Probing the Higgs coupling to the top quark at the LHC in the CMS experiment

Thomas Strebler

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

ÉCOLE DOCTORALE N°576

Particules Hadrons Énergie et Noyau : Instrumentation, Image, Cosmos et Simulation (PHENIICS)

Spécialité de doctorat : Physique des particules

Par

M. Thomas Strebler

Étude du couplage du boson de Higgs au quark top au LHC dans l'expérience CMS

Probing the Higgs coupling to the top quark at the LHC in the CMS experiment

Thèse présentée et soutenue à l'École Polytechnique, le 22 septembre 2017
devant le jury composé de:

Mme Lydia Roos	LPNHE, Paris	Présidente du jury
M. Bruno Mansoulié	CEA, Saclay	Rapporteur
M. Günther Dissertori	ETH, Zürich	Rapporteur
M. Emmanuel Perez	CERN, Meyrin	Examinateur
M. Alexander Tapper	IC, London	Examinateur
M. Pascal Paganini	LLR, Palaiseau	Directeur de thèse

Abstract

This thesis reports a study of the Higgs boson production in association with top quarks and decaying into τ leptons in proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV recorded with the CMS detector at the CERN Large Hadron Collider (LHC).

This work has been carried out in the context of the Run 2 of the LHC, marked by an increase in the center-of-mass energy from 8 to 13 TeV together with an increase in the instantaneous luminosity of the collisions with respect to Run 1. To cope with this new data-taking conditions, CMS had initiated a full upgrade of the Level-1 trigger system achieved by 2016. This new system and in particular the new Level-1 electron and photon algorithm have successfully contributed to a large number of CMS results using Run 2 data. The commissioning of this new system and its performance measured in the first data collected with this new trigger are presented in details.

An original analysis technique based on the Matrix Element Method optimized for the search of the $t\bar{t}H, H \to \tau\tau$ process is also presented, including several generic tools which can be used in a large variety of $H \to \tau\tau$ analyses. Results of the CMS analysis using this method based on 35.9 fb^{-1} collected in 2016 are presented. The background-only hypothesis is disfavored but not yet excluded by this analysis alone.

Résumé

Cette thèse présente une étude de la production associée du boson de Higgs en association avec une paire de quarks tops (processus $t\bar{t}H$), dans des collisions proton-proton (pp) à $\sqrt{s} = 13$ TeV enregistrées avec le détecteur CMS au Grand collisionneur de hadrons (LHC) du CERN. A l'ère où la physique du Higgs entre dans le domaine de la mesure de précision, le mode de production $t\bar{t}H$ est important pour caractériser les propriétés du boson de Higgs et en particulier son couplage au quark top, auquel le processus $t\bar{t}H$ offre un accès direct. Les résultats présentés concernent en particulier les canaux sensibles à une désintégration du boson de Higgs en leptons τ et viennent compléter les analyses $t\bar{t}H$ existantes, sensibles aux autres modes de désintégration du boson de Higgs.

Ce travail s'inscrit dans le contexte du Run 2 du LHC, marqué par une augmentation de l'énergie dans le centre de masse de 8 à 13 TeV accompagnée d'une augmentation de la luminosité instantanée des collisions par rapport au Run 1. Pour faire face à ces nouvelles conditions de prise de données, CMS a entrepris une amélioration complète du système de déclenchement de niveau 1 (L1 trigger) accomplie avant 2016. Le nouveau L1 trigger de CMS se base sur une architecture inédite pour un système de déclenchement, appelée *Time Multiplexed Trigger*, qui permet de traiter en parallèle l'information relative à plusieurs croisements de faisceau. Grâce à cette nouvelle architecture, des algorithmes de reconstruction et d'identification plus complexes et plus performants que ceux utilisés au Run 1 ont pu être déployés pour la prise de données du Run 2.

Dans le cadre de cette amélioration du L1 trigger de CMS, j'ai activement participé à la mise en service du système de déclenchement calorimétrique, et plus particulièrement du nouvel algorithme de sélection des électrons et photons (L1EG). J'ai notamment contribué à l'implémentation firmware de l'algorithme L1EG et à sa validation sur banc de test puis avec les données collectées pendant les dernières semaines du run de physique de 2015. J'ai ensuite finalisé la configuration de l'algorithme EG pour la prise de données en 2016 et j'ai mesuré ses performances avec les premières données collectées en 2016. Le nouveau L1 trigger et en particulier le nouvel algorithme électron et photon de niveau 1 ont ainsi contribué avec succès à un grand nombre de résultats CMS basés sur les données du Run 2, y compris les analyses $t\bar{t}H$.

En parallèle, j'ai participé à l'implémentation de la Méthode des Éléments de Matrice (MEM) optimisées pour les analyses impliquant un boson de Higgs se désintégrant en leptons τ. La MEM se base sur le calcul numérique des sections efficaces différentielles associées aux processus de signal et de bruit de fond pertinents pour l'analyse. Pour ce

faire, une intégral multi-dimensionnelle est calculée à partir de techniques Monte Carlo, en prenant en compte la description théorique des processus physiques, au travers de leurs éléments de matrice, ainsi que les effets de résolution, par le biais de *fonctions de transfert*. Les fonctions de transfert peuvent couvrir à la fois les effets liées à la reconstruction des objets dans CMS, comme dans le cas des jets, mais également les désintégrations de particules produisant des neutrinos, comme dans le cas des leptons τ. J'ai ainsi déterminé l'ensemble des fonctions de transfert applicables à un large ensemble d'analyses génériques à partir de données simulées ou à partir de calculs théoriques. J'ai également effectué tous les développements mathématiques nécessaires à la procédure d'intégration, en vue de réduire le nombre de dimensions d'intégration, afin de pouvoir mettre en place une implémentation de la MEM optimisée en termes de temps de calcul. J'ai enfin participé à la validation de son implémentation dans le cadre d'une analyse ciblant le mode de production VBF du boson de Higgs se désintégrant en paire de leptons τ.

Par la suite, j'ai optimisé l'implémentation de la MEM pour couvrir le canal avec deux leptons de même signe (électrons ou muons) et un τ hadronique ($2\ell ss + 1\tau_h$) de l'analyse $t\bar{t}H$ de CMS. La sélection d'événements avec deux leptons de même signe permet de fortement réduire la contribution de la majorité des bruits de fond du Modèle Standard et donc d'obtenir une pureté relativement élevée dans ce canal, tout en bénéficiant du rapport de branchement hadronique d'un des quarks top, qui permet de conserver un nombre d'événements de signal plus élevé que des canaux avec davantage de leptons. Ce canal offre ainsi la plus grande sensibilité au processus $t\bar{t}H$, $H \to \tau\bar{\tau}$.

Les bruits de fond résiduels principaux sont constitués par la production d'une paire de quarks top en association avec un boson électrofaible ($t\bar{t}V$) ainsi que par des événements $t\bar{t}$ avec des leptons *non-prompt*, produits dans la désintégration semi-leptonique de hadrons B. L'acceptance à ce dernier est réduite grâce à l'utilisation d'un discriminant MVA dédié à la sélection des électrons et des muons. La contamination résiduelle est estimée à l'aide d'une méthode de *fake rate*, directement à partir des données. La présence de plusieurs bruits de fond non négligeables rend cette analyse relativement complexe mais en optimisant la MEM pour discriminer simultanément différents types de bruit de fond, il m'a été possible de définir un discriminant avec des performances très satisfaisantes qui a été par la suite utilisé dans les résultats rendus publics par la collaboration CMS à l'occasion de la conférence de Moriond QCD 2017.

Cette analyse est basée sur les données collectées en 2016 par CMS, correspondant à une luminosité intégrée de 35.9 fb^{-1}. La combinaison des résultats obtenus dans la catégorie $2\ell ss + 1\tau_h$ avec d'autres catégories incluant un τ_h a permis d'établir une limite supérieure observée à 95% de niveau de confiance sur $\mu = \sigma/\sigma_{SM}$ correspondant à 2.0 fois la valeur attendue dans le cadre du Modèle Standard, pour une valeur attendue de 1.1 dans l'hypothèse bruit de fond seul. Un excès d'événements par rapport à l'hypothèse bruit de fond seul à été observée, correspondant à une intensité du signal $\mu_{obs} = 0.72^{+0.62}_{-0.53}$ pour une significance de 1.4σ. L'hypothèse bruit de fond seul est donc défavorisée mais n'est pas encore exclue par cette analyse à elle seule.

Remerciements

Durant mes trois années de thèse passées au LLR et au CERN, j'ai eu l'occasion de rencontrer, d'échanger et de collaborer avec plusieurs personnes qui ont directement ou indirectement contribué à la réussite de cette thèse. Je souhaite ici les en remercier.

Tout d'abord, merci à mes rapporteurs Bruno Mansoullié et Günther Dissertori pour leur lecture attentive de ce manuscrit ainsi qu'à Emmanuel Perez, Alex Tapper et Lydia Roos pour avoir accepté de prendre part à mon jury de thèse.

Un grand merci ensuite à tout le groupe CMS du LLR qui m'a accueilli pendant ces trois années et m'a fait profiter d'une ambiance et d'un environnement de travail exceptionnels durant ma thèse. Je mesure toute la chance que j'ai eu de travailler avec vous et j'espère que nous aurons encore l'occasion de continuer à collaborer dans le futur. Je tiens à remercier particulièrement tous les permanents pour contribuer à la vie du groupe : en premier lieu Yves mais aussi Raphaël, Matt, Philippe, Claude, Amina, Stéphanie, Roberto, Christophe, Florian, Jean-Baptiste et Alex. Je vous souhaite une très bonne continuation. Un grand merci aussi aux chevilles ouvrières du groupe, à savoir tous les doctorants et post-docs du LLR, avec qui j'ai eu l'occasion d'intéragir durant ma thèse. Je souhaite ainsi un très bon courage à Simon, Philipp, Luca, Iuri, Toni, Chiara, Marina, Jonas, Arthur, Giacomo et Olivier pour la suite de vos carrières respectives.

I'd like also to thank all the members of the Calo Layer-2 team. I really appreciated working with you on this challenging project, including the beers at R1 after the (sometimes endless) Level-1 trigger meetings. All the best to Alessandro, Tom, Sudan, Aaron, Jim, Andy, Greg, Thierry, JB, Luca, Olvier, Alex T. and Z. I really hope we will have the chance to work together again in the future, either on this trigger system or a future one.

Le développement d'une analyse basée sur la Méthode des Eléments de Matrice a été un pari ambitieux, que je n'aurais pas pu relever sans le soutien des autres MEMbers du LLR, à savoir Pascal, Florian, Arnaud, Gilles et Luca. Merci à vous tous pour vos contributions essentielles. J'espère sincèrement que les outils que nous avons développés au LLR continueront à être exploités dans le futur.

Merci également aux autres membres du LLR avec qui j'ai eu l'occasion de partager repas et cafés lors de mon séjour au LLR. J'ai grandement apprécié nos échanges qui m'ont permis de bénéficier de vos points de vue respectifs sur divers aspects de la vie du labo, de la vie politique française ou les goûts cinématographiques de chacun. Merci donc à Olivier, Michel, Matthieu, James, Qinhua, Frédéric, Emilie, Marc, Thu et Jean-Claude. Un merci particulier à Thomas pour toutes les activités extra-professionnelles que nous avons pu

partager et une bonne continuation à toi pour ta recherche de physique au sein du Modèle Standard.

Merci aussi à tous les membres du fameux Skype chat "Complaints about a shitty DAS, a mediocre CMSSW, an overly convoluted git, an unreliable CRAB, ball-breaking certificates, a poor ROOT, a relou Vidyo, MickeyMouse designs, crappy eos, and Skype itself". Cette discussion s'est avérée être une des sources d'information les plus efficaces pour ce qui concerne divers aspects de CMS, tout en entretenant une ambiance de travail très conviviale. Merci ainsi à Simon, Yacine, Christophe, JB, Luca C. et M., Nadir et Olivier pour les phrases mémorables que nous y avons échangées.

Un immense merci à ma famille pour leur soutien constant tout au long de ma thèse et pour avoir supporté mon rythme de travail qui a malheureusement parfois eu tendance à déborder sur les week-ends en famille. Merci de m'avoir toujours encouragé depuis mon enfance à poursuivre mon parcours sur la voie de la recherche de la connaissance, dont cette thèse marque une étape importante.

Enfin je souhaiterais adresser un merci particulier à trois personnes, sans qui cette thèse n'aurait pas atteint un tel niveau de réussite. Tout d'abord, merci à toi Olivier pour tout le soutien que tu m'as apporté durant cette thèse. Tu m'as non seulement fait bénéficier de ton expérience en termes d'analyse mais tu as surtout été un excellent partenaire de travail lorsque nous étions tous les deux engagés de front sur le Level-1 trigger. Je suis quasiment sûr que le prochain reboot d'Olive et Tom sera basé sur nos exploits. Quoi qu'il en soit, j'espère sincèrement que nous aurons l'occasion de continuer à collaborer directement car tu fais indéniablement partie des personnes que je considère comme les plus compétentes dans CMS.

Merci ensuite à toi Alex pour m'avoir introduit au monde merveilleux (oui, osons le terme) du trigger. Je me souviens encore quand, quelques mois après le début de ma thèse, tu m'expliquais qu'il ne fallait pas que je considère mon travail sur le trigger comme une simple tâche de service secondaire. J'espère que mes contributions à l'upgrade du Level-1 trigger t'ont montré que c'était loin d'être le cas. En tout cas, un grand merci pour ton soutien constant ainsi que pour les moments de détente appréciables que tu as organisés avec succès lors de mon séjour au CERN. J'espère également que nous continuerons à nous croiser régulièrement, que ça soit pour le travail ou autour d'un verre.

Enfin, un grand merci à Pascal, mon directeur de thèse. Quand j'ai découvert avec toi il y a cinq ans le Modèle Standard, j'étais loin de me douter que les diagrammes de Feynman allaient autant me servir durant les cinq années suivantes. Tu as su déceler en moi le potentiel nécessaire pour mener à bien une analyse ambitieuse dans CMS et je te remercie pour cette confiance. Je te souhaite une bonne continuation au-delà de CMS et j'espère que nous aurons encore souvent l'occasion de nous recroiser pour discuter et approfondir notre compréhension du Modèle Standard (ou du nouveau modèle découvert d'ici là). Je te serai toujours reconnaissant pour tout ce que j'ai appris durant cette thèse à tes côtés. Merci.

Contents

Introduction

There are many examples in physics showing that higher precision revealed new phenomena, inspired new ideas, or confirmed or dethroned well-established theories.

Wolfgang Paul

After the observation of the Higgs boson by the CMS and ATLAS collaborations in 2012 [1, 2], all of the elementary constituents needed for a consistent theory describing particle physics, namely the Standard Model, had finally been observed. This observation represented the most important achievement of LHC Run 1 and enabled to achieve one of its primary goals. After this, the high energy physics community was left without any obvious direction to look towards though. Different strategies have therefore been followed in parallel. First, analyses have been designed to look for additional particles or processes not predicted by the Standard Model, like supersymmetric particles. Other analyses tried to further improve the precision on the measurement of some Standard Model parameters, like the W boson or the top quark masses and their couplings, to be able to check more thoroughly the consistency of the experimental results with the Standard Model prediction. Finally, some of the processes predicted by the Standard Model which had not yet been observed have also been looked for. This is for instance the case of the $t\bar{t}H$ process, offering a direct probe of the Higgs boson coupling to the top quark at the tree level. This process had already been studied during LHC Run 1 and the results from the CMS and ATLAS collaborations [3] were consistently pointing towards a possible $t\bar{t}H$ excess with respect to the Standard Model expectation. The sensitivity of these analyses were however limited by the available statistics collected during Run 1 and no unambiguous observation had been possible at that time.

LHC Run 2 aroused a lot of hope and interest among particle physicists. Indeed, the increase of the center-of-mass energy of the collisions from 8 to 13 TeV together with the increase in instantaneous luminosity, beyond the LHC nominal design, has induced a large boost in sensitivity in a wide variety of analyses, sensitive to physics both within and beyond the Standard Model. However this also required significant upgrades of the detector and of the analysis techniques in order to be able to deal with the higher collision rate and the more intense pile-up conditions. One of these changes concerned the Level-1 trigger system of CMS which has been fully upgraded by 2016. Completely changing this

critical system for data-taking has been a major challenge which has been successfully met by the CMS collaboration. In this thesis, I will focus on the monitoring of this new system, and specifically on the Level 1 electron and photon algorithm, in which I played a major role.

For the $t\bar{t}H$ analyses, LHC Run 2 has also represented a major step forward thanks to the sizable increase in the cross section of this process. As the data-taking conditions have been exceptionally good in 2016, it became clear that this process would be very likely observed by the end of Run 2. $t\bar{t}H$ analyses are usually quite challenging though because of the complex final states they target. Taking into account the decays of the top quarks and of the Higgs boson, this process can indeed yield between six and ten reconstructed objects at leading order, often with no obvious variable to characterize this signal. The complexity of those analyses make them ideal use cases for multivariate discriminants, combining information related to all of the reconstructed particles. Such discriminants can be built using the so-called Matrix Element Method (MEM), which combines the theoretical information on a given physics process with the experimental information related to the detector resolution. I have been personally involved in the development of this method in the $H \to \tau\tau$ group of the Laboratoire Leprince-Ringuet, focusing first on its implementation for a VBF $H \to \tau\tau$ analysis. This gave the opportunity to develop a large variety of tools which have then been reused for the development of a MEM discriminant optimized for a $t\bar{t}H, H \to \tau\tau$ analysis.

This thesis is organized as follows. Chapter 1 gives some theoretical motivations to study the $t\bar{t}H, H \to \tau\tau$ process, briefly introducing the Standard Model and the Higgs boson phenomenology. The LHC and the CMS experiment are presented in chapter 2. Both the experimental apparatus and the basic reconstruction techniques are described. Chapter 3 focuses on the upgrade of the Level-1 electron and photon (L1EG) trigger and its commissioning. The L1EG algorithm itself is described and its performance achieved during 2016 data-taking are presented. Finally, chapter 4 offers a general introduction to the Matrix Element Method and its application to $H \to \tau\tau$ analyses, focusing in details on the various parts of its implementation I contributed to develop. Finally, chapter 5 presents its application to the search for $t\bar{t}H$ in final states with a τ lepton, including the corresponding public results released by the CMS collaboration.

Chapter 1

Theoretical motivations

Ce que l'on conçoit bien s'énonce clairement,
Et les mots pour le dire arrivent aisément.

Nicolas Boileau

The observation of a new particle at LHC, compatible with the so-called Higgs boson, and the measurement of its invariant mass enabled to determine one of the last missing parameters of the Standard Model (SM). This theory is currently the one which best describes the elementary particles and their interactions. In particular, the theoretical value of the coupling between the Higgs boson and the top quark, the two most massive SM particles, can be predicted from the measured value of the top quark mass. A measurement of that coupling at LHC would therefore provide a direct test of the SM prediction. The following chapter is dedicated to a brief overview of the SM, based on quantum field theory and gauge invariance, and to its consequences on the phenomenology of the Higgs boson and of the top quark.

1.1 Overview of the Standard Model

1.1.1 Particle content

The elementary particles described by the Standard Model (SM) are presented in Fig. 1.1. They can be divided into two categories: the fermions and the bosons. The bosons are particles with integer spin and are associated to interactions which will be introduced in the next sections. The fermions are spin-1/2 particles described with Dirac spinor fields ψ. In the case of a non-interacting fermion of mass m, those spinor fields are solutions of the Dirac equation

$$(i\partial\!\!\!/ - m)\psi = 0 \tag{1.1}$$

where natural units have been used ($\hbar = c = 1$). The operator $\partial\!\!\!/$ is defined as $\gamma^\mu \partial_\mu$, where γ^μ are matrices satisfying $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ and $\eta^{\mu\nu}$ is the Minkowski metric. The

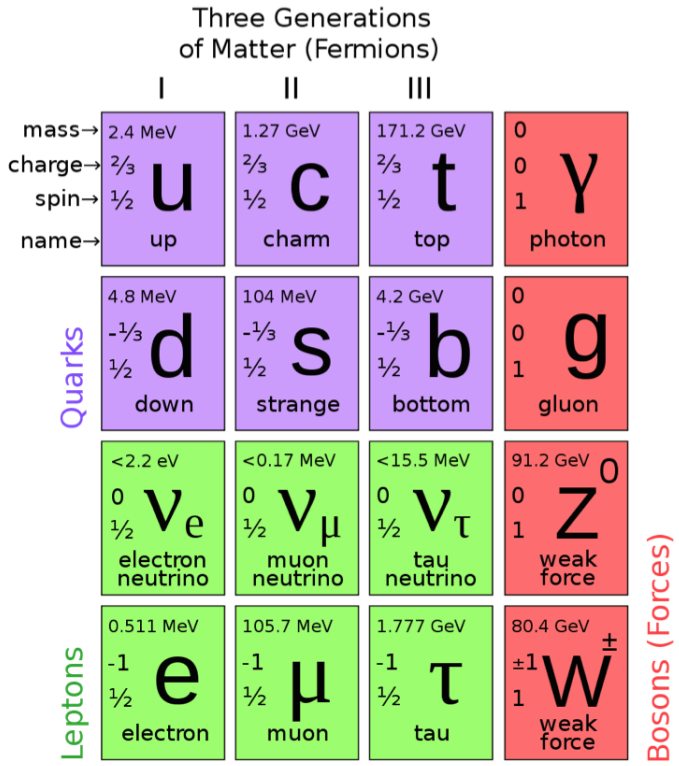

Figure 1.1: Particles described by the Standard Model [4]

Lagrangian associated to that equation is given by

$$\mathcal{L}_{f\,free} = \overline{\psi}(i\partial\!\!\!/ - m)\psi \tag{1.2}$$

with $\overline{\psi} = \psi^{\dagger}\gamma^{0}$.

The masses of the fermions cover a wide range from less than 1 eV for the neutrinos, up to 173 GeV for the top quark. The fermions are divided into three families, each composed of two leptons and of two quarks. Unlike the leptons, the quarks are never observed as free states but only in composite particles called hadrons. Being the most massive, the fermions of the 2^{nd} and the 3^{rd} families decay into fermions of the 1^{st} family, which represent therefore the ordinary matter. For every fermion, there is an associated anti-fermion of same mass, with opposite quantum numbers. The latest fermions to be observed were the top quark in 1995 [5, 6] and the τ neutrino in 2000 [7].

1.1.2 Quantum electrodynamics

The electromagnetic interaction was the first interaction to be described with a quantum field theory, called quantum electrodynamics (QED). This theory was developed by Tomonaga [8], Schwinger [9], Feynman [10] and Dyson [11] and describes the interaction of elec-

trically charged fermions with photons. The photons are described by the electromagnetic potential A^μ, used to define the electromagnetic field strength tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. In absence of fermions, the photon field obeys the Maxwell's equations which are associated to the following Lagrangian

$$\mathcal{L}_{Max} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} \tag{1.3}$$

The interaction of a fermion of electric charge q with the photons is described via the modified Lagrangian

$$\mathcal{L}_{f\,QED} = \overline{\psi}(i\not{D} - m)\psi + h.c. \tag{1.4}$$

with $\not{D} = \gamma^\mu(\partial_\mu + iqA_\mu)$ and $h.c.$ the hermitian conjugate of the first term. The total Lagrangian is then given by

$$\mathcal{L}_{QED} = \mathcal{L}_{f\,QED} + \mathcal{L}_{Max} = \overline{\psi}(i\not{\partial} - m)\psi - q\overline{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \tag{1.5}$$

One of the basic principles of QED is gauge invariance: the total Lagrangian is invariant under local $U(1)_{em}$ transformations. The action of an element $e^{-iq\chi}$ of $U(1)_{em}$ on the fields being

$$A^\mu \to A^\mu + \partial^\mu \chi \,,\; \psi \to e^{-iq\chi}\psi \tag{1.6}$$

the two terms $\mathcal{L}_{f\,QED}$ and $\mathcal{L}_{Max}$ are indeed invariant, thanks to the definition of D^μ. From Noether's theorem, this leads to a conserved quantity, which is the electric charge. Note that requiring gauge invariance strongly constrains the form of the Lagrangian. The choice of a particular gauge group ($U(1)_{em}$ in the case of QED) entirely determines the Lagrangian of the free gauge field A^μ and the choice of a particular representation of the group for the fermion fields determines the term involving the fermion fields. For instance, a mass term for the photon of the form $M_\gamma A^\mu A_\mu$ is forbidden if gauge invariance is required. In the case of an abelian group, like $U(1)_{em}$, the Lagrangian is a bit simpler than in the general case, which will be addressed in the following sections. In terms of Feynman diagram, there is a single vertex involved (see Fig. 1.2).

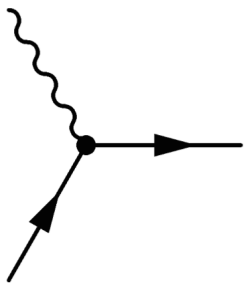

Figure 1.2: Basic vertex in QED

Taking into account the 1-loop correction to this diagram, the 1-loop β-function can be deduced

$$\beta(\alpha) = \frac{\partial \alpha}{\partial \log \mu} = \frac{2\alpha^2}{3\pi} \tag{1.7}$$

with $\alpha = q^2/4\pi$ the coupling constant and μ the energy scale. This implies that the coupling constant α increases with the energy scale μ. This phenomenon is understood as a charge screening due to virtual particle-antiparticle pairs, which tend to reduce the effective electric charge of a particle at low energy.

1.1.3 Quantum chromodynamics

Quantum chromodynamics (QCD) is the theory describing the strong interaction. The only fermions sensitive to the strong interaction are the quarks, which were proposed by Gell-Mann [12] and Zweig [13] in 1964 in order to provide a classification of hadrons. They were first observed in deep inelastic scattering experiments at the Stanford Linear Accelerator Center (SLAC) in 1968 [14, 15]. Unlike QED, QCD is a gauge theory related to a non-abelian group, $SU(3)$. The associated massless spin-1 gauge bosons are called the gluons and are described by eight independent fields G^a_μ. These fields are associated to the generators of the $SU(3)_C$ group, which can be chosen as $\lambda_i/2$, the λ's being the 8 Gell-Mann matrices

$$\begin{gathered}
\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\
\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{-2}{\sqrt{3}} \end{pmatrix}
\end{gathered} \tag{1.8}$$

The gluon field strength tensor $G^a_{\mu\nu}$ has to include an additional term with respect to QED in order to be gauge invariant under the action of $SU(3)_C$. It is defined as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu \tag{1.9}$$

with g_s the coupling constant of QCD and f_{abc} the structure constants of $SU(3)_C$ defined by

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = i f_{abc} \frac{\lambda_c}{2} \tag{1.10}$$

The quarks are associated to spinor fields in the fundamental representation of $SU(3)_C$. The free field Lagrangian for a single quark flavor q_f can therefore be written as

$$\mathcal{L}_{q\,free} = \overline{q}_f (i\gamma^\mu \partial_\mu - m) q_f = \sum_{i=1}^{3} \overline{\psi}_i (i\gamma^\mu \partial_\mu - m)\psi_i \text{ with } q_f = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} \tag{1.11}$$

The interactions with the gluon fields are simply taken into account by introducing the covariant derivative

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a \tag{1.12}$$

The total Lagrangian for the QCD sensitive fields is therefore given by

$$\mathcal{L}_{QCD} = \sum_f \overline{q}_f (i\gamma^\mu \partial_\mu - m) q_f - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a - g_s \overline{q}_f \gamma^\mu \frac{\lambda_a}{2} q_f G_\mu^a \tag{1.13}$$

This Lagrangian is by construction invariant under a local $SU(3)_C$ infinitesimal transformation parameterized by

$$U = e^{-ig_s \alpha^a \frac{\lambda_a}{2}} \approx 1 - ig_s \alpha^a \frac{\lambda_a}{2} \tag{1.14}$$

$$G_\mu^a \to G_\mu^a + g_s \alpha^b f_{abc} G_\mu^c + \partial_\mu \alpha^a \frac{\lambda_a}{2} \ , \ q_f \to U q_f \tag{1.15}$$

Note that it is also possible to include a CP-violating term in the QCD Lagrangian but this is not supported by any experimental evidence so far. The $SU(3)_C$ symmetry is associated to three color charges, usually referred to as red, blue and green. Each quark degree of freedom ψ_i carries one color whereas each gluon field G_μ^a carries a color and an anti-color. This is another difference with QED, since the photon is electrically neutral, while the gluon fields are color charged.

The vertices encoded in the Lagrangian are represented in Fig. 1.3. As for QED, there is a coupling of the quark with the gluon field but the non-abelian structure of $SU(3)_C$ also allows a gauge self-interaction. Because of that gauge self-interaction, the 1-loop β-function associated to the coupling constant $\alpha_s = g_s^2/4\pi$ is different from the QED one and is given by

$$\beta(\alpha_s) = -\frac{7\alpha_s^2}{2\pi} \tag{1.16}$$

The β-function being negative, this means that the coupling constant decreases with the energy scale, which leads to what is called asymptotic freedom, a property discovered by Politzer [16], Gross and Wilczek [17] in 1973. The low-energy limit of QCD is governed by another property called confinement. Because the interaction between two QCD-sensitive particles increases as their separation increases (which corresponds to a decreasing energy scale), the quarks and the gluons are expected never to be observed as free states but always to be bound into hadrons. This hypothesis can be accounted for by requiring that all the observable particles must be invariant under $SU(3)_C$. Consequently, when a quark or a gluon is produced, it undergoes a process called hadronization, during which quark-antiquark pairs are produced to create hadrons, which results in a bunch of collimated particles called a jet. Those objects will be described in more details in section 2.3.4.

1.1.4 Electroweak interaction

In 1934, Fermi proposed his theory to describe the β decays $n \to pe^-\overline{\nu}$ [18]. This required to introduce a new interaction called the weak interaction. It was first assumed to be a

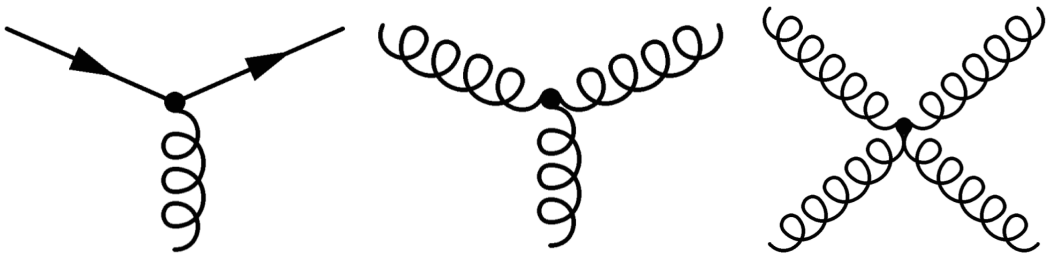

Figure 1.3: Basic vertices in QCD

point-like interaction but this theory suffers from a loss of unitarity at high energy, since it predicts interaction probability arbitrarily large as the energy increases.

Moreover, Yang and Lee suggested in 1956 [19] that parity was not conserved by weak interactions, based on studies about the decays of particles nowadays known as kaons. This was confirmed experimentally by Wu in 1957 [20] thanks to the study of nuclear β decays. This is explained by the fact that, unlike QED or QCD, the weak interaction acts differently on particles with different chiralities, a Lorentz invariant quantity which corresponds to the eigenvalue of the operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The eigenvectors of that operator associated to the eigenvalue -1 (respectively +1) are said to have a left-handed (right-handed) chirality. The properties of the γ^5 matrices enable to define the left-handed and right-handed projectors

$$P_L = \frac{1}{2}(1-\gamma^5) \text{ and } P_R = \frac{1}{2}(1+\gamma^5) \tag{1.17}$$

Each fermion field ψ can then be decomposed as the sum of a left-handed chiral field and a right-handed chiral field $\psi = P_L\psi + P_R\psi$.

Helicity corresponds to the projection of the spin along the momentum direction. Particles with a helicity $h = -1$ (respectively $h = +1$) are said to have a left-handed (right-handed) helicity. This quantity is not Lorentz invariant for a massive particle but is identical to the chirality for massless particles (or in the ultra-relativistic limit). Helicity is easier to determine experimentally and the chiral nature of the weak interaction has therefore been first observed through different results for particles with a left- and right-handed helicity.

In the 1960's, Glashow [21], Salam, Ward [22] and Weinberg [23] showed, within their electroweak theory, that the weak interaction was consistent with the existence of spin-1 massive charged $W^\pm$ bosons. The fields $W^\pm_\mu$ are coupled only with the left components of the fermion fields, which means that only the left-handed chiral particles and the right-handed chiral antiparticles can interact with the W bosons. The W bosons were discovered at CERN in 1983 in the UA1 and UA2 experiments [24, 25] and their mass is currently estimated at [26]

$$M_W = 80.385 \pm 0.015 \,\text{GeV} \tag{1.18}$$

Moreover, the weak eigenstates of the fermions interacting with the W bosons are

different from the mass eigenstates of the fermions. In the case of the quarks, that mixing is described by the CKM (Cabibbo-Kobayashi-Maskawa) matrix and in the case of the leptons, by the PMNS (Pontecorvo-Maki-Nakagawa-Sakata) matrix. In addition to the W boson, the electroweak theory also predicts the existence of neutral currents, discovered at CERN in the Gargamelle experiment in 1973 [27, 28] and associated to a spin-1 massive neutral Z boson. The Z boson was discovered a few months after the W bosons in the UA1 and UA2 experiments [29, 30] and its mass is currently estimated at [26]

$$M_Z = 91.1876 \pm 0.0021 \text{ GeV} \tag{1.19}$$

However, as explained in section 1.1.2, a massive boson is not compatible with a gauge interaction and would make the theory not renormalizable. To take into account the chiral nature of the weak interaction, Glashow, Salam and Ward proposed to introduce a new gauge group $SU(2)_L$ which acts only on the left components of the fermion fields. This leads to the following representations for the fermions[1]

$$\begin{gathered} \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \\ u_R\,,\ c_R\,,\ t_R\,,\ \nu_{eR}\,,\ \nu_{\mu R}\,,\ \nu_{\tau R}\,,\ d_R\,,\ s_R\,,\ b_R\,,\ e_R\,,\ \mu_R\,,\ \tau_R \end{gathered} \tag{1.20}$$

For simplicity, we will only consider here the doublet of fields u and d, since the structure of the Lagrangian is similar for all of them. They are denoted by

$$L = \begin{pmatrix} u_L \\ d_L \end{pmatrix},\ u_R\,,\ d_R \tag{1.21}$$

Since the left-handed and right-handed fields transform differently under an $SU(2)_L$ transformation and since the mass term of a fermion can be written as

$$m\overline{\psi}\psi = m\overline{\psi_R}\psi_L + \overline{\psi_L}\psi_R \tag{1.22}$$

the only way to have a Lagrangian invariant under $SU(2)_L$ is to require that the fermions are massless. In the next section, the mechanism of electroweak symmetry breaking, which enables to generate the mass terms required to account for the experimental mass measurements, will be described in more details. The mass term in the Lagrangian is therefore set to zero for now.

There are three gauge bosons described by the fields W^i_μ and associated to the generators of the $SU(2)_L$ group, which can be chosen as $\sigma_i/2$, the σ's being the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.23}$$

[1] In the original Standard Model formulation, no right-handed neutrino fields were present. They are added here to account for the non-zero neutrino mass, assuming Dirac neutrinos.

Similarly to QCD, the field strength tensor $W^i_{\mu\nu}$ is defined as

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g_w \epsilon_{ijk} W^j_\mu W^k_\nu \tag{1.24}$$

with g_w the coupling constant of $SU(2)_L$ and ϵ_{ijk} the structure constants of $SU(2)_L$ defined by

$$\left[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = i\epsilon_{ijk}\frac{\sigma_k}{2} \tag{1.25}$$

Defining $D_\mu = \partial_\mu + ig_w \frac{\sigma_i}{2} W^i_\mu$, one gets for the associated Lagrangian

$$\mathcal{L}_{SU(2)} = -\frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} + \overline{u_R} i\gamma^\mu \partial_\mu u_R + \overline{d_R} i\gamma^\mu \partial_\mu d_R + \overline{L} i\gamma^\mu D_\mu L \tag{1.26}$$

The conserved charge associated to the $SU(2)_L$ group is called the weak isospin and has three components T_i. The right fields u_R and d_R have a third isospin component $T_3 = 0$ as they are singlets under $SU(2)_L$ whereas the u_L and d_L fields have respectively $T_3 = +1/2$ and $T_3 = -1/2$ since they form a doublet under $SU(2)_L$. The physical $W^\pm$ bosons are associated to a linear combination of the gauge fields W^1_μ and W^2_μ

$$\begin{aligned} W^+_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu - iW^2_\mu) \\ W^-_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu) \end{aligned} \tag{1.27}$$

In order to explain the different couplings of the Z boson with respect to the W bosons, the $SU(2)_L$ group is not enough and an additional $U(1)_Y$ gauge group has to be introduced, even though the original motivation to introduce the additional $U(1)_Y$ gauge group was actually to try to unify the weak and the electromagnetic interaction. The gauge field B_μ associated to that group is described by a Lagrangian very similar to the QED Lagrangian, except that it does not involve the electric charge but another kind of charge, called the weak hypercharge Y, associated to a coupling constant g_Y

$$\mathcal{L}_{U(1)} = \sum_{u_R, d_R, L} \overline{\psi}(i\partial\!\!\!/ - Y g_Y \gamma^\mu B_\mu)\psi - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{1.28}$$

Combining the two Lagrangian, it is then possible to mix the electrically neutral B_μ and W^3_μ to obtain the electromagnetic field A_μ and the field Z_μ associated to the Z boson. This is done via a rotation of angle θ_W

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{1.29}$$

In order to get the correct electromagnetic couplings of the fermion fields with the photon field A_μ, the weak hypercharge is given by the Gell-Mann-Nishijima formula

$$Y = 2(Q - T_3) \tag{1.30}$$

and the angle θ_W is related to the coupling constants e, g_w and g_Y via

$$e = g_w \cos\theta_W = g_Y \sin\theta_W \tag{1.31}$$

At this stage, the boson fields $W_\mu^\pm$ and Z_μ are all massless and an additional field has to be introduced in order to account for their mass. This is done via the Higgs mechanism, which will be described in the next section. The development of the full electroweak Lagrangian enables to determine the interactions of the electroweak bosons and the photon with the fermions and with each other. The full development can be found for instance in [31].

1.1.5 Electroweak symmetry breaking and Higgs field

As a gauge interaction, the electroweak interaction introduced in the previous section is naturally associated with massless bosons, which are not compatible with the experimental data. A possible solution to that problem was proposed by Higgs [32], Brout, Englert [33] and Guralnik, Hagen and Kibble [34] in 1964. This solution, nowadays referred to as the Higgs mechanism, involves a doublet of two complex scalar fields, known as the Higgs field

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 + i\phi_4) \end{pmatrix} \tag{1.32}$$

where the ϕ_i's are real scalar fields. This scalar field doublet is associated to an hypercharge $Y = 1$ and is a doublet for the weak isospin. The Lagrangian used to describe the gauge interaction of that field is the following

$$\mathcal{L}_\phi^{EW} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \tag{1.33}$$

with

$$D_\mu = \partial_\mu + ig_w W_\mu^i \frac{\sigma_i}{2} + ig_Y B_\mu \frac{Y}{2} \tag{1.34}$$

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2\,,\ (\mu^2 < 0, \lambda > 0) \tag{1.35}$$

Because $\mu^2 < 0$, the ground state of the scalar field, which minimizes the potential $V(\phi)$ does not correspond to $\phi = 0$ but has to verify

$$\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.36}$$

This condition is fulfilled by an infinity of possible groundstates. Therefore, although the original Lagrangian L_ϕ^{EW} is symmetric under $SU(2)_L \times U(1)_Y$, the choice of a particular groundstate breaks that symmetry (see Fig. 1.4). This groundstate can be conveniently chosen as

$$\phi_{groundstate} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tag{1.37}$$

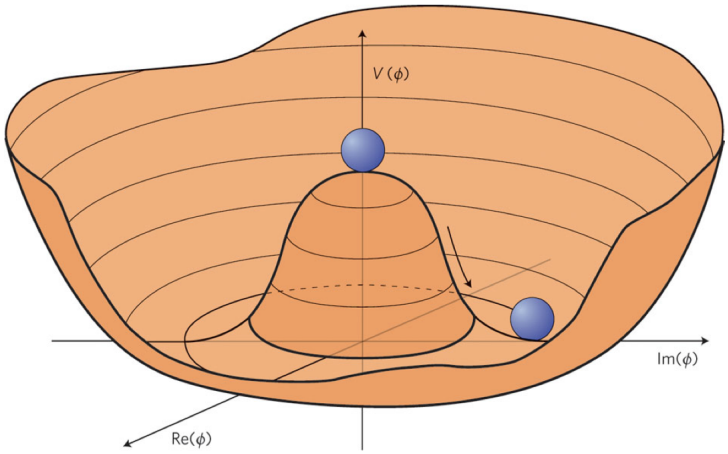

Figure 1.4: Potential of a complex scalar field $V(\phi)$ for the case $\mu^2 < 0$ in the simplified case of a $U(1)$ symmetry [35]

If the symmetries associated to the generators $\frac{\sigma_i}{2}$ and $\frac{Y}{2}$ are indeed broken, there is one remaining symmetry given by $Q = \frac{\sigma_3}{2} + \frac{Y}{2}$. This generator is nothing but the electric charge and the associated symmetry group is $U(1)_{em}$. A convenient parametrization of the Higgs field around its minimum is given by

$$\phi(x) = e^{i\theta^a(x)\sigma_a} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + h(x)) \end{pmatrix} \tag{1.38}$$

The fields θ^a are associated to massless scalar bosons, called Goldstone bosons. They are associated to the broken generators of the original $SU(2)_L \times U(1)_Y$ symmetry. These fields can be absorbed by the choice of a particular gauge, called the unitary gauge.

Combining the kinetic term of the gauge bosons of $SU(2)_L \times U(1)_Y$ with the Lagrangian of the Higgs field, we are left (after some rearrangements) with

$$\begin{aligned} \mathcal{L}^{EW}_{gauge\ free} + \mathcal{L}^{EW}_{\phi} = \quad & \tfrac{1}{2}\partial_\mu h \partial^\mu h - |\mu^2| h^2 \\ & -\tfrac{1}{4}(W^-_{\mu\nu})^\dagger W^{-\mu\nu} + \tfrac{1}{2}(\tfrac{g_w v}{2})^2 (W^-_\mu)^\dagger W^{-\mu} \\ & -\tfrac{1}{4}(W^+_{\mu\nu})^\dagger W^{+\mu\nu} + \tfrac{1}{2}(\tfrac{g_w v}{2})^2 (W^+_\mu)^\dagger W^{+\mu} \\ & -\tfrac{1}{4} Z_{\mu\nu} Z^{\mu\nu} + \tfrac{1}{2}(\tfrac{g_w v}{2\cos\theta_W})^2 Z_\mu Z^\mu \\ & -\tfrac{1}{4} A_{\mu\nu} A^{\mu\nu} \\ & +\tfrac{g_w^2 v}{2} h W^-_\mu W^{+\mu} + \tfrac{g_w^2}{4} h^2 W^-_\mu W^{+\mu} + \tfrac{g_w^2 v}{4\cos^2\theta_W} h Z_\mu Z^\mu + \tfrac{g_w^2}{8\cos^2\theta_W} h^2 Z_\mu Z^\mu \\ & +\tfrac{\mu^2}{v} h^3 + \tfrac{\mu^2}{4v^2} h^4 \end{aligned} \tag{1.39}$$

with $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$ for each of the spin-1 field. The first line of that Lagrangian describes a spin-0 field h with a mass

$$m_H = \sqrt{2}|\mu| = \sqrt{2\lambda} v \tag{1.40}$$

The associated particle is the so-called Higgs boson, observed by the ATLAS and the CMS collaborations in 2012 [1, 2] and its mass is currently measuread at [36]

$$m_H = 125.09 \pm 0.24\,\text{GeV} \tag{1.41}$$

The second, third and fourth lines describe spin-1 fields $W_\mu^\pm$ and Z_μ, associated to the W and Z bosons, with masses

$$m_{W^+} = \frac{g_w v}{2} = m_{W^-} \text{ and } m_Z = \frac{g_w v}{2\cos\theta_W} = \frac{m_W}{\cos\theta_W} \tag{1.42}$$

The fifth line describes a massless spin-1 field A_μ associated with the photon. The sixth line describes the interaction of the Higgs boson with the massive spin-1 fields and the last line describes the self-interactions of the Higgs boson field.

The Higgs mechanism only accounts for the mass of the W and Z bosons. In order to acount for the measured mass of the fermions, a gauge invariant term corresponding to Yukawa couplings of the fermion fields with the Higgs field can be introduced in the Lagrangian. For the quarks u and d, with the notations introduced in the previous section, such a term can be written as

$$\mathcal{L}_{Yuk\,d} = -g_d \overline{L}\phi d_R - g_d \overline{d_R}\phi^\dagger L \tag{1.43}$$

The gauge invariance is ensured since the hypercharge of the Higgs field and of the fermionic field satisfy the relation $y_\phi = y_L - y_{d_R}$ (this is true for every isospin doublet).

After the spontaneous symmetry breaking, $\mathcal{L}_{Yuk\,d}$ can be rewritten as

$$\mathcal{L}_{Yuk\,d} = -\frac{g_d}{\sqrt{2}}\overline{d}d(v+h) = -m_d\overline{d}d - \frac{m_d}{v}\overline{d}dh \tag{1.44}$$

with

$$m_d = \frac{g_d}{\sqrt{2}}v \tag{1.45}$$

To introduce mass term for the upper component of the isospin doublet, the charge conjugate of the Higgs field ϕ^c can be used. It is defined as

$$\phi^c = i\sigma^2\phi^* \tag{1.46}$$

The Yukawa term leading to the mass term for the u quark is then

$$\mathcal{L}_{Yuk\,u} = -g_u \overline{L}\phi^c u_R - g_u \overline{u_R}(\phi^c)^\dagger L \tag{1.47}$$

Since the hypercharge associated to ϕ^c is $y_{\phi^c} = -1$, this term is gauge invariant as well and gives after spontaneous symmetry breaking

$$\mathcal{L}_{Yuk\,u} = -m_u\overline{u}u - \frac{m_u}{v}\overline{u}uh \tag{1.48}$$

with

$$m_u = \frac{g_u}{\sqrt{2}}v \tag{1.49}$$

After spontaneous symmetry breaking (assuming Dirac neutrinos), the Yukawa terms thus lead to the following Lagrangian

$$\mathcal{L}_{Yuk} = \sum_f -m_f \overline{\psi}\psi - \frac{m_f}{v}\overline{\psi}\psi h \tag{1.50}$$

A mass term is therefore present for each fermion. Moreover each of them is coupled to the Higgs boson with a coupling proportional to its mass. Note that the fields involved in the previous Lagrangian are the mass eigenstates. As explained in section 1.1.4, those mass eigenstates are not eigenstates of the weak interaction involving the W bosons, which results in a mixing, described by the CKM matrix in the quark sector and the PMNS matrix in the lepton sector.

So far, all the measurements done in particle physics are consistent with the version of the SM presented above. Assuming massive Dirac neutrinos, the SM can be described by 25 parameters. Since the measurement of the mass of the Higgs boson observed for the first time at LHC in 2012 [1, 2], all of those parameters are measured, except in the neutrino sector. Therefore, theoretical predictions can be made for every possible process described by the SM and compared with experimental results.

1.2 Phenomenology of the Higgs boson at LHC

1.2.1 Proton-proton collisions

Before describing the production mechanisms of the Higgs boson at LHC, it is useful to explain first some properties of the proton-proton collisions. The description of these collisions can be factorized into the "hard" and the "soft" interaction [37]. The "hard" interaction corresponds to the interaction between the elementary particles called partons and can be described perturbatively as a high energy process. This is not the case for the "soft" interaction, which is associated with low-energy properties of QCD, such as confinement and hadronization. This decomposition enables to write the cross section $\sigma_{pp\to X}$, with X any final state, as

$$\sigma_{pp\to X} = \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \sum_{i,j} f_i(x_1,\mu^2) f_j(x_2,\mu^2) \hat{\sigma}_{ij\to X} \tag{1.51}$$

$x_{1,2}$ is the fraction of the total momentum of the proton carried by each parton. They can be valence quarks (u or d quarks in the case of the proton), gluons (which are constantly exchanged between the quarks) or even antiquarks from the sea (virtually produced in gluon splitting within the proton). $f_i(x,\mu^2)$ are the Parton Density Functions (PDF), which depend on the flavor i of the parton, the fraction x of the total momentum of the proton carried by the parton and the scale of the hard process μ^2. Since the proton is a composite object, the description of its inner structure is achieved thanks to the PDFs: $f_i(x,\mu^2)\mathrm{d}x$ corresponds to the number of partons of type i within the proton carrying a

momentum fraction between x and $x + \mathrm{d}x$. The last term $\hat{\sigma}_{ij\to X}$ is the cross section of the hard process $ij \to X$ and can be computed perturbatively from Feynman diagrams.

The PDFs dependence on x is not completely calculable, since they depend on "soft" QCD processes. However the dependence on μ^2 can be computed from the DGLAP equation [38, 39, 40]. The PDF dependence on x can be measured in some experiments at a given scale μ^2 and be extrapolated to the LHC collisions. Several sets of PDFs are currently available, which are determined from deep inelastic scattering experiments (HERA), fixed target experiments and hadron collisions at the Tevatron and the LHC. The PDFs from the NNPDF3.0 [41] set are presented at different energy scales in Fig. 1.5. At high x, the dominant contribution in the PDFs comes from the u and d valence quarks, whereas at low x, there is a large dominance of the gluons.

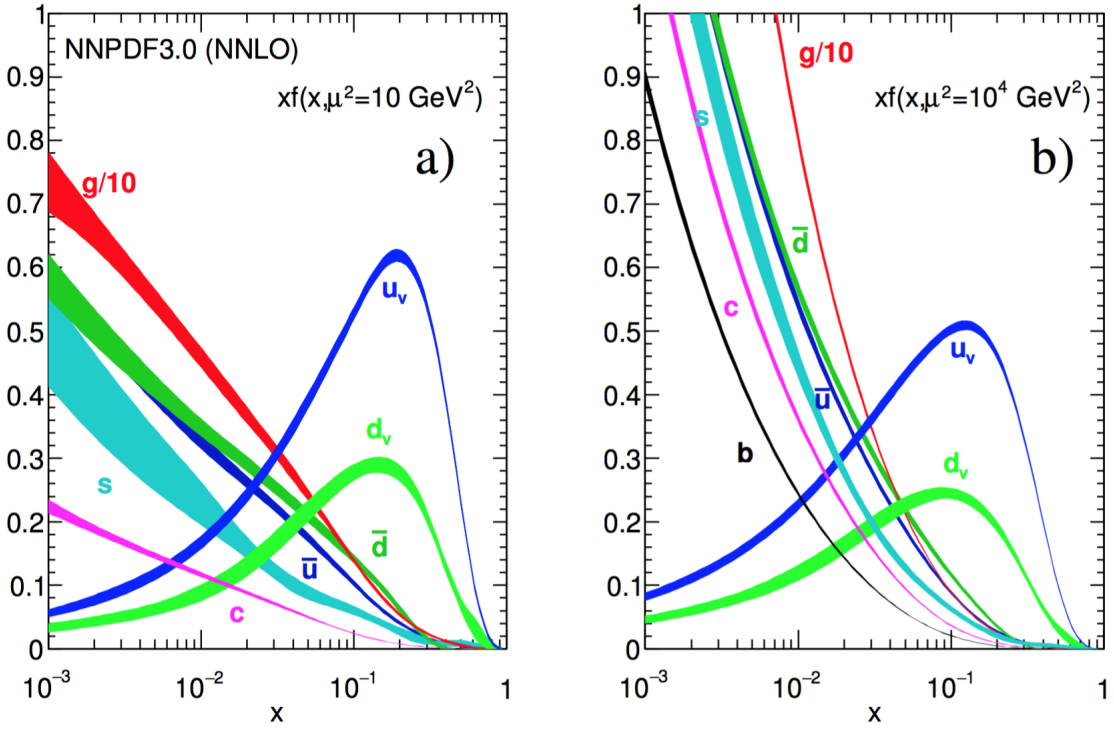

Figure 1.5: Parton Distributions Functions $xf(x, \mu^2)$ of the proton from the NNPDF3.0 set at different energy scales μ^2 [42]

1.2.2 Production of the Higgs boson at LHC

The Higgs boson being coupled to every massive particle, several production modes are possible in particle colliders. In proton-proton collisions at LHC, four main production modes are theoretically possible (see Fig. 1.6). The cross sections of the different production modes are presented in Fig. 1.7.

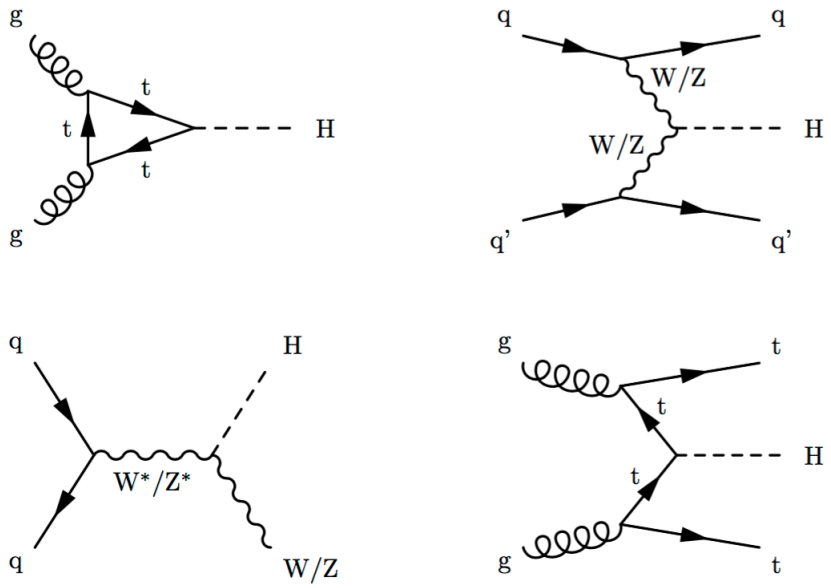

Figure 1.6: Feynman diagrams of the main Higgs production modes: gluon-gluon fusion (top left), Vector Boson Fusion (top right), VH (bottom left) and $t\bar{t}H$ (bottom right)

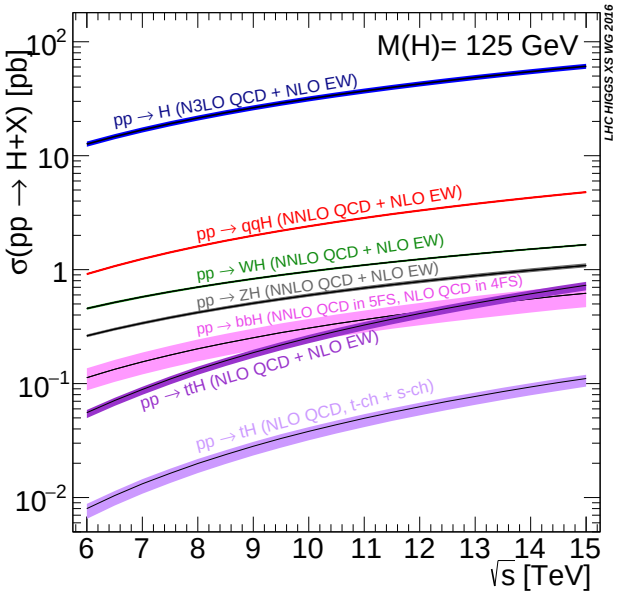

Figure 1.7: cross section of the main Higgs production modes in function of the center-of-mass energy $\sqrt{s}$ [43]

The main production mode is the gluon-gluon fusion process, in which two gluons interact to produce a Higgs boson via a loop of heavy quarks. This loop is required since the Higgs boson does not couple to the gluons which are massless. As the Higgs coupling to fermions is proportional to their mass and as the top quark is by far the most massive one, the process is dominated at leading order within the SM by a top-quark loop. Current measurements of the Higgs boson cross sections are consistent with the SM expectation for the top-quark Yukawa coupling [44]. However, any particle beyond the SM sensitive to QCD may in principle also enter this loop, therefore this production mode only enables to measure the coupling of the Higgs boson to the top quark in a model-dependent way.

For the other production modes, additional particles are produced in association with the Higgs boson. The second most likely production mode is the vector boson fusion (VBF). In that process, two quarks from the protons radiate massive vector bosons which interact to produce a Higgs boson. At LHC, the W fusion cross section is approximately three times higher than the Z fusion because of the different couplings of the quarks to the W and Z bosons. The experimental signature of that production mode is cleaner than the gluon-gluon fusion. Indeed, since that process is purely electroweak and there is no color exchange between the quarks, the QCD activity is concentrated around the outgoing quarks. Consequently, analysis categories specifically sensitive to this production mode typically require two forward jets with large pseudorapidity separation and large di-jet invariant mass, with no jet in-between.

In the case of the VH production mode or Higgs-Strahlung, a valence quark interacts with an antiquark from the sea to produce a massive vector boson which radiates a Higgs boson. The WH cross section is approximately twice higher than the ZH cross section. This difference is even increased when a leptonic decay of the vector boson is considered, since $BR(W \to \ell\nu) = 21.3\%$ whereas $BR(Z \to \ell\ell) = 6.7\%$[2].

The last of the four main production modes corresponds to the associated production of a Higgs boson with heavy quarks. In the case of top quarks, this process is referred to as $t\bar{t}H$ production mode, while in the case of b quarks this is referred to as $b\bar{b}H$ production. Those processes can be initiated either by two incoming gluons or a $q\bar{q}$ initial state (see Fig. 1.8). Those two production modes have similar cross sections at $\sqrt{s} = 13$ TeV. Since no loop is involved, unlike for the gluon-gluon fusion, the $t\bar{t}H$ production mode enables to probe directly the Higgs coupling to the top quark, which is the main focus of this thesis. During the Run I of LHC, the sensitivity to this specific production mode was not enough to claim its discovery with a 5σ significance. Searches for $t\bar{t}H$ production at LHC have previously been published for specific decay modes of the Higgs boson in Run 1 [45, 46, 47, 48]. The first combination of $t\bar{t}H$ searches in different final states has been published by the CMS collaboration on the full data set collected at $\sqrt{s} = 7$ and 8 TeV [49]. Assuming SM branching ratios, this analysis has been able to set a 95% confidence level upper limit on the $t\bar{t}H$ signal strength $\mu = \sigma/\sigma_{\mathrm{SM}}$ at 4.5, while an upper limit of 1.7 is expected in the background-only hypothesis (see Fig. 1.9). The measured value of the signal strength was of 2.8 ± 1.0 at 68% confidence level (see Fig.

[2]Decays into τ leptons are not considered here.

1.10), representing a 3.4σ excess over the background-only hypothesis and a 2σ excess over the Standard Model expectation including the $t\bar{t}H$ production. Combining Run 1 results, the ATLAS collaboration measured independently an excess in the $t\bar{t}H$ production, associated with a signal strength $\mu = 1.81 \pm 0.80$ [50]. The combination of CMS and ATLAS results on the Higgs boson production and decay rates [3] enabled to increase the measured significance over the background-only hypothesis to the 4.4σ level, while a 2.0σ significance was expected.

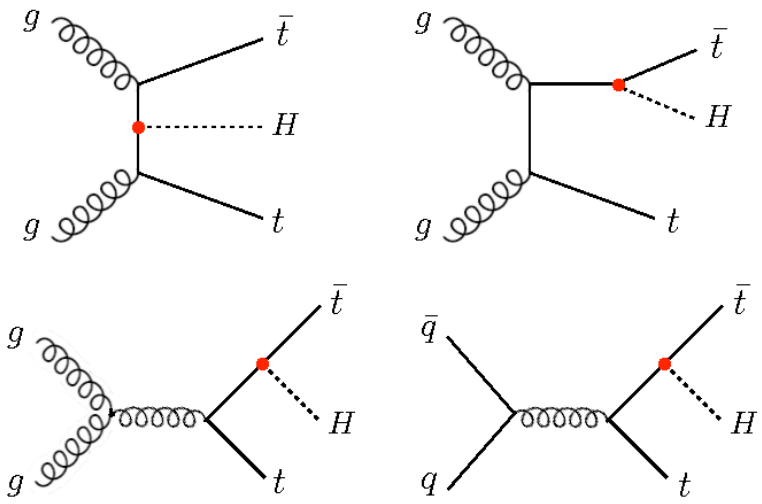

Figure 1.8: Feynman diagrams contributing to the $t\bar{t}H$ production mode

During the Run II of LHC, the center-of-mass energy has been increased to $\sqrt{s}$=13 TeV at least. This has resulted in an increase in the cross section of the different production modes as seen in Table 1.1. The $t\bar{t}H$ production has benefitted from the largest increase (by a factor 4) thanks to an opening of the phase space. Indeed, since three massive particles have to be produced, the momentum fractions of the incoming partons $x_{a,b}$ have to fulfill $(x_a + x_b)\sqrt{s} \geq M_H + 2M_t$, which is easier to achieve at $\sqrt{s}$=13 TeV than at $\sqrt{s}$=8 TeV. Despite this boost in the cross section, the $t\bar{t}H$ production will still remain the rarest production mode with a cross section of only 0.51 pb.

1.2.3 Decay channels of the Higgs boson

Since the Higgs boson couples to every massive particle, it can decay in many different channels. It can even couple indirectly, via boson or fermion loops, to photons and gluons. The corresponding branching ratios depend on the Higgs mass and are presented in Fig.

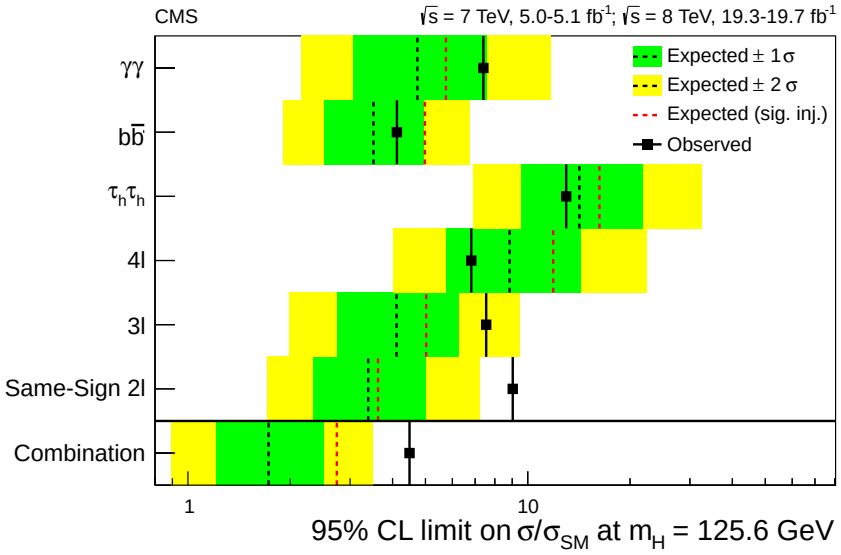

Figure 1.9: 95% CL upper limits on the signal strength parameter $\mu = \sigma/\sigma_{\mathrm{SM}}$ for the $t\bar{t}H$ process at $m_H = 125.6$ GeV, in each individual $t\bar{t}H$ channel and for all channels combined [49]

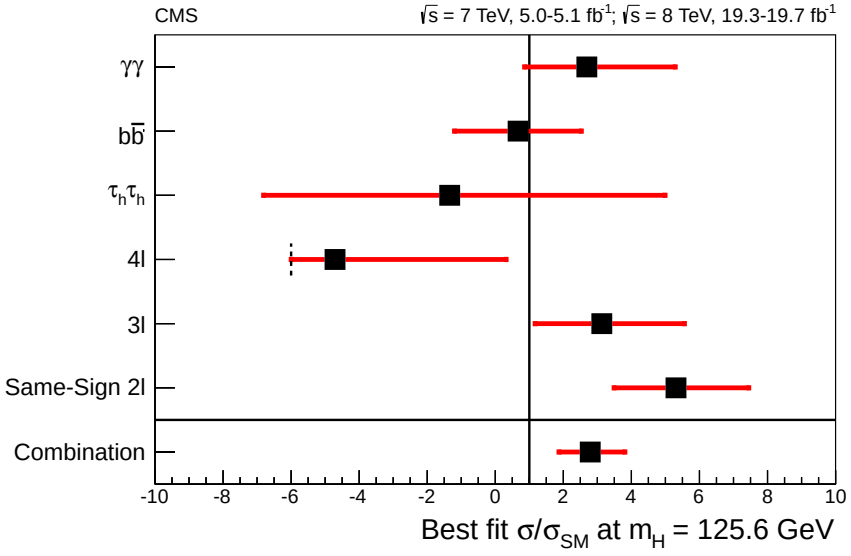

Figure 1.10: Best-fit values of $\mu = \sigma/\sigma_{\mathrm{SM}}$ for the $t\bar{t}H$ process at $m_H = 125.6$ GeV, in each individual $t\bar{t}H$ channel and for all channels combined [49]

Production mode	Cross section [pb]		Precision
	$\sqrt{s} = 8$ TeV	$\sqrt{s} = 13$ TeV	
GGF	$19.27 \pm 10.4\%$	$43.92^{+10.3\%}_{-9.9\%}$	NNLO+NNLL QCD and NLO EW
VBF	$1.58^{+2.6\%}_{-2.8\%}$	$3.75 \pm 3.3\%$	NNLO QCD and NLO EW
WH	$0.70 \pm 2.5\%$	$1.38^{+2.3\%}_{-2.7\%}$	NNLO QCD and NLO EW
ZH	$0.42 \pm 4.0\%$	$0.87 \pm 4.4\%$	NNLO QCD and NLO EW
$t\bar{t}H$	$0.13^{+8.9\%}_{-12.3\%}$	$0.51^{+10.5\%}_{-12.8\%}$	NLO QCD
$b\bar{b}H$	$0.20^{+12\%}_{-16\%}$	$0.51^{+15\%}_{-25\%}$	Santander Matching with 5FS (NNLO) and 4FS (NLO) [51]

Table 1.1: Cross section of the main Higgs production modes at $\sqrt{s} = 8$ and 13 TeV for $M_H = 125$ GeV [52]

1.11. The final states of analysis involving Higgs boson generally include leptons or photons, since these particles are rarely produced in QCD processes, which represent a major background in proton-proton collisions. These analyses can therefore benefit from lepton and photon triggers (see Section 2.2.7). Moreover the resolution on variables associated to leptons and photons is also usually much better than the one on those related to quarks or gluons.

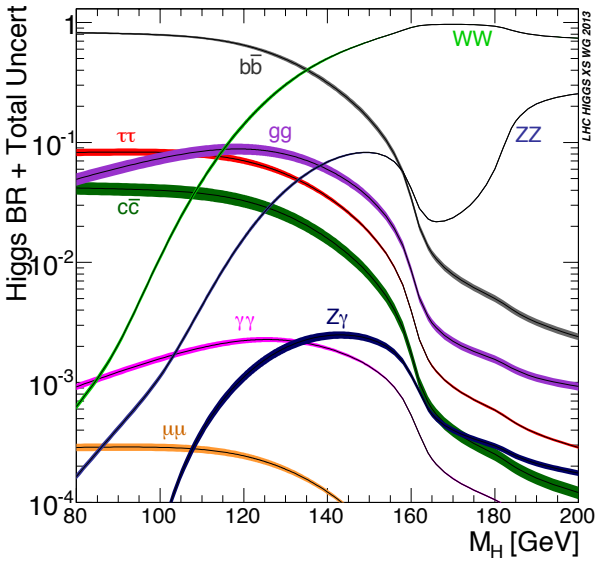

Figure 1.11: Branching ratios of the Higgs boson [43]

The highest branching ratio of the Higgs boson is for the decay $H \to b\bar{b}$ (57.7%). This decay mode suffers from a high QCD background, which can be partially reduced

using dedicated algorithms to identify jets originating from b quarks. However the QCD production of $b\bar{b}$ pairs represents a very important irreducible background which strongly limits the sensitivity of a search of the $H \to b\bar{b}$ decay targetting the gluon fusion production mode. Therefore searches for the $H \to b\bar{b}$ decay mostly focus on the VH production mode [53, 54] (see Section 1.2.2), where a massive electroweak boson is produced in association with the Higgs boson. The presence of leptons and/or missing transverse energy in the final state is then used to select events with $Z \to \ell\ell$, $Z \to \nu\nu$ or $W \to \ell\nu$.

The decay $H \to \tau\bar{\tau}$ represents the other main decay of the Higgs boson into fermions accessible at LHC, with a branching ratio of 6.4%. However the resolution on the invariant mass of the Higgs boson is limited in that channel. Indeed, since neutrinos are involved in the decay of the τ leptons, the four-momenta of the τ's cannot be exactly reconstructed and the reconstructed invariant mass has to take into account a contribution from the missing transverse energy. Some reconstruction techniques based on theoretical probability densities or on multivariate analysis can nevertheless be used. The combination of ATLAS and CMS Run 1 results yielded an excess of events over the background-only hypothesis with a 5.5σ significance [3]. Other decays of the Higgs boson into fermions may be potentially accessible at LHC, although they suffer either from a very high QCD background ($H \to c\bar{c}$, $BR = 2.7\%$) or from a very low branching ratio ($H \to \mu\bar{\mu}$, $BR = 2.2 \times 10^{-4}$).

The Higgs boson also couples to massive electroweak bosons, therefore it can decay into ZZ ($BR = 2.7\%$) or W^+W^- ($BR = 21.6\%$). Since $M_H < 2M_V$ only one of the electroweak bosons produced in this decay can be on-shell. Despite its very low branching ratio (1.3×10^{-4}), the decay mode $H \to ZZ \to 4\ell$ was one of the main discovery channels of the Higgs boson in 2012 [1, 2], thanks to its high resolution on the 4ℓ-invariant mass and to its very large signal over background ratio of $\sim 50 - 100\%$. The distribution of the 4ℓ-invariant mass observed in the preliminary CMS analysis based on 2016 data is presented in Fig. 1.12. This channel has also been used to measure properties of the Higgs boson such as its spin and CP quantum numbers [55, 56] or its decay width [57, 58]. For the $H \to WW$, the searches mainly focused on the $H \to WW \to 2\ell 2\nu$ decay channel, which has a poor resolution on the Higgs boson mass due to the presence of missing transverse energy. In that channel, the ATLAS and CMS collaboration respectively reported an observed excess over background of 6.1 and 4.3 standard deviations [59, 60].

Finally, the Higgs boson can also decay into bosons via loops of fermions (mostly top quarks) or W bosons. Although the Higgs boson does not couple to gluons or photons at tree level, since these particles are massless, an effective coupling can be generated at quantum level (see Fig. 1.13). This leads to the decays $H \to gg$, $H \to \gamma\gamma$ and $H \to Z\gamma$. Because these decays are loop-induced, they are suppressed by two extra-powers of the coupling constant (α_s or α) with respect to the tree-level decays described previously but this suppression is mitigated thanks to the large coupling of the Higgs boson to the virtual particles (namely the top quark and the W boson). This is for instance the case for the $H \to gg$ decay which has a branching ratio of 8.6%. However the sensitivity of that decay mode is very low because of its high irreducible QCD background. In contrast, the decays $H \to \gamma\gamma$ ($BR = 2.3\,10^{-3}$) and $H \to Z\gamma$ ($BR = 1.5\,10^{-3}$) have enough sensitivity to be probed at LHC. The very high resolution on the invariant di-photon mass contributed to

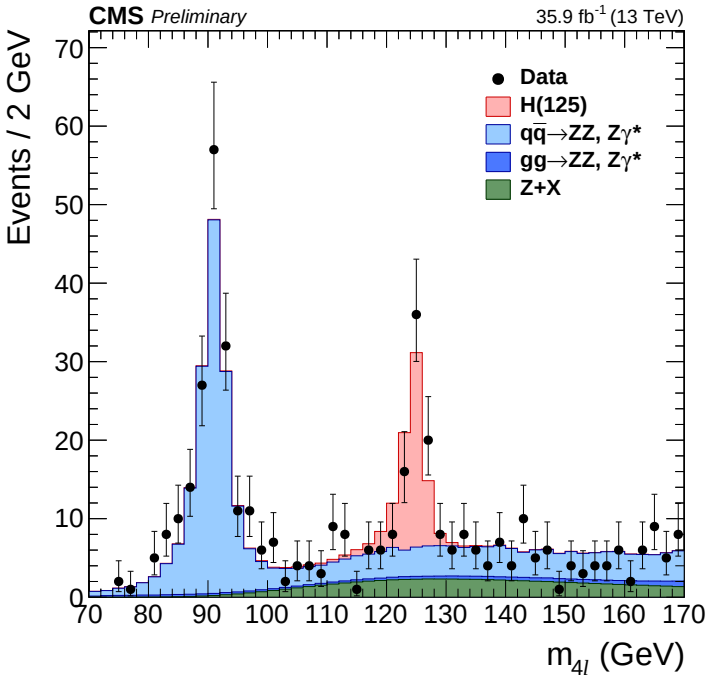

Figure 1.12: Distribution of the 4ℓ-invariant mass for the mass region $70<m_{4\ell}<170$ GeV observed in the preliminary CMS Run 2 analysis [61]

make of the $H \to \gamma\gamma$ one of the main discovery channels of the Higgs boson in 2012 [1, 2], in spite of its high background from the QCD $\gamma\gamma$ production. The diphoton invariant mass observed in the CMS Run 2 analysis is presented in Fig. 1.14. Since a top quark is involved in the loop, the measurement of the $H \to \gamma\gamma$ gives access to the top-quark Yukawa coupling, although indirectly only.

1.3 Focus on the associated production of Higgs boson with top quarks, with $H \to \tau\bar{\tau}$ decay

1.3.1 τ lepton physics

The τ lepton is the heaviest of the leptons, with a mass of [26]

$$m_\tau = 1776.82 \pm 0.16 \text{ MeV} \tag{1.52}$$

Because of its high mass, it is the only lepton which can decay into hadrons ("hadronic decay"). It can also decay into an electron or a muon ("leptonic decay"). The decay of the τ lepton occurs through the weak interaction involving an off-shell W boson, as seen in Fig. 1.15. In the case of a leptonic decay, two neutrinos are produced in association with a charged lepton, while a single neutrino is produced in a hadronic decay. The lifetime of the

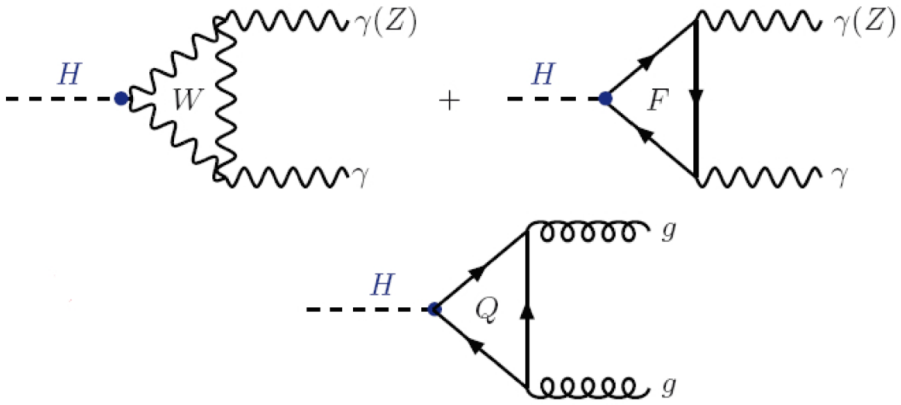

Figure 1.13: Feynman diagrams of the loop-induced decays $H \to gg$ and $H \to \gamma\gamma(Z)$

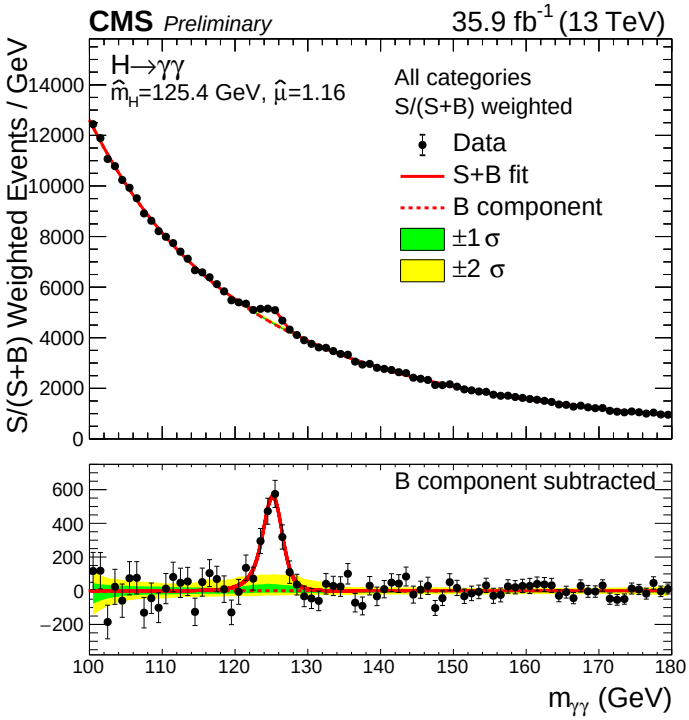

Figure 1.14: Distribution of the diphoton invariant mass observed in the preliminary CMS Run 2 analysis [62]

τ lepton being relatively short, a 30 GeV τ lepton can only cover a distance $\gamma c\tau \sim 1$ mm before decaying and it is therefore only possible to reconstruct the visible decay products of the τ leptons, hadrons or a charged lepton. The neutrinos remaining undetected, it is

not possible to evaluate exactly the momentum of the τ lepton before its decay. In the case of a hadronic decay, the set of all the reconstructed decay products is usually referred to as a "hadronic tau" and labelled as τ_h.

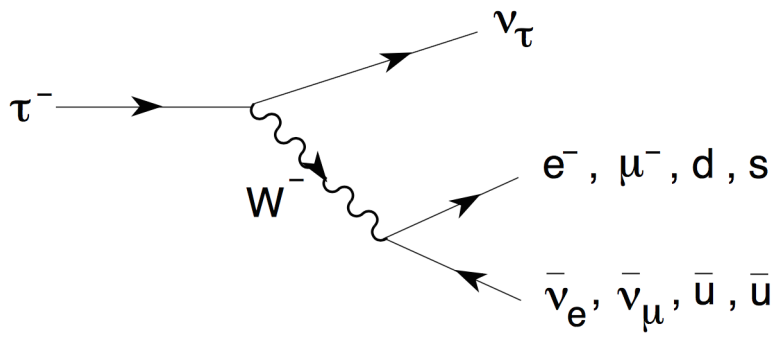

Figure 1.15: Feynman diagram of the possible τ lepton decays. The decay involving a s quark, leading to a kaon in the final state, is possible through CKM-mixing.[63]

The branching ratios of the τ lepton are detailed in Table 1.2. The leptonic decays represent around 35% of the decays, the decays into an electron or a muon sharing a similar branching ratio. The remaining 65% correspond to hadronic decays, mainly into one or three charged mesons, with or without additional neutral pions. The charged mesons are in most cases charged pions ("prongs"), the decays into kaons being CKM-suppressed. The neutral pions decay almost immediately into two narrow photons. Some of the hadronic decays occur through a meson resonance, which affects the distribution of the "visible mass" of the τ (invariant mass of its visible decay products), as seen in Fig. 1.16.

Decay mode	**Resonance**	**Mass (MeV)**	**Branching ratio**
$\tau \to e\bar{\nu}_e\nu_\tau$	-	-	17.85%
$\tau \to \mu\bar{\nu}_\mu\nu_\tau$	-	-	17.36%
$\tau \to h^-\nu_\tau$	π	139.6	11.6%
$\tau \to h^-\pi^0\nu_\tau$	ρ	770	26.0%
$\tau \to h^-\pi^0\pi^0\nu_\tau$	a_1	1260	9.5%
$\tau \to h^-h^+h^-\nu_\tau$	a_1	1260	9.8%
$\tau \to h^-h^+h^-\pi^0\nu_\tau$	-	-	4.8%
others	-	-	3.1%

Table 1.2: Branching ratios of the τ lepton decays

Since the τ's produced at LHC are very energetic, when they undergo a hadronic decay into more than one particle, the visible decay products are collimated and detected as a

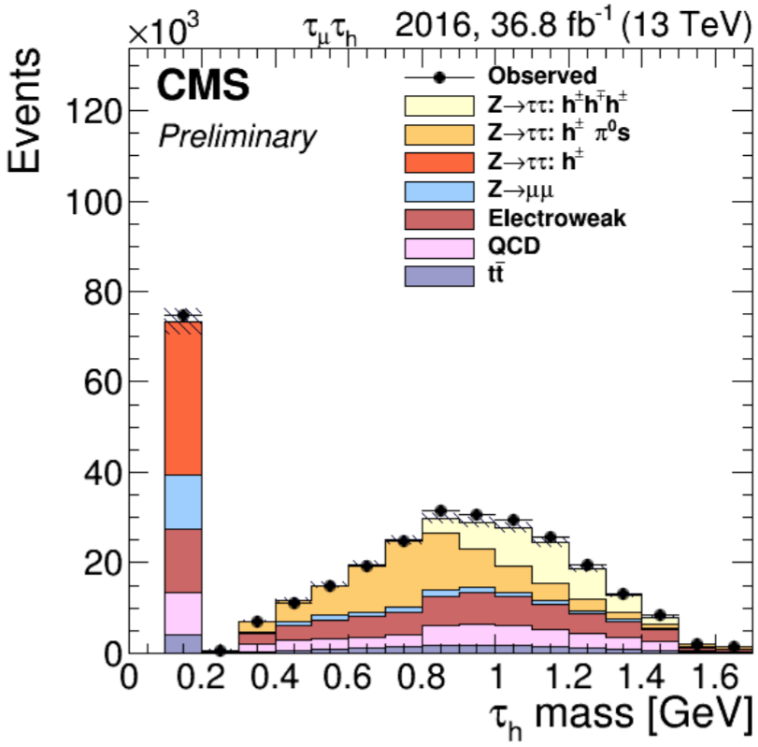

Figure 1.16: Observed and predicted distributions for the visible τ_h mass, $m_{vis}^{\tau_h}$, for events with a reconstructed μ and τ_h in CMS data. The $Z \to \tau\tau$ contribution is split according to its reconstructed decay mode. The mass distribution of the τ_h built from one charged hadron and photons peaks around the mass of the intermediate $\rho(770)$ resonance, while the one for τ_h reconstructed from three charged hadrons peaks around the mass of the intermediate $a_1(1200)$ resonance. By construction, the τ_h built from one charged hadron and no photon are reconstructed with the $\pi^{\pm}$ mass and represent the main contribution of the first bin. [64]

bunch of close-by particles in the detector. In order to differentiate hadronic τ decays from jets originating from quarks or gluons, it is necessary to exploit the granularity of the detector to reconstruct each individual decay product. The number of particles produced in those decays is indeed significantly smaller than the one produced in the hadronization of quarks or gluons. This feature is exploited by specific algorithms dedicated to the reconstruction and the identification of hadronic τ's. They will be described in Section 2.3.6.

For a resonance decaying into a pair of τ leptons, six final states are possible and their branching ratios are presented in Table 1.3. The dominant decay mode is the decay into two τ_h. However the sensitivity to that channel is reduced due to the misidentifaction of jets originating from quarks or gluons as τ_h, which can lead to a sizeable background from

QCD processes. On the other hand, although the purely leptonic final states can rely on a very clean identification of the charged leptons, they suffer from relatively low branching ratios. As for the decay modes into a charged lepton and a τ_h, they combine at the same time a high branching ratio and a lower mistag rate than the decay into two τ_h.

Decay mode	**Branching ratio**
$X \to \tau\tau \to ee + \nu'$s	3.19%
$X \to \tau\tau \to \mu\mu + \nu'$s	3.01%
$X \to \tau\tau \to e\mu + \nu'$s	6.20%
$X \to \tau\tau \to e\tau_h + \nu'$s	23.13%
$X \to \tau\tau \to \mu\tau_h + \nu'$s	22.50%
$X \to \tau\tau \to \tau_h\tau_h + \nu'$s	41.98%

Table 1.3: Branching ratios of a resonance X decaying to a pair of τ leptons

The decay of the Higgs boson into τ's is only one of the many other Standard Model processes studied at LHC which involve τ leptons. τ's can also be produced through Drell-Yan production [65], where a Z boson or a photon decays into τ's, or processes involving a W boson [66] or a top quark [67]. The τ leptons are also used as a probe for new physics beyond the Standard Model, for instance in searches for Supersymmetry [68], Z' bosons [69] or charged Higgs bosons [70].

The analysis of the angular correlations between the τ lepton and its decay products can also be used to extract some information regarding the polarization of the τ lepton. This is true for all the τ decay modes but the sensitivity is higher in the case of hadronic decay modes because they involve a single neutrino in the final state. It is then possible to exploit that property to determine the quantum numbers (J and CP) of a resonance decaying into τ leptons [71].

1.3.2 Top quark physics

The top quark is the most massive particle described within the Standard Model, with a mass currently estimated at [26]

$$M_t = 173.21 \pm 0.87\,\text{GeV} \tag{1.53}$$

The top quark has a large decay width $\Gamma = 2.0 \pm 0.5$ GeV [26], because its large mass allows for its decay into a down-type quark and an on-shell W boson. This decay width corresponds to a lifetime of only 5×10^{-25} s or $c\tau \sim 0.1$ fm. The typical QCD scale being ten times larger, the top quark does not have time to hadronize before its decay, unlike the other quarks.

Because of the large matrix element $|V_{tb}|^2$ in the CKM matrix, the top quark almost exclusively decays into a b quark and a W boson with a measured branching ratio of 91 ± 4

% [26]. The W quark can then decay either into two quarks, with a branching ratio of 67%, or into a charged lepton and a neutrino. In the case of a leptonic W decay, the three lepton flavours are approximately equally likely. The Feynman diagrams of the main top decays are presented in Fig. 1.17. Depending on the W decay mode, the top quarks can then be referred to as "hadronic" or "leptonic" tops. Leptonic tops are in principle easier to detect because of the presence of a charged lepton but because of the neutrino in the final state, it is not possible to reconstruct directly the four-momentum of the top quark. In the case of hadronic tops, jet reconstruction algorithms, which will be detailed in section 2.3.4, can give an estimate of the four-momenta of all the final-state quarks and therefore of the hadronic top. Depending on the energy of the hadronic top, standard jet algorithms may fail to reconstruct three separate jets because some of the final-state quarks are too collimated. However specific top tagging algorithms, based on the analysis of the jet substructures, can still be used to identify hadronic tops [72].

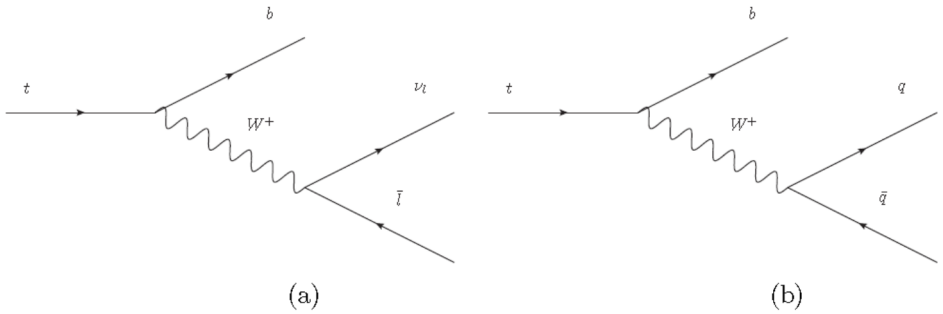

Figure 1.17: Feynman diagram of (a) leptonic top decays and (b) hadronic top decays. l stands for e, μ or τ. [73]

In the most current Standard Model processes involving top quarks, they are produced by pair via the strong interaction. Assuming a 100% branching ratio for $t \to bW$, there are therefore ten possible final states for the top pair, whose branching ratio are given in Table 1.4. The dilepton, lepton + tau and ditau final states have the lowest branching ratio (10.6% altogether), while the all-hadronic final state has the largest one but suffers from a high QCD background at LHC. The lepton + jets final state are interesting to exploit because they have at the same time a sizeable branching ratio and an energetic lepton in the final state.

The top quark has been extensively studied using Tevatron data by the CDF and DØ collaborations in various production modes [74, 75]. At LHC, the cross sections of Standard Model processes involving the top quark have also been measured in different final states [76, 77]. As the only Standard Model particle with Yukawa couplings to the Higgs boson of order one, the top quark is commonly associated to new physics. In some supersymmetric theories for instance, the scalar partner of the top quark is often assumed to be the lightest scalar superpartner and the top quark plays therefore a special role in searches for Supersymmetry at LHC [78]. It is also involved in other physics analyses

Category	Decay mode	Branching ratio
Dilepton	$tt \to bbee + \nu'$s	1.6%
	$tt \to bb\mu\mu + \nu'$s	1.6%
	$tt \to bbe\mu + \nu'$s	3.1%
Lepton + tau	$tt \to bbe\tau_h + \nu'$s	1.9%
	$tt \to bb\mu\tau_h + \nu'$s	1.9%
Ditau	$tt \to bb\tau_h\tau_h + \nu'$s	0.5%
Lepton + jets	$tt \to bbqqe + \nu'$s	17.2%
	$tt \to bbqq\mu + \nu'$s	17.0%
Tau + jets	$tt \to bbqq\tau_h + \nu'$s	9.9%
All-hadronic	$tt \to bbqqqq$	45.4%

Table 1.4: Branching ratios of a $t\bar{t}$ pair. The decays of τ leptons are taken into account here.

beyond the Standard Model such as the search for W' or Z' bosons [79, 80].

1.3.3 b quark physics

The associated production of a Higgs boson with top quarks leads to the presence of at least two b quarks in the final state, from the top quark decays. Other processes at LHC can also lead to the production of b quarks, such as the production of a Higgs boson decaying into b quarks, the production of a top quark pair or QCD processes where b quarks can be produced from gluon splitting.

With respect to other parton flavors, the jets originating from b quarks exhibit specific properties which enable to tag them. The b-tagging is then useful to reduce the important QCD background with no b quark production. The b-jet identification can also be used in multi-jet final states to determine the correct assignment of reconstructed objects to parton level objects, like in $t\bar{t}$ lepton + jets or all hadronic final states, where it can be used to distinguish jets originating from b quarks from those produced by light quarks from hadronic W decays.

The characteristic features of the b-jets originate from the specific properties of b quark hadronization and weak decays of B mesons, which are exploited in b-tagging algorithm (see Section 2.3.4). One of them is the large lifetime of B mesons of about 1.5 ps ($c\tau \approx 450\mu$m), which can be measured thanks to high resolution tracking detectors. This leads to the presence of a secondary vertex along the tracks of charged particles, whose shift in the transverse plane with respect to the primary vertex corresponding to the proton-proton interaction can be used to characterize such decays. Moreover B mesons have large masses (above 5 GeV) with respect to their decay products, which are therefore produced with a larger momentum transverse to the jet axis with respect to non b-jets, resulting in a wider opening angle. b-jets also have typically a larger multiplicity of charged particles.

Finally B mesons can also undergo semi-leptonic decays, as shown in Fig. 1.18. In about 40% of the cases, an electron or a muon is thus produced inside a b-jet. The large transverse momentum of the lepton with respect to the jet axis is used in some algorithms to tag b-jets [81]. Those leptons can also be misidentified as "prompt" leptons, produced in hard electroweak processes, such as the decay of Z or W bosons for instance. Events with such non-prompt leptons can then contribute as backgrounds for analyses targetting events with b-jets and leptons, such as $t\bar{t}H$ analyses. Dedicated identification criteria have thus been developed to distinguish prompt from non-prompt leptons and will be presented in more details in Section 5.2.1.

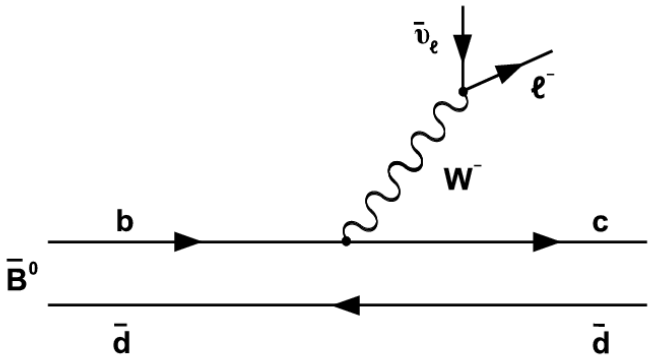

Figure 1.18: Semi-leptonic decay of a $\bar{B}_s^0$ meson into $\bar{\nu}\ell^-$ and a D^+ meson.

1.3.4 State-of-the-art $t\bar{t}H$ signal and backgrounds modeling

As one of the key processes to determine the couplings of the Higgs boson, the modeling of the $t\bar{t}H$ signal and its main backgrounds have benefited from significant efforts in order to make the theoretical predictions for those processes as accurate and precise as possible. The most recent improvements in this area are presented in [82]. The $t\bar{t}H$ signal modeling in particular has recently been improved thanks to the computation of the next-to-leading electroweak (NLO EW) corrections to this process [83, 84, 85], while the NLO QCD corrections have already been available for several years [86, 87, 88, 89]. These NLO EW corrections are important in particular to put precise constraints on the top Yukawa coupling (y_t), as they introduce additional corrections related to the Higgs boson coupling to vector bosons or to itself which modify the trivial dependence of the $t\bar{t}H$ cross section on y_t^2. The NLO cross sections for the $t\bar{t}H$ signal and for the $t\bar{t}V$ backgrounds are presented in Table 1.5 as well as their associated uncertainties.

The effect of the NLO corrections on some distributions characteristic of the $t\bar{t}H$ process is presented in Fig. 1.19. Different NLO QCD calculations from different generators interfaced with Parton Shower (PS) tools have also been compared in details. They have

Process	σ^{NLO}_{QCD}[fb]	σ^{NLO}_{QCD+EW}[fb]	K_{QCD}	δ_{EW}[%]	Scale[%]	PDF+α_S[%]
$t\bar{t}H$	498.7	507.1	1.25	1.7	+5.8 -9.2	±3.6
$t\bar{t}Z$	841.3	839.3	1.39	-0.2	+9.6 -11.3	±4.0
$t\bar{t}W$	620.6	600.8	1.50	-3.2	+12.9 -11.5	±3.4

Table 1.5: Inclusive cross sections at NLO QCD and NLO QCD+EW accuracy for $\sqrt{s}$=13 TeV of the $t\bar{t}H$ and $t\bar{t}V$ processes [82]. K_{QCD} corresponds to the ratio of the NLO QCD and LO cross sections. δ_{EW} corresponds to the relative difference between the NLO QCD+EW and NLO cross section. The scale uncertainty is estimated by varying the central value of renormalization and factorization scales $\mu_0 = M_t + M_{V/H}/2$ by a factor 2, while the PDF+α_S uncertainty is obtained following the PDF4LHC recommendations [90].

found to be overall in very good agreement, as seen in Fig. 1.20. The $t\bar{t}H$ signal and $t\bar{t}V$ background samples used in the analysis presented in Chapter 5 have been generated using MadGraph5_aMC@NLO [91] interfaced with PYTHIA 8.2 [92] to model the parton shower and fragmentation.

1.3.5 Sensitivity of $t\bar{t}H$ analyses to Beyond the Standard Model physics

A large variety of models Beyond the Standard Model (BSM) predict modified top-Higgs coupling with respect to the Standard Model. This is for instance the case in Two-Higgs-Doublet Models (2HDM) [93, 94, 95]. Those can arise in several BSM theories such as the minimal supersymmetric extension of the Standard Model (MSSM) [96], twin Higgs models [97] and certain composite Higgs Models [98]. In those classes of models, two SU(2) doublets Φ_1 and Φ_2 are introduced (instead of a single one in the Standard Model).

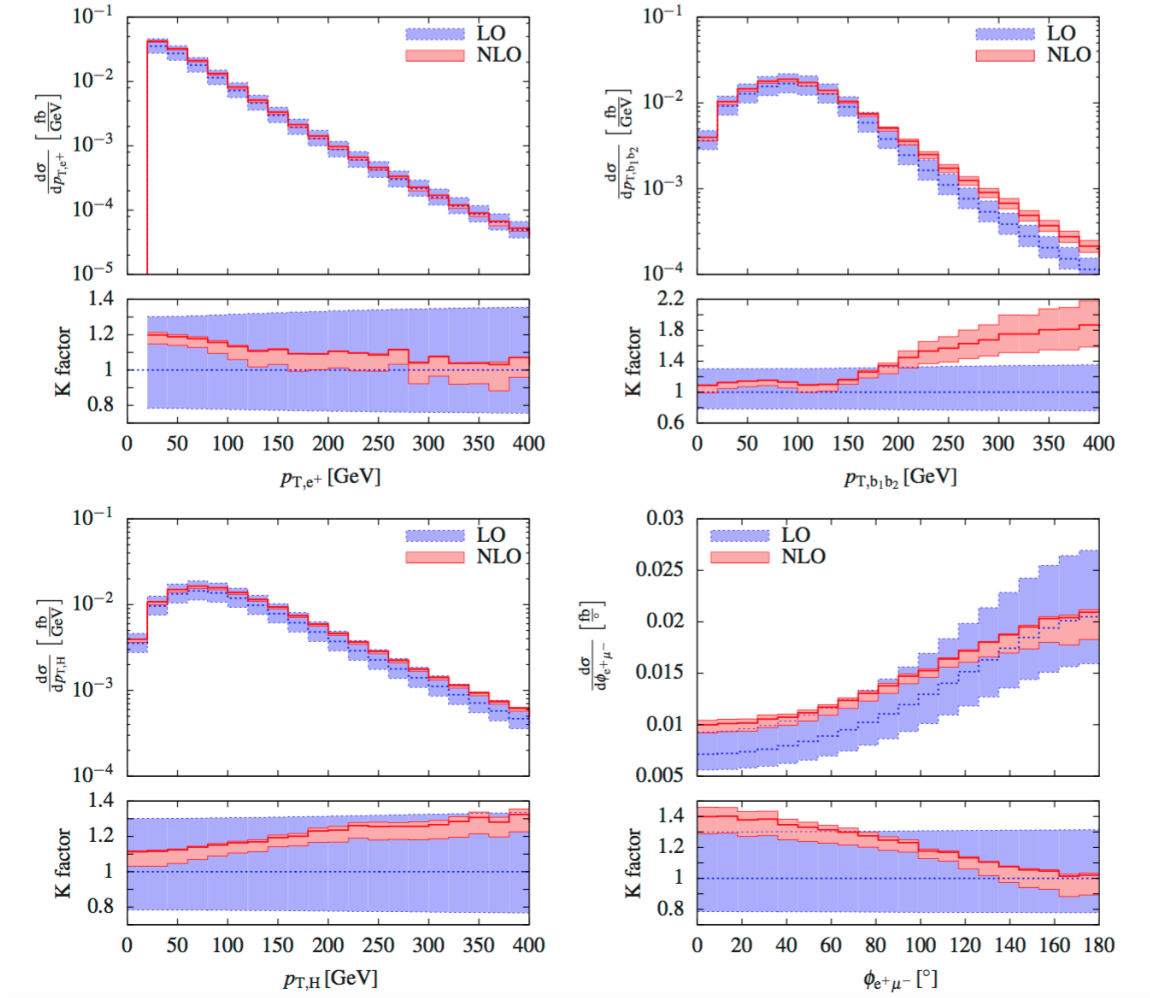

Figure 1.19: Distributions at the LHC at $\sqrt{s}$=13 TeV for the $t\bar{t}H \to b\bar{b}e^+\nu_e\mu^-\bar{\nu}_\mu H$ process: for the transverse momentum of the positron (upper left), of the b-jet pair (upper right), of the Higgs boson (lower left) and the azimuthal angle between the positron and the muon in the transverse plane (lower right). A dynamical scale $\mu_0 = (M_{T,t}M_{T,\bar{t}}M_{T,H})^{1/3}$ with $M_T = \sqrt{M^2 + p_T^2}$ is used and the bands correspond to the scale uncertainty estimated by varying the renormalization and factorization scales within a factor 2 from μ_0. [82]

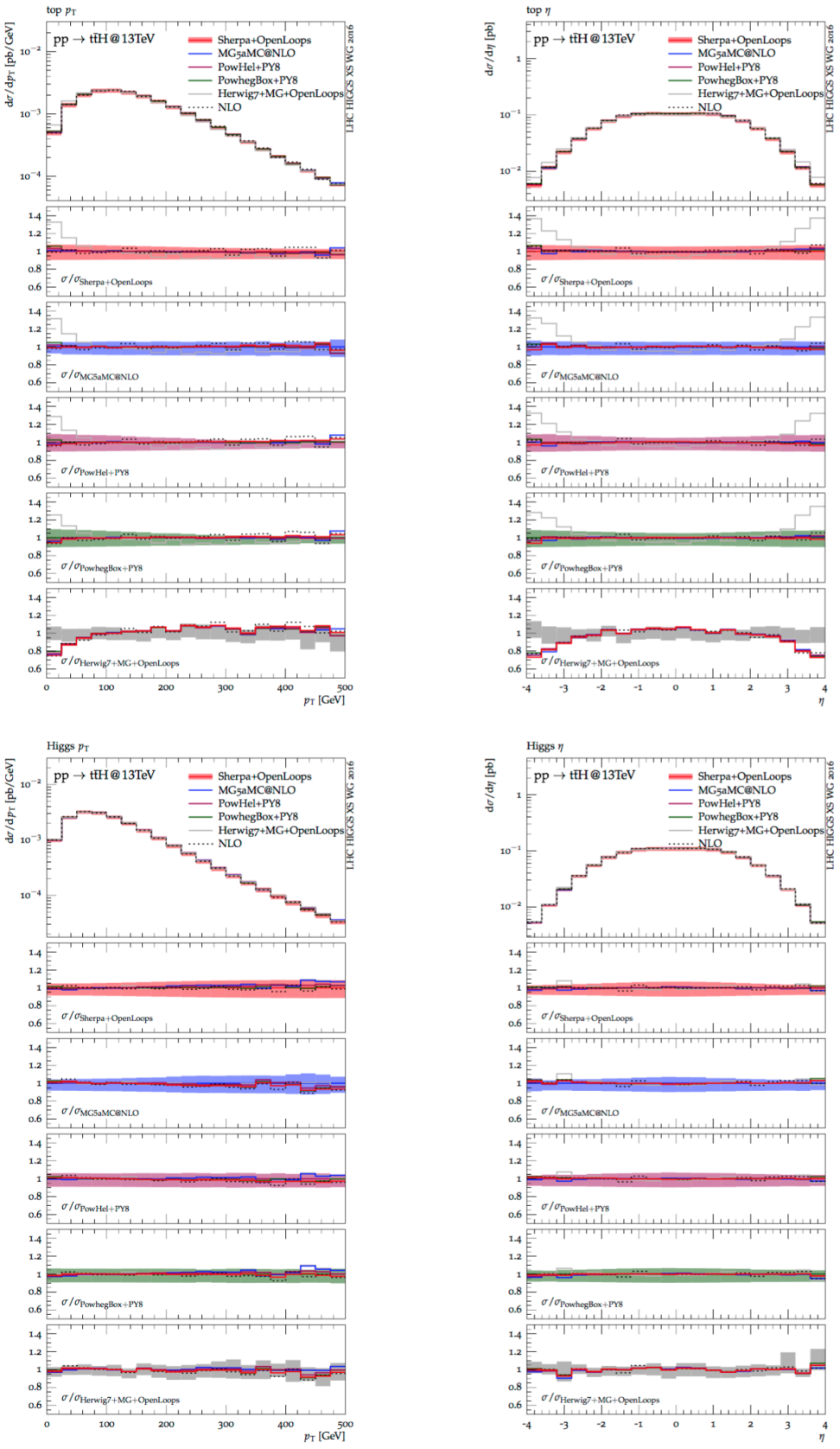

Figure 1.20: NLO QCD+PS and fixed-order NLO QCD predictions for differential $t\bar{t}H$ observables at 13 TeV. Each ratio plot shows all results normalized to one particular NLO QCD+PS prediction and the scale variation band of the latter. [82]

The most general Lagangian describing such a Higgs sector can then be written as

$$\begin{aligned} V_H = -\frac{1}{2}\left[m_{11}^2(\Phi_1^\dagger\Phi_1) + m_{22}^2(\Phi_2^\dagger\Phi_2) + \left\{m_{12}^2(\Phi_1^\dagger\Phi_2) + h.c.\right\}\right] \\ + \frac{\lambda_1}{2}(\Phi_1^\dagger\Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger\Phi_2)^2 + \frac{\lambda_3}{2}(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) \\ + \frac{\lambda_4}{2}(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) + \left\{\frac{\lambda_5}{2}(\Phi_1^\dagger\Phi_2) + \left[\lambda_6(\Phi_1^\dagger\Phi_1) + \lambda_7(\Phi_2^\dagger\Phi_2)\right](\Phi_1^\dagger\Phi_2) + h.c.\right\} \end{aligned} \quad (1.54)$$

This general potential contains 14 real free parameters (in contrast with only two real parameters in the case of a single doublet). With some additional assumptions, such as the conservation of CP symmetry in the Higgs sector, this number of parameters can however be reduced. Assuming that the vacuum expectation values of the two doublets are invariant under the $U(1)_{EM}$ gauge symmetry, they can be written as

$$\langle\Phi_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v_1\end{pmatrix}, \langle\Phi_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v_2\end{pmatrix} \quad (1.55)$$

Those two SU(2) doublets corresponds to eight scalar fields:

$$\Phi_a = \begin{pmatrix}\phi_a^+\\(v_a + \rho_a + i\eta_a)/\sqrt{2}\end{pmatrix}, a = 1, 2 \quad (1.56)$$

Among those, three correspond to Goldstone bosons which can be absorbed in the unitary gauge. After diagonalization of the mass terms of the Lagrangian, five physical Higgs bosons remain: two charged Higgses $H^\pm$, two CP-even neutral Higgses h and H and a CP-odd neutral Higgs A. A very important parameter for the study of 2HDMs is

$$\tan\beta \equiv \frac{v_2}{v_2} \quad (1.57)$$

The angle β is the rotation angle which diagonalizes the mass-squared matrices of the charged scalar fields ϕ_a^+ and pseudoscalar fields η_a, while the mass eigenstates h, H are obtained from the scalar fields ρ_a by a rotation of angle α. The phenomenology of the 2HDMs is then strongly governed by the values of those two angles. The coupling of the CP-even bosons to electroweak vector bosons is for instance given by

$$g_{hVV} = 2\frac{m_V^2}{v}\sin(\beta-\alpha),\ g_{HVV} = 2\frac{m_V^2}{v}\cos(\beta-\alpha) \quad (1.58)$$

with $v \equiv \sqrt{v_1^2 + v_2^2}$.

In the so-called type-II 2HDMs, the Φ_2 doublet has Yukawa couplings to up-type fermions while the Φ_1 doublet couples to down-type fermions. The coupling to up-type and down-type fermions of the CP-even mass eigenstates are then given by

$$g_{huu} = \frac{m_u}{v}\sin(\beta-\alpha)\left(1 + \cot\beta\cot(\beta-\alpha)\right),\ g_{Huu} = \frac{m_u}{v}\cos(\beta-\alpha)\left(1 - \cot\beta\tan(\beta-\alpha)\right) \quad (1.59)$$

$$g_{hdd} = \frac{m_d}{v}\sin(\beta-\alpha)\left(1-\tan\beta\cot(\beta-\alpha)\right),\ g_{Hdd} = \frac{m_d}{v}\cos(\beta-\alpha)\left(1+\tan\beta\tan(\beta-\alpha)\right) \tag{1.60}$$

The couplings of the h boson get compatible with the couplings of the Standard Model Higgs boson in the so-called decoupling limit with $\beta-\alpha=\pi/2$. However, away from the decoupling limit, those couplings can deviate from those predicted by the Standard Model. In particular, for low values of $\tan\beta$ the coupling of the h boson to up-type fermions gets enhanced, which would lead to an increase of the $t\bar{t}H$ cross section. The gluon-gluon fusion cross section can at the same time keep values compatible with the Standard Model in case new colored particles (such as light stop quarks) are introduced [99]. At low $\tan\beta$, the coupling of the h boson to down-type fermions gets also suppressed, which induces an increase of the branching ratio of the Higgs boson decay into electroweak bosons. Those kinds of models can then be probed in particular by studying the $t\bar{t}H, H\to WW$ process. Because of the suppressed couplings to the down-type fermions, the $t\bar{t}H, H\to b\bar{b}$ and $t\bar{t}H, H\to\tau\bar{\tau}$ processes are less sensitive to those models.

In the so-called type-I or fermiophobic 2HDMs, the Φ_2 doublet couples to all the fermions while the Φ_1 has no couplings to fermions but only to gauge bosons. For those models, charged Higgses with a mass below the top mass are not excluded yet and they can dominantly decay into $H^+\to A/HW^{+(*)}$ if $m_{H^+}>m_{A/H}$, while their fermionic decays are supressed. If $m_{A/H}\lesssim$110 GeV, A or H dominantly decays into $b\bar{b}$ or $\tau\bar{\tau}$. The process

$$pp\to t\bar{t}\to(bW^+)(\bar{b}H^-)\to(bW^+)(\bar{b}W^{-(*)}A/H) \tag{1.61}$$

can then lead to a large contamination in searches for the $t\bar{t}H, H\to b\bar{b}$ and $H\to\tau\bar{\tau}$ processes, as illustrated in Fig. 1.21 [100].

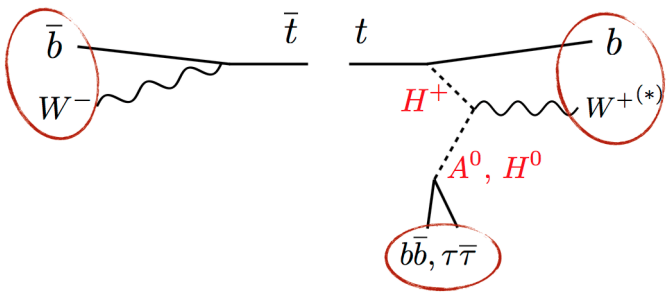

Figure 1.21: Rare top decay to $bH^\pm$ in a light type-I 2HDM, whose final states overlap with those of the Standard Model processes $t\bar{t}H, H\to b\bar{b}/\tau\bar{\tau}$ [100]

Some deviations with respect to the Standard Model are thus expected in some classes of BSM models in the different $t\bar{t}H$ final states and analyses searching for this Standard Model process can therefore also be used in principle to put some constraints on the parameters of those BSM models.

Chapter 2

The LHC and the CMS experiment

It doesn't matter how beautiful your theory is,
it doesn't matter how smart you are.
If it doesn't agree with experiment, it's wrong.

Richard Feynman

2.1 The LHC

Put into service on September 10th 2008, the Large Hadron Collider (LHC) is currently the most powerful particle collider in the world. Located under the French-Swiss border, the LHC was built at CERN in Geneva in the circular tunnel of its predecessor, the Large Electron Positron Collider (LEP). This 26.7-km-long tunnel is located at a mean depth of 100 m and hosts the two beam pipes, in which protons circulate in opposite directions. During dedicated periods, the LHC is also used to accelerate heavy ions (Pb) at a nucleon-nucleon center-of-mass energy $\sqrt{s_{NN}} = 5.1$ TeV (2.76 TeV during Run 1). The Pb-Pb and p-Pb collisions are used to cover the Heavy–Ion LHC physics programme.

Four main experiments, visible in Fig. 2.1, collect data produced in LHC collisions: ATLAS, CMS, ALICE and LHCb. ATLAS [102] and CMS [103] are multipurpose detectors that search a large panel of physics processes, from precision electroweak measurements and to the search for processes involving a Higgs boson as well as tests of supersymmetric theories or extra-dimensions. The ALICE detector [104] focuses on the analysis of heavy ion collisions to study the properties of the quark-gluon plasma. Lastly LHCb [105] is dedicated to the study of B mesons and of the CP-violation.

Before entering the LHC, the protons are accelerated in several steps by the accelerator complex shown in Fig. 2.2. The first acceleration step is achieved by the linear accelerator LINAC2, which produces 50 MeV protons. They are injected afterwards in the Proton Synchrotron Booster (PSB) and then in the Proton Synchrotron (PS), where they are accelerated up to 26 GeV. Finally the proton bunches enter the Super Proton Synchrotron where their energy is increased to 450 GeV before entering the LHC ring. Once in the

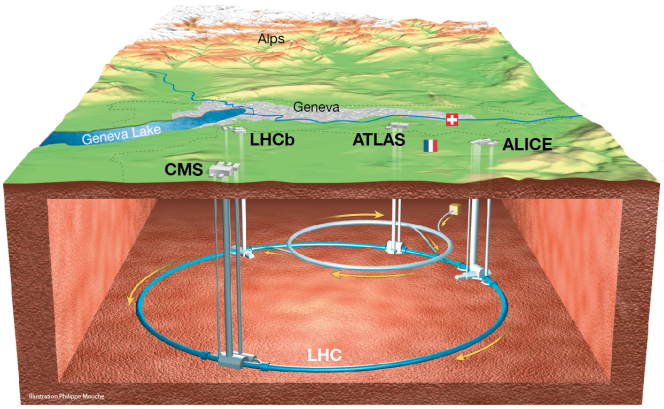

Figure 2.1: Overall view of the LHC [101]

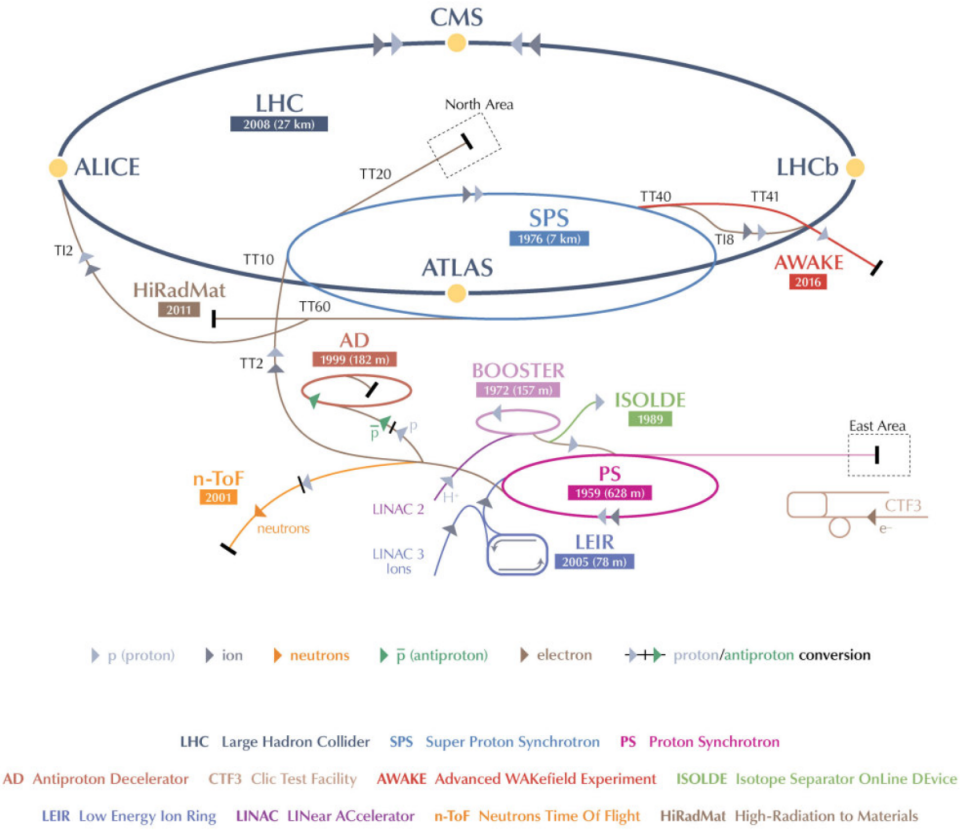

Figure 2.2: CERN's accelerator complex [106]

LHC, they are further accelerated up to their maximal energy. The linear acceleration of the protons is performed thanks to high frequency accelerating cavities whereas the bending of their trajectory is achieved with 1232 superconducting dipole magnets, which operate at a field of 8.3 Tesla and need to be cooled down to 1.9 K with superfluid helium. Additional magnets (mainly quadrupoles) are also needed in order to focus the beams.

During the LHC Run 2, the bunches contain around 10^{11} protons with an energy of 6.5 TeV (for a nominal energy of 7 TeV) and are spaced by 25 ns (50 ns at the beginning of Run 2). These bunches are characterized by a transverse normalized emittance $\epsilon_n = 3.75\mu$m and a beam β function at the collision point $\beta^* = 0.55$m. The instantaneous luminosity can be expressed as [107]

$$\mathcal{L} = \gamma \frac{n_b N^2 f_{rev}}{4\pi\beta^*\epsilon_n} R \tag{2.1}$$

with $\gamma = E_{beam}/m_p$, n_b the number of bunches in the machine (2808 for 25 ns spacing), N the number of protons per bunch, f_{rev} the revolution frequency and R a luminosity geometrical reduction factor due to a non flat crossing angle of the bunches at the collision point. The nominal instantaneous luminosity is $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ but it had been already exceeded in 2016 where it reached $1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The rate of events dN/dt (s^{-1}) for a physical process with cross section σ (cm^2) is then given by

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \tag{2.2}$$

The increase in the center-of-mass energy, from 8 TeV to 13 TeV, increases the sensitivity of the experiments for rare processes, like Higgs boson production, due to the increase in their cross-sections and in the luminosity of the collisions (both affecting background processes as well). However, this will also increase the number of particles produced in the collisions and the detectors will have therefore to extract the relevant information from a harsher environment. The inelastic collisions of two protons can be described as the interaction of one parton from each proton (see section 1.2.1). The interactions of the remaining constituents of the proton form the underlying event. During the crossing of the proton bunches, several inelastic collisions may also occur, which contribute to what is called in-time pile-up. The other contribution to pile-up is due to collisions in adjacent bunch crossings and is called out-of-time pile-up. During 2016, the average number of pile-up interactions recorded by the CMS detector was 27 but the peak pile-up reached up to 50 simultaneous collisions. The higher pile-up environment with respect to Run 1 has been partially compensated by upgrades of the detectors but the analysis techniques had also to be optimized to mitigate the influence of pile-up and underlying events.

2.2 The CMS experiment

The Compact Muon Solenoid (CMS) experiment is one of the two multipurpose detectors collecting data from the LHC collisions, the other being the ATLAS (A Toroidal LHC AppartuS) experiment. A full description of the CMS detector can be found in [103]. The

detector has a cylindrical geometry visible in Fig. 2.3. It is the heaviest detector at the LHC, with a mass of 14,000 t, a length of 28.7 m and a diameter of 15 m. The compactness of the CMS detector (half the length of ATLAS) is mainly due to its superconducting solenoid, which produces a magnetic field of 3.8 T. The design of the detector and its performances have been determined to achieve the LHC physics programme [108].

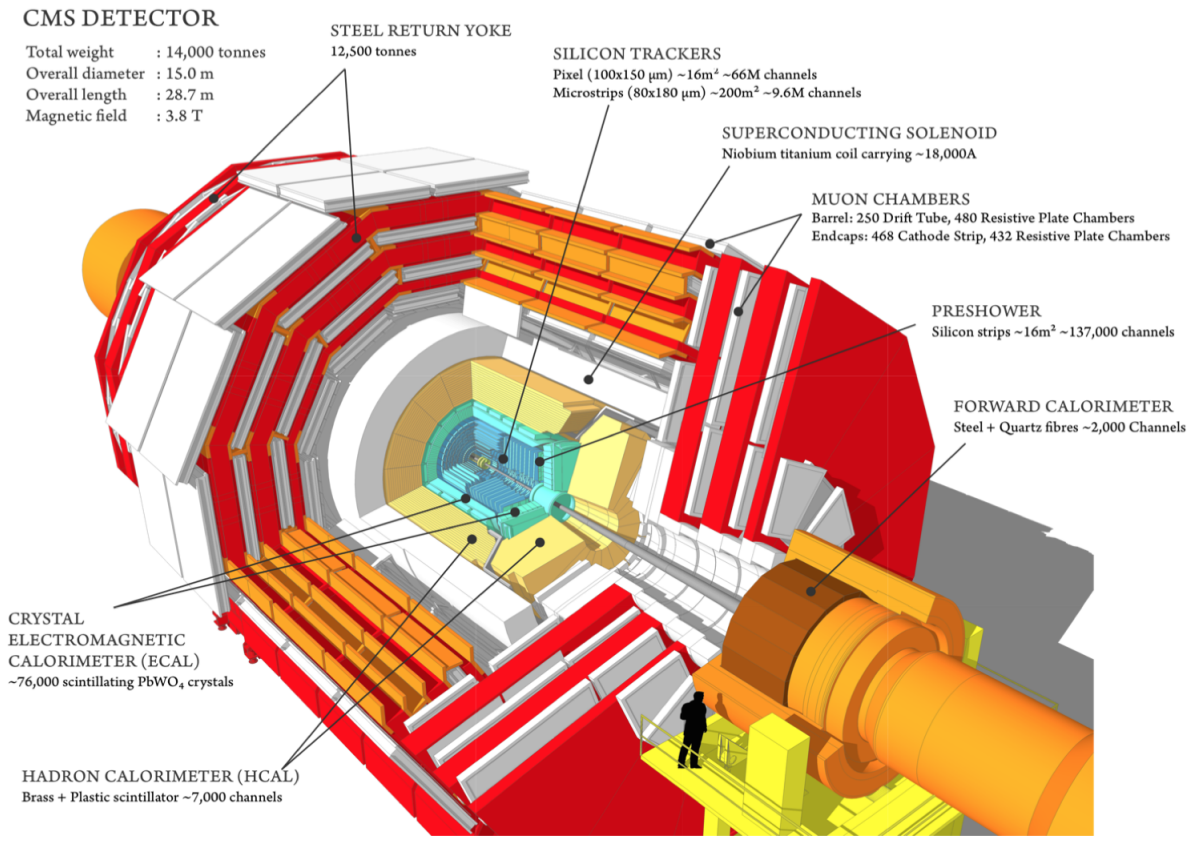

Figure 2.3: General view of the CMS detector

2.2.1 Design constraints

One of the main goals of the LHC when it was designed was the discovery of the Higgs boson, the last missing piece of the Standard Model at that time. Since its mass was very loosely constrained, the search for the Higgs boson had to span a large mass range, over which the branching ratio of the Higgs boson can significantly vary (see section 1.2.3). The CMS detector has therefore been designed to measure precisely as many decay channels as possible.

In order to be sensitive to the $H \to b\bar{b}$ and $H \to \tau\bar{\tau}$ decays, a tracker system with high performance was needed in order to reach an efficient reconstruction of all the charged

particles originating from the b quark and τ lepton decays. The $H \to \gamma\gamma$ search, which had to cope with an overwhelming continuous background, required a granular electromagnetic calorimeter with a very high energy resolution. Finally, the decay channels involving final states with leptons (e.g. $H \to ZZ \to 4\ell$, $H \to WW \to 2\ell 2\nu$) depended on an excellent lepton reconstruction, identification and isolation.

A good lepton reconstruction is also mandatory to searches of Standard Model extensions at the TeV scale. Indeed, some of these models predict the existence of new gauge bosons (e.g. Z', W') which can decay leptonically. Because of the high mass of these resonances, they decay into very energetic leptons, which requires a linear response of the electromagnetic calorimeter up to an energy of $\mathcal{O}(1\text{ TeV})$. Moreover, the presence of a very intense magnetic field of 3.8 T enables a good muon momentum resolution and unambiguous charge determination up to $p_T^\mu \sim 1$ TeV.

Other theories beyond the Standard Model (e.g. supersymmetry) predict the existence of weakly interacting particles. If they are produced in collisions, they will be associated to a significant amount of missing transverse energy $\not{E}_T$. Since this quantity is evaluated from the momentum conservation in the transverse plane, it is important to have the most accurate reconstruction as possible for every interacting particle. A precise measurement of $\not{E}_T$ can be achieved thanks to a hermetic detector with a good jet energy resolution.

The high instantaneous luminosity $\mathcal{L} \approx 10^{-34}\,\text{cm}^{-2}\text{s}^{-1}$ in collisions, necessary to have access to very rare processes, represents also a constraint in the detector design. With LHC collisions produced every 25 ns, a very fast trigger system and detector response are required to be resilient to out-of-time pile-up. In-time pile-up also produces a lot of soft particles which can degrade the detector performance, due to the energy density varying from collision to collision. In order to mitigate the influence of in-time pile-up, a very segmented tracker and high-granularity calorimeters have been built. Finally, the large flux of particles in the forward region, due to the underlying events, had to be taken into account into the design of the subdetectors in that region to resist to the radiation damage.

2.2.2 CMS geometry

The coordinate system adopted in CMS is represented in Fig. 2.4. The origin of that system is located at the interaction point, in the center of the detector. The x-axis points radially towards the center of the LHC while the y-axis is vertically oriented towards the surface. The z-axis is then defined to coincide with the beam axis, to form a direct Cartesian coordinate system. The azimuthal angle $\phi \in [-\pi, \pi]$ is measured in the (x, y) plane (transverse plane) starting from the x-axis. The polar angle θ is measured from the z-axis.

The rapidity ζ is defined as

$$\zeta = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.3}$$

which is linear under a longitudinal Lorentz boost. This motivates the use of the pseudo-

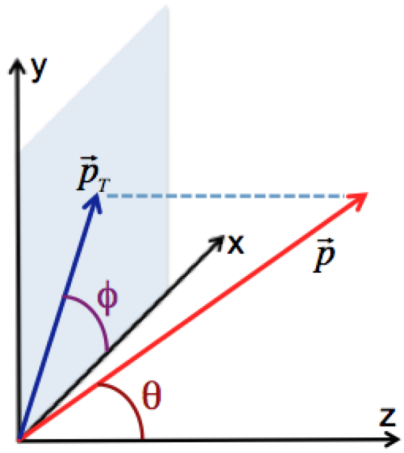

Figure 2.4: CMS coordinate system [109]

rapidity coordinate η defined as as

$$\eta = -\ln\tan\frac{\theta}{2} = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right) \tag{2.4}$$

For ultra-relativistic particles with $E \gg mc^2$, the pseudo-rapidity is approximately equal to the rapidity. The transverse momentum $\vec{p}_T$ is defined as the projection of the momentum in the transvers plane (x, y). Its magnitude is given by

$$p_T = \sqrt{p_x^2 + p_y^2} = \frac{p_z}{\sinh\eta} \tag{2.5}$$

The distance ΔR between two particles with a pseudo-rapidity difference $\Delta\eta$ and an azimuthal angle difference $\Delta\phi$ is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.6}$$

The CMS detector is divided into a central part (the barrel), which covers the region up to $|\eta| \lesssim 1.5$, and two forward regions (the endcaps), with $1.5 \lesssim |\eta| \lesssim 3$. The geometry of the detector is essentially driven by the solenoid, which produces the 3.8 Tesla magnetic field required to achieve a high momentum resolution and an efficient charge identification of the particles produced in the collision. In the barrel, the magnetic field is parallel to the direction of the beams. The tracker, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL) are located inside the magnet coil, while the muon chambers are the outermost part of the detector, outside of the coil. The muon chambers are interleaved with a return yoke, made up of iron slabs, which contain and guide the magnetic field.

2.2.3 Tracking system

The first layer crossed by the particles produced in the collisions is the tracker. This subdetector has therefore to deal with a very high flux of particles and was designed to sustain high radiation levels, a high granularity and a quick response, in order to reconstruct precisely the primary vertices (the proton interaction points) and the trajectories of every charged particle. The tracker can also identify secondary vertices along the tracks, a characteristic feature used to identify jets originating from b quarks (see section 2.3.4).

The CMS tracker forms a hermetic subdetector of 5.8 m long and 1.2 m of diameter and is entirely based on silicon detectors, which detect the charged particles via the ionization of their silicon cells. The trajectories are then built from the different hits and their associated momentum is determined from their curvature. The tracker is divided into two subsystems: the pixel detector and the silicon strip tracker (see Fig. 2.5).

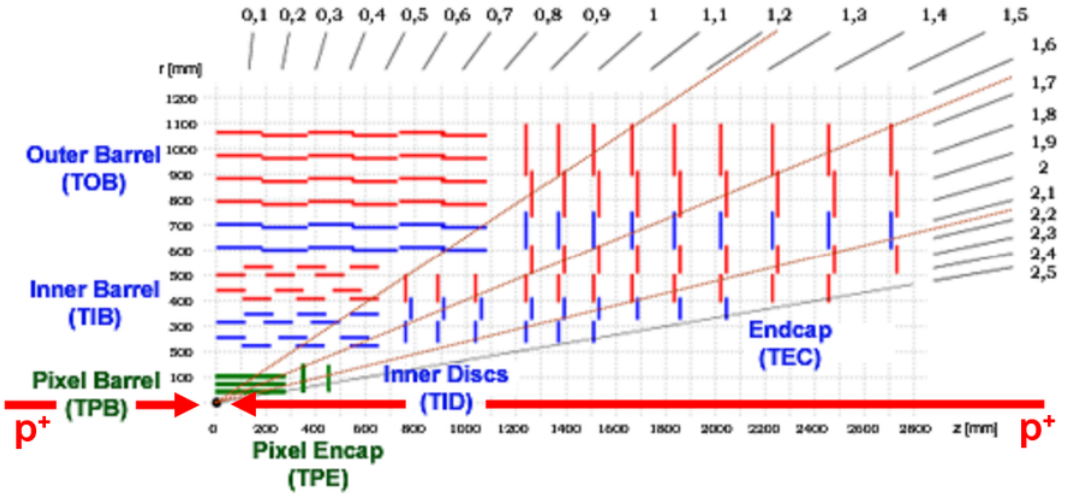

Figure 2.5: Longitudinal view of the 2016 CMS tracking system [110]

Until 2016, the pixel detector in the barrel was made of three 53-cm-long cylindrical layers (Tracker Pixel Barrel), located at r=4.4, 7.3 and 10.2 cm from the beam, while two vertical discs closed off each side of the pixel detector (Tracker Pixel Endcap), at $|z|$=34.5 and 46.5 cm between r=6 and 15 cm. Altogether, those layers represent a total of 65 million silicon channels, measuring 100x150 μm. This granularity enables to achieve a resolution of 10(20) μm in the (x, y)-plane (along the z-axis). At the beginning of 2017, the pixel detector has been upgraded to improve the efficiency and resolution on tracks and allow recovery of degradation in outer tracker layers. Its geometry has thus been modified and a fourth additional layer in the barrel and an additional third disk in the endcaps have been added [111].

The silicon strip tracker consists of up to ten layers of micro-strip sensors. The Tracker Inner Barrel (TIB) is formed of the four innermost cylindrical layers, which extend up to r=55 cm and $|z|$=65 cm and have a variable pitch between 80 and 120 μm. Each side

is enclosed by three Tracker Inner Discs (TID), between $|z|$=65 cm and 110 cm with a variable pitch between 100 and 140 μm.

Those two subsystems are surrounded by six concentric layers forming the Tracker Outer Barrel (TOB), up to r=116 cm with a pitch going from 120 to 180 μm. While the silicon strips of the TIB and the TID are 320-μm-thick, those in the TOB have a thickness of 500 μm. The spatial resolution in the transverse plane achieved in the TIB varies between 23 and 34 μm, which increases to 35-52 μm in the TOB. Along the z-coordinate, the spatial resolution is of the order of 10 times larger.

The tracker is completed by nine Tracker EndCap (TEC) discs on each side, covering the region $|\eta| < 2.5$ with $120 < |z| < 180$ cm. Among those nine discs, the five innermost are 320-μm-thick while the four outermost are 500-μm-thick with a variable pitch between 100 and 180 μm.

2.2.4 Electromagnetic calorimeter

One of the characteristic features of the CMS detector is its homogeneous crystal-made electromagnetic calorimeter (ECAL). Its main goal is to measure the energy of photons and electrons, by collecting all the energy deposit from the induced electromagnetic showers. The ECAL consists of 75,848 crystals of lead tungstate ($PbWO_4$) arranged in a quasi-projective structure, as seen in Fig. 2.6. The lead tungstate was chosen because of its high density (8.3 g/cm^3), its short radiation length (X_0 = 0.89 cm) and Molière radius (R_M = 2.2 cm), which enables to achieve a high compacity for the ECAL. This material also benefits from a short scintillation decay time, which ensures that 80% of the scintillation light is emitted within the 25 ns interval between two bunch crossings.

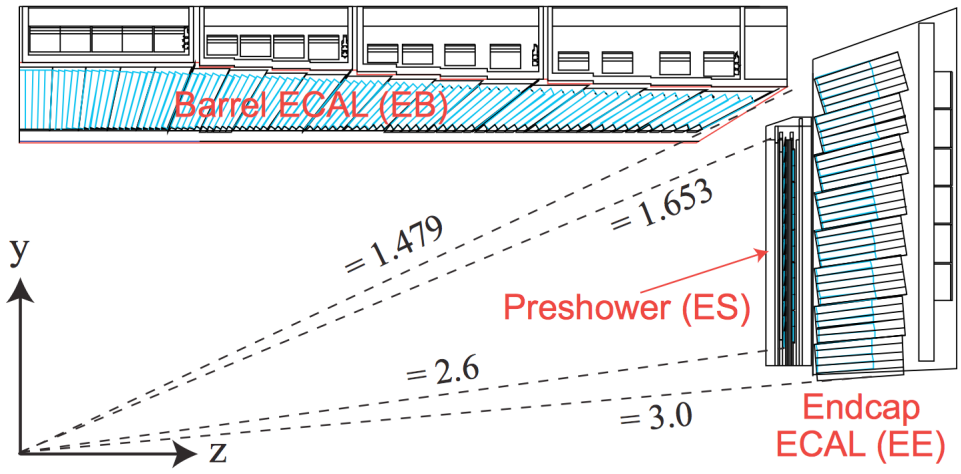

Figure 2.6: Longitudinal view of the CMS electromagnetic calorimeter [112]

The ECAL is organized in a barrel (EB) and two endcaps (EE), which cover the regions $|\eta| < 1.48$ and $1.48 < |\eta| < 3$ respectively. The barrel counts 61,200 crystals, divided

between 36 supermodules of 3t, with 1,700 crystals each, while 14,648 further crystals are located in the endcaps. Each crystal has a trapezoidal shape and covers a solid angle of 0.0174×0.0174 in (η, ϕ). Their length varies between 22 and 23 cm in the barrel and in the endcaps, which corresponds to around 25 radiation lengths X_0.

In the endcap region, the ECAL is completed by a preshower sampling calorimeter (ES), consisting of alternating layers of lead and of silicon strips, in front of the calorimeter endcaps. Since the hadronic activity is particularly high in that region, the goal of the ES is to distinguish the high-energy photons from the close pairs of low-energy photons, produced by π^0's decays.

Photodetectors are used to detect the scintillation light and amplify this signal. In the barrel, silicon avalanche photodiodes (APD) are used, while vacuum phototriodes (VPT) are deployed in the endcaps, because of the higher radiation level and magnetic field. Moreover, the light yield of the crystals is very sensitive to the temperature and a cooling system is therefore required to extract the heat from the readout electronics and keep the crystal temperature stable around $18 \pm 0.05°$C.

Although radiation-hard, the crystal transparency varies with the radiation level because of the nuclear interactions between the material and the ionizing particles. A monitoring system using laser light is therefore used. The laser monitoring is sollicitated during the abort gap of the LHC (2.6 μs) to measure the transparency change in real time. It is used to derive time-dependent calibration constants.

The intrinsic energy resolution of the ECAL barrel, measured in test beams, has the following dependence

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.7}$$

where S (2.8% GeV$^{1/2}$) is the stochastic term, N (120 MeV) the noise term and C (0.3%) the constant term [103]. The stochastic term takes into account the fluctuations in the number of produced photo-electrons, while the noise term is coming from electronics noise and pile-up events. The constant term is related to the non-uniformity of the crystal response and the error for crystal intercalibration of the calorimeter. As seen in Fig. 2.7, the energy resolution is smaller than 1% for electrons with an energy higher than 15 GeV, while it goes down to 0.6% for electrons above 40 GeV.

2.2.5 Hadron calorimeter

The hadron calorimeter (HCAL) is the last subsystem of the CMS detector located inside the solenoid coil. It measures the energy deposited by neutral and charged hadrons and helps determine the missing transverse energy (see section 2.3.7). Unlike the homogeneous ECAL, the HCAL in CMS is a sampling calorimeter. Non-ferromagnetic materials (stainless steel and brass) have been used for the supporting structure and as absorbers, because of the intense magnetic fields in the detector. As for the active medium, tiles of plastic scintillators have been used and piled-up into quasi-projective towers. The scintillation

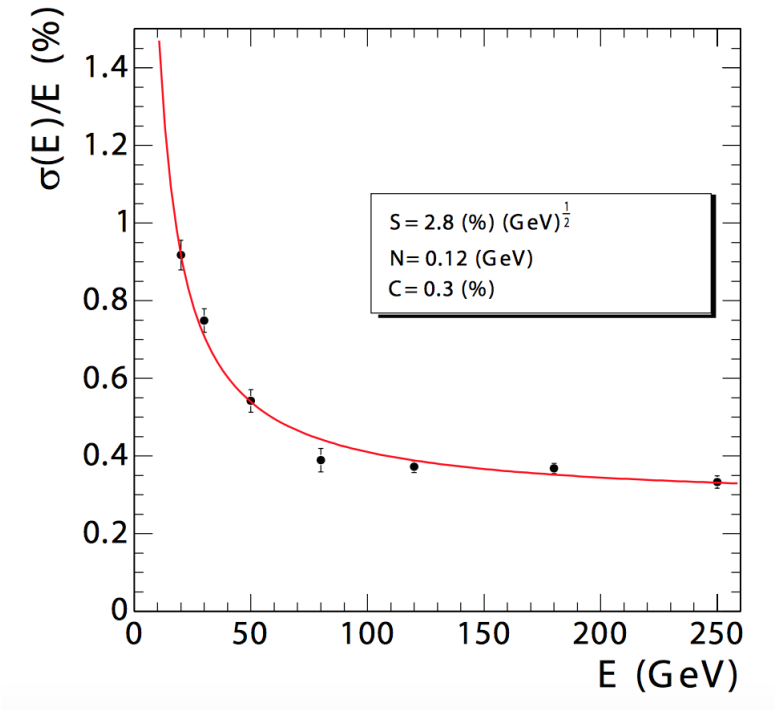

Figure 2.7: Intrinsic energy resolution of the ECAL as a function of the electron energy (measured in a beam test) [103]

light produced by the hadronic showers is channeled by optic fibers and converted into an amplified analogical signal by hybrid photodiodes (HPD).

The HCAL is divided between a barrel region (HB) and two endcaps (HE), as seen in Fig. 2.8. The barrel covers the region $|\eta| < 1.4$ and is segmented into 36 wedges of 26 t each. In order to avoid energy leak from the tails of the hadronic showers, an additional layer of active material (HO) is located outside the coil for $|\eta| < 1.26$, inside the muon system, which ensures a calorimeter depth of approximately 11 interaction lengths. The endcaps count 36 additional wedges between $|\eta| = 1.3$ and $|\eta| = 3$. The size of the towers vary between $\Delta\eta \times \Delta\phi = 0.087 \times 5°$ in the barrel and $\Delta\eta \times \Delta\phi = 0.302 \times 20°$ in the endcaps.

The barrel and the endcaps are completed by two hadronic forward calorimeters (HF), located at 11.2 m from the interaction point and covering the region $3 < |\eta| < 5$. Because of the very intense hadronic activity in that region, they have to be more radiation-hard than in the barrel and the endcaps. The materials used in the two sampling HF calorimeters are steel absorbers and embedded quartz fibers, which channel the light produced by Cerenkov radiation to photomultiplier tubes (PMT). The HF calorimeters count 18 wedges of 24 towers, whose size is $\Delta\eta \times \Delta\phi = 0.171 \times 0.171°$.

The energy resolution of the ECAL and the HCAL together has been measured from test beams with pions between 5 and 300 GeV. A reweighting of the ECAL and the HCAL

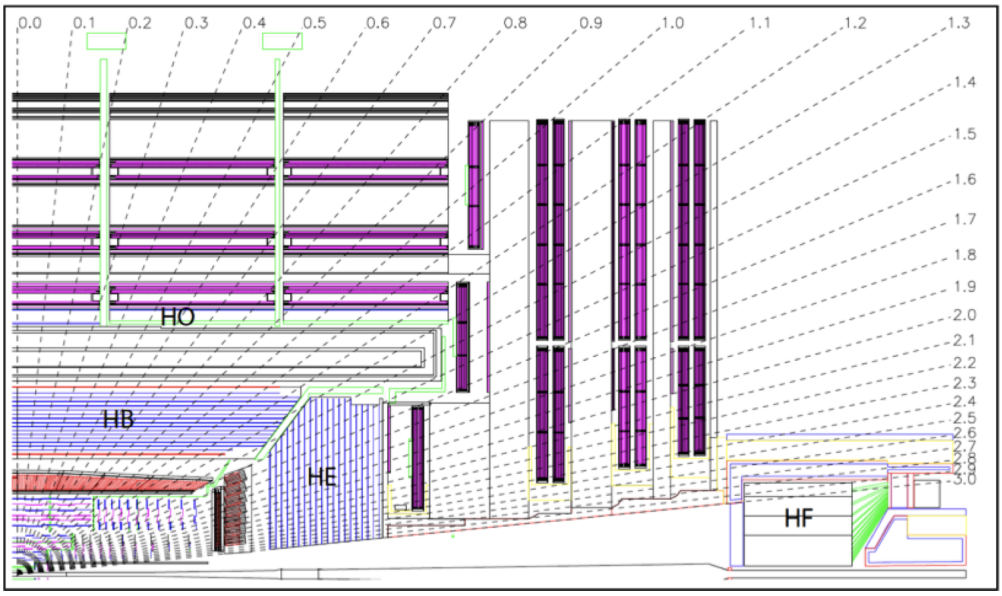

Figure 2.8: Longitudinal view of the CMS hadron calorimeter, segmented between barrel (HB), endcaps (HE), outer barrel (HO) and forward (HF) regions [103]

contributions in the barrel is applied to take into account differences in the responses for electrons and hadrons. The measured resolution is then dominated by the HCAL contribution and is given by [113]

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2 \tag{2.8}$$

2.2.6 Muon detectors

The outermost subdetector in CMS consists of the muon detectors located outside the magnet coil. The muons are indeed the only particles able to go through the whole detector (along with the undetected neutrinos), which makes their identification relatively straightforward. Different kinds of detector are part of the muon system, as seen in Fig. 2.9. The drift tubes (DT) are located inside the barrel, up to $|\eta| < 1.2$, in which the muon flux and the neutron noise are relatively low. In the endcaps, for $1.9 < |\eta| < 2.4$, Cathode Strip Chambers (CSC) are able to deal with the higher radiation level and magnetic field. Resistive Plate Chambers (RPC) are used in addition in the whole detector to ensure redundancy for muon triggers.

The barrel part of the muon system is longitudinally segmented into five dodecagonal wheels with a length of 2.7 m, each of the 12 sectors having a 30° opening. Four DTs stations and 6 RPC layers are interleaved with the three iron slabs of the return yokes of the magnet. The first DT station is located at $r = 4.5$ m, while the outermost one is

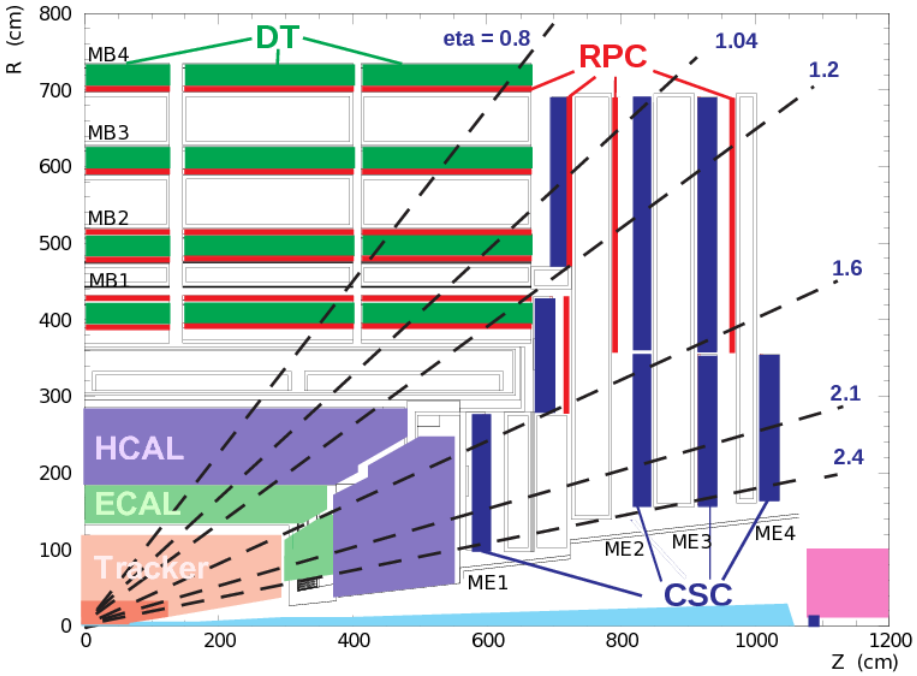

Figure 2.9: Longitudinal view of the CMS muon system, divided between drift tubes (DT), Cathode Strip Chambers(CSC) and Resistive Plate Chambers (RPC) [114]

at $r = 7$ m from the beam axis. Each DT station includes 2 or 3 *superlayers*, each one being made of four stacked layers of DT chambers. Those gaseous chambers, filled with a mixture of argon and CO_2, can record tracks from the drift time of avalanche electrons left by muon crossings.

In the three innermost DT stations, two superlayers have their wires parallel to the beam axis, to measure the trajectories along the (r, ϕ)-coordinates, whereas the third one has its wires orthogonal, to measure the trajectories along the (r, z)-coordinates. The last DT station has only two superlayers with wires installed along the beam axis. The three-dimensional trajectories of the muons can then be reconstructed from two-dimensional segments measured with different superlayers. The resolution achieved with the DTs is better than 1 mrad in ϕ and 100 (150) μm in position in the (x, y)-plane (along the z-axis).

The endcaps are both arranged into four disks perpendicular to the beam axis from $z = 7$ to 10 m, with in total six CSC layers and three RPC layers. Those disks are divided into two or more concentric rings. The CSCs are gaseous chambers filled with a mixture of argon, CO_2 and CF_4 in which closely spaced anode wires are stretched between two radially oriented cathodes. The measurements from the cathodes are used to estimate the (r, ϕ) position, while the anodes are optimized to measure the η-coordinate and the bunch crossing the detected muon comes from (because the ionization electrons can be detected much faster with the anodes than the ions on the cathodes). The CSCs have a position

resolution of 75 to 150 μm along the ϕ-coordinate and 200 μm for the r-coordinate.

The RPCs, present both in the barrel and the endcaps, are double gap chambers, operated in avalanche mode, in order to ensure good operations at high rate. Thanks to their fast response and good time resolution, they play an important role in the trigger system of CMS (see section 2.2.7) and are used to assign the muons to the correct bunch crossing.

As visible in Fig. 2.10, the p_T-resolution for the muons in the barrel with $p_T < 100$ GeV is dominated by the tracker measurement but for higher momenta the muon system response dominates and allows to achieve 10% for 1 TeV muons.

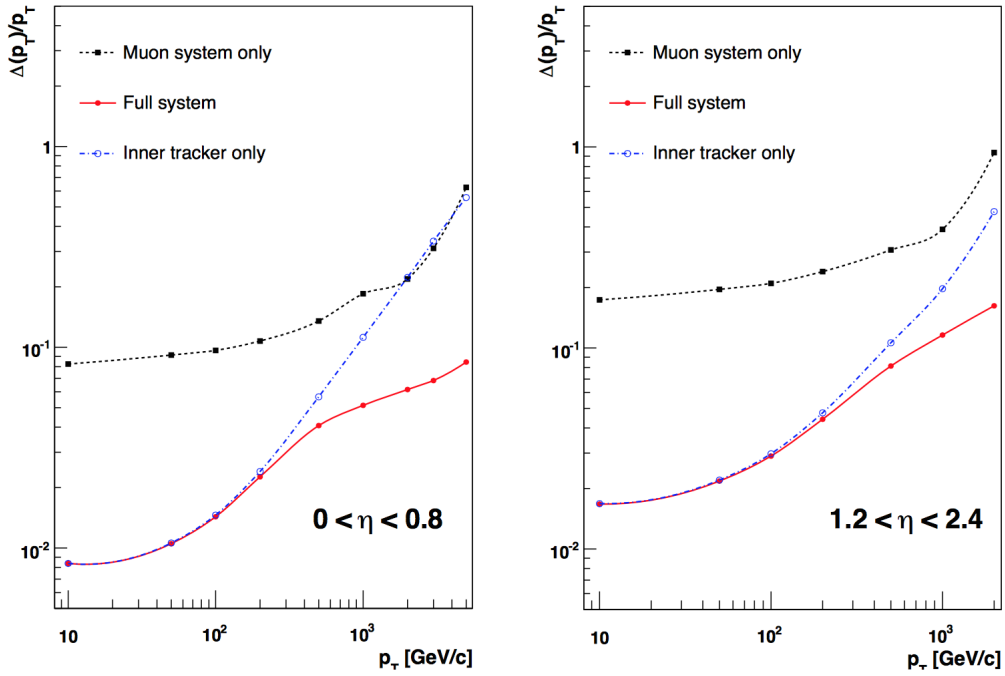

Figure 2.10: The muon transverse-momentum resolution as a function of the transverse-momentum (p_T) using the muon system only, the inner tracking only, and both [103]

2.2.7 Trigger system and data acquisition

With a full detector readout of around 1 Mb/event, it is not technologically possible to store all the data produced every day by the billions of proton collisions in the CMS detector. Hopefully most of those collisions do not correspond to the high-energy processes, like the production of Higgs bosons, but rather to elastic proton scatterings for example. It is therefore possible to reduce the bare collision rate of 40 MHz (for 25 ns bunch spacing) by selecting only events likely to involve interesting physics processes. This task is achieved by the trigger system, which has to quickly take the decision whether to store or not data

associated with a given bunch crossing. In CMS, the trigger system is organized in two successive levels: the Level 1 (L1) trigger and the High-Level Trigger (HLT) [115].

Level 1 trigger

The first level of filtering corresponds to the L1 trigger that is implemented on custom hardware processors located in the service cavern of the detector. It has to provide a rate reduction of $\mathcal{O}(10^3)$, down to an absolute rate of about 100 kHz. The L1 trigger takes as input raw data from the front-end readout electronics and looks for signs of interesting physics, like particles with a high energy.

As the L1 trigger has only 3.8 μs to take a decision, the full offline reconstruction of the physics objects cannot be applied at this level. Instead, the L1 trigger produces *L1 candidates*, corresponding to a coarse-granularity and low-energy-resolution estimate of the particle four-momenta. Because of the timing constraints, which include the fact that the L1 trigger is fully synchronized, iterative procedures like track reconstructions using the tracker information cannot be used at Level 1. Therefore the tracker and the preshower detector are not used in the L1 trigger, because their read-out do not have enough latency, and the L1 candidates are only reconstructed using calorimeter and muon system information. Note that the trigger upgrades of CMS planned for HL-LHC will however be able to make use of the tracker information within the L1 trigger [116].

Specific algorithms, directly implemented on electronic boards, are used to reconstruct the transverse energy and the position in (η, ϕ) of the different L1 candidates. They can be associated to electromagnetic objects, namely electrons or photons (L1 EG), hadronically decaying τ leptons (L1 Tau), muons (L1 Mu) or jets (L1 Jet). Those candidates are built from energy deposits in the trigger towers of the calorimeters or track segments in the muon chambers. These building bricks are called *trigger primitives* and are used to estimate the direction and the transverse momentum or energy of the candidates. Logical bits are also computed to define isolated candidates, based on the surrounding energy deposit in the calorimeters. They are then sorted according to their energy. Global quantities, such as the missing transverse energy (L1 MET) or the total hadronic transverse energy (L1 HT) are also computed and used for triggering at Level 1. During the Run 2, the L1 trigger system of CMS underwent a major upgrade to deal with new data taking conditions, in which most of the L1 algorithms have been modified. This upgrade will be specifically detailed in Chapter 3.

Different selection criteria, potentially combining different trigger primitives, can then be used to select an event. Those criteria can be for instance to require a jet with E_T larger than 180 GeV or one muon and one τ_h with an E_T larger than 18 GeV and 24 GeV respectively. Some of those selections can be prescaled to deal with high physics rate: in that case, only some fraction of events passing those selections are randomly kept. The set of all the selection criteria forms the *L1 trigger menu*, which can be modified to accommodate the different conditions of data taking in CMS. If an event is accepted by any of the selection algorithms, the associated data is transferred to the HLT.

High-Level Trigger

In order to reduce the rate down to a level of 1 kHz, compatible with the data acquisition capabilities, the HLT has to provide an additional reduction of $\mathcal{O}(10^2)$. Since the input rate has already been reduced by the L1, more sophisticated algorithms closer to the offline reconstruction can be exploited, using a software-based implementation on a computer farm of around 2000 computers. Track reconstruction can for instance be performed and a better estimation of the energy and the position of the trigger primitives can be achieved with a more optimal use of the subdetectors information. With respect to L1, high-level objects, like b-jets, can also be reconstructed and used for the trigger decision. To reduce the processing time, the HLT reconstruction starts around the L1 seeds. This results in average timing of 60 ms per event and a maximum timing of 1 s for a small fraction of events.

2.3 Particle-Flow algorithm and object reconstruction in CMS

One of the main challenges for a particle detector like CMS is the precise measurement of the energy and momentum of the hadronic jets produced in the collisions. This task could in principle be accomplished by the calorimeters only. However, the tracking system can offer a better resolution for the measurements of charged particles, depending on their energy. Only the energy of the neutral component of the jets needs to be determined using the calorimeters. Since the typical energy distribution inside a jet between the charged hadrons, the photons and the neutral hadrons is in a proportion 65:25:10, the jet reconstruction can therefore benefit from a substantial improvement. But for this, the energy deposits in the calorimeters have to be efficiently associated with the different particles, in order to avoid double counting of the charged components of the jets. This is the role of the Particle-Flow algorithm used in CMS [117], which aims at reconstructing each stable particle, to optimize the use of the information of each subdetector.

The Particle-Flow algorithm in CMS relies on its performant silicon tracker, which is able to separate the tracks in collision event with high number of charged particles. This is an essential feature required for the assignement of the calorimeter energy deposits. Benefiting from the 3.8 Tesla magnetic field, the momentum of the charged particles can also be better measured from a higher-precision track reconstruction than with the calorimeters. In the case of muons, the Particle-Flow algorithm can also rely on the complementary information from the muon detectors. The high transverse granularity of the calorimeters is also exploited to reach an effective track-to-cluster association.

The Particle-Flow algorithm produces in the end a list of reconstructed particles with their associated momentum. They can then be used as inputs for high-level algorithms which can reconstruct jets (including their substructures), hadronically decaying τ leptons or discriminate b-jets from light-quark jets.

2.3.1 Building bricks

In order to reconstruct the individual particles with the Particle-Flow algorithm, the tracker, calorimeter and muon detectors signals (*hits*) have first to be processed into higher-level objects. The tracker hits compatible with the same helix hypothesis are combined using the Combinatorial Track Finding (CTF) algorithm [118]. This algorithm reconstructs tracks with a four-step iterative procedure:

- The seed generation looks for initial track candidates with 2 or 3 hits, which give an initial estimate of the trajectory and its uncertainty.
- The track finding, based on a Kalman filter [119], extrapolates the trajectory from the initial track, consistently with the motion of a charged particle in a magnetic field. The subsequent hits consistent with the extrapolated helix are then associated to the track.
- A track fitting is performed to improve the determination of the trajectory parameters and smooth the track path.
- A track selection is then performed using a quality criteria based on the χ^2 of the fit, the number of hits and the number of missing hits in the track.

At each iteration, the hits associated to a reconstructed track are removed, which reduces the combinatorics and the fake rate. This algorithm is also applied to reconstruct tracks from the muon chambers hits. With this procedure, the track reconstruction has an overall efficiency of 99% for isolated muons and 90% for charged hadrons in jets.

In the calorimeters, the hits are gathered to build *Particle-Flow clusters*. The clustering is performed indepently in the ECAL and the HCAL and in the barrel and the endcaps, as well as in the first and second layer of the ES. (No clustering is done in the HF.) First, local energy maxima above some energy threshold are used to define seeds for the clustering. The neighboring cells with an energy above a certain clustering threshold (80 MeV in the ECAL barrel, up to 300 MeV in the ECAL endcaps and 800 MeV in the HCAL) are then aggregated to build topological clusters. Several seeds can be included into the same topological clusters and an iterative sharing of the cell energy inside those clusters is therefore performed. This leads finally to the Particle-Flow clusters which are interpreted as the energy deposits associated to single particles. In Fig. 2.11, the clusters resulting from a simple hadronic jet are represented.

2.3.2 Link algorithm

To reconstruct individual particles, the Particle-Flow algorithm must then combine the previous building bricks, in an unambiguous way, avoiding possible double counting. A link algorithm assigns therefore a link distance for each pair of elements and connect them into blocks, which combine therefore information from the different subdetectors.

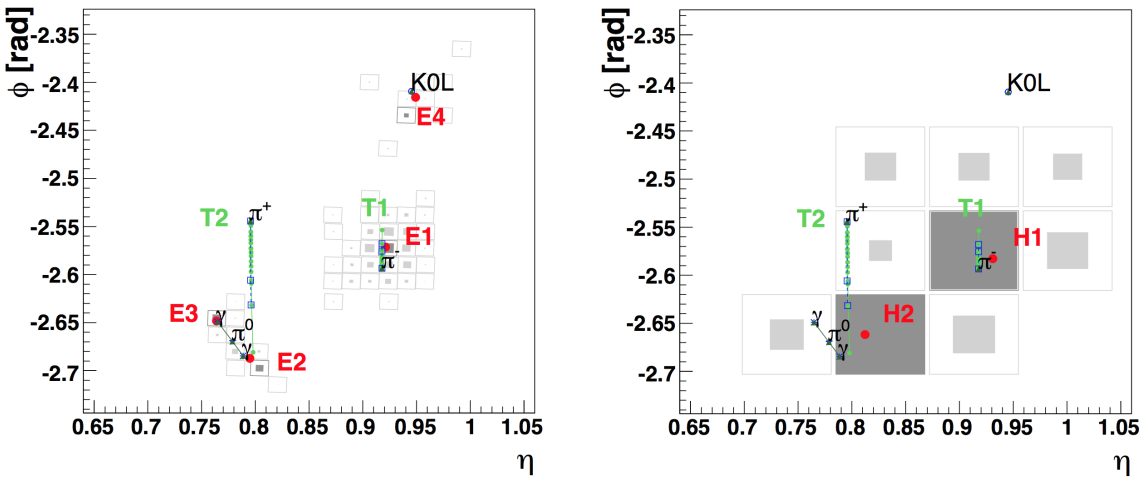

Figure 2.11: Event display in $\eta - \phi$ view of a simple hadronic jet as reconstructed in the ECAL (left) and b) in the HCAL (right). This jet consists of two charged pions (associated to tracks T_1 and T_2), a π^0 (decaying to two photons) and a K_L^0. Only the π^-, the two photons and the K_L^0 deposit some energy in the ECAL, which are gathered into three topological clusters. This leads after sharing to four Particle-Flow clusters E_1-E_4. In the HCAL, only the charged pions deposit their energy inside a single topological cluster, which can be divided into two Particle-Flow clusters H_1-H_2. [117]

To compute the link between a charged particle track from the tracker and a calorimeter cluster, the track is extrapolated from its last measured hit to the depth where the shower is expected to happen (to the two layers of the PS, to the electromagnetic shower maximum for the ECAL and to one interaction length for the HCAL). If the track lies inside the boundaries of a Particle-Flow cluster (potentially enlarged to take into account the detector geometry or multiple scattering for low-momentum particles), this cluster is linked to the track and the link distance is defined as the ΔR between the track and the cluster. To collect Bremsstrahlung photons emitted by electrons, additional ECAL clusters can be linked to a charged particle track, based on the extrapolation of the track tangents to the ECAL from each tracker layer intersection. If one of the tangents lies inside the cluster limits, it is linked to the track.

Similarly, two calorimeter clusters (one PS and one ECAL cluster or one ECAL and one HCAL cluster) are linked if the position of the cluster from the more granular calorimeter (the PS or the ECAL) is inside the boundaries of the cluster from the coarser one (the ECAL or the HCAL). In that case, the link distance is again defined as the ΔR between the two clusters.

Finally, a link between a charged particle track from the tracker and a muon track from the muon chambers is created if the χ^2 of the combined fit between the two tracks is small enough. If several global muons can be reconstructed from a given muon track, the tracker

track leading to the smallest χ^2 is chosen. The link distance is then defined as the χ^2 of the combined fit.

2.3.3 Particle reconstruction and identification

The next step of the Particle-Flow algorithm consists in the assignation of the blocks built after the linking step to single particles. During that procedure, the elements univocally assigned are progressively removed to reduce the combinatorics, until no element is left.

Muons

Muons are the first particle to be identified, because of their high identification efficiency. The muon reconstruction in CMS [120] uses two approaches to combine the tracks from the tracker (*tracker track*) and from the muon systems (*standalone-muon track*):

- **Global Muon reconstruction** (*outside-in*). For each standalone-muon track, a matching tracker track is looked for by comparing the extrapolated tracks parameters. A *global-muon track* is obtained with a combined fit of the hits associated to the two tracks. For muons with $p_T > 200$ GeV, that combined fit can improve the momentum resolution with respect to the tracker-only fit.
- **Tracker Muon reconstruction** (*inside-out*). Each tracker track with $p_T > 0.5$ GeV and $p > 2.5$ GeV is extrapolated to the muon systems. If at least one muon segment matches the extrapolated track, this track is considered as a *tracker-muon track*.

The Tracker Muon reconstruction is more efficient for low-momentum muon ($p_T < 5$ GeV) because it only requires a single muon segment in the muon chambers, while the Global Muon reconstruction targets high-p_T muons with at least two segments reconstructed in the muon chambers. Thanks to the high reconstruction efficiency of the tracker and that of the muon systems, more than 95% of the muons produced in the geometrical acceptance of the muon chambers, with a sufficiently large momentum, are reconstructed as a Global or a Tracker Muon (often as both). Standalone-muon tracks alone are typically not used in the muon reconstruction because of their worse momentum resolution and the high contamination from cosmic muons.

The reconstructed muons are then classified into three categories in the Particle-Flow reconstruction [121]. *Isolated muons* corresponds to global muons for which the sum of the p_T of the tracks and of the transverse energy of the calorimeter hits within a cone size of $\Delta R = 0.3$ centered around the muon is less than 10% of the muon p_T. Among the remaining reconstructed muons (which are typically included inside jets), the *tight muons* are then identified requiring a minimal number of hits in the muon track and compatibility of the track with the linked calorimeter clusters, according to a template derived from simulation. The tracks associated to isolated or tight muons are then removed from the blocks.

The few muons not identified as isolated or tight muons can still be identified as *loose muons*, once all the other particles have been reconstructed by the Particle-Flow algorithm. If they have a track momentum much larger than their associated energy deposit in the calorimeter, they will indeed not be reconstructed as charged hadrons. For those loose muons, the requirement on the number of hits is relaxed and a matching between the extrapolated tracker track and the hits in the muon system is required.

As shown in Fig. 2.12, the muons reconstructed in CMS can be used to reconstruct the different resonances in the di-muon mass spectrum in a range from less than 1 to several hundreds of GeV.

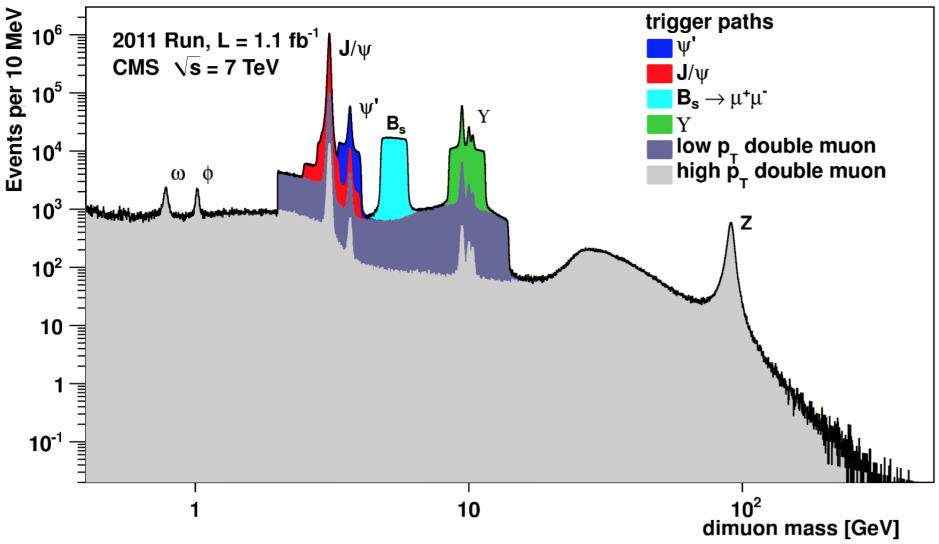

Figure 2.12: Invariant mass spectrum of opposite-sign muon pairs collected with different triggers in CMS [122]

Electrons

The next particles to be reconstructed are the electrons. The main challenge regarding the electron reconstruction in CMS is related to the production of Bremmstrahlung photons along the electron track. An electron can indeed radiate a large amount of photons in the tracker. At $\eta \approx 0$, an electron has thus lost 33% of its energy on average due to Bremmstrahlung before it reaches the ECAL and this fraction can go up to 86% at $\eta \approx 1.4$, where the amount of material budget is the highest. The electron reconstruction has to be able to assign correctly the associated energy deposits in the ECAL, otherwise this algorithm may result in a degradation of the energy resolution not only for the electrons, but for the jets and the missing transverse energy as well.

Two approaches are followed for the electron reconstruction in CMS [123]. The first one, better suited for high-p_T and isolated electrons, uses ECAL clusters as seeds for the electron

track finding. Large energy deposits in the ECAL (*superclusters*) are built by gathering hot cells around local energy maxima, in a small window in η and in an extended window in ϕ (of around ± 0.6 rad). The details of the clustering algorithm differ in the barrel and the endcaps because of the different geometry of the crystals. A large extension in ϕ is used by the supercluster algorithm are able to collect most of the Bremsstrahlung photons produced by the initial electrons, as seen in Fig. 2.13.

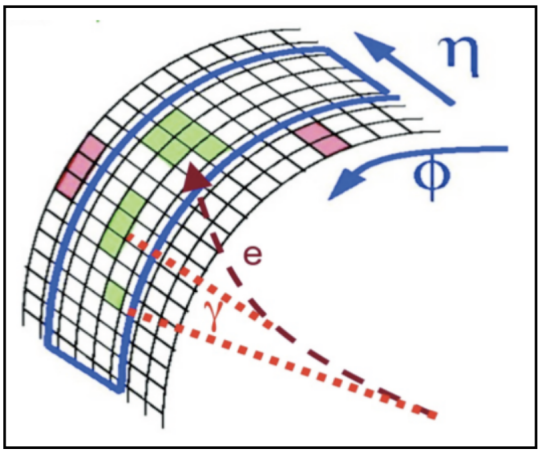

Figure 2.13: Typical energy deposit topology of an electron radiating Bremsstrahlung photons. The squares represent the ECAL crystals while the green ones are those associated with a significant energy deposit: the electron and the Bremsstrahlung photons produce each a cluster of crystals. The reconstruction procedure aims at gathering those clusters to form a supercluster [124].

The energy-weighted center of the supercluster is then assumed to be the position on the ECAL surface of an electron with the energy of the supercluster which would have produced no Bremsstrahlung. From that position, two helices (one for each charge hypothesis) are propagated back to the interaction point. If some tracker seed is compatible with those extrapolations, a dedicated track reconstruction algorithm based on a Gaussian sum filter (GSF) [125] is used. Unlike the Kalman filter algorithm, this track reconstruction is able to follow the change of curvature of the tracks due to the Bremsstrahlung radiation.

As the GSF algorithm is relatively time-consuming, the superclusters used as seeds for the ECAL-based approach must not be backed up by an HCAL cluster whose energy is more than 15% of the supercluster energy. Because of that selection, the reconstruction efficiency for electrons in jets is relatively limited with that approach alone. Low-p_T electrons also fail to be reconstructed that way because the size of the superclusters is too small to contain all of their Bremsstrahlung radiation. In order to recover those electrons, a tracker-based approach is also used for the electron reconstruction, using a loose track preselection. A dedicated search for Bremsstrahlung photons not included in superclusters is also performed by extrapolating track tangents from each tracker layer.

To be reconstructed as electrons by the Particle-Flow algorithms, the electron candidates from the two approaches have to pass in addition identification criteria based on calorimetric and track variables. Those variables are either used to define cut-based working points or are alternatively combined in a multivariate discriminator based on the output of a Boosted Decision Tree (BDT) [126], which has been trained to separate genuine electrons (selected among $Z \to ee$ candidates) from fakes (mostly from pions) [123]. The performance of different electron identifications is compared in Fig. 2.14 with ROC (Receiver Operating Characteristic) curves, which display the identifcation efficiency for electrons as a function of the mistag rate for jets faking electrons.

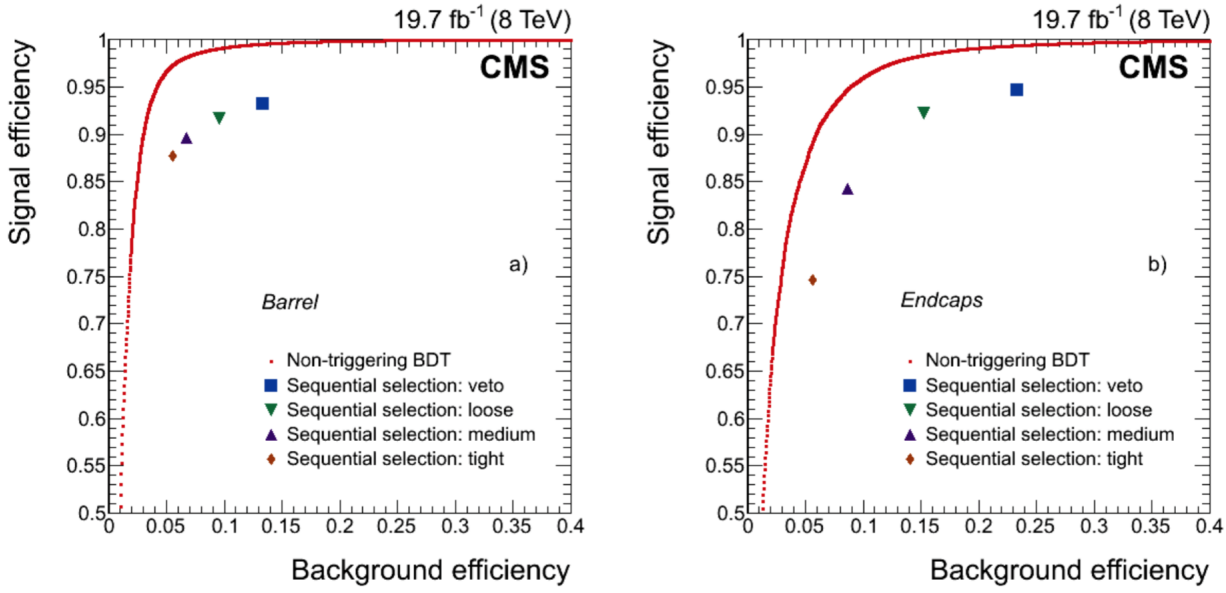

Figure 2.14: Performance of the electron identification, using the BDT discriminant (red curve) or cut-based workind points, in the barrel (left) and in the endcaps (right) [123]

Charged hadrons, neutral hadrons and photons

The remaining tracks are then associated to charged hadrons and linked to ECAL and HCAL clusters. Depending on the energy deposits in the calorimeter (E) compared to the momentum of the track (p), additional neutral particles can be created. For that purpose, a cluster calibration is performed [127], which corrects for the non-linear response of the calorimeters and for threshold effects.

For a given cluster, if its total calibrated energy is smaller than the linked track momenta ($E < \Sigma p$), a search for additional loose muons is performed. The remaining tracks are then removed progressively, starting with the ones with the highest p_T uncertainty, until either $E > \Sigma p$ or no track is left. Each remaining track gives rise to a charged hadron. Its momentum is either obtained from a fit to the track momentum and the cluster energy, if

the calorimetric energy is consistent with the track momentum within its uncertainty, or is equal to the track momentum otherwise. All the charged hadrons are assigned the charged pion mass.

On the other hand, if the track momenta are higher than the cluster energy ($\Delta E = E - \Sigma p > 0$) by more than one standard deviation of the cluster energy, additional neutral particles are created. Photon reconstruction is favored with respect to neutral hadrons, as they carry twice more energy in hadronic jets. If some ECAL cluster is present with an energy E_{em} smaller than the calorimetric excess ΔE, it is used to build a photon with energy $E_\gamma = E_{em}$ and the remaining part of the excess is assigned to a neutral hadron with energy $E_{h^0} = \Delta E - E_{em}$. Otherwise, if $E_{em} > \Delta E$, it gives rise to a photon with energy $E_\gamma = \Delta E$ and the remaining ECAL energy is assumed to come from an early shower of the charged hadron. The remaining clusters, which are not linked to tracks, are then associated to photons for ECAL clusters and to neutral hadrons with the K_L^0 mass for HCAL clusters.

2.3.4 Jet reconstruction

As mentioned in section 1.1.3, when quarks or gluons are produced in hard interactions, they will hadronize and produce bunches of collimated particles called jets. To reconstruct those high-level objects, CMS uses the anti-k_T clustering algorithm [128] with a cone size of $R = 0.4$ (0.5 during Run I). This iterative algorithm takes as inputs the particles reconstructed by the Particle-Flow algorithm and gather them into jets, whose momentum gives an estimate of the initial partons. The anti-k_T algorithm works as follows. For each input object k and for each pair of objects (i, j), a beam-distance d_{kB} and a distance d_{ij} is calculated. They are defined as

$$d_{kB} = 1/p_{Tk}^2 \tag{2.9}$$

$$d_{ij} = \min(1/p_{Ti}^2, 1/p_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{2.10}$$

where p_T is the transverse momentum and $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ the angular distance squared. At a given step, the smallest d_{ij} is compared with the smallest d_{kB}. If $d_{ij} < d_{kB}$ the objects i and j are combined into a single proto-jet with four-momentum $p_{ij} = p_i + p_j$. Otherwise, the object k is removed from the process and is considered as a jet. The anti-k_T algorithm leads usually to jets with an intuitive circular shape, as seen in Fig. 2.15.

The use of the Particle-Flow algorithm to provide the inputs of the clustering algorithm enables to achieve a better resolution on the jet transverse momentum than with *calo-jets*, reconstructed from calorimeter cells only. Indeed, the Particle-Flow algorithm benefits from the precise momentum measurement in the tracker and from the granularity of the ECAL to reconstruct the charged hadrons and the photons, which carry 90% of the overall jet energy. The jet energy resolution has been measured in $Z/\gamma^* \to \mu\mu$ events using a tag-and-probe technique, the probe being a jet recoiling against the di-μ pair, and is presented in Fig. 2.16.

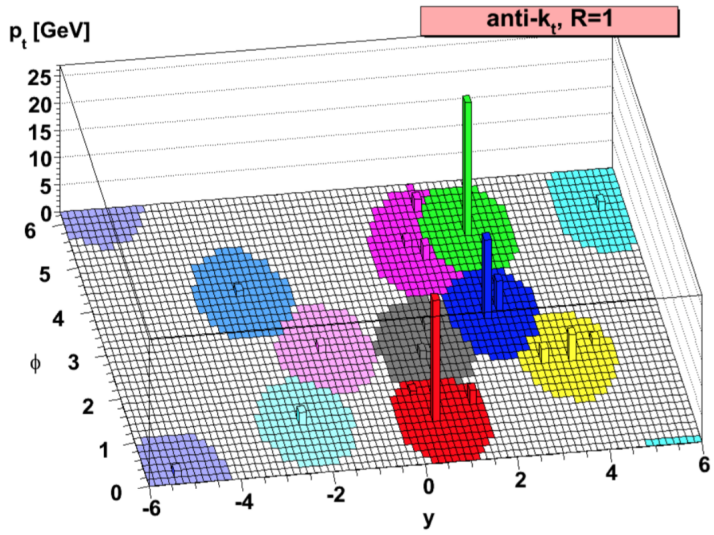

Figure 2.15: Active area of jets reclustered with the anti-k_T algorithm. A parton-level event is overlaid with soft ghost particles to illustrate the effective area of the hard jets. [128]

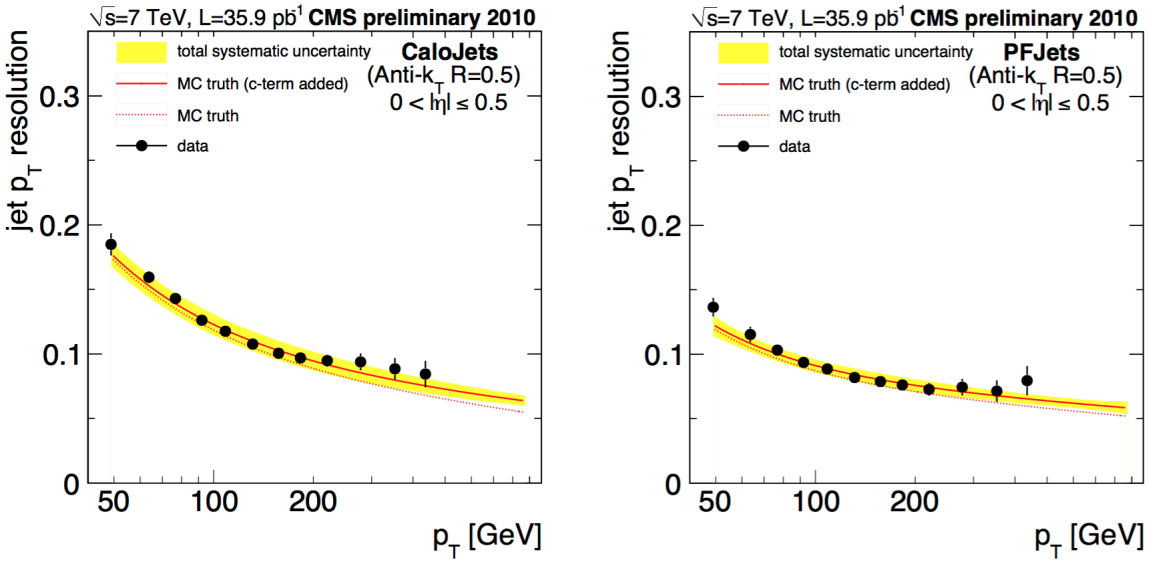

Figure 2.16: Jet transverse momentum resolution measured in Monte-Carlo simulation and 2010 data, for calo-jets (left) and Particle-Flow jets (right) [129]

The raw momentum of the jet, defined as the sum of the momenta of all its constituents, is corrected to match the momentum of the parton which initiated the jet. First an

offset correction is applied to remove the contribution from the pile-up and the detector noise. It is computed using the median transverse momentum density ρ for each event and subtracting the product of ρ times the area of the jet in the $\eta - \phi$ plane [130]. Then an additional calibration, which depends on the kinematics of the jets, is applied in three steps to correct for the response, non–linearity and inhomogeneity of the calorimeters [131].

In collisions with a high level of pile-up, fake jets may be reconstructed from the accidental clustering of neighbouring particles or from the superimposition of soft jets from different interaction vertices. To discriminate them from hard jets initiated by QCD partons, a multivariate discriminant (MVA pile-up jet ID) based on a BDT output can be applied. The BDT inputs include information from the tracks and from the jet shapes [132].

2.3.5 *b*-jet identification

As mentioned in section 1.3.3, b quark hadronization has specific characteristics with respect to light quarks or gluons, which provides a way to tag the jets originating from b quarks. Those jets are identified using the CSVv2 algorithm [133]. This algorithm relies mainly on the tracker information to identify a seconday vertex among the tracks, which is a characteristic feature of the b-jets. Several variables related to that secondary vertex identification are then used to exploit the long lifetime of B hadrons and the higher particle multiplicity and mass of b-jets compared to light-quark and gluon jets. Those variables are combined to build a discriminator value, whose distribution can be seen in Fig. 2.17 for different parton flavors.

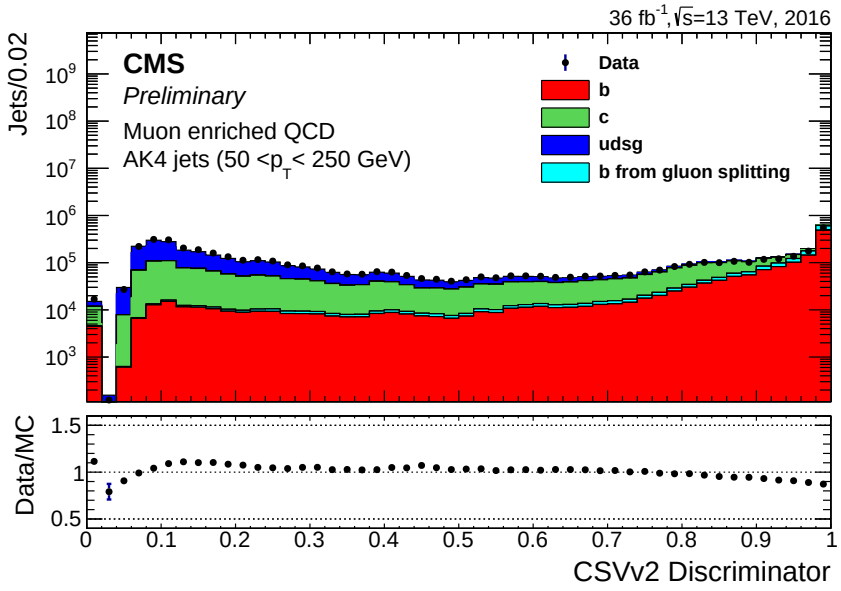

Figure 2.17: Distribution of the CSV discriminator in 2016 data [134]

Different working points can be defined with that discriminator. The associated efficiencies are detailed in Table 2.1. The mistag rates indicated correspond only to jets associated to light flavors (u, d, s quarks and gluons). The mistag rate for c quarks can be significantly higher, as its hadronization presents similar features to the one of a b quark.

Working point	**b-tagging efficiency**	**Light flavor (udsg) mistag rate**
Loose	83%	10%
Medium	69%	1%
Tight	49%	0.1%

Table 2.1: CSVv2 b-tagger efficiencies evaluated in Monte Carlo $t\bar{t}$ events [133]

2.3.6 τ reconstruction and identification

As mentioned in section 1.3.1, hadronically-decaying τ leptons have distinct features with respect to jets produced by quarks or gluons, with typically a significantly lower number of hadrons produced. In order to optimally reconstruct τ_h, dedicated algorithms have been developed. The one used in CMS is the hadrons-plus-strips (HPS) algorithm [135], which takes advantage of the individual particle reconstruction by the Particle Flow algorithm. Starting from the jets clustered with the anti-k_T algorithm with $|\eta| < 2.5$ and p_T >14 GeV (before calibration), different decay mode hypotheses are tested for the τ_h to build its 4-vector and identification criteria are applied in a second step to reduce the fake rate from QCD partons.

The possible decay modes of a τ_h include neutral pions in the final state, which quickly decay into two photons. Those photons have a high probability to be converted into an electron pair within the tracker volume. In order to take into account that possibility, the HPS algorithm clusters photons and electrons from the jet into "strips" in the $\eta - \phi$ plane. These strips are more extended along ϕ to account for the bending of the electron trajectory in the 3.8T magnetic field and their size is adjusted as function of the p_T of the electron or photon to be merged (unlike in Run 1 where their size was independent of the p_T). The clustering is performed iteratively, starting from the electron or photon with the highest p_T. Then, the subleading electromagnetic particle within the strip centered on the seed is added to the strip. The center of the strip is redefined using the direction of the sum of the 4-vectors of the clustered particles. This procedure is repeated until no electron or photon is found inside the strip and a new strip is then created starting from the remaining electromagnetic particles in the jet. In the end, strips with p_T sums of electrons and photons larger than 2.5 GeV are considered as π^0 candidates.

Those strips are then combined with the charged particles in the jet to attempt to reconstruct the different decay mode hypotheses of the τ_h. Only the charged particles with p_T>0.5 GeV and a track compatible with the production vertex of the charged particle with

highest p_T are considered, to reduce the influence of pile-up. Those elements are combined in multiple τ_h hypotheses, corresponding to combination of one or three charged particles and up to two strips. The charged particles are treated as charged pions and are assigned the mass of the π^+ while the electrons from the strips are treated as massless. The charge of a τ_h candidate is defined as the sum of all the particles included in the reconstruction hypothesis, except for electrons in the strip.

Different mass criteria are applied according to the hypothesis considered, consistently with the expected resonance involved in that decay (see section 1.3.1):

- $h^\pm h^\mp h^\pm$: combination of three charged particles with total charge ± 1, with

$$0.8 < m_{\tau_h} < 1.5\ \text{GeV} \tag{2.11}$$

- $h^\pm \pi^0 \pi^0$: combination of one charged particles with two strips, with

$$0.4 < m_{\tau_h} < \min\left(\max\left(1.2\ \text{GeV}, 1.2\ \text{GeV}\sqrt{p_{T\tau_h}/100\ \text{GeV}}\right), 4.0\ \text{GeV}\right) \tag{2.12}$$

- $h^\pm \pi^0$: combination of one charged particle with one strip, with

$$0.3 < m_{\tau_h} < \min\left(\max\left(1.3\ \text{GeV}, 1.3\ \text{GeV}\sqrt{p_{T\tau_h}/100\ \text{GeV}}\right), 4.2\ \text{GeV}\right) \tag{2.13}$$

- $h^\pm$: single charged particle with no strip

Depending on the hypothesis considered, the mass window can be enlarged as the p_T increases to take into account resolution effects.

If an hypothesis fails the mass requirement, it is discarded. This is also the case if a charged hadron or a strip is found within a signal cone of

$$\Delta R = \max\left(0.05, \min\left(3.0\ \text{GeV}/p_T, 0.1\right)\right) \tag{2.14}$$

as the decay products are expected to be more collimated for boosted τ_h. If multiple hypotheses are still possible after the mass window and signal cone requirements are applied, the one associated to the τ_h candidate with highest p_T is chosen, the others being discarded. At most, one τ_h candidate per jet is therefore produced. The associated hypothesis is then referred to as the reconstructed decay mode of the τ_h, which can be 1-prong (prong meaning charged hadron), 1-prong+π^0's (which gathers $h^\pm\pi^0$ and $h^\pm\pi^0\pi^0$) and 3-prongs.

To discriminate τ_h from QCD induced jets, the isolation of the candidates can be used as the hadronization of QCD partons is expected to produce more particles than in a genuine τ_h decay. This isolation is computed as the scalar sum of the p_T of charged particles and photons with $p_T > 0.5$ GeV, reconstructed with the PF algorithm, within an isolation cone of size $\Delta R = 0.5$, centered on the τ_h direction (excluding the τ_h constituents). For analyses targetting final states with large hadronic activity, such as $t\bar{t}H$ searches, the isolation cone

size is reduced to $\Delta R = 0.3$. To reduce the sensitivity to pile-up, only the charged particles associated to the same vertex as the τ_h are considered (within Δz<0.2 cm and Δr<0.03 cm). The photons cannot be unambiguously assigned to a given vertex and the influence of pile-up for those particles is therefore evaluated on a statistical basis via the so-called modified $\Delta\beta$ corrections:

$$I_\tau = \sum_{\text{charged, } \Delta z<0.2\,\text{cm}} p_T + \max\left(0, \sum_\gamma p_T - \Delta\beta\right) \tag{2.15}$$

The $\Delta\beta$ corrections are computed from the scalar sum of charged particles that are within a cone of size ΔR=0.8 around the τ_h direction and are associated to vertices with Δz>0.2 cm with respect to the τ_h production vertex (i.e. pile-up vertices). That sum is scaled by a factor 0.2 to take into account the relative contribution of the charged and neutral components of the pile-up:

$$\Delta\beta = 0.2 \sum_{\text{charged, } \Delta z>0.2\,\text{cm}} p_T \tag{2.16}$$

In its most simple version, the τ_h identification directly relies on thresholds applied to this isolation variable (cutoff-based isolation). A more complex MVA (Multi-Variate Analysis) discriminant, based on a BDT output, has also been developed. In addition to the aforementioned charged and neutral isolation sums, it also uses the transverse impact parameter of the leading track of the τ_h candidate as input variable. This variable is defined as the distance of closest approach in the transverse plane of the track to the τ_h production vertex. For the candidates reconstructed in the 3-prongs decay mode, it also takes into account the distance between the τ_h production point and the decay vertex. The possible p_T and η dependence of the input variables is also taken into account in the BDT training. The distribution of that variable in simulated samples is presented in Fig. 2.18.

τ_h identification efficiencies and misidentification rates, evaluated in Monte-Carlo simulated events, are presented in Fig. 2.19. The different working points are able to define working points with a sizable signal efficiency, while significantly reducing the contribution from the different reducible backgrounds.

In addition to fakes from QCD jets, the τ_h identification has also to deal with electrons or muons reconstructed as τ_h. It can be a crucial issue for analyses looking for a resonance decaying into a τ lepton pair, one of them decaying leptonically and the other hadronically. One of the main backgrounds is then the Drell-Yan production of a Z boson decaying into a pair of electrons or muons, which can have a cross section significantly higher than the one of the searched signal. Anti-lepton discriminators have therefore been developed to keep the fake rate from electrons and muons as low as possible [135].

Prompt isolated electrons, produced for instance in the decay of gauge bosons, can easily be reconstructed as 1-prong τ_h. Moreover, because of the significant material budget in the tracker, electrons have also a sizable probability to produce Bremsstrahlung photons and can be reconstructed in that case in the 1-prong+π^0's decay mode or even in the 3-prongs decay mode, in the case of a photon conversion. To discriminate genuine τ_h from electrons, a discriminator based on a BDT score is used. The variables used as

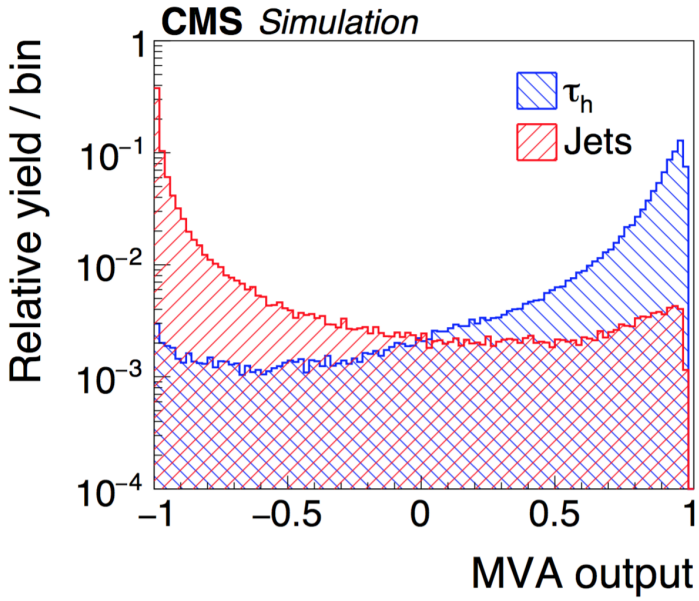

Figure 2.18: Distribution of MVA output for the τ_h identification discriminant that includes life-time information for hadronic τ decays in simulated $Z/\gamma^* \to \tau\tau$ (blue), and jets in simulated W+jets (red) events. [135]

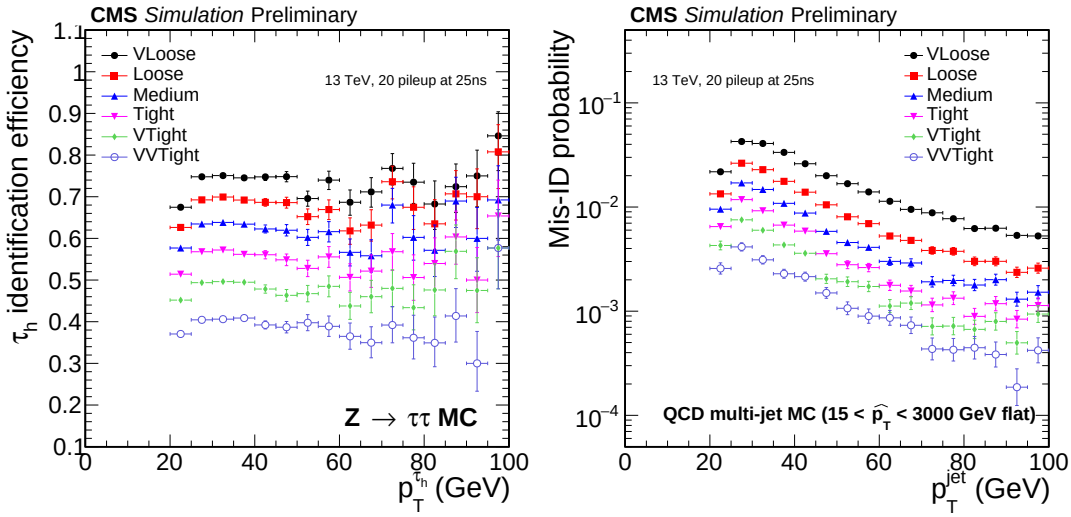

Figure 2.19: Expected efficiencies and misidentification rates of various τ_h identification working points as a function of p_T evaluated in simulation. The efficiency is shown as a function of the τ_h transverse momentum while the misidentification probability is shown as a function of the jet transverse momentum [136].

inputs are related to the distribution in energy depositions in the ECAL, to the amount of Bremsstrahlung emitted along the leading track and to the overall particle multiplicity, in order to distinguish electromagnetic from hadronic showers.

As for the muons, the discrimination is in principle a lot easier since the muons can go through the outermost parts of the detector, while the hadronic τ decay products are stopped in the calorimeters. Sometimes though, detector inefficiencies or non-instrumented regions can cause the muon reconstruction to fail and the muon can then mimic a perfect 1-prong candidate. Those cases are covered thanks to an MVA discriminant, based on a BDT using calorimetric and muon chambers information, to assess the compatibility of a candidate with a minimum ionizing particle, such as a muon.

2.3.7 Missing transverse energy

Using the Particle-Flow algorithm, the CMS detector is able to reconstruct almost all the stable particles produced in *pp* collisions. However, weakly interacting neutral particles, such as neutrinos or hypothetical dark matter candidates, cannot be directly detected, which prevents their individual reconstructions. Nevertheless, their presence can still be inferred using the conservation of the momentum in the collision. As the incoming partons from the protons can be assumed to have no transverse momentum, the vectorial sum of the momenta of all the particles produced in the final state should also share this characteristic. The missing transverse energy $\overrightarrow{\not{E}_T}$ is then defined as the imbalance in the transverse momentum of all the reconstructed particles in an event. Its magnitude is denoted as $\not{E}_T$.

In a hadron collider experiment, this quantity is the only available estimate of the undetected particles momenta. The longitudinal momentum or the energy of those particles cannot be reconstructed as the energy of the incoming partons from the protons is unknown. This limitation is not present with lepton colliders, since elementary particles are used for the collisions, whose energy is precisely measured, which enables to use the full four-momentum conservation during the collisions to evaluate the four components of the undetected particles momenta.

Different estimations of the $\overrightarrow{\not{E}_T}$ have been considered in CMS [137], based on the inputs used to compute it. The most performant algorithms usually make use of the particles reconstructed by the Particle Flow algorithm. Yet, the measurement of $\overrightarrow{\not{E}_T}$ can still be impacted by tracker inefficiency or calorimeter non linearity, which affect the other particle reconstruction. Therefore corrections are applied to mitigate those effects. The so-called "type-1" corrections are applied to propagate the jet energy corrections (see section 2.3.4) into the $\overrightarrow{\not{E}_T}$ computation, while the "type-0" corrections reduce the effects of pile-up by subtracting charged hadrons and compensating for the remaining imbalance from neutral hadrons.

The performance of the $\overrightarrow{\not{E}_T}$ reconstruction can be evaluated using $Z \to \mu\mu$ events. No real $\overrightarrow{\not{E}_T}$ is expected in those events but it can be induced by removing the muon pair from

the event reconstruction. The resolution on the $\overrightarrow{\not{E}_T}$ is then dominated by the hadronic activity as the muon resolution is very good. The scale and the resolution of the $\overrightarrow{\not{E}_T}$ are measured by comparing the response of the hadronic recoil $\overrightarrow{u_T}$ to the transverse momentum of the vector boson $\overrightarrow{q_T}$, which are related by

$$\overrightarrow{q_T} + \overrightarrow{u_T} + \overrightarrow{\not{E}_T} = \overrightarrow{0} \tag{2.17}$$

The different vectors are represented in Fig. 2.20. The hadronic recoil is decomposed into two components $u_\parallel$ and $u_\perp$ with respect to the axis defined by $\overrightarrow{q_T}$. As $\overrightarrow{\not{E}_T}$ is supposed to be null for those events, in the ideal case, $u_\parallel = q_T$ and $u_\perp = 0$ and the resolution on the two projections $\sigma(u_\parallel)$ and $\sigma(u_\perp)$ can be used to estimate the resolution on the $\overrightarrow{\not{E}_T}$.

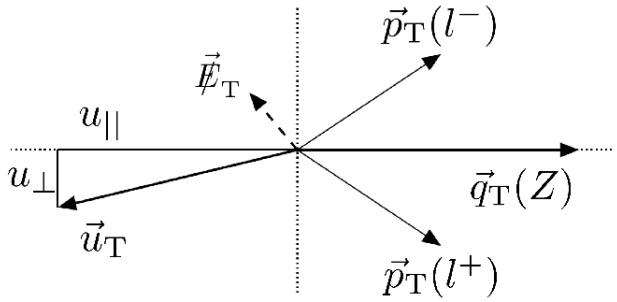

Figure 2.20: Illustration of $Z \to l^+l^-$ event kinematics in the transverse plane [138]

The resolution on the PF $\overrightarrow{\not{E}_T}$, defined from the negative vectorial sum of the momenta of all the Particle-Flow particles reconstructed in the event, is strongly degraded at higher pile-up, which induces a smearing of the longitudinal and transverse components by a few GeVs. Several algorithms have therefore been developed [138] in order to mitigate that effect, which separate the PF $\overrightarrow{\not{E}_T}$ into different contributions: particles originating from the production vertex (PV) and particles originating from pile-up interactions (PU).

One of them is the so-called No-PU PF $\overrightarrow{\not{E}_T}$, in which the contribution from pile-up particles is scaled down by a factor computed using the tracks that are not clustered into jets with p_T>30 GeV and are either associated to the primary vertex or not

$$S_F = \frac{\sum_{PV,\,charged} p_T}{\sum_{PV,\,charged} p_T + \sum_{PU,\,charged} p_T} \tag{2.18}$$

Another algorithm is based on a set of multivariate regressions computed as corrections to the hadronic recoil $\overrightarrow{u_T}$ of the PF $\overrightarrow{\not{E}_T}$. First a correction on the azimuthal angle of $\overrightarrow{u_T}$ is applied by using a BDT to match the azimuthal angle of $-\overrightarrow{q_T}$. In a second step, a correction derived from another BDT is used to predict the magnitude of the true $\overrightarrow{u_T}$. The corrected value of $\overrightarrow{u_T}$ is then added to $\overrightarrow{q_T}$ to define the negative MVA $\overrightarrow{\not{E}_T}$. Two BDT

trainings have been investigated: one optimizes the $\overrightarrow{\not{E}_T}$ resolution (MVA PF $\overrightarrow{\not{E}_T}$) while the other is trained to reach unity response for the variable $u_{\parallel}/q_T$ (MVA Unity PF $\overrightarrow{\not{E}_T}$).

Five $\overrightarrow{\not{E}_T}$ estimators are used to compute the input variables for the BDTs:

1. the negative vectorial sum of the momenta of all PF particles in the transverse plane (PF $\overrightarrow{\not{E}_T}$),
2. the negative vectorial sum of the momenta of all charged PF particles that have been associated to the selected primary vertex,
3. the negative vectorial sum of the momenta of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles within jets that have passed the MVA pileup jet ID (see section 2.3.4),
4. the negative vectorial sum of the momenta of all charged PF particles that have not been associated to the hard-scatter vertex and all neutral PF particles within jets that have failed the MVA pileup jet ID,
5. the negative vectorial sum of the momenta of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles (also those that have not been clustered into jets) plus the positive vectorial sum of the momenta of all neutral PF particles within jets that have failed the MVA pileup jet ID.

For each $\overrightarrow{\not{E}_T}$, the hadronic recoil $\overrightarrow{u_T}$ is computed. The input variables of the BDTs are then the magnitude and azimuthal angle of all five types of $\overrightarrow{u_T}$, the scalar sum of transverse momenta of all PF particles of each $\overrightarrow{\not{E}_T}$ variable, the momenta of the two highest p_T jets in the event, and the number of primary vertices. The BDTs can therefore benefit from each missing transverse energy estimation: (2) is almost pile-up insensitive, (3) does not depend much on pile-up and provides a better estimation of the energy scale, (4) estimates the pile-up contribution and (5) takes into account the contribution from unclustered particles.

As seen in Fig. 2.21, the MVA PF $\overrightarrow{\not{E}_T}$ significantly improves the resolution on the $\overrightarrow{\not{E}_T}$ and the pile-up dependence is reduced. Other $\overrightarrow{\not{E}_T}$ estimations based on the PUPPI algorithm [139] are also under study in CMS for Run 2.

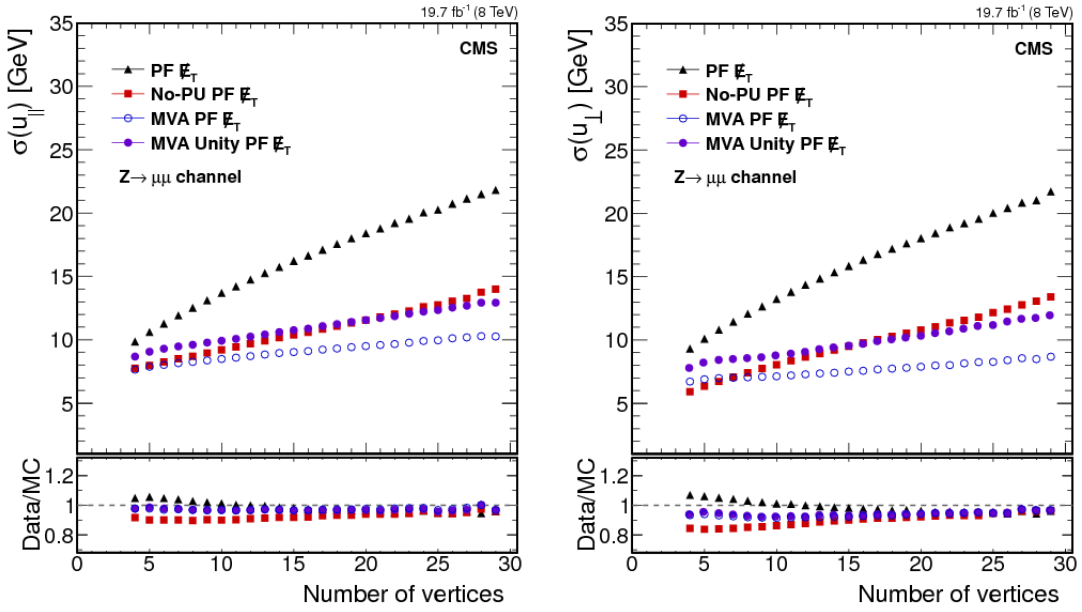

Figure 2.21: Longitudinal (left) and transverse (right) recoil component resolution as a function of the number of reconstructed vertices for different PF $\overrightarrow{\not\!\!E_T}$ algorithms in $Z \to \mu\mu$ events [138]

Chapter 3

Level-1 electron and photon trigger upgrade and commissioning

Don't lower your expectations to meet your performance.
Raise your level of performance to meet your expectations.

Ralph Marston

In Run 2, not only the center-of-mass energy of the LHC collisions has been increased to 13 TeV, but the instantaneous luminosity delivered has even reached values beyond the nominal design of the machine of 10^{34} cm^{-2}s^{-1}. In order to guarantee a successful and ambitious physics programme in this intense environment, the CMS trigger and data acquisition system had to be upgraded. The Level-1 electron and photon trigger (Level-1 EG) hardware and architecture has been redesigned to maintain the current thresholds even in presence of more extreme conditions. The trigger thresholds used for data taking play an important role for analyses involving a Higgs boson decaying into τ leptons, as their visible decay products can have relatively low p_T due to the presence of neutrinos. I had the opportunity to play a major role in the commissioning of this new trigger system and to measure the performance of the upgraded Level-1 EG trigger. In this chapter, the whole upgrade of the CMS Level-1 trigger system will be introduced. The implementation of the new Level-1 EG trigger in particular will then be detailed. Finally, the performance of this trigger, measured in the first 2016 collision data, will be presented.

3.1 Phase 1 Level-1 trigger upgrade

3.1.1 General principles of a trigger system

With a collision rate of 40 MHz, the data volume produced by the CMS detector can reach up to 1000 Tb/s. This amount of data is far beyond the technical capabilities of any data acquisition system based on computing. However, the rate of physics processes targeted by physics analyses (either within or beyond the Standard Model) can be more than ten

orders of magnitude lower than the one of inelastic scattering, as presented in Fig. 3.1. The aim of a trigger system is therefore to increase the event purity, by being as efficient as possible on the signals while reducing the QCD background down to a level compatible with data acquisition. The reduced amout of data can then be processed by a computer farm with a thousand of processing units.

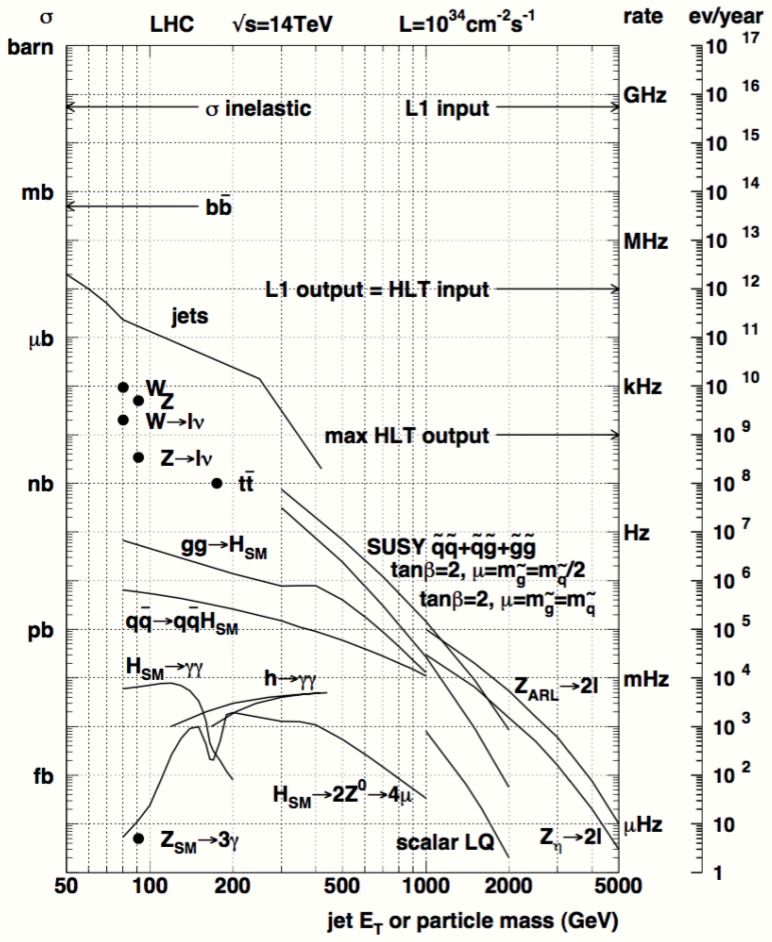

Figure 3.1: Cross sections for various processes at LHC for collisions at $\sqrt{s} = 14$ TeV. The corresponding rates are indicated for an instaneous luminosity $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}s^{-1}$. [124]

In high energy physics experiments, the rate reduction is generally performed in several steps. As mentioned in section 2.2.7, in CMS, the first level of the trigger system, called "Level-1 trigger" (L1), is based on custom electronic boards, while the second level, called

"High Level Trigger" (HLT), uses a dedicated computer farm [115]. The Level-1 trigger decision is based on coarse detector information while the HLT benefits from the full granularity of the detector. The conception of the CMS trigger system takes of course into account physics considerations related to the physics programme of CMS but has also to deal with technical constraints related to the detector itself as well as the available technology when being designed.

Physics considerations

For what regards the trigger, physics analyses must determine how they can optimally make use of an allocated bandwidth in terms of rate. For that purpose, the nature of the physics objects required for the analysis and their optimal selection criteria must be assessed. The search for the Higgs boson provides a good benchmark for this. At the time of the conception of CMS, since the mass of the Higgs boson was not known, it was important to provide a good coverage in terms of trigger of all its possible decay modes, as the corresponding branching ratios can significantly vary depending on the mass value (see section 1.2.3). The main discovery channels were $H \to ZZ \to 4\ell$, $H \to WW \to 2\ell 2\nu$ and $H \to \gamma\gamma$. For the decay into massive electroweak bosons, isolated leptons and possibly missing transverse energy are expected in the final state. The trigger selection had therefore to be able to identify leptons with p_T thresholds as low as possible with a high efficiency, to target leptons from an offshell Z or W boson decay. The triggers used for these analyses were requiring either the presence of at least a single electron or muon or a tighter selection with at least two isolated leptons, with lower p_T thresholds resulting in a higher efficiency. For what regards the $H \to \gamma\gamma$ search, the L1 trigger used was in fact the same as requiring two electrons, as no distinction is made between electrons and photons at this level since only the calorimeter information is used in that case.

The $H \to \tau\tau$ and $H \to bb$ analyses must deal with a large rate from the QCD background and it is therefore important to exploit every available leverage for rate reduction. In the case of $H \to \tau\tau$, it is not realistic to require the presence of only a single τ_h as the p_T thresholds would be too high to efficiently collect that signal. Instead the presence of at least two τ_h or at least one τ_h and one isolated lepton is required. Additional handles can be used in the case of associated productions modes like VBF, where the presence of two forward jets can be exploited, or VH, where the trigger can rely on the presence of leptons from the electroweak boson decay. As an example, the combination of triggers used in the $2\ell ss + 1\tau_h$ category of the 2016 $t\bar{t}H$ analysis in final states with τ leptons (see Chapter 5) is presented in Table 3.1.

Different kinds of objects, more or less corresponding to those used in the offline analyses but with a lower resolution, are reconstructed at the trigger level. In the busy hadronic environment of LHC, isolated leptons provide good candidates to trigger on, as they are not produced in most of QCD events, which enables to significantly reduce the rate. Muon candidates are produced with a relatively low cross section with respect to other trigger objects, as they are reconstructed based on information from the muon detectors, which are located behind the solenoid and are therefore less impacted by hadronic activity than

Trigger type	HLT path	Lowest unprescaled L1 seed
Single electron	HLT_Ele27_eta2p1_WPLoose_Gsf OR HLT_Ele27_WPTight_Gsf	L1_SingleIsoEG34er OR L1_SingleIsoEG36 OR L1_SingleEG38
Single muon	HLT_Iso(Tk)Mu22(_eta2p1) OR HLT_Iso(Tk)Mu24	L1_SingleMu20er OR L1_SingleMu22
Double electron	HLT_Ele23_ Ele12_CaloIdL_ TrackIdL_IsoVL_DZ	L1_DoubleEG_25_12 OR L1_DoubleEG_24_17 OR L1_SingleIsoEG34er OR L1_SingleIsoEG36 OR L1_SingleEG38
Muon+electron	HLT_Mu23_TrkIsoVVL_ Ele8_CaloIdL_TrackIdL_IsoVL(_DZ) HLT_Mu8_TrkIsoVVL_ Ele23_CaloIdL_TrackIdL_IsoVL(_DZ)	L1_Mu20_IsoEG6 OR L1_SingleMu20er OR L1_SingleMu22 L1_Mu5_IsoEG20 OR L1_Mu5_EG23
Double muon	HLT_Mu17_TrkIsoVVL_ Mu8_TrkIsoVVL_DZ	L1_DoubleMu_12_5

Table 3.1: Triggers used in the $2\ell ss + 1\tau_h$ category of the 2016 $t\bar{t}H$ analysis in final states with τ leptons. The events must first trigger the L1 seed then the corresponding HLT path. Only the L1 seeds enabled for all luminosity conditions are indicated here. The numbers indicated in the trigger paths refers to the p_T thresholds. In addition, various quality criteria, isolation requirement or eta-restriction (er) can be applied. For L1 seeds such as L1_SingleIsoEG34er, er means eta-restricted and corresponds to an $|\eta| < 2.1$ cut.

the calorimeters. Algorithms involving muons have therefore typically lower p_T thresholds than the calorimeter triggers. Electrons and photons candidates are reconstructed based on the information from ECAL. As the tracker information is not available at Level-1, no distinction is made between the two for Level-1 triggers. The rate associated to those objects can be further reduced by requiring isolation criteria based on the surrounding energy deposits in the calorimeters.

For what regards hadronic objects, the distinction between those produced in hard and soft scattering is relatively difficult with the limited information available at trigger level, which is why their associated p_T thresholds are much higher than those for leptons. It is possible to reduce the associated rate by requiring the presence of multiple jets in the events or by targetting specific topologies (like the VBF one). Low p_T thresholds can still be used in trigger paths with monitoring purposes but they must be prescaled. For instance, in the case of a prescale of 1000, the trigger system for instance only keeps one triggered event every one thousand, which allows to reduce the associated rate. The jets produced by hadronic τ's are typically narrower than those produced by QCD partons and this characteristic can be exploited at the trigger level using dedicated reconstruction techniques to produce τ candidates. Combined with an isolation requirement, it is thus

for instance possible to use at Level-1 the unprescaled path L1_DoubleIsoTau36er, while the lowest unprescaled double jet trigger is L1_DoubleJetC112. Finally, global quantities such that the missing transverse energy $\not{E}_T$ or the total hadronic activity H_T can also be used to trigger.

Technical considerations

The main points regarding the technical specification of the trigger system are the rate reduction, imposed by the capabilities of the data acquisition (DAQ) system, together with the available latency. The current DAQ in the CMS experiment can typically sustain a data taking up to 1.5 kHz of events. On the other hand, at a luminosity of 10^{34} cm$^{-2}s^{-1}$ and with on average more than 20 interactions per bunch crossing, the input rate of events in the detector is of the order of 1 GHz, which requires a rate reduction by a factor 10^6. As already mentioned in section 2.2.7, to meet that challenge, the CMS trigger is organized in two stages. The Level-1 trigger first reduces the rate down to 100 kHz. Because of the considerable amount of data produced by the detector (1000 Tb/s), the Level-1 trigger must be based on custom electronic boards. The Level-1 output data flux is thus reduced to 2.5 Gb/s sent to the HLT, which further skims it down to 225 Mb/s.

The latency of the trigger system corresponds to the maximum time allowed for the system to take the decision to accept or discard an event. This parameter depends on different factors including:

- the required time to read the information from the detector: this must take into account the transit of data from the detector through optical fibers, limited by their bandwidth.
- the time needed by the trigger algorithms to process the information and produce physical quantities which are used to trigger on the events
- the capability of the data buffer in each subdetector: it corresponds here to the maximal depth in terms of number of events which can be stored in the Front End (FE) Electronics.
- the time needed to distribute the L1 Accept signal to the FE of each subdetector

The CMS Level-1 trigger has a fixed latency of 3.2 μs. This constraint is imposed by the data buffer of the tracker and the preshower detectors in the endcaps. Since the electronic boards used by the Level-1 trigger system are located in the service cavern, outside of the experimental cavern, 62 m of optical fibers are required, which already consumes a non-negligible part of the latency budget. In the end, 1 μs only are allocated to the trigger algorithms, constraint which has to be taken into account in their design.

In addition to the rate reduction and the latency, the flexibility of the trigger system is important as well. Indeed, with the evolving conditions of LHC collisions, the trigger system must be configurable to make it as optimal as possible for every luminosity and pile-up conditions. As the Level-1 trigger is hardware-based, this is a priori more difficult

for the Level-1 than for the software-based HLT. Still, the Level-1 system benefits from the programmable logic available nowadays in FPGAs (Field Programmable Gate Arrays). However, these FPGAs cannot be used in a radiation-hard environment. As during the CMS conception, the available optical fibers did not meet the needs of the experiment in terms of radiation-hardness, bandwidth and cost, the chosen solution consists of ASICs (Application-Specific Integrated Circuit) on the detector and of FPGAs on the electronic boards in the service cavern.

3.1.2 CMS Level-1 Trigger architecture in Run 1

The architecture of the Level-1 trigger system used by CMS during Run 1 is presented in Fig. 3.2. As can be seen, data produced by the muon subsystems and by the calorimeters are first treated separately by the Muon Trigger and the Calorimeter Trigger. Those subsystems take as inputs *trigger primitives*, which corresponds to a coarse reconstruction of energy deposits in the subdetectors. In the case of the muon subsystems, those trigger primitives consist of track segments mainly produced by the Drift tubes (DT) and the Cathode Strip Chambers (CSC) (see section 2.2.6). The Resistive Plate Chambers (RPC) produce complementary trigger information with a very good timing resolution. Muon candidates (L1Mu) are then reconstructed from that information and the Global Muon Trigger (GMT) collects them, removes potential duplicates and orders them by decreasing order of p_T. For what regards the Calorimeter Trigger, the trigger primitives are produced from ECAL, HCAL and HF data and are processed to produce electron-photon (L1EG), jets (L1Jet), hadronic τ's (L1Tau) candidates as well as global quantities such as total hadronic energy (L1HT) or missing transverse energy (L1MET). All the L1 candidates produced by the Muon and the Calorimeter Trigger are then sent to the Global Trigger (GT), which checks if any of the conditions implemented in the full set of L1 algorithms used for data-taking (L1 menu) is fulfilled. If so, a L1 Accept signal is produced and the event is then sent to the HLT to be further processed.

One of the challenges the CMS L1 trigger system has to meet is the massive amount of data produced by the detector, which has to be processed during the 3.2 μs of latency allocated to the L1 trigger. As the amount of data to transfer in a given time is limited by the bandwidth of the system, the strategy adopted in Run 1 was then to progressively reduce the amount of data to be transferred to be compatible with the technical constraints. This reduction is achieved thanks to a regionalized implementation. This regionalization already starts at the level of the trigger primitive generation. In the case of ECAL for instance, the trigger primitives are for instance built in the barrel from 5×5 arrays of crystals (while in the endcaps, the geometry is more complex, as presented in Fig. 3.3). They already take into account some identification criteria, especially regarding the removal of anomalous ECAL signals known as spikes [124]. In the Regional Calorimeter Trigger (RCT), the detector is segmented in areas of $\Delta\eta \times \Delta\phi = 0.35 \times 0.35$, corresponding to 4×4 trigger towers (combination of an ECAL and HCAL trigger primitive). Each of those RCT regions are then processed in parallel to produce L1 EG candidates and to compute the total energy deposit in those regions, which are then sent to the Global Calorimeter

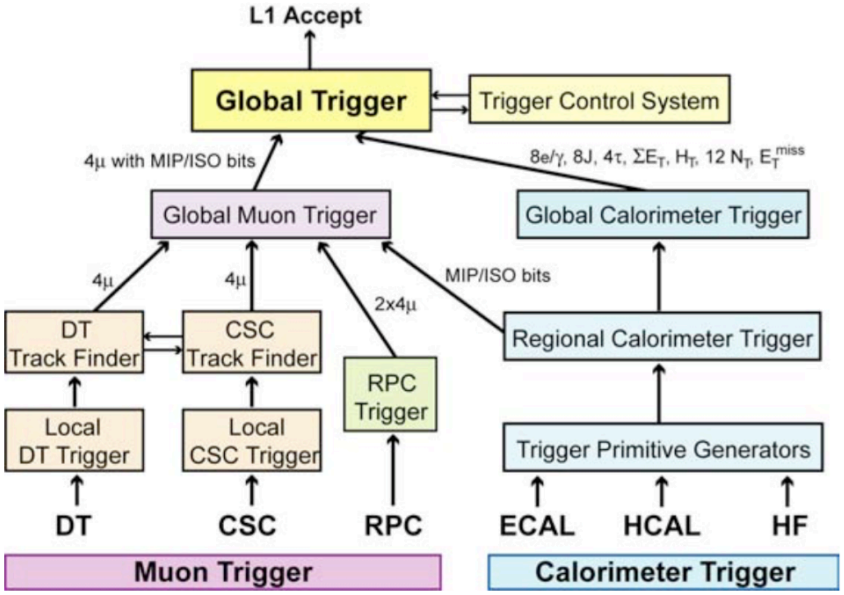

Figure 3.2: CMS Level-1 trigger architecture during Run 1. The different subsystems of the Muon Trigger are presented on the left while those for the Calorimeter Trigger are displayed on the right. Level-1 candidates are sent by the Global Muon Trigger (GMT) and the Global Calorimeter Trigger (GCT) to the Global Trigger (GT) which delivers the Level-1 trigger decision. [140]

Trigger (GCT). The GCT takes care of the E_T ordering of the L1EG candidates and of the reconstruction of L1Tau and L1Jet candidates together with the computations of L1 global quantities.

3.1.3 Upgrade of the CMS Level-1 Trigger for Run 2

With the luminosity increase planned by LHC during Run 2 and 3 potentially up to 2.10^{34} $cm^{-2}s^{-1}$ (twice the design luminosity!), CMS undertook a major upgrade of its Level-1 Trigger system to be able to deal with those harsh data-taking conditions. If nothing had been done, the trigger thresholds required to stay within the Level-1 rate of 100 kHz would have been increased significantly, which would have significantly impacted the physics programme of CMS.

For what regards the Level-1 Calorimeter Trigger, one of the major changes introduced with this upgrade has been the improvement of the granularity of this trigger system: the upgrade has enabled the trigger reconstruction algorithms to get access to the individual trigger towers instead of the regional information used previously. The upgrade had thus positive consequences in several areas:

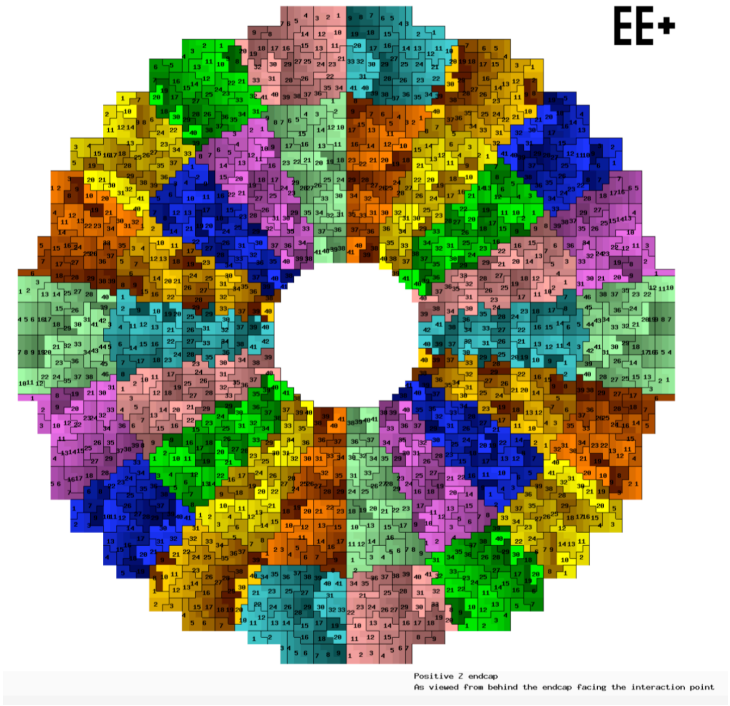

Figure 3.3: Trigger towers map in the ECAL endcaps. The trigger towers are bounded by full black lines. [124]

- For the electrons and photons, the Run 1 algorithm reconstructed L1EG candidates from a fixed window of 2x1 trigger towers. In the case of Bremsstrahlung radiation, this window was actually too small to recover the whole energy deposit from electrons or photons. With the new Run 2 trigger, it has been possible to adapt the reconstruction to improve the resolution on L1EG candidates. Moreover, an improved isolation criteria, including a pile-up mitigation, has also been deployed.

- For the hadronic τ's, the rate reduction with respect to QCD jets was achieved using a dedicated topological pattern identification on the calorimeter energy deposit. However, this resulted in a limited trigger efficiency below 60% (which was mitigated combining L1Tau and L1Jet algorithms). The new L1Tau algorithm makes use of a reconstruction and an isolation criteria similar to the ones for L1EG candidates, which achieves a much higher efficiency than the Run 1 algorithm for all the possible τ_h decay modes.

- For the jets and calorimeter sums, the new L1 trigger better granularity enabled to improve the energy resolution, especially in the HF, and introduced pile-up subtraction techniques absent in Run 1.

- The new L1 trigger also introduced the possibility to implement sophisticated correlation conditions, in particular based on invariant mass computations already possible at hardware level. Those improvements have been exploited to design for instance VBF triggers, improving significantly the sensitivity to that production mode of the Higgs boson.

Another major change introduced in the new Level-1 Calorimeter Trigger, which enabled all the aforementioned improvements, concerns the hardware architecture. The Run 1 regionalized approach has indeed been replaced with a trigger architecture called Time Multiplexed Trigger (TMT) [141]. Instead of splitting and duplicating information based on detector regions to process it in parallel, the TMT processes together all the information coming from the whole detector. This enables for instance to determine the energy density in the event to mitigate the effect of pile-up. However such a treatment of the information results in an increased latency. The TMT includes therefore several processors which can treat sequentially the events corresponding to different bunch crossings. Moreover the TMT does not require more latency as the trigger algorithms can process data as they arrive, without having to wait for the reception of the whole information.

This new architecture is illustrated in Fig. 3.4. It uses a first layer of CTP7 electronic boards and a second Layer of MP7 boards, presented in Fig. 3.5. The Layer-1 boards receive in parallel the information from the calorimeters and reorganize it to send it sequentially to the Layer-2: at each clock count, data corresponding to a whole η slice is sent to a single MP7 board. The data from the next event is then sent to another MP7. Overall 9 MP7 boards are used, which corresponds to a latency of 9×25 ns for running the L1 trigger algorithms. Thanks to that latency, a certain flexibility in the complexity of the algorithms is allowed. An additional MP7 board is then used as demultiplexer (Demux) to reorder and format data before sending it to the Global Trigger.

In order to ensure the constant availability of the trigger system during data taking, the following strategy has been adopted by CMS. An upgraded version of the Calorimeter Trigger, called Stage-1, has been designed and used for data-taking in 2015. It was based on the same regionalized architecture as the Run 1 Calorimeter Trigger. The new version of the Calorimeter Trigger, to be used for the remaining of the Phase 1 and called Stage-2, has been finalized and tested in parallel with the Stage-1 trigger until the end of data-taking in 2015. I have been deeply involved into the final implementation of the Stage-2 L1EG algorithm and its commissioning during this phase. The detailed implementation of the Stage-2 L1EG algorithm will be presented in the next section. The parallel running enabled to show that the Stage-2 trigger was able to run with the expected performance and it has therefore been used from 2016 onward for data-taking.

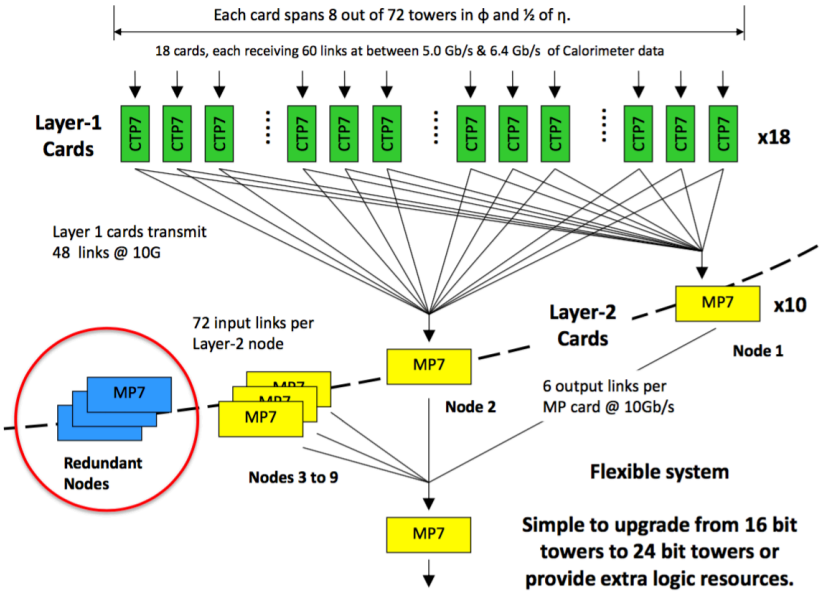

Figure 3.4: TMT architecture used in the upgraded CMS Level-1 Calorimeter Trigger. See text for explanations. [140]

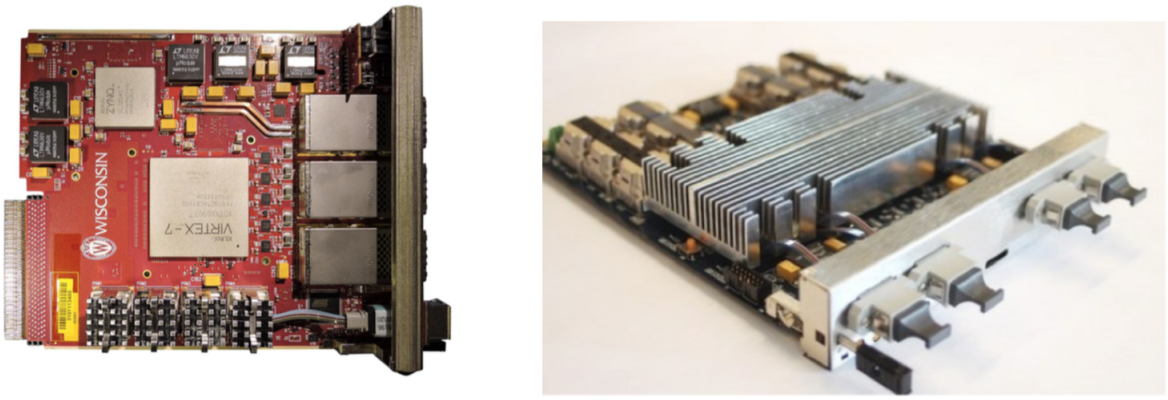

Figure 3.5: CTP7 [142] (left) and MP7 [143] (right) boards used in the upgraded CMS Level-1 Calorimeter Trigger.

3.2 Implementation of the Stage-2 Level-1 EG algorithm

3.2.1 Reconstruction

As mentioned previously, the inputs of the Calorimeter Trigger correspond to the ECAL and HCAL trigger towers. Those trigger towers are first processed by the Layer-1 to organize the information in a way compatible with the TMT architecture. Due to bandwidth limitations, information from ECAL and HCAL has to be compressed before being sent to Layer-2, where the L1EG algorithm is implemented. The information associated to each trigger tower is sent under the form of a 16-bit number:

- one bit corresponding to the ECAL fine-grain bit used for the L1EG identication (see section 3.2.2)
- one bit corresponding to the HCAL feature bit (currently unused)
- one bit indicating whether the ECAL energy in the trigger tower (E) is larger than the HCAL energy (H)
- one bit indicating whether $E = 0$ or $H = 0$
- three bits corresponding to $\log_2(E/H)$ (binary logarithm of the E/H ratio) if $E \geq H$ or $\log_2(H/E)$ if $H > E$
- nine bits corresponding to the total transverse energy $E + H$

Knowing the value of $E+H$ and H/E, it is possible to reconstruct if needed the individual components E and H. The position of a trigger tower is indexed by two numbers $i\eta$ (going from -28 to 28 for ECAL) and $i\phi$ (going from 1 to 72). Each CTP7 sends information related to 4 trigger towers along $i\phi$ (corresponding to a 20° sector). At each clock count, data corresponding to a whole ϕ slice of 72 towers is sent for each half of the detector ($\eta < 0$ and $\eta > 0$). The $i\eta$ position is therefore inferred from the clock and the L1EG algorithm can start running without waiting for data from the whole detector to be sent.

The reconstruction of the L1EG candidates used in the Stage-2 algorithm is inspired by the superclustering algorithm used for the offline reconstruction of electrons (see section 2.3.3). This algorithm showed indeed good performance to collect the potential Bremsstrahlung radiation emitted by electrons when going through the tracker material. For that reason, even though electrons produce compact electromagnetic showers, their energy deposits in the ECAL can have different sizes. The study of the footprint of electrons in the ECAL, presented in Fig. 3.6, showed that an appropriate shape was provided by a 3x3 window around the tower with maximal energy (the seed), extended by one or two towers in the ϕ direction. Moreover, the size of the footprints is most of the time contained within two trigger towers along η.

Based on those observations, the following clustering algorithm has been implemented:

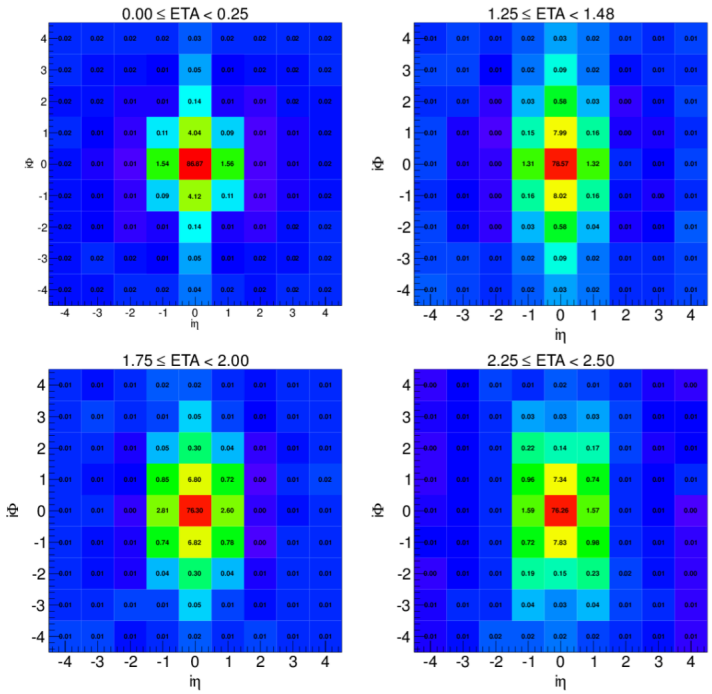

Figure 3.6: "Footprints" of the energy deposit from electrons in different η regions of the detector, observed in Run 1 data. The granularity shown here corresponds to the trigger towers (5x5 ECAL crystals). The z axis represents the fraction of energy deposited by the electron in each tower around the maximum. [140]

- Potential seeds are first defined as trigger towers associated to a local maximum of energy deposit in a 3×9 window in $i\eta \times i\phi$ around them. They are also required to have an energy above 2 GeV to limit the influence of detector noise. To avoid situations where two towers with the same energy in a 3×9 window would veto each other, a dedicated set of large and strict inequalities, shown in Fig. 3.7 (left), is used to define whether a tower is a potential seed.

- The neighboring towers in a 3×3 window around the seed are then clustered together with the seed if they have an energy above 1 GeV. The cluster is potentially extended along the ϕ direction by one or two towers if they have an energy above 1 GeV and can be linked to the seed (i.e. if the intermediary tower is also clustered).

- The maximal size of the cluster is then reduced down to 2 towers along η by removing the η-side with the lower energy.

After this procedure, the maximal size of a cluster corresponds therefore to the one presented in Fig. 3.7 (right). These clusters are used to produce not only L1EG candidates but L1Tau candidates as well. Note that for that reason, the clustering is performed using

$E + H$ as the energy of the towers and not E only. When combined with a L1EG identification based on H/E, the performance of the L1EG algorithm is actually equivalent in the two cases.

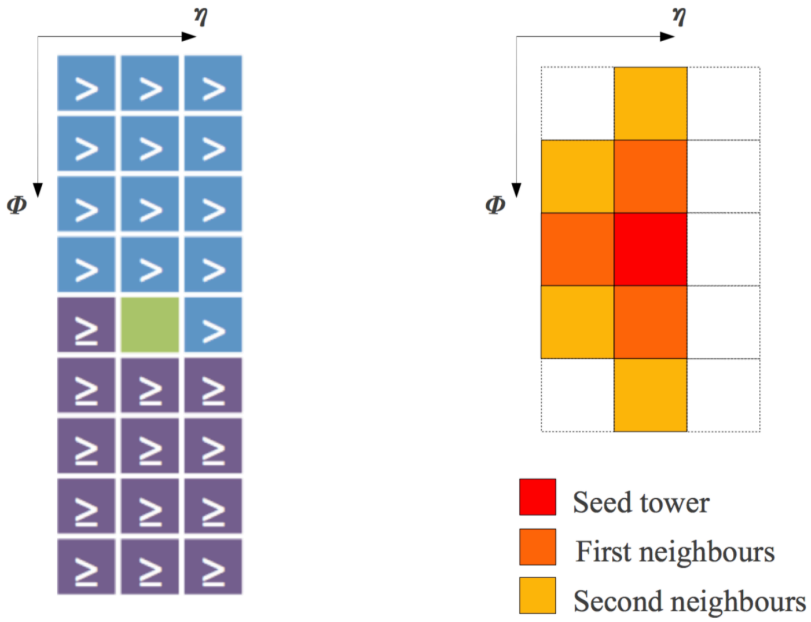

Figure 3.7: Left: Description of the filtering window used to define a local energy maximum in the trigger towers. Right: Maximal cluster shape allowed for a L1EG candidate. Each block represents a trigger tower and the cluster is at most two-tower wide along η and five-tower wide along ϕ. [140]

As previously mentioned, the dynamic clustering enables to improve the energy resolution on electrons and photons with respect to a sliding window algorithm. However, background candidates are mainly produced from QCD processes and pile-up interactions and tend to create wider energy deposits within jets. The dynamic clustering algorithm will therefore collect all the neighboring energy and would in principle produce background EG candidates with higher energy than a sliding window algorithm including only up to two trigger towers for instance, since background candidates are more likely to have wide shapes. The difference in the topological distributions of the clusters for real electrons and for background candidates is presented in Fig. 3.8. To tackle this issue, a trimming is applied on the EG candidates: this step corresponds to an effective reduction of the size of the clusters. The trimming is based on a Look-Up Table (LUT), which simply corresponds to the firmware implementation of a multi-dimensional array. This trimming LUT takes as inputs the η position and the shape of the cluster and returns a reduced shape, where some of the original towers have been removed from the original cluster. The largest shapes, more characteristic from background events, are thus reduced, which effectively limits the energy of background candidates. The trimming LUT has been determined to keep 90%

of the real electrons unaffected.

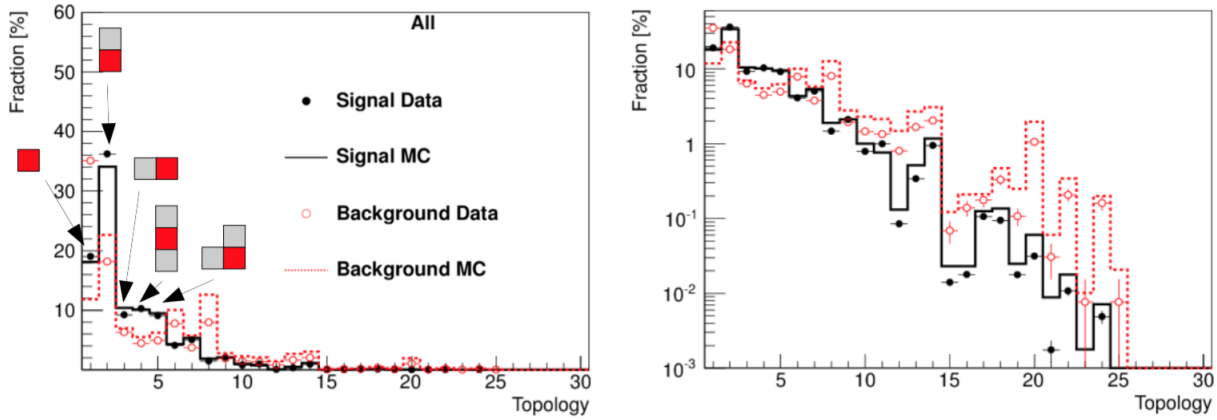

Figure 3.8: Distribution of the shapes obtained with the dynamic clustering algorithm for real electrons (signal) and for the background, in 8 TeV data and in Monte-Carlo simulation, in linear scale (left) and logarithmic scale (righ) [140]

In order to keep the benefit from the improved energy resolution due to the dynamic clustering, a dedicated calibration is applied on the L1EG candidates. The energy of the candidates, obtained by summing the energy of the towers after trimming, is multiplied by a correction factor encoded in another LUT, with the η position, the untrimmed shape and the uncalibrated energy as inputs. The correction factors are determined based on real electrons reconstructed offline in data such that the L1 transverse energy gets as close as possible to the transverse energy of the electrons offline calorimeter superclusters (see section 2.3.3). Those correction factors are typically larger than one, especially in the case of the large shapes, more affected by trimming. This calibration enables therefore to get a uniform L1 energy response in the detector and to compensate for the loss of energy due to the trimming of real electrons. The improvement in the energy resolution with respect to the Run 1 algorithm is presented in Fig. 3.9. Since background EG candidates have an energy distribution within the cluster different from real electrons and since the calibration is tuned on real electrons, the background EG candidates gets undercorrected with respect to the real EG candidates, which enables to keep their energy lower than before trimming and therefore to reduce the rate at a given L1 threshold.

The position resolution has also been considerably improved with the new L1EG algorithm used from Run 2. In the Run 1 algorithm, the resolution in position was limited by the size of the RCT, corresponding to 4×4 sets of trigger towers. With the dynamic clustering, this resolution can even go below the size of a trigger tower. Indeed, the default position attributed to an EG candidate corresponds to the one of its seed tower but using the energy distribution within the cluster, it is actually possible to refine it. For instance, if the upper half of the cluster corresponds to a higher energy deposit than the lower half,

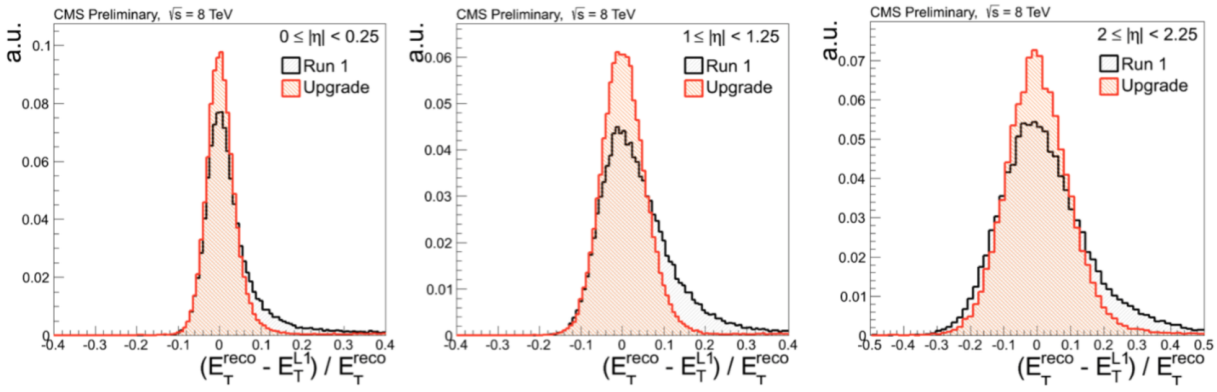

Figure 3.9: Transverse energy resolution of the L1EG candidates with respect to offline electrons in different regions of the detector, for the Run 1 and the upgraded Run 2 algorithm, in 8 TeV data. [140]

the position of the EG candidate is determined accordingly to be shifted by a quarter of trigger tower in that direction with respect to the center of the seed tower, which makes this procedure close to an energy-weighted position. A sketch of the different allowed positions together with the obtained L1EG resolution in position are presented in Fig. 3.10.

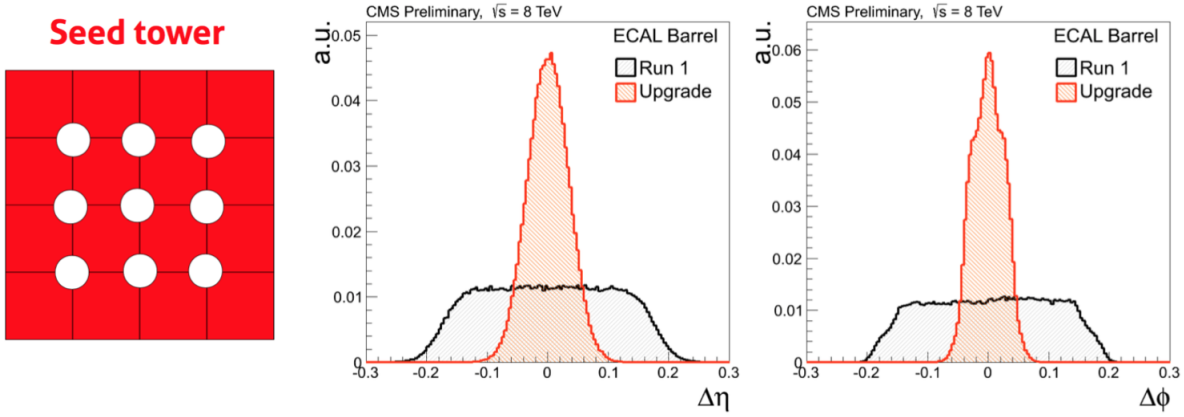

Figure 3.10: Left: Possible refined positions of a L1EG candidate within the seed tower. Right: Position resolution of the L1EG candidates with respect to offline electrons in the barrel, for the Run 1 and the upgraded Run 2 algorithm, in 8 TeV data. The resolution of the Run 1 algorithm is limited by the size of the RCT regions. [140]

3.2.2 Identification

Without any identification criteria, the rate of the L1EG algorithm would be driven by hadronic candidates and would be very similar for instance to the one of the L1Tau algorithm. Knowing that the lowest unprescaled L1 threshold affordable in 2016 for the SingleTau algorithm has been 120 GeV, dedicated identification criteria for the L1EG algorithm must be used to maintain trigger thresholds low enough to cover the electroweak processes which represents an important part of the CMS physics programme. Those are based on three vetos:

- **Fine Grain**: the Fine Grain is a veto computed at the level of the ECAL trigger primitives [124]. It is based on the compactness of the electromagnetic shower in the seed tower.
- **H/E**: a cut on the H/E ratio is determined to reject the large contribution from hadronic background candidates. Two higher bounds have been separately determined for the barrel and the endcaps to keep unaffected 99.5% of the electrons. This cut has been relaxed for candidates above 128 GeV to ensure a 100% efficiency at high p_T.
- **Shape**: this veto is based on the topological shape of the clusters. As mentioned previously, background candidates have typically a larger footprint. The largest possible shapes can therefore almost unambiguously be attributed to those background candidates and be vetoed. The shape veto is based on a LUT using as inputs the η position, the untrimmed shape and the uncalibrated energy of the clusters and is tuned to have a minimal efficiency of 99% on real electrons, this efficiency increasing with E_T to reach 100% for electrons above 70 GeV.

Those vetoes are only applied to L1EG candidates below 128 GeV, in order to ensure a 100% efficiency for high E_T electrons and photons. All the L1EG candidates passing those vetoes are then sorted independently in each η-half of the detector and the six leading ones are further sent to the Global Trigger to be used in the triggers included in the L1 menu.

3.2.3 Isolation

Most of the electrons and photons produced in electroweak processes are expected to be produced without hadronic activity around them (so-called "prompt" electrons or photons). This is for instance the case for electrons produced in W or Z boson decays or for photons produced in Higgs boson decays. An additional identification criteria, based on the energy deposit around the L1EG candidate, is therefore implemented in the Level-1 trigger and can be optionally used to define Level-1 seeds (L1IsoEG) with lower E_T thresholds than the inclusive L1EG seeds.

This isolation criteria is defined based on a 9x6 TT region around the seed of the L1EG candidate, presented in Fig. 3.11, in which the electromagnetic and hadronic energy from all the towers is summed. A footprint of the L1EG candidate is defined as a 5x2 ECAL

region backed by a 2x1 HCAL region (to include potential energy leakage in HCAL of the electron or photon). The energy from this footprint is then subtracted from the total energy deposited in the 9x6 region to define the isolation energy. The isolation criteria requires that this isolation energy is lower than some customized threshold implemented in a LUT, using as inputs the η position, the uncalibrated E_T of the L1EG candidate and an internal variable called n_{TT}.

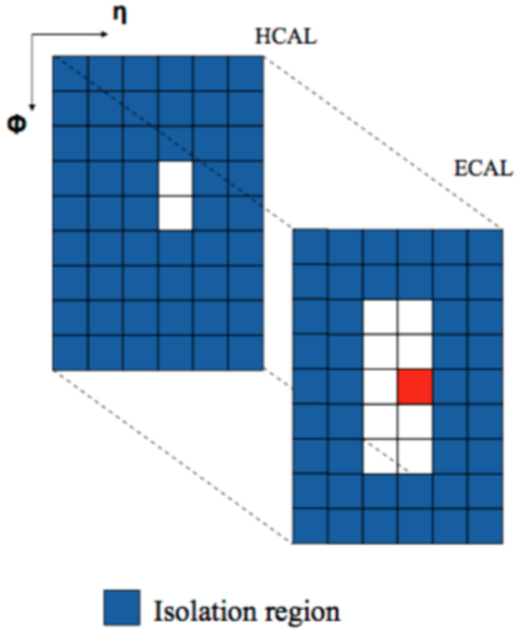

Figure 3.11: L1EG isolation region, decomposed between ECAL and HCAL components. The red tower corresponds to the seed, while the white towers are included in the EG footprint (see text). The blue towers are the ones summed to define the isolation energy.

The n_{TT} variable allows to introduce a dependence to pile-up in the thresholds to apply. Indeed, the isolation energy includes some pile-up contribution which increases with the number of additional pile-up interactions and to avoid a reduction of the triggering efficiency with pile-up, this effect has to be taken into account. The variable n_{TT} is then defined as the number of trigger towers in the 8 central slices of the detector with a non-zero energy deposit. This variable has been shown to exhibit a strong correlation with the number of reconstructed pile-up vertices, as shown in Fig. 3.12. The pile-up contribution to the isolation energy is then taken into account by progressively relaxing the isolation thresholds as n_{TT} increases.

In practice, the isolation cut has been derived based on data in a given n_{TT} range. For each value of E_T, a target isolation efficiency has been imposed and the corresponding cut on the isolation energy has been derived for each possible value of η. An linear extrapolation to any possible value of n_{TT} has then been used, based on the correlation of the isolation

energy with n_{TT}. To ensure a high efficiency at high E_T, the isolation requirement is also relaxed with the L1EG E_T, by increasing the target efficiency together with E_T. This energy dependence can be adapted to different luminosity conditions in order to optimally combine rate reduction and triggering efficiency.

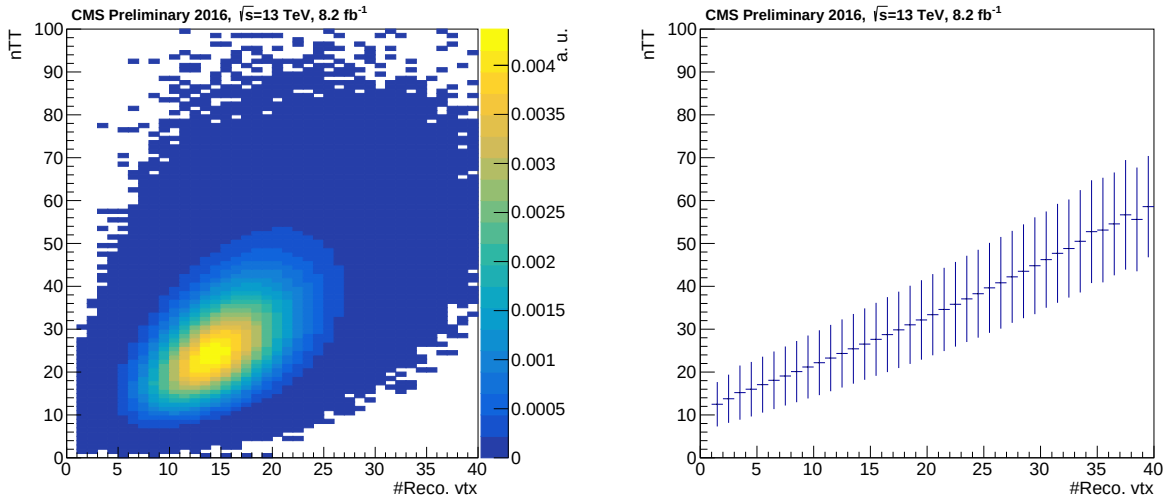

Figure 3.12: Distribution of n_{TT} and the number of reconstructed vertices (left) and average n_{TT} values as the function of the number of reconstructed vertices (right) in 13 TeV data.

3.3 Commissioning of the Stage-2 Level-1 EG trigger

3.3.1 2015 commissioning

The new Level-1 Stage-2 trigger was scheduled to be used for data-taking in CMS from 2016 onwards. 2015 has therefore been dedicated to the commissioning of the new system, which had to demonstrate that it would be ready for LHC restart after the winter shutdown. It has to be understood that for practical reasons, the trigger algorithms are first developed with a software implementation, mainly in C++. This software implementation is called "emulation". When I started my work on the upgrade of the L1EG trigger, the corresponding emulator was mostly finalized and had demonstrated already very good performance, compatible with the expected 2016 LHC conditions. Still, as mentioned previously, the Level-1 trigger in CMS is implemented on custom electronic boards and a firmware implementation had therefore to be developed to reproduce the software emulation. This firmware implementation is in fact much more complex than the software implementation as it has to be compatible with the TMT architecture and it has to be optimized such that all the calorimeter trigger algorithms are able to fit on the FPGA of a single MP7 board, which has a limited amount of resources. This implementation has

been done in tight collaboration with engineers in electronics using the VHDL programming language (Very high speed integrated circuit Hardware Description Language). The VHDL implementation can then be compiled into an executable which can be uploaded on the FPGA. This compilation is done using the Vivado software [144] and requires a powerful computer, on which the conversion is typically performed in twelve hours. Once this is done, the firmware can be uploaded on an MP7 board and tested using data to check the outputs of the algorithm. The test bench used at LLR for the tests of the L1EG firmware is visible in Fig. 3.13.

Figure 3.13: μTCA testbench used at LLR. On the left, the Vadatech μTCA crate, with the MHC board in the middle of the crate (used to download firmware) and the MP7 board on the right of the crate. On the right, the Dell computer used for the compulation of VHDL programs and the production of FPGA firmwares. [140]

The Level-1 EG firmware implementation has been first developed at LLR and I have been in charge of validating the firmware implementation with respect to the software implementation. The procedure adopted was to send to the board test patterns, generated from simulated events with two electrons or two gluons. The test patterns corresponded to data, as if produced by the Layer-1 (see section 3.1.3). Those data can then be processed on the FPGA and the outputs of the MP7 can be compared to the outputs of the software emulator. The possible discrepancies had to be understood and fixed, either by modifying the firmware or the software implementation, while maintaining the same level of physics performance. Fixing those discrepancies can be quite challenging as from the firmware side, only the outputs of the MP7 are available, i.e. the list of L1EG candidates produced with their position and energy, and the intermediary variables defined in the software emulation cannot be accessed in the FPGA. Fortunately, a virtual test bench simulation has been developed, which gives the possibility to check some of those variables directly using the VHDL implementation. After several iterations, it has been possible by the spring of 2015 to obtain a quasi-perfect agreement between emulator and firmware for the L1EG candidates, as seen in Fig. 3.14.

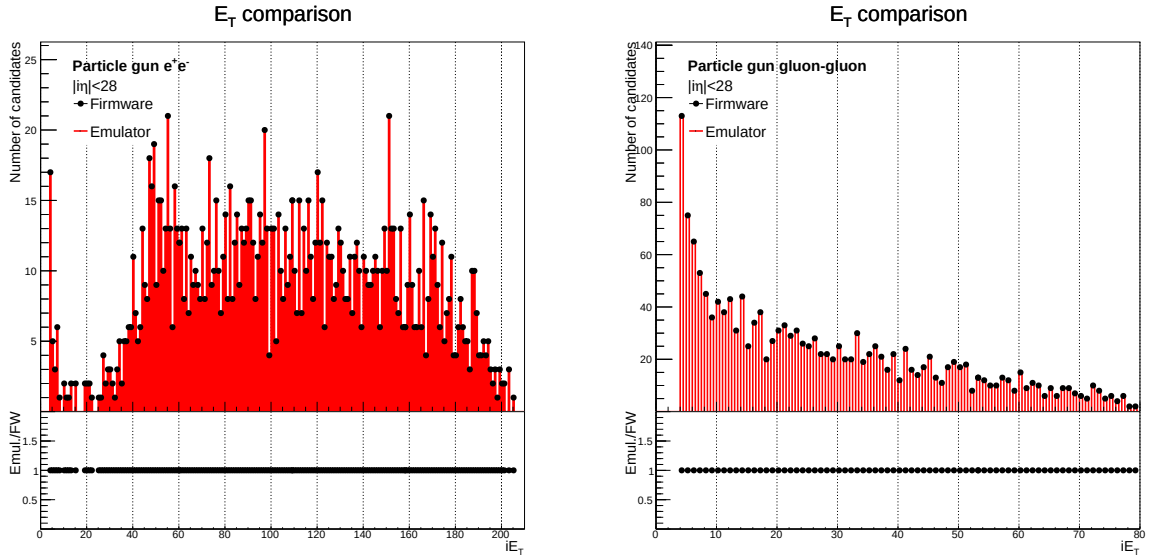

Figure 3.14: Comparison of the E_T (in firmware units) of the L1EG candidates between the emulator and the firmware outputs, for simulated events with two electrons (left) and two gluons (right).

At this point, the firmware implementation for the L1EG and L1Tau algorithms had been developed independently from the L1Jet and L1Sums algorithms. As all of those algorithms were intended to fit in the FPGA of a single MP7 board, some parts of those implementations have been revisited to better optimize the resource shared among the different blocks. For example, the original size of the isolation window for the L1EG algorithm was 5x9 trigger towers, while the L1Jet candidates were based on a 9x9 window, using intermediary sums of 3x9 windows. The L1EG isolation window has then been extended to a 6x9 window, computed as the sum of two 3x9 windows to better make use of quantities already computed in some other blocks. The structure of the whole Level-1 calorimeter trigger firmware after re-optimization is presented in Fig. 3.15. The horizontal links show how some results of the different flows are re-used in different blocks.

The last months of 2015 LHC proton-proton collisions have been dedicated to the validation of this new firmware. A parallel running had been designed for the Level-1 trigger system for that purpose. One trigger data-flow was going through the upgraded Stage-1 system, which had been deployed for the 2015 data-taking. The Level-1 trigger decision was based on the outputs of this system. Another parallel trigger data-flow was set up for monitoring purposes and was going through the Stage-2 system. All the Stage-2 data have been recorded for offline analysis, while an online Data Quality Monitoring (DQM) system had been deployed to monitor the Stage-2 system during data-taking. A sample of online DQM plots is presented in Fig. 3.16. The list of runs which have been analysed during the parallel running is presented in Table 3.2.

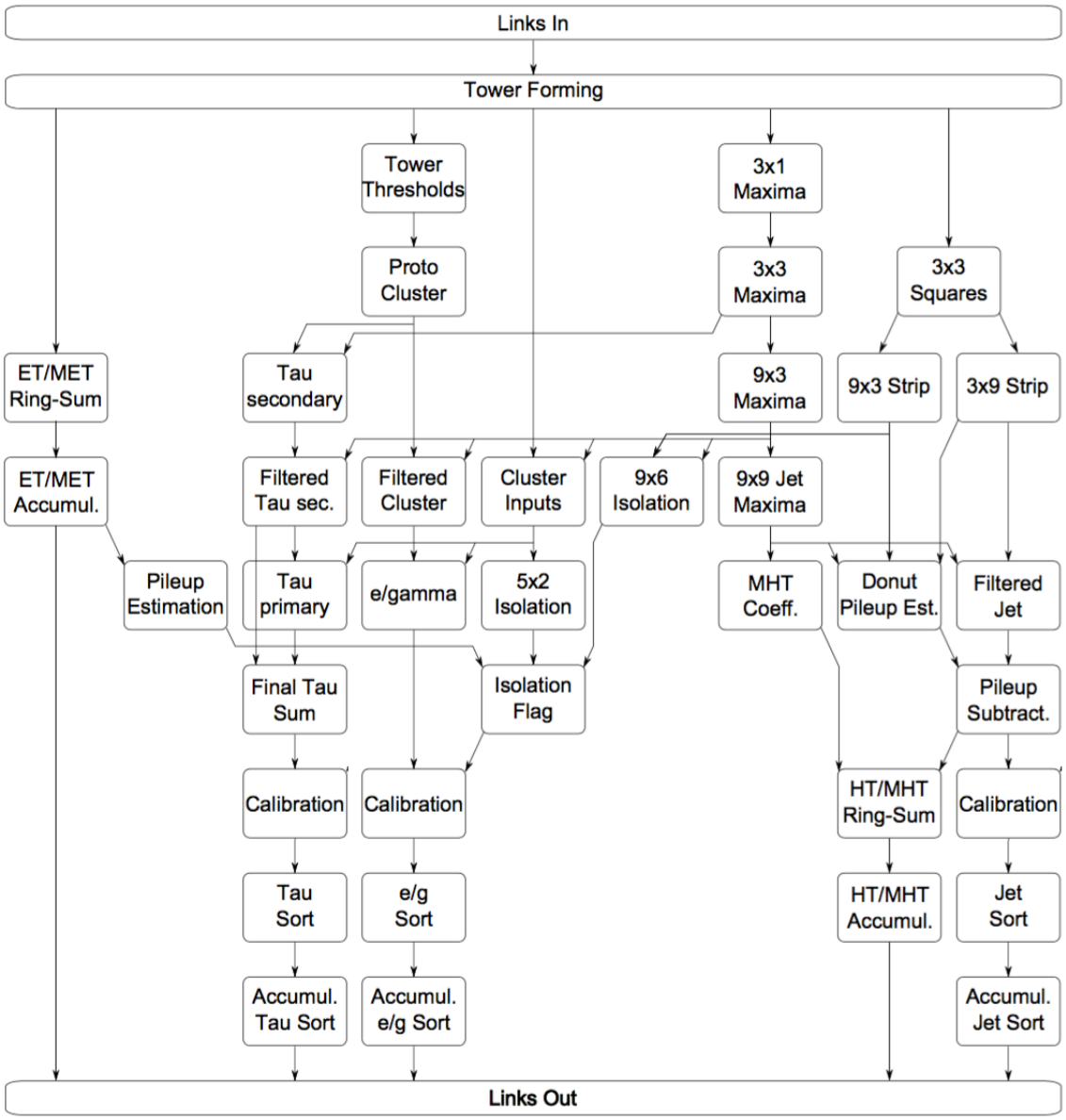

Figure 3.15: Final structure of the Level-1 calorimeter trigger, including all the calorimeter trigger algorithms (L1MET, L1EG, L1Tau, L1Jet and L1HT). The data flow goes from the top to the bottom. The different blocks correspond to computing steps run in parallel. Three main data streams can be identified, corresponding to the L1MET (left), the L1EG+L1Tau (middle) and L1HT+L1Jet (right). Five outputs streams are produced and sent to the Demux, corresponding to the L1MET, the L1Tau, the L1EG, the L1HT and the L1Jet candidates.

Figure 3.16: Sample of distributions available in the online DQM of the Level-1 Calorimeter Layer-2. On the top raw, from left to right: the distributions of L1 HTT, the 2D distribution of the E_T of L1IsoEG candidates as function of the bunch crossing, the E_T-weighted 2D distribution of the $\eta-\phi$ position of L1IsoEG candidates and the η distribution of L1IsoEG candidates. On the bottom raw, from left to right: the 2D distribution of the $\eta-\phi$ position of L1IsoEG candidates, the ϕ distribution of L1IsoEG candidates, the E_T distribution of of L1IsoEG candidates and the η distribution of L1Tau candidates.

This parallel running gave the opportunity to validate the Stage-2 firmware implementation in real data-taking conditions by comparing the firmware outputs with the corresponding emulator quantities. The corresponding validation plots are presented in Fig. 3.17, obtained from data collected with single electron triggers. A perfect agreement is obtained for the energy of the L1EG candidates. In the run 260576 (ECAL only inputs), a bump at 20 GeV (=40 in firmware units) is observed in the falling E_T distribution of the L1EG candidates. This corresponds to a bias due to the 20 GeV threshold used for the single electron trigger in that run. In the run 260627, inputs from ECAL and HCAL were used and this bias is not visible. The reason is that at this time, the H/E identification criteria for the L1EG electrons was not implemented in the firmware version deployed for the parallel running. Therefore any calorimeter energy deposit could produce a valid L1EG candidate. Since the hadronic deposits are produced in larger amount than the electromag-

Run number	Time	Luminosity	Notes
258741	30 min.	4.7 pb^{-1}	ECAL
258742	4.5 h	59 pb^{-1}	ECAL
259884	1 h	6.7 pb^{-1}	ECAL + HCAL
260424	4.3 h	63 pb^{-1}	ECAL
260490	30 min.	5.3 pb^{-1}	ECAL + HCAL
260493	20 min.	3.7 pb^{-1}	ECAL + HCAL
260576	3 h	48 pb^{-1}	ECAL
260627	12 h	180 pb^{-1}	ECAL + HCAL + HF
Total	26 h	370 pb^{-1}	

Table 3.2: List of runs recorded in the parallel running of the Stage-2 system. Those runs correspond to a total of over 7 billions events, which have been used for offline analysis.

netic ones and are not biased by the single electron trigger thresholds, the energy spectrum does not exhibit the same feature as the one from the run 260576. The energy spectrum stops at a firmware value of 300 due to the energy saturation of the L1EG candidates.

Concerning the position of the L1EG candidates, the agreement between emulator and firmware is quasi-perfect in the run 260576, while some discrepancies are observed in the run 260627, especially a trend with Φ. Those discrepancies are due to a known difference present at that time between the emulator and the firmware implementation of the sorting of the L1EG candidates, when some candidates have the same energy: in particular, the emulator implementation favors the candidates with higher Φ while the sorting in the firmware is uncorrelated with Φ. Since a limited number of L1EG candidates is produced (12 at most), the L1EG candidates may differ between emulator and firmware. Note that the energy spectrum is not affected as those differences only affect events where the trailing L1EG candidates have the same energy. The difference is mostly visible in the run 260627 because the L1EG multiplicity in that run was higher, since ECAL and HCAL inputs were used. The small remaining differences between emulator and firmware were fixed during the subsequent LHC end-of-year technical stop.

3.3.2 Preparation of the 2016 data-taking

Before the LHC restart in 2016, I tuned all the parameters (mostly encoded in LUTs) to make the L1EG algorithm ready for 2016 data-taking. This final optimization has been mostly based on data taken during the parallel running and the performance has been checked in data and simulation. The calibration has thus been rederived using an MVA-based regression to obtain the LUT encoding the correction factors to applied to the raw L1EG E_T. The improved L1EG resolution with the Stage-2 trigger has thus been confirmed, as visible in Fig. 3.18. The isolation LUT has also been rederived using 2016 data.

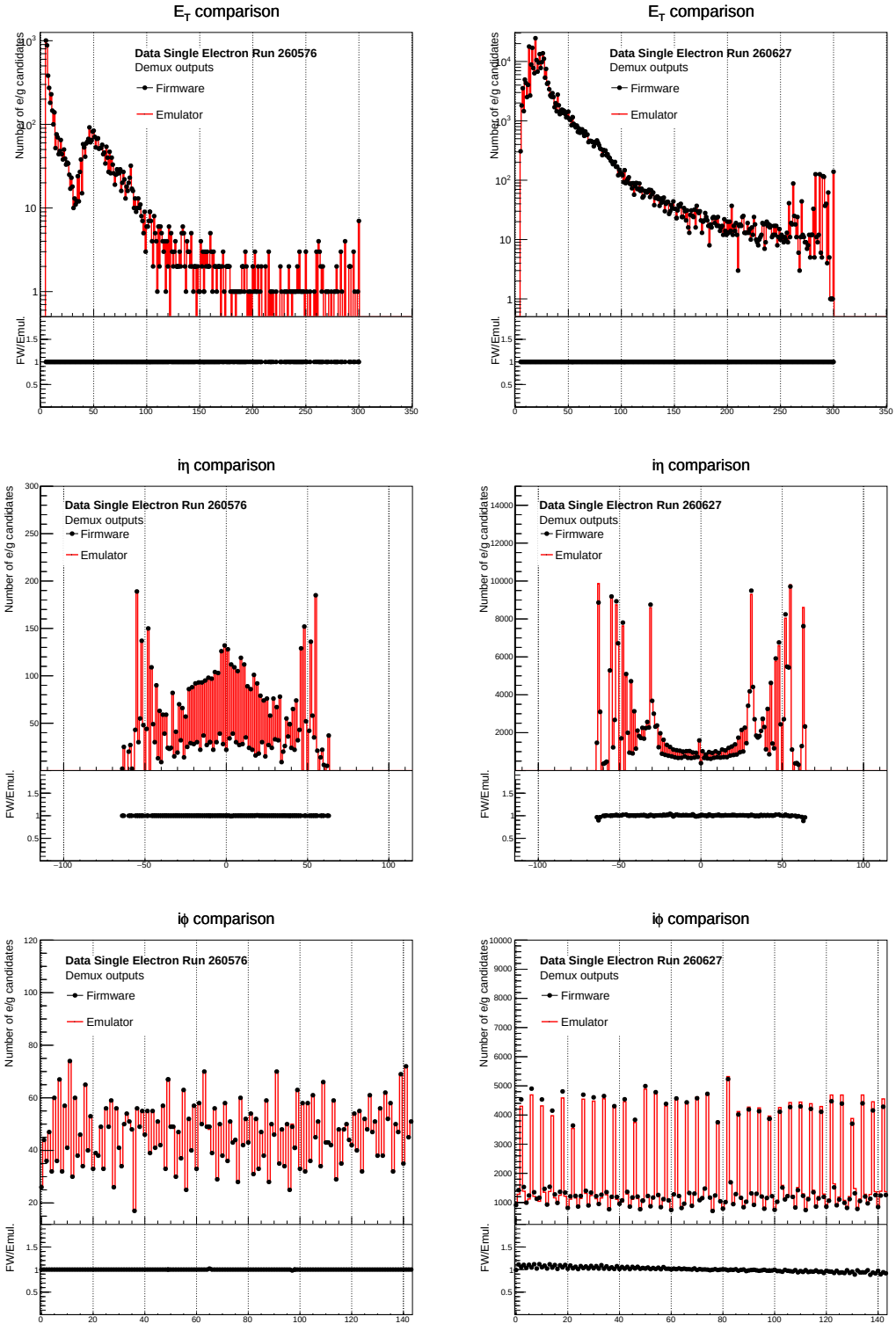

Figure 3.17: Comparison of L1EG firmware and emulator outputs for the E_T (top), η (middle) and ϕ (bottom) variables, from run 260576 (ECAL only, left column) and run 260627 (ECAL+HCAL, right column).

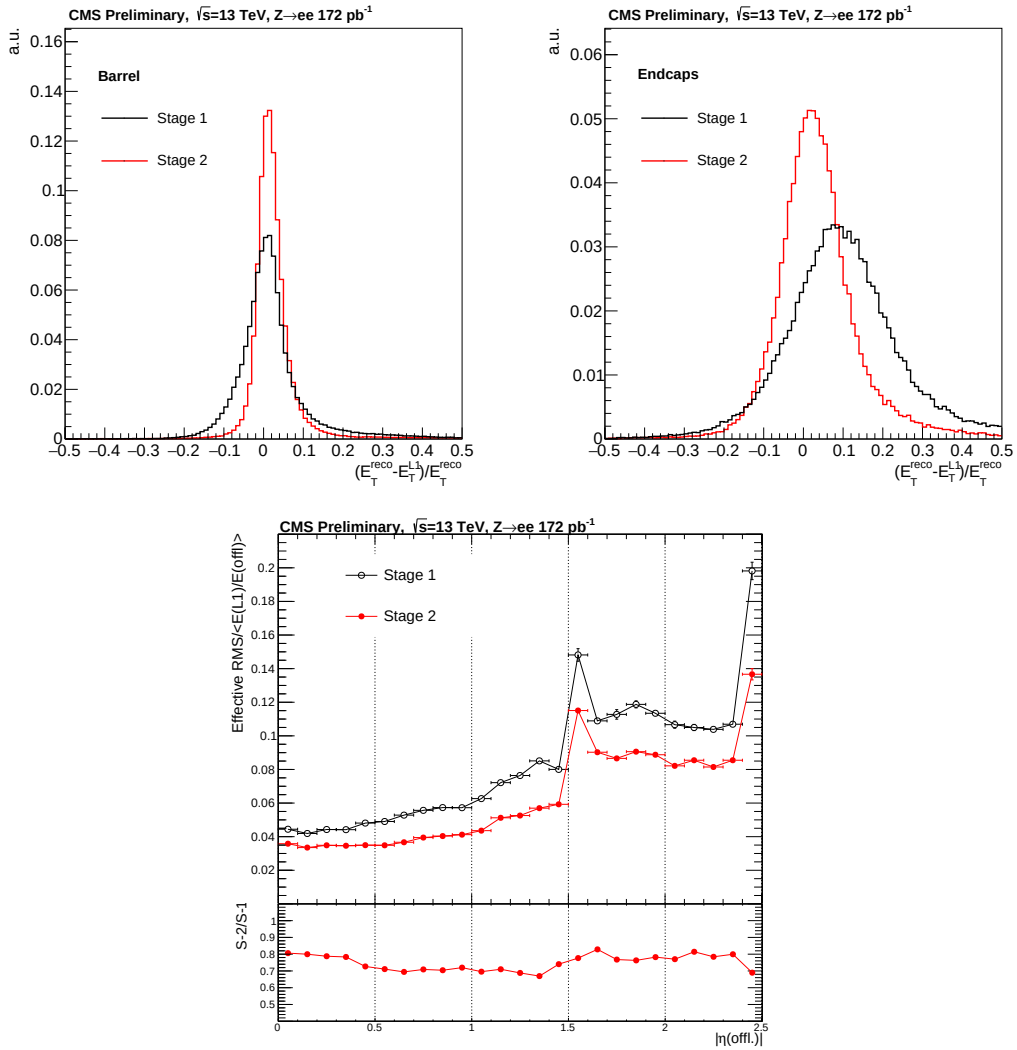

Figure 3.18: L1EG relative energy response in the barrel (top left), in the endcaps (top right) and L1EG resolution as function of the pseudorapidity (bottom) obtained with the Stage-1 system (black) and the re-optimized calibration of the Stage-2 system (red) in 2015 data. To define the L1EG resolution, an effective RMS is defined from a fit of the relative energy response with a double Crystal-Ball function and is divided by the average energy response.

Measuring the efficiency of a trigger in data requires actually some care to avoid any bias in the measurement. Indeed, data are recorded using one of the available triggers. If the efficiency of the single electron trigger was to be measured in data using events triggered with the same single electron trigger, the measurement would be biased: for instance, it would not be possible to spot any inefficiency of this trigger, since the events where the

single electron trigger has not fired would not be included in the data sample considered. One possible solution would be to use events triggered with a dedicated ZeroBias (i.e. random) trigger or an uncorrelated trigger (for instance a muon trigger). The problem in that case is that a very low number of offline electrons is expected to be found in such events and a very large number of events would have to be collected to get a reasonable statistical error on the efficiency measurement. A solution widely adopted in experiments on particle colliders is the so-called tag-and-probe method. This method takes advantage of a physical process in which the particle of interest (the probe) is produced in association with another particle (the tag). The tag is then used to trigger the event while the probe is used to provide an unbiased measurement of the trigger efficiency.

In the case of electrons, the process commonly chosen for the Tag-and-Probe method is the decay of a Z boson into a pair of electrons. The events used are selected among those triggered with a single electron trigger. Among those, additional requirements are applied to select those compatible with a Z boson decay: those events must contain two offline reconstructed electrons with opposite charge and an invariant mass compatible with the Z boson mass. Additional electron identification criteria are also applied to increase the purity in real electrons in those events. One of the electrons is then flagged as the tag while the other is flagged as the probe. A selected event can therefore have two interpretations in terms of Tag-and-Probe, depending on which electron is chosen as the tag. Both interpretations can be considered for the efficiency measurement to reduce the statistical error. The tag must then be compatible with a triggering electron, using some matching criteria. In the case of the L1EG trigger, the position of the offline electron supercluster (see section 2.3.3) was thus required to correspond to the seed trigger tower of the L1EG cluster or one of its neighboring towers.

For the L1EG efficiency measurements presented here, the events are selected among those firing the HLT_Ele30WP60_Ele8_Mass55 and HLT_Ele30WP60_SC4_Mass55 paths. Those paths require the presence at the HLT level of an electron with $p_T > 30$ GeV, passing the 60%-efficiency working point of the electron ID. The event must also contain an additional electron with $p_T > 8$ GeV or an ECAL supercluster with $E_T > 4$ GeV and the invariant mass of the "di-electron" system must be larger than 55 GeV. Those HLT paths are seeded by Level-1 SingleEG triggers and a geometrical matching of the offline tag electron is required with an L1EG candidate compatible with the Level-1 SingleEG seeds. The electrons are required to be within the ECAL fiducial volume (with $|\eta| < 2.5$, excluding the crack between the barrel and the endcaps). The pair of opposite charge electrons must have an invariant mass between 60 and 120 GeV. The tag electron is required to have $E_T > 30$ GeV and to pass the medium working point of the offline electron identification while the probe electron must only pass the loose working point of the electron identification.

One of the usual tools used to evaluate the performance of the L1EG trigger is the so-called "turn-on" curve, representing the L1 SingleEG efficiency for a given threshold as function of the E_T of the offline electron ECAL supercluster (see section 2.3.3). The SingleEG20 turn-on curves obtained after re-emulation on 2015 data with the new L1EG parameters to be used for the 2016 data-taking is presented in Fig. 3.19. The turn-on curves present a sharp increase centered on the L1 threshold value related to the good

L1EG energy resolution. The turn-on is sharper in the barrel than in the endcaps because of the better energy resolution in the barrel. Moreover, the efficiency of the SingleIsoEG20 trigger is lower than for the inclusive SingleEG20 trigger but gets closer to 1 as E_T increases due to the progressive relaxation of the isolation criteria.

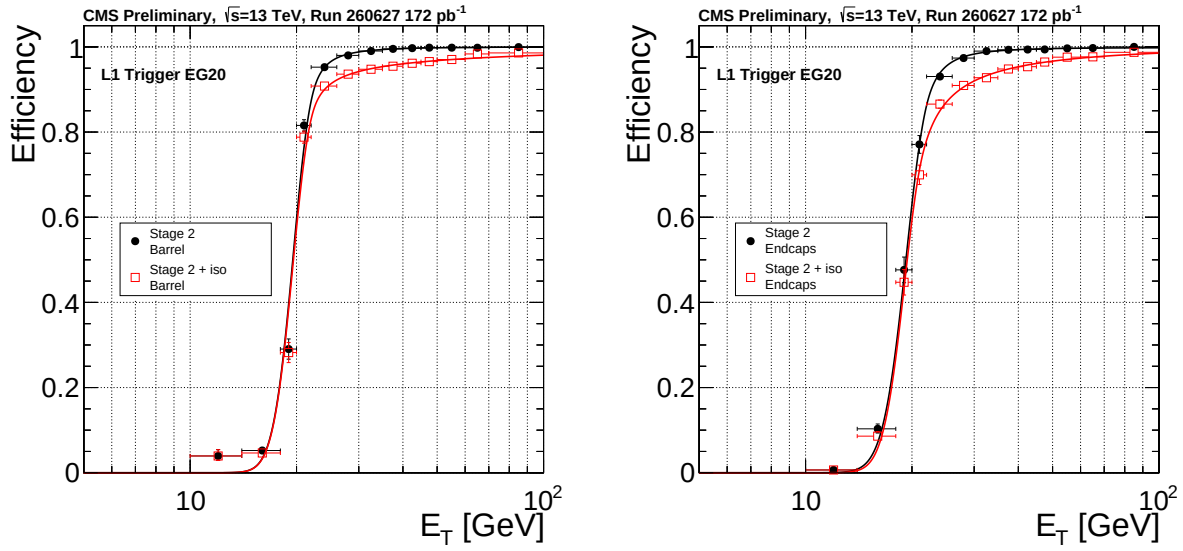

Figure 3.19: L1EG turn-on curves for a 20 GeV-threshold obtained in 2015 data with the re-optimized set of L1EG parameters in the barrel (left) and in the endcaps (right), without (black) and with (red) isolation requirement. The lines correspond to the result of an unbinned maximum likelihood with a Crystal-Ball integral function.

The updated sets of parameters have then been used first for data-taking in cosmic runs a few weeks before the LHC restarts in spring 2016. The so-called "splashes" have then been another important test for the upgraded Stage-2 trigger. The splashes correspond to injection tests performed by the LHC, where the proton beams collide with the collimator blocks in front of the detectors, which produces a very large multiplicity of very energetic particles. The splashes produce a much harsher environment than what is produced in usual proton-proton collisions and are mainly used for detector monitoring and timing studies. For the first run of splashes, it has been decided to use the new L1EG trigger for data-taking (with a 5 and 20 GeV thresholds), with only the two innermost rings of ECAL trigger towers unmasked for that test, in order to reproduce conditions compatible with central collisions. Unfortunately, this test was only a half-success as the L1_SingleEG20 algorithm did not fire. The L1_SingleEG5 triggered though in what was believed to be the "aftermath" of some splashes, catching the trailing particles produced in the splash event. After a prompt analysis of the 2015 splashes data, I've been able to identify a feature of the algorithm explaining why the L1EG algorithm failed to trigger in the configuration deployed for the splashes. During splashes, all the trigger towers are saturated because of the very large energy deposit in the whole detector (all the towers are produced with an energy of 127.5 GeV). The algorithm used to define L1EG seeds from local energy maxima (see Fig. 3.7 left) fails then to produce a valid seed if all the towers have the same energy because of the ϕ symmetry of the detector. In normal conditions during proton-proton collisions, this is not a problem as the electrons and photons produce localized energy deposits. (It is actually more appropriate to trigger on those events with L1 sums algorithm.) To take that feature into account, I proposed to mask 4 consecutive towers in the two rings to break the ϕ symmetry and to use additional L1 algorithms, including the L1 sums, as back-ups. With that feature implemented, the L1EG algorithm successfully triggered on the second run of splashes. Event displays of those splash events are presented in Fig. 3.20. The L1EG trigger had then demonstrated its complete readiness for the 2016 data-taking and was monitored in the following weeks until the beginning of the 2016 LHC proton-proton collision run.

3.4 Measurement of the Stage-2 Level-1 EG trigger performance in 2016 data

During the first months of 2016 data-taking, I continued the monitoring of the L1EG trigger performance. I have been in charge of the weekly data certification for the L1EG. Benefitting from the large luminosity collected, I have been able to measure the performance of the L1EG trigger in the first 0.5 fb^{-1} collected in 2016 [145] then with the first 8.2 fb^{-1} collected in 2016 [146]. Those results are shown in Fig. 3.21 - 3.26. The rate measurements have been performed using ZeroBias data, while for the efficiency measurements, the Tag-and-Probe method detailed in the previous section is used.

All those results are in agreement with the expected level of performance of the up-

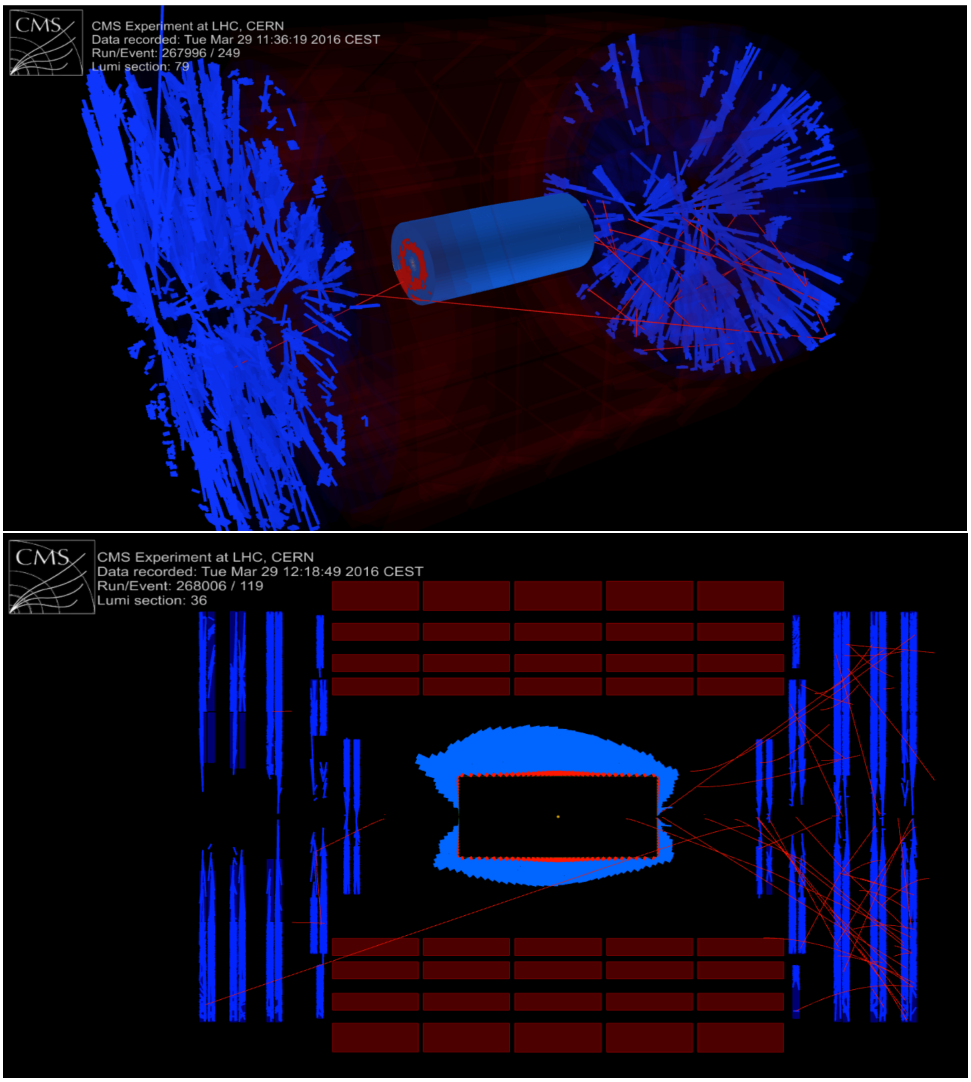

Figure 3.20: Event displays of splash events triggered with the upgraded L1EG algorithm. The light red (blue) areas correspond to ECAL (HCAL) energy deposits, while the red lines correspond to reconstructed muon tracks.

graded Level-1 EG trigger. The position and energy resolutions, presented in Fig. 3.21 and 3.22, are thus significantly improved with respect to the Run 1 L1EG trigger. The different position and energy resolutions between barrel and endcaps can be explained by the different geometries of the trigger towers, which are larger and have a more complex geometry in the endcaps (see Fig. 3.3). The energy resolution is also considerably deteriorated in the transition region between barrel and endcaps because of the large amount of material, which induces a large amount of energy loss by Bremsstrahlung. These good position and energy resolutions will be possibly exploited in the future to implement invariant mass triggers, which should be developed as the LHC luminosity increases.

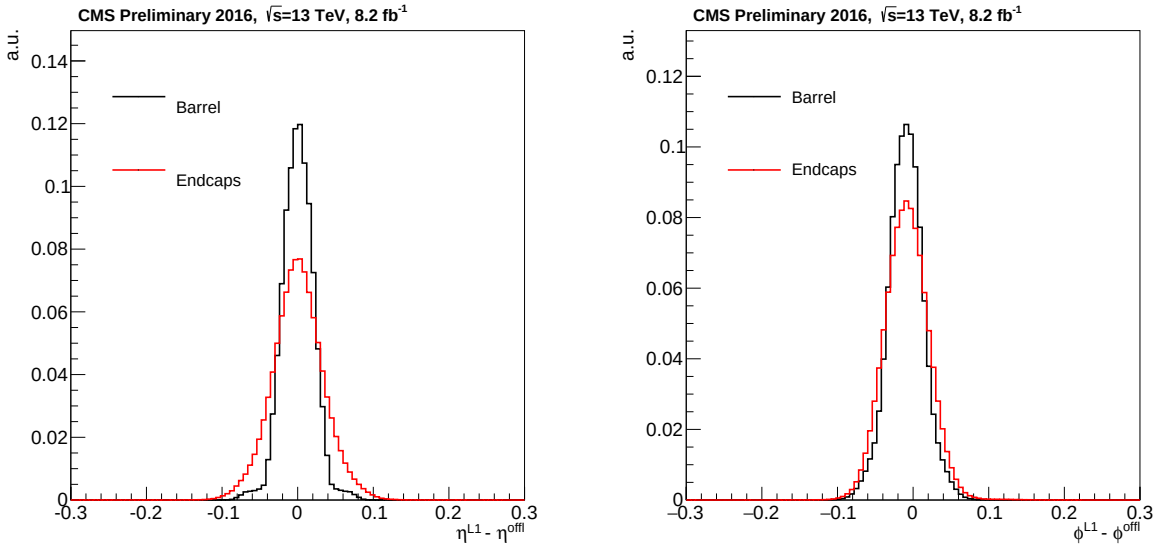

Figure 3.21: Difference in pseudo-rapidity η (left) and azimuthal angle ϕ (right) for L1 EG candidates with respect to the offline reconstructed electron supercluster, in the barrel (black) and in the endcaps (red), measured in 2016 data.

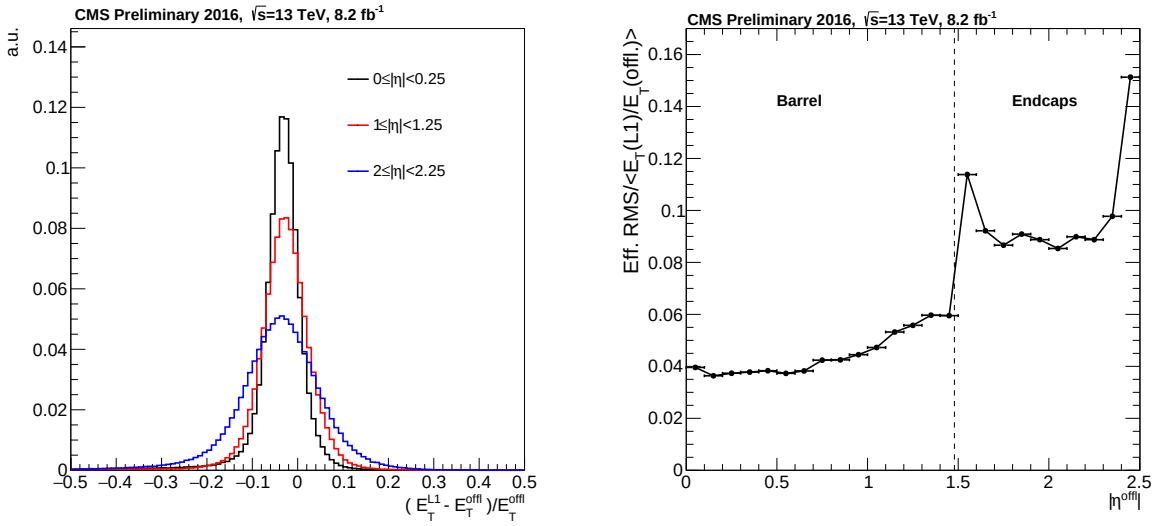

Figure 3.22: L1EG relative energy response in different η bins (left) and L1EG resolution as function of the pseudorapidity (right) measured in 2016 data. To define the L1EG resolution, an effective RMS is defined from a fit of the relative energy response with a double Crystal-Ball function and is divided by the average energy response.

As seen in Fig. 3.23 and 3.24, the turn-on curve without isolation requirement exhibits a sharp increase, in particular in the barrel, thanks to the very good energy resolution of the L1EG trigger. The isolation requirement, slightly tighter in the barrel, induces a slower turn-on but the lower efficiency is compensated by a lower rate, as visible in Fig. 3.25. The rate is thus reduced with respect to the inclusive L1EG trigger by up to a factor 2 for a 24 GeV threshold. The isolation relaxation enables to still get a very high efficiency at high energy and explains why the rate with isolation gets closer to the inclusive rate at high thresholds (above those used for data-taking). As the LHC instantaneous luminosity gets higher, it will be possible to re-optimize the isolation requirement with a new LUT to keep reasonable SingleEG thresholds. Moreover, the trigger efficiency stays relatively stable across different pile-up conditions, as seen in Fig. 3.26. This is the case even when the isolation requirement is applied, thanks to the use of an online pile-up estimate.

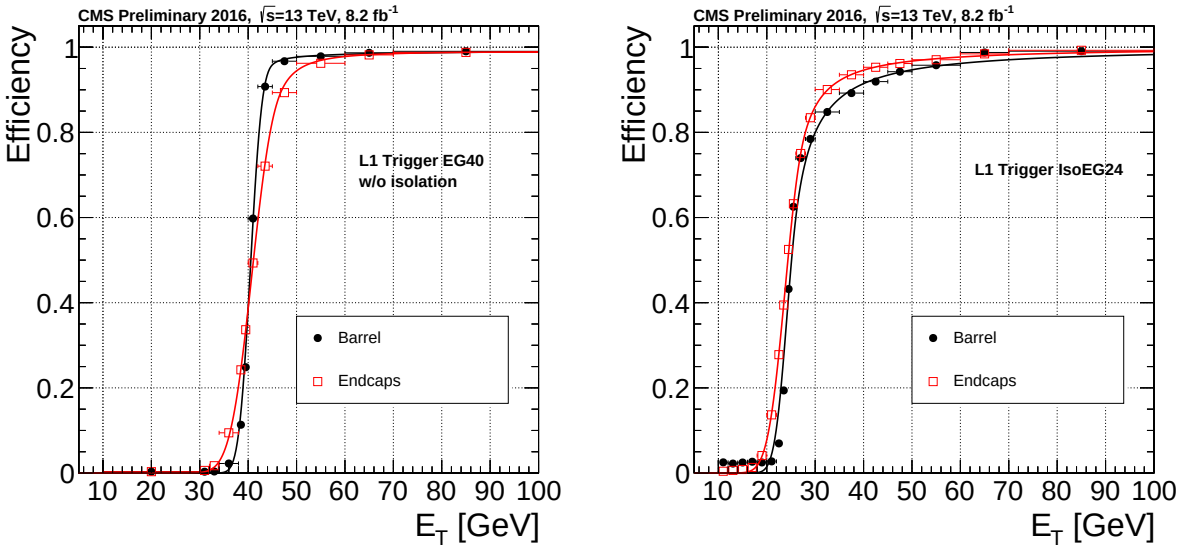

Figure 3.23: L1EG turn-on curves measured in 2016 data for a 40 GeV-threshold without isolation requirement (left) and for a 24 GeV-threshold with isolation requirement (right), in the barrel (black) and in the endcaps (red). The lines correspond to the result of an unbinned maximum likelihood with a Crystal-Ball integral function.

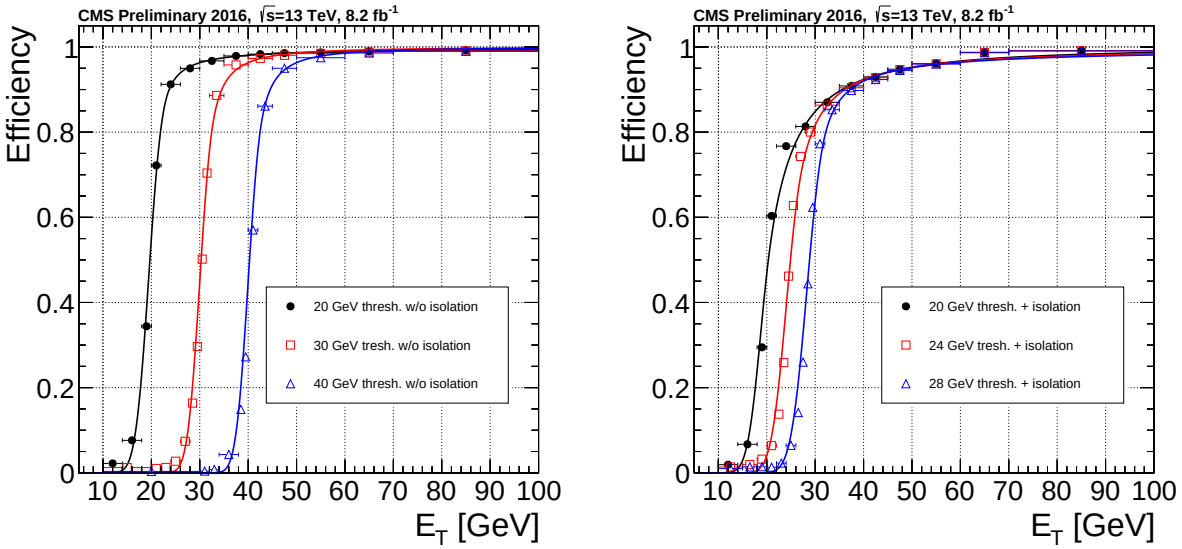

Figure 3.24: L1EG turn-on curves measured in 2016 data for different thresholds without isolation requirement (left) and with isolation requirement (right). The lines correspond to the result of an unbinned maximum likelihood with a Crystal-Ball integral function.

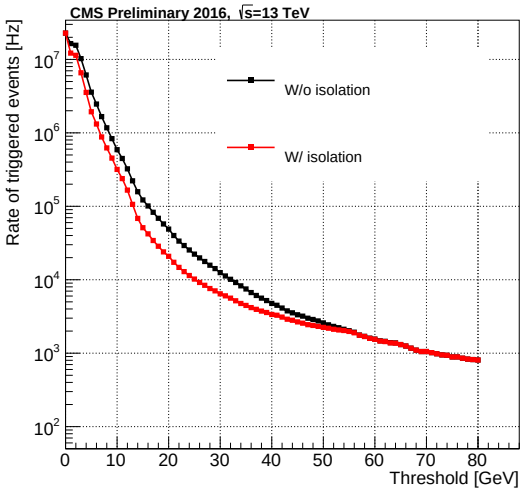

Figure 3.25: Rate of triggered events from ZeroBias data from 2016 run with initial luminosity of $7.2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$, without isolation requirement (black) and with isolation requirement (red). The lowest unprescaled SingleEG thresholds in that run were SingleEG40 (4.8 kHz) and SingleIsoEG24 (11.4 kHz).

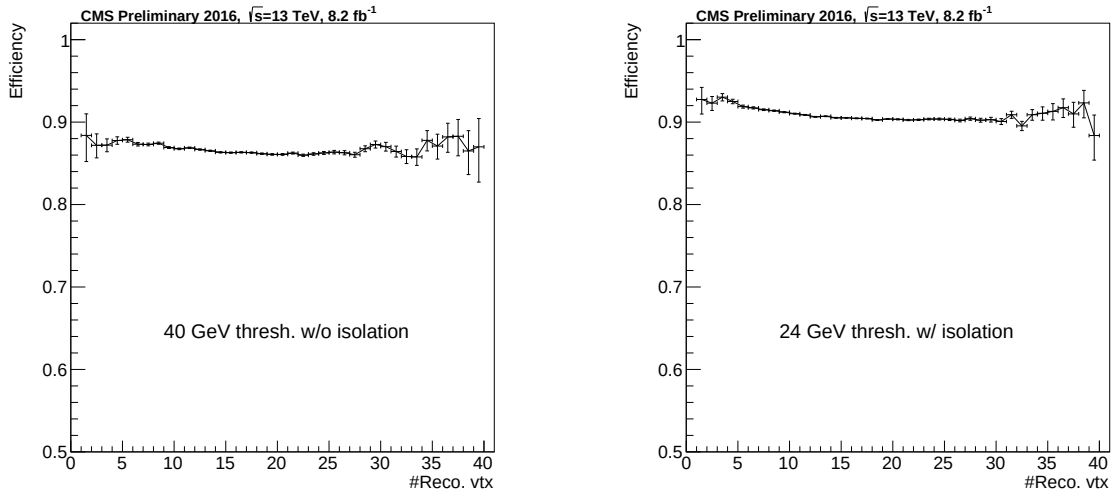

Figure 3.26: Trigger efficiency as function of the number of reconstructed vertices measured in 2016 data, for a 40 GeV-threshold without isolation requirement (left) and for a 24 GeV-threshold with isolation requirement (right). The efficiency in the left (right) plot is defined as the fraction of offline reconstructed electrons with $E_T > 40$ GeV ($E_T > 24$ GeV) associated with a L1EG (isolated) candidate passing the corresponding threshold.

3.5 Preparation of the 2017 data-taking

In the last months of my PhD, I also contributed to develop some improvements before the 2017 data-taking. These improvements were motivated by the following observation. In the original L1EG algorithm, the H/E identification was based only on the information related to the seed trigger tower of the cluster. However, as seen in Fig. 3.27, the neighbouring towers for background candidates can also be associated with a sizable hadronic energy deposit, even after applying the H/E criteria for the seed tower. A possible extension of the H/E criteria to include also the neighboring towers has therefore been investigated.

Applying the same H/E criteria as for the seed to all the neighboring towers was found to cause a large reduction of efficiency, because of low-energy towers with hadronic activity. A loose H/E criteria, corresponding to H/E<1, is therefore used for towers with transverse energy above 5 GeV, while no requirement is applied to towers with lower energy. This new extended H/E criteria is then applied in addition to the already existing one used for the seed tower. The inefficiency due to this new criteria has been measured to be very small, while bringing a rate reduction of up to 20% for SingleEG triggers, as seen in Fig. 3.28. This improvement benefited to all the triggers using EG objects and helped to keep the trigger thresholds for 2017 at a similar level as in 2016, in spite of an expected increase in instantaneous luminosity by 30%, up to $2 \times 10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$. This new feature also demonstrated the advantage of using tower-level information together with the flexibility in implementation of the Stage-2 calorimeter trigger, allowing incremental changes to be

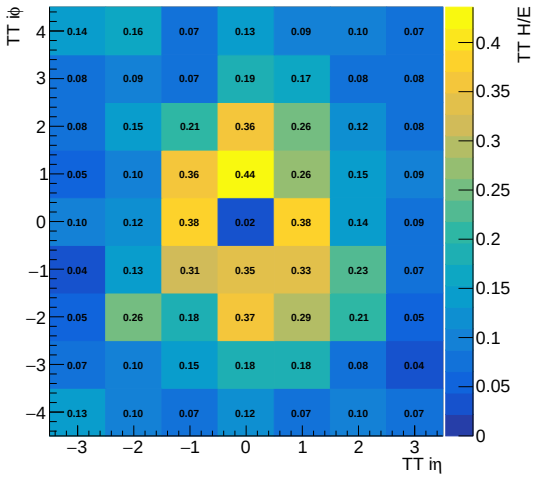

Figure 3.27: H/E ratio (z-scale) distribution for L1EG candidates in ZeroBias data. The position of the seed trigger tower corresponds to (iη,iϕ)=(0,0).

implemented in the firmware to improve the performance of all the trigger objects.

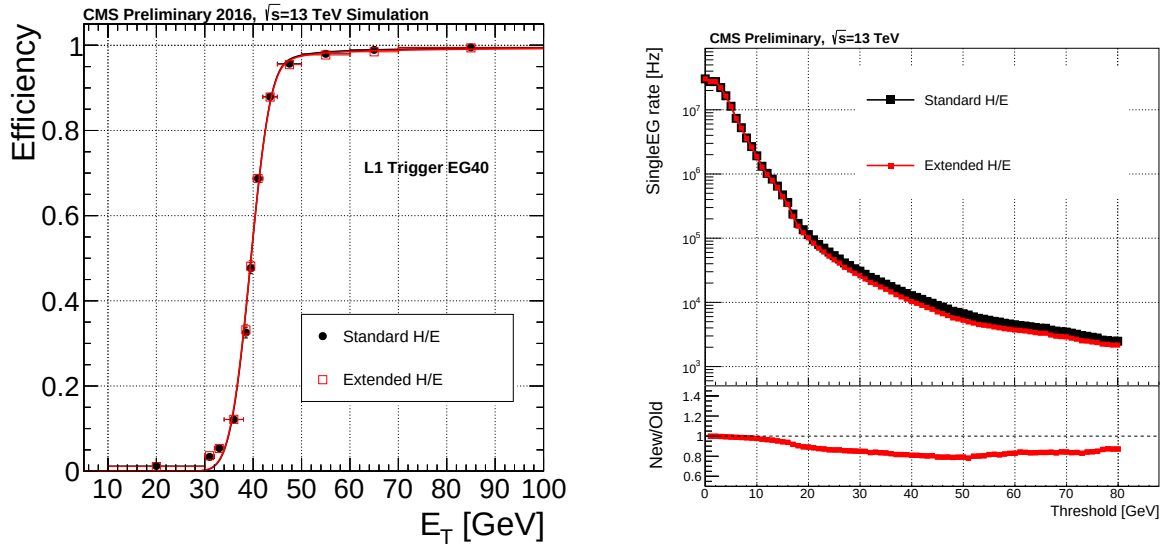

Figure 3.28: L1EG turn-on curves measured in simulation for a 40 GeV-threshold without isolation requirement (left) and SingleEG rate as function of the threshold (right) with the standard H/E used in 2016 (black) and with the new extended H/E developed for 2017 (red).

3.6 Conclusion

The upgrade of the Level-1 trigger and in particular of the L1EG trigger represented a critical achievement for the CMS collaboration in view of the end of the LHC Phase 1. Thanks to the new trigger architecture deployed and the new implementation of the L1EG algorithm, the performance of the L1EG trigger has been substantially improved. The expected level of performance of the upgraded L1EG trigger has thus been achieved and benefited to a wide range of physics analyses in CMS.

The experience gained with the Phase I trigger upgrade has shown that significant gains in performance can be achieved by bringing the reconstruction techniques at Level 1 closer to those used offline, as shown with the dynamic clustering used for L1EG. For the Phase II trigger upgrade, the Level 1 trigger will benefit from a major improvement thanks to the tracker information which will be available at Level 1. This gives in principle the possibility to implement reconstruction techniques at the trigger level very close to the offline Particle Flow algorithm, together with more powerful pile-up mitigation techniques, which will be the biggest challenge for Phase 2 with an average pile up expected to go above 120. In that context, the experience gained with the TMT architecture will probably be beneficial in order to develop sophisticated algorithms using inputs with high granularity and having access to the whole detector information.

Chapter 4

Matrix Element Method for $H \to \tau\tau$ analyses

What is the Matrix?

Neo

Physics analyses pursued in high energy physics typically involve several particles in the final state, associated to a large number of observables. One of the main issues to tackle is then to identify relevant variables which can be used to discriminate the signal from the background processes. Multivariate analysis has therefore become increasingly relevant in particle physics, as the signal process searched in LHC experiments for instance has become more and more complex. One of the developed approaches is the so-called Matrix Element Method (MEM).

Analyses involving a Higgs boson decaying into τ leptons with complex final states including jets, leptons, hadronically decaying τ's and missing transverse energy can typically benefit from dedicated multivariate discriminants. A lot of the MEM tools are actually common to all those analyses and will be presented in details in this chapter. The 8 TeV analysis focusing on the Vector Boson Fusion production of a Higgs boson decaying into τ leptons presented in [147] has been the first one to use a MEM implementation optimized for a resonance decaying into τ leptons. It benefited from the work detailed in this chapter where an implementation of the MEM optimized for the search of the $t\bar{t}H, H \to \tau\tau$ process is detailed. The results of this analysis will be presented in the last chapter of this thesis.

4.1 Introduction to the Matrix Element Method

Among all the available multivariate approaches currently available in high energy physics, the Matrix Element Method (MEM) is probably the one which tries to exploit the most the link between theory and event reconstruction in the detector. One of its first applications has been the measurement of the top quark mass at the Tevatron [148]. It has been since

adapted to other analyses, such as the $H \to 4\ell$ search in CMS [56], in the form of the "MELA" approach, or the search for a Higgs boson decaying into b quarks in the ttH production mode [149].

The goal of the MEM is to provide a discriminating variable $\mathcal{L}(\mathbf{y})$ which optimally uses the available information from the objects reconstructed in the detector, associated to a set of observables $\mathbf{y}$. According to the Neyman-Pearson lemma [150], the optimal test statistic for comparing two hypotheses H_0 and H_1 is provided by the likelihood ratio

$$\Lambda(\mathbf{y}) = \frac{L(H_0|\mathbf{y})}{L(H_1|\mathbf{y})} \tag{4.1}$$

where $L(H_0|\mathbf{y})$ (respectively $L(H_1|\mathbf{y})$) is the likelihood for the hypothesis H_0 (H_1) as a function of the data $\mathbf{y}$. In the case of the MEM, H_1 would be that a given event has been produced by a signal process while H_0 would be that this event has been produced via a background process. The MEM aims then at providing an estimate for a given event of the likelihoods $L(H_1|x)$ and $L(H_0|x)$.

To do so, the MEM makes use of the differential cross sections and decay widths predicted by the theory, by using the matrix element $\mathcal{M}_\Omega(\mathbf{x})$ associated to a given process Ω (hence the name), which depends on a set of variables $\mathbf{x}$ which describes the initial- and the final-state particles at a parton level. The squared matrix element is then convoluted with a function $W(\mathbf{y}|\mathbf{x})$ associated to the event reconstruction in the detector and called transfer function. It represents the probability that a set of observables $\mathbf{y}$ is measured given a parton-level configuration $\mathbf{x}$.

More specifically, the MEM aims at computing the differential cross section of the process Ω with respect to the observables $\mathbf{y}$, while integrating over the unmeasured or poorly measured parton-level quantities $\mathbf{x}$, as well as the Bjorken fractions of the incoming partons x_a and x_b. The differential cross section, also called weight in the case of the MEM, is then computed as

$$w_\Omega(\mathbf{y}) = \frac{1}{\sigma_\Omega} \sum_p \int d\mathbf{x} dx_a dx_b \frac{f_i(x_a, Q) f_j(x_b, Q)}{x_a x_b s} \delta^2(x_a P_a + x_b P_b - \sum p_k) |\mathcal{M}_\Omega(\mathbf{x})|^2 W(\mathbf{y}||\mathbf{x}) \tag{4.2}$$

The factor σ_Ω can be fixed by the condition $\int d\mathbf{y} w_\Omega(\mathbf{y}) = 1$ and corresponds to the product of the cross section of the process Ω with the acceptance of the analysis. Because of that normalization condition, all the overall multiplicative factors, which do not depend on $\mathbf{x}$ or $\mathbf{y}$ in the integrand can be omitted, as they would anyway simplify with σ_Ω. The sum $\sum_p$ corresponds to a sum over all the possible assignements of the reconstructed objects to the parton-level objects p (a so-called permutation). In some simple cases, it is actually possible to include those permutations in the computation of the matrix element squared $|\mathcal{M}_\Omega(\mathbf{x})|^2$, as it will be explained in Section 4.3. The functions $f_i(x, Q)$ are the parton distribution functions associated to the incoming parton of flavor i, depending on their Bjorken fraction x and the scale of the process Q.

The four-vectors $P_{a,b}$ are those of the colliding protons at a center-of-mass energy $\sqrt{s}$, while the four-vector $\sum p_k = P$ is the sum over all the final-state particles. The conservation

of energy and of longitudinal momentum is ensured thanks to the 2-dimensional δ function. The integrations over x_a and x_b can then be simplified, imposing

$$x_{a,b} = \frac{P^0 \pm |P^3|}{\sqrt{s}} \tag{4.3}$$

The total transverse momentum of the system is not constrained to be exactly zero to account for a possible boost in the transverse plane of the final-state system, due for instance to initial-state or final-state QCD radiations, which would require a proper treatment of higher order corrections in the matrix element. The approach adopted here is instead to use a leading order matrix element for the hard process and to constrain loosely the total transverse momentum of the system with a transfer function relating the measured recoil of the system to the recoil at parton level.

As the reconstructions in CMS of the different objects are mostly uncorrelated, the overall transfer function $W(\mathbf{y}|\mathbf{x})$ can be split into a product of different transfer functions associated to each individual objects. The different terms involved will be described in details in the next sections. Finally, the integration is performed over the phase-space spanned by the final-state particles $d\mathbf{x}$. In the general case, this can be decomposed as a product over the final-state particles $d\mathbf{x} = d\prod \mathbf{x}_k$ with

$$d\mathbf{x}_k = \frac{d^3\vec{k}}{(2\pi)^3 2E_k} \tag{4.4}$$

Because of the non-analytic parametrization of the function to integrate, the integration must be performed numerically event by event. This can be done using the VEGAS algorithm [151], briefly described in the next section. The hard scattering amplitude can be evaluated thanks to dedicated code generated using standard high-energy physics Monte-Carlo generators, such as MadGraph [152]. Finally, the transfer functions can be defined either analytically from theoretical considerations or parametrically, using in situ estimates for the detector resolution.

Each weight $w_\Omega(\mathbf{y})$ represents the probability to measure the set of observables $\mathbf{y}$ given the process Ω. Those weights are then usually combined in a likelihood ratio of the form

$$\mathcal{L}(\mathbf{y}) = \frac{w_S(\mathbf{y})}{w_S(\mathbf{y}) + \sum\limits_{B\in\mathcal{B}} w_B(\mathbf{y})} \tag{4.5}$$

with S the signal process and $B \in \mathcal{B}$ the set of background processes considered in the MEM. This variable, whose value is between 0 and 1, can then be interpreted as the probability that a given event associated to $\mathbf{y}$ originates from the process S, rather than from one of the background processes B.

The Matrix Element Method offers in principle a large number of advantages which can justify its use in particle physics analyses. First of all, it is in principle universal and can be applied to a wide range of processes for which a theoretical prediction exists for their amplitude (even though if the detailed implementation of the MEM may have to

be specialized for each use case). The combination of this theoretical information with the detector resolution is then done in an optimal way. Moreover, in contrast with other multivariate methods, like Boosted Decision Trees [126], no training is required for the MEM, which is really beneficial when no large statistics Monte-Carlo samples are available for the processes investigated. On the other hand, the numerical integration required to compute the final discriminating variable makes this method heavily CPU time consuming, even though progress in the implementation of this method using GPU's makes its use more practical. In addition, the MEM cannot be used directly in the context of model-independent searches but it is perfectly suited for searches of Standard Model processes with a single signal hypothesis.

4.2 Numerical integration with the VEGAS algorithm

To understand the necessary steps required to implement the computation of a MEM discriminant, it is necessary to present how a numerical integration can be performed using Monte-Carlo methods, in particular with the VEGAS algorithm [151]. The goal of such a method is to compute a n-dimensional integral

$$I = \int_{\Omega} d^n\mathbf{x} f(\mathbf{x}) \tag{4.6}$$

with $f(\mathbf{x})$ the function to integrate and Ω the integration space. Generic Monte-Carlo integration methods aim at approximating this integral as an average of the values taken by the function $f(x)$ over a large number N of points $\mathbf{x}_i$ (also called integration points or shots) within Ω, randomly chosen with a density $\rho(\mathbf{x})$

$$I \approx S = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{\rho(\mathbf{x}_i)} \tag{4.7}$$

Successive estimates $(S_j)_{1\leq j\leq m}$ are then made with the same number of integration points N using formula (4.7) to define a cumulative estimate $\bar{S}$

$$I \approx \bar{S} = \bar{\sigma}^2\sum_{j=1}^{m}\frac{S_j}{\sigma_j^2} \tag{4.8}$$

with σ_j the approximate uncertainty in S_j as an estimate of I

$$\sigma_j = \frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)^2}{\rho(\mathbf{x}_i)} - S_j^2\right) \tag{4.9}$$

and $\bar{\sigma}$ the uncertainty in $\bar{S}$

$$\frac{1}{\bar{\sigma}^2} = \sum_{j=1}^{m}\frac{1}{\sigma_j^2} \tag{4.10}$$

A χ^2 per degree of freedom is also computed as

$$\chi^2/dof = \frac{1}{m-1}\sum_{j=1}^{m}\frac{(S_j - \bar{S})^2}{\sigma_j^2} \tag{4.11}$$

which converges towards 1 when N increases, if the algorithm works as expected.

In the simplest form of Monte-Carlo integration, the integration points are chosen uniformly-distributed. With the VEGAS algorithm, the density $\rho(\mathbf{x})$ is modified for each of the successive estimate, in order to minimize σ_j^2 (Eq. 4.9) using the information from the previous sampling. Theoretically, σ_j^2 is minimized for

$$\rho((x)) = \frac{|f(\mathbf{x})|}{\int\limits_{\Omega} d^n\mathbf{x} f(\mathbf{x})} \tag{4.12}$$

In particular, it implies that the integration points are concentrated where the integrand is largest in magnitude. To realize this, the VEGAS algorithm starts from a uniform sampling at the first iteration, dividing the space to integrate into hypercubes using a rectangular grid. The integration points are then used to histogram the function and to redefine the integration grid, by concentrating the hypercubes around the region where $|f(\mathbf{x})|$ is the largest. After several iterations, the sampling converges in an adaptive way and the uncertainty $\bar{\sigma}$ decreases as $1/\sqrt{Nm}$.

To implement the numerical integration of the MEM weights using the VEGAS algorithm, it is therefore necessary to define precisely what are the integration variables to be used. Then for each integration point (defined by a given value for each integration variable), the integrand must be computed, which means that its dependence with respect to the integration variables must be explicitly defined. Moreover, the definition of the integration space must also be provided.

4.3 Scattering amplitude

The scattering amplitude for a process Ω can be generically written as

$$|\mathcal{M}_\Omega(\mathbf{x})|^2 = \sum_{flavours} f_1(x_1,Q) f_2(x_2,Q) \bar{\sum_{spins}} |\mathcal{M}(p_1 p_2 \to X)|^2 \tag{4.13}$$

with p_i the incoming QCD partons. An averaged sum is done over the spins of the different particles. (A potential refinement of the analyses could take into account polarization effect of the τ leptons.) The matrix element squared associated to different parton flavors (u, d, c, s and g) are summed incoherently, weighted by the associated PDFs, as those different processes do not interfere at the quantum level. Summing here the matrix elements associated to different flavors, rather than integrating each of them independently, allows to reduce the number of calls to the VEGAS algorithm and is done for practical purposes.

For the incoming partons, the convention chosen is to define the partons 1 and 2 as those with four-vectors

$$q_1 = \left(x_a \frac{\sqrt{s}}{2}, 0, 0, x_a \frac{\sqrt{s}}{2}\right), q_2 = \left(x_b \frac{\sqrt{s}}{2}, 0, 0, -x_b \frac{\sqrt{s}}{2}\right) \tag{4.14}$$

In the case of matrix elements associated to the VBF production of a Higgs boson decaying into τ leptons or the production of a Z boson in association with two jets, the number of possible flavor combinations for the incoming and outgoing QCD partons is of the order of several hundreds. A sample of those is presented in Fig. 4.1. A brute force implementation requires therefore a very large amount of operations, which would lead to a sizable computation time for the numerical integration. However, from first principles, it is easy to actually figure out that all the matrix elements do not have to be individually computed. Indeed, when considering for instance the VBF process $uc \to ucH$, it appears that

$$\mathcal{M}\left(u(q_1)c(q_2) \to u(q_3)c(q_4)H\right) = \mathcal{M}\left(c(q_1)u(q_2) \to c(q_3)u(q_4)H\right) \tag{4.15}$$

based on the universality of the couplings between the first and the second quark generation. In principle, some processes are still expected to have different matrix elements though, due for instance to the different couplings of the W and Z bosons to the quarks. Still, it is possible to group those into different categories of similar matrix elements for which a single matrix element can be computed with minimal effect on the overall matrix element computation. The distributions of some simplified PDF-weighted squared matrix elements are presented in Fig. 4.2. The so-called "v2" has given very close values of the matrix elements while reducing the computing time by one order of magnitude.

In the case of processes producing top quarks, the number of matrix elements is much smaller because the initial states are in most cases reduced to gluon-gluon or quark-antiquark configurations, as seen in Fig. 4.3. Thus, for the $t\bar{t}H, H \to \tau\tau$ MEM implementation, no simplification of the matrix element computation has been required and an exact sum over all the possible partonic initial states has been used.

4.4 Transfer functions

4.4.1 General considerations

The transfer function $W(\mathbf{y}||\mathbf{x})$ includes information on the resolution of the detector and reconstruction techniques. It relates the parton-level variables $\mathbf{x}$ with the set of observables $\mathbf{y}$. Those observables[1] are typically:

- the energy or the transverse momentum ($\hat{E}_j$, $\hat{p}_{Tj}$) and the direction ($\hat{\vec{e}}_j$) of the jets
- the three-momentum of the charged leptons ($\hat{\vec{l}}$)

[1]Measured quantities are represented with a hat symbol

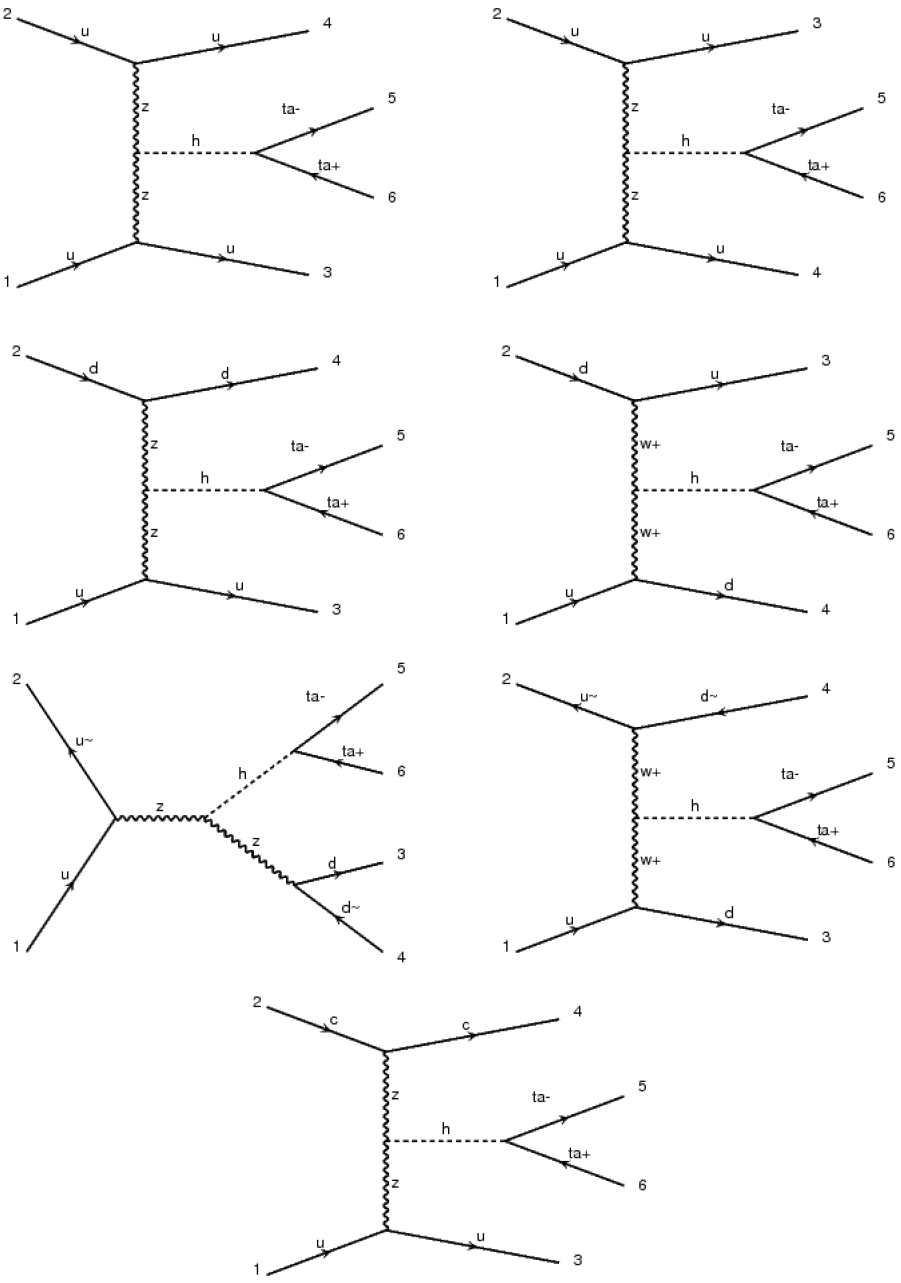

Figure 4.1: Feynman diagrams of some of the processes contributing to the VBF production of a Higgs boson decaying into τ leptons

- the three-momentum of the visible decay products from hadronic τ decays ($\hat{\vec{\pi}}$)

- the missing transverse energy ($\hat{\vec{E}}_T^{miss}$)

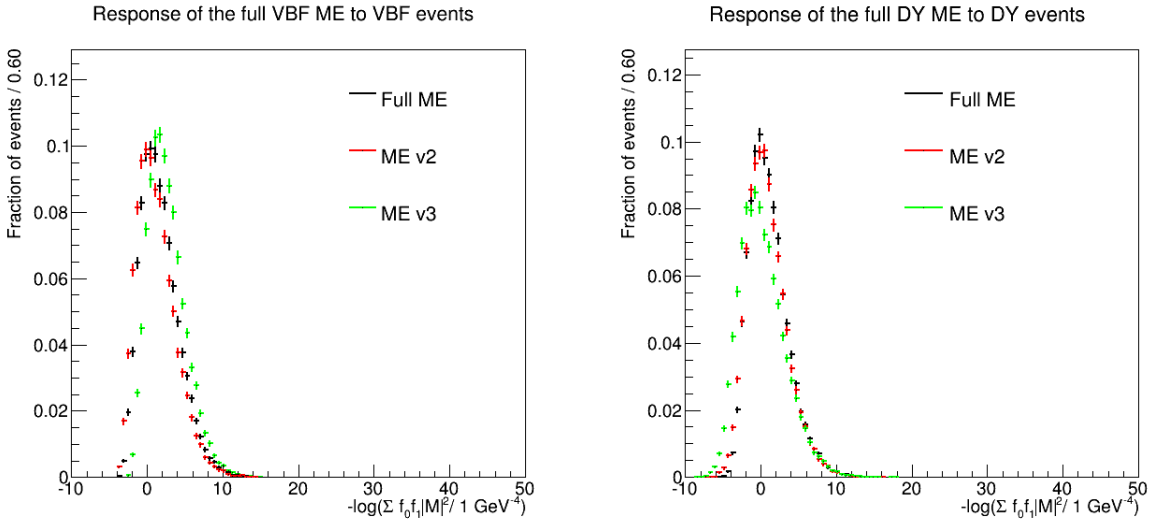

Figure 4.2: Distributions of the PDF-weighted squared matrix element for the VBF process evaluated on VBF Monte-Carlo events (left) and for the Drell-Yan (DY) production of a Z boson in association with two jets in DY Monte-Carlo events (right))

Given the performance of the CMS reconstruction, based on the Particle-Flow algorithm [127], the following assumptions can be made:

- The direction of the quarks is assumed to be perfectly measured with the direction of the jets, the angular resolution for jets with $\hat{p}_{Tj} > 20$ GeV being smaller than 0.03 rad [127].
- The momentum of the charged leptons is perfectly measured, both in direction and in magnitude.
- The momentum of the visible decay products from hadronic τ decays is perfectly measured, both in direction and in magnitude.

Based on those assumptions, the transfer function $W(\mathbf{y}||\mathbf{x})$ can be decomposed into

$$\begin{aligned} W(\mathbf{y}||\mathbf{x}) \quad = \quad & \prod_{q\in\mathcal{A}} \delta(\hat{\vec{e}}_j - \vec{e_q}) T_j(\hat{p}_{Tj}|p_{Tq},\eta_q) \prod_{q\notin\mathcal{A}} A_q(p_{Tq},\eta_q) \\ & \prod_{\pi} \delta(\hat{\vec{\pi}} - \vec{\pi}) T_h(\hat{\pi}|\tau_h) \prod_{l} \delta(\hat{\vec{l}} - \vec{l}) T_\ell(\hat{\ell}|\tau_\ell) T_{E_T}(\hat{\vec{\rho}}_T|\vec{P}_T) \end{aligned} \tag{4.16}$$

The previous assumptions have been translated into δ-functions. $T_\ell(\ell|\tau)$ relates the momentum of the τ decaying into a charged lepton and neutrinos to the momentum of its visible decay product ℓ. $T_h(\tau_h|\tau)$ relates the momentum of the τ decaying into hadrons and neutrinos to the momentum of its visible decay products τ_h. For some processes for which the reconstructed hadronic τ decay corresponds to a misreconstructed lepton, another

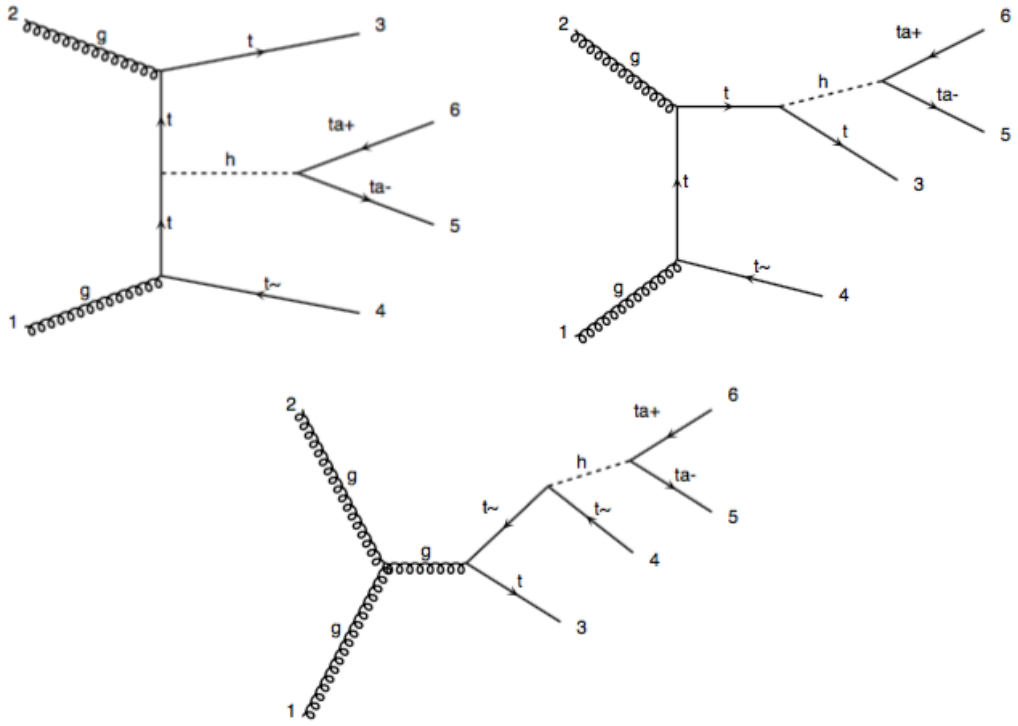

Figure 4.3: Feynman diagrams of some of the processes contributing to the ttH production of a Higgs boson decaying into τ leptons

transfer function $T_{\ell\to\tau}(\tau_h|\ell)$ is used to relate the p_T of the reconstructed τ_h to the one of the original lepton. In the case of processes with non-prompt leptons from semi-leptonic B hadron decays, a transfer function $T_{non-prompt}(\ell|b)$ is used to relate the p_T of the lepton to the one of the original b quark. The notation $q \in \mathcal{A}$ ($q \notin \mathcal{A}$) indicates those quarks which are (not) associated to at least one of the jets. For $q \in \mathcal{A}$, $T_j(p_{Tj}|p_{Tq}, \eta_q)$ relates the p_T of reconstructed jets to the p_T of the associated quark, while for $q \notin \mathcal{A}$, $A_q(p_{Tq}, \eta_q)$ represents the probability that no jet is reconstructed given the kinematics of the quark. Finally T_{E_T} relates the opposite of the measured recoil $\vec{\rho}_T = \sum_{\ell,\tau_h,j} \vec{p}_T + \vec{E}_T^{miss}$ to the recoil $\vec{P}_T$ associated with the point in phase space considered for the integration. The different parts will be described in the next sections. Note that, since the transfer functions are interpreted as conditional probabilities of measuring $\mathbf{y}$ given $\mathbf{x}$, the convention adopted for their normalization is such that

$$\int d\mathbf{y} W(\mathbf{y}||\mathbf{x}) = 1, \forall \mathbf{x} \tag{4.17}$$

4.4.2 Jets

The transfer function for the jets corresponds to the conditional probability to measure a jet with a transverse momentum p_{Tj}, given a quark with a transverse momentum p_{Tq} and pseudo-rapidity η_q. The η dependence is introduced to better model the different energy resolutions of the jets in the barrel and the endcaps. This energy transfer function can be very well modelled using a sum of two Gaussian distributions

$$T_j(p_{Tj}|p_{Tq},\eta_q) = f\mathcal{G}\left(p_{Tj}|\mu_1(p_{Tq},\eta_q),\sigma_1(p_{Tq},\eta_q)\right) + (1-f)\mathcal{G}\left(p_{Tj}|\mu_2(p_{Tq},\eta_q),\sigma_2(p_{Tq},\eta_q)\right) \tag{4.18}$$

$$\mathcal{G}(p_{Tj}|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{p_{Tj}-\mu}{\sigma}\right)^2\right) \tag{4.19}$$

with $\mu_i(p_{Tq},\eta_q)$ and $\sigma_i(p_{Tq},\eta_q)$ the mean and the width of the Gaussian distributions. The distribution of the jet transverse momentum matched to quarks in different p_T bins is presented in Fig. 4.4, both for light quarks and for b quarks. The p_T distribution for jets from b quarks shows a larger tail towards low p_T values due to the production of neutrinos in semi-leptonic decays of B hadrons. The parametrization of the jet transfer functions has been derived from a Monte-Carlo $t\bar{t}$ sample in two η bins ($|\eta| < 1.5$, $1.5 < |\eta| < 2.4$). The values of μ_i and σ_i are extracted from a fit of the p_T distribution of the jets in bins of p_{Tq} (requiring a geometrical matching of $\Delta R < 0.4$) with a double Gaussian distribution.

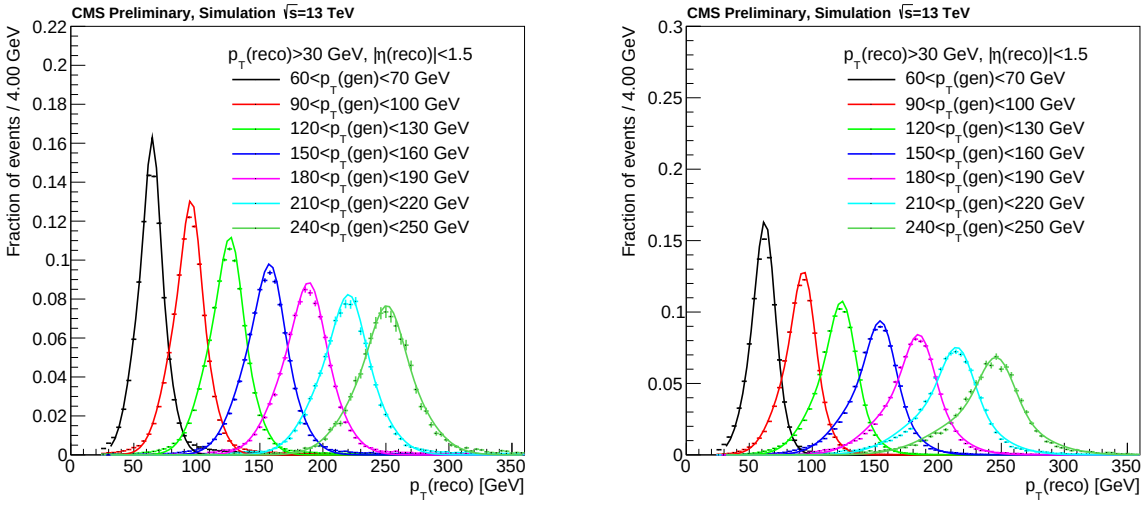

Figure 4.4: p_T distribution for jets matched to light quarks (left) and to b quarks (right) in different bins of quark p_T for jets with $|\eta| < 1.5$ obtained in a $t\bar{t}$ Monte-Carlo sample. The solid line represents the result of a fit with a double Gaussian distribution.

The p_{Tq} dependence of μ_i is then obtained fitting a linear function through the values obtained previously, while for σ it is obtained from the result of a fit with the sum in

Flavor	$\|\eta\|$	m_1/m_2	n_1/n_2	a_1/a_2	b_1/b_2	c_1/c_2	f
light quarks	[0.0,1.5]	0.99/0.94	-2.33/9.39	0.0/0.0	1.53/0.78	0.13/0.05	0.81
	[1.5,2.4]	0.94/0.82	10.79/51.65	4.26/2.56	0.58/1.99	0.0/0.0	0.80
b quarks	[0.0,1.5]	0.99/0.97	-6.24/-8.12	0.0/0.0	1.07/0.0	0.05/0.23	0.66
	[1.5,2.4]	0.98/0.94	-6.24/-10.31	0.0/3.86	1.14/0.47	0.0/0.21	0.52

Table 4.1: Numerical values of the coefficients used to parametrize the jet transfer functions

quadrature of a noise, stochastic and constant term. The relative weights f is taken constant as a function of p_{Tq}. A different parametrization is used for jets matched to light quarks and to b quarks.

$$\mu_i(p_{Tq}, \eta_q) = m_i(\eta_q) p_{Tq} + n_i(\eta_q) \tag{4.20}$$

$$\sigma_i(p_{Tq}, \eta_q) = a_i(\eta_q) \oplus b_i(\eta_q)\sqrt{p_{Tq}} \oplus c_i(\eta_q) p_{Tq} \tag{4.21}$$

The numerical values of the different parameters is presented in Table 4.1.

4.4.3 Quark acceptance

The transfer function for a quark q not reconstructed as a jet ($q \notin \mathcal{A}$) is given by

$$A_q(p_T, \eta) = \begin{cases} 1 \text{ if } |\eta| > \eta^c \\ 0 \text{ else if } \Delta R_{q\ell} < R_\ell^c \text{ and } \frac{E}{E_\ell} > f^c \\ 1 \text{ else if } \min_{j'} \Delta R_{jj'} < R_j^c \\ \int_0^{p_T^c} dp'_T T_j(p'_T | p_T, \eta) \text{ else} \end{cases} \tag{4.22}$$

The numerical values of the parameters in Eq. (4.22) are reported in Table 4.2. A uniform prior ($A_q = 1$) is assigned to a quark getting out of the detector acceptance ($|\eta| > \eta^c$), η^c corresponding to the η acceptance of the jets in the analysis. On the opposite, if the quark is close to one of the reconstructed charged lepton by less than some cut value R_ℓ^c and its energy is in excess of $f^c E_\ell$, where R_ℓ^c corresponds to the size of the cone veto used to remove jets overlapping with isolated leptons and f^c is the relative isolation cut applied on the reconstructed lepton, such an occurrence would have been vetoed by the selection cuts and is therefore not considered in the integration ($A_q = 0$).

If instead, the quark direction is closer to one of the jets j' by less than the jet algorithm radius R_j^c, that quark is assumed to have been merged with another quark q' into the same jet. In this case, the acceptance function for q is set to 1, while the transfer function $T_j(p_{Tj'}|p_{Tq'}, \eta_{q'})$ is replaced by $T_j(p_{Tj'}|p_{Tqq'}, \eta_{q'})$, using the p_T of the qq' pair instead of the p_T of the single quark q'. Otherwise, the acceptance function is taken as the cumulative distribution of the jet transfer function, evaluated at the transverse momentum threshold p_T^c.

η^c	R_j^c	R_ℓ^c	f^c	p_T^c
2.4	0.4	0.4	0.4	25 GeV

Table 4.2: Numerical values of the coefficients used to parametrize the quark acceptance

4.4.4 Hadronic τ transfer function

As hadronic τ decays include one neutrino in the final state, the four-vector of the τ lepton before decay cannot be directly deduced from its visible decay products. To take into account the contribution from neutrinos, integrations over some variables related to the di-τ system, which will be described in section 4.5.1, are performed. To account for the non-flat distributions of the energy of the visible decay products a transfer function, based on the τ differential decay width, is used.

1-prong decay mode

In the case of the 1-prong decay $\tau \to \nu\pi$, the kinematics is the one of a two-body decay. In that case, the energy transfer function is given by

$$T_{h1}(E_\pi|E_\tau) \propto \frac{d\Gamma}{dE_\pi} \propto \frac{E_\pi}{|\vec{\pi}|} \int \frac{1}{E_\tau} \frac{|\vec{\pi}|^2 d\Omega_\pi}{(2\pi)^3 2E_\pi} \frac{d^3\vec{\nu}}{(2\pi)^3 2E_\nu} \delta^4(\tau - \pi - \nu) |\mathcal{M}_\Gamma(\tau \to \nu\pi)|^2 \quad (4.23)$$

with $|\mathcal{M}_\Gamma(\tau \to \nu\pi)|^2$ the matrix-element corresponding to 1-prong decay. The spin-averaged matrix element is given by

$$\frac{1}{2}\bar{\Sigma}|\mathcal{M}(\tau \to \pi\nu)|^2 = G_F^2 |V_{ud}|^2 m_\tau^4 f_\pi^2 \left(1 - \frac{m_\pi^2}{m_\tau^2}\right) \quad (4.24)$$

The different factors do not depend on the four-vectors of the particle and are therefore dropped in the computation, as mentioned in section 4.1. From the energy-momentum conservation $\delta^4(\tau - \pi - \nu)$, we then deduce that

$$\cos\theta_{\tau\pi} = \frac{2E_\tau E_\pi - (m_\tau^2 + m_\pi^2)}{2|\vec{\tau}||\vec{\pi}|} \quad (4.25)$$

After integration and requiring the normalization condition

$$\int T_{h1}(E_\pi|E_\tau) dE_\pi = 1 \quad (4.26)$$

this leads to

$$T_{h1}(E_\pi|E_\tau) = \frac{1}{E_\tau \left(1 - \frac{m_\pi^2}{m_\tau^2}\right)} \quad (4.27)$$

This transfer function corresponds to a flat distribution of the variable $z = E_\pi/E_\tau$. Now, assuming that the momentum of the pion is perfectly measured, both in magnitude and in direction, the overall transfer function for the hadronic τ is given by

$$T_{h1}(E_\pi|E_\tau)\delta(\hat{\vec{\pi}} - \vec{\pi})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \delta(\hat{\vec{\pi}} - \vec{\pi})\frac{|\vec{\tau}|^2}{E_\tau^2\left(1 - \frac{m_\pi^2}{m_\tau^2}\right)}d|\vec{\tau}|d\cos\theta_{\tau\pi}d\phi_{\tau\pi}$$

We have to take into account the constraint on $\cos\theta_{\tau\pi}$ given by Eq. (4.25). The measurement of the momentum of the pion enables thus to remove the integration over $\cos\theta_{\tau\pi}$, which finally leads to

$$T_{h1}(\pi|\tau)\delta(\hat{\vec{\pi}} - \vec{\pi})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \frac{|\vec{\tau}|^2}{E_\tau^2\left(1 - \frac{m_\pi^2}{m_\tau^2}\right)}d|\vec{\tau}|d\phi_{\tau\pi} \tag{4.28}$$

Moreover, for a physical solution, we must have $-1 \leq \cos\theta_{\tau\pi} \leq 1$. This constraint leads to

$$|s^-| \leq |\vec{\tau}| \leq s^+ \tag{4.29}$$

with

$$s^\pm = \frac{(m_\tau^2 + m_\pi^2)|\vec{\pi}| \pm E_\pi(m_\tau^2 - m_\pi^2)}{2m_\pi^2} \tag{4.30}$$

In the ultra-relativistic regime Eq. (4.29) is simply equivalent to the constraint

$$E_\pi \leq E_\tau \leq \frac{m_\tau^2}{m_\pi^2}E_\pi \tag{4.31}$$

Other hadronic decay modes

The other hadronic τ decay modes are in principle not as straightforward. Indeed, since they involve hadronic resonances, a simple analytic computation is not possible as for the decay into 1 prong. The constraints given by Eq. (4.25) and (4.29) still remain valid. However, unlike for the 1-prong case, the invariant mass of the visible decay products cannot be assumed to be fixed but has to be defined as $m_\pi^2 = E_\pi^2 - |\vec{\pi}|^2$. The effect on the distribution of the invariant mass of the hadronic system and on the variable $z = E_\pi/E_\tau$ is visible in Fig. 4.5.

Even if the distribution of the invariant mass of the hadronic decay products cannot be directly computed, it is reasonable to assume that for a given value of m_π, the matrix element associated to the τ decay keeps the structure of a 2-body decay ($\tau \to \pi\nu$, with a subsequent decay of π) and is therefore independent on other kinematics quantities. If this is indeed the case, as for the 1-prong case, a flat distribution of the variable $z = E_\pi/E_\tau$ between m_π^2/m_τ^2 and 1 is expected for a fixed value of m_π. This is indeed what is observed in Fig. 4.6. Convoluting these distributions with the observed mass spectrum of the visible decay products enables then to recover the overall z spectrum for the different decay modes, as seen in Fig. 4.7. The modeling of the hadronic tau transfer function giving an excellent agreement with the observed z distribution, the same formula as for the 1-prong case is therefore used for the other decay modes.

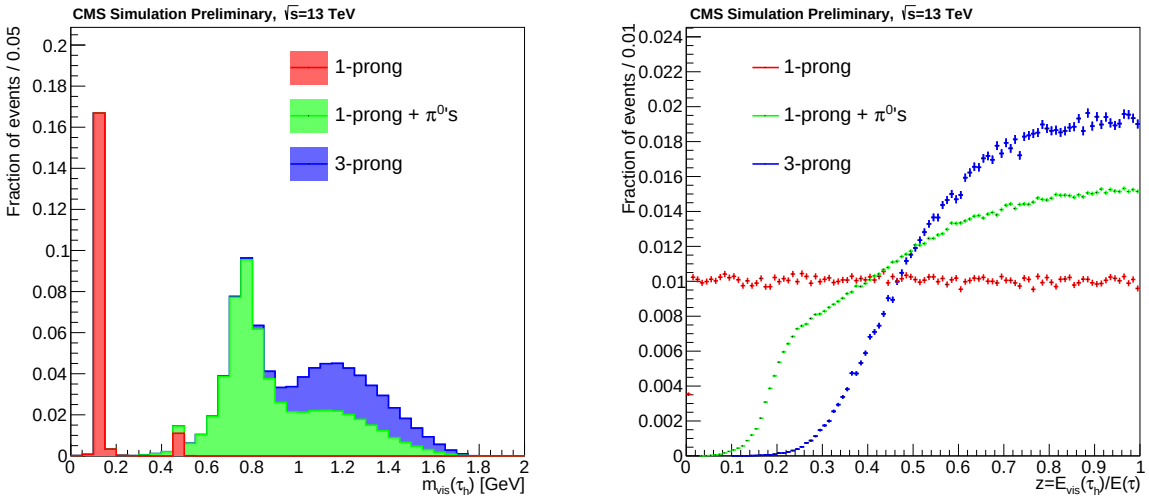

Figure 4.5: Distribution of the invariant mass of the hadronic τ visible decay products (left) and of the variable $z = E_{vis}/E_\tau$ (right) for different τ_h decay modes at generator level in $H \to \tau\tau$ simulated events.

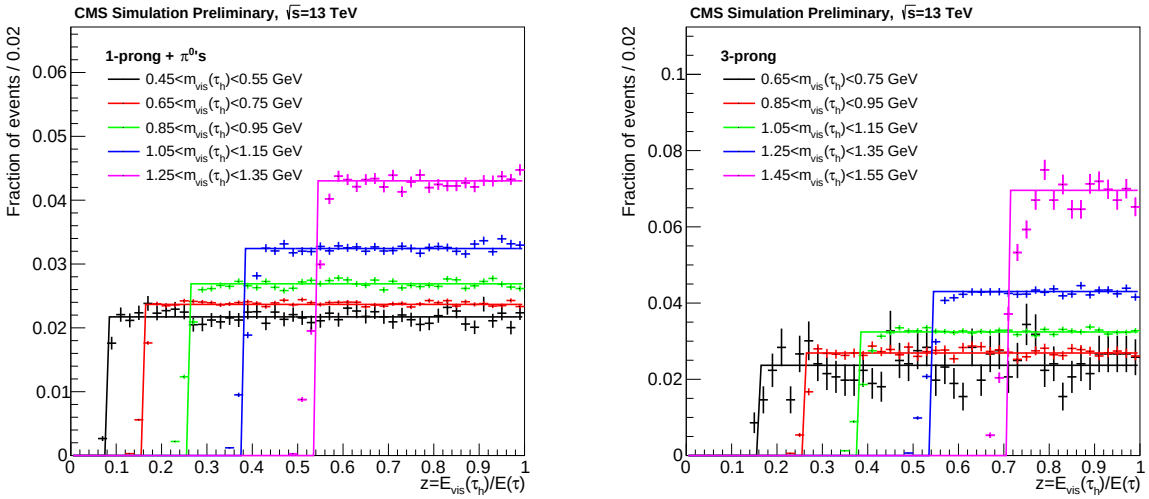

Figure 4.6: Distribution of the variable $z = E_{vis}/E_\tau$ for different bins of m_{vis} for 1-prong+π^0's (left) and 3-prong (right) τ_h in $H \to \tau\tau$ simulated events. The lines correspond to the analytic expression of the transfer function used.

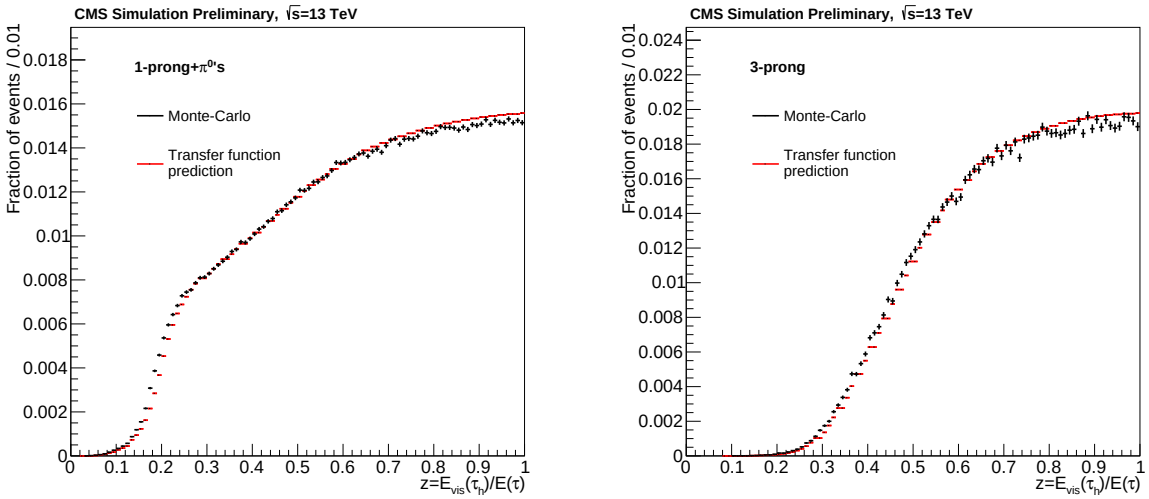

Figure 4.7: Distribution of the variable $z = E_{vis}/E_\tau$ for 1-prong+π^0's (left) and 3-prong (right) τ_h in $H \to \tau\tau$ simulated events. The black distribution correspond to the one directly observed in Monte-Carlo while the red distribution is the one obtained with the convolution of the analytic transfer function with the m_{vis} distribution.

4.4.5 Leptonic τ transfer function

In the case of a leptonic τ decay, the transfer function is given by

$$T_l(\ell|\tau) \propto \frac{d^3\Gamma}{d^3\vec{\ell}} \propto \frac{1}{E_\tau}\int \frac{1}{(2\pi)^3 2E_\ell}\frac{d^3\vec{\nu}}{(2\pi)^3 2E_\nu}\frac{d^3\vec{\bar{\nu}}}{(2\pi)^3 2E_{\bar{\nu}}}\delta^4(\tau-\ell-\nu-\bar{\nu})|\mathcal{M}_\Gamma(\tau\to\ell\nu\bar{\nu})|^2 \quad (4.32)$$

with $|\mathcal{M}_\Gamma(\tau \to \ell\nu\bar{\nu})|^2$ the matrix-element corresponding to the leptonic decay. The spin-averaged matrix element is given by

$$\frac{1}{2}\bar{\Sigma}|\mathcal{M}(\tau\to\ell\nu\bar{\nu})|^2 = 64G_F^2(\nu.\ell)(\bar{\nu}.\tau) \quad (4.33)$$

One gets then

$$T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell}\int \frac{d^3\vec{\nu}}{2E_\nu}\frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}}(\nu.\ell)(\bar{\nu}.\tau)\delta^4(\tau-\ell-\nu-\bar{\nu}) \quad (4.34)$$

One can show that[2]

$$T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell}[(\ell.\tau)(m_\tau^2+m_\ell^2-2l.\tau)+2(l.\tau-m_\ell^2)(m_\tau^2-\ell.\tau)] \quad (4.35)$$

[2]The detailed computation is available in Appendix A.

Moreover, the requirement that E_ν and $E_{\bar{\nu}}$ are positive leads to

$$\cos\theta_{\tau\ell} \geq \frac{2E_\tau E_\ell - m_\tau^2 - m_\ell^2}{2|\vec{\tau}||\vec{\ell}|} \equiv \alpha \tag{4.36}$$

This constraint has to be taken into account in the integration. Indeed, since we must have $\cos\theta_{\tau\ell} \leq 1$, this puts a constraint over the integration range for $|\vec{\tau}|$. The same kind of computation as for the hadronic case leads to the constraint

$$t^- \leq |\vec{\tau}| \leq t^+ \tag{4.37}$$

with

$$t^\pm = \frac{(m_\tau^2 + m_\ell^2)|\vec{\ell}| \pm E_\ell(m_\tau^2 - m_\ell^2)}{2m_\ell^2} \tag{4.38}$$

Contrary to the hadronic case (Eq. (4.29), there is no absolute value for the lower bound t^- Again, in the ultra-relativistic regime Eq. (4.37) is equivalent to the constraint

$$E_\ell \leq E_\tau \leq \frac{m_\tau^2}{m_\ell^2} E_\ell \tag{4.39}$$

For what regards the direction of the τ lepton with respect to the charged lepton ℓ, it is in principle only constrained by Eq. (4.36) but is not fixed beyond that. However, as the τ leptons from Higgs or Z decay are relatively boosted, the directions of the τ is very close to the one of the charged lepton ℓ, as seen in Fig. 4.8. Based on that observation, it is assumed that the direction of the τ before decay is the same as the one of the charged lepton ℓ. This enables to remove two integration variables associated to the direction of the τ in the Matrix Element Method integration. The difference in the direction being relatively small, this is expected to lead to a negligible difference in the MEM final discriminant.

Within the collinear approximation and assuming $m_\ell = 0$, the only remaining part in the leptonic τ transfer function is therefore the energy transfer function given by

$$T_{lE}(\ell|\tau) \propto \frac{d\Gamma}{dE_\ell} = \int d\Omega_\ell E_\ell^2 \frac{d^3\Gamma}{d^3\vec{\ell}} \tag{4.40}$$

It can be shown that this leads finally to[3]

$$T_{lE}(\ell|\tau) = \frac{1}{3E_\tau}(1-z)\left(5 + 5z - 4z^2\right) \tag{4.41}$$

with $z = E_\ell/E_\tau$. Together with the integration related parts, this gives therefore

$$T_{lE}(\ell|\tau)\delta(\vec{e}_\tau - \hat{\vec{e}}_\ell)\delta(\hat{\vec{\ell}} - \vec{\ell})\frac{d^3\vec{\tau}}{(2\pi)^3 2E_\tau} \propto \frac{1}{3}(1-z)\left(5 + 5z - 4z^2\right)\frac{|\vec{\tau}|^2}{E_\tau^2} d|\vec{\tau}| \tag{4.42}$$

This result has been confirmed in simulation, as presented in Fig. 4.9.

[3]The demonstration is given in Appendix A.

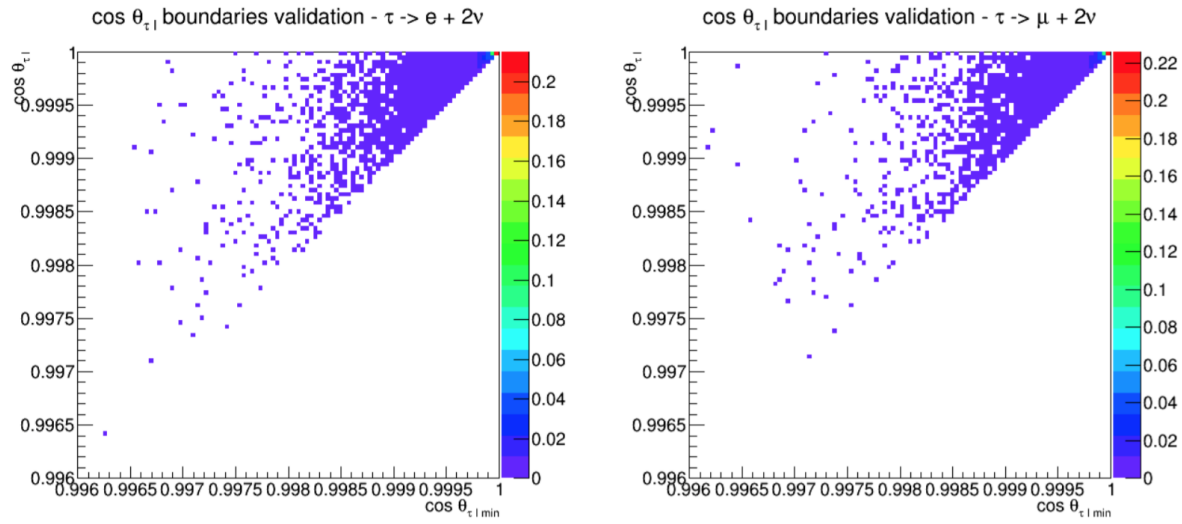

Figure 4.8: Angular distributions for the leptonic τ decay in the electron (left) and muon (right) decay channels. The distributions are obtained from generator level in a VBF sample. $\cos\theta_{\tau\ell min}$ is defined by Eq. (4.36)

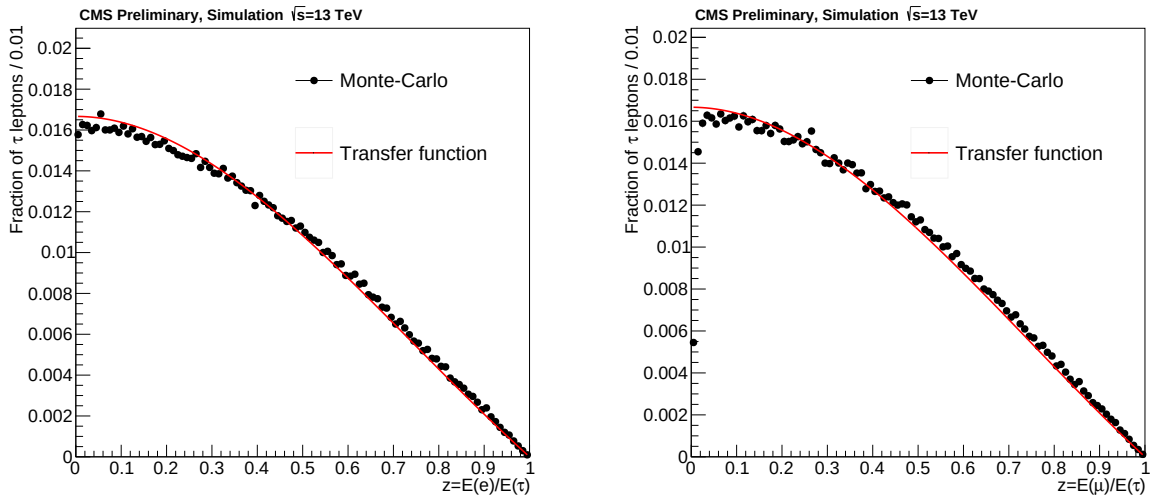

Figure 4.9: Distribution of the variable $z = E_\ell/E_\tau$ for τ decaying into neutrinos and an electron (left) or a muon (right).

4.4.6 Fake τ_h from leptons

As mentioned in section 2.3.6, some of the reconstructed τ_h can actually originate from electrons (or less often from muons), even after application of anti-lepton discriminants. Moreover, depending on the analysis, it is not always beneficial in terms of sensitivity to

use those anti-lepton discriminants if the signal has also some fraction of events with such leptons faking τ_h. As the reconstruction algorithm for τ_h is not optimized for those objects, the p_T of the reconstructed τ_h does not exactly match the one of the misreconstructed lepton, as shown in Fig. 4.10. To take that effect into account, dedicated transfer functions for fake τ_h from leptons have been developed. In the case of electrons, that transfer function can be described using a Breit-Wigner distribution, while in the case of muons a Crystal-Ball distributions has been found to be more adapted.

$$T_{e\to\tau}(\tau_h|e) \propto \frac{1}{1+\left(\frac{p_{T\tau}-\mu}{\sigma}\right)^2} \tag{4.43}$$

$$T_{\mu\to\tau}(\tau_h|\mu) \propto \begin{cases} 0 \text{ if } \left|\frac{p_{T\tau}-\mu}{\sigma}\right| > 5 \\ \exp\left(-0.5\left(\frac{p_{T\tau}-\mu}{\sigma}\right)^2\right) \text{ if } \frac{p_{T\tau}-\mu}{\sigma} < \alpha \\ \left(\frac{n}{\alpha}\right)^n \exp(-0.5\alpha^2)\left(\frac{n}{\alpha}-\alpha+\frac{p_{T\tau}-\mu}{\sigma}\right)^{-n} \text{ else} \end{cases} \tag{4.44}$$

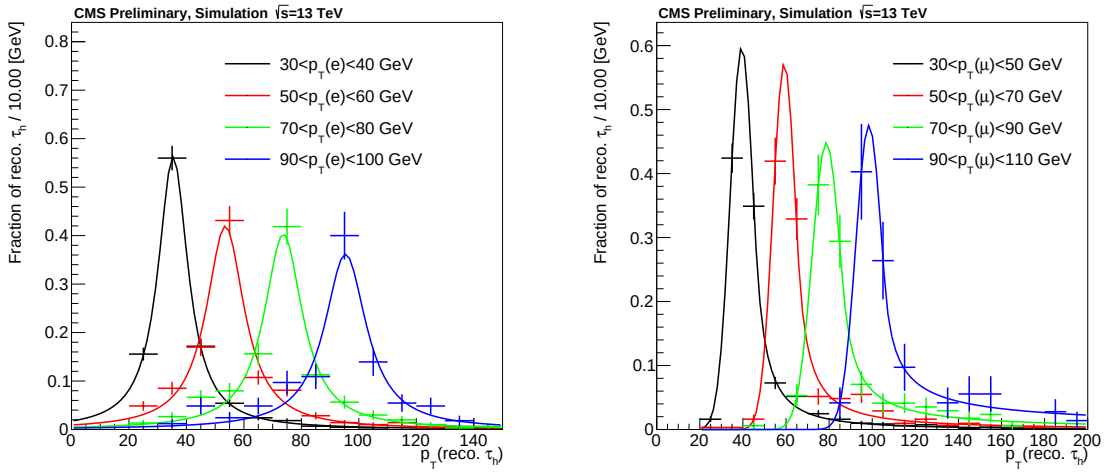

Figure 4.10: p_T distribution for reconstructed τ_h matched to generator level electrons (left) and muons (right) in different bins of lepton p_T, obtained in a $t\bar{t}Z$ Monte-Carlo sample. The solid line represents the result of a fit with a Breit-Wigner distribution for electrons and Crystal-Ball distribution for muons.

The parameters μ, σ, n and α depend on the p_T of the lepton. This parametrization is extracted from a fit of those parameters, with the following p_T dependence

$$\mu(p_T) = ap_T + b \tag{4.45}$$

$$\sigma(p_T) = cp_T + d \tag{4.46}$$

$$n(p_T) = ep_T + f \tag{4.47}$$

The numerical values of the different parameters are presented in Table 4.3.

Lepton flavor	a	b	c	d	e	f	α
Electrons	0.96	0.84	0.04	4.69	-	-	-
Muons	0.97	1.17	0.02	4.46	-0.015	2.29	1.2

Table 4.3: Numerical values of the coefficients used to parametrize the lepton to fake τ_h transfer functions

4.4.7 Non-prompt leptons from b quarks

As mentioned in section 1.3.3, non-prompt leptons can be produced inside b-jets from semi-leptonic decays of B hadrons. If such a lepton passes all the selection and isolation criteria, jets overlapping with such objects are usually not considered for the analysis. In order to evaluate the original b quark momentum, the reconstructed lepton can still be used as a proxy. A dedicated transfer function can therefore be used to relate the momenta of the b quark and the non-prompt lepton. For this purpose, the direction of the lepton is assumed to be the one of the b quark while a dedicated transfer function relating the p_T of the fake lepton to the one of the b quark has been used. As shown in Fig. 4.11, a Gaussian distribution provides a good description of this transfer function. The following parametrization is used for the Gaussian parameters.

$$\mu(p_{Tb}) = mp_{Tb} + n \tag{4.48}$$

$$\sigma(p_{Tb}) = a \oplus b\sqrt{p_{Tb}} \oplus cp_{Tb} \tag{4.49}$$

The numerical values of the different parameters is presented in Table 4.4.

m	n	a	b	c
0.58	-10.74	6.26	0.0	0.14

Table 4.4: Numerical values of the coefficients used to parametrize the fake lepton transfer functions

4.4.8 Recoil transfer function

As mentioned in section 4.4.1, the recoil transfer function relates the total measured transverse momentum $\hat{\vec{\rho}}_T = \hat{\vec{E}}_T^{miss} + \sum_{\ell,\pi,jets} \hat{\vec{p}}_T$ to its parton level equivalent $P_T = \sum_{\tau,p} \vec{p}_T$. It is used both to account for the presence of potential initial or final state radiation not described by the matrix element at leading order, as well as the detector resolution on the missing transverse energy.

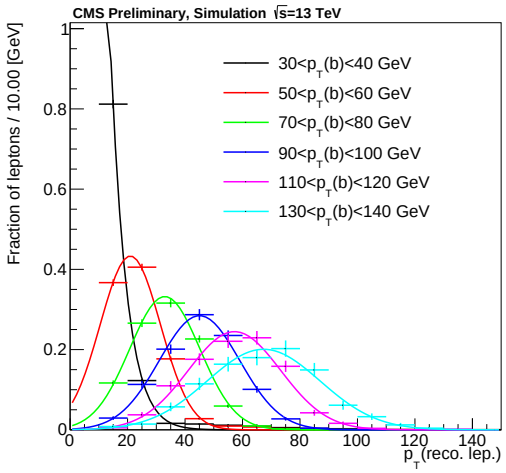

Figure 4.11: p_T distribution for reconstructed leptons matched to b quarks in different bins of b quark p_T, obtained in a $t\bar{t}$ Monte-Carlo sample. The solid line represents the result of a fit with a Gaussian distribution.

This transfer function is obtained from the "Particle-Flow significance" algorithm, documented in [131] and summarized here. The particles reconstructed with the Particle-Flow algorithm are grouped into exclusive collections of objects, depending on whether they are clustered inside jets, are isolated (e.g. leptons) or unclustered. The measurement of the transverse momentum $\vec{E}_T^i$ of a given particle i is then associated to a covariance matrix

$$\mathbf{U}_i = \begin{pmatrix} \sigma^2_{E_T i} & 0 \\ 0 & \sigma^2_{\phi i} \end{pmatrix} \tag{4.50}$$

evaluated in the reference frame with axes parallel and perpendicular to $\vec{E}_T^i$. The U_{00} and U_{11} elements depend in general on the p_T and η of the object considered. The covariance matrices are then summed together after a rotation into the global (x, y, z) CMS reference frame to build the transverse energy covariance matrix

$$\mathbf{V} = \sum_i R(\phi_i)^{-1} \mathbf{U}_i R(\phi_i) \tag{4.51}$$

This covariance matrix, evaluated on an event-by-event basis, is then used to compute the recoil transfer function, defined as

$$T_{E_T}(\hat{\vec{\rho}}_T | \vec{P}_T) = \frac{1}{\sqrt{2\pi|\mathbf{V}|}} \exp\left(-\frac{1}{2} \left(\hat{\vec{\rho}}_T - \vec{P}_T \right)^T \mathbf{V}^{-1} \left(\hat{\vec{\rho}}_T - \vec{P}_T \right) \right) \tag{4.52}$$

4.5 Dimensionality reduction and kinematic reconstruction

The computation of the multi-dimensional integral defined in Eq. (4.2) requires in principle 3 dimensions of integration per particle in the final state. In the case of a $t\bar{t}H$, $t \to b\ell\nu$, $\bar{t} \to \bar{b}q\bar{q}$, $H \to \tau\bar{\tau} \to \ell\tau_h + \nu's$ weight, this represents for instance 33 dimensions. Such a high number of dimensions would require a very large number of integration points to make the numerical integration convergent. Fortunately, some of the transfer functions enable to reduce the dimensionality: this is for instance the case for leptons where the detector resolution makes reasonable the assumption that the measured momentum corresponds to the genuine momentum of the lepton. Due to the resonant nature of the processes taken into account, the dimensionality can be further reduce by assuming that some of the intermediate particles involved are on their mass shell. This is the case for Higgs, Z and W bosons, for top quarks and for τ leptons. Off-shell contributions would anyway be strongly suppressed by a small value of the matrix element and would therefore not have a significant contribution to the integral.

4.5.1 Higgs and Z boson decaying into τ leptons

For a Higgs (or a Z) boson decaying into τ leptons, the phase space associated to that decay is

$$d\Phi_H \propto \frac{d^3\vec{\tau}}{2E_\tau}\frac{d^3\vec{\bar{\tau}}}{2E_{\bar{\tau}}} \tag{4.53}$$

Without loss of generality, it is assumed here that the τ lepton decays into $\tau \to \ell + \nu_\tau + \bar{\nu}_\ell$ while the $\bar{\tau}$ lepton decays into $\bar{\tau} \to \tau_h + \bar{\nu}_\tau$. With the approach of effective τ transfer functions presented in sections 4.4.4 and 4.4.5, the direction of the τ lepton is assumed to be the one of the charged lepton ℓ, which leaves the following integration variables

$$d\Phi_H \propto \frac{|\vec{\tau}|^2}{E_\tau}d|\vec{\tau}|\frac{|\vec{\bar{\tau}}|^2}{E_{\bar{\tau}}}d\vec{\bar{\tau}}d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \tag{4.54}$$

with the angles $\theta_{\bar{\tau}\pi}$ and $\phi_{\bar{\tau}\pi}$ defined with respect to the visible decay products of the hadronically-decaying τ.

To take into account the resonant mass spectrum, the following change of variables is introduced.

$$d|\vec{\tau}|d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \to d|\vec{\tau}|d\cos\theta_{\bar{\tau}\pi}d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}} \tag{4.55}$$

This requires to be able to determine $\vec{\bar{\tau}}$ as a function of the variables used for the final integration. For that purpose, we can introduce the following Cartesian system of coordinates $(\vec{e}_\pi, \vec{e}_x, \vec{e}_y)$ with $\vec{e}_\pi$ the direction of the visible decay products of the $\bar{\tau}$, $\vec{e}_x = \frac{1}{\sin\theta_{\tau\pi}}(\vec{e}_\tau - \cos\theta_{\tau\pi}\vec{e}_\pi)$ and $\vec{e}_y = \vec{e}_\pi \wedge \vec{e}_x$ (see Fig. 4.12). In that frame, $\vec{\bar{\tau}}$ is given by

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\cos\theta_{\bar{\tau}\pi}\vec{e}_\pi + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\vec{e}_x + \sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\vec{e}_y) \tag{4.56}$$

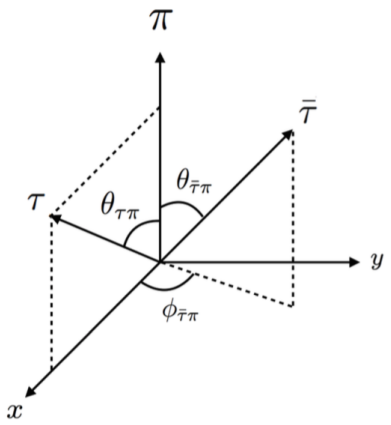

Figure 4.12: Cartesian system used to describe a $\tau_l \bar{\tau}_h$ decay

Introducing now the system of non-Cartesian coordinates $(\vec{e}_\pi, \vec{e}_\tau, \vec{e}_y)$, $\vec{\bar{\tau}}$ can be written as

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\alpha\vec{e}_\pi + \beta\vec{e}_\tau + \gamma\vec{e}_y) \tag{4.57}$$

With that parametrization,

$$\alpha = \frac{\cos\theta_{\bar{\tau}\pi} - \cos\theta_{\tau\pi}\cos\theta_{\tau\bar{\tau}}}{\sin^2\theta_{\tau\pi}} \ , \ \beta = \frac{\cos\theta_{\tau\bar{\tau}} - \cos\theta_{\tau\pi}\cos\theta_{\bar{\tau}\pi}}{\sin^2\theta_{\tau\pi}} \tag{4.58}$$

$$\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi} \tag{4.59}$$

This system has a solution if and only if $\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi} \geq 0$, with α and β defined by the previous equations as function of $\cos\theta_{\tau\bar{\tau}}$ and $\cos\theta_{\tau\pi}$. However, this is not always guaranteed. In principle, to take into account all of the kinematically allowed configurations, the integration domain should cover at least the whole region of phase space for which $\gamma^2 \geq 0$. Although not solvable analytically, this constraint could be tested prior to the integration process for different values of the integration variables. We can therefore determine in advance what will be the contributing region in the phase space for the integration and reduce the integration range consequently.

The so-called "boost" associated to the outgoing particles, defined by $\vec{P}_T = \sum \vec{p}_T$, is then computed for the two signs. At leading order, all the outgoing particles are indeed expected to be perfectly balanced in the transverse plane, having thus $P_T = 0$. The sign of γ is consequently chosen as the one leading to the smallest boost in magnitude. This procedure to determine the sign of γ has been proven to be very efficient at generator level (see Fig. 4.13).

The value of $|\vec{\bar{\tau}}|$ can be determined from the following equation

$$m^2_{\tau\bar{\tau}} = (\tau + \bar{\tau})^2 = 2m^2_\tau + 2(E_\tau E_{\bar{\tau}} - |\vec{\tau}||\vec{\bar{\tau}}|\cos\theta_{\tau\bar{\tau}}) \tag{4.60}$$

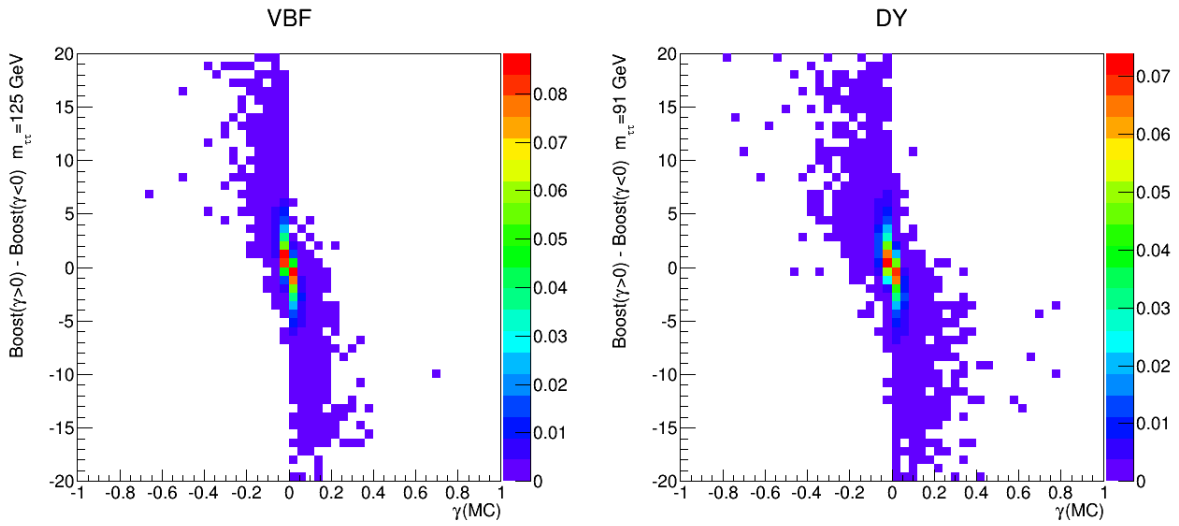

Figure 4.13: Distribution of $\gamma(MC) = \vec{e}_\tau.\vec{e}_y$ and $P_T(\gamma > 0) - P_T(\gamma < 0)$ in leading-order Monte-Carlo samples for VBF (left) and DY+2 jets (right). All the quantities have been evaluated using the kinematic reconstruction detailed previously, fixing the integration variables at their generator-level values. A clear anti-correlation can be observed between the sign of γ and the sign of $P_T(\gamma > 0) - P_T(\gamma < 0)$.

Defining $M^2 = \frac{1}{2}m^2_{\tau\bar{\tau}} - m^2_\tau$, we get

$$|\vec{\bar{\tau}}| = \frac{M^2|\vec{\tau}|\cos\theta_{\tau\bar{\tau}} + E_\tau\sqrt{(M^2)^2 - m^2_\tau\left(m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}\right)}}{m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}} \tag{4.61}$$

with the additional constraint

$$|\vec{\tau}| \leq \frac{m_{\tau\bar{\tau}}}{|\sin\theta_{\tau\bar{\tau}}|}\sqrt{\frac{m^2_{\tau\bar{\tau}}}{4m^2_\tau} - 1} \tag{4.62}$$

With the change of variables introduced in Eq. (4.55) and taking into account the additional constraint from Eq. (4.25) and the on-shell hypothesis used for the di-τ pair, the phase space of the Higgs boson is finally reduced to a 2-dimensional phase space[4]

$$d\Phi_H \propto \frac{|\vec{\tau}|^2}{E_\tau}\frac{|\vec{\bar{\tau}}|^2}{E_{\bar{\tau}}}\frac{1}{\gamma|\sin\theta_{\tau\pi}|\left|E_\tau\frac{|\vec{\bar{\tau}}|}{E_{\bar{\tau}}} - |\vec{\tau}|\cos\theta_{\tau\bar{\tau}}\right|}d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}} \tag{4.63}$$

The integration boundaries are determined taking both into account the constraint from Eq. (4.37) and the constraint $\gamma^2 > 0$. Those constraints are illustrated for a single

[4]The detailed computation is available in Appendix A.

event in Fig. 4.14. If for a given permutation, no $\gamma^2 > 0$ region is found (which can happen for instance if $m(\ell\tau_h) > 125$ GeV in the case of a signal hypothesis with a $H \to \tau\bar{\tau}$ decay), the numerical integration is skipped for that permutation and the corresponding weight is assigned a value of 0, as the kinematical configuration is not compatible with the process hypothesis.

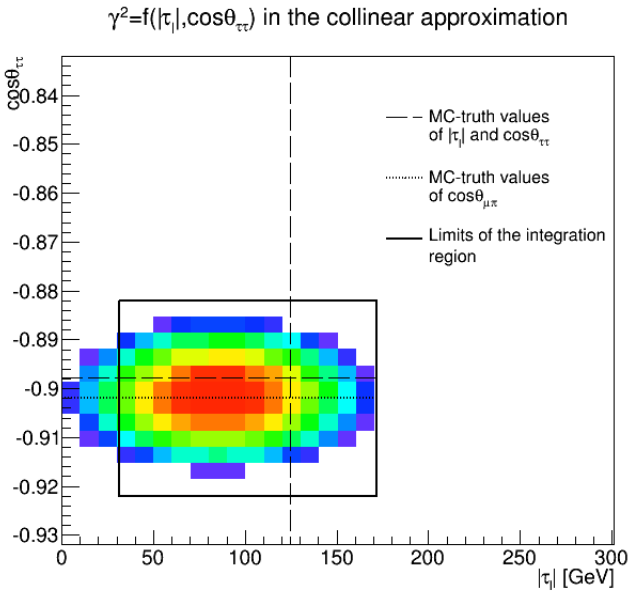

Figure 4.14: Region where $\gamma^2 > 0$ (colored region, the color scale corresponding to the value of γ^2) for a single VBF event in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$ space. The black rectangle corresponds to the limits of the integration region, which is determined prior to the integration. The lower limit for $|\vec{\tau}|$ takes into account t^- from Eq. (4.37), which explains why the rectangle does not cover the whole region $\gamma^2 > 0$.

4.5.2 Hadronic top

The phase space associated to a hadronic top can be written as

$$d\Phi_{t_h} \propto dt^2 dW^2 \delta(t - W - b)\delta(W - q - \bar{q}) \frac{d^3\vec{t}}{2E_t} \frac{d^3\vec{W}}{2E_W} \frac{d^3\vec{b}}{2E_b} \frac{d^3\vec{q}}{2E_q} \frac{d^3\vec{\bar{q}}}{2E_{\bar{q}}} \tag{4.64}$$

Taking into account the mass constraints of the top quark and the W boson, this can be reduced to[5]

$$d\Phi_{t_h} \propto \frac{|\vec{b}|E_{\bar{q}}}{1 - \cos\theta_{q\bar{q}}} \frac{1}{\left|\frac{E_b}{|\vec{b}|}\vec{W}.\vec{e}_b - E_W\right|} dE_q \tag{4.65}$$

[5]The detailed computation is available in Appendix A.

with

$$E_{\bar{q}} = \frac{M_W^2}{2E_q(1-\cos\theta_{q\bar{q}})} \tag{4.66}$$

$$\vec{W} = E_q\vec{e}_q + E_{\bar{q}}\vec{e}_{\bar{q}} \text{ and } E_W = \sqrt{M_W^2 + E_q^2 + E_{\bar{q}}^2 + 2E_qE_{\bar{q}}\cos\hat{\theta}_{q\bar{q}}} \tag{4.67}$$

$$E_b^{\pm} = \frac{E_W\Delta M^2 \pm |\vec{W}.\vec{e}_b|\sqrt{(\Delta M^2)^2 - m_b^2(E_W^2 - (\vec{W}.\vec{e}_b)^2)}}{E_W^2 - (\vec{W}.\vec{e}_b)^2} \tag{4.68}$$

and $\Delta M^2 = (M_t^2 - M_W^2 - m_b^2)/2$. In Eq. (4.68), some unphysical solutions may have been introduced, depending on the sign chosen, which can be removed by requiring the following condition to be fulfilled by one of the solutions $E_b^{\pm}$,

$$\vec{W}.\vec{e}_b > 0 \text{ and } E_b^{\pm} > \frac{\Delta M^2}{E_W} \text{ or } \vec{W}.\vec{e}_b < 0 \text{ and } E_b^{\pm} < \frac{\Delta M^2}{E_W} \tag{4.69}$$

If two admissible (positive) solutions exist, the largest one is chosen.

The MEM also allows in principle to account for jets from one of the light quarks not being reconstructed. In that case, the phase space integration is simply expanded by adding two integration variables corresponding to the direction of the missing jet. Up to that change, all of the previous developments are still valid and the reconstruction of the hadronic top can be performed in the same way.

In all cases, the integration range for E_q is determined by considering the 95% confidence interval of the associated jet transfer function. Moreover, in the case where all jets are reconstructed, a lower limit $M_t^{low}(bjj)$ and an upper limit $M_t^{up}(bjj)$ are determined for the mass of the hadronic top candidate, by considering the invariant mass of the jet triplet with their transverse momentum shifted to the edge of their respective 95% confidence interval. If the following condition is not fulfilled

$$M_t^{low}(bjj) < M_t = 173.21 \text{ GeV} < M_t^{up}(bjj) \tag{4.70}$$

the numerical integration is skipped as the jet triplet is not compatible with a hadronic top decay and a null value is assigned to the integral. In the case of an event with a missing jet, a lower limit $M_t^{low}(bj)$ is defined only from the two considered jets, and the following condition has to be fulfilled to get a non-zero integral

$$M_t^{low}(bj) < M_t = 173.21 \text{ GeV} \tag{4.71}$$

Those compatibility checks enable to avoid the computation of integrals for spurious permutations mainly arising from combinatorial background (when considering wrong jet assignements for instance). If taken into account, the kinematic reconstruction introduced above would assign to the quarks p_T very far from the value of their associated jet. In that case, the integrals would have a negligible value due to the jet transfer functions and neglecting that computation does not change much the final result while speeding up the computation.

4.5.3 Leptonic top

The phase space associated to a leptonic top can be written as

$$d\Phi_{t_l} \propto dt^2 dW^2 \delta(t - W - b)\delta(W - \ell - \nu)\frac{d^3\vec{t}}{2E_t}\frac{d^3\vec{W}}{2E_W}\frac{d^3\vec{b}}{2E_b}\frac{d^3\vec{\ell}}{2E_\ell}\frac{d^3\vec{\nu}}{2E_\nu} \tag{4.72}$$

Taking into account the mass constraints of the top quark and the W boson, this can be reduced to[6]

$$d\Phi_{t_l} \propto \frac{|\vec{b}|E_\nu^2}{E_\ell}\frac{1}{\left|\frac{E_b}{|\vec{b}|}\vec{W}.\hat{\vec{e}}_b - E_W\right|}d\Omega_\nu \tag{4.73}$$

with

$$E_\nu = \frac{M_W^2}{2E_\ell(1 - \cos\theta_{\ell\nu})} \tag{4.74}$$

$$\vec{W} = \vec{\ell} + E_\nu\vec{e}_\nu \text{ and } E_W(E_\nu, \vec{e}_\nu) = \sqrt{M_W^2 + E_\ell^2 + E_\nu^2 + 2E_\ell E_\nu \cos\theta_{l\nu}} \tag{4.75}$$

$$E_b^\pm = \frac{E_W\Delta M^2 \pm |\vec{W}.\vec{e}_b|\sqrt{(\Delta M^2)^2 - m_b^2(E_W^2 - (\vec{W}.\vec{e}_b)^2)}}{E_W^2 - (\vec{W}.\vec{e}_b)^2} \tag{4.76}$$

and $\Delta M^2 = (M_t^2 - M_W^2 - m_b^2)/2$. Again in Eq. (4.76), some unphysical solutions may have been introduced, depending on the sign chosen, which can be removed by requiring the same conditions as in Eq. (4.69). If two admissible (positive) solutions exist, the largest one is chosen.

Similarly to the case of the hadronic top with a missing jet, a lower limit $M_t^{low}(b\ell)$ is defined by shifting the energy of the b jet according to the 95% lower edge of the jet transfer function and considering the invariant mass of the b jet and the lepton. If the following condtion is not fulfilled

$$M_t^{low}(b\ell) < M_t = 173.21 \text{ GeV} \tag{4.77}$$

the integration is skipped for the permutation and the value of the integral is set to zero.

In the case of a top decaying into $t \to b\nu\tau, \tau \to \tau_h\nu$ ("tauonic top"), the reduction of the dimensionality and the kinematic reconstruction is very similar to the one of the leptonic top, the four-momentum of the τ lepton replacing the one of the charged lepton ℓ. The $\tau \to \tau_h\nu$ decay is taken into account using the hadronic τ transfer function introduced in 4.4.4 and adding an additional integration over $d|\vec{\tau}|$. The top mass compatibility check is the same as for a leptonic top, using the measured τ_h momentum instead of the charged lepton momentum.

[6]The detailed computation is available in Appendix A.

4.5.4 Z boson decaying into leptons

In the case of a Z boson decaying into two leptons with one of them misreconstructed as a τ_h (labelled as ℓ_τ in that section), an integration over the p_T of the lepton is in principle required because of the finite resolution associated to the p_T measurement of such an object. However, this can be avoided if the Z boson is assumed to be on-shell. In that case, the phase-space associated to the Z boson is given by

$$d\Phi_Z \propto dZ^2 \delta(Z - \ell - \ell_\tau) \frac{d^3\vec{Z}}{2E_Z} \frac{d^3\vec{\ell}}{2E_\ell} \frac{d^3\vec{\ell_\tau}}{2E_{\ell_\tau}} \tag{4.78}$$

Taking into account the on-shell condition of the Z boson, the perfect momentum measurement assumed for the lepton ℓ, the perfect direction measurement assumed for ℓ_τ, it is reduced after integration over $\vec{Z}$ to

$$d\Phi_Z \propto \delta(E_Z(\ell, \ell_\tau) - E_\ell - E_{\ell_\tau}) \frac{E_{\ell_\tau}}{E_Z E_\ell} dE_{\ell_\tau} \tag{4.79}$$

with

$$E_Z(\ell, \ell_\tau) = \sqrt{M_Z^2 + E_\ell^2 + E_{\ell_\tau}^2 + 2E_\ell E_{\ell_\tau} \cos\theta_{\ell\ell_\tau}} \tag{4.80}$$

Integrating over E_{ℓ_τ} one gets

$$d\Phi_Z \propto \frac{1}{\left|\frac{\partial E_Z}{\partial E_{\ell_\tau}} - 1\right|} \frac{E_{\ell_\tau}}{E_Z E_\ell} \tag{4.81}$$

with E_{ℓ_τ} defined by

$$E_{\ell_\tau} = \frac{M_Z^2}{2E_\ell(1 - \cos\theta_{\ell\ell_\tau})} \tag{4.82}$$

This leaves finally no remaining integration and a simple Jacobian term

$$d\Phi_Z \propto \frac{E_{\ell_\tau}^2}{E_\ell} \tag{4.83}$$

4.6 Integration of the MEM in the VBF $H \to \tau\tau$ categories

Based on the previous tools developed in this thesis, a full analysis using the Matrix Element Method for signal extraction in the VBF $H \to \tau\tau$ categories has been presented in [147], focusing on the $H \to \tau\tau \to \mu\tau_h$ decay channel. The results presented in that reference are summarized here for completeness, to assess the improvement brought by the MEM with respect to the CMS $H \to \tau\tau$ analysis of Run I [153]. This analysis includes several categories targeting the two main production modes of the Higgs boson, the gluon-gluon fusion and the VBF production. Those categories are defined in particular based

on the jet multiplicity in the event. The so-called VBF categories include events with at least two jets with $p_T > 30$ GeV. To take into account the VBF topology, additional requirements are applied on the invariant mass of the two leading jets (m_{jj}) as well as the difference in pseudo-rapidity of those jets ($\Delta\eta_{jj}$). Moreover, due to the electroweak nature of the interaction between the two partons in the VBF process, additional hadronic activity is expected to be suppressed in the region between the two VBF jets. For this reason, a so-called "central jet veto", requiring that no additional jet with $p_T > 30$ GeV is present between the two leading jets, is also applied in the VBF categories. All the categories used in [153] are presented in Fig. 4.15.

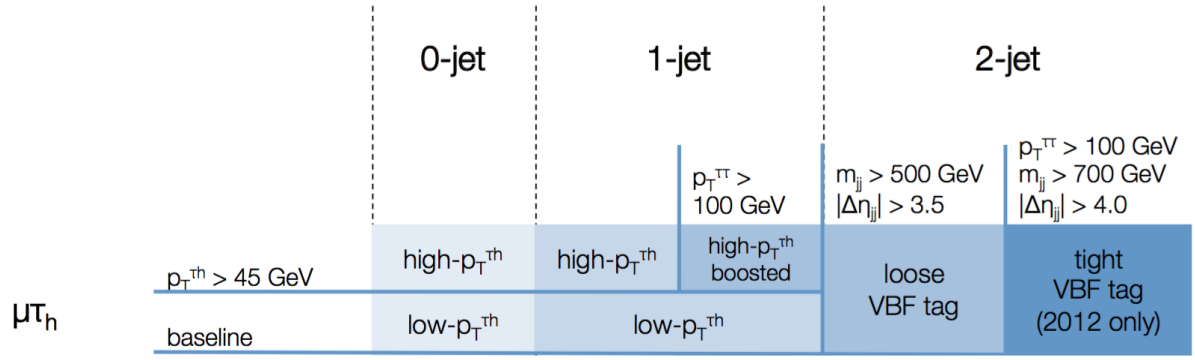

Figure 4.15: Categories used in the $H \to \tau\tau \to \mu\tau_h$ channel in the CMS $H \to \tau\tau$ Run I analysis [153]

4.6.1 Validation of the Matrix Element Method implementation

The full reconstruction of the di-τ system, starting from the variables integration variables $d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}}$, has been validated in a VBF Monte-Carlo sample. For that purpose, the reconstruction procedure detailed previously in section 4.5.1 has been used with the variables $|\vec{\tau}|$, $\cos\theta_{\tau\bar{\tau}}$ and $m^2_{\tau\bar{\tau}}$ set to their actual values. The results of that validation are presented in Fig. 4.16-4.18. As can be seen, an excellent agreement is observed in the quantities associated both to the leptonic τ and the hadronic τ, which shows the validity of the procedure described above. The small disagreements observed between the actual values in Monte-Carlo and the reconstructed values can be explained by the various approximations performed in the previous computations (including the collinear approximation for the leptonic τ).

As an additional cross-check, the dependence on the functions to integrate with respect to $m_{\tau\tau}$ has been checked. For that purpose, a VBF event and a Drell-Yan event have been chosen randomly in the Monte-Carlo samples used to evaluate those processes. All of the integration variables have then been set to their actual values except for $m_{\tau\tau}$ which has been varied. The result of that check is presented in Fig. 4.19. As can be seen, the

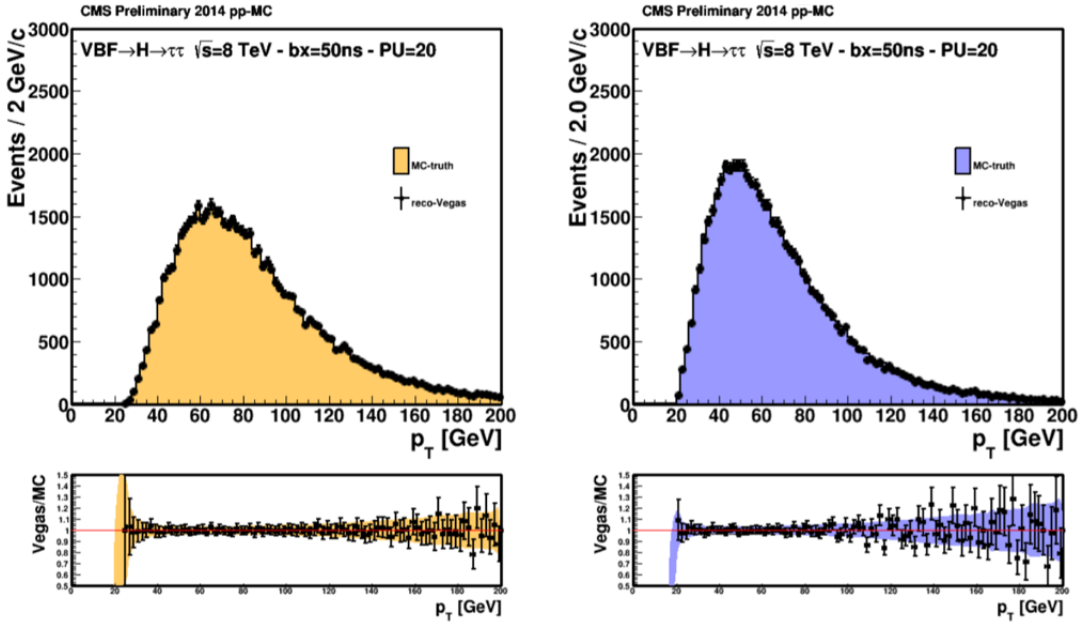

Figure 4.16: p_T distribution of the leptonic (left) and hadronic (right) τ, directly from the Monte-Carlo (filled) and as reconstructed by the MEM algorithm (dots) [147]

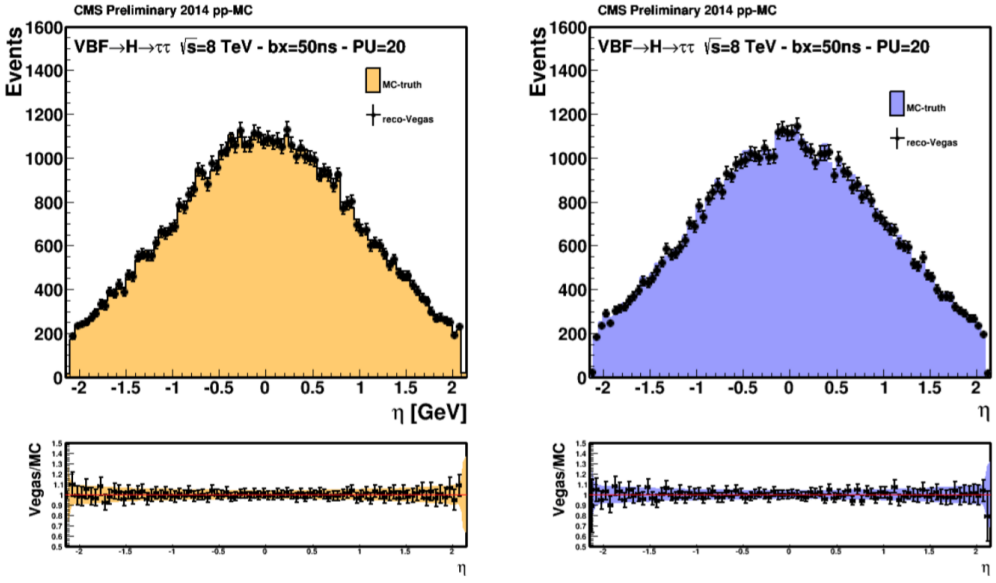

Figure 4.17: η distribution of the leptonic (left) and hadronic (right) τ, directly from the Monte-Carlo (filled) and as reconstructed by the MEM algorithm (dots) [147]

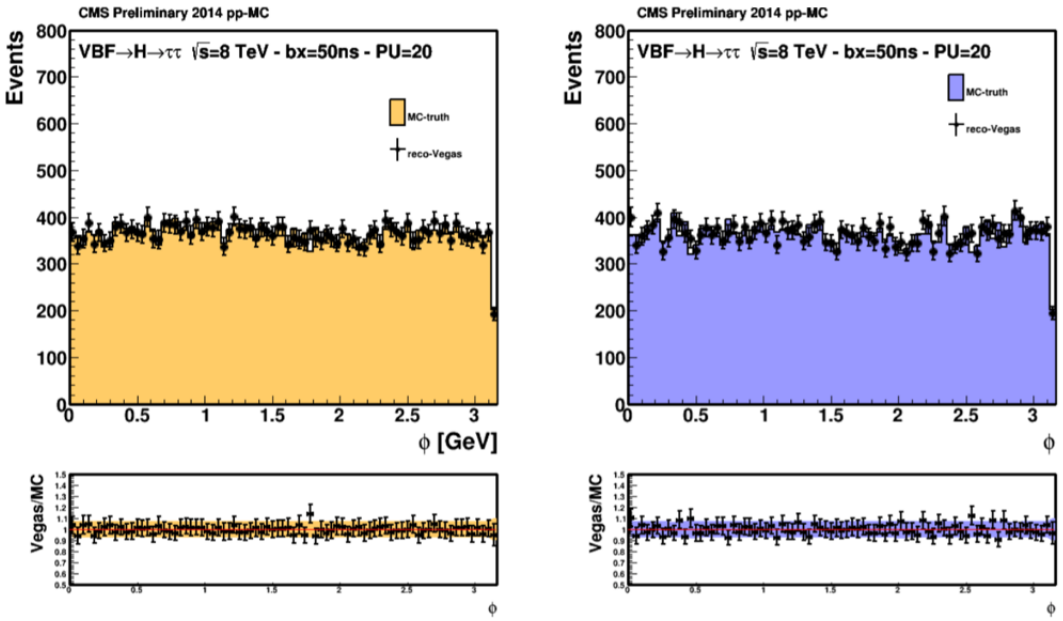

Figure 4.18: ϕ distribution of the leptonic (left) and hadronic (right) τ, directly from the Monte-Carlo (filled) and as reconstructed by the MEM algorithm (dots) [147]

VBF integrand (used to compute the VBF weight) peaks at 125 GeV while the Drell-Yan integrand (used to compute the Drell-Yan weight) peaks at 91 GeV, which is the expected behavior due to the resonant nature of the matrix elements.

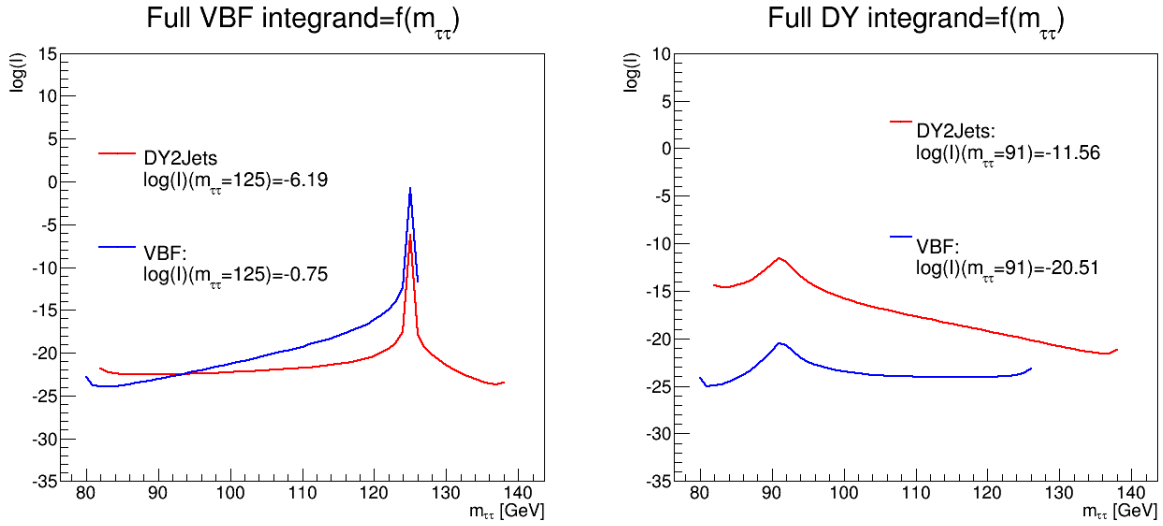

Figure 4.19: VBF (left) and Drell-Yan (right) integrand as a function of $m_{\tau\tau}$ for a VBF (blue) and a Drell-Yan (red) event

Based on those results, the VBF and Drell-Yan MEM weights (w_{VBF} and w_{DY}) have been computed in the VBF and the Drell-Yan samples, for events with one muon, one τ_h and two jets. The associated distributions are presented in Fig. 4.20 and some discrimination is already visible with those individual weights. The expected ordering is also observed, namely that the VBF weight is higher on average for VBF events while the Drell-Yan weight is higher on average for Drell-Yan events. To combine those weights, a likelihood ratio is defined as introduced in section 4.1

$$\mathcal{L} = \frac{w_{VBF}}{w_{VBF} + k w_{DY}} \tag{4.84}$$

The factor k is present to adjust the relative values of the VBF and Drell-Yan weights. Its numerical value however does not impact the discrimination power of the likelihood ratio, as any cut on $\mathcal{L}$ is equivalent to a cut on the ratio w_{DY}/w_{VBF} which does not depend on k. In the following results, its numerical value has been chosen to get a crossing point of the VBF and Drell-Yan distributions around $\mathcal{L} = 0.5$. The distribution of the likelihood ratio is presented in Fig. 4.21.

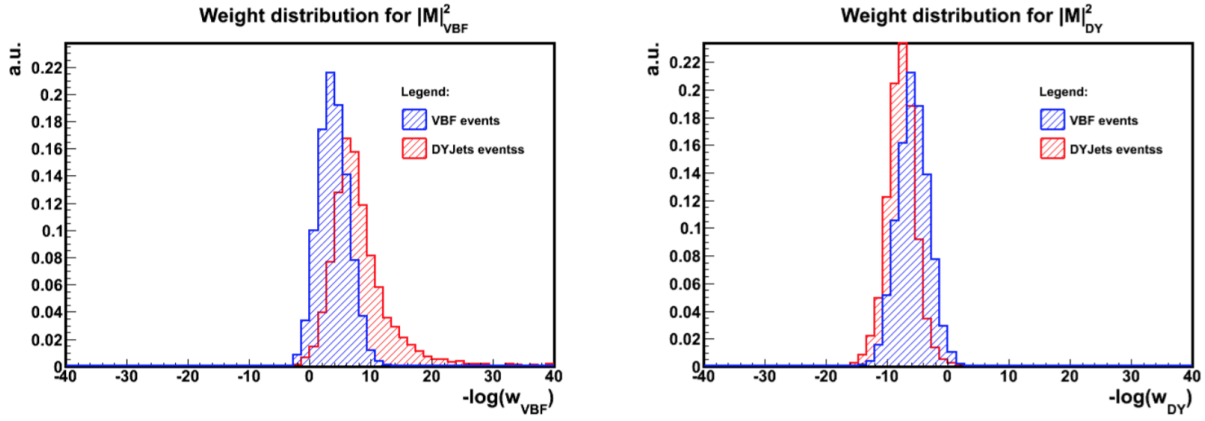

Figure 4.20: VBF (left) and Drell-Yan (right) weight distributions in the VBF (blue) and Drell-Yan (red) samples [147]

To further validate the MEM implementation, the correlation of the likelihood ratio with VBF-associated variables has been investigated, namely the invariant mass of the two leading jets m_{jj} and their pseudo-rapidity difference $\Delta\eta_{jj}$. Due to the VBF topology with two forward jets, those variables are expected to be higher on average in VBF events than in Drell-Yan events. Selecting VBF-like events based on the value of the MEM likelihood ratio should therefore reflect those characteristics and it is indeed what is observed in Fig. 4.22.

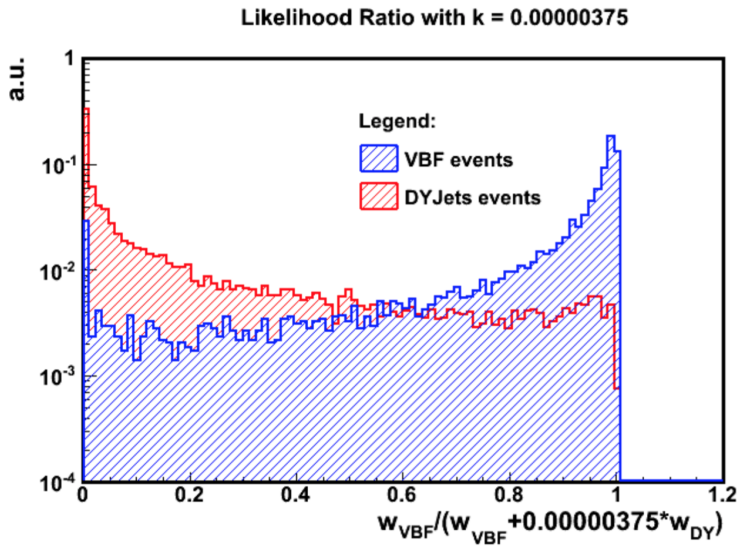

Figure 4.21: MEM likelihood ratio distribution in the VBF (blue) and Drell-Yan (red) samples [147]

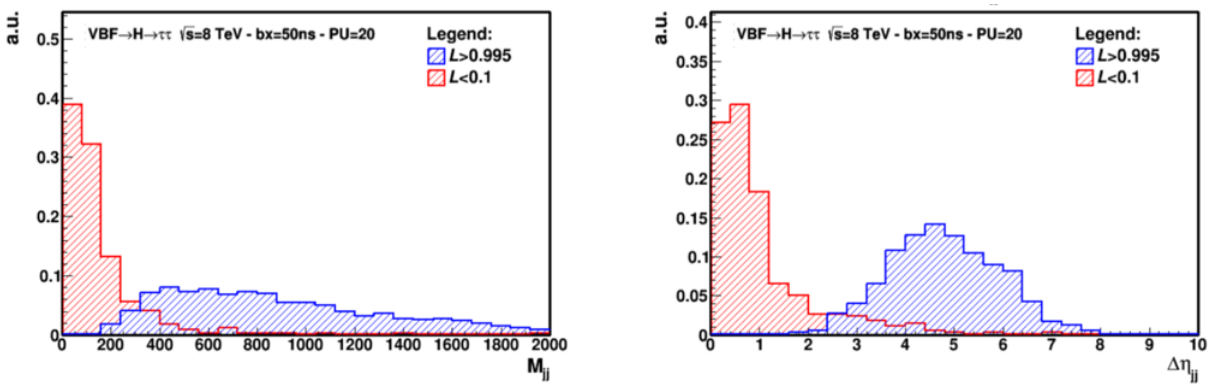

Figure 4.22: m_{jj} (left) and $\Delta\eta_{jj}$ (right) distributions evaluated in the VBF sample for events with $\mathcal{L} < 0.1$ (red) and $\mathcal{L} > 0.995$ (blue)

4.6.2 Results

After the analysis selections, the main background in all the categories is the Drell-Yan $Z \to \tau\tau$ production. Another important background corresponds to the electroweak production of a W boson decaying into a muon and a neutrino in association with jets, among which one fakes a τ_h. The remaining backgrounds consist essentially in $t\bar{t}$ production and in pure QCD processes. Most of those backgrounds are evaluated using data-driven techniques and the interested reader can find more details about the background estimation techniques and the associated systematics in [147] and [153].

To compare the performance of the MEM with the default analysis, the following strategy has been adopted. In the VBF categories, a binned maximum likelihood fit is performed for the signal extraction using either the MEM distribution or the distribution of the di-τ mass estimate. Because the number of events in those categories is relatively low, the background shapes and yields are poorly constrained if those categories only are considered. Therefore, the 0- and 1-jet categories have also been taken into account in the analysis, using the distribution of the di-τ mass estimate for the signal extraction as in [153].

The detailed statistical procedure applied for this analysis is detailed in section 5.5, as it is the same as the one used for the $t\bar{t}H$ analysis targetting final states with τ_h. The expected and observed 95% C.L. exclusion limits are displayed in Fig. 4.23. As expected in the MEM analysis, the highest sensitivity is achieved for m_H =125 GeV, which is consistent with the fact that the implementation of the MEM has been optimized for that mass value. The significance of the observed excess is presented in Fig. 4.24. In both cases, the MEM analysis shows an improvement with respect to the default analysis, demonstrating the potential of this method. The expected 95% C.L. exclusion limits and significance at m_H =125 GeV are summarized in Table 4.5.

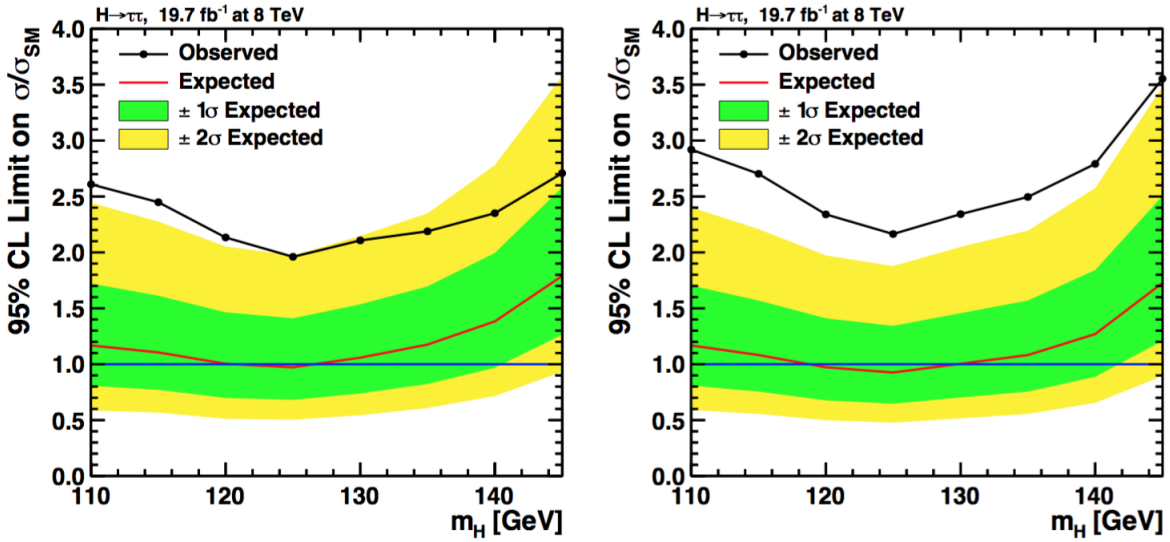

Figure 4.23: 95% C.L. upper limit on $\mu = \sigma/\sigma_{SM}$ as a function of the Higgs boson mass m_H in the $H \to \tau\tau \to \mu\tau_h$ channel for the default analysis (left) and the analysis using the MEM in the VBF categories (right). The observed limit (black dots) is represented together with the expected limit (red line). The bands show the expected one- and two-standard-deviation probability intervals around the expected limit. [147]

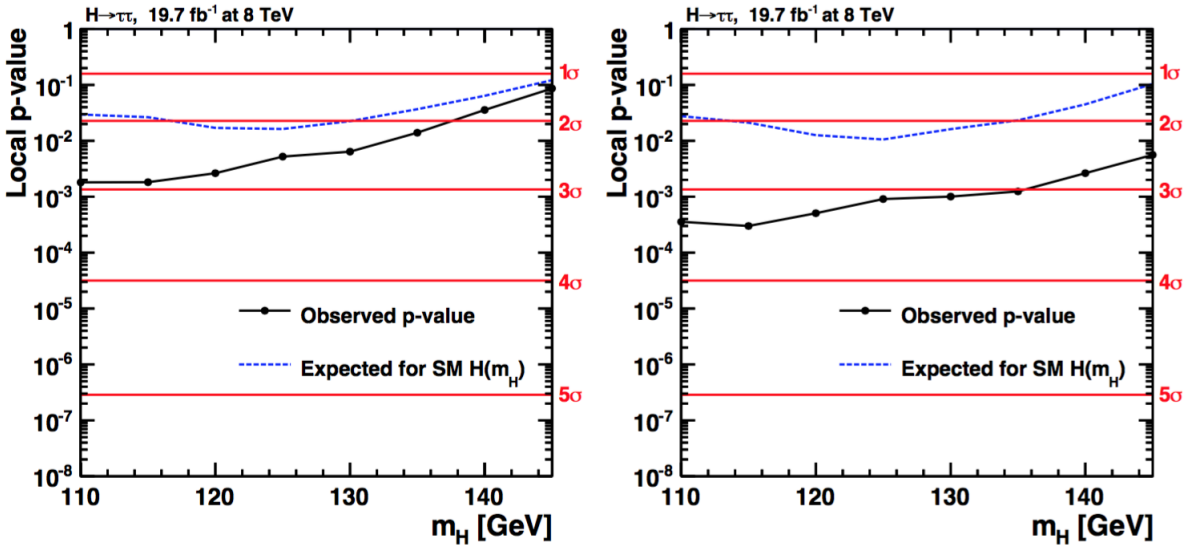

Figure 4.24: Local p-value and significance in number of standard deviations as a function of the Higgs boson mass m_H for the default analysis (left) and the analysis using the MEM in the VBF categories (right). The observation (solid line) is compared to the expectation (dashed line) for a Standard Model Higgs boson with mass m_H. [147]

	Default analysis	**Analysis with MEM in the VBF categories**
Expected 95% C.L. upper limit on $\mu = \sigma/\sigma_{SM}$ (m_H=125 GeV)	0.97	0.93
Expected significance (m_H=125 GeV)	2.1σ	2.3σ

Table 4.5: Expected 95% C.L. upper limit and expected significance at m_H=125 GeV for the default analysis and the analysis using the MEM in the VBF categories in the $H \to \tau\tau \to \mu\tau_h$ decay channel.

4.7 Conclusion

In this thesis, significant contributions to the development of generic tools useful for the implementation of the Matrix Element Method have been developed in the context of analyses with a Higgs boson decaying into τ leptons. The 8 TeV VBF $H \to \tau\tau$ analysis presented in [147] has thus been the first analysis where the MEM has been successfully implemented for the search of a resonance decaying into τ leptons. The different conclusive checks presented in this chapter gave confidence in the MEM implementation presented here and the observed improvement with respect to the default 8 TeV $H \to \tau\tau$ analysis showed the potential of this method, which can now be applied to more complex final states. This is the purpose of the next chapter where the MEM will be applied to the search for the associated production of the Standard Model Higgs boson with top quarks in final states with a τ lepton.

Chapter 5

Search for the $t\bar{t}H$ process in final states with a τ lepton

Rien ne donne le repos que la recherche sincère de la vérité.

Blaise Pascal

The study of the $t\bar{t}H$ process offers a good opportunity to study the coupling of the Higgs boson to the top quark. Thanks to the substantial increase in the cross-section of this process and the large luminosity collected, this process is expected to be observed by the end of LHC Run 2. In order to optimize this search, all the different possible Higgs decay modes are investigated. Due to the different final states, the analysis techniques and signal extraction methods need to be specifically optimized for each final state though.

In this thesis, the development of the first dedicated analysis targeting the $t\bar{t}H$ production of a Higgs boson decaying into τ leptons using 2016 data is presented. My work specifically focused on the most sensitive category of this analysis, covering events with two same-sign leptons and one hadronically-decaying τ. Indeed, this category benefited from the development of a discriminant based on the Matrix Element Method, making use of the tools presented in the previous chapter. In addition, my contributions to the combination of this category with other categories sensitive to the $t\bar{t}H, H \to \tau\bar{\tau}$ production are presented in this chapter as well as the specific details of this analysis presented in the Physics Analysis Summary [154].

5.1 Introduction

Analyses focusing on the search for the $t\bar{t}H$ production in final states with several leptons are mostly sensitive to the $H \to WW/ZZ$ decay modes. Still, as mentioned in section 1.3.5, some BSM models predict different modifications of the signal strengths of the $t\bar{t}H$ production for different decay modes of the Higgs boson, due to the modification of several

Higgs boson couplings and not only the top-Higgs coupling. In that sense, the search for the $t\bar{t}H$ production in different decay modes provides additional checks of the Standard Model, which can be confronted with different BSM models. The contribution from $H \to \tau\bar{\tau}$ decays can in particular be enhanced by requiring the presence of reconstructed τ_h's in the final state.

The $t\bar{t}H, H \to \tau\bar{\tau}$ process produces three massive objects, with uncorrelated decay modes. For instance, one light charged lepton (electron or muon) can be produced in one of the top quark decays while another one is produced in the subsequent decay of one of the τ lepton. One of the advantages of having those uncorrelated decays is that in 50% of the cases, the leptons will have the same charge, while in most of the Standard Model processes where two leptons are produced they have opposite charges. Selecting events with two same-sign leptons and one τ_h enables thus to strongly reduce the contribution of most of the Standard Model backgrounds. The contribution from events where a pair of vector bosons is produced (diboson) can be reduced by requiring the presence of b-tagged jets. The main irreducible backgrounds consist essentially in events where a top quark pair is produced in association with a vector boson ($t\bar{t}V$). As mentioned in section 4.4.7, additional leptons can also be produced in semileptonic decays of B hadrons. This makes the $t\bar{t}$ production a sizable source of background, which can be reduced by applying tight isolation and identification criteria for the leptons but still yields a non-negligible contribution, due to its relatively large cross section in comparison to $t\bar{t}H$. The object and event selections in the two same-sign leptons + one τ_h ($2\ell ss + 1\tau_h$) category, as well as the background estimation techniques will be presented in section 5.2 and 5.3.

Even after the event selection, the background yield remains sizable in comparison to the signal yield and the signal extraction can be improved with the use of dedicated variables with some discriminating power between the $t\bar{t}H$ signal and the different backgrounds. I developed for that purpose a Matrix Element Method (MEM) discriminant optimized to separate the $t\bar{t}H, H \to \tau\bar{\tau}$ signal component from the irreducible $t\bar{t}Z$ and the reducible $t\bar{t}$ backgrounds. This discriminant benefits from all the tools presented in the previous chapter and its specificities will be presented in details in section 5.4. Finally, the results obtained with events with two same-sign leptons and one τ_h have been combined with other categories sensitive to the $t\bar{t}H, H \to \tau\bar{\tau}$ production in order to measure the signal strength of this process. The statistical procedure used for the signal extraction will be introduced in section 5.5 and the results will be presented in section 5.6.

5.2 Objects and event selection

As mentioned in section 2.3, the information provided by all CMS subdetectors is used by a Particle-Flow algorithm [127, 155, 156, 121, 117] to identify and reconstruct individual particles in the event, namely muons, electrons, photons, charged and neutral hadrons. These particles are then used to reconstruct jets, τ_h candidates and the missing transverse energy vector $\vec{\not{E}}_T$, as well as to quantify the isolation of leptons. Those objects are then used for the event selection and the computation of the MEM discriminant.

5.2.1 Electrons and muons

The identification of electrons and muons is performed in two stages. Basic electron (muon) identification criteria and loose isolation criteria are applied in the first stage to separate genuine leptons from jet backgrounds. In the second stage, the leptons originating from decays of W and Z bosons and from τ lepton decays are separated from leptons produced in the decays of bottom quarks or from misreconstructed hadrons. The former are referred to as "prompt" leptons while the latter as "non-prompt" leptons. Dedicated multivariate algorithms have been developed in the context of the $t\bar{t}H$ multilepton analysis [157, 158] to distinguish the prompt leptons from the non-prompt leptons.

Basic muon identification

Muons reconstructed by the Particle-Flow algorithm are used in the analysis. As described in section 2.3.3, muon candidates are reconstructed combining the information from both the silicon tracker and the muon spectrometer in a global fit. Different identification selections are then performed using the quality of the geometrical matching between the tracker and the muon system measurements, defining a loose and medium muon identification. Only muons within the muon system acceptance $|\eta| < 2.4$ and minimum p_T cuts of 5 GeV are considered.

Basic electron identification

As described in section 2.3.3, electrons are reconstructed using tracking and electromagnetic calorimeter information by combining ECAL superclusters and Gaussian sum filter (GSF) tracks. Electrons are required to have $|\eta| < 2.5$ to ensure that they are within the tracking volume and a minimum p_T of 7 GeV. The electron identification is performed using a multivariate discriminant built with variables related to the shape of the electromagnetic shower reconstructed in ECAL and to the associated tracks. A loose selection based on η-dependent cuts on this discriminant is used to preselect electron candidates. This discriminant is used as well for the lepton multivariate selection to separate prompt leptons from non-prompt leptons.

In order to remove electron candidates that are due to photon conversions, it is also required that the electron track is associated to a hit in each layer of the pixel detector that is crossed by the track, except for at most one layer. Electron candidates are also rejected in case there exists a track of opposite charge near the electron track that, if paired with the electron track, can be fitted to a common vertex within the volume of the tracking detector. For reasons which will be described in section 5.3.2, additional identification criteria are also applied for electrons with p_T greater than 30 GeV to mimic the identification applied at the trigger level for single electron triggers.

Lepton vertexing

Collision vertices are reconstructed using a deterministic annealing algorithm [159, 160]. The reconstructed vertex position is required to be compatible with the location of the LHC beam in the $x-y$ plane. The tracks associated to each vertex are clustered using the anti-k_T algorithm [128]. For each vertex, a weighted p_T^2 is then computed using those jets, remaining single tracks including identified leptons and the missing transverse momentum associated to the vertex, to account for neutral particles. The primary vertex (PV) is then selected as the one with the largest p_T^2.

In order to reject pile-up or misreconstructed tracks, and more importantly to reject non-prompt leptons from B hadron decays, the following impact parameter variables are also considered: the impact parameter in the transverse plane d_{xy}, the impact parameter along the z axis d_z and the signed impact parameter, in three dimensions, of the lepton track with respect to the PV, divided by its uncertainty, which corresponds to its significance d/σ_d.

Electron and muon isolation

Electrons and muons in signal events are expected to be isolated, while leptons from B hadrons decays, as well as from in-flight decays of pions and kaons, are often reconstructed within jets. Isolated leptons are distinguished from leptons in jets by means of the sum of scalar p_T values of charged particles, neutral hadrons, and photons, that are reconstructed within a narrow cone centered on the lepton direction. The size R of the cone shrinks inversely proportional with the p_T of the lepton in order to increase the efficiency for leptons reconstructed in events with high hadronic activity to pass the isolation criteria. The narrow cone size has the further advantage that it reduces the effect of pileup. Efficiency loss due to pileup is further reduced by considering only charged particles originating from the lepton production vertex in the isolation sum. Residual contributions of pileup to the neutral component of the isolation of the lepton is taken into account by means of so-called effective area corrections:

$$I_\ell = \sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutrals}} p_T - \rho\,\mathcal{A}\left(\frac{R}{0.3}\right)^2\right), \tag{5.1}$$

where ρ represents the energy density of neutral particles reconstructed within the geometric acceptance of the tracking detectors, computed as described in Refs. [161, 162]. The size of the cone is given by:

$$R = \begin{cases} 0.05 & \text{if } p_T > 200 \text{ GeV} \\ 10 \text{ GeV}/p_T & \text{if } 50 < p_T < 200 \text{ GeV} \\ 0.20 & \text{if } p_T < 50 \text{ GeV} \end{cases}. \tag{5.2}$$

The effective area $\mathcal{A}$ is obtained from the simulation, by studying the correlation between I_ℓ and ρ, and is determined in bins of η, separately for electrons and muons.

Lepton MVA discriminator

To discriminate prompt leptons from non-prompt leptons produced in the decays of B hadrons, additional variables are used in addition to the vertexing and isolation variables. These additional variables are related to the jet reconstructed closest to the lepton. To avoid over-correction of the jet due to the clustering of the prompt lepton, the jet energy corrections are only applied to the hadronic part of the jets. The additional variables related to the jets are then the ratio between the p_T of the lepton and the p_T of the jet, the value of the CSV b-tagging discriminator (see section 2.3.5), the number of charged tracks of the jets and the transverse momentum of the lepton with respect to the jet.

In order to optimally combine the discrimination from all those variables, a multivariate discriminant has been trained using prompt leptons passing loose preselections from a $t\bar{t}H$ signal Monte Carlo sample and non-prompt leptons from a $t\bar{t}$+jets Monte Carlo sample, separately for electrons and muons. This MVA uses as input variables the vertexing, isolation and jet-related variables described so far, the p_T and η of the lepton and two additional variables that contribute to make it robust also in the rejection of leptons from light jets misidentification: the electron MVA ID discriminator and the muon segment-compatibility variables. The distributions of some of those variables obtained for prompt and non-prompt leptons in a $t\bar{t}$ Monte Carlo sample are presented in Fig. 5.1.

The *tight* lepton selection, used in the signal region, corresponds to the requirement lepMVA>0.75. The closest jet must also fail the medium working point of the CSV discriminator (CSV<0.8484). A so-called *fakeable* selection is also used for the data-driven estimation of backgrounds with non-prompt leptons, presented in details in section 5.3.2. This selection is obtained by relaxing the lepMVA requirement. Some additional requirements are applied to leptons failing the tight selection based on the jet variables, in order to reduce the systematic uncertainty on the fake lepton background related to the jet flavor composition.

5.2.2 Hadronic τ decays

The hadronic τ decays are reconstructed using the hadrons-plus-strips algorithm. The identification of the τ_h candidate is then based on an MVA discriminant, whose inputs variables have been described in section 2.3.6. However, with an isolation cone of $\Delta R = 0.5$, as it is the cases in most analyses using τ_h, it has been found that this isolation cone size leads to a sizable inefficiency for the τ_h selection in $t\bar{t}H, H \to \tau\bar{\tau}$ events, due to the high hadronic activity in those events. The cone size for the isolation variable used as input of the MVA has thus been reduced to $\Delta R = 0.3$, which has been determined to be the optimal one in terms of signal efficiency and background rejection.

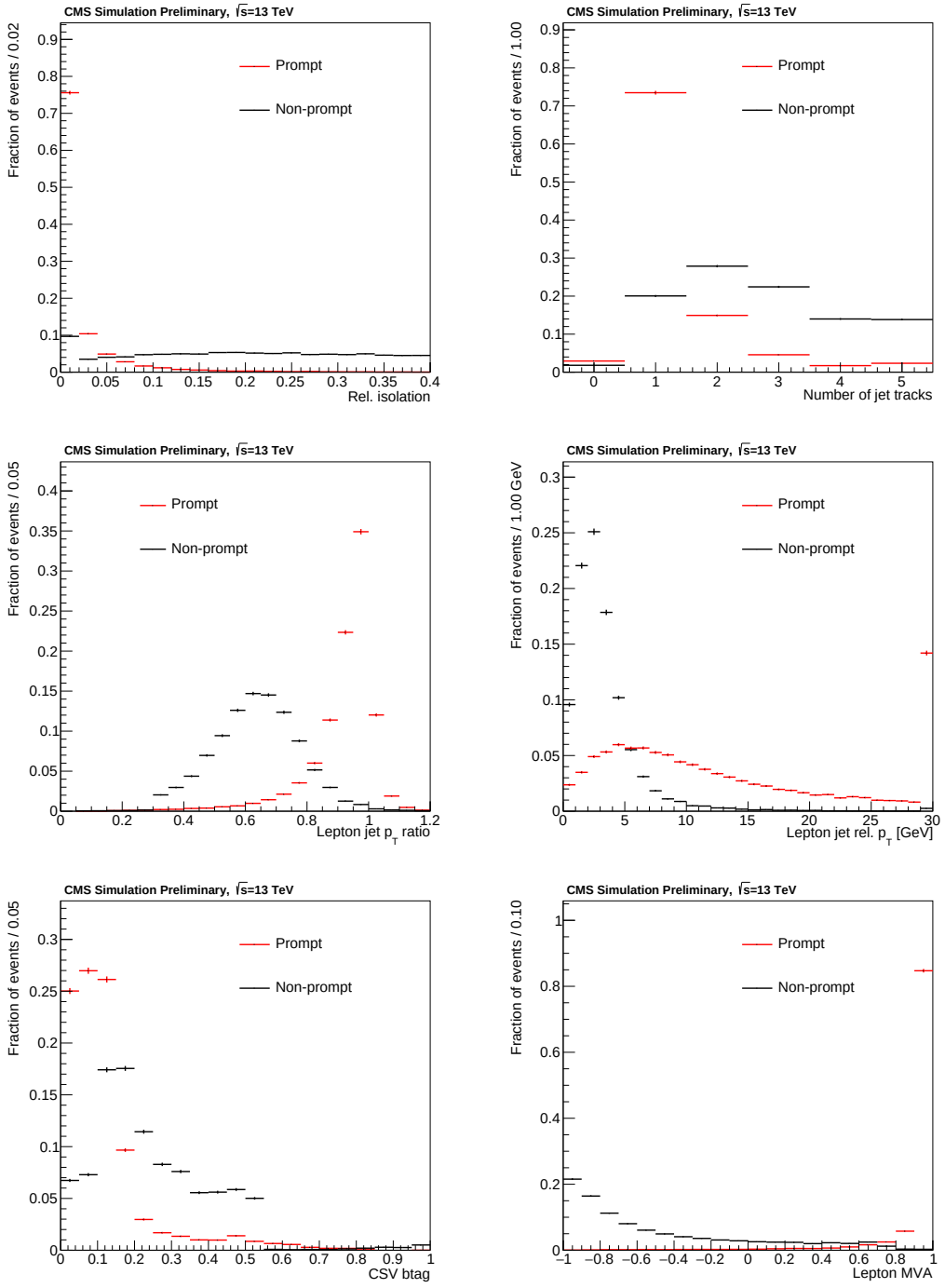

Figure 5.1: Distribution of some of the input variables and of the MVA used for the lepton selection, obtained in $t\bar{t}$ Monte Carlo events with two same-sign fakeable leptons. Prompt and non-prompt leptons are defined based on a generator-level matching with a lepton produced in the decay of a W boson. The variables presented are: the relative isolation of the lepton (top left), the number of charged tracks in the nearby jet (top right), the ratio of the lepton and nearby jet p_T (middle left), the transverse momentum of the lepton with respect to the nearby jet (middle right), the CSV b-tagging discriminator of the nearby jet (bottom left) and the lepton MVA.

τ_h candidates are then required to have a $p_T > 20$ GeV and $|\eta| < 2.3$ and not to overlap within $\Delta R = 0.3$ with any muon or electron passing loose identification criteria. No dedicated anti-lepton discriminator are used beyond this overlap removal as they do not improve the sensitivity of the analysis. The medium working point of the MVA Tau ID is used in the $2\ell ss + 1\tau_h$ category, corresponding to an average τ_h selection efficiency around 50%.

5.2.3 Jets and missing energy

Jets are reconstructed using the anti-k_T algorithm [128] with a distance parameter $R = 0.4$. Charged particles not originating from the primary vertex are excluded from the jet clustering. Fake jets, mainly arising from calorimeter noise, are rejected by requiring reconstructed jets to pass a set of loose jet identification criteria. As mentioned in section 2.3.4, jet energy corrections depending on p_T and η are applied. Jets are only considered if they have a $p_T > 25$ GeV and $|\eta| < 2.4$. In addition, they have to be separated from any fakeable lepton and loose τ_h candidates. Jets originating from the hadronization of b quarks are identified by the CSV algorithm (see section 2.3.5). The loose and medium CSV working points are used to select jets for this analysis.

The type-1 PF $\vec{\not{E}}_T$ is used as estimate of the missing transverse energy (see section 2.3.7). Its magnitude is referred to as E_T^{miss}. To mitigate the influence of pile-up, the missing transverse energy is complemented by the observable H_T^{miss}, defined as the magnitude of the vectorial transverse momentum sum of leptons, τ_h, and jets:

$$H_T^{miss} = |\sum_{\text{leptons}} \vec{p}_{T\ell} + \sum_{\tau_h} \vec{p}_{T\tau} + \sum_{\text{jets}} \vec{p}_{Tj}| \,. \tag{5.3}$$

The H_T^{miss} variable has a worse resolution than E_T^{miss} but is more robust with respect to pile-up as it does not rely on the soft part of the event. A linear combination of those two variables

$$E_T^{miss}LD = 0.6 \times E_T^{miss} + 0.4 \times H_T^{miss} \,, \tag{5.4}$$

is used for the event selection, based on the fact that E_T^{miss} and H_T^{miss} are less correlated in events with instrumental missing energy with respect to events with genuine missing energy.

5.2.4 Event selection in the $2\ell ss + 1\tau_h$ category

Events in the $2\ell ss + 1\tau_h$ category are selected by a combination of triggers based on the presence of one or two leptons, either electrons or muons, in the event. The p_T threshold of the single electron (muon) trigger amounts to 27 GeV (22 GeV). The p_T thresholds of the double electron (muon) triggers amount to 23 GeV (17 GeV) for the leading and 12 GeV (8 GeV) for the subleading lepton. Events containing electron plus muon pairs pass the trigger selection if either the electron satisfies the condition $p_T > 23$ GeV and the muon

$p_T > 8$ GeV or the muon satisfies the condition $p_T > 23$ GeV and the electron $p_T > 8$ GeV. The complete list of HLT paths and L1 seeds can be found in Table 3.1.

The $2\ell ss + 1\tau_h$ category includes events with exactly two tight leptons together with at least one τ_h. The two selected leptons are required to have the same charge. The leading lepton must have a transverse momentum larger than 25 GeV, while the subleading lepton must be above 10 GeV for a muon and 15 GeV for an electron. At least one τ_h is required to pass the medium working point of the Tau ID and the leading selected τ_h must have an opposite charge with respect to the leptons. In addition, selected events are also required to contain at least three jets, including either at least two jets passing the loose CSV working point or at least one jet passing the medium CSV working point.

To reduce the acceptance to events where the charge of one of the leptons is mismeasured, additional requirements are applied on the quality of the charge assignement. For the electrons, the consistency between the independent measurements of the charge from the ECAL supercluster and from the tracker is required, while for the muons the track transverse momentum must measured with a relative uncertainty lower than 20%. Those requirements are further referred to as *tight-charge* criteria. The background from electrons from Z boson decays, where the charge of one electron is mismeasured, is further suppressed by vetoing events where the di-electron invariant mass is within a 10 GeV-window around the Z mass. For the same reason, the $E_T^{miss}LD$ is required to be larger than 30 GeV for di-electron events. Events containing lepton pairs of mass less than 12 GeV are rejected as well, as these events are not well modeled in the Monte Carlo samples.

5.3 Signal and backgrounds estimation

5.3.1 Monte Carlo estimation

Samples

The shape templates for the $t\bar{t}H$ signal and the backgrounds with two same-sign prompt leptons are estimated using Monte Carlo simulated samples. The signal events are generated for a H boson mass of $m_H = 125$ GeV. The dominant irreducible backgrounds consist of the production of a top quark pair in association with a Z or a W boson. The other irreducible backgrounds include the diboson electroweak production as well as some rare SM processes, such as triboson production, the production of a single top quark in association with a Z boson or the production of same-sign WW boson pairs. The background estimation includes also a contribution from $t\bar{t}$ events with real photons, where a photon undergoes an asymmetric conversion $\gamma \to e^+e^-$ in which one electron or positron carries most of the energy of the photon, while the other is of low energy and fails to get reconstructed. For all those processes, contributions from events both with a genuine τ_h or a fake τ_h (from jets or leptons) are considered.

Monte Carlo samples have been generated using next-to-leading-order matrix elements implemented in the generators MadGraph5_aMC@NLO [91] and POWHEG [163, 164, 165], interfaced with PYTHIA 8.2 [92] to model the parton shower and fragmentation.

The samples are generated with the NNPDF3.0 set of parton distribution functions (PDF) [166, 167, 168]. The decays of τ leptons, including polarization effects, are modelled by PYTHIA. Minimum bias events generated with PYTHIA are overlaid on all simulated events, according to the luminosity profile of the analyzed data and for a pp inelastic cross section of 69.2 mb. In the analyzed dataset, approximately 30 inelastic pp interactions (pileup) occur per bunch crossing on average. All generated events are passed through a detailed simulation of the CMS apparatus, based on GEANT4 [169], and are reconstructed using the same version of the CMS event reconstruction software as used for data.

Data-to-Monte Carlo corrections

To correct for small data-to-Monte Carlo mismodellings, so-called "scale factors" are computed as ratios of reconstructions and/or identification efficiencies between data and Monte Carlo. They are used to reweight the Monte Carlo events and correct for those differences. Corrections are applied for the following efficiencies:

- **Trigger efficiency**: The combined trigger efficiency of single and double lepton triggers has been measured in data collected with a MET trigger, avoiding thus any bias in the measurement. The data/MC scale factors have been measured to be 1.01 ± 0.02 for events with two electrons, 1.01 ± 0.01 for events with one electron and one muon and 1.00 ± 0.01 for events with two muons.

- **Lepton identification efficiency**: The lepton identification efficiency has been measured in data separately for muons and electrons, using $Z \to \ell\ell$ events selected using the tag-and-probe method (see section 3.3.2). Scale factors are computed as a function of the lepton p_T and η and the event scale factor is computed as the product of individual lepton scale factors.

- τ_h **identification efficiency and mistag rate**: The τ_h identification efficiency has been measured in data, using $Z \to \tau\tau \to \mu\tau_h$ tag-and-probe events [136], and has been found to be compatible with the τ_h identification efficiency obtained in simulation. The mistag rate due to fake τ_h from jets has been measured in $t\bar{t} \to b\bar{b}e\mu\nu\nu$ events and a corresponding scale factor is applied to simulated events with a fake τ_h from a jet.

- **b-tag efficiency and mistag rate**: The full shape of the CSV discriminant used for b-jet identification is corrected in simulation to match the one observed in data. This shape is measured in $t\bar{t}$ events for b quarks and in $Z \to \ell\ell$+jets events for light quarks [170]. The event scale factor is then computed as the product of the individual jet scale factors.

5.3.2 Non-prompt leptons

The background contribution due to events with at least one non-prompt leptons is evaluated from data using a loose-to-tight extrapolation developed in the CMS $t\bar{t}H$ multilepton

analysis [157, 158]. This method requires to measure the fake rate for non-prompt leptons passing the tight lepton MVA selection, which is then used to extrapolate the contamination of such events in the signal region.

Fake rate measurement

The region used to measure the fake rate is enriched in QCD jet events and obtained selecting events with one loose lepton and a jet well separated from the lepton. Events are selected at trigger level using prescaled single lepton triggers. At low muon p_T, a jet with p_T greater than 40 GeV is also required in the trigger. Studies based on simulated samples have shown that the trigger requirements could affect the fake rate, which represents an issue as events in the signal region can be triggered by single or double lepton triggers. In particular, a trigger with a given threshold can only be used to measure the fake rate for leptons with significantly higher p_T. Moreover, single electron triggers used in the signal region include tight isolation criteria which are not used for double lepton triggers or for the triggers used in the fake rate measurement region. Selections reproducing the HLT isolation cuts are therefore used for electrons with $p_T > 30$ GeV to avoid any bias in the fake rate.

The fake rate f is then measured as the fraction of fakeable leptons (as defined in section 5.2.1) passing the tight MVA selection as a function of p_T, η and flavor of the lepton, after subtraction of backgrounds with prompt leptons. The lepton identification criteria used on the fakeable object affect mostly the fake rate for fake leptons from hadrons, while a cut on the CSV discriminant of the jet associated to the fakeable lepton can alter the fake rate for non-prompt leptons from b-jets. The fakeable selection has thus been tuned in order to reduce the flavor dependency of the fake rate and thus reduce the associated systematic uncertainty.

Extrapolation to the signal region

The contamination from events with non-prompt leptons in the signal region is evaluated using a so-called "application region", defined with the same selections as the signal region, except that at least one of the leptons must pass the fakeable selection while failing the tight MVA selection. The background prediction due to events with fake leptons can then be obtained by weighting the events in the application region with:

- $f/(1-f)$ for events with a single failing lepton, where f is the fake rate of the failing lepton
- $-f_1 f_2/((1-f_1)(1-f_2))$ for events with two failing leptons, where f_1 and f_2 are the fake rates of the two failing leptons

The second case allows to correct for the contamination from events with two non-prompt leptons in events with a single failing lepton. Contamination from events with two prompts leptons is subtracted from the application region using simulation in order to avoid double-counting in the signal region.

5.3.3 Charge mis-identified leptons

The background where the charge of one of the lepton is misreconstructed arises mainly from $t\bar{t} \to b\bar{b}\ell^+\ell^-\nu\nu$+jets events. The charge misidentification rate for electrons and muons has been measured in data as a function of p_T and η, using $Z \to ee$ and $Z \to \mu\mu$ events. This rate is found to be negligible for muons, while it varies from around 0.02% in the barrel up to around 0.4% in the endcaps for electrons. This background is estimated from data, using events that pass all selection criteria of the signal region, except that the two leptons are required to be of opposite charge. The events in this application region are then reweighed by the charge misidentification rate associated to the lepton with opposite-sign with respect to the selected τ_h.

5.4 Matrix Element Method in the $2\ell ss + 1\tau_h$ category

5.4.1 Implementation and object assignement

After the event selection, the dominant background in the $2\ell ss + 1\tau_h$ category is the production of a Z boson in association with a pair of top quarks ($t\bar{t}Z$). The reconstructed τ_h can either correspond to a genuine hadronic decay of a τ (from a $t\bar{t}Z, Z \to \tau\tau$ event) or a misreconstructed electron or muon (from a $t\bar{t}Z, Z \to \ell\ell$ event). A sizable background contribution also originates from events with a non-prompt lepton, mainly from $t\bar{t}$ events where the non-prompt lepton is produced in a b quark hadronization and is isolated enough to pass the tight lepton MVA selection. In that case, the associated b-jet is often not considered in the analysis because it overlaps with the tight lepton, as illustrated in Fig. 5.2.

Consequently, the following processes[1] have been considered in the computation of a Matrix Element Method (MEM) discriminant, using the various tools presented in Chapter 4.

- $gg/q\bar{q} \to t\bar{t}H, H \to \tau\bar{\tau}$ with the subsequent decays $t \to b\ell^+\nu$, $\bar{t} \to \bar{b}q\bar{q}'$, $\bar{\tau} \to \ell^+\bar{\nu}_\tau\nu_\ell$ and $\tau \to \tau_h\nu_\tau$
- $gg/q\bar{q} \to t\bar{t}Z, Z \to \tau\bar{\tau}$ with the subsequent decays $t \to b\ell^+\nu$, $\bar{t} \to \bar{b}q\bar{q}'$, $\bar{\tau} \to \ell^+\bar{\nu}_\tau\nu_\ell$ and $\tau \to \tau_h\nu_\tau$
- $gg/q\bar{q} \to t\bar{t}Z, Z \to \ell^+\ell^-$ with the subsequent decays $t \to b\ell^+\nu$ and $\bar{t} \to \bar{b}q\bar{q}'$ and ℓ^- being reconstructed as a τ_h
- $gg/q\bar{q} \to t\bar{t}$ with the subsequent decays $t \to b\ell^+\nu$ and $\bar{t} \to \bar{b}\tau\bar{\nu}_\tau$, $\tau \to \tau_h\nu_\tau$ and an additional ℓ^+ lepton produced in the semi-leptonic decay of a B hadron from one of the b quarks

Some of the corresponding Feynman diagrams are presented in Fig. 5.3.

[1]For all the processes, the corresponding case where the two charged leptons have a negative charge can be easily deduced.

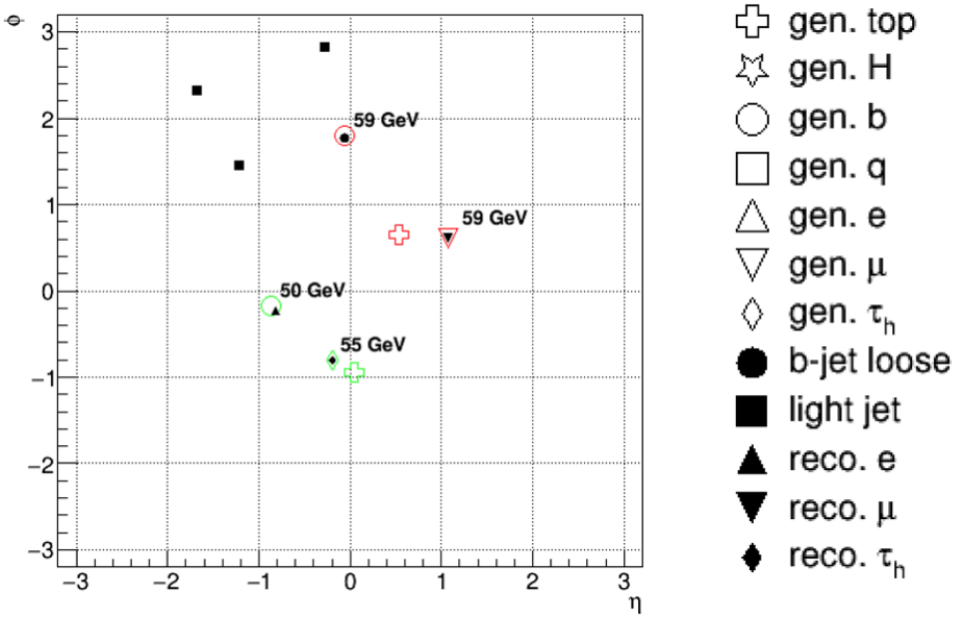

Figure 5.2: Event display of a $t\bar{t}$ event entering the $2\ell ss+1\tau_h$ category, with one of the tight leptons produced in the hadronization of a b quark and vetoing the associated jet. The particles at generator level are represented with empty markers while the reconstructed objects are represented with filled markers. The generator objects with identical color are from the same decay chain. In that event, an electron is produced in the hadronization of the b quark around $(\eta, \phi) = (-1, 0)$ and passes the tight lepton selection. No jet is then considered within $\Delta R = 0.4$ around this tight electron.

The jet multiplicity expected in a leading order $t\bar{t}H$ event is in principle of four jets, including two b-tagged jets. However, due to next-to-leading order QCD contributions and to pile-up, the jet multiplicity can often be higher than the leading order value expected for the MEM computation. To avoid computing the MEM weights for every possible combination of four jets, which would be very CPU consuming, the following assignment algorithm is used to select the MEM inputs. First the two jets with the highest CSV value are chosen to be the ones associated to the b quarks (later called "b-jets"). This has been shown to lead to a correct assignment in 62% of Monte Carlo $t\bar{t}H$ events, with $t \to b\ell\nu, t \to bqq$ and $H \to \tau\tau$ and one jet associated to each quark. The two jets to be associated to the W quarks are then chosen among the remaining jets (later called "light jets") as the pair of jets with the invariant mass closest to the mass of the W boson ("W-tagged jets"). This whole procedure leads to a correct assignment in 35% of the $t\bar{t}H$ with one jet associated to each quark.

In that case, there are two same-sign leptons and two jets associated to b quarks in the MEM inputs. For a $t\bar{t}H$ event, there is therefore an ambiguity to assign one lepton to the

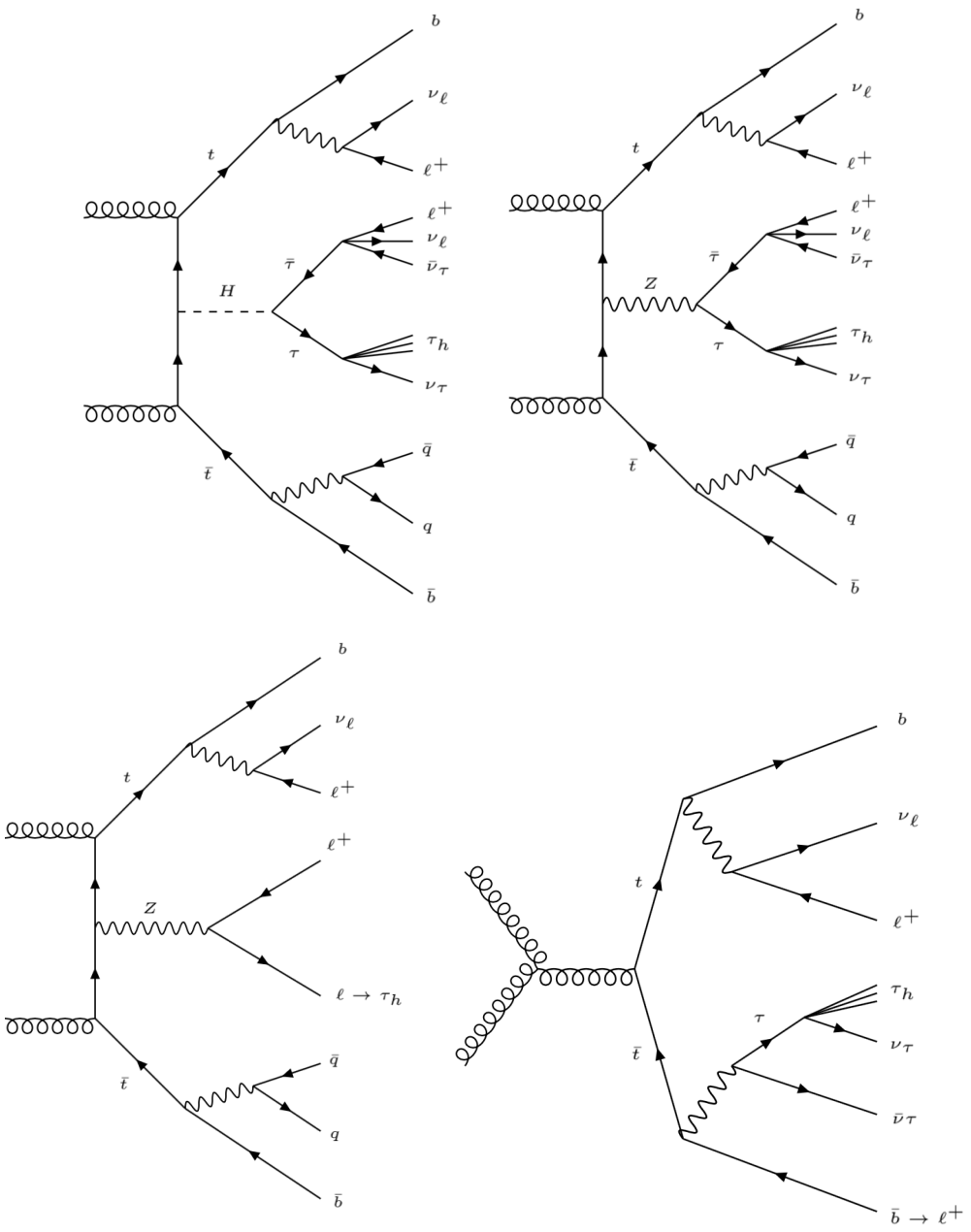

Figure 5.3: Sample of Feynman diagrams of the processes for which a dedicated MEM weight is computed in the implementation presented here. The $q\bar{q}$ initial states are considered as well in the computation of the MEM weights.

leptonic top decay and one to the Higgs decay and another ambiguity to assign one b-jet to the leptonic top decay and one to the hadronic top decay. To take those ambiguities into account, the computation of the MEM weights under each of the signal and background

hypotheses is done for each of the four possible permutations and they are summed to define a total weight. Note that in the case of the $t\bar{t}$ background weight, only one b-jet and no pair of W-tagged jets is expected but there are still four permutations taken into account as one has to decide which of the two b-jets is to be used in the computation. For each permutation, all the compatibility checks based on the reconstructed Higgs and top masses presented in section 4.5 are applied so the computation may be skipped in some cases.

Moreover, sometimes no jet is associated to one of the light quarks, either because the hadronization products gets out of the detector acceptance or because they have been merged with another jet during the reconstruction. To cover those cases, two hypotheses are taken into account in the MEM computation: either all the four jets expected are reconstructed ("no missing jet hypothesis") or one of the two jets from the W boson decay is not reconstructed ("missing jet hypothesis"). In the missing jet hypothesis, only three jets are taken as inputs and the MEM weights computation includes an integration over the direction of the missing jet. In that case, as it is not possible to know a priori which jet comes from a hadronic W decay, the computation of the weights is done for each light jet in the event (each time taking into account the four permutations due to the ambiguity in the lepton and b-jets assignement). Again among all the a priori $4 \times n$(light jets) computations, only those passing all the compatibility checks are taken into account.

To determine which of the two hypotheses is to be used in the computation, the following procedure is applied. If the invariant mass of the two jets associated to the W boson is lower than 60 GeV or larger than 100 GeV, the missing jet hypothesis is used. This means that no pair of jets compatible with a hadronic W boson decay has been found in the event. Otherwise, all the compatibility checks are applied to the four permutations. If one of the permutations passes all those checks for the $t\bar{t}H$ signal hypothesis, the computation is done under the "no missing jet hypothesis" while if all of the permutations fail those checks, the computation is done under the "missing jet hypothesis". This last possibility targets cases where the two W-tagged jets happen to have an invariant mass close to the one of the W boson but in fact do not originate from a W boson decay. In that case, it is unlikely that the event passes the compatibility checks related to the hadronic top and this event enters the missing jet hypothesis. The two exclusive subcategories of events (no missing jet and missing jet) are treated as independent categories for the signal extraction.

5.4.2 Performance

As presented in section 4.1, the MEM weights are combined to build a so-called MEM likelihood ratio defined as

$$LR(k(t\bar{t}Z, Z \to \tau\tau), k(t\bar{t}Z, Z \to \ell\ell), k(t\bar{t})) = \frac{w(t\bar{t}H)}{w(t\bar{t}H) + \sum\limits_{\mathcal{B}} k(\mathcal{B})w(\mathcal{B})} \tag{5.5}$$

The closer to 1 this likelihood ratio is, the more signal-like is the event. Reicever Operating Characteristic (ROC) curves are used for performance assessment: they represent the

background efficiency as a function of the signal efficiency for different selection working points defined from a cut on the likelihood ratio. Monte Carlo samples are used to estimate the corresponding efficiencies. Those efficiencies are computed for events entering the two same-sign leptons and one τ_h category, mentioned in section 5.2.4. Due to a limited statistics in the $t\bar{t}$ dilepton Monte Carlo sample after those selections, the shape templates for that background are obtained by considering events where at least one of the leptons fail the tight MVA selection and reweighting them using their fake rate, as it is done for the data-driven evaluation of that background presented in section 5.3.2.

Note that the ROC curve associated to a given definition of the likelihood ratio is invariant by a uniform scaling of all the $k(\mathcal{B})$ values and thus, in theory, one can fix one of the $k(\mathcal{B})$ to 1. However, for plotting purposes, specific values are chosen here. Specific likelihood ratios are defined as the ones where all the $k(\mathcal{B})$'s but one are equal to zero. They are expected to give the largest discrimination between $t\bar{t}H$ and the corresponding background. A combined likelihood ratio is tested as well where all the $k(\mathcal{B})$'s are non-zero. The values of the $k(\mathcal{B})$'s in the combined likelihood ratio have been checked to give a discrimination very close to the optimal one. The values of the $k(\mathcal{B})$'s used for the different likelihood ratios are presented in Table 5.1.

Category	Likelihood ratio	$k(t\bar{t}Z, Z \to \tau\tau)$	$k(t\bar{t}Z, Z \to \ell\ell)$	$k(t\bar{t})$
No missing jet	Specifc $t\bar{t}Z, Z \to \tau\tau$	0.5	-	-
	Specifc $t\bar{t}Z, Z \to \ell\ell$	-	1.0	-
	Specifc $t\bar{t}$	-	-	10^{-15}
	Combined	0.1	0.2	10^{-18}
Missing jet	Specifc $t\bar{t}Z, Z \to \tau\tau$	0.05	-	-
	Specifc $t\bar{t}Z, Z \to \ell\ell$	-	0.1	-
	Specifc $t\bar{t}$	-	-	10^{-12}
	Combined	0.05	0.5	$5 \cdot 10^{-15}$

Table 5.1: Values of $k(\mathcal{B})$ used for the different versions of the likelihood ratio

The discrimination obtained in each of those categories of events with the different likelihood ratios is presented in the ROC curves in Fig. 5.4 and the distributions of the combined likelihood ratio are presented in Fig 5.5. It can be seen that the specific likelihood ratios have the expected behavior and provide a significant discrimination between the $t\bar{t}H$, $H \to \tau\bar{\tau}$ signal and the background for which they are optimized. However, using a single specific likelihood ratio would have poor performance to discriminate the $t\bar{t}H$, $H \to \tau\bar{\tau}$ signal from the backgrounds for which the likelihood ratio is not optimized. The combined likelihood ratio, although less performant against a given background than the optimal specific likelihood ratio, is still relatively performant against each of the background components and has therefore been chosen as the discriminating variable used for signal extraction in the $2\ell ss + 1\tau_h$ category.

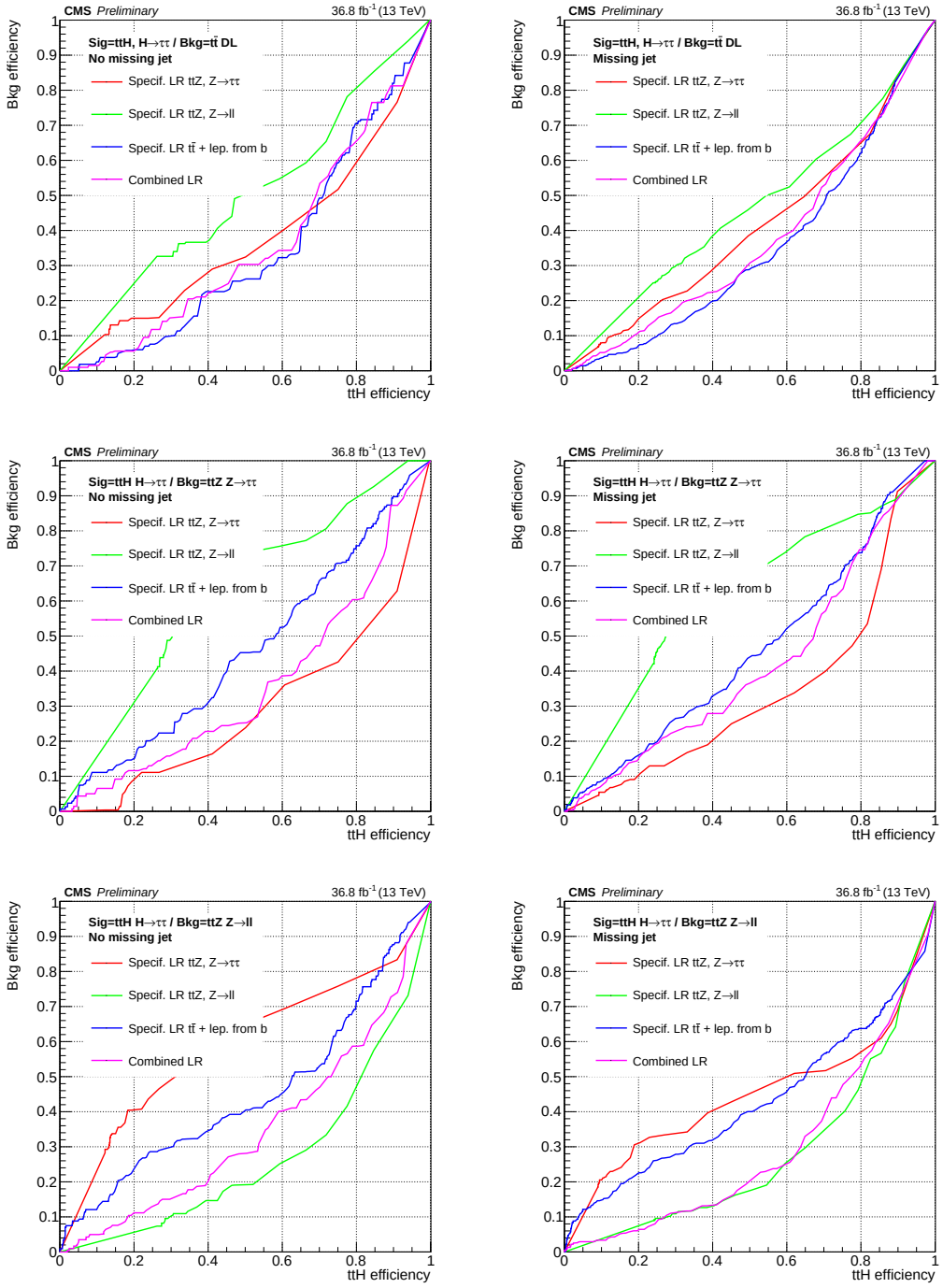

Figure 5.4: ROC curves obtained with the different specific and combined likelihood ratios (LR) obtained in the "no missing jet" subcategory (left) and in the "missing jet" subcategory (right). The discrimination is presented for the signal and backgrounds for which the MEM has been optimized: $t\bar{t}H$, $H \to \tau\bar{\tau}$ signal against $t\bar{t}$ dilepton (top), $t\bar{t}Z, Z \to \tau\tau$ (middle) and $t\bar{t}Z, Z \to \ell\ell$ (bottom) backgrounds. The $t\bar{t}$ dilepton background is evaluated here from Monte Carlo.

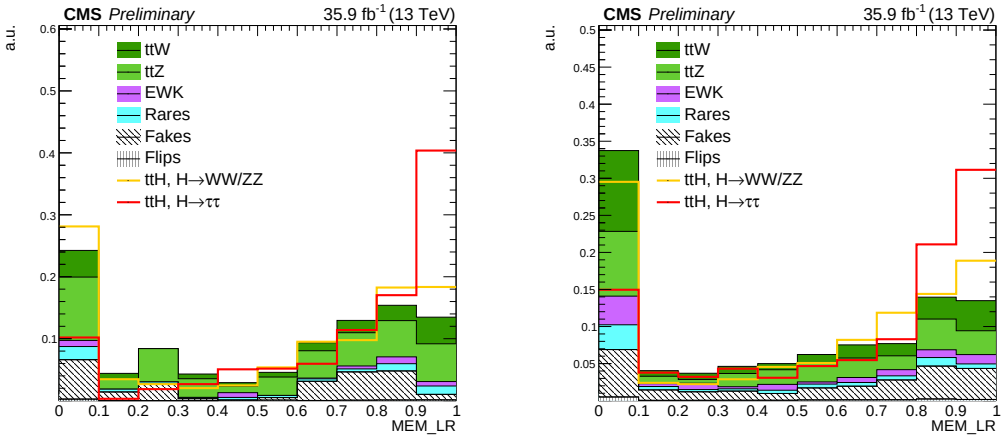

Figure 5.5: Combined likelihood ratio distributions in the "no missing jet" subcategory (left) and in the "missing jet" subcategory (right). The area of the inclusive background and of each signal component is normalized to unity.

It can be seen in Fig. 5.5 that the $t\bar{t}H, H \to WW/ZZ$ component of the the signal has a likelihood ratio distribution closer to the background than the $t\bar{t}H, H \to \tau\tau$ component of the signal. Although the $t\bar{t}H, H \to \tau\tau$ represents the dominant signal contribution, the $t\bar{t}H, H \to WW/ZZ$ component represents still around 40% of the inclusive $t\bar{t}H$ signal and therefore improvements in the treatment of this signal component could probably be used to improve the sensitivity of the analysis. Keeping a MEM based approach, a possibility would be to compute a MEM weight for the $t\bar{t}H, H \to WW$ hypothesis but this would face two main challenges. The first one is that the $t\bar{t}H, H \to WW$ contribution actually originates both from events where a genuine τ_h is produced (either as a Higgs or as a top decay product) and from events where the reconstructed τ_h is a fake produced in a light quark hadronization. Those two components yield similar contributions and the optimal approach would probably be to design a different weight for each one. The second challenge is that the $H \to WW$ decay is a priori less kinematically constrained than the $H \to \tau\tau$ decay, due in particular to the fact that one of the W bosons is offshell. The phase space associated to the Higgs decay in the integration would then require two additional dimensions with respect to the one used for the $H \to \tau\tau$ decay and two different weight computations depending on which W is off-shell. All of this would in principle make the computation more CPU intensive but developments of GPU-based implementations of the MEM could actually make this realistic in the near future. Another possibility would be to use multivariate discriminants trained with MEM variables as inputs using the inclusive $t\bar{t}H$ signal but this requires large statistics samples to be made available for training.

5.5 Statistical interpretation and systematic uncertainties

5.5.1 Likelihood model

The statistical procedure used for the signal extraction in the general case of a binned maximum likelihood fit using the distribution of a discriminating variable is presented in this section. The fit is performed both under the signal + background hypothesis (H_1) and under the background only hypothesis (H_0). The statistical approach presented here corresponds to the modified CL_s method [171]. The results of an observation can be represented as an array of data yields in each bin of the distribution of the variable considered in each category. The expected yield in the bin number i is parametrized as

$$\nu_i(\mu, \vec{\theta}) = \mu s_i(\vec{\theta}) + b_i(\vec{\theta}) \tag{5.6}$$

where s_i and b_i represents respectively the expected amount of signal and background in that bin, depending on some nuisance parameters $\vec{\theta}$. In the case of Higgs searches, the signal yield is fixed to its Standard Model prediction and a signal strength modifier μ is introduced, potentially modifying by the same scale the signal rate from all production modes. The Standard Model prediction corresponds by construction to $\mu = 1$, while the absence of signal corresponds to $\mu = 0$. The likelihood associated to an observation of n_i events in each bin is given by the following product of Poisson distributions

$$\mathcal{L}(data|\mu, \vec{\theta}) = \prod_i \frac{\left(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})\right)^{n_i}}{n_i!} e^{-(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))} \cdot \rho(\vec{\theta}|\hat{\vec{\theta}}) \tag{5.7}$$

with $\rho(\vec{\theta}|\hat{\vec{\theta}})$ the probability density function associated to the nuisance parameters $\vec{\theta}$, given their best estimates $\hat{\vec{\theta}}$.

5.5.2 Systematic uncertainties and nuisance parameters

In the statistical model used here, each source of systematic uncertainty is associated with a nuisance parameter θ_j. Those nuisance parameters are typically constrained by auxiliary measurements that restrict their values within confidence intervals. Those auxiliary measurements can be for instance event counting in sidebands used to constrain the background yields in the signal region or efficiency measurements using a tag-and-probe method.

The nuisance parameters associated to a normalization uncertainty (like the uncertainty on a cross section) are associated to a log-normal distribution

$$\rho(\theta|\hat{\theta}) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{\left(\ln(\theta/\hat{\theta})\right)^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta} \tag{5.8}$$

with $\kappa = 1 + \epsilon$ and ϵ the relative scale of the uncertainty. Such nuisance parameters can affect several processes in different channels or categories, in general with different scales. The errors originating from the same nuisance parameter are then taken as fully correlated or fully anti-correlated.

In the case of an estimation from a sideband, the uncertainty has a statistical origin and is modeled with a Gamma distribution

$$\rho(n|N) = \frac{1}{\alpha}\frac{(n/\alpha)^N}{N!}\exp(-n/\alpha) \tag{5.9}$$

where N is the number of events in the control region, used to estimate the number of events in the signal region n with an extrapolation factor α such that $n = \alpha N$. The extrapolation factor α can itself be associated with a log-normal systematic uncertainty.

Finally some nuisance parameters can affect the shape of the distribution of the variable used for the signal extraction. This is for instance the case of the jet or the τ energy scale. Those are taken into account using a vertical template morphing technique: a shape uncertainty is modeled by defining a family of alternative shape templates governed by a parameter λ. The templates corresponding to $\lambda = \pm 1$ correspond to the templates obtained by shifting the relevant nuisance parameter by ± 1 standard deviation around their nominal value, while the nominal template is recovered for $\lambda = 0$. The family of templates is then defined by a quadratic interpolation between the up and down template for $|\lambda| < 1$, and linear beyond. The pdf constraining the λ parameter is taken to be the normal distribution with mean 0 and $\sigma = 1$.

The systematic uncertainties taken into account are detailed below. Effects arising from the same source cause correlations across different event categories, which are taken into account in the statistical analysis.

Normalization uncertainty

- **Luminosity**
  The uncertainty in the integrated luminosity amounts to 2.6%. This value is obtained from dedicated Van-der-Meer scans and stability of detector response during the data taking.

- **Trigger efficiencies**
  In the $2\ell ss + 1\tau_h$ category the uncertainty in the trigger efficiency amounts to 1% (2%) for events containing either two muons or one electron and one muon (two electrons). This corresponds to the uncertainty associated to the measurement of the trigger efficiency.

- **Identification and isolation efficiency for electrons, muons and τ_h**
  The uncertainty for electrons and muons to pass the loose (tight) lepton selection criteria amount to 2% (3%) while the uncertainty to reconstruct and identify hadronic τ decays amounts to 5%. This corresponds to the uncertainty associated to the measurement of those efficiencies with tag-and-probe methods.

- **Charge mismeasurement**
  The yield of the charge "flip" background in the $2\ell ss+1\tau_h$ category has an associated uncertainty of 30%, based on the statistical uncertainty in the charge misidentification measurement and on the closure test performed on simulated events.

- **Signal and background yields**
  The uncertainties in the SM $t\bar{t}H$ cross section, computed at NLO accuracy, amount to $^{+5.8\%}_{-9.1\%}$ due to missing higher orders and to $\pm 3.6\%$ due to the uncertainties in the PDF and α_s [52]. The corresponding uncertainties with $t\bar{t}W$ ($t\bar{t}Z$) backgrounds are 12% (11%) for the higher order corrections and 4% (3%) for the PDF and α_s. The rate of the WZ+jets background is assigned an uncertainty of 100%, while the rate of other (small) irreducible backgrounds is assigned an uncertainty of 50%.

- **Jet and τ_h energy scales**
  Uncertainties in the energy scale of jets are parametrized as function of jet p_T and η and typically vary between 1% and 4%. The energy scale of τ_h is attributed an uncertainty of 3%. Those uncertainties are taken into account in the $2\ell ss + 1\tau_h$ category by computing the variation obtained in event yields after selections when varying the jet and τ_h energy scales in the Monte Carlo samples. The impact on the shape of the MEM discriminant has been checked to be negligible.

Shape uncertainty

- **b-tagging efficiency**
  Uncertainties in b-tagging efficiencies and mistag rates are applied as function of jet p_T and η and typically amount to 3% and 10%, respectively.

- **Modelling of $t\bar{t}H$ signal and $t\bar{t}W$, $t\bar{t}Z$ backgrounds**
  Uncertainties of theoretical origin on the modelling of the $t\bar{t}H$ signal and of the $t\bar{t}W$ and $t\bar{t}Z$ backgrounds are estimated by varying the renormalization and factorization scales within a factor two relative to their nominal values.

- **"Fake" backgrounds**
  The uncertainties in the measurement of the jet to τ_h misidentification rate and fake rate for non-prompt leptons are treated as p_T- and η-dependent effects. They are determined from the statistical uncertainty in the measurement of the tight-to-loose ratios and from testing the closure of the method in simulated background events [158].

- **Bin-by-bin uncertainty** Additional uncertainties arise from the limited amount of Monte Carlo events available to model the shape of the distribution in the discriminating observable for the $t\bar{t}H$ and for the background processes. Limitations in number of events on the shape templates are accounted for by the approach described in [172, 173]. Nuisance parameters are added to the likelihood function $\mathcal{L}$

which allow the number of events expected in a given bin to vary within statistical uncertainties during the fit.

5.5.3 Measured signal strength, limits setting and significance

Results are presented in terms of the measured value of the signal strength μ_{obs}, i.e. the value of the signal strength μ which maximizes the likelihood function $\mathcal{L}$, in terms of an upper limit on μ at 95% confidence level (CL) and in terms of significance.

The uncertainty in μ_{obs} is obtained by determining lower and upper bounds, μ_{min} and μ_{max}, for which the value of -2 $\ln \mathcal{L}$ exceeds the maximum by one unit, corresponding to a coverage probability of 68%. The differences $\delta_+ = \mu_{max} - \mu_{obs}$ and $\delta_- = \mu_{obs} - \mu_{min}$ represent the uncertainty on μ that arises from the combination of statistical and systematic uncertainties. The nuisance parameters are profiled, that is, their values are chosen such that the likelihood function $\mathcal{L}$ reaches its local maximum, subject to the constraint that μ equals $\mu_{\rm min}$ and $\mu_{\rm max}$, respectively.

Given the likelihood $\mathcal{L}$ defined in Eq. (5.7), a test statistics q_μ is defined using the profile likelihood ratio $\lambda(\mu)$ defined as

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \tag{5.10}$$

with $\hat{\hat{\vec{\theta}}}(\mu)$ the conditional maximum likelihood estimator of $\vec{\theta}$, i.e. the value of $\vec{\theta}$ maximizing $\mathcal{L}$ for a given value of μ, while $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the usual maximum likelihood estimators. In addition, the following requirement is imposed

$$0 \leq \hat{\mu} \leq \mu \tag{5.11}$$

that ensures the signal rate to be positive ($\hat{\mu} \geq 0$) and that one-sided confidence interval are obtained ($\hat{\mu} \leq \mu$) when setting limits on μ. This condition is enforced by defining the test statistics q_μ as

$$q_\mu = \begin{cases} -2\log \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\vec{\theta}}}(0))} & \text{if } \hat{\mu} < 0 \\ -2\log \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{if } \mu < \hat{\mu} \end{cases} \tag{5.12}$$

By construction, q_μ is positive and the larger it is, the less compatible with the $\mu s + b$ hypothesis the observation is.

To establish confidence levels on μ, the probability density function $f(q_\mu|\mu s + b)$ is needed as a function of μ. This can be obtained from the distribution of the test statistics q_μ in a set of pseudo-experiments generated for the corresponding hypothesis using Monte

Carlo techniques. This pdf can then be used to define CL_{s+b} and CL_b as

$$CL_{s+b} = P(q_\mu \leq q_\mu^{obs}|\mu s + b) = \int_{-\infty}^{q_\mu^{obs}} f(q_\mu|\mu s + b) dq_\mu \tag{5.13}$$

$$CL_b = P(q_\mu \leq q_\mu^{obs}|b) = \int_{-\infty}^{q_\mu^{obs}} f(q_\mu|b) dq_\mu \tag{5.14}$$

It is possible to directly use CL_{s+b}, to set limits on μ. In that case, if $CL_{s+b} > 0.95$ for instance, the cross section $\sigma = \mu\sigma_{SM}$ is said to be excluded at a 95% confidence level. However this method suffers from the following inconvenient: in the case where the signal yield is small with respect to the background yield, a negative background fluctuation could lead to a value of CL_{s+b} close to 1, because the background fluctuation is larger than the signal yield. To prevent from excluding values of μ for which there is no actual sensitivity, the CL_s method is preferred. The CL_s is defined as

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{5.15}$$

As $CL_b < 1$, $CL_s > CL_{s+b}$ and less μ hypotheses are rejected, which makes it more conservative and more robust against downwards background fluctuations. Based on the CL_s method, the sensitivity of an experiment can then be assessed based on the median expected exclusion limit on μ in the background only hypothesis, μ_{exp}^{95}, together with the intervals where μ_{obs}^{95} is expected to lie in 68% (1σ) and 95% (2σ) of the cases.

In case an excess is observed, the observed value μ_{obs}^{95} would significantly deviate from the expected value μ_{exp}^{95}. To assess the significance of that excess, the p–value for the observation of q_0^{obs} in the null hypothesis is used, defined using the pdf of the test statistics q_0

$$q_0 = \begin{cases} -2\log\frac{\mathcal{L}(0,\hat{\hat{\theta}}(0))}{\mathcal{L}(\hat{\mu},\hat{\hat{\theta}})} \text{ if } \hat{\mu} \geq 0 \\ 0 \text{ if } \hat{\mu} < 0 \end{cases} \tag{5.16}$$

$$p_0 = P(q_0 \geq q_0^{obs}|b) = \int_{q_0^{obs}}^{+\infty} f(q_0|b) dq_0 \tag{5.17}$$

The significance of a discovery is often reported in the form of a Gaussian probability $Z = \Phi^{-1}(1 - p_0)$ with Φ the inverse cumulative function of the normal distribution. By convention, an evidence is claimed when $Z > 3\sigma$ and we speak of a discovery if $Z > 5\sigma$.

5.6 Results

5.6.1 $2\ell ss + 1\tau_h$ category

The pre-fit distributions of the main event observables (lepton and jet multiplicity, p_T spectra and E_T^{miss} spectrum) are presented in Fig. 5.6 and 5.7. Data and the expected

signal and background already shows a good agreement, within the statistical uncertainty, at the pre-fit level. An event display of an event entering the $2\ell ss{+}1\tau_h$ category is presented in Fig. 5.8. The number of events observed in the $2\ell ss + 1\tau_h$ category is compared to the SM expectation before and after the maximum likelihood fit in Tables 5.2 and 5.3. The expected yield of signal events is computed for the case that the $t\bar{t}H$ signal is produced at the rate predicted by the SM. The signal and background rates are computed for the values of nuisance parameters obtained from the maximum likelihood fit. The event yields observed in data are in agreement with the SM expectation.

The post-fit distributions of the MEM discriminant are shown in Fig. 5.9 for the no-missing-jet and missing-jet subcategories. The measured value of the $t\bar{t}H$ signal strength obtained in the $2\ell ss + 1\tau_h$ category alone is $\mu_{obs} = 0.81^{+0.78}_{-0.65}$ while the observed (expected) 95% CL upper limit is 2.25 (1.37). As a reference, the expected 95% CL upper limit obtained with a simple counting experiment in this category only amounts to 1.74, which shows the benefit from using a discriminant based on the MEM for the signal extraction.

Process	$2\ell ss + 1\tau_h$	
	"no-missing-jet"	"missing-jet"
$t\bar{t}H$, $H \to \tau\tau$	1.89 ± 0.28	3.93 ± 0.49
$t\bar{t}H$, $H \to WW$	1.40 ± 0.19	2.86 ± 0.35
$t\bar{t}H$, $H \to ZZ$	0.09 ± 0.02	0.08 ± 0.01
$t\bar{t}Z$	3.01 ± 0.46	8.18 ± 1.24
$t\bar{t}W$	1.06 ± 0.21	6.95 ± 1.04
Electroweak	0.19 ± 0.14	3.36 ± 2.52
Fake	1.50 ± 0.47	7.05 ± 2.26
Charge flip	0.05 ± 0.10	0.38 ± 0.10
Other	0.45 ± 0.19	2.22 ± 0.95
Total expected background	6.26 ± 0.78	28.14 ± 4.07
Observed data	8	41

Table 5.2: Number of events events selected in the different categories compared to the SM expectation for the $t\bar{t}H$ signal and background processes. The event yields expected for the ttH signal and for the backgrounds are shown as estimated before the maximum likelihood fit. Quoted uncertainties represent the combination of statistical and systematic uncertainties.

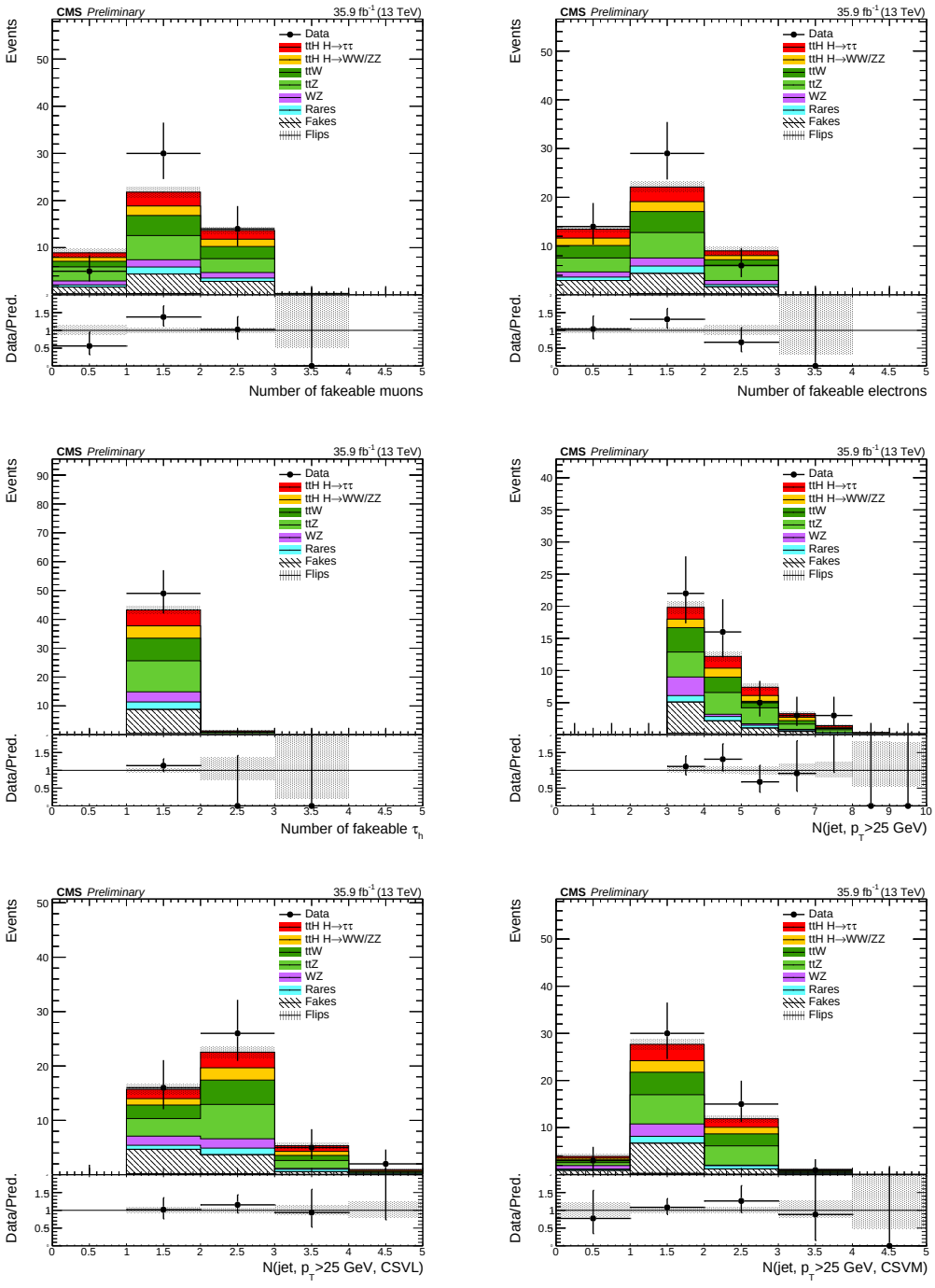

Figure 5.6: Number of muons and electrons passing the fakeable object selections (top), number of reconstructed τ_h passing the fakeable object selections and jet multiplicity (middle) and multiplicity of jets passing the loose and the medium working points of the CSV tagger (bottom) in the $2\ell ss + 1\tau_h$ category. The uncertainty displayed here is statistical only.

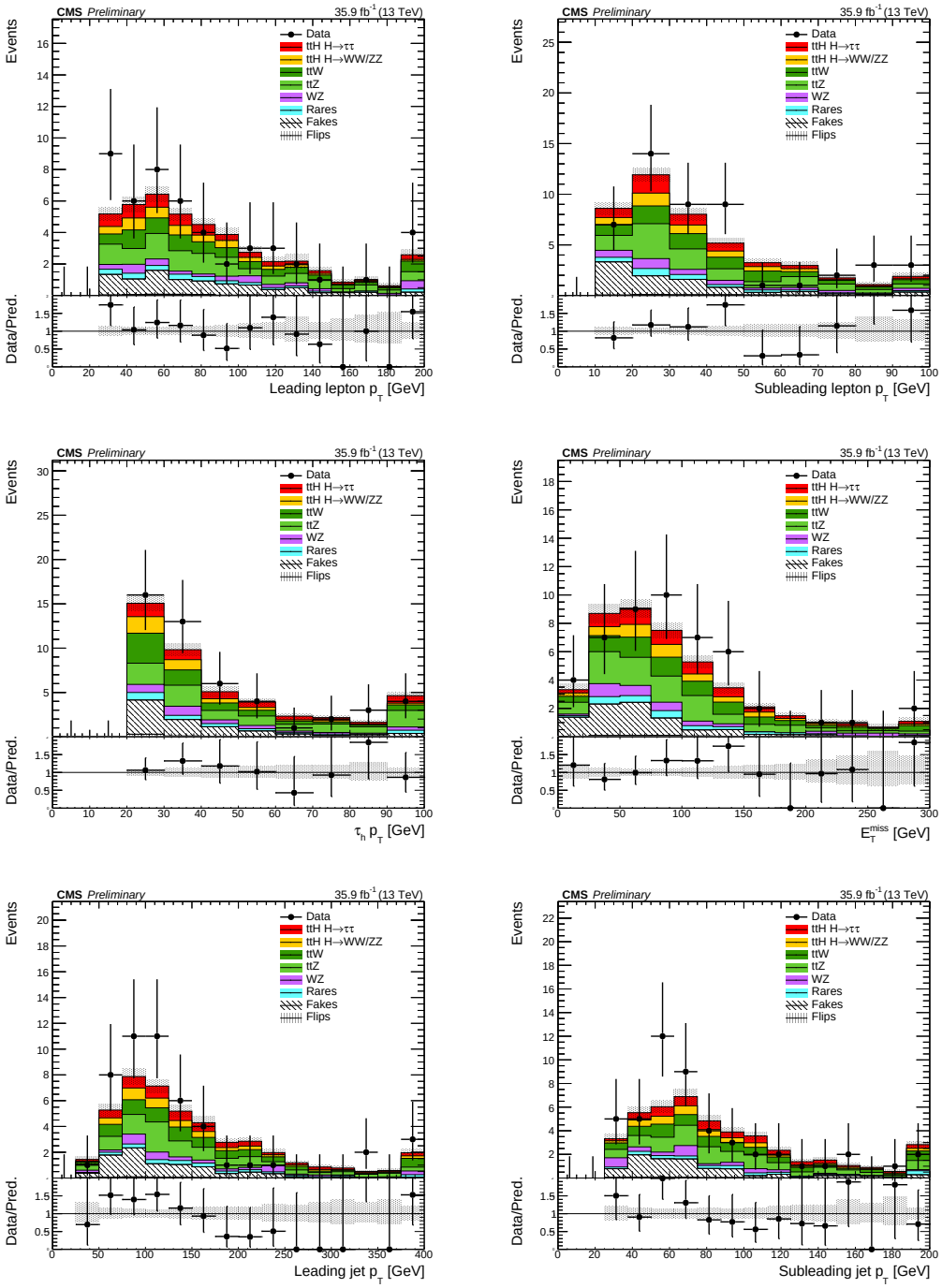

Figure 5.7: Lepton transverse momentum spectra (top), τ_h transverse momentum and missing transverse energy spectra (middle) and leading and subleading jet transverse momentum spectra (bottom) in the $2\ell ss + 1\tau_h$ category. The uncertainty displayed here is statistical only.

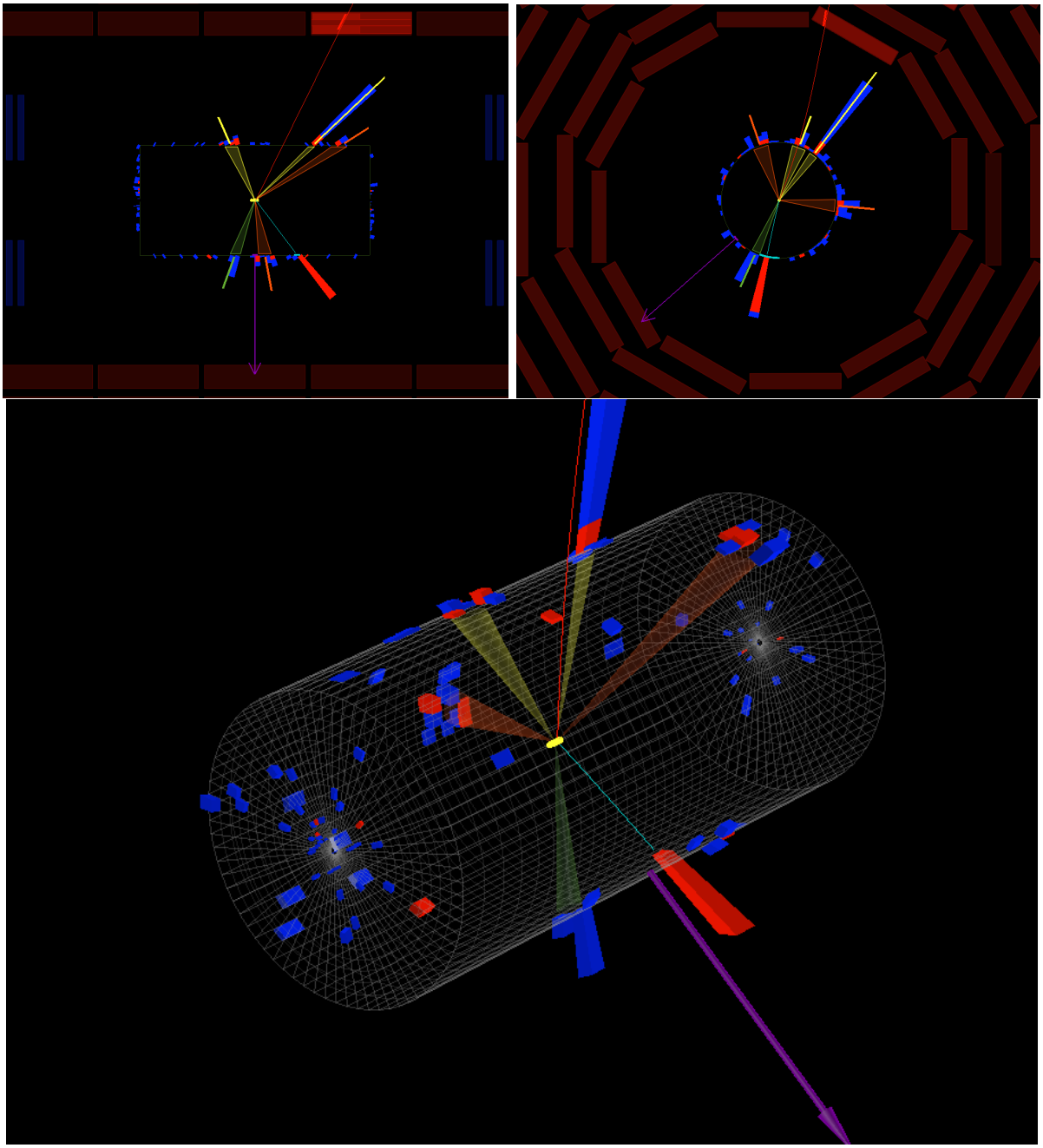

Figure 5.8: Event display of an event entering the $2\ell ss + 1\tau_h$ category in 2016 data. The red (blue) rectangles correspond to ECAL (HCAL) energy deposits. The red (cyan) line represents a reconstructed muon (electron) and the green shaded area a reconstructed τ_h. The orange shaded areas correspond to reconstructed jets identified as b-jets while the yellow shaded areas are for other jets. The purple arrow represents the E_T^{miss}.

Process	$2\ell ss + 1\tau_h$	
	"no-missing-jet"	"missing-jet"
$t\bar{t}H$, $H \to \tau\tau$	1.90 ± 1.28	3.94 ± 2.64
$t\bar{t}H$, $H \to WW$	1.41 ± 1.00	2.87 ± 1.93
$t\bar{t}H$, $H \to ZZ$	0.09 ± 0.06	0.08 ± 0.06
$t\bar{t}Z$	3.07 ± 0.47	8.33 ± 1.31
$t\bar{t}W$	1.10 ± 0.20	7.18 ± 1.06
Electroweak	0.21 ± 0.19	3.73 ± 4.42
Fake	1.66 ± 0.54	7.80 ± 2.69
Charge flip	0.05 ± 0.10	0.39 ± 0.11
Other	0.50 ± 0.26	2.44 ± 1.24
Total expected background	6.59 ± 0.83	29.87 ± 4.73
Observed data	8	41

Table 5.3: Number of events events selected in the different categories compared to the SM expectation for the $t\bar{t}H$ signal and background processes. The event yields expected for the ttH signal and for the backgrounds are shown for the values of nuisance parameters obtained from the maximum likelihood fit. Quoted uncertainties represent the combination of statistical and systematic uncertainties.

5.6.2 Combination with other $t\bar{t}H$ categories with τ_h

Other categories with τ_h are sensitive to the $t\bar{t}H$ production. My results obtained in the $2\ell ss + 1\tau_h$ category have thus been combined with the results obtained in the $1\ell + 2\tau_h$ category and the $3\ell + 1\tau_h$ category in [154]. The $1\ell + 2\tau_h$ targets mostly $t\bar{t}H$ events with the $t\bar{t}$ pair decaying into the lepton+jets final state while the Higgs boson decaying into $H \to \tau_h\tau_h$. It has to deal with a large $t\bar{t}$ background with fake τ_h against which a dedicated BDT has been trained. The $3\ell + 1\tau_h$ is more similar to the $2\ell ss + 1\tau_h$ category, except it is mostly sensitive to dileptonic decays of the $t\bar{t}$ pair. In this category, two BDT's have been trained inclusively in 3ℓ events against the irreducible $t\bar{t}V$ background and the reducible $t\bar{t}$ background. Based on the expected signal-to-background ratio the output of the BDT's is mapped into a single discriminant, referred to as D_{MVA}, which is used for the signal extraction. The observed event yields and the distributions of the discriminant used for signal extraction in those two categories are presented in Table 5.4 and Fig. 5.10. Comparing the event yields for those two categories and the ones for the $2\ell ss+1\tau_h$ category, it is clear that the latter is the most sensitive to the $t\bar{t}H$ signal.

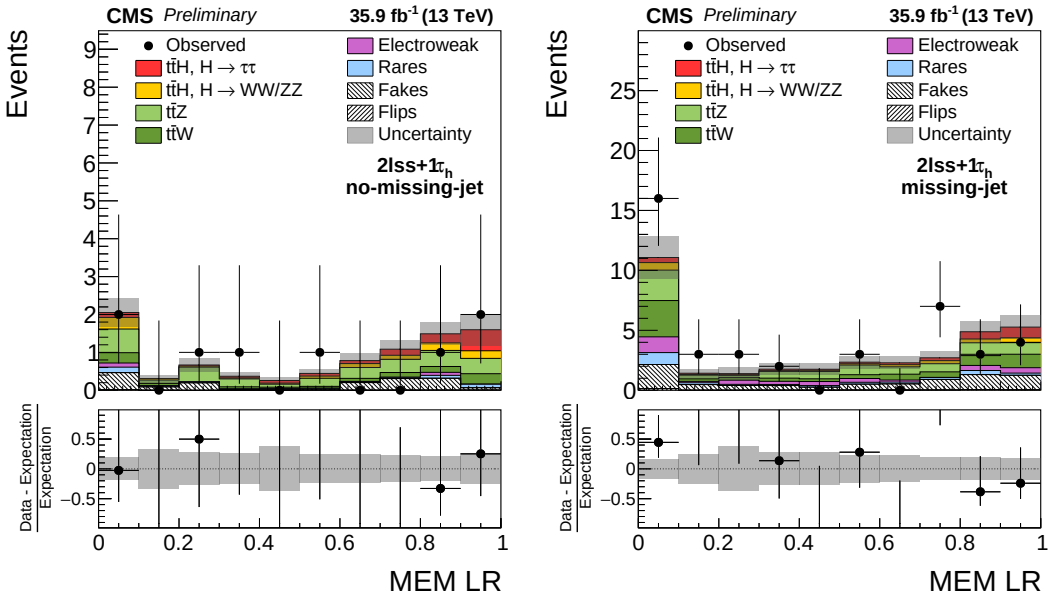

Figure 5.9: Distributions in the discriminating observables used for the signal extraction in the no-missing-jet (left) and (missing-jet) subcategories of the $2\ell ss{+}1\tau_h$ category, compared to the SM expectation for the $t\bar{t}H$ signal and for background processes. The distributions expected for the $t\bar{t}H$ signal and for the backgrounds are shown for the values of the signal strength and nuisance parameters obtained from the maximum likelihood fit. The lowest bin of the MEM discriminant in the missing-jet subcategory collects events for which the kinematics of the reconstructed objects is not compatible with the $t\bar{t}H, H \to \tau\tau$ signal hypothesis.

Process	$1\ell + 2\tau_h$	$3\ell + 1\tau_h$
$t\bar{t}H$, $H \to \tau\tau$	2.84 ± 1.35	1.01 ± 0.65
$t\bar{t}H$, $H \to WW$	0.07 ± 0.04	0.63 ± 0.29
$t\bar{t}H$, $H \to ZZ$	0.02 ± 0.01	0.09 ± 0.04
$t\bar{t}Z$	4.07 ± 0.56	3.78 ± 0.62
$t\bar{t}W$	0.21 ± 0.05	0.24 ± 0.05
Electroweak	1.10 ± 1.05	0.32 ± 0.05
Fake	20.98 ± 3.87	1.07 ± 0.34
Other	0.54 ± 0.23	0.24 ± 0.08
Total expected background	26.91 ± 3.84	5.65 ± 0.85
Observed data	24	7

Table 5.4: Number of events events selected in the $1\ell+2\tau_h$ and $3\ell+1\tau_h$ categories compared to the SM expectation for the $t\bar{t}H$ signal and background processes. The event yield in the $1\ell+2\tau_h$ category is given in the signal-like region MVA>0.2 of the output of the BDT that is used for the signal extraction. Quoted uncertainties represent the combination of statistical and systematic uncertainties.

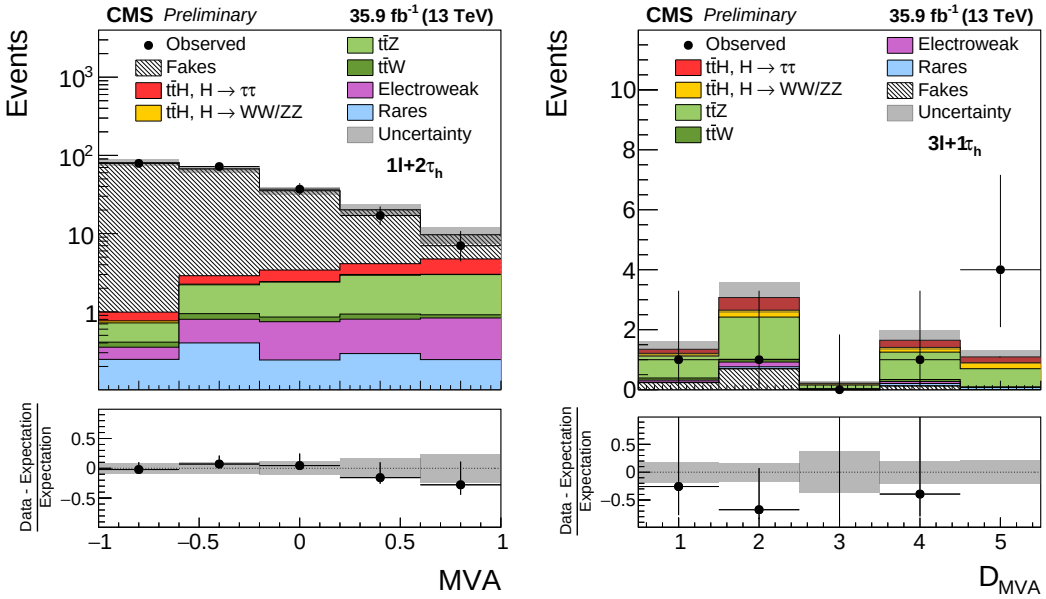

Figure 5.10: Distributions of the discriminating observables used for the signal extraction in the $1\ell + 2\tau_h$ (left) and $3\ell + 1\tau_h$ (right) categories, compared to the SM expectation for the $t\bar{t}H$ signal and for background processes. The distributions expected for the ttH signal and for the backgrounds are shown for the values of the signal strength and nuisance parameters obtained from the maximum likelihood fit [154].

Upper limits on the signal rate, computed at 95% CL, are shown in Fig. 5.11. Numerical values are given in Table 5.5. The limits are computed for each of the categories $2\ell ss+1\tau_h$, $3\ell+1\tau_h$, and $1\ell+2\tau_h$ individually and for their combination. The limits expected in the case $\mu = 1$ are computed using an Asimov dataset [2], while the limits expected in the case $\mu = 0$ are computed for the values of nuisance parameters obtained from a maximum likelihood fit of the background-only hypothesis to the data. The observed limit computed for the combination of all three event categories amounts to 2.0 times the SM $t\bar{t}H$ production rate. The observed limit is compatible with the expected limit in case a $t\bar{t}H$ signal with $\mu = 1$ is present in the data.

Signal rates μ have been computed for each of the categories $2\ell ss + 1\tau_h$, $3\ell + 1\tau_h$, and $1\ell + 2\tau_h$ individually and for their combination, taking into account the correlation of the systematic uncertainties between categories. The results are shown in Fig. 5.12. Numerical values are given in Table 5.6. The signal rates measured in each of the three categories are compatible with each other and with the SM expectation. For the combination of all three event categories, the observed (expected) signal rate is $\mu = 0.72^{+0.62}_{-0.53}$ ($1.00^{+0.67}_{-0.57}$) times the SM $t\bar{t}H$ production rate, with an observed (expected) significance of 1.4σ (1.8σ). The breakdown of the uncertainty is presented in Fig. 5.13 and as can be seen, the statistical uncertainty is already close to the systematics uncertainty.

In the light of these results, the background-only ($\mu = 0$) hypothesis is disfavored, but not yet excluded. A possible combination with other final states, such as $H \to bb$, $H \to \gamma\gamma$ or multilepton final states without τ_h, would help to bring unambiguous evidence of the $t\bar{t}H$ production and give the opportunity to precisely measure the top-Higgs coupling at tree-level.

Category	Observed limit	Expected limit ($\mu = 0$)	Expected limit ($\mu = 1$)
$1\ell + 2\tau_h$	2.6	$3.4^{+1.6}_{-1.0}$	4.4
$2\ell ss + 1\tau_h$	2.4	$1.4^{+0.6}_{-0.4}$	2.4
$3\ell + 1\tau_h$	4.0	$2.7^{+1.3}_{-0.8}$	3.8
Combined	2.0	$1.1^{+0.6}_{-0.3}$	2.2

Table 5.5: 95% CL upper limits on the $t\bar{t}H$ signal rate, in units of the SM $t\bar{t}H$ production rate, obtained in each of the categories $2\ell ss + 1\tau_h$, $3\ell + 1\tau_h$, and $1\ell + 2\tau_h$ individually and for the combination of all three event categories. The observed limit is compared to the limits expected for the background-only hypothesis ($\mu = 0$) and for the case that a $t\bar{t}H$ signal of SM production rate is present in the data ($\mu = 1$). The $\pm 1\sigma$ uncertainty intervals of the limits expected in case of the background-only hypothesis are also given in the table.

[2] corresponding to the sum of the signal and background expectations

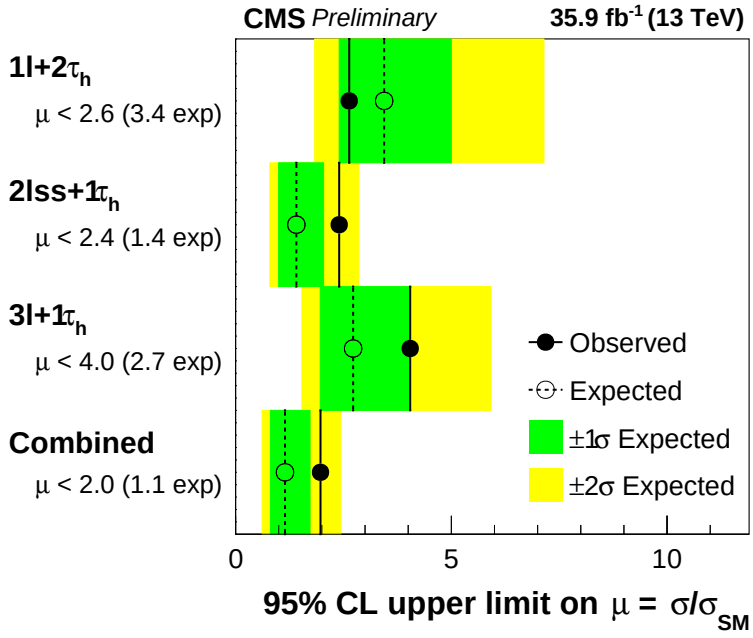

Figure 5.11: 95% CL upper limits on the $t\bar{t}H$ signal rate, obtained in each of the categories $2\ell ss + 1\tau_h$, $3\ell + 1\tau_h$, and $1\ell + 2\tau_h$ individually and for the combination of all three event categories. The expected limits are computed for the background-only ($\mu = 0$) hypothesis.

Category	Measured signal rate $\pm 1\sigma$	Expected signal rate $\pm 1\sigma$
$1\ell + 2\tau_h$	$-1.20^{+1.50}_{-1.47}$	$1.00^{+1.75}_{-1.57}$
$2\ell ss + 1\tau_h$	$0.86^{+0.79}_{-0.66}$	$1.00^{+0.77}_{-0.64}$
$3\ell + 1\tau_h$	$1.22^{+1.34}_{-1.00}$	$1.00^{+1.41}_{-1.06}$
Combined	$0.72^{+0.62}_{-0.53}$	$1.00^{+0.67}_{-0.57}$

Table 5.6: Signal rates μ, in units of the SM $t\bar{t}H$ production rate, measured and expected in each of the categories $2\ell ss+1\tau_h$, $3\ell+1\tau_h$, and $1\ell+2\tau_h$ individually and for the combination of all three categories.

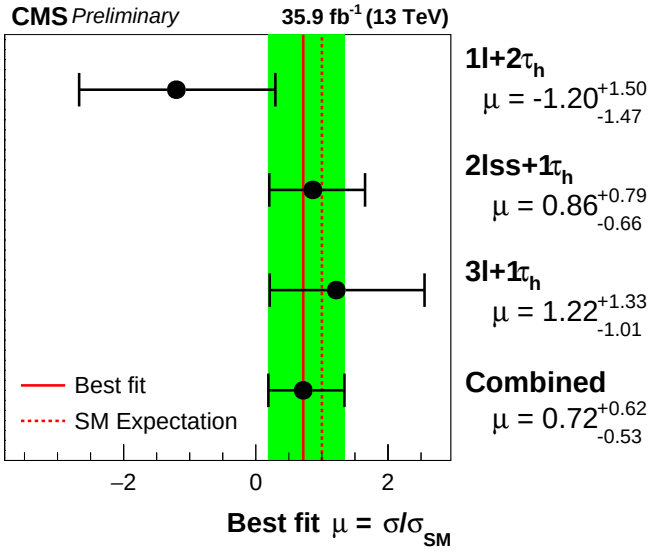

Figure 5.12: Signal rates μ, in units of the SM $t\bar{t}H$ production rate, measured in each of the categories $1\ell + 2\tau_h$, $2\ell ss + 1\tau_h$ and $3\ell + 1\tau_h$ individually and for the combination of all three categories. The expected signal rate is obtained using an Asimov dataset with $\mu = 1$.

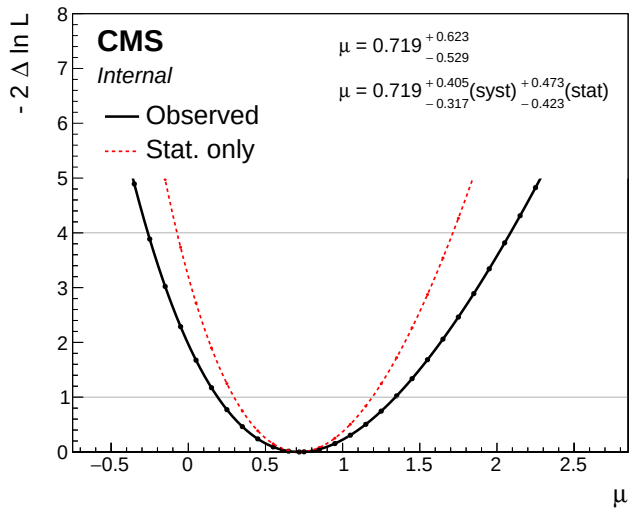

Figure 5.13: Negative log-likelihood as a function of the signal strength μ for the combination of the three categories, including both systematics and statistical uncertainty (black) and only the statistical uncertainty with the systematics assigned to their best-fit values (red dashed).

Conclusion

Success is knowing that your contribution is what helps the collective.

Adrian Grenier

After the observation of the Higgs boson by the CMS and ATLAS collaborations in 2012, the physics of the Higgs boson enters now the era of precision measurements. The measurement of the Higgs boson properties and especially of its couplings to other Standard Model particles and to itself is becoming now one of the major goals of LHC experiments. In that context, the increase in the center-of-mass energy and in the instantaneous luminosity of the collisions achieved for LHC Run 2 offer the possibility to study now new processes which were still statistically limited during Run 1. The search for the $t\bar{t}H$ production in particular is now getting sensitive enough to constrain the coupling of the Higgs boson to the top quarks, the two most massive elementary particles in the Standard Model.

Although the LHC is turning now into a real factory, producing several inverse femtobarns of data every week, this puts unprecedented constraints on the trigger systems of the detectors. To deal with this harsher data-taking conditions, considerable efforts have been put into hardware and firmware developments to be able to maintain trigger thresholds low enough to keep sensitivity to electroweak processes. The Phase 1 trigger upgrade completed by CMS in 2016 represented in that context a major success, including the new Level-1 EG algorithm, which contributed to a large variety of CMS analyses based on 2016 data. All the new available features of this algorithm presented in this thesis have been optimized for the 2016 data-taking and have positively impacted the performance of this trigger. Thanks to the flexibility of the firmware implementation, future improvements of the L1EG algorithm can certainly be considered until the end of LHC Phase 1, together with the design of more and more complex correlation conditions at the Level 1 trigger in order to optimally select events.

Thanks to the large luminosity collected thanks to the new CMS trigger, analysis of rare processes can be considered for Run 2, such as the search for the $t\bar{t}H$ process in final states with a τ lepton. Due to the complexity of this final state, defining a variable with significant discrimination power to be used for signal extraction is not an easy task but this challenge has been met thanks to advanced techniques, such as the Matrix Element Method. The implementation I designed has thus demonstrated very good performance in the search for the $t\bar{t}H, H \to \tau\tau$ process in the $2\ell ss + 1\tau_h$ channel. Combining the results with other $t\bar{t}H, H \to \tau\tau$ sensitive channels allowed to measure a signal strength of $\mu = 0.72^{+0.62}_{-0.53}$, with an observed significance of 1.4σ. Although there is not yet an unambiguous evidence for the $t\bar{t}H$ process in that channel alone, combining this result with those from complementary analyses targeting different final states will definitely contribute to getting closer to a 5σ discovery of the $t\bar{t}H$ process, which should happen by the end of LHC Run 2.

Appendices

Appendix A

Matrix Element Method computations

A.1 Leptonic τ transfer function

A.1.1 Full differential decay width $d^3\Gamma/d^3\vec{\ell}$

The aim of this section is to show the announced result in Eq. (4.35), namely that starting from

$$T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell} \int \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} (\nu.\ell)(\bar{\nu}.\tau)\delta^4(\tau - \ell - \nu - \bar{\nu}) \tag{A.1}$$

one gets

$$T_l(\ell|\tau) \propto \frac{1}{E_\tau E_\ell} [(\ell.\tau)(m_\tau^2 + m_\ell^2 - 2l.\tau) + 2(l.\tau - m_\ell^2)(m_\tau^2 - \ell.\tau)] \tag{A.2}$$

The integral in A.1 can be rewritten as

$$\begin{aligned} \int (\nu.\ell)(\bar{\nu}.\tau)\delta^4(\tau - \ell - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} &= l^\alpha \tau^\beta \int \nu_\alpha \bar{\nu}_\beta \delta^4(\tau - \ell - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} \\ &\equiv l^\alpha \tau^\beta I_{\alpha\beta}(\tau - \ell) \end{aligned} \tag{A.3}$$

By Lorentz covariance, the integral must be of the form

$$I_{\alpha\beta}(Q) = A(Q^2)(g_{\alpha\beta}Q^2 + 2Q_\alpha Q_\beta) + B(Q^2)(g_{\alpha\beta}Q^2 - 2Q_\alpha Q_\beta) \tag{A.4}$$

Since

$$(g^{\alpha\beta}Q^2 + 2Q^\alpha Q^\beta)(g_{\alpha\beta}Q^2 - 2Q_\alpha Q_\beta) = 0 \tag{A.5}$$

$$(g^{\alpha\beta}Q^2 + 2Q^\alpha Q^\beta)(g_{\alpha\beta}Q^2 + 2Q_\alpha Q_\beta) = 12(Q^2)^2 \tag{A.6}$$

$$(g^{\alpha\beta}Q^2 - 2Q^\alpha Q^\beta)(g_{\alpha\beta}Q^2 - 2Q_\alpha Q_\beta) = 4(Q^2)^2 \tag{A.7}$$

we have

$$12(Q^2)^2 A(Q^2) = \int (\nu.\bar{\nu}Q^2 + 2(\nu.Q)(\bar{\nu}.Q))\delta^4(Q - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} \tag{A.8}$$

For $Q = \nu + \bar{\nu}$, we have $Q^2 = 2\nu.\bar{\nu} = 2\nu.Q = 2\bar{\nu}.Q$. Therefore

$$12(Q^2)^2 A(Q^2) = (Q^2)^2 \int \delta^4(Q - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} \tag{A.9}$$

Since the integral is Lorentz-invariant, it can be evaluated in the frame where $\vec{Q} = \vec{0}$. We have then

$$12A(Q^2) = \int \delta(Q^0 - 2|\vec{\nu}|) \frac{|\vec{\nu}|^2 d|\vec{\nu}| d^2\Omega_\nu}{4|\vec{\nu}|^2} = \frac{\pi}{2} \tag{A.10}$$

Moreover

$$4(Q^2)^2 B(Q^2) = \int (\nu.\bar{\nu}Q^2 - 2(\nu.Q)(\bar{\nu}.Q)) \delta^4(Q - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} = 0 \tag{A.11}$$

So in the end,

$$\begin{aligned} \int (\nu.\ell)(\bar{\nu}.\tau) \delta^4(\tau - \ell - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} &\propto l^\alpha \tau^\beta (g_{\alpha\beta}(\tau - \ell)^2 + 2(\tau - \ell)_\alpha (\tau - \ell)_\beta) \\ &\propto (\ell.\tau)(m_\tau^2 + m_\ell^2 - 2\ell.\tau) + 2(\ell.\tau - m_l^2)(m_\tau^2 - \ell.\tau) \end{aligned} \tag{A.12}$$

This integration not only gives the announced result but it also imposes some constraints on τ. To determine explicitly those constraints, let's begin the computation differently.

$$\begin{aligned} &\int (\nu.\ell)(\bar{\nu}.\tau) \delta^4(\tau - \ell - \nu - \bar{\nu}) \frac{d^3\vec{\nu}}{2E_\nu} \frac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} \\ &= \int (\nu.\ell)(\bar{\nu}.\tau) \delta^4(\tau - \ell - \nu - \bar{\nu}) \theta(\nu^0) \delta(\nu^2) d^4\nu \theta(\bar{\nu}^0) \delta(\bar{\nu}^2) d^4\bar{\nu} \\ &= \int (\nu.\ell)(\bar{\nu}.\tau) \theta(\nu^0) \delta(\nu^2) \theta(E_\tau - E_\ell - E_\nu) \delta((\tau - \ell - \nu)^2) d^4\nu \end{aligned} \tag{A.13}$$

with $\theta(x)$ the Heaviside function (equal to 0 for $x < 0$ and to 1 for $x \leq 0$). We have, using $\nu^2 = 0$,

$$(\tau - \ell - \nu)^2 = (\tau - \ell)^2 - 2(\tau - \ell).\nu = (\tau - \ell)^2 - 2|\vec{\nu}|(E_\tau - E_\ell - |\vec{\tau} - \vec{\ell}| \cos\theta_{\nu,\tau-\ell}) \tag{A.14}$$

Therefore

$$\begin{aligned} & \int (\nu.\ell)(\bar{\nu}.\tau) \delta^4(\tau - \ell - \nu - \bar{\nu}) \tfrac{d^3\vec{\nu}}{2E_\nu} \tfrac{d^3\vec{\bar{\nu}}}{2E_{\bar{\nu}}} \\ = & \int (\nu.\ell)(\bar{\nu}.\tau) \theta(\nu^0) \delta((\nu^0)^2 - \vec{\nu}^2) \theta(E_\tau - E_\ell - |\vec{\nu}|) \\ & \delta((\tau - \ell)^2 - 2|\vec{\nu}|(E_\tau - E_\ell - |\vec{\tau} - \vec{\ell}| \cos\theta_{\nu,\tau-\ell})) d^4\nu \\ = & \int (\nu.\ell)(\bar{\nu}.\tau) \theta(|\vec{\nu}|) \theta(E_\tau - E_\ell - |\vec{\nu}|) \\ & \delta\left((\tau - \ell)^2 - 2|\vec{\nu}|(E_\tau - E_\ell - |\vec{\tau} - \vec{\ell}| \cos\theta_{\nu,\tau-\ell})\right) \tfrac{d^3\vec{\nu}}{2|\vec{\nu}|} \end{aligned} \tag{A.15}$$

The computation will not be pushed further (we already know the result of the integral). However, the constraints are to be provided by the θ functions, which only ensure that the energy of the neutrinos are positive. The remaining δ-function imposes that

$$|\vec{\nu}| = \frac{(\tau - \ell)^2}{2(E_\tau - E_\ell - |\vec{\tau} - \vec{\ell}| \cos\theta_{\nu,\tau-\ell})} \tag{A.16}$$

The θ functions impose that $0 \leq |\vec{\nu}| \leq E_\tau - E_\ell$. In the integral, all the possible values of $\cos\theta_{\nu,\tau-\ell}$ between -1 and 1 will contribute. For $\cos\theta_{\nu,\tau-\ell} = -1$, we have

$$|\vec{\nu}| = \frac{(\tau-\ell)^2}{2(E_\tau - E_\ell + |\vec{\tau}-\vec{\ell}|)} \geq 0 \tag{A.17}$$

This imposes that $(\tau-\ell)^2 \geq 0$. This implies that $|\vec{\nu}|$ is maximal for $\cos\theta_{\nu,\tau-\ell} = 1$ and we have then

$$|\vec{\nu}| = \frac{(E_\tau - E_\ell)^2 - |\vec{\tau}-\vec{\ell}|^2}{2(E_\tau - E_\ell - |\vec{\tau}-\vec{\ell}|)} = \frac{1}{2}((E_\tau - E_\ell) + |\vec{\tau}-\vec{\ell}|) \leq E_\tau - E_\ell \tag{A.18}$$

Therefore we must have $|\vec{\tau}-\vec{\ell}| \leq E_\tau - E_\ell$, which is the same as $(\tau-\ell)^2 \geq 0$. Writing $(\tau-\ell)^2 = m_\tau^2 + m_\ell^2 - 2E_\tau E_\ell + 2|\vec{\tau}||\vec{\ell}|\cos\theta_{\tau\ell}$, this implies that

$$\cos\theta_{\tau\ell} \geq \frac{2E_\tau E_\ell - m_\tau^2 - m_\ell^2}{2|\vec{\tau}||\vec{\ell}|} \equiv \alpha \tag{A.19}$$

This constraint has to be taken into account in the integration.

A.1.2 Energy transfer function

For $m_\ell = 0$, the energy transfer function is given by

$$T_{lE}(\ell|\tau) = \int d\Omega_\ell E_\ell^2 T_l(\ell|\tau) \tag{A.20}$$

From Eq. (A.2), we have

$$\begin{aligned} T_l(\ell|\tau) &\propto \left(1 - \frac{|\vec{\tau}|}{E_\tau}\cos\theta_{\tau l}\right)\left(3m_\tau^2 - 4E_\ell E_\tau\left(1 - \frac{|\vec{\tau}|}{E_\tau}\cos\theta_{\tau\ell}\right)\right) \\ &= 3m_\tau^2 - 4E_\ell E_\tau + \left(8E_\ell|\vec{\tau}| - 3m_\tau^2\frac{|\vec{\tau}|}{E_\tau}\right)\cos\theta_{\tau\ell} - 4\frac{E_\ell}{E_\tau}|\vec{\tau}|^2\cos^2\theta_{\tau l} \end{aligned} \tag{A.21}$$

Therefore

$$T_{lE}(\ell|\tau) \propto \int_{\alpha}^{1} d(\cos\theta_{\tau l}) E_l^2 T_l(\ell|\tau)$$

$$= (1-\alpha)E_l^2 \left(3m_\tau^2 - 4E_l E_\tau + \frac{1}{2}\left(8E_l|\vec{\tau}| - 3m_\tau^2\frac{|\vec{\tau}|}{E_\tau}\right)(1+\alpha) - \frac{4}{3}\frac{E_l}{E_\tau}|\vec{\tau}|^2(1+\alpha+\alpha^2)\right) \tag{A.22}$$

Using Eq. (A.19), we get then

$$T_{lE}(\ell|\tau) \propto \frac{m_\tau^2 - 2E_l(E_\tau - |\vec{\tau}|)}{2E_\tau|\vec{\tau}|}\left(\frac{5}{6}m_\tau^2 E_l(E_\tau - |\vec{\tau}|) - \frac{8}{3}E_l^2 E_\tau(E_\tau - |\vec{\tau}|) + \frac{5}{12}m_\tau^4 + \frac{4}{3}m_\tau^2 E_l^2\right) \tag{A.23}$$

In the limit where $m_\tau \ll E_\tau$, we have

$$E_\tau - |\vec{\tau}| = E_\tau - E_\tau\left(1 - \frac{m_\tau^2}{E_\tau^2}\right)^{1/2} = E_\tau - E_\tau\left(1 - \frac{1}{2}\frac{m_\tau^2}{E_\tau^2} - \frac{1}{8}\frac{m_\tau^4}{E_\tau^4} + o\left(\frac{m_\tau^4}{E_\tau^4}\right)\right) \tag{A.24}$$

and

$$T_{lE}(\ell|\tau) \propto \frac{m_\tau^2}{2E_\tau^2}\left(1 - \frac{E_l}{E_\tau}\right)\left(\frac{5}{12}m_\tau^4\frac{E_l}{E_\tau} - \frac{4}{3}m_\tau^2 E_l^2 - \frac{1}{3}m_\tau^4\frac{E_l^2}{E_\tau^2} + \frac{5}{12}m_\tau^4 + \frac{4}{3}m_\tau^2 E_l^2 + o(m_\tau^4)\right)$$

which gives

$$T_{lE}(l|\tau) \propto \frac{1}{E_\tau^2}(1-z)\left(5 + 5z - 4z^2\right) \tag{A.25}$$

with $z = E_l/E_\tau$. Imposing the normalization condition, $\int dE_\ell T_{lE}(l|\tau) = 1$ finally leads to

$$T_{lE}(l|\tau) \propto \frac{1}{3E_\tau}(1-z)\left(5 + 5z - 4z^2\right) \tag{A.26}$$

A.2 Change of variables for dimensionality reduction

A.2.1 Higgs and Z boson decaying into τ leptons

For a Higgs (or a Z) boson decaying into τ leptons, the phase space associated to that decay is

$$d\Phi_H \propto \frac{|\vec{\tau}|^2}{E_\tau}d|\vec{\tau}|\frac{|\vec{\bar{\tau}}|^2}{E_{\bar{\tau}}}d\vec{\bar{\tau}}\,d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \tag{A.27}$$

with the angles $\theta_{\bar{\tau}\pi}$ and $\phi_{\bar{\tau}\pi}$ defined with respect to the visible decay products of the hadronically-decaying τ. To take into account the resonant spectrum in the di-τ invariant mass distribution, it is beneficial to constrain the integration over the variable

$$m_{\tau\bar{\tau}}^2 = (\tau + \bar{\tau})^2 = 2m_\tau^2 + 2(E_\tau E_{\bar{\tau}} - |\vec{\tau}||\vec{\bar{\tau}}|\cos\theta_{\tau\bar{\tau}}) \tag{A.28}$$

To introduce an integration over $m^2_{\tau\bar{\tau}}$ fulfilling the constraint in Eq. (A.28), a simple possibility would be to introduce a change of variable $d\phi_{\bar{\tau}\pi} \to d\cos\theta_{\tau\bar{\tau}}$. This requires to be able to determine $\vec{\bar{\tau}}$ as a function of $(|\vec{\bar{\tau}}|, \cos\theta_{\tau\bar{\tau}}, \cos\theta_{\bar{\tau}\pi})$. For that, we can introduce the Cartesian system of coordinates $(\vec{e}_\pi, \vec{e}_x, \vec{e}_y)$ with $\vec{e}_\pi$ the direction of π, $\vec{e}_x = \frac{1}{\sin\theta_{\tau\pi}}(\vec{e}_\tau - \cos\theta_{\tau\pi}\vec{e}_\pi)$ and $\vec{e}_y = \vec{e}_\pi \wedge \vec{e}_x$ (see Fig.A.1). In that frame, $\vec{\bar{\tau}}$ is given by

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\cos\theta_{\bar{\tau}\pi}\vec{e}_\pi + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\vec{e}_x + \sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\vec{e}_y) \tag{A.29}$$

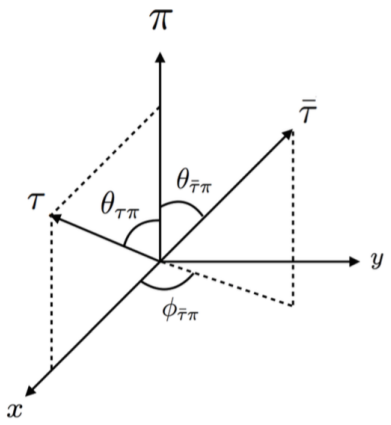

Figure A.1: Cartesian system used to describe a $\tau_l\bar{\tau}_h$ decay

We have then

$$\cos\theta_{\tau\bar{\tau}} = \cos\theta_{\bar{\tau}\pi}\cos\theta_{\tau\pi} + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi} \tag{A.30}$$

Therefore

$$d\phi_{\bar{\tau}\pi} = \frac{1}{\left|\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}}\right|} d(\cos\theta_{\tau\bar{\tau}}) \tag{A.31}$$

with

$$\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} = -\sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi} \tag{A.32}$$

In the system of coordinates (non-Cartesian) $(\vec{e}_\pi, \vec{e}_\tau, \vec{e}_y)$, $\vec{\bar{\tau}}$ can be written as

$$\vec{\bar{\tau}} = |\vec{\bar{\tau}}|(\alpha\vec{e}_\pi + \beta\vec{e}_\tau + \gamma\vec{e}_y) \tag{A.33}$$

With that parametrization

$$\gamma = \vec{e}_{\bar{\tau}}.\vec{e}_y = \sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi} \tag{A.34}$$

One gets then

$$1 = \alpha^2 + \beta^2 + \gamma^2 + 2\alpha\beta\cos\theta_{\tau\pi} \tag{A.35}$$

$$\cos\theta_{\bar{\tau}\pi} = \alpha + \beta\cos\theta_{\tau\pi} \ , \ \cos\theta_{\tau\bar{\tau}} = \alpha\cos\theta_{\tau\pi} + \beta \tag{A.36}$$

Therefore

$$\alpha = \frac{\cos\theta_{\bar{\tau}\pi} - \cos\theta_{\tau\pi}\cos\theta_{\tau\bar{\tau}}}{\sin^2\theta_{\tau\pi}} \ , \ \beta = \frac{\cos\theta_{\tau\bar{\tau}} - \cos\theta_{\tau\pi}\cos\theta_{\bar{\tau}\pi}}{\sin^2\theta_{\tau\pi}} \tag{A.37}$$

$$\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi} \tag{A.38}$$

This system has a solution if and only if $\gamma^2 = 1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi} \geq 0$, with α and β defined by the previous equations in function of $\cos\theta_{\tau\bar{\tau}}$ and $\cos\theta_{\tau\pi}$. However, this is not always guaranteed. Indeed, we are here looking for a possible three-momentum $\vec{\bar{\tau}}$, which fulfills at the same time constraints on $\cos\theta_{\tau\bar{\tau}}$ and $\cos\theta_{\bar{\tau}\pi}$ (which depend on $|\vec{\bar{\tau}}|$). In geometrical terms, this is equivalent to look for the intersections of two cones whose axes are given by $\vec{e}_\tau$ and $\vec{e}_\pi$ and whose opening angles are $\theta_{\tau\bar{\tau}}$ and $\theta_{\bar{\tau}\pi}$. Now, the axes being fixed, it is possible that the opening angles are too small for the cones to intersect. This would be the case for instance if the momentum $|\vec{\bar{\tau}}|$ chosen is too high, because in that case $\bar{\tau}$ should be at the same time almost collinear with τ and π.

If the constraint $\gamma^2 \geq 0$ is fulfilled, it is possible to define

$$\gamma = \pm\sqrt{1 - \alpha^2 - \beta^2 - 2\alpha\beta\cos\theta_{\tau\pi}} \tag{A.39}$$

The so-called "boost" associated to the outgoing particles, defined by $\vec{P}_T = \vec{\tau}_T + \vec{\bar{\tau}}_T + \vec{q}_T + \vec{q'}_T$, is then computed for the two signs. The sign of γ is consequently chosen as the one leading to the smallest boost in magnitude.

The constraint $\gamma^2 \geq 0$ could in principle be translated on the integration range for one of the integration variable but it may not be possible to do it analytically. Since the region $\gamma^2 \geq 0$ has a non-zero measure in the phase-space, the following strategy has been first tested: if the choice of the point in the phase space is such that the inequality is not fulfilled then this point has a null contribution to the integral. However this solution turns out to be relatively inefficient, since when picking random values of $|\vec{\tau}|$ and $|\vec{\bar{\tau}}|$, most of the time the constraints on the angles are such that there is no intersection of the two cones of direction $\vec{e}_\tau$ and $\vec{e}_\pi$ with opening angles $\theta_{\tau\bar{\tau}}$ and $\theta_{\bar{\tau}\pi}$. Only for a very small region in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space this constraint is fulfilled, as shown in Fig. A.2.

Therefore, rather than a change of variables

$$d|\vec{\tau}|d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \rightarrow d|\vec{\tau}|d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}dm^2_{\tau\bar{\tau}} \tag{A.40}$$

the change of variables

$$d|\vec{\tau}|d|\vec{\bar{\tau}}|d\cos\theta_{\bar{\tau}\pi}d\phi_{\bar{\tau}\pi} \rightarrow d|\vec{\tau}|d\cos\theta_{\bar{\tau}\pi}d\cos\theta_{\tau\bar{\tau}}dm^2_{\tau\bar{\tau}} \tag{A.41}$$

is introduced instead. It has the following advantage: since the direction of the τ leptons is close to the direction of their visible decay products, $\cos\theta_{\ell\pi}$ gives a good estimator of $\cos\theta_{\tau\bar{\tau}}$. In that case, it is possible to integrate $\cos\theta_{\tau\bar{\tau}}$ in a small window around $\cos\theta_{\ell\pi}$ such that $\gamma^2 \geq 0$ in most of the integration domain.

In principle, to take into account all of the kinematically allowed configurations, the integration domain should cover at least the whole region of phase-space for which $\gamma^2 \geq 0$.

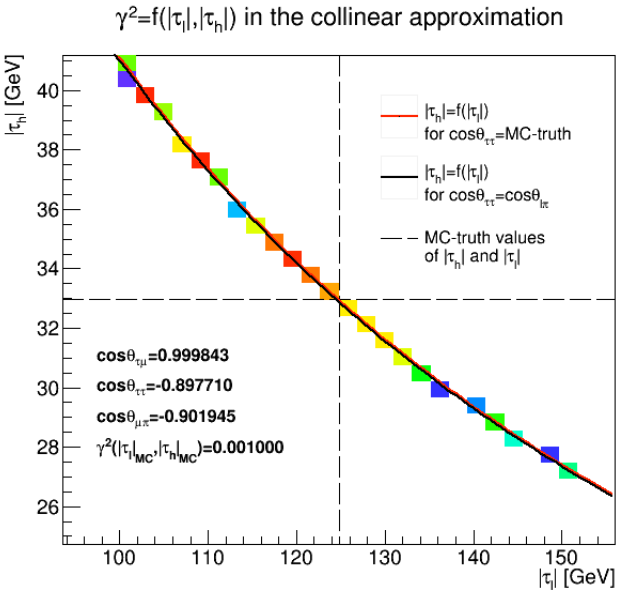

Figure A.2: Region where $\gamma^2 > 0$ (colored region, the color scale corresponding to the value of γ^2) for a single VBF event in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space. The red line corresponds to the function $|\vec{\tau}_h| = f(|\vec{\tau}_l|)$ determined from Eq. (A.28) using $m_{\tau\bar{\tau}}$ =125 GeV and the true value of $\cos\theta_{\tau\bar{\tau}}$, while for the black line $\cos\theta_{\tau\bar{\tau}} = \cos\theta_{\ell\pi}$ was used.

Although not solvable analytically, this constraint could be tested prior to the integration process for different values of the integration variables. We can therefore determine in advance what will be the contributing region in the phase space for the integration and reduce the integration range consequently. The advantages of the new change of variables is that this region is now more extended in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$ space than previously in the $(|\vec{\tau}|, |\vec{\bar{\tau}}|)$ space and we know its approximate location around $\cos\theta_{\tau\bar{\tau}} \approx \cos\theta_{\ell\pi}$, as visible in Fig. A.3.

The change of variables presented in Eq. (A.41) introduces the following Jacobian term

$$d|\vec{\bar{\tau}}|d\phi_{\bar{\tau}\pi} = \frac{1}{\begin{vmatrix} \frac{\partial m^2_{\tau\bar{\tau}}}{\partial|\vec{\bar{\tau}}|} & \frac{\partial \cos\theta_{\tau\bar{\tau}}}{\partial|\vec{\bar{\tau}}|} \\ \frac{\partial m^2_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} & \frac{\partial \cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} \end{vmatrix}} d\cos\theta_{\tau\bar{\tau}} dm^2_{\tau\bar{\tau}} \tag{A.42}$$

with

$$\cos\theta_{\tau\bar{\tau}}\left(|\vec{\bar{\tau}}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi}\right) = \cos\theta_{\bar{\tau}\pi}\cos\theta_{\tau\pi} + \sin\theta_{\bar{\tau}\pi}\cos\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi} \tag{A.43}$$

$$m^2_{\tau\bar{\tau}}\left(|\vec{\bar{\tau}}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi}\right) = 2m^2_\tau + 2\left(E_\tau E_{\bar{\tau}} - |\vec{\tau}||\vec{\bar{\tau}}|\cos\theta_{\tau\bar{\tau}}(|\vec{\bar{\tau}}|, \cos\theta_{\bar{\tau}\pi}, \phi_{\bar{\tau}\pi})\right) \tag{A.44}$$

One gets then

$$\frac{\partial\cos\theta_{\tau\bar{\tau}}}{\partial|\vec{\bar{\tau}}|} = 0$$

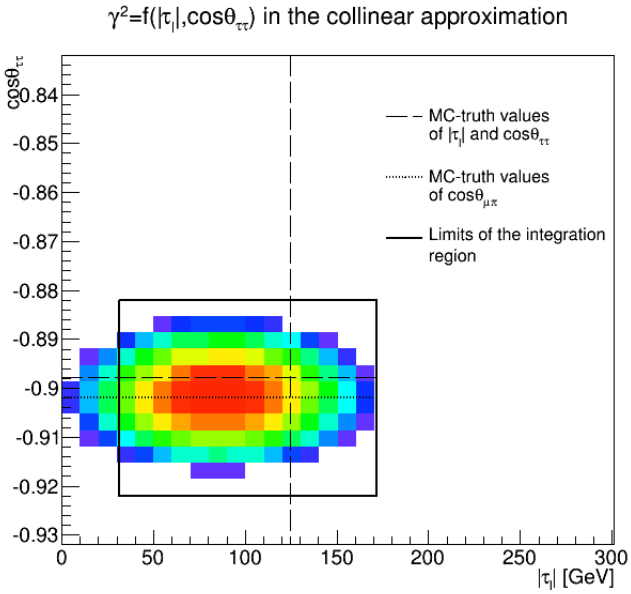

Figure A.3: Region where $\gamma^2 > 0$ (colored region) for a single VBF event in the $(|\vec{\tau}|, \cos\theta_{\tau\bar{\tau}})$ space. The black rectangle corresponds to the limits of the integration region, which is determined prior to the integration. The lower limit for $|\vec{\tau}|$ takes into account t^- from Eq. (4.37), which explains why the rectangle does not cover the whole region $\gamma^2 > 0$.

$$\frac{\partial m^2_{\tau\bar{\tau}}}{\partial|\vec{\tau}|} = 2\left(E_\tau \frac{|\vec{\tau}|}{E_{\bar{\tau}}} - |\vec{\tau}|\cos\theta_{\tau\bar{\tau}}\right) \tag{A.45}$$

$$\frac{\partial \cos\theta_{\tau\bar{\tau}}}{\partial\phi_{\bar{\tau}\pi}} = -\sin\theta_{\bar{\tau}\pi}\sin\phi_{\bar{\tau}\pi}\sin\theta_{\tau\pi} = -\gamma\sin\theta_{\tau\pi}$$

Moreover, defining $M^2 = \frac{1}{2}m^2_{\tau\bar{\tau}} - m^2_\tau$, we get

$$|\vec{\tau}| = \frac{M^2|\vec{\tau}|\cos\theta_{\tau\bar{\tau}} + E_\tau\sqrt{(M^2)^2 - m^2_\tau\left(m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}\right)}}{m^2_\tau + |\vec{\tau}|^2\sin^2\theta_{\tau\bar{\tau}}} \tag{A.46}$$

with

$$|\vec{\tau}| \leq \frac{m_{\tau\bar{\tau}}}{|\sin\theta_{\tau\bar{\tau}}|}\sqrt{\frac{m^2_{\tau\bar{\tau}}}{4m^2_\tau} - 1} \equiv u^+ \tag{A.47}$$

Taking into account the additional constraint from Eq. (4.25) and the on-shell hypothesis used for the di-τ pair, the phase space of the Higgs (or Z) boson is finally reduced to a 2-dimensional phase space

$$d\Phi_H \propto \frac{|\vec{\tau}|^2}{E_\tau}\frac{|\vec{\tau}|^2}{E_{\bar{\tau}}}\frac{1}{\gamma|\sin\theta_{\tau\pi}|\left|E_\tau\frac{|\vec{\tau}|}{E_{\bar{\tau}}} - |\vec{\tau}|\cos\theta_{\tau\bar{\tau}}\right|}d|\vec{\tau}|d\cos\theta_{\tau\bar{\tau}} \tag{A.48}$$

A.2.2 Hadronic top

The phase space associated to a hadronic top can be written as

$$d\Phi_{t_h} \propto dt^2 dW^2 \delta(t-W-b)\delta(W-q-\bar{q})\frac{d^3\vec{t}}{2E_t}\frac{d^3\vec{W}}{2E_W}\frac{d^3\vec{b}}{2E_b}\frac{d^3\vec{q}}{2E_q}\frac{d^3\vec{\bar{q}}}{2E_{\bar{q}}} \tag{A.49}$$

Assuming the top quark and the W bosons are on-shell removes the integration on dt^2 and dW^2. Moreover, the direction of the quarks are assumed to be the ones of the associated jets, which removes the integrations over the quark directions. In practice, this enables to replace $d^3\vec{q}/E_q$ with $|\vec{q}|dE_q$. After integration over $\vec{W}$, one gets

$$\begin{aligned} d\Phi_{t_h} \propto |\vec{b}|E_q E_{\bar{q}}\delta(E_t - E_W(E_q,E_{\bar{q}}) - E_b)\delta(\vec{t} - E_q\vec{e}_q - E_{\bar{q}}\vec{e}_{\bar{q}} - \vec{b}) \\ \delta(E_W(E_q,E_{\bar{q}}) - E_q - E_{\bar{q}})dE_b dE_q dE_{\bar{q}}\frac{d^3\vec{t}}{E_t E_W(E_q,E_{\bar{q}})} \end{aligned} \tag{A.50}$$

with $\vec{W} = E_q\vec{e}_q + E_{\bar{q}}\vec{e}_{\bar{q}}$ and

$$E_W(E_q,E_{\bar{q}}) = \sqrt{M_W^2 + E_q^2 + E_{\bar{q}}^2 + 2E_qE_{\bar{q}}\cos\hat{\theta}_{q\bar{q}}} \tag{A.51}$$

Integrating over $d^3\vec{t}$, one gets

$$\begin{aligned} d\Phi_{t_h} \propto \frac{|\vec{b}|E_qE_{\bar{q}}}{E_t(E_q,E_{\bar{q}},E_b)E_W(E_q,E_{\bar{q}})}\delta(E_t(E_q,E_{\bar{q}},E_b) - E_W(E_q,E_{\bar{q}}) - E_b) \\ \delta(E_W(E_q,E_{\bar{q}}) - E_q - E_{\bar{q}})dE_b dE_q dE_{\bar{q}} \end{aligned} \tag{A.52}$$

with

$$E_t(E_q,E_{\bar{q}},E_b) = \sqrt{M_t^2 + E_b^2 - m_b^2 + E_W^2 - M_W^2 + 2\sqrt{E_b^2 - m_b^2}\vec{W}.\hat{\vec{e}}_b} \tag{A.53}$$

Integrating over E_b, one gets

$$d\Phi_{t_h} \propto \frac{|\vec{b}|E_qE_{\bar{q}}}{E_t(E_q,E_{\bar{q}},|\vec{b}|)E_W(E_q,E_{\bar{q}})}\frac{1}{\left|\frac{\partial E_t}{\partial E_b} - 1\right|}\delta(E_W(E_q,E_{\bar{q}}) - E_q - E_{\bar{q}})dE_qdE_{\bar{q}} \tag{A.54}$$

with

$$\frac{\partial E_t}{\partial E_b} = \frac{1}{E_t}\left(E_b + \frac{E_b}{|\vec{b}|}\vec{W}.\vec{e}_b\right) \tag{A.55}$$

and E_b fixed by

$$E_t(E_q,E_{\bar{q}},E_b) = E_W + E_b \tag{A.56}$$

This leads to

$$E_b^{\pm} = \frac{E_W\Delta M^2 \pm |\vec{W}.\vec{e}_b|\sqrt{(\Delta M^2)^2 - m_b^2(E_W^2 - (\vec{W}.\vec{e}_b)^2)}}{E_W^2 - (\vec{W}.\vec{e}_b)^2} \tag{A.57}$$

with $\Delta M^2 = (M_t^2 - M_W^2 - m_b^2)/2$. To derive Eq. (A.57), some unphysical solutions may have been introduced, depending on the sign chosen. In order to remove them, we can rewrite the original equation (A.56) as

$$\frac{\Delta M^2}{E_W} + \frac{\vec{W}.\vec{e}_b}{E_W}\sqrt{E_b^2 - m_b^2} = E_b \tag{A.58}$$

Thus the following condition is required to be fulfilled by one of the solutions $E_b^{\pm}$,

$$\vec{W}.\vec{e}_b > 0 \text{ and } E_b^{\pm} > \frac{\Delta M^2}{E_W} \quad \text{or} \quad \vec{W}.\vec{e}_b < 0 \text{ and } E_b^{\pm} < \frac{\Delta M^2}{E_W} \tag{A.59}$$

If two admissible (positive) solutions exist, the largest one is chosen.

Then integrating over $E_{\bar{q}}$, one gets

$$d\Phi_{t_h} \propto \frac{|\vec{b}|E_q E_{\bar{q}}}{E_t(E_q, E_{\bar{q}}, |\vec{b}|)E_W(E_q, E_{\bar{q}})} \frac{1}{\left|\frac{\partial E_t}{\partial E_b} - 1\right|} \frac{1}{\left|\frac{\partial E_W}{\partial E_{\bar{q}}} - 1\right|} dE_q \tag{A.60}$$

with

$$\frac{\partial E_W}{\partial E_{\bar{q}}} = \frac{1}{E_W}\left(E_{\bar{q}} + E_q \cos\theta_{q\bar{q}}\right) \tag{A.61}$$

and $E_{\bar{q}}$ fixed by

$$E_{\bar{q}} = \frac{M_W^2}{2E_q(1 - \cos\theta_{q\bar{q}})} \tag{A.62}$$

This finally leads to

$$d\Phi_{t_h} \propto \frac{|\vec{b}|E_{\bar{q}}}{1 - \cos\theta_{q\bar{q}}} \frac{1}{\left|\frac{E_b}{|\vec{b}|}\vec{W}.\vec{e}_b - E_W\right|} \, dE_q \tag{A.63}$$

Once E_q is fixed, $E_{\bar{q}}$ is fixed by Eq. (A.62). The four-vector of the W boson is then fixed by $W = q + \bar{q}$. This determines E_b via Eq. (A.57) and (A.59) and we have $\vec{b} = \sqrt{E_b^2 - m_b^2}\vec{e}_b$. The four-vector of the top is then given by $t = b + q + \bar{q}$.

A.2.3 Leptonic top

The phase space associated to a leptonic top can be written as

$$d\Phi_{t_l} \propto dt^2 dW^2 \delta(t - W - b)\delta(W - \ell - \nu)\frac{d^3\vec{t}}{2E_t}\frac{d^3\vec{W}}{2E_W}\frac{d^3\vec{b}}{2E_b}\frac{d^3\vec{\ell}}{2E_\ell}\frac{d^3\vec{\nu}}{2E_\nu} \tag{A.64}$$

Assuming the top quark and the W bosons are on-shell removes the integration on dt^2 and dW^2. Moreover, the direction of the b quark is assumed to be the one of the associated jet

and the momentum of the lepton is assumed to exactly correspond to its measured value. In that case, after integration over $\vec{W}$, one gets

$$d\Phi_{t_l} \propto \frac{|\vec{b}|dE_b}{E_\ell E_W(E_\nu,\vec{e}_\nu)}\frac{d^3\vec{\nu}}{E_\nu}\frac{d^3\vec{t}}{E_t}\delta(E_t-E_W(E_\nu,\vec{e}_\nu)-E_b)\delta(\vec{t}-\vec{\ell}-E_\nu\vec{e}_\nu-\vec{b})\delta(E_W(E_\nu,\vec{e}_\nu)-E_\ell-E_\nu) \tag{A.65}$$

with

$$\vec{W} = \vec{\ell} + E_\nu\vec{e}_\nu \tag{A.66}$$

$$E_W(E_\nu,\vec{e}_\nu) = \sqrt{M_W^2 + E_\ell^2 + E_\nu^2 + 2E_\ell E_\nu\cos\theta_{l\nu}} \tag{A.67}$$

Integrating over $d^3\vec{t}$, one gets

$$\begin{aligned} d\Phi_{t_l} \propto\ & \frac{|\vec{b}|dE_b}{E_\ell E_W(E_\nu,\vec{e}_\nu)E_t(E_\nu,\vec{e}_\nu,E_b)}\frac{d^3\vec{\nu}}{E_\nu} \\ & \delta(E_t(E_\nu,\vec{e}_\nu,E_b) - E_W(E_\nu,\vec{e}_\nu) - E_b)\delta(E_W(E_\nu,\vec{e}_\nu) - E_\ell - E_\nu) \end{aligned} \tag{A.68}$$

with

$$E_t(E_\nu,\vec{e}_\nu,E_b) = \sqrt{M_t^2 + |\vec{b}|^2 + |\vec{W}|^2 + 2|\vec{b}|\vec{W}.\vec{e}_b} \tag{A.69}$$

Integrating over E_b, we get

$$d\Phi_{t_l} \propto \frac{|\vec{b}|}{E_\ell E_W(E_\nu,\vec{e}_\nu)E_t(E_\nu,\vec{e}_\nu,E_b)}\frac{1}{\left|\frac{\partial E_t}{\partial E_b}-1\right|}\delta(E_W(E_\nu,\vec{e}_\nu) - E_\ell - E_\nu)E_\nu dE_\nu d\Omega_\nu \tag{A.70}$$

with

$$\frac{\partial E_t}{\partial E_b} = \frac{1}{E_t}\left(E_b + \frac{E_b}{|\vec{b}|}\vec{W}.\vec{e}_b\right) \tag{A.71}$$

and E_b fixed, as in the hadronic case, by

$$E_b = \frac{E_W\Delta M^2 \pm |\vec{W}.\vec{e}_b|\sqrt{(\Delta M^2)^2 - m_b^2(E_W^2 - (\vec{W}.\vec{e}_b)^2)}}{E_W^2 - (\vec{W}.\vec{e}_b)^2} \tag{A.72}$$

with $\Delta M^2 = (M_t^2 - M_W^2 - m_b^2)/2$. The following condition is required to be fulfilled by one of the solutions $E_b^\pm$,

$$\vec{W}.\vec{e}_b > 0 \text{ and } E_b^\pm > \frac{\Delta M^2}{E_W} \quad \text{or} \quad \vec{W}.\vec{e}_b < 0 \text{ and } E_b^\pm < \frac{\Delta M^2}{E_W} \tag{A.73}$$

If two admissible (positive) solutions exist, the largest one is chosen.

Integrating over E_ν, one gets

$$d\Phi_{t_l} \propto \frac{|\vec{b}|E_\nu}{E_\ell E_W(E_\nu,\vec{e}_\nu)E_t(E_\nu,\vec{e}_\nu,E_b)}\frac{1}{\left|\frac{\partial E_t}{\partial E_b}-1\right|}\frac{1}{\left|\frac{\partial E_W}{\partial E_\nu}-1\right|}d\Omega_\nu \tag{A.74}$$

with

$$\frac{\partial E_W}{\partial E_\nu} = \frac{1}{E_W}\left(E_\nu + E_\ell \cos\theta_{\ell\nu}\right) \tag{A.75}$$

and E_ν fixed by

$$E_\nu = \frac{M_W^2}{2E_\ell(1-\cos\theta_{\ell\nu})} \tag{A.76}$$

This finally leads to

$$d\Phi_{t_l} \propto \frac{|\vec{b}|E_\nu^2}{E_\ell}\frac{1}{\left|\frac{E_b}{|\vec{b}|}\vec{W}.\hat{\vec{e}}_b - E_W\right|}d\Omega_\nu \tag{A.77}$$

Once $\vec{e}_\nu$ is fixed, E_ν is fixed by Eq. (A.76). The four-vector of the W boson is then fixed by $W = \ell + \nu$. This determines E_b via Eq. (4.76) and (A.73) and $\vec{b} = \sqrt{E_b^2 - m_b^2}\hat{\vec{e}}_b$. The four-vector of the top is then given by $t = b + \ell + \nu$.

Bibliography

[1] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys.Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[2] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys.Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[3] Georges Aad et al. "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". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[4] CMS Collaboration. URL: http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/2.jpg?itok=pPAIaie2.

[5] CDF Collaboration. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. URL: http://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[6] D0 Collaboration. "Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.8 TeV". In: *Phys. Rev. Lett.* 74 (13 Mar. 1995), pp. 2422–2426. DOI: 10.1103/PhysRevLett.74.2422. URL: http://link.aps.org/doi/10.1103/PhysRevLett.74.2422.

[7] DONUT Collaboration. "Observation of tau neutrino interactions". In: *Phys.Lett.* B504 (2001), pp. 218–224. DOI: 10.1016/S0370-2693(01)00307-0. arXiv: hep-ex/0012035 [hep-ex].

[8] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields". In: *Progress of Theoretical Physics* 1.2 (1946), pp. 27–42. DOI: 10.1143/PTP.1.27. eprint: http://ptp.oxfordjournals.org/content/1/2/27.full.pdf+html. URL: http://ptp.oxfordjournals.org/content/1/2/27.short.

[9] Julian Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". In: *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439. URL: http://link.aps.org/doi/10.1103/PhysRev.74.1439.

[10] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769. URL: http://link.aps.org/doi/10.1103/PhysRev.76.769.

[11] F. J. Dyson. "The Radiation Theories of Tomonaga, Schwinger, and Feynman". In: *Phys. Rev.* 75 (3 Feb. 1949), pp. 486–502. DOI: 10.1103/PhysRev.75.486. URL: http://link.aps.org/doi/10.1103/PhysRev.75.486.

[12] Murray Gell-Mann. "A Schematic Model of Baryons and Mesons". In: *Phys.Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[13] G. Zweig. "An SU(3) model for strong interaction symmetry and its breaking. Version 2". In: *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*. Ed. by D.B. Lichtenberg and Simon Peter Rosen. 1964, pp. 22–101. URL: http://inspirehep.net/record/4674/files/cern-th-412.pdf.

[14] E. D. Bloom et al. "High-Energy Inelastic $e-p$ Scattering at 6° and 10°". In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 930–934. DOI: 10.1103/PhysRevLett.23.930. URL: http://link.aps.org/doi/10.1103/PhysRevLett.23.930.

[15] M. Breidenbach et al. "Observed Behavior of Highly Inelastic Electron-Proton Scattering". In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935. URL: http://link.aps.org/doi/10.1103/PhysRevLett.23.935.

[16] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346. URL: http://link.aps.org/doi/10.1103/PhysRevLett.30.1346.

[17] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343. URL: http://link.aps.org/doi/10.1103/PhysRevLett.30.1343.

[18] E. Fermi. "Versuch einer Theorie der β-Strahlen. I". In: *Zeitschrift für Physik* 88.3 (1934), pp. 161–177. ISSN: 0044-3328. DOI: 10.1007/BF01351864. URL: http://dx.doi.org/10.1007/BF01351864.

[19] T. D. Lee and C. N. Yang. "Question of Parity Conservation in Weak Interactions". In: *Phys. Rev.* 104 (1 Oct. 1956), pp. 254–258. DOI: 10.1103/PhysRev.104.254. URL: http://link.aps.org/doi/10.1103/PhysRev.104.254.

[20] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: http://link.aps.org/doi/10.1103/PhysRev.105.1413.

[21] Sheldon L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl.Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[22] Abdus Salam and John C. Ward. "Electromagnetic and weak interactions". In: *Physics Letters* 13 (Nov. 1964), pp. 168–171. DOI: 10.1016/0031-9163(64)90711-5.

[23] Steven Weinberg. "A Model of Leptons". In: *Phys.Rev.Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[24] UA1 Collaboration. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122.1 (1983), pp. 103 –116. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91177-2.

[25] UA2 Collaboration. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 122.5–6 (1983), pp. 476 –485. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91605-2.

[26] K. A. Olive et al. "Review of Particle Physics". In: *Chin. Phys.* C38 (2014), p. 090001. DOI: 10.1088/1674-1137/38/9/090001.

[27] F.J. Hasert et al. "Search for elastic muon-neutrino electron scattering". In: *Physics Letters B* 46.1 (1973), pp. 121 –124. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/0370-2693(73)90494-2.

[28] F.J. Hasert et al. "Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment". In: *Physics Letters B* 46.1 (1973), pp. 138 –140. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/0370-2693(73)90499-1.

[29] UA1 Collaboration. "Experimental Observation of Lepton Pairs of Invariant Mass around 95 GeV/c^2 at the CERN SPS Collider". In: *Phys. Lett. B* 126.5 (1983), pp. 398 –410. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)90188-0.

[30] UA2 Collaboration. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys.Lett.* B129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.

[31] Pascal Paganini. "An introduction to the Standard Model of Particle Physics". In: *Cours de l'Ecole Polytechnique* (2015).

[32] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[33] François Englert and Robert Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[34] Gerald S. Guralnik, Carl R. Hagen, and Thomas W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[35] Luis Alvarez-Gaume and John Ellis. "Eyes on a prize particle". In: *Nature Physics* 7.1 (2011), pp. 2 –3. DOI: 10.1038/nphys1874.

[36] ATLAS and CMS collaborations. "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[37] John M. Campbell, J.W. Huston, and W.J. Stirling. "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics". In: *Rept.Prog.Phys.* 70 (2007), p. 89. DOI: 10.1088/0034-4885/70/1/R02. arXiv: hep-ph/0611148 [hep-ph].

[38] Guido Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl.Phys.* B126 (1977), p. 298. DOI: 10.1016/0550-3213(77)90384-4.

[39] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov.Phys.JETP* 46 (1977), pp. 641–653.

[40] V.N. Gribov and L.N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov.J.Nucl.Phys.* 15 (1972), pp. 438–450.

[41] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[42] C. Patrignani and Particle Data Group. "Review of Particle Physics". In: *Chinese Physics C* 40.10 (2016), p. 100001. URL: http://stacks.iop.org/1674-1137/40/i=10/a=100001.

[43] LHC Higgs cross section working group. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[44] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 1306 (2013), p. 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex].

[45] CMS Collaboration. "Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC". In: *JHEP* 1305 (2013), p. 145. DOI: 10.1007/JHEP05(2013)145. arXiv: 1303.0763 [hep-ex].

[46] ATLAS Collaboration. "Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector". In: *Phys.Lett.* B740 (2015), pp. 222–242. DOI: 10.1016/j.physletb.2014.11.049. arXiv: 1409.3122 [hep-ex].

[47] Georges Aad et al. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Eur. Phys. J.* C75.7 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex].

[48] Georges Aad et al. "Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector". In: *Phys. Lett.* B749 (2015), pp. 519–541. DOI: 10.1016/j.physletb.2015.07.079. arXiv: 1506.05988 [hep-ex].

[49] CMS Collaboration. "Search for the associated production of the Higgs boson with a top-quark pair". In: *JHEP* 1409 (2014), p. 087. DOI: 10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106. arXiv: 1408.1682 [hep-ex].

[50] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment". In: *Eur. Phys. J.* C76.1 (2016), p. 6. DOI: 10.1140/epjc/s10052-015-3769-y. arXiv: 1507.04548 [hep-ex].

[51] Robert Harlander, Michael Kramer, and Markus Schumacher. "Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach". In: (2011). arXiv: 1112.3478 [hep-ph].

[52] S Heinemeyer et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. Tech. rep. arXiv:1307.1347. CERN-2013-004. Geneva, 2013. URL: http://cds.cern.ch/record/1559921.

[53] CMS collaboration. "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks". In: *Phys.Rev.* D89.1 (2014), p. 012003. DOI: 10.1103/PhysRevD.89.012003. arXiv: 1310.3687 [hep-ex].

[54] ATLAS collaboration. "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector". In: *JHEP* 1501 (2015), p. 069. DOI: 10.1007/JHEP01(2015)069. arXiv: 1409.6212 [hep-ex].

[55] ATLAS Collaboration. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". In: *Phys.Lett.* B726 (2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex].

[56] CMS Collaboration. "Measurement of the properties of a Higgs boson in the four-lepton final state". In: *Phys.Rev.* D89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[57] Vardan Khachatryan et al. "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs". In: *Phys. Lett.* B736 (2014), pp. 64–85. DOI: 10.1016/j.physletb.2014.06.077. arXiv: 1405.3455 [hep-ex].

[58] Georges Aad et al. "Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector". In: *Eur. Phys. J.* C75.7 (2015), p. 335. DOI: 10.1140/epjc/s10052-015-3542-2. arXiv: 1503.01060 [hep-ex].

[59] CMS collaboration. "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states". In: *JHEP* 1401 (2014), p. 096. DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].

[60] ATLAS collaboration. "Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector". In: (2014). arXiv: 1412.2641 [hep-ex].

[61] *Measurements of properties of the Higgs boson decaying into four leptons in pp collisions at sqrts = 13 TeV*. Tech. rep. CMS-PAS-HIG-16-041. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256357.

[62] *Measurements of properties of the Higgs boson in the diphoton decay channel with the full 2016 data set*. Tech. rep. CMS-PAS-HIG-16-040. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2264515.

[63] Antonio Pich. "Precision Tau Physics". In: *Prog. Part. Nucl. Phys.* 75 (2014), pp. 41–85. DOI: 10.1016/j.ppnp.2013.11.002. arXiv: 1310.7922 [hep-ph].

[64] "Tau-Id performance with full 2016 dataset using $Z \to \tau_\mu \tau_h$ events". In: (Jan. 2017). URL: http://cds.cern.ch/record/2243476.

[65] CMS collaboration. "Measurement of the Inclusive Z Cross Section via Decays to Tau Pairs in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 08 (2011), p. 117. DOI: 10.1007/JHEP08(2011)117. arXiv: 1104.1617 [hep-ex].

[66] CMS collaboration. *Measurement of the $W \to \tau\nu$ cross–section in pp collisions at $\sqrt{s} = 7$ TeV*. Tech. rep. CMS-PAS-EWK-11-019. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1403095.

[67] CMS collaboration. *Measurement of the $t\bar{t}$ production cross section in the tau+jets channel in pp collisions at $\sqrt{s} = 7$ TeV*. Tech. rep. CMS-PAS-TOP-11-004. Geneva: CERN, 2012. URL: http://cds.cern.ch/record/1446652.

[68] CMS Collaboration. "Search for physics beyond the standard model in events with τ leptons, jets, and large transverse momentum imbalance in pp collisions at $\sqrt{s} =$ 7 TeV". In: *Eur.Phys.J.* C73 (2013), p. 2493. DOI: 10.1140/epjc/s10052-013-2493-8. arXiv: 1301.3792 [hep-ex].

[69] CMS collaboration. "Search for high mass resonances decaying into τ^- lepton pairs in *pp* collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Lett.* B716 (2012), pp. 82–102. DOI: 10.1016/j.physletb.2012.07.062. arXiv: 1206.1725 [hep-ex].

[70] CMS collaboration. *Search for charged Higgs bosons with the $H^+ \to \tau^+\nu_\tau$ decay channel in the fully hadronic final state at $\sqrt{s} = 8$ TeV*. Tech. rep. CMS-PAS-HIG-14-020. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1950346.

[71] B. K. Bullock, Kaoru Hagiwara, and Alan D. Martin. "Tau polarization and its correlations as a probe of new physics". In: *Nucl. Phys.* B395 (1993), pp. 499–533. DOI: 10.1016/0550-3213(93)90045-Q.

[72] Thomas Strebler et al. "Resonance Searches with an Updated Top Tagger". In: *JHEP* 06 (2015), p. 203. DOI: 10.1007/JHEP06(2015)203. arXiv: 1503.05921 [hep-ph].

[73] Gregor Kasieczka and André Schöning. "Search for Resonances Decaying into Top Quark Pairs Using Fully Hadronic Decays in pp Collisions with ATLAS at $\sqrt{s} = 7$ TeV". Presented 02 May 2013. PhD thesis. Heidelberg U., 2013. URL: http://cds.cern.ch/record/1622296.

[74] CDF collaboration. "Search for single top quark production in $\bar{p}p$ collisions at $\sqrt{s} = 1.96$ TeV in the missing transverse energy plus jets topology". In: *Phys. Rev.* D81 (2010), p. 072003. DOI: 10.1103/PhysRevD.81.072003. arXiv: 1001.4577 [hep-ex].

[75] DØ collaboration. "Measurement of the $t\bar{t}$ production cross section using dilepton events in $p\bar{p}$ collisions". In: *Phys. Lett.* B704 (2011), pp. 403–410. DOI: 10.1016/j.physletb.2011.09.046. arXiv: 1105.5384 [hep-ex].

[76] ATLAS collaboration. "Measurement of the top pair production cross section in 8 TeV proton-proton collisions using kinematic information in the lepton+jets final state with ATLAS". In: *Phys. Rev.* D91.11 (2015), p. 112013. DOI: 10.1103/PhysRevD.91.112013. arXiv: 1504.04251 [hep-ex].

[77] CMS collaboration. "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J.* C75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].

[78] CMS collaboration. "Search for top-squark pairs decaying into Higgs or Z bosons in pp collisions at $\sqrt{s}$=8 TeV". In: *Phys. Lett.* B736 (2014), pp. 371–397. DOI: 10.1016/j.physletb.2014.07.053. arXiv: 1405.3886 [hep-ex].

[79] CMS collaboration. "Search for W' $\to$ tb decays in the lepton + jets final state in pp collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 05 (2014), p. 108. DOI: 10.1007/JHEP05(2014)108. arXiv: 1402.2176 [hep-ex].

[80] CMS collaboration. *Search for resonant* $\mathrm{t\bar{t}}$ *production in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*. Tech. rep. arXiv:1506.03062. CMS-B2G-13-008. CERN-PH-EP-2015-126. Comments: Submitted to Phys. Rev. D. Geneva: CERN, June 2015. URL: http://cds.cern.ch/record/2023105.

[81] Andrea Bocci et al. *Tagging b jets with electrons and muons at CMS*. Tech. rep. CMS-NOTE-2006-043. Geneva: CERN, Mar. 2006. URL: http://cds.cern.ch/record/951390.

[82] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[83] Yu Zhang et al. "QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider". In: *Phys. Lett.* B738 (2014), pp. 1–5. DOI: 10.1016/j.physletb.2014.09.022. arXiv: 1407.1110 [hep-ph].

[84] S. Frixione et al. "Weak corrections to Higgs hadroproduction in association with a top-quark pair". In: *JHEP* 09 (2014), p. 065. DOI: 10.1007/JHEP09(2014)065. arXiv: 1407.0823 [hep-ph].

[85] S. Frixione et al. "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons". In: *JHEP* 06 (2015), p. 184. DOI: 10.1007/JHEP06(2015)184. arXiv: 1504.03446 [hep-ph].

[86] L. Reina and S. Dawson. "Next-to-leading order results for t anti-t h production at the Tevatron". In: *Phys. Rev. Lett.* 87 (2001), p. 201804. DOI: 10.1103/PhysRevLett.87.201804. arXiv: hep-ph/0107101 [hep-ph].

[87] S. Dawson et al. "Associated top quark Higgs boson production at the LHC". In: *Phys. Rev.* D67 (2003), p. 071503. DOI: 10.1103/PhysRevD.67.071503. arXiv: hep-ph/0211438 [hep-ph].

[88] S. Dawson et al. "Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections". In: *Phys. Rev.* D68 (2003), p. 034022. DOI: 10.1103/PhysRevD.68.034022. arXiv: hep-ph/0305087 [hep-ph].

[89] W. Beenakker et al. "Higgs radiation off top quarks at the Tevatron and the LHC". In: *Phys. Rev. Lett.* 87 (2001), p. 201805. DOI: 10.1103/PhysRevLett.87.201805. arXiv: hep-ph/0107081 [hep-ph].

[90] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[91] 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". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[92] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[93] T. D. Lee. "A Theory of Spontaneous T Violation". In: *Phys. Rev.* D8 (1973). [,516(1973)], pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226.

[94] Pierre Fayet. "A Gauge Theory of Weak and Electromagnetic Interactions with Spontaneous Parity Breaking". In: *Nucl. Phys.* B78 (1974), pp. 14–28. DOI: 10.1016/0550-3213(74)90113-8.

[95] Ricardo A. Flores and Marc Sher. "Higgs Masses in the Standard, Multi-Higgs and Supersymmetric Models". In: *Annals Phys.* 148 (1983), p. 95. DOI: 10.1016/0003-4916(83)90331-7.

[96] Savas Dimopoulos and Howard Georgi. "Softly Broken Supersymmetry and SU(5)". In: *Nucl. Phys.* B193 (1981), pp. 150–162. DOI: 10.1016/0550-3213(81)90522-8.

[97] Z. Chacko et al. "Natural little hierarchy from a partially goldstone twin Higgs". In: *JHEP* 01 (2006), p. 126. DOI: 10.1088/1126-6708/2006/01/126. arXiv: hep-ph/0510273 [hep-ph].

[98] J. Mrazek et al. "The Other Natural Two Higgs Doublet Model". In: *Nucl. Phys.* B853 (2011), pp. 1–48. DOI: 10.1016/j.nuclphysb.2011.07.008. arXiv: 1105.5403 [hep-ph].

[99] Marcin Badziak and Carlos E. M. Wagner. "Enhancing the Higgs associated production with a top quark pair". In: *JHEP* 05 (2016), p. 123. DOI: 10.1007/JHEP05(2016)123. arXiv: 1602.06198 [hep-ph].

[100] Daniele S. M. Alves et al. "Charged Higgs Signals in $t\bar{t}H$ Searches". In: (2017). arXiv: 1703.06834 [hep-ph].

[101] Philippe Mouche. *Overall view of the LHC. Vue d'ensemble du LHC.* General Photo. June 2014.

[102] ATLAS collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[103] CMS collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[104] ALICE collaboration. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[105] LHCb collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[106] Fabienne Marcastel. *CERN's Accelerator Complex. La chaîne des accélérateurs du CERN.* General Photo. Oct. 2013.

[107] L Rossi and O Brüning. *High Luminosity Large Hadron Collider A description for the European Strategy Preparatory Group.* Tech. rep. CERN-ATS-2012-236. Geneva: CERN, Aug. 2012.

[108] CMS Collaboration. "CMS technical design report, volume II: Physics performance". In: *J.Phys.* G34 (2007), pp. 995–1579. DOI: 10.1088/0954-3899/34/6/S01.

[109] Ivo Nicolas Naranjo Fong. "Tau lepton reconstruction and search for Higgs bosons decaying to tau pairs in the CMS experiment at the LHC". PhD thesis. Ecole Polytechnique, 2014.

[110] V. Halyo, P. LeGresley, and P. Lujan. "Massively Parallel Computing and the Search for Jets and Black Holes at the LHC". In: *Nucl.Instrum.Meth.* A744 (2014), pp. 54–60. DOI: 10.1016/j.nima.2014.01.038. arXiv: 1309.6275 [physics.comp-ph].

[111] Anirban Saha. "Phase 1 upgrade of the CMS pixel detector". In: *Journal of Instrumentation* 12.02 (2017), p. C02033. URL: http://stacks.iop.org/1748-0221/12/i=02/a=C02033.

[112] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Tech. rep. Geneva, 2006. URL: http://cds.cern.ch/record/922757.

[113] CMS ECAL/HCAL Collaborations. "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c". In: *Journal of Physics: Conference Series* 160.1 (2009), p. 012056. URL: http://stacks.iop.org/1742-6596/160/i=1/a=012056.

[114] Min Suk Kim et al. "CMS reconstruction improvement for the muon tracking by the RPC chambers". In: *PoS* RPC2012 (2012), p. 045. DOI: 10.1088/1748-0221/8/03/T03001. arXiv: 1209.2646 [physics.ins-det].

[115] CMS Collaboration. *The CMS trigger system*. Tech. rep. CERN-EP-2016-160. CMS-TRG-12-001-003. arXiv:1609.02366. Comments: Submitted to J. Instrum. Geneva: CERN, Sept. 2016. URL: https://cds.cern.ch/record/2212926.

[116] G. Boudoul. "A level-1 tracking trigger for the CMS upgrade using stacked silicon strip detectors and advanced pattern technologies". In: *JINST* 8 (2013), p. C01024. DOI: 10.1088/1748-0221/8/01/C01024.

[117] Albert M Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". Submitted to *JINST*. 2017.

[118] CMS collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[119] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[120] CMS collaboration. "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[121] CMS collaboration. *Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV*. Tech. rep. CMS-PAS-PFT-10-003. 2010. Geneva: CERN, 2010. URL: http://cds.cern.ch/record/1279347.

[122] Tommaso Dorigo. "Recent Results of the CMS Experiment". In: *EPJ Web Conf.* 70 (2014), p. 00021. DOI: 10.1051/epjconf/20147000021. arXiv: 1210.6809 [hep-ex].

[123] CMS collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[124] Nadir Daci. "Sélection des électrons et recherche du boson de Higgs se désintégrant en paires de leptons tau avec l'expérience CMS au LHC". PhD thesis. Université Paris Sud - Paris XI, 2013. URL: http://tel.archives-ouvertes.fr/tel-00953409.

[125] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN, Jan. 2005. URL: http://cds.cern.ch/record/815410.

[126] Byron P. Roe et al. "Boosted decision trees, an alternative to artificial neural networks". In: *Nucl. Instrum. Meth.* A543.2-3 (2005), pp. 577–584. DOI: 10.1016/j.nima.2004.12.018. arXiv: physics/0408124 [physics].

[127] CMS collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. 2009. Geneva: CERN, Apr. 2009. URL: https://cds.cern.ch/record/1194487.

[128] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k(t) jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[129] CMS collaboration. *Jet Energy Resolution in CMS at $\sqrt{(s)}$=7 TeV*. Tech. rep. CMS-PAS-JME-10-014. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1339945.

[130] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[131] CMS collaboration. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[132] CMS collaboration. *Pileup Jet Identification*. Tech. rep. CMS-PAS-JME-13-005. Geneva: CERN, 2013. URL: http://cds.cern.ch/record/1581583.

[133] *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: http://cds.cern.ch/record/2138504.

[134] "Heavy flavor identification at CMS with deep neural networks". In: (Mar. 2017). URL: http://cds.cern.ch/record/2255736.

[135] Vardan Khachatryan et al. "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS". In: *JINST* 11.01 (2016), P01019. DOI: 10.1088/1748-0221/11/01/P01019. arXiv: 1510.07488 [physics.ins-det].

[136] CMS Collaboration. *Performance of reconstruction and identification of τ leptons in their decays to hadrons and ν_τ in LHC Run 2*. CMS Physics Analysis Summary CMS-PAS-TAU-16-002. 2016. URL: http://cds.cern.ch/record/2196972.

[137] CMS collaboration. "Missing transverse energy performance of the CMS detector". In: *Journal of Instrumentation* 6 (Sept. 2011), p. 9001. DOI: 10.1088/1748-0221/6/09/P09001. arXiv: 1106.5048 [physics.ins-det].

[138] CMS collaboration. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV". In: *JINST* 10.02 (2015), P02006. DOI: 10.1088/1748-0221/10/02/P02006. arXiv: 1411.0511 [physics.ins-det].

[139] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 59. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[140] Alexandre Zabi. "Opérations et optimisations du système de déclenchement calorimétrique de l'expérience CMS au LHC". Habilitation à Diriger des Recherches. Université Paris-Sud XI, 2016.

[141] M. Baber et al. "Development and testing of an upgrade to the CMS level-1 calorimeter trigger". In: *Journal of Instrumentation* 9.01 (2014), p. C01006. URL: http://stacks.iop.org/1748-0221/9/i=01/a=C01006.

[142] A. Svetek et al. "Construction, Testing, Installation, Commissioning and Operation of the CMS Calorimeter Trigger Layer-1 CTP7 Cards". In: *TWEPP Proceedings* (2015).

[143] Imperial College London. 2013. URL: http://www.hep.ph.ic.ac.uk/mp7.

[144] URL: http://www.xilinx.com/products/design-tools/vivado.html.

[145] "Level-1 Electron and Photon trigger performance in 2016 data". In: (June 2016). URL: https://cds.cern.ch/record/2160349.

[146] "CMS L1 Calorimeter Trigger performance in 2016 data". In: (July 2016). URL: https://cds.cern.ch/record/2202966.

[147] Luca Mastrolorenzo. "Search for the Higgs boson decaying into τ lepton pairs with the Matrix Element Method and τ trigger optimization in the CMS experiment at the LHC". Presented 23 Oct 2015. PhD thesis. Ecole Polytechnique, Sept. 2015. URL: https://cds.cern.ch/record/2124426.

[148] D$\varnothing$ collaboration. "A precision measurement of the mass of the top quark". In: *Nature* 429 (2004), pp. 638–642. DOI: 10.1038/nature02589. arXiv: hep-ex/0406031 [hep-ex].

[149] CMS collaboration. "Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method". In: *Eur. Phys. J.* C75.6 (2015), p. 251. DOI: 10.1140/epjc/s10052-015-3454-1. arXiv: 1502.02485 [hep-ex].

[150] E. S. Pearson J. Neyman. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933), pp. 289–337. ISSN: 02643952. URL: http://www.jstor.org/stable/91247.

[151] G. P. Lepage. "A New Algorithm for Adaptive Multidimensional Integration". In: *Journal of Computational Physics* 27 (May 1978), p. 192. DOI: 10.1016/0021-9991(78)90004-9.

[152] 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". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[153] CMS collaboration. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". In: *JHEP* 1405 (2014), p. 104. DOI: 10.1007/JHEP05(2014)104. arXiv: 1401.5041 [hep-ex].

[154] *Search for the associated production of a Higgs boson with a top quark pair in final states with a τ lepton at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-HIG-17-003. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257067.

[155] CMS Collaboration. *Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector*. CMS Physics Analysis Summary CMS-PAS-PFT-10-001. 2010. URL: http://cdsweb.cern.ch/record/1247373.

[156] CMS Collaboration. *Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV*. CMS Physics Analysis Summary CMS-PAS-PFT-10-002. 2010. URL: http://cds.cern.ch/record/1279341.

[157] CMS Collaboration. *Search for $t\bar{t}H$ production in multilepton final states at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-HIG-15-008. 2015. URL: http://cds.cern.ch/record/2141078.

[158] CMS Collaboration. *Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-HIG-17-004. 2017. URL: Inpreparation.

[159] E. Chabanat and N. Estre. "Deterministic annealing for vertex finding at CMS". In: *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP'04, Interlaken, Switzerland, September 27-October 1, 2004*. 2005, p. 287. URL: http://doc.cern.ch/yellowrep/2005/2005-002/p287.pdf.

[160] Wolfgang Waltenberger, Rudolf Frühwirth, and Pascal Vanlaer. "Adaptive vertex fitting". In: *J. Phys. G: Nucl. Part. Phys.* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[161] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The catchment area of jets". In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[162] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B 659 (2008), p. 119. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[163] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[164] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[165] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[166] R. D. Ball et al. "Parton distributions with QED corrections". In: *Nucl. Phys.* B 877 (2013), p. 290. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv: 1308.0598 [hep-ph].

[167] R. D. Ball et al. "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO". In: *Nucl. Phys.* B 855 (2012), p. 153. DOI: 10.1016/j.nuclphysb.2011.09.024. arXiv: 1107.2652 [hep-ph].

[168] R. D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[169] S. Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nucl. Instrum. Meth.* A 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[170] Serguei Chatrchyan et al. "Identification of b quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[171] A. L. Read. "Modified frequentist analysis of search results (the CL_s method)". In: CERN-OPEN-2000-205 (2000). URL: https://cds.cern.ch/record/451614.

[172] R. J. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.

[173] J. S. Conway. "Incorporating nuisance parameters in likelihoods for multisource spectra". In: (2011). arXiv: 1103.0354 [physics.data-an].

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Titre: Étude du couplage du boson de Higgs au quark top au LHC dans l'expérience CMS
Mots clés: *Higgs, quark top, LHC, CMS*
Résumé:

Cette thèse présente une étude de la production associée du boson de Higgs en association avec des quarks tops, suivie d'une désintégration du boson de Higgs en leptons τ, dans des collisions proton-proton (pp) à une énergie dans le centre de masse de $\sqrt{s} = 13$ TeV enregistrées avec le détecteur CMS au Grand collisionneur de hadrons (LHC) du CERN.

Ce travail s'inscrit dans le contexte du Run 2 du LHC, marqué par une augmentation de l'énergie dans le centre de masse de 8 à 13 TeV accompagnée d'une augmentation de la luminosité instantanée des collisions par rapport au Run 1. Pour faire face à ces nouvelles conditions de prise de données, CMS a entrepris une amélioration complète du système de déclenchement de niveau 1 accomplie avant 2016. Ce nouveau système et en particulier le nouvel algorithme électron et photon de niveau 1 ont contribué avec succès à un grand nombre de résultats CMS basés sur les données du Run 2. La mise en service de ce nouveau système ainsi que ces performances mesurées dans les premières données collectées par ce nouveau système de déclenchement sont présentées en détails.

Une technique d'analyse novatrice basée sur la Méthode des Éléments de Matrice optimisée pour la recherche du processus $t\bar{t}H, H \to \tau\tau$ est également présentée, ainsi que plusieurs outils génériques pouvant âtre utilisés dans une large variété d'analyses $H \to \tau\tau$. Les résultats de l'analyse CMS utilisant cette méthode basés sur 35.9 fb^{-1} collectés en 2016 sont présentés. L'hypothèse "bruit de fond seul" est défavorisée mais n'est pas encore exclue par cette analyse à elle seule.

Title: Probing the Higgs coupling to the top quark at the LHC in the CMS experiment
Key words: *Higgs, top quark, LHC, CMS*
Abstract:

This thesis reports a study of the Higgs boson production in association with top quarks and decaying into τ leptons in proton-proton (pp) collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV recorded with the CMS detector at the CERN Large Hadron Collider (LHC).

This work has been carried out in the context of the Run 2 of the LHC, marked by an increase in the center-of-mass energy from 8 to 13 TeV together with an increase in the instantaneous luminosity of the collisions with respect to Run 1. To cope with this new data-taking conditions, CMS had initiated a full upgrade of the Level-1 trigger system achieved by 2016. This new system and in particular the new Level-1 electron and photon algorithm have successfully contributed to a large number of CMS results using Run 2 data. The commissioning of this new system and its performance measured with the first data collected with this new trigger are presented in details.

A novel analysis technique based on the Matrix Element Method optimized for the search of the $t\bar{t}H, H \to \tau\tau$ process is also presented, including several generic tools which can be used in a large variety of $H \to \tau\tau$ analyses. Results of the CMS analysis using this method based on 35.9 fb^{-1} collected in 2016 are presented. The background-only hypothesis is disfavored but not yet excluded by this analysis alone.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France