasonably close to the optimal discriminant and can be applied with success to a large variety of signals.

Performance

The performance of the methods for resonant and nonresonant signals are summarized in Figure 4.27, where the curves show the signal efficiency and background rejection as a function of the selection applied to the BDT output. Curves for resonant signals for $\mathrm{m_X}$ values between 270 and 900 GeV are reported for both LM and HM trainings to compare their performance outside the respective mass training range. The overall performance of the BDT methods is largely improved with respect to the selection $m_{\mathrm{T}} < 30\,\mathrm{GeV}$ (star marker in the figure) that has been used in the Run I search for low mass HH resonant production [143]. It can also be observed how the LM BDT method efficiently



(a) LM training (b) HM training

Figure 4.25 – Overtraining tests for LM (a) and HM (b) BDT trainings. The red and blue curves denote the signal and background events, respectively. Points represent the training sample and solid histograms indicate the test sample.

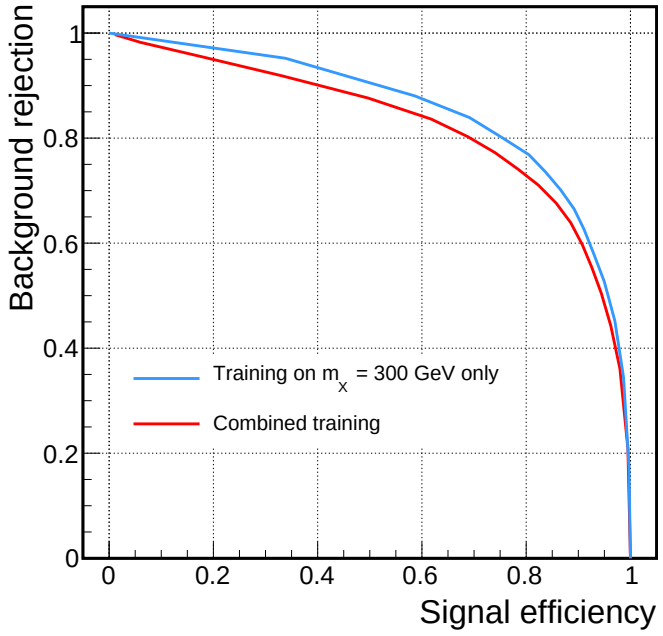
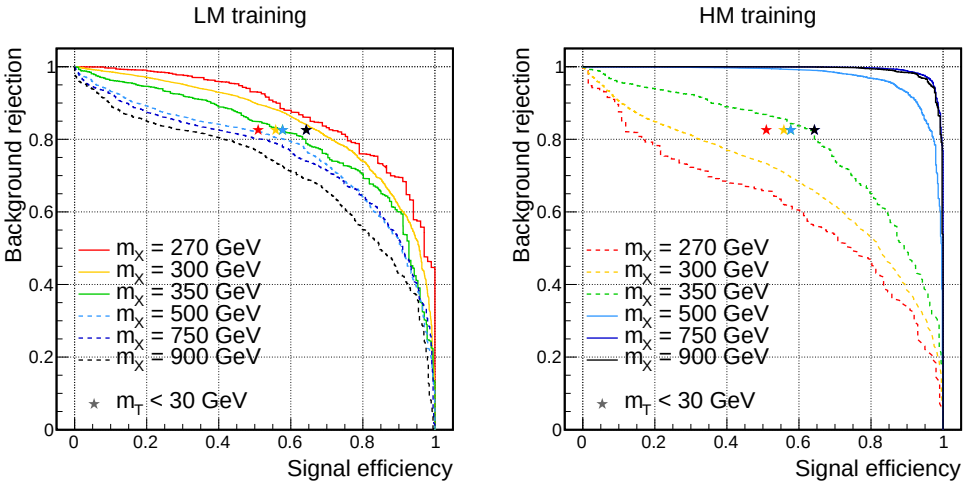


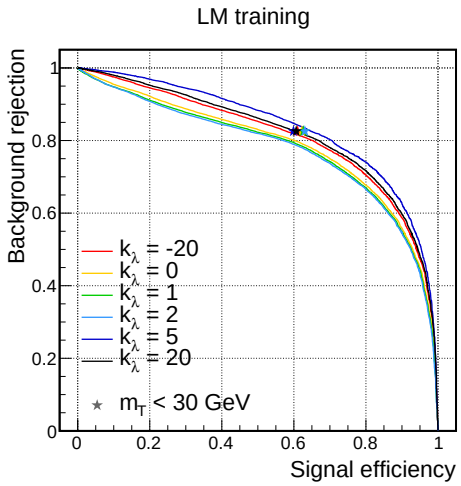
Figure 4.26 – Comparison of the signal efficiency versus background rejection for $m_X = 300\,\text{GeV}$. The red line represents the training performed on all the signal samples with $m_X \leq 350\,\text{GeV}$ (LM training in the text) and the blue line indicates a training on the $m_X = 300\,\text{GeV}$ signal only.

performs for the nonresonant HH signals. This is due to the similar kinematic properties of nonresonant production and the combination of resonant production signals in the LM region.



(a) Resonant signals - LM

(b) Resonant signals - HM



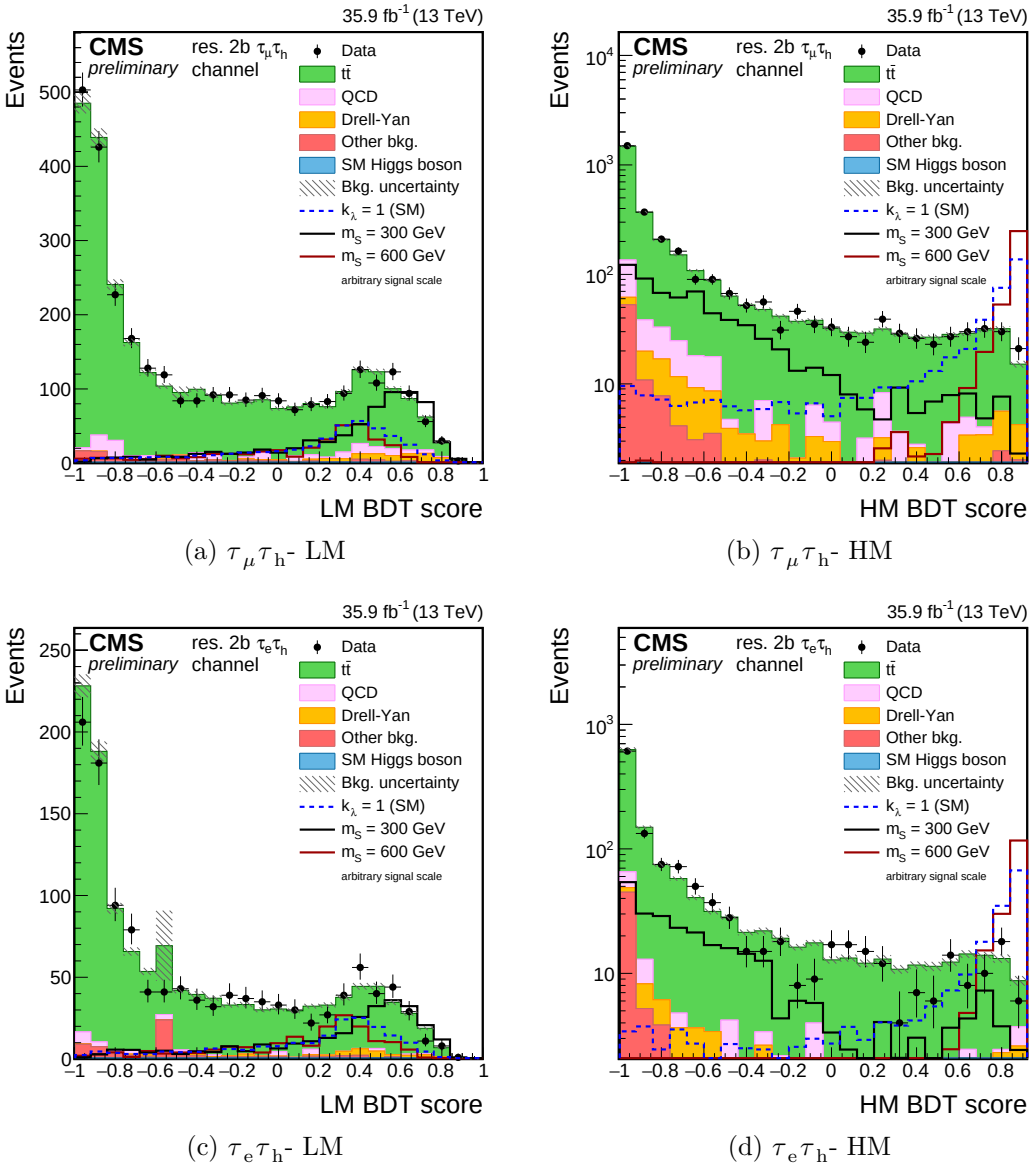
(c) Nonresonant signals - LM

Figure 4.27 – $t\bar{t}$ rejection as a function of signal efficiency for resonant signals for LM (a) and HM (b) trainings and for nonresonant signals (c). Solid and dashed lines indicate m_X values respectively inside and outside the mass training range of the BDT. The star markers denote the performance of a selection $m_T < 30\,\text{GeV}$.

The distributions of the BDT outputs in $\tau_\mu\tau_h$ and $\tau_e\tau_h$ events classified in the 2b resolved category and satisfying the invariant mass selections are shown in Figure 4.28.

Choice of the working point

The selection on the BDT output s has been determined by comparing the sensitivity of the search using the observables that are described in Section 6.2 of Chapter 6, expressed in terms of the 95% CL upper limits on the production cross section times branching fraction. The gain achieved with the usage of the BDT discriminant is defined as the



(a) $\tau_\mu\tau_h$- LM (b) $\tau_\mu\tau_h$- HM

(c) $\tau_e\tau_h$- LM (d) $\tau_e\tau_h$- HM

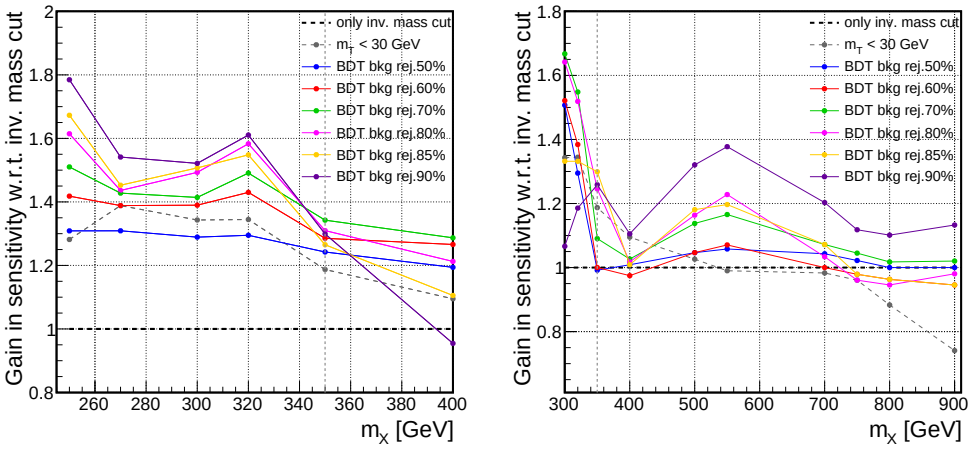
Figure 4.28 – BDT output distribution for the $\tau_\mu\tau_h$ final state (upper row) and the $\tau_e\tau_h$ final state (bottom row). The LM BDT output is shown on the left and the HM BDT output is shown on the right.

ratio of the upper limits obtained in absence and in presence of such requirement, so that a gain > 1 indicates a better sensitivity.

The expected gain as a function of the resonance mass m_X and of the anomalous trilinear coupling k_λ are shown in Figure 4.29 for the LM and HM BDT, where different selections on the BDT score are compared. These results show that the best sensitivity to resonant production is achieved with a tight selection rejecting approximately 90% of the $t\bar{t}$ background for both the LM and HM regimes. The corresponding signal efficiency ranges between 65% and 95% depending on the mass hypothesis. The selection applied on the BDT output corresponds to $s > 0.477$ and $s > 0.0188$ for the LM and HM regimes respectively. For nonresonant production, the best sensitivity is achieved with a selection that rejects approximately 70% of the $t\bar{t}$ background for a signal efficiency of about 80%,

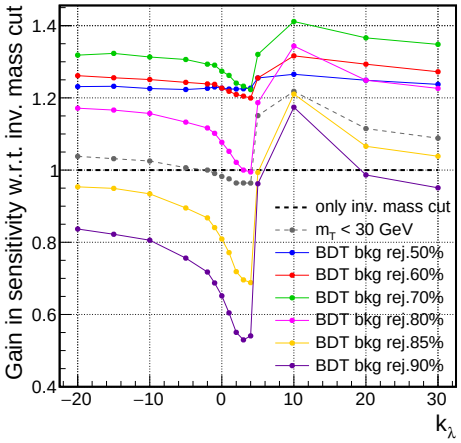
corresponding to a selection $s > -0.0764$ on the LM BDT output. The difference with respect to the resonant case is due to the signal kinematics as well as to the usage of a different observable for the signal extraction.

The gains achieved with these optimal BDT selections are compared for the resonant and nonresonant searches in Figure 4.30. A clear transition between the LM and HM regimes at $m_X = 350\,\text{GeV}$, corresponding to the change of the BDT training region, is observed. The performance of the BDT is also shown to be superior to the one of a $m_T < 30\,\text{GeV}$ selection, and the combination of LM and HM trainings results in a significant improvement in the sensitivity to resonant production over the entire mass range studied, with a gain ranging between 10 and 80%. In the case of nonresonant production, the usage of the BDT discriminant improves the sensitivity between 20 and 40% depending on the anomalous coupling hypothesis studied.



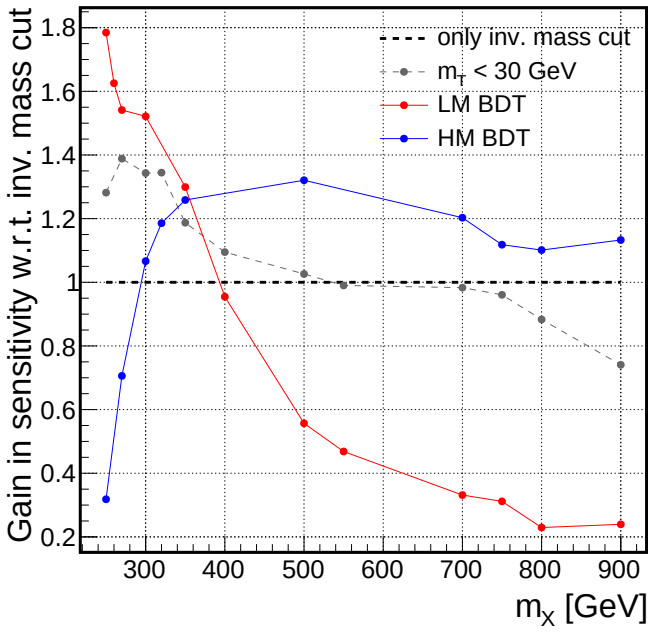
(a) Resonant signals - LM

(b) Resonant signals - HM

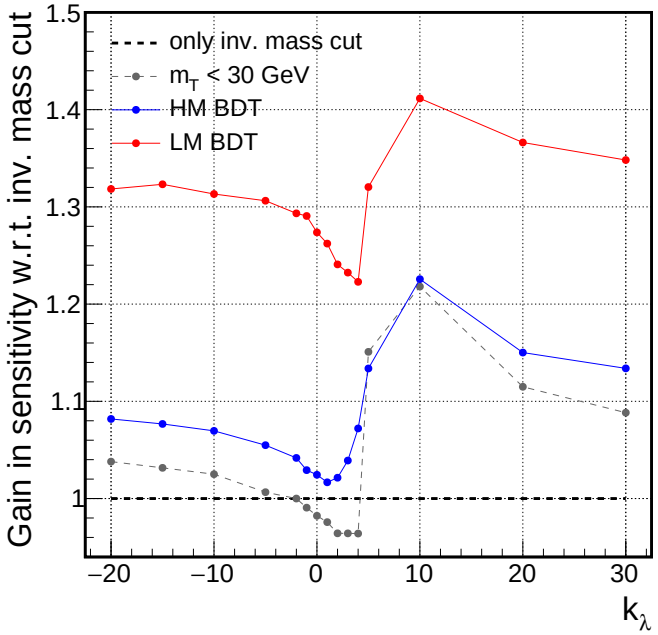


(c) Nonresonant signals - LM

Figure 4.29 – Comparison of the expected gains in the sensitivity in the 2b0j resolved $\tau_\mu\tau_h$ category for different selections on the BDT score. Resonant signals for the LM (a) and HM (b) trainings and nonresonant signals (c) for the LM training are shown.



(a) Resonant signals



(b) Nonresonant signals

Figure 4.30 – Expected gain in the sensitivity for the 2b resolved $\tau_\mu\tau_h$ category for different selections applied after the invariant mass requirements. The gain is defined as the ratio of the 95% confidence level upper limits obtained in absence and presence of the corresponding selection, so that a gain > 1 corresponds to an improved sensitivity. (a): resonant search, where a clear transition between the LM and HM BDT optimal performance around 350 GeV is observed. (b): nonresonant search.

Effect of the BDT on the observables

It is interesting to observe the effect of the application of the BDT discriminant on the observables used for the signal extraction. Without entering in the details of their definition, that are explained in Section 6.2 of Chapter 6, these observables reconstruct the $\tau\tau$ and bb four-body mass (m_{HH}^{KinFit}) or generalize the transverse mass for a decay with multiple invisible particles (m_{T2}). Resonant signals appear as a "bump" in the distribution of the former while nonresonant ones enhance the tails of the latter. The impact of the BDT method on their distributions is thus important to quantify because directly related to the sensitivity of the search.

The distributions of m_{HH}^{KinFit} and m_{T2} before and after the BDT selections are illustrated in Figure 4.31 for the $\tau_\mu\tau_h$ final state and 2b resolved event category. In the resonant case the LM method tends to suppress high mass contributions and, inversely, the HM BDT selection removes most of the low mass events, with a significant improvement in the signal to background ratio in the regions of interest. This is an important observation and indicates that the impact of the method cannot be determined only from the signal and background efficiency curves previously shown, and justifies the detailed optimization based on the sensitivity of the search. Similarly, the application of the LM training to the nonresonant search reduces the $t\bar{t}$ contamination at low m_{T2} and has small impact on the signal distribution. The signal to background ratio is improved in the sensitive region $m_{T2} \gtrsim 150\,\text{GeV}$ with a consequent increase in the sensitivity.

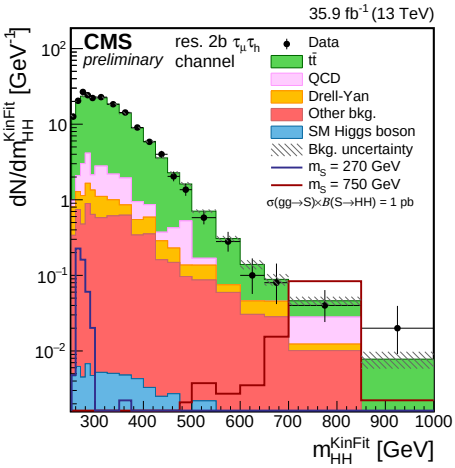
4.7 Selection efficiency

This section summarizes the criteria applied to select the collision events and their efficiency for the signal processes. These criteria correspond to the trigger requirements described in Section 4.2, $H \to \tau\tau$ and $H \to bb$ object preselections described in Sections 4.3 and 4.4 and summarized in Tables 4.5 and 4.6, and the invariant mass requirements and multivariate discriminant for resolved $\tau_\mu\tau_h$ and $\tau_e\tau_h$ event categories described in Section 4.6. For every final state considered $\tau_i\tau_h$ ($i = \mu, e, h$), the efficiencies are defined with respect to an inclusive production of $HH \to bb\tau\tau \to \tau_i\tau_h$ events.

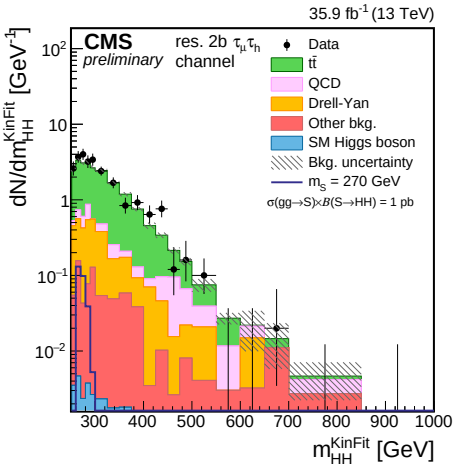
The efficiencies of the selections are studied for the different signals investigated in the search. Spin-0 and spin-2 resonant signals are shown in Figure 4.32. Nonresonant signals for anomalous k_λ/k_t couplings and for the shape benchmarks are shown in Figure 4.33.

The two effects that mostly limit the efficiency are the trigger and the selection of the $\tau\tau$ candidate pair. The former is mainly due to the limited acceptance of the p_T selections applied at trigger level, and is particularly severe for the $\tau_h\tau_h$ final state because of the higher thresholds. For the same reason the trigger efficiency increases a function of m_X and for nonresonant signals with a harder Higgs boson p_T spectrum, such as $0 < k_\lambda/k_t < 2$ or the shape benchmark 2. The latter source of inefficiency is mainly due to the additional p_T selection and isolation criteria applied on the reconstructed τ_h objects in the $H \to \tau\tau$ candidate selection. It should be recalled that the isolation requirements, although apparently inefficient, were chosen as those giving the best sensitivity of the analysis and are required to suppress the very large multijet background.

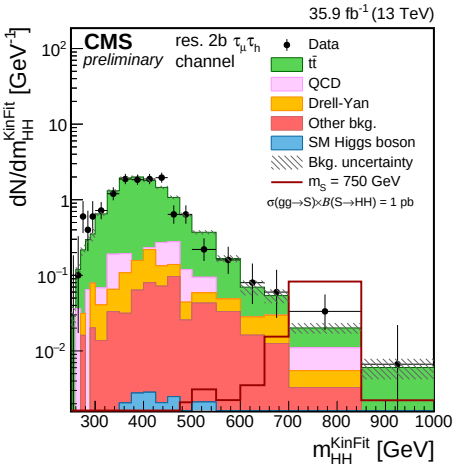
The efficiency curves for the b tagging criteria and subsequent requirements include the three bb event categories. The large efficiency reduction associated to the invariant mass selection is thus mainly due to 1b1j resolved events where the untagged jet does not



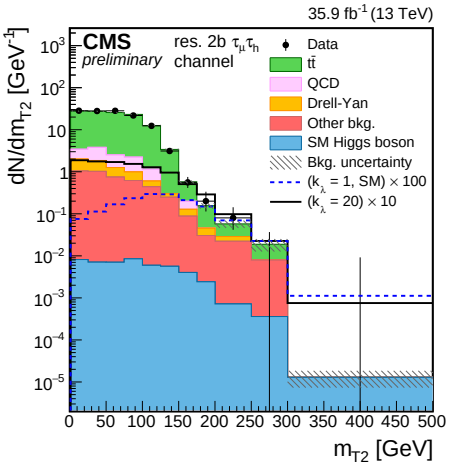
(a) m_{HH}^{KinFit}, before BDT



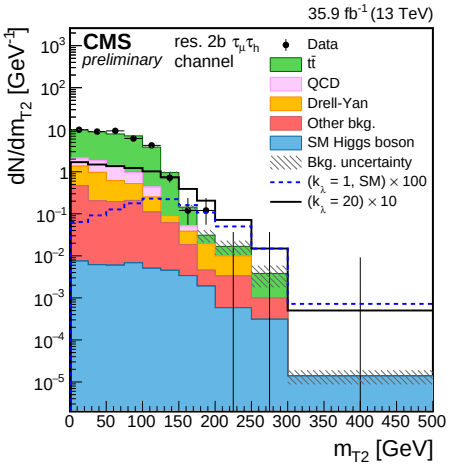
(b) m_{HH}^{KinFit}, after LM BDT



(c) m_{HH}^{KinFit}, after HM BDT



(d) m_{T2}, before BDT



(e) m_{T2}, after BDT

Figure 4.31 – Distribution of m_{HH}^{KinFit} before any BDT requirement (a) and after the LM (b) and HM (c) BDT requirements, and distribution of m_{T2} before (a) and after (e) the BDT requirement.

$\tau_\mu\tau_\mathrm{h}$	
μ	Tight ID Vertex $\Delta_{xy} < 0.045\,\mathrm{cm}$, $\Delta_z < 0.2\,\mathrm{cm}$ Isolation $\mathcal{I}_\mu^\mathrm{rel} < 0.15$ $p_\mathrm{T} > 23\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.1$
τ_h	Decay mode identification Vertex $\Delta_{xy} < 0.045\,\mathrm{cm}$, $\Delta_z < 0.2\,\mathrm{cm}$ $p_\mathrm{T} > 20\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.3$ Very loose anti-e Tight anti-μ Medium MVA isolation ($\star$)
Pair	$\Delta R(\mu, \tau_\mathrm{h}) > 0.1$ Opposite electric charge ($\star$)

$\tau_\mathrm{e}\tau_\mathrm{h}$	
e	80% eff. MVA ID Vertex $\Delta_{xy} < 0.045\,\mathrm{cm}$, $\Delta_z < 0.2\,\mathrm{cm}$ Isolation $\mathcal{I}_\mathrm{e}^\mathrm{rel} < 0.1$ $p_\mathrm{T} > 27\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.1$
τ_h	Decay mode identification Vertex $\Delta_{xy} < 0.045\,\mathrm{cm}$, $\Delta_z < 0.2\,\mathrm{cm}$ $p_\mathrm{T} > 20\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.3$ Tight anti-e Loose anti-μ Medium MVA isolation ($\star$)
Pair	$\Delta R(\mathrm{e}, \tau_\mathrm{h}) > 0.1$ Opposite electric charge ($\star$)

$\tau_\mathrm{h}\tau_\mathrm{h}$	
Both τ_h	Decay mode identification Vertex $\Delta_{xy} < 0.045\,\mathrm{cm}$, $\Delta_z < 0.2\,\mathrm{cm}$ Medium MVA isolation ($\star$) $p_\mathrm{T} > 40\,\mathrm{GeV}$ $p_\mathrm{T} > 45\,\mathrm{GeV}$ ($\star$) $\lvert\eta\rvert < 2.1$ Very loose anti-e Loose anti-μ
Pair	$\Delta R(\tau_\mathrm{h}, \tau_\mathrm{h}) > 0.1$ Opposite electric charge ($\star$)

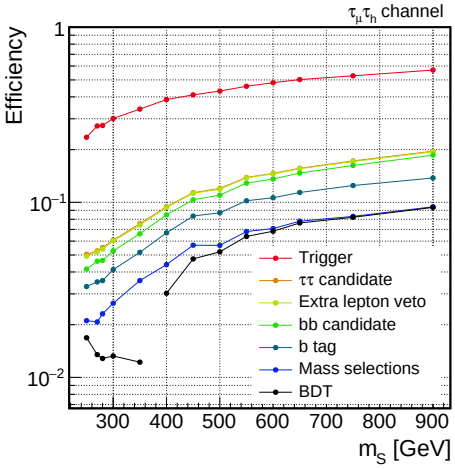
Table 4.5 – Summary table of the selection requirements applied to define the $\mathrm{H} \to \tau\tau$ pair candidate. Lines marked with ($\star$) are not part of the preselection and are applied after the reconstructed candidates pair has been selected.

Common
$p_{\mathrm{T}}^{1} > 20\,\mathrm{GeV}$ and $\lvert\eta^{1}\rvert < 2.4$
$p_{\mathrm{T}}^{2} > 20\,\mathrm{GeV}$ and $\lvert\eta^{2}\rvert < 2.4$
j^{1} and j^{2} PF identification
1b1j resolved category
$\mathrm{CSV}^{1} > 0.8484$ (Medium WP)
$\mathrm{CSV}^{2} > 0.8484$ (Medium WP)
No AK8 jet satisfying m_{AK8}, p_{T}, and substructure criteria
2b resolved category
$\mathrm{CSV}^{1} > 0.8484$ (Medium WP)
$\mathrm{CSV}^{2} < 0.8484$ (Medium WP)
No AK8 jet satisfying m_{AK8}, p_{T}, and substructure criteria
boosted category
$\mathrm{CSV}^{1} > 0.5426$ (Loose WP)
$\mathrm{CSV}^{2} > 0.5426$ (Loose WP)
One AK8 jet of $m_{\mathrm{AK8}} > 30\,\mathrm{GeV}$, $p_{\mathrm{T}} > 170\,\mathrm{GeV}$
Two subjets with $\Delta R(\mathrm{jet}, \mathrm{subjet}) < 0.4$

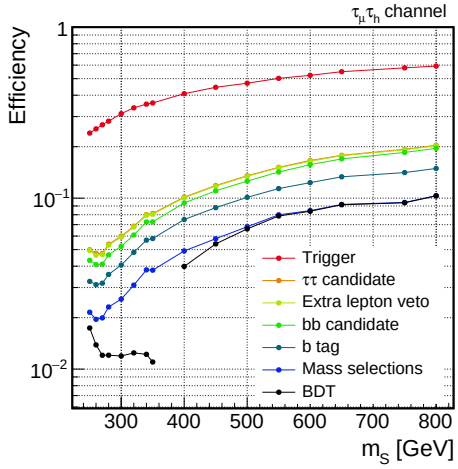
Table 4.6 – List of the selections applied to define the $\mathrm{H} \to \mathrm{bb}$ candidate. The indices 1 and 2 denote the highest and second-to-highest b tagged jets (based on the associated CSV discriminant output) selected amongst those that satisfy the selections labelled as "common".

correspond to a genuine b jet from a $H \to bb$ decay and is rejected by the selection. The difference with respect to the 2b resolved and boosted event categories can be observed from the summary of the expected event yields in the three categories, illustrated in Figure 4.34 for $\tau_h\tau_h$ events.

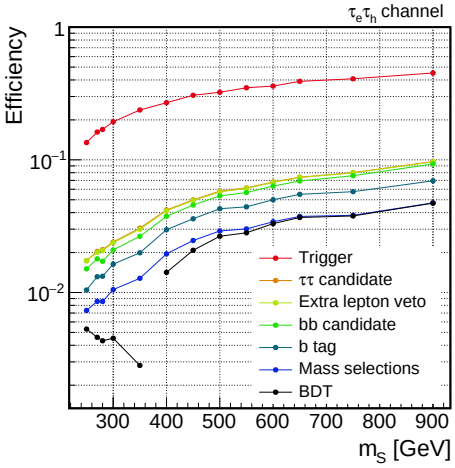
Finally, the transition between the LM and HM BDT selections at $m_X = 350\,\text{GeV}$ can be noticed for resonant signals. For mass values close to such value, the BDT selection efficiency appears limited, but the resulting sensitivity is improved with respect to alternative selections as shown previously in Figure 4.30a. For nonresonant signals, the efficiency of the BDT selection is higher and more uniform because of the lower thresholds applied on the LM BDT score.



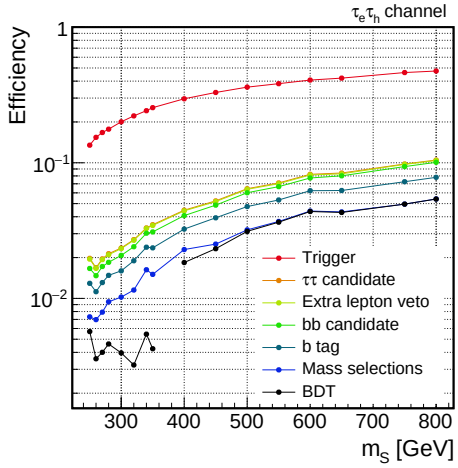
(a) spin-0, $\tau_\mu\tau_h$



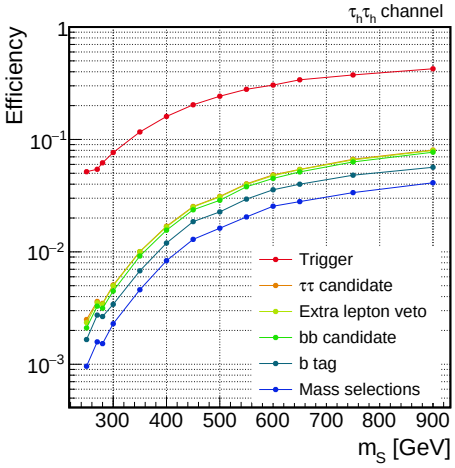
(b) spin-2, $\tau_\mu\tau_h$



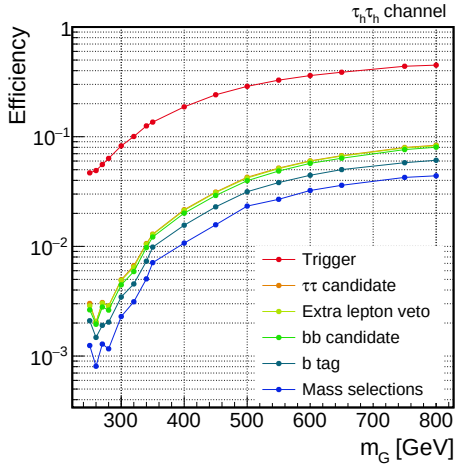
(c) spin-0, $\tau_e\tau_h$



(d) spin-2, $\tau_e\tau_h$

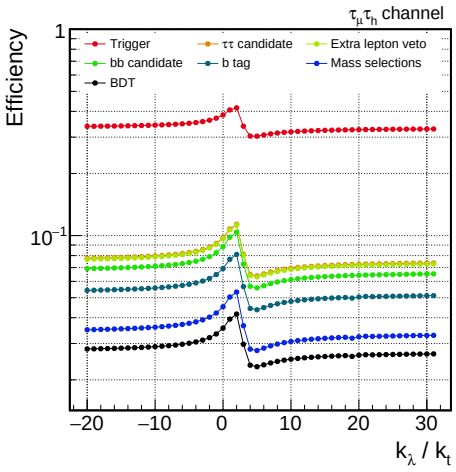


(e) spin-0, $\tau_h\tau_h$

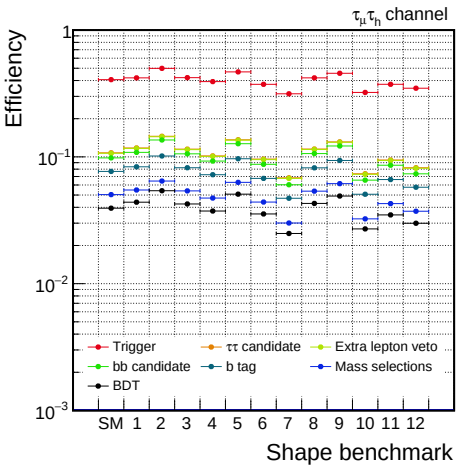


(f) spin-2, $\tau_h\tau_h$

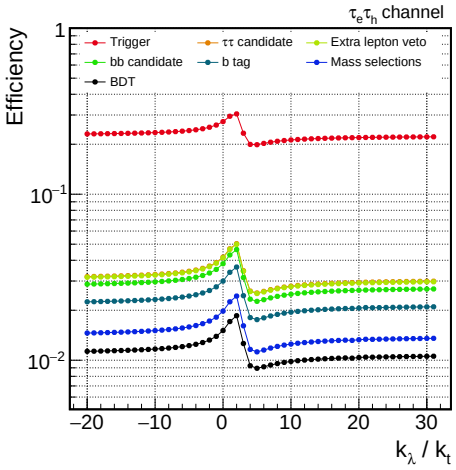
Figure 4.32 – Efficiency for spin-0 (left column) and spin-2 (right column) signals in the $\tau_\mu\tau_h$ (top row), $\tau_e\tau_h$ (central row), and $\tau_h\tau_h$ (bottom row) final states. The cumulated efficiency of the different selections described in the text are shown.



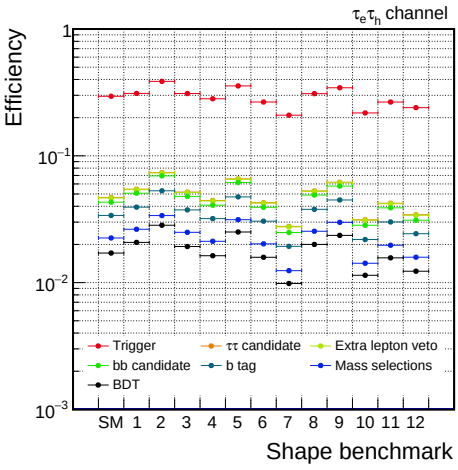
(a) k_λ/k_t, $\tau_\mu\tau_\mathrm{h}$



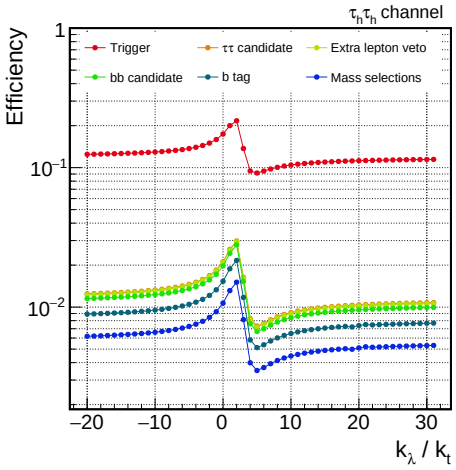
(b) shape benchmarks, $\tau_\mu\tau_\mathrm{h}$



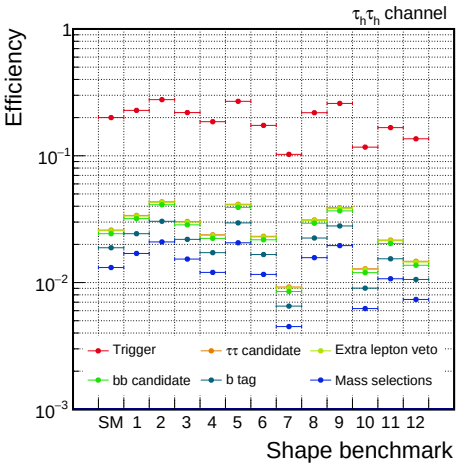
(c) k_λ/k_t, $\tau_\mathrm{e}\tau_\mathrm{h}$



(d) shape benchmarks, $\tau_\mathrm{e}\tau_\mathrm{h}$

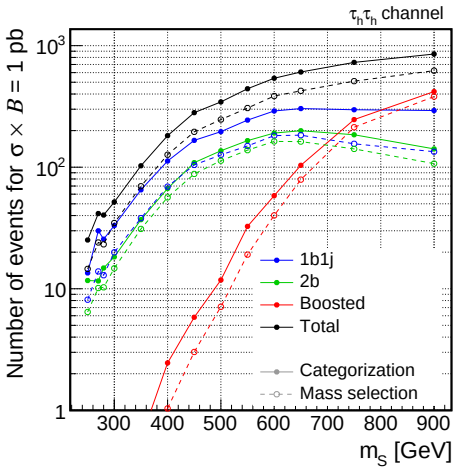


(e) k_λ/k_t, $\tau_\mathrm{h}\tau_\mathrm{h}$

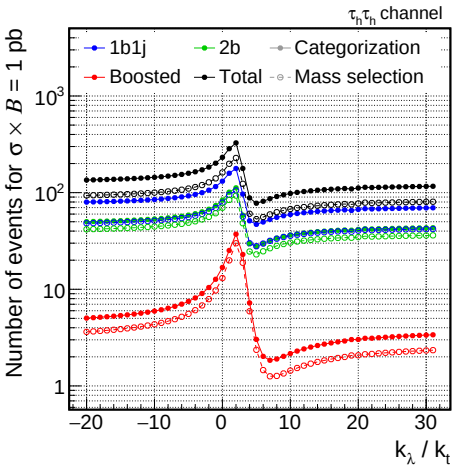


(f) shape benchmarks, $\tau_\mathrm{h}\tau_\mathrm{h}$

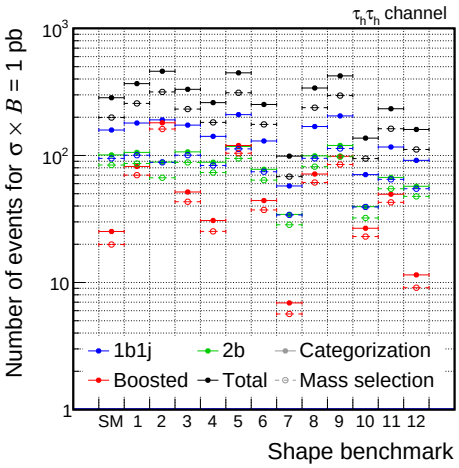
Figure 4.33 – Efficiency of the selections as a function of the anomalous k_λ/k_t hypothesis (left column) and for the shape benchmarks (right column) in the $\tau_\mu\tau_\mathrm{h}$ (top row), $\tau_\mathrm{e}\tau_\mathrm{h}$ (central row), and $\tau_\mathrm{h}\tau_\mathrm{h}$ (bottom row) final states.



(a) Resonant



(b) Nonresonant - k_λ/k_t



(c) Nonresonant - shape benchmarks

Figure 4.34 – Expected event yield in the three bb event categories before (solid lines) and after (dashed lines) the invariant mass selections. The 1b1j (blue), 2b (green) and boosted (red) categories are indicated, together with their sum (black). Panels (a), (b), and (c) represent respectively the resonant spin-0 signal, nonresonant production for anomalous k_λ/k_t couplings, and nonresonant production shape benchmarks. The event yields are normalized to $\sigma(gg \rightarrow HH) \times \mathcal{B}(HH \rightarrow bb\tau\tau) = 1$ pb and are illustrated for the $\tau_h\tau_h$ final state. The same repartition of the events among the three categories is observed in the other final states.

Chapter 5

Modelling of physics processes

An accurate modelling of the signal and background processes is crucial to explore the selected data events. Simulations performed with the Monte Carlo (MC) technique are an essential tool to optimize the analysis strategy, to compare the observed data to the predictions, and to look for the presence of a signal. Some processes are however imperfectly modelled either in their hard scatter physics (such as the emission of a vector boson in association with jets in a leading order simulation) or in their interaction within the detector (such as the probability of misidentification of a quark or gluon jet as a τ_{h}). Approaches consisting in the estimation or correction of these background processes in signal-free data regions, usually referred to as "data-driven" methods, are thus adopted to improve their description in the signal regions.

Residual discrepancies between the MC simulation and the observed data, as well as statistical uncertainties affecting the data-driven methods, result in an imperfect knowledge of the modelling of signal and background processes. These are evaluated and taken into account as systematic uncertainties, that are part of the statistical interpretation of the data.

This chapter presents the modelling of the $\mathrm{HH} \to \mathrm{bb}\tau\tau$ signal in both its resonant and nonresonant production modes, and the MC simulation and data-driven estimations of the background sources that affect this decay channel. The systematic uncertainties are also detailed together with their effect on the sensitivity of the search.

5.1 Properties of Monte Carlo simulation

The production of a MC simulation sample accounts for both the physics and experimental effects and is performed in different steps. The first one is the generation of the hard scatter interaction, and different event generators are used depending on the process studied. The search presented in this thesis uses the MadGraph5_aMC@NLO 2.3.2 [144] and the POWHEG 2.0 [145, 146] generators, with the NNPDF3.0 [147] parton distribution function (PDF) set. Both programs are interfaced with PYTHIA 8.212 [148], that simulates quark hadronization and fragmentation effects, underlying event, and multiple parton interactions (pileup, PU). Events are processed through a simulation of the CMS detector response based on the Geant4 software [149, 150] and undergo the same reconstruction algorithms used for data that are described in Section 2.3 of Chapter 2.

The creation of the MC datasets used in the following is a computing-intensive operation that requires several weeks to be completed. MC samples are usually produced before knowing the exact profile of the number of pileup interactions observed in the data. This difference is shown in Figure 5.1 for the full 2016 dataset. Because of the dependence of

the detector response and of the performance of reconstruction algorithms on the number of pileup interactions, simulated events are weighted according to the ratio of the two distributions. The sum of the weights over an entire sample is scaled to unity to preserve its normalization.

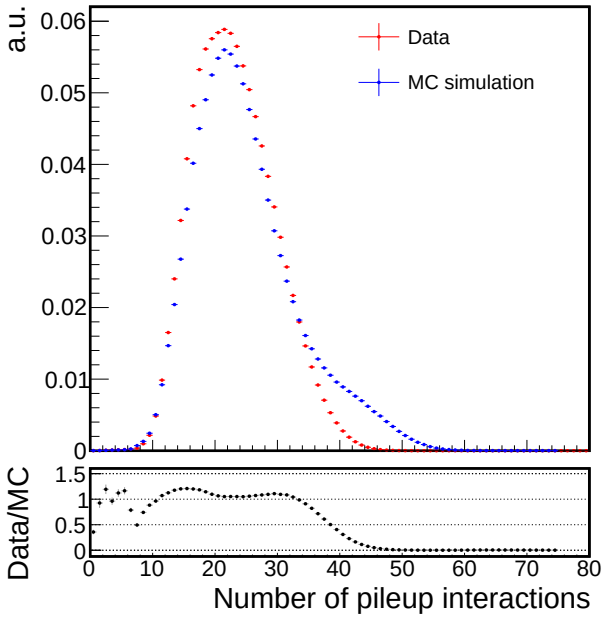


Figure 5.1 – Comparison of the number of pileup interactions observed in 2016 data (red) with the profile used for the MC simulation (blue). The data distribution is obtained assuming a minimum bias effective cross section of 69.2 mb and is derived for the 2016 dataset corresponding to an integrated luminosity of 35.9 fb^{-1}. The ratio shown below is used to derive a correction weight for events in the MC simulation.

5.2 HH signals

The resonant and nonresonant HH production mechanisms are modelled using a MC simulation.

Resonant HH production via gluon-gluon fusion is simulated with MadGraph5_aMC@NLO 2.3.2 at the leading order (LO) precision. Samples are generated for the production of a resonance X decaying to $\text{bb}\tau\tau$. The mass of the resonance (m_X) ranges between 250 and 900 GeV and both the spin-0 and spin-2 hypotheses are considered. A narrow resonance width (i.e. negligible in comparison to the detector resolution) is assumed in the event generation. The number of events generated for every mass value and spin hypothesis ranges between 5×10^4 and 4.5×10^5, corresponding to a sufficient event statistics for an accurate modelling of the signal processes.

For the nonresonant production mechanism, it is of utmost importance to model a large variety of HH signals. In addition to the SM prediction, BSM scenarios in the context of the effective Lagrangian parametrization presented in Section 1.3.2 of Chapter 1 must be explored. This effective Lagrangian approach defines a five-dimensional parameter space where the signal depends on five Higgs boson couplings: λ_{HHH}, y_t, c_2, c_g, $\text{c}_{2\text{g}}$. Deviations from the SM couplings are expressed as $k_\lambda = \lambda_{\text{HHH}}/\lambda_{\text{HHH}}^{\text{SM}}$ and $k_\text{t} = y_\text{t}/y_\text{t}^{\text{SM}}$.

The generation of a separate sample, including the hard scatter process, hadronization and detector response, for all the BSM Higgs boson couplings combinations studied in this search is clearly not feasible. An event weighting technique is therefore used to model a specific combination of BSM couplings starting from a general collection of nonresonant HH events.

The event weighting technique is applied to samples produced for the shape benchmark signals discussed in Section 1.3.2 of Chapter 1, generated with MadGraph5_aMC@NLO 2.3.2 at LO precision. As these samples represent characteristic signal distributions arising from the different combinations of the five BSM couplings, their usage ensures that a large variety of HH kinematic properties is simulated with enough event statistics. The combination of these samples corresponds to a total of about 1.8×10^6 events.

The event weighting technique starts from the consideration that the HH production via gluon fusion is a $2 \to 2$ scattering process. At LO and before the hadronization effects the two Higgs bosons are produced back-to-back in the azimuthal direction with the same transverse momentum. Effects from the PDF can be factored out as the Lorentz boost of the HH pair. Also, the azimuthal angle of the HH pair can be ignored as isotropy is assumed in this direction. Consequently, the HH production in the centre-of-mass frame of the collision is determined by two parameters, that are chosen as the invariant mass of the Higgs boson pair, m_{HH}, and the absolute value of the cosine of the polar angle of one Higgs boson with respect to the beam axis, $|\cos\theta^*|$. The simulated events in each shape benchmark sample are combined and the resulting event distribution in the m_{HH} and $|\cos\theta^*|$ variables is represented as a bidimensional histogram with 55 bins in the first variable and 4 bins in the second one. An histogram with the same binning is realized for the SM HH sample only. After normalizing the two histograms to a unitary area, the content of each bin j is denoted as f^j_{comb} and f^j_{SM}, respectively. The weight associated to an event is determined depending on its m_{HH} and $|\cos\theta^*|$ values using the ratio of the HH differential cross sections in the corresponding bin j for the target BSM coupling combination to the SM.

As discussed in Section 1.3.2 of Chapter 1, the ratio of the total HH cross section to the SM prediction, R_{HH}, is parametrized with the formula (1.47). This is a generic expression for the interference of the five diagrams arising from the effective Lagrangian parametrization, and is thus valid for every bin j of the bidimensional $(m_{\mathrm{HH}}, |\cos\theta^*|)$ histogram, provided the coefficients A_i become a function of the bin number j itself, A^j_i. Denoting the cross section in such bin as σ^j_{HH}, the formula reads:

$$
\begin{aligned}
R^j_{\mathrm{HH}} = \frac{\sigma^j_{\mathrm{HH}}}{\sigma^{j,\mathrm{SM}}_{\mathrm{HH}}} \overset{\text{LO}}{=} & A^j_1 k_{\mathrm{t}}^4 + A^j_2 \mathrm{c}_2^2 + \left(A^j_3 k_{\mathrm{t}}^2 + A^j_4 \mathrm{c}_{\mathrm{g}}^2\right) k_\lambda^2 + A^j_5 \mathrm{c}_{2\mathrm{g}}^2 \\
& + \left(A^j_6 \mathrm{c}_2 + A^j_7 k_\lambda k_{\mathrm{t}}\right) k_{\mathrm{t}}^2 + \left(A^j_8 k_{\mathrm{t}} k_\lambda + A^j_9 \mathrm{c}_{\mathrm{g}} k_\lambda\right) \mathrm{c}_2 \\
& + A^j_{10} \mathrm{c}_2 \mathrm{c}_{2\mathrm{g}} + \left(A^j_{11} \mathrm{c}_{\mathrm{g}} k_\lambda + A^j_{12} \mathrm{c}_{2\mathrm{g}}\right) k_{\mathrm{t}}^2 \\
& + \left(A^j_{13} k_\lambda \mathrm{c}_{\mathrm{g}} + A^j_{14} \mathrm{c}_{2\mathrm{g}}\right) k_{\mathrm{t}} k_\lambda + A^j_{15} \mathrm{c}_{\mathrm{g}} \mathrm{c}_{2\mathrm{g}} k_\lambda
\end{aligned}
\tag{5.1}
$$

Using generated HH events, the ratio R^j_{HH} is computed for different combinations of the five couplings and interpolated as a function of them to extract the values of the A^j_i coefficients. The method is thus a generalization of the one used for the parametrization

of the total HH cross section already discussed. The details on the fit procedure and the verification of its quality, as well as the values of the A_i^j coefficients, are given in Ref. [151].

The event weight ω used to model a generic BSM combination of Higgs boson couplings is consequently fully determined from this parametrization and from the bidimensional histograms of event numbers described above. A value Ω is defined as:

$$\begin{aligned}\Omega(k_\lambda, k_\mathrm{t}, \mathrm{c}_2, \mathrm{c}_\mathrm{g}, \mathrm{c}_{2\mathrm{g}}; j) &\equiv \frac{1}{f^j_\mathrm{comb}} \cdot \frac{\sigma^j_\mathrm{HH}(k_\lambda, k_\mathrm{t}, \mathrm{c}_2, \mathrm{c}_\mathrm{g}, \mathrm{c}_{2\mathrm{g}})}{\sigma_\mathrm{HH}(k_\lambda, k_\mathrm{t}, \mathrm{c}_2, \mathrm{c}_\mathrm{g}, \mathrm{c}_{2\mathrm{g}})} \\ &= \frac{f^j_\mathrm{SM}}{f^j_\mathrm{comb}} \cdot \frac{R^j_\mathrm{HH}(k_\lambda, k_\mathrm{t}, \mathrm{c}_2, \mathrm{c}_\mathrm{g}, \mathrm{c}_{2\mathrm{g}})}{R_\mathrm{HH}(k_\lambda, k_\mathrm{t}, \mathrm{c}_2, \mathrm{c}_\mathrm{g}, \mathrm{c}_{2\mathrm{g}})}\end{aligned} \tag{5.2}$$

where it has been indicated explicitly the dependence on the five couplings of the ratio of the total (R_HH) and differential (R^j_HH) cross section to the SM prediction, completely determined from the A_i^j coefficients. The event weight ω is defined from Ω by normalizing it to the sum over all the n simulated MC signal events considered:

$$\omega = \frac{\Omega}{\sum_n \Omega} \tag{5.3}$$

In this way, the application of ω only modifies the differential event distribution but not the normalization.

The validity of the procedure is controlled with a closure test where the distribution obtained with the application of the formula (5.3) is compared to the generated samples. This is illustrated in Figure 5.2, where the good agreement of the two methods is shown for two separate signal samples.

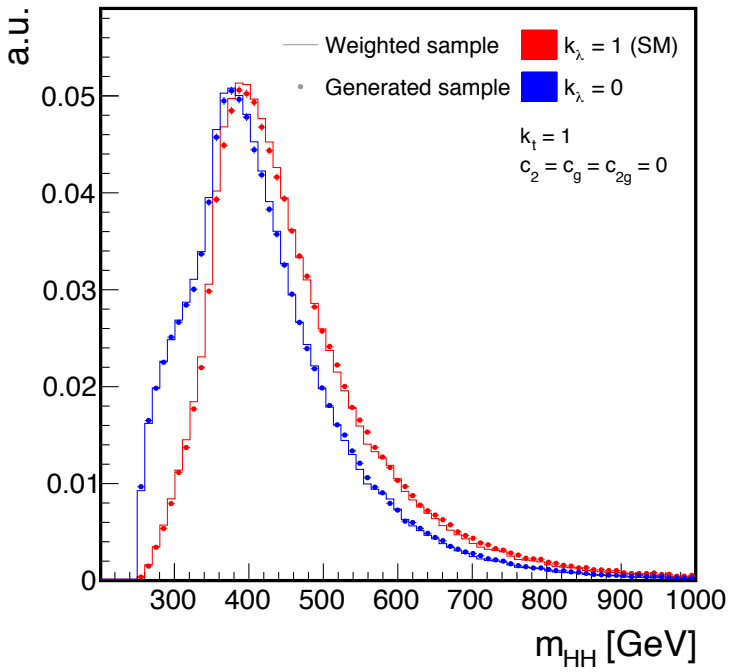


Figure 5.2 – Comparison of the m_HH distribution obtained in a MC sample generated with a specific set of Higgs boson couplings (points) and with the event weighting procedure (solid lines).

5.3 $t\bar{t}$ background

The $t\bar{t}$ production, representing the major source of background for a $HH \to bb\tau\tau$ signal, is modelled with a MC simulation at NLO precision with POWHEG 2.0. To optimise the coverage of the decay phase space, an inclusive event sample is complemented by separate samples generated for the fully leptonic ($t\bar{t} \to b\bar{b}\ell^-\bar{\nu}_\ell\ell^+\nu_\ell$) and semileptonic ($t\bar{t} \to b\bar{b}\ell\nu_\ell qq'$) decay modes. These two decay modes represent, respectively, the 10.5% and 43.8% of the $t\bar{t}$ decays. Events are normalized to the theoretical cross section at next-to-next-to-leading order (NNLO) precision [152, 153], that amounts to $831.8^{+46.4}_{-50.9}$ pb. The value is computed assuming a top quark mass of 172.5 GeV, and the quoted uncertainty combines the scale, PDF, α_S, and top quark mass uncertainties as detailed in the references.

The POWHEG MC simulation of $t\bar{t}$ production is known to describe accurately the experimental data, and the agreement has been recently improved thanks to the usage of the underlying event tune CUETP8M2 [154] in PYTHIA, and verified in independent analyses of $t\bar{t}$ production [155, 156, 157]. In the context of this search, verifications are performed with the observed data in $t\bar{t}$ dominated regions. Some of these comparisons were already shown for the input variables of the BDT discriminant in Section 4.6.2 of Chapter 4, where the good agreement in different kinematic properties of the $t\bar{t}$ simulation is observed. An additional comparison in performed in a region where the m_{bb} and $m_{\tau\tau}$ are required to fail the invariant mass cut (4.6) described in Section 4.6.1 of Chapter 4 but satisfy:

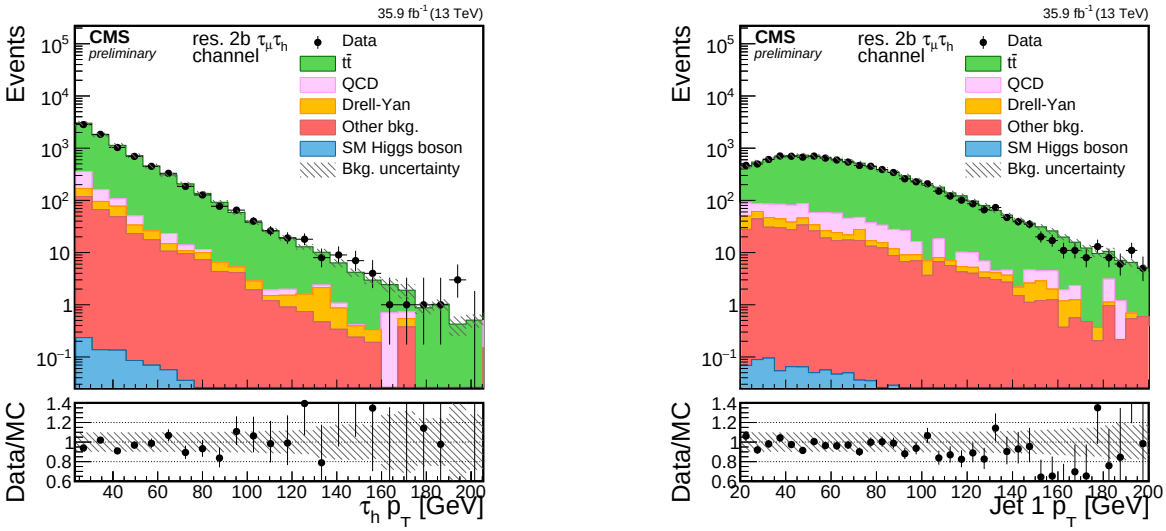
$$\sqrt{\left(m_{\tau\tau} - 116\,\text{GeV}\right)^2 + \left(m_{bb} - 111\,\text{GeV}\right)^2} < 60\,\text{GeV} \tag{5.4}$$

that defines a $t\bar{t}$ dominated mass sideband. The good agreement of the MC prediciton with the observed data is illustrated for the τ_h and b jet p_T distributions in Figure 5.3, and is shown before the maximum likelihood fit of the nuisance parameter to the data described in Section 6.3 of Chapter 6. The MC simulation models well these important kinematic variables, that are highly correlated with the fitted observables described in Section 6.2 of Chapter 6.

Given the importance of the $t\bar{t}$ background in the $bb\tau\tau$ decay channel, a procedure to account for residual uncertainties in its p_T distribution is applied as described in Section 5.7.

5.4 Multijet background

Multijet QCD events where one or two jets are misidentified as a τ_h candidate represent a challenging background to be modelled and estimated. Two factors limit our capability to model the multijet background using MC samples. First, the small probability for a quark or gluon jet to be identified as a τ_h candidate (ranging between 10^{-2} and 10^{-3}), combined with the small probability for two additional jets in the event to satisfy b tagging criteria (of 10^{-2} for the medium working point for gluon and light flavour jet misidentification), would require the generation of a very large sample to ensure the presence of a sufficient number of simulated events in the signal regions. Secondly, misidentification of quark and gluon jets as τ_h is known to be imperfectly modelled by the MC simulation because entirely dependent on detector effects. These two problems are solved by adopting a data-driven



(a) τ_{h} p_{T} (b) b jet p_{T}

Figure 5.3 – Distribution of the τ_{h} (a) and highest b tagging score jet (b) in a mass sideband region of the $\tau_{\mu}\tau_{\mathrm{h}}$ final state in association with two b tagged jets. The background expectation is shown before the maximum likelihood fit of the nuisance parameter to the data.

approach, consisting in the estimation of the multijet background contribution from jet-enriched regions that are populated by events independent from those entering the signal region (the regions are said to be "orthogonal").

Description

The data-driven estimation is performed in this search with the so-called ABCD method, that consists in the usage of a data sideband (B) where the multijet background is estimated and extrapolated to the signal region (A), with an extrapolation factor k being computed in two orthogonal regions C and D. The signal region A is defined by the event selection criteria detailed in Chapter 4, that include the presence of a τ candidate pair ($\tau_{\mu}\tau_{\mathrm{h}}$, $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$, or $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ depending on the final state) with opposite-sign (OS) electric charge, and where the τ_{h} candidates satisfy the medium working point of the MVA isolation criterion. The region B is defined with the same isolation criteria but inverting the electric charge requirement (same-sign or SS region). In the regions C and D, the τ_{h} isolation criterion is modified. In the $\tau_{\mu}\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states, the selected τ_{h} candidate must satisfy the very loose working point of the MVA isolation criterion (corresponding to a misidentification rate between 5% and 0.5% depending on the jet p_{T} [97]) but fail the medium one. In the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state, the same change of the isolation criterion is applied only to the lowest p_{T} τ_{h} candidate selected. OS and SS electric charge requirements are applied in the C and D regions, respectively. The definition of these four regions is schematically represented in Figure 5.4.

In the B, C, and D regions thus defined, the number of multijet background events (N_i, $i = \mathrm{B}, \mathrm{C}, \mathrm{D}$) is estimated by subtracting to the total number of data events observed (N_i^{data}) the yields of the residual backgrounds, estimated using the MC simulation (N_i^{bkg}):

$$N_i = N_i^{\mathrm{data}} - N_i^{\mathrm{bkg}} \tag{5.5}$$

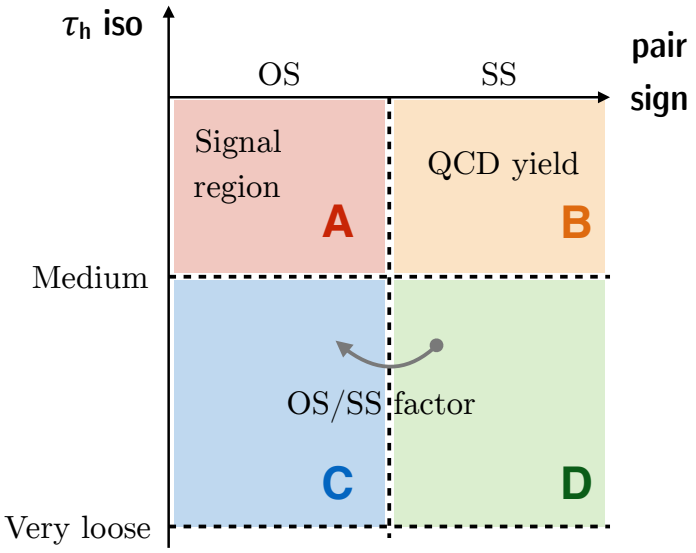


Figure 5.4 – Schematic representation of the regions used for the QCD multijet background estimation. The region A represents the signal region, defined by the application of the medium working point of the τ_h isolation criterion and opposite electric charge sign of the selected lepton pair. The multijet background yield is estimated from the region B, where the pair charge sign requirement is inverted. The factor to extrapolate the yield from the region B to the region A is computed in the regions C and D, defined by the presence of τ_h candidates that satisfy the very loose but fail the medium working point of the isolation criterion.

The multijet background yield in the signal region A (N_A) is estimated as:

$$N_A = N_B \times \frac{N_C}{N_D} \tag{5.6}$$

The differential distribution of the multijet background, or "shape" in the following, is similarly estimated from an SS region. Because the number of events is typically small in the B region previously defined, a B' region with a relaxed τ_h isolation criterion is used. Different values of the MVA isolation relaxation have been tested, and the final definition of B' corresponds to the presence of one τ_h candidate satisfying the very loose MVA isolation working point, which corresponds to the union of the B and D regions. The multijet background shape is computed as a subtraction of the residual backgrounds from the data in each bin of the distributions considered, and it is normalized to the value obtained from the formula (5.6).

Validation

The method is validated using A, B, C, and D regions where the two jets fail the medium working point of the b tagging discriminant. These regions have a negligible presence of signal and are largely contaminated from the multijet background. Focusing on the $\tau_h\tau_h$ final state where the multijet background contamination is more important, the number of events observed in the B, C, and D regions is summarized in Table 5.1; it can be seen that contamination from residual background is negligible. The event distributions of

the p_{T} of the leading τ_{h} candidate are shown in Figures 5.5b, 5.5c, and 5.5d fort the B, C and D regions, respectively. The factor k is computed as the ratio of the multijet background in regions C and D, giving a value of 1.51 ± 0.07. Here the uncertainty is obtained by propagating the statistical uncertainty from the observed data and assuming a Poissonian fluctuation; uncertainties in the other background processes are neglected because of their small contribution. The multijet normalization is thus computed by multiplying the estimation from region B by this factor, while its shape is obtained from region B', shown in Figure 5.5e. The resulting background template corresponds well to the data observed in the region A, as shown in Figure 5.5a.

Region	Obs. data	Other bkg. sum
B	469	29
C	1217	153
D	720	16

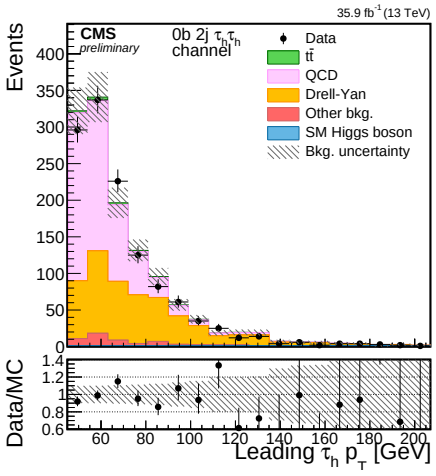
Table 5.1 – Total event yields observed and expectation from simulated SM backgrounds in the B, C, and D regions used for the validation of the multijet estimation method in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state in association to two anti b-tagged jets. Uncertainties in the normalization of the residual background sum are negligible in this case and thus not reported.

The large number of events in the SS regions also allows for an additional verification of the hypothesis that no bias in the shape exists when relaxing the isolation criterion in the definition of the region B' with respect to B. The isolation criterion on the lowest p_{T} τ_{h} candidate is progressively relaxed and the shapes obtained with the very loose, loose, and medium MVA isolation WPs are compared in Figure 5.6. The three shapes are found to be compatible within the uncertainties and no bias or trend is observed as the isolation criterion is relaxed. This complements the previous studies and confirms the validity of the B' region definition to estimate the multijet background shape.

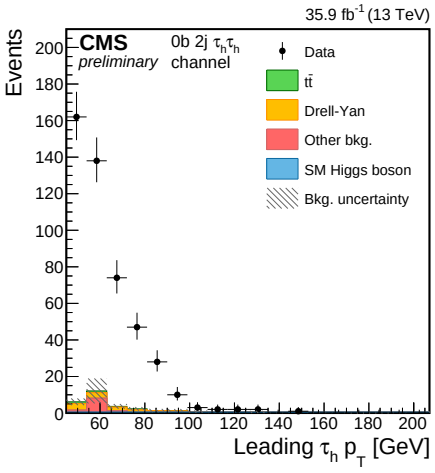
The validity of the method is similarly verified for the $\tau_{\mu}\tau_{\mathrm{h}}$ amd $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states in regions defined by the presence of two jets that fail the medium WP of the b tag discriminant. The large contamination from W boson production in association with jets is suppressed by requiring the transverse mass m_{T} of the electron or muon to be smaller than 30 GeV. After this selection, about 20% of the events are due to the multijet background in the $\tau_{\mu}\tau_{\mathrm{h}}$ final state. In the $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final state, the multijet background is subdominant with respect to the Drell-Yan background because of the highest lepton p_{T} threshold and larger probability of electron misidentification as a τ_{h}. The multijet contribution is increasde to about 15% of the events by removing the invariant mass selection. The good agreement of the background modelling with the observed data in the two final states is shown in Figure 5.7 for the lepton and jet p_{T} distributions.

Inclusion in the search

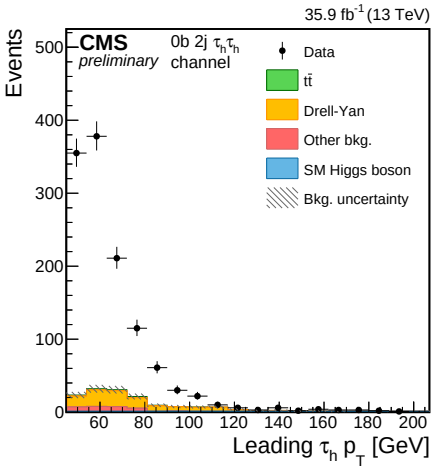
Once the multijet estimation method is applied in the signal regions of the search, defined by the presence of one or two b-tagged jets, a larger contamination from $\mathrm{t\bar{t}}$ background is expected in the regions B, C, and D with respect to the validation regions shown in Figure 5.5. The multijet estimation of Eq. (5.6) is thus expressed as a parametric function of the observed data and of the residual background yields in the three regions, which



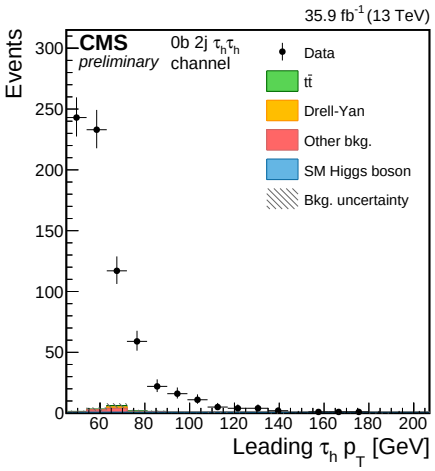
(a) A region (signal region)



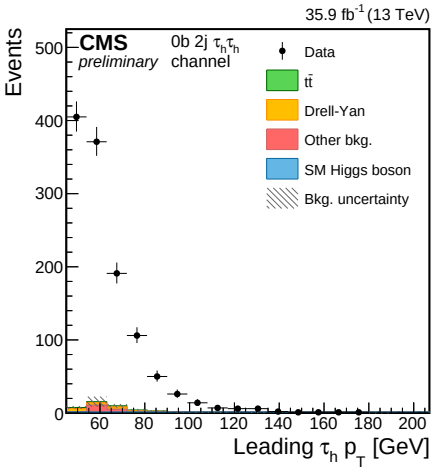
(b) B region (SS, iso)



(c) C region (OS, anti-iso)



(d) D region (SS, anti-iso)



(e) B' region (shape)

Figure 5.5 – Event distribution in anti b-tagged data sideband used for the verification of the multijet background estimation method. The multijet contribution is estimated as the difference between the data and the other backgrounds in the region B (b) and then scaled by the ratio of the multijet contribution in regions C (c) and D (d). The shape of the multijet background is estimated from the B' region (e). The agreement with the observed data in the signal region A is shown in (a).

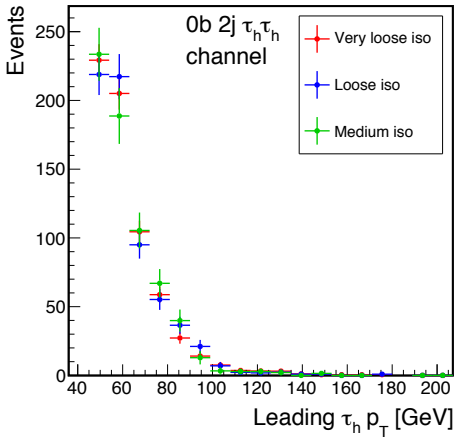
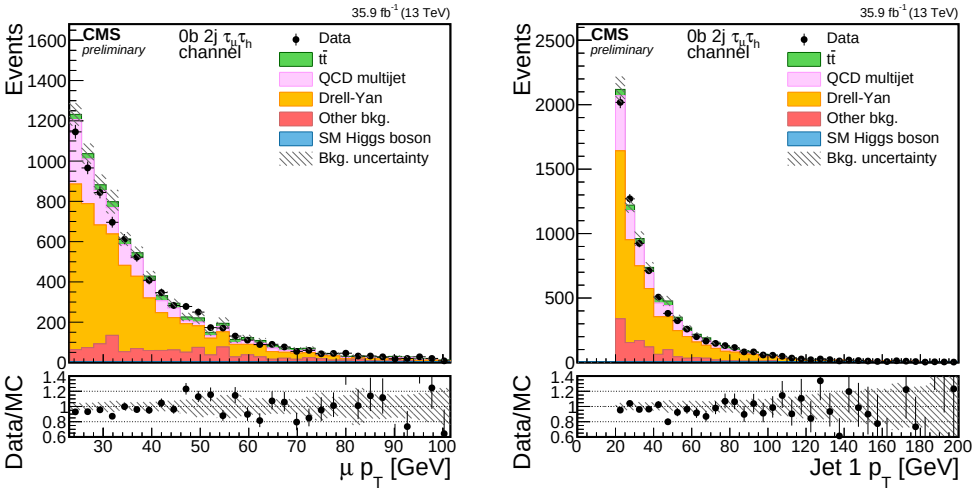
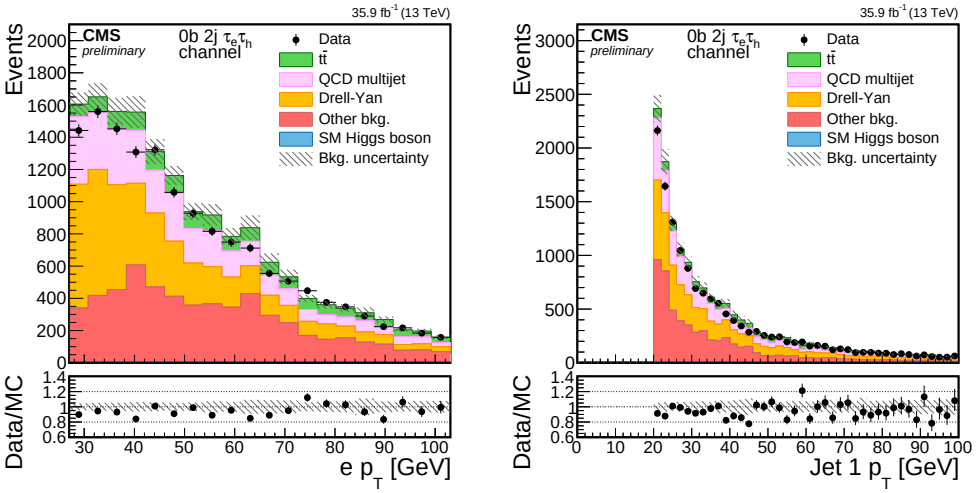


Figure 5.6 – Comparison of the shapes of the multijet background estimated in a SS region where the lowest p_T τ_h must satisfy the very loose (red), loose (blue), and medium (green) WP of the MVA isolation criterion. The red curve corresponds to the region B' used for the shape determination. No bias or trend is observed when progressively relaxing the isolation criterion.



(a) μ p_T (b) highest b tag score jet p_T



(c) e p_T (d) highest b tag score jet p_T

Figure 5.7 – Distribution of the lepton (left column) and jet (right column) p_T in the $\tau_\mu\tau_h$ (top row) and $\tau_e\tau_h$ (bottom row) final states. Selected events have two jets that fail the medium WP of the b tagging algorithm and an electron or muon with $m_T < 30\,\text{GeV}$. For the $\tau_e\tau_h$ final state, events are not required to satisfy any invariant mass requirement.

are included in the maximum likelihood fit detailed in Section 6.3 of Chapter 6. In this way, the correlation between the $\mathrm{t\bar{t}}$ and multijet processes introduced by the estimation method are taken into account.

5.5 Drell-Yan background

The estimation of the $\mathrm{Z}/\gamma^* \to \ell\ell$ + jets (with $\ell = \mathrm{e}, \mu, \tau$) background, also denoted as Drell-Yan background, makes use of a MC simulation based on the MadGraph5_aMC@NLO 2.3.2 generator used at LO precision. To increase the number of the selected events, an inclusive simulated sample is complemented by exclusive samples where the emission of 1,2,3, and 4 jets at the matrix element (ME) level is required, and by two samples where one and two b quarks must be emitted at the ME level. The matching between the jets generated at ME and from the parton shower simulation is realized in the MLM merging scheme [158]. The generation is optimized by requiring the lepton pair in the $Z/\gamma^* \to \ell\ell$ process to satisfy the invariant mass requirement $m_{\ell\ell} > 50\,\mathrm{GeV}$, as all events below such threshold are anyway rejected by the selections that define the signal regions.

The MC samples are combined and normalized to the theoretical cross section known at the NNLO precision, computed with the FEWZ program [159]. The total cross section amounts to $\sigma(\mathrm{Z}/\gamma^* \to \ell\ell) = 5765 \pm 99\,\mathrm{pb}$, where the error combines the PDF, α_S, and numerical integration errors.

The requirement on the presence of two jets in the event and the application of b tagging criteria defines a narrow phase space where only a small fraction of the generated Drell-Yan events is selected, although their contribution to the signal regions of the $\mathrm{bb}\tau\tau$ decay channel is sizeable because of the large total cross section. As an example, the fraction of events with ≥ 2 jets emitted at the ME level is only about 10% and the fraction of events where 2 b quarks are emitted is 0.2%. The usage of the LO precision in the MC simulation allows for the generation of large samples that populate these specific kinematic regions with ≥ 2 jets of heavy quark flavour, but has the drawback to imperfectly model the normalization of jet emission in different flavours, as it can be seen in analyses of SM Z boson production in association with jets [160]. In CMS Run I data analyses this problem was avoided with the usage of the so-called embedding technique, where $\mathrm{Z}/\gamma^* \to \mu\mu$ events are selected in real data and the muon candidates are substituted by τ leptons in a dedicated simulation; this ensures that other properties of the event such as the jet emission are directly modelled from data. The embedding technique was unfortunately not ready for 2015 and 2016 analyses because of technical issues, and an alternative data-driven approach has thus been developed to correct the yield of Drell-Yan production in association to jets.

The method similarly uses $\mathrm{Z} \to \mu\mu$ production in association to two jets to compare the measured data with the MC simulation and correct for the jet emission. Events in the $\mu\mu$ control region are recorded with the same single muon trigger used for the $\tau_\mu\tau_\mathrm{h}$ final state, and must contain two muons, one with $p_\mathrm{T} > 23\,\mathrm{GeV}$ and $|\eta| < 2.1$ and the other with $p_\mathrm{T} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$. Both muons must satisfy the tight working points of the identification algorithm and the relative isolation criterion $\mathcal{I}_\mu^\mathrm{rel} < 0.15$. The trigger, identification, and isolation efficiency corrections are applied. The muon pair must have an invariant mass $m_{\mu\mu}$ satisfying the selection $60 < m_{\mu\mu} < 120\,\mathrm{GeV}$. To further suppress the $\mathrm{t\bar{t}}$ background, the magnitude of the missing transverse momentum vector of the

event is required to be smaller than 45 GeV. The selected events must also contain two jets of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.4$, that satisfy the particle-flow identification criterion. A requirement on the invariant mass m_{jj} of the jet pair $80 < m_{\mathrm{jj}} < 160\,\mathrm{GeV}$ is applied to ensure that these jets have similar kinematics to those selected in the signal region.

The data sample thus selected is divided in three regions depending on the number of b tagged jets (0, 1, or 2, respectively denoted as 0b2j, 1b1j, 2b0j); the medium working point of the b tagging algorithm is used. The MC Drell-Yan event sample is similarly split into three components depending on the number of generator level b jets (0, 1, or 2 genuine b jets, respectively denoted as $\mathrm{DY_{0b}}$, $\mathrm{DY_{1b}}$, $\mathrm{DY_{2b}}$), that are defined by clustering the stable generated particles (excluding the neutrinos) with the anti-k_{T} algorithm, and by inferring their flavour from the clustered candidates. The $m_{\mu\mu}$ distribution is then simultaneously fitted in the three regions, allowing the three Drell-Yan components and the total background template to float around their initial values estimated from the simulation. As a result, three correction factors are derived for $\mathrm{DY_{0b}}$, $\mathrm{DY_{1b}}$, and $\mathrm{DY_{2b}}$.

It should be noted that this procedure only aims at correcting the total yield of the $\mathrm{DY_{0b}}$, $\mathrm{DY_{1b}}$, and $\mathrm{DY_{2b}}$ processes, without modifying their differential distribution. The shape of variables related to the jet emission, such as the Z boson p_{T} reconstructed as the vectorial sum of the two selected muons p_{T} vectors, is correctly modelled in the MC simulation. As shown in Figure 5.8, the ratio between the MC simulation and the data is approximatively flat as a function of the Z boson p_{T}. Additional discrepancies have a negligible impact on the background modelling because small with respect to systematic bin-by-bin uncertainties defined in Section 5.7.

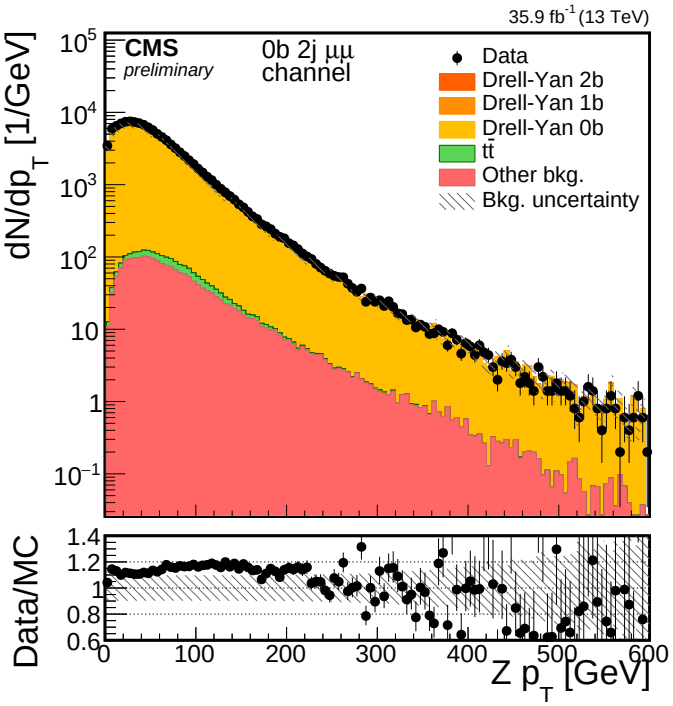


Figure 5.8 – Reconstructed p_{T} distribution of the selected $\mathrm{Z} \to \mu\mu$ events in the $\mu\mu$ data sideband in association to two jets that do not satisfy the b-tagging criterion. The ratio of the data to the MC modelling is approximately flat as a function of p_{T}, but significantly different from unity.

The distribution of the $m_{\mu\mu}$ variable in the three $\mu\mu$ control regions are shown in

Figures 5.9a, 5.9c and 5.9e for the 0b2j, 1b1j, and 2b0j categories. As expected, the 0b2j region is dominated by $\mathrm{DY_{0b}}$ production, while data in the 1b1j and 2b0j regions can constrain the contribution from $\mathrm{DY_{1b}}$, and $\mathrm{DY_{2b}}$ respectively.

The simultaneous fit results in the following correction factors κ for the three background components:

$$\begin{aligned} \kappa(\mathrm{DY_{0b}}) &= 1.1272 \pm 0.0017 \\ \kappa(\mathrm{DY_{1b}}) &= 1.173 \pm 0.015 \\ \kappa(\mathrm{DY_{2b}}) &= 1.146 \pm 0.029 \end{aligned} \tag{5.7}$$

The agreement with the observed data after the application of the three factors is shown in Figures 5.9b, 5.9d, and 5.9f for the three regions.

It should be noted that the 0b2j, 1b1j, and 2b0j regions are not independent and that, in particular, sizeable contamination from $\mathrm{DY_{0b}}$ is observed in 1b1j and 2b0j. The fit procedure consequently introduces a correlation between the three correction factors κ and, in addition, a fourth source of uncertainty in the Drell-Yan normalization is represented by the residual background contribution, that is left floating in the fit. The correlation between these effects is described by the following covariance matrix:

$$\mathrm{cov}_{ij} = \begin{pmatrix} 2.952 \cdot 10^{-6} & -7.515 \cdot 10^{-6} & 7.45 \cdot 10^{-6} & -1.00 \cdot 10^{-5} \\ -7.515 \cdot 10^{-6} & 0.0002393 & -0.000214 & -4.622 \cdot 10^{-5} \\ 7.45 \cdot 10^{-6} & -0.000214 & 0.0008489 & -0.0002045 \\ -1.002 \cdot 10^{-5} & -4.622 \cdot 10^{-5} & -0.0002045 & 0.0005443 \end{pmatrix} \tag{5.8}$$

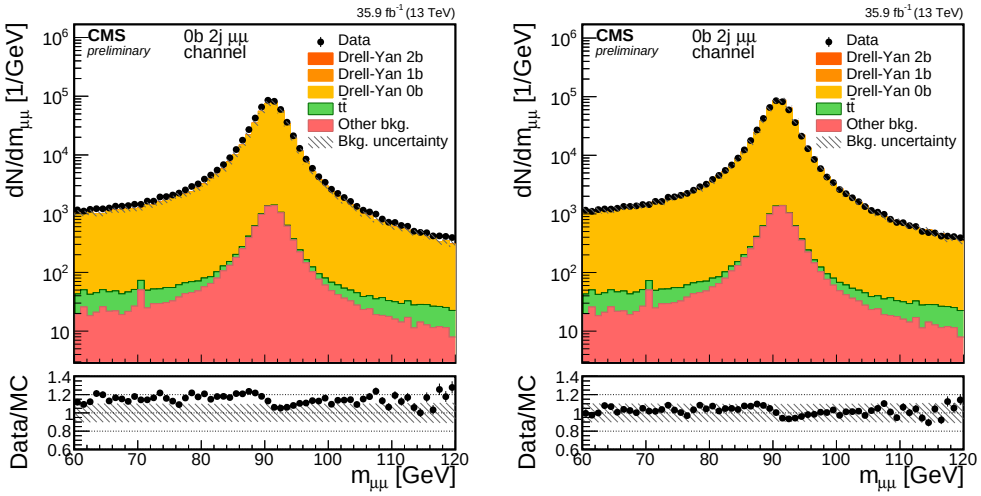
where indices i and j run between 0 and 3, with 0,1,2 representing $\mathrm{DY_{0b}}$, $\mathrm{DY_{1b}}$, and $\mathrm{DY_{2b}}$ respectively, and 3 representing the residual background contribution. The correlated effect of the three factors κ on the Drell-Yan background normalization is taken into account with the usage of this matrix as described in Section 5.7.

5.6 Other backgrounds

Other background processes that affect the $\mathrm{bb}\tau\tau$ decay channel are W bosons in association with jets, single top quarks in association to a vector boson, pair production of vector bosons, electroweak production of a vector boson in association with jets, and SM single Higgs boson production in association to a Z boson. As their total contribution is smaller than about 3% in the most sensitive event categories, and never accounts for more than 10% of the the total event yields in any signal regions of this search, their estimation relies solely on the MC simulation.

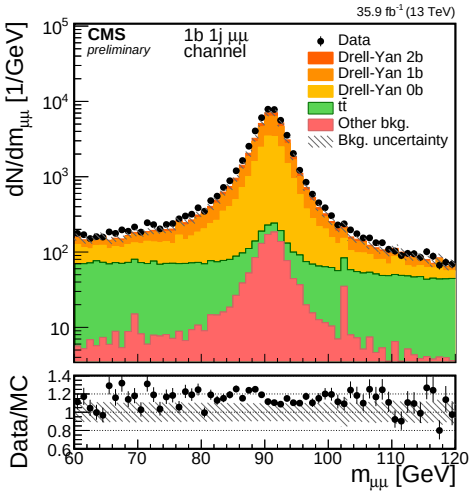
The $\mathrm{W} \to \ell\nu_\ell$ production (with $\ell = \mathrm{e}, \mu, \tau$) is simulated with the MadGraph5_aMC@NLO 2.3.2 generator at LO precision. As for the $\mathrm{Z}/\gamma^* \to \ell\ell$ simulation, an inclusive sample is complemented with samples generated in exclusive regions of the transverse momentum scalar sum of the partons emitted at the matrix element level (H_T). The merging with the parton shower is realized in the MLM scheme. The sample is normalized to the theoretical cross section computed at the NNLO precision with the FEWZ program and amounts to $\sigma(\mathrm{W} \to \ell\nu_\ell) = 6.15^{+0.24}_{-0.23} \times 10^4\,\mathrm{pb}$.

The single top quark production in the W channel is simulated with POWHEG 2.0 at NLO precision, and normalized to the theoretical cross section at NNLO approximate

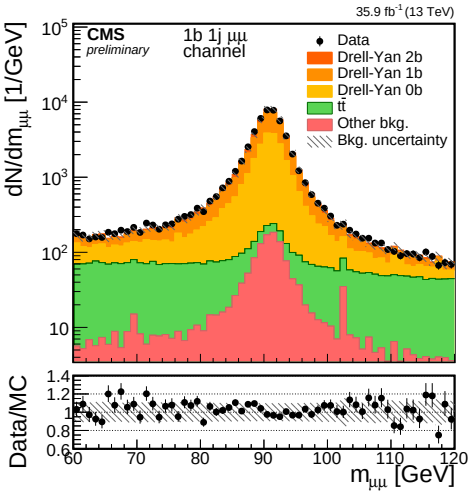


(a) 0b2j, no correction

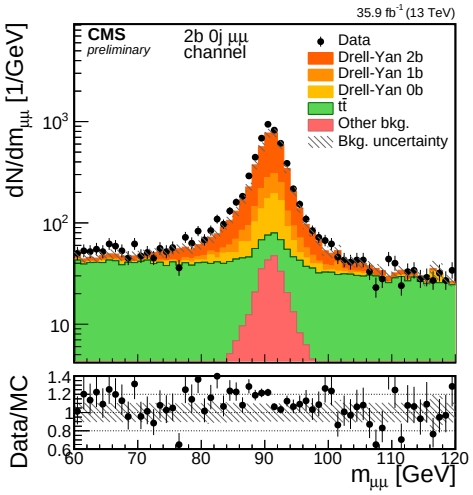
(b) 0b2j, with correction



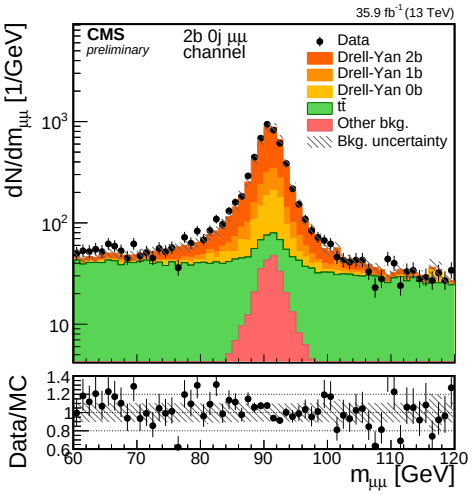
(c) 1b1j, no correction



(d) 1b1j, with correction



(e) 2b0j, no correction



(f) 2b0j, with correction

Figure 5.9 – Event distribution in the $m_{\mu\mu}$ variable before (left column) and after (right column) the application of the correction of the Drell-Yan background yield. The first, second, and third row denote the 0b2j, 1b1j, and 2b0j event categories, respectively.

precision that amounts to $\sigma(\mathrm{tW}) = 71.7 \pm 3.8\mathrm{pb}$ [161]. The uncertainty combines scale, α_S, and PDF uncertainties.

Vector boson pair (also denoted as diboson) samples include the ZZ, ZW, and WW production and are realized with the POWHEG 2.0 and MADGRAPH5_aMC@NLO 2.3.2 generators. For the ZZ process, separate samples are generated for the $\ell\ell\nu_\ell\nu_\ell$, $\ell\ell qq$, $\ell\ell\ell'\ell'$, and $qqq'q'$ final states, where $\ell = \mathrm{e}, \mu, \tau$. They are normalized to the theoretical cross section know at NNLO [162], that for the inclusive ZZ production amounts to $\sigma(\mathrm{ZZ}) = 16.5^{+0.5}_{-0.4}\,\mathrm{pb}$, and the quoted uncertainties account for the variations of renormalization and factorization scales. Samples for the ZW process are separately generated in the $\ell\ell\ell'\nu_{\ell'}$, $\nu\nu\ell\nu_\ell$, $qq\ell\nu_\ell$, and $\ell\ell qq'$ final states and normalized to the inclusive NNLO cross section $\sigma(\mathrm{WZ}) = 45.0 \pm 2.5\,\mathrm{pb}$. Finally, simulated WW samples are generated for the $\ell\nu_\ell\ell'\nu_{\ell'}$, $\ell\nu_\ell qq'$, and $qq'qq'$ final states and normalized to the inclusive NNLO cross section $\sigma(\mathrm{WW}) = 118.7^{+3.0}_{-2.6}\,\mathrm{pb}$.

Electroweak production of W^+, W^-, and Z boson decaying to final state leptons in association with two jets is simulated with MADGRAPH5_aMC@NLO 2.3.2 at LO. The three processes are normalized to the cross sections at LO obtained from the MC generator, corresponding to $\sigma^{\mathrm{EWK}}(\mathrm{W}^+) = 25.69 \pm 0.51$ pb, $\sigma^{\mathrm{EWK}}(\mathrm{W}^-) = 20.25 \pm 0.41$ pb, and $\sigma^{\mathrm{EWK}}(\mathrm{Z}) = 3.987 \pm 0.080$ pb, where an error of 2% is assumed together with the numerical integration error from the generator.

The production of a single SM Higgs boson in association to a Z boson is considered as a background to this search. Other production modes via gluon or vector boson fusion, or production in association to a top quark pair, are largely suppressed by the invariant mass requirements and are not taken into account. The simulation is realized with POWHEG 2.0 and assumes a Higgs boson of mass 125 GeV. The decay modes $\mathrm{Z}(\to \ell\ell)\mathrm{H}(\to \mathrm{bb})$, $\mathrm{Z}(\to qq)\mathrm{H}(\to \mathrm{bb})$, and $\mathrm{Z}(\to \mathrm{any})\mathrm{H}(\to \tau\tau)$ are simulated. The samples are normalized to the theoretical cross section computed at NNLO precision of the QCD corrections and at the NLO precision of electroweak corrections, that amounts to $\sigma(\mathrm{ZH}) = 0.884^{+0.036}_{-0.031}\,\mathrm{pb}$ [21].

A complete list of the simulated processes described in this section, of the number of simulated events, and of the corresponding theoretical cross section, is reported in Table 5.2.

5.7 Systematic uncertainties

The effects arising from the imperfect knowledge of the detector response, discrepancies between the simulation and the data, and uncertainties in the theoretical prediction influence the modelling of signal and background processes that was described above. These effects are taken into account as systematic uncertainties, that are included in the statistical analysis of the data in the form of nuisance parameters as described in Section 6.3 of Chapter 6. They are separately treated as "normalization" and "shape" uncertainties: the former affect the number of expected events associated to a specific process, while the latter modify the differential distribution of the discriminating observables described in Section 6.2 of Chapter 6.

5.7.1 Normalization uncertainties

The normalization uncertainties affecting the signal and background modelling are different for simulated processes and for data-driven estimates. In the former case, uncertainties

Process	MC gen.	N. evts. [$\times 10^6$]	σ [pb]
$W \to \ell\nu_\ell$ + jets	MG5		
H_T inclusive		87.7	$\sigma_W = 6.15^{+0.24}_{-0.23} \times 10^4$
$70 \leq H_T < 100\,\text{GeV}$		10.1	$3.0 \times 10^{-2} \times \sigma_W$
$100 \leq H_T < 200\,\text{GeV}$		79.4	$2.8 \times 10^{-2} \times \sigma_W$
$200 \leq H_T < 400\,\text{GeV}$		39.7	$7.6 \times 10^{-3} \times \sigma_W$
$400 \leq H_T < 600\,\text{GeV}$		7.8	$1.0 \times 10^{-3} \times \sigma_W$
$600 \leq H_T < 800\,\text{GeV}$		18.7	$2.5 \times 10^{-4} \times \sigma_W$
$800 \leq H_T < 1200\,\text{GeV}$		7.8	$1.2 \times 10^{-4} \times \sigma_W$
$120 \leq H_T < 2500\,\text{GeV}$		6.9	$2.8 \times 10^{-5} \times \sigma_W$
$H_T > 2500\,\text{GeV}$		2.6	$6.8 \times 10^{-7} \times \sigma_W$
$Z\gamma^* \to \ell\ell$ + jets	MG5		
Jet inclusive		49.1	$\sigma_Z = 5765 \pm 99$
1 jet		62.6	$2.04 \times 10^{-1} \times \sigma_Z$
2 jets		20.0	$6.74 \times 10^{-2} \times \sigma_Z$
3 jets		5.9	$2.05 \times 10^{-2} \times \sigma_Z$
4 jets		4.2	$1.11 \times 10^{-2} \times \sigma_Z$
1 b		1.5	$1.41 \times 10^{-2} \times \sigma_Z$
2 b		2.6	$2.33 \times 10^{-3} \times \sigma_Z$
Electroweak W^+ + jj	MG5	4.9	25.69 ± 0.51
Electroweak W^- + jj	MG5	4.8	20.25 ± 0.41
Electroweak Z + jj	MG5	1.0	3.987 ± 0.080
$t\bar{t}$	POW		
Inclusive		155.2	$\sigma_{t\bar{t}} = 831.8^{+46.4}_{-50.9}$
$t\bar{t} \to b\bar{b}\ell\nu_\ell qq'$		152.7	$0.438 \times \sigma_{t\bar{t}}$
$t\bar{t} \to b\bar{b}\ell^-\nu_\ell\ell^+\bar{\nu}_\ell$		79.1	$0.105 \times \sigma_{t\bar{t}}$
Single t, W channel	POW	6.9	35.9 ± 1.9
Single $\bar{t}$, W channel	POW	7.0	35.9 ± 1.9
$WW \to \ell\nu_\ell\ell'\nu_{\ell'}$	POW	2.0	$12.18^{+0.30}_{-0.27}$
$WW \to \ell\nu_\ell qq'$	POW	9.0	$50.0^{+1.2}_{-1.1}$
$WW \to qq'qq'$	POW	2.0	$51.7^{+1.3}_{-1.1}$
$ZZ \to \ell\ell\nu_\ell\nu_\ell$	POW	8.8	$0.564^{+0.017}_{-0.014}$
$ZZ \to \ell\ell qq$	MG5	15.3	$3.22^{+0.10}_{-0.08}$
$ZZ \to \ell\ell\ell'\ell'$	MG5	10.7	$1.212^{+0.037}_{-0.029}$
$ZZ \to qqq'q'$	MG5	30.5	$7.06^{+0.21}_{-0.17}$
$ZW \to \ell\ell\ell'\nu_{\ell'}$	POW	1.99	4.43 ± 0.25
$ZW \to \nu\nu\ell\nu_\ell$	MG5	1.7	3.03 ± 0.17
$ZW \to qq\ell\nu_\ell$	MG5	24.2	10.71 ± 0.60
$ZW \to \ell\ell qq'$	MG5	26.5	5.60 ± 0.31
ZH	POW		$\sigma_{ZH} = 0.884^{+0.036}_{-0.031}$
$Z(\to \text{any})H(\to \tau\tau)$		0.57	$0.0632 \times \sigma_{ZH}$
$Z(\to \ell\ell)H(\to bb)$		5.0	$0.0583 \times \sigma_{ZH}$
$Z(\to qq)H(\to bb)$		0.50	$0.403 \times \sigma_{ZH}$

Table 5.2 – Simulated processes and their corresponding generator, number of events, and normalization cross section. In the table, MG5 and POW denote MadGraph5_aMC@NLO 2.3.2 and powheg 2.0, respectively.

on trigger, reconstruction and identification efficiency of the final state objects, on the integrated luminosity, and on the theoretical cross section must be taken into account. In the latter case, uncertainties arise from the limited number of events used to perform the estimate. The systematic uncertainties considered in this search and their values are listed in the following.

- The integrated luminosity is known with an uncertainty of 2.5% [163], which is obtained from dedicated Van der Meer scans and from the stability of detector response during the data taking. Because the luminosity is used to normalize all the MC samples, its uncertainty is assumed to be fully correlated among the various final states and applies to all the simulated processes, but it is not applied to the multijet and Drell-Yan backgrounds because their data-driven estimation or correction already accounts for luminosity effects.

- Muon, electron, and τ_{h} trigger, reconstruction and identification efficiencies are measured using $\mathrm{Z} \to \mu\mu$, $\mathrm{Z} \to \mathrm{ee}$, and $\mathrm{Z} \to \tau\tau \to \tau_{\mathrm{h}}\nu_\tau\mu\nu_\mu\nu_\tau$ events collected at $\sqrt{s} = 13\,\mathrm{TeV}$, as described in Chapters 3 and 4. The corresponding uncertainties affect separately the $\tau_\mu\tau_{\mathrm{h}}$, $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final states, and are thus considered as uncorrelated for electron and muons and correlated for τ_{h} candidates. The uncertainties for the individual objects amount to 3% for electrons, 2% for muons, and 6% for τ_{h}.

- Uncertainties are considered on the τ_{h} energy scale, that is defined as the average of the ratio of the reconstructed and simulated energies. The energy scale is measured in $\mathrm{Z} \to \tau\tau \to \tau_{\mathrm{h}}\nu_\tau\mu\nu_\mu\nu_\tau$ events by fitting the event distributions of observables sensitive to its value, namely the reconstructed mass of the τ_{h} candidate and the invariant mass of the τ_{h} and μ system [97]. Different scales are observed and range between 0.2 and 2.3% depending on the decay mode of the τ lepton and on the fitted observable. A conservative approach is adopted to not modify the scale of simulated τ_{h} candidates and consider an uncertainty of 3% associated to it. The overall impact is estimated by varying the τ_{h} energy within the uncertainty and observing the changes to the signal and background acceptances from the τ_{h} selections. These include the selection on $m_{\tau\tau}$, for which the SVfit algorithm is run multiple time after applying the energy scale shift on the $\tau\tau$ pair. The resulting impact on the overall normalization ranges from 3 to 10% depending on the process and final state considered. This effect is fully correlated with a corresponding shape uncertainty in the distribution of $\mathrm{m_{T2}}$ and $\mathrm{m_{HH}^{KinFit}}$.

- Uncertainties arising from the imperfect knowledge of the jet measured energy are defined as the residual uncertainties from the jet energy correction procedure detailed in Ref. [101]. Their impact is estimated by varying the jet energy scale in the simulation within the uncertainties and assessing the changes induced in the process normalization due to the p_{T} and invariant mass selections, and have an impact of about 2% for the signal processes and of about 4% for the dominant backgrounds. It should be noted that 27 independent sources affect the jet energy scale, and only their combination in quadrature is considered here. As done for the τ_{h} energy scale, this normalization uncertainty is fully correlated with a corresponding shape uncertainty on the $\mathrm{m_{T2}}$ and $\mathrm{m_{HH}^{KinFit}}$ distributions.

- Uncertainties from the b tagging efficiency in the simulation are estimated by propagating the uncertainty on the values of the scale factors described in Section 4.4.2 of Chapter 4. The resulting uncertainty has a value ranging from 2 to 6% for the samples with genuine b jets in the final state.

- Uncertainties in the theoretical cross section used to normalize the $\mathrm{t\bar{t}}$, W+jets, single top quark, single Higgs boson, and diboson backgrounds are considered and range from 1 to 6%. The values are summarized in Table 5.2.

- The uncertainties in the three correction factors derived in the control regions with 0, 1, and 2 b-tagged jets for the $\mathrm{Z}/\gamma^* \to \ell\ell$ background are propagated from the control regions to the signal region. The correlation between them is taken into account by diagonalizing the covariance matrix in (5.8) and by introducing four nuisance parameters defined as the matrix eigenvalues, that impact the three $\mathrm{Z}/\gamma^* \to \ell\ell$ background components according to the eigenvectors of the same matrix. The resulting uncertainties are in the range 0.1-2.5%.

- An uncertainty in the multijet background normalization arises from the limited number of events observed in the same-sign sidebands used for its estimation that is described in Section 5.4. These statistical uncertainties in the number of events are propagated to the signal region in the simultaneous fit procedure, and they range between 5 and 30% depending on the final state and category.

- The uncertainties in the theoretical nonresonant signal cross section arising from scale variations result in an uncertainty in its normalization of $+4.3\%/-6.0\%$ while effects from other theoretical uncertainties such as uncertainties on α_s, PDFs and finite top quark mass effects at NNLO amount to a further 5.9% uncertainty [21]. This uncertainty is only considered when upper limits are derived on the signal strength and quoted with respect to the SM HH cross section, while they are not included in the model-independent limits on $\sigma(\mathrm{gg} \to \mathrm{HH}) \times \mathcal{B}(\mathrm{HH} \to \mathrm{bb}\tau\tau)$.

- Uncertainties in the theoretical branching fractions of the Higgs boson decays to a $\tau\tau$ and a bb pair are $\pm 1.6\%$ and $+1.2\%/-1.3\%$ respectively. As with the previous theoretical uncertainty, they are only considered when quoting a limit on HH production with respect to the SM theoretical prediction.

5.7.2 Shape uncertainties

The shape uncertainties considered are introduced to take into account the limited knowledge in the description of background processes that arise from both imperfect simulation and limited number of events from the simulation or the data sidebands. The following shape uncertainties are considered.

- Differential $\mathrm{t\bar{t}}$ p_T measurements described in [157, 156] show a systematic difference between the observed p_T spectrum and the NNLO prediction from samples generated with POWHEG and PYTHIA 8. This ratio is interpolated with a functional form $f(p_\mathrm{T}) = \exp(a - b \cdot p_\mathrm{T})$, where $a = 0.0615$ and $b = 0.005\,\mathrm{GeV}^{-1}$. The uncertainty affecting the kinematic distribution in the simulation of the $\mathrm{t\bar{t}}$ background is estimated by defining an alternative shape that is obtained by varying

the p_{T} distribution of the top quark and antiquark generated at the matrix element level according to the fit function. The weight w assigned to each event is thus $w = \sqrt{f(p_{\mathrm{T}}^{\mathrm{t}}) f(p_{\mathrm{T}}^{\bar{\mathrm{t}}})}$. This alternative shape represents a variation with respect to the nominal shape obtained directly from the simulation. Its impact on the distribution of the fitted observables that are described in more detail in Section 6.2 of Chapter 6 is shown in Figures 5.10a and 5.10b, where it can be seen that it mostly affects the high-mass tails of the distribution of the two variables.

- Uncertainties due to the limited number of simulated events entering individual bins of the distributions are considered if larger than 10%. In this case, alternative shapes are computed for a variation of each bin content corresponding to the uncertainty. All the alternative shapes are considered as uncorrelated across bins in the individual shapes. Similarly, in the multijet control region alternative shapes are taken into account by considering a Poissonian fluctuation for every bin of the multijet background shape. These uncertainties are referred to as bin-by-bin (bbb) uncertainties.

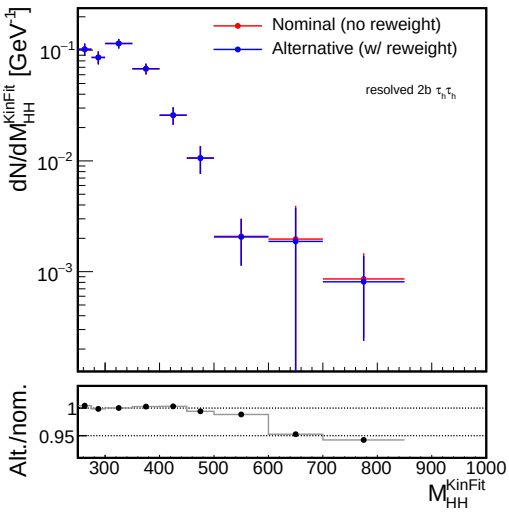
- Uncertainties due to the τ_{h} and jet energy scales are taken into account and are fully correlated with the associated normalization uncertainties. Their impact on the fitted observables is illustrated in the case of the m_{T2} variable distribution for SM HH events in Figures 5.11a and 5.11b. Uncertainties in the energy scales of other objects have negligible impacts on the simulated event distributions given the resolution and event binning chosen and are thus not taken into account.

The sources of systematic uncertainties and their values discussed are summarized in Table 5.3. The table also reports the CMS-specific names associated to the uncertainty sources that are referred to in the next section.

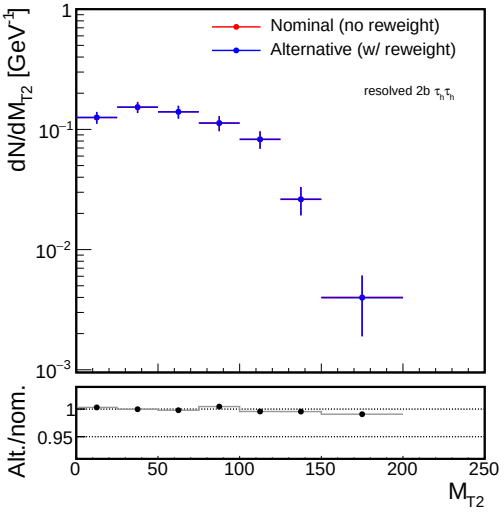
5.7.3 Impact of the systematic uncertainties

The effects of the systematic uncertainties described in the previous section on the sensitivity of the analysis are estimated by verifying the changes induced in the signal strength parameter μ (corresponding to the signal enhancement with respect to a normalization fixed to 1 pb in this case) when each source is varied by 1σ inside its associated uncertainty. Adopting the notation and following the statistical methods further discussed in Section 6.3 of Chapter 6, the impact of a nuisance parameter θ associated to a systematic uncertainty is defined as the shift $\Delta\mu$ induced on the signal strength μ by fixing θ to its postfit value plus or minus 1σ while the other nuisance parameters are estimated as the minimum of a profiled likelihood. This method thus measures the correlation of a θ with the signal strength μ. The maximum likelihood fit is performed on an "Asimov dataset" [164], i.e. a set of pseudo-random events distributed according to the likelihood function, where all the observed quantities (nuisance parameters and signal strength) are set to their expected values. The usage of the Asimov dataset allows one to verify the impact of the systematic uncertainties before looking at the actual data, and to verify the validity of the statistical model: by definition, the value and error of θ estimated after the maximum likelihood fit must correspond to the expected value and error.

In total, 197 nuisance parameters are introduced when performing the combined fit over the three $\tau\tau$ final state and the three bb categories. Most of these are constituted

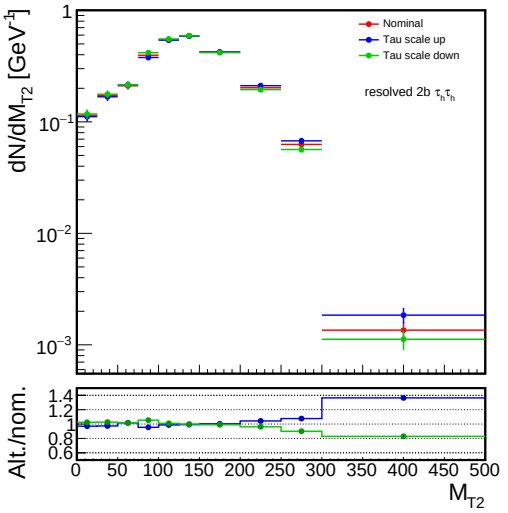


(a) Effect on m_{HH}^{KinFit}

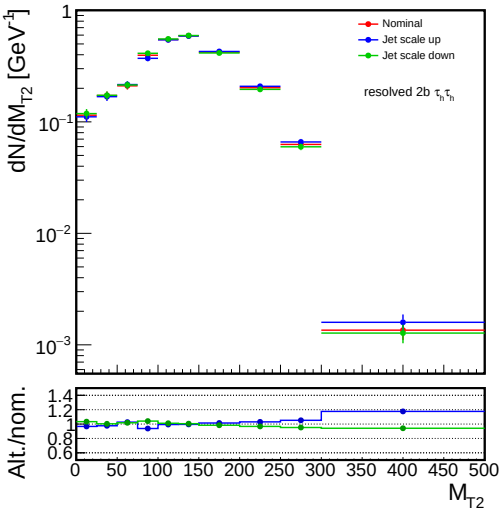


(b) Effect on m_{T2}

Figure 5.10 – Comparison of the nominal and alternative shapes defining the shape uncertainty for the m_{HH}^{KinFit} (a) and m_{T2} (b) observables.



(a) τ_h energy scale



(b) Jet energy scale

Figure 5.11 – Effect of the uncertainty on τ_h (a) and jet (b) energy scale for the m_{T2} distribution of SM HH events in the 2b $\tau_h\tau_h$ signal region of the nonresonant search.

Systematic uncertainty	Value	Processes	CMS name
	Normalization		
Luminosity	2.5%	all but mj, DY	`lumi`
Lepton trg. and reco.	2-6%	all but mj	`eff_m`, `eff_t`, `eff_e`
τ energy scale	3-10%	all but mj	`scale_t`
Jet energy scale	2-4%	all but mj	`scale_j`
b tag efficiency	2-6%	all but mj	`eff_btag`
Bkg. cross section	1-6%	all but mj, DY	`QCDscale_*`
DY SF	0.1-2.5%	DY	`DY*b_SFUnc`
Multijet norm.	5-30%	mj	(fit from data)
$\mathcal{B}(\mathrm{H} \to \tau\tau)$	$\pm 1.6\%$	signals	`BR_Htt`
$\mathcal{B}(\mathrm{H} \to \mathrm{bb})$	$+1.2\% / -1.3\%$	signals	`BR_Hbb`
Scale unc.	$+4.3\% / -6.0\%$	signals	`QCD_ggHH`
Theory unc.	5.9%	signals	`pdf_ggHH`
	Shape		
Top p_{T} rew.	-	$\mathrm{t\bar{t}}$	`top`
Bin-by-bin	-	all MC (low stat.), mj	`*_bin_*`
τ energy scale	-	all but mj	`scale_t`
Jet energy scale	-	all but mj	`scale_j`

Table 5.3 – Systematic uncertainties affecting the normalization and the distribution of the different processes. The last column reports the CMS-specific name used in the statistical interpretation, and * denotes any combination of symbols. The mj and DY symbols denote respectively the multijet and the Drell-Yan $\mathrm{Z}/\gamma^* \to \ell\ell$ backgrounds.

of shape uncertainties related to single bins of the distributions. The Asimov dataset is generated under the hypothesis that both signal and background are present, and the signal strength is fixed to 0.061, corresponding to the sensitivity on its value that is expected from an upper limit procedure detailed in Section 6.3 of Chapter 6; this choice makes it possible to observe effects from systematic uncertainties affecting both the signal and the background processes. The 30 systematic uncertainties with the largest associated impact are summarized in Figure 5.12. The left panel in the figure indicates the source of systematic uncertainty, using the internal notation of the CMS experiment as introduced in Table 5.3. The central panel reports the "pull" of each nuisance parameter, defined as the difference between its maximum likelihood estimation and its nominal value, divided by the uncertainty. No parameter is significantly pulled away from its nominal value or overconstrained by the fit procedure, indicating the quality of the statistical model. Finally, the third panel indicates the impact of each nuisance parameter on the signal strength value.

As expected, τ_{h} related uncertainties, i.e. its energy scale and efficiency, have the largest impact on the sensitivity, because they affect all the final states and categories simultaneously. Theoretical uncertainties in the normalization of the signal and background processes, as well as bin-by-bin uncertainties in bins where a large signal purity is expected have a smaller but still sizeable impact. However, the parameter with the largest impact affect the signal strength by about 8% and are thus subdominant with respect to statistical uncertainties, that with an integrated luminosity of 35.9 fb^{-1} limit the sensitivity of the search. The role of the statistical uncertainties for higher integrated luminosities is discussed in Chapter 7 in the context of the result extrapolation.

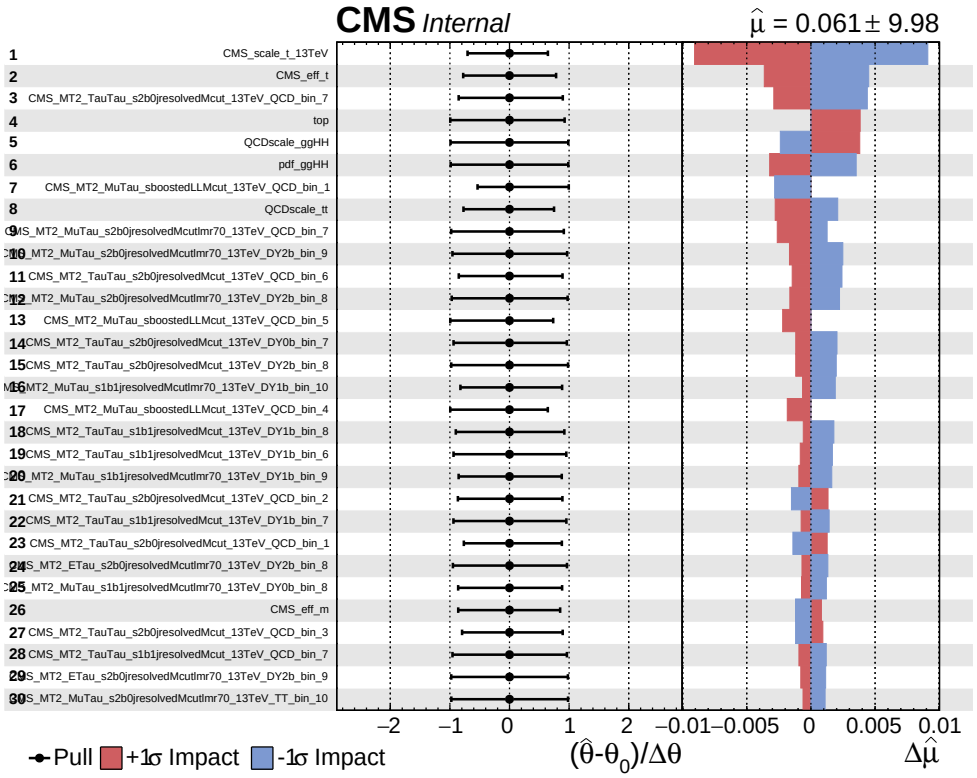


Figure 5.12 – Highest impact systematic uncertainties and relative impact on the signal strength. The central panel denotes the deviation of the value of θ from a maximum likelihood estimation with respect to the nominal value, divided by its uncertainty ("pull"), while the right panel reports the impact of each nuisance parameter. Blue and red bars denote respectively a positive and negative 1σ shift of the parameter, while the x-axis denotes the changes induced in the value of the signal strength μ.

Chapter 6

Results on HH $\rightarrow b\bar{b}\tau^+\tau^-$

Exploring HH production with the CMS data requires the definition of sensitive observables and of a statistical method to evaluate the presence of a signal, that are the object of the first part of this chapter. The results obtained on the data collected in 2016 are then described and compared to other HH searches performed at the LHC Run II.

6.1 Dataset analysed

The results presented in this chapter are derived on the full dataset collected with the CMS experiment during the 2016 data taking. A centralized data quality certification procedure is regularly performed to ensure that only the data recorded when all the detector subsystems were active and correctly functioning are analysed. The total integrated luminosity of the certified data corresponds to 35.9 fb^{-1}. The data events are divided into datasets according to specific subsets of HLT paths fired. Each of these datasets covers a separate run period denoted by a letter from B to H. The complete list of datasets analysed for the three final states of the HH $\rightarrow$ bb$\tau\tau$ search, their CMS-specific names, and the corresponding run ranges and integrated luminosities, are reported in Table 6.1.

6.2 Discriminating observables

Different observables are used in the resonant and nonresonant analyses to search for the presence of a signal contribution over the expected background. Their choice is crucial to maximize the sensitivity of the search and to provide information on the nature of the signal in case an excess is found.

In the search for resonant HH production, the invariant mass of the bb$\tau\tau$ selected candidates, m_{HH}, is used. The choice directly stems from the expected signal signature, corresponding to a localized excess over a broadly distributed background. The presence of neutrinos in the final state deteriorates the resolution on m_{HH} and a kinematic fit is thus applied to correct for this effect, the resulting variable being denoted as $\text{m}_{\text{HH}}^{\text{KinFit}}$. The fit technique, already employed in the HH $\rightarrow$ bb$\tau\tau$ search with Run I data [143], is based on the hypothesis of the decay of two Higgs bosons of $m_{\text{H}} = 125\,\text{GeV}$ to a bb quark and a $\tau\tau$ lepton pair. This constraint is formulated as:

$$m_{\text{bb}} = m_{\tau\tau} = m_{\text{H}} \tag{6.1}$$

Dataset Name	Run range	Luminosity [fb^{-1}]
$\tau_\mu\tau_\text{h}$ final state		
/SingleMuon/Run2016B-03Feb2017_ver2-v2	273150 - 275376	5.79
/SingleMuon/Run2016C-03Feb2017-v1	275656 - 276283	2.57
/SingleMuon/Run2016D-03Feb2017-v1	276315 - 276811	4.25
/SingleMuon/Run2016E-03Feb2017-v1	276831 - 277420	4.01
/SingleMuon/Run2016F-03Feb2017-v1	277932 - 278808	3.10
/SingleMuon/Run2016G-03Feb2017-v1	278820 - 280385	7.54
/SingleMuon/Run2016H-03Feb2017_ver2-v1	281613 - 284035	8.39
/SingleMuon/Run2016H-03Feb2017_ver3-v1	284036 - 284044	0.22
$\tau_\text{e}\tau_\text{h}$ final state		
/SingleElectron/Run2016B-03Feb2017_ver2-v2	273150 - 275376	5.79
/SingleElectron/Run2016C-03Feb2017-v1	275656 - 276283	2.57
/SingleElectron/Run2016D-03Feb2017-v1	276315 - 276811	4.25
/SingleElectron/Run2016E-03Feb2017-v1	276831 - 277420	4.01
/SingleElectron/Run2016F-03Feb2017-v1	277932 - 278808	3.10
/SingleElectron/Run2016G-03Feb2017-v1	278820 - 280385	7.54
/SingleElectron/Run2016H-03Feb2017_ver2-v1	281613 - 284035	8.39
/SingleElectron/Run2016H-03Feb2017_ver3-v1	284036 - 284044	0.22
$\tau_\text{h}\tau_\text{h}$ final state		
/Tau/Run2016B-03Feb2017_ver2-v2	273150 - 275376	5.79
/Tau/Run2016C-03Feb2017-v1	275656 - 276283	2.57
/Tau/Run2016D-03Feb2017-v1	276315 - 276811	4.25
/Tau/Run2016E-03Feb2017-v1	276831 - 277420	4.01
/Tau/Run2016F-03Feb2017-v1	277932 - 278808	3.10
/Tau/Run2016G-03Feb2017-v1	278820 - 280385	7.54
/Tau/Run2016H-03Feb2017_ver2-v1	281613 - 284035	8.39
/Tau/Run2016H-03Feb2017_ver3-v1	284036 - 284044	0.22

Table 6.1 – List of the datasets analysed for each $\tau\tau$ final state. The first column reports the CMS-specific dataset name, where the set of HLT triggers, the data taking period, and the data reconstruction campaign are indicated. The second and third columns report the run intervals associated to a specific dataset and the corresponding integrated luminosity.

The neutrinos from the τ decays are assumed to be nearly collinear to the corresponding visible τ decay products (collinear approximation) and to directly produce the missing transverse momentum vector of the event. Both assumptions are well verified for the resonant HH production signals studied in this search. It is also assumed that the η and φ coordinates of the two b jets and of the reconstructed τ decay products are accurately determined and that their uncertainties are negligible with respect to those arising from

the energy reconstruction of the same objects. Combining these approximations and the constraint in Eq. (6.1), the $\mathrm{bb}\tau\tau$ system can be described by two parameters, that are chosen as the energy of one of the two b jets (E_{b_1}) and of one of the two τ leptons (E_{τ_1}). A χ^2 minimization is performed by varying these two parameters around their measured quantities, and using the resulting estimates from the fit to compute the mass of the resonance X decaying to a HH pair.

The χ^2 term for the $\mathrm{H} \to \mathrm{bb}$ decay hypothesis for the jet pair is computed as:

$$\chi^2_{\mathrm{b}_i} = \left(\frac{E^{\mathrm{fit}}_{\mathrm{b}_i} - E^{\mathrm{meas}}_{\mathrm{b}_i}}{\sigma_{\mathrm{b}_i}} \right)^2 \tag{6.2}$$

where the index $i = 1, 2$ denotes the jet, and $E^{\mathrm{fit}}_{\mathrm{b}_i}$ and $E^{\mathrm{meas}}_{\mathrm{b}_i}$ the fitted and measured energies, respectively, and σ_{b_i} the experimental resolution on $E^{\mathrm{meas}}_{\mathrm{b}_i}$.

For the $\tau\tau$ pair, the large momentum carried away from the neutrinos does not allow for setting a similar constraint, but contributes to the missing transverse momentum vector $\vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,miss}}$, where the superscript is added to indicate that this is the quantity obtained from data with the reconstruction algorithm discussed in Section 4.3.4 of Chapter 4. The missing transverse momentum can thus be exploited by comparing the expected transverse momentum of the resonance X:

$$\vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,X}} = \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,H_1}} + \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,H_2}} = \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,b_1}} + \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,b_2}} + \vec{p}^{\,\mathrm{fit}}_{\mathrm{T},\,\tau_1} + \vec{p}^{\,\mathrm{fit}}_{\mathrm{T},\,\tau_2} \tag{6.3}$$

with the one measured experimentally:

$$\vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,X}} = \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,H_1}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,H_2}} = \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,b_1}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,b_2}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T},\,\tau_1^{\mathrm{vis}}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T},\,\tau_2^{\mathrm{vis}}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,miss}} \tag{6.4}$$

By definition, the quantity in Eq. (6.4) corresponds to the measured recoil of the X resonance:

$$\vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,recoil}} = -\vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,X}} \tag{6.5}$$

that allows for writing the constraint in terms of the residual recoil vector:

$$\vec{p}^{\,\mathrm{res}}_{\mathrm{T,\,recoil}} = \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,X}} - \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,X}} = \vec{p}^{\,\mathrm{fit}}_{\mathrm{T,\,X}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{T,\,recoil}} \tag{6.6}$$

The corresponding χ^2 term for the minimization is defined as:

$$\chi^2_{\mathrm{recoil}} = \left(\vec{p}^{\,\mathrm{res}}_{\mathrm{T,\,recoil}}\right)^{\mathrm{t}} \cdot V_{\mathrm{recoil}} \cdot \vec{p}^{\,\mathrm{res}}_{\mathrm{T,\,recoil}} \tag{6.7}$$

where V_{recoil} is the covariance matrix of the reconstructed recoil vector.

All the terms for the minimization are thus determined and the complete expression reads:

$$\chi^2 = \chi^2_{\mathrm{b}_1} + \chi^2_{\mathrm{b}_2} + \chi^2_{\mathrm{recoil}} \tag{6.8}$$

Given the constraints described above and the kinematic properties of the $\mathrm{bb}\tau\tau$ system, the three terms on the right side of the equation are expressed as functions of E_{b_1} and E_{τ_1}. A minimization of Eq. (6.8) is performed by varying these two parameters, and their estimates are used to compute the value of $\mathrm{m_X}$.

The usage of the kinematic fit improves the resolution on the invariant mass of the HH system by about a factor of four over the entire mass range studied, as it can be seen from

Figure 6.1 for several values of m_X. A comparison of the distributions of m_{HH} and m_{HH}^{KinFit} is illustrated in Figure 6.2 for the $\tau_\mu\tau_h$ final state in the resolved 2b event category. The bottom panels show the ratio of the signal expectation, arbitrarily normalized to $\sigma(gg \to S) \times \mathcal{B}(S \to HH) = 10$ pb, to the background.

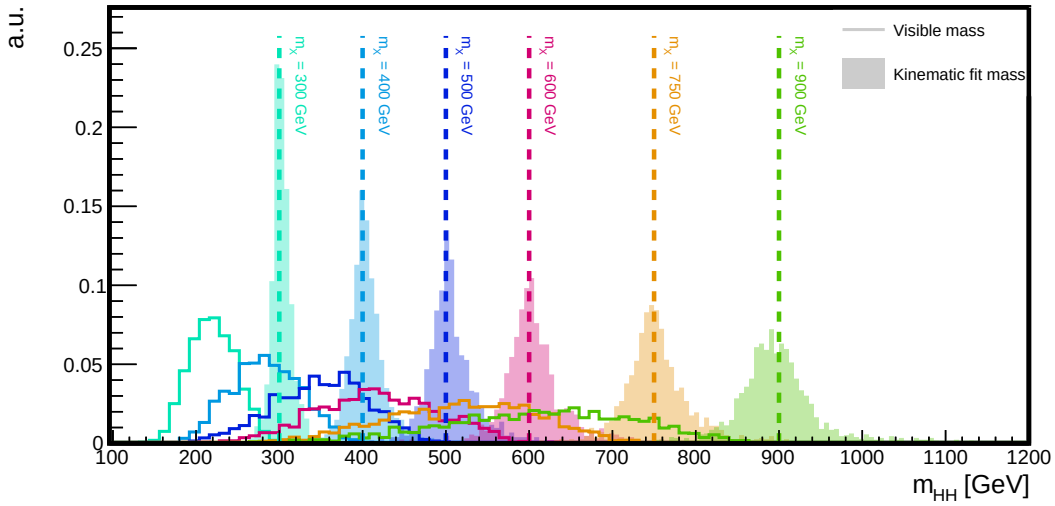
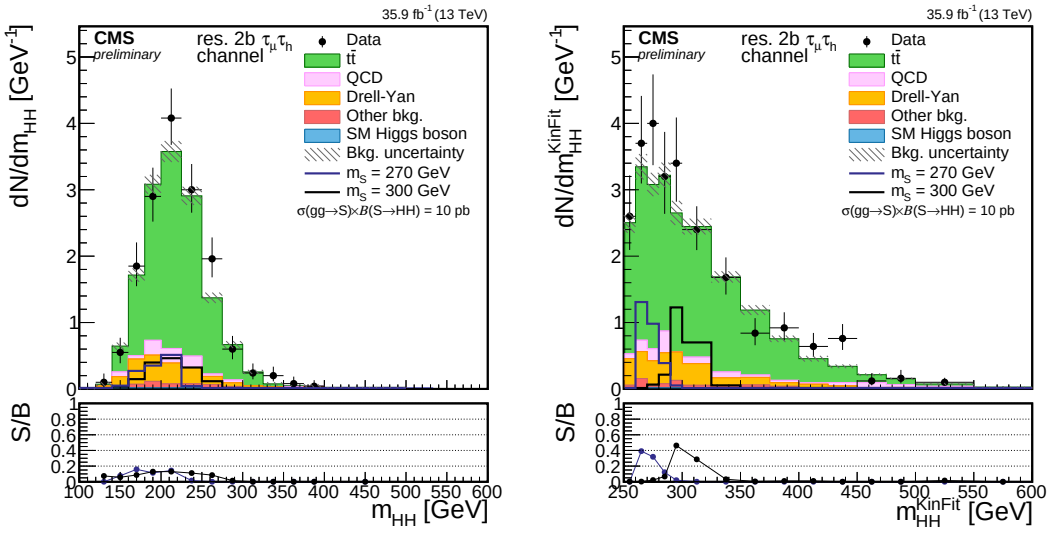


Figure 6.1 – Comparison of m_{HH} estimated as the visible $bb\tau\tau$ invariant mass (solid lines) and with the kinematic fit (shaded histograms). The usage of the kinematic fit improves the resolution by about a factor of four, and is capable of reconstructing the correct value of m_X, denoted by the vertical dashed line. Visible and kinematic fit mass histograms are normalized to a unitary area. The performance of the kinematic fit is illustrated using a combination of the events selected in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states with two b-tagged jets, and similar results are observed in the other event categories and in the $\tau_h\tau_h$ final state.



(a) Visible HH mass

(b) Kinematic fit HH mass

Figure 6.2 – Event distribution in the $\tau_\mu\tau_h$ final state, 2b resolved event category, for the LM regime. The visible HH invariant mass (a) and the HH mass reconstructed with the kinematic fit m_{HH}^{KinFit} (b) are shown. The lower panels indicate the ratio of the signal expectation to the background for $\sigma(gg \to S) \times \mathcal{B}(S \to HH) = 10$ pb.

Because of its different m_{HH} distribution, the identification of a nonresonant HH signal requires the usage of a distinct technique. In the absence of a clear peak, the signal contribution is more difficult to disentangle from the background using the m_{HH} variable, even when reconstructed with the kinematic fit. A different variable, denoted as "stransverse mass" or $\mathrm{m_{T2}}$, is thus used to achieve a better separation from the background. This variable was originally introduced for supersymmetry searches involving invisible particles in the final state [165, 166] and later proposed for HH searches in $\mathrm{bb}\tau\tau$ events [70]. It is conceived as a multi-body generalization of the transverse mass m_{T} and aims at exploiting the kinematic information of events where two equal mass particles are produced and each undergoes a two-body decay into a visible and an invisible particle, a situation illustrated in Figure 6.3. In this case, the $\mathrm{m_{T2}}$ variable is defined as the largest mass of the parent particle that is compatible with the kinematic constraints of the event, and is a function of the mass of the invisible particle that is assumed when evaluating these constraints.

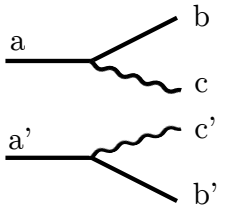


Figure 6.3 – Schematic representation of a two-body decay of two equal mass particles $a^{(\prime)}$ to a visible particle $b^{(\prime)}$ and an invisible particle $c^{(\prime)}$. The $\mathrm{m_{T2}}$ variable is defined as the maximum mass of a that is compatible with the kinematic constraints of the event from the measured b and b' momenta and the missing transverse momentum from c and c'.

In the case of the $\mathrm{bb}\tau\tau$ final state, where the dominant background is $\mathrm{t\bar{t}}$ production, the parent particle is interpreted as the top quark that decays into a bottom quark and a W boson. Following the description in [70], the symbols $\vec{b}$, $\vec{b}'$ denote the momenta of the two selected b jets and m_b, $m_{b'}$ their invariant masses. The momenta of the other particles produced in the top quark decay, i.e. the measured leptons and the neutrinos, are globally denoted as $\vec{c}$ and $\vec{c}'$, and their masses are set to $m_c = m^{\mathrm{vis}}(\tau_1)$ and $m_{c'} = m^{\mathrm{vis}}(\tau_2)$, where m^{vis} denotes the invariant mass of the measured leptons or τ_{h}. The "transverse energy" e of an object of transverse momentum p_{T} and mass m is defined as:

$$e = \sqrt{m^2 + p_{\mathrm{T}}^2} \tag{6.9}$$

Under this notation, the transverse mass is:

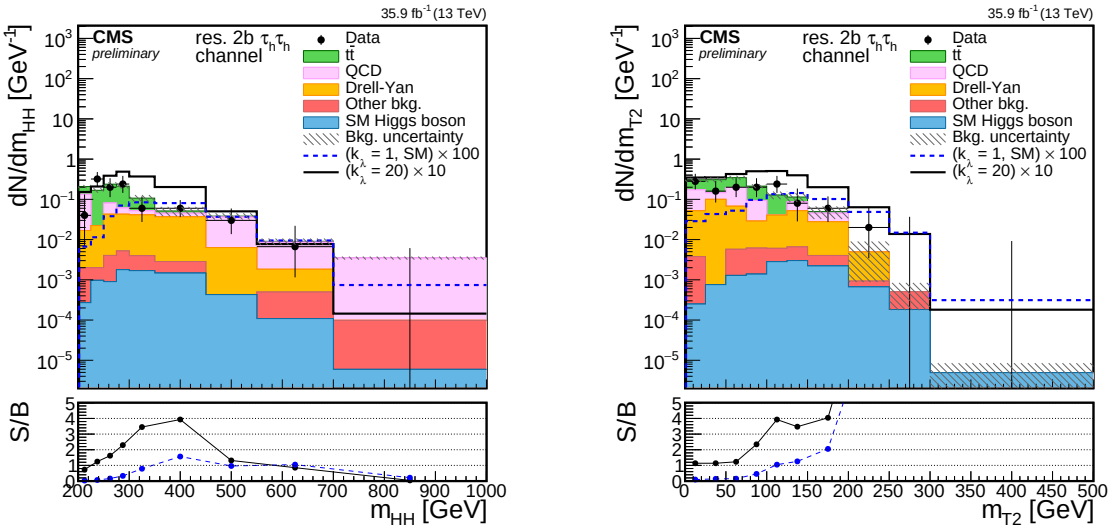
$$\begin{aligned} m_{\mathrm{T}}\left(\vec{b}_{\mathrm{T}}, \vec{c}_{\mathrm{T}}, m_b, m_c\right) &= \sqrt{m_b^2 + m_c^2 + 2\left(e_b e_c - \vec{b}_{\mathrm{T}} \cdot \vec{c}_{\mathrm{T}}\right)} \\ &= \sqrt{\left(e_b + e_c\right)^2 - \left(\vec{b}_{\mathrm{T}} + \vec{c}_{\mathrm{T}}\right)^2} \end{aligned} \tag{6.10}$$

a definition that corresponds to the one already introduced in Eq. (4.8) of Chapter 4 in the limit of two massless b and c particles. The $\mathrm{m_{T2}}$ variable is defined from these quantites as:

$$\mathrm{m_{T2}}\left(m_b, m_{b'}, \vec{b_{\mathrm{T}}}, \vec{b'_{\mathrm{T}}}, \vec{\Sigma}_{\mathrm{T}}, m_c, m_{c'}\right) = \min_{\vec{c_{\mathrm{T}}} + \vec{c_{\mathrm{T}}}' = \vec{\Sigma}_{\mathrm{T}}} \left\{\max\left(m_{\mathrm{T}}, m'_{\mathrm{T}}\right)\right\} \tag{6.11}$$

where the constraint in the minimization is over the measured lepton momenta and the missing transverse momentum, i.e. $\vec{\Sigma}_{\mathrm{T}} = \vec{p_{\mathrm{T}}}^{\,\mathrm{vis}}(\tau_1) + \vec{p_{\mathrm{T}}}^{\,\mathrm{vis}}(\tau_2) + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The implementation provided in [167] is used to perform the minimization of Eq. (6.11).

The $\mathrm{m_{T2}}$ variable is by construction bounded above by the top quark mass m_{t} for the irreducible background process $\mathrm{t\bar{t}} \to \mathrm{bb\,WW} \to \mathrm{bb}\,\tau\nu_\tau\tau\nu_\tau$, while it can assume larger values for the HH signal where the τ and the b jet do not originate from the same parent particle. Detector resolution effects and other decay modes of the $\mathrm{t\bar{t}}$ system (such as jets from the W boson misidentified as τ_{h}) result in an extension of the tail of the $\mathrm{m_{T2}}$ distribution in $\mathrm{t\bar{t}}$ events beyond the m_{t} value. With the usage of the $\mathrm{m_{T2}}$ variable, the presence of a signal contribution would appear as an enhancement of the event yield in the tails of its distribution, largely improving the sensitivity with respect to the usage of m_{HH}. This is illustrated in Figure 6.4, where the two variables are compared for the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state in association to two b jets. The usage of the $\mathrm{m_{T2}}$ variable instead of m_{HH} leads to an improvement of the sensitivity of about 40% for SM HH production and of about 15% for an anomalous coupling of $k_\lambda = 20$.



(a) m_{HH} (b) $\mathrm{m_{T2}}$

Figure 6.4 – Expected event distribution in the resolved 2b $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state for the m_{HH} (a) and $\mathrm{m_{T2}}$ (b) variables. The lower panel shows the ratio of the signal expectation to the background. The better separation achieved with the usage of $\mathrm{m_{T2}}$ can be observed.

6.3 Statistical interpretation

Once the selection of the events, the estimation of the backgrounds, and the evaluation of the systematic uncertainties have been established, a statistical procedure is needed to evaluate the presence or absence of a signal in the observed data. The statistical framework used for this search corresponds to the frequentist approach adopted by the ATLAS and CMS collaborations in the context of the Higgs analyses combination [168].

6.3.1 Likelihood function and nuisance parameters

In the following, the expected signal event yield is denoted as s and the total background event yield is denoted as b. As binned distributions are used in the search for HH production presented here, s and b are vectors. They represent the event yield expectations in all the bins of the distributions considered in the three final states and event categories. For a model-independent search the signal normalization is arbitrarily fixed ($\sigma \times \mathcal{B} = 1\,\text{pb}$ in this case) and is scaled by a signal strength modifier factor μ. The predictions of s and b are subject to the systematic uncertainties that have been discussed in Section 5.7 of Chapter 5. Any of these sources of systematic uncertainties is represented as a "nuisance parameter" θ_i The entire set of θ_i considered is collectively denoted as θ, so that $s \equiv s(\theta)$ and $b \equiv b(\theta)$. An example of θ_i is the uncertainty on the integrated luminosity, that is known with a precision of 2.5% and affects the normalization of both the signal and background estimations. The shape uncertainties related to the τ_h and jet energy scales do not impact the total values of s and b but affect their repartition among the bins considered. In general, nuisance parameters affect the statistical model but do not provide an interesting information, in contrast with the strength μ of the signal that we are looking for.

The estimation of the nuisance parameters from a priori considerations or auxiliary measurements (with events independent from those entering the HH signal regions) is denoted as $\tilde{\theta}_i$. Our degree of belief on what is the real value of θ_i given its $\tilde{\theta}_i$ estimation is represented with the Bayesian probability density function $\rho(\theta_i|\tilde{\theta}_i)$. Thanks to the Bayes' theorem, this is reinterpreted as a frequentist probability $p(\tilde{\theta}_i|\theta_i)$ with the usage of a prior that can be shown to be flat with the functional forms of ρ that are described later in this section. Because all the systematic uncertainty sources are assumed to be uncorrelated, the combined pdf $p(\tilde{\theta}|\theta)$ is the product of the individual p_i, i.e. $p(\tilde{\theta}|\theta) = \prod_i p_i(\tilde{\theta}_i|\theta_i)$

Given an observation n, the likelihood function $\mathcal{L}$ is written as:

$$\mathcal{L}(n, \tilde{\theta}|\mu, \theta) \equiv P\left(n|\mu s + b\right) \cdot p(\tilde{\theta}|\theta) \tag{6.12}$$

where P denotes the evaluation of the probability density function of the events, that depends on μs and b, on the data n. The latter can refer to either the measured data, or to the values obtained from a "pseudo-experiment", i.e. a set of pseudo-random numbers that are sampled from the expected distributions given μs, b, and θ. For binned distributions, P is expressed as the product of the Poisson probabilities for every bin j considered:

$$P\left(n|\mu \cdot s + b\right) = \prod_j \frac{\left(\mu s_j + b_j\right)^{n_j}}{n_j!} e^{-\left(\mu s_j + b_j\right)} \tag{6.13}$$

Two cases are considered to select a functional form of $\rho(\theta_i|\tilde{\theta}_i)$. In case the uncertainty arises from an independent measurement, such as those related to the luminosity or trigger efficiencies, the log-normal function is used and is defined as:

$$\rho(\theta_i|\tilde{\theta}_i; \kappa) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{\left(\ln(\theta_i/\tilde{\theta}_i)\right)^2}{2\left(\ln\kappa\right)^2}\right)\frac{1}{\theta} \tag{6.14}$$

For small uncertainties the log-normal distribution with $\kappa = 1 + \varepsilon$ is asymptotically

identical to a Gaussian distribution of width ε, but has the advantage to correctly describe positively defined observables by going to zero a $\theta_i = 0$.

Uncertainties on the background estimated from the data in control regions are treated with a separate approach. It is the case of the multijet background estimation with same-sign $\tau\tau$ events that was described in Section 5.4 of Chapter 5. The estimation consists in computing the number of events n in the signal regions from the number of events N in the control region. A transfer factor α is used, resulting in $n = \alpha N$. In this case, the uncertainty on n is described by the gamma distribution:

$$\rho(n|N;\alpha) = \frac{1}{\alpha}\frac{(n/\alpha)^N}{N!}\exp(-n/\alpha) \tag{6.15}$$

that represents the propagation of the statical uncertainties on $N = n/\alpha$ from the control region to the signal region.

6.3.2 Hypothesis testing

The likelihood formalism sets the basis for an hypothesis testing procedure that is used to quantify the presence or absence of a signal in the observed data. The procedure makes use of two hypotheses in which a signal contribution is present or absent, respectively denoted as $H_{\mu s+b}$ and H_b. To set a limit on the presence of a signal, one has to find a value of μ for which the $H_{\mu s+b}$ hypothesis can be excluded in favour of the H_b one. Inversely, if a signal excess is present and needs to be quantified, the compatibility of the observed data with the H_b hypothesis is expressed in terms of a probability or p-value. The hypothesis test is done by means of the "test statistics", a quantity that is derived from the likelihood of Eq. (6.12).

Setting an upper limit

The test statistic q_μ used to set exclusion limits is computed as:

$$q_\mu \equiv -2\ln\frac{\mathcal{L}(n,\tilde{\theta}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(n,\tilde{\theta}|\hat{\mu},\hat{\theta})}, \qquad \text{with } 0 \leq \hat{\mu} \leq \mu \tag{6.16}$$

The symbol $\hat{\theta}_\mu$ in the numerator denotes the conditional maximum likelihood estimator of θ, i.e. the value of θ that maximises the likelihood of Eq. (6.12) for a fixed μ, and thus $\hat{\theta}_\mu$ is a function of μ itself. The symbols $\hat{\mu}$ and $\hat{\theta}$ in the denominator denote instead the global minimum of the likelihood function, i.e. the values obtained when minimization is performed on both parameters simultaneously. The constraint $\hat{\mu} \geq 0$ indicates that the signal contribution cannot be negative, and the other constraint $\hat{\mu} \leq \mu$ ensures that upwards fluctuations of the data, if larger than expected for a signal of strength μ, are not regarded as evidence against the signal hypothesis itself. The definition implies that larger values of q_μ represent increasing incompatibility between the data n and the hypothesized value of μ.

This definition of the q_μ test statistics slightly differs from what has been used in searches at the LEP or Tevatron colliders. It has been adopted for LHC experiments because of the asymptotic properties of Eq. (6.16), that make it possible to derive the q_μ distribution under the $H_{\mu s+b}$ and H_b hypotheses with analytical formulae [164] instead of

using pseudo-experiments when a large number of background events is expected. This asymptotic approximation is used in deriving the results presented in this chapter. It has been verified for a few signal hypotheses that these formulae give the same result as a a complete computation based on pseudo-experiments.

Exclusion limits are computed using q_μ in the modified frequentist criterion $\mathrm{CL_s}$ [169, 170]. Given an observed value of the test statistic q_μ^{obs} (that is obtained by computing the value of Eq. (6.16) using the data n) the probability for q_μ to be equal or larger than q_μ^{obs} under the $H_{\mu s+b}$ and the H_b hypotheses is defined as:

$$\begin{aligned} \mathrm{CL_{s+b}}(\mu) &\equiv P(q_\mu \geq q_\mu^{\mathrm{obs}} | H_{\mu s+b}) \\ \mathrm{CL_b}(\mu) &\equiv P(q_\mu \geq q_\mu^{\mathrm{obs}} | H_b) \end{aligned} \tag{6.17}$$

where the dependence of the two probabilities on the value of μ has been explicitly indicated. The $\mathrm{CL_s}$ quantity is defined as:

$$\mathrm{CL_s}(\mu) \equiv \frac{\mathrm{CL_{s+b}}(\mu)}{\mathrm{CL_b}(\mu)} \tag{6.18}$$

A signal of strength μ is said to be excluded at a confidence level (CL) of α if $\mathrm{CL_s}(\mu) < 1 - \alpha$. It is customary to compute exclusions limits for $\alpha = 95\%$, with the value of μ being varied until the condition is satisfied. The usage of the $\mathrm{CL_s}$ criterion is preferable with respect to an exclusion based on $\mathrm{CL_{s+b}}(\mu)$, because the construction in Eq. (6.18) protects against underfluctuations in the background that could lead to the exclusion of a signal even if the latter is absent. It should also be noted that the $\mathrm{CL_s}$ criterion results in a limit on the value of μ, that is converted into a limit in the signal cross section by a simple rescale of the signal normalization initially fixed.

Quantifying and excess

The test statistics q_0 used to quantify an event excess is:

$$q_0 \equiv -2 \ln \frac{\mathcal{L}(n, \tilde{\theta} | 0, \hat{\theta}_0)}{\mathcal{L}(n, \tilde{\theta} | \hat{\mu}, \hat{\theta})} \qquad \text{with } \hat{\mu} \geq 0 \tag{6.19}$$

a definition that corresponds to the test statistics used for upper limits with a signal strength $\mu = 0$.

The significance of an excess is measured as the probability of q_0 to be larger or equal than the observed q_0^{obs} under the assumption that no signal is present:

$$p \equiv P(q_0 \geq q_0^{\mathrm{obs}} | H_b) \tag{6.20}$$

This quantity, or p-value, corresponds to the probability of the background to fluctuate and give an excess that is at least large as the one observed in the data. It is by definition independent on the normalization of the signal, although the value of $\hat{\mu}$ retains the information on the magnitude of the excess compared to the expectation. The p-value is usually converted to a Z significance by expressing it as a one-sided Gaussian integral:

$$p = \int_Z^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x \tag{6.21}$$

It is usual to regard a significance of 3σ ($Z = 3$, $p = 1.3 \times 10^{-3}$) as an "evidence" and a significance of 5σ ($Z = 5$, $p = 2.8 \times 10^{-7}$) as an "observation" of a signal. It should be noted that when searches for different signals are performed simultaneously, such as when looking at the presence of a resonance under different m_X mass hypotheses, the p value underestimates the probability to observe a significant fluctuation for any of the hypothesized signals. A procedure to correct for the combined probability associated to this multiple testing and to compute a "global" p-value, in contrast with the "local" p-value introduced above, is detailed in Ref. [168].

6.3.3 Validation of the statistical model

The statistical procedure described above defines an objective way to quantify the presence of a signal or to put an upper limit on its magnitude using the measured data. Both the CL_s and p-value quantities are defined in a frequentist approach that makes it possible to interpret them as probabilities of an observed result under well defined hypotheses. Still, a danger lurks beneath the method and can distort the inference that we draw from the data: it is represented by the physicist that develops the analysis methods. Usually the physicist is not himself a danger, but he might introduce a bias when defining the event selections and the analysis methods. If the choice is based on the observed data, the probability to observe an excess cannot be computed a priori with the methods discussed in the previous section, as it becomes a function of the data themselves. A typical example is a fine tuning of the kinematic selections to maximise the number of observed signal-like events in real data: in this case, the probability to observe an excess of events is artificially increased when defining the selections themselves. In other words, a definition of the analysis strategy that is based on the actual observed data introduces a bias that cannot be taken into account in the statistical procedure defined above.

To prevent that from happening, a so-called "blind analysis" procedure is followed. It consists in defining, developing, and validating the analysis strategy using only the MC simulation or data in signal-free control regions. Only once the procedure established, it is possible to look at the data in the signal region (the thrilling "unblinding" moment) and to establish if a signal is present. Blind analyses are a standard procedure in searches performed by the CMS Collaboration and the one presented in this thesis is no exception. Every time a new dataset was available, the search was developed and optimized in a blind way and the unblinding was preceded by detailed checks of the quality of the physics modelling.

Important checks performed are the control of the impact of the nuisance parameters and of their "pulls", defined as the difference between the $\hat{\theta}$ estimation from the likelihood maximisation and the initial value of the nuisance parameter, normalized to the value of the uncertainty. This verification is shown in Figure 6.5 using the nonresonant signal in a combined likelihood maximisation simultaneously performed on the three final states and three event categories of the search. Only the uncertainties with the larger impact on the analysis are shown. The dominant uncertainties are related to τ reconstruction efficiency and energy scale, and are followed in importance by top p_T reweighting, signal and background normalization, jet energy scale, and bin-by-bin uncertainties in bins with a large signal to background ratio. These observation are in agreement with those previously formulated in Section 5.7.3 of Chapter 5, where the maximum likelihood fit

was performed on the Asimov dataset instead of the observed data. Similar results are observed for other signal hypotheses.

The only uncertainties that are significantly overconstrained are the jet and τ scale uncertainties. Both effects are expected from the treatment of these two systematic uncertainties discussed in Section 5.7.1 of Chapter 5. A conservative τ scale uncertainty of 3% has been assumed, and is larger than the one measured to cover discrepancies between the different τ decay modes. Also, the 27 different sources of uncertainties are considered together as a single nuisance parameter, which in turn leads to an expected overconstrain of the uncertainty that has no side effects on the search. No other unexpected deviations or overconstraints of the nuisance parameters are observed.

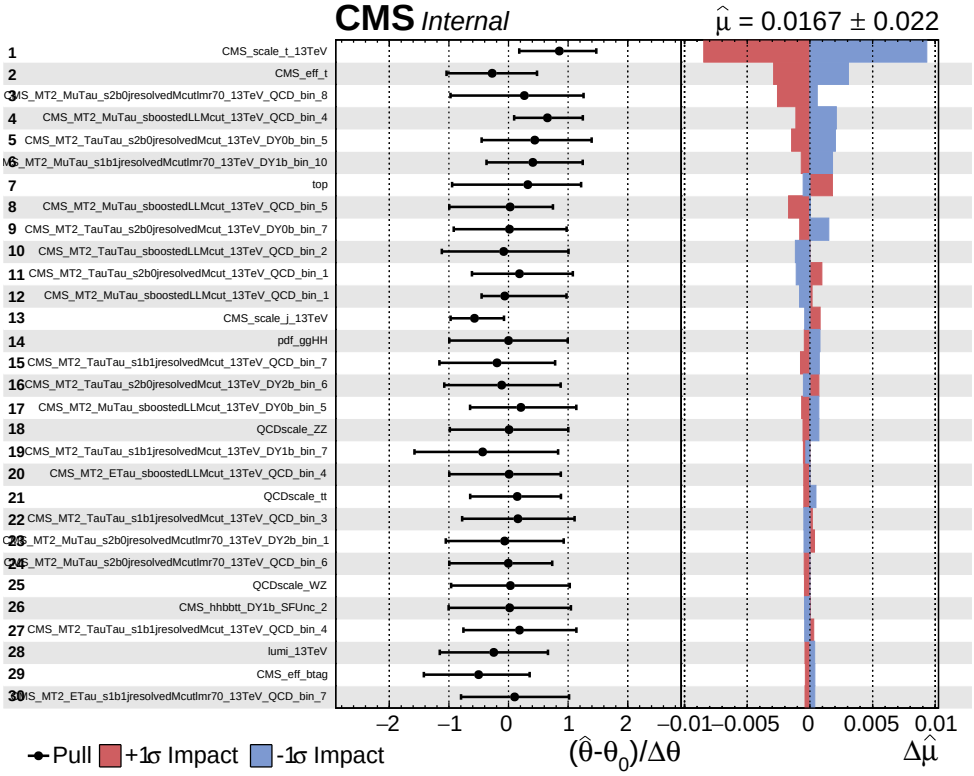


Figure 6.5 – Highest impact systematic uncertainties, corresponding impact on the signal strength μ, and value and error of the corresponding nuisance parameter after the maximum likelihood fit to the observed data.

Further verification of the correct modelling of the data is provided by the goodness-of-fit test with the "saturated" method [171]. It is defined as a generalization of the χ^2 method for data that are not normally distributed, such as the number of events in the binned distributions used in this search. The test is constructed as a likelihood ratio, where the alternate hypothesis is taken as the one that matches exactly the data. The distribution of this test statistics is obtained from pseudo-experiments generated from the modelled background and is compared to the one computed with the observed data. The result is detailed in Figure 6.6. The observed value is found to lie in the bulk of the expected distribution, showing that the background models well the data.

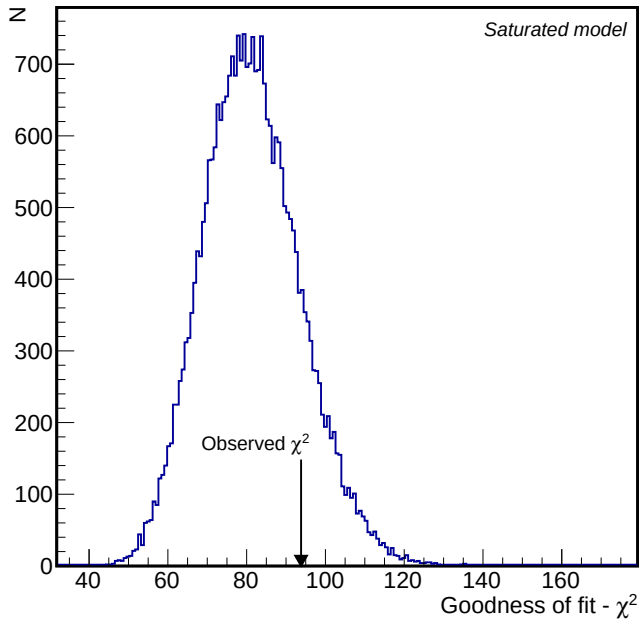


Figure 6.6 – Distribution of the χ^2 goodness of fit variable computed with the saturated model (blue line) and observed value (black arrow) from the simultaneous fit to the data performed on all the final states and event categories.

6.4 Results

In both the resonant and nonresonant searches, model-independent results are derived under general assumptions on the signal kinematics. These results can thus be reinterpreted in a variety of physics models, a few of which are explored and discussed in this section.

6.4.1 Event yields and distributions

As discussed in Chapter 4, both the the resonant and nonresonant searches explore three $\tau\tau$ final states ($\tau_\mu\tau_\mathrm{h}$, $\tau_\mathrm{e}\tau_\mathrm{h}$, $\tau_\mathrm{h}\tau_\mathrm{h}$) and make use of three event categories of quality and topology of the selected bb candidates (resolved 1b1j, resolved 2b, boosted). Multivariate methods are used to reject the background in the $\tau_\mu\tau_\mathrm{h}$ and $\tau_\mathrm{e}\tau_\mathrm{h}$ final states, and in the resonant search two different discriminants are used for the low-mass (LM) and high-mass (HM) regions, corresponding to signal mass hypotheses $\mathrm{m_X} \leq 350\,\mathrm{GeV}$ and $\mathrm{m_X} > 350\,\mathrm{GeV}$ respectively.

The expected and observed event yields in the signal regions of the $\tau_\mu\tau_\mathrm{h}$ final state are summarized in Tables 6.2 and 6.5 for the resonant and nonresonant searches, respectively. The table also illustrates the expected event yields for a few signal processes. Similarly, observed event yields for the resonant and nonresonant searches in the $\tau_\mathrm{e}\tau_\mathrm{h}$ final state are shown in Tables 6.3 and 6.6. For the $\tau_\mathrm{h}\tau_\mathrm{h}$ final state, the resonant and nonresonant signal regions are identically defined, and the event yields are summarized in Table 6.4. In all the tables the background event yields are computed by fixing the nuisance parameters to their best estimate from a fit to the data under the background-only hypothesis, following

the procedure detailed in Section 6.3. The fit is performed simultaneously on all the final states and signal regions considered. The distribution of the m_{HH}^{KinFit} and m_{T2} variables for these events are shown in Figures 6.7, 6.8, and 6.9 for the three final states. The binning of the distributions has been determined before looking at the actual data as the best compromise between the resolution on m_{HH}^{KinFit} or m_{T2} and the statistical fluctuations of the background from limited event statistics in the simulation or in the data sidebands. Displays of two selected events in the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ final states are shown in Figures 6.10 and 6.11.

No evidence for the presence of a signal is found in the final states and event categories studied. The data are thus used to set upper limits on the resonant and nonresonant HH production cross section.

$\tau_\mu\tau_h$ final state					
Process	res. 1b1j		res. 2b0j		boosted
	LM	HM	LM	HM	
$t\bar{t}$	523.1 ± 19.2	507.4 ± 15.5	263.5 ± 11.2	267.1 ± 8.4	18.2 ± 1.0
QCD	266.2 ± 29.2	-	24.5 ± 2.7	19.0 ± 3.7	6.3 ± 1.6
Z+jets	373.8 ± 15.9	160.0 ± 6.8	40.8 ± 1.7	16.5 ± 1.0	3.7 ± 0.1
W+jets	45.0 ± 2.1	14.1 ± 1.3	1.5 ± 0.07	2.8 ± 0.1	0.76 ± 0.04
single top	38.3 ± 3.2	36.8 ± 2.5	7.6 ± 0.7	10.8 ± 0.7	2.3 ± 0.2
diboson	7.5 ± 0.5	7.5 ± 0.6	1.5 ± 0.1	1.4 ± 0.1	0.75 ± 0.05
EWK W/Z	4.6 ± 0.2	5.1 ± 0.3	0.77 ± 0.04	0.85 ± 0.05	0.15 ± 0.01
SM Higgs	0.72 ± 0.04	0.97 ± 0.06	0.46 ± 0.02	0.68 ± 0.04	0.14 ± 0.01
Tot. exp. bkg.	1259 ± 39	732 ± 17	340 ± 12	319 ± 9	32.2 ± 1.9
Expected signal for $\sigma(gg \to S) \times \mathcal{B}(S \to HH \to bb\tau\tau) = 1$ pb					
$m_X = 300$ GeV	59.6	11.5	47.3	10.2	0.6
$m_X = 600$ GeV	38.6	264.2	36.5	237.1	55.4
$m_X = 900$ GeV	23.0	176.3	12.2	127.9	419.6
Observed data	1252	782	363	318	28

Table 6.2 – Observed and expected event yields in different signal regions of the resonant search for the $\tau_\mu\tau_h$ final state. Quoted uncertainties are obtained after fixing the nuisance parameters to their estimation from a maximum likelihood fit under the background-only hypothesis.

6.4.2 Resonant production

Model-independent upper limits at the 95% CL are shown in Figure 6.12 for the combination of all the final states and event categories. Limits under both the spin-0 (resonance S) and spin-2 (resonance G) hypotheses are presented. The numerical values of the limits are similar under the two hypotheses, with a slightly better sensitivity observed for the latter spin hypothesis, because of the larger acceptance to spin-2 signals, as discussed in Section 4.7 of Chapter 4.

$\tau_e\tau_h$ final state					
Process	res. 1b1j		res. 2b0j		boosted
	LM	HM	LM	HM	
$t\bar{t}$	187.5 ± 6.8	227.4 ± 7.3	95.2 ± 4.0	118.7 ± 4.0	8.1 ± 0.4
QCD	62.7 ± 6.9	16.8 ± 3.3	6.8 ± 2.1	-	7.34 ± 2.2
Z+jets	106.7 ± 5.0	59.6 ± 2.2	8.2 ± 0.7	8.3 ± 0.4	0.69 ± 0.03
W+jets	10.4 ± 0.9	10.3 ± 1.1	0.029 ± 0.001	0.099 ± 0.004	0.45 ± 0.02
single top	14.6 ± 1.2	15.9 ± 1.2	2.2 ± 0.2	4.2 ± 0.4	0.68 ± 0.05
diboson	3.7 ± 0.2	3.9 ± 0.4	0.56 ± 0.06	0.61 ± 0.06	0.27 ± 0.02
EWK W/Z	1.2 ± 0.1	0.63 ± 0.02	0.093 ± 0.004	0.43 ± 0.01	0.14 ± 0.01
SM Higgs	0.26 ± 0.01	0.48 ± 0.03	0.14 ± 0.01	0.29 ± 0.02	0.10 ± 0.01
Tot. exp. bkg.	387 ± 11	335 ± 9	113 ± 5	133 ± 4	17.7 ± 2.2
Expected signal for $\sigma(gg \to S) \times \mathcal{B}(S \to HH \to bb\tau\tau) = 1$ pb					
$m_X = 300$ GeV	21.2	6.8	16.2	5.1	0.1
$m_X = 600$ GeV	15.5	127.5	16.1	118.5	28.0
$m_X = 900$ GeV	10.6	100.3	5.1	57.3	213.8
Observed data	388	316	114	123	7

Table 6.3 – Observed and expected event yields in different signal regions of the resonant search for the $\tau_e\tau_h$ final state. Quoted uncertainties are obtained after fixing the nuisance parameters to their estimation from a maximum likelihood fit under the background-only hypothesis.

$\tau_h\tau_h$ final state			
Process	res. 1b1j	res. 2b0j	boosted
$t\bar{t}$	33.6 ± 1.5	16.5 ± 1.1	0.068 ± 0.004
QCD	40.6 ± 7.9	14.5 ± 2.8	0.012 ± 0.012
Z+jets	48.7 ± 6.2	9.1 ± 1.0	2.2 ± 0.1
W+jets	1.11 ± 0.06	-	0.031 ± 0.002
single top	4.2 ± 0.3	0.026 ± 0.002	-
diboson	2.3 ± 0.4	0.57 ± 0.08	0.33 ± 0.03
EWK W/Z	0.78 ± 0.04	-	0.15 ± 0.01
SM Higgs	0.63 ± 0.08	0.38 ± 0.05	0.14 ± 0.01
Tot. exp. bkg.	132 ± 10	41 ± 3	2.9 ± 0.1
Expected signal for $\sigma(gg \to S) \times \mathcal{B}(S \to HH \to bb\tau\tau) = 1$ pb in resonant case			
$m_X = 300$ GeV	20.48	15.03	0.08
$m_X = 600$ GeV	185.27	165.44	40.51
$m_X = 900$ GeV	126.17	105.13	379.10
$k_\lambda = 1$ (SM)	0.24	0.21	0.05
$k_\lambda = 20$	9.20	7.88	0.60
Observed data	140	33	3

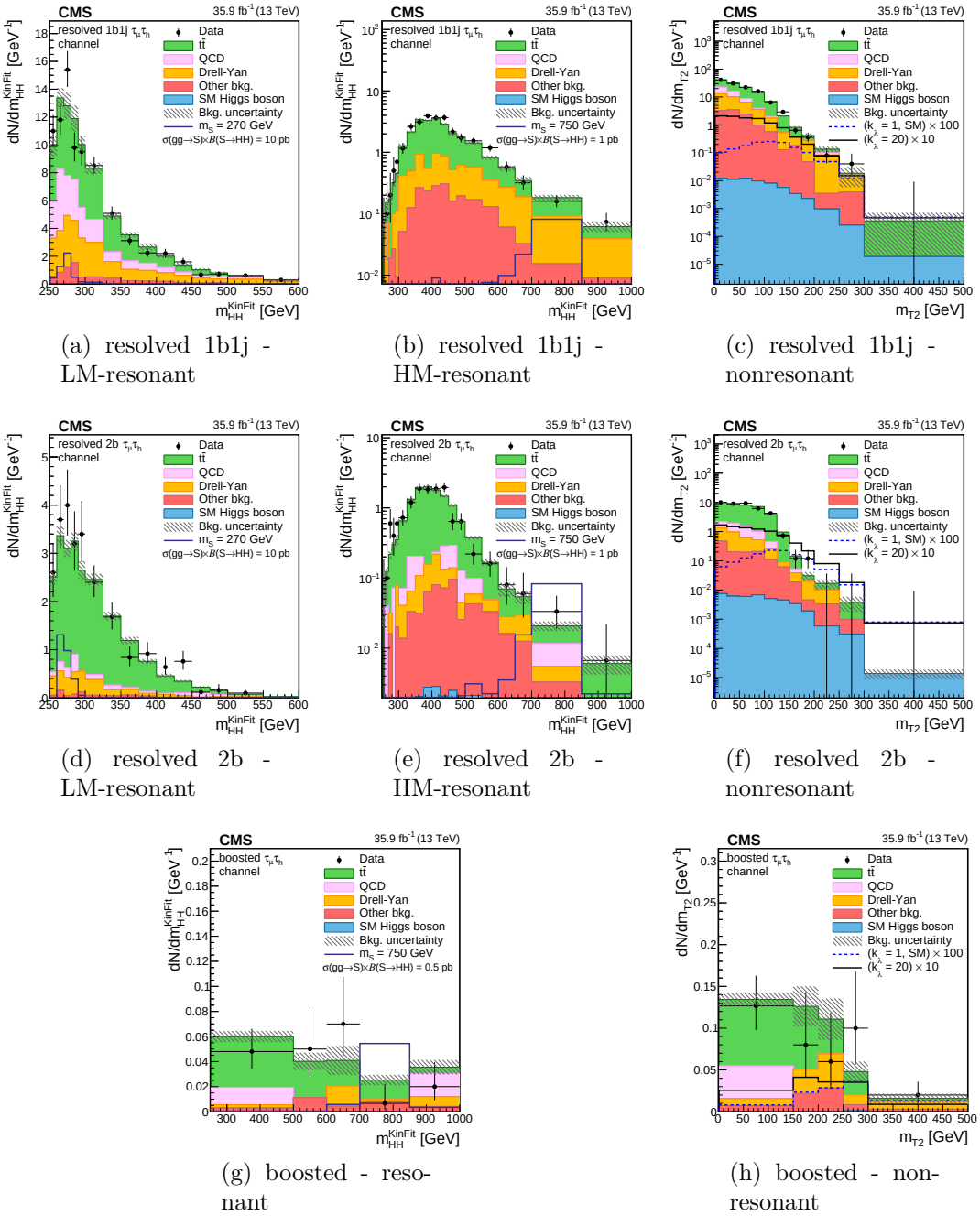
Table 6.4 – Observed and expected event yields in different signal regions of the $\tau_h\tau_h$ final state (the definition is the same for both resonant and nonresonant searches). Quoted uncertainties are obtained after fixing the nuisance parameters to their estimation from a maximum likelihood fit under the background-only hypothesis.

$\tau_\mu\tau_h$ final state			
Process	res. 1b1j	res. 2b0j	boosted
$t\bar{t}$	1617.6 ± 38.7	802.2 ± 22.4	20.0 ± 0.9
QCD	443.9 ± 38.2	80.9 ± 7.0	5.6 ± 1.9
Z+jets	629.6 ± 22.3	64.8 ± 2.9	7.1 ± 0.3
W+jets	124.7 ± 6.7	4.9 ± 0.2	0.95 ± 0.04
single top	121.9 ± 7.8	22.0 ± 1.5	2.5 ± 0.2
diboson	18.3 ± 1.2	2.9 ± 0.3	0.89 ± 0.06
EWK W/Z	9.4 ± 0.5	1.2 ± 0.1	0.15 ± 0.01
SM Higgs	1.7 ± 0.1	1.1 ± 0.1	0.18 ± 0.01
Tot. exp. bkg.	2967 ± 60	980 ± 24	38 ± 2
Expected signal			
$k_\lambda = 1$ (SM)	0.38	0.33	0.08
$k_\lambda = 20$	25.75	20.88	1.12
Observed data	3020	996	35

Table 6.5 – Observed and expected event yields in different signal regions of the nonresonant search for the $\tau_\mu\tau_h$ final state. Quoted uncertainties are obtained after fixing the nuisance parameters to their estimation from a maximum likelihood fit under the background-only hypothesis.

$\tau_e\tau_h$ final state			
Process	res. 1b1j	res. 2b0j	boosted
$t\bar{t}$	631.8 ± 16.3	311.1 ± 9.3	8.9 ± 0.4
QCD	135.9 ± 11.7	6.7 ± 2.1	6.5 ± 2.1
Z+jets	213.3 ± 7.0	20.2 ± 0.8	2.2 ± 0.1
W+jets	70.2 ± 3.2	0.42 ± 0.02	0.47 ± 0.02
single top	48.9 ± 3.2	10.5 ± 0.8	0.82 ± 0.05
diboson	7.9 ± 0.5	1.1 ± 0.1	0.42 ± 0.03
EWK W/Z	3.3 ± 0.1	0.91 ± 0.03	0.33 ± 0.02
SM Higgs	0.69 ± 0.04	0.41 ± 0.03	0.12 ± 0.01
Tot. exp. bkg.	1112 ± 22	351 ± 10	19.7 ± 2.1
Expected signal			
$k_\lambda = 1$ (SM)	0.16	0.14	0.04
$k_\lambda = 20$	10.28	8.26	0.55
Observed data	1057	355	11

Table 6.6 – Observed and expected event yields in different signal regions of the nonresonant search for the $\tau_e\tau_h$ final state. Quoted uncertainties are obtained after fixing the nuisance parameters to their estimation from a maximum likelihood fit under the background-only hypothesis.



(a) resolved 1b1j - LM-resonant

(b) resolved 1b1j - HM-resonant

(c) resolved 1b1j - nonresonant

(d) resolved 2b - LM-resonant

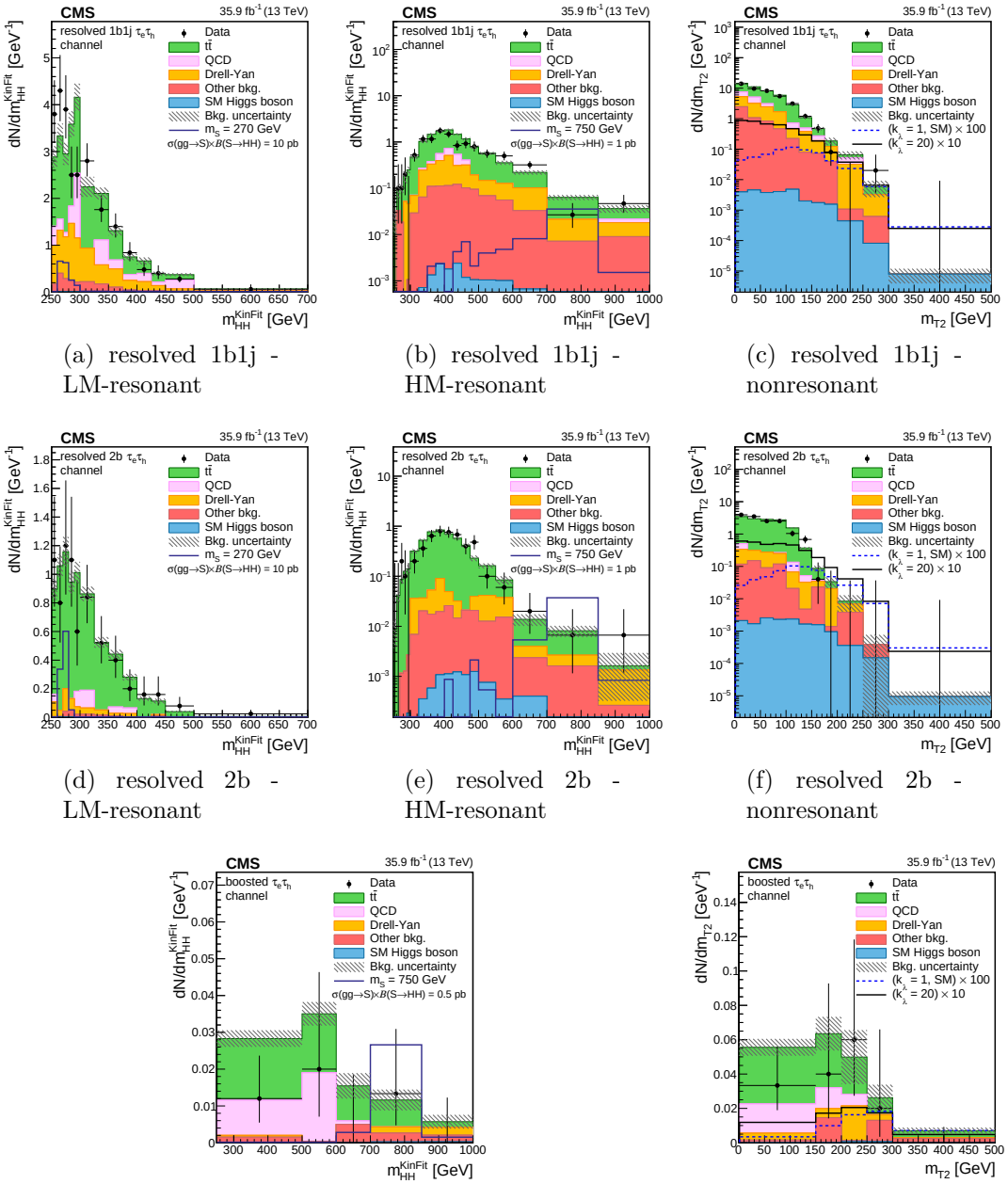
(e) resolved 2b - HM-resonant

(f) resolved 2b - nonresonant

(g) boosted - resonant

(h) boosted - nonresonant

Figure 6.7 – Distributions of the events observed in the signal regions of the $\tau_\mu\tau_h$ final state. The first, second, and third rows show the resolved 1b1j, 2b, and boosted regions, respectively. Panels (a),(b),(d),(e),(g) show the distribution of the m_{HH}^{KinFit} variable in the low-mass (LM) and high-mass (HM) signal regions, and panels (c),(f),(h) show the distribution of the m_{T2} variable. Data are represented by points with error bars and expected signal contributions are represented by the solid (BSM HH signals) and dashed (SM nonresonant HH signal) lines. Expected background contributions (shaded histograms) and associated systematic uncertainties (dashed areas) are shown as obtained after the maximum likelihood fit to the data under the background-only hypothesis. The background histograms are stacked while the signal histograms are not stacked.



(a) resolved 1b1j - LM-resonant

(b) resolved 1b1j - HM-resonant

(c) resolved 1b1j - nonresonant

(d) resolved 2b - LM-resonant

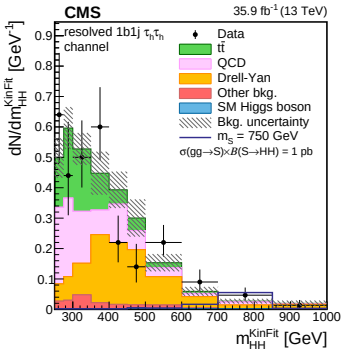
(e) resolved 2b - HM-resonant

(f) resolved 2b - nonresonant

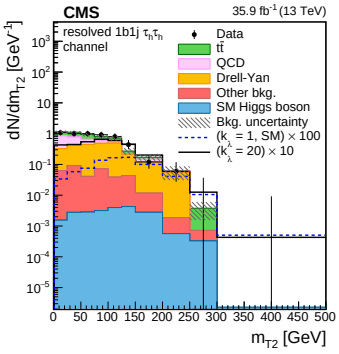
(g) boosted - resonant

(h) boosted - nonresonant

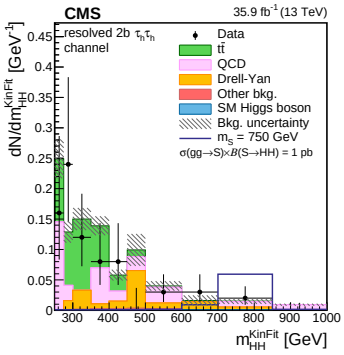
Figure 6.8 – Distributions of the events observed in the signal regions of the $\tau_e\tau_h$ final state. The first, second, and third rows show the resolved 1b1j, 2b, and boosted regions, respectively. Panels (a),(b),(d),(e),(g) show the distribution of the m_{HH}^{KinFit} variable in the low-mass (LM) and high-mass (HM) signal regions, and panels (c),(f),(h) show the distribution of the m_{T2} variable. Data are represented by points with error bars and expected signal contributions are represented by the solid (BSM HH signals) and dashed (SM nonresonant HH signal) lines. Expected background contributions (shaded histograms) and associated systematic uncertainties (dashed areas) are shown as obtained after the maximum likelihood fit to the data under the background-only hypothesis. The background histograms are stacked while the signal histograms are not stacked.



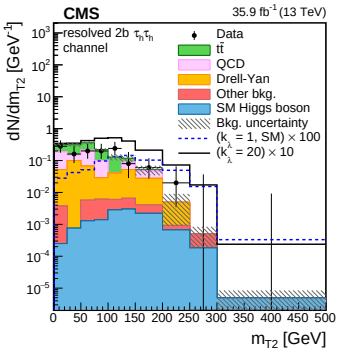
(a) resolved 1b1j - resonant



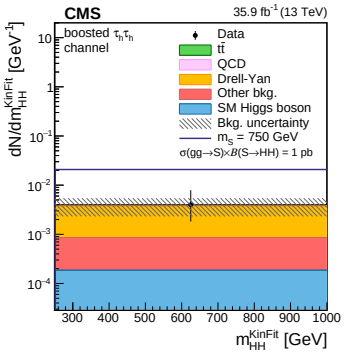
(b) resolved 1b1j - nonresonant



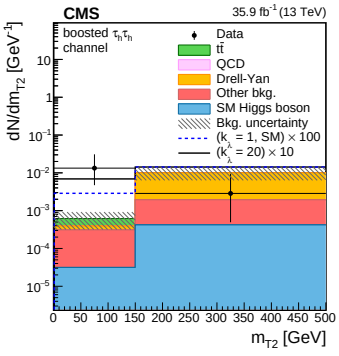
(c) resolved 2b - resonant



(d) resolved 2b - nonresonant

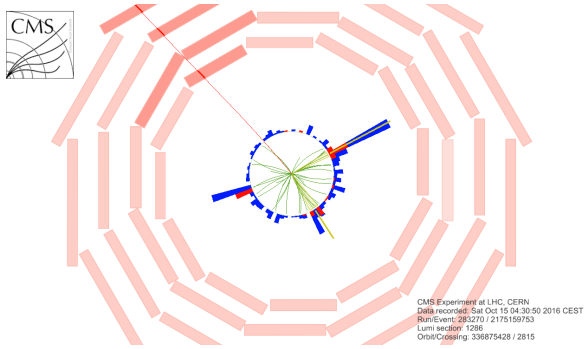


(e) boosted - resonant

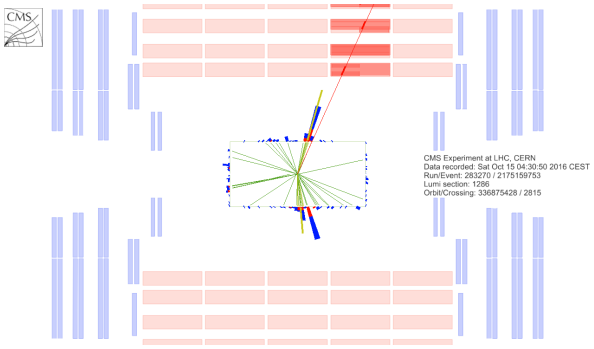


(f) boosted - nonresonant

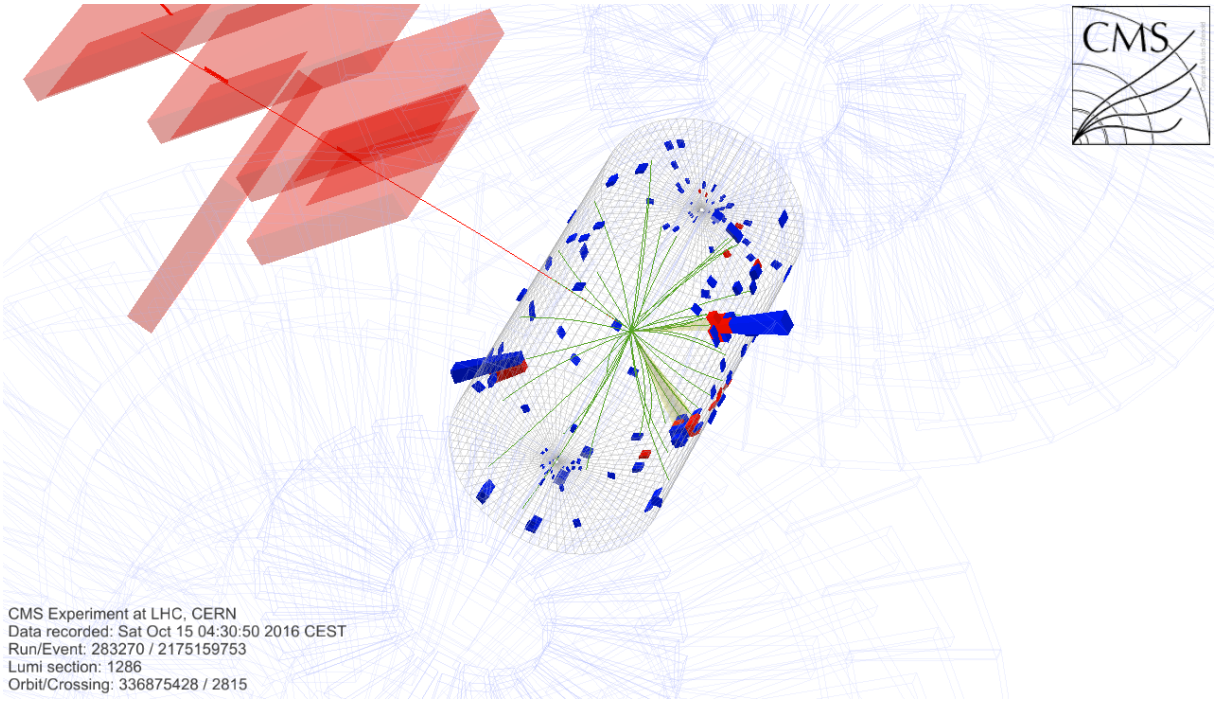
Figure 6.9 – Distributions of the events observed in the signal regions of the $\tau_h\tau_h$ final state. The first, second, and third rows show the resolved 1b1j, 2b, and boosted regions, respectively. Panels (a),(c),(e) show the distribution of the m_{HH}^{KinFit} variable and panels (b),(d),(f) show the distribution of the m_{T2} variable. Data are represented by points with error bars and expected signal contributions are represented by the solid (BSM HH signals) and dashed (SM nonresonant HH signal) lines. Expected background contributions (shaded histograms) and associated systematic uncertainties (dashed areas) are shown as obtained after the maximum likelihood fit to the data under the background-only hypothesis. The background histograms are stacked while the signal histograms are not stacked.



(a) Transverse view

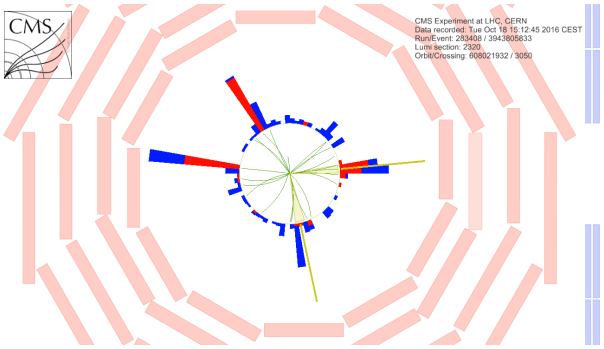


(b) Lateral view

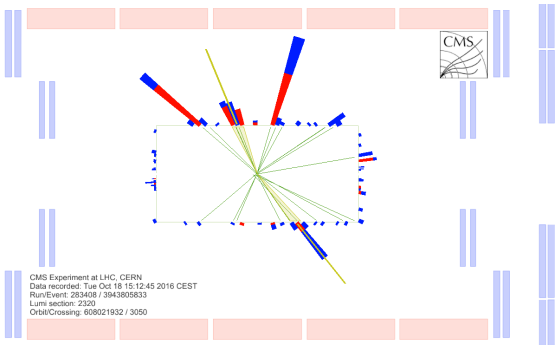


(c) Tridimensional view

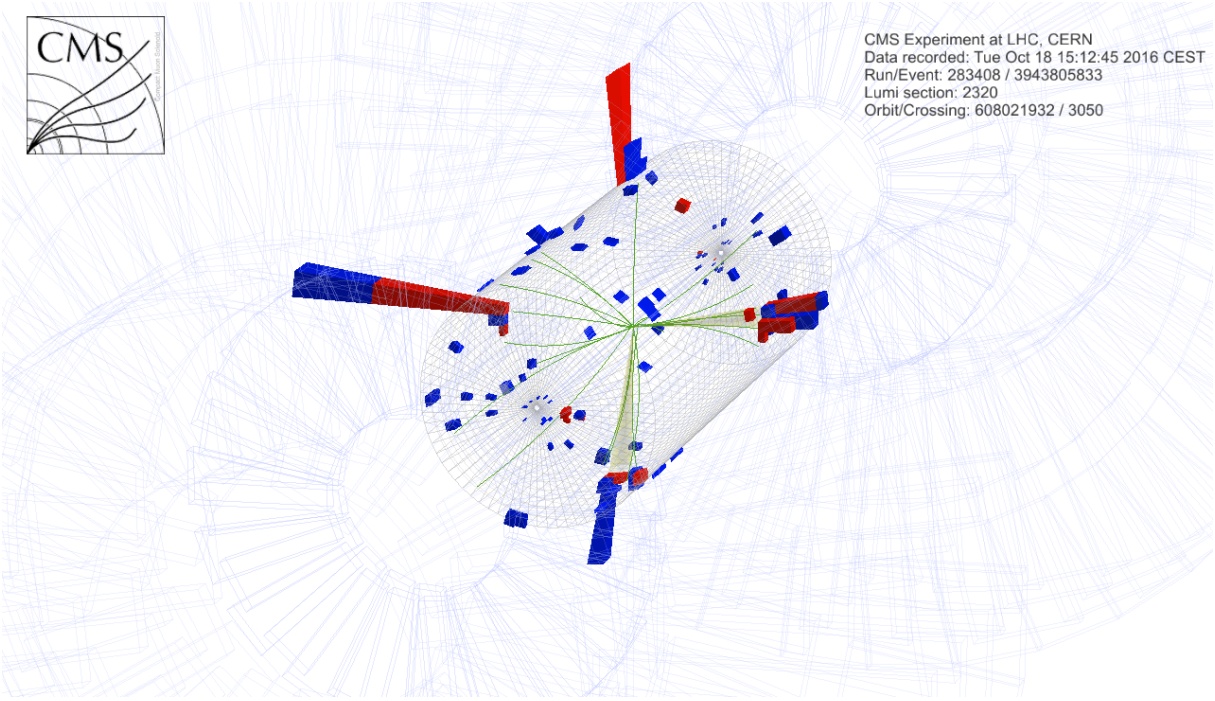
Figure 6.10 – Display of one event selected in the $\tau_\mu \tau_h$ final state, 2b event category. The event is selected in the signal region of the nonresonant search.



(a) Transverse view

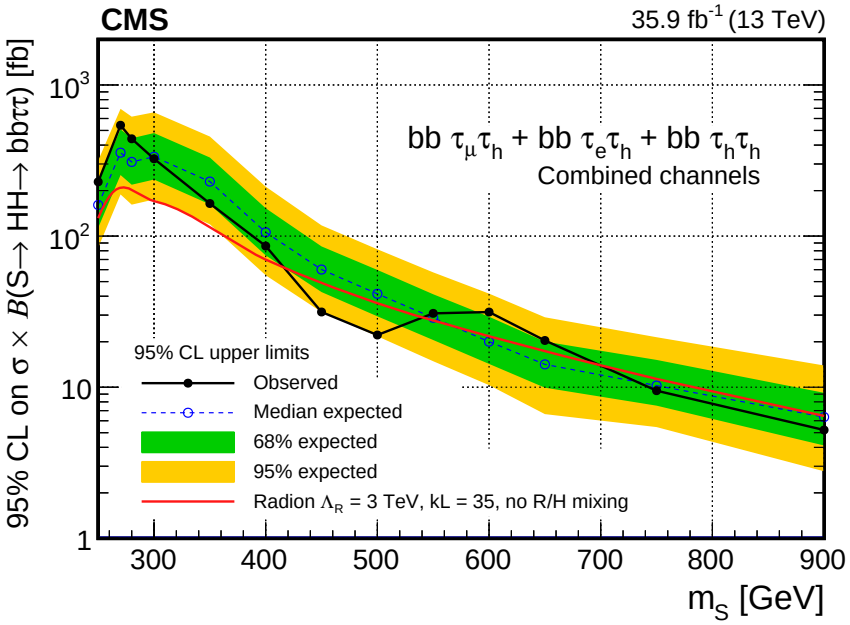


(b) Lateral view

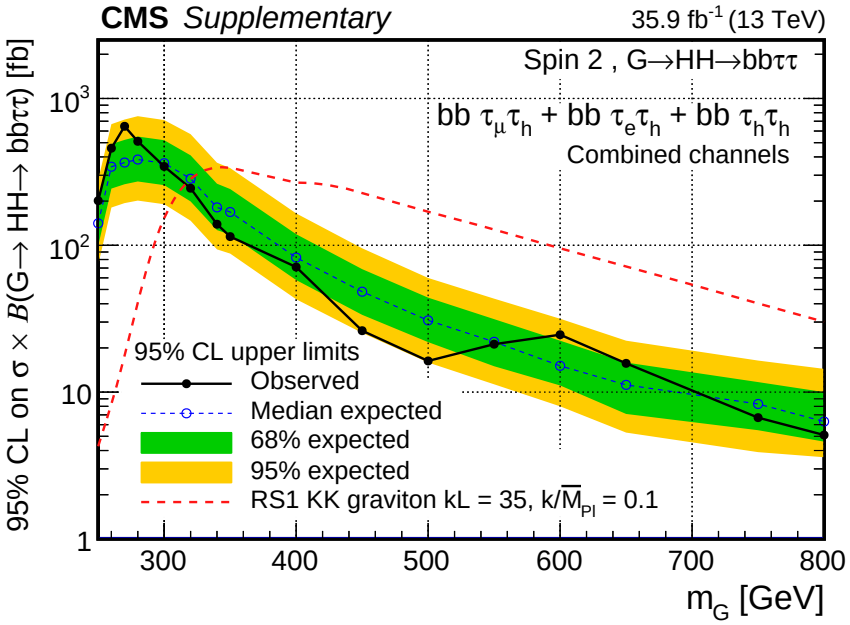


(c) Tridimensional view

Figure 6.11 – Display of one event selected in the $\tau_h\tau_h$ final state, 2b event category.



(a) Spin-0

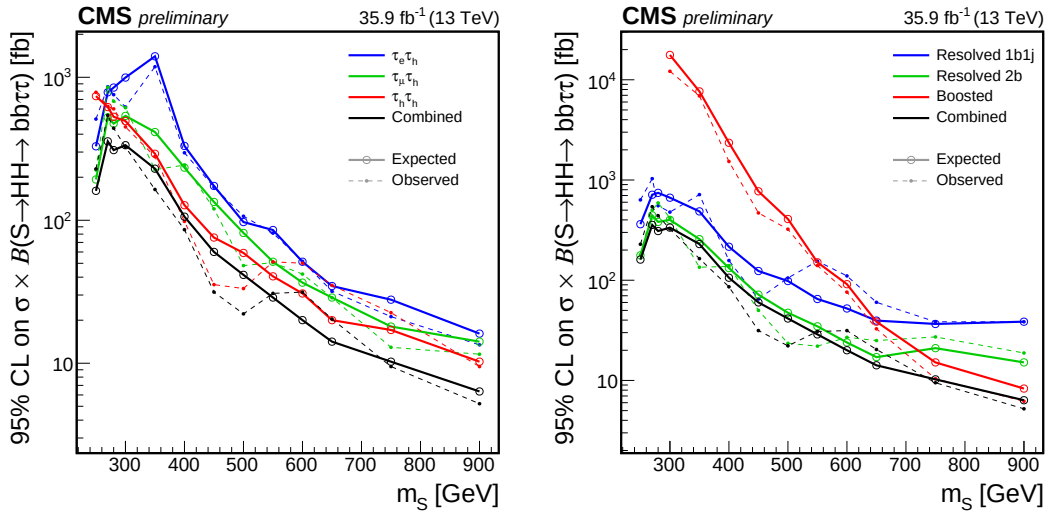


(b) Spin-2

Figure 6.12 – 95% CL upper limits on $\sigma(gg \to X) \times \mathcal{B}(X \to bb\tau\tau)$ for a spin-0 (a) and spin-2 (b) resonance. The green and yellow bands indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The red curves indicate the theoretical prediction for the production of a radion (spin-0) or of a graviton (spin-2) decaying to a HH pair.

The values of the cross section times branching fraction excluded range from about 500 to 5 fb depending on the value of m_X. Better exclusion is achieved at high m_X results thanks to both the larger signal acceptance and the reduced background contamination. The reduction of the sensitivity around $m_X = 280\,\text{GeV}$ is a consequence of the shape of the $t\bar{t}$ background, that peaks about this value, as it can be observed in Figures 6.7, 6.8, and 6.9.

The separate contributions to the sensitivity from the different final states and event categories is illustrated in Figure 6.13 in the case of a spin-0 resonance. Similar conclusions are valid in the case of spin-2 resonances. As expected, the $\tau_h\tau_h$ final state has the largest sensitivity over most of the m_X range studied thanks to the best signal purity with respect to the other final states. There is however a sizeable contribution from the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states, that benefit from lower p_T thresholds and thus dominate the combined sensitivity for $m_X \lesssim 300\,\text{GeV}$. For resonances of mass $m_X \lesssim 650\,\text{GeV}$ the best sensitivity is achieved in the resolved 2b category thanks to the larger signal purity with respect to the resolved 1b1j. For higher resonance masses, the contribution of the boosted category becomes dominant. The migration of events from the resolved 2b to the boosted category results in the reduction of the sensitivity of the former, but improves the overall sensitivity of the search.



(a) Sensitivity per final state

(b) Sensitivity per category

Figure 6.13 – Comparison of the 95% CL upper limits separately computed for the three final states (a) and the three event categories (b).

These model-independent results are reinterpreted in the context of the two MSSM scenarios already described in Section 1.3.1 of Chapter 1. These models, denoted as the hMSSM and low-tan β-high scenarios, assume that the observed Higgs boson of mass 125 GeV is the MSSM CP-even lighter scalar, conventionally denoted as h. The spin-0 resonance S represents then the CP-even heavier scalar, denoted as H in the context of the model (and not to be confused here with the SM Higgs). The standard MSSM notation is adopted for the following paragraph to describe the interpretation of the resonant search in these two scenarios. The symbols used are:

- h for the CP-even lighter scalar, usually assumed to be the observed Higgs boson of mass $m_h = 125\,\text{GeV}$

- H for the CP-even heavier scalar. In the model independent search, H is represented by the spin-0 state S
- A for the CP-odd scalar
- $\tan\beta$ for the ratio of the vacuum expectation values of the two Higgs doublets of the model

The predictions of the hMSSM and low-$\tan\beta$-high scenarios are tested against the observed results in the $\mathrm{bb}\tau\tau$ final state; the theoretical uncertainties on the cross section and branching fractions of the models are introduced as additional nuisance parameters and vary as a function of m_A and $\tan\beta$. The exclusion set on these parameters is shown in Figure 6.14 for the two models: a region in the parameter space corresponding to values of m_A between 230 and 360 GeV and $\tan\beta \lesssim 2$ is excluded at 95% CL. The reduced sensitivity of the model-independent sensitivity around $\mathrm{m_X} = 280\,\mathrm{GeV}$ produces two distinct regions in this interpretation, that is particularly evident in the expected excluded contour of the hMSSM model.

6.4.3 Nonresonant production

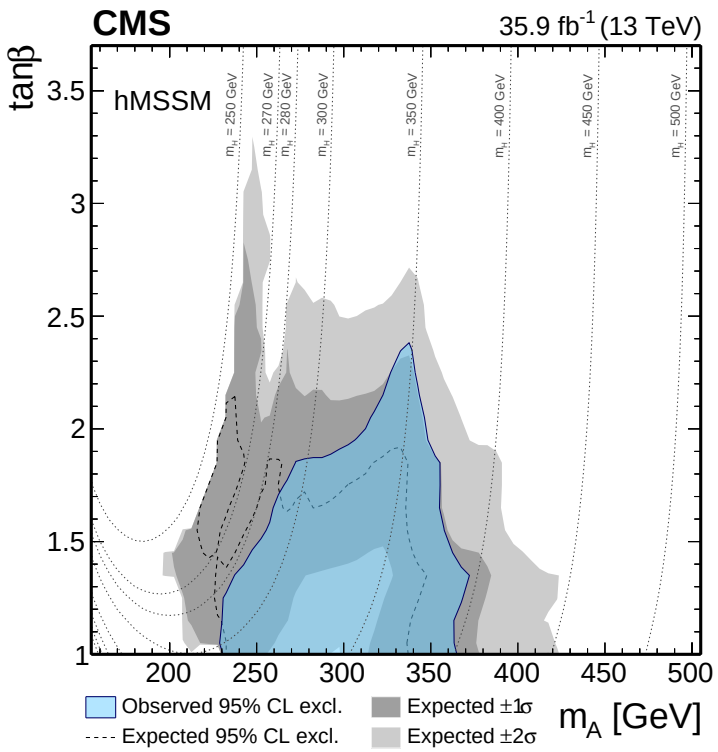
Model-independent limits are derived on the nonresonant HH production cross section times branching fraction. As discussed in Section 1.3.2 of Chapter 1, signals are parametrized with five Higgs boson couplings that are derived with an effective Lagrangian approach. The choice of the coupling combination largely affects the signal kinematics and, consequently, the distribution of the $\mathrm{m_{T2}}$ variable. Two sets of results are derived: the first one is a test of the k_λ and k_t couplings assuming $\mathrm{c_2} = \mathrm{c_g} = \mathrm{c_{2g}} = 0$, while the second one covers additional BSM topologies using the twelve shape benchmarks.

In the first case, the kinematics of the signal produced via gluon fusion is totally determined by the two diagrams discussed in Section 1.2 of Chapter 1. Because the amplitudes of the triangle ($A_\triangleright$) and square ($A_\square$) diagrams depend on the k_λ and k_t coupling modifiers as $A_\triangleright \sim k_\lambda k_\mathrm{t} a_\triangleright$ and $A_\square \sim k_\mathrm{t}^2 a_\square$ respectively, one parameter can be factored out in the differential cross section expression and the signal event distribution only depends on the ratio of the two couplings:

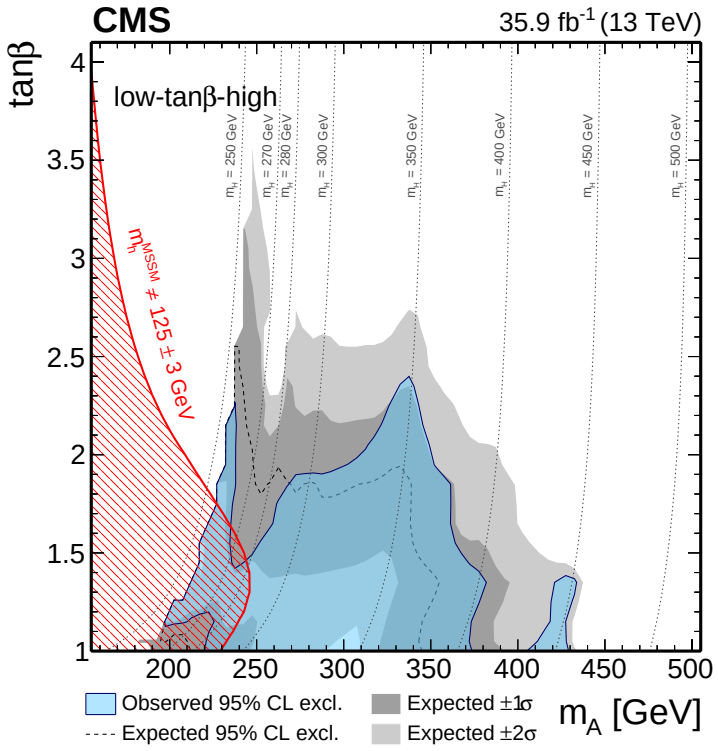
$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}x} &\sim \frac{\mathrm{d}}{\mathrm{d}x}\left(|A_\triangleright(x) + A_\square(x)|^2\right) \\ &\sim \frac{\mathrm{d}}{\mathrm{d}x}\left(k_\lambda^2 k_\mathrm{t}^2 |a_\triangleright(x)|^2 + k_\mathrm{t}^4 |a_\square(x)|^2 + 2k_\lambda k_\mathrm{t}^3 a_\triangleright(x)\cdot a_\square(x)\right) \\ &\sim \left(\frac{k_\lambda}{k_\mathrm{t}}\right)^2 \frac{\mathrm{d}\left(|a_\triangleright(x)|^2\right)}{\mathrm{d}x} + \frac{\mathrm{d}\left(|a_\square(x)|^2\right)}{\mathrm{d}x} + 2\frac{k_\lambda}{k_\mathrm{t}}\frac{\mathrm{d}\left(a_\triangleright(x)\cdot a_\square(x)\right)}{\mathrm{d}x}\end{aligned} \tag{6.22}$$

where x is a generic observable.

Limits on the HH production cross section times $\mathrm{HH} \to \mathrm{bb}\tau\tau$ branching fraction as a function of the k_λ/k_t ratio are shown in Figure 6.15 and compared to the theoretical predictions for $k_\mathrm{t} = 1$. The observed constraints on k_λ, assuming the other couplings to be equal to their SM prediction ($k_\mathrm{t} = 1$, $\mathrm{c_2} = \mathrm{c_g} = \mathrm{c_{2g}} = 0$) are $-18 < k_\lambda < 26$, with the expected ones being $-14 < k_\lambda < 22$. We can observe in the exclusion limit the typical pattern already encountered several times before in this thesis that



(a) hMSSM



(b) low-tan β-high

Figure 6.14 – Interpretation of the exclusion limit in the context of the hMSSM (a) and low-tan β-high (b) models. The blue transparent curve denotes the regions of the tan β and m_A parameters excluded by the observation, while the dashed line and the grey bands denote the expected exclusion and its associated 68 and 95% exclusion intervals. The dotted lines indicate trajectories in the plane corresponding to equal values of the mass of the CP-even heavier scalar of the model m_H. In the low-tan β-high interpretation, the red area represent points of the parameter space that are not compatible with the 125 GeV Higgs boson mass within 3 GeV.

results from the interference between the two diagrams contributing to the process. In the vicinity of the maximal interference point $k_\lambda/k_t = 2.46$ small modifications of the ratio of the two couplings produce large effects on the signal kinematics and consequently on the sensitivity. For $|k_\lambda/k_t| \gtrsim 10$ the triangle diagram becomes dominant and the values of the excluded cross section for positive and negative k_λ/k_t become asymptotically equal. Including the theoretical uncertainties, the observed 95% CL upper limit on $\sigma_{HH}^{SM} \times \mathcal{B}(HH \to bb\tau\tau)$ (SM production with $k_\lambda = k_t = 1$) corresponds to 75.4 fb while the expected 95% CL upper limit amounts to 61.0 fb. These values correspond to about 30 and 25 times the SM prediction, respectively. The sensitivity in the $bb\tau\tau$ channel alone exceeds the combined Run I exclusion of CMS searches, that set an observed 95% CL upper limit of 43 times the SM cross section for an expected exclusion of 47 times σ_{HH}^{SM} [74].

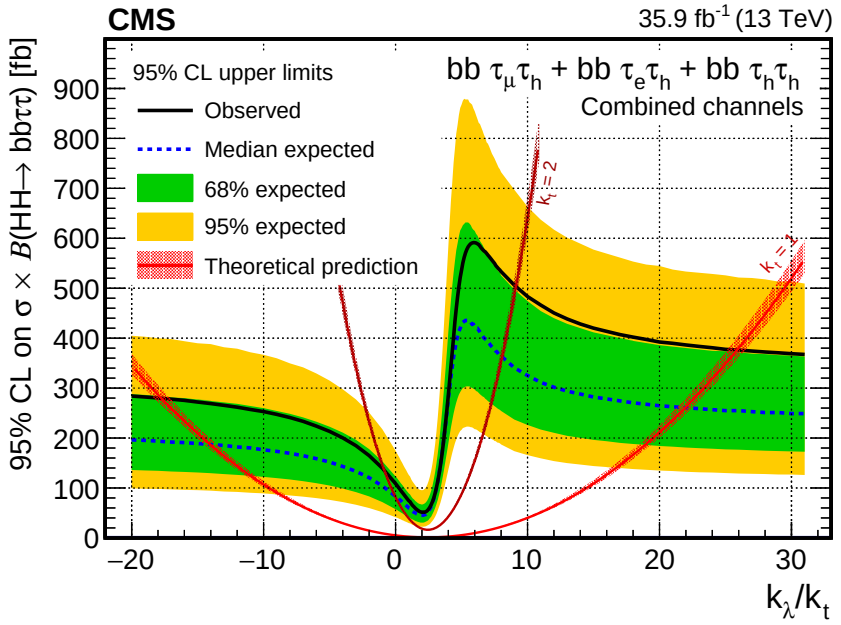
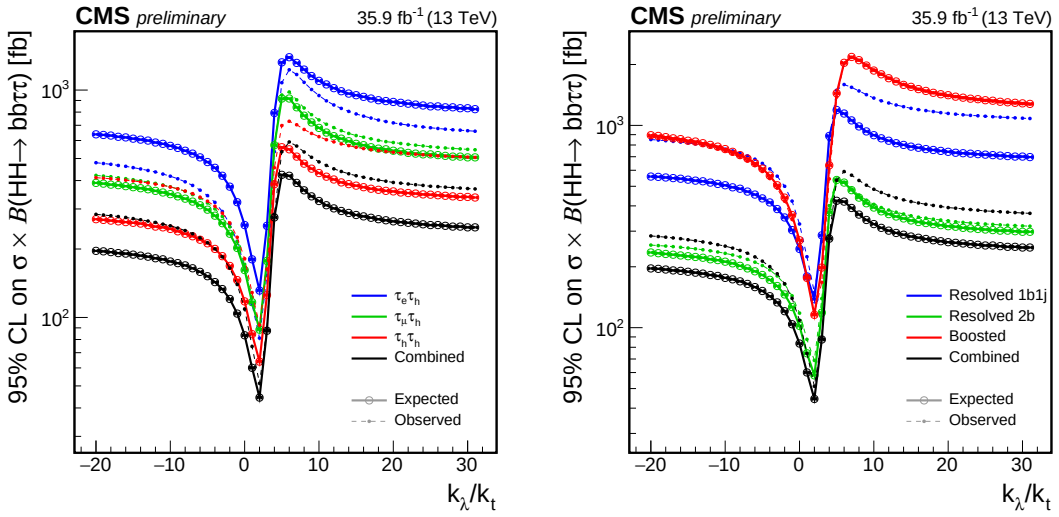


Figure 6.15 – 95% CL confidence level upper limits on $\sigma(gg \to HH) \times \mathcal{B}(HH \to bb\tau\tau)$ for the combination of the three final states and three event categories. The two red lines denote the theoretical cross section times the branching fraction for a value of k_t of 1 (SM prediction) and 2.

The separate contributions of the three final states and of the three categories are detailed in Figure 6.16. Differently from the resonant case, here the 2b resolved category of the $\tau_h\tau_h$ final state has the best sensitivity for all the values of k_λ/k_t studied. The contribution from the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states is nevertheless relevant and their inclusion improves the sensitivity by about 30%. The contribution of the boosted category is important in the interference region because of the increase of high m_{HH} events that populate this category.

These results are used to set constraints on anomalous Higgs boson couplings. The exclusion achieved is illustrated in Figure 6.17 as a function of k_λ and k_t. The SM branching fractions of the Higgs boson to a bb and a $\tau\tau$ pair are assumed for all the the couplings tested, and the other couplings are assumed to vanish ($c_2 = c_g = c_{2g} = 0$). It should be noted that the assumption on the branching fractions might not be true for



(a) Sensitivity per final state (b) Sensitivity per category

Figure 6.16 – Comparison of the 95% CL expected upper limits as a function of the k_λ/k_t, separately shown for the three final states (a) and the three event categories (b).

specific BSM models, and is chosen here to illustrate the simultaneous exclusion on k_λ and k_t. Exclusions for other model-specific branching fractions can be performed from the model-independent results of Figure 6.15. Because the HH cross section depends on the k_λ and k_t parameters through terms proportional to $k_\lambda^2 k_t^2$, k_t^4, and $k_\lambda k_t^3$, the exclusion is symmetric under a $(k_\lambda,\ k_t) \leftrightarrow (-k_\lambda, -k_t)$ transformation. As a consequence, the search is sensitive on the relative sign of the two couplings: if the SM λ_{HHH} coupling is assumed ($k_\lambda = 1$), the intervals $k_t < -2$ and $k_t > 2.2$ are excluded by the observation. Sensitivity to the k_t sign is typically achieved in searches for Higgs boson production in association with a single top quark, as seen for example in the search performed in the multilepton final state [172] where the k_t coupling is constrained to be $-1.25 < k_t < 1.6$. These results in the HH final state show that the sign of the k_t coupling can be probed, although with less sensitivity, in a completely independent final state and could in future provide important verifications in case deviations from the SM prediction on the sign of the k_t coupling are observed.

A second set of results is presented as the upper limit on the cross section for the signal shape benchmarks, specific points of the five-dimensional effective Lagrangian parametrization of BSM physics that represent characteristic kinematic properties of the HH system. The 95% CL upper limits on the cross section of these shape benchmarks is illustrated in Figure 6.18, where they are compared to the upper limit on the SM signal and $k_\lambda = 0$ signals. The variations in the numerical values of the excluded cross sections directly depend on the m_{HH} distribution of the shape benchmarks. Better sensitivity is observed for benchmarks with an enhanced cross section at high m_{HH} values.

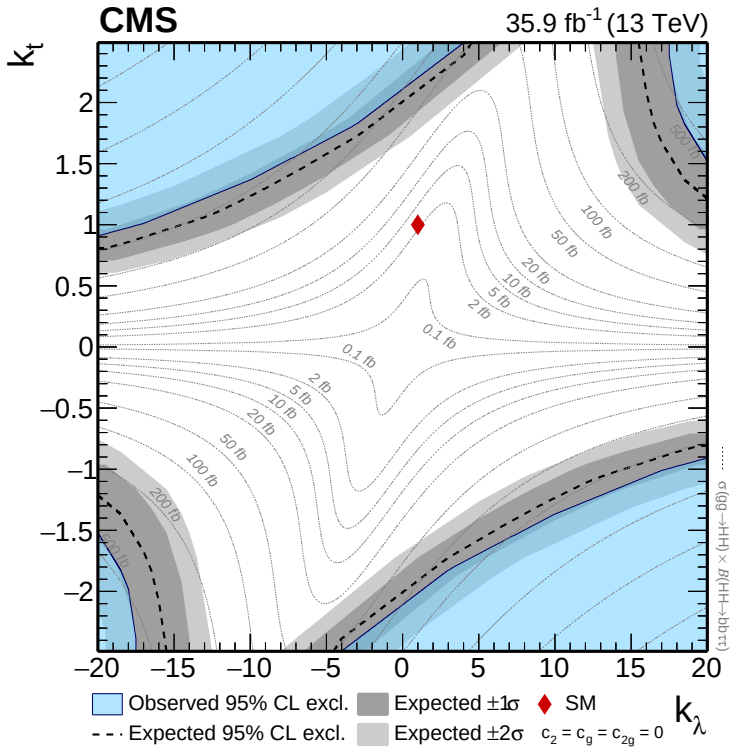


Figure 6.17 – Test of k_λ and k_t anomalous couplings. The blue region denotes the parameters excluded at 95% CL with the observed data, while the dashed black line and the grey regions denote the expected exclusions and the 1σ and 2σ bands. The dotted lines indicate trajectories in the plane with equal values of cross section times branching fraction that are displayed in the associated labels. The SM couplings, corresponding to $k_\lambda = k_\mathrm{t} = 1$, are indicated by the diamond-shaped marker. Theory predictions, expected and observed limits are symmetric through a $(k_\lambda, k_\mathrm{t}) \leftrightarrow (-k_\lambda, -k_\mathrm{t})$ transformation.

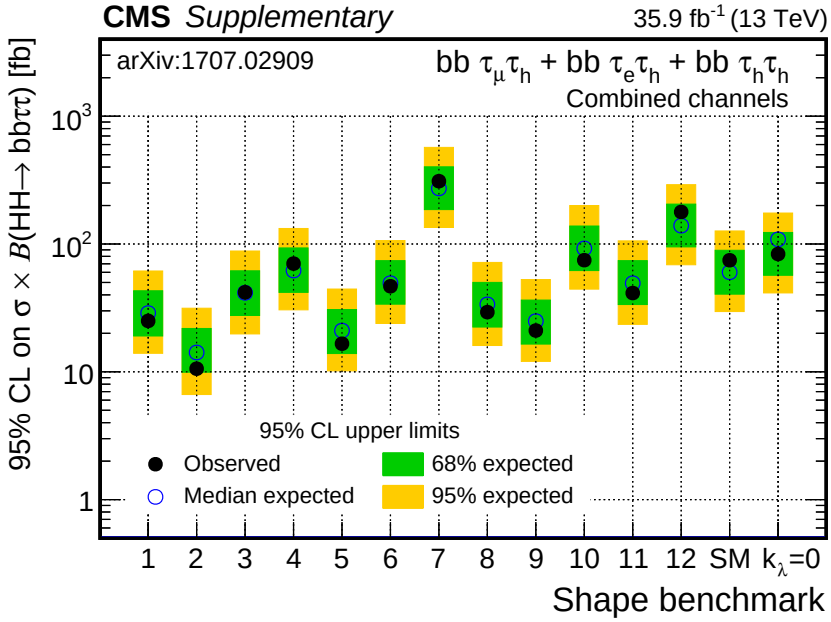


Figure 6.18 – 95% CL upper limits on $\sigma(\mathrm{gg} \to \mathrm{HH}) \times \mathcal{B}(\mathrm{HH} \to \mathrm{bb}\tau\tau)$ for the twelve shape benchmarks, the SM and the $k_\lambda = 0$ signals. Numerical differences in the upper exclusion limit are directly related to the changes in the signal kinematics from BSM Higgs boson couplings.

6.5 Comparison of the results

In this section, the results described above, obtained by analysing an integrated luminosity of 35.9 fb^{-1}, are briefly compared to previous ones obtained on smaller datasets to illustrate the most important improvements introduced in the analysis strategy and their impact on the sensitivity. The results are subsequently put in the broader context of HH searches performed at the LHC Run II, where they are compared to searches performed by the ATLAS and CMS Collaborations in other final states.

6.5.1 Earlier Run II bb$\tau\tau$ results

The analysis strategy progressively evolved with the integrated luminosity (L) to exploit the larger size of the dataset. The techniques developed also had to face different pileup and instantaneous luminosity conditions. Three sets of $\text{HH} \to \text{bb}\tau\tau$ results were derived in Run II.

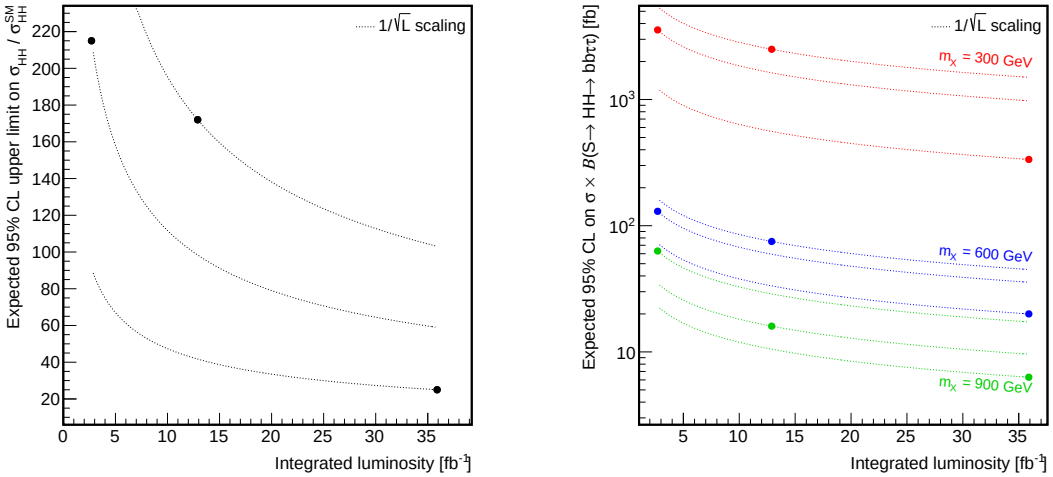
- The "Moriond 2016" [122, 123] search, based on $L = 2.7\ \text{fb}^{-1}$. Tau leptons are identified using the charged hadron p_{T} sum described in Section 4.3.3 of Chapter 4, that was observed to perform well under the relatively low pileup conditions experienced in 2015. No event categorization is used in both the resonant and nonresonant searches, and signal regions are defined by the presence of two b-tagged jets using the medium and loose working points, respectively. The visible four body mass of the selected candidates, m_{HH}, is used to search for a nonresonant signal contribution.

- The "ICHEP 2016" [126, 127] search made use of an integrated luminosity $L = 12.9\ \text{fb}^{-1}$. The bb categorization is introduced in the resonant to enhance the sensitivity at high mass, allowing to fully profit of the larger dataset and of the increase in the centre-of-mass energy with respect to Run I. The LHC peak instantaneous luminosity rose to the design value of $1.0 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ with up to 35 pileup interactions per event. This unexpected performance, the evolving nature of the dataset (collisions were ongoing when the analysis methods were developed), and dynamic inefficiencies observed in the strip tracker, related to the high detector occupancy, limited our possibility to fully optimized the methods. In particular, τ_{h} and b misidentification rates were observed to be larger than expected, resulting in a sizeable contamination from QCD multijet events.

- The "Moriond 2017" search [133], based on an integrated luminosity $L = 35.9\ \text{fb}^{-1}$, has been discussed in detail in this thesis. The bb categorization is applied to the nonresonant search and multivariate methods extended to the resonant search. Invariant mass and isolation selections are optimized, and the m_{T2} variable is introduced in the nonresonant search to enhance its sensitivity.

More details on the Moriond 2016 and ICHEP 2016 searches are given in the Appendix A. The comparison of the sensitivity for the SM nonresonant signal, and for resonant HH production with $\text{m}_{\text{X}} = 300, 600$, and 900 GeV, is shown in Figure 6.19. The evolution of the upper limits is compared with a scaling inversely proportional to the square root of the integrated luminosity $1/\sqrt{L}$, that is typically observed when the sensitivity is limited by the statistical uncertainties, as it is the case of this search. The general

trend observed is an underperformance of the ICHEP 2016 search, which arises from the large contamination from multijet background, in turn caused by a worse performance of the τ isolation and b tagging criteria than expected. The effect is particularly relevant for the low mass resonant and SM nonresonant signals. No impact on the sensitivity is observed for high mass signals because of the limited multijet contamination in the high mass tails of the m_{HH}^{KinFit} variable, as well as for the introduction of the boosted event category. These effects were addressed with the optimizations performed for the Moriond 2017 analysis, were the changes introduced in the analysis strategy resulted in improvements larger than those expected solely from the increase of the integrated luminosity.



(a) SM HH (b) Resonant HH

Figure 6.19 – Comparison of the expected 95% CL upper limits achieved with the analysis of 2.7, 12.9, and 35.9 fb^{-1} (full circles). The dotted lines indicate the scaling of the upper limit as the square root of the integrated luminosity ratio.

6.5.2 LHC searches at $\sqrt{s} = 13\,\text{TeV}$

Searches for resonant and nonresonant HH production using data collected at $\sqrt{s} = 13\,\text{TeV}$ have been performed by both the ATLAS and CMS Collaborations in several final states. Higgs boson pair production is a lively subject of research and some of these searches are still ongoing or being updated using the total amount of data recorded by the two experiments in 2016. The bbbb channel was explored by both the ATLAS [173] and CMS [174, 175, 176] collaborations, using different bb reconstruction techniques to improve the sensitivity to a large range of m_X values. In particular, the high branching fraction of the bbbb decay channel, combined with the usage of boosted event topologies, makes possible to explore values of m_X up to 3 TeV. Only the CMS collaboration investigated HH production in the bb$\tau\tau$ and bbVV $\to$ bb$\ell\nu_\ell\ell\nu_\ell$ ($\ell = e, \mu$) [177] decay channels, while only the ATLAS collaboration probed $WW^*\gamma\gamma \to qq'\ell\nu_\ell$ ($\ell = e, \mu$) decays [178]. Finally, both collaborations performed a search for HH production in the bb$\gamma\gamma$ final state [73, 179]. No evidence for HH production has been found in any of these final states. It is interesting to compare the sensitivity achieved by these searches and to highlight the role of the bb$\tau\tau$ final state in this context.

Nonresonant production

A summary of the 95% CL upper limits on the SM HH cross section is shown in Table 6.7. Although a different integrated luminosity has been analysed in the various channels, this comparison shows that the bb$\tau\tau$ decay channel is one of the most sensitive to SM HH production together with the bb$\gamma\gamma$ and bbbb ones. This indicates the necessity to explore several final states to enhance the sensitivity to this rare process. With results established in the bb$\tau\tau$, bb$\gamma\gamma$, and bbVV decay channels, and upcoming results in the bbbb one, the CMS collaboration is now working on their combination. A naïve sum in quadrature of the results gives a sensitivity that is about 12 times the SM prediction. Assuming that similar sensitivities are achieved in searches performed by the ATLAS Collaboration, the combination of the results of two experiments opens very interesting possibilities of probing SM HH production at the LHC. Prospects for measurements of Higgs boson pair production are further discussed in Chapter 7.

		bbbb	bbVV	bb$\tau\tau$	bb$\gamma\gamma$	WW$^*\gamma\gamma$
ATLAS	Int. luminosity [fb^{-1}]	13.3	-	-	3.2	13.3
	Obs. (exp.) lim. on $\sigma_{\mathrm{HH}}^{\mathrm{SM}}$	29 (38)	-	-	117 (161)	747 (386)
	Anomalous couplings	-	-	-	-	-
	Reference	[173]	-	-	[73]	[178]
CMS	Int. luminosity [fb^{-1}]	2.3	35.9	35.9	35.9	-
	Obs. (exp.) lim. on $\sigma_{\mathrm{HH}}^{\mathrm{SM}}$	342 (308)	79 (89)	31 (25)	19 (16)	-
	Anomalous couplings	-	✓	✓	✓	-
	Reference	[176]	[177]	[133]	[179]	-

Table 6.7 – Summary of the sensitivities to nonresonant SM HH production in the searches performed by the ATLAS and CMS collaborations using the data collected at $\sqrt{s} = 13\,\mathrm{TeV}$. The integrated luminosity analysed and the observed and expected 95% CL upper limits on SM HH production with respect to the SM HH cross section $\sigma_{\mathrm{HH}}^{\mathrm{SM}}$ are shown for five final states. The table also indicates whether anomalous Higgs boson couplings were explored, and the reference to the published results. The CMS results in the bb$\tau\tau$ final state are to those discussed in this thesis.

In the context of anomalous Higgs boson couplings, a comparison of the upper limits set on the shape benchmarks signals is illustrated in Figure 6.20 for the bb$\tau\tau$ and bb$\gamma\gamma$ final states, the only ones at present where these signals have been probed. We observe that the bb$\tau\tau$ final state achieves the best sensitivity in several BSM hypotheses with a significant enhancement of the HH cross section at high m_{HH}: it is for example the case of the shape benchmarks number 2, 5, and 9. Inversely, the bb$\gamma\gamma$ final state is more sensitive to cases where softer m_{HH} spectra are expected, such as the shape benchmark number 7, the SM case or the one where all the couplings but the y_{t} one vanish. This differences in the sensitivity arise from the interplay between the background contamination and the signal branching fraction, as well as from the different techniques used for signal extraction. In the bb$\tau\tau$ search, the usage of $\mathrm{m_{T2}}$ (highly correlated with m_{HH}) makes it very sensitive to changes in the kinematic distributions induced by anomalous couplings. Inversely, the bb$\gamma\gamma$ search looks for a signal contribution in the bidimensional event distribution of the bb and $\gamma\gamma$ invariant masses, thus being sensitive to the signal kinematics only from

changes in the signal acceptance. This comparison highlights the role of the $bb\tau\tau$ decay channel in the exploration of BSM physics via HH production, and shows the necessity to probe several final states to fully cover all the possible BSM signal topologies.

Resonant production

The complementarity of the HH final states currently explored at $\sqrt{s} = 13\,\text{TeV}$ is also illustrated in the case of resonant HH production in Figure 6.21, where the 95% CL upper limits from ATLAS and CMS Run II searches are compared. We again observe that the three most sensitive final states are $bb\gamma\gamma$, $bb\tau\tau$, and bbbb, that give an important contribution to the sensitivity in the low, intermediate, and high m_X regimes, respectively. This is expected from the different branching fraction and background contamination in these three final states: the $bb\gamma\gamma$ decay channel profits of the very small background in the low m_X region, while the $bb\tau\tau$ and bbbb ones can take advantage of their larger branching fraction at higher values of m_X where the background contribution is reduced. The differences observed between the ATLAS and CMS bbbb results have been investigated and arise from the better b tagging performance in the first case. This is due to the additional layer of pixel tracking detector that has been installed before the restart of the Run II data taking. The CMS experiment also recently completed an upgrade of its inner pixel tracking detector, that is operational since the beginning of the 2017 data taking. Improvements in the b tagging performance of the CMS experiment are consequently expected and could largely impact not only the HH searches in the bbbb final state, but in general other channels containing a $H \to bb$ decay, including $bb\tau\tau$.

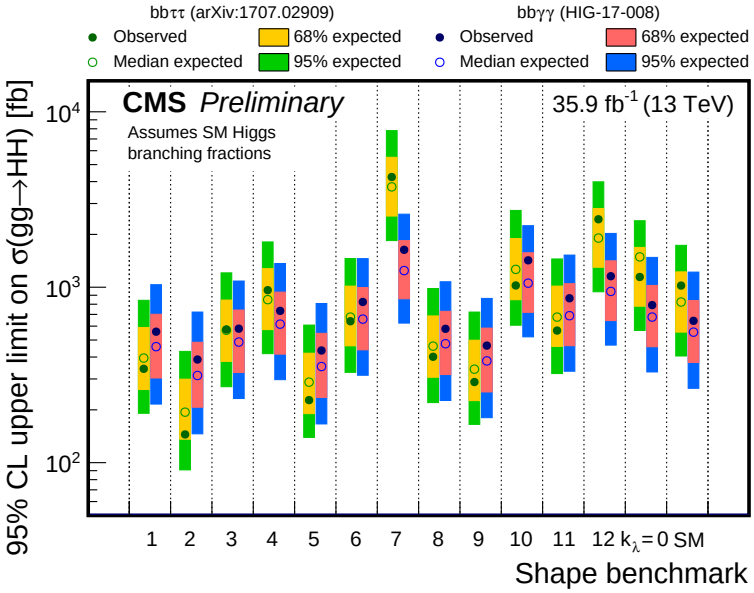


Figure 6.20 – Comparison of the 95% CL upper limits obtained in the $\mathrm{bb}\tau\tau$ (left, green and yellow) and $\mathrm{bb}\gamma\gamma$ (rigth, blue and red) final states for different nonresonant signal models. The numbers from 1 to 12 denote the corresponding shape benchmark, while "SM" and "$k_\lambda = 0$" denote respectively the HH production under the SM and $k_\lambda = \mathrm{c}_2 = \mathrm{c}_\mathrm{g} = \mathrm{c}_{2\mathrm{g}} = 0$ hypotheses.

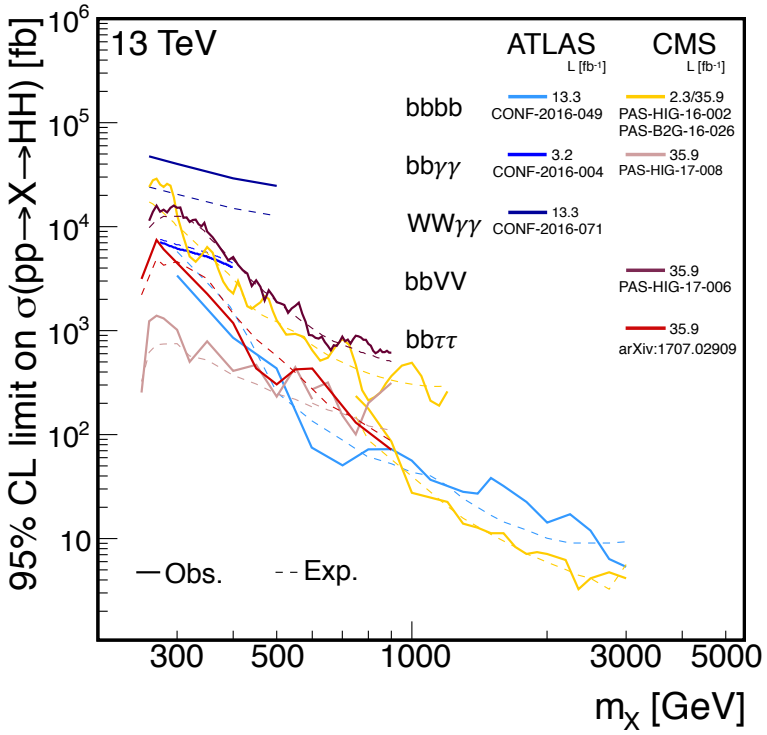


Figure 6.21 – Comparison of the 95% CL upper limit on resonant HH production in the different channels explored at $\sqrt{s} = 13\,\mathrm{TeV}$ by the ATLAS and CMS Collaborations. The solid and dashed lines denote the observed and expected upper limits, respectively. The ATLAS (CONF-2016-049) and CMS high mass (PAS-B2G-16-026) bbbb limits refer to the spin-2 hypothesis while the others are computed for a spin-0 resonance.

Chapter 7

Future prospects for HH searches

The high luminosity phase of the LHC (HL-LHC) is planned to follow a LHC Run III where 300 fb^{-1} of pp collisions are expected to be collected, and to start after the completion of the LHC upgrades during the long shutdown in 2024-2025. The goal is to increase the integrated luminosity further by a factor of 10, collecting a dataset of about 3000 fb^{-1} in a decade of pp collisions at $\sqrt{s} = 14\,\text{TeV}$. It is an unprecedented opportunity to study very rare phenomena, where HH production represents a top priority subject of research. An early understanding of the sensitivity to observe HH production and measure the Higgs boson trilinear coupling is thus important to organize future searches.

Two concurring effects must be taken into account in this context. First, HL-LHC collision conditions will be very different from those experienced in the 2016 data taking: the average pileup of 140 and significant radiation damages will heavily impact the performance of the detector. However, upgrades of the tracker and calorimeter subdetectors and of the trigger system are planned to reduce their consequences and improve the overall CMS performance. As a consequence, object reconstruction and analysis techniques are expected to change and possibly to improve owing to both the experience gained with Run II and Run III data and the upgraded detector. Systematic uncertainties affecting the analyses are expected to be sensibly reduced: experimental effects will be better controlled thanks to the large amount of data collected. Theoretical uncertainties, although more difficult to quantify, are also expected to be reduced thanks to improvements taking place in the next decade. Reasonable guesses of these effects and of their impact on the analyses sensitivities are made and combined into different "scenarios" that can be regarded as more or less conservative.

Two complementary approaches are followed to apply these scenarios to a specific data analysis. In the first one, a parametric simulation of the detector is performed to take into account the machine conditions and CMS detector performance. Because reconstruction algorithms for HL-LHC conditions are not fully developed yet, the simulation relies on parametric object resolutions, efficiencies and misidentification rates that are applied to the generated particles either directly or using the DELPHES 3 [180] software. A second approach consists in extrapolating the Run II results to higher integrated luminosities, taking into account changes in signal and background expected yields and in systematic uncertainties. The two approaches clearly have distinct advantages and drawbacks. While a parametric simulation makes it possible to optimize the analysis strategy for the HL-LHC conditions (for example by modifying the object selections), some experimental effects are poorly modelled: an important limitation of DELPHES 3 for the $\text{bb}\tau\tau$ search is the absence of modelling of jet misidentification as a τ_h object. Inversely, these experimental effects are better accounted for in an extrapolation of existing results, but

neither an optimization of the analysis strategy nor a modification of the selections can be performed: the results thus suffer from the limitations imposed by the small volume of data presently analysed.

The parametric simulation approach has been applied to the study of HH production by both the ATLAS and CMS collaborations. The significance predicted for SM HH production is of 1.05σ in the $\mathrm{bb}\gamma\gamma$ final state [181] (ATLAS) and 0.9σ in the $\mathrm{bb}\tau\tau$ final state [182] (CMS).

The results presented in this chapter follow the second approach and extrapolate the results of the Run II HH search in the $\mathrm{bb}\tau\tau$ final state to the integrated luminosities expected for the LHC Run II and HL-LHC. A first estimation of the sensitivity is presented using the results based on a dataset of 2.7 fb^{-1}, that we realized in the context of the third workshop of the European Committee for Future Accelerators (ECFA) held in 2016, and allows for the comparison of several HH decay channels. Because of the large improvements subsequently achieved in the $\mathrm{bb}\tau\tau$ search, updated extrapolations based on the results derived on the full 2016 dataset are subsequently presented.

7.1 ECFA 2016 extrapolation

The results presented in this section were a first attempt toward establishing the sensitivity to HH production at the HL-LHC after the very first Run II results from the CMS experiment. They are part of a more vast effort to project the performance of Higgs boson analyses in HL-LHC conditions that has been realized in the context of the 2016 ECFA workshop and which results are documented in [183].

The $\mathrm{bb}\tau\tau$ results derived on a dataset of 2.7 fb^{-1} are used for this extrapolation. They correspond to the Moriond 2016 nonresonant search, documented in Ref. [122] and summarized in the Appendix A of this thesis. An expected 95% CL upper limit of about 220 times the SM prediction is achieved. Extrapolations are also performed using the results of the $\mathrm{bb}\gamma\gamma$, bbVV, and bbbb, that were derived for integrated luminosities between 2.3 and 2.7 fb^{-1}.

As discussed above, the definition of the projection scenarios must account for both the harsher collision conditions and the foreseen CMS upgrades. The baseline assumption of the projection is that the performance of the CMS detector will be unchanged with respect to the one observed in 2016. Although this could appear as a very strong assumption, it is motivated by the performance studies of the upgraded detector [182]. Moreover, the possible deterioration of the detector performance is expected to be compensated by improvements in the object reconstruction and data analysis techniques.

Two possible scenarios are constructed from this baseline and differ in the assumptions on the magnitude of the systematic uncertainties affecting the analyses:

- Scenario S1: all the systematic uncertainties are assumed to be constant as a function of the integrated luminosity, and the performance of the detector is assumed to be unchanged with respect to the one observed in the 2016 data analysis. Uncertainties arising from data-driven methods are assumed to be negligible thanks to the large amount of data analysed.

- Scenario S2: improvements are assumed in the detector performance and in the theoretical understanding of the physics processes. Systematic uncertainties on the

trigger, identification and selection of leptons are assumed to be reduced down to 1% for muons and electrons and down to 2% for τ_{h}. The uncertainty in b tagging efficiency is assumed to be reduced down to 2% for genuine b jets and the jet energy scale determination is expected to reach a precision of 1%. The integrated luminosity is assumed to be known with a precision of 1.5%. Shape uncertainties affecting the $\mathrm{bb}\tau\tau$ final state, such as top p_{T} differential distribution for $\mathrm{t\bar{t}}$ events and bin-by-bin uncertainties, are assumed to be negligible. The uncertainty on the multijet background is also assumed to vanish thanks to the large amount of data available to constrain it. Theoretical uncertainties are also reduced by a factor of 2.

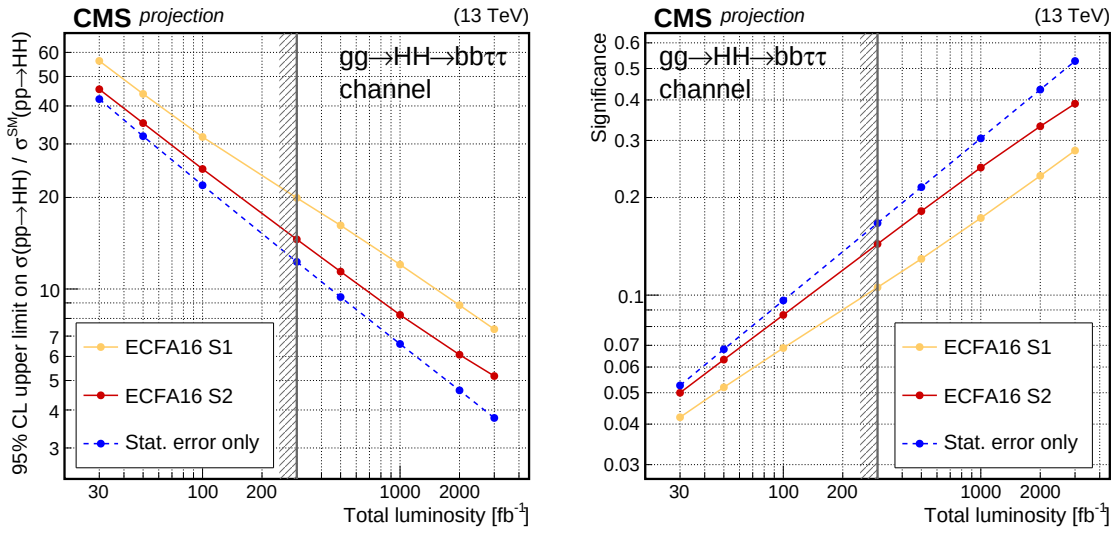
The S1 and S2 scenarios are expected to bracket any future performance of the CMS detector. If scenario S1 is conservative and assumes that no changes will occur from what experienced in Run II, scenario S2 makes important assumptions on the effects of the upgraded detector and on the improvements in theoretical predictions.

In both scenarios, improvements in the background rejection are also assumed: a reduction by a factor of two of the jets rate misidentification as τ_{h} candidates is expected to reduce the multijet contribution by a factor of four and the $\mathrm{t\bar{t}}$ contribution by a factor of two, where the latter is determined according the fractions of jet misidentified as τ_{h} in simulated $\mathrm{t\bar{t}}$ events. This impacts the sensitivity of the search and improves it by about 15%. The assumptions in background rejection might seem quite strong, but were verified in the analysis of the full 2016 dataset where, as it will be shown later, the sensitivity has already exceeded the most optimistic projections. It should also be noted that the normalization of the processes is maintained to the corresponding cross section at $\sqrt{s} = 13\,\mathrm{TeV}$, and a significant gain can result from the increase to 14 TeV.

The expected significance and the 95% CL upper limit on the SM prediction as a function of the integrated luminosity analysed are shown in Figure 7.1 for the scenarios S1 and S2, and in presence of statistic uncertainties only. These figures show that even in the most optimistic scenario S2, a significance below 0.4σ is expected, or equivalently a limit of 5 times the SM prediction. The sizeable impact of systematic uncertainties can be appreciated by comparing S1 and S2 curves, although statistical uncertainties are the main responsible for the limited sensitivity, which is observed to continuously improve with the integrated luminosity.

A comparison with other HH final states is illustrated in Figure 7.2 for the expected uncertainty on the measured HH signal strength $\mu = \sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{\mathrm{SM}}$, and in Table 7.1 where the limit on σ_{HH} and the expected significance are also quoted. The $\mathrm{bb}\tau\tau$ decay channel appears to be one of the least sensitive in this projection, and its expected significance is lower than the one of 0.9σ previously estimated with a parametric simulation [182]. The causes have been studied in detail and shown to be a consequence of the different assumptions in the background contamination and analysis improvements. In particular, the multijet background is assumed to be negligible and different discriminating variables are used in the parametric simulation, which results in a better signal purity and thus in a higher significance. Less conservative assumptions are also made in the case of the $\mathrm{bb}\gamma\gamma$ and bbVV decay channels, by assuming that all the backgrounds can be estimated from data with negligible systematic uncertainties, as well as by hypothesizing sizeable improvements in the analysis performance.

These results may seem discouraging because they indicate that little sensitivity to HH production can be achieved even with the very large luminosities expected by the end of HL-LHC operations. Observing HH production will be extremely challenging and no clear



(a) 95% CL upper limit on $\sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{\mathrm{SM}}$

(b) Significance

Figure 7.1 – Projected 95% CL upper limit on SM HH production (a) and significance (b) as a function of the integrated luminosity analysed. The yellow and red solid lines denote the S1 and S2 scenarios, respectively, while the dashed blue lines denotes the expectation in presence of statistic uncertainties only. The dashed vertical line corresponding to an integrated luminosity of 300 fb^{-1} represents the beginning of the HL-LHC phase, and the assumptions made are not strictly valid below this value.

"golden channel" can be identified: many different final states will need to be explored and combined to improve the sensitivity. These projections should however be regarded as conservative and represent a lower limit on the sensitivity that can be achieved to HH production: while the two scenarios well enclose the future CMS detector performance in terms of systematic uncertainties, they do not represent the improvements in terms of object reconstruction and identification and analysis techniques that larger integrated luminosities will make it possible to achieve.

We can observe that with the data collected in 2016, corresponding to an integrated luminosity of about a factor 10 larger than the one used for the projection, the predicted sensitivity has already been exceeded. While the extrapolation of Figure 7.1a at an integrated luminosity of 35.9 fb^{-1} sets a limit of about 40 times the SM prediction in the most optimistic scenarios, the results discussed in Chapter 6 set an upper limit of 25 times $\sigma_{\mathrm{HH}}^{\mathrm{SM}}$, an improvement of about 40%. With the sensitivity currently achieved, a combination of the most sensitive channels at the end of the Run III with about 100 fb^{-1} of pp collisions will likely result in a limit below 10 times the SM HH cross section. These remarkable improvements arise from the optimization of the search that was made possible thanks to the larger dataset analysed and the better understanding of background contamination and detector effects, and give new hope for HH observation at the HL-LHC.

The above considerations motivate an update of the extrapolation using the results the latest 2016 results in the bb$\tau\tau$ final state. Expected improvements in the analysis performance and the corresponding projected sensitivity are presented in the following part of this chapter.

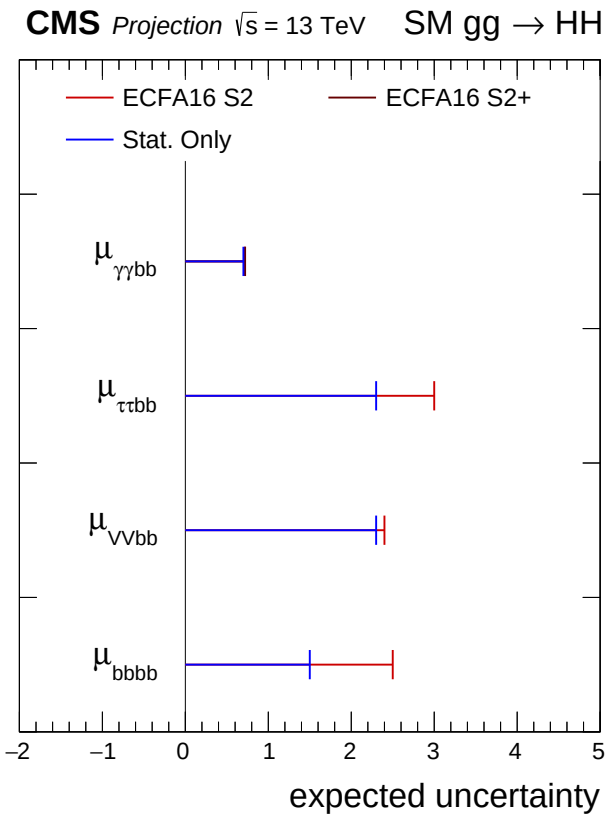


Figure 7.2 – Expected uncertainty on the measured HH signal strength $\mu = \sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{\mathrm{SM}}$ for the four HH decay channels studied in the extrapolation. The red line denotes the S2 scenario described in the text; for the $\mathrm{bb}\gamma\gamma$ decay channel, a S2+ scenario is used that takes into account additional effects of pileup and detector upgrades.

7.2 Extrapolation using the full 2016 dataset

The projection of the results obtained with the full 2016 dataset are presented in this section. The performance scenarios are defined in close analogy with those of the ECFA projection, and possible improvements in the signal efficiency and background rejection are estimated. Such improvements can arise from both the upgrades of the CMS detector and from changes in the analysis strategy to target the largest sources of signal inefficiency and the main background contaminations, and they are discussed starting from the results of Chapter 6. As before, the extrapolation focuses on SM HH production and on the determination of the λ_{HHH} coupling, but the reader should not forget the broad BSM physics program accessible through both resonant and nonresoant HH production.

7.2.1 Data analysis improvements

Better background rejection and signal efficiency are key factors to enhance the sensitivity of the search, and the main factors limiting them are studied from the results presented in Chapter 6. It should however be noted that large improvements can also be obtained with other changes to the analysis strategy, such as the usage of more sensitive observables. For example, the larger amount of events collected in the high luminosity phase will make it possible to perform a signal extraction based on the simultaneous values of $\mathrm{m_{T2}}$ and

Channel	Median expected limits on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$		Significance (Z-value)		Uncertainty as a fraction of μ	
	ECFA16 S2	Stat. only	ECFA16 S2	Stat. only	ECFA16 S2	Stat. only
$HH \to bb\gamma\gamma$(S2+)	1.44	1.37	1.43	1.47	0.72	0.71
$HH \to bb\tau\tau$	5.2	3.9	0.39	0.53	2.6	1.9
$HH \to bbVV$	4.8	4.6	0.45	0.47	2.4	2.3
$HH \to bbbb$	7.0	2.9	0.39	0.67	2.5	1.5

Table 7.1 – Comparison of the 95% CL upper limit on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$, significance, and uncertainty on μ for the four HH decay channels results extrapolated to a luminosity of 3000 fb^{-1}.

other observables such as the output of the BDT discriminant. These possibilities are clearly very hard to evaluate quantitatively in the context of an extrapolation and will not be explored in the following. The discussion will consequently assume that the general analysis strategy is unchanged.

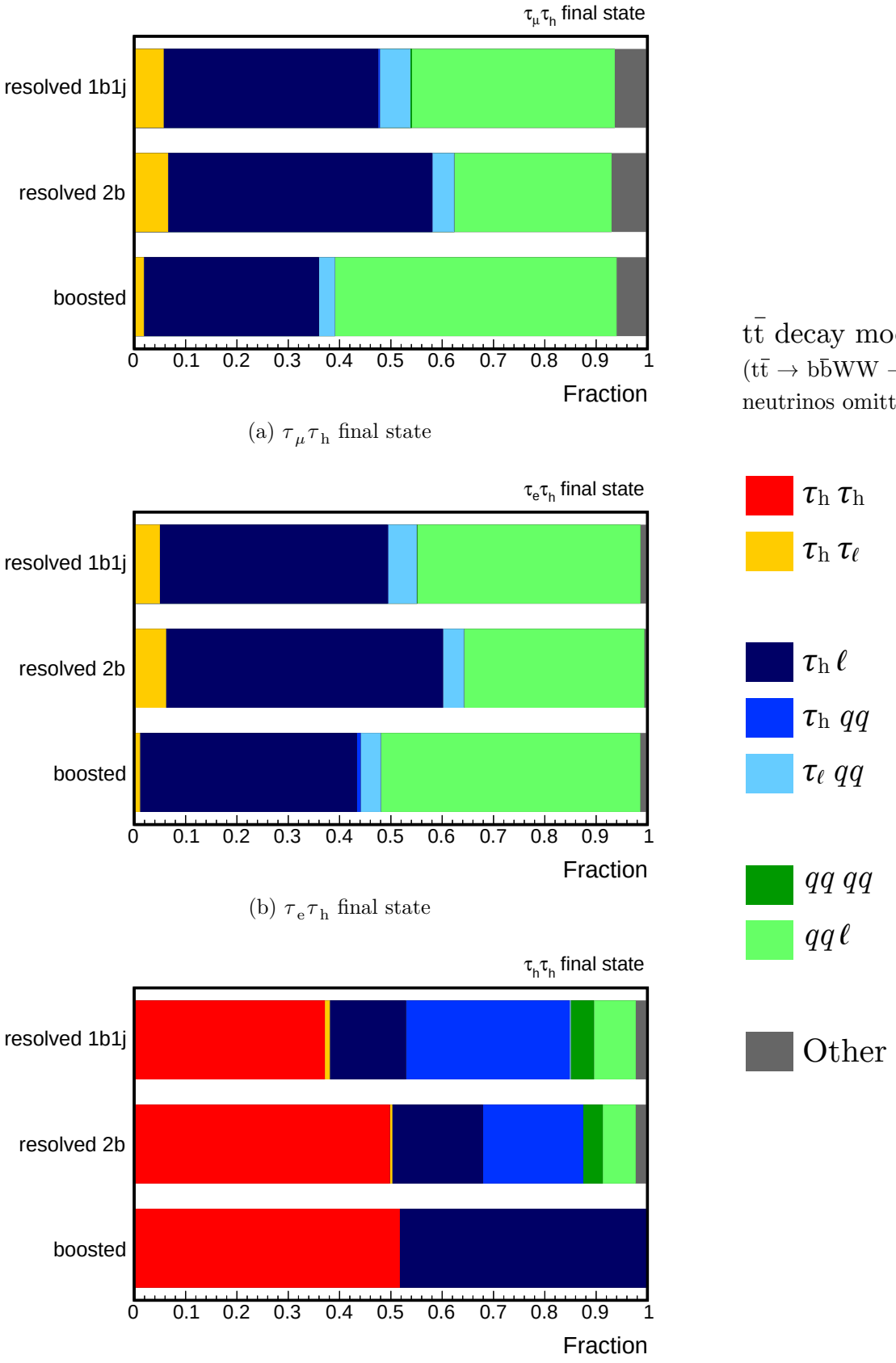
Background rejection

The $t\bar{t}$ and Drell-Yan processes constitute the major background sources. The development of improved rejection strategies is based on the following considerations.

- Different contamination in the flavour of the jets produced in the Drell-Yan background (DY_{2b}, DY_{1b}, and DY_{0b}) is observed depending on the event category. The contamination from DY_{2b} accounts for about 70% of the events in the 2b resolved event categories, but its fraction is only about 10% and 20% in the 1b1j resolved and boosted categories, respectively. The 1b1j resolved category shows an almost equal contamination from DY_{1b} and DY_{0b}, while the boosted category is dominated by DY_{0b} that represents about 80% of the total background contribution. These numbers are approximately constant for all the $\tau\tau$ final states. Such differences in these fractions are expected from the b tagging criteria applied. DY_{0b} contamination in the boosted category may be reduced by tightening the working point of the discriminant (currently the loose one is used because of the small number of events expected).

- The dominant Drell-Yan contribution is realized through the $Z/\gamma^* \to \tau\tau$ decay mode, with the two τ leptons subsequently decaying to $\tau_\mu\tau_h$, $\tau_e\tau_h$, or $\tau_h\tau_h$ depending on the final state considered. Between 85% and 90% of the selected Drell-Yan events correspond to this decay, with minor contamination from prompt electron or muon production; these numbers are consistent between the various bb event categories explored. This shows the effectiveness of the object quality selection and additional lepton veto, and indicate that a further suppression of this background must rely on the kinematic properties of the selected $\tau\tau$ and bb candidates.

- The contamination from $t\bar{t}$ events differs according to the final state considered. The fraction of generated $t\bar{t}$ decays that contaminate the three bb categories are separately shown for the three $\tau\tau$ final states in Figure 7.3. It can be observed that prompt production of a muon or an electron in the $t\bar{t}$ decay dominates the $\tau_\mu\tau_h$



(a) $\tau_\mu\tau_h$ final state

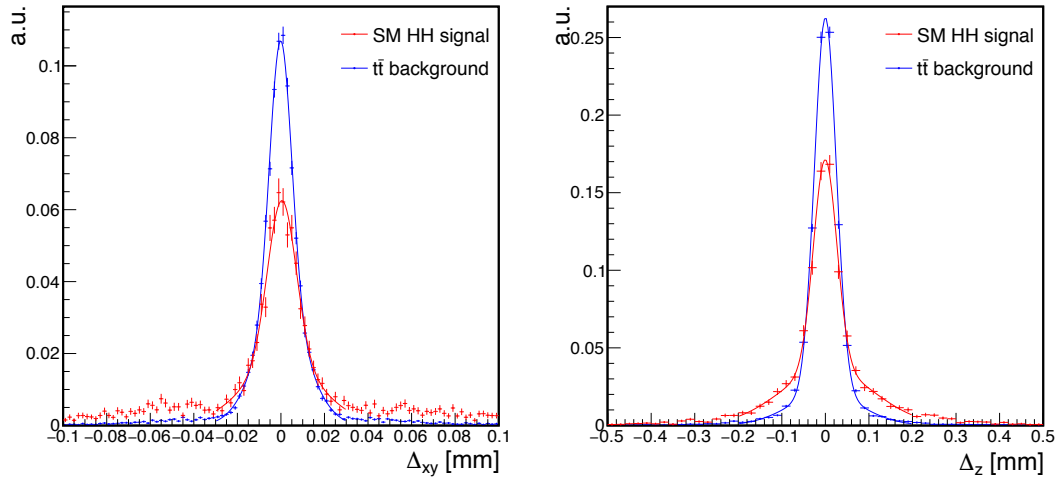
(b) $\tau_e\tau_h$ final state

(c) $\tau_h\tau_h$ final state

Figure 7.3 – Decay modes of the selected $t\bar{t}$ events that contaminate the three signal regions of the $\tau_\mu\tau_h$ (a), $\tau_e\tau_h$ (b), and $\tau_h\tau_h$ (c) final states. The symbols ℓ and qq denote respectively $W \to \ell\nu_\ell$ and $W \to qq$ decays, while τ_ℓ and τ_h indicate a $W \to \tau\nu_\tau$ decay with the τ lepton subsequently decaying leptonically or semileptonically, respectively.

and $\tau_e\tau_h$ final states, and that between 35% and 50% of the events contain at least one $W \to qq'$ decay where one jet is misidentified as a τ_h. The situation is different in the $\tau_h\tau_h$ final state, where a fraction of the events ranging between 35% and 50% is due to an irreducible $t\bar{t} \to b\bar{b}WW \to b\bar{b}\tau\tau \to b\bar{b}\tau_h\tau_h$ background, while most of the residual events contains a genuine τ_h decay and a misidentified jet.

The different $t\bar{t}$ background contamination is at the origin of the superior performance of the $\tau_h\tau_h$ final state with respect to the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ ones, and indicates that the sensitivity of the latter can be enhanced by minimizing the contribution from prompt leptons. Although the separation of a $\tau \to \ell\nu_\ell\nu_\tau$ decay from a prompt lepton production is challenging, help can come from the distance of the muon or electron track from the primary vertex. As shown in Figure 7.4 for selected muon candidates, leptons in $t\bar{t}$ events are typically closer to the primary interaction vertex than signal ones; the usage of this variable in the BDT method could help to reduce the $t\bar{t}$ contamination.



(a) xy plane (b) z direction

Figure 7.4 – Distance of the selected muon track from the primary vertex in the direction transverse (a) and parallel (b) to the beam pipe. The comparison is done between the events selected in the resolved 2b events category of the $\tau_\mu\tau_h$ final state. The distributions are normalized to a unitary area.

Further reduction of the $t\bar{t}$ contribution can be achieved by targeting those events where a quark jet is misidentified as a τ_h. Because this jet is produced from a quark in a $t \to bW \to bqq'$ decay, these events can be identified from the presence of additional jets with a specific topology. The invariant mass of the selected τ_h and jet candidates, and the three-body invariant mass obtained by combining the b jet, are strong signatures of these specific decays, and top-tagging algorithms [184] can be applied to suppress their contribution.

Finally, the $t\bar{t}$ contribution in the $\tau_h\tau_h$ final state can be suppressed together with the one from the Drell-Yan background by adapting to this final state the multivariate methods already used in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ decay channels. While this could not be developed until now because of the small number of events expected in this final state, the increase of the integrated luminosity will allow for a precise control of all the input variables and the verification of their correct modelling in the simulation.

Signal efficiency

The curves reported in Section 4.7 of Chapter 4 show that the signal efficiency is limited since the very first selection represented by the trigger, an effect that is particularly severe in the $\tau_h\tau_h$ final state. This is caused by the high p_T thresholds that reject a large fraction of the signal events. For SM HH production, rather soft leptons and τ_h candidates are expected, as it can be seen from the distributions obtained from generated HH $\rightarrow$ bb$\tau\tau$ decays shown in Figure 7.5.

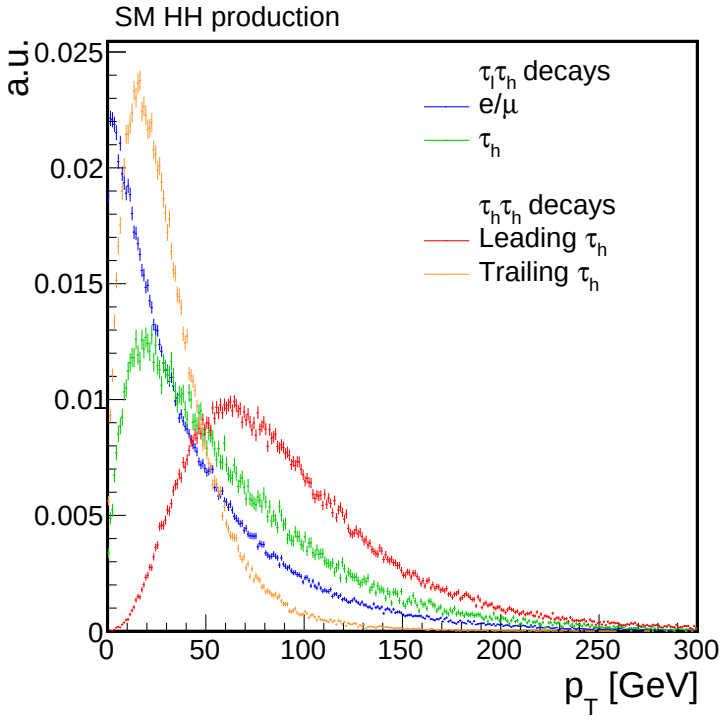


Figure 7.5 – p_T distribution of lepton and τ_h candidates from $\tau\tau$ decays in simulated SM HH events. The distribution are computed on generated particles. Green and blue lines refer to $\tau_\mu\tau_h$ and $\tau_e\tau_h$ decays, that have been added together because the same distributions are expected. Red and orange lines refer to $\tau_h\tau_h$ decays.

The L1 trigger upgrade and the improvements in τ_h reconstruction discussed in Chapter 3 open new opportunities to address this limitation and increase the signal efficiency. In particular, dedicated L1 seeds and HLT paths can be designed to reduce the trigger thresholds by requiring the simultaneous presence of two jets together with the $\tau\tau$ object pair in the events. The usage of b tagging criteria at the HLT, already well established and used in other HH searches such as the one in the bbbb final state, can be particularly important in this context: because of the rapidly decreasing p_T spectrum of the leptons and τ_h candidates, even a reduction of a few GeV on the threshold can result in large improvements in the signal acceptance. For the HL-LHC another trigger upgrade is planned to make track information available at the hardware level. The algorithms are still being discussed but the potential of this upgrade is very large, so that low trigger thresholds even in the HL-LHC conditions do not seem unrealistic at the moment.

Improvements in the b tagging performance that results from the recent upgrade of the pixel tracker detector are also expected to positively impact the signal acceptance.

Studies performed on MC simulation [86] showed that an improvement in the identification of genuine b jets of about 15% for the same misidentification rate are expected. As the detector is now installed and fully commissioned, its impact on the b tagging is being verified with 2017 data.

7.2.2 Performance scenarios

Three performance scenarios are defined. Assumptions are made on both the systematic uncertainties and on the contamination from background processes to bracket possible future performance of the detector and of the analysis. The scenarios are defined as follows:

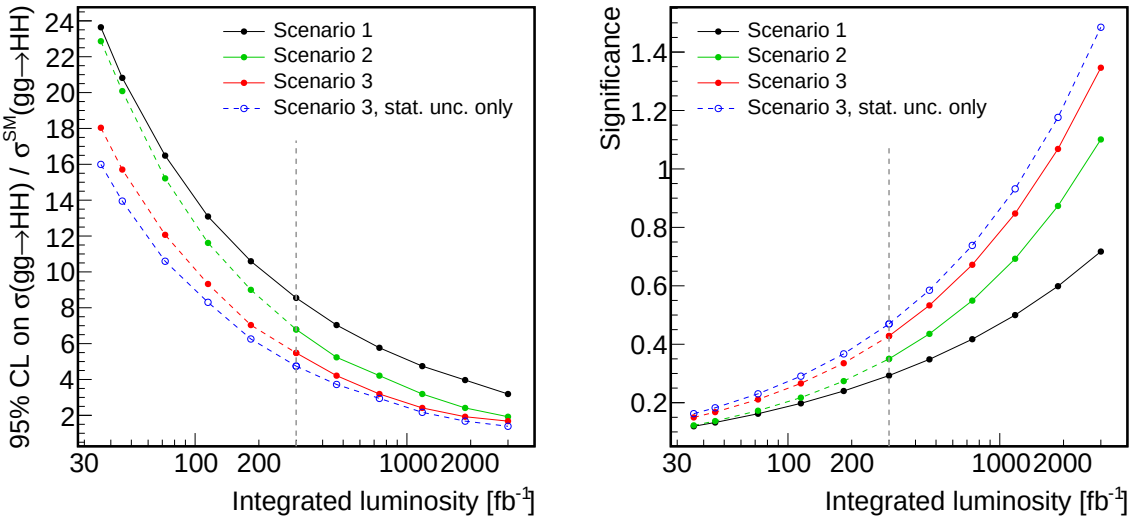
- Scenario 1: everything is assumed to be unchanged and only the integrated luminosity is varied in the extrapolation. This is the most conservative scenario.
- Scenario 2: the same reduction in the systematic uncertainties as the S2 ECFA scenario detailed in Section 7.1 are assumed, corresponding to smaller uncertainties on the object efficiencies and scale, on the integrated luminosity, and on the theoretical predictions. Bin-by-bin, top p_{T}, and Drell-Yan correction factors uncertainties are assumed to be negligible because of the large amount of data available to constrain them.
- Scenario 3: in addition to the reduction in the systematic uncertainties from the previous scenario, improvements in the b tagging are assumed. For the same misidentification rate, the b jet identification efficiency is assumed to increase by 15% (from 65% to 75% for the medium working point), resulting in an increase of about 30% of all the processes containing two b quarks and of 15% for those containing one b quark. The contamination from the $\mathrm{t\bar{t}}$ background is also assumed to be reduced by 30% because of the possible improvements previously discussed.

It should be noted that further improvements are expected from the change of the centre-of-mass energy from 13 to 14 TeV, where both the SM signal and $\mathrm{t\bar{t}}$ background cross sections are expected to increase by 18%, and possibly from improved trigger algorithms. These are not taken into account here because of the changes that they would induce in the $\mathrm{m_{T2}}$ distribution of the simulated events.

7.2.3 Extrapolation of the results

The projected 95% CL upper limit on $\sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{\mathrm{SM}}$ and the signal significance under the three performance scenarios are shown in Figure 7.6.

These projections demonstrate that, even in the conservative scenario 1, the sensitivity from the ECFA extrapolation shown in Figure 7.1 has been exceeded: a limit of about 3 times the SM prediction, for a significance around 0.7σ, are expected by the end of the HL-LHC for the $\mathrm{bb}\tau\tau$ channel alone. The sensitivity is largely improved in the scenario 2: while systematic uncertainties have little impact for small integrated luminosities, they become relevant for larger datasets. An important remark is that the individual importance of the different sources of systematic uncertainties changes depending on the integrated luminosity. In particular, bin-by-bin uncertainties have little impact with respect to other uncertainties for luminosities below 300 fb^{-1}, but they become the dominant effect for



(a) 95% CL upper limit on $\sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{\mathrm{SM}}$

(b) Significance

Figure 7.6 – Projected 95% CL upper limit on SM HH production (a) and significance (b) as a function of the integrated luminosity analysed. The black, green, and red lines denote the scenarios 1, 2, and 3, respectively, while the blue line denotes the scenario 3 in presence of statistic uncertainties only. The dashed vertical line corresponding to an integrated luminosity of 300 fb^{-1} represents the beginning of the HL-LHC phase, and the assumptions made are not strictly valid below this value.

larger integrated luminosities: at 3000 fb^{-1}, their removal is responsible alone for half of the improvement observed between scenario 1 and 2. As these uncertainties are expected to be sizably reduced in the future because of the large datasets collected (and possibly large MC samples with an equivalent luminosity), it is likely that this assumption of the scenario 2 will be verified. Finally, a sizeable effect is observed from the assumed improvements in signal efficiency and background rejection: in this more optimistic scenario, a sensitivity of about 1.3σ, and an upper limit of about 1.7 times the SM prediction, are expected.

The comparison of Table 7.2 illustrates the better sensitivity achieved with respect to the ECFA projections. Such improvement follows from the more advanced and optimised analysis methods that could be developed with a larger integrated luminosity analysed. Nevertheless, they confirm that evidence of HH production will be difficult to achieve in a single channel at the HL-LHC. The combination of several decay channels will be crucial to establish the existence of this important process by the end of the HL-LHC operations.

If HH production seems at reach for the HL-LHC, a precise determination of the trilinear Higgs boson coupling will be more difficult to achieve. The expected 95% CL upper limit on σ_{HH} as a function of the k_λ/k_{t} ratio is shown in Figure 7.7. Assuming that the y_{t} coupling corresponds to the SM prediction ($k_{\mathrm{t}} = 1$), these results show that even in the most optimistic scenarios only loose constraints on the value of λ_{HHH} will be set. While negative couplings could be excluded in the scenario 3, values of k_λ up to 8 will be very difficult to probe because of the soft Higgs boson pair p_{T} spectrum and, consequently, small signal acceptance and sensitivity from the m_{T2} variable. This indicates that dedicated methods will need to be developed to enhance the sensitivity in

Scenario	Median expected limits on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$		Significance (Z-value)	
	ECFA	Full 2016 dataset	ECFA	Full 2016 dataset
S1	7.4	3.2	0.28	0.72
S2	5.2	1.9	0.39	1.10
S3	–	1.7	–	1.35

Table 7.2 – Projected sensitivity to HH production in the $bb\tau\tau$ decay channel for an integrated luminosity $L = 3000$ fb^{-1}. The 95% CL upper limit on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ and the significance are compared for the ECFA extrapolation (based on $L = 2.7$ fb^{-1}) and for the extrapolation of the results obtained with the full 2016 dataset (based on $L = 35.9$ fb^{-1}).

the low momentum region to improve the sensitivity to this important coupling; some of the improvements discussed in Section 7.2.1 can help to achieve this goal.

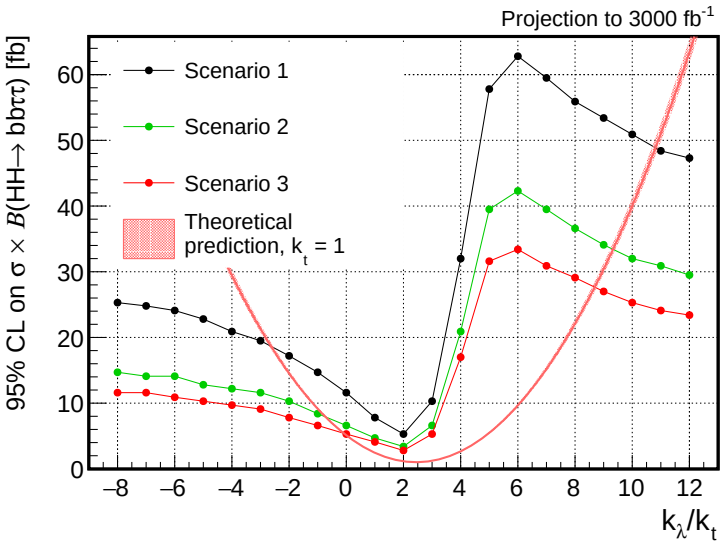


Figure 7.7 – 95% CL upper limit on σ_{HH} as a function of k_λ/k_t for the three extrapolation scenarios. The theoretical prediction for $k_t = 1$ is superposed.

7.3 General prospects for HH searches

The extrapolations presented in this chapter focused on the measurement of the HH cross section and on the determination of λ_{HHH} in the context of the SM. The clean theoretical prediction for σ_{HH} and the importance of λ_{HHH} in the SM formulation provide an ideal benchmark for the evaluation of HH analyses sensitivity, which will represent one the top priority research subjects in future HL-LHC operations.

Nevertheless, it would be reductive to consider HH searches uniquely as a test of the SM, a goal that can be achieved only by the end of the HL-LHC. As detailed in this thesis, these searches are sensitive to the existence of physics beyond the SM, and represent one of

the major channels for BSM physics discovery at the LHC. It is thus important to remark the versatility of HH searches and their broad reach for different BSM physics models via either resonant or nonresonant production. Short and medium-term LHC operations will possibly reveal something new about high energy physics before sensitivity to $\sigma_{\mathrm{HH}}^{\mathrm{SM}}$ is achieved with the large datasets of the HL-LHC.

The analysis of an integrated luminosity of $100 - 150$ fb^{-1}, that is expected to be collected in the next two years at $\sqrt{s} = 13$ TeV in the Run II, will improve the current sensitivity to HH production by about a factor of 2. Even more interestingly, the Run III operations are foreseen to deliver a total integrated luminosity of about 300 fb^{-1} at a larger centre-of-mass energy of 14 TeV. These data will allow for setting precise constraints on resonant HH production or, hopefully, for its discovery in the LHC data. They will also make it possible to study in details the five-dimensional parameter space of the effective HH Lagrangian, providing access also to BSM models beyond the direct LHC reach.

In the unfortunate eventuality that no clear signs of BSM physics are identified by the end of the Run III, HH will continue to be a key research subject at the HL-LHC for the determination of λ_{HHH}, with a further motivation to search for the well-hidden presence of BSM physics in deviations of this coupling. In general, the projections of the $\mathrm{HH} \to \mathrm{bb}\tau\tau$ search discussed above show that by the end of the HL-LHC, measuring σ_{HH} will be challenging. My opinion is however that these projection are quite conservative and should not be interpreted as an impossibility to observe this process at the HL-LHC. They indicate for sure that this is a tremendous challenge and that much effort will be required to improve the detector performance, the understanding of the systematic uncertainties, and the analysis techniques. However, we should not expect only adiabatic improvements as assumed in these projections: new ideas can be developed and tested in the fertile ground of large recorded datasets and can impact the sensitivity of HH searches much more than postulated in these projections.

Moreover, these projections for the $\mathrm{bb}\tau\tau$ decay channel should be put in the broader context of HH searches performed at the LHC. The current Run II results indicate that at least three decay channels ($\mathrm{bb}\gamma\gamma$, $\mathrm{bb}\tau\tau$, and bbbb) are expected to yield similar sensitivities. A naïve combination in quadrature of these results gives a 95% CL upper limit of approximately $12 \times \sigma_{\mathrm{HH}}^{\mathrm{SM}}$. Assuming a scaling of the sensitivity as $\sqrt{L}$ and a similar performance of ATLAS searches, the combined results of the two experiments at the end of the HL-LHC operations will be sensitive to SM HH production. This can be also observed for the ECFA extrapolation presented in Section 7.1, that clearly indicates the importance of combining several final states as well as the results of both experiments. In the preparation for the HL-LHC operations, these decay channels represent an important physics case for the upcoming upgrades of the detectors. Moreover, the larger integrated luminosity will also open up the possibility to study very rare but pure HH decay modes, such as $\mathrm{HH} \to \mathrm{WWWW}$, $\mathrm{HH} \to \mathrm{WW}\tau\tau$, or $\mathrm{HH} \to \mathrm{ZZ}\tau\tau$.

In conclusion, the exploration of HH production will be a collective effort involving several decay channels, with the $\mathrm{HH} \to \mathrm{bb}\tau\tau$ one being expected to have a key role in this broad and important search programme. Their combination, as well as a combination of ATLAS and CMS results, will be crucial to observe HH production. The measurement of σ_{HH} both in the context of the SM and of physics beyond it represents in this sense a very interesting, stimulating, and rewarding challenge for the future LHC and HL-LHC operations.

Conclusions

With the mass of the Higgs boson precisely known and its properties in good agreement with the theoretical predictions, HH production is becoming increasingly important as a probe of the scalar sector of the SM. The presence of new resonances, anomalous couplings, and eventually the Higgs boson self-coupling λ_{HHH} can be simultaneously studied in HH searches.

The restart of the LHC, marked by a centre-of-mass energy increase from 8 to 13 TeV and a higher instantaneous luminosity, that exceeded in 2016 its original design value, represent an unprecedented possibility to explore HH production, but constitutes an experimental challenge for the CMS trigger system. An upgrade of the L1 trigger was performed to address this issue and consisted in the complete replacement of the electronics. In particular, the powerful FPGA now mounted in the calorimeter trigger boards allowed for the development of a sophisticated τ reconstruction algorithm. The development and optimization for the LHC Run II collisions have been realized and the algorithm performance has been validated and measured with data. The results show the significant increase in the CMS triggering capabilities for events containing τ leptons, with a large positive impact for its physics programme.

With a solid trigger strategy, the search for HH production in the $\mathrm{b\bar{b}}\tau^{+}\tau^{-}$ decay channel has been developed and optimized for Run II conditions. Benefiting of a sizeable branching fraction and limited contamination from SM backgrounds, this decay channel is one of the most sensitive to HH production. Three sets of results were published following the quick pace of the Run II LHC operations. The search covered the three main $\tau\tau$ decay modes containing at least one semileptonic τ decay. Multivariate methods were introduced for the first time in the $\tau_{\mu}\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ decay modes. A dedicated event categorization was put in place to simultaneously probe the resonant and nonresonant production mechanisms for a variety of signal hypotheses. The results were found in agreement with the SM background expectation, and the 95% CL upper limits that were set improve significantly the previous results from Run I searches. In particular, the observed upper limit of about $30 \times \sigma_{\mathrm{HH}}^{\mathrm{SM}}$, for an expected limit of about $25 \times \sigma_{\mathrm{HH}}^{\mathrm{SM}}$, represents today one of the most sensitive results on HH production. These results should be regarded in the broader context of CMS HH searches, with four separate decay channels currently probed at 13 TeV, a few others HH searches in preparation, and an upcoming combination of the results.

As this thesis comes to its end after three years, the exploration of the TeV scale has just started. The data collected in the LHC Run II and III, and the subsequent operations of the HL-LHC, will elucidate many open questions on the SM and on its possible BSM extensions. We do not know what these data will reveal us, and whether the SM will continue to resist to experimental tests or if clear signs of BSM physics will appear. For sure, HH searches will be at the forefront of the rich physics programme of the CMS experiment. Run II and Run III data will give an unprecedented insight on the presence

of resonant BSM physics and on anomalous coupling structures of the Higgs boson. If no deviations from the SM are found, the large datasets collected during the HL-LHC operations will ultimately open the way to the measurement of σ_{HH} and the determination of λ_{HHH}. Extrapolations of the current Run II results to HL-LHC luminosities have been performed addressing the latter scenario. These result show that σ_{HH} can possibly be measured at the LHC and that significant constraints the value of λ_{HHH} can be set. Further improvements of the sensitivity are possible and are expected to be achieved with the larger dataset available, some of which have been accounted for in the extrapolation scenarios. Most importantly, these extrapolations show that no "golden channel" exists for a standalone HH production observation, and that several decay channels will need to be explored and combined. From the near future of Run II operations to the end of HL-LHC operations, HH searches will continue to be one of the main paths to the exploration of the SM scalar sector and of the physics at the TeV scale.

Appendix A

Earlier Run II HH $\rightarrow$ $\mathrm{b\bar{b}}\tau^{+}\tau^{-}$ searches

This appendix gives an overview of the Run II searches for HH $\rightarrow$ bb$\tau\tau$ production that were derived in the context of this thesis using the 2015 CMS dataset of 2.7 fb^{-1} and on the first part of the 2016 dataset of 12.9 fb^{-1}. In the following, the two searches are referred to as the "Moriond 2016 search" and the "ICHEP 2016 search" accordingly to the conference in which the corresponding results were presented for the first time. The discussion focuses on the event selection and categorization, and on the results obtained. The selections related to the object quality (compatibility with the primary vertex and identification criteria) are the same discussed in Chapter 4 and thus not repeated here. The MC simulation, the background modelling methods, and the evaluation of the systematic uncertainties also closely follow what has been described in Chapter 5 and are consequently not detailed further.

A.1 Moriond 2016 search

The Moriond 2016 search constituted the first exploration of HH production at 13 TeV, and was based on an integrated luminosity of 2.7 fb^{-1} collected with the CMS experiment in 2015. The corresponding results are documented in Refs. [122] and [123] for the nonresonant and resonant HH production mode, respectively, with additional information provided in the CMS internal analysis notes [124, 125].

A.1.1 Event selection

The three final states containing at least one semileptonic τ decay ($\tau_\mu\tau_\mathrm{h}$, $\tau_\mathrm{e}\tau_\mathrm{h}$, $\tau_\mathrm{h}\tau_\mathrm{h}$) are explored. The dataset analysed corresponds to 2.7 fb^{-1} of data collected in 2015. Because of issues experienced with the HF calorimeter, affecting about 0.4 fb^{-1}, inputs from this subdetector are not used in the computation of the $\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$ vector. It was verified that its magnitude and directions are properly estimated for the selected bb$\tau\tau$ events and well modelled in the MC simulation.

Most of the data in the $\tau_\mu\tau_\mathrm{h}$ and $\tau_\mathrm{e}\tau_\mathrm{h}$ final states are collected with a trigger requiring the presence of, respectively, a single muon or electron in the event. For a first part of the data, corresponding to 76 pb^{-1}, a cross-trigger requiring the additional presence of a τ_h is used. The $\tau_\mu\tau_\mathrm{h}$ events thus recorded must contain a muon candidate of $p_\mathrm{T} > 19\,\mathrm{GeV}$ and $|\eta| < 2.1$ and a τ_h candidate of $p_\mathrm{T} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$. The selected muon must satisfy the relative isolation requirement $\mathcal{I}^\mathrm{rel} < 0.1$ while the selected τ_h must have an

absolute charged hadron isolation $\sum p_{\mathrm{T}}^{\mathrm{ch}} < 3\,\mathrm{GeV}$; both isolation criteria are described in Chapter 4. Similarly, $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ events must contain an electron candidate of $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ and $|\eta| < 2.1$, which satisfies $\mathcal{I}^{\mathrm{rel}} < 0.1$. The presence of a τ_{h} candidate that satisfies the same selections as in the $\tau_\mu\tau_{\mathrm{h}}$ final state is also required. Finally, events selected in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state are selected with a double-τ_{h} trigger, and must contain two τ_{h} candidates reconstructed offline with $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ and $|\eta| < 2.1$, that satisfy $\sum p_{\mathrm{T}}^{\mathrm{ch}} < 2\,\mathrm{GeV}$. In all the three final states, the two leptons that form the $\mathrm{H} \to \tau\tau$ candidate are required to have opposite electric charge, and events containing additional isolated leptons are rejected.

The events selected with the previous criteria must also contain two AK4 jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. For the nonresonant search, both jets in the $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final state must satisfy the loose b tagging WP corresponding to a misidentification efficiency of gluon and light flavour quark jets of 10%. A multivariate discriminant is applied on these events. The variables used as inputs to the method are a subset of those used for the HM BDT described in Section 4.6.2 of Chapter 4. In particular, the $\Delta\varphi(l, p_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell)$, and $m_{\mathrm{T}}(\tau_{\mathrm{h}})$ variables are not used and the method is trained on a combination of nonresonant events corresponding to $k_\lambda = 1$ (SM) and $k_\lambda = 2.45$ (maximal interference of the two diagrams contributing to gluon fusion HH production). For the resonant search, no multivariate method is applied in the $\tau_\mu\tau_{\mathrm{h}}$ and $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$ final states and the two jets must satisfy the medium b tagging WP, corresponding to a misidentification rate of 1%. In the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state, the two jets must satisfy the loose b tagging WP in both the resonant and nonresonant searches.

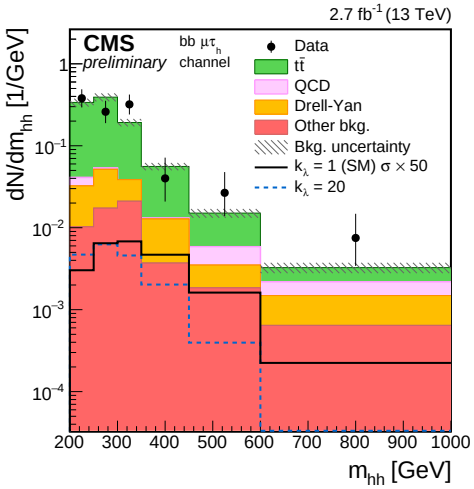
An invariant mass requirement on the bb and $\tau\tau$ invariant masses (the latter reconstructed with the SVfit algorithm) is applied on the selected events as $80 < m_{\mathrm{bb}}(m_{\tau\tau}) < 160\,\mathrm{GeV}$.

A.1.2 Results

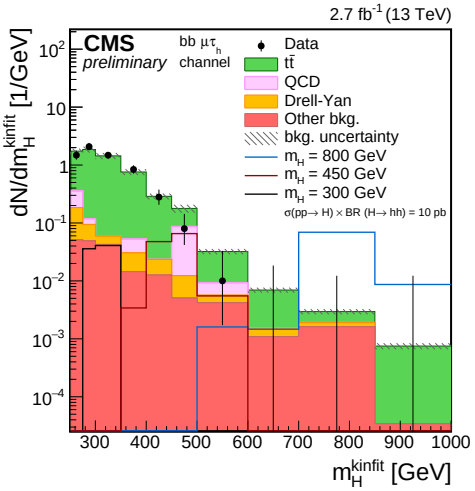
The $\mathrm{m}_{\mathrm{HH}}^{\mathrm{KinFit}}$ variable described in Section 6.2 of Chapter 6 is used for the resonant search, while the four-body mass of the visible $\tau\tau$ decay products and of the selected bb candidates, m_{HH}, is used for the nonresonant search. The corresponding event distributions are shown in Figure A.1 and the observed and expected number of events is summarized in Table A.1. The 95% CL upper limits on nonresonant production as a function of k_λ are summarized in Figure A.2a, and the 95% CL upper limits on resonant production as a function of m_{X} for the spin-0 resonance hypothesis are shown in Figure A.2b. The observed and expected 95% CL upper limits on $\sigma(\mathrm{gg} \to \mathrm{HH})$ for the SM signal amount to 8.8 and 7.2 pb, respectively. These values correspond to approximately 260 and 215 times the SM prediction.

A.2 ICHEP 2016 search

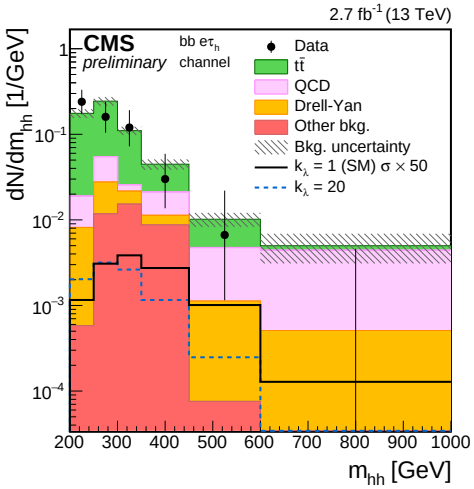
The ICHEP 2016 search was based on the first part of the dataset collected in 2016, corresponding to an integrated luminosity of 12.9 fb^{-1}. The results for the nonresonant and resonant search are documented in Refs. [126] and [127], respectively, and in the supporting CMS internal analysis notes [128, 129].



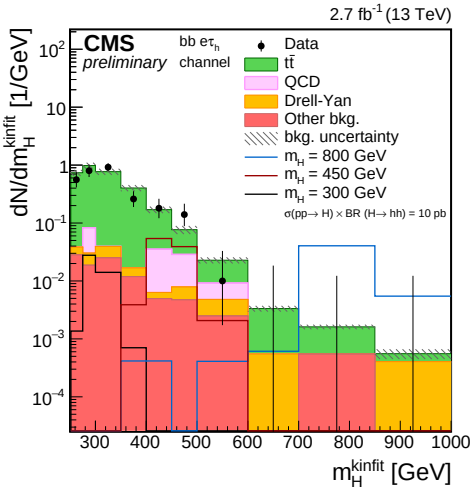
(a) $\tau_\mu\tau_h$, nonresonant



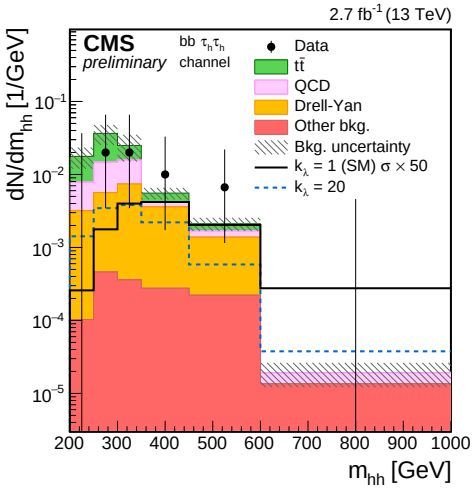
(b) $\tau_\mu\tau_h$, resonant



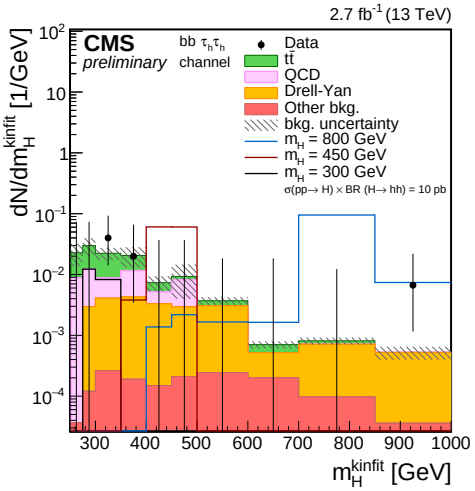
(c) $\tau_e\tau_h$, nonresonant



(d) $\tau_e\tau_h$, resonant



(e) $\tau_h\tau_h$, nonresonant

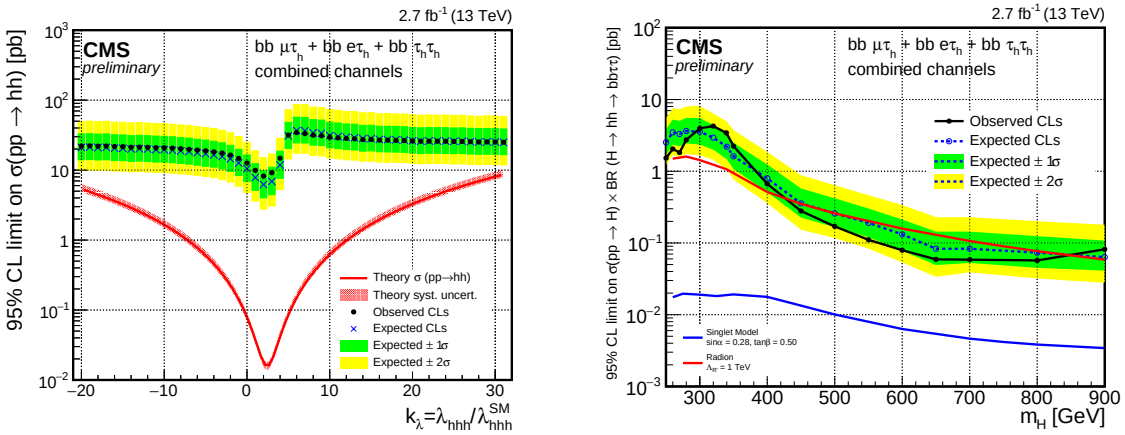


(f) $\tau_h\tau_h$, resonant

Figure A.1 – Event distribution of the m_{HH} (left column) and $\mathrm{m}_{\mathrm{HH}}^{\mathrm{KinFit}}$ (right column) variables for the $\tau_\mu\tau_h$ (top row), $\tau_e\tau_h$ (central row), and $\tau_h\tau_h$ (bottom row) final states. The two variables are used for the nonresonant and resonant searches, respectively. The events are selected from the 2.7 fb^{-1} dataset collected in 2015.

Process	$\tau_\mu\tau_h$		$\tau_e\tau_h$		$\tau_h\tau_h$
	nonres.	res.	nonres.	res.	
$t\bar{t}$	45.5 ± 5.7	203.3 ± 25.4	25.0 ± 3.2	106.8 ± 13.6	2.2 ± 0.3
QCD	1.2 ± 1.0	10.0 ± 3.3	5.2 ± 2.3	4.2 ± 1.8	1.4 ± 1.0
Z+jets	5.2 ± 1.7	7.2 ± 1.8	2.1 ± 0.7	2.1 ± 0.5	1.2 ± 0.3
W+jets	0.9 ± 0.2	1.4 ± 0.1	1.1 ± 0.2	0.5 ± 0.1	–
single top	2.0 ± 0.2	5.2 ± 0.5	1.1 ± 0.1	3.2 ± 0.3	–
di-boson	0.4 ± 0.1	0.3 ± 0.1	0.08 ± 0.02	0.11 ± 0.02	0.11 ± 0.02
Tot. exp. bkg.	55.2 ± 6.0	227.4 ± 25.7	34.6 ± 4.0	116.9 ± 13.7	4.9 ± 1.1
$m_X = 300\,\text{GeV}$		4.0		2.0	0.8
$m_X = 600\,\text{GeV}$		14.1		8.2	10.9
$m_X = 900\,\text{GeV}$		18.4		10.5	13.3
$k_\lambda = 1$ $(\times 10^{-2})$	3.2		1.7		2.1
$k_\lambda = 15$ $(\times 10^{-1})$	7.8		4.1		5.3
DATA	59	224	30	110	4

Table A.1 – Observed and expected event yields for the Moriond 2016 $HH \to bb\tau\tau$ search. Values for the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states are separately reported for the resonant and nonresonant searches, that differ for the selections applied, while the same signal region is defined for the $\tau_h\tau_h$ final state. The background values and the errors correspond to the nuisance parameters obtained from a maximum likelihood fit to the observed data under the background-only hypothesis. The expected yields of a few resonant and nonresonant processes are quoted. The former are normalized to $\sigma(pp \to X) \times \mathcal{B}(X \to HH \to bb\tau\tau) = 1$ pb while the latter are normalized to the theoretical cross section.



(a) Nonresonant search (b) Resonant search

Figure A.2 – 95% CL upper limits derived in the Moriond 2016 search for nonresonant production as a function of k_λ (a) and as for resonant production as a function of (b).

A.2.1 Event selection

As in the Moriond 2016 analysis, events in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states are recorded with a single muon and single lepton trigger, respectively, while events in the $\tau_h\tau_h$ final state are recorded with a double-τ_h trigger. Because of the higher instantaneous luminosity experienced in 2016, higher p_T thresholds are applied on muon and electrons and, consequently, they are also increased in the offline selection. Muons in the $\tau_\mu\tau_h$ final state must have $p_T > 23\,\text{GeV}$, and electrons in the $\tau_e\tau_h$ final state must have $p_T > 27\,\text{GeV}$. In both cases, the leptons must satisfy the additional requirements $|\eta| < 2.1$ and $\mathcal{I}^{rel} < 0.1$. In contrast, the same trigger thresholds as in 2015 were maintained for the double-τ_h trigger thanks to the deployment of the upgraded L1 trigger with the τ algorithm described in Chapter 3. The other selections applied on the τ_h candidates selected in three final states are the same as those of the Moriond 2016 search discussed above.

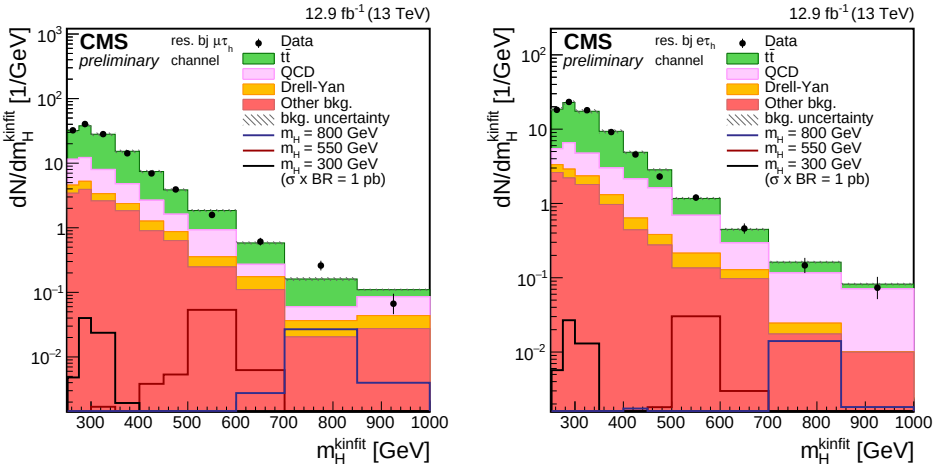
The nonresonant ICHEP 2016 search follows the same event selection strategy as the Moriond 2016 one, applying the methods previously developed to the larger integrated luminosity analysed. The resonant search is instead improved with the introduction of the bb event categorization. As discussed in Section 4.5.2 of Chapter 4, two resolved categories (1b1j and 2b) and one boosted category are defined. The medium WP of the b tagging discriminant is used in the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ final states while the loose one is used in the $\tau_h\tau_h$ final state because of the smaller number of events expected. All events are required to satisfy the invariant mass selection $80 < m_{\tau\tau} < 160\,\text{GeV}$. Events in the two resolved categories are required to satisfy in addition $80 < m_{bb} < 160\,\text{GeV}$ while events in the boosted category must have $90 < m_{AK8} < 160\,\text{GeV}$, where the symbol m_{AK8} denotes the AK8 jet invariant mass estimated with the soft drop jet grooming algorithm.

Results

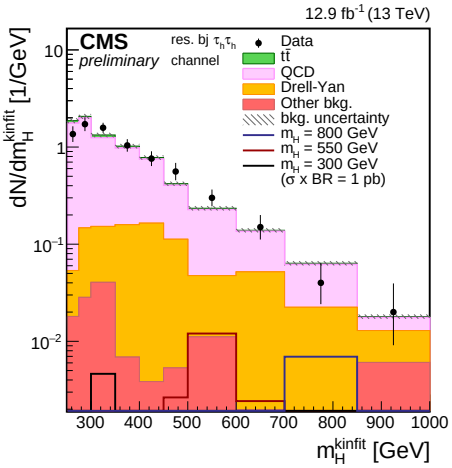
The distributions of the m_{HH} variable, used to search for the presence of a nonresonant signal, is shown in Figure A.3, and the corresponding number of events is summarized in Table A.2. The distributions of the m_{HH}^{KinFit} variable used in the resonant search are shown in Figure A.4 for the different categories of the search, and the expected and observed event yields are shown in Table A.3.

Model independent limits on nonresonant production are set both as a function of k_λ/k_t and for the twelve shape benchmarks, as shown in Figures A.5a and A.5b, respectively. The results are also used to set 95% CL exclusion limits on anomalous k_λ and k_t couplings as illustrated in Figure A.5c. The parametric signal modelling detailed in Section 5.2 of Chapter 5 was not yet finalized when these results were released. Consequently, the event weighting procedure was applied for a smaller number of points and based on the bidimensional m_{HH} and $|\cos\theta^*|$ distribution obtained from generated events for the (k_λ, k_t) combinations denoted with a circle in the figure. For the SM signal, the observed and expected 95% CL upper limits on $\sigma(gg \to HH) \times \mathcal{B}(HH \to bb\tau\tau)$ amount to 508 and 420 fb, respectively. These values correspond to approximately 200 and 170 times the SM prediction.

The 95% CL upper limits on the resonant production cross section as a function of the mass hypothesis m_X are shown in Figure A.6.



(a) $\tau_\mu\tau_\mathrm{h}$ (b) $\tau_\mathrm{e}\tau_\mathrm{h}$

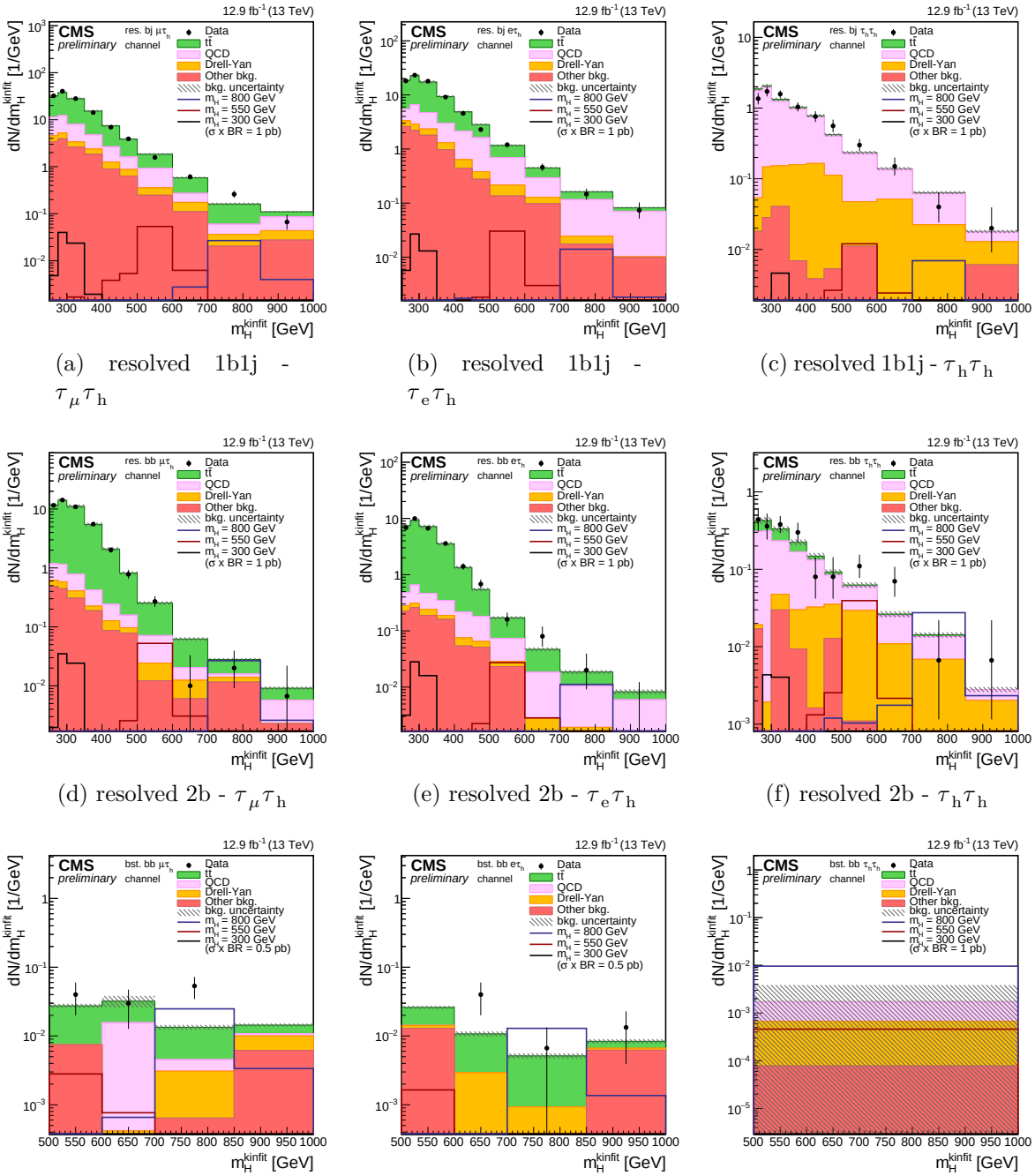


(c) $\tau_\mathrm{h}\tau_\mathrm{h}$

Figure A.3 – m_{HH} distribution for events selected in the nonresonant ICHEP 2016 search in the $\tau_\mu\tau_\mathrm{h}$ (a), $\tau_\mathrm{e}\tau_\mathrm{h}$ (b), and $\tau_\mathrm{h}\tau_\mathrm{h}$ (c) final states.

Process	$\tau_\mu\tau_\mathrm{h}$	$\tau_\mathrm{e}\tau_\mathrm{h}$	$\tau_\mathrm{h}\tau_\mathrm{h}$
$\mathrm{t\bar{t}}$	368.1 ± 37.2	228.5 ± 23.4	15.3 ± 1.7
multijet	52.2 ± 6.5	55.7 ± 4.6	45.7 ± 4.1
Z+jets	31.5 ± 3.0	18.7 ± 1.9	10.3 ± 1.1
W+jets	13.0 ± 1.0	11.0 ± 0.9	1.4 ± 0.1
single top	11.6 ± 1.0	10.7 ± 1.0	1.5 ± 0.2
di-boson	3.1 ± 0.4	1.4 ± 0.2	0.7 ± 0.1
Total expected background	480.0 ± 37.9	326.0 ± 24.4	74.8 ± 4.6
$k_\lambda = 1$	0.24	0.13	0.12
$k_\lambda = 20$	7.8	4.8	4.1
DATA	464	317	84

Table A.2 – Observed and expected event yields for the ICHEP 2016 nonresonant $\mathrm{HH} \to \mathrm{bb}\tau\tau$ search. The background values and the errors correspond to the nuisance parameters obtained from a maximum likelihood fit to the observed data under the background-only hypothesis.



(a) resolved 1b1j - $\tau_\mu\tau_h$

(b) resolved 1b1j - $\tau_e\tau_h$

(c) resolved 1b1j - $\tau_h\tau_h$

(d) resolved 2b - $\tau_\mu\tau_h$

(e) resolved 2b - $\tau_e\tau_h$

(f) resolved 2b - $\tau_h\tau_h$

(g) boosted - $\tau_\mu\tau_h$

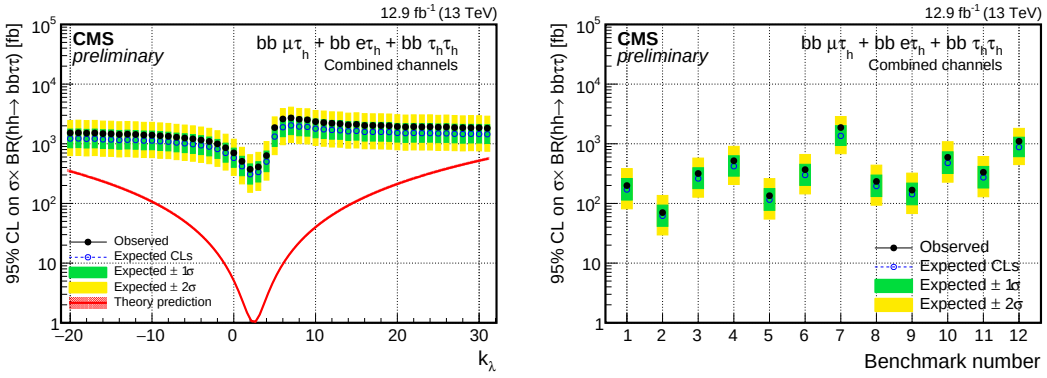
(h) boosted - $\tau_e\tau_h$

(i) boosted - $\tau_h\tau_h$

Figure A.4 – m_{HH}^{KinFit} distribution of events selected in the ICHEP 2016 resonant search. Events are separately shown in the 1b1j (top row), 2b (central row), and boosted category (bottom row), for the $\tau_\mu\tau_h$ final state (left column), $\tau_e\tau_h$ (central column), and $\tau_h\tau_h$ (right column).

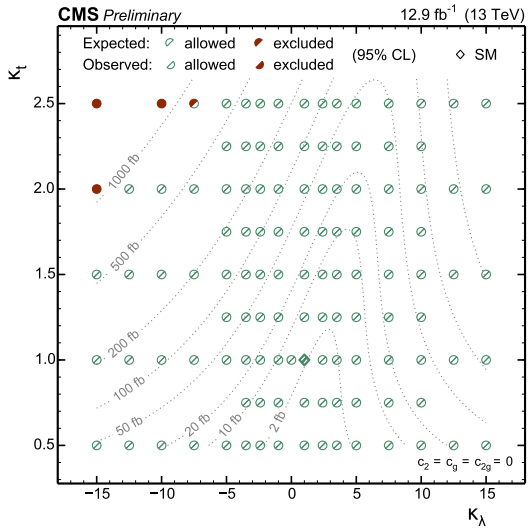
Process	$\tau_\mu\tau_h$	$\tau_e\tau_h$	$\tau_h\tau_h$
resolved 1b1j category			
$t\bar{t}$	3184.1 ± 358.7	1957.4 ± 226.5	15.2 ± 1.9
multijet	880.9 ± 30.6	573.3 ± 24.8	261.8 ± 16.5
Z+jets	180.6 ± 17.6	107.7 ± 10.9	44.3 ± 6.0
W+jets	283.5 ± 26.7	162.4 ± 15.9	2.9 ± 0.3
single top	220.7 ± 22.2	147.3 ± 15.3	1.3 ± 0.2
di-boson	20.6 ± 2.7	12.6 ± 1.7	2.1 ± 0.3
Total expected background	4770.5 ± 362.1	2960.7 ± 229.2	327.6 ± 17.7
$m_X = 300\,\text{GeV}$	34.2	21.5	4.4
$m_X = 600\,\text{GeV}$	90.9	47.7	25.8
$m_X = 900\,\text{GeV}$	63.4	29.6	21.9
DATA	4755	2938	333
resolved 2b category			
$t\bar{t}$	1501.6 ± 169.1	975.7 ± 112.9	15.4 ± 1.9
multijet	72.4 ± 8.7	53.7 ± 7.4	44.2 ± 6.2
Z+jets	18.7 ± 2.4	9.6 ± 1.2	9.8 ± 1.6
W+jets	12.8 ± 1.2	11.4 ± 1.1	1.4 ± 0.1
single top	45.1 ± 4.5	24.5 ± 2.6	1.5 ± 0.2
di-boson	2.2 ± 0.3	1.2 ± 0.2	0.53 ± 0.08
Total expected background	1652.8 ± 169.5	1076.2 ± 113.2	72.8 ± 6.7
$m_X = 300\,\text{GeV}$	30.1	22.1	4.5
$m_X = 600\,\text{GeV}$	84.8	48.0	68.5
$m_X = 900\,\text{GeV}$	56.7	24.1	58.3
DATA	1638	1065	82
boosted category			
$t\bar{t}$	10.7 ± 1.2	6.2 ± 0.6	0.045 ± 0.006
multijet	2.1 ± 0.8	–	0.52 ± 1.06
Z+jets	1.3 ± 0.1	1.3 ± 0.1	0.30 ± 0.04
W+jets	0.9 ± 0.1	1.0 ± 0.1	–
single top	0.44 ± 0.04	1.1 ± 0.1	–
di-boson	0.6 ± 0.1	0.1 ± 0.01	0.08 ± 0.01
Total expected background	16.1 ± 1.5	9.8 ± 0.7	0.94 ± 1.06
$m_X = 300\,\text{GeV}$	–	–	–
$m_X = 600\,\text{GeV}$	19.5	11.1	9.0
$m_X = 900\,\text{GeV}$	148.8	72.9	102.2
DATA	21	11	0

Table A.3 – Observed and expected event yields for the ICHEP 2016 resonant $HH \to bb\tau\tau$ search. The background values and the errors correspond to the nuisance parameters obtained from a maximum likelihood fit to the observed data under the background-only hypothesis. Signal yields are normalized to $\sigma(pp \to X) \to \mathcal{B}(X \to HH) = 1$ pb.



(a) k_λ scan

(b) Shape benchmarks



(c) $(k_\lambda, k_\mathrm{t})$ exclusion

Figure A.5 – 95% CL upper limits as a function of the anomalous trilinear coupling k_λ (a) and for the twelve shape benchmarks (b) obtained from the ICHEP 2016 nonresonant analysis. (b) 95% CL excluded regions of the $(k_\lambda, k_\mathrm{t})$ plane assuming $\mathrm{c_2} = \mathrm{c_g} = \mathrm{c_{2g}} = 0$. Open green semicircles denote points compatible with the current data while red full semicircles denote points excluded with the current data, with the two halves of the circles denoting the expected and observed exclusion as reported in the legend. The diamond shaped marker indicates to the prediction of the SM. The dotted lines indicate trajectories in the plane with equal HH production cross section, and are labelled with the corresponding value of $\sigma(\mathrm{gg} \to \mathrm{HH}) \times \mathcal{B}(\mathrm{HH} \to \mathrm{bb}\tau\tau)$.

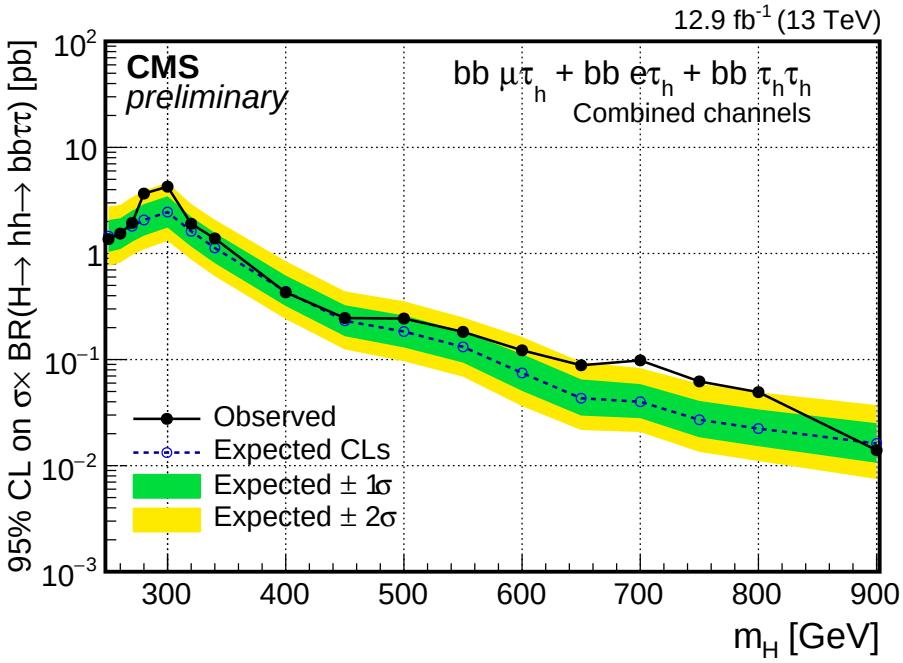


Figure A.6 – 95% CL upper limits on $\sigma(\mathrm{gg} \to \mathrm{X} \to \mathrm{HH}) \times \mathcal{B}(\mathrm{HH} \to \mathrm{bb}\tau\tau)$ as a function of the resonance mass m_{X}. These results correspond to the resonant ICHEP 2016 search.

Bibliography

[1] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[2] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[4] J. Goldstone, "Field theories with « Superconductor » solutions", *Il Nuovo Cimento (1955-1965)* **19** (1961) 154, doi:10.1007/BF02812722.

[5] G. Hooft, "Renormalizable Lagrangians for massive Yang-Mills fields", *Nuclear Physics B* **35** (1971) 167, doi:http://dx.doi.org/10.1016/0550-3213(71)90139-8.

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[8] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[9] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at sqrt(s) = 13 TeV", arXiv:1706.09936.

[10] ATLAS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel with 36.1 fb^{-1} *pp* collision data at the center-of-mass energy of 13 TeV with the ATLAS detector", ATLAS Conference Note ATLAS-CONF-2017-045, CERN, 2017.

[11] CMS Collaboration Collaboration, "Search for Higgs boson production in association with top quarks in multilepton final states at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-17-004, CERN, 2017.

[12] CMS Collaboration, "Measurements of properties of the Higgs boson in the diphoton decay channel with the full 2016 data set", CMS Physics Analysis Summary CMS-PAS-HIG-16-040, CERN, 2017.

[13] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **710** (2012) 26, doi:10.1016/j.physletb.2012.02.064, arXiv:1202.1488.

[14] CMS Collaboration, "Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs", *Phys. Rev. Lett.* **110** (2013) 081803, doi:10.1103/PhysRevLett.110.081803, arXiv:1212.6639.

[15] CMS Collaboration, "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", *Eur. Phys. J. C* **75** (2015) 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.

[16] ATLAS and CMS Collaborations, "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 $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[17] CMS Collaboration, "Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC", CMS Physics Analysis Summary CMS-PAS-HIG-16-043, CERN, 2017.

[18] ATLAS Collaboration, "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector", ATLAS Conferente Note ATLAS-CONF-2017-041, CERN, 2017.

[19] CMS Collaboration, "Evidence for the decay of the Higgs Boson to Bottom Quarks", CMS Physics Analysis Summary CMS-PAS-HIG-16-044, CERN, 2017.

[20] ATLAS Collaboration, "Search for the dimuon decay of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", arXiv:1705.04582.

[21] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[22] G. Degrassi, M. Fedele, and P. P. Giardino, "Constraints on the trilinear Higgs self coupling from precision observables", *JHEP* **04** (2017) 155, doi:10.1007/JHEP04(2017)155, arXiv:1702.01737.

[23] G. Degrassi, P. P. Giardino, F. Maltoni, and D. Pagani, "Probing the Higgs self coupling via single Higgs production at the LHC", *JHEP* **12** (2016) 080, doi:10.1007/JHEP12(2016)080, arXiv:1607.04251.

[24] U. Haisch, "Yukawas and trilinear Higgs terms from loops", in *52nd Rencontres de Moriond on EW Interactions and Unified Theories (Moriond EW 2017) La Thuile, Italy, March 18-25, 2017*. 2017. arXiv:1706.09730.

[25] W. Bizon, M. Gorbahn, U. Haisch, and G. Zanderighi, "Constraints on the trilinear Higgs coupling from vector boson fusion and associated Higgs production at the LHC", *JHEP* **07** (2017) 083, doi:10.1007/JHEP07(2017)083, arXiv:1610.05771.

[26] S. Di Vita et al., "A global view on the Higgs self-coupling", 2017. arXiv:1704.01953.

[27] E. Glover and J. van der Bij, "Higgs boson pair production via gluon fusion", *Nuclear Physics B* **309** (1988) 282, doi:http://dx.doi.org/10.1016/0550-3213(88)90083-1.

[28] J. Baglio et al., "The measurement of the Higgs self-coupling at the LHC: theoretical status", *JHEP* **04** (2013) 151, doi:10.1007/JHEP04(2013)151, arXiv:1212.5581.

[29] R. Frederix et al., "Higgs pair production at the LHC with NLO and parton-shower effects", *Phys. Lett. B* **732** (2014) 142, doi:10.1016/j.physletb.2014.03.026, arXiv:1401.7340.

[30] F. Maltoni, G. Ridolfi, and M. Ubiali, "b-initiated processes at the LHC: a reappraisal", *JHEP* **07** (2012) 022, doi:10.1007/JHEP04(2013)095,10.1007/JHEP07(2012)022, arXiv:1203.6393. [Erratum: JHEP04,095(2013)].

[31] J. Elias-Miró et al., "Higgs mass implications on the stability of the electroweak vacuum", *Phys. Lett. B* **709** (2012) 222, doi:http://dx.doi.org/10.1016/j.physletb.2012.02.013.

[32] F. L. Bezrukov and M. Shaposhnikov, "The Standard Model Higgs boson as the inflaton", *Phys. Lett. B* **659** (2008) 703, doi:10.1016/j.physletb.2007.11.072, arXiv:0710.3755.

[33] F. Bezrukov, "The Higgs field as an inflaton", *Class. Quant. Grav.* **30** (2013) 214001, doi:10.1088/0264-9381/30/21/214001, arXiv:1307.0708.

[34] T. Binoth and J. J. van der Bij, "Influence of strongly coupled, hidden scalars on Higgs signals", *Z. Phys. C* **75** (1997) 17, doi:10.1007/s002880050442, arXiv:hep-ph/9608245.

[35] R. M. Schabinger and J. D. Wells, "Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the CERN Large Hadron Collider", *Phys. Rev. D* **72** (2005) 093007, doi:10.1103/PhysRevD.72.093007, arXiv:hep-ph/0509209.

[36] B. Patt and F. Wilczek, "Higgs-field Portal into Hidden Sectors", 2006. arXiv:hep-ph/0605188.

[37] D. López-Val and T. Robens, "Δr and the W-boson mass in the singlet extension of the standard model", *Phys. Rev. D* **90** (2014) 114018, doi:10.1103/PhysRevD.90.114018, arXiv:1406.1043.

[38] T. Robens and T. Stefaniak, "LHC Benchmark Scenarios for the Real Higgs Singlet Extension of the Standard Model", *Eur. Phys. J. C* **76** (2016) 268, doi:10.1140/epjc/s10052-016-4115-8, arXiv:1601.07880.

[39] S. Dawson and I. M. Lewis, "NLO corrections to double Higgs boson production in the Higgs singlet model", *Phys. Rev. D* **92** (2015) 094023, doi:10.1103/PhysRevD.92.094023, arXiv:1508.05397.

[40] G. C. Branco et al., "Theory and phenomenology of two-Higgs-doublet models", *Phys. Rept.* **516** (2012) 1, doi:10.1016/j.physrep.2012.02.002, arXiv:1106.0034.

[41] E. Bagnaschi et al., "Benchmark scenarios for low $\tan\beta$ in the MSSM", LHC Cross Section Working Group Note LHCHXSWG-2015-002, CERN, 2015.

[42] A. Djouadi et al., "The post-Higgs MSSM scenario: Habemus MSSM?", *Eur. Phys. J. C* **73** (2013) 2650, doi:10.1140/epjc/s10052-013-2650-0, arXiv:1307.5205.

[43] A. Djouadi et al., "Fully covering the MSSM Higgs sector at the LHC", *JHEP* **06** (2015) 168, doi:10.1007/JHEP06(2015)168, arXiv:1502.05653.

[44] S. Heinemeyer, "Benchmark Scenario for low $\tan\beta$ in the MSSM: First preliminary interim recommendation/suggestion". https://twiki.cern.ch/twiki/pub/LHCPhysics/HXSWG3LowTanB/benchmark5-v0.pdf, 2014.

[45] CMS Collaboration, "Summary results of high mass BSM Higgs searches using CMS run-I data", CMS Physics Analysis Summary CMS-PAS-HIG-16-007, CERN, 2016.

[46] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The Hierarchy problem and new dimensions at a millimeter", *Phys. Lett. B* **429** (1998) 263, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.

[47] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, "Searching for the Kaluza-Klein graviton in bulk RS models", *JHEP* **09** (2007) 013, doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.

[48] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, "Warped gravitons at the CERN LHC and beyond", *Phys. Rev. D* **76** (2007) 036006, doi:10.1103/PhysRevD.76.036006, arXiv:hep-ph/0701186.

[49] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[50] A. Oliveira, "Gravity particles from Warped Extra Dimensions, predictions for LHC", 2014. arXiv:1404.0102.

[51] V. Barger and M. Ishida, "Randall-Sundrum Reality at the LHC", *Phys. Lett. B* **709** (2012) 185, doi:10.1016/j.physletb.2012.01.073, arXiv:1110.6452.

[52] A. Efrati and Y. Nir, "What if $\lambda_{hhh} \neq 3m_h^2/v$", 2014. arXiv:1401.0935.

[53] L. Di Luzio, R. Gröber, and M. Spannowsky, "Maxi-sizing the trilinear Higgs self-coupling: how large could it be?", 2017. arXiv:1704.02311.

[54] U. Baur, T. Plehn, and D. L. Rainwater, "Determining the Higgs boson selfcoupling at hadron colliders", *Phys. Rev. D* **67** (2003) 033003, doi:10.1103/PhysRevD.67.033003, arXiv:hep-ph/0211224.

[55] B. W. Lee, C. Quigg, and H. B. Thacker, "Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass", *Phys. Rev. Lett.* **38** (1977) 883, doi:10.1103/PhysRevLett.38.883.

[56] A. Falkowski and R. Rattazzi, "Which EFT?", 2017. To appear.

[57] A. Carvalho et al., "Analytical parametrization and shape classification of anomalous HH production in the EFT approach", 2016. arXiv:1608.06578.

[58] F. Goertz, A. Papaefstathiou, L. L. Yang, and J. Zurita, "Higgs boson pair production in the D=6 extension of the SM", *JHEP* **04** (2015) 167, doi:10.1007/JHEP04(2015)167, arXiv:1410.3471.

[59] A. Carvalho et al., "Higgs Pair Production: Choosing Benchmarks With Cluster Analysis", *JHEP* **04** (2016) 126, doi:10.1007/JHEP04(2016)126, arXiv:1507.02245.

[60] S. Dawson and C. W. Murphy, "Standard Model EFT and Extended Scalar Sectors", arXiv:1704.07851.

[61] J. de Blas, M. Chala, M. Perez-Victoria, and J. Santiago, "Observable Effects of General New Scalar Particles", *JHEP* **04** (2015) 078, doi:10.1007/JHEP04(2015)078, arXiv:1412.8480.

[62] H. Bélusca-Maïto et al., "Higgs EFT for 2HDM and beyond", *Eur. Phys. J. C* **77** (2017) 176, doi:10.1140/epjc/s10052-017-4745-5, arXiv:1611.01112.

[63] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Observable contributions of new exotic quarks to quark mixing", *JHEP* **09** (2000) 011, doi:10.1088/1126-6708/2000/09/011, arXiv:hep-ph/0007316.

[64] F. del Aguila, J. de Blas, and M. Perez-Victoria, "Effects of new leptons in Electroweak Precision Data", *Phys. Rev. D* **78** (2008) 013010, doi:10.1103/PhysRevD.78.013010, arXiv:0803.4008.

[65] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, "The Strongly-Interacting Light Higgs", *JHEP* **06** (2007) 045, doi:10.1088/1126-6708/2007/06/045, arXiv:hep-ph/0703164.

[66] R. Contino et al., "Effective Lagrangian for a light Higgs-like scalar", *JHEP* **07** (2013) 035, doi:10.1007/JHEP07(2013)035, arXiv:1303.3876.

[67] R. Grober and M. Muhlleitner, "Composite Higgs Boson Pair Production at the LHC", *JHEP* **06** (2011) 020, doi:10.1007/JHEP06(2011)020, arXiv:1012.1562.

[68] R. Contino, "The Higgs as a Composite Nambu-Goldstone Boson", in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, p. 235. 2011. arXiv:1005.4269. doi:10.1142/9789814327183_0005.

[69] M. J. Dolan, C. Englert, and M. Spannowsky, "Higgs self-coupling measurements at the LHC", *JHEP* **10** (2012) 112, doi:10.1007/JHEP10(2012)112, arXiv:1206.5001.

[70] A. J. Barr, M. J. Dolan, C. Englert, and M. Spannowsky, "Di-Higgs final states augMT2ed – selecting hh events at the high luminosity LHC", *Phys. Lett. B* **728** (2014) 308, doi:10.1016/j.physletb.2013.12.011, arXiv:1309.6318.

[71] ATLAS Collaboration, "Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector", *Phys. Rev. D* **92** (2015) 092004, doi:10.1103/PhysRevD.92.092004, arXiv:1509.04670.

[72] ATLAS Collaboration, "Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector", *Phys. Rev. Lett.* **114** (2015) 081802, doi:10.1103/PhysRevLett.114.081802, arXiv:1406.5053.

[73] ATLAS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state using pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", ATLAS Conference Note ATLAS-CONF-2016-004, CERN, 2016.

[74] CMS Collaboration, "A search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 8$TeV", 2017. arXiv:1707.00350. Submitted to *Phys. Rev. D*.

[75] CMS Collaboration, "Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV", *Phys. Rev. D* **94** (2016) 052012, doi:10.1103/PhysRevD.94.052012, arXiv:1603.06896.

[76] CMS Collaboration, "Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \to ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV", *Phys. Rev. D* **90** (2014) 112013, doi:10.1103/PhysRevD.90.112013, arXiv:1410.2751.

[77] CMS Collaboration, "Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks", *Eur. Phys. J. C* **76** (2016) 371, doi:10.1140/epjc/s10052-016-4206-6, arXiv:1602.08762.

[78] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to $b\bar{b}$ and $\tau^+\tau^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-EXO-15-008, CERN, 2015.

[79] The LHC Study Group, "LHC: the Large Hadron Collider accelerator project", Technical Report CERN-AC-93-03-LHC, CERN, 1993.

[80] The LHC Study Group, "The Large Hadron Collider: conceptual design", Technical Report CERN-AC-95-05-LHC, CERN, 1995.

[81] C. De Melis, "The CERN accelerator complex. Complexe des accélérateurs du CERN", OPEN-PHO-ACCEL-2016-009, CERN, 2016.

[82] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[83] G. Apollinari et al., "High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report". CERN Yellow Reports: Monographs. CERN, 2015. doi:10.5170/CERN-2015-005.

[84] "The HiLumi LHC Design Study". https://hilumilhcds.web.cern.ch/, 2017.

[85] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[86] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade", CERN-LHCC-2012-016. CMS-TDR-11, CERN, 2012.

[87] P. Adzic et al., "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *JINST* **2** (2007) P04004.

[88] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016. [JINST8,9009(2013)].

[89] CMS HCAL/ECAL Collaboration, "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J. C* **60** (2009) 359, doi:10.1140/epjc/s10052-009-0959-5,10.1140/epjc/s10052-009-1024-0. [Erratum: Eur. Phys. J.C61,353(2009)].

[90] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC", *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

[91] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", 2017. arXiv:1706.04965. Submitted to *JINST*.

[92] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[93] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[94] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[95] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in CMS", *JINST* **7** (2012) P01001.

[96] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *JINST* **11** (2016) P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.

[97] CMS Collaboration, "Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2", CMS Physics Analysis Summary CMS-PAS-TAU-16-002, CERN, 2016.

[98] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[99] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[100] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[101] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[102] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017.

[103] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", CMS Physics Analysis Summary CMS-PAS-JME-16-004, CERN, 2016.

[104] G. Antchev et al., "First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s}$ =7 TeV", *Europhys. Lett.* **96** (2011) 21002, doi:10.1209/0295-5075/96/21002, arXiv:1110.1395.

[105] CMS Collaboration, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems", CERN-LHCC-2000-038, CERN, 2000.

[106] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CERN-LHCC-2013-011. CMS-TDR-12, CERN, 2013.

[107] L. Cadamuro, "The CMS Level-1 trigger system for LHC Run II", *JINST* **12** (2017) C03021.

[108] A. Zabi et al., "The CMS Level-1 Calorimeter Trigger for the LHC Run II", *JINST* **12** (2017) C01065.

[109] A. Svetek et al., "The Calorimeter Trigger Processor Card: the next generation of high speed algorithmic data processing at CMS", *JINST* **11** (2016) C02011.

[110] Imperial College of London, "MP7". `http://www.hep.ph.ic.ac.uk/mp7/`, 2013.

[111] E. Hazen et al., "The AMC13XG: a new generation clock/timing/DAQ module for CMS MicroTCA", *JINST* **8** (2013) C12036.

[112] Boston University, "AMC13". `http://bucms.bu.edu/twiki/bin/view/BUCMSPublic/HcalDTC`, 2016.

[113] A. Zabi et al., "Triggering on electrons, jets and tau leptons with the CMS upgraded calorimeter trigger for the LHC RUN II", *JINST* **11** (2016) C02008.

[114] D. Acosta et al., "The CMS Modular Track Finder boards, MTF6 and MTF7", *JINST* **8** (2013) C12034.

[115] B. Kreis et al., "Run 2 Upgrades to the CMS Level-1 Calorimeter Trigger", *JINST* **11** (2016) C01051, `doi:10.1088/1748-0221/11/01/C01051`, `arXiv:1511.05855`.

[116] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, `doi:10.1088/1748-0221/12/01/P01020`, `arXiv:1609.02366`.

[117] CMS Collaboration, "Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions", *JHEP* **10** (2014) 160, `doi:10.1007/JHEP10(2014)160`, `arXiv:1408.3316`.

[118] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **05** (2014) 104, `doi:10.1007/JHEP05(2014)104`, `arXiv:1401.5041`.

[119] L. Cadamuro, "The CMS Level-1 Tau algorithm for the LHC Run II", CMS Conference Report CMS-CR-2015-214, CERN, 2015. Proceedings of the European Physical Society Conference on High Energy Physics.

[120] CMS Collaboration, "L1 calorimeter trigger upgrade: tau performance", CMS Detector Performance Note CMS-DP-2015-009, CERN, 2015.

[121] L. Mastrolorenzo, "The CMS Level-1 Tau identification algorithm for the LHC Run II", *Nuclear and Particle Physics Proceedings* **273** (2016) 2518, `doi:http://dx.doi.org/10.1016/j.nuclphysbps.2015.09.444`. 37th International Conference on High Energy Physics (ICHEP).

[122] CMS Collaboration, "Search for non-resonant Higgs boson pair production in the $\mathrm{b\bar{b}}\tau^+\tau^-$ final state", CMS Physics Analysis Summary CMS-PAS-HIG-16-012, CERN, 2016.

[123] CMS Collaboration, "Search for resonant Higgs boson pair production in the $\mathrm{b\bar{b}}\tau^+\tau^-$ final state", CMS Physics Analysis Summary CMS-PAS-HIG-16-013, CERN, 2016.

[124] CMS Collaboration, "Search for non-resonant Higgs boson pair production in the bb$\tau\tau$ final state", CMS Analysis Note AN-2015/315, CERN, 2015. Access restricted to CMS members.

[125] CMS Collaboration, "Search for resonant Higgs boson pair production in the bb$\tau\tau$ final state", CMS Analysis Note AN-2016/006, CERN, 2016. Access restricted to CMS members.

[126] CMS Collaboration, "Search for non-resonant Higgs boson pair production in the bbtautau final state using 2016 data", CMS Physics Analysis Summary CMS-PAS-HIG-16-028, CERN, 2016.

[127] CMS Collaboration, "Search for resonant Higgs boson pair production in the $\mathrm{b\bar{b}}\tau^+\tau^-$ final state using 2016 data", CMS Physics Analysis Summary CMS-PAS-HIG-16-029, CERN, 2016.

[128] CMS Collaboration, "Search for non-resonant Higgs boson pair production in the bb$\tau\tau$ final state with 2016 data", CMS Analysis Note AN-2016/232, CERN, 2016. Access restricted to CMS members.

[129] CMS Collaboration, "Search for resonant Higgs boson pair production in the bb$\tau\tau$ final states with 2016 datasets", CMS Analysis Note AN-2016/186, CERN, 2016. Access restricted to CMS members.

[130] CMS Collaboration, "Search for pair production of Higgs bosons in the two tau leptons and two bottom quarks final state using proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-17-002, CERN, 2017.

[131] CMS Collaboration, "Search for Higgs boson pair production in the bb$\tau\tau$ final state with the CMS detector at the LHC", CMS Analysis Note AN-2017/007, CERN, 2017. Access restricted to CMS members.

[132] L. Cadamuro, "Search and prospects for HH production", CMS Conference Report CMS-CR-2017-126, CERN, 2017. To be published in the proceedings of the 52$^{\text{nd}}$ Rencontres de Moriond.

[133] CMS Collaboration, "Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at sqrt(s) = 13 TeV", 2017. arXiv:1707.02909. Submitted to Phys. Lett. B.

[134] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.

[135] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.", *Ann. Statist.* **29** (2001) 1189, doi:10.1214/aos/1013203451.

[136] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[137] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST* **5** (2010) T03022, doi:10.1088/1748-0221/5/03/T03022, arXiv:0911.4994.

[138] CMS Collaboration, "Search for new physics in same-sign dilepton events in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **76** (2016) 439, doi:10.1140/epjc/s10052-016-4261-z, arXiv:1605.03171.

[139] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop", *JHEP* **2014** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[140] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider", *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.

[141] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, 2016.

[142] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, "Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques", *J. Phys. Conf. Ser.* **513** (2014) 022035, doi:10.1088/1742-6596/513/2/022035.

[143] CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$", *Phys. Lett. B* **755** (2016) 217, doi:10.1016/j.physletb.2016.01.056, arXiv:1510.01181.

[144] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 79, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[145] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[146] J. M. Campbell, R. K. Ellis, P. Nason, and E. Re, "Top-pair production and decay at NLO matched with parton showers", *JHEP* **04** (2015) 114, doi:10.1007/JHEP04(2015)114, arXiv:1412.1828.

[147] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[148] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[149] S. Agostinelli et al., "Geant4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:http://dx.doi.org/10.1016/S0168-9002(03)01368-8.

[150] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[151] A. Carvalho, "Tools for search for non-resonant GF di-Higgs at LHC". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/Higgs/NonResonantHHAtLHC, 2017.

[152] M. Czakon and A. Mitov, "NNLO+NNLL top-quark-pair cross sections". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO, 2015.

[153] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[154] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, CERN, 2016.

[155] CMS Collaboration, "Measurement of particle level differential ttbar cross sections in the dilepton channel at sqrt(s) = 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-007, CERN, 2016.

[156] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Phys. Rev. D* **95** (2017) 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.

[157] CMS Collaboration, "Measurement of the differential cross section for $\mathrm{t\bar{t}}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-011, CERN, 2016.

[158] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J.* **C53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[159] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[160] CMS Collaboration, "Measurements of differential production cross sections for a Z boson in association with jets in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **04** (2017) 022, doi:10.1007/JHEP04(2017)022, arXiv:1611.03844.

[161] N. Kidonakis, "Top Quark Production", 2014. arXiv:1311.0283.

[162] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[163] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.

[164] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[165] C. G. Lester and D. J. Summers, "Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders", *Phys. Lett. B* **463** (1999) 99, doi:10.1016/S0370-2693(99)00945-4, arXiv:hep-ph/9906349.

[166] A. Barr, C. Lester, and P. Stephens, "A variable for measuring masses at hadron colliders when missing energy is expected; m(T2): the truth behind the glamour", *J. Phys. G* **29** (2003) 2343, doi:10.1088/0954-3899/29/10/304, arXiv:hep-ph/0304226.

[167] C. G. Lester and B. Nachman, "Bisection-based asymmetric MT2 computation: a higher precision calculator than existing symmetric methods", *JHEP* **03** (2015) 100, doi:10.1007/JHEP03(2015)100, arXiv:1411.4312.

[168] ATLAS and CMS Collaborations and LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, 2011.

[169] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[170] A. L. Read, "Presentation of search results: The CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[171] R. D. Cousins, "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms". http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf, 2013.

[172] CMS Collaboration, "Search for production of a Higgs boson and a single top quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV", Physics Analysis Summary CMS-PAS-HIG-17-005, CERN, 2017.

[173] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton−proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", ATLAS Conference Note ATLAS-CONF-2016-049, CERN, 2016.

[174] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 13 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-16-002, CERN, 2016.

[175] CMS Collaboration, "Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-B2G-16-026, CERN, 2017.

[176] CMS Collaboration, "Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state with 13 TeV CMS data", CMS Physics Analysis Summary CMS-PAS-HIG-16-026, CERN, 2016.

[177] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the bblnulnu final state in proton-proton collisions at sqrt(s) = 13 TeV", arXiv:1708.04188.

[178] ATLAS Collaboration, "Search for Higgs boson pair production in the final state of $\gamma\gamma WW^*(\to l\nu jj)$ using 13.3 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector", ATLAS Conference Note ATLAS-CONF-2016-071, CERN, 2016.

[179] CMS Collaboration, "Search for Higgs boson pair production in the final state containing two photons and two bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-17-008, CERN, 2017.

[180] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.

[181] ATLAS Collaboration, "Study of the double Higgs production channel $H(\to b\bar{b})H(\to \gamma\gamma)$ with the ATLAS experiment at the HL-LHC", CMS Physics Analysis Summary ATL-PHYS-PUB-2017-001, CERN, 2017.

[182] CMS Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS Detector", Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, CERN, 2015.

[183] CMS Collaboration, "Projected performance of Higgs analyses at the HL-LHC for ECFA 2016", Technical Report CMS-PAS-FTR-16-002, CERN, 2017.

[184] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks", *Phys. Rev. Lett.* **101** (2008) 142001, doi:10.1103/PhysRevLett.101.142001, arXiv:0806.0848.

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Titre : Recherche de la production de paires de bosons de Higgs dans le canal de désintégration $b\bar{b}\tau^+\tau^-$ avec le détecteur CMS auprès du LHC

Mots clefs : Boson de Higgs, Expérience CMS, Collisionneur LHC, Modèle standard et au-delà, Déclenchement

Résumé : Cette thèse présente une recherche pour la production de paires de bosons de Higgs (HH) en utilisant les données de collisions proton-proton à $\sqrt{s} = 13\,\text{TeV}$ enregistrées avec l'expérience CMS auprès du LHC au CERN. Les évènements avec les deux bosons de Higgs se désintégrant en une paire de quarks b et de leptons τ ($\text{HH} \to b\bar{b}\tau^+\tau^-$) sont utilisés pour l'exploration des mécanismes de production résonante et non-résonante. La production de HH donne accès à l'auto-couplage trilińeaire du boson de Higgs et pourrait révéler la présence de physique au-delà du modèle standard.
Un effort important a été consacré au développement d'un algorithme pour la reconstruction des leptons τ se désintégrant en hadrons (τ_h) et un neutrino au sein du déclenchement calorimétrique de premier niveau de l'expérience. Ceci a été amélioré pour faire face à l'augmentation de l'énergie dans le centre de masse et de la luminosité instantanée de collisions au Run II du LHC. L'algorithme se fonde sur une technique avancée de regroupement dynamique de l'énergie et utilise des critères dédiés pour la réduction du bruit de fond. Sa structure, son optimisation, sa mise en place, la vérification de son fonctionnement pour le redémarrage à 13 TeV au LHC, et la mesure de sa performance sont présentés ici. Cet algorithme est un élément essentiel dans la recherche de la production HH.
L'investigation du processus $\text{HH} \to b\bar{b}\tau^+\tau^-$ utilise les trois canaux de désintégration du système $\tau^+\tau^-$ avec au moins un τ_h dans l'état final. La sélection et la catégorisation des évènements sont conçues pour optimiser la sensibilité de la recherche, et des techniques d'analyse multivariée sont mises en place pour distinguer le signal du bruit de fond. Les résultats sont présentés en utilisant une luminosité intégrée de $35.9\,\text{fb}^{-1}$. Ils sont compatibles, compte tenu des incertitudes expérimentales, avec les prédictions du modèle standard pour les bruits de fond. Des limites supérieures à la production résonante et non-résonante de HH sont évaluées et permettent de contraindre l'espace des paramètres du modèle standard supersymétrique minimal et les couplages anormaux du boson de Higgs. Les limites supérieures observées et attendues correspondent respectivement à environ 30 et 25 fois la prédiction du modèle standard, et représentent l'un des résultats les plus sensibles à la production de HH jamais atteints au LHC.
Les perspectives pour l'observation de la production de HH au LHC sont enfin discutées. Les résultats actuels sont extrapolés pour une luminosité intégrée de $3000\,\text{fb}^{-1}$ en considérant différentes hypothèses pour la performance du détecteur et de l'analyse.

Title : Search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel with the CMS detector at the LHC

Keywords : Higgs boson, CMS experiment, LHC collider, Standard model and beyond, Trigger

Abstract : This thesis describes a search for Higgs boson pair (HH) production using proton-proton collision data collected at $\sqrt{s} = 13\,\text{TeV}$ with the CMS experiment at the CERN LHC. Events with one Higgs boson decaying into two b quarks and the other decaying into two τ leptons ($\text{HH} \to b\bar{b}\tau^+\tau^-$) are explored to investigate both resonant and nonresonant production mechanisms. HH production gives access to the Higgs boson trilinear self-coupling and is sensitive to the presence of physics beyond the standard model.
A considerable effort has been devoted to the development of an algorithm for the reconstruction of τ leptons decays to hadrons (τ_h) and a neutrino for the Level-1 calorimeter trigger of the experiment, that has been upgraded to face the increase in the centre-of-mass energy and instantaneous luminosity conditions expected for the LHC Run II operations. The algorithm implements a sophisticated dynamic energy clustering technique and dedicated background rejection criteria. Its structure, optimisation and implementation, its commissioning for the LHC restart at 13 TeV, and the measurement of its performance are presented. The algorithm is an essential element in the search for HH production.
The investigation of the $\text{HH} \to b\bar{b}\tau^+\tau^-$ process explores the three decay modes of the $\tau^+\tau^-$ system with one or two τ_h in the final state. A dedicated event selection and categorisation is developed and optimised to enhance the sensitivity, and multivariate techniques are applied for the first time to these final states to separate the signal from the background. Results are derived using an integrated luminosity of $35.9\,\text{fb}^{-1}$. They are found to be consistent, within uncertainties, with the standard model background predictions. Upper limits are set on resonant and nonresonant HH production and constrain the parameter space of the minimal supersymmetric standard model and anomalous Higgs boson couplings. The observed and expected upper limits are about 30 and 25 times the standard model prediction respectively, corresponding to one of the most stringent limits set so far at the LHC.
Finally, prospects for future measurements of HH production at the LHC are evaluated by extrapolating the current results to an integrated luminosity of $3000\,\text{fb}^{-1}$ under different detector and analysis performance scenarios.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France