NNT : 2025IPPAX096

Thèse de doctorat

Optimization of electron identification and measurement of the Higgs boson self-coupling via off-shell decays to four leptons using the CMS detector

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'Ecole polytechnique

École doctorale n°626 de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Physique

Thèse présentée et soutenue à Palaiseau, le 16/07/2026, par

ANDRO PETKOVIC

Composition du Jury :

Emanuele Di Marco INFN Rome	Président
Marco Delmastro LAPP Annecy	Rapporteure
Sandra Kortner Max Planck Institute for Physics, (Munich)	Rapporteure
Mia Tosi Universita' degli Studi di Padova	Examinateur
Kresimir Kumericki University of Zagreb, PMF	Examinateur
Christophe Ochando LLR (Palaiseau)	Directeur de thèse
Toni Sculac University of Split, PMF	Directeur de thèse

Abstract

The work presented in this thesis can be divided into two main parts. The first part concerns the determination of electron efficiencies used in the $H \to ZZ \to 4\ell$ analysis. This includes the training of the electron identification algorithm for the 2022 data-taking period and the derivation of the electron identification efficiency scale factors for the same year. In addition, a new tool for the measurement of electron reconstruction efficiency in the low-p_T region ($5 < p_T < 10$ GeV) has been developed using $J/\Psi \to e^-e^+$ decays. The development of this method is motivated by the significant contribution of low-p_T electrons to the systematic uncertainties in the four-lepton final state, as well as in multilepton analyses in general.

The central physics result of this thesis is the measurement of the Higgs boson self-coupling modifier (κ_λ) in the $H \to ZZ \to 4\ell$ decay channel in the off-shell region. The analysis uses proton–proton collision data collected by the CMS detector during Run 2 of the LHC (2016–2018) at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The measurement is performed within the framework of the Standard Model effective field theory, where κ_λ is probed through insertions of a dimension-six operator that affects both the event yields and the shapes of observables.

Unlike direct determinations based on Higgs boson pair production, the constraint on κ_λ is obtained from loop-level corrections in single-Higgs processes. For the first time, the off-shell region of the $H \to ZZ \to 4\ell$ phase space is used to perform such a measurement. In this region, the invariant-mass distribution of the four-lepton system is sensitive to the presence of the dimension-six operator, leading to a characteristic modification of the off-shell $m_{4\ell}$ lineshape. Limits on κ_λ are obtained from a three-dimensional profiled fit to $m_{4\ell}$ and to background and interference discriminants, with the signal strength modifier treated as a free parameter. The observed (expected) 68% CL interval on κ_λ is 2^{+17}_{-15} (1^{+21}_{-18}), with a corresponding 95% CL interval of [−21, 27] ([−22, 26]), consistent with the SM expectation.

Résumé étendu

La compréhension des constituants fondamentaux de la matière et des lois qui régissent leurs interactions constitue l'une des questions centrales de la physique depuis l'Antiquité. Les premières idées sur la structure de la matière, issues de réflexions philosophiques sur des particules indivisibles, ont progressivement évolué vers la vision moderne selon laquelle le comportement de la nature aux plus petites échelles est décrit par la théorie quantique des champs. Au cours du développement de la physique, il est apparu à plusieurs reprises que des structures considérées comme élémentaires possédaient en réalité une sous-structure, ce qui a stimulé le développement de nouveaux modèles théoriques et la construction d'instruments expérimentaux toujours plus puissants.

La description actuelle des particules élémentaires et de leurs interactions est donnée par le Modèle standard de la physique des particules. Cette théorie repose sur le principe de l'invariance de jauge locale et décrit, dans un cadre unifié, les interactions électromagnétique, faible et forte. Au cours des dernières décennies, les prédictions du Modèle standard ont été confirmées par un grand nombre de mesures de précision, ce qui en fait l'une des théories les plus abouties de l'histoire de la physique. Un élément essentiel de cette théorie est le mécanisme de brisure de la symétrie électrofaible, qui explique comment les particules élémentaires acquièrent une masse sans violer l'invariance de jauge. La confirmation expérimentale de ce mécanisme a été obtenue en 2012 avec la découverte du boson de Higgs au Grand collisionneur de hadrons (Large Hadron Collider, LHC), validant ainsi le dernier élément manquant du contenu particulaire prédit par le Modèle standard.

Malgré son succès remarquable, le Modèle standard n'est pas considéré comme une théorie complète de la nature. Il n'inclut pas la gravitation, n'explique pas la nature de la matière noire et ne répond pas à plusieurs questions ouvertes liées à la structure du secteur scalaire et à la stabilité du potentiel de Higgs. C'est pourquoi l'un des principaux objectifs de la physique des particules moderne est la recherche d'écarts par rapport aux prédictions du Modèle standard susceptibles d'indiquer l'existence d'une nouvelle physique. De tels écarts peuvent apparaître par la production directe de nouvelles particules à haute énergie, mais aussi sous la forme de modifications faibles de processus connus, révélées par des mesures de haute précision.

Le secteur de Higgs joue un rôle particulièrement important dans ces recherches. Bien que la masse du boson de Higgs et plusieurs de ses couplages aient déjà été mesurés, la forme du potentiel scalaire n'a pas encore été entièrement testée expérimentalement. Dans le Modèle standard, ce potentiel détermine l'intensité de l'auto-couplage du boson de Higgs, qui peut être étudié dans des processus comportant un ou plusieurs bosons de Higgs dans l'état final. La mesure directe de l'auto-couplage est très difficile expérimentalement, car elle repose sur des processus rares possédant de faibles sections efficaces de production. Cependant, une sensibilité à cet auto-couplage peut également être obtenue de manière indirecte, à travers des mesures précises de la production et de la désintégration d'un seul boson de Higgs, où les corrections quantiques dépendent de la structure du secteur scalaire. De tels effets peuvent être décrits dans le cadre de la théorie effective du Modèle standard (SMEFT), dans laquelle les écarts possibles au Modèle standard sont paramétrés à l'aide d'opérateurs de dimension supérieure.

Les plus hautes énergies de collision actuellement accessibles sont atteintes au LHC, ce qui permet une étude détaillée des propriétés du boson de Higgs et d'autres processus à des échelles d'énergie jusque-là inaccessibles. Après la découverte du boson de Higgs lors de la première période de fonctionnement du LHC, les données collectées pendant les deuxième et troisième périodes permettent des tests toujours plus stricts du Modèle standard. Dans ce domaine, la sensibilité des mesures ne dépend pas seulement de la quantité de données, mais aussi d'une compréhension très fine du fonctionnement du détecteur, de la reconstruction des particules et de la détermination des efficacités d'identification. Les états finaux contenant des électrons et des muons sont particulièrement importants car ils permettent une reconstruction précise de la cinématique des événements. Parmi eux, le canal de désintégration $H \to ZZ \to 4\ell$ se distingue comme l'un des plus précis pour l'étude des propriétés du boson de Higgs.

Les recherches présentées dans ce travail ont été menées dans le cadre de l'expérience CMS au LHC et portent sur des mesures liées au boson de Higgs dans l'état final à quatre leptons. Un accent particulier est mis sur la détermination des efficacités de reconstruction et d'identification des électrons, qui sont essentielles pour les mesures de précision dans ce canal, ainsi que sur l'étude de la sensibilité des processus à un seul boson de Higgs aux modifications de l'auto-couplage du champ de Higgs dans le cadre du formalisme SMEFT. Ces recherches s'inscrivent dans un effort plus large visant à tester le Modèle standard avec la plus grande précision possible et à rechercher d'éventuels signes de physique au-delà de ce cadre théorique.

Introduction théorique

Le Modèle standard de la physique des particules est une théorie quantique des champs fondée sur la symétrie de jauge locale $SU(3)_C \times SU(2)_L \times U(1)_Y$, qui décrit les interactions forte, faible et électromagnétique entre les particules élémentaires. Le contenu fermionique de la théorie se compose de trois générations de quarks et de leptons, tandis que les interactions sont médiées par des bosons de jauge associés aux générateurs du groupe de symétrie. La dynamique du système est déterminée par un lagrangien comprenant les termes cinétiques des champs fermioniques et de jauge, ainsi que les interactions définies par la dérivée covariante.

Les termes de masse des bosons faibles et des fermions ne peuvent pas être écrits d'une manière respectant la symétrie de jauge. Ce problème est résolu en introduisant un doublet scalaire complexe dont le potentiel conduit à une brisure spontanée de la symétrie électrofaible. Après le choix de la valeur moyenne dans le vide du champ scalaire, les bosons de jauge faibles acquièrent une masse, tandis que le photon reste sans masse. Les fermions acquièrent leur masse par des interactions de Yukawa avec le champ scalaire. Le spectre de la théorie contient alors une excitation scalaire physique, le boson de Higgs, dont les couplages aux fermions et aux bosons de jauge sont déterminés par leurs masses. Le potentiel du champ scalaire contient également des termes décrivant l'auto-couplage du champ de Higgs, qui fixe la forme du secteur scalaire de la théorie.

Bien que le Modèle standard décrive avec succès un très grand nombre de résultats expérimentaux, il peut être considéré comme une théorie effective valable jusqu'à une certaine échelle d'énergie. Dans le cas général, les effets d'une nouvelle physique à plus haute énergie peuvent être décrits en étendant le lagrangien par des opérateurs de dimension supérieure construits à partir des champs connus. Une telle approche conduit à une théorie effective des champs, dans laquelle le lagrangien s'écrit comme un développement en puissances inverses d'une échelle caractéristique de nouvelle physique Λ,

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{C_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \cdots , \tag{1}$$

où $\mathcal{O}_i^{(d)}$ sont des opérateurs de dimension d, et $C_i^{(d)}$ les coefficients de Wilson correspondants. Les opérateurs doivent respecter toutes les symétries du Modèle standard, tandis que les contributions des dimensions supérieures sont supprimées par l'échelle Λ.

Dans le cas où il n'existe pas de degrés de liberté légers supplémentaires, la description la plus générale des écarts possibles est obtenue dans le cadre de la théorie effective du Modèle standard (SMEFT). Dans ce formalisme, seuls les champs du Modèle standard sont utilisés, et le lagrangien est étendu par tous les opérateurs de dimension supérieure à quatre invariants sous $SU(3)_C \times SU(2)_L \times U(1)_Y$. Les contributions les plus importantes aux énergies accessibles au LHC proviennent des opérateurs de dimension six, de sorte que le lagrangien peut s'écrire

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^2}\mathcal{O}_i^{(6)}. \tag{2}$$

Les opérateurs de dimension six peuvent modifier les couplages des fermions et des bosons de jauge, affecter les interactions du champ scalaire et modifier la forme du potentiel de Higgs. Les modifications du secteur scalaire sont particulièrement importantes car elles affectent l'auto-couplage du champ de Higgs. Les opérateurs contenant plusieurs facteurs du doublet scalaire peuvent modifier les termes cubique et quartique du potentiel, ce qui conduit à des changements dans les processus impliquant un ou plusieurs bosons de Higgs.

Bien que l'auto-couplage du boson de Higgs soit directement lié aux processus de production de paires de bosons de Higgs, les mêmes opérateurs peuvent également contribuer aux processus à un seul boson de Higgs par des corrections quantiques. De telles contributions modifient les sections efficaces et les distributions cinématiques ; des contraintes sur l'auto-couplage peuvent donc être obtenues à partir de mesures précises de la production et de la désintégration d'un seul boson de Higgs. Ainsi, le formalisme SMEFT permet une étude systématique des écarts possibles au Modèle standard et constitue le cadre théorique naturel des analyses présentées dans ce travail.

L'expérience CMS au LHC

Le Grand collisionneur de hadrons (Large Hadron Collider, LHC) du CERN est le plus grand accélérateur de particules au monde. Il est installé dans un tunnel souterrain de 27 km de circonférence, à la frontière entre la Suisse et la France. Au LHC, des faisceaux de protons sont accélérés à très haute énergie à l'aide d'un système d'aimants supraconducteurs dipolaires et quadrupolaires, ainsi que de cavités radiofréquence qui assurent l'augmentation progressive de l'énergie des particules. Les protons sont d'abord accélérés dans une chaîne d'accélérateurs plus petits, puis injectés dans l'anneau principal du LHC, où ils sont portés à une énergie de plusieurs TeV par faisceau. Deux faisceaux de protons circulent en sens opposés et entrent en collision en des points d'interaction précis où sont installées les grandes expériences de détection.

L'une des principales expériences du LHC est CMS (Compact Muon Solenoid), un détecteur polyvalent conçu pour l'étude d'un large éventail de processus physiques dans les collisions proton–proton et d'ions lourds. L'expérience a été construite dans le but de mesurer avec précision la production et la désintégration de particules connues, ainsi que de rechercher de nouveaux phénomènes au-delà du Modèle standard. Le nom du détecteur reflète sa conception compacte et le puissant solénoïde supraconducteur qui génère un champ magnétique de 3.8 T, l'un des plus intenses jamais utilisés dans un détecteur de particules.

Le détecteur CMS possède une géométrie cylindrique autour du point de collision et se compose de plusieurs sous-systèmes concentriques, chacun ayant un rôle spécifique dans la reconstruction des particules. La partie la plus interne est constituée du trajectographe en silicium, qui permet une mesure précise des trajectoires des particules chargées dans un champ magnétique intense. Il est entouré d'un calorimètre électromagnétique, construit à partir de cristaux de tungstate de plomb, qui mesure l'énergie des électrons et des photons via le développement de gerbes électromagnétiques. À l'extérieur se trouve le calorimètre hadronique, destiné à mesurer l'énergie des hadrons par absorption et développement de gerbes hadroniques dans des couches d'absorbeur et de matériau actif.

L'ensemble du détecteur interne est placé à l'intérieur du solénoïde supraconducteur, tandis qu'à l'extérieur de l'aimant se trouve le système de détection des muons. Le système muonique est constitué de détecteurs gazeux intégrés dans le retour de flux magnétique en fer, qui sert à la fois de retour du champ magnétique et d'absorbeur supplémentaire pour les hadrons. En raison de leur grande pénétration, les muons traversent les parties internes du détecteur et peuvent être identifiés dans le système externe avec une grande précision.

Afin de traiter le très grand nombre de collisions produites au LHC, CMS utilise un système de déclenchement à plusieurs niveaux qui sélectionne en temps réel les événements d'intérêt physique. À partir d'une fréquence de collision de l'ordre de plusieurs dizaines de MHz, le système de déclenchement réduit le taux d'enregistrement à quelques centaines de Hz, permettant ainsi le stockage et l'analyse ultérieure des données. La reconstruction des événements est réalisée en combinant les informations de tous les sous-systèmes du détecteur, ce qui permet l'identification des électrons, muons, photons, hadrons et jets, ainsi que la détermination de l'impulsion transverse manquante.

L'énergie élevée des collisions, la grande quantité de données collectées et l'excellente résolution du détecteur font de l'expérience CMS un instrument essentiel pour les mesures de précision en physique des particules, notamment pour l'étude détaillée des propriétés du boson de Higgs et la recherche d'éventuels écarts au Modèle standard.

Électrons

Dans l'analyse du canal de désintégration $H \to ZZ \to 4\ell$, dont l'état final contient quatre leptons chargés, une reconstruction et une identification fiables des électrons sont essentielles pour la détermination précise des propriétés du boson de Higgs. Le canal à quatre leptons est l'un des canaux expérimentaux les plus propres pour l'étude du boson de Higgs, mais sa sensibilité dépend fortement de la capacité du détecteur et des algorithmes de reconstruction à reconstruire de manière fiable tous les leptons de l'événement. Étant donné que le rejet d'un seul électron entraîne la perte de tout l'événement candidat, une reconstruction et une identification de haute qualité des électrons sont indispensables pour préserver le signal.

Les électrons de faible impulsion transverse jouent un rôle particulièrement important. Dans l'état final à quatre leptons, une fraction significative des électrons se trouve précisément dans cette région, où le rapport signal sur bruit de fond est moins favorable qu'à plus haute impulsion transverse. Par conséquent, une détermination précise de l'efficacité dans cette région est cruciale pour le contrôle des incertitudes systématiques dans l'analyse $H \to ZZ \to 4\ell$.

Efficacité de reconstruction des électrons de faible impulsion transverse

L'approche standard utilisée dans CMS pour déterminer l'efficacité de reconstruction des électrons repose sur la résonance du boson Z. Cependant, en raison de la grande masse du boson Z, cette méthode n'est pas optimale pour l'étude des électrons de faible impulsion transverse. C'est pourquoi, dans ce travail, la méthode tag-and-probe a été appliquée à la résonance $J/\psi \to e^+e^-$, ce qui a permis de mesurer l'efficacité de reconstruction dans la région inférieure à 10 GeV. Le signal J/ψ est obtenu à partir des désintégrations $B \to KJ/\psi$, suivies de la désintégration du J/ψ en une paire d'électrons, ce qui permet d'obtenir un échantillon riche en électrons de faible impulsion transverse.

À cette fin, l'échantillon de données B-parking a été utilisé ; il a été spécifiquement conçu pour collecter un grand nombre d'événements contenant des désintégrations de mésons B. Dans la région de faible p_T, une grande fraction des électrons est reconstruite par une procédure tracker-driven, ce qui rend l'approche standard fondée sur les superamas insuffisamment efficace. Dans ce travail, l'efficacité a été déterminée en utilisant des traces comme objets probe, en vérifiant si elles sont reconstruites avec succès comme électrons. Il a été montré que la résonance J/ψ peut être clairement isolée du bruit de fond et qu'elle permet une mesure fiable de l'efficacité de reconstruction dans la région de faible impulsion transverse. Il s'agit ainsi de la première démonstration, au sein de CMS, d'une mesure de l'efficacité de reconstruction des électrons dans la région $p_T < 10$ GeV, ce qui constitue une étape importante vers la future détermination des facteurs d'échelle correspondants.

Efficacité d'identification des électrons

Dans l'analyse $H \to ZZ \to 4\ell$, il est nécessaire de distinguer les électrons prompts des objets de bruit de fond, tels que les hadrons mal reconstruits comme électrons, les électrons issus de désintégrations semi-leptoniques d'hadrons contenant des quarks lourds ou les électrons produits par conversion de photons. À cette fin, un algorithme d'identification multivarié fondé sur la méthode des arbres de décision à gradient boosté a été entraîné pour l'année 2022.

L'entraînement a été effectué séparément dans plusieurs régions d'impulsion transverse et de pseudorapidité afin de prendre en compte la dépendance des variables discriminantes à la cinématique et à la géométrie du détecteur. Les variables utilisées comprennent la forme des gerbes électromagnétiques dans le calorimètre, des informations issues du système de trajectographie, des variables de compatibilité entre la trace et le superamas, ainsi que des variables d'isolation. Il a notamment été montré que les variables d'isolation basées sur les amas fournissent une meilleure séparation entre signal et bruit de fond qu'une approche basée sur l'isolation particulaire. Une telle stratégie a conduit à une plus faible sélection du bruit de fond pour une même efficacité cible du signal, et a donc été retenue pour l'algorithme final d'identification. Le nouveau modèle d'identification pour 2022 a été intégré dans le logiciel CMS et approuvé pour être utilisé dans l'analyse $H \to ZZ \to 4\ell$.

Facteurs d'échelle pour l'identification des électrons

Dans ce travail, les facteurs d'échelle d'identification des électrons pour l'année 2022 ont été déterminés au moyen de la méthode tag-and-probe appliquée à la résonance du boson Z.

La mesure a été réalisée de manière différentielle dans plusieurs régions d'impulsion transverse et de pseudorapidité. Les plus faibles bins en p_T, ainsi que la région de transition entre la partie tonneau et la partie bouchon du calorimètre électromagnétique, ont constitué un défi particulier, car les incertitudes y sont plus élevées en raison d'une réponse du détecteur plus complexe. Pour déterminer les valeurs centrales et les incertitudes associées, une méthode RMS a été utilisée, consistant à moyenner les résultats obtenus avec différents choix de fonctions de densité de probabilité décrivant la distribution de la masse invariante.

Les facteurs d'échelle obtenus pour l'année 2022 présentent le comportement attendu : ils sont, dans la majeure partie de l'espace de phase, proches de l'unité, tandis que les incertitudes les plus importantes apparaissent dans la région de faible impulsion transverse et dans les zones géométriquement défavorables du détecteur. Ces facteurs d'échelle ont été approuvés et utilisés dans les premières mesures de Run 3 du processus $H \to ZZ \to 4\ell$, de sorte que ce travail contribue directement aux résultats physiques de l'expérience CMS.

Mesure de l'auto-couplage du boson de Higgs dans la région hors couche de masse

Le résultat central de cette thèse est la mesure du modificateur d'auto-couplage du boson de Higgs (κ_λ) dans le canal de désintégration $H \to ZZ^* \to 4\ell$ hors couche de masse. L'analyse utilise les données de collisions proton–proton collectées par le détecteur CMS pendant la deuxième période de fonctionnement du LHC (2016–2018) à une énergie dans le centre de masse de 13 TeV, correspondant à une luminosité intégrée totale de 138 fb^{-1}. La mesure a été réalisée dans le cadre du formalisme SMEFT, dans lequel κ_λ est sondé par l'ajout d'un opérateur de dimension six, qui affecte à la fois les rendements totaux d'événements et les distributions des variables observées.

Contrairement aux mesures directes fondées sur la production de paires de bosons de Higgs, la contrainte sur κ_λ obtenue dans ce travail provient de corrections de boucle dans des processus impliquant un seul boson de Higgs. Pour la première fois, la région hors couche de masse de l'espace de phase est utilisée pour une telle mesure. Dans cette région, la distribution de la masse invariante du système à quatre leptons présente une sensibilité accrue à la présence de l'opérateur de dimension six, ce qui conduit à une modification caractéristique de la forme de la distribution $m_{4\ell}$.

Corrections SMEFT

Les écarts par rapport aux prédictions du Modèle standard sont introduits dans le cadre du formalisme SMEFT. Les corrections au processus de fusion gluonique $gg \to H^* \to ZZ$ sont calculées d'après les résultats de [1], où les contributions de l'opérateur $\mathcal{O}_6$ apparaissent dans la production, la propagation et la désintégration du boson de Higgs.

La correction à la production du Higgs est donnée par le vertex renormalisé :

$$\hat{\Gamma}^{\mu\nu}_{ggh} = -\frac{\alpha_s}{\pi v}\left(\eta^{\mu\nu} p_1 \cdot p_2 - p_1^{\nu} p_2^{\mu}\right)\left[\frac{\delta Z_h}{2}\mathcal{F}_1 + \frac{\lambda \bar{c}_6}{(4\pi)^2}\mathcal{F}_2\right]. \tag{3}$$

La correction du propagateur est obtenue à partir de la self-énergie renormalisée :

$$\hat{\Sigma}(\hat{s}) = \Sigma(\hat{s}) + (\hat{s} - m_h^2)\delta Z_h - \delta m_h^2. \tag{4}$$

La correction à la désintégration $H \to ZZ$ est donnée par :

$$\hat{\Gamma}^{\mu\nu}_{hZZ} = \frac{2m_Z^2}{v}\left[\eta^{\mu\nu}\mathcal{G}_1 + p_1^{\nu} p_2^{\mu}\mathcal{G}_2\right]. \tag{5}$$

Les contributions proportionnelles à δZ_h issues de la production et de la désintégration s'annulent avec la contribution correspondante provenant du propagateur, de sorte que l'amplitude prend finalement une forme dans laquelle les corrections de production et de désintégration varient linéairement avec le paramètre c_6, tandis que le propagateur est modifié quadratiquement en c_6.

Dans le cadre de cette thèse, ce formalisme a également été étendu au secteur électrofaible, de sorte que des corrections analogues ont été implémentées pour les processus de fusion de bosons vecteurs et de production associée avec un boson vecteur.

Repondération par poids MELA et construction des discriminants

Dans le canal de désintégration $H \to ZZ \to 4\ell$, la cinématique de l'événement permet de calculer la probabilité qu'il provienne d'un processus physique donné. À cette fin, on utilise la Matrix Element Likelihood Approach (MELA), dans laquelle, pour chaque événement, on calcule une probabilité proportionnelle au carré de l'élément de matrice, évalué pour une hypothèse donnée à partir de la cinématique reconstruite de l'événement.

Pour des hypothèses physiques ayant les mêmes états initiaux et finaux, mais différant par la structure des couplages, les événements générés pour une hypothèse donnée peuvent être repondérés vers une autre hypothèse à l'aide du rapport des éléments de matrice. Cette procédure s'applique aux échantillons du Modèle standard comme aux échantillons décrivant des couplages anormaux, ce qui permet d'obtenir des prédictions pour différentes valeurs du paramètre c_6 sans devoir générer de nouveaux échantillons Monte Carlo.

À partir des probabilités MELA, on définit des discriminants permettant de séparer le signal, l'interférence et le bruit de fond. Ces discriminants sont construits à partir des probabilités associées aux processus de signal, de bruit de fond et d'interférence ; ils sont utilisés pour la catégorisation des événements et pour accroître la sensibilité de l'analyse, en particulier dans la région hors couche de masse où l'interférence entre signal et bruit de fond est importante.

Modèle statistique et résultats

L'analyse a été réalisée à l'aide d'une approche dans laquelle les distributions attendues des variables observées sont décrites par des histogrammes obtenus par repondération MELA. Chaque histogramme représente un *template* pour la contribution du signal, du bruit de fond ou de l'interférence, et le modèle statistique est construit en comparant ces *templates* aux données observées. Le paramètre d'intérêt est le modificateur de l'auto-couplage du boson de Higgs κ_λ, qui modifie les rendements attendus dans les différents bins.

Pour chaque bin du modèle statistique, le nombre attendu d'événements est paramétré par

$$\nu_i = \mu S_i + \sqrt{\mu}\, I_i + B_i + \mathcal{O}_{\lambda,i}, \tag{6}$$

où S_i désigne la contribution du signal, B_i celle du bruit de fond, I_i l'interférence entre le signal et le bruit de fond, et $\mathcal{O}_{\lambda,i}$ la contribution supplémentaire provenant des corrections SMEFT dépendant du paramètre κ_λ.

La fonction de vraisemblance est construite comme un produit de probabilités de Poisson sur tous les bins, avec l'inclusion de paramètres supplémentaires décrivant les incertitudes systématiques. Les limites sur le paramètre κ_λ sont déterminées à l'aide de la méthode du maximum de vraisemblance profilé, où, pour chaque valeur de κ_λ, les autres paramètres du modèle sont déterminés par maximisation de la vraisemblance.

Il est obtenu que la valeur attendue du paramètre d'auto-couplage est $\kappa_\lambda = 1^{+21}_{-18}$ au niveau de confiance de 68%, avec un intervalle $[-21, 27]$ au niveau de confiance de 95%. Le résultat observé est $\kappa_\lambda = 2^{+17}_{-15}$ au niveau de confiance de 68%, avec un intervalle $[-19, 29]$ au niveau de confiance de 95%.

Conclusion et perspectives

Cette thèse présente le travail réalisé au cours des études doctorales menées entre 2022 et 2026 dans le cadre de l'expérience CMS au LHC. Après la découverte du boson de Higgs pendant la période Run 1, le programme de physique du LHC est entré dans une ère de mesures de précision, dont l'objectif est de tester le Modèle standard avec une sensibilité croissante et de rechercher de faibles écarts susceptibles d'indiquer l'existence d'une physique au-delà du Modèle standard. Les grands ensembles de données collectés durant Run 2, ainsi que les données de Run 3, permettent d'étudier les propriétés du boson de Higgs avec la meilleure précision jamais atteinte et marquent la transition vers l'ère du High-Luminosity LHC. Au moment de la rédaction de cette thèse, le Run 3 approche de sa fin, avec encore quelques mois de prise de données avant le prochain long arrêt de l'accélérateur.

Les analyses présentées dans cette thèse sont réalisées dans l'état final $H \to ZZ \to 4\ell$, qui constitue l'un des plus sensibles pour la mesure des propriétés du boson de Higgs. Dans ce domaine, la précision des résultats dépend fortement de la qualité de la reconstruction et de l'identification des leptons, ainsi que de la détermination fiable des efficacités associées. C'est pourquoi une partie importante du travail a été consacrée à la détermination des efficacités de reconstruction et d'identification des électrons, y compris l'entraînement de l'algorithme d'identification, le calcul des facteurs d'échelle et la mesure de l'efficacité de reconstruction dans la région de faible impulsion transverse, particulièrement importante pour l'état final à quatre leptons.

Le principal résultat physique de cette thèse est la mesure de l'auto-couplage du boson de Higgs dans le cadre du formalisme SMEFT, paramétrée par le modificateur κ_λ, à partir de processus impliquant un seul boson de Higgs. Bien que l'auto-couplage ne soit directement accessible que dans la production de paires de bosons de Higgs, les corrections de boucle dans les processus à un seul boson de Higgs permettent une sensibilité indirecte à la forme du potentiel scalaire. Le canal $H \to ZZ \to 4\ell$ offre un environnement expérimental particulièrement propre, ce qui permet des mesures sensibles à de tels effets.

Avec la poursuite de l'exploitation du LHC en Run 3 et les préparatifs en vue de l'ère du High-Luminosity LHC, l'étude du secteur de Higgs restera l'un des domaines centraux de la physique des particules. L'augmentation de la sensibilité des mesures, l'étude de processus rares et la recherche d'écarts par rapport aux prédictions du Modèle standard seront essentielles pour comprendre si le Modèle standard constitue une description complète de la nature ou seulement une approximation effective d'une théorie plus fondamentale. Le travail présenté dans cette thèse apporte une contribution modeste, mais significative, à cet effort collectif.

Prošireni sažetak

Razumijevanje temeljnih sastavnica tvari i zakona koji upravljaju njihovim međudjelovanjem jedno je od središnjih pitanja fizike još od antičkih vremena. Prve ideje o građi materije, koje potječu iz filozofskih razmatranja o nedjeljivim česticama, postupno su se razvile u suvremenu sliku prema kojoj se ponašanje prirode na najmanjim skalama opisuje kvantnom teorijom polja. Tijekom razvoja fizike više se puta pokazalo da strukture koje su se smatrale elementarnima zapravo imaju unutarnju građu, što je poticalo razvoj novih teorijskih modela i izgradnju sve snažnijih eksperimentalnih uređaja.

Današnji opis elementarnih čestica i njihovih međudjelovanja daje Standardni model (eng., Standard Model, SM) fizike čestica. Ova teorija temelji se na principu lokalne baždarne simetrije i unutar jedinstvenog formalizma opisuje elektromagnetsku, slabu i jaku interakciju. Tijekom posljednjih desetljeća predviđanja SM-a potvrđena su velikim brojem preciznih mjerenja, čime je on postao jedna od najuspješnijih teorija u povijesti fizike. Ključni dio teorije predstavlja mehanizam narušavanja elektroslabe simetrije, koji objašnjava kako elementarne čestice dobivaju masu bez narušavanja baždarne invarijantnosti. Eksperimentalna potvrda ovog mehanizma ostvarena je 2012. godine otkrićem Higgsova bozona na Velikom hadronskom sudaraču (eng., *Large Hadron Collider*, LHC), čime je potvrđen posljednji nedostajući element čestičnog sadržaja SM-a.

Unatoč iznimnom uspjehu, SM se ne smatra potpunom teorijom prirode. Ne uključuje gravitaciju, ne objašnjava prirodu tamne tvari niti daje odgovor na niz otvorenih pitanja vezanih uz strukturu skalarnog sektora i stabilnost Higgsova potencijala. Zbog toga je jedan od glavnih ciljeva suvremene fizike čestica traženje odstupanja od predviđanja SM-a koja bi mogla ukazivati na postojanje nove fizike. Takva odstupanja mogu se pojaviti kroz izravnu produkciju novih čestica pri visokim energijama, ali i kroz male promjene u poznatim procesima koje postaju vidljive u preciznim mjerenjima.

Higgsov sektor ima posebno važnu ulogu u tim istraživanjima. Iako su masa Higgsova bozona i mnoga njegova vezanja već izmjerena, oblik skalarnog potencijala još uvijek nije u potpunosti eksperimentalno ispitan. U SM-u taj potencijal određuje jačinu samovezanja Higgsova bozona, koja se može proučavati u procesima s jednim ili više Higgsovih bozona u konačnom stanju. Izravno mjerenje samovezanja eksperimentalno je vrlo zahtjevno jer se oslanja na rijetke procese s malim produkcijskim presjecima. Međutim, osjetljivost na samovezanje može se dobiti i neizravno, kroz precizna mjerenja produkcije i raspada jednog Higgsova bozona, gdje kvantne korekcije ovise o strukturi skalarnog sektora. Takvi učinci mogu se opisati u okviru efektivne teorije polja Standardnog modela (eng., *Standard Model effective field theory*, SMEFT), u kojoj se moguća odstupanja od Standardnog modela parametriziraju pomoću operatora viših dimenzija.

Najveće energije sudara danas se postižu na LHC-u u CERN-u, što omogućuje detaljno proučavanje svojstava Higgsova bozona i drugih procesa na dosad nedostižnim energijskim skalama. Nakon otkrića Higgsova bozona u prvom periodu rada LHC-a, podaci prikupljeni tijekom drugog i trećeg perioda rada omogućuju sve preciznija testiranja SM-a. U tom području osjetljivost mjerenja ne ovisi samo o količini podataka, nego i o vrlo dobrom razumijevanju rada detektora, rekonstrukcije čestica i određivanja učinkovitosti identifikacije. Konačna stanja koja sadrže elektrone i mione posebno su važna jer omogućuju preciznu rekonstrukciju kinematike događaja. Među njima se raspadni kanal $H \to ZZ \to 4\ell$, zbog odličnog omjera signala i pozadine, ističe kao jedan od najpreciznijih za proučavanje svojstava Higgsova bozona.

Istraživanja predstavljena u ovom radu provedena su u sklopu eksperimenta kompaktnog mionskog solenoida (eng., *Compact Muon Solenoid*, CMS) na LHC-u i usmjerena su na mjerenja povezana s Higgsovim bozonom u četveroleptonskom konačnom stanju. Poseban naglasak stavljen je na određivanje učinkovitosti rekonstrukcije i identifikacije elektrona, koje su ključne za precizna mjerenja u ovom kanalu, kao i na proučavanje osjetljivosti procesa s jednim Higgsovim bozonom na promjene samovezanja Higgsova polja u okviru SMEFT formalizma. Ova istraživanja dio su šireg napora da se SM ispita s najvećom mogućom točnošću i da se pronađu eventualni znakovi fizike izvan njegovih okvira.

Teorijski uvod

SM fizike čestica je kvantna teorija polja zasnovana na lokalnoj baždarnoj simetriji $SU(3)_C \times SU(2)_L \times U(1)_Y$, koja opisuje jaku, slabu i elektromagnetsku interakciju među elementarnim česticama. Fermionski sadržaj teorije sastoji se od tri generacije kvarkova i leptona, dok međudjelovanja posreduju baždarni bozoni povezani s generatorima grupe simetrija. Dinamika sustava određena je Lagrangijanom koji uključuje kinetičke članove fermionskih i baždarnih polja, kao i međudjelovanja definirana kovarijantnom derivacijom.

Maseni članovi za slabe baždarne bozone i fermione ne mogu se zapisati na način koji poštuje baždarnu simetriju. Ovaj problem rješava se uvođenjem kompleksnog skalarnog dubleta čiji potencijal dovodi do spontanog narušenja elektroslabe simetrije. Nakon odabira vakuumskog očekivanog iznosa skalarnog polja, slabi baždarni bozoni dobivaju masu, dok foton ostaje bez mase. Fermioni dobivaju masu kroz Yukawina međudjelovanja sa skalarnim poljem. U spektru teorije ostaje jedna fizička skalarna pobuda, Higgsov bozon, čija su vezanja prema fermionima i baždarnim bozonima određena njihovim masama. Potencijal skalarnog polja sadrži i članove koji opisuju samovezanje Higgsova polja, što određuje oblik skalarnog sektora teorije.

Iako SM uspješno opisuje veliki broj eksperimentalnih rezultata, može se promatrati kao efektivna teorija koja vrijedi do određene energijske skale. U općenitom slučaju, učinci nove fizike na višim energijama mogu se opisati proširenjem Lagrangijana dodatnim operatorima viših dimenzija izgrađenima od poznatih polja. Takav pristup vodi do efektivne teorije polja, u kojoj se Lagrangijan zapisuje kao razvoj po inverznim potencijama karakteristične skale nove fizike Λ,

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{C_i^{(5)}}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{C_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \cdots , \tag{7}$$

gdje su $\mathcal{O}_i^{(d)}$ operatori dimenzije d, a $C_i^{(d)}$ pripadni Wilsonovi koeficijenti.

U slučaju da nema dodatnih lakih stupnjeva slobode, najopćenitiji opis mogućih odstupanja dobiva se u okviru SMEFT-a. U tom formalizmu koriste se samo polja SM-a, a Lagrangijan se proširuje svim operatorima dimenzije veće od četiri koji su invarijantni na $SU(3)_C \times SU(2)_L \times U(1)_Y$. Najvažniji doprinosi na energijama dostupnim na LHC-u dolaze od operatora dimenzije šest, pa se Lagrangijan može zapisati kao

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^2} \mathcal{O}_i^{(6)}. \tag{8}$$

Operatori dimenzije šest mogu mijenjati vezanja fermiona i baždarnih bozona, utjecati na međudjelovanja skalarnog polja te modificirati oblik Higgsova potencijala. Promjene u skalarnom sektoru posebno su važne jer utječu na samovezanje Higgsova polja. Operatori koji sadrže više faktora skalarnog dubleta mogu promijeniti kubični i kvartični član potencijala, što dovodi do promjena u procesima koji uključuju jedan ili više Higgsovih bozona.

Iako je samovezanje Higgsova bozona izravno povezano s procesima produkcije para Higgsovih bozona, isti operatori mogu doprinijeti i procesima s jednim Higgsovim bozonom kroz kvantne korekcije. Takvi doprinosi mijenjaju udarne presjeke i kinematičke distribucije, pa se ograničenja na samovezanje mogu dobiti iz preciznih mjerenja produkcije i raspada jednog Higgsova bozona. Zbog toga SMEFT formalizam omogućuje sustavno proučavanje mogućih odstupanja od SM-a i predstavlja prirodan teorijski okvir za analize predstavljene u ovom radu.

CMS eksperiment na LHC-u

LHC u CERN-u najveći je akcelerator čestica na svijetu. Smješten je u podzemnom tunelu opsega 27 km, na granici između Švicarske i Francuske. U LHC-u se snopovi protona ubrzavaju do vrlo visokih energija pomoću sustava supravodljivih dipolnih i kvadrupolnih magneta te radiofrekvencijskih šupljina koje osiguravaju postupno povećavanje energije čestica. Protoni se najprije ubrzavaju u nizu manjih akceleratora, a zatim se ubrizgavaju u glavni prsten LHC-a, gdje se ubrzavaju do energije od nekoliko TeV po snopu. Dva snopa protona kruže u suprotnim smjerovima i sudaraju se u točno određenim točkama u kojima su smješteni veliki detektorski eksperimenti.

Jedan od glavnih eksperimenata na LHC-u je CMS, opći detektor namijenjen proučavanju širokog spektra fizikalnih procesa u sudarima proton–proton i teških iona. Eksperiment je konstruiran s ciljem preciznog mjerenja produkcije i raspada poznatih čestica, kao i traženja novih fenomena izvan SM-a. Naziv detektora odražava njegovu kompaktnu konstrukciju i snažan supravodljivi solenoid koji stvara magnetsko polje od 3.8 T, jedno od najjačih polja ikada korištenih u detektorima čestica, kao i sposobnost precizne detekcije miona.

CMS detektor ima cilindričnu geometriju oko točke sudara i sastoji se od više koncentričnih podsustava, od kojih svaki ima posebnu ulogu u rekonstrukciji čestica. Najunutarniji dio čini silicijski detektor tragova, koji omogućuje precizno mjerenje putanja nabijenih čestica u jakom magnetskom polju. Okružen je elektromagnetskim kalorimetrom, izgrađenim od kristala olovnog volframata, koji mjeri energiju elektrona i fotona putem razvoja elektromagnetskih pljuskova. Izvan njega nalazi se hadronski kalorimetar, koji služi za mjerenje energije hadrona kroz apsorpciju i razvoj hadronskih pljuskova u slojevima apsorbera i aktivnog materijala.

Cijeli unutarnji detektor smješten je unutar supravodljivog solenoida, dok se izvan magneta nalazi sustav za detekciju miona. Mionski sustav sastoji se od plinskih detektora ugrađenih u željeznu povratnu jarmu magneta, koja služi kao magnetski povratni tok i kao dodatni apsorber za hadrone. Zbog velike prodornosti, mioni prolaze kroz unutarnje dijelove detektora i mogu se identificirati u vanjskom sustavu s visokom preciznošću.

Kako bi se obradio velik broj sudara koji se događaju u LHC-u, CMS koristi višestupanjski sustav okidanja (eng., *trigger*) koji u stvarnom vremenu odabire događaje od fizikalnog interesa. Od početne frekvencije sudara od reda veličine desetaka MHz, sustav okidanja smanjuje brzinu zapisa na nekoliko stotina Hz, što omogućuje pohranu i daljnju analizu podataka. Rekonstrukcija događaja provodi se kombiniranjem informacija iz svih podsustava detektora, čime se omogućuje identifikacija elektrona, miona, fotona, hadrona i mlazova (eng., *jets*), kao i određivanje nedostatka transverzalnog momenta.

Visoka energija sudara, velika količina prikupljenih podataka i vrlo dobra rezolucija detektora čine CMS eksperiment ključnim instrumentom za precizna mjerenja u fizici čestica, uključujući detaljno proučavanje svojstava Higgsova bozona i traženje mogućih odstupanja od SM-a.

Elektroni

U analizi kanala raspada $H \to ZZ \to 4\ell$, u kojem konačno stanje sadrži četiri nabijena leptona, pouzdana rekonstrukcija i identifikacija elektrona ključne su za precizno određivanje svojstava Higgsova bozona. Kanal s četiri leptona jedan je od eksperimentalno najčišćih kanala za proučavanje Higgsova bozona, ali njegova osjetljivost snažno ovisi o sposobnosti detektora i rekonstrukcijskih algoritama da pouzdano rekonstruiraju sve leptone u događaju. Budući da odbacivanje i jednog elektrona znači gubitak cijelog kandidata događaja, kvalitetna rekonstrukcija i identifikacija elektrona nužna je za očuvanje signala.

Posebno važnu ulogu imaju elektroni s niskim transverzalnim momentom. U konačnom stanju sa četiri leptona značajan dio elektrona nalazi se upravo u tom području, gdje je omjer signala i pozadine nepovoljniji nego pri većim vrijednostima transverzalnog momenta. Stoga je precizno određivanje učinkovitosti u tom području ključno za kontrolu sistematskih nesigurnosti u analizi $H \to ZZ \to 4\ell$.

Učinkovitost rekonstrukcije elektrona s niskim transverzalnim momentom

Standardni pristup u CMS-u za određivanje rekonstrukcijske učinkovitosti elektrona temelji se na čestici Z bozona. Međutim, zbog velike mase Z bozona takva metoda nije optimalna za proučavanje elektrona s niskim transverzalnim momentom. Zbog toga je u ovom radu metoda označi i testiraj (eng., *Tag and Probe*, TnP) primijenjena na česticu $J/\psi \to e^+e^-$, čime je omogućeno mjerenje učinkovitosti rekonstrukcije u području transverzalnog momenta ispod 10 GeV. Signal J/ψ dobiven je iz raspada $B \to KJ/\psi$, pri čemu J/ψ dalje raspada u par elektrona i pozitrona, što omogućuje dobivanje uzorka bogatog elektronima niskog transverzalnog momenta.

U tu svrhu iskorišten je B-parking skup podataka, koji je posebno dizajniran za prikupljanje velikog broja događaja s raspadima B-mezona, pri čemu se obrada događaja odgađa za kasnije kako bi se zaobišla ograničenja računalnih resursa tijekom prikupljanja podataka. U području niskog p_T, velik dio elektrona rekonstruira se postupkom baziranom na detektoru tragova (eng., *tracker-driven*), zbog čega standardni pristup zasnovan na superklasterima nije dovoljno učinkovit. U ovom radu učinkovitost je određena korištenjem tragova kao testnih (eng., *probe*) objekata, pri čemu se ispituje jesu li oni uspješno rekonstruirani kao elektroni. Pokazano je da se čestica J/ψ može jasno izdvojiti iz pozadine te da omogućuje pouzdano mjerenje učinkovitosti rekonstrukcije u području niskog transverzalnog momenta. Time je prvi put u CMS-u demonstrirano mjerenje učinkovitosti rekonstrukcije elektrona u području $p_T < 10$ GeV, što predstavlja važan korak prema budućem određivanju odgovarajućih faktora skaliranja.

Algoritam za identifikaciju elektrona

U analizi $H \to ZZ \to 4\ell$ potrebno je razlikovati elektrone iz primarnih interakcija od pozadinskih objekata, poput hadrona pogrešno rekonstruiranih kao elektroni, elektrona iz poluleptonskih raspada hadrona s teškim kvarkovima ili elektrona nastalih konverzijom fotona. U tu je svrhu za 2022. godinu treniran multivarijatni identifikacijski algoritam, temeljen na metodi gradijentno pojačanih stabala odlučivanja.

Trening je proveden odvojeno u više područja transverzalnog momenta i pseudorapiditeta, kako bi se u obzir uzela ovisnost diskriminacijskih varijabli o kinematici i geometriji detektora. Korištene su varijable oblika elektromagnetskog pljuska u kalorimetru, informacije iz sustava tragova, varijable usklađenosti između traga i superklastera te izolacijske varijable. Posebno je pokazano da klaster-bazirane izolacijske varijable daju bolju separaciju signala i pozadine od pristupa temeljenog na čestičnoj izolaciji. Takav pristup rezultirao je manjim propuštanjem pozadine za istu ciljanu učinkovitost signala te je zato odabran za konačni identifikacijski algoritam. Novi identifikacijski model za 2022. integriran je u CMS softver te se koristi u $H \to ZZ \to 4\ell$ analizi.

Faktori skaliranja za identifikaciju elektrona

U ovom radu određeni su identifikacijski faktori skaliranja za identifikaciju elektrona za 2022. godinu, korištenjem TnP metode na čestici Z bozona.

Mjerenje je provedeno diferencijalno u više područja transverzalnog momenta i pseudorapiditeta. Poseban izazov predstavljali su najniži p_T binovi i prijelazno područje između bačvastog (eng., *barrel*) i krajnjeg (eng., *endcap*) dijela elektromagnetskog kalorimetra, gdje su nesigurnosti veće zbog složenijeg odziva detektora. Za određivanje središnjih vrijednosti i pripadnih nesigurnosti korištena je metoda kvadratnog usrednjavanja (eng., *Root Mean Square*), kojom se usrednjuje rezultat baziran na različitim odabirima funkcija gustoće vjerojatnosti koji opisuju distribuciju invarijantne mase.

Dobiveni faktori skaliranja za 2022. godinu pokazuju očekivano ponašanje: u većini faznog prostora bliski su jedinici, dok su veće nesigurnosti prisutne u području niskog transverzalnog momenta i u geometrijski nepovoljnim dijelovima detektora. Ti su faktori skaliranja odobreni i korišteni u prvim Run 3 mjerenjima procesa $H \to ZZ \to 4\ell$, čime ovaj rad izravno doprinosi fizikalnim rezultatima CMS eksperimenta.

Mjerenje samovezanja Higgsovog bozona u području van-masene ljuske

Središnji rezultat ove disertacije je mjerenje modifikatora samovezanja Higgsovog bozona (κ_λ) u kanalu raspada $H \to ZZ \to 4\ell$ van masene ljuske. Analiza koristi podatke proton–proton sudara prikupljene detektorom CMS tijekom drugog rada LHC-a (2016.–2018.) pri energiji sudara centra mase od 13 TeV, što odgovara ukupnom integriranom luminozitetu od 138 fb^{-1}. Mjerenje je provedeno u okviru SMEFT formalizma, u kojem se κ_λ ispituje kroz dodavanje operatora dimenzije šest, koji utječe i na ukupni udarni presjek i na distribucije promatranih varijabli.

Za razliku od izravnih mjerenja temeljenih na produkciji para Higgsovih bozona, ograničenje na κ_λ u ovom radu dobiva se iz korekcija u procesima s jednim Higgsovim bozonom. Po prvi put se za takvo mjerenje koristi područje faznog prostora van masene ljuske. U tom području raspodjela invariantne mase četveroleptonskog sustava pokazuje povećanu osjetljivost na prisutnost operatora dimenzije šest, što dovodi do karakteristične promjene oblika $m_{4\ell}$ raspodjele.

SMEFT korekcije

Odstupanja od predviđanja SM-a uvedena su u okviru formalizma SMEFT-a. Korekcije na proces gluonske fuzije $gg \to H \to ZZ$ računate su prema rezultatu [1] pri čemu se doprinosi operatora $\mathcal{O}_6$ pojavljuju u produkciji, propagaciji i raspadu Higgsovog bozona.

Korekcija na Higgs produkciju dana je renormaliziranim vrhom:

$$\hat{\Gamma}^{\mu\nu}_{ggh} = -\frac{\alpha_s}{\pi v}\left(\eta^{\mu\nu} p_1 \cdot p_2 - p_1^\nu p_2^\mu\right)\left[\frac{\delta Z_h}{2}\mathcal{F}_1 + \frac{\lambda \bar{c}_6}{(4\pi)^2}\mathcal{F}_2\right]. \tag{9}$$

Korekcija propagatora dobiva se iz renormalizirane samo - energije:

$$\hat{\Sigma}(\hat{s}) = \Sigma(\hat{s}) + (\hat{s} - m_h^2)\delta Z_h - \delta m_h^2. \tag{10}$$

Korekcija na raspad $H \to ZZ$ dana je sa:

$$\hat{\Gamma}^{\mu\nu}_{hZZ} = \frac{2m_Z^2}{v}\left[\eta^{\mu\nu}\mathcal{G}_1 + p_1^\nu p_2^\mu \mathcal{G}_2\right]. \tag{11}$$

Doprinosi proporcionalni δZ_h iz produkcije i raspada poništavaju se s odgovarajućim doprinosom iz propagatora, tako da amplituda na kraju poprima oblik u kojem korekcije produkcije i raspada skaliraju linearno s parametrom c_6, dok se propagator mijenja kvadratno sa c_6.

U okviru ove teze navedeni formalizam proširen je i na elektroslabi sektor, tako da su analogne korekcije implementirane i za procese fuzije vektorskih bozona i pridružene produkcije s vektorskim bozonom.

MELA težinsko reskaliranje i izgradnja diskriminanti

U kanalu raspada $H \to ZZ \to 4\ell$ kinematika događaja omogućuje izračun vjerojatnosti da događaj potječe iz određenog fizikalnog procesa. Za tu svrhu koristi se izračun vjerovatnosti temeljen na matričnim elementima (eng., *Matrix Element Likelihood Approach,* MELA), u kojem se za svaki događaj računa vjerojatnost proporcionalna kvadratu matričnog elementa pri čemu se matrični element evaluira za zadanu hipotezu koristeći rekonstruiranu kinematiku događaja.

Za fizikalne hipoteze koje imaju ista početna i konačna stanja, a razlikuju se samo u strukturi vezanja, događaji generirani za jednu hipotezu mogu se težinski reskalirati na drugu hipotezu pomoću omjera matričnih elemenata. Ovaj postupak se primjenjuje na uzorke SM-a i na uzorke koji opisuju anomalna vezanja, što omogućuje dobivanje predviđanja za različite vrijednosti parametra c_6 bez potrebe za generiranjem novih Monte Carlo uzoraka.

Na temelju MELA vjerojatnosti definiraju se diskriminante koje razdvajaju signal, interferenciju i pozadinu. Diskriminante se konstruiraju iz vjerojatnosti za signalni, pozadinski i interferencijski proces te se koriste za kategorizaciju događaja i povećanje osjetljivosti analize, posebno u području van masene ljuske gdje je interferencija između signala i pozadine značajna.

Statistički model i rezultati

Analiza je provedena koristeći pristup u kojem se očekivane distribucije promatranih varijabli opisuju histogramima dobivenima MELA težinskim reskaliranjem. Svaki histogram predstavlja predložak za signalni, pozadinski ili interferencijski doprinos, a statistički model konstruira se usporedbom tih predložaka s opaženim podacima. Parametar od interesa je modifikator samovezanja Higgsovog bozona κ_λ, koji mijenja očekivane prinose u pojedinim binovima.

Za svaki bin statističkog modela očekivani broj događaja parametriziran je kao

$$\nu_i = \mu S_i + \sqrt{\mu}\, I_i + B_i + \mathcal{O}_{\lambda,i}, \tag{12}$$

gdje S_i označava signalni doprinos, B_i pozadinu, I_i interferenciju između signala i pozadine, μ modifikator snage signala a $\mathcal{O}_{\lambda,i}$ dodatni doprinos koji potječe od SMEFT korekcija ovisnih o parametru κ_λ.

Vjerojatnosna funkcija konstruirana je kao produkt Poissonovih vjerojatnosti po svim binovima, uz uključivanje dodatnih parametara koji opisuju sistematske nesigurnosti. Limiti na parametru κ_λ određeni su metodom profiliranog maksimuma vjerojatnosti, pri čemu se za svaku vrijednost κ_λ ostali parametri modela određuju maksimizacijom funkcije vjerojatnosti.

Dobiveno je da očekivana vrijednost parametra samovezanja iznosi $\kappa_\lambda = 1^{+21}_{-18}$ na razini pouzdanosti od 68%, uz interval $[-22,\ 26]$ na razini pouzdanosti od 95%. Opaženi rezultat iznosi $\kappa_\lambda = 2^{+17}_{-15}$ na razini pouzdanosti od 68%, uz interval $[-21,\ 27]$ na razini pouzdanosti od 95%.

Zaključak i perspektive

Ova disertacija prikazuje rad proveden tijekom doktorskog studija u razdoblju 2022.–2026. u okviru eksperimenta CMS na sudaraču LHC. Nakon otkrića Higgsovog bozona tijekom Run 1 razdoblja, fizikalni program LHC-a ušao je u razdoblje preciznih mjerenja, u kojem je cilj ispitivati SM s rastućom osjetljivošću te tražiti mala odstupanja koja bi mogla upućivati na postojanje fizike izvan SM-a. Veliki skupovi podataka prikupljeni tijekom Run 2 razdoblja, zajedno s Run 3 podacima, omogućuju proučavanje svojstava Higgsovog bozona s dosad najvećom preciznošću i predstavljaju prijelaz prema razdoblju visokog luminoziteta (eng., *High-Luminosity*, HL) LHC-a. U trenutku pisanja ove disertacije Run 3 se približava kraju, s još nekoliko mjeseci prikupljanja podataka prije sljedećeg dugog zastoja akceleratora.

Analize prikazane u ovoj disertaciji provedene su u konačnom stanju $H \to ZZ \to 4\ell$, koje je jedno od najosjetljivijih za mjerenje svojstava Higgsovog bozona. U tom području točnost rezultata snažno ovisi o kvaliteti rekonstrukcije i identifikacije leptona, kao i o pouzdanom određivanju pripadajuće efikasnosti. Zbog toga je značajan dio rada bio posvećen određivanju efikasnosti rekonstrukcije i identifikacije elektrona, uključujući treniranje identifikacijskog algoritma, računanje faktora skaliranja te mjerenje efikasnosti rekonstrukcije u području malog transverzalnog momenta, koje je posebno važno za četveroleptonsko konačno stanje.

Glavni fizikalni rezultat disertacije je mjerenje samovezanja Higgsovog bozona u okviru SMEFT formalizma, parametrizirano modifikatorom κ_λ, koristeći procese s jednim Higgsovim bozonom. Iako je samovezanje izravno dostupno tek u produkciji para Higgsovih bozona, korekcije u procesima s jednim Higgsovim bozonom omogućuju indirektnu osjetljivost na oblik skalarnog potencijala. Kanal $H \to ZZ \to 4\ell$ pruža posebno čisto eksperimentalno okruženje, što omogućuje mjerenja osjetljiva na takve efekte.

S nastavkom rada LHC-a u Run 3 i pripremama za razdoblje rada HL LHC-a, istraživanje Higgsovog sektora ostat će jedno od središnjih područja fizike čestica. Daljnje povećanje osjetljivosti mjerenja, proučavanje rijetkih procesa i potraga za odstupanjima od predviđanja SM-a bit će ključni za razumijevanje predstavlja li SM potpun opis prirode ili samo efektivnu aproksimaciju dublje teorije. Rad prikazan u ovoj disertaciji čini mali, ali značajan doprinos tom zajedničkom nastojanju.

Contents

Introduction

Understanding the fundamental constituents of matter and the laws that govern their interactions has been one of the central goals of physics for centuries. Early ideas about the structure of matter, originating from philosophical concepts of indivisible particles, gradually evolved into the modern picture in which the behavior of nature at the smallest scales is described by quantum field theory (QFT). Over time, experimental discoveries have repeatedly revealed that structures once thought to be elementary are in fact composed of smaller building blocks, motivating the development of increasingly powerful theoretical frameworks and experimental tools.

The current description of elementary particles and their interactions is given by the Standard Model (SM) of particle physics. This theory is based on the principle of local gauge invariance and successfully describes the electromagnetic, weak, and strong interactions within a single consistent framework. Over the past several decades, the predictions of the SM have been tested with high accuracy in a wide range of experiments, establishing it as one of the most successful theories in the history of science. A crucial element of the theory is the mechanism responsible for the electroweak symmetry breaking, which explains how elementary particles acquire mass without violating gauge invariance. The experimental confirmation of this mechanism came with the discovery of the Higgs boson, with a mass of approximately 125 GeV, at the Large Hadron Collider (LHC) in 2012 [2, 3], completing the particle content predicted by the SM.

Although the SM provides an accurate description of currently accessible phenomena, it is widely believed to be only an effective theory valid up to a certain energy scale. It does not incorporate gravity, it does not explain the origin of dark matter, and it leaves several fundamental questions unanswered, such as the structure of the scalar sector and the stability of the Higgs potential. For this reason, one of the main objectives of modern particle physics is to search for deviations from SM predictions that could indicate the presence of new physics. Such effects may appear either through the direct production of new particles or through small modifications of known processes that become visible in high-precision measurements.

The Higgs sector plays an important role in these studies. While the mass of the Higgs boson and many of its couplings have already been measured, the form of the scalar potential remains only partially tested experimentally. In the SM, the Higgs potential determines the strength of the Higgs boson self-interaction, which can be probed through processes involving one or more Higgs bosons. Direct measurements of the Higgs self-coupling modifier are experimentally difficult because they rely on rare processes with very small production rates. However, indirect sensitivity can also be obtained from precise studies of single-Higgs production and decay, where loop corrections depend on the structure of the scalar sector. These effects can be described using the Standard Model effective field theory framework, in which possible deviations from the SM are parameterized through higher-dimensional operators.

The LHC provides the highest-energy particle collisions currently available and makes it possible to study the Higgs boson with unprecedented detail. After the discovery of the Higgs boson during the first run of the LHC, the data collected in Run 2 and the ongoing Run 3 allow increasingly stringent tests of the SM. In this regime, the sensitivity of the measurements depends not only on the size of the dataset, but also on the precise understanding of detector performance, object reconstruction, and efficiency corrections. Final states containing electrons and muons are especially important, since they offer clean experimental signatures and allow accurate reconstruction of the event kinematics. Among these, the decay channel $H \to ZZ \to 4\ell$ provides one of the most powerful probes of the Higgs boson properties, owing to its excellent signal-to-background ratio.

The work presented in this thesis is performed within the Compact Muon Solenoid (CMS) experiment at the LHC and focuses on studies related to Higgs boson measurements in the four-lepton final state. It includes the determination of electron reconstruction and identification efficiencies, which are essential for accurate measurements in this channel, as well as an investigation of the sensitivity of off-shell single-Higgs processes to modifications of Higgs self-coupling within the Standard Model effective field theory framework. These studies contribute to the broader effort of testing the SM with increasing accuracy and exploring the possible presence of beyond Standard Model (BSM) physics.

This manuscript is structured as follows. Chapter 1 summarizes the theoretical framework relevant for this work. It introduces the SM of particle physics, the mechanism of electroweak symmetry breaking, and the main Higgs boson production and decay modes at the LHC. In addition, the Standard Model effective field theory framework is presented, which provides the theoretical basis for Higgs self-coupling measurement discussed in this thesis.

Chapter 2 describes experimental apparatus used to acquire data for this analysis. It summarizes the general detector design and the main subsystems. Data processing and simulation are also discussed.

Chapter 3 describes the treatment of electrons as physics objects in Compact Muon Solenoid, including triggering, reconstruction, identification, calculation of energy corrections and derivation of scale factors. Measurement of the reconstruction efficiency in the low-p_T regime is also discussed.

Chapter 5 describes the core analysis of this thesis: Higgs self-coupling measurement in the off-shell region. It includes detailed description of all analysis steps as well as overview of a broader set of off-shell measurements. Limits on Higgs self-coupling modifier are reported, using full Run 2 dataset.

Chapter 1

Theoretical Foundations of Particle Physics

This chapter summarizes the theoretical framework relevant for the analyses presented in this thesis. It begins with an overview of the SM of particle physics in Section 1.1, including its gauge structure, particle content, and the basic form of the Lagrangian describing the strong, weak, and electromagnetic interactions. The mechanism of electroweak symmetry breaking and the origin of particle masses are discussed in Section 1.2. This section introduces the scalar sector of the SM, spontaneous symmetry breaking, the generation of gauge-boson and fermion masses, and the properties of the physical Higgs boson. Section 1.3 describes the main Higgs boson production and decay processes at the LHC, together with their role in testing the SM and probing possible deviations from its predictions. Section 1.5 introduces the effective field theory framework used to parameterize possible effects of BSM physics. The general structure of the effective Lagrangian, the relevant higher-dimensional operators in the Higgs sector, and the conditions for the validity of the effective field theory expansion are discussed.

1.1 The Standard Model

The SM of particle physics is a QFT that describes the electromagnetic, weak, and strong interactions of elementary particles. It is based on the local gauge symmetry group

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{1.1}$$

where $SU(3)_C$ corresponds to quantum chromodynamics (QCD), while the electroweak interaction is described by the group $SU(2)_L \times U(1)_Y$ [4, 5, 6, 7].

The particle content of the SM consists of fermions, which form the matter fields, and gauge bosons, which mediate the interactions. The fermions are arranged in three generations with identical quantum numbers but different masses. Each generation contains two quarks and two leptons, organized in left-handed weak isospin doublets and right-handed singlets [7, 8]. For every fermion, there exists a corresponding antiparticle with the same mass but opposite quantum numbers, forming the antimatter sector of the SM. The first generation consists of the up and down quarks, the electron, and the electron neutrino. The second and third generations contain heavier copies of these particles, namely the charm and strange quarks, the muon and associated neutrino, and the top and bottom quarks together with the tau lepton and associated neutrino. All known matter is composed of first-generation fermions, while the heavier generations can be produced in high-energy collisions [7].

The interactions between fermions are mediated by gauge fields associated with the generators of the symmetry group. The eight gluon fields correspond to the $SU(3)_C$ symmetry and mediate the strong interaction. In the electroweak sector, the gauge symmetry $SU(2)_L \times U(1)_Y$ introduces three gauge fields W^1_μ, W^2_μ, and W^3_μ and one gauge field B_μ. These fields are not directly observable. After electroweak symmetry breaking, the charged combinations of W^1_μ and W^2_μ form the physical $W^\pm$ bosons, while the neutral fields W^3_μ and B_μ mix to produce the Z boson and the photon [9, 7].

The dynamics of the SM are defined by a Lagrangian that contains kinetic terms for the fermion and gauge fields, interaction terms determined by the gauge symmetry, and additional terms responsible for fermion masses and electroweak symmetry breaking. The mechanism that generates particle masses while preserving gauge invariance is provided by the Higgs mechanism [9, 7].

1.1.1 Fermions

The matter fields of the SM are spin-$\frac{1}{2}$ fermions, which are divided into two categories: quarks and leptons. They are arranged in three generations with identical quantum numbers under the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$, but with different masses [7, 9].

Each generation contains two quarks and two leptons. The left-handed components of the fermion fields transform as doublets under $SU(2)_L$, while the right-handed components transform as singlets. Quarks carry color charge and therefore participate in the strong interaction, while leptons are color neutral and do not couple to gluons.

The three generations of fermions are listed in Table 1.1. The first generation consists of the lightest particles and forms all stable matter observed in nature, while the heavier generations can be produced in high-energy collisions.

Generation	Quarks	Leptons	Electric charge
1	u, d	e, ν_e	$+\frac{2}{3}, -\frac{1}{3}, -1, 0$
2	c, s	μ, ν_μ	$+\frac{2}{3}, -\frac{1}{3}, -1, 0$
3	t, b	τ, ν_τ	$+\frac{2}{3}, -\frac{1}{3}, -1, 0$

Table 1.1: Fermion content of the SM grouped into three generations.

The interactions of fermions are determined by the gauge symmetry of the SM. Quarks interact through the strong, weak, and electromagnetic forces, while leptons interact only through the weak and electromagnetic interactions. The coupling of fermions to the gauge bosons is fixed by their gauge charges and is fully specified by the SM Lagrangian [8, 7].

1.1.2 Gauge Bosons

The interactions in the SM are mediated by gauge bosons associated with the generators of the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. Each symmetry gives rise to a set of vector gauge fields that couple to the fermions according to their gauge charges [9, 7].

The strong interaction is described by quantum chromodynamics (QCD), which is based on the non-Abelian gauge group $SU(3)_C$. This symmetry leads to eight massless gauge bosons, known as gluons, which carry color charge and interact with quarks as well as with each other.

The electroweak interaction is described by the gauge group $SU(2)_L \times U(1)_Y$. The corresponding gauge fields are three weak fields W_μ^1, W_μ^2, W_μ^3 associated with $SU(2)_L$, and the hypercharge field B_μ associated with $U(1)_Y$. These fields couple to left-handed fermion doublets and right-handed singlets according to their weak isospin and hypercharge quantum numbers.

After electroweak symmetry breaking, the weak and hypercharge gauge fields mix to form the physical electroweak bosons: the charged W^+ and W^- bosons, the neutral Z boson, and the photon γ. The photon mediates the electromagnetic interaction, while the $W^\pm$ and Z bosons mediate the weak interaction. The gluons remain massless and continue to mediate the strong interaction [8, 7].

The couplings of the gauge bosons to fermions and to each other are fixed by the gauge symmetry and are fully determined by the SM Lagrangian.

1.1.3 SM Lagrangian

The dynamics of the SM are determined by a Lagrangian that is invariant under the local gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. It contains kinetic terms for fermion and gauge fields, interaction terms dictated by the gauge symmetry, and additional terms responsible for fermion masses and electroweak symmetry breaking [9, 7].

The interaction between fermions and gauge bosons is introduced through the covariant derivative,

$$D_\mu = \partial_\mu - ig_s T^a G_\mu^a - ig\frac{\tau^i}{2}W_\mu^i - ig'\frac{Y}{2}B_\mu, \tag{1.2}$$

where G_μ^a, W_μ^i, and B_μ are the gauge fields associated with the groups $SU(3)_C$, $SU(2)_L$, and $U(1)_Y$, respectively, and g_s, g, and g' are the corresponding coupling constants. The matrices T^a and τ^i are the generators of the $SU(3)$ and $SU(2)$ groups, and Y denotes the weak hypercharge of the field [8, 9].

The kinetic and interaction terms of the fermion fields can be written as

$$\mathcal{L}_{\text{fermion}} = \sum_f \bar{\psi}_f i\gamma^\mu D_\mu \psi_f, \tag{1.3}$$

where the sum runs over all fermion fields. The gauge boson kinetic terms are given by

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} - \frac{1}{4}W_{\mu\nu}^i W^{i\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.4}$$

where the field strength tensors are defined in the usual way for non-Abelian gauge theories [9].

Fermion mass terms cannot be written directly in a gauge-invariant way using only the fermion fields. In the SM, fermion masses arise from Yukawa interactions with a scalar field, which lead to mass terms after electroweak symmetry breaking.

1.2 Higgs Mechanism

1.2.1 Scalar Field and Lagrangian

In the SM, the generation of particle masses while preserving gauge invariance is achieved through the Higgs mechanism. This mechanism introduces a complex scalar field that transforms as a doublet under the $SU(2)_L$ gauge group and carries weak hypercharge $Y = 1$ [10, 11, 12, 9].

The scalar field can be written as

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.5}$$

where ϕ^+ and ϕ^0 are complex scalar fields. The Lagrangian for the scalar sector contains a kinetic term and a potential term,

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi), \tag{1.6}$$

where the covariant derivative is defined as

$$D_\mu = \partial_\mu - ig\frac{\tau^i}{2}W^i_\mu - ig'\frac{Y}{2}B_\mu. \tag{1.7}$$

The most general renormalizable scalar potential consistent with the gauge symmetry is

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2, \tag{1.8}$$

where μ^2 and λ are real parameters.

1.2.2 Spontaneous Symmetry Breaking

The form of the scalar potential

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2 \tag{1.9}$$

depends on the sign of the parameter μ^2. For $\mu^2 < 0$ and $\lambda > 0$, the potential has a continuum of degenerate minima at non-zero values of the scalar field. This leads to spontaneous symmetry breaking, in which the ground state of the theory does not share the full gauge symmetry of the Lagrangian [9, 8]. The characteristic shape of the scalar potential is shown in Fig. 1.1.

The minimum of the potential is obtained for

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}, \tag{1.10}$$

where v is the vacuum expectation value of the scalar field. A convenient choice of the vacuum state is

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.11}$$

which preserves the electromagnetic $U(1)_{\mathrm{EM}}$ symmetry while breaking the electroweak symmetry.

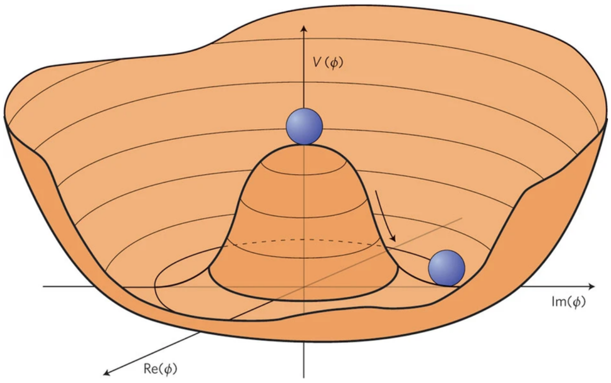


Figure 1.1: Shape of the Higgs potential for $\mu^2 < 0$, showing the degenerate minima at non-zero field value.

The scalar field can be expanded around the vacuum expectation value as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{1.12}$$

where $h(x)$ represents a real scalar field corresponding to a physical excitation around the vacuum. The remaining degrees of freedom are absorbed by the gauge bosons through the Higgs mechanism, providing the longitudinal polarization states of the massive vector bosons.

Spontaneous symmetry breaking therefore allows the gauge bosons and fermions to acquire mass without explicitly breaking gauge invariance.

1.2.3 Mass Generation for Gauge Bosons

After spontaneous symmetry breaking, the scalar field acquires a non-zero vacuum expectation value,

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.13}$$

which leads to mass terms for the electroweak gauge bosons through the kinetic term of the scalar field,

$$(D_\mu \Phi)^\dagger (D^\mu \Phi). \tag{1.14}$$

Substituting the vacuum expectation value into the covariant derivative generates quadratic terms in the gauge fields that can be interpreted as mass terms. The charged weak bosons are defined as

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right), \tag{1.15}$$

and their masses are given by

$$m_W = \frac{1}{2} g v. \tag{1.16}$$

The neutral gauge fields W_μ^3 and B_μ mix to form the physical Z boson and the photon,

$$\begin{aligned} Z_\mu &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu, \\ A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu, \end{aligned} \tag{1.17}$$

where θ_W is the weak mixing angle. The masses of the neutral bosons are

$$m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}, \qquad m_\gamma = 0. \tag{1.18}$$

The vacuum expectation value v is related to the Fermi constant G_F measured in muon decay,

$$v = \left(\sqrt{2} G_F\right)^{-1/2} \approx 246 \text{ GeV}. \tag{1.19}$$

As a result of the Higgs mechanism, the weak gauge bosons acquire mass while the photon remains massless, preserving the $U(1)_{\text{EM}}$ gauge symmetry. This mechanism provides a consistent description of electroweak interactions that is in agreement with precision experimental measurements [9, 8, 7].

1.2.4 Fermion Masses and Yukawa Couplings

In the SM, fermion mass terms cannot be written directly in a gauge-invariant way using only the fermion fields. A Dirac mass term of the form

$$m \bar{\psi} \psi \tag{1.20}$$

mixes left- and right-handed components of the fermion field, which transform differently under the $SU(2)_L \times U(1)_Y$ gauge symmetry. As a result, such a term would explicitly break the gauge invariance of the theory [9, 8].

Fermion masses are generated through Yukawa interactions between the fermion fields and the scalar doublet. The gauge-invariant Yukawa Lagrangian for a fermion field can be written as

$$\mathcal{L}_{\text{Yukawa}} = -y_f \, \bar{\psi}_L \Phi \psi_R + \text{h.c.}, \tag{1.21}$$

where y_f is the Yukawa coupling constant, ψ_L and ψ_R are the left- and right-handed fermion fields, Φ is the scalar doublet, and h.c. denotes the Hermitian conjugate of the preceding term, which is added to ensure that the Lagrangian is Hermitian. After spontaneous symmetry breaking, the scalar field acquires a vacuum expectation value, and the Yukawa interaction generates a fermion mass term,

$$m_f = \frac{y_f v}{\sqrt{2}}. \tag{1.22}$$

In this way, the masses of all charged fermions arise from their coupling to the scalar field. The values of the Yukawa couplings are free parameters of the SM and must be determined experimentally. The large hierarchy of fermion masses corresponds to a wide range of Yukawa coupling strengths.

Neutrinos are massless in the minimal SM, since no right-handed neutrino fields are included. The observation of neutrino oscillations indicates that neutrinos have non-zero masses, which may require a different mass generation mechanism [7].

The Yukawa interactions also determine the couplings of the scalar boson to fermions, which are proportional to the fermion masses. These couplings play an important role in Higgs boson production and decay processes studied at the Large Hadron Collider.

1.2.5 Higgs Boson and Scalar Potential

Substituting expression 1.12 into the scalar potential 1.9, and expanding around the vacuum expectation value leads to

$$V(h) = \frac{1}{2}m_h^2 h^2 + \lambda v h^3 + \frac{\lambda}{4}h^4, \tag{1.23}$$

where the Higgs boson mass is given by

$$m_h^2 = 2\lambda v^2. \tag{1.24}$$

Rewriting the potential in terms of the Higgs self-couplings,

$$V(h) = \frac{1}{2}m_h^2 h^2 + \frac{\lambda_3}{3!}h^3 + \frac{\lambda_4}{4!}h^4, \tag{1.25}$$

one identifies the trilinear and quartic couplings as

$$\lambda_3 = 6\lambda v = \frac{3m_h^2}{v}, \qquad \lambda_4 = 6\lambda = \frac{3m_h^2}{v^2}. \tag{1.26}$$

The first term in Eq. 1.23 corresponds to the mass of the scalar particle, while the higher-order terms describe self-interactions of the Higgs field. The couplings of the Higgs boson to fermions and gauge bosons are proportional to the particle masses and are therefore fixed once the vacuum expectation value and the Higgs boson mass are known [9, 8, 7].

The interactions of the Higgs boson with fermions arise from the Yukawa terms and are given by

$$g_{hff} = \frac{m_f}{v}, \tag{1.27}$$

while the couplings to the weak gauge bosons are

$$g_{hWW} = \frac{2m_W^2}{v}, \qquad g_{hZZ} = \frac{2m_Z^2}{v}. \tag{1.28}$$

These relations are a direct consequence of the Higgs mechanism and represent a key prediction of the SM. Precise measurements of the Higgs boson production and decay rates therefore provide stringent tests of the scalar sector of the theory.

1.3 Higgs Boson Production and Decay at the LHC

At the LHC, Higgs bosons are produced through several mechanisms with markedly different cross sections and event topologies [13, 14, 15, 7]. The dominant production mode is gluon fusion, while vector boson fusion, associated production with a vector boson, and production in association with top quarks provide experimentally distinct signatures and sensitivity to different Higgs boson couplings to gauge bosons and fermions [13, 14, 15, 7]. A schematic overview of the main production processes is shown in Figs. 1.2–1.5.

1.3.1 Higgs Boson Production Modes

Gluon Fusion

Gluon fusion (ggF) is the dominant Higgs boson production mechanism at the LHC [15, 13, 7]. At leading order, the Higgs boson is produced through a heavy-quark loop, with the top-quark contribution providing the largest amplitude because of the large top Yukawa coupling [15, 13]. Although the process is loop-induced, its large cross section makes it the primary source of inclusive Higgs boson events at hadron colliders. The corresponding leading-order Feynman diagram is shown in Fig. 1.2.

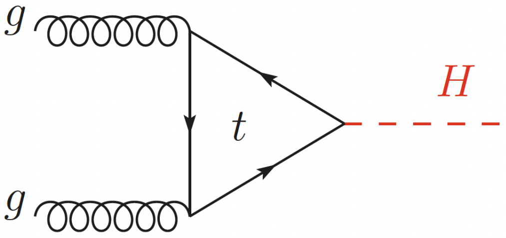


Figure 1.2: Leading-order Feynman diagram of Higgs boson production through ggF [16].

Vector Boson Fusion

Vector boson fusion (VBF) proceeds through the fusion of electroweak gauge bosons radiated from the incoming quarks [15, 13]. This production mode is characterized by two energetic forward jets with a large rapidity separation and relatively little hadronic activity in the central region [15, 13, 7]. These features make VBF experimentally important despite its smaller cross section compared to gluon fusion. A leading-order Feynman diagram for this process is shown in Fig. 1.3.

Associated Production with a Vector Boson

In associated production with a vector boson, usually denoted VH, the Higgs boson is produced together with a W or Z boson [14, 13]. Although the production rate is lower than that of gluon fusion, the presence of a leptonically decaying vector boson can significantly improve event selection and background rejection [14, 13, 7]. This production mode plays an important role in measurements of Higgs boson couplings and in studies of decay channels with large backgrounds. A leading-order Feynman diagram for VH production is shown in Fig. 1.4.

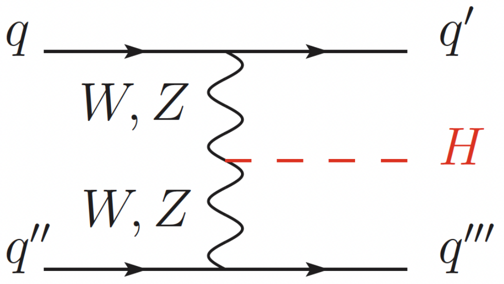


Figure 1.3: Leading-order Feynman diagram of Higgs boson production through VBF [16].

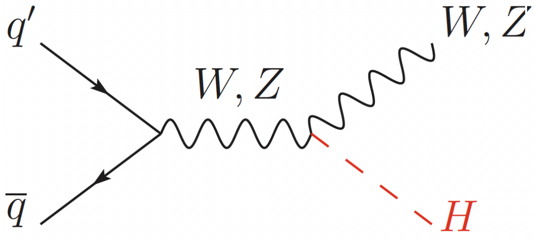


Figure 1.4: Leading-order Feynman diagram of Higgs boson production through VH process [16].

Associated Production with Top Quarks

The production of a Higgs boson in association with a top-quark pair (ttH) directly probes the top Yukawa coupling [13, 7]. Its cross section is smaller than those of ggF and VBF, but the process is of particular importance because it gives direct access to the interaction between the Higgs boson and the top quark. Single-top associated Higgs production is also possible, though with an even smaller rate, as shown in Fig. 1.5.

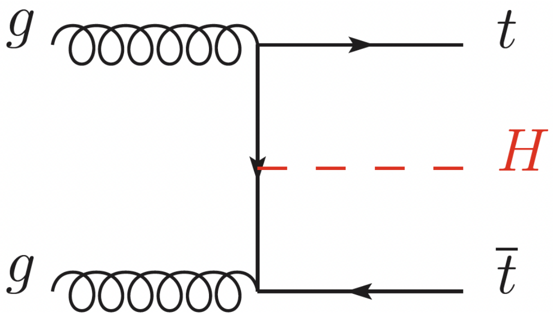


Figure 1.5: LO Feynman diagram of Higgs boson production via ttH [16].

The relative contributions of the different production mechanisms at the LHC are shown in Fig. 1.6. Gluon fusion dominates the total production cross section, while vector boson fusion, associated production with vector bosons, and production in association with top quarks provide smaller but experimentally important contributions [13].

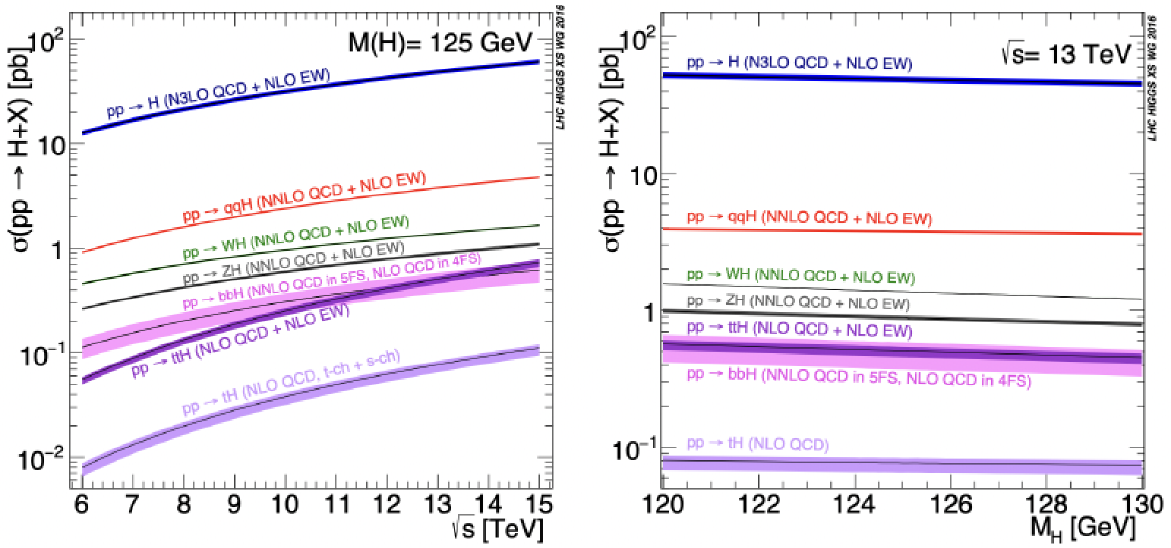


Figure 1.3: Higgs boson production cross section as a function of the centre-of-mass energy $\sqrt{s}$ (left), and as a function of the m_H (right), for different production mechanisms[20]

.

Figure 1.6: Higgs boson production cross section as a function of the centre-of-mass energy $\sqrt{s}$ (left), and as a function of the Higgs boson mass m_H (right), for different production mechanisms [13].

1.3.2 Higgs Boson Decay Modes

For a Higgs boson mass of approximately 125 GeV, the SM predicts a rich decay pattern into fermions and gauge bosons. The dominant decay mode is $H \to b\bar{b}$, followed by decays to pairs of W bosons, gluons, tau leptons, charm quarks, and Z bosons. Although the branching fraction for $H \to \gamma\gamma$ is small, this decay channel is experimentally important because of its clean final-state signature and excellent mass resolution [13, 15, 7]. Similarly, the decay channel $H \to ZZ \to 4\ell$ has a small branching fraction but provides clean experimental signature. Table 1.2 summarizes representative SM branching ratios for a Higgs boson with mass 125.09 GeV [1].

The experimental sensitivity to Higgs boson production and decay depends on a balance between production rate, branching fraction, detector performance, and background composition. As a result, channels with smaller branching fractions may still play a central role in precision measurements if they provide clean experimental signatures. This is the case for the diphoton and four-lepton final states, which remain among the most powerful channels for studies of Higgs boson properties at the LHC [13, 7].

---

[1] Measurements of the Higgs boson mass evolve over time as more data are collected and analyses are refined. Earlier combined results yield a value of $m_H \approx 125.09$ GeV [17], while more recent CMS measurements report values around $m_H \approx 125.38$ GeV [18]

Decay mode	$\mathcal{B}$ [%]
$H \to b\bar{b}$	$58.09^{+0.72}_{-0.73}$
$H \to W^+W^{-*}$	21.53 ± 0.33
$H \to gg$	8.18 ± 0.42
$H \to \tau^+\tau^-$	6.27 ± 0.10
$H \to c\bar{c}$	$2.88^{+0.16}_{-0.06}$
$H \to ZZ^*$	2.641 ± 0.040
$H \to \gamma\gamma$	0.2270 ± 0.0047
$H \to Z\gamma$	0.1541 ± 0.0090
$H \to \mu^+\mu^-$	$0.02171^{+0.00036}_{-0.00037}$

Table 1.2: Branching fractions of the major Higgs boson decay modes for $m_H = 125.09$ GeV. Theoretical uncertainties include contributions from the Higgs boson partial widths, the value of α_s, and the quark masses [13].

1.3.3 Sensitivity of Higgs Processes to BSM Physics

Precision measurements of Higgs boson production and decay provide a powerful test of the SM and offer sensitivity to new physics at energy scales beyond the direct reach of the LHC. Since the couplings of the Higgs boson to fermions and gauge bosons are predicted by the SM in terms of particle masses, any deviation from the expected production cross sections or branching fractions could indicate the presence of additional interactions or particles [15, 13, 7].

New physics effects can modify Higgs boson observables through virtual corrections, even if the new particles are too heavy to be produced directly. Such effects can change the rates of single-Higgs production and decay, alter kinematic distributions, or affect processes involving multiple Higgs bosons. Loop-induced processes, such as gluon fusion production and the decay $H \to \gamma\gamma$, are particularly sensitive to contributions from heavy particles because they occur through quantum loops already in the SM [15].

A consistent and model-independent way to describe possible deviations from SM predictions is provided by the effective field theory (EFT) framework, in which the SM Lagrangian is extended by higher-dimensional operators suppressed by the scale of new physics [19, 20]. In this approach, modifications of Higgs boson interactions can be parametrized through additional terms in the Lagrangian, allowing experimental measurements to be interpreted without assuming a specific underlying theory.

1.4 Off-shell Region and Virtual Particles

In QFT, interactions are described by Feynman diagrams, which represent the sum over all possible time orderings of a given process. While the external lines correspond to physical initial and final states, the internal lines represent intermediate contributions to the amplitude. These internal contributions are interpreted as the exchange of virtual particles.

Such contributions arise from the perturbative expansion of the theory, in which observables are expressed as a series in the interaction strength (coupling constant). Each term in this expansion corresponds to Feynman diagrams of increasing complexity, and the full amplitude is obtained as a sum over diagrams with different numbers of interaction vertices and loops.

Virtual particles are not physical, observable states but rather mathematical constructs emerging from this perturbative expansion. They carry four-momentum fixed by momentum conservation at interaction vertices, but they are not required to satisfy the mass–energy relation of real particles. Their invariant mass q^2 is determined by the kinematics of the external particles and can differ from the physical mass of the corresponding particle.

In the SM, the physical mass of a particle is defined by the pole of its renormalized propagator and corresponds to the mass measured in experiments. A particle is said to be on-shell when its invariant mass equals this pole mass, while contributions at invariant masses different from the pole mass are referred to as off-shell. Virtual particles appearing as internal lines are therefore generically off-shell and should be understood as intermediate quantum fluctuations rather than directly observable states. However, in the vicinity of a propagator pole, the contribution is dominated by configurations where the invariant mass is close to the physical mass, and the intermediate state behaves as a resonant, on-shell particle.

An important example is provided by Higgs boson mediated processes, in which the Higgs boson appears as an internal propagator. When the invariant mass of the final-state system is close to the Higgs pole mass, $m_{VV} \simeq m_H$, the contribution of the Higgs propagator is enhanced. In this region, the process is referred to as on-shell Higgs production. For invariant masses away from the pole, the contribution is smaller and corresponds to off-shell Higgs production.

1.4.1 Interference in the Off-shell Region

In the $ZZ \to 4\ell$ final state, the gluon-induced production receives contributions from both the ggF signal and $gg \to ZZ$ background. The total amplitude can therefore be written as

$$\mathcal{A}_{gg\to ZZ} = \mathcal{A}_{\mathrm{ggF}} + \mathcal{A}_{ggZZ}, \tag{1.29}$$

and the corresponding cross section contains an interference term,

$$|\mathcal{A}_{gg\to ZZ}|^2 = |\mathcal{A}_{\mathrm{ggF}}|^2 + |\mathcal{A}_{ggZZ}|^2 + 2\,\mathrm{Re}\big(\mathcal{A}_{\mathrm{ggF}}\mathcal{A}^*_{ggZZ}\big)\,. \tag{1.30}$$

The ggF contribution is proportional to the Higgs propagator,

$$\mathcal{A}_{\mathrm{ggF}}(\hat{s}) \propto \frac{1}{\hat{s} - m_H^2 + i m_H \Gamma_H}, \tag{1.31}$$

where $\hat{s} = m_{ZZ}^2$ and Γ_H denotes the total decay width of the Higgs boson, which characterizes its finite lifetime and determines the spread of the resonance around the pole mass.

The behavior of the interference term is governed by the real and imaginary parts of this propagator. Near the Higgs pole, $\hat{s} \simeq m_H^2$, the propagator is dominated by its imaginary part, and the ggF amplitude is approximately imaginary. Since the $gg \to ZZ$ background amplitude is predominantly real in this region, the interference term is suppressed.

In the off-shell region, $\hat{s} \gg m_H^2$, the propagator becomes predominantly real,

$$\frac{1}{\hat{s} - m_H^2 + i m_H \Gamma_H} \simeq \frac{1}{\hat{s} - m_H^2}, \tag{1.32}$$

and the ggF amplitude is correspondingly real. The interference term is then dominated by the product of the real parts of the ggF and $gg \to ZZ$ amplitudes. In this region, these contributions have opposite signs over a large part of phase space, leading to destructive interference. As a result, the interference term significantly reduces the total cross section and becomes comparable in magnitude to the ggF contribution at high invariant masses.

1.4.2 The Higgs Boson Lineshape in the Off-shell Region

In the case of the Higgs boson, although its pole mass is approximately $m_H \approx 125$ GeV, virtual Higgs exchange can contribute at significantly higher invariant masses. In particular, the region above the ZZ production threshold provides access to Higgs-mediated amplitudes far from the resonance peak.

This behaviour is illustrated in Fig. 1.7, from Passarino *et al.* [21]. The black curve, corresponding to the $gg \to H \to ZZ$ process, exhibits a clear enhancement at the $2m_Z$ threshold, where both Z bosons can be produced on-shell. A second enhancement appears near $2m_t$, reflecting the kinematic threshold associated with the top-quark loop in gluon-fusion production.

Another important feature of this region is strong destructive interference between signal and continuum background. This is depicted on Figure 1.8 where signal+background contribution (in cyan) sits higher then signal+background+interference (in black). When the continuum $gg \to ZZ$ background is included, the structure near the $2m_t$ threshold becomes significantly less pronounced. After passing the $2m_Z$ threshold, invariant mass tends to drop in a more slowly-varying regime, which starts at ~ 220 GeV. For this reason, $m_{4\ell} \geq 220$ GeV is taken as a definition of off-shell region.

Off-shell Higgs boson production accounts to approximately 14% of overall Higgs events. Although suppressed relative to the on-shell contribution, this high-mass regime provides unique sensitivity to Higgs properties through interference effects and potential modifications of couplings. Inclusion of this region also opens up a phase-space for higher order-loop effects. This is exploited in the κ_λ measurement, where addition of SMEFT effects modify the Higgs lineshape at $\sim 2m_H$.

1.5 Effective Field Theory

1.5.1 Motivation for Effective Field Theory

The SM provides an accurate description of particle interactions up to the energy scales that have been probed experimentally. However, it is widely expected to be an effective theory valid only up to a certain energy scale, beyond which additional degrees of freedom or new interactions may appear. Possible extensions of the SM are motivated by several open questions, including the hierarchy problem, the nature of dark matter, the origin of neutrino masses, and the absence of a quantum theory of gravity [7, 20].

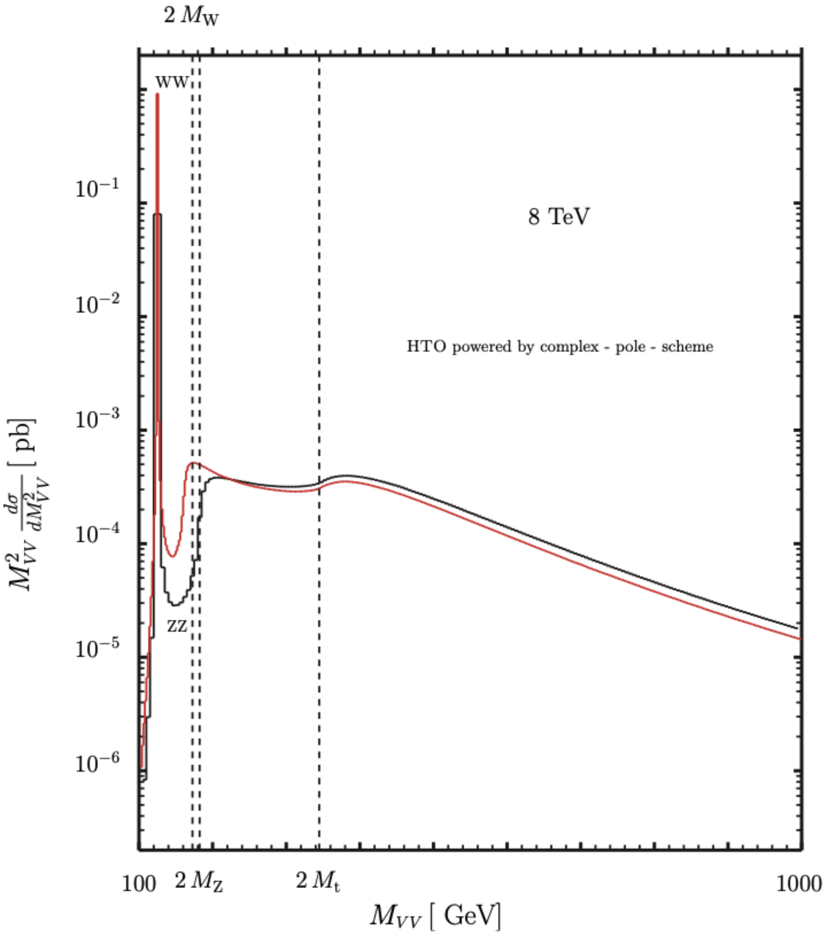


Figure 1.7: $gg \to H \to VV$ signal invariant mass distributions for ZZ (black) and WW (red). Dashed lines, from left to right, indicate the $2m_W$, $2m_Z$, and $2m_t$ thresholds. Numerical results are obtained with the program HTO (G. Passarino, unpublished) that allows for the study of the Higgs boson lineshape, in gluon-gluon fusion, using complex poles. Results are obtained for $\sqrt{s} = 8$ TeV [21].

If new particles exist at energy scales larger than those directly accessible at the Large Hadron Collider, their effects can still be observed indirectly through deviations from the SM predictions. A convenient framework to describe such effects is provided by effective field theory (EFT) [22, 19].

In the EFT approach, the known particles are treated as the low-energy degrees of freedom, while the effects of heavy states are encoded in additional interaction terms in the Lagrangian. These terms are organized as an expansion in inverse powers of a characteristic scale of new physics Λ, which represents the energy scale at which the effective description breaks down.

Effective field theories allow precision measurements at relatively low energies to be interpreted in a consistent theoretical framework without specifying the details of the underlying high-energy theory. This approach is widely used in the interpretation of Higgs boson measurements, where small deviations from SM predictions may arise from virtual contributions of heavy particles.

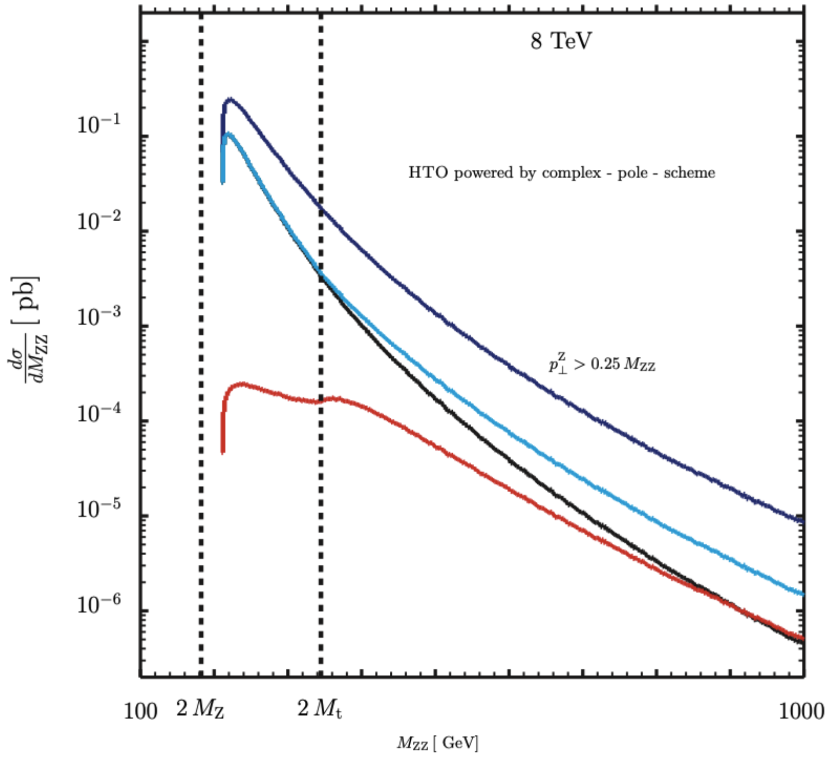


Figure 1.8: Invariant mass distributions for ZZ processes. $gg \to H \to ZZ$ signal is shown in red. $gg \to ZZ$ total contribution is shown in black, $gg \to ZZ$ signal + background is shown in cyan. The blue line includes the $qq \to ZZ$. Transverse momentum cut $p_T^Z > 0.25 M_{ZZ}$ is applied on Z bosons. Numerical results are obtained with the program HTO (G. Passarino, unpublished) that allows for the study of the Higgs boson lineshape, in gluon-gluon fusion, using complex poles. Results are obtained for $\sqrt{s} = 8$ TeV [21].

1.5.2 General Effective Field Theory Lagrangian

In an effective field theory, the effects of heavy particles with masses much larger than the energy scale of the experiment are described by additional interaction terms involving only the light degrees of freedom. These heavy states are integrated out, leading to a Lagrangian that contains the SM terms together with higher-dimensional operators suppressed by powers of the new physics scale Λ [22, 19, 20].

The effective Lagrangian can be written as an expansion in inverse powers of Λ,

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{C_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i \frac{C_i^{(7)}}{\Lambda^3}\mathcal{O}_i^{(7)} + \cdots , \tag{1.33}$$

where $\mathcal{L}_{\text{SM}}$ is the SM Lagrangian, $\mathcal{O}_i^{(d)}$ are gauge-invariant operators of mass dimension $d > 4$, and $C_i^{(d)}$ are dimensionless Wilson coefficients that encode the strength of the corresponding interactions.

The operators in the effective Lagrangian must respect the symmetries of the low-energy theory. In the case of the SM effective field theory, the operators are constructed to be invariant under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry. The lowest-dimension operator beyond the SM has dimension five and leads to Majorana neutrino masses, while the leading contributions to most collider observables arise from dimension-six operators [20, 7].

The Wilson coefficients depend on the details of the underlying high-energy theory and can in general be complex. Experimental measurements can be used to constrain these coefficients, providing indirect information about the scale and structure of new physics.

1.5.3 Standard Model Effective Field Theory

In the case where no additional light particles are present beyond those of the SM, possible deviations from SM predictions can be described within the Standard Model Effective Field Theory (SMEFT) [19, 23, 20]. In this framework, the SM Lagrangian is extended by higher-dimensional operators constructed from SM fields and required to respect the full $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry.

The effects of heavy new physics appear through these operators, which modify amplitudes either at loop level or through interference with the SM contributions, leading to changes in event rates and kinematic distributions. Such effects can therefore be probed through precision measurements.

A complete set of dimension-six operators consistent with the SM symmetries [23] is commonly referred to as the Warsaw basis. The corresponding Wilson coefficients parameterize the strength of possible deviations from the SM and can be measured experimentally.

In Higgs boson studies, SMEFT provides a framework to describe modifications of Higgs couplings and of the scalar potential. Operators affecting the Higgs self-interaction can induce corrections not only to double-Higgs production, but also to single-Higgs processes through higher-order contributions. This approach is used in this thesis to study the sensitivity of single-Higgs observables to modifications of the Higgs self-coupling.

1.5.4 Higher-dimensional Operators in the Higgs Sector

In the SMEFT framework, higher-dimensional operators can modify the interactions of the Higgs field with fermions, gauge bosons, and with itself. These operators are constructed from the SM fields and must respect the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry. The leading contributions to Higgs boson observables at the LHC typically arise from dimension-six operators, which are suppressed by two powers of the new physics scale Λ [19, 20].

Operators involving the scalar doublet can affect both the couplings of the Higgs boson to other particles and the structure of the scalar potential. Examples of dimension-six operators containing the Higgs field include

$$\mathcal{O}_H = (\Phi^\dagger \Phi)^3, \qquad \mathcal{O}_{\Phi\Box} = (\Phi^\dagger \Phi)\,\Box\,(\Phi^\dagger \Phi), \qquad \mathcal{O}_{\Phi D} = \left(\Phi^\dagger D_\mu \Phi\right)^* \left(\Phi^\dagger D^\mu \Phi\right). \tag{1.34}$$

where Φ denotes the scalar doublet and D_μ is the covariant derivative. Such operators modify the normalization of the Higgs field, the couplings to gauge bosons, and the form of the scalar potential after electroweak symmetry breaking.

Additional operators involving fermions can change the Yukawa interactions, leading to deviations in Higgs boson production and decay rates. Similarly, operators containing gauge field strengths can alter loop-induced processes such as gluon fusion production and the decay $H \to \gamma\gamma$, which are particularly sensitive to new physics contributions [15, 20].

The presence of higher-dimensional operators therefore leads to measurable deviations in Higgs boson observables, including production cross sections, branching fractions, and kinematic distributions. Precision measurements of these quantities provide constraints on the corresponding Wilson coefficients and can probe energy scales well above those directly accessible at the LHC.

Modifications of the scalar potential are of special interest, since they affect interactions involving multiple Higgs bosons and can also induce corrections to single-Higgs processes through higher-order contributions.

1.5.5 Effects of Higher-dimensional Operators in Higgs Boson Production and Decay

Higher-dimensional operators in the effective field theory framework can modify Higgs boson observables in several ways. These modifications may affect production cross sections, decay branching fractions, and kinematic distributions, and can therefore be probed through precision measurements at the LHC. Since the SM predicts the Higgs boson couplings in terms of particle masses, any deviation from the expected rates can provide indirect evidence for new physics at energy scales larger than those directly accessible in collider experiments [20, 13, 7].

Single-Higgs production processes are sensitive to new physics through both tree-level and loop-level contributions. The dominant production mode at the LHC, gluon fusion, occurs through a heavy-quark loop already in the SM and is therefore particularly sensitive to modifications of the top Yukawa coupling or to the presence of additional heavy particles in the loop. Similarly, the decay channel $H \to \gamma\gamma$ is loop-induced and can receive contributions from new charged particles, making it an important probe of physics beyond the SM [15].

Higher-dimensional operators can also modify the couplings of the Higgs boson to the electroweak gauge bosons, affecting vector boson fusion and associated production processes. These effects can change both the total production rates and the kinematic properties of the final state, allowing experimental measurements to constrain the corresponding Wilson coefficients.

Interactions involving multiple Higgs bosons provide additional sensitivity to the structure of the scalar sector. Operators that modify the scalar potential lead to changes in processes with two or more Higgs bosons in the final state, but they can also contribute to single-Higgs observables through higher-order corrections. As a result, precision measurements of single-Higgs production and decay can provide indirect constraints on interactions that are directly accessible only in multi-Higgs production.

The interpretation of experimental measurements in terms of Wilson coefficients requires a consistent treatment of higher-order corrections and a careful consideration of the validity range of the effective field theory.

1.5.6 Validity of the Effective Field Theory Approach

The effective field theory description relies on the assumption that the scale of new physics Λ is significantly larger than the energy scale of the processes under consideration. Under this condition, the effects of heavy particles can be represented by a series of higher-dimensional operators suppressed by powers of Λ, and the expansion in Eq. (5.1) can be truncated after a finite number of terms [20, 13].

In most applications to Higgs boson physics at the LHC, the effective Lagrangian is truncated at dimension six, keeping only the leading corrections beyond the SM. This approximation is justified if the typical energy scale E of the process satisfies

$$E \ll \Lambda, \tag{1.35}$$

so that contributions from operators of dimension eight and higher are suppressed. If the energy of the interaction becomes comparable to the new physics scale, the expansion may no longer converge and the effective field theory description can lose its validity [20].

Another important requirement is that the Wilson coefficients remain within the perturbative regime. Large values of the coefficients can lead to unphysical results such as violation of unitarity at high energies. For this reason, constraints derived within the EFT framework must be interpreted together with assumptions about the scale of new physics and the range of validity of the expansion.

1.6 Higgs Boson Self-coupling

The measurement of the Higgs boson self-coupling is a central objective of the current collider physics program, as it provides direct access to the structure of the scalar sector and sensitivity to BSM physics. Experimentally, deviations from the SM prediction are commonly parameterized in the κ-framework by the Higgs self-coupling modifier

$$\kappa_\lambda = \frac{\lambda_3}{\lambda_3^{\mathrm{SM}}}, \tag{1.36}$$

which quantifies the ratio of the trilinear Higgs coupling to its SM value.

The most direct probe of κ_λ is Higgs boson pair production. An example of such a process is shown in Fig. 1.9. The corresponding amplitude is proportional to the product of the top Yukawa coupling modifier κ_t and κ_λ. As a result, measurements of the double-Higgs production cross section can be used to constrain κ_λ.

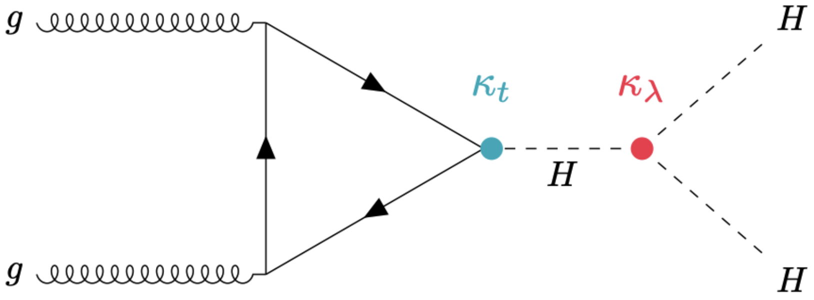


Figure 1.9: Double Higgs production via the ggF process [24].

However, the double-Higgs production cross section at the LHC is very small, making a direct measurement of κ_λ experimentally challenging even with the large integrated luminosity collected by CMS.

In addition to Higgs boson pair production, κ_λ affects single-Higgs production and decay processes through higher-order electroweak corrections. Although this dependence is indirect, the much larger single-Higgs production rates provide additional statistical power. Precision measurements in single-Higgs channels can therefore place competitive constraints on κ_λ, providing a complementary probe of the Higgs self-coupling.

One such measurement, performed in the on-shell Higgs region, is described in Sec. 4.6. In this thesis, κ_λ is also probed through off-shell single-Higgs production by incorporating SMEFT corrections. This measurement is performed here for the first time and is presented in Sec. 5.5.

All these complementary experimental approaches are well motivated by the fundamental importance of determining the Higgs self-coupling. From a theoretical perspective, the measurement of κ_λ is essential for several reasons:

- **Higgs potential and electroweak symmetry breaking:** As introduced in Eq. 1.23, the Higgs self-coupling directly determines the shape of the scalar potential after electroweak symmetry breaking. In the SM, the Higgs self-interactions is uniquely fixed by the Higgs boson mass. A measurement of the Higgs self-coupling provides direct experimental access to the structure of the Higgs potential and constitutes a fundamental test of the mechanism of electroweak symmetry breaking.

- **Sensitivity to heavy new physics:** Within EFT, the effects of heavy new physics are described by higher-dimensional operators. The dimension six operator modifies the Higgs self-coupling already at leading order, while its impact on single-Higgs processes typically appears only at loop level. Consequently, sizeable deviations in κ_λ can arise even when other Higgs observables remain consistent with their SM predictions, making the Higgs self-coupling a uniquely sensitive probe of high-scale physics.

- **Electroweak vacuum stability:** The Higgs self-coupling is directly related to the quartic coupling, whose renormalization group evolution determines the stability of the electroweak vacuum. Depending on its behavior at high energy scales, the vacuum may be stable, metastable, or unstable. Precise knowledge of the Higgs self-coupling at the electroweak scale provides indirect information about the behavior of the theory at energy scales far beyond those accessible in collider experiments.

- **Electroweak phase transition and cosmology:** The Higgs potential governs the nature of the electroweak phase transition in the early Universe. In the SM, this transition is a smooth crossover, meaning that the Higgs field evolves continuously from the symmetric phase to the broken phase without the presence of a potential barrier separating distinct minima. In contrast, many extensions of the SM predict a strongly first-order phase transition, in which a barrier develops in the Higgs potential and the transition proceeds via bubble nucleation. Such dynamics provide the necessary conditions for electroweak baryogenesis. These modifications of the Higgs potential are typically accompanied by sizeable deviations in the Higgs self-coupling.

- **Explicit BSM scenarios:** A wide range of extensions of the Standard Model predict sizeable deviations in the Higgs self-coupling, including singlet extensions, two-Higgs-doublet models, supersymmetric models, and composite Higgs scenarios. In many of these models, κ_λ can deviate at $\mathcal{O}(1)$, even when modifications of other Higgs couplings remain comparatively small. This makes the Higgs self-coupling a particularly powerful observable for distinguishing between different new physics scenarios.

1.7 Anomalous Couplings

In different BSM scenarios, Higgs boson couplings to vector bosons can also deviate from the SM predictions. The scattering amplitude of the Higgs boson with two vector bosons can be written as

$$\begin{aligned}\mathcal{A}(HVV) \propto & \left[a_1^{VV} + \frac{\kappa_1^{VV} q_1^2 + \kappa_2^{VV} q_2^2}{(\Lambda_1^{VV})^2} + \frac{\kappa_3^{VV} (q_1+q_2)^2}{(\Lambda_Q^{VV})^2} \right] m_{V_1}^2 \, \epsilon_{V_1}^* \epsilon_{V_2}^* \\ & + a_2^{VV} \, f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + a_3^{VV} \, f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu} .\end{aligned} \tag{1.37}$$

Here, ϵ_i and q_i denote the polarization vectors and four-momenta of the vector bosons V_i, respectively. The tensor

$$f_{\mu\nu}^{(i)} = \epsilon_i^\mu q_i^\nu - \epsilon_i^\nu q_i^\mu \tag{1.38}$$

represents the field-strength tensor associated with the vector boson, while $\tilde{f}_{\mu\nu}$ denotes its dual.

The coefficients a_1^{VV}, a_2^{VV}, and a_3^{VV} parameterize different tensor structures of the Higgs–vector boson interaction. The term proportional to a_1^{VV} corresponds to the SM coupling. The a_2^{VV} term represents a CP-even anomalous interaction arising from higher-dimensional operators, while a_3^{VV} corresponds to a CP-odd (CP-violating) interaction.

The parameters κ_i^{VV} correspond to momentum-dependent corrections to the SM interaction, suppressed by the energy scales Λ_1^{VV} and Λ_Q^{VV}, which characterize the scale of possible BSM physics.

Contributions of differnet a_i^{VV} terms are usually expressed via fractional contributions f_{a_i} and phase ϕ_{a_i}, which are defined as:

$$f_{a_i} = \frac{|a_i|^2\,\sigma_i}{\sum_{j=1,2,3,\ldots}|a_j|^2\sigma_j}, \qquad \phi_{a_i} = \arg\left(\frac{a_i}{a_1}\right). \tag{1.39}$$

Following this convention, pure SM behaviour would be expressed as $f_{a_1} = 1$. This can be extended to anomalous a_1^{VV} and a_2^{VV}. For example, BSM CP-even contribution to the cross-section of 50% and SM contribution of 50% would be expressed as $f_{a_2} = 0.5$ and $f_{a_1} = 0.5$.

Chapter 2

The CMS Experiment at the LHC

This chapter provides an overview of the experimental framework in which the measurements presented in this thesis are performed. It begins with a brief historical overview of the European Organization for Nuclear Research (CERN) and the physics motivation for high-energy collider experiments in Section 2.1, followed by a description of the LHC and its accelerator complex in Section 2.2. The main experiments operating at the LHC are briefly introduced in Section 2.3.

The CMS detector, which is the primary instrument used for the data recorded for this analysis, is described in Section 2.4. This section summarizes the general detector design and the main subsystems responsible for tracking, calorimetry, muon detection, and event selection, together with the trigger and data acquisition system that reduces the event rate to a manageable level. The basic principles of particle detection and reconstruction in CMS are also outlined. Muons and jets are introduced, as they are important physics objects in the analysis. A dedicated chapter (Chapter 3) is devoted to electron reconstruction and identification, as a substantial part of this thesis focuses on their study.

The processing of recorded data and the different data formats used in CMS analyses are discussed in Section 2.5. Finally, the simulation of proton–proton collisions and the detector response is described in Section 2.6, including the Monte Carlo generators, detector simulation, and reconstruction of simulated events. These elements together define the experimental and computational environment required for the analyses presented in the following chapters.

2.1 Short history of CERN and Physics Motivation

CERN was founded in 1954 with the goal of promoting scientific collaboration in Europe and advancing the understanding of fundamental physics. The creation of CERN was motivated both by the need for large-scale research facilities that could not be realized by individual countries alone and by the desire to foster peaceful scientific cooperation after the Second World War. Since its establishment, CERN has developed into one of the world's leading laboratories for particle physics, operating a sequence of increasingly powerful accelerators that have enabled major discoveries about the structure of matter and the fundamental forces of nature [25, 26].

The main scientific objective of experiments at CERN is the study of elementary particles and their interactions, which are described by SM of particle physics. Despite its remarkable predictive power, the SM does not provide answers to several fundamental questions, such as the origin of the mass hierarchy, the nature of dark matter, or the unification of forces. These open problems motivate the construction of particle accelerators capable of reaching higher energies and probing shorter distance scales.

Over the decades, CERN has operated a series of accelerator facilities, each extending the accessible energy range. One of the early milestones was the Proton Synchrotron (PS), which began operation in 1959 and enabled precision studies of hadronic interactions. It was followed by the Super Proton Synchrotron (SPS), commissioned in 1976, which later operated as a proton–antiproton collider. Experiments at SPS, operated in proton–antiproton collider mode as the $Sp\bar{p}S$, led to the discovery of the W and Z bosons in 1983 by the UA1 and UA2 collaborations, providing experimental confirmation of the electroweak theory developed by Glashow, Weinberg, and Salam [27, 28, 7].

Another major facility was the Large Electron–Positron Collider (LEP), which operated from 1989 to 2000 in the same tunnel that is currently used by the LHC. LEP performed high-precision measurements of electroweak observables and provided stringent tests of the SM at the quantum level. The precision achieved at LEP allowed indirect constraints to be placed on the masses of the top quark and the Higgs boson before their direct observation [29, 7].

The current flagship accelerator at CERN is the LHC, the most powerful particle accelerator ever built. The LHC was designed to collide protons at unprecedented center-of-mass energies in order to explore the mechanism of electroweak symmetry breaking and to search for physics beyond the SM [26]. One of its primary goals was the discovery of the Higgs boson, the particle associated with the scalar field responsible for generating the masses of elementary particles. In 2012, the ATLAS and CMS experiments reported the observation of a new boson with a mass near 125 GeV, consistent with the Higgs boson predicted by the SM [3, 2]. This discovery completed the particle content of the SM, while at the same time opening a new era of precision measurements aimed at testing its validity and searching for possible deviations.

Today, CERN continues to push the energy and precision frontiers of particle physics. The LHC and its experiments provide a unique environment for studying rare processes, measuring the properties of the Higgs boson with high precision, and searching for phenomena that could reveal physics beyond the SM. These goals motivate ongoing upgrades of both the accelerator complex and the detectors, ensuring that CERN remains at the forefront of fundamental research.

2.2 LHC Design and Accelerator Complex

The LHC is the highest-energy particle accelerator currently in operation at CERN. It is part of a complex chain of accelerators that successively increase the energy of the particle beams before injection into the LHC ring. The CERN accelerator complex, shown in Fig. 2.1, consists of several pre-accelerators that prepare the proton beams and deliver them to the LHC at the required energy [26].

Protons used in the LHC are obtained from hydrogen gas. The accelerator chain starts with a source of hydrogen ions, which are accelerated in a linear accelerator (Linac). After passing through the Linac, the ions are injected into the Proton Synchrotron Booster

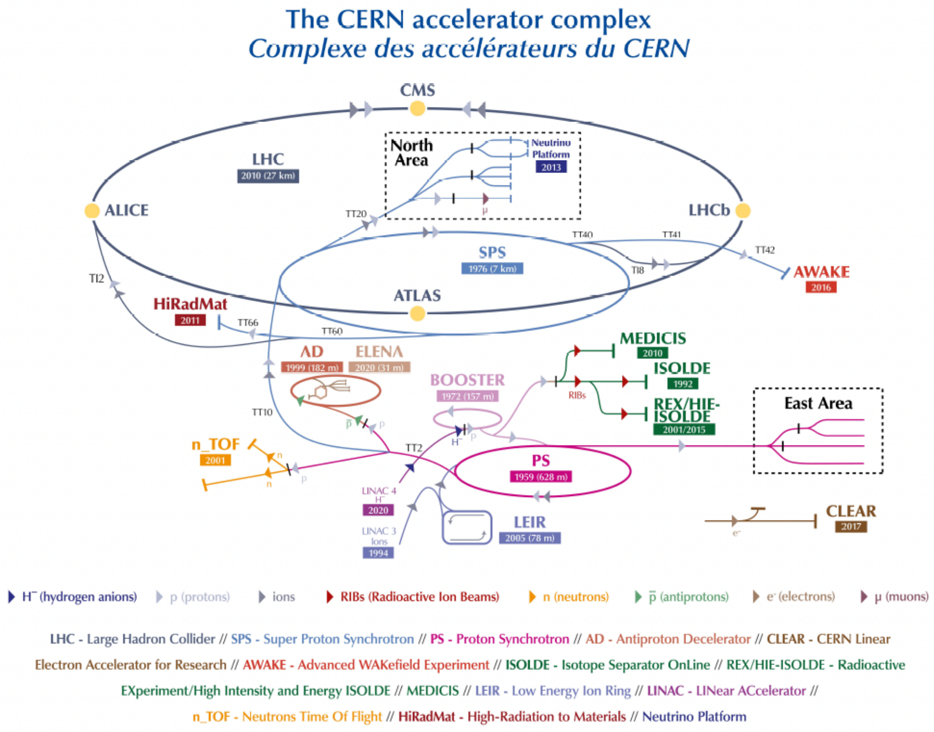


Figure 2.1: Schematic layout of the CERN accelerator complex, including the injector chain and the LHC ring.

(PSB), where the electrons are stripped off, leaving only protons. The proton beam is then transferred to the PS, where it is accelerated to an energy of 25 GeV. The next stage of acceleration takes place in the SPS, which increases the beam energy to 450 GeV. Finally, the protons are injected into the LHC, where they are accelerated to their nominal energy of several teraelectronvolts per beam.

The LHC is located in a circular tunnel with a circumference of approximately 27 km, previously used by the LEP. Inside the tunnel, two proton beams circulate in opposite directions in separate beam pipes under ultra-high vacuum conditions. The beams are guided and focused by a system of superconducting magnets operating at cryogenic temperatures. Dipole magnets bend the particle trajectories along the circular path, while quadrupole magnets focus the beams to maintain a small transverse size at the interaction points.

Proton bunches are arranged in a sequence along each beam, and collisions occur when bunches from the two counter-rotating beams cross at the interaction points. The rate of collisions is determined by the instantaneous luminosity, which is defined as

$$\frac{dN}{dt} = L\,\sigma, \tag{2.1}$$

where σ is the cross section of the process under study and L is the instantaneous luminosity. The luminosity depends on the beam parameters and, for Gaussian beam profiles, can be written as

$$L = \frac{N_b^2\, n_b\, f_{\text{rev}}\, \gamma_r}{4\pi\, \varepsilon_n\, \beta^*} F, \tag{2.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic Lorentz factor, ε_n is the normalized transverse emittance, β^* is the beta function at the interaction point, and F is a geometrical factor accounting for the crossing angle of the beams [26]. Typical instantaneous luminosities achieved at the LHC during Run 2 were of the order of 10^{34} $\text{cm}^{-2}\text{s}^{-1}$. In Run 3, further improvements in beam performance have led to peak luminosities exceeding 2×10^{34} $\text{cm}^{-2}\text{s}^{-1}$,

Several large experiments are installed at the interaction points of the LHC. The two general-purpose detectors, ATLAS and CMS, are designed to study a wide range of physics processes, including Higgs boson production, precision measurements of SM parameters, and searches for new particles. The ALICE experiment is optimized for heavy-ion collisions, while LHCb focuses on the study of heavy-flavour physics and CP violation.

The first proton beams were circulated in the LHC in 2008. After an initial shutdown due to a technical incident, stable operation began in 2009. During Run 1, the LHC delivered proton–proton collisions at center-of-mass energies of 7 and 8 TeV, which led to the discovery of the Higgs boson in 2012. In Run 2, the collision energy was increased to 13 TeV, allowing for more precise measurements and extended searches for new physics. After a long shutdown and upgrades of the accelerator and detectors, Run 3 started in 2022 with improved performance and higher luminosity. Future upgrades, known as the High-Luminosity LHC (HL-LHC), aim to significantly increase the instantaneous luminosity, enabling precision studies of rare processes and detailed measurements of the Higgs sector. The main running periods of the LHC are summarized in Table 2.1.

Run	**Period**	**Collision energy**
Run 1	2010–2012	7–8 TeV
Run 2	2015–2018	13 TeV
Run 3	2022–2026	13.6 TeV
HL-LHC	∼2029–2040 (planned)	14 TeV

Table 2.1: Summary of the LHC running periods, their time spans, and corresponding center-of-mass energies.

2.3 Experiments at the LHC

The LHC hosts several major experiments located at different interaction points along the accelerator ring. These experiments are designed to study a wide range of physics processes, from precision measurements of SM parameters to the search for new particles and the investigation of strongly interacting matter at extreme energy densities. The four main experiments currently operating at the LHC are ATLAS, CMS, ALICE, and LHCb. In this section, a brief overview of the experiments other than CMS is given.

2.3.1 ATLAS

The ATLAS (A Toroidal LHC ApparatuS) experiment is one of the two general-purpose detectors at the LHC. It is designed to study proton–proton collisions at the highest available energies and to explore a broad physics program including the Higgs boson, electroweak processes, top-quark physics, and searches for physics beyond the SM. The detector has a cylindrical geometry with nearly full coverage in solid angle around the interaction point. It consists of an inner tracking detector, electromagnetic and hadronic calorimeters, and a large muon spectrometer based on toroidal magnetic fields.

A schematic view of the ATLAS detector is shown in Fig. 2.2.

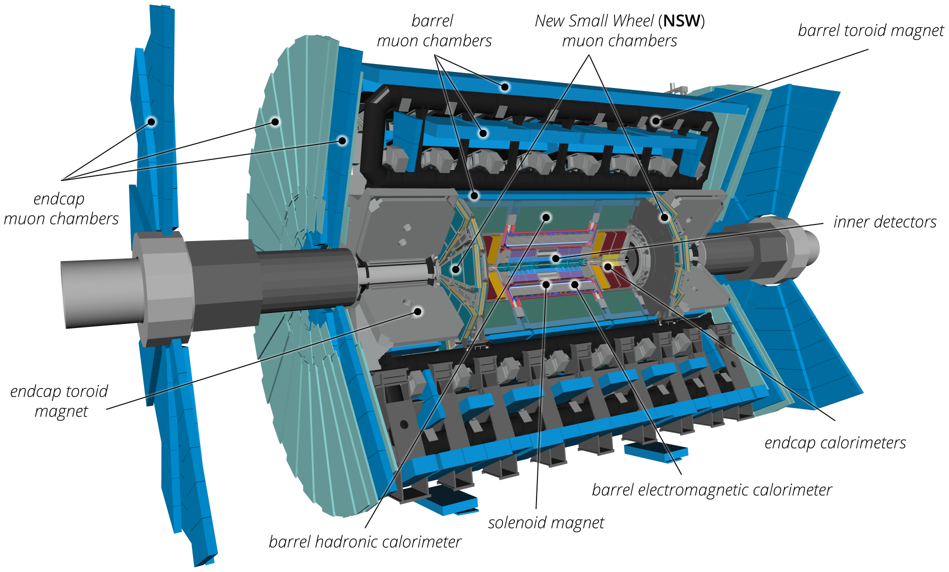


Figure 2.2: Schematic view of the ATLAS detector [30].

2.3.2 ALICE

The ALICE (A Large Ion Collider Experiment) detector is dedicated to the study of heavy-ion collisions. Its main goal is to investigate the properties of strongly interacting matter at extremely high energy density, where a state known as the quark–gluon plasma is expected to be formed. In addition to heavy-ion collisions, ALICE also records proton–proton interactions for reference measurements.

The detector is optimized for tracking and particle identification in events with very high particle multiplicity. It includes a large Time Projection Chamber, silicon tracking detectors, transition radiation detectors, and several subsystems designed for precise particle identification over a wide momentum range.

A schematic view of the ALICE detector is shown in Fig. 2.3.

2.3.3 LHCb

The LHCb (LHC beauty) experiment is designed to study the physics of heavy-flavour hadrons, in particular particles containing b and c quarks. Its main physics goals include precision measurements of CP violation, studies of rare decays, and searches for indirect effects of BSM physics.

Unlike the general-purpose detectors, LHCb has a forward geometry optimized for the detection of particles produced at small angles with respect to the beam direction.

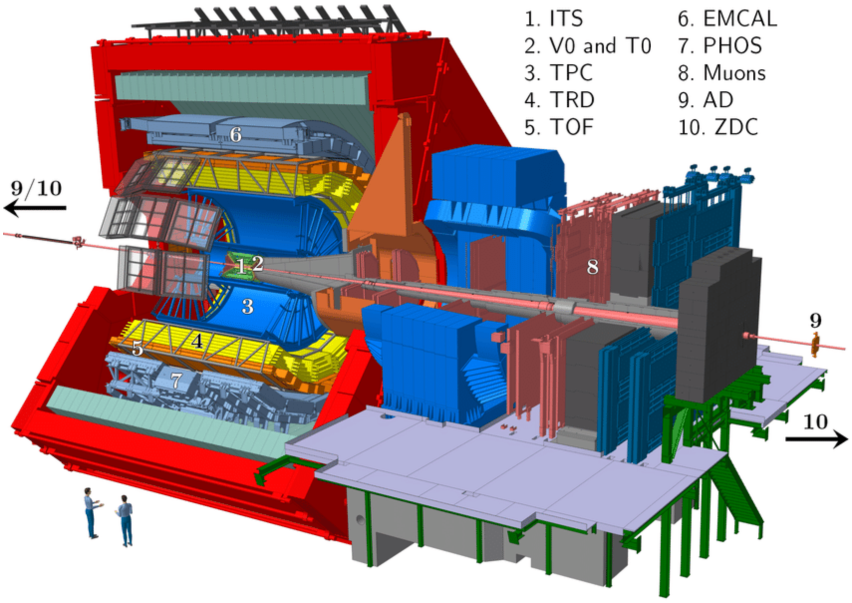


Figure 2.3: Schematic view of the ALICE detector [31].

This configuration provides high efficiency for heavy-flavour production, which is strongly boosted in the forward region. The detector includes a high-precision tracking system, particle identification detectors based on Cherenkov radiation, calorimeters, and a muon system.

A schematic view of the LHCb detector is shown in Fig. 2.4.

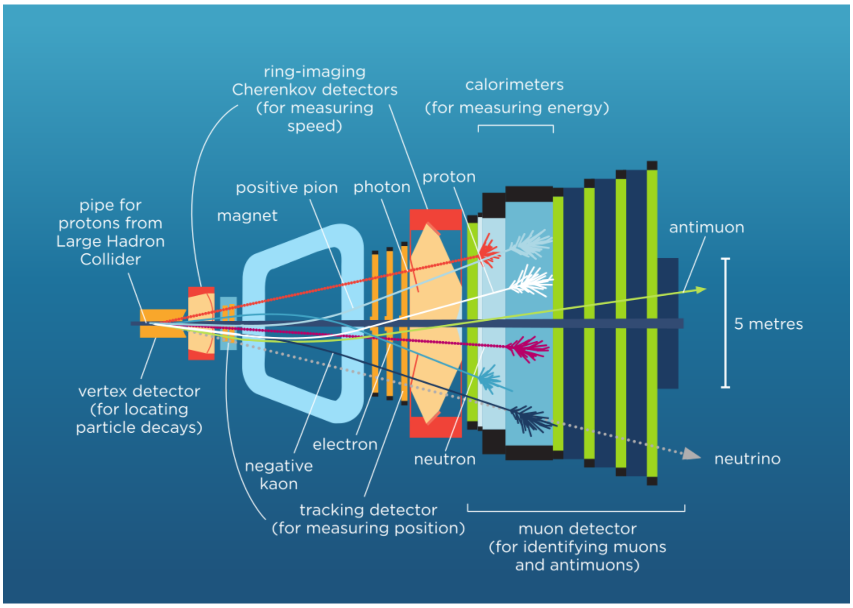


Figure 2.4: Schematic view of the LHCb detector [32].

2.3.4 Other experiments

In addition to the four main experiments, several smaller and more specialized experiments are operating at the LHC:

- **LHCf** (Large Hadron Collider forward) and **TOTEM** (TOTal Elastic and diffractive cross section Measurement) are made from detectors put at hundred of meters on each side of the ATLAS and CMS interaction point, respectively, and are meant to study particles produced at very small angles, inaccessible for ATLAS and CMS.

- **MoEDAL** (Monopole and Exotics Detector at the LHC) shares the same cavern with LHCb and is designed to directly search for the magnetic monopole.

- **FASER** (ForwArd Search ExpeRiment) is designed to search for light and extremely weakly interacting particles and is located along the beam trajectory, 480 metres downstream of the ATLAS detector. It started taking data in 2022.

2.4 The CMS Detector: Design and Particle Detection

The CMS detector is one of the two multi-purpose experiments at the LHC, located at Interaction Point 5. The detector was conceived with the primary objectives of discovering the Higgs boson, characterising its properties and those of the electroweak sector, searching for signs of physics beyond the SM, and studying high-energy heavy-ion collisions.

The name of the CMS detector reflects some of its distinctive features. Despite its total weight of approximately 14 000 tons, the detector is relatively compact compared to ATLAS, measuring about 15 m in height and 21 m in length. A strong emphasis in the design is placed on the detection of muons, which provide a clean and distinctive signature for many interesting physics processes. Muons can be efficiently identified, and high-energy muons often originate from the decay of heavy resonances, making them especially important for searches for new particles.

The central element of the detector is a powerful superconducting solenoid magnet, providing a magnetic field of 3.8 T, one of the strongest ever built for a particle physics experiment. Inside the solenoid, and therefore immersed in the magnetic field, are located the silicon tracking system and the electromagnetic and hadronic calorimeters. The muon detection system is placed outside the solenoid, embedded in the iron return yoke of the magnet.

This configuration allows excellent performance of the inner tracking detector, thanks to the combination of a high magnetic field and a silicon-based tracking system, while keeping the amount of material in front of the calorimeters relatively low. A consequence of the compact design is a more limited space available for the hadronic calorimeter, which required a careful optimisation of the detector layout in order to maintain good energy resolution for hadrons.

A schematic representation of the CMS detector is shown in Fig. 2.5.

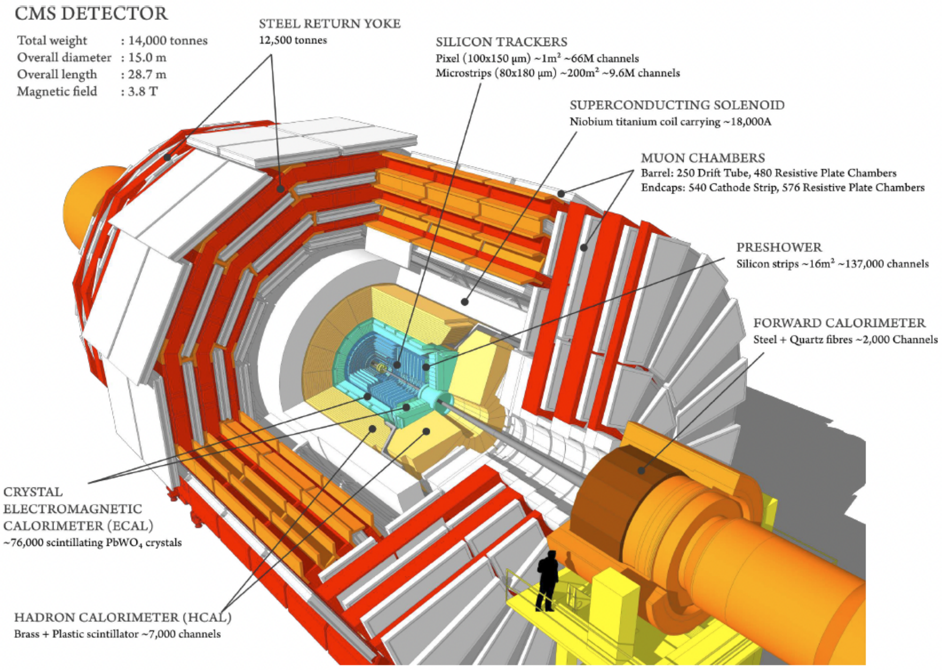


Figure 2.4: Pictorial representation of the CMS detector.

Figure 2.5: Schematic representation of the CMS detector.

2.4.1 Coordinate System

The CMS experiment adopts a right-handed coordinate system with the origin placed at the nominal collision point inside the detector. The z-axis is defined along the beam direction, the x-axis points toward the centre of the LHC ring, and the y-axis points vertically upward, completing the right-handed system. Because of the cylindrical geometry of the detector, it is natural to use a polar coordinate system. The azimuthal angle ϕ is measured in the x–y plane from the x-axis, the radial coordinate is denoted by r, and the polar angle θ is measured from the z-axis in the y–z plane. The relation between Cartesian and polar coordinates is

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta \end{aligned} \tag{2.3}$$

A schematic representation of the CMS coordinate system is shown in Fig. 2.6.

While this coordinate system is well suited for describing the detector geometry, it is not ideal for describing proton–proton collisions at a hadron collider. Protons are composite particles made of quarks and gluons, collectively called partons, and the actual interaction occurs between two partons carrying unknown fractions of the proton momentum. As a consequence, the centre-of-mass frame of the parton collision is generally boosted along the beam axis.

For this reason, variables that are invariant under Lorentz boosts along the z-axis are commonly used. Quantities defined in the plane perpendicular to the beam direction are denoted with the subscript T. The transverse momentum is defined as

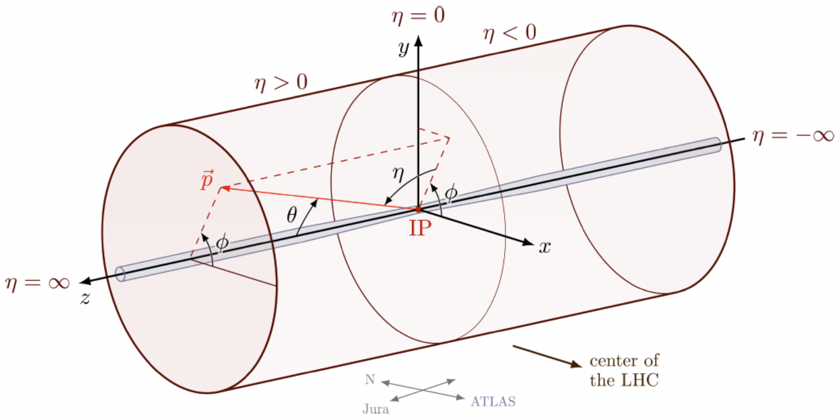


Figure 2.6: CMS coordinate system.

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.4}$$

and the transverse energy is

$$E_T^2 = m^2 + p_T^2. \tag{2.5}$$

The use of transverse quantities is central at hadron colliders, where the longitudinal momentum of the interacting partons is unknown, while the total transverse momentum of the initial state is approximately zero. Momentum conservation in the transverse plane therefore provides a strong constraint on event kinematics. The missing transverse momentum (p_T^{miss}) is defined as the negative vector sum of the transverse momenta of all reconstructed particles in the event. A non-zero value indicates the presence of particles that escape detection, such as neutrinos, which do not interact with the detector material. The measurement of p_T^{miss} is essential for identifying processes involving neutrinos and other weakly interacting particles.

Another useful variable is the rapidity y, defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{2.6}$$

For ultra-relativistic particles, the rapidity is well approximated by the pseudorapidity η, which depends only on the polar angle:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \tag{2.7}$$

The pseudorapidity is purely geometrical and is particularly convenient because differences in η are invariant under boosts along the beam axis. The angular separation between two particles in the detector is commonly expressed as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.8}$$

Based on the pseudorapidity coordinate, the detector is divided into a central region called the barrel and two forward regions called the endcaps. The exact transition value in η depends on the specific subdetector.

2.4.2 CMS Subsystems

The CMS detector is composed of several subdetectors arranged in concentric layers around the interaction point. Each subsystem is optimized for measuring specific properties of the particles produced in collisions, and the combination of their information allows a precise reconstruction of the event.

Moving outward from the interaction point, the detector consists of the silicon tracking system, the electromagnetic calorimeter (ECAL), the hadron calorimeter (HCAL), the superconducting solenoid magnet, and the muon detection system embedded in the iron return yoke. The tracking detector reconstructs charged particle trajectories and determines their momenta in the magnetic field. The calorimeters measure the energy of particles interacting with dense material, while the outer muon system detects muons that traverse the inner detector layers. The solenoid magnet provides the strong magnetic field required for accurate momentum measurements and determines the overall geometry of the detector. The magnetic field also allows for the separation of charged and neutral particles, as only charged particles are deflected, which is essential for particle identification.

Each of these subsystems uses different detection technologies and is designed to achieve optimal performance for the corresponding type of measurement. A more detailed description of the individual components is given below.

Superconducting Solenoid Magnet

A key element of the CMS detector is the superconducting solenoid magnet, which provides the strong magnetic field required for precise momentum measurements of charged particles. The solenoid is located at the center of the detector and surrounds the tracking system and the calorimeters. Its design determines the overall geometry of the experiment and strongly influences the layout of all other subdetectors.

The CMS solenoid produces a uniform magnetic field of 3.8 T. The corresponding operating current is approximately 18 160 A, and the stored magnetic energy is about 2.3 GJ. The magnet has a length of approximately 13 m and an inner diameter of about 6 m. Despite the large field strength, the detector remains relatively compact compared to other general-purpose experiments, which is reflected in the name Compact Muon Solenoid.

The magnetic field bends the trajectories of charged particles, allowing their momenta to be determined from the curvature of their tracks reconstructed in the inner tracking detector. A strong magnetic field improves the momentum resolution, especially for high-energy particles, which is essential for precision measurements and searches for rare processes.

The solenoid coil is made of superconducting niobium–titanium (NbTi) material and is operated at a cryogenic temperature of approximately 4 K. Outside the coil, a massive iron return yoke is used to close the magnetic flux. This iron structure also serves as the mechanical support for the muon detection system, which is installed in several layers within the return yoke. The combination of the solenoid and the iron yoke makes it possible to achieve both a strong magnetic field in the inner detector and an efficient identification of muons in the outer region.

Silicon Tracker

The silicon tracking system is the innermost subdetector of the CMS experiment and is responsible for reconstructing the trajectories of charged particles produced in proton–proton collisions. The tracker is located inside the superconducting solenoid and operates in the uniform magnetic field of 3.8 T, which allows the determination of particle momenta from the curvature of their trajectories. Precise tracking is essential for the reconstruction of primary and secondary vertices, for the identification of heavy-flavour decays, and for accurate measurements of particle kinematics.

The CMS tracker is entirely based on silicon detector technology in order to withstand the high particle flux and radiation levels expected at the LHC. The detector extends to a radius of about 1.1 m and a length of 5.8 m along the beam direction, covering the pseudorapidity range up to approximately $|\eta| < 2.5$. It is composed of two main subsystems: the pixel detector close to the interaction point and the silicon strip detector at larger radii. A schematic layout of the CMS tracking system is shown in Fig. 2.7.

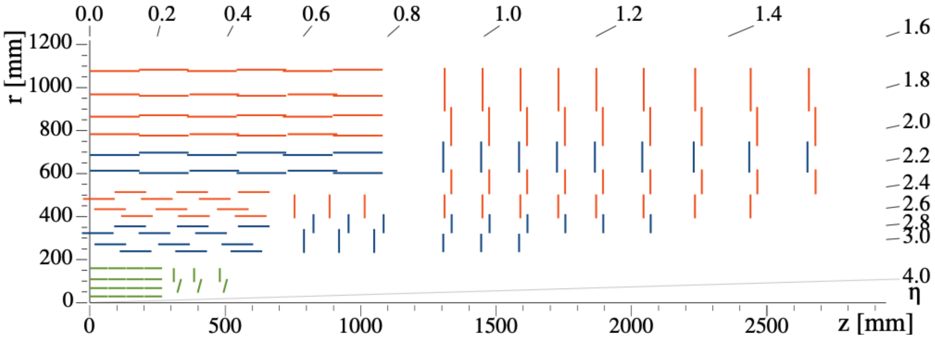


Figure 2.7: Sketch of one quarter of the tracking system in the r–z view. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively [33].

The pixel detector forms the innermost part of the tracker and provides the highest spatial resolution. In the Phase-1 configuration used during Run 2 and Run 3, the pixel detector consists of four barrel layers located at radii between approximately 3 cm and 16 cm, together with three forward disks on each side of the interaction point. The pixel size is about $100 \times 150\ \mu\text{m}^2$, allowing a hit resolution of the order of 10 μm in the transverse plane. The high granularity of the pixel detector is crucial for efficient track reconstruction in events with large particle multiplicity and for precise determination of the interaction vertices.

Surrounding the pixel detector is the silicon strip tracker, which provides multiple position measurements along the particle trajectory. The strip tracker is divided into several substructures, including the Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disks (TID), and Tracker Endcaps (TEC). In total, the strip detector contains ten barrel layers and several disks in the forward regions, providing up to about ten measurement points per track. The strip pitch varies between approximately 80 and 180 μm, depending on the detector region.

The full tracking system contains more than 200 m^2 of silicon sensors and over 100 million readout channels, making it one of the largest silicon tracking systems ever built for a particle physics experiment. The combination of a large number of precise measurements and the strong magnetic field results in an excellent momentum resolution. For charged particles with transverse momentum of about 10 GeV in the central region, the relative momentum resolution is typically around 1%, while for higher momenta the resolution degrades gradually due to the limited lever arm of the detector.

In addition to momentum measurement, the tracker provides precise reconstruction of the primary interaction vertex, with a resolution of the order of tens of micrometres. This capability is essential in LHC conditions, where multiple proton–proton interactions occur in the same bunch crossing. The reconstruction of displaced secondary vertices allows the identification of hadrons containing heavy quarks, which plays a key role in many physics analyses.

The high precision, radiation hardness, and fast readout of the silicon tracker make it a crucial component of the CMS detector, providing the foundation for accurate event reconstruction.

Electromagnetic Calorimeter

The ECAL of the CMS detector is designed for precise measurements of the energy and position of electrons and photons. It is located inside the superconducting solenoid, between the silicon tracker and the HCAL, and operates in the 3.8 T magnetic field. The performance of the ECAL is especially important for physics channels involving photons and electrons, such as the reconstruction of neutral bosons and Higgs boson decays.

The CMS ECAL is a homogeneous calorimeter built from lead tungstate ($PbWO_4$) scintillating crystals. This material was chosen because of its high density (8.28 g/cm^3), short radiation length $X_0 = 0.89$ cm, and small Molière radius of about 2.2 cm, which allow the calorimeter to be compact while maintaining good lateral containment of electromagnetic showers. The scintillation signal of $PbWO_4$ is fast, with a decay time of about 10 ns, making it suitable for operation at the LHC bunch spacing of 25 ns. The light yield is relatively low compared to other scintillators, which requires the use of photodetectors with internal gain.

A schematic view of the ECAL detector is shown in Fig. 2.8.

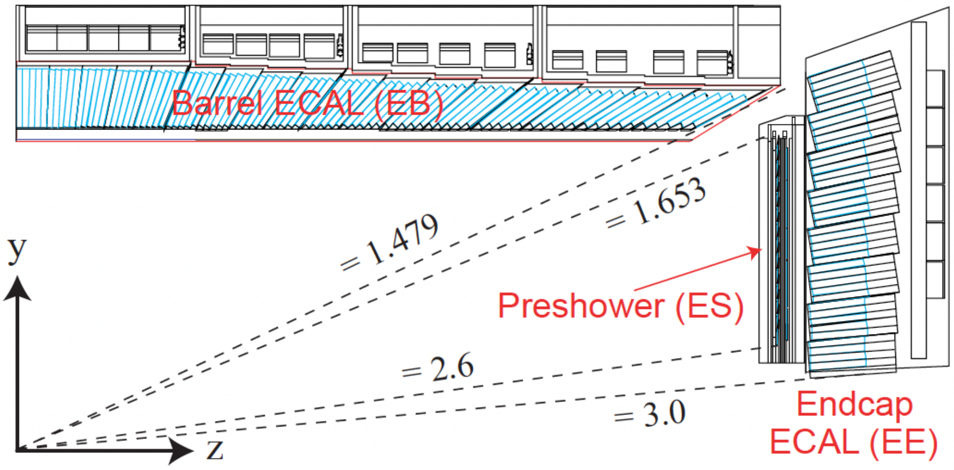


Figure 2.8: Graphical representation of a quarter of the ECAL [34].

The ECAL consists of a barrel section (EB) and two endcaps (EE). The barrel covers the pseudorapidity range $|\eta| < 1.48$ and is composed of 61 200 crystals arranged in 36 supermodules. Each crystal has a length of 23 cm, corresponding to about 25 radiation lengths, which is sufficient to contain electromagnetic showers up to high energies. The endcaps extend the coverage to $|\eta| < 3.0$ and contain 7 324 crystals each. In the region $1.65 < |\eta| < 2.6$, a preshower detector (ES) is installed in front of the endcaps to improve the separation between single photons and photons originating from neutral pion decays. The preshower consists of two layers of silicon strip sensors interleaved with lead absorbers, providing additional spatial information before the shower develops in the crystals.

The scintillation light produced in each crystal is detected by photodetectors mounted on the rear side of the crystals. In the barrel region, avalanche photodiodes (APDs) are used. An APD is a silicon semiconductor device operated at high reverse bias voltage, where the primary photoelectrons generated by incident light are multiplied through impact ionization in a strong electric field. This internal gain, typically of order 50, compensates for the relatively small light yield of $PbWO_4$. APDs are insensitive to magnetic fields and therefore can operate inside the 3.8 T solenoid without performance degradation. Two APDs are coupled to each barrel crystal in order to improve the signal-to-noise ratio and provide redundancy.

In the endcap region, vacuum phototriodes (VPTs) are used instead of APDs because of the higher radiation levels. A VPT is a vacuum photodetector consisting of a photocathode, an anode, and a dynode structure that provides moderate gain. VPTs are more radiation tolerant than semiconductor devices and can operate in strong magnetic fields when properly aligned with the field direction. Their gain is lower than that of APDs, but sufficient for the larger signals produced in the endcap region.

During operation at the LHC, the transparency of the $PbWO_4$ crystals degrades under irradiation due to the formation of colour centres[1], which absorb part of the scintillation light and reduce the signal reaching the photodetectors. This effect is partially reversible during periods without irradiation but varies with time and accumulated dose, especially in the endcap regions where the radiation levels are higher. To correct for this, the ECAL is equipped with a laser monitoring system that continuously measures the transparency of each crystal by injecting light of known intensity. The observed variations are used to derive time-dependent correction factors, which are applied to the reconstructed energy to maintain a stable calorimeter response at the per-mille level.

The energy measurement in the ECAL is based on the development of electromagnetic showers. When a high-energy electron enters the calorimeter, it loses energy mainly through bremsstrahlung in the electric field of atomic nuclei, emitting photons. High-energy photons interact through pair production, creating electron–positron pairs. These secondary particles in turn radiate photons and produce additional pairs, resulting in a cascade of particles. The shower develops over a characteristic scale determined by the radiation length X_0, and the number of particles in the shower increases until the average particle energy falls below the critical energy, after which ionization losses dominate. Because the ECAL crystals have a depth of about 25 radiation lengths, the electromagnetic shower is almost completely contained inside the detector, allowing the deposited energy

---

[1] Colour centres are defects in the crystal lattice created by radiation damage, which introduce energy levels in the band gap and absorb part of the scintillation light, reducing the optical transparency of the material.

to be measured with high precision.

The energy resolution of the ECAL can be parametrized as

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \tag{2.9}$$

where S is the stochastic term, N is the electronic noise term, and C is the constant term arising from calibration uncertainties and non-uniformities. For the barrel calorimeter, the resolution for high-energy electrons and photons is better than 1%, which is essential for precision measurements such as the determination of the Higgs boson mass in the diphoton channel.

Hadron Calorimeter

The HCAL is a hermetic sampling calorimeter designed to measure the energy of hadrons produced in proton–proton collisions. It complements the ECAL by detecting particles that undergo hadronic interactions, such as charged and neutral hadrons, and plays a crucial role in the reconstruction of jets and missing transverse momentum. The combination of ECAL and HCAL allows the measurement of the total visible energy in the event, which is essential for identifying neutrinos and possible signatures of BSM physics.

A schematic representation of the HCAL layout inside the CMS detector is shown in Fig. 2.9.

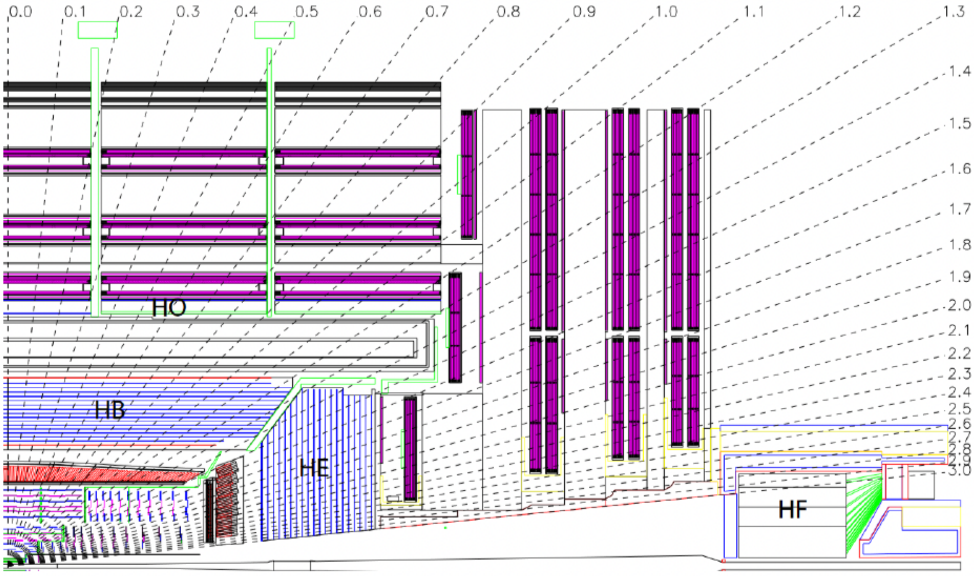


Figure 2.9: Graphical representation of a quarter of the CMS detector, with the HCAL sections labeled [35].

Unlike the ECAL, which is a homogeneous calorimeter, the HCAL is a sampling calorimeter consisting of alternating layers of absorber material and active scintillator tiles. The absorber plates are made of brass or steel and are used to induce hadronic showers, while the plastic scintillators detect the energy deposited by charged particles produced in the shower. When a high-energy hadron enters the calorimeter, it interacts with nuclei in the absorber material through strong interactions, producing secondary hadrons and nuclear fragments. These particles generate a cascade known as a hadronic shower. The characteristic scale of hadronic shower development is given by the nuclear interaction length λ_I, which is significantly larger than the radiation length X_0 relevant for electromagnetic showers. In CMS, for example, the electromagnetic calorimeter made of lead tungstate has a radiation length of $X_0 \approx 0.9$ cm, while the brass absorber used in the hadron calorimeter has a nuclear interaction length of $\lambda_I \approx$ 16–20 cm. Because of the large interaction length, the HCAL must be thicker than ECAL to contain the shower.

The HCAL is divided into several sections according to the detector geometry and pseudorapidity coverage. The barrel HCAL (HB) covers the central region $|\eta| < 1.3$ and is located outside the ECAL, inside the solenoid magnet. The outer radius of the HB is about 1.77 m, while the inner radius of the solenoid is approximately 2.95 m, which limits the available space for absorber material. As a consequence, the total depth of the HB corresponds to about 7 interaction lengths, which is not sufficient to fully contain high-energy hadronic showers.

To improve the containment of hadronic showers in the central region, the outer HCAL (HO) is placed outside the solenoid magnet, using the iron return yoke as an additional absorber. The HO consists of several layers of plastic scintillators installed inside the iron structure and covers the region $|\eta| < 1.26$. By adding the HO, the total depth of the calorimeter in the central region increases to about 11 interaction lengths, which significantly improves the energy resolution for hadrons and jets.

The hadron endcap calorimeter (HE) extends the coverage to higher pseudorapidity, in the range $1.3 < |\eta| < 3.0$. It is located behind the ECAL endcaps and is built using the same sampling technique as the HB, with alternating layers of absorber and scintillator. The geometry of the HE absorbers is designed to minimize cracks between the barrel and endcap regions and to maintain good hermeticity of the detector. The total depth of the HE reaches approximately 10 interaction lengths, providing sufficient containment of hadronic showers in the forward region.

In the very forward region, where the particle flux and radiation levels are much higher, the hadron forward calorimeter (HF) is used. The HF is located at a distance of about 11.2 m from the interaction point on both sides of the detector and covers the region up to $|\eta| \approx 5.2$. Because of the extreme radiation environment, the HF uses quartz fibers embedded in a steel absorber instead of plastic scintillators. Charged particles produced in the shower generate Cherenkov light in the quartz fibers, which is collected and used to measure the deposited energy. This design provides good radiation hardness and fast response, which are required for operation in the forward region.

The energy resolution of the HCAL is limited by fluctuations in the development of hadronic showers and by the sampling nature of the detector. The combined energy resolution of ECAL and HCAL for hadrons can be parametrized as

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{84.7\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + (7.4\%)^2}. \tag{2.10}$$

Although the resolution is worse than for electromagnetic particles, the HCAL provides the necessary hermeticity and energy measurement required for jet reconstruction and for the determination of missing transverse momentum, which is one of the key observables in many CMS physics analyses.

Muon System

The muon system is the outermost subdetector of the CMS experiment and is designed for the identification and precise momentum measurement of muons in the pseudorapidity range $|\eta| < 2.4$. It is embedded in the steel flux-return yoke of the superconducting solenoid, where the magnetic field is approximately 2 T. Because muons are minimally ionizing particles and can traverse the inner detector and calorimeters with relatively small energy loss, their detection requires dedicated tracking chambers placed outside the calorimeter system. The magnetic field in the return yoke bends the muon trajectories in the opposite direction with respect to the inner tracker, providing an independent momentum measurement that can be combined with the tracker information to improve the overall resolution.

The muon system is based on gaseous detectors, which allow the coverage of a large surface area with good spatial resolution and moderate cost. It consists of several types of chambers optimized for different regions of pseudorapidity and operating conditions: Drift Tube (DT) chambers, Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Gas Electron Multiplier (GEM) detectors.

In the barrel region, $|\eta| < 1.3$, the muon detection is performed using DT chambers. The DT system consists of about 250 chambers arranged in four stations interleaved with the layers of the iron return yoke. Each chamber is built from multiple drift cells with transverse dimensions of approximately 42×13 mm^2. A gold-plated stainless-steel anode wire of 50 μm diameter is placed at the centre of each cell and is operated at a voltage of about 3.6 kV. The cells are filled with a gas mixture of 85% argon and 15% CO_2. When a muon crosses the chamber, it ionizes the gas and the resulting electrons drift toward the anode wire, producing a signal that allows the determination of the hit position with high precision. Several layers of drift cells, arranged in different orientations, provide both spatial and timing information, enabling accurate reconstruction of the muon trajectory.

In the endcap region, where the particle rate is higher and the magnetic field is stronger and less uniform, CSCs are used instead of DTs. The CSC system covers the region $0.9 < |\eta| < 2.4$ and consists of multi-wire proportional chambers with finely segmented cathode strips. The strips measure the position of the ionization signal with high precision, while the anode wires provide timing information. The chambers operate with a gas mixture of approximately 50% CO_2, 40% argon, and 10% CF_4, which ensures fast response and stable operation in the high-rate environment. Different chamber sizes are used depending on the detector region, in order to maintain good spatial resolution and efficiency across the full pseudorapidity range.

In addition to the DT and CSC tracking detectors, the muon system includes RPCs, which provide fast timing information and are primarily used for triggering. RPCs are installed in both the barrel and endcap regions up to $|\eta| < 1.9$. Each RPC consists of a double-gap structure formed by two resistive Bakelite plates separated by a gas gap of about 2 mm. The plates are operated at high voltage, typically around 9.6 kV, and the gas mixture contains about 95.2% freon, 4.5% isobutane, and 0.3% sulphur hexafluoride. When a charged particle crosses the detector, it ionizes the gas and initiates an avalanche multiplication process, producing a fast electrical signal that is collected by readout strips. The excellent time resolution of RPCs makes them well suited for the first-level muon trigger.

To improve the performance in the forward region, GEM detectors are being installed as part of the muon system upgrade. The GEM detectors cover the region $1.6 < |\eta| < 2.4$, where the particle flux and radiation levels are highest. A GEM detector consists of thin polymer foils coated with metal on both sides and perforated with a high density of microscopic holes. A strong electric field inside the holes amplifies the electrons produced by ionization, resulting in a fast and precise signal. The first GEM station was installed during the Long Shutdown 2 and is used during Run 3, while the full system is planned to be completed before the start of the High-Luminosity LHC operation.

The combination of DT, CSC, RPC, and GEM detectors provides efficient muon identification, precise momentum measurement, and reliable triggering over the full CMS acceptance. This makes the muon system a key component for many physics analyses, especially those involving high-momentum muons in the final state.

2.4.3 The CMS Trigger and Data Acquisition System

The amount of data produced by the CMS detector is too large to be fully recorded. A single event has a size of about 1 MB, and with the LHC bunch crossing frequency of 40 MHz this would correspond to a data rate of approximately 40 TB/s. Since only a very small fraction of collisions contains processes of physical interest, the CMS experiment uses a dedicated trigger and data acquisition system to select relevant events in real time and reduce the data rate to a manageable level. The trigger and data acquisition system, together with the event builder and storage infrastructure, form the Trigger and Data Acquisition (TriDAS) system.

The trigger must take a decision within a very short time after each bunch crossing, using only partial detector information. This requires a compromise between speed and precision: a fast decision can only use coarse detector information, while more accurate reconstruction requires more time. To satisfy these constraints, the CMS trigger is organized in two stages: the Level-1 trigger (L1) and the High-Level Trigger (HLT).

Level-1 Trigger The L1 trigger is a hardware-based system that processes data from every bunch crossing at the full LHC frequency of 40 MHz. The decision must be taken within a latency of about 3.8 μs, which limits the amount of information that can be used. For this reason, the L1 trigger relies primarily on data from the calorimeters and the muon detectors, while the silicon tracker is generally not used due to both its longer readout time and the limited computing resources available at this stage to perform complex tracking algorithms. The inclusion of tracking information at the hardware trigger level is foreseen for the HL LHC upgrade.

The L1 trigger is implemented using custom electronics based on Field Programmable Gate Arrays (FPGA) and Application Specific Integrated Circuits (ASIC). The input to the trigger consists of trigger primitives produced by the front-end electronics of the subdetectors. These primitives represent coarse measurements of energy deposits in the calorimeters or hits in the muon chambers. For example, in the ECAL several neighbouring crystals are summed together to form one trigger primitive.

From these inputs, the L1 trigger reconstructs physics objects such as muons, electrons, photons, jets, and global quantities including missing transverse energy and the scalar sum of transverse energies. These objects are processed by the global trigger, which compares their properties with a predefined trigger menu consisting of a set of algorithms with thresholds on transverse momentum, energy, or object multiplicity. If at least one algorithm is satisfied, the event is accepted as a Level-1 Accept (L1A). The output rate of the L1 trigger is limited to about 100 kHz.

High-Level Trigger Events accepted by the L1 trigger are passed to the HLT, which is a software-based system running on a large computing farm. The HLT receives events at a rate of about 100 kHz and reduces it to an output rate of the order of 1 kHz for prompt reconstruction and storage. This rate is primarily limited by the available computing resources for event processing rather than a strict hardware constraint. The average processing time per event is about 200 ms.

Unlike the L1 trigger, the HLT has access to the full detector information and can perform reconstruction using algorithms similar to those used in the offline analysis. This allows a much more precise selection of events based on reconstructed physics objects. The HLT farm consists of tens of thousands of CPU cores, which process events in parallel and apply several hundred trigger algorithms corresponding to different physics analyses. The accepted events are grouped into primary datasets according to the trigger paths that selected them.

During Run 2 and Run 3, additional data-taking strategies such as data scouting and data parking were introduced. In scouting, only a reduced event content is stored, allowing significantly higher trigger rates. In parking, events are recorded with full information but processed with a delay or using additional computing resources, relaxing the constraints associated with prompt reconstruction. In practice, during Run 3, a significant fraction of parked data was processed in parallel with prompt data taking as computing resources allowed.

The distribution of the HLT output bandwidth among different physics groups depends on the priorities of the data-taking period. An example of the bandwidth allocation is shown in Fig. 2.10.

2.4.4 Particle Detection Logic

At the CMS detector, different types of particles interact with the subdetectors in different ways. Because the detector consists of several layers optimized for different measurements, a given particle can leave signals in more than one subsystem. The pattern of these signals allows the particle type and its kinematic properties to be reconstructed.

Muons are minimally ionizing particles and therefore lose only a small amount of energy while traversing the tracker and the calorimeters. They produce hits in the silicon tracker that are used to reconstruct their trajectories, but typically deposit only a small amount of energy in the electromagnetic and HCALs. Because of their large penetration power, muons reach the outer muon chambers, where additional measurements are performed, allowing a precise determination of their momentum.

Electrons leave hits in the tracker, which are used to reconstruct their tracks, and then develop electromagnetic showers in the ECAL, where most of their energy is deposited. The ECAL measurement provides a precise estimate of the electron energy, while the tracker determines the momentum and the charge. The combination of these measurements allows efficient identification of electrons.

Photons are electrically neutral and therefore do not produce hits in the tracker unless

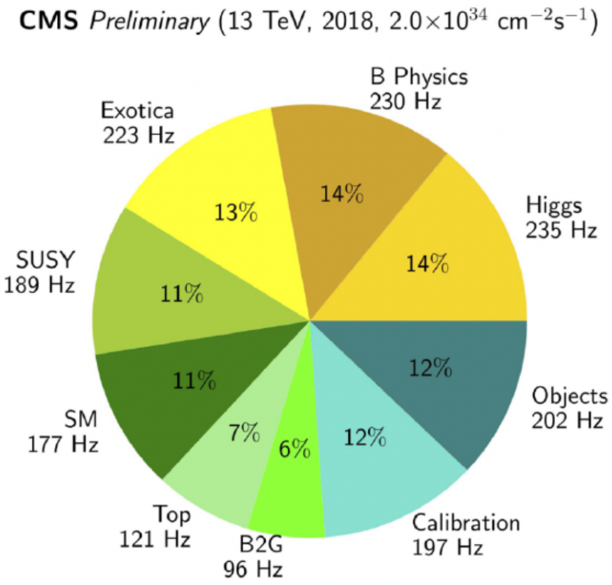


Figure 2.10: Bandwidth occupancy of the High-Level Trigger (HLT) for different physics categories during 2018 data taking [36]. The labels indicate the trigger rate allocated to each category in Hz and the corresponding fraction of the total HLT bandwidth. The categories include Higgs boson searches (Higgs), SM measurements (SM), searches for supersymmetry (SUSY) and other exotic phenomena (Exotica), heavy-flavour physics (B Physics), top-quark studies (Top), beyond-two-generation signatures (B2G), calibration streams (Calibration), and object-based triggers used for performance studies (Objects).

they convert into an electron–positron pair in the detector material. In most cases, photons are detected only through the electromagnetic showers they produce in the ECAL. The absence of a matching track in the tracker is an important criterion used to distinguish photons from electrons.

Charged hadrons leave tracks in the silicon tracker and then interact with the calorimeters, producing both electromagnetic and hadronic showers. A fraction of their energy is deposited in the ECAL through electromagnetic processes, while the rest is measured in the HCAL through hadronic interactions. The combined information from the tracker and the calorimeters is required to obtain an accurate measurement of their energy.

Neutral hadrons do not produce signals in the tracker because they carry no electric charge. They can only be detected through the hadronic showers they create in the HCAL. As a consequence, the energy resolution for neutral hadrons is worse than for charged particles, since no tracker information is available.

Neutrinos interact only through the weak interaction and have a very large interaction length, making them effectively undetectable in the CMS detector. Their presence can only be inferred indirectly through an imbalance of the total transverse momentum in the event.

The different interactions of particles with the CMS subdetectors are illustrated in Fig. 2.11.

Because no single subdetector provides optimal measurements for all particles, the reconstruction of physics objects in CMS relies on combining information from multi-

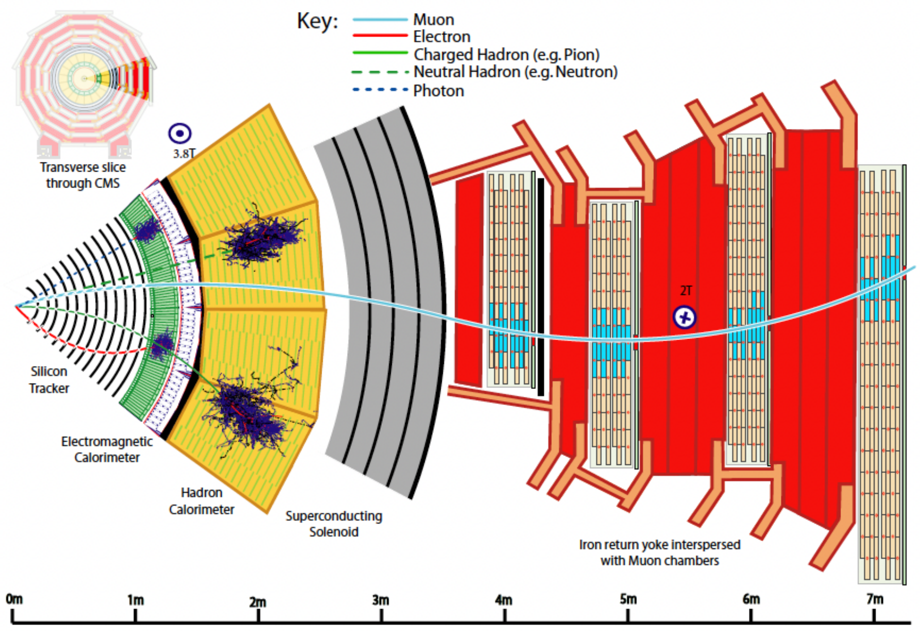


Figure 2.11: Illustration of the interactions of different particle types in the CMS detector [37].

ple detector components. The tracker provides excellent spatial resolution and precise momentum measurement for charged particles, while the calorimeters provide accurate energy measurements, especially at high energy. The muon system adds an independent measurement for muons and improves the overall momentum resolution.

To exploit the full detector information, CMS uses the Particle-Flow (PF) algorithm, which reconstructs individual particles by combining signals from the tracker, calorimeters, and muon system. The PF algorithm uses reconstructed tracks, calorimeter clusters, and muon chamber hits as input elements. These elements are linked together based on their spatial and kinematic compatibility, forming particle candidates that correspond to charged hadrons, neutral hadrons, photons, electrons, or muons.

Tracks of charged particles are reconstructed using an iterative tracking procedure. Calorimeter energy deposits are grouped into clusters, starting from cells with energy above a given threshold and adding neighbouring cells to form topological clusters. The PF algorithm then associates tracks with calorimeter clusters and muon hits, allowing a consistent reconstruction of the particle content of the event.

The combined use of all subdetectors significantly improves the resolution of reconstructed objects, especially for jets and missing transverse momentum, and is therefore essential for most physics analyses performed with the CMS detector.

2.4.5 Muons

Muons, together with electrons, define the fully reconstructed four-lepton final state used in the analysis presented in this thesis. Their reconstruction efficiency, momentum resolution, and associated systematic uncertainties directly affect the signal acceptance and the four-lepton invariant mass resolution. In the $H \to ZZ \to 4\ell$ analysis, muons are considered down to transverse momentum of 5 GeV, making the modelling of the low-p_{T} region particularly important.

Muon reconstruction in CMS combines information from the silicon tracker and the outer muon system through dedicated algorithms. The standard reconstruction provides high efficiency over the full acceptance ($|\eta| < 2.4$) and across the transverse momentum range relevant for Higgs boson measurements.

Muon candidates are reconstructed using complementary approaches. Tracks are first reconstructed either in the inner tracker or independently in the muon system (standalone muons). Standalone muons are matched to tracker tracks to form global muons, for which a combined fit improves the momentum resolution, especially at high transverse momentum ($p_{\mathrm{T}} \gtrsim 200$ GeV). In parallel, tracker muons are reconstructed by extrapolating inner tracker tracks to the muon system and requiring a matching muon segment. This approach provides higher efficiency at low momentum, where multiple scattering in the detector material can prevent the reconstruction of global muons. The combination of global and tracker muons ensures efficient reconstruction over the full momentum range.

The reconstruction and identification efficiencies are measured in data using the Tag and Probe method. For muons with $p_{\mathrm{T}} > 20$ GeV, the efficiencies are derived from $Z \to \mu^+\mu^-$ events, which provide a clean and high-statistics sample across the central and forward detector regions. In the low-p_{T} regime, relevant for subleading leptons in the four-lepton final state, identification and reconstruction efficiencies are measured using $J/\psi \to \mu^+\mu^-$ decays. This complementary approach ensures coverage of the full transverse momentum range.

As can be seen in Figure 2.12, muon identification efficiency is close to unity over the full range.

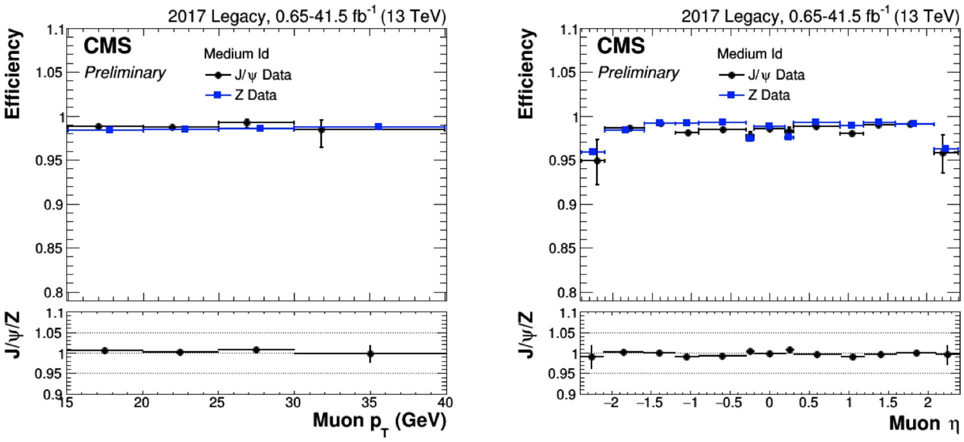


Figure 2.12: Efficiency of the medium muon identification[2] measured in the 2017 legacy data-taking period as a function of the muon p_T (left) and η (right). The efficiencies are obtained with the TnP method using $Z \to \mu\mu$ events (blue) and $J/\psi \to \mu\mu$ events (black). The lower panels show the ratio of the efficiencies measured with the two resonances. Error bars represent the corresponding uncertainties [38].

---

[2] The medium muon identification is optimized for prompt muons and muons from heavy-flavor decays. It requires the muon to pass the loose identification and to have a tracker track with hits in more than 80% of the crossed tracker layers. Additional requirements on the segment compatibility, global-muon fit

Muon identification working points are defined to balance efficiency and background rejection. Identification relies on track quality variables, requirements on impact parameters, and consistency between measurements in the tracker and the muon system. Isolation requirements suppress non-prompt muons originating from heavy-flavour decays or hadron misidentification.

The relative isolation variable is constructed from the scalar sum of transverse momenta of charged hadrons, neutral hadrons, and photons within a cone around the muon direction, with corrections applied to mitigate pileup effects.

The muon momentum is determined from the curvature of the track in the magnetic field. At low and intermediate transverse momentum, the resolution is dominated by the silicon tracker performance, while at higher transverse momentum the inclusion of muon system information in the global fit improves the resolution, and effects from alignment and multiple scattering become increasingly relevant [39].

Residual biases in the momentum scale are corrected using dedicated calibration procedures. The Rochester method [40] is employed to correct for detector effects, magnetic field modelling, and residual misalignments. The method determines event-by-event momentum corrections by comparing reconstructed dimuon invariant mass distributions in data and simulation. Scale and resolution parameters are extracted using $J/\psi \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ decays to ensure agreement of the mass peak position and width. The resulting multiplicative corrections are applied to the muon transverse momentum, in both the low- and high-p_T regions.

2.4.6 Jets

Jets are reconstructed hadronic objects originating from the fragmentation and hadronization of quarks and gluons. In the $H \to ZZ \to 4\ell$ analysis, jets are not part of the Higgs boson decay itself, but they play a central role in event categorization and in measurements differential in jet multiplicity and jet kinematic observables. In particular, the number of reconstructed jets, their transverse momentum, and the presence of b-tagged jets are used to enhance sensitivity to different Higgs boson production mechanisms, such as vector boson fusion, and associated production.

The PF candidates are clustered into jets using the anti-k_T sequential recombination algorithm with a distance parameter of $R = 0.4$, as implemented in the FastJet package [41]. In this algorithm, particles are merged according to distance measures defined in transverse momentum and angular space,

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = \frac{1}{p_{T,i}^2},$$

where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the distance between jets in η–ϕ space. Particles are iteratively combined starting from the smallest distance. Because the metric is weighted by the inverse square of the transverse momentum, high-p_T particles are clustered first, resulting in approximately conical jets. The anti-k_T algorithm is infrared and collinear safe, meaning that the addition of soft jets or collinear splittings [3] does not significantly

---

quality, and tracker–standalone matching are applied when available. The selection is tuned to achieve an efficiency of about 99.5% for muons from simulated W and Z boson decays.

[3] the splitting of a particle into two nearly parallel particles

alter the reconstructed jet structure, ensuring stability of jet observables and compatibility with perturbative QCD calculations.

The reconstructed jet energy differs from the true particle-level energy due to detector response, the different calorimeter response to electromagnetic and hadronic shower components, and additional energy from pileup interactions. A factorized jet energy correction (JEC) procedure is therefore applied to restore the correct energy scale. The correction chain includes offset corrections removing pileup contributions, relative corrections accounting for the η dependence of the detector response, absolute corrections restoring the p_T scale, and residual corrections applied to data. The corrections are derived from simulation and validated with in situ measurements using dijet, γ+jet, and Z+jet events. In particular, the missing transverse momentum projection fraction (MPF) method is used to determine the jet response by exploiting transverse momentum balance in γ+jet events. In this method, the response is defined as

$$R_{\mathrm{MPF}} = 1 + \frac{\vec{E}_T^{\mathrm{miss}} \cdot \hat{n}_\gamma}{p_T^\gamma}, \tag{2.11}$$

where $\vec{E}_T^{\mathrm{miss}}$ is the missing transverse momentum vector, p_T^γ is the transverse momentum of the photon, and $\hat{n}_\gamma$ is the unit vector in the direction of the photon in the transverse plane. After calibration, the average jet response is consistent with unity within uncertainties [42].

The remaining uncertainty on the calibrated jet energy scale, referred to as the jet energy scale (JES) uncertainty, is parameterized as a function of jet p_T and η and propagated to physics observables. For Run 2 data at $\sqrt{s} = 13$ TeV, the JES uncertainty is typically at the level of 1–2% for central jets with $p_T > 30$ GeV [42].

The jet energy resolution (JER) quantifies fluctuations of the reconstructed jet energy around its true value. It is measured using transverse momentum balance in dijet and γ+jet events. For central jets with $p_T \approx 100$ GeV, the resolution is of order 10% in Run 2 data [42]. To ensure agreement between data and simulation, jets in simulated events are smeared to reproduce the resolution observed in data.

Jets originating from bottom quarks are identified using multivariate algorithms that exploit the long lifetime of b hadrons. In Run II, the DeepCSV and DeepJet algorithms were widely used, combining information from track impact parameters, secondary vertices, and jet kinematics within deep neural networks [43]. Working points are defined according to target efficiencies, typically around 70% for b jets, with misidentification probabilities for light-flavour jets at the percent level. The efficiency and mistag rates are measured in data, and data-to-simulation scale factors are derived and applied to simulated events.

2.5 Data Processing at CMS

After the events are selected by the trigger system, the data are processed through several reconstruction steps before they can be used for physics analysis. At each step, the event content is transformed into a different data format, optimized for a specific stage of processing. These formats differ in size, level of detail, and intended use, ranging from raw detector information to compact datasets suitable for analysis.

The main CMS data formats are summarized below:

- **RAW**

  The RAW format contains the detector readout after the trigger selection. It includes the full information from all subdetectors, such as digitized signals from the tracker, calorimeters, and muon chambers. The event size is typically of the order of 1 MB. RAW data are stored at the CERN Tier-0 computing center and are used as the input for the first reconstruction step.

- **RECO**

  The RECO format contains the result of the full event reconstruction. In this step, the detector signals are converted into reconstructed physics objects such as tracks, calorimeter clusters, muons, electrons, photons, jets, and missing transverse momentum. The PF algorithm is applied to combine information from different subdetectors and reconstruct individual particles. The RECO format contains detailed information needed for detector studies and advanced analyses, but its size is relatively large, typically several hundred kilobytes per event.

- **AOD**

  The Analysis Object Data (AOD) format is a reduced version of the RECO format that keeps only the information needed for most physics analyses. Low-level detector details are removed, while reconstructed physics objects and their main properties are retained. The event size is typically around 200 kB. AOD datasets are distributed to Tier-1 and Tier-2 computing centers for analysis.

- **MiniAOD**

  The MiniAOD format was introduced to reduce storage requirements while keeping the information necessary for the majority of analyses. It contains Particle-Flow candidates, high-level reconstructed objects, and a limited set of additional variables used for calibrations and systematic studies. The event size is about 30–50 kB, making it suitable for large-scale data processing and analysis. Most CMS analyses performed during Run 2 and Run 3 use MiniAOD as the starting point.

- **NanoAOD**

  The NanoAOD format is a highly reduced dataset designed for fast analysis. It contains only high-level physics objects and a minimal set of variables stored in flat tables, which allows efficient reading and processing. The typical event size is of the order of 1 kB. NanoAOD is commonly used for final analysis steps, large-scale statistical studies, and machine learning applications, where compact data representation is essential.

The use of multiple data formats allows CMS to balance storage requirements, reconstruction complexity, and analysis flexibility. Detailed formats are used in the early processing stages, while compact formats are used for physics analyses, enabling efficient handling of the large data samples produced by the LHC.

2.6 Monte Carlo Simulation in CMS

Monte Carlo (MC) simulation plays a crucial role in the CMS experiment. Simulated events are used to study detector performance, estimate backgrounds, determine signal efficiencies, and compare theoretical predictions with measured data. The simulation chain consists of several steps, including the generation of the hard scattering process, parton showering and hadronization, detector simulation, and event reconstruction. Different software tools are used for each step of the simulation.

2.6.1 Matrix Element Generators

The hard scattering process is calculated using matrix element generators based on perturbative quantum chromodynamics (QCD) and electroweak theory. An essential ingredient of these calculations are the parton distribution functions (PDFs) [4], which describe the probability of finding a parton carrying a given fraction of the proton momentum. PDFs are determined from global fits to experimental data and are required to relate parton-level calculations to proton–proton collisions.

One of the commonly used programs is MCFM [44], which provides next-to-leading-order (NLO) calculations for many SM processes. It is often used to compute precise cross sections and kinematic distributions.

Another widely used generator is POWHEG [45], which implements the POWHEG method for matching NLO matrix element calculations with parton shower simulations. This approach allows the generation of events with NLO accuracy while preserving the advantages of parton shower algorithms.

In many CMS analyses, matrix elements are generated using MADGRAPH [46], which allows the simulation of complex final states with multiple particles. MADGRAPH can be combined with parton shower generators to produce fully simulated events[5].

For Higgs boson studies, the JHUGEN generator [47] is widely used to model signal processes with detailed control over the spin and parity structure of the interactions. It provides matrix element calculations for Higgs boson production and decay, including anomalous couplings and interference effects.

2.6.2 Parton Shower and Hadronization

After the hard scattering process, the outgoing partons undergo parton showering, followed by hadronization. This step is typically simulated using PYTHIA [48], which models the evolution of partons through successive emissions of gluons and quark–antiquark pairs, and the formation of hadrons using phenomenological models. PYTHIA also simulates most particle decays, as well as the underlying event and multiple parton interactions, which are important for a realistic description of proton–proton collisions.

---

[4] In this thesis, PDF denotes both parton distribution functions and probability density functions; the meaning is inferred from the context.

[5] The original MADGRAPH framework has since evolved into MADGRAPH5_AMC@NLO, which provides both leading-order and next-to-leading-order matrix element calculations. The name MADGRAPH will be retained within the thesis.

2.6.3 Detector Simulation

The interaction of particles with the detector material is simulated using the GEANT 4 toolkit [49]. GEANT 4 provides a detailed description of particle interactions with matter, including electromagnetic and hadronic processes, multiple scattering, energy loss, and nuclear interactions. In CMS, GEANT 4 is used to simulate the full detector geometry, including the tracker, calorimeters, muon system, and magnetic field.

2.6.4 Digitization

After the simulation of particle interactions in the detector, the detector response is modeled through a digitization step. In this stage, the energy deposits and particle hits produced by GEANT 4 are converted into electronic signals, taking into account the response of the detector readout electronics, noise, and finite resolution effects. The output of the digitization step is a set of simulated signals that closely resemble those recorded in real data and can be processed by the same reconstruction algorithms.

Chapter 3

Electrons

> *As part of my thesis work, I developed a novel tool for the measurement of low-p_T electron reconstruction efficiency, reaching down to 5 GeV, by exploiting the J/ψ resonance. I trained the electron identification for 2022 data-taking period and computed electron identification scale factors for the same year, both of which have been approved and are used in $H \to ZZ \to 4\ell$ measurements, including the first Run 3 cross-section measurement [50].*

In this thesis, the properties of the Higgs boson are inferred from the 4ℓ decay channel, where ℓ denotes an electron (e) or a muon (μ). Consequently, the precise reconstruction and identification of electrons are essential for accurate measurements. At CMS, electron reconstruction involves combining energy deposits in the electromagnetic calorimeter with tracks from the silicon tracker to determine the electron's position, energy, charge, and momentum[1]. The reconstruction procedure is not perfectly efficient, and the corresponding efficiency can differ between MC simulation and real data. Therefore, the simulation must be corrected by applying electron reconstruction scale factors (SFs).

These SFs are usually determined using the Z boson resonance, but this choice is not optimal for low-p_T electrons. Instead, an approach using the lower-mass resonance J/ψ is proposed and developed within this thesis to ensure reliable efficiency measurements. This is particularly important for the $H \to ZZ \to 4\ell$ analysis, where low-p_T leptons often appear as subleading decay products, and uncertainties related to their reconstruction efficiency constitute one of the dominant systematic effects.

After reconstruction, an identification algorithm is applied to separate prompt electrons, originating from the primary vertex (PV), from background sources. This is achieved using a set of discriminating variables, combined either through sequential selection cuts or multivariate techniques. Differences in identification efficiency between data and simulation are corrected by applying dedicated identification SFs.

This chapter describes the electron reconstruction and identification procedures used in this thesis. The reconstruction algorithms employed offline and at the trigger level are presented in Sections 3.1 and 3.2, while the energy scale and resolution corrections are discussed in Section 3.3. The general procedure for efficiency measurements, including the Tag and Probe and Root Mean Square methods, is described in Section 3.4. The measurement of the low-p_T electron reconstruction efficiency using $J/\Psi \to e^- e^+$ decays

---

[1] In CMS jargon, both electrons and positrons are referred to as electrons

is presented in Section 3.5. The electron identification training for the 2022 data-taking period is discussed in Section 3.6, and the computation of the 2022 identification SF is described in Section 3.7.

3.1 Offline Electron Reconstruction

Electron and photon reconstruction in CMS is integrated into the PF framework [51], and is based on the same basic building blocks as for other particles. The reconstruction procedure begins with the identification of energy deposits in the ECAL. Individual ECAL crystals with significant energy are grouped into basic clusters, corresponding to localized electromagnetic showers. However, electrons frequently emit bremsstrahlung photons while traversing the tracker material. As a consequence, the energy associated with a single electron may be distributed across multiple clusters, primarily along the azimuthal direction due to the strong magnetic field of the CMS detector. To recover the full electron energy, CMS employs dedicated clustering algorithms that merge several nearby clusters into a single composite object, referred to as a supercluster (SC) [52].

The default superclustering algorithm, commonly referred to as the *mustache* algorithm, defines a characteristic η–ϕ region based on the expected bremsstrahlung emission pattern derived from simulation. All clusters falling within this region are combined to form the supercluster. An example of the mustache SC distribution can be seen on Figure 3.1, for simulated electrons with $1 < E_T^{seed} < 10$GeV. Also, a refined superclustering procedure incorporates tracking information to further improve the recovery of radiated energy and energy deposits arising from photon conversions. The resulting supercluster provides an initial estimate of the electron energy before track association is done.

The standard algorithm for track reconstruction at CMS, the Kalman Filter (KF) [51], estimates particle trajectories by iteratively combining measurements from the silicon tracker with a predictive model of motion, accounting for detector resolution. However, the KF cannot accurately follow electron trajectories when there are large deviations due to bremsstrahlung. To properly account for radiative energy losses, the Gaussian Sum Filter (GSF) tracking algorithm is used. Since GSF track fitting is computationally intensive, it cannot be applied to all reconstructed hits in the tracker. Therefore, electron track reconstruction begins with the identification of hit patterns likely to belong to an electron trajectory, a process known as "seeding," which can be either "ECAL-driven" or "tracker-driven."

- **ECAL-driven seeding:** Uses ECAL superclusters to define initial seeds, which are propagated inward into the tracker to find matching hits. This is effective for electrons that have undergone significant bremsstrahlung and may have missing hits in the tracker. It performs better for high E_T isolated electrons, such as the ones coming from $Z \rightarrow e^-e^+$ decay.

- **Tracker-driven seeding:** Uses tracks reconstructed with the standard tracker algorithms as seeds, which are then extrapolated outward to match ECAL superclusters. This ensures reconstruction efficiency for electrons with small energy loss or those in complex detector regions. It performs better for non-isolated and low-p_T electrons, such as the ones coming from $J/\Psi \rightarrow e^-e^+$ decay. This is exploited in part of the thesis dedicated to low-p_T electron reconstruction efficiency measurements.

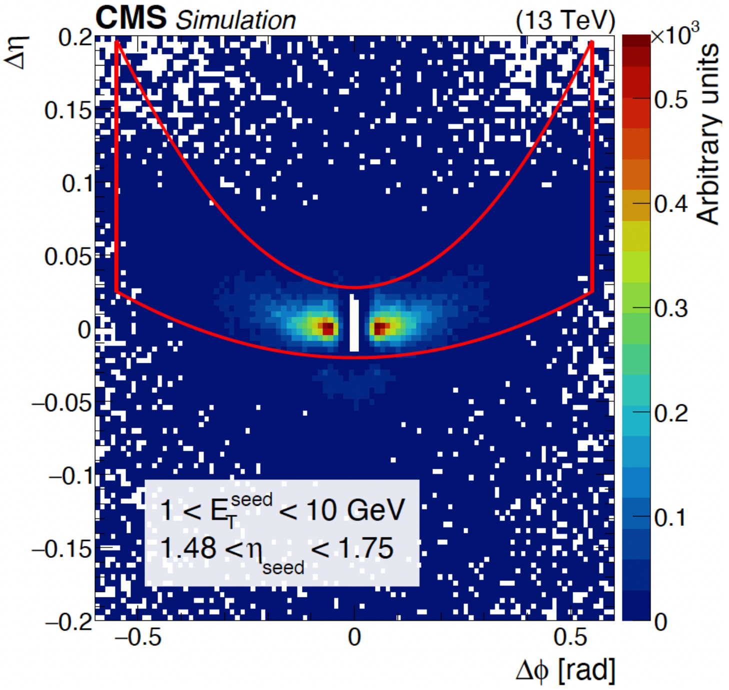


Figure 3.1: Distribution of $\Delta\eta = \eta_{\text{seed-cluster}} - \eta_{\text{cluster}}$ versus $\Delta\phi = \phi_{\text{seed-cluster}} - \phi_{\text{cluster}}$ for simulated electrons with $1 < E_{\mathrm{T}}^{\text{seed}} < 10\,\text{GeV}$ and $1.48 < \eta_{\text{seed}} < 1.75$. The z axis represents the occupancy of the number of PF clusters matched with the simulation around the seed. The red line contains approximately the set of clusters selected by the mustache algorithm [52].

The final set of electron seeds is obtained by combining both ECAL-driven and tracker-driven seeds, after which the electron trajectory is reconstructed using KF algorithm [51]. Electron candidates are constructed by matching GSF tracks with the SCs, based on a Boosted Decision Tree (BDT) output. The BDT is built using track information, supercluster observables, and track–cluster matching variables.

Bremsstrahlung photons can convert into electron-positron pairs within the tracker material, producing additional tracks and spatially separated energy deposits in the ECAL. Such effects complicate the reconstruction of the original electron and can lead to an incomplete measurement of its energy if not properly accounted for.

To mitigate these effects, the CMS electron reconstruction includes a dedicated bremsstrahlung and conversion recovery procedure. After the initial association between the GSF track and the ECAL supercluster, the electron trajectory is extrapolated to the ECAL surface. Tangents drawn from the track to the calorimeter are used to search for additional ECAL clusters compatible with bremsstrahlung photons or conversion products. Clusters satisfying the required geometric and energy-momentum compatibility criteria are merged into the electron supercluster, improving the recovery of radiated energy [52].

The impact of bremsstrahlung radiation on the reconstructed electron is quantified using the bremsstrahlung fraction, f_{brem}, defined as

$$f_{\text{brem}} = \frac{p_{\text{in}} - p_{\text{out}}}{p_{\text{in}}}, \tag{3.1}$$

where p_{in} and p_{out} are the magnitudes of the electron momentum at the innermost and outermost valid states of the GSF track, respectively. This observable provides a direct measure of the relative momentum loss due to bremsstrahlung emission along the electron trajectory.

The distribution of f_{brem} exhibits a strong dependence on the detector material budget and therefore varies as a function of pseudorapidity. Because of higher tracker material budget in the endcaps, f_{brem} has a higher peak at large values, close to 1, compared to the distribution in the barrel detector [52], as seen on Figure 3.2.

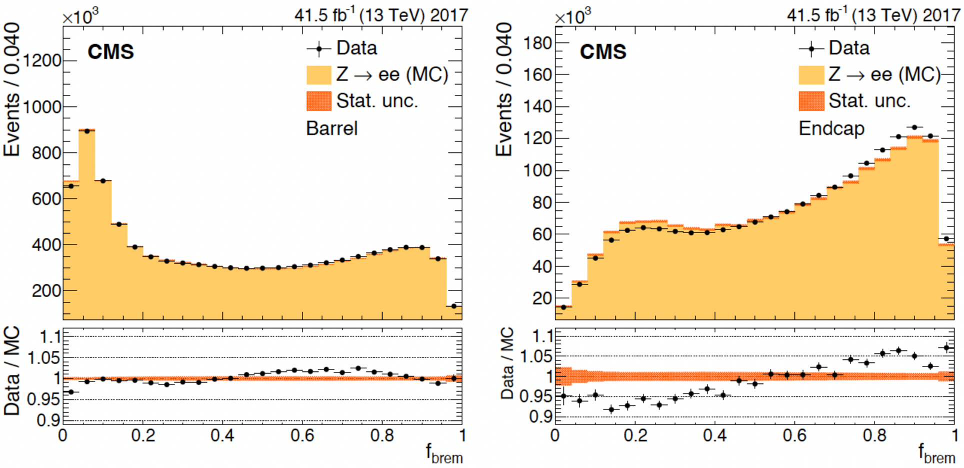


Figure 3.2: Distribution of the bremsstrahlung fraction f_{brem} for electrons reconstructed in the barrel (left) and in the endcaps (right). Black points represent collision data collected by CMS in 2017, while the yellow histograms correspond to simulated $Z \to ee$ events. Orange shaded band indicates the statistical uncertainty of the simulation, and the lower panels show the ratio of data to simulation [52].

The recovery of bremsstrahlung photons and conversion products significantly improves the accuracy of the reconstructed electron energy and momentum, especially in regions with substantial material in front of the ECAL. This procedure is therefore an essential component of the overall electron reconstruction strategy in CMS and plays an important role in precision measurements involving electrons.

In addition to the offline reconstruction, CMS performs an online reconstruction of electrons as part of the trigger system. The goal of the online reconstruction is to efficiently identify events containing electrons in real time, while satisfying the strict latency constraints of the CMS trigger.

3.2 Online Electron Reconstruction

Due to the limited computing resources available online, simplified reconstruction and selection criteria are applied relative to the offline case. Despite these simplifications, the online algorithms are designed to closely follow the offline electron reconstruction strategy.

At L1, electron and photon candidates are reconstructed from calorimeter trigger towers, each of which aggregates the transverse energy measured in a group of ECAL crystals. In the barrel region, a trigger tower corresponds to a 5×5 array of lead-tungstate crystals, while in the endcaps the tower geometry is more complex to accommodate the differing detector layout and segmentation [53]. Trigger primitives from these towers are used as input to the L1 electron/photon (e/γ) clustering algorithm. The algorithm identifies a "seed" tower as a local transverse energy maximum above a programmable threshold (typically $E_T > 2\,\mathrm{GeV}$) and dynamically builds a cluster by including neighbouring towers with E_T above a lower threshold (for example $E_T \geq 1\,\mathrm{GeV}$), without imposing a fixed cluster shape, to recover energy lost through bremsstrahlung and shower spread. The raw transverse energy of the L1 electromagnetic candidate is defined as the sum of all tower E_T values in the cluster [54]. Isolation and identification criteria at L1 exploit shower-shape observables and the ratio of hadronic to electromagnetic energy (H/E) to suppress background, but L1 cannot distinguish between electrons and photons in the absence of tracker information [53].

Online reconstruction at the HLT begins from the L1 seed and uses fine-granularity ECAL crystal information to build electromagnetic clusters within a restricted region of interest around that seed. The HLT clustering procedure processes only ECAL readout channels overlapping a matrix of crystals centred on the L1 candidate in order to limit computing time. Superclusters are then formed using offline clustering algorithms adapted for the HLT environment [55]. For electron candidates, the ECAL supercluster is associated with a reconstructed track whose direction is consistent with the cluster position, whereas photon candidates lack such a track match. The HLT selection applies identification criteria based on the transverse shower profile in ECAL, the energy deposited in HCAL behind the ECAL cluster, and - for electrons - the quality of the track–supercluster association. In addition, isolation criteria are computed from energy deposits surrounding the candidate in the ECAL, HCAL, and tracker detectors, and minimal transverse energy thresholds are imposed on the supercluster E_T to reduce trigger rates [55].

3.3 Energy Corrections

Electron and photon energy, deposited in ECAL, can be lost in several ways: shower leakage, through losses in gaps between modules, in dead crystals or in tracker. To correct for these losses, and to improve energy resolution, a multivariate technique is used to correct the energy estimation for these effects. BDT regression uses simulated events to correct the energy of e/γ object [56].

The electron and photon energy regressions are trained on simulated events containing two electrons or two photons per event, generated with a flat transverse momentum spectrum. In these samples, the true energy of each electron or photon is known, and a geometric matching criterion of $\Delta R < 0.1$ is applied to associate reconstructed objects with their true counterparts. The target of the regression is defined as the ratio of the true energy to the reconstructed energy of the electron or photon.

The electron energy is then calibrated through a sequential application of three regressions. The first regression (step 1) provides a correction to the SC energy. The second regression (step 2) estimates the SC energy resolution, taking into account the additional spread observed in data due to real detector conditions. Finally, the third regression (reffered to as the $E - p$ combination) combines the SC and electron track information to produce the corrected momentum, yielding the final value as the weighted combination of the two. This improves the prediction at low-p_T where tracker has a better resolution and also in the gap regions between supermodules or between endcap and barrel. Relative resolutions of tracker and ECAL, together with $E - p$ combination are shown on Figure 3.3.

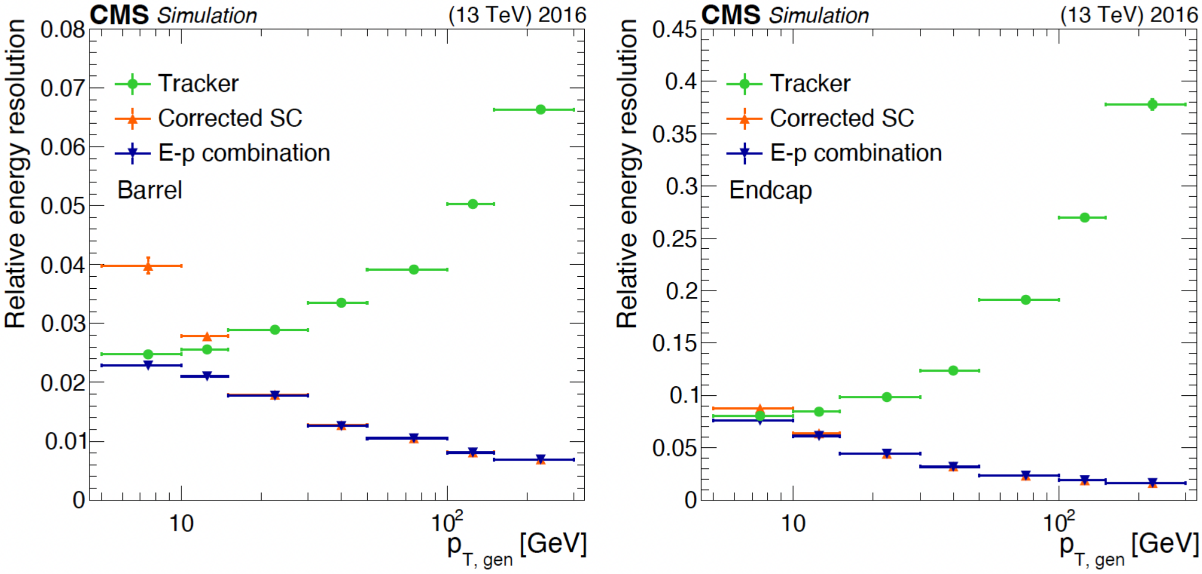


Figure 3.3: Relative electron resolution versus electron pT, as measured by the ECAL ("corrected SC"), by the tracker, and seen in the E-p combination after the step 3 regression, as found in 2016 MC samples for barrel (left) and endcap (right) electrons. Vertical bars on the markers represent the uncertainties coming from the fit procedure [52].

These regressions result in a significant improvement of electron and photon energy measurements, as can be seen on Figure 3.4

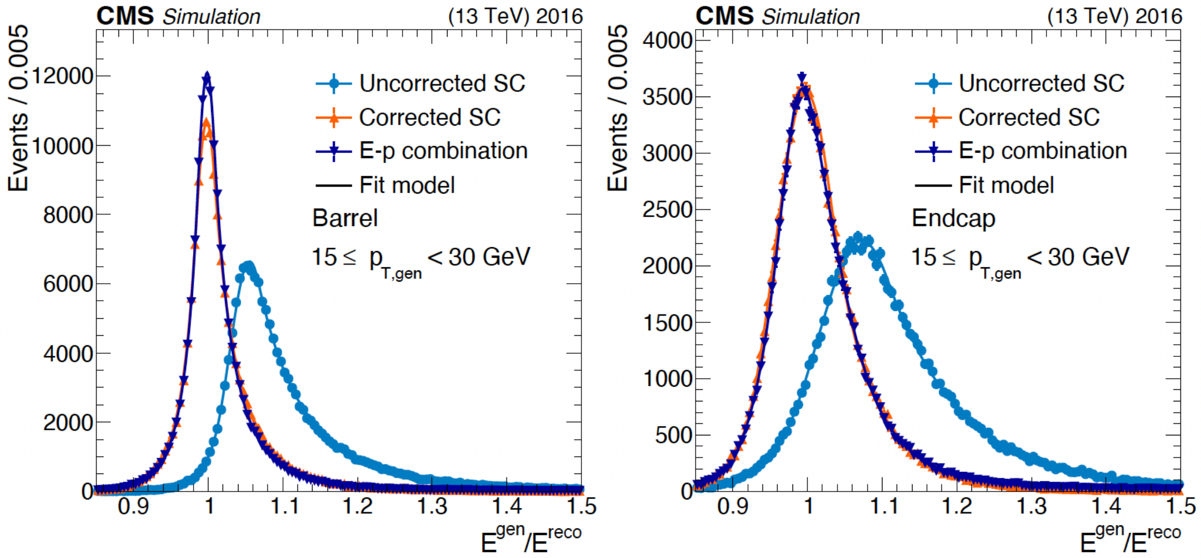


Figure 3.4: Ratio of the true to the reconstructed electron energy in the pT range 15–30 GeV with and without regression corrections, with a DSCB function fit overlaid, in 2016 MC samples for barrel (left) and endcap (right) electrons. Vertical bars on the markers represent the statistical uncertainties of the MC samples [52].

After applying regression corrections, small differences remain between the data and simulation in both the electron and photon energy scales and resolutions. In particular, the energy resolution in the simulation is better than that observed in data. Two dedicated approaches, referred to as the *fit method* and the *spreading method* [57], were developed during Run 1 to extract scale and spreading corrections using $Z \to ee$ events.

In the fit method, an analytic fit is performed to the reconstructed Z-boson invariant mass distribution, $m_{e^-e^+}$. The fit model is built from the convolution of a Breit-Wigner (BW) function, describing the Z lineshape, with a one-sided Crystal Ball (OSCB) function. The invariant mass spectra obtained from data and from simulated events are fitted independently, and the fitted parameters are compared to extract a relative energy scale offset. The parameters of the OSCB function, which account for calorimeter resolution effects and bremsstrahlung losses, are left free in the fit.

The spreading method, on the other hand, uses the simulated $m_{e^-e^+}$ distribution as a probability density function (PDF) in a maximum-likelihood fit to the data, allowing the additional smearing required to match the energy resolution observed in data to be determined.

The corrections derived from the two methods are combined to provide an optimized fine-tuning of the electron and photon energy scale and resolution in simulation. The resulting performance is validated using $Z \to ee$ events.

The performance of energy corrections in data, including regression, is illustrated by the reconstructed $Z \to e^-e^+$ mass distribution before and after corrections, as shown in Fig. 3.5. The regression leads to a clear improvement in the dielectron mass resolution for electrons from Z-boson decays in both the barrel and endcap regions. In addition, the absolute energy scale correction moves the peak of the reconstructed dielectron mass distribution closer to the world-average Z-boson mass value.

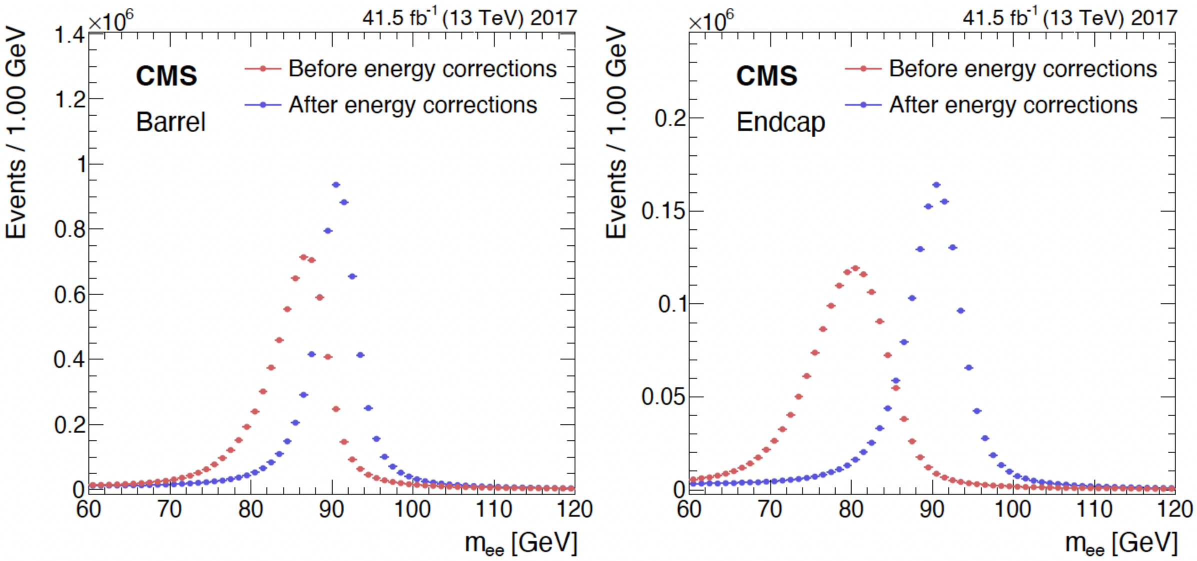


Figure 3.5: Reconstructed $m_{e^-e^+}$ distribution in data before and after applying the energy corrections.

3.4 Scale Factor Calculation Procedure

In CMS analyses, most studies rely on MC simulation to model the signal and some background processes. To ensure that the selection criteria applied to both data and simulation have the same effect, the MC must be corrected if differences exist. For electrons, reconstruction and identification efficiencies can differ between simulation and real data. The main challenge for MC is an accurate description of the material budget within the tracker volume, which strongly affects electron propagation in the CMS magnetic field. While the mechanical structure of the tracker is modeled reliably, the description of services and cabling for high voltage and data transmission is more complex and therefore harder to reproduce with high precision. This small residual inaccuracy in the material modeling represents the dominant source of electron mismodeling in the MC. To account for these differences and correct the simulation, dedicated SFs are computed. An overview of the procedure used to extract these SFs is provided here.

3.4.1 Tag and Probe Method

SFs are defined as ratio of efficiencies evaluated for data and for MC:

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}}. \tag{3.2}$$

Efficiencies in Eq. 3.2 are computed using the Tag and Probe (TnP) method [58]. First, electrons from a well known ressonance are selected, for example $Z \to e^- e^+$. Simulated electrons for these measurements are usually found in the Drell-Yann samples. The Drell-Yan (DY) process is the production of a lepton–antilepton pair in hadron–hadron collisions through an intermediate electroweak gauge boson. At the LHC, it proceeds primarily via quark-antiquark annihilation,

$$q\bar{q} \to \gamma^*/Z \to \ell^+\ell^-,$$

where $\ell = e, \mu, \tau$. The virtual photon contribution dominates at low dilepton invariant mass, while the Z boson contribution dominates near the Z pole. It is one of the cleanest SM channels and is widely used, not only for efficiency measurements but for detector calibration, and precision tests of electroweak physics as well.

After selecting a sample of electron candidates, additional requirements are imposed using the TnP method. One electron in the $e^- e^+$ pair, referred to as the *tag*, is required to satisfy tight identification and trigger criteria in order to ensure a high-purity signal sample. The second electron, called the *probe*, is required to satisfy only minimal selection criteria to remain unbiased and is used to evaluate the efficiency of a specific selection. The probe can either *pass* or *fail* the criteria under study.

The efficiency is then defined as

$$\epsilon = \frac{N_P}{N_P + N_F}, \tag{3.3}$$

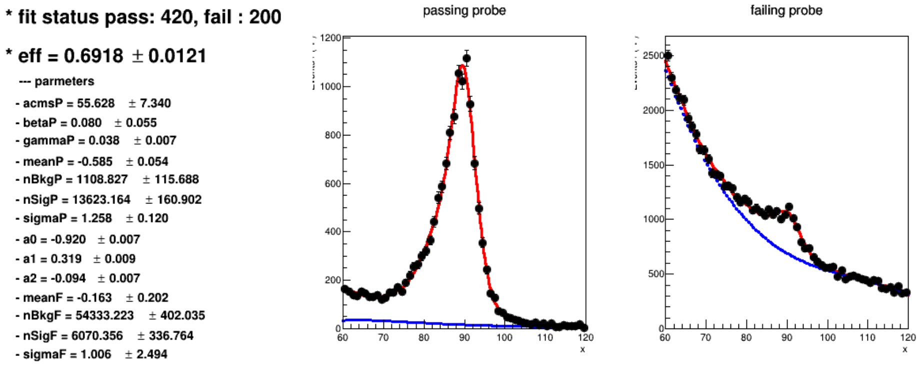
where N_P is the number of passing probes and N_F is the number of failing probes.

In data, the TnP sample contains not only signal events but also a background contribution. Therefore, the raw probe counts cannot be used directly. Instead, the signal yields in the passing and failing categories are extracted from a fit to the dilepton invariant mass distribution. This is performed using an extended maximum likelihood fit, in which the invariant mass spectrum is modeled as a sum of signal and background PDFs.

For a given category (pass or fail), the extended likelihood can be written as

$$\mathcal{L}(N_S, N_B) = e^{-(N_S+N_B)} \prod_{i=1}^{N} \left[N_S\, f_S(m_i) + N_B\, f_B(m_i) \right], \tag{3.4}$$

where m_i is the invariant mass of the i-th tag-probe pair, N is the total number of events in the fitted sample, N_S and N_B are the fitted signal and background yields, and $f_S(m)$ and $f_B(m)$ are the normalized signal and background PDFs, respectively. The efficiency is then computed using the extracted signal yields N_S^{pass} and N_S^{fail} in the two categories, and the corresponding statistical uncertainty is taken from the fit and propagated to 3.2. Example of the fitting procedure is given on Figure 3.6. The functions describing signal and background are fitted to invariant mass data. Area below the blue curve represents the background yield while the area between blue and red curves represents the signal yield. For MC, there is no fitting included since all TnP pairs constitute signal.



```
* fit status pass: 420, fail : 200

* eff = 0.6918 ± 0.0121
   --- parmeters
  - acmsP = 55.628   ± 7.340
  - betaP = 0.080   ± 0.055
  - gammaP = 0.038   ± 0.007
  - meanP = -0.585   ± 0.054
  - nBkgP = 1108.827   ± 115.688
  - nSigP = 13623.164   ± 160.902
  - sigmaP = 1.258   ± 0.120
  - a0 = -0.920   ± 0.007
  - a1 = 0.319   ± 0.009
  - a2 = -0.094   ± 0.007
  - meanF = -0.163   ± 0.202
  - nBkgF = 54333.223   ± 402.035
  - nSigF = 6070.356   ± 336.764
  - sigmaF = 1.006   ± 2.494
```

Figure 3.6: Example showing the fitting procedure in the TnP method for passing probes (left) and for failling probes (right). Data are shown as black points with error bars representing statistical uncertainties. The signal component is shown in red, while the background contribution is shown in blue.

3.4.2 Root Mean Square Method

The procedure described in Sec. 3.4.1 is used to calculate the nominal values of the SFs and their statistical uncertainties. In order to derive the corresponding systematic uncertainty and the final SF values, the so-called Root Mean Square (RMS) method is employed [59].

From Eq. 3.4, it can be seen that there is a level of freedom in how the likelihood function is defined, depending on the choice of the signal and background PDFs, f_S and f_B. The RMS method accounts for this by repeating the fit using different PDF choices and then combining the resulting SFs. In this analysis, four PDF variations are considered [2]:

---

[2] To quantify the dependence of the result on the choice PDF, two different PDFs are used for both signal and background, based on empirical studies.

- **nominal:** f_S and f_B are chosen based on the invariant mass shapes for signal and background; this configuration is referred to as nominal
- **alternative signal:** f_B is kept unchanged, while an alternative choice for f_S is used with respect to the nominal case
- **alternative background:** f_S is kept unchanged, while an alternative choice for f_B is used with respect to the nominal case
- **alternative signal and background:** both f_S and f_B are modified with respect to the nominal case

The final SF value is taken as the average of the four results:

$$SF = \frac{SF^{nom} + SF^{altSig} + SF^{altBkg} + SF^{altSigBkg}}{4}. \tag{3.5}$$

The corresponding systematic deviation for each variation is defined as

$$\mathrm{syst}_i = SF - SF_i, \tag{3.6}$$

where the index $i = 1, 2, 3, 4$ corresponds to the four different PDF choices.

The RMS contribution is then computed as

$$RMS = \sqrt{\frac{\sum_{i=1}^{4} \mathrm{syst}_i^2}{N-1}}, \tag{3.7}$$

Finally, the total uncertainty on the SF is calculated as

$$\sigma_{SF} = \sqrt{\left(\frac{RMS}{\sqrt{N}}\right)^2 + stat_{data}^2 + alt_{MC}^2}, \tag{3.8}$$

where alt_{MC} represents the variation induced by using a different MC sample, $stat_{data}$ represents statistical uncertainty from the nominal fit and $N = 4$.

The main advantage of the RMS method is the significant reduction of systematic uncertainties associated with the electron identification efficiency, especially in the low p_T region. These are reduced by approximately 30–40% [59] compared to the previous approach, in which systematic uncertainties were evaluated as the difference between the nominal and alternative measurements.

3.5 Electron Reconstruction Efficiency for Low p_T Electrons

Measuring the reconstruction efficiency of low-p_T electrons is crucial for multilepton analyses. To illustrate this, the p_T spectrum of the four electrons in the $H \to ZZ \to 4e$ channel is shown in Fig. 3.7.

$H \to ZZ \to 4\ell$ analysis uses electrons down to $p_T = 7$ GeV and, as can be seen in Fig. 3.7, there is a significant fraction of selected electrons with $p_T < 10$ GeV, a region that will be referred to as the low-p_T region.

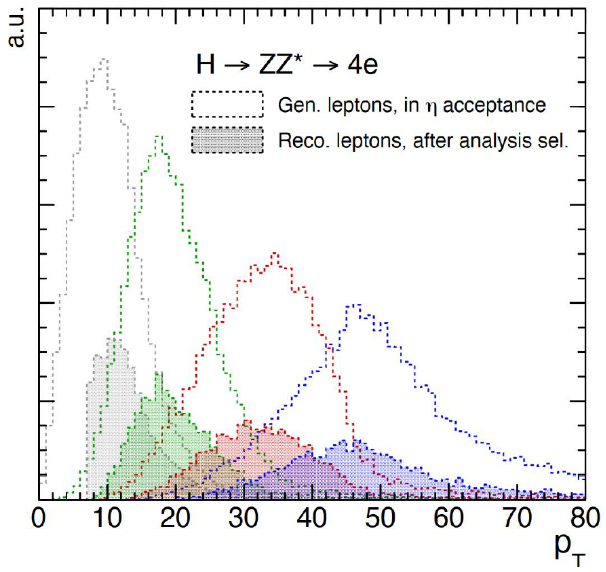


Figure 3.7: p_T spectrum of the four electrons in the $H \to ZZ \to 4e$ channel, obtained from simulated signal events. The distributions are shown for generator-level leptons within the detector acceptance (blank histograms) and for reconstructed leptons after the full analysis selection (filled histograms). The four electrons are ordered according to their transverse momentum, from the lowest to the highest.

To measure the electron reconstruction efficiency at CMS, the Z resonance is typically used. However, due to its relatively large mass, $m_Z = 91.2$ GeV, it is not an optimal source of low-p_T electrons. This is clearly illustrated in Fig. 3.8, where the invariant mass fit cannot reliably extract the signal contribution for failing probes in the low-p_T region. Therefore, a lower-mass resonance is required.

A suitable alternative is the J/ψ resonance, which has an invariant mass of approximately $m_{J/\psi} = 3.1$ GeV. The J/ψ is a bound state of a charm quark and an anti-charm quark ($c\bar{c}$), and it provides an abundant source of low-p_T electrons through the decay $J/\psi \to e^+e^-$. These events are accessed using the B-parking dataset.

3.5.1 B-parking Dataset

The search for new physics often motivates scenarios in which hypothetical particles have low masses and very weak couplings. Such processes are difficult to observe at the LHC due to the large rate of SM backgrounds [60, 61]. To keep the overall trigger rate manageable, the standard data acquisition strategy typically relies on relatively high trigger thresholds in order to suppress SM events. As a consequence, potentially interesting signal events with low energy and momentum can be rejected already at the trigger level. Low-mass BSM particles decaying into final states with soft jets or low-energy lepton pairs therefore represent a challenging signature at the LHC.

The traditional analysis workflow at the LHC follows a well-established paradigm: proton–proton collision events are selected online by the trigger system, stored to disk in raw data format, and then reconstructed and analyzed offline. The offline reconstruction is designed to produce the highest-quality physics objects for analysis. Since it does not

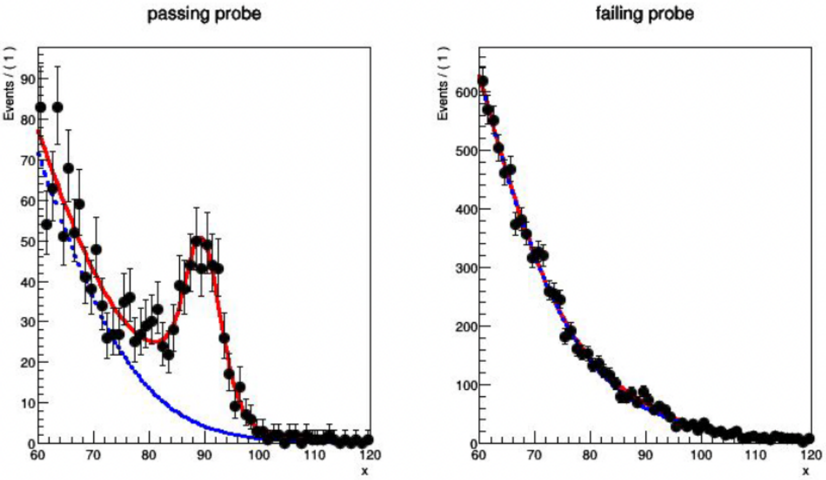


Figure 3.8: Distribution of the invariant mass of TnP electron pairs for passing probes (left) and failing probes (right), using $Z \to e^- e^+$ events in data for year 2022, in the range $5 < p_T^{\text{probe}} < 7$ GeV. Data are shown as black points with error bars representing statistical uncertainties. The signal component is shown in red, while the background contribution is shown in blue. The same color scheme is used for all subsequent $Z \to e^- e^+$ plots showing invariant mass distributions used for the determination of the electron identification efficiency.

need to run at the same pace as the data acquisition system and is not constrained by the strict low-latency requirements of the trigger, it can employ more sophisticated and computationally expensive algorithms.

Data parking is a relatively new strategy at the LHC, inspired by approaches commonly used in fixed-target experiments, where raw data are recorded and processed for analysis at a later time. In addition, the data-parking strategy allows lower trigger thresholds, increasing the experimental acceptance for low-mass physics processes. This leads to a substantially higher event collection rate, which can exceed the available computing resources for prompt reconstruction. In such cases, the parked data stream is transferred directly to tape storage without being processed and is kept in raw format until sufficient computational resources become available for reconstruction, for example during periods between data-taking [62].

A natural strategy to obtain a large and unbiased sample of low-p_T electrons would be to select $J/\psi \to e^+ e^-$ events directly with dedicated electron triggers. However, during Run 2 this approach was not feasible at the trigger level due to rate limitations and the comparatively challenging online reconstruction of low-energy electrons in the presence of large SM backgrounds. Therefore, an alternative dataset was required.

For the measurement of low-p_T electron reconstruction efficiencies, Run 2 CMS data collected with the B-parking strategy are used. The trigger strategy was originally designed to maximise the number of $B \to K J/\psi$ decays and is based on the selection of $b\bar{b}$ events using a "tag-side" logic. It requires the presence of a single muon, exploiting the fact that semileptonic decays to muonic final states account for approximately 20% of all b-hadron decays. Here, B denotes a meson containing a bottom quark (b) bound with a light quark (u, d, s, or c), while K denotes a kaon, i.e. a meson containing a strange quark (s) bound with a light quark (u or d). Both B mesons and kaons can be charged or neutral.

The B-parking dataset collected in Run 2 contains approximately 10^{10} b-hadron decays. In comparison, other datasets highly enriched in b hadrons collected during the same period comprise at most 5×10^8 unbiased b decays. The B-parking strategy therefore provides an unprecedented statistical sample for precision studies of low-momentum final states.

The B-parking dataset was primarily designed to enable precision measurements in B-hadron decays, in particular tests of lepton universality in rare decays involving low-momentum lepton pairs. In this analysis, however, the same dataset is exploited for a different purpose: the measurement of the electron reconstruction efficiency in the low-p_T regime. The B-parking triggers provide a source of low-p_T electrons originating from $J/\psi \to e^+e^-$ decays, where the J/ψ meson is produced in B-hadron decays such as $B \to KJ/\psi$.

The B-parking triggers were deployed with multiple configurations throughout Run 2, corresponding to different instantaneous luminosity conditions. The configurations differ in the applied L1 and HLT muon p_T thresholds, as well as the impact-parameter significance requirement at HLT. The main trigger settings used for the B-parking data-taking period are summarised in Table 3.1.

Setting	Peak L_{inst} [$10^{34}\,\text{cm}^{-2}\text{s}^{-1}$]	L1 p_T [GeV]	HLT p_T [GeV]	HLT IP$_{\text{sig}}$ threshold	Purity	Peak rate [kHz]
1	1.7	12	12	6	0.92	1.5
2	1.5	10	9	6	0.87	2.8
3	1.3	9	9	5	0.86	3.0
4	1.1	8	8	5	0.83	3.7
5	0.9	7	7	4	0.59	5.4

Table 3.1: Summary of the tag-side muon trigger requirements at L1 and HLT: the L1 and HLT muon p_T thresholds and the HLT muon impact-parameter significance requirement, IP$_{\text{sig}}$. Also shown are the trigger purity and the peak trigger rate. All quantities are given as a function of the peak instantaneous luminosity, L_{inst} [63]. Trigger purity represents the fraction of events accepted by a trigger that actually contain the targeted physics signature .

3.5.2 Computation of Low-p_T Electron Reconstruction Efficiency

The standard approach for deriving reconstruction SFs is based on identifying superclusters in the event and checking whether they are successfully reconstructed as electrons. Supercluster–electron pairs are then formed around a known resonance, and the reconstruction efficiency is measured, typically in bins of the supercluster transverse energy and pseudorapidity. However, this strategy becomes ineffective in the low-p_T region, where a large fraction of electrons are reconstructed as tracker-driven electrons. For this reason, a novel approach is adopted: general tracks are used as probes, and the efficiency is determined by evaluating whether these tracks are reconstructed as electrons.

On top of the high amount of background in the B-parking dataset, another issue comes from the low branching ratio $\mathcal{B}(B^+ \to K^+e^+e^-) = 4.5 \times 10^{-7}$ [16]. To address this challenge, a stringent event selection is implemented to isolate the target signal topology, as illustrated in Fig. 3.9.

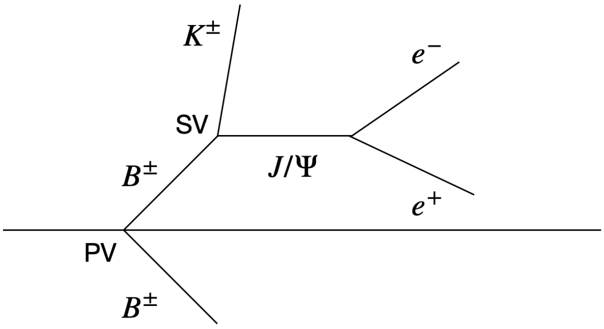


Figure 3.9: Topology of the $B \to KJ/\psi$ decay. A pair of B mesons is produced at the PV. One of them, shown here, decays at the secondary vertex (SV) into a kaon and a J/ψ, followed by the decay $J/\psi \to e^-e^+$.

First, the tag electron is required to satisfy basic p_T and MVA (multivariate score) requirements: $p_T > 7$ GeV and MVA > 0.95. These thresholds are intentionally kept relatively mild in order to avoid unnecessary signal loss. The probe is required to have $p_T > 5$ GeV, corresponding to the lowest transverse momentum for which the efficiency is measured.

The TnP candidates are required to originate from the same secondary vertex (SV). In addition, exactly three tracks must be associated with the SV, with a total charge of ±1 and an invariant mass $m_{inv}^{tracks} > 4.5$ GeV, tuned to select B-hadron decays. The invariant mass of the tag-and-probe pair is required to lie in the range 2 GeV $< m_{inv}^{tag,probe} < 4$ GeV, corresponding to the region around the J/ψ resonance.

Further requirements are applied to control the angular separation between the decay products, the scalar sum of track p_T associated with the SV, and the quality of the SV fit. The full selection is summarized in Table 3.2.

After applying the full selection and forming the TnP pairs, an appropriate choice of PDFs is required to extract the signal yield from the invariant mass fit. In the passing-probe category, the selection is highly exclusive, suppressing the background almost completely. The dominant contribution in this region is the J/ψ peak, while an additional resonant structure around 3.7 GeV is also visible. This corresponds to the J/ψ' — an excitation of the J/ψ charmonium state — and is treated as background with respect to the J/ψ signal.

The invariant mass distribution for passing probes is therefore modeled using two Double-Sided Crystal Ball (DSCB) functions, one for the J/ψ resonance and one for the

Cut number	Criteria
1.	Tag electron: $p_T > 7$ GeV,
2.	$MVA > 0.95$
3.	Probe track $p_T > 5$ GeV
4.	Opposite charge for tag and probe
5.	2 GeV $< m_{inv}^{tag,probe} <$ 4 GeV
6.	Tag and probe come from the same secondary vertex
7.	Secondary vertex has 3 tracks,
8.	total charge ±1,
9.	$m_{inv}^{tracks} > 4.5$ GeV
10.	Secondary vertex fit probability > 0.001
11.	p_T sum (tag,probe,kaon) > 3 GeV
12.	δR (tag,probe) > 0.03
13.	min δR (tag, probe, kaon) > 0.03

Table 3.2: Summary of the tag-and-probe selection used to extract $J/\psi \to e^+e^-$ candidates from $B^\pm \to K^\pm J/\psi$ decays in the B-parking dataset.

J/ψ'. The DSCB function is defined as:

$$\mathrm{DSCB}(y;\mu,\sigma,\alpha_L,n_L,\alpha_R,n_R) = \begin{cases} N\exp\Big(-\dfrac{\xi(y)^2}{2}\Big), & -\alpha_L \le \xi(y) \le \alpha_R, \\ N\exp\Big(-\dfrac{\alpha_L^2}{2}\Big)\Big(\dfrac{\alpha_L}{n_L}\big(\dfrac{n_L}{\alpha_L} - \alpha_L - \xi(y)\big)\Big)^{-n_L}, & \xi(y) < -\alpha_L, \\ N\exp\Big(-\dfrac{\alpha_R^2}{2}\Big)\Big(\dfrac{\alpha_R}{n_R}\big(\dfrac{n_R}{\alpha_R} - \alpha_R + \xi(y)\big)\Big)^{-n_R}, & \xi(y) > \alpha_R, \end{cases} \tag{3.9}$$

where N is the normalization constant, $\xi(y) = (y-\mu)/\sigma$, the variables μ and σ are the parameters of the Gaussian core, and the α_R (α_L) and n_R (n_L) parameters control the right (left) tails of the function. The DSCB is well suited for passing probes because the selection is highly exclusive and the background is negligible, allowing a sharp Gaussian peak with asymmetric tails to describe the resonant signal accurately.

For failing probes, where the background contribution is significant, the signal is modeled with a Gaussian core plus exponential tail (see 3.10), while the background is modeled with a third-order polynomial. Unlike the DSCB function, this parametrization provides a better description of distributions exhibiting a less pronounced peak.

$$f(x;\,\bar{x},\sigma,k) = \begin{cases} \exp\Big[-\dfrac{1}{2}\Big(\dfrac{x-\bar{x}}{\sigma}\Big)^2\Big], & \dfrac{x-\bar{x}}{\sigma} \ge -k, \\ \exp\Big[\dfrac{k^2}{2} + k\Big(\dfrac{x-\bar{x}}{\sigma}\Big)\Big], & \dfrac{x-\bar{x}}{\sigma} < -k. \end{cases} \tag{3.10}$$

The *GaussExp* function is used here because the failing-probe distribution contains both signal and a sizable background; the Gaussian core captures the main signal peak, while the exponential tail accounts for the effects that skew the distribution. As is common in efficiency measurements, the background is modeled with a polynomial, since its exact shape is not well known.

The fits for passing and failing probes for 2018 dataset are shown in Figures 3.10 and 3.11. Although the $10 < p_T < 20$ GeV region is not the primary focus of this study, it provides a useful cross-check, as the electron reconstruction efficiency in this range is well measured using Z boson decays.

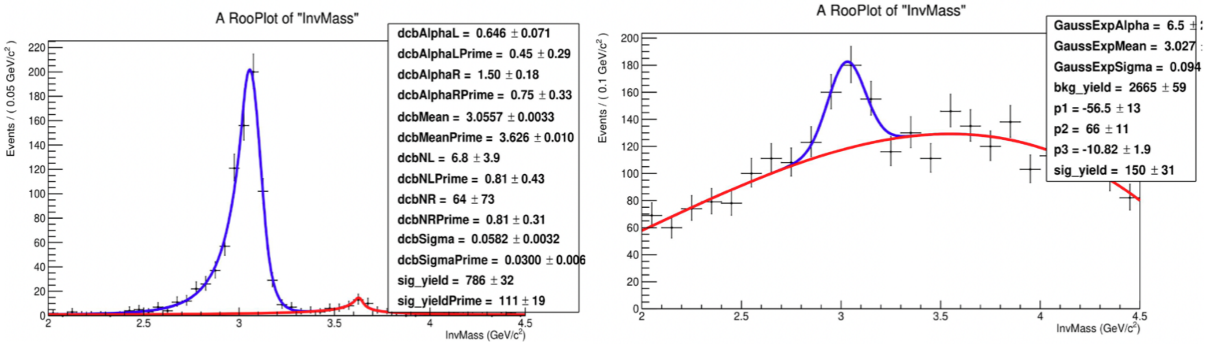


Figure 3.10: Distribution of the invariant mass of TnP electron pairs for passing probes (left) and failing probes (right), using $J/\psi \to e^- e^+$ events in 2018 data, in the range $5 < p_T^{\text{probe}} < 10$ GeV. Data are shown as black markers with error bars corresponding to statistical uncertainties. The signal component is shown in blue, while the background contribution is shown in red.

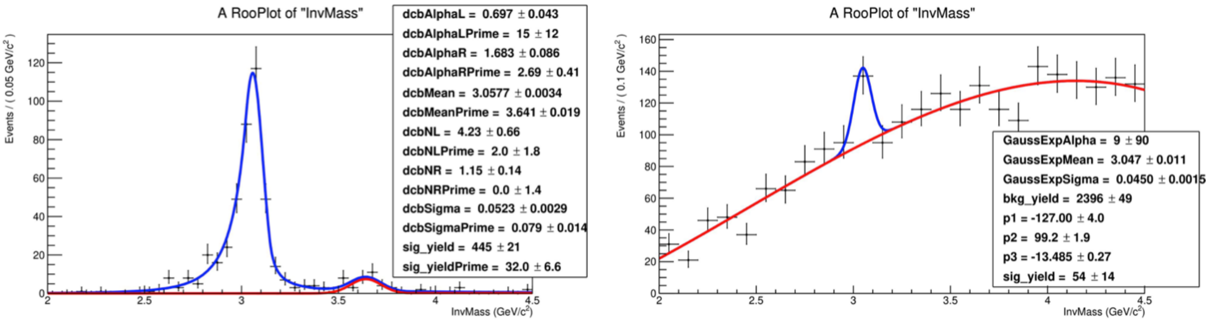


Figure 3.11: Distribution of the invariant mass of TnP electron pairs for passing probes (left) and failing probes (right), using $J/\psi \to e^- e^+$ events in 2018 data, in the range $10 < p_T^{\text{probe}} < 20$ GeV. Data are shown as black markers with error bars corresponding to statistical uncertainties. The signal component is shown in blue, while the background contribution is shown in red.

In comparison to the known issues affecting the $Z \to e^- e^+$ fits in the low-p_T region, the invariant mass distribution shown in Fig. 3.10 exhibits a well-resolved J/Ψ peak, allowing for a straightforward determination of the reconstruction efficiency.

The electron reconstruction efficiencies, together with their statistical uncertainties, extracted using the J/ψ and Z resonances are shown in Fig. 3.12. The J/ψ measurement provides the main result in the 5–10 GeV range, while the overlap region 10–20 GeV is used as a cross-check between the two methods. Below 10 GeV, the J/ψ-based result is consistent with the expectation derived from the Z measurement, and good agreement is observed in the 10-20 GeV region. This is the first time the electron reconstruction efficiency has been measured in CMS in the $p_T < 10$ GeV regime. Currently, in $H \to ZZ \to 4\ell$ analysis, the reconstruction scale factor below 10 GeV is typically taken to be equal to that measured in the 10–20 GeV bin, with a large uncertainty assigned to account for this, implicitly assuming that at lower p_T SFs remain the same.

While MC simulation was not available in the AOD format used for this study and therefore scale factors could not be derived, the results provide a proof of concept and demonstrate good consistency with the standard measurement procedure, indicating strong potential for use in future analysis.

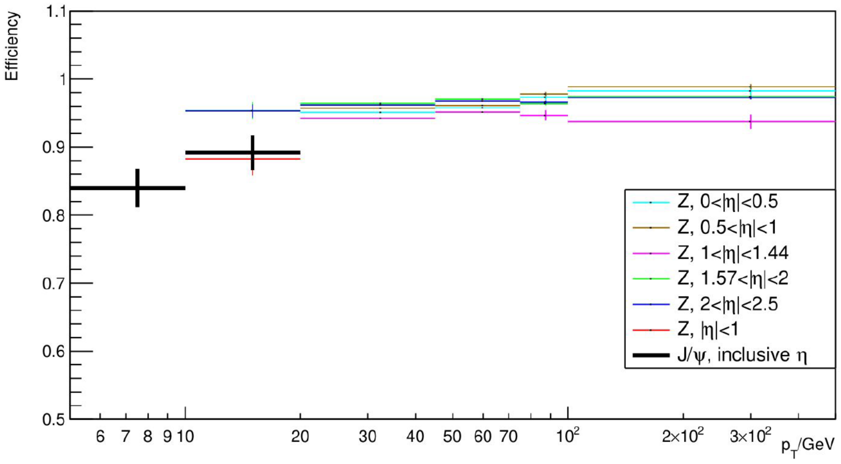


Figure 3.12: Electron reconstruction efficiency as a function of the probe transverse momentum, measured with 2018 data, using $Z \to e^-e^+$ events in different η regions (cyan: $0 < |\eta| < 0.5$, brown: $0.5 < |\eta| < 1$, magenta: $1 < |\eta| < 1.44$, green: $1.57 < |\eta| < 2$, blue: $2 < |\eta| < 2.5$, red: $|\eta| < 1$) and using $J/\psi \to e^-e^+$ events (black, inclusive η).

3.6 Electron Identification using Boosted Decision Tree

To study the final states in $H \to ZZ \to 4\ell$ analysis, it is important to properly distinguish prompt electrons, from the background. To achieve this goal, two types of algorithms are usually used at CMS. The first one is a cut based-identification (ID), where series of cuts are applied on sensitive variables. The second one uses multivariate technique BDT, and is described in this chapter.

3.6.1 XGBoost Algorithm

BDTs are widely used multivariate methods for classification and regression. A decision tree separates events by applying a sequence of binary cuts on the input variables, such that events with similar properties end up in the same terminal node, called a *leaf.* Each leaf is assigned a score, which provides the final prediction for events falling into that region of phase space. While a single decision tree is simple and interpretable, its performance is typically limited and can be sensitive to statistical fluctuations in the training sample.

Boosting improves the performance by combining many decision trees into an ensemble. In gradient boosting, the trees are trained sequentially, where each new tree is constructed to correct the mistakes of the previous ones. The final BDT score is obtained from the sum of the contributions from all trees. In this work, the XGBoost implementation of gradient-boosted decision trees [64] is used.

The performance of the XGBoost BDT depends on a set of parameters, such as the maximum tree depth, the learning rate, and the number of trees. These parameters are optimized using Bayesian optimization [65], which efficiently explores the parameter space by iteratively selecting new configurations based on the performance of previously tested ones. This approach typically converges to a near-optimal configuration with significantly fewer trials than a grid search.

Parameters, controlling electron BDT training are described below:

- **Minimum leaf weight (`min_child_weight`):** Controls the minimum sum of event weights (or number of events) required in a leaf. It prevents overfitting by ensuring that leaves contain enough events.

- **Feature fraction (`colsample_bytree`):** Fraction of input variables (features) randomly chosen to build each tree. Lower values introduce more randomness and can improve generalization.

- **Maximum tree depth (`max_depth`):** Maximum depth of each decision tree. Shallow trees are simpler, while deeper trees can capture more complex patterns but may overfit.

- **Row fraction (`subsample`):** Fraction of training events used to build each tree. Using less than all events introduces randomness, which can improve robustness.

- **Minimum loss reduction (`gamma`):** Minimum loss reduction required to make a split on a leaf. Higher values make the tree more conservative in splitting.

- **L1 regularization (`reg_alpha`):** L1 regularization term on the leaf scores. L1 adds a penalty proportional to the absolute value of the leaf weights, helping reduce overfitting.

- **L2 regularization (`reg_lambda`):** L2 regularization term on the leaf scores. L2 adds a penalty proportional to the square of the leaf weights, which also helps control overfitting and stabilizes training.

3.6.2 Boosted Decision Tree Training Variables

BDT training variables are chosen to maximise the separation between prompt electrons and the background, which can originate from photon conversions, hadrons misidentified as electrons, secondary electrons and semileptonic decays of b or c quarks [52]. Variables used for identification training can be grouped into 5 categories: calorimetric (SC in ECAL), tracker, SC - track matching, isolation and pileup (PU).

Electron isolation quantifies the activity in the vicinity of an electron candidate and is one of the variables used in electron identification training. Isolation variables can be categorized into particle-based and cluster-based types. Particle-based isolation sums the transverse momentum or energy of PF candidates in a cone around the electron direction. The charged hadron isolation corresponds to the scalar sum of transverse momentum of charged hadron candidates, the neutral hadron isolation corresponds to the sum of transverse energy of neutral hadron candidates, and the photon isolation corresponds to the sum of transverse energy of photon candidates within the cone. These particle-based sums provide a detailed reconstruction-level view of the surrounding activity by combining tracker and calorimeter information.

Cluster-based isolation uses calorimeter energy deposits and tracker activity directly without full PF reconstruction. The ECAL cluster isolation measures the transverse energy of electromagnetic calorimeter clusters, while the HCAL cluster isolation measures the transverse energy of hadronic calorimeter clusters within the cone of $\Delta R = 0.3$ around the electron. Tracker isolation corresponds to the sum of transverse momenta of tracks within a cone of $\Delta R = 0.3$. These cluster-based variables are generally more robust against PF reconstruction fluctuations and pileup effects, providing a more direct representation of detector-level energy deposits.

Studies performed by the EGamma Physics Object Group have shown that electron identification in Run 3 achieves improved performance when cluster-based isolation variables are employed rather than particle-based sums [66]. Motivated by this observation, the $H \rightarrow ZZ \rightarrow 4l$ electron identification, trained as part of this thesis, has been constructed using cluster-based isolation variables to exploit the improved discriminating power observed in the Physics Object Group (POG) studies.

Here, the key variables used in training are described:

- **Hadronic over electromagnetic energy ratio (H/E):** The H/E ratio is a longitudinal shower-shape variable, defined as the ratio of the energy deposited in the HCAL (H) to that in the ECAL (E) by a given object. It is a key variable for electron and photon identification, as it effectively discriminates between electromagnetic objects, which deposit minimal energy in the HCAL, and hadronic jets, which typically deposit substantial HCAL energy. For electrons, H is taken as the HCAL energy within a cone of $\Delta R < 0.15$ around the electron's position in the ECAL.

Another approach to reject jets with a high electromagnetic fraction uses the shape of the electromagnetic shower in the ECAL. Shower shape variables are defined to capture the differences in the geometrical distribution of energy deposits produced by prompt electrons versus those originating from hadrons in jets. Most important are:

- **$\boldsymbol{\sigma_{i\eta i\eta}}$:** a lateral shower-shape variable that quantifies the spread of an electromagnetic shower in the η direction. It is defined as the logarithm of the energy-weighted root-mean-square (RMS) of the shower in units of crystals. This variable is among the most important for electron and photon identification, due to its strong discriminating power between electrons/photons (signal) and jets (background). It is also sensitive to ECAL noise, particularly for low-to medium-energy electrons and photons. The definition of this variable got changed in Run 3 wrt. Run 2 to cope with increased ECAL noise [67]. Showers produced by genuine electrons or photons are generally narrower than those from jets. However, if crystals containing noise are included in the calculation of $\sigma_{i\eta i\eta}$, the apparent shower width of real electromagnetic objects increases, making it more similar to that of jets. As a result, the discriminating power of $\sigma_{i\eta i\eta}$ decreases with increasing ECAL noise. By excluding crystals that fall below the noise threshold, the discrimination power can be restored. Procedure for noise cleaning is depicted on Figure 3.13. Redefinition of $\sigma_{i\eta i\eta}$ shows ID performance improvements in the endcaps, as seen on Figure 3.14.

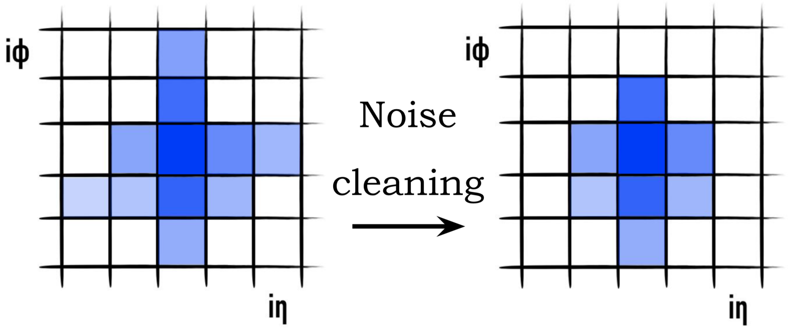


Figure 3.13: (Left) All crystals in the 5x5 area around the most energetic crystal are considered for the $\sigma_{i\eta i\eta}$ definition. This leads to a broader shower (Right). Crystals not passing noise thresholds are dropped. This leads to a narrower shower [67].

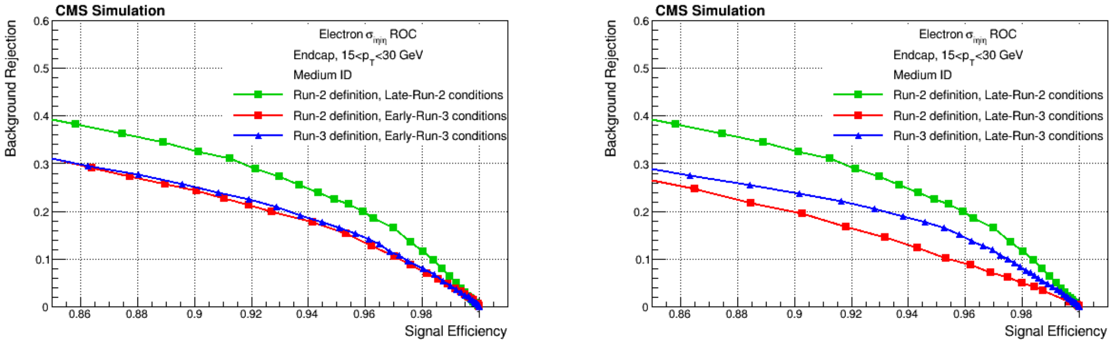


Figure 3.14: The redefinition of $\sigma_{i\eta i\eta}$ leads to significant performance improvements in the endcaps under Late-Run-3 conditions (right plot), where the ECAL noise is relatively high. For Early-Run-3 conditions (left plot), where the noise level is lower, the noise-cleaning procedure provides a smaller improvement [67].

Several tracker-supercluster matching variables are used to suppress hadronic backgrounds and improve electron identification:

- **$\mathbf{1/E - 1/p}$:** This variable compares the energy E measured in the ECAL with the momentum p measured by the tracker. For a prompt electron, the ECAL energy and track momentum are expected to match, yielding $\frac{1}{E} - \frac{1}{p} \approx 0$. For hadrons or jets misidentified as electrons, the ECAL energy and tracker momentum often do not match, resulting in a larger deviation from zero, making this a powerful discriminator against hadronic backgrounds.

- **$|\Delta\boldsymbol{\eta}_{\text{seed}}| = |\boldsymbol{\eta}_{\text{seed}} - \boldsymbol{\eta}_{\text{track}}|$:** Here, η_{seed} is the η position of the seed crystal in the supercluster, and η_{track} is the track η extrapolated from the innermost tracker layer.

- **$|\Delta\eta_{\mathrm{SC}}| = |\eta_{\mathrm{SC}} - \eta_{\mathrm{track}}|$:** Similar to the previous variable, this uses the energy-weighted position of the full supercluster (η_{SC}) instead of the seed crystal.

Another important background to consider are non-prompt electrons produced by photon conversion in the tracker. When photon conversions take place inside the tracker volume, the first hit of the electron tracks from the converted photons is not likely to be located in the innermost tracker layer. As no missing inner hits are expected for prompt electrons, this variable helps discriminate against photon conversions. All variables used in HZZ electron ID training for 2022 are summarized in Table 3.3

3.6.3 Electron Identification Training for 2022 Data-taking Period

Electron ID training, targeting the $H \to ZZ \to 4\ell$ analysis for the year 2022, is performed using the 2022 Ultra Legacy (UL) DY dataset. The signal class is defined as the set of reconstructed (RECO) electrons matched to prompt generated (GEN) electrons within a cone of size $\Delta R < 0.1$. The background class is defined as the set of RECO electrons that either have no matching GEN electron or are matched to non-prompt GEN electrons, typically originating from jets. RECO electrons matched to GEN electrons from leptonic τ decays are excluded from both the signal and background samples. With this definition of signal and background, the distributions of the input variables used for the training are shown in Figs. 3.15–3.18.

Phase space is split into two p_T regions: $5 < p_T < 10$ GeV and $p_T > 10$ GeV, and three η regions: inner barrel ($|\eta| < 0.8$), outer barrel ($0.8 < |\eta| < 1.479$), and endcap ($|\eta| > 1.479$). XGBoost BDT, described in 3.6.1 is used for training. The corresponding parameter space, explored with Bayesian optimization, is reported in Table 3.4.

The classifier performance is compared for trainings using particle-based isolation variables and cluster-based isolation variables in six p_T–η regions using receiver operating characteristic (ROC) curves, which show the background efficiency as a function of the signal efficiency, as shown in Fig. 3.19, 3.20 and 3.21. Given its superior performance across all regions, the cluster-based ID is retained and integrated into the CMS software (CMSSW).

The HZZ electron identification working point - a boolean flag indicating whether an object is classified as an electron, is chosen to match the signal efficiency target established in Run 2. At the analysis stage, the key metric is the amount of background passing this selection for the given signal efficiency. These background rates are summarized in Table 3.5 and compared with the corresponding values from Run 2 (year 2017).

As in the ROC curves, the Table 3.5 shows that cluster-based training results in lower background leakage compared to training using particle-based isolation variables. The 2022 cluster-based training even slightly outperforms the 2017 results.

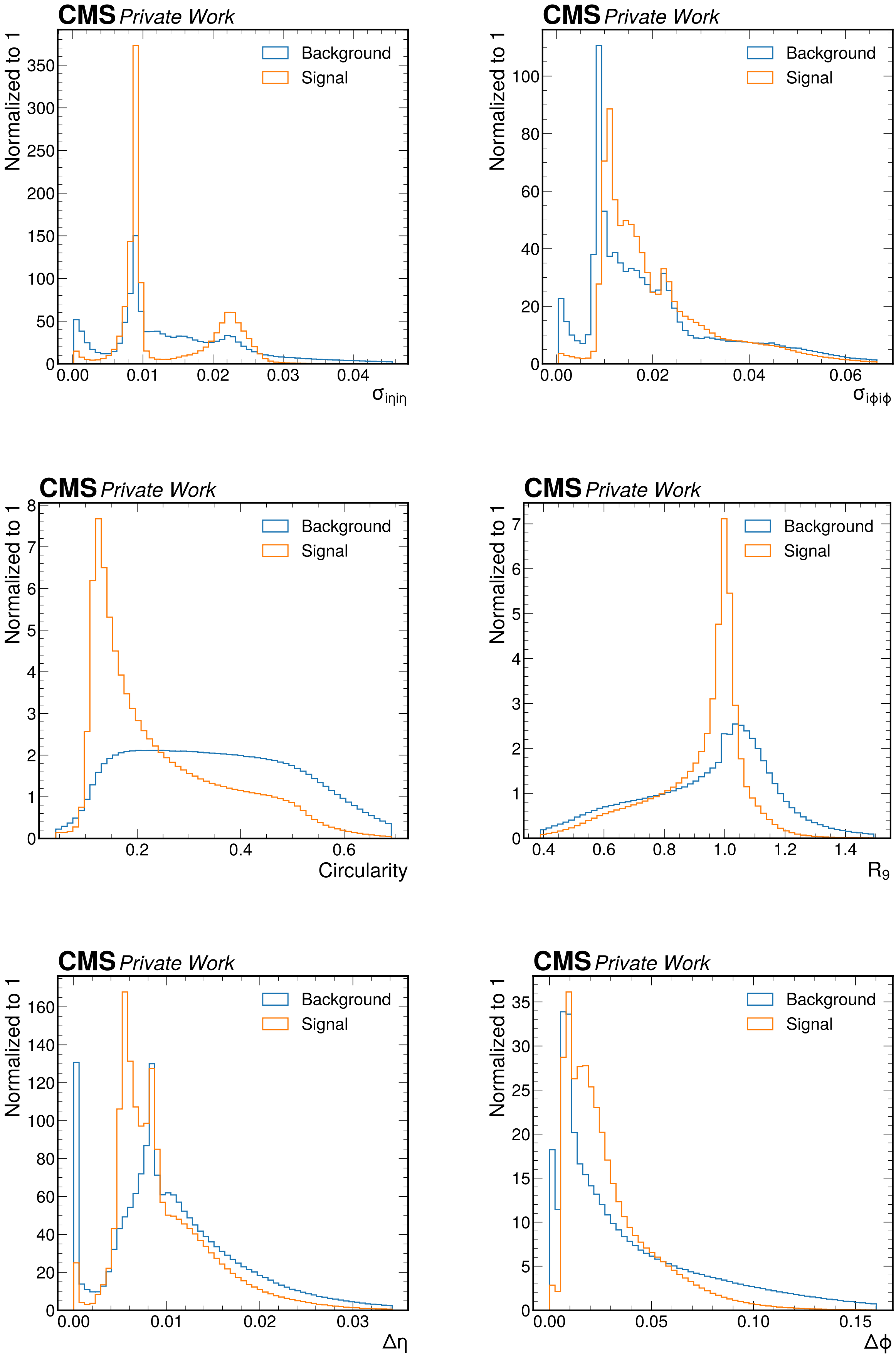


Figure 3.15: Distribution of variables used for electron identification training for Signal (orange) and Background (blue). Both distributions are normalized to unity. Some variables are shown in logarithmic scale to better emphasize their separation power. Variables are described in Table 3.3.

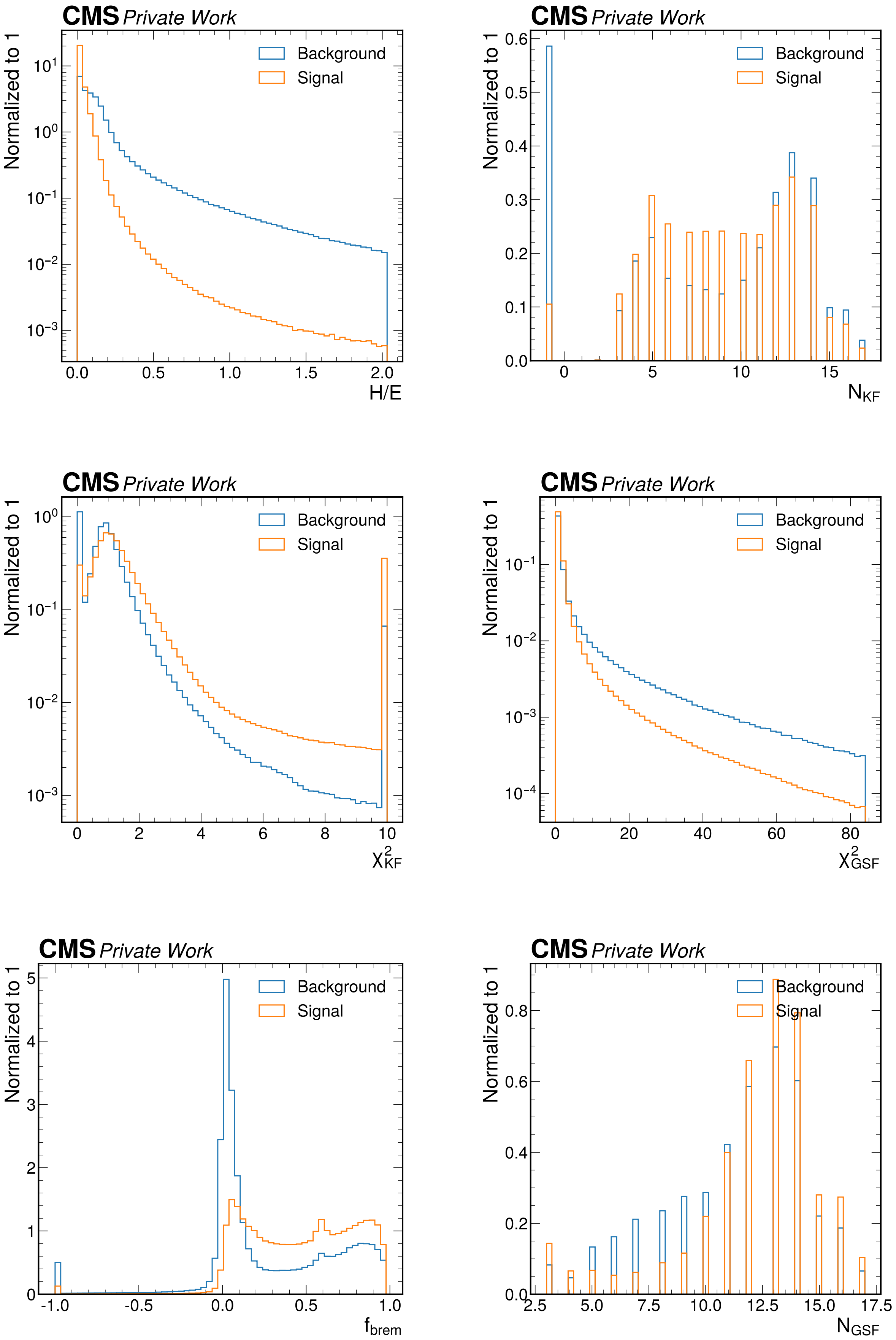


Figure 3.16: Distribution of variables used for electron identification training for Signal (orange) and Background (blue). Both distributions are normalized to unity. Some variables are shown in logarithmic scale to better emphasize their separation power. Variables are described in Table 3.3.

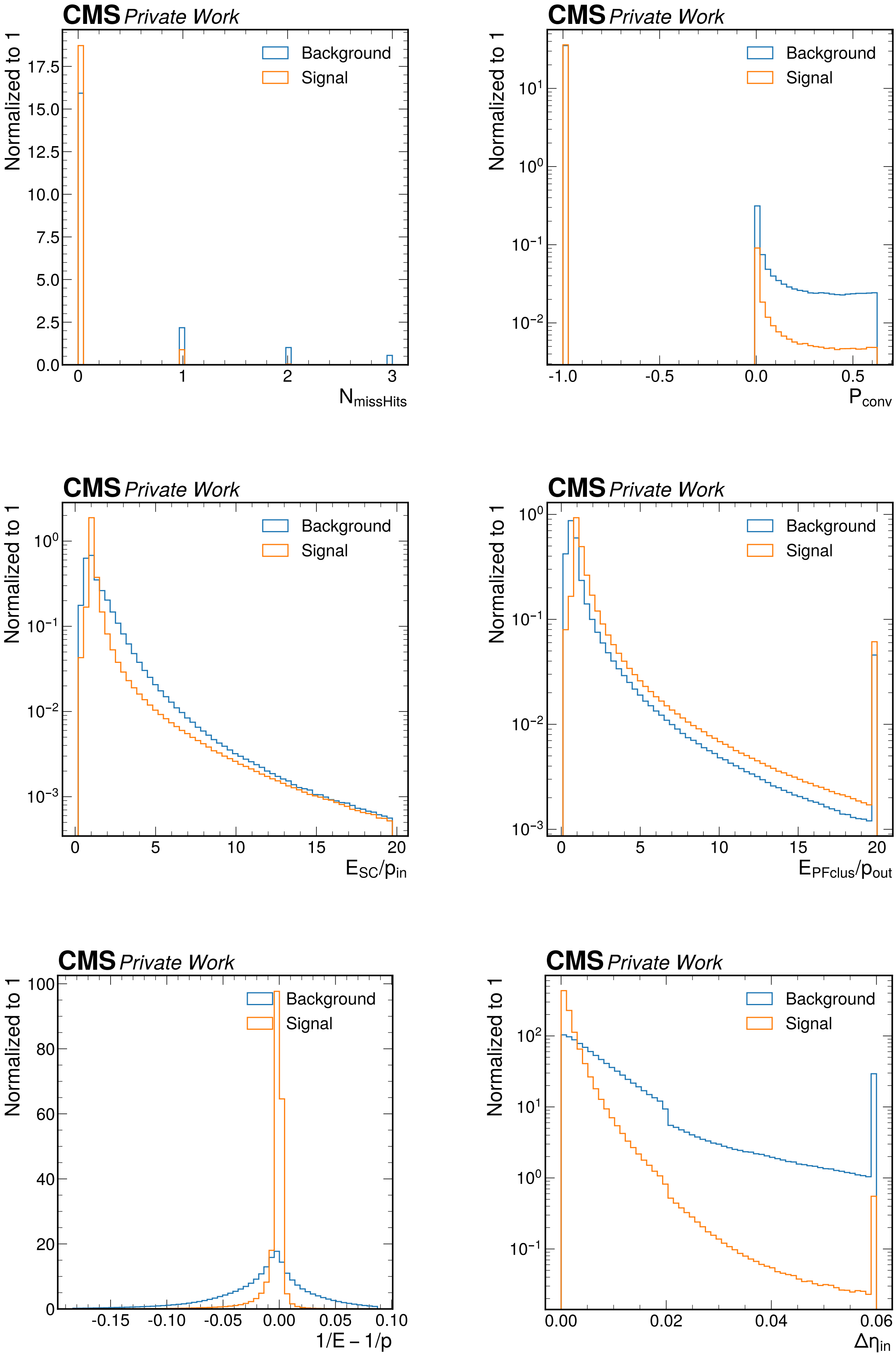


Figure 3.17: Distribution of variables used for electron identification training for Signal (orange) and Background (blue). Both distributions are normalized to unity. Some variables are shown in logarithmic scale to better emphasize their separation power. Variables are described in Table 3.3.

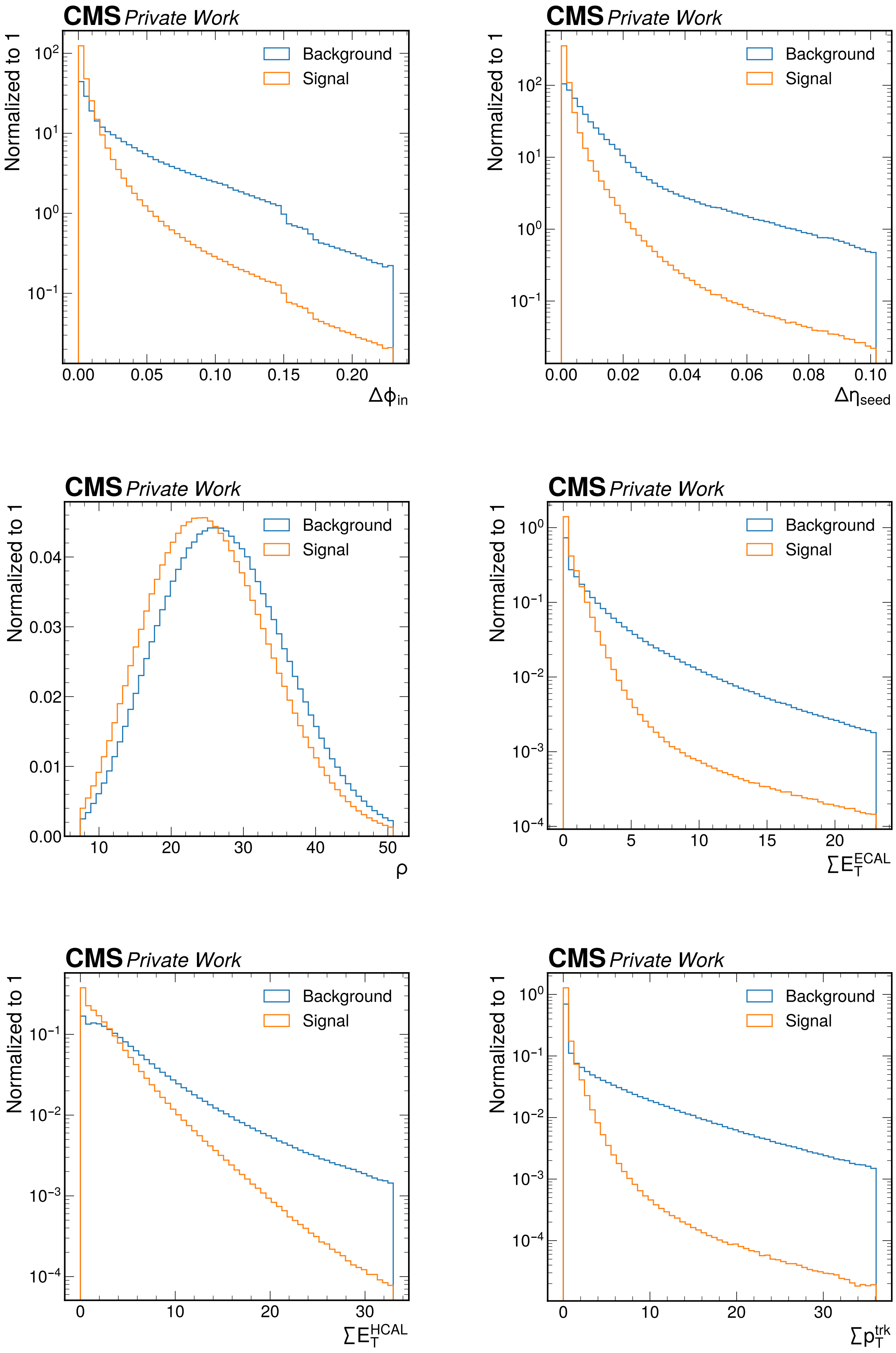


Figure 3.18: Distribution of variables used for electron identification training for Signal (orange) and Background (blue). Both distributions are normalized to unity. Some variables are shown in logarithmic scale to better emphasize their separation power. Variables are described in Table 3.3.

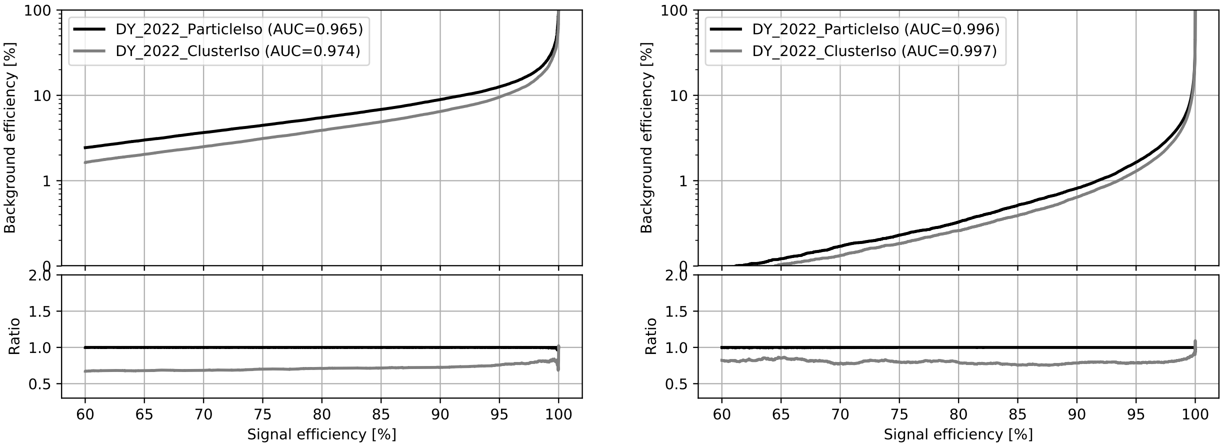


Figure 3.19: ROC curves in the inner barrel region for $5 < p_T < 10$ GeV (left) and $p_T > 10$ GeV (right), for the training using PF (black) and Cluster based (grey) isolation sums.

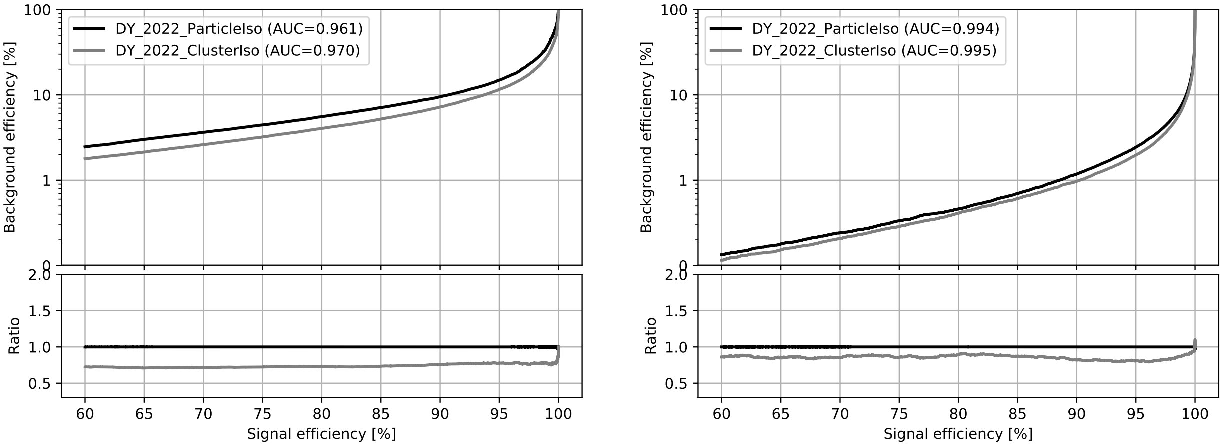


Figure 3.20: ROC curves in the outer barrel region for $5 < p_T < 10$ GeV (left) and $p_T > 10$ GeV (right), for the training using PF (black) and Cluster based (grey) isolation sums.

SC in the ECAL	
$\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$	Variance of the log-energy-weighted crystal energy distribution along η and ϕ
Δ_η, Δ_ϕ	Effective supercluster width along η and ϕ
H/E	Ratio of HCAL energy in a cone of radius $\Delta R = 0.15$ around the SC direction to the SC energy
Circularity	$(E_{5\times5} - E_{5\times1})/E_{5\times5}$
R_9	$E_{3\times3}/E_{\mathrm{SC}}$
$E_{\mathrm{ES}}/E_{\mathrm{raw}}$	Relative contribution of ECAL preshower energy to the uncorrected SC energy
Reconstruction of electron tracks	
f_{brem}	Bremsstrahlung fraction estimated from the momentum loss of the GSF track
N_{KF}, N_{GSF}	Number of hits associated with the KF and GSF tracks
χ^2_{KF}, χ^2_{GSF}	Reduced χ^2 of the KF and GSF track fits
N_{missHits}	Number of missing inner-tracker hits (expected to be zero for prompt electrons)
P_{conv}	Conversion probability derived from the χ^2 of the conversion fit
SC - track matching	
$E_{\mathrm{SC}}/p_{\mathrm{in}}$	Ratio of SC energy to the momentum of the inner track
$E_{\mathrm{PFclus}}/p_{\mathrm{out}}$	Ratio of the closest ECAL PF cluster energy at the extrapolated track position to the outer-track momentum
$1/E - 1/p$	Consistency between the SC energy and the electron track momentum
$\Delta\eta_{\mathrm{in}}$, $\Delta\phi_{\mathrm{in}}$	Angular separation between the SC barycentre and the inner-track extrapolation in η and ϕ
$\Delta\eta_{\mathrm{seed}}$	Separation in η between the seed cluster and the outer-track extrapolation
Cluster isolation	
$\sum_{\mathrm{ECAL}} E_T$	Scalar E_T sum of ECAL PF clusters in a cone of radius $\Delta R = 0.3$ around the electron direction
$\sum_{\mathrm{HCAL}} E_T$	Scalar E_T sum of HCAL PF clusters in a cone of radius $\Delta R = 0.3$ around the electron direction
$\sum_{\mathrm{trk}} p_T$	Scalar p_T sum of track transverse momenta in a cone of radius $\Delta R = 0.3$ around the electron direction
PU resilience	
ρ	Median transverse energy density from pileup in the event.

Table 3.3: Variables used in HZZ electron ID training for 2022

Hyperparameter	**Range**
`min_child_weight`	[1, 20]
`colsample_bytree`	[0.1, 1]
`max_depth`	[2, 15]
`subsample`	[0.5, 1]
`gamma`	[0, 10]
`reg_alpha`	[0, 10]
`reg_lambda`	[0, 10]

Table 3.4: Hyperparameters for XGBoost electron training and their ranges explored with Bayesian optimization. Sample fraction used for training is 75%.

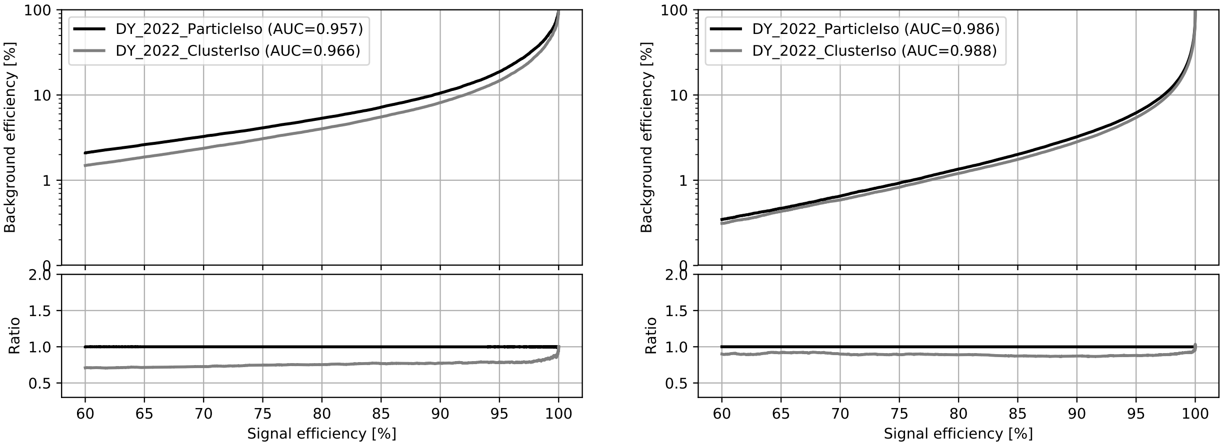


Figure 3.21: ROC curves in the endcap region for $5 < p_T < 10$ GeV (left) and $p_T > 10$ GeV (right), for the training using PF (black) and Cluster based (grey) isolation sums.

Region	Signal efficiency target	2017UL background efficiency	2022 bkg. eff. (particle iso)	2022 bkg. eff. (cluster iso)
EB1_5	81.6%	5.2%	5.9%	4.2%
EB1_10	97.4%	2.9%	2.8%	2.3%
EB2_5	80.3%	4.8%	5.6%	4.1%
EB2_10	96.7%	3.9%	3.5%	2.8%
EE_5	74.4%	3.3%	4.0%	3.0%
EE_10	96.6%	7.4%	8.5%	7.5%

Table 3.5: Background efficiencies for HZZ electron ID at target signal efficiencies in different p_T–η regions: inner barrel (EB1), outer barrel (EB2), endcap (EE) and for $5 < p_T < 10$ GeV and $p_T > 10$ GeV.

3.7 Electron Identification Scale Factors for 2022 Data-taking Period

Electron ID SFs, targeting $H \rightarrow ZZ \rightarrow 4\ell$ analysis for 2022 data taking period, were derived for the electron BDT which was trained on 2018 MC samples. This choice was made because, at the time of the computation, the 2022 HZZ ID was not yet available in centrally produced samples and therefore was not ready for use in the analysis. In 2022, CMS experienced an ECAL cooling leak that affected detector conditions and required changes in the ECAL operation and calibrations. Because of this, the 2022 data are split into preEE (before the leak/condition change) and postEE (after), since the detector response and corresponding corrections differ between the two periods. SFs are not measured only for the HZZ ID, but for the entire HZZ electron selection, which includes the requirements $|SIP| < 4$, $|d_z| < 1$, and $|d_{xy}| < 0.5$. Here, SIP denotes the 3D impact parameter significance, $|IP_{3D}/\sigma(IP_{3D})|$, which quantifies how significantly the lepton track is displaced from the PV. The variable d_z represents the longitudinal distance between the lepton track and the PV along the beam axis, while d_{xy} measures the transverse distance in the x–y plane. SFs are calculated using TnP and RMS methods, in the p_T bins of $[7, 15, 20, 35, 50, 100, 500]$ GeV and η bins of $[-2.5, -2.0, -1.566, -1.442, -0.8, 0, 0.8, 1.442, 1.566, 2.0, 2.5]$. SFs in η bins are then symmetrized to retrieve the final $|\eta|$ values. TnP pairs are selected using following requirements: $p_T^{tag} > 30$ GeV, $|\eta^{tag} < 2.17|$, tag and probe have opposite charge. Tighter requirement is additionally imposed in lowest p_T bin: $p_T^{tag} > 50$ GeV.

The default choice of PDFs used for fitting is:

- **nominal signal:** Z invariant mass templates from simulation is convolved with Gaussian
- **nominal background:** *RooCMSShape*, falling exponential multiplied by an error function, defined in 3.11

$$\mathrm{f}(x;\alpha,\beta,\gamma,\mu) = \mathrm{erfc}\big(\beta(\alpha - x)\big)\ \exp\big(-\gamma\,(x-\mu)\big) \tag{3.11}$$

- **alternative signal:** Z invariant mass templates from simulation are convolved with *CBExGaussShape*, convolution of a CB shape with a Gaussian resolution, defined in 3.12

$$\mathrm{g}(x;m_0,\sigma,\alpha,n,\sigma_2) = \int_{-\infty}^{\infty} \mathrm{CB}(x';m_0,\sigma,\alpha,n)\,\frac{1}{\sqrt{2\pi}\,\sigma_2}\exp\Big(-\frac{(x-x')^2}{2\sigma_2^2}\Big)\,dx' \tag{3.12}$$

where CB is one-sided Crystall Ball function defined as:

$$\mathrm{CB}(x;m_0,\sigma,\alpha,n) = \begin{cases} \exp\Big(-\frac{(x-m_0)^2}{2\sigma^2}\Big), & \text{if } \frac{x-m_0}{\sigma} > -\alpha, \\ A\left(B - \frac{x-m_0}{\sigma}\right)^{-n}, & \text{if } \frac{x-m_0}{\sigma} \le -\alpha, \end{cases} \tag{3.13}$$

$$A = \left(\frac{n}{|\alpha|}\right)^n \exp\Big(-\frac{|\alpha|^2}{2}\Big), \quad B = \frac{n}{|\alpha|} \tag{3.14}$$

- **alternative background:** Exponential

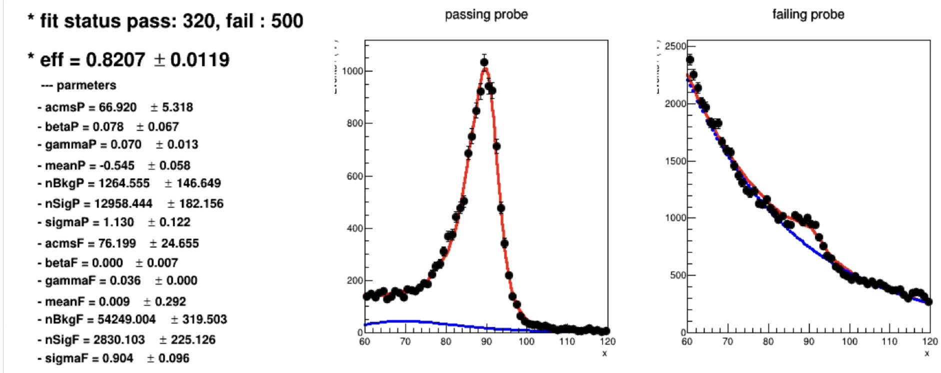
In the low p_T range of $[7, 15]$ GeV, the default background functions fail to model the steep slope observed in the left tail of the invariant mass distribution. To account for this behavior, a third-order Chebyshev polynomial is employed as the nominal background PDF, while a Γ distribution is used as an alternative background model in this bin. This is illustrated in Figures 3.22, 3.23, 3.24 and 3.25. The mathematical forms of these functions are given by:

$$T_3(x) = a_0 + a_1 x + a_2(2x^2 - 1) + a_3(4x^3 - 3x) \tag{3.15}$$

where $T_3(x)$ is the third-order Chebyshev polynomial, and a_0, a_1, a_2, a_3 are free parameters of the fit.

$$f(x; \gamma, \beta, \mu) = \frac{(x-\mu)^{\gamma-1}\, e^{-(x-\mu)/\beta}}{\Gamma(\gamma)\, \beta^{\gamma}} \tag{3.16}$$

where γ is the shape parameter, β is the scale parameter, μ is the location parameter, and $\Gamma(\gamma)$ is the Gamma function.



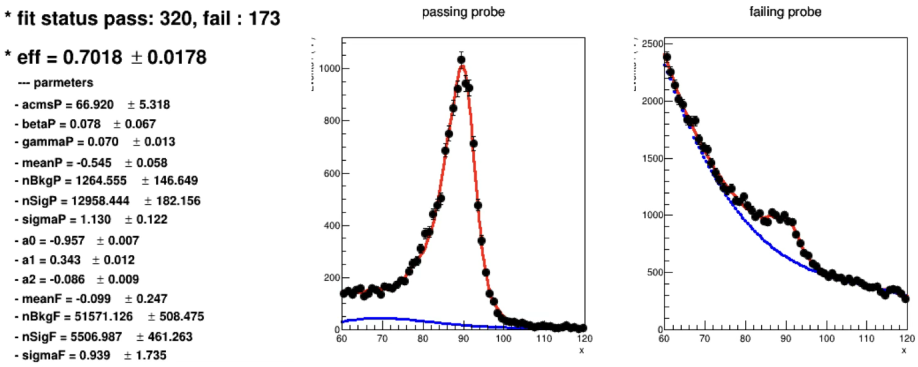
```
* fit status pass: 320, fail : 500

* eff = 0.8207 ± 0.0119
  --- parmeters
 - acmsP = 66.920   ± 5.318
 - betaP = 0.078   ± 0.067
 - gammaP = 0.070   ± 0.013
 - meanP = -0.545   ± 0.058
 - nBkgP = 1264.555   ± 146.649
 - nSigP = 12958.444   ± 182.156
 - sigmaP = 1.130   ± 0.122
 - acmsF = 76.199   ± 24.655
 - betaF = 0.000   ± 0.007
 - gammaF = 0.036   ± 0.000
 - meanF = 0.009   ± 0.292
 - nBkgF = 54249.004   ± 319.503
 - nSigF = 2830.103   ± 225.126
 - sigmaF = 0.904   ± 0.096
```

Figure 3.22: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, fitted using the RooCMSShape function as the nominal background PDF. In the region $-1.44 < \eta < -0.8$ and $7 < p_T < 15$ GeV, the model does not provide a satisfactory description of the low-mass tail in the failing-probe distribution.

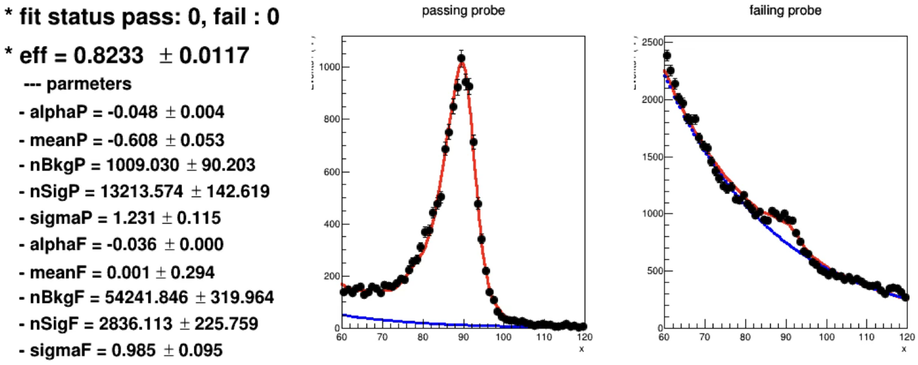
Apart from low-p_T bins, fits are also difficult in the crack region between barrel and endcap, which is defined as $1.44 < |\eta| < 1.57$. One such fit is shown on Figure 3.26. On the other hand, high p_T region exhibits clear peaks and low background, resulting in low uncertainty in efficiency estimate (see Figure 3.27).

SFs with their uncertainties are shown in Figs. 3.28–3.33. As expected, efficiency rises with p_T and SFs are mostly flat around 1. Lower p_T bins and, especially crack region exhibit higher relative uncertainties. These SFs have been approved and used in the first Run 3 $H \to ZZ \to 4\ell$ cross-section measurement [50]. Additionally, there is an ongoing effort within the HZZ group to exploit the J/ψ resonance for low-p_T electron identification SF measurements, similarly to the reconstruction study described in Sec. 3.5, with promising initial results.



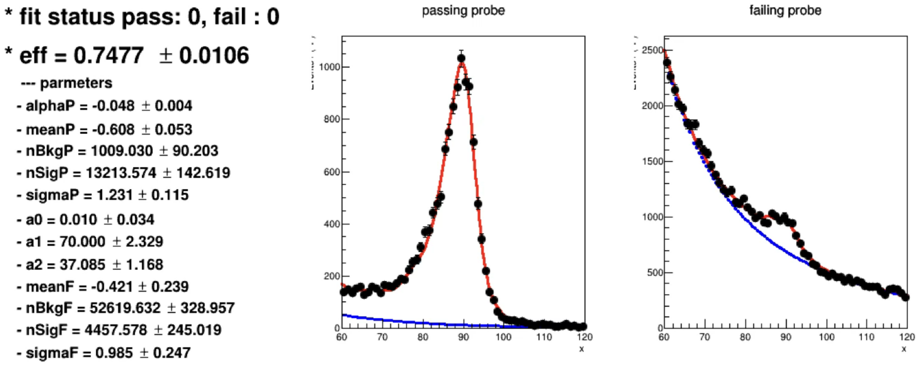
```
* fit status pass: 320, fail : 173
* eff = 0.7018 ± 0.0178
 --- parmeters
 - acmsP = 66.920  ± 5.318
 - betaP = 0.078  ± 0.067
 - gammaP = 0.070  ± 0.013
 - meanP = -0.545  ± 0.058
 - nBkgP = 1264.555  ± 146.649
 - nSigP = 12958.444  ± 182.156
 - sigmaP = 1.130  ± 0.122
 - a0 = -0.957  ± 0.007
 - a1 = 0.343  ± 0.012
 - a2 = -0.086  ± 0.009
 - meanF = -0.099  ± 0.247
 - nBkgF = 51571.126  ± 508.475
 - nSigF = 5506.987  ± 461.263
 - sigmaF = 0.939  ± 1.735
```

Figure 3.23: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, fitted using a third-order Chebyshev polynomial as the nominal background PDF. In the region $-1.44 < \eta < -0.8$ and $7 < p_T < 15$ GeV, this parametrization provides a good description of the low-mass tail in the failing-probe distribution.



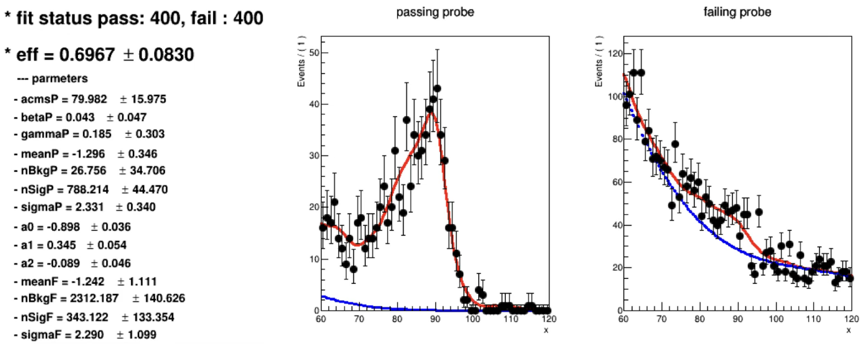
```
* fit status pass: 0, fail : 0
* eff = 0.8233 ± 0.0117
 --- parmeters
 - alphaP = -0.048 ± 0.004
 - meanP = -0.608 ± 0.053
 - nBkgP = 1009.030 ± 90.203
 - nSigP = 13213.574 ± 142.619
 - sigmaP = 1.231 ± 0.115
 - alphaF = -0.036 ± 0.000
 - meanF = 0.001 ± 0.294
 - nBkgF = 54241.846 ± 319.964
 - nSigF = 2836.113 ± 225.759
 - sigmaF = 0.985 ± 0.095
```

Figure 3.24: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, fitted using an exponential function as an alternative background PDF. In the region $-1.44 < \eta < -0.8$ and $7 < p_T < 15$ GeV, the model does not correctly reproduce the low-mass tail in the failing-probe distribution.



```
* fit status pass: 0, fail : 0
* eff = 0.7477 ± 0.0106
 --- parmeters
 - alphaP = -0.048 ± 0.004
 - meanP = -0.608 ± 0.053
 - nBkgP = 1009.030 ± 90.203
 - nSigP = 13213.574 ± 142.619
 - sigmaP = 1.231 ± 0.115
 - a0 = 0.010 ± 0.034
 - a1 = 70.000 ± 2.329
 - a2 = 37.085 ± 1.168
 - meanF = -0.421 ± 0.239
 - nBkgF = 52619.632 ± 328.957
 - nSigF = 4457.578 ± 245.019
 - sigmaF = 0.985 ± 0.247
```

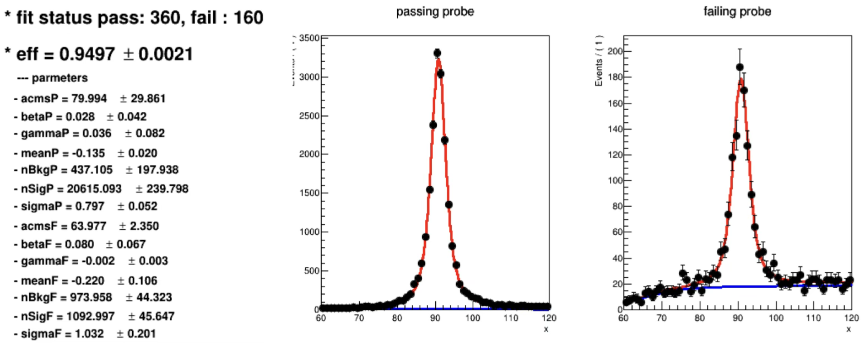
Figure 3.25: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, fitted using a Γ distribution as an alternative background PDF. In the region $-1.44 < \eta < -0.8$ and $7 < p_T < 15$ GeV, this parametrization provides a good description of the low-mass tail in the failing-probe distribution.



```
* fit status pass: 400, fail : 400

* eff = 0.6967 ± 0.0830
  --- parmeters
 - acmsP = 79.982    ± 15.975
 - betaP = 0.043   ± 0.047
 - gammaP = 0.185    ± 0.303
 - meanP = -1.296    ± 0.346
 - nBkgP = 26.756    ± 34.706
 - nSigP = 788.214    ± 44.470
 - sigmaP = 2.331    ± 0.340
 - a0 = -0.898   ± 0.036
 - a1 = 0.345   ± 0.054
 - a2 = -0.089   ± 0.046
 - meanF = -1.242    ± 1.111
 - nBkgF = 2312.187    ± 140.626
 - nSigF = 343.122    ± 133.354
 - sigmaF = 2.290    ± 1.099
```

Figure 3.26: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, shown for the region $-1.57 < \eta < -1.44$ and $7 < p_T < 15$ GeV, together with the nominal fit used for the efficiency measurement.



```
* fit status pass: 360, fail : 160

* eff = 0.9497 ± 0.0021
  --- parmeters
 - acmsP = 79.994    ± 29.861
 - betaP = 0.028   ± 0.042
 - gammaP = 0.036    ± 0.082
 - meanP = -0.135    ± 0.020
 - nBkgP = 437.105    ± 197.938
 - nSigP = 20615.093    ± 239.798
 - sigmaP = 0.797    ± 0.052
 - acmsF = 63.977    ± 2.350
 - betaF = 0.080   ± 0.067
 - gammaF = -0.002    ± 0.003
 - meanF = -0.220    ± 0.106
 - nBkgF = 973.958    ± 44.323
 - nSigF = 1092.997    ± 45.647
 - sigmaF = 1.032    ± 0.201
```

Figure 3.27: Invariant mass distribution of TnP electron pairs for passing probes (left) and failing probes (right) in 2022 data, shown for the region $0 < \eta < 0.8$ and $50 < p_T < 100$ GeV, together with the nominal fit used for the efficiency measurement.

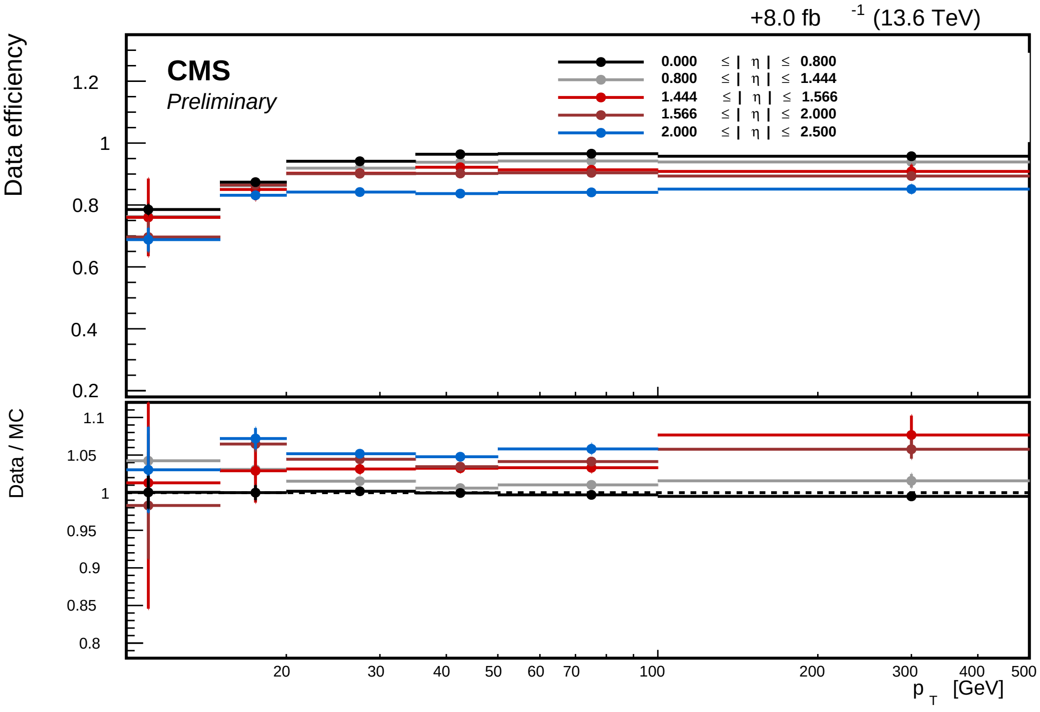


Figure 3.28: Electron selection efficiencies and SFs for the 2022 preEE data-taking period as a function of p_T in several $|\eta|$ regions. In the upper panel, points with error bars represent efficiencies and their corresponding uncertainties measured in data, shown in different colors for different $|\eta|$ intervals: black ($0.000 \leq |\eta| \leq 0.800$), gray ($0.800 \leq |\eta| \leq 1.444$), red ($1.444 \leq |\eta| \leq 1.566$), brown ($1.566 \leq |\eta| \leq 2.000$), and blue ($2.000 \leq |\eta| \leq 2.500$). The lower panel shows the corresponding SFs, with error bars indicating uncertainties.

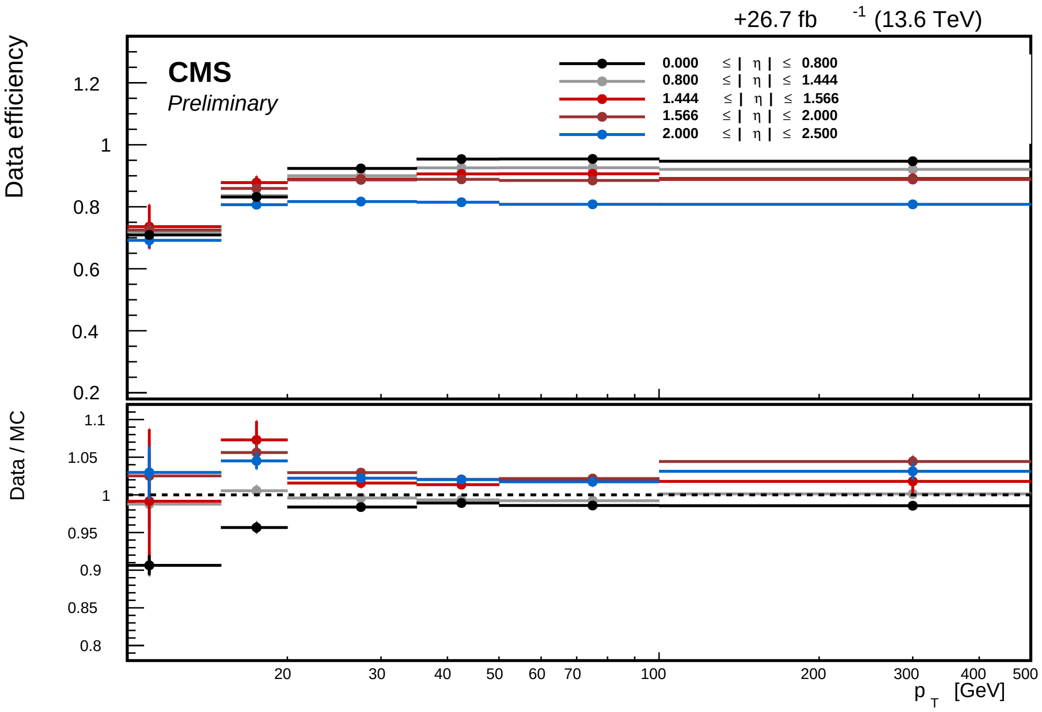


Figure 3.29: Electron selection efficiencies and SFs for the 2022 postEE data-taking period as a function of p_T in several $|\eta|$ regions. In the upper panel, points with error bars represent efficiencies and their corresponding uncertainties measured in data, shown in different colors for different $|\eta|$ intervals: black ($0.000 \leq |\eta| \leq 0.800$), gray ($0.800 \leq |\eta| \leq 1.444$), red ($1.444 \leq |\eta| \leq 1.566$), brown ($1.566 \leq |\eta| \leq 2.000$), and blue ($2.000 \leq |\eta| \leq 2.500$). The lower panel shows the corresponding SFs, with error bars indicating uncertainties.

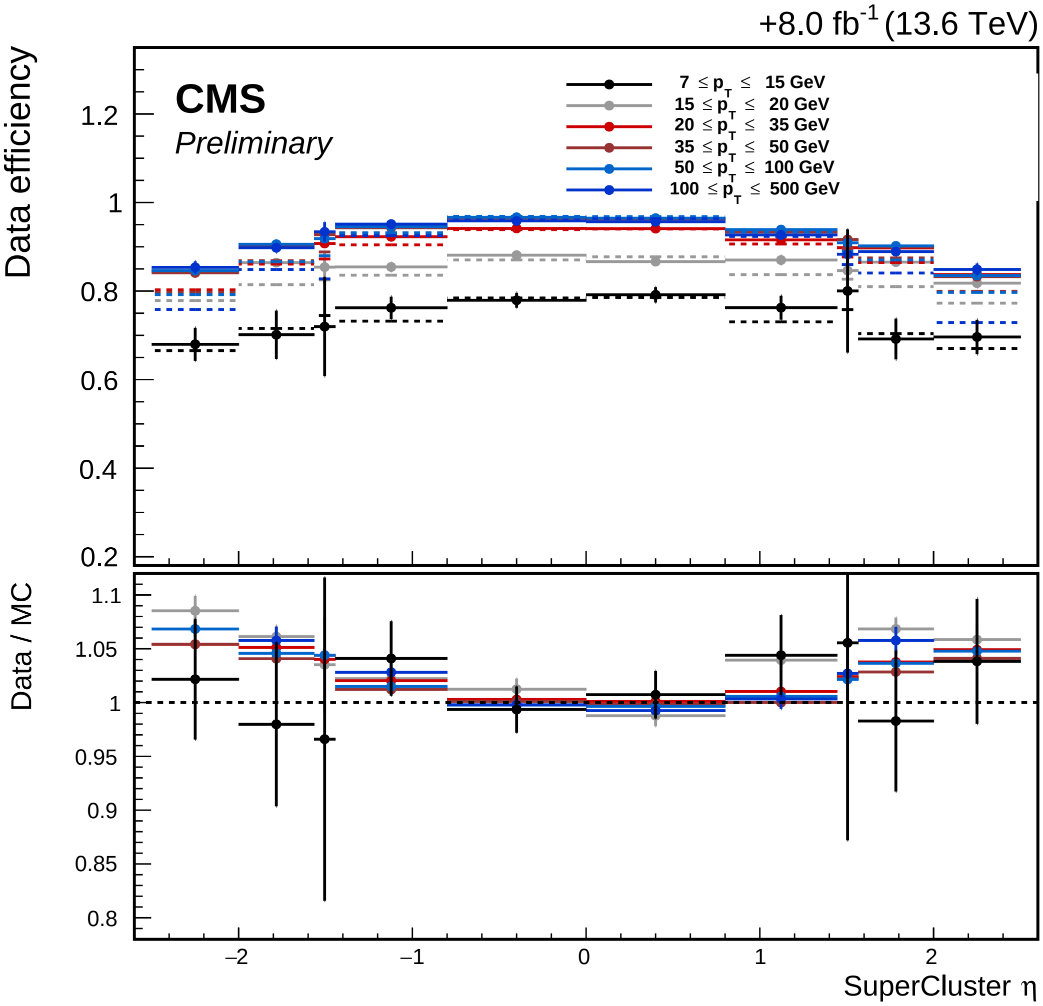


Figure 3.30: Electron selection efficiencies and SFs for the 2022 preEE data-taking period as a function of $|\eta|$ in several p_T ranges. In the upper panel, points with error bars represent efficiencies and their corresponding uncertainties measured in data, shown in different colors for different p_T intervals: black ($7 \leq p_T \leq 15$ GeV), gray ($15 \leq p_T \leq 20$ GeV), red ($20 \leq p_T \leq 35$ GeV), brown ($35 \leq p_T \leq 50$ GeV), light blue ($50 \leq p_T \leq 100$ GeV), and dark blue ($100 \leq p_T \leq 500$ GeV). The lower panel shows the corresponding SFs, with error bars indicating uncertainties.

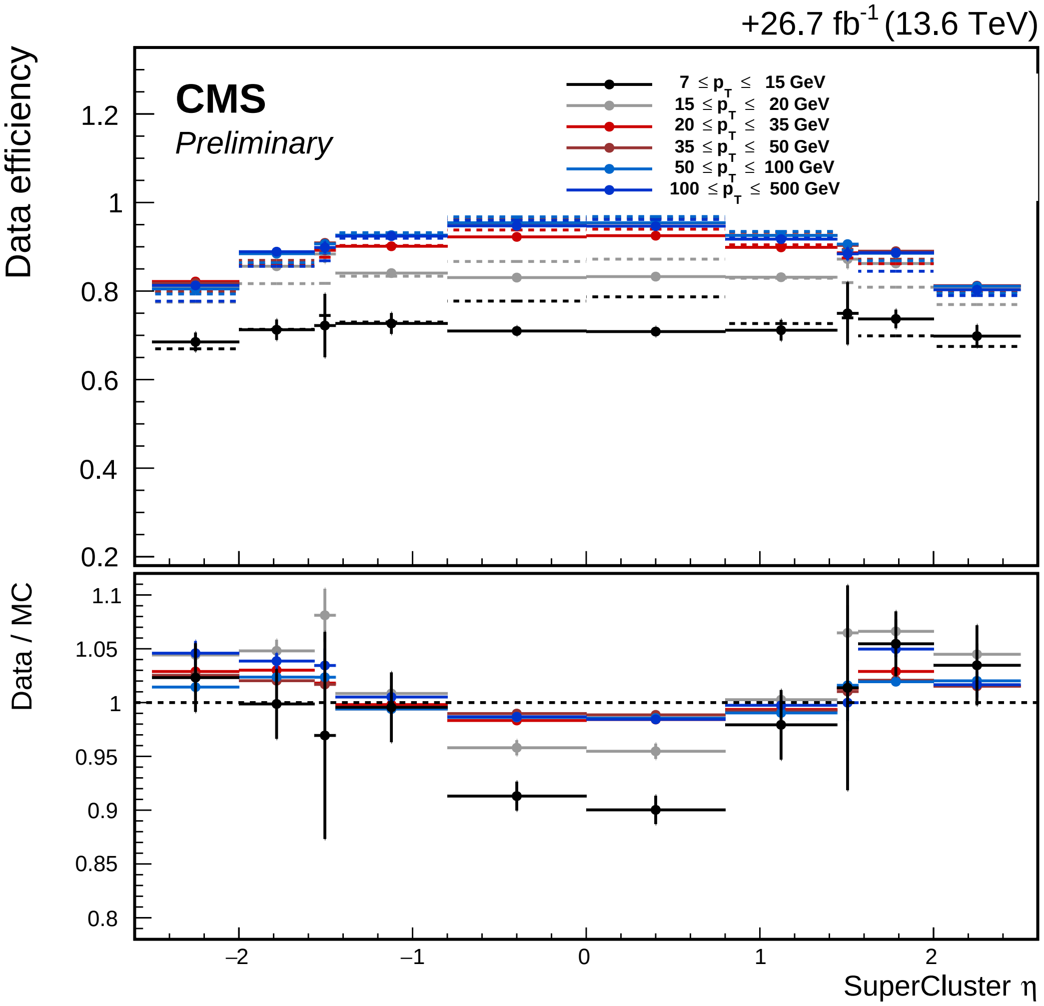


Figure 3.31: Electron selection efficiencies and SFs for the 2022 postEE data-taking period as a function of $|\eta|$ in several p_T ranges. In the upper panel, points with error bars represent efficiencies and their corresponding uncertainties measured in data, shown in different colors for different p_T intervals: black ($7 \leq p_T \leq 15$ GeV), gray ($15 \leq p_T \leq 20$ GeV), red ($20 \leq p_T \leq 35$ GeV), brown ($35 \leq p_T \leq 50$ GeV), light blue ($50 \leq p_T \leq 100$ GeV), and dark blue ($100 \leq p_T \leq 500$ GeV). The lower panel shows the corresponding SFs, with error bars indicating uncertainties.

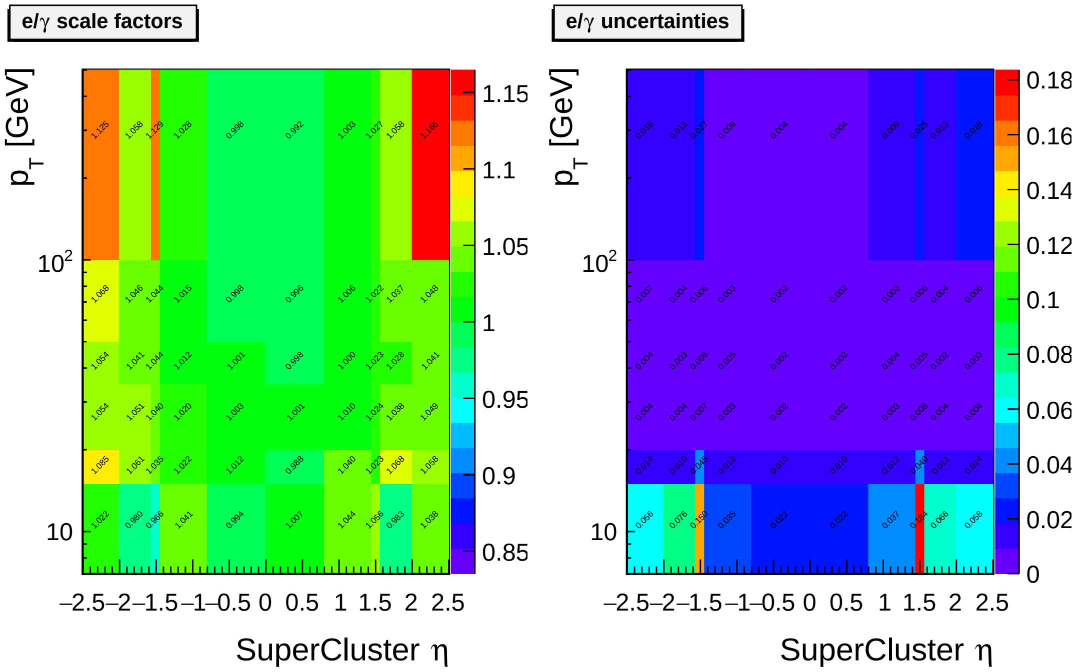


Figure 3.32: Two-dimensional electron selection SFs for the 2022 preEE data-taking period as a function of p_T and eta. The left panel shows the scale factor values in each bin, while the right panel shows the corresponding uncertainties. The color scale indicates the numerical value in each bin, as shown by the color bar.

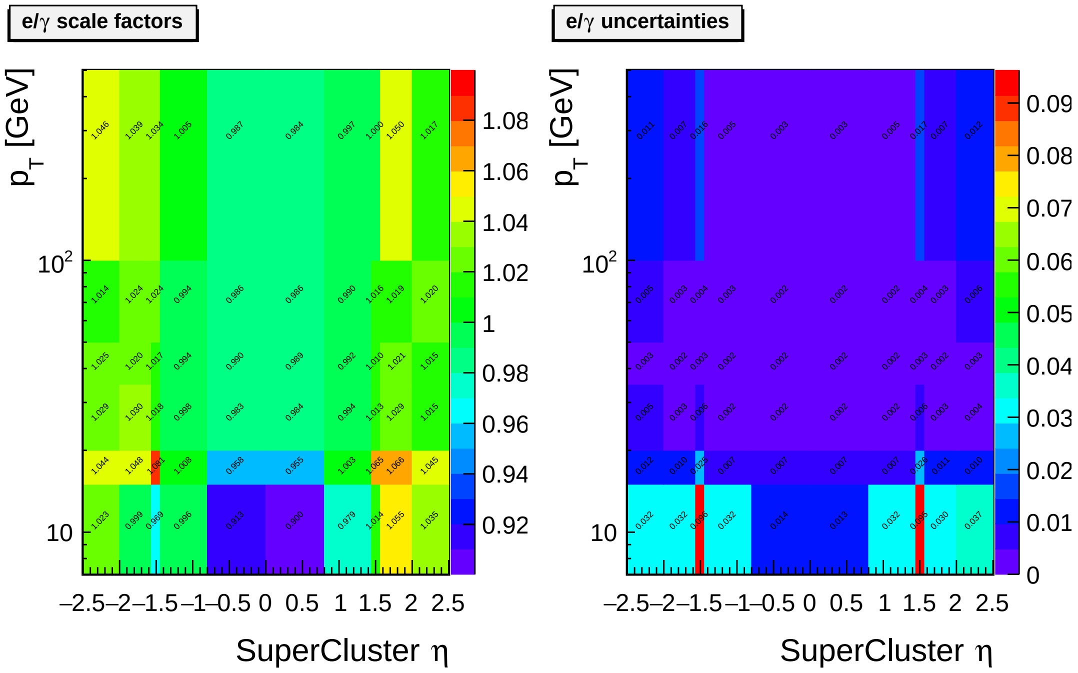


Figure 3.33: Two-dimensional electron selection SFs for the 2022 postEE data-taking period as a function of p_T and eta. The left panel shows the scale factor values in each bin, while the right panel shows the corresponding uncertainties. The color scale indicates the numerical value in each bin, as shown by the color bar.

Chapter 4

The Golden Channel

This chapter presents the common analysis methods, physics objects, and event selection employed by the group studying Higgs boson properties in the $H \to ZZ \to 4\ell$ decay channel. Although the branching fraction of this process is relatively small (approximately 2.6%), the exceptional cleanliness of the final state makes it powerful for precision measurements.

In this channel, the Higgs boson is reconstructed from four leptons (electrons or muons), which form a narrow resonance in the four-lepton invariant mass distribution around 125 GeV. The dominant background arises from continuum ZZ production, whose invariant mass spectrum is smooth in the signal region. This characteristic enables the extraction of the Higgs boson signal with a high signal-to-background ratio.

The excellent lepton reconstruction performance of the CMS detector leads to a four-lepton mass resolution of approximately 1–2 GeV. Such precision allows accurate measurements of the Higgs boson mass. Furthermore, since all final-state particles are experimentally detectable, the complete event kinematics can be reconstructed. This enables the construction of kinematic discriminants that enhance the separation between signal and background processes.

The combination of a clean experimental signature, excellent mass resolution, manageable backgrounds, and full kinematic reconstruction capability makes the $H \to ZZ \to 4\ell$ decay mode a cornerstone of precision Higgs boson measurements at the LHC. Owing to these unique experimental advantages, it is commonly referred to as the *golden channel.*

The $H \to ZZ \to 4\ell$ channel played a key role in the observation of the Higgs boson in 2012 [2, 3]. Subsequent measurements in this channel provided precise determinations of the Higgs boson mass and constraints on its spin-parity quantum numbers [68]. With the increase of the available data set, the focus shifted toward precision measurements of signal strengths, fiducial and differential cross sections, and constraints on anomalous Higgs couplings [69, 70, 71, 72, 73]. Early Run III measurements at $\sqrt{s} = 13.6$ TeV continued this program [74], and the HL-LHC is expected to further improve the statistical precision of these studies.

The event selection used to build the HZZ candidate sample is described in Section 4.1, while the systematic uncertainties considered in the analysis are discussed in Section 4.4. Section 4.2 presents the treatment of reducible backgrounds, with emphasis on the data-driven methods used to estimate their contribution. The construction of kinematic observables used to enhance the separation between signal and background is outlined in Section 4.3. Differential cross-section measurements in this channel are presented in Section 4.5, and their interpretation within the on-shell κ-framework is given in Section 4.6.

4.1 Event Selection

The first step of the event selection is the trigger requirement. The data used in the $H \to ZZ \to 4\ell$ analysis were collected during Run 2 of the LHC (2016–2018), corresponding to a total validated integrated luminosity of 138 fb^{-1}. Events are recorded in primary datasets (PDs), each defined by a specific set of high-level trigger (HLT) paths. An event is included in a given PD if it satisfies at least one of the HLT paths associated with that dataset. Although each HLT path belongs to a single PD, an event may appear in multiple PDs if it fires more than one trigger.

The analysis makes use of the following primary datasets: `DoubleMuon`, `MuonEGamma`, `SingleMuon`, `DoubleEGamma` (2016–2017), `SingleElectron` (2016–2017), and `EGamma` (2018). For the 2018 data-taking period, the `EGamma` dataset replaces the previous `DoubleEGamma` and `SingleElectron` datasets by combining their trigger paths into a single PD. The dataset names reflect the type of lepton triggers they contain, such as dimuon, dielectron, single-lepton, or mixed electron–muon triggers.

Since events can be recorded in multiple PDs, a prioritisation scheme is applied to avoid double counting. Events are assigned according to the following hierarchy:

- Events passing dielectron or triple-electron triggers are taken from the `EGamma` (or `DoubleEGamma`) dataset.
- Events passing dimuon or triple-muon triggers, and not selected by electron triggers, are taken from the `DoubleMuon` dataset.
- Events passing mixed lepton triggers (electron–muon combinations) and not selected by the previous categories are taken from the `MuonEGamma` dataset.
- Events passing single-electron triggers and failing all previous criteria are taken from the `EGamma` (or `SingleElectron`) dataset.
- Events passing single-muon triggers and not selected by any of the above categories are taken from the `SingleMuon` dataset.

The minimal transverse momentum thresholds required by the individual HLT paths depend on the specific trigger and are summarised in Table 4.1. In general, dilepton and multilepton triggers require electrons and muons with transverse momentum thresholds that vary based on a data-taking period.

	e/e (GeV)	μ/μ (GeV)	e/μ, μ/e (GeV)
2016	17/12	17/8	17/8, 8/23
2017	23/12	17/8	23/8, 12/23
2018	23/12	17/8	23/8, 12/23

Table 4.1: Lepton p_{T} HLT thresholds for each data-taking year.

The object selection requirements are optimised to ensure high signal efficiency for the $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ final state.

Electrons and muons are required to satisfy $\mathrm{p_T}$ and $|\eta|$ thresholds consistent with the detector acceptance and trigger turn-on regions. The relatively low $\mathrm{p_T}$ thresholds (7 GeV for electrons and 5 GeV for muons) maximise the acceptance for four-lepton final states, where subleading leptons can be soft due to the kinematics of off-shell Z bosons.

Impact parameter constraints (d_{xy}, d_z) and the significance of the impact parameter (SIP $<$ 4) suppress non-prompt leptons originating from heavy-flavour decays. Isolation requirements further reduce contamination from jets misidentified as leptons while retaining prompt leptons from Z boson decays.

Electron identification relies on a multivariate discriminator (XGBoost), which separates prompt electrons from fakes. For muons, a dedicated high-$\mathrm{p_T}$ identification is applied above 200 GeV to maintain performance in boosted topologies.

Jets are reconstructed with $\mathrm{p_T} > 30$ GeV and $|\eta| < 4.7$ to ensure reliable calibration and pileup mitigation. The lepton–jet separation requirement $\Delta R(\ell/\gamma, j) > 0.4$ prevents double counting of energy deposits and ensures proper object cleaning. The tight pileup jet identification suppresses jets originating from additional proton–proton interactions, which is especially important in high pileup conditions.

The application of the medium DeepCSV working point for b-tagging enables categorisation of events according to production mechanisms (e.g. $\mathrm{t\bar{t}H}$) while keeping mistag rates under control.

Finally, to account for energy loss of final-state leptons due to Final State Radiation (FSR), a dedicated correction procedure is applied. Candidate FSR photons are selected among PF photons with transverse momentum $p_T^\gamma > 2$ GeV and pseudorapidity $|\eta^\gamma| < 2.4$, satisfying a relative isolation requirement $I^\gamma_{\mathrm{rel}}(\Delta R < 0.3) < 1.8$. Photons are associated with the closest selected lepton and are required to satisfy $\Delta R(\ell, \gamma) < 0.5$ and $\Delta R(\ell, \gamma)/(p_T^\gamma)^2 < 0.012\ \mathrm{GeV}^{-2}$. Photons overlapping with electrons are removed, and in the case of multiple candidates, the most compatible photon is retained. Selected FSR photons are excluded from lepton isolation sums.

Full offline object selection requirements are summarized in Table 4.2.

After selection of physics objects, Z boson candidates are reconstructed by combining pairs of opposite sign tight leptons. The dilepton invariant mass, including the contribution from FSR photons, is required to satisfy $12 < m_{\ell^+\ell^-(\gamma)} < 120$ GeV. The selected Z candidates are subsequently paired to form ZZ candidates. Among the two dilepton pairs, the one with invariant mass closest to the nominal Z boson mass is denoted as $\mathrm{Z_1}$, while the other is labelled $\mathrm{Z_2}$. If two Z candidates share at least one lepton, the corresponding ZZ candidate is discarded. To prevent overlapping configurations, all lepton pairs originating from different Z candidates are required to satisfy $\Delta R(\ell, \ell) > 0.02$. The Higgs boson candidate is then selected from the reconstructed ZZ combinations according to the following criteria:

- The invariant mass of $\mathrm{Z_1}$ must satisfy $m_{\mathrm{Z_1}} > 40$ GeV. This requirement suppresses background processes that do not contain a genuine on-shell Z boson, such as $\mathrm{t\bar{t}}$ and $\mathrm{W}\gamma + \mathrm{jets}$, while preserving sensitivity to decay into two off-shell Z bosons.

- The leading and subleading leptons are required to satisfy $\mathrm{p_T} > 20$ GeV and $\mathrm{p_T} > 10$ GeV, respectively, ensuring operation in the plateau region of the trig-

Objects	**Selection criteria**
Electrons	**Soft:** $p_T^e > 7$ GeV, $\lvert\eta_e\rvert < 2.5$ $d_{xy} < 0.5$ cm, $d_z < 1.0$ cm **Loose:** Soft requirements + SIP < 4 **Tight:** Loose requirements + Identification and isolation based on XGBoost discriminator
Muons	**Soft:** Global or tracker muon $p_T^\mu > 5$ GeV, $\lvert\eta_\mu\rvert < 2.4$ $d_{xy} < 0.5$ cm, $d_z < 1.0$ cm **Loose:** Soft requirements + SIP < 4 **Tight:** Loose requirements + PF muon ID (tracker high-p_T muon ID if $p_T^\mu > 200$ GeV) $I_{rel}^\mu(\Delta R < 0.3) < 0.35$
FSR photons	$p_T^\gamma > 2$ GeV, $\lvert\eta_\gamma\rvert < 2.4$ $I_{rel}^\gamma(\Delta R < 0.3) < 1.8$ $\Delta R(\ell, \gamma) < 0.5$ $\Delta R(\ell, \gamma)/(p_T^\gamma)^2 < 0.012\ \mathrm{GeV}^{-2}$
Jets	$p_T^j > 30$ GeV, $\lvert\eta_j\rvert < 4.7$ $\Delta R(\ell/\gamma, j) > 0.4$ Cut-based jet ID, tight PU ID
b-jets	Medium DeepCSV working point applied additionally

Table 4.2: Summary of the object selection criteria used in the $H \to ZZ \to 4\ell$ analysis, including the definition of soft, loose, and tight leptons.

ger efficiency.

- The invariant mass of all possible opposite sign lepton pairs formed from the four selected leptons must satisfy $m_{\ell^+\ell'^-} > 4$ GeV. The FSR recovery is not included in this requirement. This selection suppresses backgrounds arising from low-mass hadronic resonances and bound quark states.

- In the 4e and 4μ final states, alternative dilepton pairings are considered. The two additional pairs, denoted as Z_a and Z_b, are constructed by combining one lepton from Z_1 and one from Z_2. The event is rejected if

$$|m_{Z_a} - m_{Z_0}| < |m_{Z_1} - m_{Z_0}| \quad \text{and} \quad m_{Z_2} < 12\ \text{GeV},$$

where m_{Z_0} denotes the nominal Z boson mass. The FSR recovery is included in the invariant mass calculation. This requirement removes events in which the alternative pairing resembles an on-shell Z boson accompanied by a low-mass dilepton resonance.

- The invariant mass of the four-lepton system must satisfy $m_{4\ell} > 70$ GeV.

4.2 Reducible Background

Among the various background contributions in the $H \to ZZ \to 4\ell$ analysis, the reducible component is the most challenging to model accurately. This background arises when at least one selected lepton does not originate from the decay of a prompt Z boson. The dominant sources are: (i) light-flavour hadrons misidentified as leptons, (ii) non-prompt leptons from semileptonic decays of heavy-flavour hadrons, and (iii) a smaller contribution from photon conversions in the detector material. Although these leptons are commonly referred to as “fake leptons”, only the first case corresponds to genuine misidentification, while the latter two involve real but non-prompt leptons.

The main SM processes contributing to this topology are Drell–Yan (Z+jets), $t\bar{t}$, WW+jets, $Z\gamma$+jets, and WZ+jets production. Since the dominant contribution is associated with Z+jets, the reducible background in this analysis is collectively denoted as Z+X. The typical event signature consists of two prompt leptons forming a Z boson candidate and one or two additional non-prompt or misidentified leptons.

Due to limited statistics and imperfect modelling of higher-order effects in simulation, the Z+X contribution in the signal region cannot be reliably estimated using Monte Carlo samples alone. Instead, a data-driven approach based on control regions orthogonal to the signal region is employed.

4.2.1 Fake Rate Determination

The reducible background is estimated using the fake rate method. The fake rate, denoted by f, represents the probability that a non-prompt or misidentified lepton passing a loose selection also satisfies the full tight lepton requirements used in the signal region.

The fake rate is measured in a dedicated control region referred to as Z+L. This region consists of events containing a well-identified Z_1 candidate and exactly one additional loose lepton. To suppress contamination from processes with genuine prompt leptons, a requirement on missing transverse momentum $p_T^{\text{miss}} < 25$ GeV is imposed, and the invariant mass of the additional lepton and the oppositely charged lepton from the Z_1 candidate must exceed 4 GeV.

Fake rates are measured separately for electrons and muons, in bins of lepton transverse momentum and pseudorapidity, and independently for each data-taking period. Once determined, the fake rates are applied to events in method-specific Z+LL control regions to predict the reducible background contribution in the signal region.

4.2.2 Opposite-sign Method

In the opposite-sign (OS) implementation, the Z+L control region used to determine the fake rates includes an additional constraint on the invariant mass of the leading Z candidate, $|m_{Z_1} - m_Z| < 7$ GeV. This requirement suppresses contributions from asymmetric photon conversions in the detector material, which predominantly affect the low dilepton mass region and can bias the fake rate measurement. WZ contamination that survives the p_T^{miss} requirement is subtracted using simulation.

The OS approach defines two Z+LL control regions: the 2P2F and the 3P1F regions. In the 2P2F region, two leptons satisfy the full tight selection and form the Z_1 candidate, while the remaining two leptons fail the tight identification and/or isolation requirements. This region is mainly populated by processes containing two prompt leptons, such as Drell–Yan and $t\bar{t}$ production. The 3P1F region, instead, contains events with three tightly selected leptons (two forming the Z_1 candidate) and one failing lepton. This topology is typical of processes with three prompt leptons, such as WZ and $Z\gamma^*$ production. By construction, the 3P1F region partially overlaps with the 2P2F region.

The contribution of the 2P2F region to the signal region is estimated by weighting each event with the probability that both failing leptons would satisfy the tight selection. Denoting the fake rates of the two additional leptons as f_3 and f_4, the predicted yield is

$$N^{Z+X}_{\mathrm{SR,\,2P2F}} = \sum_{i}^{N_{\mathrm{2P2F}}} \left(\frac{f_3^i}{1-f_3^i} \cdot \frac{f_4^i}{1-f_4^i} \right), \tag{4.1}$$

where N_{2P2F} denotes the number of events in the 2P2F control region. Similarly, the contribution from the 3P1F region is obtained by

$$N^{Z+X}_{\mathrm{SR,\,3P1F}} = \sum_{i}^{N_{\mathrm{3P1F}}} \left(\frac{f_4^i}{1-f_4^i} \right). \tag{4.2}$$

When combining the two contributions, care must be taken to correct for events that populate both control regions, in order to avoid double counting. In addition, the small irreducible ZZ contribution present in the 3P1F region is subtracted using simulated samples.

The final Z+X estimate obtained with the OS method is therefore constructed by summing the corrected 2P2F and 3P1F contributions, after subtracting the overlap term and the irreducible ZZ component.

4.2.3 Same-sign Method

In the same-sign (SS) implementation, the Z+L control region used for the determination of the fake rates is defined with a wider invariant mass requirement for the leading Z candidate, namely $40 < m_{Z_1} < 120$ GeV. As in the OS approach, the WZ contribution is subtracted using simulated samples. The enlarged Z mass window increases the fraction of asymmetric photon conversions, particularly in the Z+e control region. In such events, one of the conversion legs may not be reconstructed, leading to an apparent additional electron that populates the Z+L region. This modifies the composition of the control region relative to the Z+LL region and can bias the fake rate measurement. To account for this effect, an additional correction is applied to the electron fake rates in the SS method.

The correction procedure is performed in two steps. First, the Z+e control region is subdivided into categories characterized by different requirements on $|m_{Z_1} - m_Z|$ and $|m_{eZ_1} - m_Z|$, thereby varying the relative contribution of photon conversions. For each category, the fake rate and the mean number of missing hits in the pixel detector are determined in bins of the transverse momentum and pseudorapidity of the additional lepton. Since the number of missing hits is sensitive to conversion electrons, a linear dependence of the fake rate on this observable is observed. In a second step, the mean number of missing hits is measured in the corresponding Z+LL control region, and the corrected fake rate is extracted from the derived linear relation. The limited statistics available in these control samples constitute one of the dominant sources of uncertainty in the SS background estimate.

The Z+LL control region in the SS method, denoted as $2\mathrm{P}2\mathrm{L}_{\mathrm{SS}}$, requires two tightly selected leptons forming the Z_1 candidate and two additional SS loose leptons. Unlike the OS method, failing leptons are not explicitly required. The SS condition ensures orthogonality to the signal region, allowing the use of loose leptons while avoiding signal contamination. As a consequence, the SS control region is significantly more populated than the corresponding OS regions.

The reducible background contribution in the signal region is obtained by weighting each event in the $2\mathrm{P}2\mathrm{L}_{\mathrm{SS}}$ region according to

4.2.4 Uncertainty and Choice of Method for Reducible Background Estimation

Off-shell analysis presented in this thesis utilizes both yields and shapes of the observables. For this purpose, SS method is used. This choice is mainly motivated by lower number of events in the OS CRs. When evaluating the Z+X background in poorly populated kinematic bins, such as those at large transverse momentum, there is a potential risk of negative bins.

The uncertainty in the normalization of the Z+X background is evaluated by combining statistical and systematic contributions. The statistical component arises from the number of events in the Z+L and Z+LL CRs. The dominant systematic uncertainty originates from differences in the composition of the CRs where the fake rates are computed (Z+L) and the CRs where the fake rates are applied (Z+LL). This contribution is computed with MC samples. An additional sub-dominant contribution accounts for variations of of the expected yield considering up and down variation of the fake rate. All contributions are combined in quadrature to obtain the total uncertainty.

4.3 Kinematic Observables

To isolate specific contributions to the $H \to ZZ \to 4\ell$ process, it is necessary to exploit the kinematic information from the four-lepton final state. Owing to the scalar nature of the Higgs boson and the sequential decay through two intermediate Z bosons, the process can be completely described by five angular observables and two invariant masses at fixed four-lepton invariant mass $m_{4\ell}$.

A schematic representation of the $gg/q\bar{q} \to H \to ZZ \to 4\ell$ process is shown in Fig. 4.1. The five angles characterize both the production and decay of the Higgs boson, while the invariant masses m_{Z_1} and m_{Z_2} correspond to the two reconstructed Z boson candidates. The Z boson labeled Z_1 is defined as the dilepton pair with invariant mass closest to the nominal Z boson mass, while Z_2 denotes the remaining dilepton pair.

The five angular observables are defined as follows:

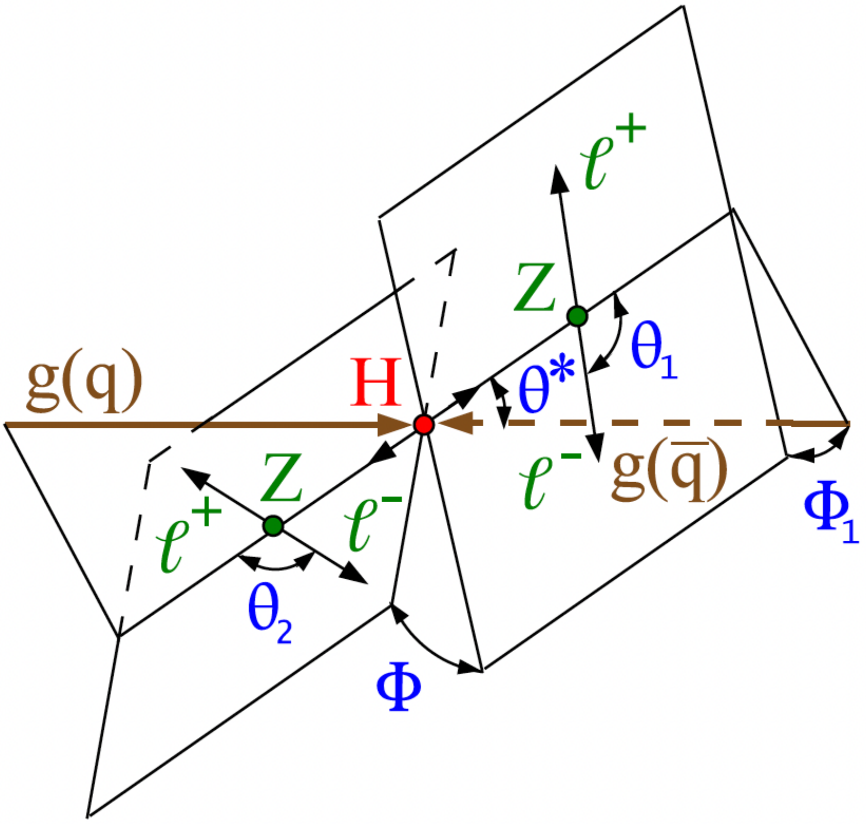


Figure 4.1: Schematic representation of the $gg/q\bar{q} \to H \to ZZ \to 4\ell$ process. The five decay angles and the two dilepton invariant masses (m_{Z_1} and m_{Z_1}) fully characterize the kinematics of the event at fixed $m_{4\ell}$.

- $\boldsymbol{\theta^*}$: defined in the Higgs boson rest frame as the angle between the direction of the incoming partons and the $H \to ZZ$ decay axis.
- $\boldsymbol{\phi_1}$: defined in the Higgs boson rest frame as the angle between the plane formed by the incoming partons and the decay axis, and the decay plane of the Z_1 boson.
- $\boldsymbol{\theta_1}$: defined in the Z_1 rest frame as the angle between the direction of the Z_1 boson in the Higgs rest frame and the momentum of the negatively charged lepton from its decay.
- $\boldsymbol{\theta_2}$: defined analogously in the Z_2 rest frame as the angle between the Z_2 direction in the Higgs rest frame and the momentum of the negatively charged lepton from its decay.
- $\boldsymbol{\phi}$: defined in the Higgs boson rest frame as the angle between the decay planes of the two Z bosons.

The angles θ^*, θ_1, and θ_2 are defined in the interval $[0, \pi]$, while ϕ_1 and ϕ take values in $[-\pi, \pi]$. In CMS analyses, the cosine of the polar angles, $\cos\theta^*$, $\cos\theta_1$, and $\cos\theta_2$, is typically used, as these quantities arise naturally from scalar products of unit vectors.

Collectively, the decay observables are denoted as

$$\Omega(\theta^*, \theta_1, \theta_2, \phi, \phi_1, m_{Z_1}, m_{Z_2} \mid m_{4\ell}),$$

which define the full set of degrees of freedom describing the $H \to ZZ \to 4\ell$ process.

Within the SM, the distributions of $\cos\theta_1$, $\cos\theta_2$, and ϕ are symmetric. In the presence of CP-violating interactions, these distributions can exhibit characteristic asymmetries. In particular, a CP-violating term with a purely real phase relative to the SM coupling leads to asymmetry in the ϕ distribution, while a purely imaginary phase induces asymmetries in $\cos\theta_1$ and $\cos\theta_2$. For a spin-zero particle such as the SM Higgs boson, the parton-level distributions of $\cos\theta^*$ and ϕ_1 are expected to be flat, although detector effects can distort these shapes.

4.3.1 Matrix Element Likelihood Approach

The kinematic configuration of the $H \to ZZ \to 4\ell$ process, described by the set of observables

$$\Omega(\theta^*, \theta_1, \theta_2, \phi, \phi_1, m_{Z_1}, m_{Z_2} \mid m_{4\ell}),$$

fully characterizes each event at fixed four-lepton invariant mass. To exploit this multidimensional information in a statistically optimal way, matrix-element–based techniques are employed.

In perturbative QFT, the differential cross section is proportional to the squared matrix element, which determines the probability amplitude for a given process. This observation connects directly to the Neyman–Pearson lemma, which states that the likelihood ratio provides the most powerful test statistic for distinguishing between two hypotheses at fixed size. In collider physics, the likelihood for a given event corresponds to the differential cross section evaluated at the observed kinematic point. Consequently, the optimal discriminator between competing physics scenarios can be constructed from ratios of their matrix elements.

The Matrix Element Likelihood Approach (MELA) [75] implements this principle by computing leading-order matrix-element probabilities for signal and background hypotheses using the same theoretical ingredients employed in the MC simulation. For a reconstructed event characterized by invariant mass $m_{4\ell}$ and kinematic configuration Ω, MELA evaluates the probability that the event originates from a given physics process.

Because these probabilities are derived directly from the matrix element, they depend on the complete set of kinematic observables and therefore retain the full leading-order dynamical information of the process. An important feature of this approach is its flexibility: when two hypotheses share identical initial and final states, their difference resides solely in the matrix element. In such cases, events generated under one hypothesis can be reweighted to another by taking ratios of the corresponding matrix-element probabilities. This allows efficient exploration of alternative coupling structures or BSM scenarios without the need for separate simulated samples for each hypothesis.

Using the MELA probabilities, a discriminant between two hypotheses with probabilities P_a and P_b is defined as

$$D_{\text{alt}}(\Omega) = \frac{P_a(\Omega)}{P_a(\Omega) + P_b(\Omega)}, \tag{4.3}$$

which is bounded between 0 and 1 and quantifies the relative compatibility of the event with the two scenarios.

If the two hypotheses interfere at amplitude level, an additional observable sensitive to the interference contribution can be constructed:

$$D_{\text{int}}(\Omega) = \frac{P_{\text{int}}(\Omega)}{2\sqrt{P_a(\Omega)P_b(\Omega)}}. \tag{4.4}$$

This quantity lies in the interval $[-1, 1]$. The interference probability is obtained from the sum of amplitudes,

$$P_{a+b} = |A_a + A_b|^2 = P_a + P_b + P_{\text{int}}, \tag{4.5}$$

where A_a and A_b denote the corresponding matrix-element amplitudes. The interference contribution is extracted as

$$P_{\text{int}} = P_{a+b} - P_a - P_b.$$

Matrix-element discriminants constructed in this manner are widely used in $H \to ZZ \to 4\ell$ analyses, enabling precision measurements of Higgs boson properties and tests of alternative interaction hypotheses. They have also been extensively used in the off-shell Higgs self-coupling measurement, which will be further discussed in Section 5.2.

4.4 Systematic Uncertainties

Systematic uncertainties are typically classified into two broad categories: experimental and theoretical. Experimental uncertainties originate from detector effects, object reconstruction, calibration procedures, and background estimation methods. Theoretical uncertainties arise from the precision of perturbative calculations, the modeling of parton distribution functions, and higher-order corrections applied to signal and background processes.

Within the maximum-likelihood framework, systematic effects are incorporated as nuisance parameters that are profiled in the fit. Normalization uncertainties are generally implemented using log-normal probability density functions. For a nuisance parameter θ constrained around its nominal value $\hat{\theta}$, the log-normal prior is written as

$$p(\theta \mid \hat{\theta}) = \frac{1}{\sqrt{2\pi}\,\ln\kappa} \exp\left(-\frac{\ln^2(\theta/\hat{\theta})}{2\,\ln^2\kappa}\right)\frac{1}{\theta}, \tag{4.6}$$

where the parameter κ determines the width of the distribution. The log-normal form is appropriate for multiplicative corrections and results in variations of the overall normalization of the corresponding process. Shape uncertainties are implemented through alternative templates that modify the distributions of the discriminating observables.

Although individual $H \to ZZ$ analyses may exhibit minor differences in implementation of nuisance parameters, the underlying sources of systematic uncertainty are broadly consistent. The list presented below corresponds to the off-shell analysis described in this thesis. The quantitative impact of these uncertainties and their breakdown within the fit are discussed in the following chapter.

- **Integrated luminosity (normalization)**

  Uncertainty in the determination of the integrated luminosity for the different Run 2 data-taking periods, affecting the normalization of all simulated processes. The corresponding uncertainties are between 0.18%–0.64%.

- **QCD scale uncertainties (normalization and shape)**

  Residual theoretical uncertainty associated with variations of the renormalization (μ_R) and factorization (μ_F) scales in perturbative calculations. For VV, ttH, and vector-boson-fusion Higgs production, these uncertainties affect the overall normalization at the level of 1%–4%. For ggF signal, interference and background and for $q\bar{q} \to ZZ$, they additionally induce shape variations in the kinematic distributions. These uncertainties are evaluated by taking the envelope [1] of variations of μ_R and μ_F.

- **Parton distribution functions and α_s (normalization and shape)**

  Uncertainties due to the limited knowledge of parton distribution functions and the strong coupling constant α_s. For Higgs production via vector-boson fusion and associated processes, they affect the normalization at the level of 1%–2%, while for ggF signal, interference and background and for $q\bar{q} \to ZZ$, they additionally induce shape variations in the kinematic distributions. These uncertainties are evaluated using PDF eigenvector variations and changes in α_s.

- **Electroweak corrections (shape)**

  Uncertainty associated with higher-order electroweak corrections applied to the $q\bar{q} \to ZZ$ background, affecting the shape of the relevant observables. It is estimated by varying the size of the applied electroweak correction. It is evaluated by taking the envelope of variations of μ_R and μ_F.

- **$gg \to ZZ$ correction factor (normalization)**

  Uncertainty on the higher-order correction factor applied to the gluon-induced ZZ background, amounting to approximately 10% for the dominant contribution, and about 5% for interference terms. This contribution is described in Sec. 5.3.2.

- **Higgs boson branching fraction (normalization)**

  Theoretical uncertainty on the branching fraction $\mathcal{B}(H \to ZZ \to 4\ell)$, of 2%. It is obtained by combining in quadrature the contributions from missing higher-order corrections and parametric uncertainties on the SM input parameters, such as particle masses.

- **Lepton reconstruction and identification (normalization)**

  Uncertainty in electron and muon reconstruction, identification and trigger efficiencies. The normalization uncertainties are approximately 3.5%–6% for electrons and 1%–2% for muons. These are evaluated using methods described in Sec. 3.4.

---

[1] The envelope is defined as the largest upward and downward deviations with respect to the nominal prediction among all considered variations.

- **Z+X background normalization (normalization)**

  Data-driven uncertainty in the normalization of the reducible Z+X background, ranging from approximately 3.5% to 10.4% depending on the final state and data-taking period. It is estimated using methods described in Sec. 4.2.

- **Jet energy scale and resolution (shape)**

  Uncertainty in the calibration of reconstructed jet energies, affecting primarily the shape of kinematic distributions, with a subleading impact on normalization. It is evaluated by varying jet energy corrections and resolutions within their uncertainties.

- **b-tagging efficiency (normalization)**

  Uncertainty in the identification efficiency and mistag rate of b-jets, amounting to 0.5%.

- **Parton shower and underlying event modeling (normalization and shape)**

  Uncertainty associated with variations of the parton shower tune and radiation modeling. It affects both normalization (typically below 1%) and shape of distributions, and is evaluated by comparing alternative PYTHIA tunes and radiation settings.

4.5 Differential Measurements of the Higgs Boson Production Cross Section

Differential measurements of the Higgs boson production in the $H \to ZZ \to 4\ell$ final state provide a detailed characterization of the production dynamics. In particular, the transverse momentum of the Higgs boson, p_T^H, is directly sensitive to higher-order QCD effects, initial-state radiation, and loop-induced contributions in gluon-fusion production.

Recent measurements at $\sqrt{s} = 13.6$ TeV with 2022 data [50] have extracted the differential cross section as a function of p_T^H by performing fits to the four-lepton invariant mass distribution in bins of Higgs transverse momentum. Electron identification efficiency SFs derived in this thesis were incorporated in this measurement to correct simulated events to data, contributing to the precision of the extracted cross sections. Corresponding data/MC agreement and differential cross section distributions are shown on Figures 4.2 and 4.3.

Differential cross sections also provide important inputs for coupling interpretations. In the on-shell determination of the Higgs self-coupling modifier, presented in the next section, the production cross section is parameterized in bins of p_T^H, allowing potential deviations of the Higgs self-coupling to manifest through modifications of the transverse-momentum spectrum.

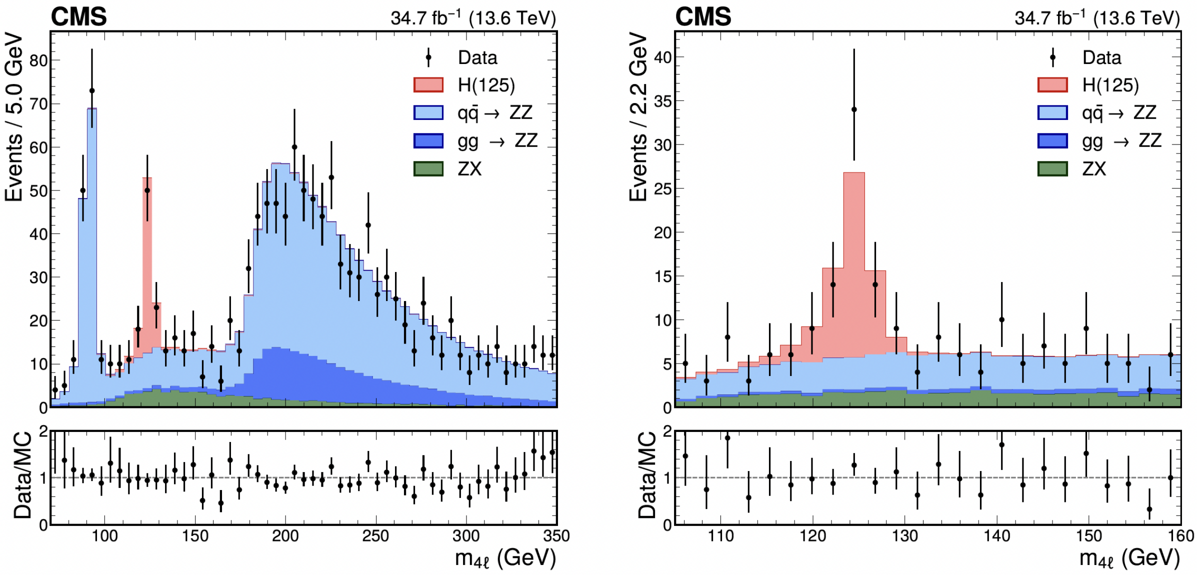


Figure 4.2: Distribution of the four-lepton invariant mass $m_{4\ell}$ in the ranges 70–350 GeV (left) and 105–160 GeV (right). The black points with error bars represent the observed data, while the colored histograms show the signal (red) and the individual background contributions. The post-fit normalization for all processes is obtained from the measurement performed in the range $105 < m_{4\ell} < 160$ GeV and subsequently propagated to the other distributions and mass ranges by scaling with the ratio of post-fit to pre-fit yields. The lower panels display the ratio of the data to the post-fit MC prediction.

4.6 On-shell Measurement of Higgs Self-coupling Modifier

Measurements of the Higgs boson self-coupling modifier are of fundamental importance, as they directly probe the structure of the Higgs potential and the mechanism of electroweak symmetry breaking. In the SM, the most direct sensitivity to κ_λ arises from Higgs boson pair production.

In this Section, the sensitivity to κ_λ in the on-shell regime is studied using differential distributions in the Higgs transverse momentum, p_T^H, following Ref. [76]. Representative NLO electroweak Feynman diagrams involving the Higgs self-coupling are shown in Fig. 4.4.

It is assumed that possible BSM effects modify only the Higgs self-coupling, while all other Higgs couplings remain SM-like. Under this assumption, the impact of κ_λ is incorporated through a parameterisation of the signal strength modifier μ in each p_T^H bin.

An anomalous Higgs self-coupling affects both the overall normalisation and the shape of kinematic distributions. The transverse momentum of the Higgs boson is especially sensitive to such effects and allows the contribution from a modified self-coupling to be disentangled from other deviations.

The scaling function for each production mode i and decay channel j is written as

$$\mu_{ij}(\kappa_\lambda) = \mu_{ij}^{\text{prod}}(\kappa_\lambda) \times \mu^{\text{dec}}(\kappa_\lambda), \tag{4.7}$$

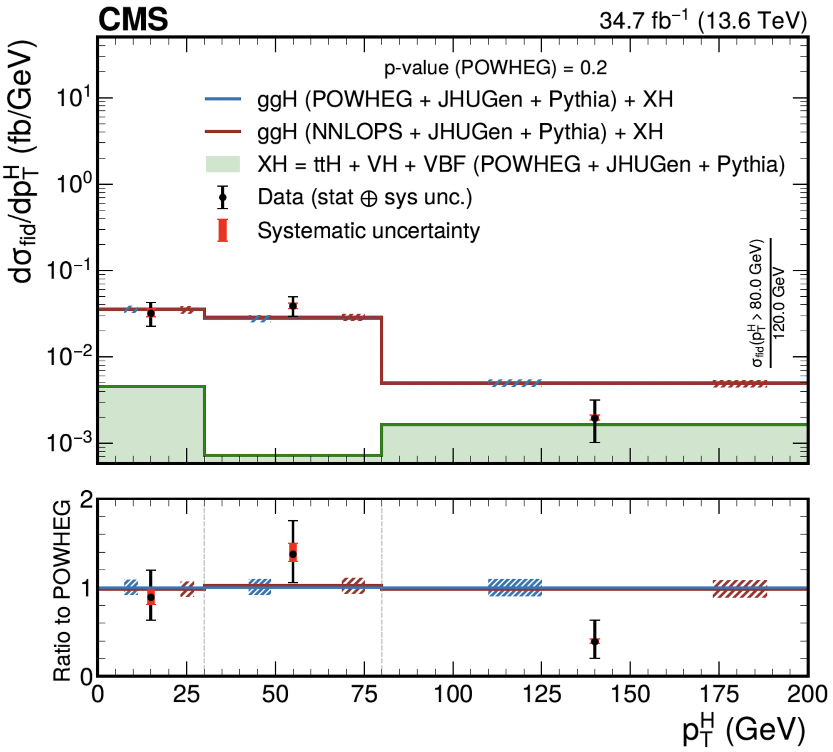


Figure 4.3: Differential cross section measured in the $H \to ZZ \to 4\ell$ channel at $\sqrt{s} = 13.6$ TeV as a function of the Higgs boson transverse momentum p_T^H. The acceptance and theoretical uncertainties in each bin are evaluated using gg $\to H$ predictions from two generators, normalized to next-to-next-to-leading-order accuracy. The subdominant signal contributions from VBF, VH, and $t\bar{t}H$ production (denoted as XH) are fixed to their SM expectations. The measured cross sections are compared to predictions obtained with Powheg (blue) and NNLOPS (dark red). The hatched bands indicate the theoretical uncertainties associated with the predictions. Black markers represent the measured cross section in each bin, with error bars corresponding to the total uncertainty, while red boxes indicate the systematic component. The highest transverse-momentum bin corresponds to $p_T^H > 80$ GeV and is normalized to a bin width of 120 GeV. The lower panel displays the ratio of the measured cross section to the Powheg prediction, together with the ratio of the NNLOPS prediction to Powheg. The compatibility between data and theoretical predictions is quantified using a p-value of 0.2 [50].

where the fiducial cross section and branching ratios are normalised to their SM expectations.

The production component is parameterised as

$$\mu_{ij}^{\text{prod}}(\kappa_\lambda) = \frac{1 + \kappa_\lambda C_{1,ij} + \delta Z_H}{(1 - (\kappa_\lambda - 1)\delta Z_H)(1 + C_{1,ij} + \delta Z_H)}. \tag{4.8}$$

The term δZ_H denotes the universal correction arising from Higgs wave-function renormalisation and is common to all production processes:

$$\delta Z_H = -\frac{9}{16\sqrt{2}\pi^2}\left(\frac{2\pi}{3\sqrt{3}} - 1\right) G_\mu m_H^2. \tag{4.9}$$

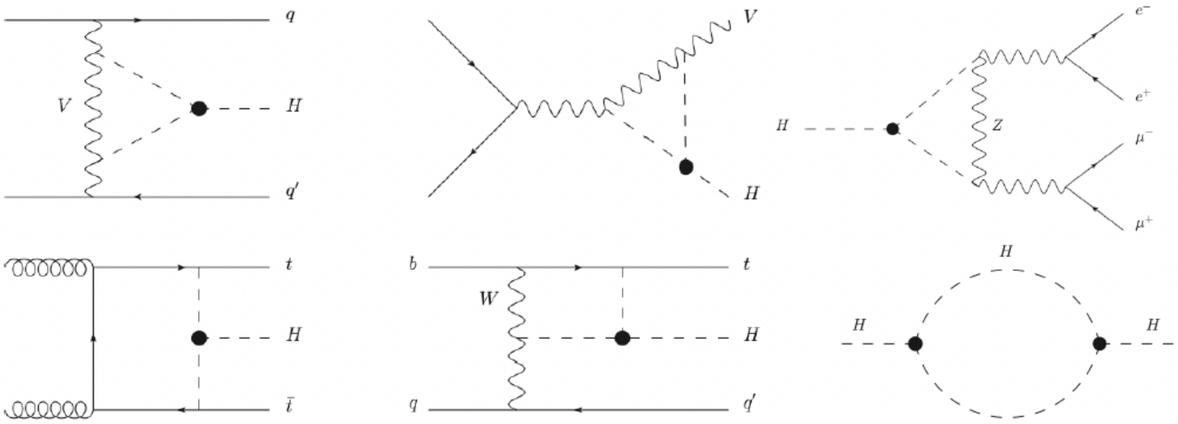


Figure 4.4: Representative NLO electroweak Feynman diagrams involving the Higgs self-coupling. The left and central panels illustrate contributions to single-Higgs production, the top-right panel shows the decay into four leptons, and the bottom-right panel depicts the one-loop contribution to the Higgs self-energy [76, 77].

The coefficients $C_{1,ij}$ describe the process- and kinematic-dependent contributions from the $\mathcal{O}(\lambda)$ corrections. They are defined as

$$C_{1,ij} = \frac{2\,\mathrm{Re}\Big(\mathcal{A}_{\mathrm{LO}}\mathcal{A}^{\lambda\,*}_{\mathrm{NLO}}\Big)}{|\mathcal{A}_{\mathrm{LO}}|^2}, \tag{4.10}$$

where $\mathcal{A}$ denotes the corresponding matrix element. The coefficient $C_{1,ij}$ therefore quantifies the relative strength of the λ-dependent NLO correction with respect to the leading-order contribution.

The values of $C_{1,ij}$ used in this analysis are summarised in Table 4.3. They are provided in five $p_{\mathrm{T}}^{\mathrm{H}}$ intervals for the main production modes. For ggF, the coefficient is taken to be constant across all bins due to the absence of a fully differential theoretical prediction. The sensitivity generally increases at higher transverse momenta, particularly for the $\mathrm{t\bar{t}}H$ production mode.

$p_{\mathrm{T}}^{\mathrm{H}}$ bin	ggH	VBFH	VH	$\mathrm{t\bar{t}}H$
0–45	0.66	0.66	1.66	5.31
45–80	0.66	0.66	1.23	4.73
80–120	0.66	0.64	0.77	3.92
120–200	0.66	0.58	0.35	2.79
200–∞	0.66	0.55	0.02	1.42

Table 4.3: Values of the $C_{1,ij}$ coefficients (in %) for the different production modes in the five $p_{\mathrm{T}}^{\mathrm{H}}$ bins [78]. The bin boundaries are given in GeV.

Scaling of Higgs production cross-section, as a function of κ_λ, is depicted on Figure 4.5.

The decay component is parameterised as:

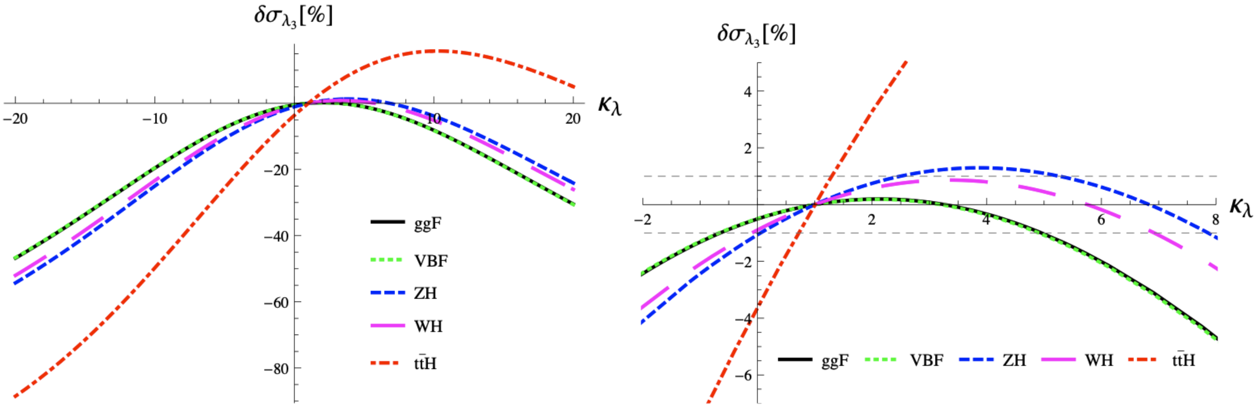


Figure 4.5: Percentage change in SM cross-section, induced by varying κ_λ for ggH (solid black), VBF (dashed green), ZH (dashed blue), WH (solid magenta) and $t\bar{t}H$ (dashed red) Higgs production mechanisms. Plot on the left side shows the scaling on $-20 < \kappa_\lambda < 20$ range while plot on the right side shows scaling on $-2 < \kappa_\lambda < 8$ range [77].

$$\mu^{\text{dec}}(\kappa_\lambda) = 1 + (\kappa_\lambda - 1)\frac{C_1^{\Gamma_{ZZ}} - C_1^{\Gamma_{\text{tot}}}}{1 + (\kappa_\lambda - 1)C_1^{\Gamma_{\text{tot}}}}, \tag{4.11}$$

where the C_1 coefficients encode the $\mathcal{O}(\lambda)$ corrections to the partial and total decay widths $C_1^{\Gamma_{ZZ}} = 0.0083$ and $C_1^{\Gamma_{\text{tot}}} = 2.5 \times 10^{-3}$. Corresponding scalings are shown on Figure 4.6.

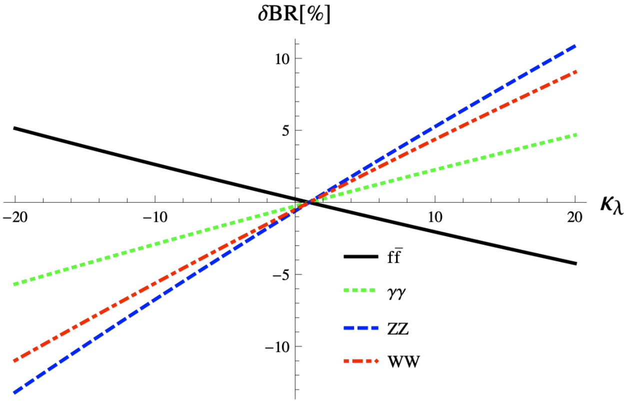


Figure 4.6: Percentage change in decay branching ration for $f\bar{f}$ (solid black), $\gamma\gamma$ (dashed green), ZZ (dashed blue), WW (dashed red) decay channels [77].

The profile negative log-likelihood as a function of κ_λ is shown on Figure 4.7. The observed (expected) range of the parameter κ_λ at the 95% confidence level is $-5.4\,(-7.6) < \kappa_\lambda < 14.9\,(17.7)$.

Finally, $H \to ZZ \to 4\ell$ constraint is compared to the most sensitive direct searches for Higgs boson pair production. As shown in Fig. 4.8, the current limits from double-Higgs channels generally provide stronger constraints on κ_λ. This is expected, as di-Higgs production offers the most direct sensitivity to the trilinear coupling.

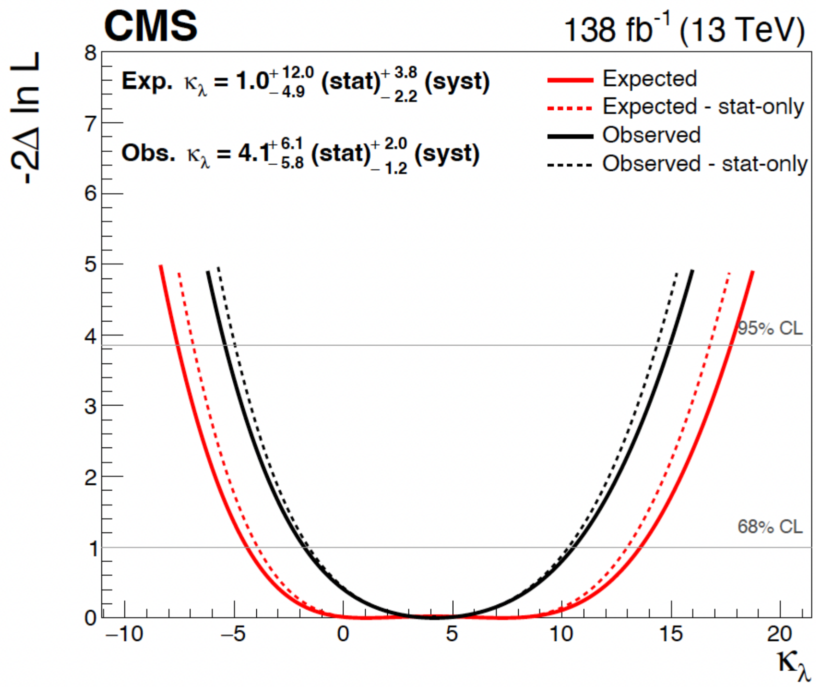


Figure 4.7: On-shell κ_λ observed and expected scans, with and without systematic uncertainties [78].

Nevertheless, it is important to explore complementary regions of phase space and alternative sensitivity mechanisms. Indirect probes, including both the on-shell approach presented in this chapter and the off-shell strategy discussed later in this thesis, access different kinematic regimes and theoretical structures. Although individually less constraining at present, such measurements probe directions in parameter space that are not fully covered by direct searches and may provide valuable input in a global combination.

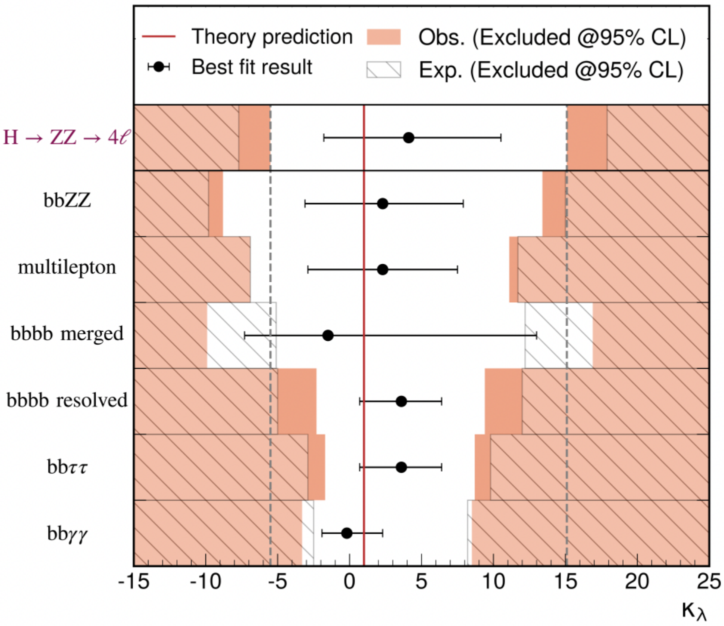


Figure 4.8: Comparison of the $H \to ZZ \to 4\ell$ result with the most stringent direct constraints from double-Higgs production searches performed by the CMS Collaboration. The constraint obtained in the bbZZ final state (four leptons plus two b jets) is taken from [79]. The result for Higgs boson pairs decaying to $WWWW$, $WW\tau\tau$, and $\tau\tau\tau\tau$ (multilepton final states) is taken from [80]. The constraint from the production of highly boosted Higgs boson pairs decaying to bottom quarks (bbbb merged topology) is taken from [81]. The measurement in the four bottom quark final state (bbbb resolved topology) is taken from [82]. The result in the bb$\tau\tau$ final state is taken from [83], while the constraint from the bb$\gamma\gamma$ final state is taken from [84].

Chapter 5

Higgs Boson Properties from the Off-shell Region

As part of this thesis, I performed and completed the study of the Higgs boson self-coupling modifier κ_λ in the off-shell region using the full Run 2 dataset. I extended the SMEFT corrections from [1], originally derived for gluon fusion production, to the electroweak sector. I implemented both the gluon-fusion and electroweak SMEFT corrections within the analysis framework, and carried out the first measurement of κ_λ in the off-shell regime, establishing a novel and previously unexplored probe of the Higgs self-interaction. The study of the self-coupling constitutes one component of a broader analysis that probes additional Higgs boson properties from the off-shell region, including Yukawa couplings, possible compositeness effects, and the Higgs boson width. The complete analysis has been reviewed and approved by the CMS Collaboration.

This chapter presents measurements of the Higgs boson production in the high-mass region using the four-lepton decay channel. The analysis is based on the full Run 2 proton–proton collision dataset collected by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}.

The motivation for studying this region of phase space, along with its distinctive features in the $m_{4\ell}$ distribution, is discussed in Sec. 1.4.2. The MC datasets used in the analysis are introduced in Sec. 5.1. Observables and event categorization are described in Sec. 5.2, while the modeling of signal, background, and interference contributions is presented in Sec. 5.3. The statistical framework is outlined in Sec. 5.4.

Section 5.5 details the strategy for the measurement of κ_λ, including the implementation of SMEFT corrections in MC generators and the analysis framework, as well as the final fits and estimation of confidence intervals. Additional measurements in the off-shell region are presented in Sec. 5.6, covering Higgs compositeness, Yukawa couplings, and the total width. Feymann diagrams in this chapter have been made using the *TikZ-Feynman* package [85].

5.1 Datasets and MC

Off-shell analysis presented in this thesis uses the full Run 2 data with UL processing. The datasets used in the analysis are listed in Table 5.1. The main signal production modes considered in this analysis are ggF and electroweak production (EW), which includes VBF and VH [1]. The ggF signal is modeled using both MCFM–JHUGEN leading-order (LO) samples and POWHEG next-to-leading-order (NLO) high-mass samples. The EW signal is modeled using MCFM–JHUGEN LO samples as well as PHANTOM LO samples. The dominant irreducible background from $q\bar{q} \to ZZ \to 4\ell$ is simulated using POWHEG. The MC samples used in the analysis are summarized in Table 5.2. MC modelling is discussed in detail in Section 5.3.

Dataset name
/SingleMuon/Run2018A-UL2018_MiniAODv2-v3
/DoubleMuon/Run2018A-UL2018_MiniAODv2-v1
/EGamma/Run2018A-UL2018_MiniAODv2-v1
/MuonEG/Run2018A-UL2018_MiniAODv2-v1
/SingleMuon/Run2018B-UL2018_MiniAODv2-v2
/DoubleMuon/Run2018B-UL2018_MiniAODv2-v1
/EGamma/Run2018B-UL2018_MiniAODv2-v1
/MuonEG/Run2018B-UL2018_MiniAODv2-v1
/SingleMuon/Run2018C-UL2018_MiniAODv2-v2
/DoubleMuon/Run2018C-UL2018_MiniAODv2-v1
/EGamma/Run2018C-UL2018_MiniAODv2-v1
/MuonEG/Run2018C-UL2018_MiniAODv2-v1
/SingleMuon/Run2018D-UL2018_MiniAODv2-v3
/DoubleMuon/Run2018D-UL2018_MiniAODv2-v1
/EGamma/Run2018D-UL2018_MiniAODv2-v1
/MuonEG/Run2018D-UL2018_MiniAODv2-v1

Table 5.1: Data samples used in the analysis.

All MC samples are interfaced with PYTHIA 8 [86] to simulate parton showering and multi-parton interactions, using version 8.230 for all years with the CUETP8M1 tune [87] for the simulation of the 2016 data-taking period, and the CP5 tune [88] for the simulation of the 2017 and 2018 data taking periods. The NNPDF 3.0 parton distribution functions are used [89]. Simulated events include the contribution from additional proton-proton interactions within the same or adjacent bunch crossings (pileup) and are weighted to reproduce the observed pileup distribution in data. The MC samples are further processed through a dedicated simulation of the CMS detector based on GEANT4 [90].

---

[1] Vector boson scattering (VBS) is also an electroweak process, but is treated as a background in this analysis. In addition, the process $q\bar{q} \to ZZ$ constitutes the dominant background in the off-shell region, and does not interfere with the VBF, VH, or VBS contributions.

ggF (MCFM/JHUGen LO)
/GluGluTo*ToZZTo*_M125_*_TuneCP5_13TeV-mcfm701-pythia8/*/MINIAODSIM
ggF (POWHEG NLO high-mass)
/GluGluHToZZTo4L_M*_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8/*/MINIAODSIM
EW (MCFM/JHUGen LO)
/VBFTo*ToZZTo*_M125_GaSM_TuneCP5_withDipoleRecoil_13TeV-jhugenv730-mcfm701-pythia8/*/MINIAODSIM
EW (Phantom LO)
/VBFTo*ToZZTo*muJJ_TuneCP5_withDipoleRecoil_13TeV-phantom-pythia8/*/MINIAODSIM
$q\bar{q} \to ZZ \to 4\ell$ background (POWHEG NLO QCD + LO EW)
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/*/MINIAODSIM

Table 5.2: MC samples used in the analysis.

5.2 Observables and Event Categorization

A common strategy in Higgs analyses is to categorize events based on their kinematic properties, as different regions of phase space are sensitive to different observables and production modes. For the off-shell analysis, these observables are $m_{4\ell}$ and kinematic dicriminants, as described in 4.3.1. A set of discriminants, denoted $\mathcal{D}_{\text{2jet}}$, is constructed following Eq. 4.3. In this formulation, P_{sig} corresponds to the signal probability for the VBF (WH or ZH) production hypothesis in the VBF-tagged (or VH-tagged) category, while P_{alt} represents the probability for Higgs boson production in association with two jets via gluon fusion. These discriminants exploit the full kinematic information of the Higgs boson candidate and its associated jets. They are labeled to indicate both the event topology ("2jet") and the targeted production mechanism (VBF, WH, or ZH), which is discriminated against the dominant gluon fusion process: $\mathcal{D}^{\text{VBF}}_{\text{2jet}}$, $\mathcal{D}^{\text{ZH}}_{\text{2jet}}$, and $\mathcal{D}^{\text{WH}}_{\text{2jet}}$.

When more than two jets satisfy the selection criteria, the two jets with the highest transverse momentum are used in the matrix-element calculations. The $\mathcal{D}_{\text{2jet}}$ discriminants therefore separate the target production mode in each category from gluon fusion production, using the kinematic properties of the Higgs boson candidate and the two associated jets.

Sequential selection criteria are applied to define the event categories:

- **VBF-2jet category:** Events are required to contain exactly four leptons. In addition, events must have either two or three jets with at most one identified as originating from a b quark (b-tagged jet), or at least four jets with no b-tagged jets. Furthermore, events must satisfy $\mathcal{D}^{\text{VBF}}_{\text{2jet}} > 0.5$.
- **VH-hadronic category:** Events must contain exactly four leptons and either two or three jets, or at least four jets with no b-tagged jets. They are required to satisfy $\max\left(\mathcal{D}^{\text{WH}}_{\text{2jet}}, \mathcal{D}^{\text{ZH}}_{\text{2jet}}\right) > 0.5$.
- **Untagged category:** All remaining events are assigned to this category.

Categorization cuts and corresponding observables are summarized in Table 5.3

Category	VBF-tagged	VH-tagged	Untagged
Selection	$\mathcal{D}_{\rm 2jet}^{\rm VBF} > 0.5$	$\mathcal{D}_{\rm 2jet}^{\rm ZH}$ or $\mathcal{D}_{\rm 2jet}^{\rm WH} > 0.5$	Rest of events
Observables	$m_{4\ell}$, $\mathcal{D}_{\rm bkg}^{\rm VBF+dec}$, $\mathcal{D}_{\rm bsi}^{\rm VBF+dec}$	$m_{4\ell}$, $\mathcal{D}_{\rm bkg}^{\rm VH+dec}$, $\mathcal{D}_{\rm bsi}^{\rm VH+dec}$	$m_{4\ell}$, $\mathcal{D}_{\rm bkg}^{\rm kin}$, $\mathcal{D}_{\rm bsi}^{gg,{\rm dec}}$

Table 5.3: Summary of observables and cuts used for event categorization. All discriminants are calculated with the JHUGEN signal and MCFM background matrix elements. The VH interference discriminant in the VH-tagged category is defined as the average of the corresponding ZH and WH discriminants.

Observables $\mathcal{D}_{\rm bkg}^{kin}$ and $\mathcal{D}_{\rm bsi}$ are defined differently in each event category, as summarized in Table 5.4 [2].

In the Untagged category, the kinematic discriminant $\mathcal{D}_{\rm bkg}^{\rm kin}$ separates the ggF Higgs signal from the dominant $q\bar{q} \to ZZ \to 4\ell$ background, using the corresponding matrix-element probabilities $\mathcal{P}_{\rm ggH}$ and $\mathcal{P}_{q\bar{q}\to ZZ}$. The interference discriminant $\mathcal{D}_{\rm bsi}^{gg,{\rm dec}}$ is constructed from the ggF signal–background interference term, $\mathcal{P}_{\rm int}^{gg}$, together with the ggF signal probability $\mathcal{P}_{\rm ggH}$ and the gluon-induced background probability $\mathcal{P}_{gg\to ZZ}$.

In the VBF-tagged category, the background discriminant $\mathcal{D}_{\rm bkg}^{\rm VBF+dec}$ is constructed using the EW Higgs production probabilities $\mathcal{P}_{\rm VBF}$, $\mathcal{P}_{\rm ZH}$, and $\mathcal{P}_{\rm WH}$, and separates them from alternative processes including vector boson scattering (VBS, see the Feymann diagram on Fig. 5.11(b)), ZZZ, WZZ, and QCD background with 2 jets [3], which are collectively denoted as $\mathcal{P}_{\rm alt}^{\rm EW+QCD}$. The corresponding interference discriminant $\mathcal{D}_{\rm bsi}^{\rm VBF+dec}$ is built from the interference between the VBF Higgs signal and the VBS background, using the probabilities $\mathcal{P}_{\rm VBF}$, $\mathcal{P}_{\rm VBS}$, and the corresponding interference term $\mathcal{P}_{\rm int}^{\rm VBF}$ between VBS background and VBF signal.

In the VH-tagged category, the background discriminant $\mathcal{D}_{\rm bkg}^{\rm VH+dec}$ is defined analogously, using $\mathcal{P}_{\rm VBF}$, $\mathcal{P}_{\rm ZH}$, and $\mathcal{P}_{\rm WH}$ against the same set of alternative processes. The interference discriminant $\mathcal{D}_{\rm bsi}^{\rm VH+dec}$ is constructed separately for ZH and WH production, using the interference between each signal process (ZH, WH) and its corresponding $q\bar{q} \to ZZ + Z/W$ interfering background. Final discriminant is computed as the average of ZH and WH discriminants.

For the $D_{\rm bkg}^{\rm kin}$ observables, $m_{4\ell}$ is also included in the computation. In the VBF-tagged and VH-tagged categories, additional information from the kinematic properties of the two associated jets is incorporated, enhancing the discrimination of the targeted signal production against both background processes and Higgs boson production via gluon fusion. The Higgs boson mass is fixed to $m_{\rm H} = 125.38$ GeV in the computation of all probabilities.

Example of the discriminating power of $\mathcal{D}_{\rm bkg}^{\rm kin}$ is shown in Figure 5.1.

---

[2] The expressions shown in Table 5.4 omit normalization constants.

[3] The contributions denoted as ZZZ and WZZ correspond to non-resonant electroweak production of a vector boson in association with a ZZ pair, arising from multi-boson production diagrams without an intermediate Higgs boson. The QCD $ZZ + 2$ jets background refers to ZZ production with two jets, including contributions from $q\bar{q}$ and gg processes.

Category	**Discriminant definition**
Untagged	$\mathcal{D}_{\text{bkg}}^{\text{kin}} = \dfrac{\mathcal{P}_{\text{ggH}}}{\mathcal{P}_{\text{ggH}} + \mathcal{P}_{q\bar{q}\to ZZ}}$ $\mathcal{D}_{\text{bsi}}^{gg,\text{dec}} = \dfrac{\mathcal{P}_{\text{int}}^{gg}}{2\sqrt{\mathcal{P}_{\text{ggH}}\,\mathcal{P}_{gg\to ZZ}}}$
VBF-tagged	$\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}} = \dfrac{\mathcal{P}_{\text{VBF}} + \mathcal{P}_{\text{ZH}} + \mathcal{P}_{\text{WH}}}{\mathcal{P}_{\text{VBF}} + \mathcal{P}_{\text{ZH}} + \mathcal{P}_{\text{WH}} + \mathcal{P}_{\text{alt}}^{\text{EW+QCD}}}$ $\mathcal{D}_{\text{bsi}}^{\text{VBF+dec}} = \dfrac{\mathcal{P}_{\text{int}}^{\text{VBF}}}{2\sqrt{\mathcal{P}_{\text{VBF}}\,\mathcal{P}_{\text{VBS}}}}$
VH-tagged	$\mathcal{D}_{\text{bkg}}^{\text{VH+dec}} = \dfrac{\mathcal{P}_{\text{VBF}} + \mathcal{P}_{\text{ZH}} + \mathcal{P}_{\text{WH}}}{\mathcal{P}_{\text{VBF}} + \mathcal{P}_{\text{ZH}} + \mathcal{P}_{\text{WH}} + \mathcal{P}_{\text{alt}}^{\text{EW+QCD}}}$ $\mathcal{D}_{\text{bsi}}^{\text{VH+dec}} = \frac{1}{2}\left[\dfrac{\mathcal{P}_{\text{int}}^{\text{ZH}}}{2\sqrt{\mathcal{P}_{\text{ZH}}\,\mathcal{P}_{\text{ZZZ}}}} + \dfrac{\mathcal{P}_{\text{int}}^{\text{WH}}}{2\sqrt{\mathcal{P}_{\text{WH}}\,\mathcal{P}_{\text{WZZ}}}}\right]$

Table 5.4: Definitions of the kinematic and interference discriminants in different event categories.

5.3 Modelling of the Off-shell Production

The off-shell signal and background ggF process is simulated using MCFM 7.0.1 [44, 92, 93, 94]. The generation of the Higgs boson signal, including both SM and anomalous coupling (AC) hypotheses, as well as its interference with the continuum background, is performed within the JHUGEN 7.3.0 framework [95, 96, 97, 98, 99]. Samples are produced at LO and subsequently reweighted using MELA to obtain predictions corresponding to the SM $gg \to H \to ZZ$ signal, the continuum $gg \to ZZ$ background, and their interference. The MELA package is additionally used to reweight events to alternative coupling hypotheses probed in this analysis, including scenarios with modified κ_λ parameters within SMEFT framework.

Off-shell ggF production is generated at LO without associated jets at the matrix-element level, so any reconstructed jets in these samples arise only from the parton shower in PYTHIA.

In the VBF-tagged and VH-tagged categories, where the event selection and the discriminant observables explicitly depend on the kinematics of the two leading jets, the template shapes are taken from ggF samples generated with POWHEG including up to two associated jets. The corresponding templates are then scaled such that their total yields match the yields predicted by MCFM.

The EW processes are simulated following a similar strategy. The MCFM and JHUGEN generators are used to model EW SM and anomalous coupling contributions. These samples are reweighted with MELA to obtain the EW SM signal (VBF and VH), the EW interfering background (dominated by VBS but also includes ZZZ and WZZ), and their interference.

To summarize, off-shell anomalous coupling and SM ggF samples produced with

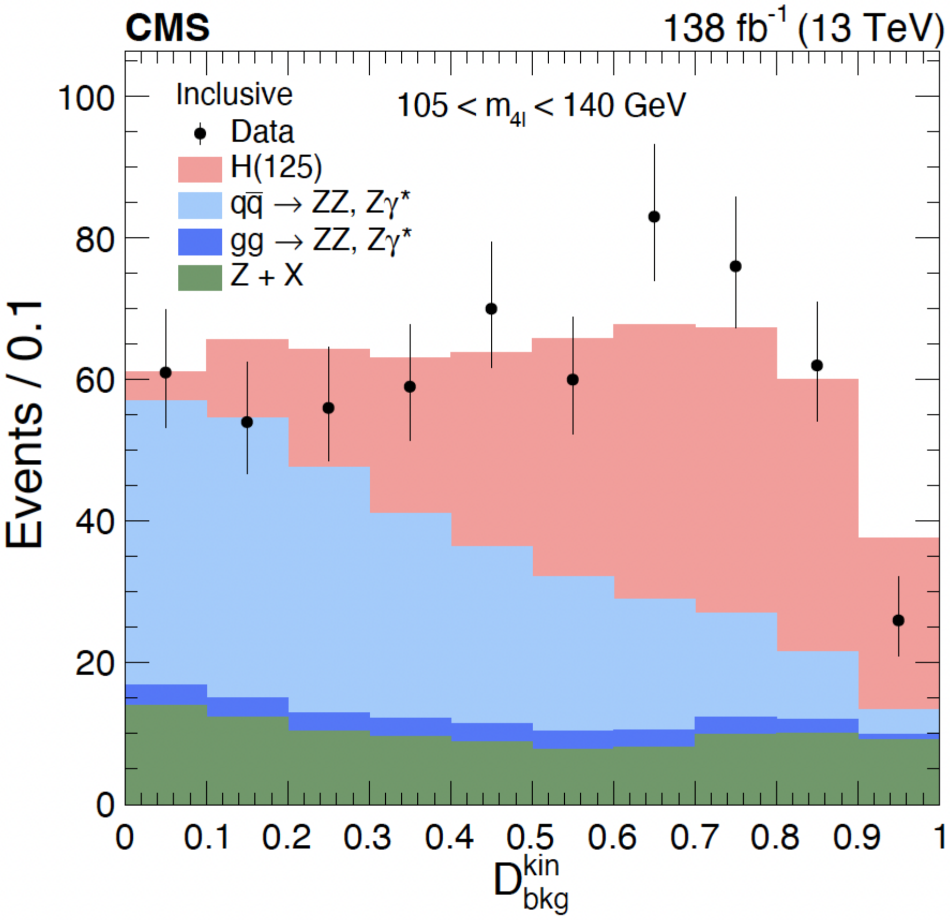


Figure 5.1: Distributions of the observed data (points) and predicted contributions (stacked histograms) of the kinematic discriminant $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ for the four-lepton system in the inclusive final state. The Higgs boson signal and the three dominant background processes are represented by different colors. The vertical bars on the data points indicate the statistical uncertainties [91].

MCFM, together with POWHEG samples listed in Table 5.2, are reweighted using MELA to obtain ggF signal, background, interference, and SMEFT contributions. In the Untagged category, both shapes and yields from MCFM are used for ggF production. For the tagged categories, the yields are taken from MCFM, while the shapes are modeled using POWHEG. For EW production, MCFM EW off-shell anomalous coupling and SM samples from Table 5.2 are similarly reweighted using MELA to obtain EW signal, background, interference, and SMEFT contributions. In addition, EW PHANTOM samples are included to recover approximately 10% of events that were removed from the original gridpacks due to additional jet requirements [4].

These constitute the set of samples used for reweighting. The non-interfering $q\bar{q} \to ZZ$ background is simulated at NLO accuracy using POWHEG.

The remaining background, $Z + X$, is estimated directly from data, using methods described in Section 4.2.

---

[4] A gridpack is a pre-generated set of parton-level events from a MC generator, enabling efficient and consistent large-scale event production without rerunning the full simulation.

5.3.1 Cross-feed from On-shell to Off-shell Region

In the off-shell region, processes other than genuine off-shell Higgs boson production can contribute to the selected four-lepton final state. Certain on-shell Higgs boson production modes may be reconstructed in the off-shell region due to the presence of additional leptons in the event. This on-shell contribution only accounts to 0.1% of events selected in the off-shell region, but is still considered in the analysis.

A representative example arises in associated ZH production. If the Higgs boson decays to $2\ell + X$, while the associated Z boson decays leptonically, the event can contain four reconstructed leptons satisfying the analysis selection. In such configurations, the four-lepton system does not necessarily originate from a single off-shell Higgs boson decay, but may instead correspond to an on-shell Higgs boson accompanied by additional leptons from associated production. Depending on the decay topology, this can lead to a reconstructed four-lepton invariant mass above 220 GeV, thereby populating the off-shell signal region.

The dominant on-shell Higgs boson contribution of this type originates from $Z(\ell\ell)H(2\ell+X)$ production. These contributions are referred to as *cross-feed* processes, as they migrate from on-shell Higgs production into the off-shell analysis region.

The cross-feed contribution from on-shell $ZH(125)$ production is estimated using dedicated on-shell simulation samples in which the Higgs boson decays to WW or ZZ. At generator level, a lepton filter is applied in order to ensure that the final state contains four leptons of any flavor. The dominant cross-feed contributions arise from $H(125) \to 2\ell 2\nu$ and $H(125) \to 2\ell 2q$ decays produced in association with $Z \to 2\ell$. Additional on-shell production modes, such as $t\bar{t}H$, are also evaluated and included where relevant, while other rare decay channels are found to give negligible contributions.

Dedicated templates are constructed from the on-shell ZH samples to model the cross-feed as a separate on-shell process. To avoid double counting, these cross-feed contributions are removed from the off-shell EW signal modeling using generator-level information. Specifically, simulated $H \to 2\ell 2q$ events in the off-shell EW JHUGEN sample with a Higgs boson mass within 100 MeV of the mass pole are excluded. This procedure achieves nearly 100% efficiency in removing on-shell cross-feed events while eliminating a negligible fraction of non-signal $H(125)$ events. In addition to the ZH templates, the $t\bar{t}H$ contribution is modeled separately and incorporated into the statistical interpretation.

5.3.2 Corrections to Off-shell Simulation

The primary MC samples used to model signal and its interference with background are off-shell ggF and EW production, performed with the JHUGEN +MCFM generators. Such a complex modeling, including the effects of anomalous contributions, has been available only at LO in both QCD and EW.

To account for missing higher-order corrections, K factors are applied to the simulated samples. A K factor quantifies the impact of higher-order perturbative effects and is defined as the ratio of the cross section evaluated at a given higher order to the corresponding LO prediction,

$$K = \frac{\sigma_{\mathrm{N^nLO}}}{\sigma_{\mathrm{LO}}}.$$

In ggF Higgs boson production, higher-order QCD corrections provide the dominant contribution to the K factor, while EW corrections are typically smaller but become increasingly relevant at high energies, especially in the off-shell regime. Background processes such as $q\bar{q} \to ZZ$ and $gg \to ZZ$ similarly receive significant QCD and loop-induced contributions beyond LO accuracy.

These corrections are implemented as multiplicative weights applied to the generated events. Depending on the process, the K factors may be treated either as overall normalization factors or as differential corrections in observables, such as $m_{4\ell}$.

Gluon-fusion process k factor

To incorporate higher-order QCD corrections, signal cross section calculations are performed at LO, NLO, and next-to-next-to-leading order (NNLO) using the MCFM and HNNLO programs [100]. Corresponding NNLO to LO k-factors are used to reweight $m_{4\ell}$ distribution and are shown on Figure 5.2. The on-shell $gg \to H$ production cross section is reported in the LHC Higgs Cross Section Working Group Yellow Report 4 (YR4) [17]. To match this reference prediction, an additional mass-independent K factor accounting for NNLO-to-N^{3}LO QCD and NLO EW corrections is applied to the simulated samples. This correction has a value of 1.154.

While the LO-to-NNLO K factor is computed for the ggF signal process, it is applied not only to the SM signal contribution but also to the continuum $gg \to 4\ell$ background, to the signal–background interference, and to the SMEFT templates corresponding to modified κ_λ interactions. Strictly speaking, the signal K factor is exact only for the signal itself, while for the background and interference it represents an approximation. However, approximate NLO calculations available for both the background and interference processes demonstrate that their LO-to-NLO K factors agree with that of the signal within about 10% in the $m_{4\ell} > 220$ GeV region relevant to this analysis [101, 102, 103, 104].

This observation supports the assumption that higher-order QCD corrections factorize and can be applied uniformly to all components of the $gg \to ZZ$ amplitude. In this approach, the N^{3}LO QCD correction is assumed to rescale the full amplitude, including the SM signal, the SM background, their interference, and any additional contributions arising from SMEFT effects, such as those induced by modified κ_λ coupling. Such a treatment is standard in SMEFT-based Higgs analyses, where genuine higher-order QCD calculations explicitly including beyond-SM effects are generally unavailable. The same prescription has been adopted in previous off-shell Higgs boson measurements, including Higgs width studies [105].

To account for this assumption, an additional 10% uncertainty is assigned to the background component. Since the interference term depends linearly on the signal and background amplitudes, a corresponding 5% uncertainty is applied to the interference, reflecting the square-root scaling of the cross-section variation ($\sqrt{1+0.10} \approx 1.05$).

EW process k factor

YR4 provides cross-section for VBF and VH at NNLO in QCD and NLO in EW. Analogously to the ggF case, the on-shell cross section of the simulated EW samples is compared to the YR4 prediction in order to determine a universal correction factor. The derived K factor, equal to 1.039, is applied to the EW signal VBF and VH, signal–background interference, and EW interfering background contribution.

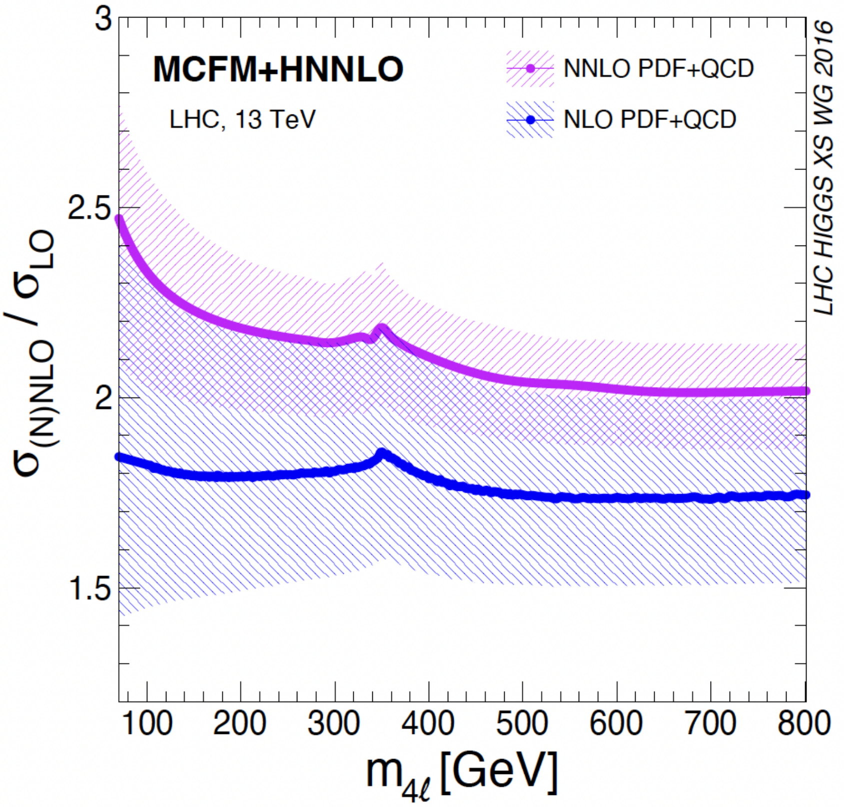


Figure 5.2: LO-to-NNLO and LO-to-NLO K factors as functions of the invariant mass for gluon-fusion Higgs boson production ($gg \to H$) at $\sqrt{s} = 13$ TeV [17]. LO-to-NNLO (in purple) is used in off-shell analysis to reweight $m_{4\ell}$ distribution.

The $q\bar{q} \to 4\ell$ background is simulated at NLO in QCD and LO in EW theory using POWHEG. The fully differential cross section for this process has been computed at NNLO in QCD [106], and a differential NLO-to-NNLO QCD K factor, evaluated as a function of $m_{4\ell}$, is applied to the POWHEG sample. The corresponding K factor is illustrated in Figure 5.3 (left).

To account for EW effects, an LO-to-NLO EW K factor [107] is applied to events containing two on-shell Z bosons. The mass dependence of this correction is shown in Figure 5.3 (right).

5.4 Statistical Framework

This analysis follows a template-based approach, where PDFs describing the observables are represented by histograms. To set up a statistical model, $3D$ templates in $(m_{4\ell}, \mathcal{D}^{\text{kin}}_{\text{bkg}}, \mathcal{D}_{\text{bsi}})$ are created for signal, background and interference contributions.

As this analysis employs three observables across three event categories, and given the presence of destructive interference effects in the off-shell region, a fully three-dimensional template construction can lead to a large number of poorly populated bins. The $m_{4\ell}$ binning is summarized in Table 5.5, where 21 mass bins are defined. In previous off-shell measurements [91], the analysis used 21 $m_{4\ell}$ bins together with 20 $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ bins and 10 $\mathcal{D}_{\text{bsi}}$ bins, resulting in a total of 4200 bins per category. In practice, many of these bins were

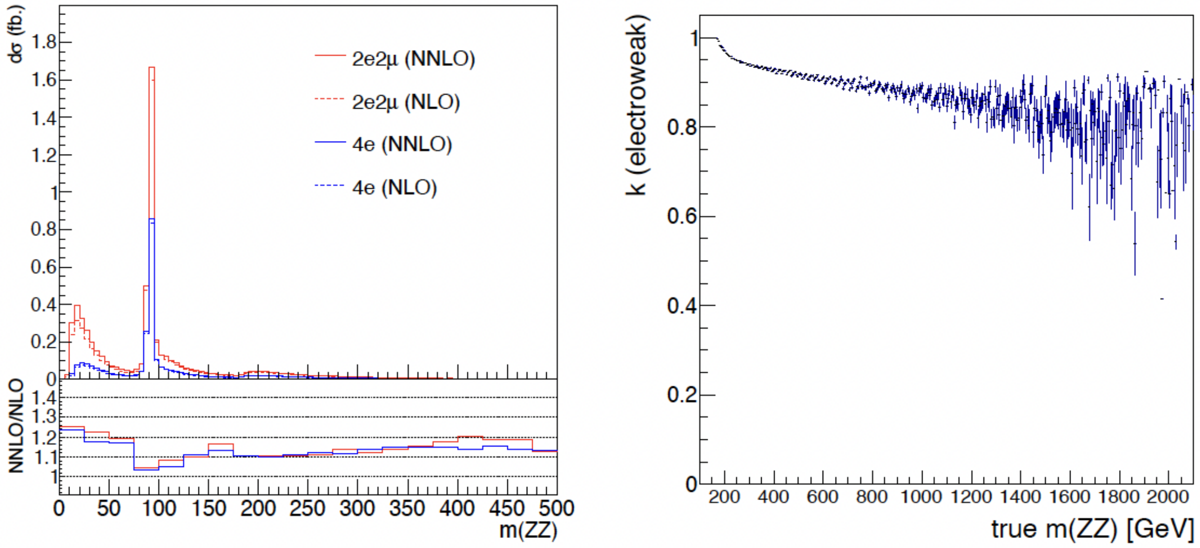


Figure 5.3: NLO-to-NNLO QCD K factor for the $q\bar{q} \rightarrow ZZ$ background (left) and LO-to-NLO EW K factor as a function of the invariant mass (right) at $\sqrt{s} = 13$ TeV [108].

sparsely populated, requiring ad hoc merging with neighboring bins to stabilize the fit.

Although the combination of $\mathcal{D}_{\rm bkg}^{\rm kin}$ and $\mathcal{D}_{\rm bsi}$ retains the full discriminating power when used together with the $m_{4\ell}$ distribution, performing a three-dimensional template fit in such a high-dimensional space can compromise fit stability. To address this limitation in a systematic way, the MiLoMerge procedure [109] is employed. Rather than merging bins manually after template construction, MiLoMerge optimally combines the two-dimensional observable $(\mathcal{D}_{\rm bkg}^{\rm kin}, \mathcal{D}_{\rm bsi})$ into a single one-dimensional discriminant.

Because the kinematic properties of off-shell Higgs boson production evolve with $m_{4\ell}$, the merging is performed independently within each of the 21 mass bins. For every $m_{4\ell}$ interval, 40 optimized bins are defined for the combined discriminant. The binning is constructed to maximize the separation among SM signal, background, and interference contributions. This approach achieves comparable performance to the previous analysis while significantly reducing the total number of bins, improving the robustness and stability of the likelihood fit.

As part of the template construction procedure, an averaging strategy is applied to ensure stable and reliable modeling of the multidimensional observable space. Reweighting requires a sufficiently populated target phase space in order to obtain accurate predictions. To illustrate this procedure, off-shell anomalous couplings samples, that are used in reweighting, are shown in Tables 5.6, 5.7, 5.8 and 5.9 [5]. The names of the processes listed in the tables reflect the nature of the off-shell anomalous couplings introduced in Sec. 1.7. For example, the sample `ggTo4e_0PHf05ph0H125Contin` can be interpreted as follows.

The prefix `ggTo4e` denotes the production process $gg \rightarrow H \rightarrow ZZ \rightarrow 4e$. The label `0PH` corresponds to a spin-0 Higgs boson with a CP-even (scalar) interaction structure, associated with the coupling a_1. The modifier `f05` indicates a fractional contribution of

---

[5] All of the samples are collectively referred to as anomalous couplings samples, even tough some of them are pure SM. To keep the reasonable size of the Table, only a subset of samples is shown.

Category	$m_{4\ell}$ bins [GeV]
Untagged	220–230, 230–240, 240–250, 250–260, 260–270, 270–280, 280–290, 290–300, 300–320, 320–340, 340–350, 350–370, 370–390, 390–410, 410–450, 450–490, 490–530, 530–570, 570–670, 670–1000, 1000–13000
VBF-tagged	220–230, 230–240, 240–250, 250–270, 270–280, 280–300, 300–320, 320–340, 340–360, 360–370, 370–390, 390–410, 410–440, 440–460, 460–490, 490–530, 530–610, 610–680, 680–840, 840–1200, 1200–13000
VH-tagged	220–230, 230–240, 240–250, 250–260, 260–270, 270–280, 280–300, 300–320, 320–340, 340–350, 350–370, 370–400, 400–430, 430–460, 460–490, 490–520, 520–560, 560–620, 620–680, 680–990, 990–13000

Table 5.5: $m_{4\ell}$ binning (in GeV) for the three event categories.

the CP-odd component of $f_{a_3} = 0.5$, defined at the cross-section level. The term `ph0` specifies a relative phase $\phi_{a_3} = 0$ between the CP-even (a_1) and CP-odd (a_3) amplitudes. The substring `H125` denotes a Higgs boson mass of 125 GeV, while `Contin` indicates that the sample also contains continuum background.

These samples were generated for the 2016, 2016APV, 2017, and 2018 data-taking periods. The 2016 dataset is split into pre-VFP (2016APV) and post-VFP (2016) eras, corresponding to different detector conditions before and after the resolution of a pixel detector readout timing issue[6].

Index	Process	Period	Events
1	`ggTo4e_0PHH125Contin`	2018	969268
2	`ggTo2e2mu_0MH125Contin_10GaSM`	2016APV	488000
3	`ggTo4e_0PHf05ph0H125Contin`	2016	981664
4	`ggTo4e_0PHf05ph0H125Contin`	2018	984996
5	`ggTo4mu_0PHH125Contin`	2018	384879
6	`ggTo2e2mu_0PL1f05ph0H125`	2017	500000
7	`ggTo2e2mu_0PL1f05ph0H125Contin_10GaSM`	2017	500000
8	`ggTo4e_0PL1H125`	2016	984000
9	`ggTo2e2mu_0Mf05ph0H125Contin`	2016APV	500000
10	`ggTo4e_0PL1f05ph0H125`	2016	993000
	...		
225	`ggTo4mu_0MH125`	2016	524000

Table 5.6: MCFM samples used for reweighting of ggF production. The first column shows the sample index, the second the simulated process, the third the corresponding data-taking period, and the fourth the number of events.

---

[6] VFP (Vertex Finding Problem) refers to a period in early 2016 data-taking during which a pixel detector readout timing issue affected tracking and vertex reconstruction performance.

Index	Process	Period	Events
1	ggH600	2016APV	499999
2	ggH155	2016APV	499994
3	ggH750	2016	500000
4	ggH126	2016APV	999999
5	ggH3000	2016	356000
6	ggH500	2016APV	500000
7	ggH180	2018	408000
8	ggH190	2017	499680
9	ggH900	2018	482000
10	ggH700	2016APV	465996
	...		
152	ggH175	2017	447994

Table 5.7: POWHEG high-mass samples used for reweighting of ggF production. The first column shows the sample index, the second the simulated process, the third the corresponding data-taking period, and the fourth the number of events.

Index	Process	Period	Events
1	VBFToHiggs0PHToZZTo4l_M125_GaSM	2018	1857601
2	VBFToHiggs0Mf075ph0ToZZTo4l_M125_GaSM	2016	1936503
3	VBFToHiggs0MContinToZZTo4l_M125_GaSM	2018	2591336
4	VBFToHiggs0MContinToZZTo4l_M125_GaSM	2017	2965836
5	VBFToHiggs0MContinToZZTo4l_M125_GaSM	2016APV	1498884
6	VBFToHiggs0L1f05ph0ToZZTo4l_M125_GaSM	2016APV	1490125
7	VBFToHiggs0PHToZZTo4l_M125_GaSM	2017	2987904
8	VBFToHiggs0L1ContinToZZTo4l_M125_GaSM	2016APV	1452519
9	VBFToHiggs0L1f05ph0ToZZTo4l_M125_GaSM	2018	2857197
10	VBFToHiggs0L1ContinToZZTo4l_M125_GaSM	2018	2994564
	...		
41	VBFToHiggs0L1ContinToZZTo4l_M125_GaSM	2017	2990915

Table 5.8: MCFM samples used for reweighting of EW production. The first column shows the sample index, the second the simulated process, the third the corresponding data-taking period, and the fourth the number of events.

In principle, a single sample from each table (Tables 5.6–5.9) could be used and reweighted to the target hypothesis. However, this approach would lead to statistical fluctuations and limited phase-space coverage. Instead, a weighted averaging procedure is employed. The weighted average is computed per mass bin within each sample. Rather than weighting an entire sample uniformly, the procedure is applied independently in each mass bin. The weight assigned in the averaging corresponds to the raw number of MC events entering the template from a given sample.

In addition, samples that produce extreme outlying values after reweighting are excluded for the corresponding hypothesis. To maintain consistency with the mass-bin

Index	Process	Period	Events
1	`VBFTo4eJJ_Contin_phantom128`	2016APV	238100
2	`VBFTo2e2muJJ_0PMH125_phantom128`	2018	488989
3	`VBFTo4eJJ_0PMH125_phantom128`	2018	482057
4	`VBFTo2e2muJJ_Contin_phantom128`	2017	370902
5	`VBFTo2e2muJJ_0PMH125_phantom128`	2016	499980
6	`VBFTo4muJJ_0PMH125_phantom128`	2016	472540
7	`VBFTo2e2muJJ_0PMH125_phantom128`	2016APV	175991
8	`VBFTo2e2muJJ_Contin_phantom128`	2016	499855
9	`VBFTo2e2muJJ_0PMH125Contin_phantom128`	2016	498849
10	`VBFTo4muJJ_Contin_phantom128`	2016APV	122970
	...		
36	`VBFTo4eJJ_0PMH125Contin_phantom128`	2016APV	214940

Table 5.9: PHANTOM samples used for reweighting of EW production. The first column shows the sample index, the second the simulated process, the third the corresponding data-taking period, and the fourth the number of events.

weighting, a separate blacklist is constructed for each hypothesis and for each mass bin. The selection is based on the median of the distribution, which is less sensitive to outliers than the mean. Samples are removed if they satisfy

$$I(B) > Q_3 + 1.5 \times \mathrm{IQR} \quad \text{or} \quad I(B) < Q_1 - 1.5 \times \mathrm{IQR}, \tag{5.1}$$

where $I(B)$ is the integral of bin B, Q_1 and Q_3 denote the first and third quartiles, respectively, and the interquartile range is defined as $\mathrm{IQR} = Q_3 - Q_1$, which is a common condition for outlier removal in statistics [110]. An example of this procedure is shown in Fig. 5.4. The target hypothesis corresponds to the interference between the SM ggF signal and background for the 2016 data-taking period in the $2e2\mu$ final state, evaluated in the mass bin $[360, 370)$ GeV. Samples with yields lying outside the green band satisfy the condition defined in Eq. 5.1 and are excluded from the averaging in this hypothesis and mass bin.

After applying all corrections and event selections to the UL MC samples, the final templates used in the statistical analysis are obtained. The expected off-shell event yields for each process and for each event category, together with the observed data yields, are summarized in Table 5.10. The corresponding distributions of the observables entering the template construction are shown in Figure 5.5, where the SM expectation is compared to data in the off-shell region. The four-lepton invariant mass $m_{4\ell}$ and the combined discriminant $D_{\mathrm{bkg}}^{\mathrm{kin}} \times D_{\mathrm{bsi}}$ are displayed for the Untagged, VBF-tagged, and VH-tagged categories. The scenario without off-shell Higgs production is also shown to highlight the effect of destructive interference. This can be seen from Fig. 5.5 as $\mu_{\mathrm{off}} = 0$ consistently has higher yield then a stacked histogram with off-shell Higgs production. Modifications of these distributions induced by SMEFT effects will be discussed in Section 5.5.

As shown in Table 5.10, the dominant contribution in all categories arises from the $q\bar{q} \to ZZ$ background. A strong destructive interference is observed in the off-shell region, with a magnitude exceeding that of the signal contribution. This behavior is a

Figure 5.4: Yields for different anomalous couplings samples, corresponding to the interference between the SM ggF signal and background for the 2016 data-taking period in the $2e2\mu$ final state, evaluated in the mass bin $[360, 370)$ GeV. Samples outside of green band are excluded from the averaging in this hypothesis and mass bin

characteristic feature of off-shell Higgs boson production, where the interference between signal and background amplitudes plays a significant role. The cross-feed contribution remains small in all categories. Overall, the observed data are in good agreement with the SM expectation.

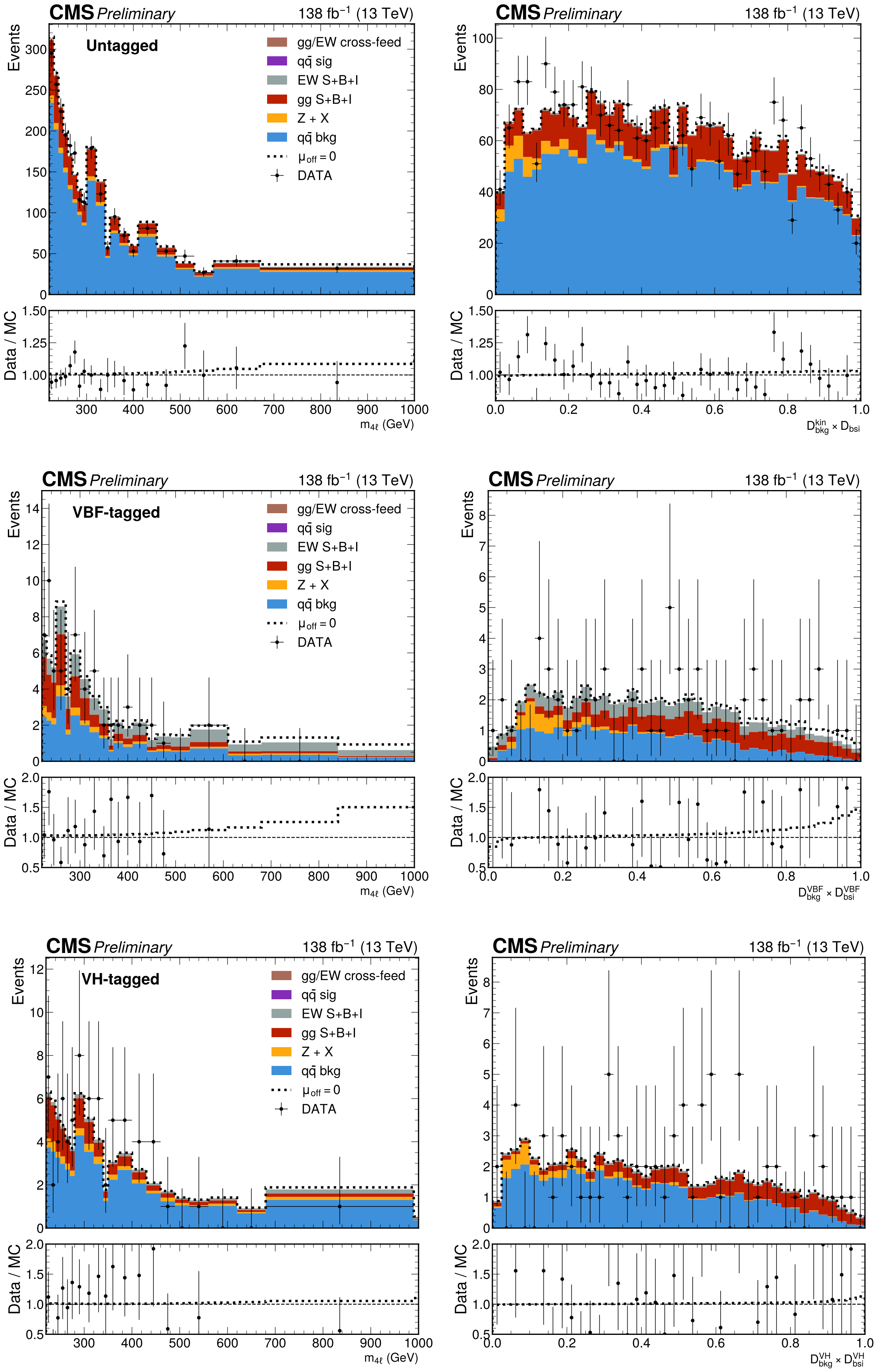


Figure 5.5: Distributions of the four-lepton invariant mass $m_{4\ell}$ up to 1 TeV (left) and the combined discriminant (right) in the Untagged (top), VBF-tagged (middle), and VH-tagged (bottom) categories. The data are shown as points with statistical uncertainties, while the stacked histograms represent the expected SM contributions. The lower panels display the ratio of data to the total prediction. The scenario with no off-shell signal contribution is indicated by the dotted line corresponding to $\mu_{\text{off}} = 0$.

	Untagged	VBF-tagged	VH-tagged	Total
$gg \to ZZ$ background	496.00	18.83	15.29	530.10
$gg \to H \to ZZ$ signal	40.77	1.71	1.46	43.94
ggF interference	-69.60	-2.89	-2.43	-74.92
$q\bar{q} \to ZZ$ background	1895.00	25.90	46.00	1967.00
EW background	18.59	16.64	2.31	37.54
EW signal	2.61	1.94	0.15	4.71
EW interference	-4.69	-4.16	-0.09	-8.95
Z+X background	75.47	4.32	4.50	84.29
Cross-feed	3.54	0.09	0.18	3.81
Total SM expectation	2458.00	62.39	67.37	2588.00
Observed	2420.00	63.00	71.00	2554.00

Table 5.10: Expected and observed SM yields in the off-shell region for the VBF-tagged, VH-tagged, and Untagged categories.

5.5 Higgs Self-coupling Measurement

The measurement of κ_λ in the off-shell region is a central component of this thesis. This section describes the full analysis procedure. It begins with a discussion of the phenomenology of ggF SMEFT corrections that modify κ_λ and their extension to the EW production modes. Since this implementation is used in a CMS analysis for the first time, it requires a dedicated validation. A series of studies is therefore performed to verify the correctness of the implementation.

Once the framework is validated, BSM templates are constructed and the corresponding physics model is defined. Likelihood scans and various fit metrics are then shown. Finally, the statistical interpretation of the results is discussed, with emphasis on the construction of confidence intervals.

5.5.1 SMEFT Corrections in Gluon Fusion

The SMEFT corrections in ggF are computed following Ref. [1], which also served as the basis for extending these corrections to the EW sector. Higgs boson self-coupling deviations with respect to the SM are parameterized by introducing higher-dimensional operators within the framework of the Strongly Interacting Light Higgs (SILH) basis [111]. In this approach, the effective Lagrangian is written as

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_{i,H} \frac{c_i}{v^2} \mathcal{O}_i, \tag{5.2}$$

where c_i are the Wilson coefficients and $\mathcal{O}_i$ denote the corresponding higher-dimensional operators.

Corrections to the LO SM process arising from the insertion of the operator $\mathcal{O}_6$ are considered. The relevant parameters in Eq. 5.2 are the Wilson coefficient c_6, evaluated at the EW scale $\mu = \mathcal{O}(v)$, and the operator $\mathcal{O}_6$, which is built from SM Higgs doublet H:

$$\mathcal{O}_6 = -\lambda |H|^6. \tag{5.3}$$

Calculation of the $\mathcal{O}(\lambda)$ corrections to the SM LO process $gg \to H \to ZZ$ (Figure 5.6) that arise due to insertion of the operator $\mathcal{O}_6$ can be grouped into three categories:

- corrections to ggF Higgs production: Two-loop topologies and the wave-function renormalisation of the Higgs boson field yields renormalised ggF vertex (Figure 5.7)
- corrections to the Higgs propagator: Wave-function renormalisation of the Higgs boson field and corrections to bare Higgs self-energy and the mass counterterm yield renormalised contribution to the self-energy of the Higgs field (Figure 5.8)
- corrections to the Higgs decay: Two-loop contribution and a Higgs wave-function renormalisation yield the renormalised HZZ vertex (Figure 5.9)

Introduction of dimension 6 operator aslo affects the Higgs width, but the effect is negligible in the offshell region and is therefore ommitted.

Based on the calcuclation presented in [1], it can be seen that sensitivity to the parameter c_6 in the process $gg \to H \to ZZ \to 4l$ arises through NLO EW corrections. Besides

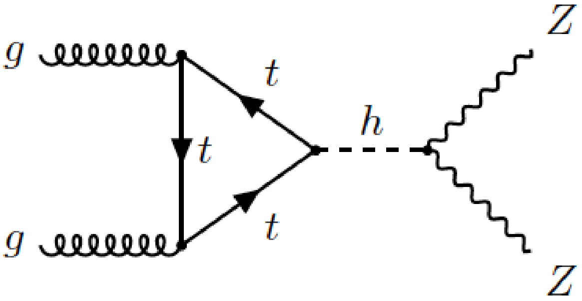


Figure 5.6: SM LO $gg \to H \to ZZ$ [1].

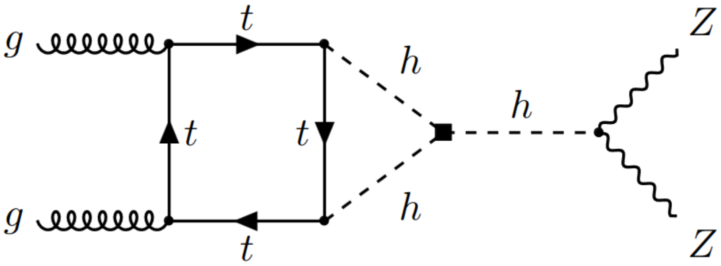


Figure 5.7: $\mathcal{O}(\lambda)$ correction to the $gg \to H \to ZZ$ production [1].

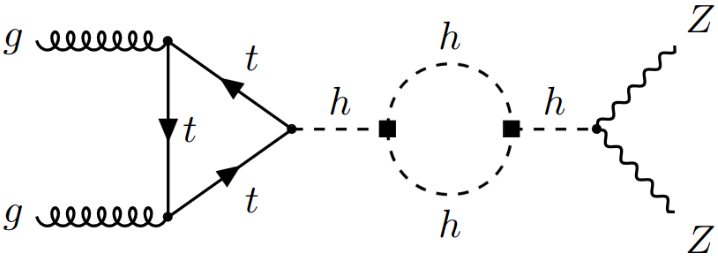


Figure 5.8: $\mathcal{O}(\lambda)$ correction to the $gg \to H \to ZZ$ propagation [1].

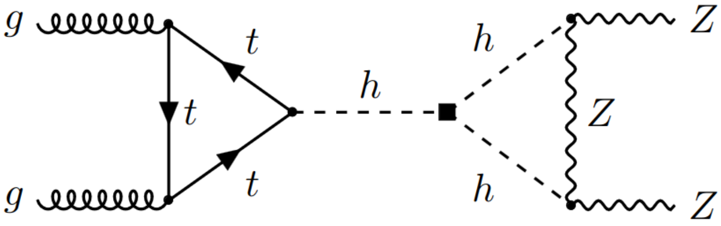


Figure 5.9: $\mathcal{O}(\lambda)$ correction to the $gg \to H \to ZZ$ decay [1].

affecting Higgs production, propagation and decay, $\mathcal{O}_6$ affects the overall size of the cross section and the shape of kinematic distributions with m_{4l} playing a significant role [1].

Effect of including $\mathcal{O}(\lambda)$ corrections to the Feynman diagrams shown in Figures 5.7, 5.8, and 5.9 is introduced. The renormalised vertex that describes the Higgs production is given as:

$$\hat{\Gamma}^{\mu\nu}_{ggh}(p_1,p_2) = -\frac{\alpha_s\delta^{a_1a_2}}{\pi v}\left(\eta^{\mu\nu}p_1\cdot p_2 - p_1^{\nu}p_2^{\mu}\right)\left[\frac{\delta Z_h}{2}\mathcal{F}_1 + \frac{\lambda\bar{c}_6}{(4\pi)^2}\mathcal{F}_2\right], \tag{5.4}$$

where α_s denotes the strong coupling constant, a_1 and a_2 are colour indices and $\eta^{\mu\nu}$ is a Minkowski metric. Additionally,

$$\delta Z_h = N_h\bar{c}_6\left(\bar{c}_6+2\right), \quad N_h = \frac{\lambda}{(4\pi)^2}\left(9-2\sqrt{3}\pi\right) \simeq -1.54\cdot 10^{-3}, \tag{5.5}$$

is the one-loop correction to the Higgs boson wave function associated to insertions of the operator $\mathcal{O}_6$ [77, 112] and

$$\mathcal{F}_1 = \frac{m_t^2}{m_h^2}\left[2-\left(m_h^2-4m_t^2\right)C_0(\hat{s},0,0,m_t^2,m_t^2,m_t^2)\right], \tag{5.6}$$

represents the one-loop triangle diagram with internal top quarks and m_t is the top-quark mass. Numerical result for the non-factorisable two-loop form factor $\mathcal{F}_2$, presented in [113, 114] is used.

The Higgs propagator also receives corrections from insertions of the operator $\mathcal{O}_6$ (Figure 5.8). The resulting renormalised contribution to the self-energy of the Higgs takes the form:

$$\hat{\Sigma}(\hat{s}) = \Sigma(\hat{s}) + (\hat{s}-m_h^2)\delta Z_h - \delta m_h^2, \tag{5.7}$$

where the one-loop corrections to the Higgs wave function have already been given in (5.5), and the bare Higgs self-energy and the mass counterterm in the on-shell scheme are given by the following expressions:

$$\Sigma(\hat{s}) = \frac{\lambda\bar{c}_6}{(4\pi)^2}(\bar{c}_6+2)\,9m_h^2\,B_0(\hat{s},m_h^2,m_h^2), \tag{5.8}$$

$$\delta m_h^2 = \frac{\lambda\bar{c}_6}{(4\pi)^2}(\bar{c}_6+2)\,9m_h^2\,B_0(m_h^2,m_h^2,m_h^2). \tag{5.9}$$

Here, the B_0 functions are two-point Passarino-Veltman scalar integrals defined as in [115].

The full $\mathcal{O}(\lambda)$ correction to the Higgs decay $h \to ZZ$ receives a two-loop contribution (5.9) as well as a counterterm contribution involving Higgs wave-function renormalisation. In the notation of [116], the relevant renormalised vertex describing the $h(p_1+p_2) \to Z(p_1)Z(p_2)$ transition reads:

$$\hat{\Gamma}^{\mu\nu}_{hZZ}(p_1,p_2) = \frac{2m_Z^2}{v}\left[\eta^{\mu\nu}\mathcal{G}_1 + p_1^{\nu}p_2^{\mu}\mathcal{G}_2\right], \tag{5.10}$$

where m_Z denotes the Z-boson mass. The $\mathcal{O}(\lambda)$ corrections to the one-loop form factors $\mathcal{G}_1$ and $\mathcal{G}_2$ are given by:

$$\mathcal{G}_1 = \frac{\delta Z_h}{2} - \frac{\lambda c_6}{(4\pi)^2}\left\{12\left[m_Z^2C_0((p_1+p_2)^2,p_1^2,p_2^2,m_h^2,m_h^2,m_Z^2)\right.\right. \tag{5.11}$$

$$-C_{00}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_Z^2)\Big] + 3B_0((p_1+p_2)^2, m_h^2, m_h^2)\Bigg\}\,,$$

$$\mathcal{G}_2 = \frac{\lambda c_6}{(4\pi)^2}\, 12 \Big[C_1((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_Z^2) + C_{11}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_Z^2) \tag{5.12}$$

$$+C_{12}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_Z^2)\Big]\,,$$

and the tensor coefficients C_1, C_{00} and C_{11} of the three-point Passarino-Veltman integrals are defined as in the publications [115, 116].

5.5.2 Wave-function Renormalization

The SMEFT corrections appearing in Eqs. 5.4, 5.7, and 5.10 include contributions from the Higgs wave-function renormalization (WFR), denoted by δZ_h. These contributions cancel in the full amplitude, a feature that will be exploited in the construction of the physics model. To make this cancellation explicit, it is convenient to rewrite the production, propagator, and decay terms in a form where the SM contributions are factored out.

Production

The SM production vertex can be written as

$$\hat{\Gamma}^{\mu\nu}_{ggh,\mathrm{SM}}(p_1,p_2) = \frac{\alpha_s \delta^{a_1 a_2}}{\pi v}\left(\eta^{\mu\nu}\, p_1\cdot p_2 - p_1^\nu p_2^\mu\right)\mathcal{F}_1. \tag{5.13}$$

The full production amplitude includes contraction with the gluon polarization vectors $\epsilon_{1\mu}$ and $\epsilon_{2\nu}$,

$$\mathcal{M}_{\mathrm{prod}} = \epsilon_{1\mu}\,\epsilon_{2\nu}\,\hat{\Gamma}^{\mu\nu}_{ggh}. \tag{5.14}$$

Factoring out the SM contribution, this becomes

$$\mathcal{M}_{\mathrm{prod}} = \mathcal{M}^{\mathrm{SM}}_{\mathrm{prod}}\left[1 + \frac{\delta Z_h}{2} + \frac{\lambda \bar{c}_6}{(4\pi)^2}\frac{\mathcal{F}_2}{\mathcal{F}_1}\right], \tag{5.15}$$

where $\mathcal{M}^{\mathrm{SM}}_{\mathrm{prod}} = \epsilon_{1\mu}\epsilon_{2\nu}\hat{\Gamma}^{\mu\nu}_{ggh,\mathrm{SM}}$. The WFR contribution enters as $+\delta Z_h/2$.

Propagator

Including the renormalized Higgs self-energy $\hat{\Sigma}(\hat{s})$, the propagator is [7]

$$\Delta_h(\hat{s}) = \frac{i}{\hat{s} - m_h^2 + \hat{\Sigma}(\hat{s})}. \tag{5.16}$$

Substituting the explicit form of $\hat{\Sigma}(\hat{s})$ and expanding to first order in $\hat{\Sigma}(\hat{s})/(\hat{s}-m_h^2)$ yields

$$\Delta_h(\hat{s}) \simeq \frac{i}{\hat{s}-m_h^2}\left[1 - \delta Z_h - \frac{\Sigma(\hat{s}) - \delta m_h^2}{\hat{s}-m_h^2}\right]. \tag{5.17}$$

---

[7] Propagator also contains width-related term. It is, however, omitted since it's contribution in off-shell region is negligible.

Factoring out the SM propagator,

$$\Delta_h(\hat{s}) = \Delta_h^{\mathrm{SM}}(\hat{s}) \left[1 - \delta Z_h - \frac{\Sigma(\hat{s}) - \delta m_h^2}{\hat{s} - m_h^2} \right], \tag{5.18}$$

where $\Delta_h^{\mathrm{SM}}(\hat{s}) = i/(\hat{s} - m_h^2)$. The WFR contribution appears with a factor $-\delta Z_h$.

Decay

The HZZ decay amplitude is obtained by contracting the vertex with the polarization vectors of the final-state bosons, $\epsilon_{1\mu}^*$ and $\epsilon_{2\nu}^*$,

$$\mathcal{M}_{\mathrm{dec}} = \epsilon_{1\mu}^* \, \epsilon_{2\nu}^* \, \hat{\Gamma}_{HZZ}^{\mu\nu}(p_1, p_2). \tag{5.19}$$

The decay vertex 5.10 can be written as

$$\hat{\Gamma}_{HZZ}^{\mu\nu}(p_1, p_2) = \hat{\Gamma}_{HZZ,\mathrm{SM}}^{\mu\nu} \, \mathcal{G}_1 + \frac{2m_Z^2}{v} \, p_1^\nu p_2^\mu \, \mathcal{G}_2, \tag{5.20}$$

where

$$\hat{\Gamma}_{HZZ,\mathrm{SM}}^{\mu\nu} = \frac{2m_Z^2}{v} \, \eta^{\mu\nu}. \tag{5.21}$$

Factoring out the SM contribution, the decay amplitude becomes

$$\mathcal{M}_{\mathrm{dec}} = \mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}} \left[\frac{\delta Z_h}{2} + \mathcal{O}_{\mathrm{dec}}(p_1, p_2) \right], \tag{5.22}$$

where $\mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}} = \epsilon_{1\mu}^* \epsilon_{2\nu}^* \hat{\Gamma}_{HZZ,\mathrm{SM}}^{\mu\nu}$ and $\mathcal{O}_{\mathrm{dec}}(p_1, p_2)$ represents remaining terms needed to recover decay vertex in Eq. 5.10. Again, the WFR contribution enters as $+\delta Z_h/2$.

Cancellation in the full amplitude

The δZ_h contribution arises from three diagrams, shown in Figs. 5.7, 5.8, and 5.9, corresponding to insertions in the production vertex, propagator, and decay vertex, respectively. The total δZ_h contribution to the amplitude can therefore be written as

$$\mathcal{M}^{\delta Z_h} = \mathcal{M}_1^{\delta Z_h} + \mathcal{M}_2^{\delta Z_h} + \mathcal{M}_3^{\delta Z_h}. \tag{5.23}$$

Each contribution can be written as a product of production, propagator, and decay terms. For the first diagram (Fig. 5.7), the correction is applied to the production vertex, while the propagator and decay remain SM:

$$\mathcal{M}_1^{\delta Z_h} = \mathcal{M}_{\mathrm{prod}}^{\delta Z_h} \, \Delta_h^{\mathrm{SM}} \, \mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}}. \tag{5.24}$$

Using Eq. 5.15, this becomes

$$\mathcal{M}_1^{\delta Z_h} = \left(\frac{\delta Z_h}{2} \, \mathcal{M}_{\mathrm{prod}}^{\mathrm{SM}} \right) \Delta_h^{\mathrm{SM}} \mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}} = \frac{\delta Z_h}{2} \, \mathcal{M}^{\mathrm{SM}}. \tag{5.25}$$

where $\mathcal{M}^{\mathrm{SM}} = \mathcal{M}_{\mathrm{prod}}^{\mathrm{SM}} \Delta_h^{\mathrm{SM}} \mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}}$

For the second diagram (Fig. 5.8), the correction is applied to the propagator:

$$\mathcal{M}_2^{\delta Z_h} = \mathcal{M}_{\mathrm{prod}}^{\mathrm{SM}} \, \Delta_h^{\delta Z_h} \, \mathcal{M}_{\mathrm{dec}}^{\mathrm{SM}}. \tag{5.26}$$

Using Eq. 5.18, one obtains

$$\mathcal{M}_2^{\delta Z_h} = \mathcal{M}_{\text{prod}}^{\text{SM}} \left(-\delta Z_h \, \Delta_h^{\text{SM}}\right) \mathcal{M}_{\text{dec}}^{\text{SM}} = -\delta Z_h \, \mathcal{M}^{\text{SM}}. \tag{5.27}$$

For the third diagram (Fig. 5.9), the correction is applied to the decay vertex:

$$\mathcal{M}_3^{\delta Z_h} = \mathcal{M}_{\text{prod}}^{\text{SM}} \, \Delta_h^{\text{SM}} \, \mathcal{M}_{\text{dec}}^{\delta Z_h}. \tag{5.28}$$

Using Eq. 5.22, this becomes

$$\mathcal{M}_3^{\delta Z_h} = \mathcal{M}_{\text{prod}}^{\text{SM}} \Delta_h^{\text{SM}} \left(\frac{\delta Z_h}{2} \, \mathcal{M}_{\text{dec}}^{\text{SM}}\right) = \frac{\delta Z_h}{2} \, \mathcal{M}^{\text{SM}}. \tag{5.29}$$

Summing all contributions, one obtains

$$\mathcal{M}^{\delta Z_h} = \left(\frac{\delta Z_h}{2} - \delta Z_h + \frac{\delta Z_h}{2}\right) \mathcal{M}^{\text{SM}} = 0. \tag{5.30}$$

Thus, δZ_h contributions cancel out at amplitude level.

5.5.3 MC Implementation

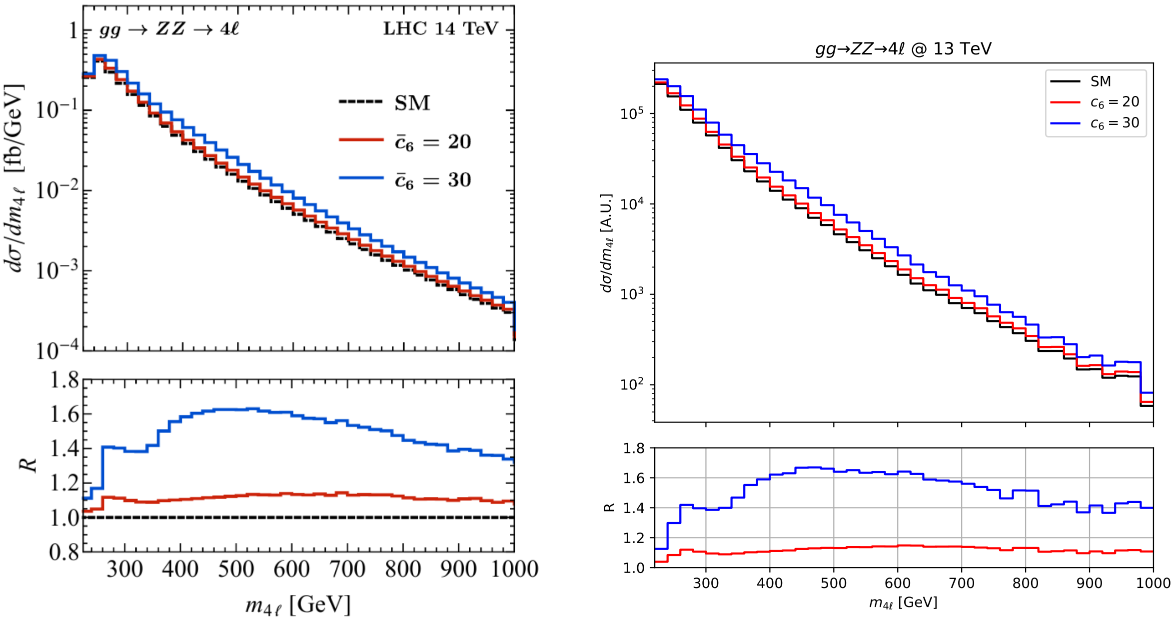
MC implementation [1] relies on the MCFM 8.0 code [117], where the c_6 corrections to the Higgs production, propagation and decay were incorporated for the process $gg \to H \to ZZ$. At the matrix element level, the implementation includes all contributions up to $\mathcal{O}(\lambda^2)$, which includes both the BSM and LO SM $gg \to H \to ZZ$ Feynman diagram. Based on the modified version 8.0 of MCFM, amplitudes were first propagated to MCFM version 7.0, which is integrated into the JHUGEN framework [47]. Subsequently, the amplitudes were transferred to the MELA package [95, 96, 97, 98, 47], enabling event reweighting. Since the measurement of the κ_λ parameter, including ggF SMEFT corrections based on [1], is performed for the first time, a series of validation tests is carried out to ensure the correctness and stability of the implementation.

Comparison to reported MC predictions

As an initial validation step, the invariant mass distributions for different c_6 hypotheses are compared to the corresponding results reported in [1]. The SM $gg \to ZZ$ (including Higgs boson) sample is reweighted using MELA to obtain predictions for $c_6 = 20$ and $c_6 = 30$. These reweighted distributions are then compared to the reference predictions, applying the same GEN-level event selection summarized in Table 5.11.

Figure 5.10 compares the $m_{4\ell}$ distributions from [1], shown in panel (a), with those obtained using MELA reweighting, shown in panel (b). Overall, good agreement is observed, especially in the lower panels, which display the ratio of the c_6 hypotheses to the SM prediction.

A minor difference in the setup originates from the choice of centre-of-mass energy: the reference results in [1] are obtained at $\sqrt{s} = 14$ TeV, while the reweighted distributions are produced at $\sqrt{s} = 13$ TeV, consistent with the Run 2 dataset used in the analysis. However, in the ratio of BSM to SM contributions, a significant cancellation of energy-dependent effects is expected. The observed agreement in these ratios therefore provides a robust validation of the reweighting procedure.



(a) (b)

Figure 5.10: GEN-level $m_{4\ell}$ distributions for the process $gg \to ZZ$ (including the Higgs boson) reported in [1] at $\sqrt{s} = 14$ TeV (a), normalized to the cross section, compared to the corresponding distributions obtained using MELA reweighting at $\sqrt{s} = 13$ TeV (b), shown in arbitrary units (A.U.). The distributions correspond to the SM prediction (black), and the BSM hypotheses $c_6 = 20$ (red) and $c_6 = 30$ (blue). The lower panels show the ratio of the BSM predictions to the SM distribution.

Table 5.11: GEN-level cuts used for event selection.

Selection criteria
$p_{T,l_1} > 20$ GeV
$p_{T,l_2} > 15$ GeV
$p_{T,l_3} > 10$ GeV
$p_{T,l_4} > 6$ GeV
$\|\eta_l\| < 2.5$
220 GeV $< m_{4\ell} <$ 1000 GeV
50 GeV $< m_{Z_1} <$ 106 GeV
50 GeV $< m_{Z_2} <$ 115 GeV

Cross section comparison between MCFM8.0 and JHUGen-MCFM7.0

The second test compares cross sections obtained with MCFM 8.0 and JHUGEN-MCFM 7.0 to verify that the BSM amplitudes are correctly implemented in the generator used for the analysis. Each contribution to the BSM matrix element can be accessed and tested independently by controlling the set of coefficients $\{t_1, t_2, t_3, t_4, t_5, t_6, t_7, w_1, w_2, w_3, w_4, w_5\}$. The coefficients $\{t_4, t_5, w_3\}$ modify the production part of the matrix element, $\{t_1, w_1\}$ correspond to the propagation term, and $\{t_2, t_3, w_2\}$ affect the decay contribution. The remaining coefficients $\{t_6, t_7, w_4, w_5\}$ control the Higgs boson width. Setting the coefficients within a chosen subset to 1 (0) switches the corresponding BSM contribution on (off). For example, choosing $t_1 = 1$ and $w_1 = 1$ while keeping all other coefficients equal to zero activates only the BSM effects in the propagation term of the matrix element. After implementing the BSM matrix elements in JHUGEN–MCFM, each BSM contribution was enabled individually and the corresponding cross section was computed for the process $gg \to H_{\text{SM+BSM}} \to ZZ \to 2e2\mu$. The results of these calculations are presented in Table 5.12, where a comparison between the predictions of MCFM 8.0 and JHUGEN–MCFM is shown. In these calculations the Higgs boson width was fixed to its SM value, corresponding to setting the coefficients $\{t_6, t_7, w_4, w_5\}$ to zero, while the SMEFT parameter c_6 was set to 15.

A good agreement is observed between the cross section values obtained with the two generators for each individual BSM contribution, as well as for the SM prediction. Small discrepancies are present. These differences do not originate from different matrix elements used in the calculations, but rather reflect the sensitivity of the predicted cross sections to the choice of PDF sets, renormalisation and factorisation schemes, compilers, and other technical parameters. Based on the level of agreement observed for all BSM contributions, it can be concluded that the BSM matrix elements have been correctly propagated to the JHUGEN–MCFM generator, allowing their use in further studies.

5.5.4 SMEFT Corrections in the EW Production

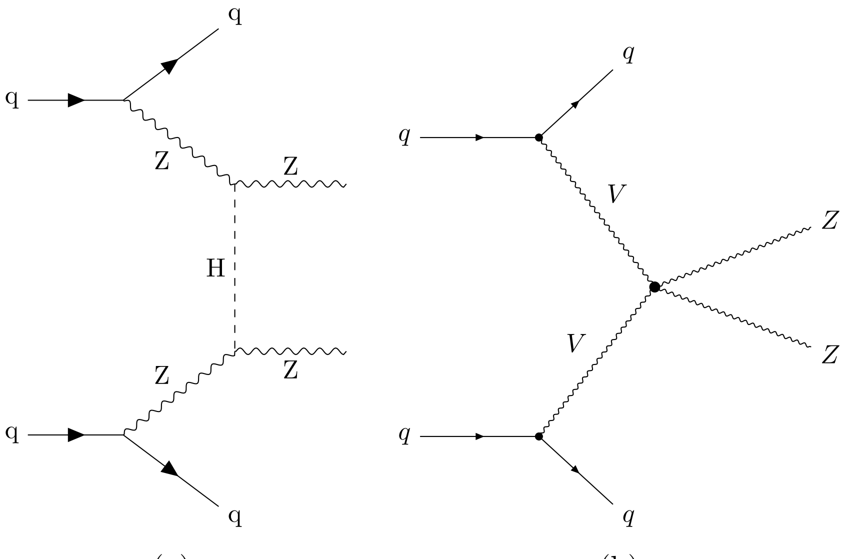
While the SMEFT corrections from [1] affect the dominant ggF production mode, their extension to EW Higgs production has been implemented as part of this thesis. Among EW processes, the LO VBF s-channel and LO VH production modes are illustrated in Figs. 1.3 and 1.4, respectively. In addition to the s-channel, VBF production also happens via t- and u-channel diagrams, with the t-channel topology shown in Fig. 5.11(a). The

MCFM 8.0	
Contribution	σ [fb]
Production	$0.703822 \cdot 10^{-1} \pm 0.87191 \cdot 10^{-5}$
Propagation	$0.177317 \pm 0.23768 \cdot 10^{-4}$
Decay	$0.713107 \cdot 10^{-1} \pm 0.91856 \cdot 10^{-5}$
All contributions	$0.869220 \cdot 10^{-1} \pm 0.11098 \cdot 10^{-4}$

JHUGEN-MCFM	
Contribution	σ [fb]
Production	$0.714484 \cdot 10^{-1} \pm 0.816460 \cdot 10^{-5}$
Propagation	$0.179944 \pm 0.208984 \cdot 10^{-4}$
Decay	$0.723839 \cdot 10^{-1} \pm 0.829225 \cdot 10^{-5}$
All contributions	$0.882184 \cdot 10^{-1} \pm 0.119974 \cdot 10^{-4}$

Table 5.12: Contributions to the ggF cross section for each BSM component of the matrix element calculated with MCFM 8.0 and JHUGEN-MCFM. Cross section was computed for value $c_6 = 15$. Mass range used in comparison is $130 < m_{4\ell} < 1000$ GeV. Width BSM contributions are set to 0.

background that interferes with VBF production mode arises from VBS, illustrated in Fig. 5.11(b).

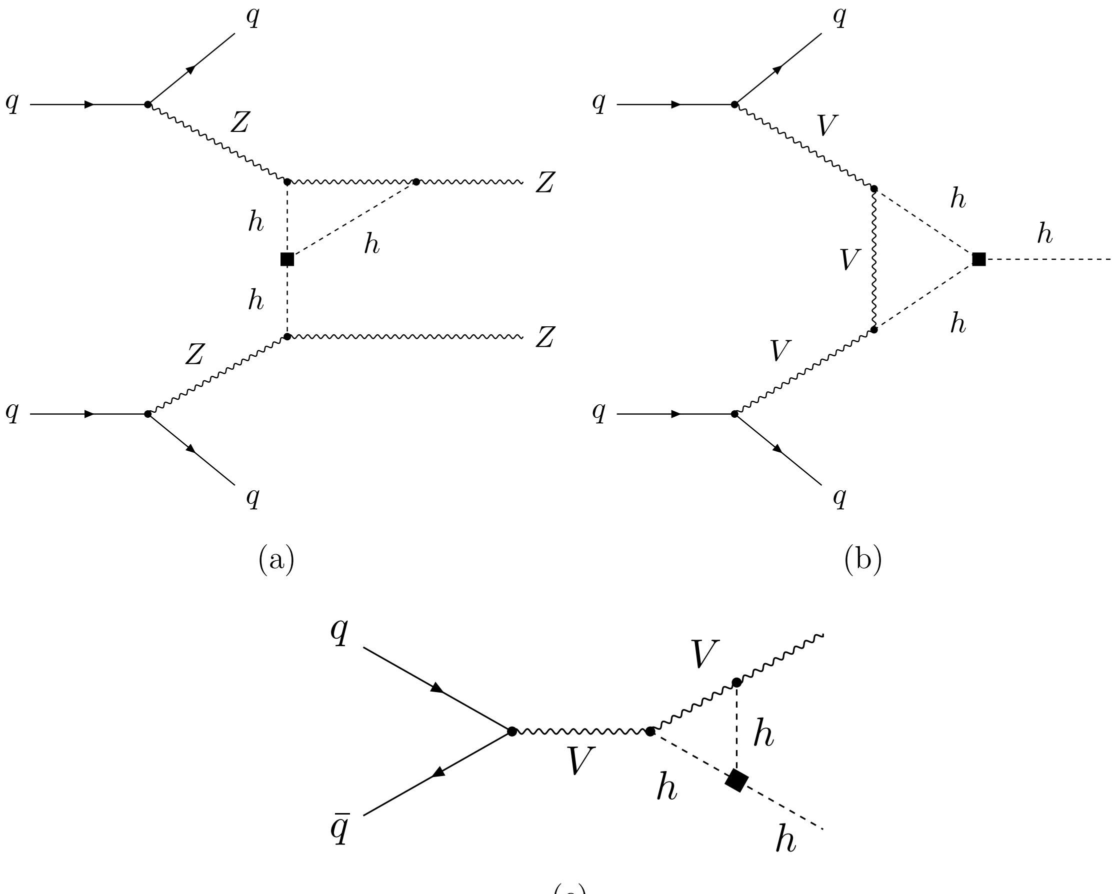


(a) (b)

Figure 5.11: (a) LO VBF t-channel Feynman diagram. (b) Vector boson scattering (VBS) background.

SMEFT corrections, that have been applied atop of the LO EW production are depicted on Fig. 5.12.

Inspecting Fig. 5.12 and comparing it to the decay correction shown in Fig. 5.9, one observes that they share a similar structure, namely a triangle diagram involving two Higgs bosons and one vector boson. This structural similarity can be exploited to derive



(a) (b)

(c)

Figure 5.12: $\mathcal{O}_6$ corrections to VBF t-channel (a), VBF s-channel (b), and VH production (c).

the EW production correction. The expression in Eq. 5.10 can be used directly, with the only modification being that the vector boson is not restricted to Z, but can also be W:

$$\hat{\Gamma}^{\mu\nu}(p_1, p_2) = \frac{2m_{W/Z}^2}{v}\left[\eta^{\mu\nu}\mathcal{G}_1 + p_1^\nu p_2^\mu \mathcal{G}_2\right] \tag{5.31}$$

with the coefficients $\mathcal{G}_1$ and $\mathcal{G}_2$ are modified as follows:

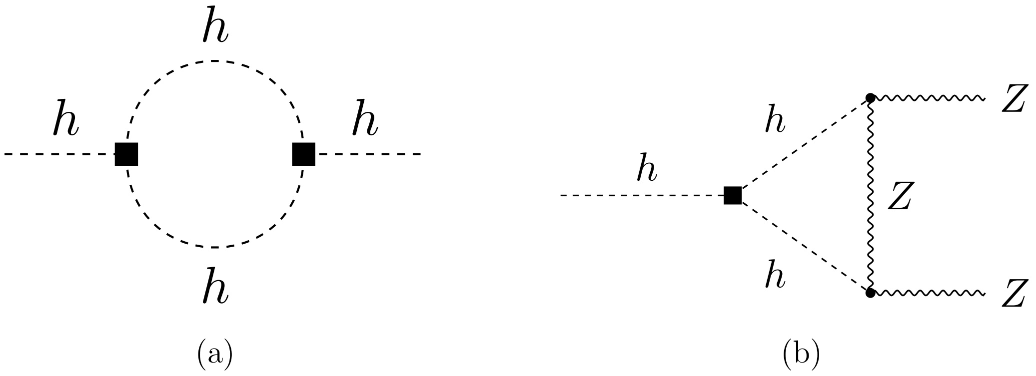
$$\mathcal{G}_1 = \frac{\delta Z_h}{2} - \frac{\lambda c_6}{(4\pi)^2}\left\{12\left[m_{Z/W}^2 C_0((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_{Z/W}^2)\right.\right. \tag{5.32}$$

$$\left.\left.-C_{00}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_{Z/W}^2)\right] + 3B_0((p_1+p_2)^2, m_h^2, m_h^2)\right\},$$

$$\mathcal{G}_2 = \frac{\lambda c_6}{(4\pi)^2}\, 12\left[C_1((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_{Z/W}^2) + C_{11}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_{Z/W}^2)\right. \tag{5.33}$$

$$\left.+C_{12}((p_1+p_2)^2, p_1^2, p_2^2, m_h^2, m_h^2, m_{Z/W}^2)\right].$$

Corresponding propagator and decay corrections, which have been isolated on Fig. 5.13, are identical to the ones from 5.8 and 5.9, and therefore follow Eqs. 5.7 and 5.10.



(a) (b)

Figure 5.13: $\mathcal{O}_6$ corrections to propagator (a) and decay (b).

Following the same reasoning as in Sec. 5.5.2, WFR terms cancel out in EW aswell, since production, propagation and decay contribute as $\frac{\delta Z_h}{2}$, $-\delta Z_h$ and $\frac{\delta Z_h}{2}$ times SM amplitude, respectively.

The EW process has been implemented in MCFM and MELA such that individual contributions are controlled via parameters t_i, w_i, as in the case of gg fusion. Their effect is summarized in Table 5.13

5.5.5 Physics Model

This section introduces the κ_λ parametrization of the templates, incorporating corrections to the SM LO ggF and EW process arising from production, propagation, and decay effects. The model is based on Eqs. 5.4, 5.31, 5.7 and 5.10.

Sector	Component	Coefficients	Symbolic form	Original Eq.
ggF	Production	t_4, t_5, w_3	$(t_4F_1)(w_3\delta Z_H) + t_5F_2$	Eq. 5.4
	Propagation	t_1, w_1	$t_1\big(f(c_6) + w_1\delta Z_H\big)$	Eq. 5.7
	Decay	t_2, t_3, w_2	$t_2G_1\big(f(c_6) + w_2\delta Z_H\big) + t_3G_2$	Eq. 5.10
EW	Production	t_4, t_5, w_2	$t_4G_1\big(f(c_6) + w_2\delta Z_H\big) + t_5G_2$	Eq. 5.31
	Propagation	t_1, w_1	$t_1\big(f(c_6) + w_1\delta Z_H\big)$	Eq. 5.7
	Decay	t_2, t_3, w_2	$t_2G_1\big(f(c_6) + w_2\delta Z_H\big) + t_3G_2$	Eq. 5.10
Both (ggF, EW)	Width	w_4, w_5, t_6, t_7	$\Gamma_H(w_4, w_5, t_6, t_7)$	omitted

Table 5.13: Summary of SMEFT coefficients controlling production, propagation, decay, and width contributions in gg and EW Higgs production. The symbolic forms indicate how each set of coefficients enters the corresponding matrix element components, while the final column is reserved for references to the original equations.

In the MCFM implementation, the c_6 coefficient modifies only the correction diagrams. This allows identification of $c_6 \approx \kappa_\lambda$. At the amplitude level, ggF and EW share many features. After δZ_h cancellation, they both have production and decay part linear in κ_λ and propagator proportional to $\kappa_\lambda(\kappa_\lambda+2)$. This is exploited when constructing physics model. Production and decay corrections are merged into one linear term, while propagator is kept as a single term with it's scaling as a 2nd order polynomial.

The total amplitude includes four components: SM LO signal, interfering background, a combined production and decay correction term (scaling linearly with κ_λ), and the propagation correction [8]. The full amplitude can be expressed as:

$$\begin{aligned}
A^2 &= |A_{SM_{LO}} + A_{BKG} + A^{tot}_{prod} + A^{tot}_{prop} + A^{tot}_{dec}|^2 = \\
&= |A_{SM_{LO}} + A_{BKG} + A_{prod} + A^{\delta Z_H}_{prod} + A_{prop} + A^{\delta Z_H}_{prop} + A_{dec} + A^{\delta Z_H}_{dec}|^2 = \\
&= |A_{SM_{LO}} + A_{BKG} + (A_{prod} + A_{dec})(\kappa_\lambda) + A_{prop}(\kappa_\lambda^2 + 2\kappa_\lambda)|^2 = \\
&= A^2_{SM_{LO}} + A^2_{BKG} + (A_{prod} + A_{dec})^2 + A^2_{prop} + I(A_{SM_{LO}}, A_{BKG}) + I(A_{SM_{LO}}, A_{prod} + A_{dec}) + \\
&\quad + I(A_{SM_{LO}}, A_{prop}) + I(A_{BKG}, A_{prod} + A_{dec}) + I(A_{BKG}, A_{prop}) + I(A_{prod} + A_{dec}, A_{prop})
\end{aligned} \tag{5.34}$$

Expression 5.34 consists of 10 terms. Of these, 3 are independent of κ_λ, namely $A_{SM_{LO}}$, A_{BKG}, and $I(A_{SM_{LO}}, A_{BKG})$, which correspond to the SM LO signal, the background, and their interference, respectively. The remaining 7 terms depend on κ_λ, either through interference effects or directly via amplitudes induced by SMEFT corrections. These include interference between SMEFT corrections aswell. For example, the term $I(A_{prod} + A_{dec}, A_{prop}$ corresponds to the interference between linear parts of merged production and decay and propagator correction with δZ_h contribution removed. Since $A_{prod} + A_{dec}$ scale as κ_λ and A_{prop} scales as $\kappa_\lambda(\kappa_\lambda + 2)$ their interference scales as $\kappa^2_\lambda(\kappa_\lambda + 2)$.

---

[8] In the analysis, there are 2 SMEFT amplitude expressions. The first one is for ggF, where $A_{SM_{LO}}$ and A_{BKG} refer to LO $gg \to H$ signal and $gg \to ZZ$ background. In EW case, $A_{SM_{LO}}$ refers to LO contribution from VBF and VH while A_{BKG} refers to VBS

5.5.6 Scaling Verification

Using MELA, all off-shell anomalous coupling (AC) samples are reweighted to these ten components, across a set of κ_λ values for both the ggF and EW production modes. For each value of κ_λ and for each component in Eq. 5.34, a yield is computed and normalized with the corresponding contribution evaluated at $\kappa_\lambda = 1$. Scaling tests are conducted in order to verify that SMEFT amplitudes correctly scale according to Eqs. 5.4, 5.31, 5.7 and 5.10. As an example, the ggF propagation term is considered first. Since the propagation amplitude scales quadratically with κ_λ, the yield scales as a fourth-order polynomial. Figure 5.14 shows a fit to the function $f(\kappa_\lambda) = a + b\kappa_\lambda + c\kappa_\lambda^2 + d\kappa_\lambda^3 + e\kappa_\lambda^4$, with coefficients $a \approx 0$, $b \approx 0$, $c \approx 0.44$, $d \approx 0.44$, and $e \approx 0.11$.

The theoretical prediction for the propagation yield (with δZ_h cancellation) is $[\kappa_\lambda(\kappa_\lambda + 2)]^2 = \kappa_\lambda^4 + 4\kappa_\lambda^3 + 4\kappa_\lambda^2$. Normalized to the SM case ($\kappa_\lambda = 1$), the scaling becomes:

$$\frac{Yield(\kappa_\lambda)}{Yield(\kappa_\lambda = 1)} = \frac{\kappa_\lambda^4 + 4\kappa_\lambda^3 + 4\kappa_\lambda^2}{1 + 4 + 4} = 0.11\,\kappa_\lambda^4 + 0.44\,\kappa_\lambda^3 + 0.44\,\kappa_\lambda^2 \tag{5.35}$$

This matches the fit results and provides a strong validation of both the c_6 MELA implementation and the overall framework.

As a further example, the interference between the linear (production + decay) and propagation terms in ggF is analyzed. The theoretical scaling is given by $I(\text{prod+dec}, \text{prop}) \sim \kappa_\lambda^2(\kappa_\lambda + 2) = \kappa_\lambda^3 + 2\kappa_\lambda^2$, which leads to the normalized form:

$$\frac{Yield(\kappa_\lambda)}{Yield(\kappa_\lambda = 1)} = \frac{\kappa_\lambda^3 + 2\kappa_\lambda^2}{1 + 2} = 0.33\,\kappa_\lambda^3 + 0.66\,\kappa_\lambda^2 \tag{5.36}$$

The fit shown in Figure 5.15 confirms this behavior.

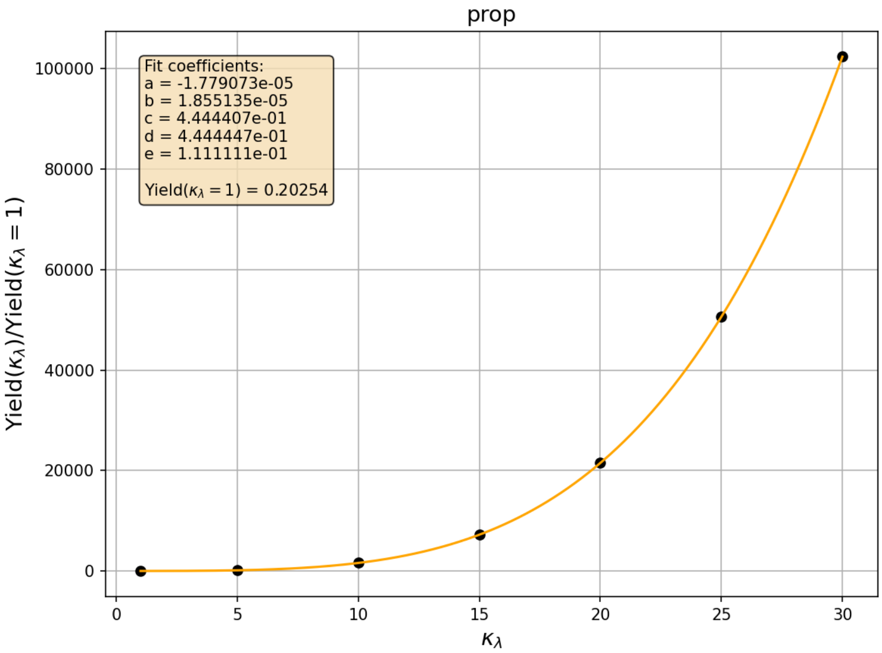


Figure 5.14: Propagation correction, κ_λ parametrization

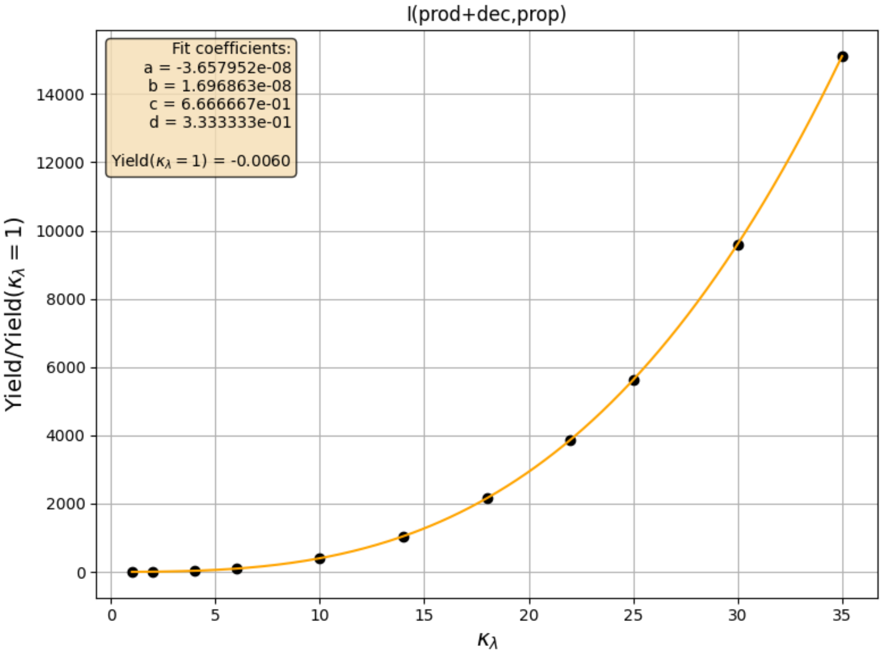
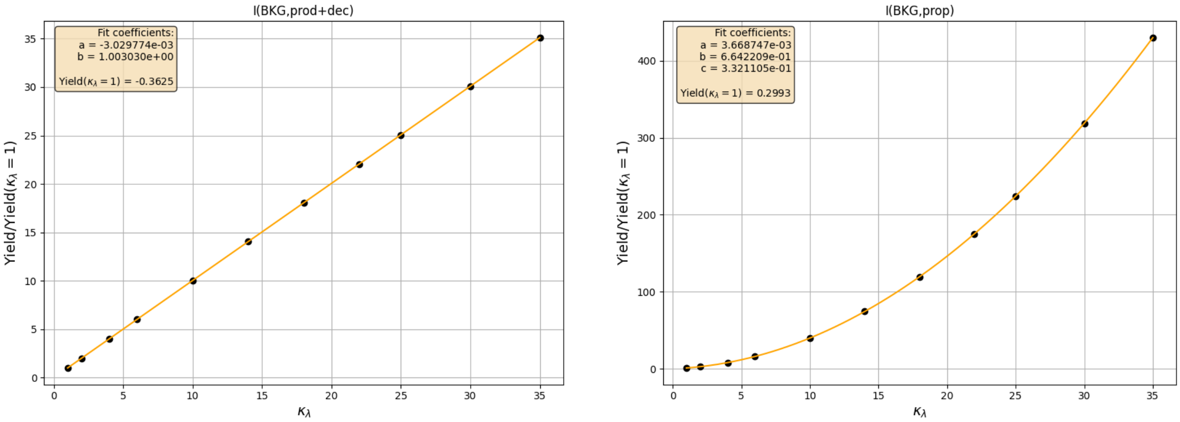


Figure 5.15: Interference between production + decay correction and propagation correction, κ_λ parametrization

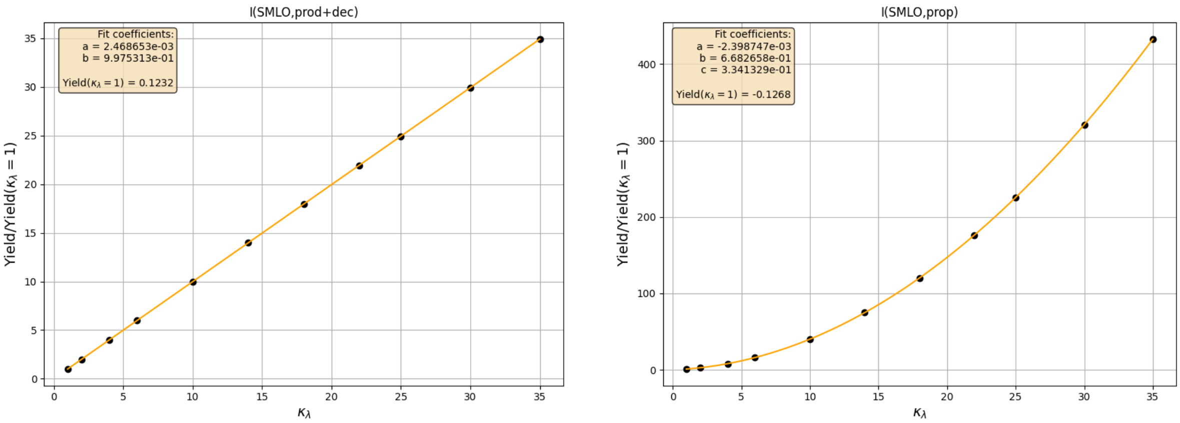
All 14 κ_λ-depending components (7 for ggF and 7 for EW) have been subjected to this test. Corresponding scalings are displayed in Figures 5.16–5.21, where fitted coefficients are consistent with theoretical predictions within numerical precision for all of the components.



(a) (b)

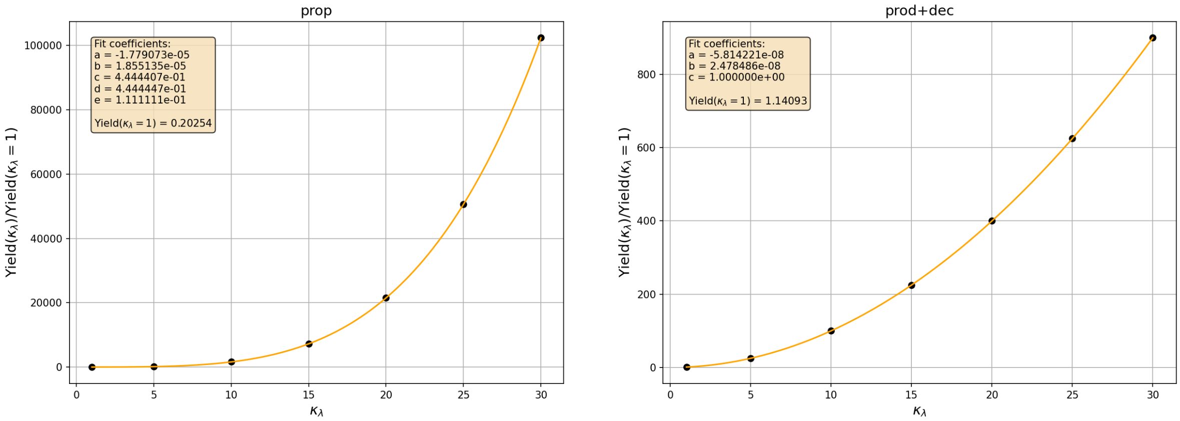
Figure 5.16: ggF: Polynomial fit to the yield ratio evaluated at different κ_λ values for (a) interference between background and production+decay correction, and (b) interference between background and propagation correction.

As all scalings are verified, it is sufficient to reweight all samples to $\kappa_\lambda = 1$ and then apply analytical parametrizations to construct the physics model. This approach reduces computational cost by eliminating the need for reweighting to multiple κ_λ values.



(a) (b)

Figure 5.17: ggF: Polynomial fit to the yield ratio evaluated at different κ_λ values for (a) interference between SM LO signal and production+decay correction, and (b) interference between SM LO signal and propagation correction.



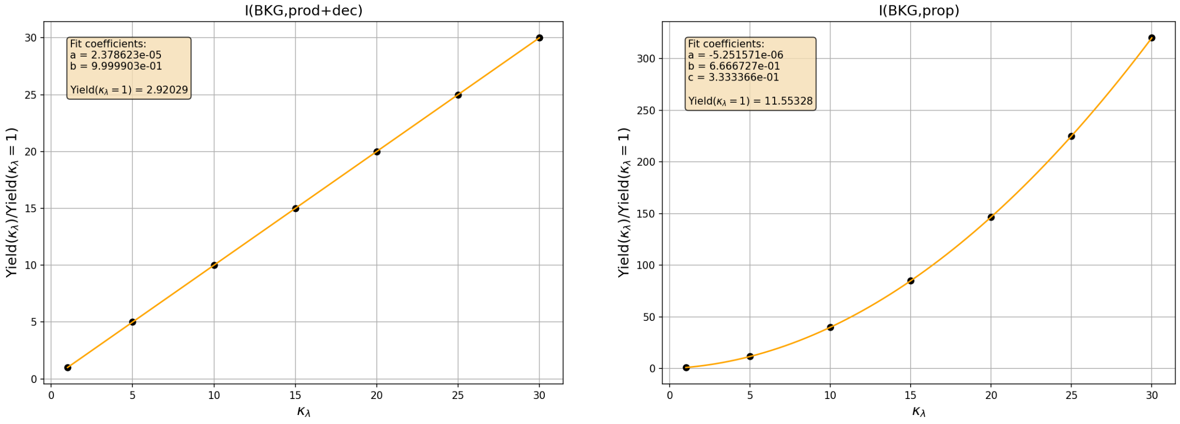
(a) (b)

Figure 5.18: EW: Polynomial fit to the yield ratio evaluated at different κ_λ values for (a) propagation correction, and (b) production+decay correction.

All scalings wrt. to $\kappa_\lambda = 1$ are summarized in Table 5.14.

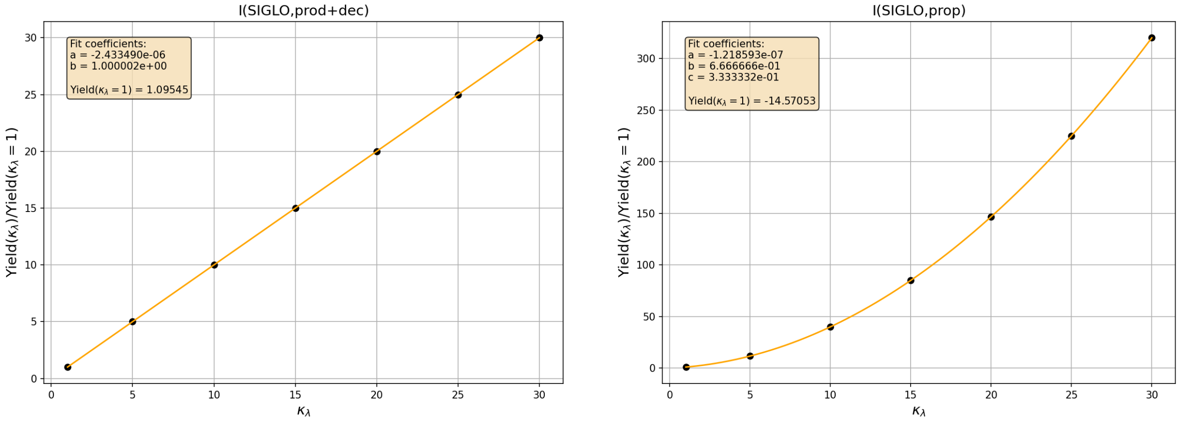
Template	$\boldsymbol{f(\kappa_\lambda)}$
$A^2_{SM_{LO}}$, A^2_{BKG}, $I(A_{SM_{LO}}, A_{BKG})$	1
$I(A_{SM_{LO}}, A_{prod} + A_{dec})$, $I(A_{BKG}, A_{prod} + A_{dec})$	κ_λ
$(A_{prod} + A_{dec})^2$	κ_λ^2
$I(A_{SM_{LO}}, A_{prop})$, $I(A_{BKG}, A_{prop})$	$\frac{1}{3}\kappa_\lambda^2 + \frac{2}{3}\kappa_\lambda$
$I(A_{prod} + A_{dec}, A_{prop})$	$\frac{1}{3}\kappa_\lambda^3 + \frac{2}{3}\kappa_\lambda^2$
A^2_{prop}	$\frac{1}{9}\kappa_\lambda^4 + \frac{4}{9}\kappa_\lambda^3 + \frac{4}{9}\kappa_\lambda^2$

Table 5.14:
Functional forms of the κ_λ parametrization for different templates.



(a) (b)

Figure 5.19: EW: Polynomial fit to the yield ratio evaluated at different κ_λ values for (a) interference between background and production+decay correction, and (b) interference between background and propagation correction.



(a) (b)

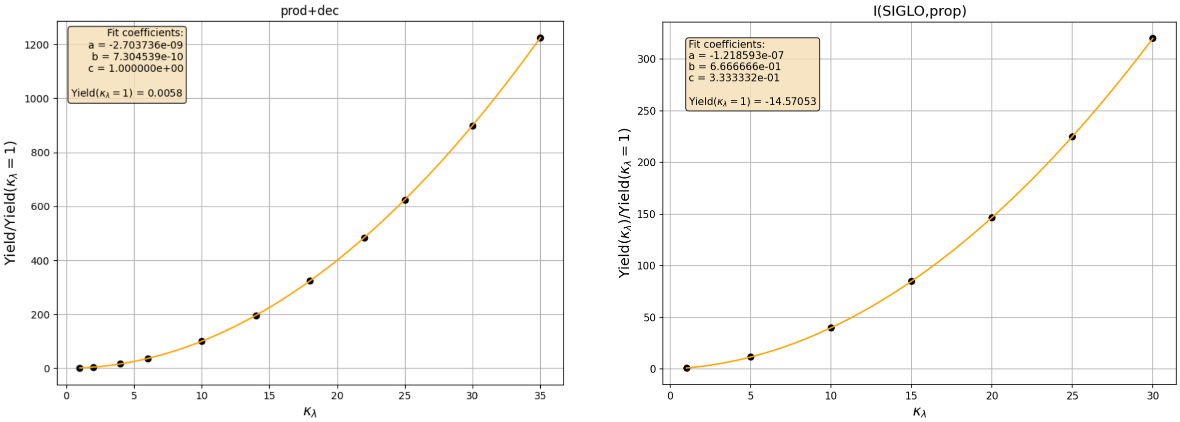
Figure 5.20: EW: Polynomial fit to the yield ratio evaluated at different κ_λ values for (a) interference between SM LO signal and production+decay correction, and (b) interference between SM LO signal and propagation correction.

5.5.7 Closure Test

After scaling verification, further validation tests are conducted to ensure correctness of the model. The first such test is the yield closure test, where an alternative, but yield-equivalent physics model is created. In this alternative model, all signal-related terms are grouped into a single template. The total amplitude is expressed as

$$A^2 = \left| A_{SM_{LO}} + A_{BKG} + A^{tot}_{prod} + A^{tot}_{prop} + A^{tot}_{dec} \right|^2 = \left| A_{SM_{LO}+\mathcal{O}(\lambda)} + A_{BKG} \right|^2 \tag{5.37}$$

Expanding this expression, the alternative model consists of three terms: $A^2_{SM_{LO}+\mathcal{O}(\lambda)}$, the interference term $I(A_{SM_{LO}+\mathcal{O}(\lambda)}, A_{BKG})$, and A^2_{BKG}. Reweighting is performed to



(a) ggF (b) EW

Figure 5.21: Polynomial fit to the yield ratio evaluated at different κ_λ values for the remaining contributions: (a) production+decay correction in ggF, and (b) interference between SM LO signal and propagation correction in EW production.

match the processes used in the "3-term model", and the resulting yields are compared with those of the "10-term model", which is used in the main analysis.

It is important to note that δZ_h terms are retained in the alternative model. Since this comparison is only needed as a consistency check, a single sample is sufficient for testing. For ggF the chose sample is `MC_2018/OffshellAC/gg/ggTo2e2mu_0MH125_MCFM701`, reweighted to $c_6 = 14$.

Process	**Yield**
BKG	105.62
SM LO	8.67
prod+dec	0.25
prop	1.98
I(BKG, SM LO)	-14.79
I(BKG, prod+dec)	-1.10
I(BKG, prop)	4.85
I(SM LO, prod+dec)	0.37
I(SM LO, prop)	-2.06
I(prod+dec, prop)	-1.36
Total	**102.43**

Table 5.15: Yields for 10-term ggF model

Process	**Yield**
SM LO + correction	7.85
BKG	105.62
Interference	-11.04
Total	**102.43**

Table 5.16: Yields for 3-term model ggF

As shown in Tables 5.15 and 5.16, both models yield the same total value. Closure test is also performed for EW production. Here, the yield from the 10-term model is compared to the inclusive model, where direct reweighting is applied using the combination of all terms. This test is carried out on the sample `MC_2018/OffshellAC/VBF/VBFToHiggs0PMToZZTo4l_M125_GaSM` for a benchmark value of $c_6 = 20$ and yields are computed in arbitrary units, keeping conistent scale between different terms.

As shown in Tables 5.17 and 5.18, EW yields are consistent between different models.

Process	**Yield**
BKG	14.58
SM LO	1.75
prod+dec	0.16
prop	1.56
I(BKG, SM LO)	-3.35
I(BKG, prod+dec)	0.02
I(BKG, prop)	0.61
I(SM LO, prod+dec)	0.01
I(SM LO, prop)	-0.76
I(prod+dec, prop)	-0.95
Total	**13.62**

Table 5.17: Yields for 10-term EW model

Process	**Yield**
Inclusive	**13.62**

Table 5.18: Yields for inclusive EW model

This validation serves two purposes: (1) it confirms that the δZ_H contributions cancel out, since the yields remain unchanged regardless of their presence, and (2) it demonstrates that reweighting is robust to different groupings of amplitude terms as the final yield is unaffected by how the components are organized.

5.5.8 Shape Tests

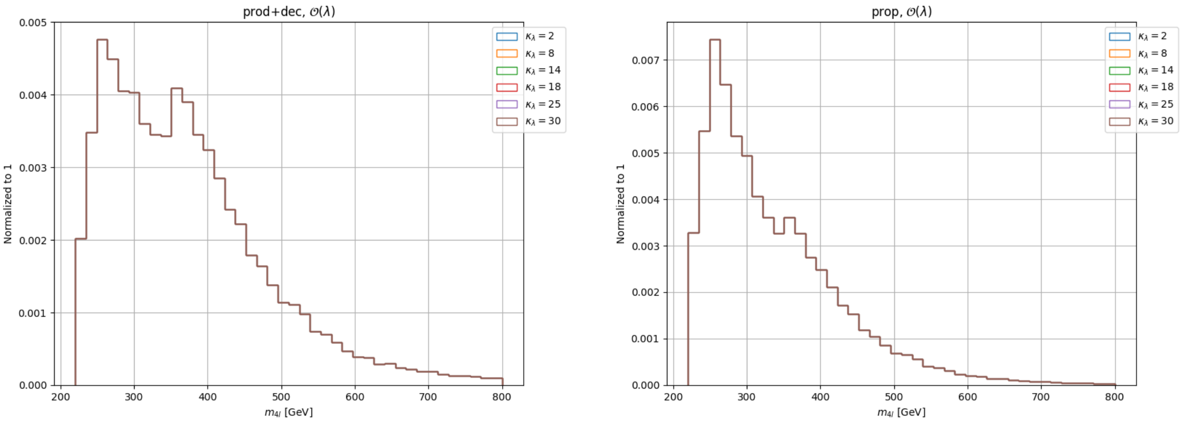
If the κ_λ measurement were based solely on the overall yield, it would not be necessary to decompose the total amplitude into the 10 components defined in Eq. 5.34. In that case, one could instead reweight to an inclusive model containing the SM LO signal, background, and SMEFT contributions, and scale it using a fourth-order polynomial to obtain the prediction for a given value of κ_λ.

However, varying κ_λ affects not only the total yield but also the shapes of the observables. Applying a single parametric scaling to the entire template would reproduce the yield correctly, but would fail to describe the shape dependence. By decomposing the amplitude into its fundamental components, each contribution retains a fixed shape as κ_λ varies, while its normalization changes. Combination of these components with different κ_λ-dependent scalings then induces a shape dependence in the final distribution.

Following the yield validation, a shape validation is carried out to ensure that the specific template can be parametrized with a single scaling function. This requirement implies that for each template, the shape of an observable should remain unchanged under variations in κ_λ; only the normalization should vary.

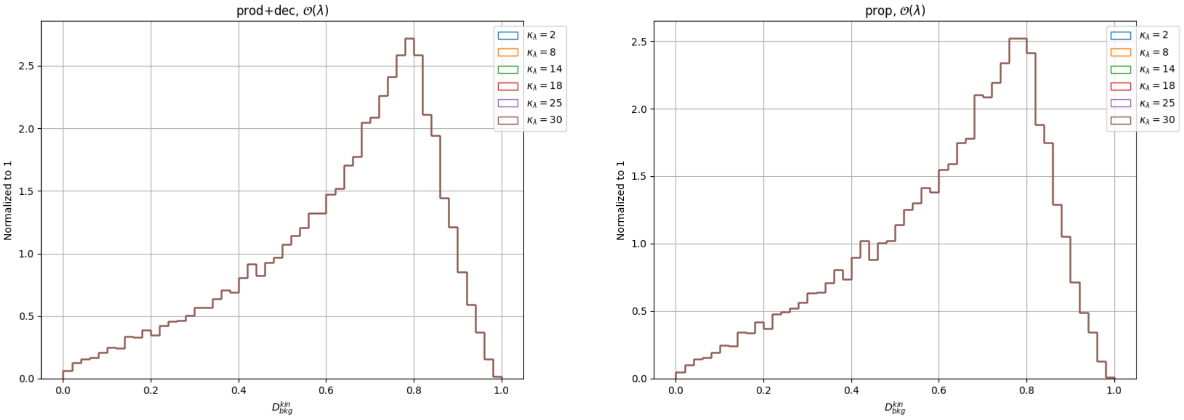
To test this, templates are reweighted to various values of κ_λ and the resulting distributions are normalized and compared. If shape preservation holds, normalized distributions for all observables should overlap. These tests are conducted for all 7 κ_λ-dependent components both for ggF and EW and for all observables. Chosen distributions are shown on Figs. 5.22–5.25.

The shape preservation holds for all the templates and as seen for example in Figs. 5.22–5.25, the shape of observables is preserved across different κ_λ values, for linear component of production combined with decay and propagation processes.



(a) (b)

Figure 5.22: ggF: $m_{4\ell}$ distributions for (a) linear component of production and decay with δZ_H contributions removed and (b) propagation with δZ_H contributions removed.



(a) (b)

Figure 5.23: ggF: $D_{\text{bkg}}^{\text{kin}}$ distributions for (a) linear component of production and decay and (b) propagation, with δZ_H contributions removed.

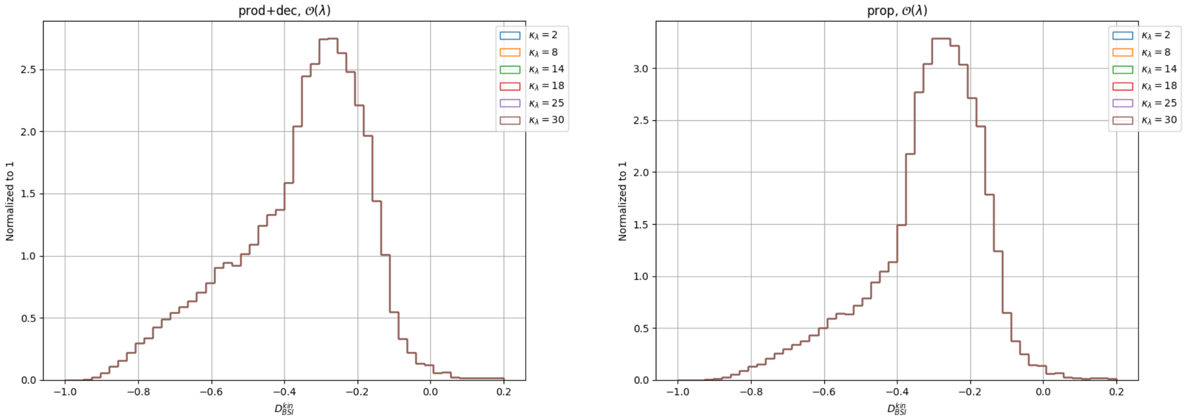
5.5.9 Effect of Modified Higgs Self-coupling on Shapes of Observables

Insertion of the $\mathcal{O}_6$ operator affects not only the overall cross section but also the shapes of the observables used in the analysis. The most visible effect appears in the $m_{4\ell}$ distribution, where a bump is observed near the $2m_H$ threshold. The non-local discriminant ($nlD = D_{bkg}^{kin} \times D_{bsi}$) has a weaker dependence on κ_λ. This observable is still important since D_{bkg}^{kin} provides separation against the SM background.

Figures 5.26,5.27 and 5.28 show the $m_{4\ell}$ signal, background and interference distributions for the Untagged, VBFtagged and VHtagged categories. The binning shown corresponds to the one used in analysis [9]

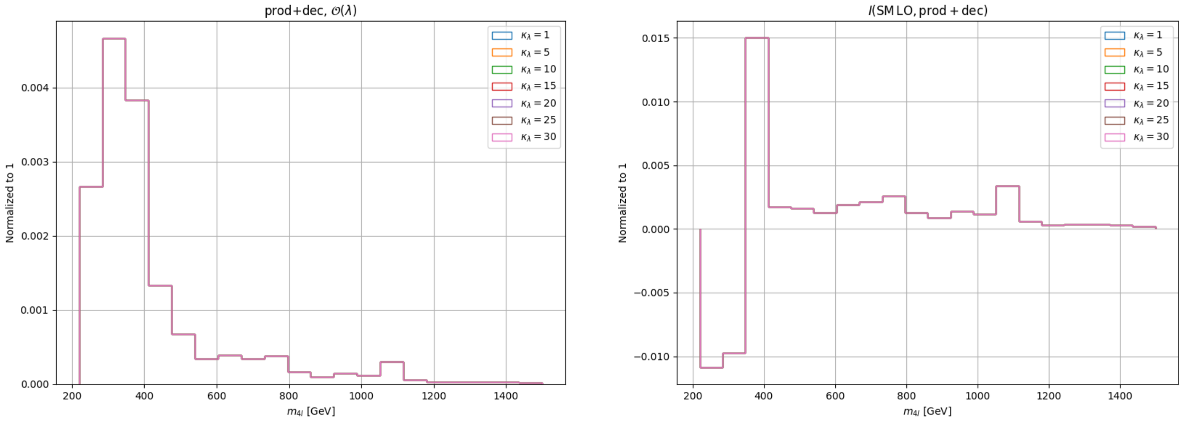
---

[9] Analysis considers invariant mass up to 13 TeV. For visual clarity, and to emphasize $2m_H$ threshold, invariant mass plots are kept to $220 < m_{4\ell} < 570$ GeV range.



(a) (b)

Figure 5.24: ggF: D_{bsi} distributions for (a) linear component of production and decay and (b) propagation, with δZ_H contributions removed.



(a) (b)

Figure 5.25: EW: Shape tests of $m_{4\ell}$ distributions for (a) linear component of production and decay contributions (excluding δZ_H) and (b) interference between SM LO signal and propagation.

Figures 5.29–5.31 show the distributions of the non-local discriminant for the three event categories.

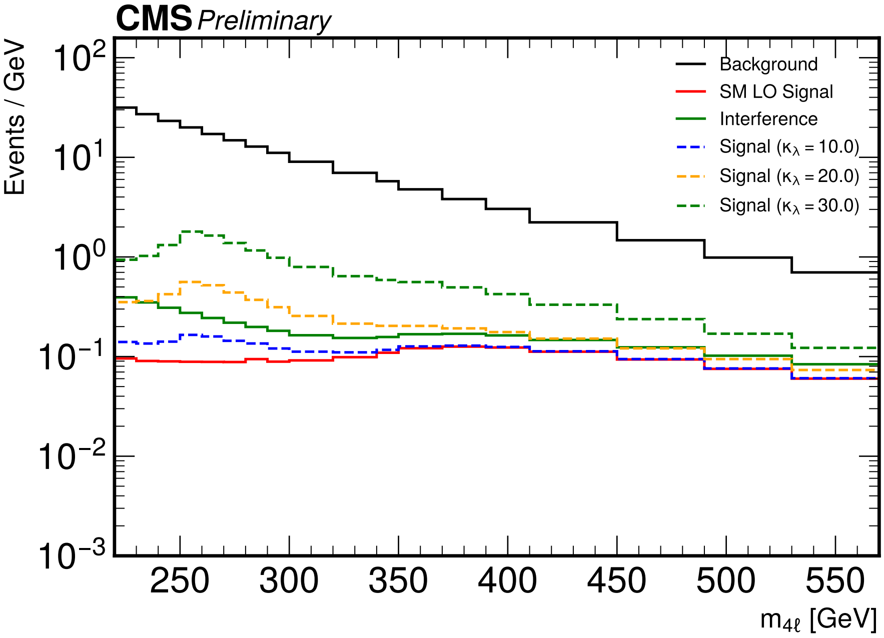


Figure 5.26: $m_{4\ell}$ distribution in the Untagged category. Background sources, shown in black, include ggZZ, qqZZ, VBS and Z+X. The SM LO signal, shown in red, is composed of LO ggF and EW production. The interference contribution, depicted in green, corresponds to the sum of the interference between LO ggF and ggZZ, and between LO EW production and VBS. Dashed lines indicate the effect of different values of κ_λ ($\kappa_\lambda = 10, 20, 30$).

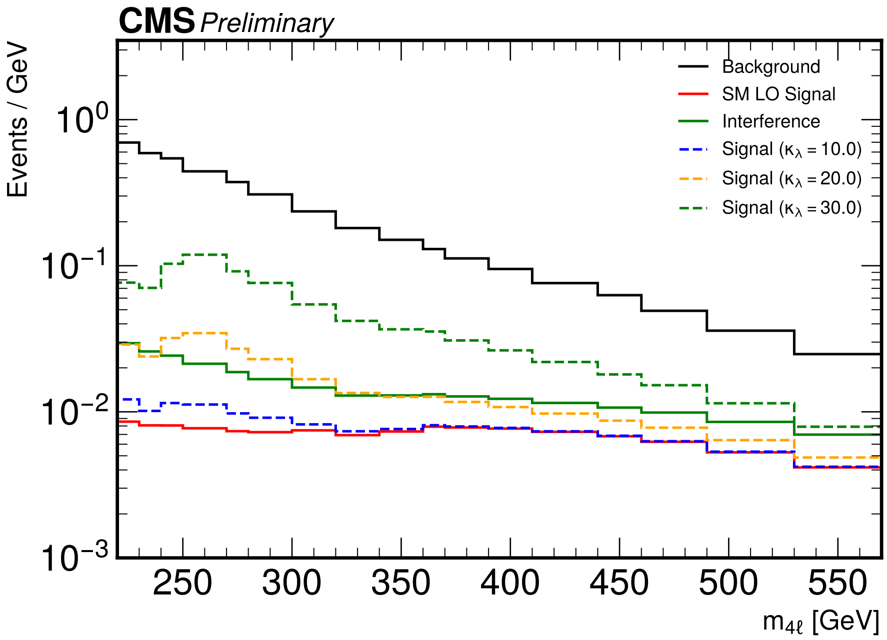


Figure 5.27: $m_{4\ell}$ distribution in the VBFtagged category. Background sources, shown in black, include ggZZ, qqZZ, VBS and Z+X. The SM LO signal, shown in red, is composed of LO ggF and EW production. The interference contribution, depicted in green, corresponds to the sum of the interference between LO ggF and ggZZ, and between LO EW production and VBS. Dashed lines indicate the effect of different values of κ_λ ($\kappa_\lambda = 10, 20, 30$).

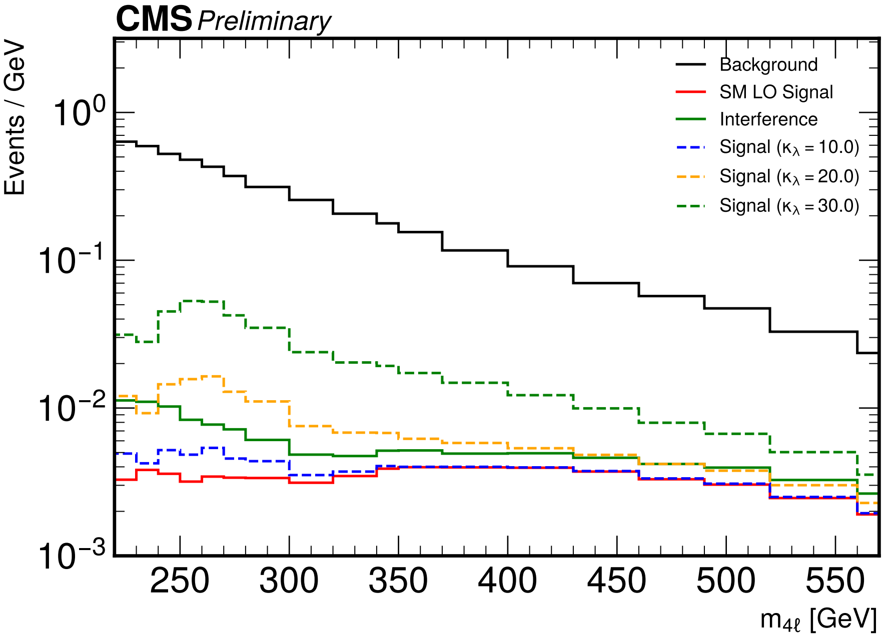


Figure 5.28: $m_{4\ell}$ distribution in the VHtagged category. Background sources, shown in black, include ggZZ, qqZZ, VBS and Z+X. The SM LO signal, shown in red, is composed of LO ggF and EW production. The interference contribution, depicted in green, corresponds to the sum of the interference between LO ggF and ggZZ, and between LO EW production and VBS. Dashed lines indicate the effect of different values of κ_λ ($\kappa_\lambda = 10, 20, 30$).

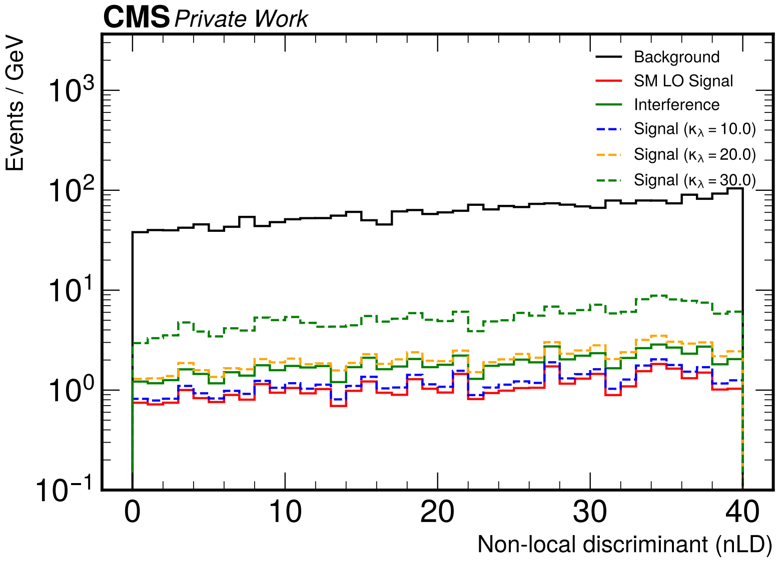


Figure 5.29: nLD distribution in the Untagged category for SM LO signal, background, interference and $\kappa_\lambda = 10, 20, 30$.

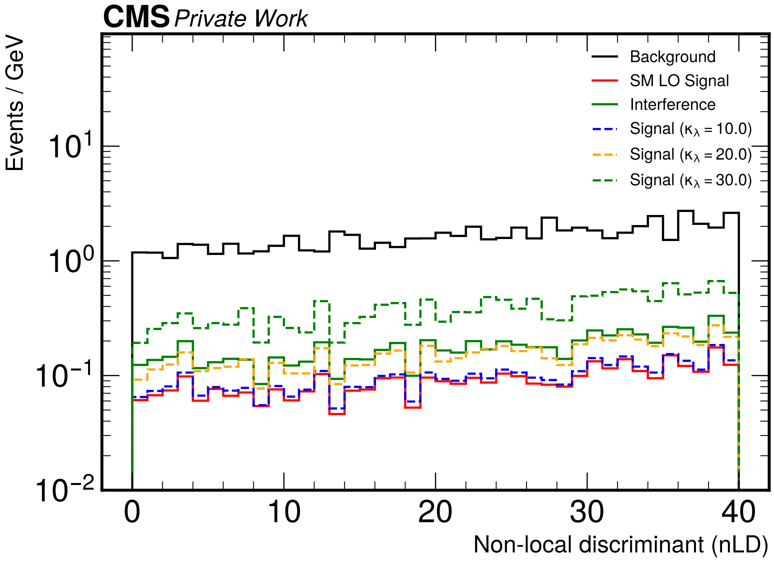


Figure 5.30: nLD distribution in the VBFtagged category for SM LO signal, background, interference and $\kappa_\lambda = 10, 20, 30$.

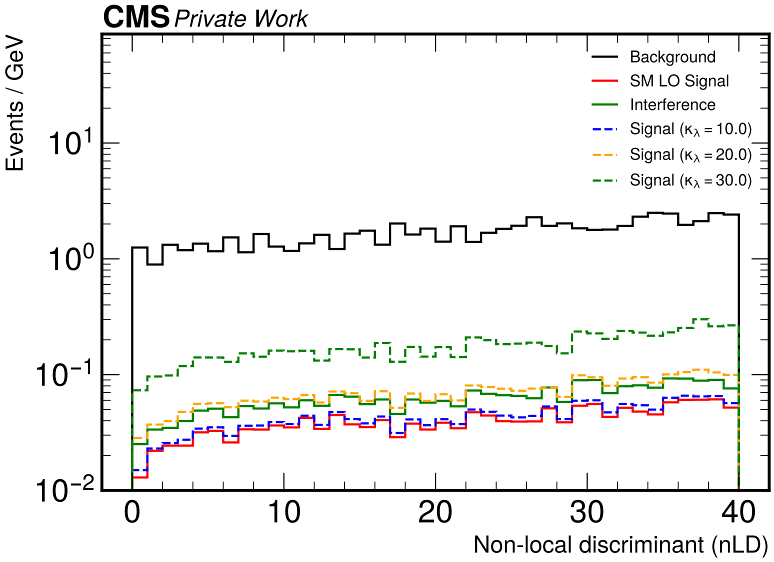


Figure 5.31: nLD distribution in the VHtagged category for SM LO signal, background, interference and $\kappa_\lambda = 10, 20, 30$.

5.5.10 Confidence Intervals

This analysis follows a template-based approach, where PFDs describing the observables are represented by histograms. These templates define the expected distributions of signal, background and interference processes and are used to construct the statistical model from which the parameter of interest (POI), κ_λ, is extracted. The extraction is performed using the Maximum Likelihood method.

The likelihood function quantifies the probability of observing the experimental data given a particular set of model parameters. In this analysis, the likelihood connects the predicted signal and background yields to the observed event distributions.

Given an experimental dataset $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ consisting of n observed events, the likelihood function can be written as

$$\mathcal{L}(\vec{x}|\vec{\Phi}) = \prod_{i=1}^{n} f(x_i|\vec{\Phi}), \tag{5.38}$$

where $f(x|\vec{\Phi})$ denotes the PDF for observing an event x given the model parameters $\vec{\Phi}$.

In this analysis the observables are described by binned distributions. Each bin corresponds to a region of phase space and is modeled using a Poisson distribution,

$$P(k|\nu) = \frac{e^{-\nu}\nu^k}{k!}, \tag{5.39}$$

where ν denotes the expected number of events in the bin and k the number of observed events. The likelihood is therefore constructed as the product of Poisson probabilities across all bins,

$$\mathcal{L}(\vec{x}|\mu, \kappa_\lambda, \vec{\theta}) = \prod_{i=1}^{n} P(x_i|\nu_i(\mu, \kappa_\lambda, \vec{\theta})) \prod_{k} p(\theta_k|\hat{\theta}_k). \tag{5.40}$$

Here $\vec{\theta}$ represents nuisance parameters associated with systematic uncertainties. Each nuisance parameter θ is constrained around its nominal value $\hat{\theta}$ through a corresponding PDF $p(\theta|\hat{\theta})$.

In many Higgs measurements the parameter of interest is expressed as the signal strength modifier μ, defined as the ratio between the measured signal cross section and the prediction of the SM,

$$\mu = \frac{\sigma}{\sigma_{\mathrm{SM}}}. \tag{5.41}$$

In such cases the expected yield in each bin can be written as

$$\nu_j = \mu \cdot S_j + B_j, \tag{5.42}$$

where S_j and B_j represent the expected signal and background contributions in bin j.

In this analysis the expected yields are modified to account for the strong interference present in the off-shell region and for additional BSM contributions arising from the SMEFT dependence on κ_λ. The expected yield in each bin can therefore be written as

$$\nu_j = \mu \cdot s_j + \sqrt{\mu} \cdot I_j + b_j + \mathcal{O}_{\lambda,j}, \tag{5.43}$$

where I_j and $\mathcal{O}_{\lambda,j}$ represent the interference and BSM contributions. The interference term scales as $\sqrt{\mu}$. This can be understood from the structure of the squared matrix element,

$$A^2 = |A_{\text{SIG}} + A_{\text{BKG}}|^2 = A_{\text{SIG}}^2 + 2A_{\text{SIG}}A_{\text{BKG}} + A_{\text{BKG}}^2. \tag{5.44}$$

If the signal contribution is scaled as $A_{\text{SIG}}^2 \to \mu A_{\text{SIG}}^2$, the amplitude itself scales as $\sqrt{\mu}$, which leads to the interference term scaling as $\sqrt{\mu}$.

The parameter of interest in this analysis is κ_λ, which modifies the signal yield through the physics model described in the previous sections. The statistical model therefore depends on κ_λ, the signal strength modifier μ, and the nuisance parameters $\vec{\theta}$.

The likelihood is evaluated using a profile likelihood approach. In this procedure κ_λ is scanned while the remaining parameters (μ and $\vec{\theta}$) are determined by maximizing the likelihood at each value of κ_λ. The likelihood therefore takes the form

$$\mathcal{L}(\kappa_\lambda, \mu, \vec{\theta}). \tag{5.45}$$

For each point in the κ_λ scan the minimizer finds the values of μ and $\vec{\theta}$ that maximize the likelihood, resulting in the profiled likelihood function.

Two types of scans are considered in this analysis: expected and observed. The expected scan is obtained using an Asimov dataset [118]. The Asimov dataset is a representative dataset in which statistical fluctuations are surpressed and the number of events in each bin is set equal to the expected value of the statistical model.

The observed scan is obtained using the data collected by the experiment.

Confidence intervals on κ_λ are determined using the profile likelihood ratio

$$-2\Delta \ln \mathcal{L}(\kappa_\lambda) = -2 \left[\ln \mathcal{L}(\kappa_\lambda) - \ln \mathcal{L}_{\text{min}} \right]. \tag{5.46}$$

According to Wilks' theorem [119, 118], this test statistic asymptotically follows a χ^2 distribution with one degree of freedom. As a result, confidence intervals can be obtained from the condition

$$-2\Delta \ln \mathcal{L} = 1 \quad (68\% \text{ CL}), \qquad -2\Delta \ln \mathcal{L} = 4 \quad (95\% \text{ CL}). \tag{5.47}$$

These criteria correspond to the commonly used 1σ and 2σ confidence intervals for a single parameter of interest.

The fits in this analysis are performed using the CMS `Combine` framework [120], which provides tools for constructing likelihood-based statistical models and performing parameter estimation and hypothesis testing. It can be also used for evaluation of fit diagnostics and several fit metrics.

Frequentist approach

Inspired by ATLAS measurement of width in off-shell region [121], shown on Fig. 5.32, 68% and 95% confidence level (CL) curves are established using frequentist approach, and compared to flat lines from Wilk's theorem. The procedure goes as follows:

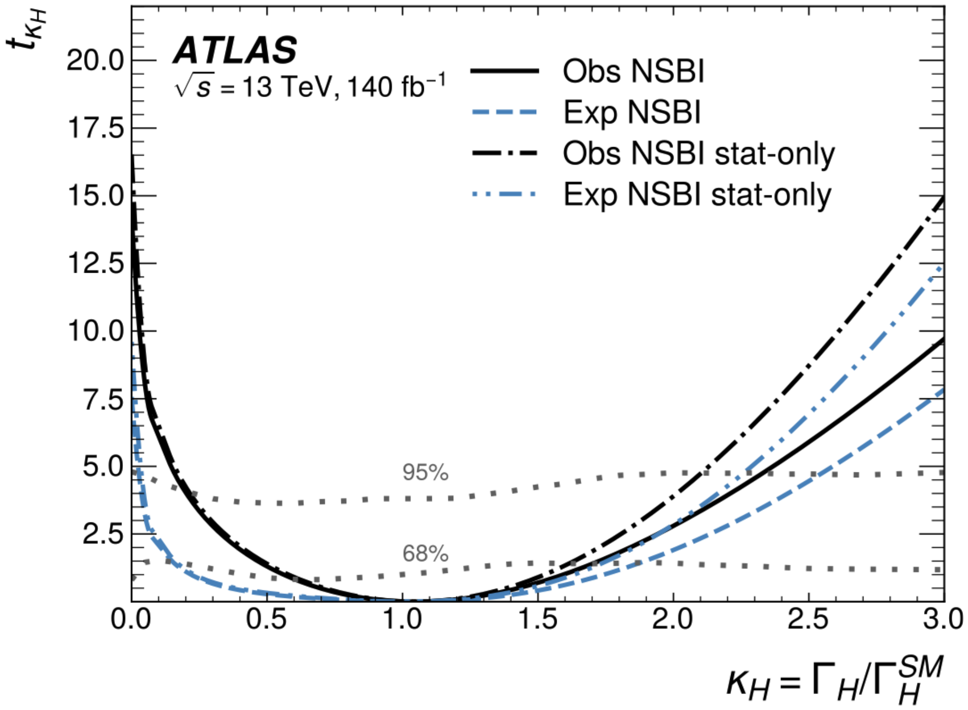


Figure 5.32: Values of the test statistic t_{κ_H} as a function of $\kappa_H = \Gamma_H/\Gamma_H^{\mathrm{SM}}$, obtained with an Asimov dataset (expected, dotted black) and with data (observed, solid black). The dash-dotted curves show the statistics-only results where all nuisance parameters are fixed to their best-fit values $\hat{\alpha}$. The dotted gray lines indicate the 68% and 95% confidence intervals derived using the Neyman construction. Results obtained with neural simulation-based inference (NSBI) are shown [121].

- For each value of κ_λ, 10000 toy datasets are generated
- For each toy, an independent scan over κ_λ is performed; the value of 2ΔNLL (with respect to the toy's best-fit point) is evaluated at the κ_λ value used to generate the toy
- The 68^{th} and 95^{th} percentiles of the resulting 2ΔNLL distribution are extracted
- The procedure is repeated for all considered κ_λ values

To make fits easier μ is fixed to 1 and systematics are frozen. Construction is depicted on Figure 5.33. As can be seen from Figure 5.33, Wilk's theorem appears to be generally more conservative or on point with frequentist approach. For this reason, CI estimation using Wilk's theorem is adopted in the analysis.

5.5.11 Impact plots and Uncertainty Breakdown

Before making scans, it is common in the CMS analysis to evaluate the impact of systematic uncertainties to the POI. Impact plots summarize the effect of nuisance parameters (NPs) on the POI. In this analysis the POI is κ_λ. The impact of a nuisance parameter θ on κ_λ is defined as the shift $\Delta\kappa_\lambda$ obtained when θ is fixed to its post-fit value shifted by

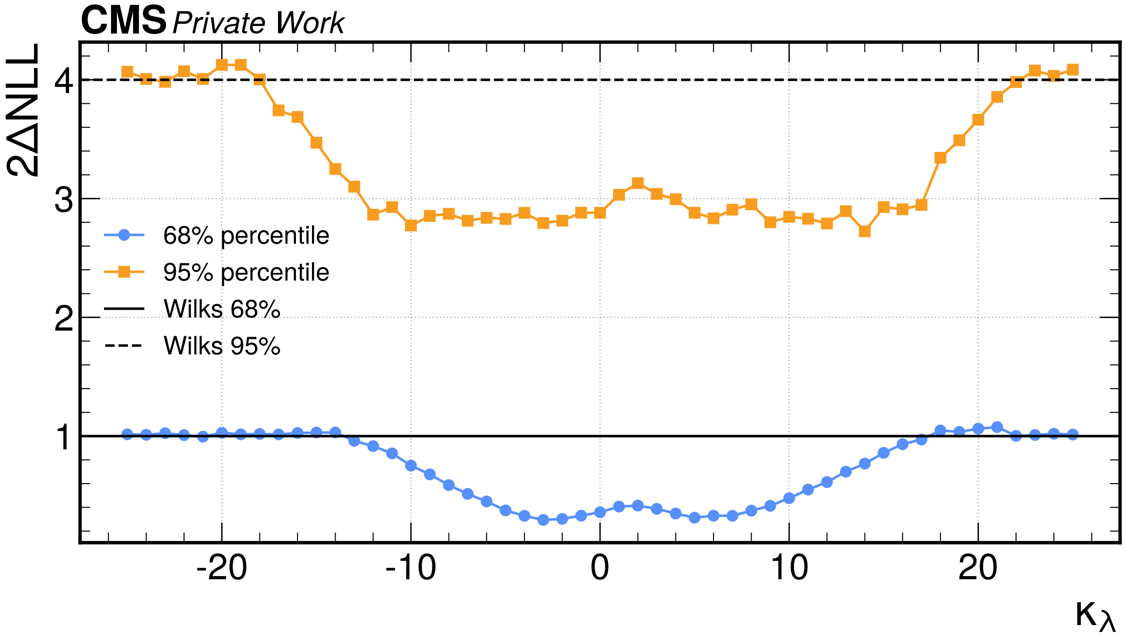


Figure 5.33: Confidence intervals derived using frquentist approach, compared to Wilk's theorem.

$\pm 1\sigma$, while all other parameters are profiled in the likelihood fit. This procedure quantifies how variations of a given systematic uncertainty propagate to the extracted value of κ_λ and provides a ranking of nuisance parameters according to their influence on the measurement.

Impact plots for 30 leading NPs are shown on Fig. 5.34 for Asimov dataset and on Fig. 5.35 for observed data.

The left panel in the summary plot shows the value of $(\nu - \nu_0)/\Delta_\nu$ where ν and ν_0 are the post and pre-fit values of the nuisance parameter and Δ_ν is the pre-fit uncertainty. The asymmetric error bars show the post-fit uncertainty divided by the pre-fit uncertainty meaning that parameters with error bars smaller than ±1 are constrained in the fit. The pull (the difference between the post-fit and nominal value of a nuisance parameter expressed in units of its pre-fit uncertainty) is additionally shown.

As shown in Figs. 5.34 and 5.35, uncertainties associated with the k-factors have a significant impact on the measurement (*EWcorr_qqZZ* and *kf_ggZZ_back*). This behavior is typical for off-shell analyses. The $ggZZ$ background process carries a k-factor uncertainty of 10%, while the $qqZZ$ process constitutes the dominant contribution in the analysis. As a result, uncertainties affecting these processes propagate strongly to the parameter of interest.

The electron efficiency uncertainty (*CMS_eff_e*) also has a noticeable impact on the measurement, reflecting the sensitivity of the analysis to lepton reconstruction and identification performance. In addition, sizeable contributions arise from PYTHIA-related renormalization and factorization scale uncertainties, *QCDscale_muF_ggH* and *QCDscale_muR_qqbar*.

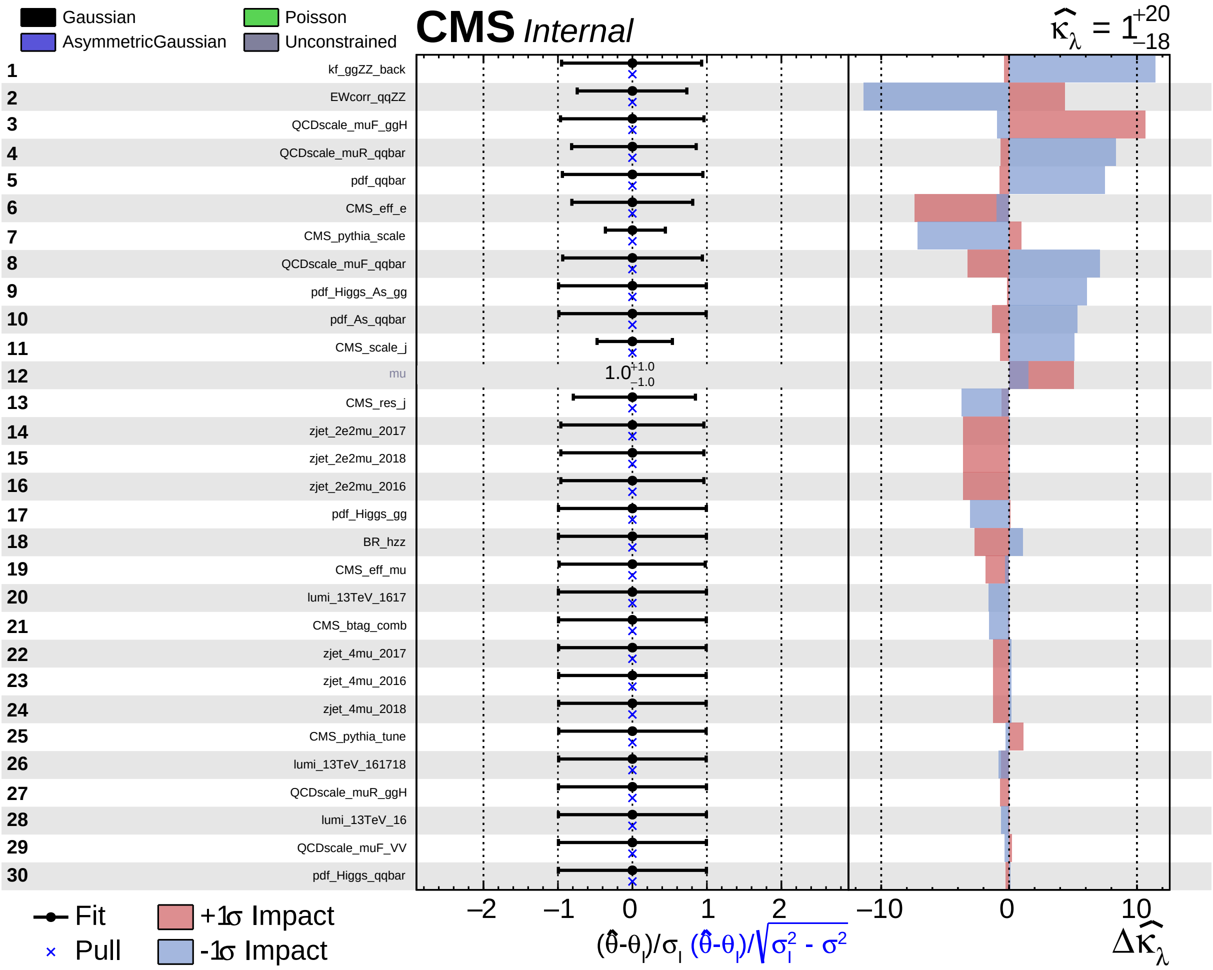


Figure 5.34: Impact plot for Asimov dataset, nuisance parameters 1–30

This analysis is however, heavily statistically dominated. This can be seen from Fig. 5.36, comparing total uncertainty (black) and contribution of the statistical component (dashed magenta). Contributions from the selected nuisance parameters are also reported. Selected nuisance parameters are freezed in the fit, after which their contribution is estimated by subtraction in quadrature as $\sigma_{\text{nuisance}} = \sqrt{\sigma^2_{\text{total}} - \sigma^2_{\text{freeze nuisance}}}$

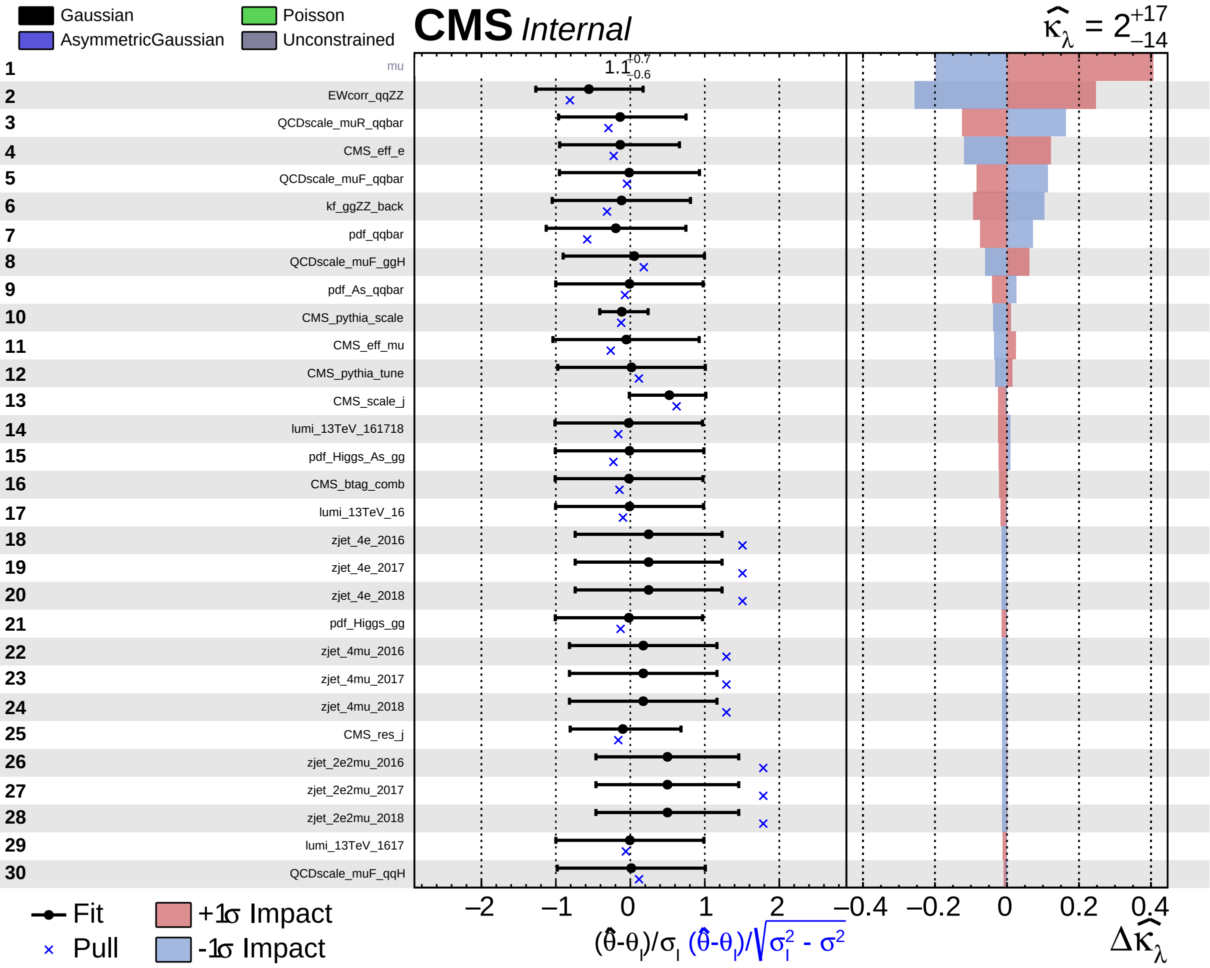


Figure 5.35: Impact plot for data, nuisance parameters 1–30

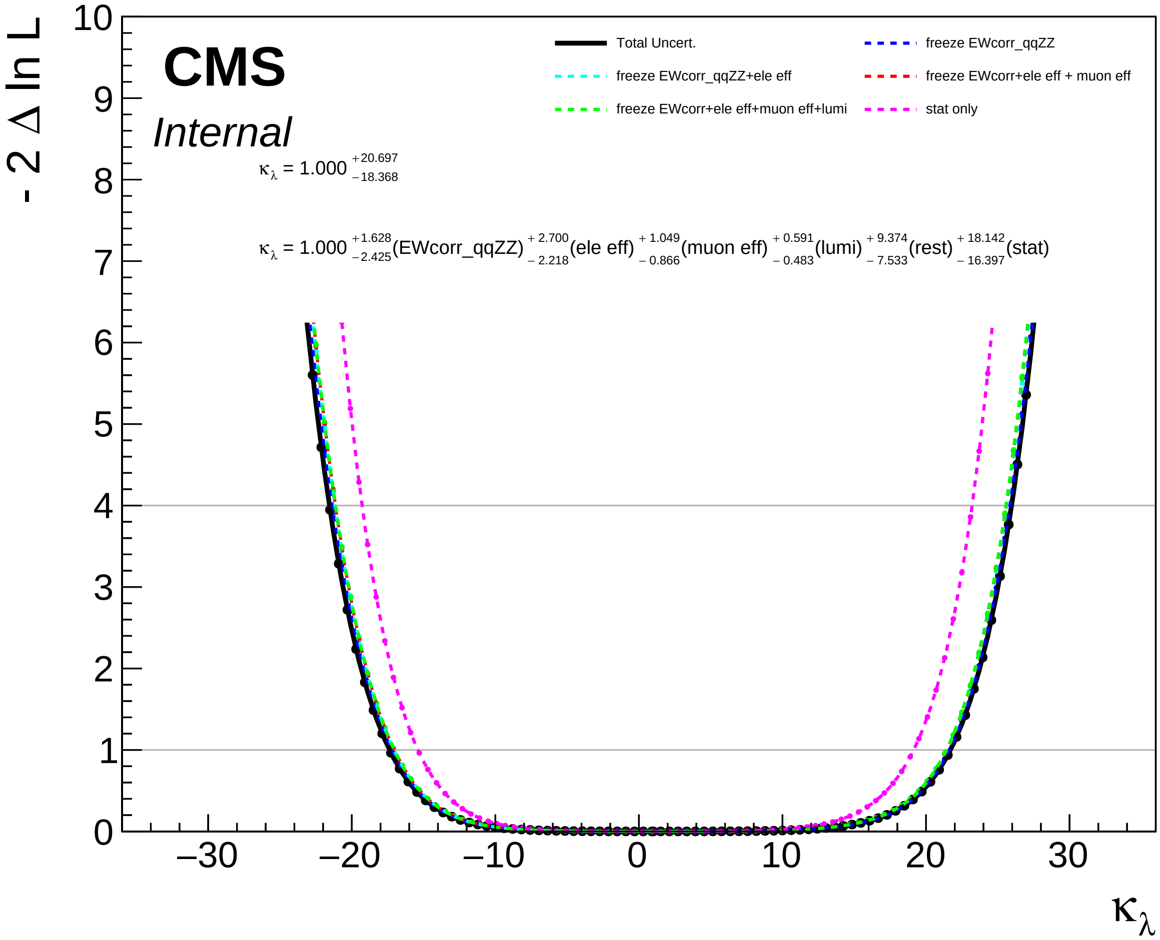


Figure 5.36: Breakdown of the uncertainties in the κ_λ measurement obtained from the profile likelihood scan. The black curve shows the full uncertainty, while the magenta dashed curve corresponds to the statistical-only result. The remaining curves are used to estimate the contribution of selected nuisance parameters: $qqZZ$ EW correction (*EWcorr_qqZZ*, blue), electron efficiency (*CMS_eff_e*, cyan), muon efficiency (*CMS_eff_mu*, red), and luminosity (*lumi*, green).

5.5.12 Likelihood Scan of Higgs Self-coupling Modifier

A likelihood fit is performed, with parameter μ floating. The resulting scan is shown in Fig. 5.37. Table 5.19 summarizes the 68% and 95% confidence intervals, for both the expected and observed scans.

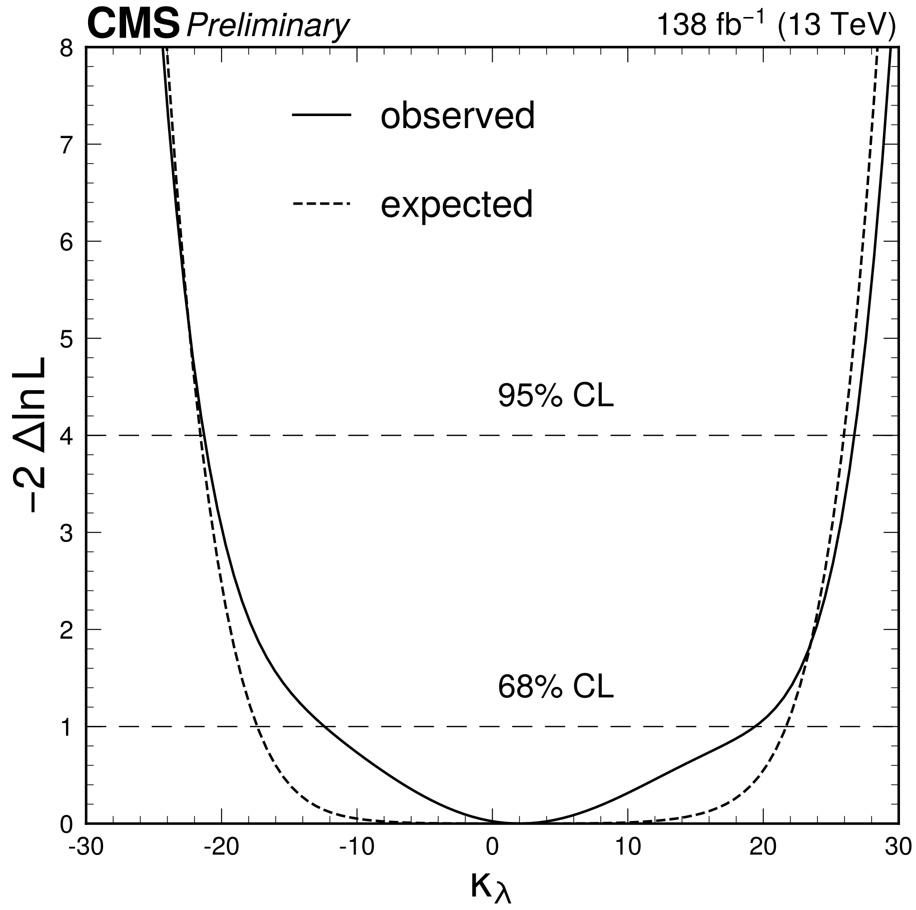


Figure 5.37: κ_λ observed and expected scans

	Expected	Observed
κ_λ	1^{+21}_{-18} [−22, 26]	2^{+17}_{-15} [−21, 27]

Table 5.19: Summary of the κ_λ measurement. The central values and uncertainties correspond to the 68% CL intervals, while the 95% CL intervals are shown in square brackets.

Although this limit is not as stringent as those obtained from double Higgs measurements, such as the bb$\tau\tau$ final state [83], which yields a 95% CL observed (expected) constraint of $-1.7 < \kappa_\lambda < 8.7$ ($-2.9 < \kappa_\lambda < 9.8$), it represents the first measurement of this parameter in the off-shell region and opens the possibility for future combinations in upcoming LHC runs.

To summarize, this analysis builds upon SMEFT corrections to ggF production derived in [1]. These corrections were propagated to the MCFM version used in the analysis and implemented in MELA to enable event reweighting. A comprehensive set of validation and closure tests was performed to ensure the correctness of the implementation. An analogous implementation of SMEFT effects in the EW sector was then developed and incorporated into the analysis.

The resulting model was integrated into the full analysis framework, enabling the first constraint on the Higgs self-coupling modifier κ_λ from the off-shell region. Although the current sensitivity is limited, this result constitutes the first probe of this parameter in this regime and establishes a foundation for future improvements and combinations in global fits with upcoming LHC data.

5.6 Other Off-shell Measurements

This section briefly describes other measurements that are part of the off-shell analysis and were carried out by other members of the analysis team. These include measurements of Higgs compositeness, Yukawa couplings, and the Higgs boson width.

5.6.1 Higgs Structure

This measurement probes the compositeness of the Higgs boson. The strategy is analogous to those applied to hadronic systems, such as the proton, whose internal structure was revealed in deep inelastic scattering: at large momentum transfer q^2, deviations from point-like behavior indicate substructure.

In composite Higgs models, new strong dynamics emerge at the TeV scale, leading to a momentum-dependent form factor that modifies Higgs boson interactions. As in the proton case, increasing q^2 probes shorter distance scales, reducing the probability for the Higgs boson to remain a single bound object. This effect is described by a form factor

$$F(q^2) = \left[\frac{1}{1 + |q^2|/\Lambda_H^2}\right]^n, \tag{5.48}$$

where Λ_H denotes the compositeness scale and is related to the effective size of the Higgs boson via

$$d = \hbar c/\Lambda_H. \tag{5.49}$$

Finite values of Λ_H correspond to a composite object, while the SM limit is recovered for $\Lambda_H \to \infty$. The form factor is implemented as a modification of the HZZ vertex, affecting both production and decay processes.

On-shell and off-shell regions are combined and the results can be seen in Fig. 5.38, and Table. 5.20. In the most general case using off-shell production alone, the value for Λ_H is above 301.04 GeV at the 95% confidence level and minimizes to a finite value at 1500 GeV. Combining on- and off-shell while assuming Γ_H to be fixed at SM value of 4.14 MeV finds that Λ_H is above 839.07 GeV at the 95% confidence level.

5.6.2 Yukawa Couplings

The Yukawa couplings of the Higgs boson enter the off-shell region mainly through the gluon-fusion production process, which proceeds via quark loops. The dominant contribution arises from the top quark, but the analysis also considers possible contributions from light quarks, $q \in \{u, d, s, c\}$, and from heavier states, denoted collectively by

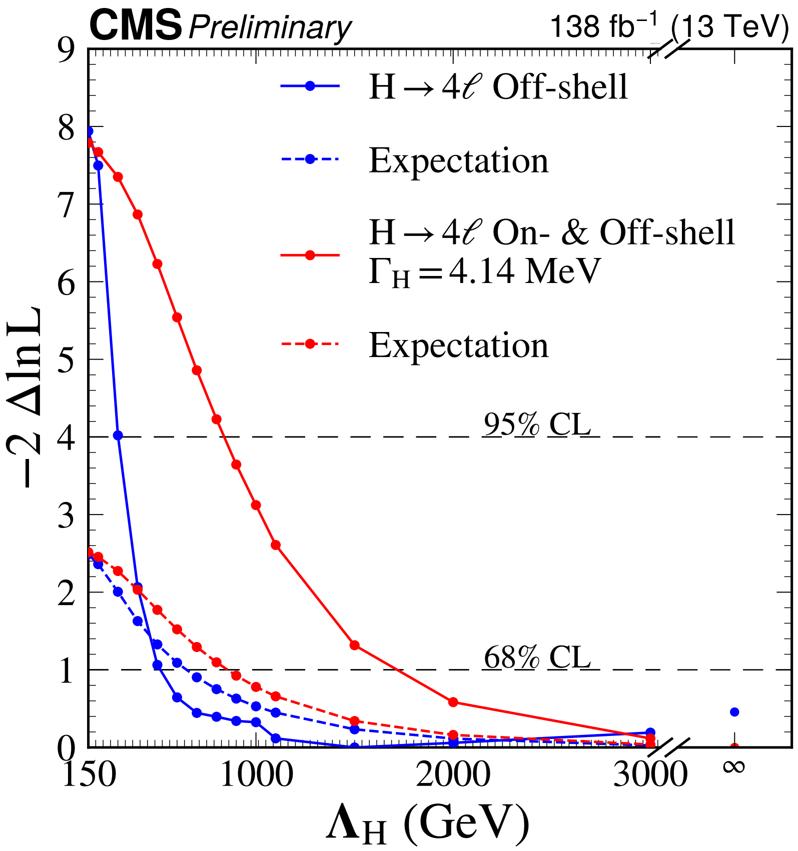


Figure 5.38: The observed (solid) and expected (dashed) profile likelihood projections from the fit for Λ_H using off-shell (blue) and on- and off-shell (orange) production. The markers on each line show the values of Λ_H that were profiled.

Channel	Best Fit (GeV)	Allowed Values (68%) [95%] (GeV)
4ℓ off-shell	1500	(> 515.10), $[> 301.04]$
Expected	∞	(> 648.27), $[> 0]$
4ℓ on- and off-shell	∞	(> 1716.00), $[> 839.07]$
Expected	∞	(> 856.209), $[> 0]$

Table 5.20: Summary of the Λ_H values measured in observation and expectation, showing the central value, and the 68% CL (in parenthesis) and 95% CL (in square brackets) for $H \to ZZ \to 4\ell$.

$Q' \in \{t, Q, \tilde{t}, \tilde{Q}\}$. Here, the non-tilde states correspond to CP-even interactions, while the tilde states denote CP-odd contributions. In addition, EW production terms are included and are controlled by the couplings to vector bosons.

On-shell and off-shell regions are combined, with the main benefit being the removal of degeneracies present in the on-shell cross section, which is proportional to κ_Z^2/Γ_H. When considering on-shell production only, the restriction $|\kappa_Z| \leq 1$ must be imposed, since the cross section would otherwise be unbounded, as an increase in $|\kappa_Z|$ could be compensated by Γ_{BSM}. The off-shell region is independent of Γ_H and exhibits shape sensitivity in the $m_{4\ell}$ distribution, allowing this constraint to be removed. The combination of on- and off-shell regions reduces degeneracies in κ_Z, after which the on-shell region drives the sensitivity to the Higgs boson coupling to light quarks.

Figs. 5.39 and 5.40 show the combined on- and off-shell scans, with different parameters floating. Numerical results are shown Table 5.21. Heavy quark couplings are fixed to their SM values throughout these scans, while Yukawa couplings to light quarks q and Γ_{BSM} are floated.

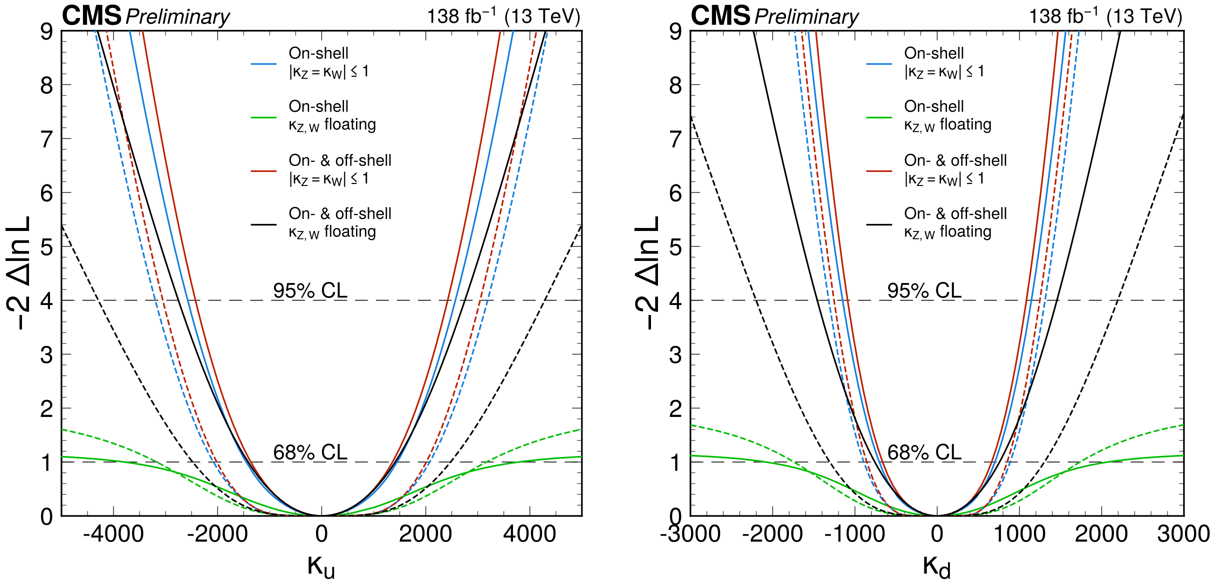


Figure 5.39: Observed (solid) and expected (dashed) likelihood scans for light-quark Yukawa couplings. The blue curve corresponds to the on-shell analysis assuming custodial symmetry with $|\kappa_{Z,W}| < 1$. The red curve shows the corresponding result including both on-shell and off-shell regions. The green curve represents the on-shell analysis without the custodial symmetry assumption, with $\kappa_{Z,W}$ treated as free parameters. The black curve shows the corresponding result including both on-shell and off-shell regions. Left: κ_u scan. Right: κ_d scan.

Kappa term	Expected	Observed
κ_u	$1.0^{+2.5\times10^3}_{-2.5\times10^3}[-4.3\times10^3,\ 4.3\times10^3]$	$0.3^{+1.4\times10^3}_{-1.4\times10^3}[-2.8\times10^3,\ 2.8\times10^3]$
κ_d	$1.0^{+1.3\times10^3}_{-1.3\times10^3}[-2.2\times10^3,\ 2.2\times10^3]$	$0.34^{+760}_{-760}[-1.46\times10^3,\ 1.46\times10^3]$
κ_s	$1.0^{+77}_{-81}[-122,\ 117]$	$0.51^{+54}_{-56}[-96,\ 93]$
κ_c	$1.0^{+5.2}_{-7.0}[-10.2,\ 8.8]$	$0.12^{+4.9}_{-5.7}[-8.9,\ 7.8]$

Table 5.21: Summary of the light-Yukawa coupling measurements, showing the 68% CL (central values with uncertainties) and 95% CL (in square brackets) intervals for $H \to ZZ \to 4\ell$. In these scans, the parameters κ_u, κ_d, κ_s, κ_c, κ_Z, and κ_W are floated, while $\kappa_t = \kappa_b = 1$ and all heavy-quark BSM couplings are fixed to their SM values.

Scans are also performed for heavy quarks while fixing the couplings to light quarks

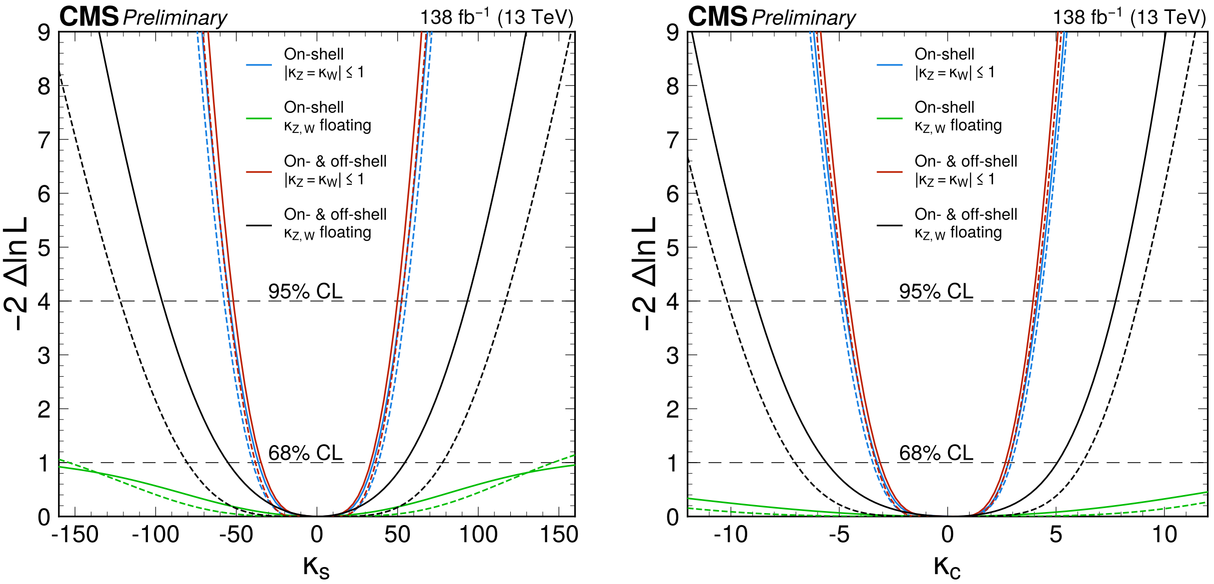


Figure 5.40: Observed (solid) and expected (dashed) likelihood scans for light-quark Yukawa couplings. The blue curve corresponds to the on-shell analysis assuming custodial symmetry with $|\kappa_{Z,W}| < 1$. The red curve shows the corresponding result including both on-shell and off-shell regions. The green curve represents the on-shell analysis without the custodial symmetry assumption, with $\kappa_{Z,W}$ treated as free parameters. The black curve shows the corresponding result including both on-shell and off-shell regions. Left: κ_s scan. Right: κ_c scan.

and the b quark to their SM values. Heavy Yukawa couplings and Γ_{BSM} are floated during the fit.

Figs. 5.41 and 5.42 show the combined scans, with numerical results in Table 5.22.

Kappa term	Expected	Observed
κ_Q	$0.00^{+1.13}_{-1.16}[-2.43, 2.40]$	$-0.04^{+0.29}_{-0.22}[-0.82, 0.80]$
κ_t	$\left(1.00^{+2.87}_{-0.89}\right) \cup \left(-1.00^{+0.90}_{-2.95}\right)$ $[-5.46, 5.41]$	$0.51^{+0.49}_{-0.20}[0.09, 3.11]$
$\widetilde{\kappa}_t$	$0.00^{+0.90}_{-0.90}[-1.71, 1.71]$	$0.00^{+0.27}_{-0.27}[-1.00, 1.00]$
$\widetilde{\kappa}_Q$	$0.00^{+0.92}_{-0.83}[-1.77, 1.77]$	$0.00^{+0.15}_{-0.15}[-0.65, 0.65]$

Table 5.22: Summary of the heavy-Yukawa coupling measurements, showing the 68% CL (central values with uncertainties) and 95% CL (in square brackets) intervals for $H \to ZZ \to 4\ell$. In these scans, the parameters κ_t, κ_Q, $\widetilde{\kappa}_t$, $\widetilde{\kappa}_Q$, κ_Z, and κ_W are floated, while $\kappa_b = 1$ and all light-quark Yukawa couplings are fixed to their SM values.

5.6.3 Higgs Width

Higgs width measurement is based on the the off-shell method, originally proposed by Caola and Melnikov [122] and later implemented by the CMS Collaboration [105]. This approach exploits Higgs boson production in two kinematic regimes: the on-shell region,

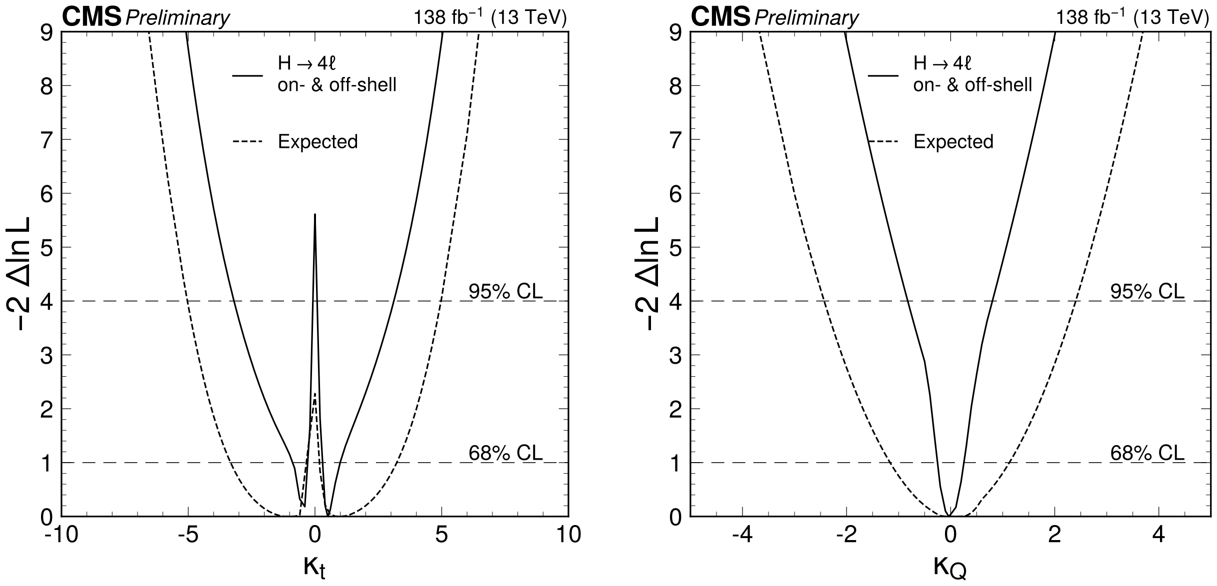


Figure 5.41: Observed (solid) and expected (dashed) likelihood scans for heavy-quark Yukawa couplings with on-shell and off-shell regions combined. Left: κ_t scan. Right: κ_Q scan.

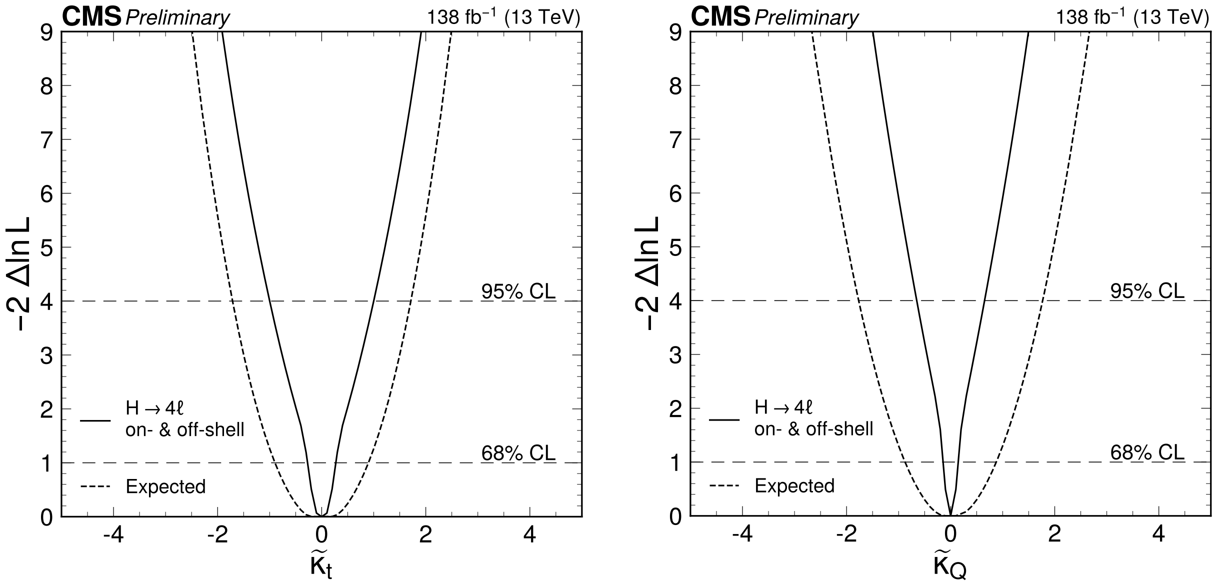


Figure 5.42: Observed (solid) and expected (dashed) likelihood scans for CP-odd heavy-quark Yukawa couplings with on-shell and off-shell regions combined. Left: $\widetilde{\kappa}_t$ scan. Right: $\widetilde{\kappa}_Q$ scan.

defined around the resonance mass (in this case taken as $105 < m_{4\ell} < 140$ GeV), and the off-shell region, characterized by four-lepton invariant masses above 220 GeV.

A central result of [122] is that the total Higgs boson width, Γ_H, can be constrained by comparing production rates in the on-shell and off-shell regions. In the $H \to ZZ \to 4\ell$ channel, a direct determination of the width from the resonance lineshape is fundamentally limited by the experimental mass resolution, which is much larger than the predicted SM width. As a consequence, an on-shell measurement alone does not provide sufficient sensitivity to Γ_H.

The off-shell method overcomes this limitation by exploiting the different dependence of the cross section on Γ_H in the two kinematic regimes. The relevant proportionalities can be written as

$$\sigma^{\text{on-shell}}_{VV\to H\to 4\ell} \propto \mu_{VVH}\,\frac{1}{\Gamma_H}, \qquad \sigma^{\text{off-shell}}_{VV\to H\to 4\ell} \propto \mu_{VVH}, \tag{5.50}$$

where μ_{VVH} denotes the signal strength modifier, defined as the ratio of the observed to the SM expectation for on-shell 4ℓ events. Taking the ratio of off-shell to on-shell production therefore leads to

$$\Gamma_H \propto \frac{\sigma^{\text{off-shell}}_{VV\to H\to 4\ell}}{\sigma^{\text{on-shell}}_{VV\to H\to 4\ell}}. \tag{5.51}$$

Because the off-shell cross section does not scale inversely with the width, this comparison enables an indirect determination of Γ_H that is not constrained by the detector resolution of the Higgs mass peak.

This approach also benefits from distinct kinematic features of Higgs boson decays. For an on-shell Higgs boson with $m_H \approx 125$ GeV, decay into two Z bosons requires at least one Z boson to be produced off-shell, since the Higgs mass lies below the threshold $2m_Z \approx 182$ GeV. In contrast, in the off-shell regime ($m_{4\ell} > 220$ GeV), the intermediate Higgs state can decay into two fully on-shell Z bosons.

The presence of two on-shell vector bosons together with sizable interference between Higgs-mediated and continuum amplitudes makes the high-invariant-mass region sensitive to the Higgs boson width and to potential modifications of the underlying interactions.

The width of the Higgs boson may deviate from the SM expectation of 4.1 MeV [17] if the Higgs boson decays to yet-unknown particles beyond the SM. As such, the direct measurement of the Higgs boson utilizing the off-shell method searches for these possible decays. For example, if the Higgs boson were to decay to potential dark matter candidates, this would be reflected as an increase in the value of Γ_H.

Constraints on Γ_H are obtained from a simultaneous fit to $H \to ZZ \to 4\ell$ events in both the on-shell and off-shell regions, allowing different subsets of parameters to float. The resulting likelihood scans for Γ_H are shown in Fig. 5.43, with the corresponding observed and expected constraints summarized in Table 5.23.

Couplings floated	Expected Γ_H (MeV)	Observed Γ_H (MeV)
SM-like 4ℓ on - and off-shell	$4.1^{+4.2}_{-3.8}$ [< 12.8]	$3.0^{+3.5}_{-2.2}$ [0.3, 10.5]
Z, W separated 4ℓ on - and off-shell	$4.1^{+4.7}_{-3.9}$ [< 13.6]	$5.4^{+4.2}_{-3.7}$ [0.2, 14.9]
Heavy Yukawa floating 4ℓ on - and off-shell	$4.1^{+7.9}_{-3.9}$ [< 19.2]	$5.0^{+7.3}_{-4.4}$ [0.2, 19.2]
Light Yukawa floating 4ℓ on - and off-shell	$4.1^{+10.0}_{-3.8}$ [< 31.3]	$5.5^{+7.3}_{-3.8}$ [0.3, 23.1]
Heavy + Light floating 4ℓ on - and off-shell	$4.1^{+10.9}_{-3.9}$ [< 33.8]	$5.1^{+7.5}_{-4.5}$ [0.2, 23.2]

Table 5.23: Summary of the total Higgs boson width Γ_H measurement, showing the 68% CL (central values with uncertainties) and 95% CL (in square brackets) for $H \to ZZ \to 4\ell$ for the SM and for models with selected anomalous couplings floated.

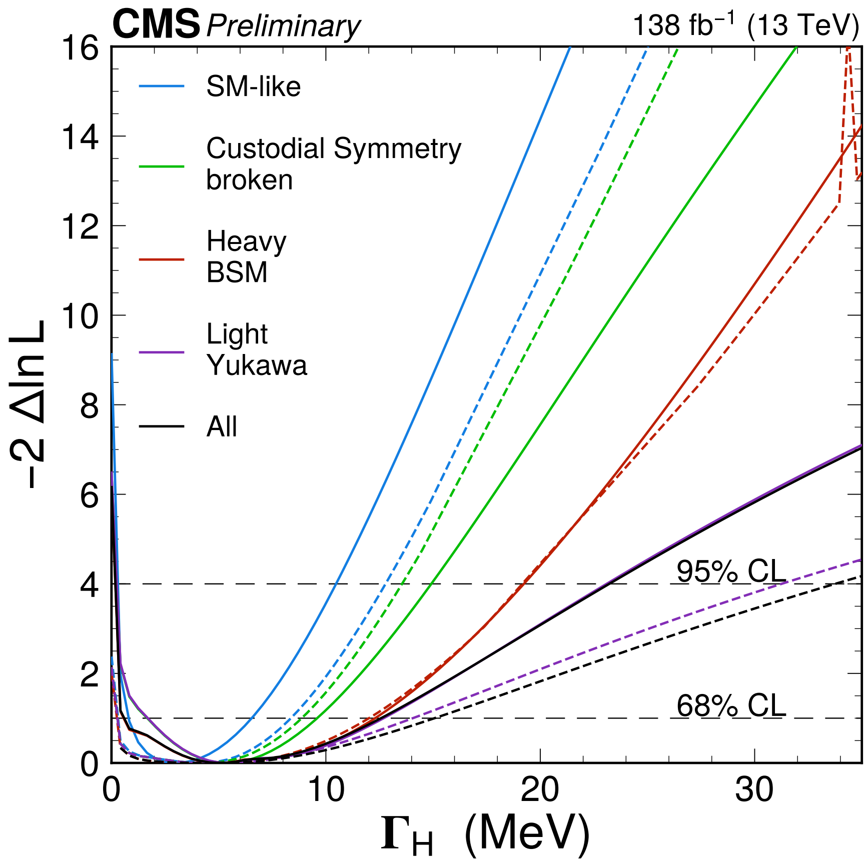


Figure 5.43: The observed (solid) and expected (dashed) values of Γ_H for various sets of anomalous couplings. Blue line represents configuration where κ_t, κ_Z and κ_W are allowed to float while enforcing $\kappa_Z = \kappa_W$ with all couplings fixed to their SM values. Green line additionally removes assumption of custodial symmetry. Red line represents configuration where the terms κ_Q, $\widetilde{\kappa}_t$, and $\widetilde{\kappa}_Q$ are floated in addition to those floated when custodial symmetry is broken. Purple line represents the configuration where κ_u, κ_d, κ_s, κ_c are floated in addition to those floated when custodial symmetry is broken. The black line represents configuration where κ_Z, κ_W and both heavy and light quark couplings are floated with custodial symmetry assumption removed.

Conclusion

This thesis presents the work performed during my PhD studies in the period 2022–2026 within the CMS experiment at the LHC. Following the discovery of the Higgs boson during Run 1, the LHC physics programme has entered an era of high-accuracy measurements, in which the goal is to test the SM with increasing sensitivity and to search for small deviations that could indicate the presence of BSM physics. The large datasets collected during Run 2, together with the ongoing Run 3 data taking, allow Higgs boson properties to be studied with great accuracy, placing strong constraints on possible extensions of the scalar sector.

At the time of writing this thesis, Run 3 is approaching its end, with only a few months of data taking remaining before the start of the next long shutdown. The combined Run 2 and Run 3 datasets represent the most sensitive probe of the Higgs sector so far and mark the transition toward the HL-LHC era.

The analyses presented in this thesis are performed in the $H \to ZZ \to 4\ell$ final state, which has been one of the most powerful channels for Higgs boson studies at the LHC. In this regime, the accuracy of the measurement strongly depends on the quality of the reconstruction and identification of final-state leptons, as well as on the determination of the corresponding efficiency corrections.

In order to reach the level of accuracy required for these studies, a significant part of the work focused on the determination of electron reconstruction and identification efficiencies used in CMS analyses. This includes the training of the electron identification algorithm, the computation of identification efficiency scale factors, and measurements of reconstruction efficiencies at low transverse momentum. These improvements are especially important for the four-lepton final state, where low-p_T electrons contribute to the signal acceptance and where reliable efficiency corrections are necessary to control systematic uncertainties in Higgs measurements.

The main physics result of this thesis is the measurement of the Higgs boson self-coupling within the SMEFT framework, parameterized by the κ_λ modifier, using single-Higgs production. Although the Higgs self-coupling can be accessed directly only through double-Higgs production, single-Higgs processes receive loop-level contributions that allow indirect constraints on the scalar potential. The $H \to ZZ \to 4\ell$ channel provides a clean environment for these studies and is well suited for measurements sensitive to higher-order effects.

As the LHC continues its operation in Run 3 and prepares for the HL-LHC era, the exploration of the Higgs sector will remain one of the central goals of experimental particle physics. Further studies of Higgs boson properties, searches for rare processes, and increasing sensitivity to possible deviations from the SM will be essential for understanding whether the SM provides a complete description of nature or only an effective approximation of a more fundamental theory. The work presented in this thesis contributes a small but meaningful part to this collective effort.

Bibliography

[1] Ulrich Haisch and Gabriël Koole, "Off-shell Higgs production at the LHC as a probe of the trilinear Higgs coupling", in: *Journal of High Energy Physics* 2022.2 (Feb. 2022), ISSN: 1029-8479, DOI: `10.1007/jhep02(2022)030`, URL: `http://dx.doi.org/10.1007/JHEP02(2022)030` (Cited on pages 7, 16, 129, 145, 146, 150, 151, 152, 176).

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", in: *Phys. Lett. B* 716 (2012), pp. 30–61, DOI: `10.1016/j.physletb.2012.08.021`, arXiv: `1207.7235 [hep-ex]` (Cited on pages 23, 48, 111).

[3] 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. B* 716 (2012), pp. 1–29, DOI: `10.1016/j.physletb.2012.08.020`, arXiv: `1207.7214 [hep-ex]` (Cited on pages 23, 48, 111).

[4] Sheldon L. Glashow, "Partial-symmetries of weak interactions", in: *Nucl. Phys.* 22 (1961), pp. 579–588, DOI: `10.1016/0029-5582(61)90469-2` (Cited on page 25).

[5] Steven Weinberg, "A Model of Leptons", in: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266, DOI: `10.1103/PhysRevLett.19.1264` (Cited on page 25).

[6] Abdus Salam, "Weak and Electromagnetic Interactions", in: *Elementary Particle Theory*, Almqvist and Wiksell, 1968 (Cited on page 25).

[7] Particle Data Group, "Review of Particle Physics", in: *Prog. Theor. Exp. Phys.* (2024) (Cited on pages 25, 26, 27, 30, 31, 32, 33, 34, 35, 37, 40, 41, 48).

[8] Francis Halzen and Alan D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Wiley, 1984 (Cited on pages 25, 26, 27, 28, 30, 31).

[9] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, Westview Press, 1995 (Cited on pages 26, 27, 28, 30, 31).

[10] Peter W. Higgs, "Broken symmetries and the masses of gauge bosons", in: *Phys. Rev. Lett.* 13 (1964), pp. 508–509, DOI: `10.1103/PhysRevLett.13.508` (Cited on page 28).

[11] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", in: *Phys. Rev. Lett.* 13 (1964), pp. 321–323 (Cited on page 28).

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", in: *Phys. Rev. Lett.* 13 (1964), pp. 585–587 (Cited on page 28).

[13] Daniel de Florian et al., "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", in: *CERN Yellow Report* (2016), arXiv: `1610.07922 [hep-ph]` (Cited on pages 32, 33, 34, 35, 41, 42).

[14] Stefan Dittmaier et al., “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, in: *CERN Yellow Report* (2011), arXiv: `1101.0593 [hep-ph]` (Cited on page 32).

[15] Abdelhak Djouadi, “The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model”, in: *Phys. Rept.* 457 (2008), pp. 1–216, DOI: `10.1016/j.physrep.2007.10.004`, arXiv: `hep-ph/0503172 [hep-ph]` (Cited on pages 32, 34, 35, 41, 42).

[16] Particle Data Group et al., “Review of Particle Physics”, in: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01, ISSN: 2050-3911, DOI: `10.1093/ptep/ptac097`, eprint: `https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf`, URL: `https://doi.org/10.1093/ptep/ptac097` (Cited on pages 32, 33, 86).

[17] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector”, in: *CERN* (2016), DOI: `10.23731/CYRM-2017-002`, arXiv: `1610.07922 [hep-ph]` (Cited on pages 34, 136, 137, 182).

[18] CMS Collaboration, “A measurement of the Higgs boson mass in the diphoton decay channel”, in: *Physics Letters B* 805 (June 2020), p. 135425, ISSN: 0370-2693, DOI: `10.1016/j.physletb.2020.135425`, URL: `http://dx.doi.org/10.1016/j.physletb.2020.135425` (Cited on page 34).

[19] W. Buchmuller and D. Wyler, “Effective Lagrangian analysis of new interactions and flavor conservation”, in: *Nucl. Phys. B* 268 (1986), pp. 621–653 (Cited on pages 35, 38, 40, 41).

[20] Ilaria Brivio and Michael Trott, “The Standard Model as an Effective Field Theory”, in: *Phys. Rept.* 793 (2019), pp. 1–98 (Cited on pages 35, 37, 40, 41, 42).

[21] Nikolas Kauer and Giampiero Passarino, “Inadequacy of zero-width approximation for a light Higgs boson signal”, in: *JHEP* 08 (2012), p. 116, DOI: `10.1007/JHEP08(2012)116`, arXiv: `1206.4803 [hep-ph]` (Cited on pages 37, 38, 39).

[22] Steven Weinberg, “Baryon and Lepton Nonconserving Processes”, in: *Phys. Rev. Lett.* 43 (1979), pp. 1566–1570, DOI: `10.1103/PhysRevLett.43.1566` (Cited on pages 38, 40).

[23] B. Grzadkowski et al., “Dimension-Six Terms in the Standard Model Lagrangian”, in: *JHEP* 10 (2010), p. 085, DOI: `10.1007/JHEP10(2010)085`, arXiv: `1008.4884 [hep-ph]` (Cited on page 40).

[24] ATLAS Collaboration, “Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} =$ 13 TeV”, in: *Physics Letters B* 843 (Aug. 2023), p. 137745, ISSN: 0370-2693, DOI: `10.1016/j.physletb.2023.137745`, URL: `http://dx.doi.org/10.1016/j.physletb.2023.137745` (Cited on page 43).

[25] CERN, *The History of CERN*, 2024, URL: `https://home.cern/about/who-we-are/our-history` (Cited on page 47).

[26] L. Evans and P. Bryant, “LHC Machine”, in: *JINST* 3 (2008), S08001, DOI: `10.1088/1748-0221/3/08/S08001` (Cited on pages 47, 48, 50).

[27] G. Arnison et al., “Experimental observation of isolated large transverse energy electrons with associated missing energy at sqrt(s)=540 GeV”, in: *Phys. Lett. B* 122 (1983), p. 103 (Cited on page 48).

[28] M. Banner et al., “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy”, in: *Phys. Lett. B* 122 (1983), p. 476 (Cited on page 48).

[29] The LEP Collaborations, “Precision Electroweak Measurements on the Z Resonance”, in: *Phys. Rept.* 427 (2006), p. 257 (Cited on page 48).

[30] ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3”, in: *JINST* 19.05 (2024), 233 pages in total, author list starting page 214, 116 figures, 15 tables, published in JINST. All figures including auxiliary figures are available at http://atlas.web.cern.ch/ 2019-02/, P05063, DOI: `10.1088/1748-0221/19/05/P05063`, arXiv: `2305.16623`, URL: `https://cds.cern.ch/record/2859916` (Cited on page 51).

[31] ALICE Collaboration, *Real-time data processing in the ALICE High Level Trigger at the LHC*, Dec. 2018, DOI: `10.48550/arXiv.1812.08036` (Cited on page 52).

[32] Antimatter Matters, *LHC experiments overview*, Accessed: 2026-03-12, 2024, URL: `http://antimattermatters.s3-website-eu-west-1.amazonaws.com/lhc.html` (Cited on page 52).

[33] *The Phase-2 Upgrade of the CMS Tracker*, tech. rep., Geneva: CERN, 2017, DOI: `10.17181/CERN.QZ28.FLHW`, URL: `https://cds.cern.ch/record/2272264` (Cited on page 57).

[34] CMS Collaboration, *CMS Physics*, Technical design report. CMS, There is an error on cover due to a technical problem for some items, Geneva: CERN, 2006, URL: `https://cds.cern.ch/record/922757` (Cited on page 58).

[35] “The CMS experiment at the CERN LHC”, in: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004, DOI: `10.1088/1748-0221/3/08/S08004`, URL: `https://doi.org/10.1088/1748-0221/3/08/S08004` (Cited on page 60).

[36] Somnath Choudhury, “Performance of the High-Level Trigger System at CMS in LHC Run-2”, in: *IEEE Transactions on Nuclear Science* 68.8 (2021), pp. 2035–2042, DOI: `10.1109/TNS.2021.3087618` (Cited on page 65).

[37] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, in: *JINST* 12 (2017), P10003, DOI: `10.1088/1748-0221/12/10/P10003`, arXiv: `1706.04965` (Cited on page 66).

[38] “Muon Reconstruction and Identification Performance with Run-2 data”, in: (2020), URL: `https://cds.cern.ch/record/2727091` (Cited on page 67).

[39] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, in: *JINST* 13 (2018), P06015, DOI: `10.1088/1748-0221/13/06/P06015`, arXiv: `1804.04528 [physics.ins-det]` (Cited on page 68).

[40] A. Bodek et al., “Extracting muon momentum scale corrections for hadron collider experiments”, in: *Eur. Phys. J. C* 72 (2012), p. 2194, DOI: `10.1140/epjc/s10052-012-2194-8`, arXiv: `1208.3710 [hep-ex]` (Cited on page 68).

[41] M. Cacciari, G. P. Salam, and G. 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] (Cited on page 68).

[42] CMS Collaboration, *Jet energy scale and resolution measurement with Run 2 legacy data collected by CMS at 13 TeV*, tech. rep. CMS-DP-2021/033, CMS Detector Performance Summary, CMS Detector Performance Group, 2021, URL: https://cds.cern.ch/record/2792322 (Cited on page 69).

[43] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, in: *JINST* 13 (2018), P05011, DOI: 10.1088/1748-0221/13/05/P05011, arXiv: 1712.07158 [hep-ex] (Cited on page 69).

[44] J. M. Campbell and R. K. Ellis, “MCFM for the Tevatron and the LHC”, in: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 10–15, DOI: 10.1016/j.nuclphysbps.2010.08.011 (Cited on pages 71, 133).

[45] S. 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 (Cited on page 71).

[46] J. Alwall et al., “MadGraph 5 : Going Beyond”, in: *JHEP* 06 (2011), p. 128, DOI: 10.1007/JHEP06(2011)128 (Cited on page 71).

[47] Andrei V. Gritsan et al., “New features in the JHU generator framework: Constraining Higgs boson properties from on-shell and off-shell production”, in: *Physical Review D* 102.5 (Sept. 2020), ISSN: 2470-0029, DOI: 10.1103/physrevd.102.056022, URL: http://dx.doi.org/10.1103/PhysRevD.102.056022 (Cited on pages 71, 150).

[48] T. Sjostrand et al., “An Introduction to PYTHIA 8.2”, in: *Comput. Phys. Commun.* 191 (2015), pp. 159–177, DOI: 10.1016/j.cpc.2015.01.024 (Cited on page 71).

[49] S. Agostinelli et al., “GEANT4 — a simulation toolkit”, in: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303, DOI: 10.1016/S0168-9002(03)01368-8 (Cited on page 72).

[50] CMS Collaboration, “Measurements of the Higgs boson production cross section in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13.6$ TeV”, in: *Journal of High Energy Physics* 2025.5 (May 2025), ISSN: 1029-8479, DOI: 10.1007/jhep05(2025)079, URL: http://dx.doi.org/10.1007/JHEP05(2025)079 (Cited on pages 73, 102, 122, 124).

[51] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, in: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003, ISSN: 1748-0221, DOI: 10.1088/1748-0221/12/10/p10003, URL: http://dx.doi.org/10.1088/1748-0221/12/10/P10003 (Cited on pages 74, 75).

[52] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, in: *Journal of Instrumentation* 16.05 (May 2021), P05014, ISSN: 1748-0221, DOI: 10.1088/1748-0221/16/05/p05014, URL: http://dx.doi.org/10.1088/1748-0221/16/05/P05014 (Cited on pages 74, 75, 76, 78, 90).

[53] CMS Collaboration, *The CMS trigger system*, Comprehensive description of CMS Level-1 trigger including e/γ algorithms, 2017, DOI: `10.1088/1748-0221/12/01/P01020` (Cited on page 77).

[54] *Performance of the Level-1 Trigger Electrons and Photons in CMS*, CMS DP-2025/084, CERN CDS, Detailed description of L1 e/γ clustering, seed selection and isolation algorithms, 2025 (Cited on page 77).

[55] *CMS High Level Trigger performance*, CMS Public TWiki and performance notes, HLT electron and photon reconstruction and selection strategies, 2025 (Cited on page 77).

[56] CMS Collaboration, "Observation of the diphoton decay of the Higgs boson and measurement of its properties", in: *The European Physical Journal C* 74.10 (Oct. 2014), ISSN: 1434-6052, DOI: `10.1140/epjc/s10052-014-3076-z`, URL: `http://dx.doi.org/10.1140/epjc/s10052-014-3076-z` (Cited on page 77).

[57] CMS Collaboration, in: *Journal of Instrumentation* 10.06 (June 2015), P06005–P06005, ISSN: 1748-0221, DOI: `10.1088/1748-0221/10/06/p06005`, URL: `http://dx.doi.org/10.1088/1748-0221/10/06/P06005` (Cited on page 79).

[58] CMS Collaboration, "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV", in: *Journal of High Energy Physics* 2011.1 (Jan. 2011), ISSN: 1029-8479, DOI: `10.1007/jhep01(2011)080`, URL: `http://dx.doi.org/10.1007/JHEP01(2011)080` (Cited on page 80).

[59] Ana Sculac, *Measurement of electron identification efficiency and search for Higgs boson pair production in the WWZZ decay channel using the CMS detector at the LHC* (Cited on pages 81, 82).

[60] P. Agrawal et al., "Feebly-interacting particles: FIPs 2020 workshop report", in: *The European Physical Journal C* 81.11 (Nov. 2021), ISSN: 1434-6052, DOI: `10.1140/epjc/s10052-021-09703-7`, URL: `http://dx.doi.org/10.1140/epjc/s10052-021-09703-7` (Cited on page 83).

[61] C. Antel et al., *Feebly Interacting Particles: FIPs 2022 workshop report*, 2023, arXiv: `2305.01715 [hep-ph]`, URL: `https://arxiv.org/abs/2305.01715` (Cited on page 83).

[62] CMS Collaboration, "Enriching the physics program of the CMS experiment via data scouting and data parking", in: *Physics Reports* 1115 (Apr. 2025), pp. 678–772, ISSN: 0370-1573, DOI: `10.1016/j.physrep.2024.09.006`, URL: `http://dx.doi.org/10.1016/j.physrep.2024.09.006` (Cited on page 84).

[63] "Recording and reconstructing 10 billion unbiased b hadron decays in CMS", in: (2019), URL: `https://cds.cern.ch/record/2704495` (Cited on page 85).

[64] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System", in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 785–794, DOI: `10.1145/2939672.2939785`, URL: `https://arxiv.org/abs/1603.02754` (Cited on page 90).

[65] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms", in: *Advances in Neural Information Processing Systems*, vol. 25, 2012, URL: https://arxiv.org/abs/1206.2944 (Cited on page 90).

[66] CMS Collaboration, *Run 3 Electron MVA based ID training*, Internal document, CERN Indico, Available at CERN Indico (requires CERN credentials), 2023, URL: https://indico.cern.ch/event/1220628/contributions/5134878/attachments/2546114/4384580/Run%203%20Electron%20MVA%20based%20ID%20training.pdf (Cited on page 91).

[67] "Redefinition of Electron and Photon identification variables for Run-3 in the CMS Experiment", in: (2020), URL: https://cds.cern.ch/record/2724491 (Cited on pages 91, 92).

[68] CMS Collaboration, "Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs", in: *Physical Review Letters* 110.8 (Feb. 2013), ISSN: 1079-7114, DOI: 10.1103/physrevlett.110.081803, URL: http://dx.doi.org/10.1103/PhysRevLett.110.081803 (Cited on page 111).

[69] CMS Collaboration, "Measurement of differential and integrated fiducial cross sections for Higgs boson production in the four-lepton decay channel in pp collisions at $\sqrt{s}$ = 7 and 8 TeV", in: *Journal of High Energy Physics* 2016.4 (Apr. 2016), ISSN: 1029-8479, DOI: 10.1007/jhep04(2016)005, URL: http://dx.doi.org/10.1007/JHEP04(2016)005 (Cited on page 111).

[70] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", in: *Physical Review D* 92.1 (July 2015), ISSN: 1550-2368, DOI: 10.1103/physrevd.92.012004, URL: http://dx.doi.org/10.1103/PhysRevD.92.012004 (Cited on page 111).

[71] CMS Collaboration, "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV", in: *The European Physical Journal C* 81.6 (June 2021), ISSN: 1434-6052, DOI: 10.1140/epjc/s10052-021-09200-x, URL: http://dx.doi.org/10.1140/epjc/s10052-021-09200-x (Cited on page 111).

[72] CMS Collaboration, "Measurements of inclusive and differential cross sections for the Higgs boson production and decay to four-leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV", in: *Journal of High Energy Physics* 2023.8 (Aug. 2023), ISSN: 1029-8479, DOI: 10.1007/jhep08(2023)040, URL: http://dx.doi.org/10.1007/JHEP08(2023)040 (Cited on page 111).

[73] CMS Collaboration, "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state", in: *Physical Review D* 104.5 (Sept. 2021), ISSN: 2470-0029, DOI: 10.1103/physrevd.104.052004, URL: http://dx.doi.org/10.1103/PhysRevD.104.052004 (Cited on page 111).

[74] ATLAS Collaboration, "Measurement of the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ cross-sections in pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS detector", in: *The European Physical Journal C* 84.1 (Jan. 2024), ISSN: 1434-6052, DOI: `10.1140/epjc/s10052-023-12130-5`, URL: `http://dx.doi.org/10.1140/epjc/s10052-023-12130-5` (Cited on page 111).

[75] Andrei V. Gritsan et al., "Constraining anomalous Higgs boson couplings to the heavy-flavor fermions using matrix element techniques", in: *Physical Review D* 94.5 (Sept. 2016), ISSN: 2470-0029, DOI: `10.1103/physrevd.94.055023`, URL: `http://dx.doi.org/10.1103/PhysRevD.94.055023` (Cited on page 119).

[76] Fabio Maltoni et al., "Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC", in: *The European Physical Journal C* 77.12 (Dec. 2017), ISSN: 1434-6052, DOI: `10.1140/epjc/s10052-017-5410-8`, URL: `http://dx.doi.org/10.1140/epjc/s10052-017-5410-8` (Cited on pages 123, 125).

[77] G. Degrassi et al., "Probing the Higgs self coupling via single Higgs production at the LHC", in: *Journal of High Energy Physics* 2016.12 (Dec. 2016), ISSN: 1029-8479, DOI: `10.1007/jhep12(2016)080`, URL: `http://dx.doi.org/10.1007/JHEP12(2016)080` (Cited on pages 125, 126, 147).

[78] Alessandro Tarabini, *Measurement of Higgs boson fiducial cross sections with the CMS detector and electromagnetic reconstruction with the high-granularity endcap calorimeter*, May 2024 (Cited on pages 125, 127).

[79] CMS Collaboration, "Search for nonresonant Higgs boson pair production in the four leptons plus twob jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", in: *Journal of High Energy Physics* 2023.6 (June 2023), ISSN: 1029-8479, DOI: `10.1007/jhep06(2023)130`, URL: `http://dx.doi.org/10.1007/JHEP06(2023)130` (Cited on page 128).

[80] CMS Collaboration, "Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV", in: *Journal of High Energy Physics* 2023.7 (July 2023), p. 095, DOI: `10.1007/JHEP07(2023)095` (Cited on page 128).

[81] CMS Collaboration, "Search for Nonresonant Pair Production of Highly Energetic Higgs Bosons Decaying to Bottom Quarks", in: *Physical Review Letters* 131.4 (July 2023), ISSN: 1079-7114, DOI: `10.1103/physrevlett.131.041803`, URL: `http://dx.doi.org/10.1103/PhysRevLett.131.041803` (Cited on page 128).

[82] CMS Collaboration, "Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV", in: *Physical Review Letters* 129.8 (Aug. 2022), p. 081802, DOI: `10.1103/PhysRevLett.129.081802` (Cited on page 128).

[83] CMS Collaboration, "Search for nonresonant Higgs boson pair production in the final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", in: *Physics Letters B* 842 (July 2023), p. 137531, DOI: `10.1016/j.physletb.2022.137531` (Cited on pages 128, 176).

[84] CMS Collaboration, “Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, in: *Journal of High Energy Physics* 2021.3 (Mar. 2021), ISSN: 1029-8479, DOI: `10.1007/jhep03(2021)257`, URL: `http://dx.doi.org/10.1007/JHEP03(2021)257` (Cited on page 128).

[85] Joshua P. Ellis, “TikZ-Feynman: Feynman diagrams with TikZ”, in: *Computer Physics Communications* 210 (Jan. 2017), pp. 103–123, ISSN: 0010-4655, DOI: `10.1016/j.cpc.2016.08.019`, URL: `http://dx.doi.org/10.1016/j.cpc.2016.08.019` (Cited on page 129).

[86] Torbjörn Sjöstrand et al., “An introduction to PYTHIA 8.2”, in: *Comput. Phys. Commun.* 191 (2015), p. 159, DOI: `10.1016/j.cpc.2015.01.024`, arXiv: `1410.3012 [hep-ph]` (Cited on page 130).

[87] Vardan Khachatryan et al., “Event generator tunes obtained from underlying event and multiparton scattering measurements”, in: *Eur. Phys. J. C* 76 (2016), p. 155, DOI: `10.1140/epjc/s10052-016-3988-x`, arXiv: `1512.00815 [hep-ex]` (Cited on page 130).

[88] Albert M Sirunyan et al., “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, in: *Eur. Phys. J. C* 80 (2020), p. 4, DOI: `10.1140/epjc/s10052-019-7499-4`, arXiv: `1903.12179 [hep-ex]` (Cited on page 130).

[89] Richard 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]` (Cited on page 130).

[90] 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` (Cited on page 130).

[91] CMS Collaboration, “Measurement of the Higgs boson mass and width using the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$”, in: *Physical Review D* 111.9 (May 2025), p. 092014, ISSN: 2470-0029, DOI: `10.1103/PhysRevD.111.092014`, URL: `https://doi.org/10.1103/PhysRevD.111.092014` (Cited on pages 134, 137).

[92] John M. Campbell, R. Keith Ellis, and Ciaran Williams, “Vector boson pair production at the LHC”, in: *JHEP* 07 (2011), p. 018, DOI: `10.1007/JHEP07(2011)018`, arXiv: `1105.0020 [hep-ph]` (Cited on page 133).

[93] John M. Campbell, R. Keith Ellis, and Ciaran Williams, “Bounding the Higgs width at the LHC using full analytic results for $gg \to e^-e^+\mu^-\mu^+$”, in: *JHEP* 04 (2014), p. 060, DOI: `10.1007/JHEP04(2014)060`, arXiv: `1311.3589 [hep-ph]` (Cited on page 133).

[94] John M. Campbell and R. Keith Ellis, “Higgs constraints from vector boson fusion and scattering”, in: *JHEP* 04 (2015), p. 030, DOI: `10.1007/JHEP04(2015)030`, arXiv: `1502.02990 [hep-ph]` (Cited on page 133).

[95] Yanyan Gao et al., “Spin determination of single-produced resonances at hadron colliders”, in: *Phys. Rev. D* 81 (2010), p. 075022, DOI: `10.1103/PhysRevD.81.075022`, arXiv: `1001.3396 [hep-ph]` (Cited on pages 133, 150).

[96] Sara Bolognesi et al., “Spin and parity of a single-produced resonance at the LHC”, in: *Phys. Rev. D* 86 (2012), p. 095031, DOI: `10.1103/PhysRevD.86.095031`, arXiv: `1208.4018 [hep-ph]` (Cited on pages 133, 150).

[97] Ian Anderson et al., “Constraining anomalous HVV interactions at proton and lepton colliders”, in: *Phys. Rev. D* 89 (2014), p. 035007, DOI: `10.1103/PhysRevD.89.035007`, arXiv: `1309.4819 [hep-ph]` (Cited on pages 133, 150).

[98] Andrei V. Gritsan et al., “Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques”, in: *Phys. Rev. D* 94 (2016), p. 055023, DOI: `10.1103/PhysRevD.94.055023`, arXiv: `1606.03107 [hep-ph]` (Cited on pages 133, 150).

[99] Andrei V. Gritsan et al., “New features in the JHU generator framework: constraining Higgs boson properties from on-shell and off-shell production”, in: *Phys. Rev. D* 102 (2020), p. 056022, DOI: `10.1103/PhysRevD.102.056022`, arXiv: `2002.09888 [hep-ph]` (Cited on page 133).

[100] Stefano Catani and Massimiliano Grazzini, “Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and its Application to Higgs-Boson Production at the Large Hadron Collider”, in: *Phys. Rev. Lett.* 98 (22 May 2007), p. 222002, DOI: `10.1103/PhysRevLett.98.222002`, URL: `https://link.aps.org/doi/10.1103/PhysRevLett.98.222002` (Cited on page 136).

[101] Fabrizio Caola et al., “QCD corrections to ZZ production in gluon fusion at the LHC”, in: *Phys. Rev. D* 92.9 (2015), p. 094028, DOI: `10.1103/PhysRevD.92.094028`, URL: `https://doi.org/10.1103/PhysRevD.92.094028` (Cited on page 136).

[102] Kirill Melnikov and Matthew Dowling, “Production of two Z-bosons in gluon fusion in the heavy top quark approximation”, in: *Phys. Lett. B* 744 (2015), p. 43, DOI: `10.1016/j.physletb.2015.03.030`, arXiv: `1503.01274 [hep-ph]` (Cited on page 136).

[103] John M. Campbell et al., “Two loop correction to interference in gg → ZZ”, in: *Journal of High Energy Physics* 2016.8 (Aug. 2016), ISSN: 1029-8479, DOI: `10.1007/jhep08(2016)011`, URL: `http://dx.doi.org/10.1007/JHEP08(2016)011` (Cited on page 136).

[104] Fabrizio Caola et al., “QCD corrections to vector boson pair production in gluon fusion including interference effects with off-shell Higgs at the LHC”, in: *JHEP* 07 (2016), p. 087, DOI: `10.1007/JHEP07(2016)087`, arXiv: `1605.04610 [hep-ph]` (Cited on page 136).

[105] CMS Collaboration, “Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs”, in: *Physics Letters B* 736 (2014), pp. 64–85, ISSN: 0370-2693, DOI: `https://doi.org/10.1016/j.physletb.2014.06.077` (Cited on pages 136, 180).

[106] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev, “ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD”, in: *Phys. Lett. B* 750 (2015), pp. 407–410, DOI: `10.1016/j.physletb.2015.09.055`, URL: `https://www.sciencedirect.com/science/article/pii/S0370269315007303` (Cited on page 137).

[107] Anastasiya Bierweiler, Tobias Kasprzik, and Johann H. Kühn, “Vector-boson pair production at the LHC to $\mathcal{O}(\alpha^3)$ accuracy”, in: *JHEP* 2013.12 (2013), p. 071, DOI: `10.1007/JHEP12(2013)071`, URL: `https://doi.org/10.1007/JHEP12(2013)071` (Cited on page 137).

[108] Lucas Kang, *Measuring properties of the Higgs boson utilizing its off-shell production*, Aug. 2025 (Cited on page 138).

[109] Jeffrey Davis et al., *Maximizing Returns: Optimizing Experimental Observables at the LHC*, 2026, arXiv: `2601.10822 [hep-ph]`, URL: `https://arxiv.org/abs/2601.10822` (Cited on page 138).

[110] Nonso Nnamoko and Ioannis Korkontzelos, “Efficient treatment of outliers and class imbalance for diabetes prediction”, in: *Artificial Intelligence in Medicine* 104 (2020), p. 101815, ISSN: 0933-3657, DOI: `https://doi.org/10.1016/j.artmed.2020.101815`, URL: `https://www.sciencedirect.com/science/article/pii/S093336571830681X` (Cited on page 141).

[111] Gian Francesco Giudice et al., “The strongly-interacting light Higgs”, in: *Journal of High Energy Physics* 2007.06 (June 2007), pp. 045–045, ISSN: 1029-8479, DOI: `10.1088/1126-6708/2007/06/045`, URL: `http://dx.doi.org/10.1088/1126-6708/2007/06/045` (Cited on page 145).

[112] Martin Gorbahn and Ulrich Haisch, “Indirect probes of the trilinear Higgs coupling: $gg \to h$ and $h \to \gamma\gamma$”, in: *Journal of High Energy Physics* 2016.10 (Oct. 2016), ISSN: 1029-8479, DOI: `10.1007/jhep10(2016)094`, URL: `http://dx.doi.org/10.1007/JHEP10(2016)094` (Cited on page 147).

[113] Wojciech Bizoń, Ulrich Haisch, and Luca Rottoli, “Constraints on the quartic Higgs self-coupling from double-Higgs production at future hadron colliders”, in: *Journal of High Energy Physics* 2019.10 (Oct. 2019), ISSN: 1029-8479, DOI: `10.1007/jhep10(2019)267`, URL: `http://dx.doi.org/10.1007/JHEP10(2019)267` (Cited on page 147).

[114] Sophia Borowka et al., “Probing the scalar potential via double Higgs boson production at hadron colliders”, in: *Journal of High Energy Physics* 2019.4 (Apr. 2019), ISSN: 1029-8479, DOI: `10.1007/jhep04(2019)016`, URL: `http://dx.doi.org/10.1007/JHEP04(2019)016` (Cited on page 147).

[115] T. Hahn and M. Perez-Victoria, “Automated one-loop calculations in four and D dimensions”, in: *Computer Physics Communications* 118.2–3 (May 1999), pp. 153–165, ISSN: 0010-4655, DOI: `10.1016/s0010-4655(98)00173-8`, URL: `http://dx.doi.org/10.1016/S0010-4655(98)00173-8` (Cited on pages 147, 148).

[116] Wojciech Bizoń et al., “Constraints on the trilinear Higgs coupling from vector boson fusion and associated Higgs production at the LHC”, in: *Journal of High Energy Physics* 2017.7 (July 2017), ISSN: 1029-8479, DOI: `10.1007/jhep07(2017)083`, URL: `http://dx.doi.org/10.1007/JHEP07(2017)083` (Cited on pages 147, 148).

[117] Radja Boughezal et al., “Color-singlet production at NNLO in MCFM”, in: *The European Physical Journal C* 77.1 (Dec. 2016), ISSN: 1434-6052, DOI: `10.1140/epjc/s10052-016-4558-y`, URL: `http://dx.doi.org/10.1140/epjc/s10052-016-4558-y` (Cited on page 150).

[118] Glen Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, in: *European Physical Journal C* 71 (2011), p. 1554, DOI: `10.1140/epjc/s10052-011-1554-0`, arXiv: `1007.1727 [physics.data-an]` (Cited on page 170).

[119] Samuel S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, in: *Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62, DOI: `10.1214/aoms/1177732360` (Cited on page 170).

[120] Aram Hayrapetyan et al., “The CMS statistical analysis and combination tool: COMBINE”, in: *Comput. Softw. Big Sci.* 8 (2024), p. 19, DOI: `10.1007/s41781-024-00121-4`, arXiv: `2404.06614 [physics.data-an]` (Cited on page 170).

[121] ATLAS Collaboration, “Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV pp collisions with the ATLAS detector”, in: *Reports on Progress in Physics* 88.5 (2025), p. 057803, DOI: `10.1088/1361-6633/adcd9a` (Cited on pages 170, 171).

[122] Fabrizio Caola and Kirill Melnikov, “Constraining the Higgs boson width with ZZ production at the LHC”, in: *Phys. Rev. D* 88 (5 Sept. 2013), p. 054024, DOI: `10.1103/PhysRevD.88.054024`, URL: `https://link.aps.org/doi/10.1103/PhysRevD.88.054024` (Cited on pages 180, 181).