



HAL
open science

The measurement of the Higgs Boson properties and the timing calibration of the CMS detector using machine learning techniques

Victor Lohezic

► To cite this version:

Victor Lohezic. The measurement of the Higgs Boson properties and the timing calibration of the CMS detector using machine learning techniques. High Energy Physics - Experiment [hep-ex]. Université Paris-Saclay, 2023. English. NNT : 2023UPASP113 . tel-04338145

HAL Id: tel-04338145

<https://theses.hal.science/tel-04338145>

Submitted on 12 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The measurement of the Higgs Boson properties and the timing calibration of the CMS detector using machine learning techniques

*La mesure des propriétés du boson de Higgs et
l'étalonnage temporel du détecteur CMS à l'aide de
méthodes d'apprentissage automatique*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, particules hadrons énergie et noyau : instrumentation,
imagerie, cosmos et simulation (PHENIICS)

Spécialité de doctorat : Physique des particules

Graduate School : Physique. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche

Département de Physique des Particules (Université Paris-Saclay, CEA),

sous la direction de **Fabrice COUDERC**, Ingénieur-Chercheur CEA,

et la co-direction de **Mehmet-Özgür SAHIN**, Ingénieur-Chercheur CEA

Thèse soutenue à Paris-Saclay, le 28 Septembre 2023, par

Victor LOHEZIC

Composition du jury

Membres du jury avec voix délibérative

Anne-Catherine LE BIHAN

Directrice de recherche, IPHC, CNRS - Université de Strasbourg

Lucia DI CIACCIO

Professeure des universités, LAPP, Université Savoie Mont Blanc

José OCARIZ

Professeur des universités, LPNHE, Université Paris Cité

Emanuele DI MARCO

Chargé de recherche, INFN - Sapienza University of Rome

Jean-Roch VLIMANT

Chargé de recherche, California Institute of Technology

Présidente

Rapporteur & Examinatrice

Rapporteur & Examineur

Examineur

Examineur

Titre : La mesure des propriétés du boson de Higgs et l'étalonnage temporel du détecteur CMS à l'aide de méthodes d'apprentissage automatique

Mots clés : LHC, Jouvence CMS, Higgs, Distribution d'horloge, Apprentissage automatique, Intelligence artificielle

Après l'observation du boson de Higgs par les expériences ATLAS et CMS en 2012, les mesures de précision de ses propriétés sont aujourd'hui un des enjeux majeurs de la physique des hautes énergies et du Large Hadron Collider (LHC). En effet, il s'agit de tester la compatibilité de ce boson avec celui attendu par le modèle standard (MS) de la physique des particules. Dans son canal de désintégration en deux photons ($H \rightarrow \gamma\gamma$), le boson de Higgs est entièrement reconstruit, le pic de masse correspondant pouvant être mesuré avec une très bonne résolution expérimentale (autour de 1%). En conséquence, en dépit d'un taux d'embranchement très faible dans le MS (d'environ 0.2%), le canal $H \rightarrow \gamma\gamma$ fut l'un des deux canaux ayant permis la découverte du boson de Higgs, le canal de désintégration en quatre leptons étant le second. Cette thèse pose des contraintes sur couplages anormaux (CA) du boson de Higgs avec des bosons de jauge. Un classificateur en multiples catégories basé sur des méthodes d'apprentissage profond (deep learning) est développé pour utiliser l'ensemble des informations disponibles dans l'analyse $H \rightarrow \gamma\gamma$ et pour fournir la meilleure séparation possible entre le bruit de fond, les différents modes de production du boson de Higgs du MS et les productions CA du boson de Higgs.

Un bruit de fond conséquent pour les analyses $H \rightarrow \gamma\gamma$ vient des processus QCD produisant une paire diphoton. Même les événements avec seulement un, voire aucun pho-

ton, contribuent grandement à la contamination du signal si d'autres particules sont fausement identifiées comme des photons. De ce fait, une estimation précise du bruit de fond émergent de ces photons mal identifiés est nécessaire pour atteindre une extraction optimale du signal. Cette thèse décrit une nouvelle méthode pour l'estimation précise du bruit de fond. Cette méthode s'appuie sur des modèles d'apprentissage profond avancés appelés réseaux antagonistes génératifs (ou GAN), pour générer des photons mal identifiés et améliorer la description du bruit de fond associé grâce à des régions de contrôle définies dans les données.

D'autre part, le LHC subira dans les prochaines années une jouvence permettant d'augmenter sa luminosité (High Luminosity LHC, HL-LHC) d'un facteur 10 environ. En contrepartie, les conditions de prise de données seront beaucoup plus difficiles. En conséquence, le détecteur CMS sera également amélioré (jouvence Phase II) pour faire face à ces conditions. La possibilité d'associer à chaque objet reconstruit dans la collision un temps mesuré avec une grande précision constitue un enjeu majeur qui permettra d'améliorer la qualité des différentes mesures réalisées dans le canal $H \rightarrow \gamma\gamma$. Cette thèse fournit une contribution aux mesures de temps de haute résolution envisagées par CMS, en particulier sur la surveillance et la calibration ultra rapide du système de distribution d'horloge.

Title: The measurement of the Higgs Boson properties and the timing calibration of the CMS detector using machine learning techniques

Keywords: LHC, CMS upgrade, Higgs, Timing, Deep learning, Artificial intelligence

After the observation of a Higgs boson which is compatible with the predictions of the standard model (SM) of particle physics at the ATLAS and CMS detectors in 2012, the precise measurement of its properties is now one of the primary goals of high energy physics. The Higgs boson decaying into two photons ($H \rightarrow \gamma\gamma$ decay channel) provides a fully reconstructed final state and its invariant mass peak can be measured with a very good mass resolution (around 1%). Consequently, despite the small branching ratio predicted by the SM (approximately 0.2%), $H \rightarrow \gamma\gamma$ was one of the two most essential channels involved in the discovery of the Higgs boson together with its decay to four leptons. This PhD thesis establishes constraints on the Higgs boson anomalous couplings (AC) to gauge bosons. A multiclassifier based on a deep learning model is designed to use all possible ingredients of the $H \rightarrow \gamma\gamma$ analyses to provide the most optimal separation between background, SM production and AC production of the Higgs boson.

Significant backgrounds to the $H \rightarrow \gamma\gamma$ analysis originate from QCD-induced production of diphoton, or diphoton-like, pairs. Processes producing only one or no photon con-

tribute significantly to the contamination of signal if other particles are misidentified as photons. As such, a precise estimation of the background emerging from misidentified photons is necessary to reach an optimal signal extraction. This thesis describes a novel method relying on advanced machine learning models named generative adversarial networks or GAN to generate misidentified photons and improve the description of such backgrounds from data control regions.

Furthermore, the LHC will undergo a High Luminosity (HL) upgrade, delivering around ten times more integrated luminosity with the downside of imposing harsher conditions on the CMS detector. An accompanying upgrade of the CMS detector (Phase II upgrade) is foreseen to not only cope with these harsher conditions but also significantly improve the performance of the detector. One of the most critical aspects of this upgrade is the ability to tag events with very high timing resolution, which will also improve the study of the $H \rightarrow \gamma\gamma$ decay channel. This thesis provides a contribution to the timing upgrade of the CMS detector, particularly to the fast monitoring and calibration of the high-precision clock distribution.

Remerciements

J'ai souvent entendu que les petits ruisseaux font les grandes rivières; l'accomplissement de cette thèse en est une preuve supplémentaire et je tiens à remercier ici toutes ces contributions qui ont rendu ce travail possible au cours des trois dernières années.

Mes premiers remerciements iront à l'Université Paris-Saclay, à l'école doctorale PHENIICS, ainsi qu'au DPhP de l'Irfu pour m'avoir accueilli et m'avoir permis de réaliser cette thèse dans de bonnes conditions. J'ai une pensée particulière pour Georges Vasseur et Martine Oger qui s'appliquent à faciliter la vie des doctorants du DPhP notamment grâce à leur bienveillance.

Ce manuscrit a été révisé par Anne-Catherine Le Bihan, présidente du jury, Emanuele Di Marco et Jean-Roch Vlimant, examinateurs et avec une attention toute particulière par les rapporteurs Lucia Di Ciaccio et José Ocariz. Merci à eux d'avoir accepté cette tâche, merci pour leurs commentaires et pour nos interactions.

L'un de mes plus grands remerciements va naturellement à ma sacrée équipe de supervision. Thanks Özgür for your infectious passion for machine learning which was one of the reasons I set my mind on this PhD topic. Merci Fabrice pour tes cours de physique des particules d'abord, pour tous tes excellents conseils ensuite et enfin pour ta relecture plus qu'attentive de ce manuscrit. Merci Julie pour m'avoir guidé dans la découverte de certains arcanes de l'analyse $H \rightarrow \gamma\gamma$ de CMS (flashgg...) et pour ton suivi tout en étant une super cheffe de groupe. Merci à vous trois pour votre engagement de tous les instants et pour avoir su me pousser à toujours donner le meilleur de moi-même.

J'ai aussi eu la chance, pendant trois ans, de m'enrichir du contact de physiciens et physiciennes passionnants, en particulier dans le groupe CMS. Merci Marc, Gautier, Philippe, Federico, Serguei, Fabrice, Irakli, Florent et au reste du groupe pour votre accueil. Mais je n'oublie pas non plus ceux du groupe LISA qui m'ont aidé à garder l'esprit ouvert sur d'autres domaines de la physique et qui ont aussi su être de bon conseil donc merci Antoine et merci Marc (encore).

Ces trois années n'auraient certainement pas été aussi agréables sans mes collègues non-permanents du DPhP. Un grand merci d'abord aux habitués du *Friday lunch*. Je remercie aussi chaleureusement tous les adeptes de la pause café (ou fruit) qui ont rendu de longues journées plus digestes, en commençant par

mes aînés : Jelena, Chi-Hsun, Chiara et Quentin, et sans oublier ceux à qui il reste encore une année de labeur : Alexandre et Chantal. Bon courage à vous deux pour la fin mais je ne m'inquiète pas trop. Enfin, ce fut un vrai plaisir de partager mon bureau avec Polina. Je pense qu'on a formé une bonne équipe qui se serre les coudes et ça m'a bien aidé à garder le moral, merci.

Je n'oublie évidemment pas l'assemblée, que dis-je, la joyeuse troupe qui m'accompagne depuis le magistère et ceux qui nous ont rejoints depuis. Merci de tout coeur à Alexandre, Alexis C., Alexis F., Alice et Xavier. C'est une nouvelle saison qui s'achève pour nous mais vous faites partie des raisons pour lesquelles je pourrais revivre ces trois années sans hésiter. J'ai une pensée affectueuse pour mes amis de longue date également. Merci Raphaëlle pour ton amitié depuis toutes ces années, elle compte beaucoup pour moi. Merci George-Yann de t'être assuré qu'on garde contact pendant ces années bien occupées... Ça m'aurait manqué de rater ça et je sais que j'en suis ici notamment grâce aux moments qu'on a partagés en grandissant.

Je remercie infiniment ma famille dont le soutien m'a porté probablement plus que je ne sais le dire. Avoir cette famille soudée est une chance, merci à mes grands-parents, oncles, tantes, parrain, marraine, cousins, cousines. J'ajoute bien sûr une mention spéciale pour mes parents qui m'ont appris tant de valeurs si importantes et qui ont su, chacun à leur façon, me faire comprendre qu'ils seraient toujours à mes côtés. Mille mercis aussi à mes chers petits frère et soeurs, vous me donnez chaque jour une raison de vouloir vous rendre fiers. Merci Maman, Augustin, Marie-Lou, Papa, Valérie, Yuna!

Enfin et surtout, je remercie Marion de m'avoir toléré tout ce temps. Merci pour les rires, pour le soutien et pour ton aide (au-delà de ce que j'aurais osé te demander). Je suis très heureux d'avoir pu vivre cette aventure à tes côtés et j'ai hâte de découvrir les prochaines!

Résumé en français

Dix ans après la découverte du boson de Higgs, nos connaissances sur ses propriétés se sont sensiblement étoffées grâce aux performances excellentes du grand collisionneur de hadrons (LHC) du CERN ainsi qu'à celles de ses détecteurs. Les collaborations ATLAS et CMS ont annoncé la découverte d'une particule pouvant correspondre au boson de Higgs en 2012, après deux ans de prise de données venant de collisions proton-proton avec une énergie du centre de masse de 7 TeV et 8 TeV. La désintégration du boson de Higgs en deux photons ($H \rightarrow \gamma\gamma$) a fait partie de ses canaux de désintégration offrant la sensibilité la plus fine au moment de la découverte. La désintégration $H \rightarrow \gamma\gamma$ a un rapport d'embranchement assez faible, autour de 0.2%, mais son état final est entièrement reconstruit, avec une résolution proche de 1% sur la masse invariante du système diphoton. Depuis sa découverte, environ 13 fois plus de données ont été collectées à des énergies plus élevées et cette particule a été établie comme étant le boson de Higgs prédit par le modèle standard (MS) de la physique des particules. Les mesures de sa masse, de son spin, et de ses couplages ont été réalisées avec une grande précision, toutes confirmant la robustesse du MS. Augmenter la quantité de données accessibles crée une opportunité remarquable pour pousser encore plus loin notre compréhension du boson de Higgs et, pour exploiter cette opportunité au maximum, les équipes mettant en place des analyses pour les expériences du LHC fournissent un effort constant pour améliorer la conception des stratégies d'analyse.

Parallèlement, au cours de la dernière décennie, le domaine de l'apprentissage automatique (ML, de l'anglais "machine learning") a connu une croissance impressionnante grâce au développement d'outils de plus en plus performants, tels que les nombreuses variétés de réseaux de neurones profonds. Ces méthodes avancées se montrent particulièrement efficaces lorsqu'elles sont entraînées avec une grande quantité de données. De ce fait, elles deviennent un choix naturel pour les communautés de physique des hautes énergies lorsqu'il s'agit d'extraire l'information venant d'un signal enfoui sous une vaste quantité de bruit de fond.

Comme l'analyse $H \rightarrow \gamma\gamma$ fournit un accès essentiel aux propriétés du boson de Higgs, cette thèse a pour but d'améliorer les performances de l'analyse $H \rightarrow \gamma\gamma$ de l'expérience CMS en apportant, à différentes étapes de l'analyse, des méthodes basées sur l'apprentissage automatique profond. Une première méthode est proposée dans le Chapitre IV pour améliorer la description du bruit de fond de l'analyse $H \rightarrow \gamma\gamma$ grâce à des réseaux antagonistes génératifs (GAN). Les sources principales de bruit de fond de l'analyse $H \rightarrow \gamma\gamma$ sont les processus physiques produisant deux photons dans leur état final ou des particules (souvent des jets) faussement identifiées en tant que photons. Pour mieux différencier ces deux types de photons, chaque photon dans un événement se voit attribuer un score appelé le photonID, qui représente sa probabilité d'être un photon correctement identifié. Ce score permet de réaliser une sélection visant les événements dont deux photons ont un photonID élevé afin de construire une région enrichie en signal. De façon similaire, une région de contrôle peut être définie en inversant la sélection sur l'un des photonID. Cette région de contrôle est donc enrichie en bruit de fond et donne accès à une description du bruit de fond plus fidèle que les simulations Monte Carlo (MC) actuellement disponibles. Cependant, en inversant la sélection sur l'un des photonID, celui-ci a une valeur qui n'est pas dans le bon intervalle si l'on cherche à estimer le bruit de fond de la région de signal. Le rôle du GAN est de générer de nouveaux photons faussement identifiés (dont un nouveau photonID) pour offrir une estimation du bruit de fond dans la région de signal. Une stratégie d'évaluation des performances du GAN spécifique à notre méthode a été mise en place. Cette stratégie permet de tester plusieurs configurations d'entraînement et d'optimiser les hyperparamètres du GAN pour atteindre les meilleures performances. Finalement, le GAN se montre capable d'apprendre comment générer des photons faussement identifiés convaincants et corrélés correctement avec le reste de l'événement, comme illustré en Fig. 1. De plus, une comparaison de la méthode basée sur le GAN avec la méthode utilisée dans les dernières analyses $H \rightarrow \gamma\gamma$ publiées montre que des arbres de décision boostés (BDTs) entraînés avec l'échantillon de bruit de fond généré par le GAN atteignent de meilleures performances lors de la séparation du signal et du bruit de fond de l'analyse $H \rightarrow \gamma\gamma$.

L'une des propriétés du boson de Higgs étudiée dans cette thèse est son comportement sous la symétrie charge-parité (CP). Comme le boson de Higgs est supposé avoir des interactions CP paires, tous couplages pointant vers un comportement différent (couplages CP impairs par exemple) sont considérés anormaux. De tels couplages indiqueraient une nouvelle source de violation CP dans le MS (en plus des interactions électrofaibles) et une fraction (f_{a3}) des événements de signal observés contiendrait un boson de Higgs avec un comportement CP impair. Cette thèse présente les contraintes attendues sur de tels couplages anormaux en sondant les couplages du boson de Higgs avec des bosons de jauge électrofaibles (boson W ou Z). L'analyse complète englobe plusieurs modes de productions du boson de Higgs qui sont sensibles à ses couplages avec le boson W ou Z : sa production par fusion de bosons vecteurs (VBF) et sa production en association avec un boson vecteur (VH). Cette thèse présente la

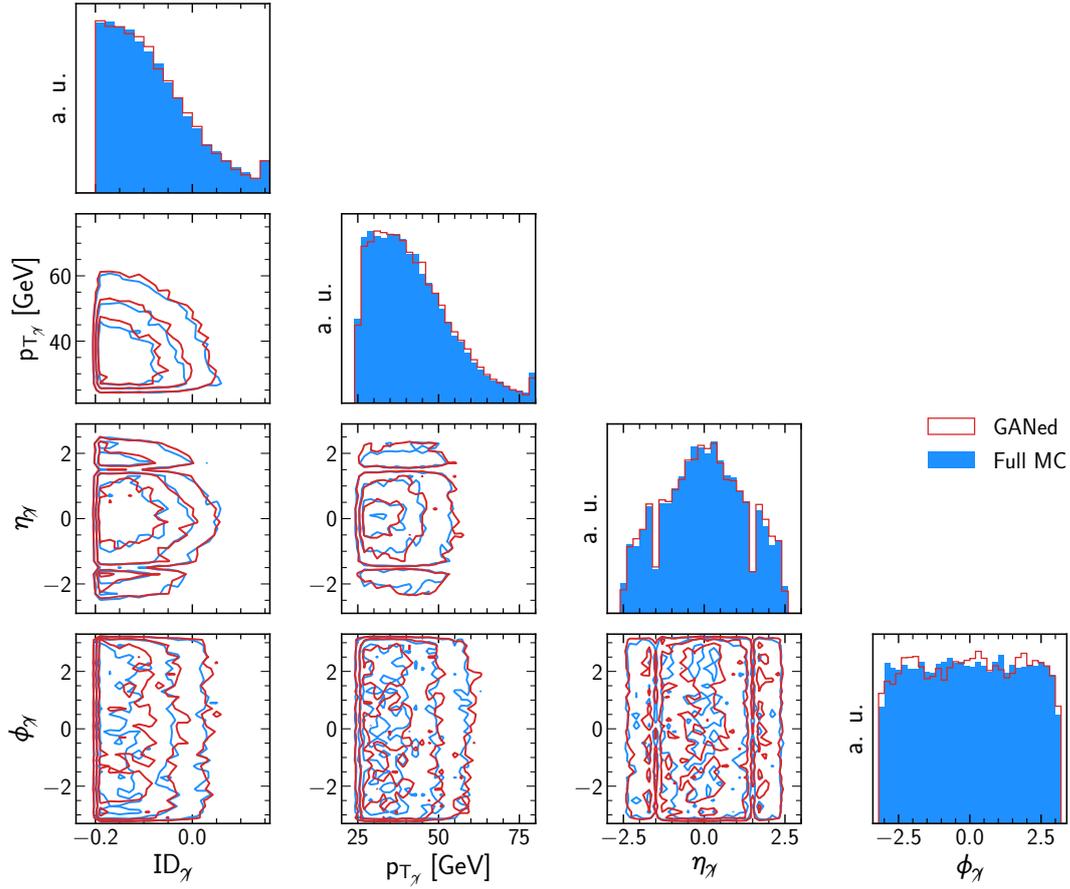


Figure 1 – Diagonale : Comparaison entre les distributions des propriétés des photons générés par le GAN depuis la région de contrôle ("GANed" en rouge) et les distributions décrites par la région de signal de la simulation MC ("Full MC" en bleu). Les propriétés montrées sont le photonID ID_γ , la quantité de mouvement transverse $p_{T,\gamma}$, la pseudorapidité η_γ et l'angle azimutal ϕ_γ du photon généré. Hors diagonale : Courbes de densités qui contiennent 25 %, 50 % et 75 % des évènements et montrant les corrélations entre paires de propriétés.

structure générale de l'analyse en portant une attention particulière à ma contribution : la mise en place d'une stratégie pour étudier la production VH où V est un boson W ou Z qui se désintègre en une paire de quarks ($V_{\text{had}}H$). Pour mettre en évidence les différences entre certains types d'évènements, nous mettons en oeuvre un réseau de neurones profond (DNN) chargé de séparer les évènements entre plusieurs classes : bruit de fond, signal avec couplages attendus par le MS, signal avec couplages anormaux. Les scores obtenus par les évènements en sortie du DNN permettent de les classifier selon plusieurs catégories pour améliorer la sensibilité de l'analyse. La définition de ces catégories est optimisée en minimisant une estimation de la limite supérieure de f_{a3} . Enfin, un test statistique est calculé pour déterminer la valeur de f_{a3} en accord avec les données analysées. Cette analyse utilise l'ensemble des données collectées entre 2016 et 2018 par le détecteur CMS. Cela correspond à 137 fb^{-1} de données provenant de collisions proton-proton à $\sqrt{s} = 13 \text{ TeV}$. Cette analyse et ses résultats préliminaires sont exposés dans le Chapitre V; le scan du test statistique est rapporté

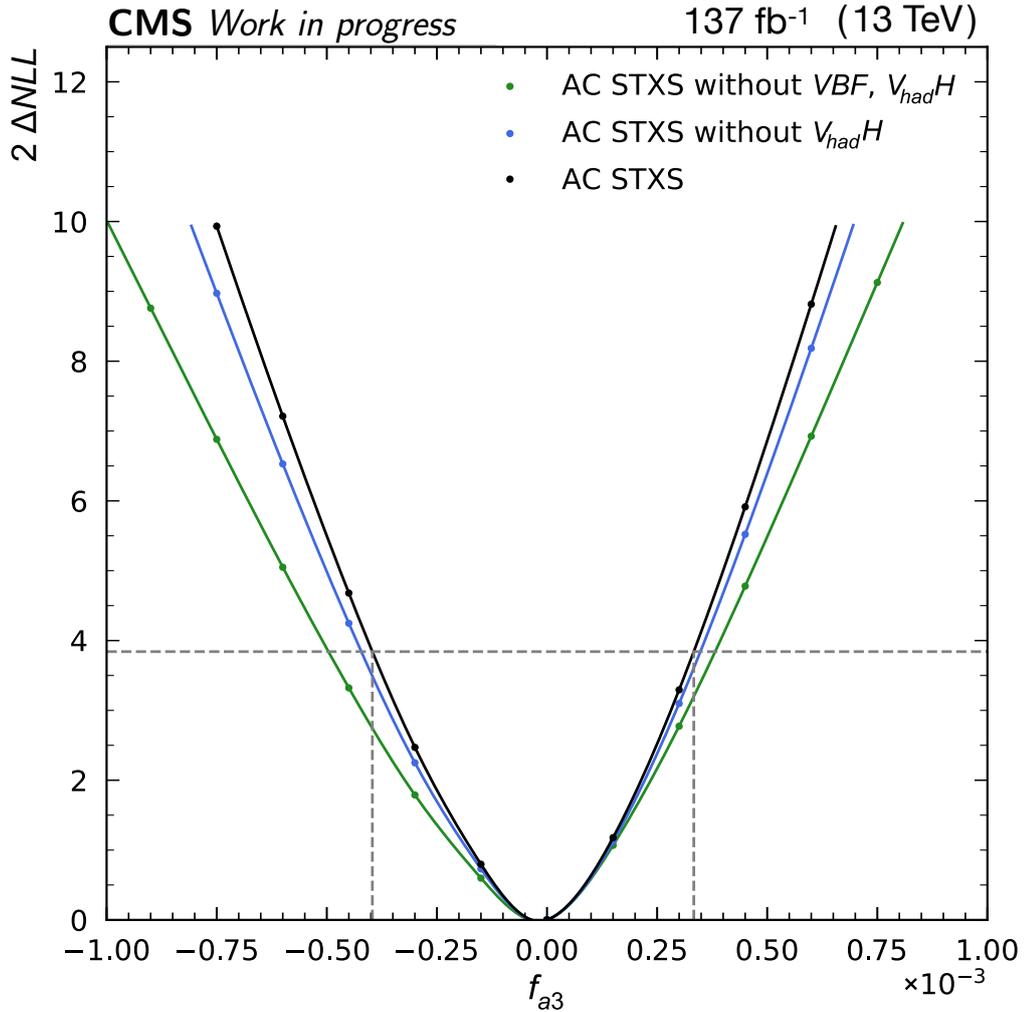


Figure 2 – Scan du test statistique ($2\Delta NLL$) pour différentes hypothèses sur la valeur de f_{a3} . Trois configurations sont présentées pour comparer les améliorations apportées par les catégories optimisées. En vert, le test statistique est calculé sans utiliser les catégories ciblant les productions VBF et $V_{had}H$. En bleu, les catégories VBF sont incluses. Enfin, le scan final utilisant toutes les catégories ($V_{had}H$ comprises) est représentée par la courbe noire.

en Fig. 2 et les contraintes attendues sur la valeur de f_{a3} forment l'intervalle de confiance à 68% suivant : $[-0.17 \times 10^{-3}, 0.13 \times 10^{-3}]$. Les limites attendues obtenues sont compétitives avec d'autres analyses ciblant des états finaux du boson de Higgs différents, notamment les analyses $H \rightarrow ZZ$ et $H \rightarrow \tau\tau$.

Bien que le LHC aura fourni une quantité de données importante (une luminosité intégrée de 300 fb^{-1} est prévue) à la fin de la période de prise de données en cours (fin 2025), le potentiel de découvertes liées à une physique au-delà du MS deviendra ensuite négligeable à la fréquence de collision actuelle. C'est pourquoi une jouvence importante est planifiée pour le LHC et ses expériences afin d'entrer en phase dite de haute luminosité qui devrait fournir une luminosité totale dix fois plus élevée après la même durée de fonctionnement. Les améliorations prévues pour cette jouvence seront mises en place lors du

prochain arrêt long (long shutdown 3) du LHC. Pour le détecteur CMS, un aspect important de cette jouvence consiste à ajouter la possibilité d'extraire le temps d'arrivée des particules dans les sous-détecteurs avec une très grande précision. Cette information devrait permettre de compenser les effets néfastes du grand nombre d'évènements simultanés causés par la fréquence de collision plus rapide. Cela sera aussi bénéfique aux analyses physiques telles que l'analyse $H \rightarrow \gamma\gamma$. Pour atteindre la résolution temporelle nécessaire (entre 30 ps et 40 ps), un signal d'horloge très précis doit être distribué et synchronisé entre toutes les parties des sous-détecteurs. Cette thèse apporte une contribution au contrôle rapide et à la calibration de la distribution du signal d'horloge de haute précision. Cette contribution prend la forme d'un environnement de simulation décrit dans le Chapitre II : *pyDDMTD*. Cet environnement apporte une simulation de ce qu'apporterait d'une surveillance du signal d'horloge basée sur un DDMTD (digital dual mixer time difference) pour contrôler la gigue introduite par la chaîne de distribution du signal. D'après les résultats d'une étude simple reposant sur les simulations réalisées avec *pyDDMTD*, une solution utilisant seulement un DDMTD pourrait déjà apporter une correction des éléments de gigue ayant une fréquence au-dessus de 10 Hz. Les simulations fournies par *pyDDMTD* pourraient maintenant être utilisées pour étudier et comparer différentes stratégies de correction basées sur des DDMTDs.

Contents

Introduction	1
I The Standard Model and the Higgs Boson	3
I.1 The standard model of particle physics	4
I.1.1 Elementary particles	5
I.1.2 Strong interaction	7
I.1.3 Electroweak interaction	8
I.2 The Higgs boson and where to find it	12
I.2.1 The Brout–Englert–Higgs mechanism	12
I.2.2 The Higgs boson ten years after its discovery	14
I.3 Physics beyond the standard model	21
II The CMS Experiment	23
II.1 The LHC and the CMS detector	24
II.1.1 The large hadron collider at CERN	24
II.1.2 A compact muon solenoid: the CMS detector	28
II.2 High-luminosity LHC and subdetectors adapted for timing in CMS phase II	38
II.2.1 A luminosity upgrade for the LHC	38
II.2.2 CMS 2.0: probing a 4th dimension	40
II.3 DDMTD simulation for precise clock monitoring in CMS subdetectors	43
II.3.1 Clock distribution	43
II.3.2 Characterisation study of the lpGBT version 1	45
II.3.3 The DDMTD system	46
II.3.4 Simulating a DDMTD: <i>pyDDMTD</i>	49
II.3.5 Realistic jitter simulation	53
II.3.6 Evaluating corrections on the clock jitter	57
II.4 Conclusion	59
III Deep Learning	61
III.1 Machine learning and neural networks	62
III.1.1 Development of machine learning	62
III.1.2 Building an artificial neuron	63
III.1.3 Neural networks	64
III.2 Teaching a machine	66
III.2.1 Learning paradigms	66
III.2.2 Loss and metrics	67

III.2.3 Gradient descent and backpropagation	68
III.3 Deep neural networks	71
III.3.1 Overtraining	71
III.3.2 Vanishing and exploding gradients	73
III.3.3 Learning rate	74
III.3.4 An example of DNN: the convolutional neural network	76
III.4 Generative adversarial networks	78
III.5 Summary	81
IV Data Driven Estimation of Background Using GAN	83
IV.1 Motivations	84
IV.2 Data-driven estimation of the background in the $H \rightarrow \gamma\gamma$ analysis of CMS	85
IV.3 Generating photons with a GAN: a brilliant method?	90
IV.3.1 Methodology	90
IV.3.2 Model architecture and training	91
IV.3.3 Performance metrics	92
IV.3.4 Hyperparameters optimisation	95
IV.4 GAN performance	100
IV.5 Application to data	104
IV.5.1 Generating a $\gamma + \text{jets}$ sample with the GAN	105
IV.5.2 Generating a $\gamma + \text{jets}$ sample with a 1D PDF	106
IV.5.3 Comparison of the two methods	108
IV.6 Conclusion	112
V Constraints on the Anomalous Couplings of the Higgs Boson in the VH Production Mode and Diphoton Final State	113
V.1 Physics motivations and analysis strategy	114
V.1.1 Modeling couplings of the Higgs boson with electroweak gauge bosons	114
V.1.2 Strategy	116
V.2 Trigger and objects definition	116
V.3 Highlighting $H \rightarrow \gamma\gamma$ events	121
V.3.1 Collected data and MC simulations	121
V.3.2 Reweighting of JHU samples	122
V.3.3 Phenomenology of $V_{\text{had}}H$ events and preselection	124
V.4 Machine learning for events discrimination	129
V.5 Optimisation of analysis categories	132
V.6 Statistical inference and results	136
V.6.1 Inference procedure	136
V.6.2 Background modelling	139
V.6.3 Signal modelling	140
V.6.4 Systematic uncertainties	142
V.6.5 Results	152
V.7 Conclusion	155
Conclusion	159
Bibliography	161

Introduction

Ten years after the discovery of the Higgs boson, our knowledge of its properties has greatly increased thanks to the excellent performance of the CERN large hadron collider (LHC) and its detectors. The ATLAS and CMS collaborations announced the discovery of a candidate Higgs boson in 2012 after two years of data taking with proton-proton collisions at a centre-of-mass energy of 7 TeV and 8 TeV. The decay of the Higgs boson into two photons ($H \rightarrow \gamma\gamma$) was one of the decay channels with the highest sensitivity at the time of discovery, together with the decay of the Higgs boson in four leptons. The $H \rightarrow \gamma\gamma$ decay has a small branching ratio of about 0.2%, but its final state is fully reconstructed, with a resolution close to 1% on the invariant mass of the diphoton system. Since the discovery, about 13 times more data have been collected at higher energies, and this boson candidate has been confirmed to correspond to the Higgs boson predicted by the standard model (SM) of particle physics. Measurements of its mass, spin, and couplings with other particles of the SM have been achieved with great precision, all confirming the robustness of the SM. Increasing the amount of data available provides a remarkable opportunity to push further our understanding of the Higgs boson, and to make the most of it, analyses teams of the LHC experiments deliver constant efforts to improve the design of analysis strategies.

Concurrently, in the last decade, the field of machine learning (ML) has known extensive growth with the development of more performing tools, particularly the many flavours of deep neural networks. These advanced methods are proven to be especially powerful when given a large amount of data during their training phase. Therefore, they are becoming a natural choice for high energy physics (HEP) communities when developing techniques to extract information from signal events buried under vast amounts of background.

As the $H \rightarrow \gamma\gamma$ analysis is a crucial probe to access properties of the Higgs boson, this thesis aims to enhance the performance of the $H \rightarrow \gamma\gamma$ analysis of the CMS experiment by including deep learning methods at different stages of the analysis. A first improvement is proposed by implementing generative adversarial networks (GAN) to better describe the background of the $H \rightarrow \gamma\gamma$ analysis. The primary background sources of the $H \rightarrow \gamma\gamma$ analysis are processes producing two photons in their final state or objects (mainly jets) misidentified

as photons. The role of the GAN is to generate new misidentified photons for events belonging to a control region in data so that these events offer a more accurate estimation of background in the signal region than the available Monte Carlo simulations.

One of the properties of the Higgs boson studied in this thesis is its behaviour with respect to the charge-parity (CP) symmetry. As the Higgs boson is expected to have a CP-even nature in the SM, any couplings hinting at a different behaviour (*e.g.* CP-odd couplings) are considered anomalous. This thesis presents the expected constraints on these anomalous couplings (AC) through the couplings of the Higgs boson with electroweak gauge bosons (W or Z boson). The complete analysis targets multiple production modes of the Higgs boson sensitive to its couplings with the W or Z boson. I am involved in the analysis through the design of the analysis strategy for the VH production where V is a W or Z boson which decays hadronically. To improve the sensitivity of the analysis, we include a multiclassifier deep neural network to separate events between background, SM-like signal and AC-like signal.

Although the LHC will have provided a significant amount of data (an integrated luminosity of 300 fb^{-1} is foreseen) at the end of the current data-taking period (end of 2025), the potential for new physics discoveries will become negligible at the current rate of collisions. This is why significant upgrades are planned for the LHC and its experiments to enter a high-luminosity phase which should provide a total luminosity ten times higher after the same operation time. These upgrades will be implemented during the long shutdown 3 of the LHC. For the CMS detector, a significant aspect of its upgrades is to include information on the time of arrival of particles in its subdetectors. This information should mitigate the side effects of the high number of simultaneous events arising from a faster collision rate. It will also benefit offline analyses such as the $H \rightarrow \gamma\gamma$ analysis. To reach the necessary timing resolution of 30 ps to 40 ps, a very clean clock signal needs to be delivered and synchronised among all parts of the subdetectors. This thesis contributes to the fast monitoring and calibration of the high-precision clock distribution.

The thesis unfolds as follows. In Chapter I, I describe the theoretical context of the SM and of Higgs boson physics at the LHC. A brief description of the LHC, the CMS detector and their upgrade for a high-luminosity phase are given in Chapter II, just before a report on my contribution to the monitoring of high-precision clock distribution in the context of the high-luminosity LHC (Section II.3 of the same chapter). Furthermore, the notions of deep learning necessary for presenting ML-based methods in the $H \rightarrow \gamma\gamma$ analysis are defined in Chapter III. These notions are applied in Chapter IV, which details the implementation of a GAN for the data-driven description of the background of the $H \rightarrow \gamma\gamma$ analysis. Finally, Chapter V presents the analysis of anomalous couplings of the Higgs boson with gauge boson in the $H \rightarrow \gamma\gamma$ channel with a focus on my contribution, *i.e.* on the VH production of the Higgs boson.

CHAPTER I

The Standard Model and the Higgs Boson

I.1	The standard model of particle physics	4
I.1.1	Elementary particles	5
I.1.2	Strong interaction	7
I.1.3	Electroweak interaction	8
I.2	The Higgs boson and where to find it	12
I.2.1	The Brout–Englert–Higgs mechanism	12
I.2.2	The Higgs boson ten years after its discovery	14
I.3	Physics beyond the standard model	21

In an attempt to describe the laws of our universe, the standard model of particle physics (SM) offers a single theoretical framework describing three of the four fundamental interactions discovered so far. The SM was established throughout the second half of the 20th century from the joint progress of quantum field theories (QFT) and high energy physics (HEP) experiments. One aspect that demonstrates the strength of the SM is its accurate prediction of many physical effects, later confirmed experimentally such as the predictions and observations of neutral weak currents (by the Gargamelle experiment in 1973), the discovery of the weak gauge bosons (UA1/UA2 experiments, 1983), of the top quark (CDF and DØ experiments, 1995), of the tau neutrino (DONUT experiment, 2000), and more recently of the Higgs boson (ATLAS and CMS experiments, 2012). While the SM has proven to be highly robust, it still leaves open questions about certain observed phenomena, and the search for physics going beyond the standard model (BSM) is one of the main ambitions of modern HEP experiments. This chapter briefly summarises the SM in Section I.1, then Section I.2 covers the Higgs sector and its experimental status.

I.1 The standard model of particle physics

The SM is based on QFT to describe three fundamental forces: the strong interaction (SI), the electromagnetic interaction (EMI), and the weak interaction (WI). These forces act on and are mediated by particles. In the context of QFT, each particle type is represented by a field taking values everywhere in space-time. When quantised, the local excitation of the fields corresponds to particles. These particles are divided into two categories based on their spin. Half-integer spin particles are called fermions, and integer spin particles are bosons. The SM characterises the behaviour of elementary particles, *i.e.* particles that do not have internal substructures to our current knowledge. Among these particles, the building blocks of matter are the elementary fermions with spin $1/2$.

In the SM, the fundamental interactions have an elegant origin based on internal symmetries on the Lagrangian. Building a theory of interacting particles can be summarised in three main steps. First, one isolates the internal global symmetries in the Lagrangian of a non-interacting field of matter (free field). A global symmetry is a differentiable transformation of the fields that do not depend on space-time and leaves the Lagrangian unchanged: for instance, a phase rotation of the fields for the EMI. These symmetries are fundamental in physics since, via Noether's theorem, they correspond to a conserved physical quantity, like the electric charge for the EMI. In the second step, one follows a recipe adapted from the building of the EMI theory, global symmetries are required to be local symmetries, *i.e.* that the field transformations depend on the space-time point. To do so, one needs to introduce a new field (known as a gauge field) as well as an interaction between these gauge fields and the original fields of matter. Finally, one provides some dynamics to the gauge fields by introducing a corresponding kinematic term in the Lagrangian. In this final step, one should stress that mass terms for gauge fields are prohibited because they would break the local symmetries (also known as gauge symmetries); thus, all gauge fields are, in essence, massless. A more thorough overview of these

steps in the context of interactions of the SM is given in Sections I.1.2 and I.1.3. In the SM, the interactions are the fundamental forces, and they are carried by the quantised gauge fields which are the gauge bosons. The set of mathematical transformations associated with the gauge symmetries of the system forms the symmetry group (or gauge group), and the total symmetry group of the SM is $SU(3)_C \times SU(2)_L \times U(1)_Y$ [1].

I.1.1 Elementary particles

The elementary particles of the SM are presented and classified in Fig. I.1.

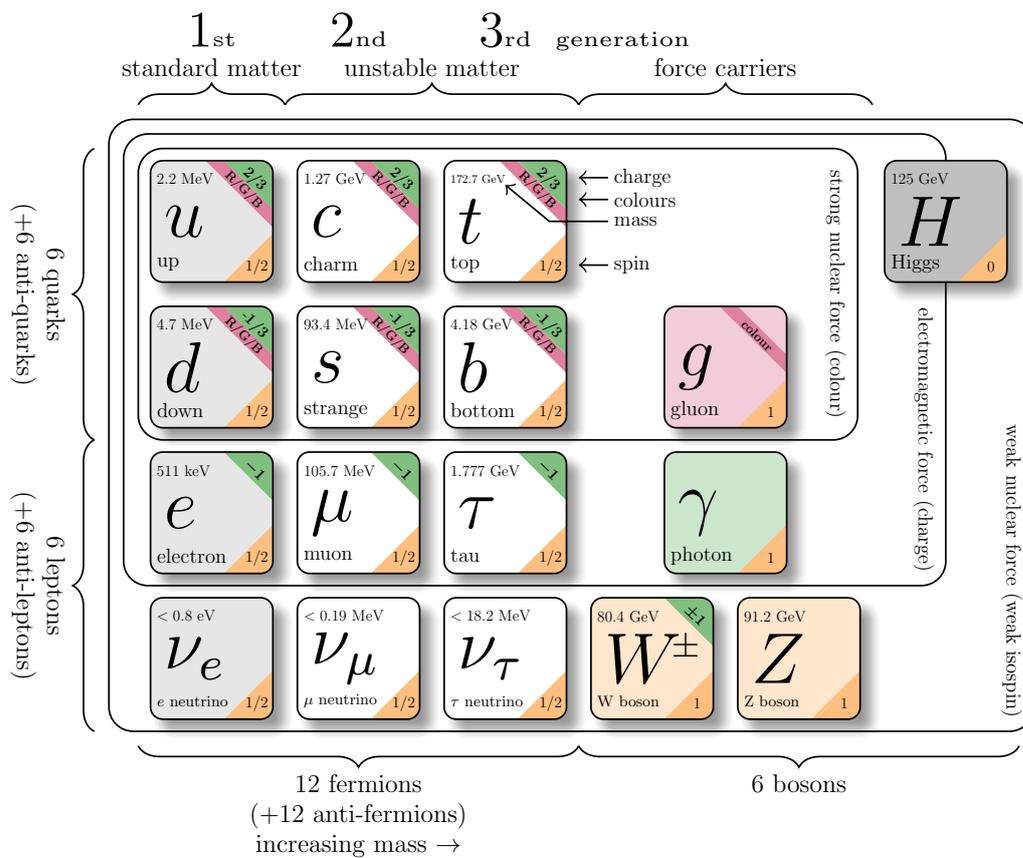


Figure I.1 – Classification of the particles described by the SM. Adapted from Ref. [2]. Values of the masses are taken from Ref. [3].

The elementary fermions of the SM have a spin $S = 1/2$ and are classified according to how they behave with respect to the fundamental interactions. This behaviour is encoded in a quantum number called a charge, with a different charge for each interaction. The dynamics of spin-1/2 massive particles are described by the Dirac equation, which implies that, for each such particle, there exists an antiparticle of the same mass but with opposite charges. Thus, from 12 fermions forming matter, there are 12 associated antifermions forming anti-matter.

The 6 fermions interacting through the SI are the quarks: up (u), down (d), strange (s), charm (c), bottom (b), and top (t). They carry a colour charge (red (R), green (G) or blue (B)) and have masses ranging between 2.2 MeV for the up quark to 172.7 GeV for the top quark. The theory describing the SI is called quantum chromodynamics (QCD) and shows that quarks cannot exist outside of colour-neutral bounded states (hadrons); this phenomenon is known as colour confinement. Hadrons can be further classified depending on their quark composition; bounded states consisting of three quarks are called baryons, and those containing one quark and one antiquark are called mesons. In the last decade, collider experiments have also discovered exotic baryons: pentaquarks (four quarks and one antiquark) and exotic mesons: tetraquarks (two quarks and two antiquarks). Therefore, when quarks are produced in collider experiments, we observe jets of hadrons in the detectors rather than quarks due to their hadronisation. Only the top quark has a mean lifetime so short ($\approx 5 \times 10^{-25}$ ps) that it decays before it can hadronise (typical SI time scale is $\approx 10^{-23}$ s). Quarks are also sensitive to the EMI and WI, so they have an electric and weak charge.

The 6 remaining fermions are uncoloured as they do not interact through the SI and are called leptons. Three of them interact electromagnetically and also through the WI: the electron e^- ($m_e = 511$ keV), the muon μ^- ($m_\mu = 105.7$ MeV), and the tau τ^- ($m_\tau = 1.777$ GeV). They are associated with three neutrinos: the electron neutrino ν_e ($m_{\nu_e} < 0.8$ eV), the muon neutrino ν_μ ($m_{\nu_\mu} < 0.19$ MeV), and the tau neutrino ν_τ ($m_{\nu_\tau} < 18.2$ MeV). As neutrinos do not have an electric charge, they interact only through the WI (and potentially gravity), making them particularly hard to detect as they pass through matter unfazed.

The SM includes bosons with a spin $S = 1$ which are the gauge bosons mediating the SI, EMI, and WI. The gauge bosons carrying the SI are the gluons corresponding to the $SU(3)_C$ symmetry group of QCD. Since the $SU(N)$ group is described by $(N^2 - 1)$ generators, the $SU(3)_C$ group has 8 generators, and that imposes the number of force-carrier of the theory: 8 gluons. Each corresponds to a different combination of colour and anticolour charges transferred between quarks. They are massless and do not interact with other gauge bosons, so they do not carry additional charges.

The four bosons γ , Z , W^+ , and W^- are emerging from a unified theory describing both the EMI and the WI: the Glashow-Salam-Weinberg (GSW) theory [4–6]. The GSW model has a $SU(2)_L \times U(1)_Y$ symmetry group (which has $3 + 1$ generators) that describes four massless gauge bosons before taking into account a spontaneous symmetry breaking leading to the Brout–Englert–Higgs (BEH) mechanism. This mechanism details how the four massless gauge bosons transform into the massive physical states Z , W^\pm , plus a massless photon that are observed experimentally. The photon mediates the EMI but does not carry any charge. The Z boson mediates the neutral WI; it has no electric charge and a mass $m_Z = 91.2$ GeV. In contrast, the W boson, which carries the charged WI, has an electric charge of one unit and a mass $m_W = 80.4$ GeV.

The Higgs boson and BEH mechanism are presented in more detail in Section 1.2

I.1.2 Strong interaction

The SI is described by the QCD gauge theory. This theory was needed to describe how the large number of hadrons experimentally observed could be explained by the existence of smaller particles: the quarks. A new quantum number, the colour $C = \{r, g, b\}$, was introduced to represent an interaction with a $SU(3)_C$ symmetry group between quarks. Thus, the quark field ψ can be represented as a triplet of Dirac spinors corresponding to the different colour charges:

$$\psi = \begin{pmatrix} \psi_r \\ \psi_g \\ \psi_b \end{pmatrix} \text{ and a Dirac adjoint defined as } \bar{\psi} = \psi^\dagger \gamma^0.$$

As mentioned already, the dynamics of spin-1/2 massive particles are described by the Dirac equation, so the most basic description of the Lagrangian of the theory has the form

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\cancel{\partial} - m)\psi, \quad (I.1)$$

where the slashed notation is used, $\cancel{\partial} = \gamma^\mu \partial_\mu$, and γ^μ are the Dirac matrices. A unitary 3×3 matrix U , with determinant 1, can be used to represent a global symmetry of $SU(3)$ through the transformation

$$\psi \rightarrow \psi' = U\psi \quad (I.2)$$

which leaves $\mathcal{L}_{\text{Dirac}}$ unchanged. This transformation can be adapted to represent a local symmetry if it depends on the space-time coordinate x : $U \rightarrow U(x)$. Since the transformation belongs to the symmetry group, it can be expressed as a function of the generators t^a of the group

$$U(x) = e^{ig\alpha^a(x)t^a}, \quad (I.3)$$

where g is the strong coupling constant and α^a are arbitrary functions of space-time. Following Einstein's notation, we imply the summation of repeated indices for both Greek and Roman indices. In the context of $SU(3)$, the generators are represented by the Gell-Mann matrices λ^a as $t^a = \lambda^a/2$. The transformation $U(x)$ can be written in its infinitesimal form

$$U(x) = 1 + ig\alpha^a(x)t^a + \mathcal{O}(\alpha^2). \quad (I.4)$$

To ensure the symmetry of the Lagrangian with respect to this local transformation, we need to introduce new vector fields A^a , which are absorbed in the definition of a sensible derivative adapted to the theory (the covariant derivative):

$$D_\mu = \partial_\mu - igA_\mu^a t^a. \quad (I.5)$$

A field is associated with each generator of $SU(3)$, so 8 in total, and they are associated with the gluon fields carrying the SI. For non-Abelian theories (*i.e.* the t^a do not commute), the commutator of the t^a generators can be written as

$$[t^a, t^b] = if^{abc}t^c, \quad (I.6)$$

where the f^{abc} are called the structure constants. It allows the definition of the transformation laws for the fields of the theory as

$$\psi \rightarrow (1 + ig\alpha^a t^a)\psi; \quad (1.7)$$

$$A_\mu^a \rightarrow A_\mu^a + \frac{1}{g}\partial_\mu\alpha^a + f^{abc}A_\mu^b\alpha^c. \quad (1.8)$$

To define the most general locally invariant Lagrangian, it must include an invariant term describing the dynamics of the A^a fields. It is defined by introducing a gluon field strength tensor G and forming a simple invariant term

$$\mathcal{L}_{\text{Gluon}} = -\frac{1}{4}(G_{\mu\nu}^a)^2, \quad (1.9)$$

with

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c. \quad (1.10)$$

Thus, the complete QCD Lagrangian is written as

$$\begin{aligned} \mathcal{L}_{\text{QCD}} &= \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Gluon}} \\ &= \bar{\psi}(i\cancel{D} - m)\psi - \frac{1}{4}(G_{\mu\nu}^a)^2. \end{aligned} \quad (1.11)$$

which describe a special case of Yang-Mills theory [7] associated with a $SU(3)$ symmetry. The interaction between quarks and gluon is described through the term

$$\mathcal{L}_{\text{QCD}} = \dots + g\bar{\psi}\gamma^\mu A_\mu^a t^a \psi + \dots$$

and the nonlinear term in the gluon field strength tensor gives rise to three- and four-gluons interactions. The only mass term of the theory is associated with the quarks

$$\mathcal{L}_{\text{QCD}} = \dots - m\bar{\psi}\psi + \dots$$

Mass terms for the gauge fields of the form $A_\mu^a A^{a\mu}$ would not keep the Lagrangian invariant under $SU(3)$ transformations, so they cannot be included in such a gauge theory, *i.e.* gluons must be massless.

I.1.3 Electroweak interaction

The unified theory describing the EMI and WI is also a Yang-Mills theory but associated with a $SU(2)_L \times U(1)_Y$ gauge group. This describes the experimental observations showing that the W^\pm bosons only interact with certain fermions, depending on their chirality. The chirality of a fermion is exhibited by the $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ operator, which has ± 1 eigenvalues: $+1$ corresponds to right-handed fermions and -1 to left-handed fermions. In the massless limit, the chirality of a particle is equivalent to its helicity (projection of the spin direction on the momentum direction). Moreover, the WI acts on pairs of fermions of different types (*e.g.* $W^+ \rightarrow u\bar{d}$ or $W^- \rightarrow e^-\bar{\nu}_e$). Thus, through the WI, left-handed fermions (or right-handed antifermions) are represented as doublets that can interact with

the WI, whereas right-handed fermions (or left-handed antifermions) are singlets that do not interact through WI. In this section, the particle fields are denoted by their symbol and the different doublets and singlets are represented as

$$\psi_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} \nu_{eL} \\ e_L^- \end{pmatrix}, \dots, \text{ and } \psi_R = u_R, d_R, e_R^-, \dots$$

Given that left- and right-handed fermions belong to different $SU(2)$ representations, the free Lagrangian for fermions cannot contain terms such as:

$$\mathcal{L}_{\text{Dirac}} = \dots - m_u(\bar{u}_L u_R + \bar{u}_R u_L) + \dots$$

This would spoil the $SU(2)_L$ invariance of the Lagrangian, so it imposes the fermions of the theory to be massless.

Then, by proceeding similarly as the SI, we can introduce two kinds of gauge fields A^a ($a = 1, 2, 3$) and B to keep the Lagrangian invariant with respect to transformations from the two symmetry groups $SU(2)_L$ and $U(1)_Y$. They are associated with the generators of these groups: three generators $T^a = \sigma^a/2$ for $SU(2)_L$, where σ^a are the Pauli matrices, and one generator $Y/2$ for $U(1)_Y$. The eigenvalues ($+1/2$ or $-1/2$) I_3 of T_3 corresponds to the weak isospin. Based on the weak isospin, fermions can be classified into two categories: up-type fermions have a value $I_3 = +1/2$, while down-type fermions correspond to $I_3 = -1/2$. The Y value is the weak hypercharge, and its combination with the weak isospin gives the electric charge Q of a fermion following Eq. I.12. A summary of the electric charge, weak isospin, and weak hypercharge of each fermion type is given in Table I.1.

$$Q = I_3 + \frac{Y}{2}. \tag{I.12}$$

Fermions	Q	I_3		Y	
		Left	Right	Left	Right
u, c, t	$+\frac{2}{3}$	$+\frac{1}{2}$	0	$+\frac{1}{3}$	$+\frac{4}{3}$
d, s, b	$-\frac{1}{3}$	$-\frac{1}{2}$	0	$+\frac{1}{3}$	$-\frac{2}{3}$
e^-, μ^-, τ^-	-1	$-\frac{1}{2}$	0	-1	-2
ν_e, ν_μ, ν_τ	0	$+\frac{1}{2}$		-1	

Table I.1 – Value of the electric charge Q , weak isospin I_3 , and weak hypercharge Y for each fermion type, with a distinction between left-handed (*Left*) and right-handed (*Right*) particles when needed.

From these gauge fields and generators, we can define the covariant derivative of the theory

$$D_\mu = \partial_\mu - ig_2 A_\mu^a T^a - ig_1 \frac{Y}{2} B_\mu. \tag{I.13}$$

And finally, we can write their associated field strength tensors:

$$A_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_2 \epsilon^{abc} A_\mu^b A_\nu^c \quad (\text{I.14})$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (\text{I.15})$$

where ϵ^{abc} is the Levi-Civita tensor, giving the commutation relation of the Pauli matrices.

We can now write the Lagrangian of the electroweak theory:

$$\begin{aligned} \mathcal{L}_{\text{EW}} &= \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{int}} + \mathcal{L}_{\text{bosons}} \\ &= \bar{\Psi}_L(i\not{D})\Psi_L + \bar{\psi}_R(i\not{D})\psi_R - \frac{1}{4}(A_{\mu\nu}^a)^2 - \frac{1}{4}(B_{\mu\nu})^2 \\ &= \bar{\Psi}_L(i\not{\partial})\Psi_L + \bar{\psi}_R(i\not{\partial})\psi_R \\ &+ g_2 \bar{\Psi}_L \gamma^\mu A_\mu^a T^a \Psi_L + g_1 \frac{Y}{2} \bar{\Psi}_L \gamma^\mu B_\mu \Psi_L + g_1 \frac{Y}{2} \bar{\psi}_R \gamma^\mu B_\mu \psi_R \\ &- \frac{1}{4}(A_{\mu\nu}^a)^2 - \frac{1}{4}(B_{\mu\nu})^2. \end{aligned} \quad (\text{I.16})$$

But as mentioned above, given the observed behaviour of the WI, we expect interaction terms of the form:

$$\mathcal{L}_{\text{int}} = \dots + \kappa \bar{u}_L \gamma^\mu W_\mu^+ d_L + \dots,$$

with a coupling constant κ , so under their current form, the gauge fields are not the physical gauge bosons we observe experimentally. The latter can be recovered by combining the A^a and B fields. The two charged W^\pm bosons are obtained from

$$W_\mu^\pm = \frac{A_\mu^1 \mp iA_\mu^2}{\sqrt{2}}, \quad (\text{I.17})$$

and similarly, to obtain the correct association of generators

$$T^\pm = T^1 \pm iT^2. \quad (\text{I.18})$$

On the other hand, the Z_μ and A_μ fields associated with the Z boson and the photon are recovered by mixing the remaining A_μ^3 and B_μ fields:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_w & -\sin \theta_w \\ \sin \theta_w & \cos \theta_w \end{pmatrix} \begin{pmatrix} A_\mu^3 \\ B_\mu \end{pmatrix}, \quad (\text{I.19})$$

where θ_w is the weak mixing angle, also called the Weinberg angle.

From these definitions, we can rewrite the covariant derivative defined in Eq. I.13 as

$$\begin{aligned} D_\mu &= \partial_\mu - i \frac{g_2}{\sqrt{2}} W_\mu^+ T^+ - i \frac{g_2}{\sqrt{2}} W_\mu^- T^- \\ &- i \left(g_2 \cos(\theta_w) T^3 - g_1 \sin(\theta_w) \frac{Y}{2} \right) Z_\mu \\ &- i \left(g_2 \sin(\theta_w) T^3 + g_1 \cos(\theta_w) \frac{Y}{2} \right) A_\mu. \end{aligned} \quad (\text{I.20})$$

Since the last term, associated with the photon field A_μ , should have as coupling constant the electric charge, we can identify the following relation using Eq. I.12:

$$e = g_1 \cos(\theta_w) = g_2 \sin(\theta_w), \quad (\text{I.21})$$

which allows the expression of the weak mixing angle θ_w with respect to the coupling constants g_1 and g_2 :

$$\cos \theta_w = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}, \quad \sin \theta_w = \frac{g_1}{\sqrt{g_1^2 + g_2^2}}, \quad \text{as well as } e = \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (\text{I.22})$$

Finally, we can write the different interaction terms between fermions and the physical gauge bosons, first for left-handed fermions e.g. with the doublet describing the u_L and d_L quarks:

$$\begin{aligned} \mathcal{L}_{\text{int}}^L &= \frac{g_2}{\sqrt{2}} \bar{u}_L \gamma^\mu W_\mu^+ d_L + \frac{g_2}{\sqrt{2}} \bar{d}_L \gamma^\mu W_\mu^- u_L \\ &+ e \left(I_u^3 + \frac{Y_u}{2} \right) \bar{u}_L \gamma^\mu A_\mu u_L + e \left(I_d^3 + \frac{Y_d}{2} \right) \bar{d}_L \gamma^\mu A_\mu d_L \\ &+ e \frac{g_2}{g_1} \left(I_u^3 - \left(\frac{g_1}{g_2} \right)^2 \frac{Y_u}{2} \right) \bar{u}_L \gamma^\mu Z_\mu u_L \\ &+ e \frac{g_2}{g_1} \left(I_d^3 - \left(\frac{g_1}{g_2} \right)^2 \frac{Y_d}{2} \right) \bar{d}_L \gamma^\mu Z_\mu d_L. \end{aligned} \quad (\text{I.23})$$

And then, for any right-handed fermion ψ_R with electric charge Q we have:

$$\mathcal{L}_{\text{int}}^R = eQ \bar{\psi}_R \gamma^\mu A_\mu \psi_R + \frac{g_1}{g_2} eQ \bar{\psi}_R \gamma^\mu Z_\mu \psi_R. \quad (\text{I.24})$$

The interactions between gauge bosons is included in the $\mathcal{L}_{\text{boson}}$ term, expressing interactions between three ($\gamma W^+ W^-$, $Z W^+ W^-$) and four ($\gamma \gamma W^+ W^-$, $Z Z W^+ W^-$, $Z \gamma W^+ W^-$, $W^+ W^- W^+ W^-$) gauge bosons.

While successfully describing the interactions between particles under the WI, the electroweak Lagrangian \mathcal{L}_{EW} reached here is missing crucial components. Indeed, to conserve the $SU(2)_L \times U(1)_Y$ symmetry assumed for this gauge theory, all fermions and bosons should be massless. However, the previous theory described the WI through a four-fermions point-like interaction (Fermi's contact interaction [8]) introducing a coupling constant G_F . With the development of QFTs, mathematical inconsistencies in Fermi's contact interaction were pointed out, and Fermi's theory was interpreted as a low-energy effective theory. Its proper mathematical description was achieved via the electroweak theory with gauge bosons mediating the WI. Yet, the short-range interaction which was implied by the contact interaction, was pointing towards very massive gauge bosons, in contradiction with the requirement of massless gauge bosons required by gauge theories. It means that the gauge symmetry of the theory should break spontaneously to give mass to certain particles; this mechanism is described in Section I.2.1.

I.2 The Higgs boson and where to find it

I.2.1 The Brout-Englert-Higgs mechanism

The $SU(2)_L$ symmetry of the electroweak theory, while offering an accurate description of the WI, is not observed experimentally as fermions and the Z and W^\pm bosons have a measurable mass. It is suggestive of a hidden symmetry of the theory. Only a $U(1)$ symmetry is observed in the electroweak sector, associated with the EMI and to a massless photon. This specific mechanism of spontaneous symmetry breaking $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$ of the electroweak theory giving mass to the bosons of the SM is called the Brout-Englert-Higgs (BEH) mechanism [9–11].

The BEH mechanism introduces a complex scalar ($S = 0$) field ϕ which is a $SU(2)_L$ doublet with hypercharge $Y = 1$ and transforms under the $SU(2)_L \times U(1)_Y$ symmetry as

$$\phi \rightarrow e^{i\alpha^a(x)T^a} e^{i\frac{\beta(x)}{2}} \phi, \quad (\text{I.25})$$

with arbitrary functions of space-time α^a and β . The dynamics of this field are described through a Lagrangian, including a kinematic and a potential term of this scalar field:

$$\mathcal{L}_{\text{BEH}} = (D_\mu \phi^\dagger)(D^\mu \phi) - V(\phi) \quad (\text{I.26})$$

where D is the covariant derivative defined in Eq. I.20 including the gauge fields of the electroweak theory, and the potential V is defined as

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda |\phi|^4, \quad (\text{I.27})$$

with two parameters μ and λ such that $\mu^2 > 0$ and $\lambda > 0$. An illustration of the shape of the potential function V is provided in Fig. I.2. The ground state of the scalar field is at the minimum of the potential function such that

$$|\phi_{\text{min}}| = \frac{v}{\sqrt{2}}, \quad \text{with } v = \sqrt{\frac{\mu^2}{\lambda}}. \quad (\text{I.28})$$

It means that the field ϕ acquires a nonzero vacuum expectation value (VEV), and the $SU(2)_L \times U(1)_Y$ is spontaneously broken. We write the VEV of the field ϕ under the form

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (\text{I.29})$$

Small excitations of the scalar field about its VEV can be parameterised using four real fields:

$$\phi(x) = \frac{1}{\sqrt{2}} e^{iT^a \theta^a(x)} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (\text{I.30})$$

where the three θ^a fields are called the Goldstone fields and H is the Higgs boson real scalar field. Since the rest of the Lagrangian is invariant under $SU(2)$ transformations, we can use the inverse transformation to remove any dependency

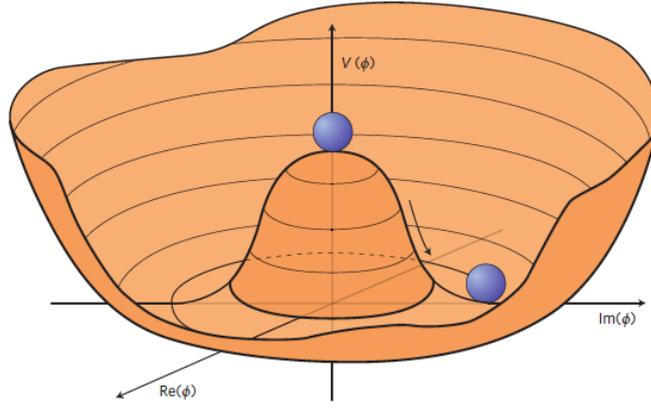


Figure I.2 – Illustration [12] of the potential described by Eq. I.27. The ball is shown at two equilibrium positions: an unstable one associated with a null field and a stable one corresponding to $|\phi| = |\phi_{\min}|$.

with respect to the Goldstone fields:

$$\phi \rightarrow e^{-iT^a \theta^a(x)} \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (\text{I.31})$$

Thus, the only physical field remaining is the scalar Higgs boson H which corresponds to the last particle predicted by the SM.

By using the form of the scalar field ϕ presented in Eq. I.31 and the covariant derivative defined in Eq. I.20, we can expand the Lagrangian \mathcal{L}_{BEH} and exhibit mass terms for the theory:

$$\mathcal{L}_{\text{BEH}} = \dots + \frac{1}{2} 2\mu^2 H^2 + \frac{g_2^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{1}{2} \frac{g_1^2 + g_2^2}{4} v^2 Z_\mu Z^\mu + \dots,$$

where we can identify

$$m_H = \sqrt{2} \mu, \quad m_W = \frac{g_2 v}{2}, \quad \text{and} \quad m_Z = \frac{\sqrt{g_1^2 + g_2^2}}{2} v. \quad (\text{I.32})$$

The rest of the Lagrangian also contains the interaction terms between the H boson and the W, Z bosons, as well as self-coupling terms of the H boson. However, as there is no term including the photon field A_μ , the photon remains massless after the symmetry breaking, as expected, and it does not interact with the Higgs field.

Finally, the observed mass of the fermions of the theory is described by the Yukawa interaction between these fermions and the scalar field doublet ϕ . For more clarity in the presentation of the following equations, we separate the fermion left-handed doublets between quark and lepton doublets:

$$Q_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix} \quad \text{and} \quad E_L^i = \begin{pmatrix} \nu_L^i \\ e_L^i \end{pmatrix} \quad (\text{I.33})$$

with i running over the different generations of fermions, so $u^i(d^i)$ corresponds to the up-type(down-type) quarks. Then, we can define the Lagrangian representing the Yukawa interaction:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_u^i \bar{Q}_L^i \tilde{\phi} u_R^i - \lambda_d^i \bar{Q}_L^i \phi d_R^i - \lambda_e^i \bar{E}_L^i \phi e_R^i + \text{h.c.}, \quad (\text{I.34})$$

where $\tilde{\phi}$ is the charge-conjugate Higgs field $\tilde{\phi} = i\sigma_2\phi$, h.c. denotes the hermitian conjugate terms and λ_x are the coupling constants of the interactions between the Higgs boson and fermions. Given the form of ϕ defined in Eq. I.31, we can expand the Lagrangian as

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} &= -\lambda_{u^i} \frac{v+H}{\sqrt{2}} \bar{u}_L^i u_R^i - \lambda_{d^i} \frac{v+H}{\sqrt{2}} \bar{d}_L^i d_R^i - \lambda_{e^i} \frac{v+H}{\sqrt{2}} \bar{e}_L^i e_R^i + \text{h.c.} \\ &= -\frac{\lambda_{u^i} v}{\sqrt{2}} \bar{u}^i u^i - \frac{\lambda_{d^i} v}{\sqrt{2}} \bar{d}^i d^i - \frac{\lambda_{e^i} v}{\sqrt{2}} \bar{e}^i e^i \\ &\quad - \frac{\lambda_{u^i}}{\sqrt{2}} H \bar{u}^i u^i - \frac{\lambda_{d^i}}{\sqrt{2}} H \bar{d}^i d^i - \frac{\lambda_{e^i}}{\sqrt{2}} H \bar{e}^i e^i\end{aligned}\tag{I.35}$$

where the fermion fields are regrouped as $u = (u_R^i, u_L^i)$, $d^i = (d_R^i, d_L^i)$, and $e^i = (e_R^i, e_L^i)$. Thus, the masses of fermions are expressed as

$$m_{u^i} = \frac{\lambda_{u^i} v}{\sqrt{2}}, \quad m_{d^i} = \frac{\lambda_{d^i} v}{\sqrt{2}}, \quad \text{and} \quad m_{e^i} = \frac{\lambda_{e^i} v}{\sqrt{2}},\tag{I.36}$$

and we can identify interactions between the Higgs boson and fermions (last line of Eq. I.35) with a coupling constant related to the mass of fermions of the form m_x/v . In this brief presentation of the Yukawa couplings, we omit the terms responsible for mixing the different generations of fermions in the WI. The terms are introduced by the CKM matrix, which has nonzero off-diagonal values. The CKM gives rise to the only known source of charge-parity (CP) symmetry violation in the SM.

The model presented here assumes neutrinos are massless, which is not what is observed experimentally. Alternative mechanisms are proposed to include the mass of neutrinos. For instance, adding a Yukawa coupling to Eq. I.34 could describe a massive Dirac neutrino. The BEH mechanism provides an explanation to the observed mass of fermions and gauge bosons, and expands the electroweak theory with a description of the dynamics of a Higgs field and its interactions with other particles $\mathcal{L}_{\text{EW}} = \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{boson}} + \mathcal{L}_{\text{int}}^L + \mathcal{L}_{\text{int}}^R + \mathcal{L}_{\text{BEH}} + \mathcal{L}_{\text{Yukawa}}$. The Higgs boson interacts with fermions and bosons with coupling constants related to their mass, granting a unique way to probe fundamental properties of particles.

I.2.2 The Higgs boson ten years after its discovery

The large hadron collider (LHC) was designed to reach energy levels sufficient to probe the properties and interactions of the Higgs field described in Section I.2.1 by producing Higgs bosons from proton-proton (pp) collisions. In 2012, after two years of data-taking at centre-of-mass energies $\sqrt{s} = 7$ TeV and 8 TeV, the ATLAS and CMS experiments of the LHC discovered the Higgs boson [13–15]. A decade later, the LHC has provided much more data to the experiments through the Run 1 (data-taking period 2010–2012) and Run 2 (2015–2018) and has reached even higher energies. For instance, the CMS experiment has recorded a total of 138 fb^{-1} of data at $\sqrt{s} = 13$ TeV during the Run 2. The goal of this section is to

provide an overview of our knowledge of the BEH sector thanks to the experiments of the LHC.

As the couplings of the Higgs boson grow linearly with the mass of fermions and quadratically with the mass of bosons, interactions of the Higgs boson with heavy particles are more likely. The most frequent production modes occurring in pp collisions at the LHC are presented in Table I.2 (left) along with their cross section (σ_X). The cross section is a representation of the probability for a given process to occur following a pp collision and is expressed here in picobarn ($1 \text{ pb} = 10^{-36} \text{ cm}^{-2}$). According to Table I.2, the total cross section of Higgs boson pro-

Production mode	Cross section (pb)	Decay channel	Branching fraction (%)
ggH	48 ± 2	bb	57.6 ± 0.7
VBF	3.8 ± 0.8	WW	22.0 ± 0.3
WH	1.36 ± 0.03	gg	8.2 ± 0.4
ZH	0.88 ± 0.04	$\tau\tau$	6.21 ± 0.09
ttH	0.50 ± 0.04	cc	2.86 ± 0.09
bbH	0.48 ± 0.09	ZZ	2.71 ± 0.04
tH	0.09 ± 0.01	$\gamma\gamma$	0.227 ± 0.005
		$Z\gamma$	0.157 ± 0.009
		ss	0.025 ± 0.001
		$\mu\mu$	0.0216 ± 0.0004

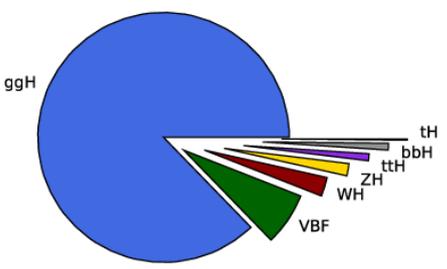
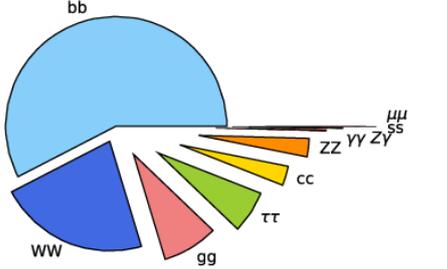



Table I.2 – Expected cross sections for the leading production modes of the H boson and expected branching fraction of its leading decay channels for proton-proton collision at $\sqrt{s} = 13 \text{ TeV}$ and for $m_H = 125.38 \text{ GeV}$. Adapted from Ref. [16].

duction at $\sqrt{s} = 13 \text{ TeV}$ is around 55 pb , so for a total of 138 fb^{-1} of collision data accumulated during the Run 2, we expect around $55\,000 \times 138 \approx 7.6$ million collisions produced a Higgs boson; the challenge is then to detect them. The dominant production mode is by far the gluon fusion (ggH or ggF), where the Higgs boson is produced via a loop of heavy quarks (mainly involving t quarks). The second-largest production comes from the fusion of vector bosons (VBF), and the Higgs boson is produced in association with two quarks. Finally, the Higgs boson can be produced in association with a single vector boson (VH with $V = Z$ or W) or a pair of heavy quarks (ttH, bbH). The Feynman diagrams associated with these production modes are shown in Fig. I.3.

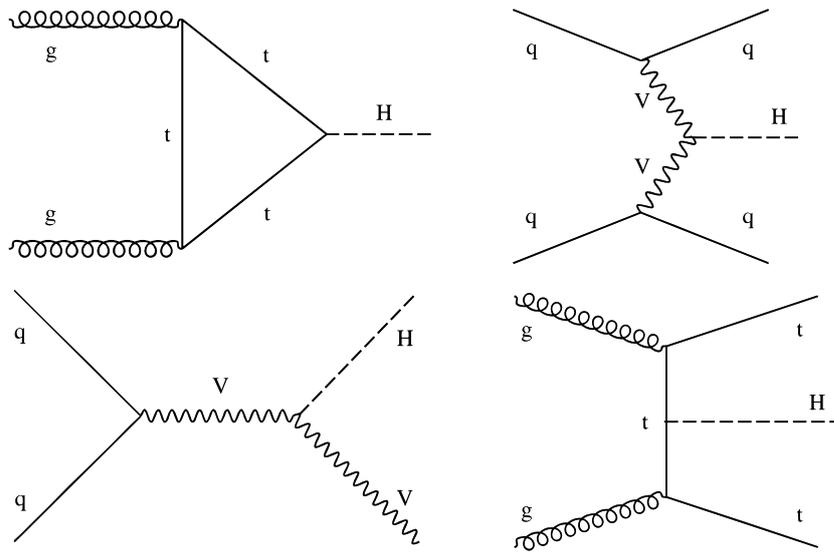


Figure I.3 – Feynman diagrams of the dominant production modes of the Higgs boson at the LHC: ggH (top left), VBF (top right), VH (bottom left), and ttH (bottom right).

After being produced, the Higgs boson decays through one of its decay channels to form the final states that will be observed by the detector. The probability of decay in a given channel is called branching fraction (\mathcal{B}_γ). The dominant decay channels and their branching fraction are reported in Table I.2, and the associated Feynman diagrams are shown in Fig. I.4. The largest \mathcal{B} is associated with

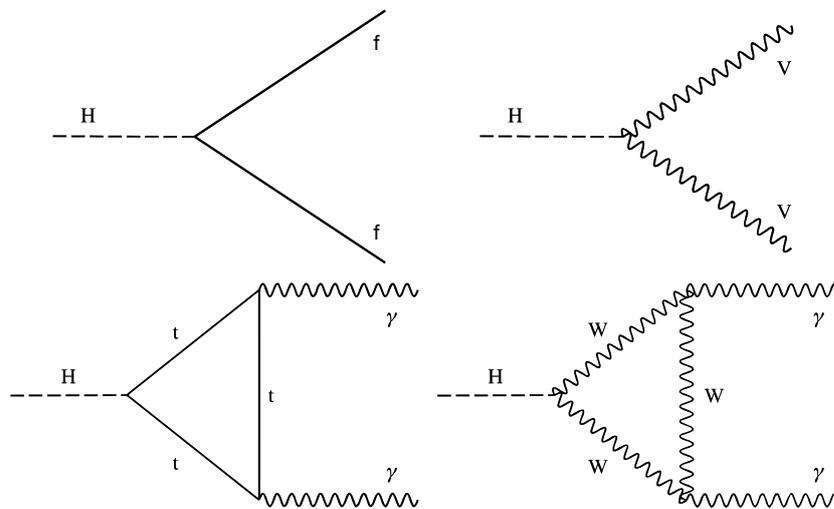


Figure I.4 – Feynman diagrams of the decay channels of the Higgs boson: into a pair of massive fermions (top left), into a pair of massive bosons (top right), and through loops of heavy particles to produce two photons (bottom).

the decay in a pair of b quarks, occurring in around 58% of the cases. However, the two most sensitive channels for measurements on the Higgs boson are its decay in a pair of Z bosons (then decaying to four leptons) and its decay in a pair of photons through loops of W bosons or t quarks. They both have a low branching fraction of around 0.23% and 0.01% for the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4l$ decay respectively, but their final state is fully reconstructed, and the LHC experiments have very good resolution on the energy resolution of leptons and

photons. These two decay channels brought a key sensitivity to the discovery of the Higgs boson, and they are currently the only channels allowing for a precise measurement of the Higgs boson mass (m_H). The most precise determination of m_H with the CMS experiment [17] is obtained for now from a combination of the measurements in these two channels for data obtained during the Run 1 plus 2016:

$$m_H = 125.38 \pm 0.14 \text{ GeV}. \quad (\text{I.37})$$

To check the consistency of the Higgs boson observed experimentally with the properties predicted by the SM, the number of observed events associated with a Higgs boson for a given production mode or a given decay channel are parameterised using the signal strength μ_X^Y defined as

$$\mu_X^Y = \frac{\sigma_X \mathcal{B}_Y}{\sigma_X^{\text{SM}} \mathcal{B}_Y^{\text{SM}}}. \quad (\text{I.38})$$

Measurements of the signal strength parameters are presented in Figs. I.5 and I.6 for different production modes (μ_X assuming $\mathcal{B}_Y = \mathcal{B}_Y^{\text{SM}}$, left) and decay channels (μ^Y assuming $\sigma_X = \sigma_X^{\text{SM}}$, right). The results obtained by the CMS experiment over the Run 2 (Fig. I.6) are compared to ATLAS and CMS results during Run 1 (Fig. I.5). Thanks to the large amount of data collected during Run 2, new production mode (tH) and decay channels ($\mu\mu$, $Z\gamma$) are analysed. Moreover, for the production modes and decay channels already considered during Run 1, the CMS experiment has increased its precision on the results by more than a factor 2. All measurements are compatible with the predictions of the SM.

Another parameterisation can be introduced at the coupling constant level to take into account the fact that a deviation of the coupling constant would affect both production modes and decay channels in a correlated way. This modifier κ_X is introduced in the couplings of fermions and bosons probed by the current analyses. The results of the fit of the κ parameters are shown in Figs. I.7 and I.8. The value of the coupling parameters is drawn as a function of the particle mass. An agreement is observed between measurements and expectations of the SM for the couplings of the Higgs boson already with data of Run 1 (Fig. I.7). This agreement is getting much clearer with the results of Run 2. Analyses of the Run 2 data recorded by the CMS experiment resulted in the observation of the Higgs boson decay to a pair τ leptons [19], to a pair of b quarks [20] and first evidence of the Higgs decay to a pair of muons [21], showing that this agreement holds over more than three orders of magnitude of mass. This is powerful proof that the observed Higgs boson is indeed the one introduced by the BEH mechanism.

One last argument to confirm the nature of the boson observed experimentally is to examine its spin-parity nature. Indeed, as the SM predicts the Higgs boson to be a scalar field, it should be the only elementary particle with quantum numbers $J^P = 0^+$, where J is the spin of the particle and P is the parity symmetry. Many alternative J^P hypotheses are tested and confronted with the SM hypothesis using Run 1 data [22]. The different alternatives can be compared using the kinematic variables of the decay products as they differ depending on the spin-parity hypothesis. The results are detailed in Fig. I.9. A test statistics

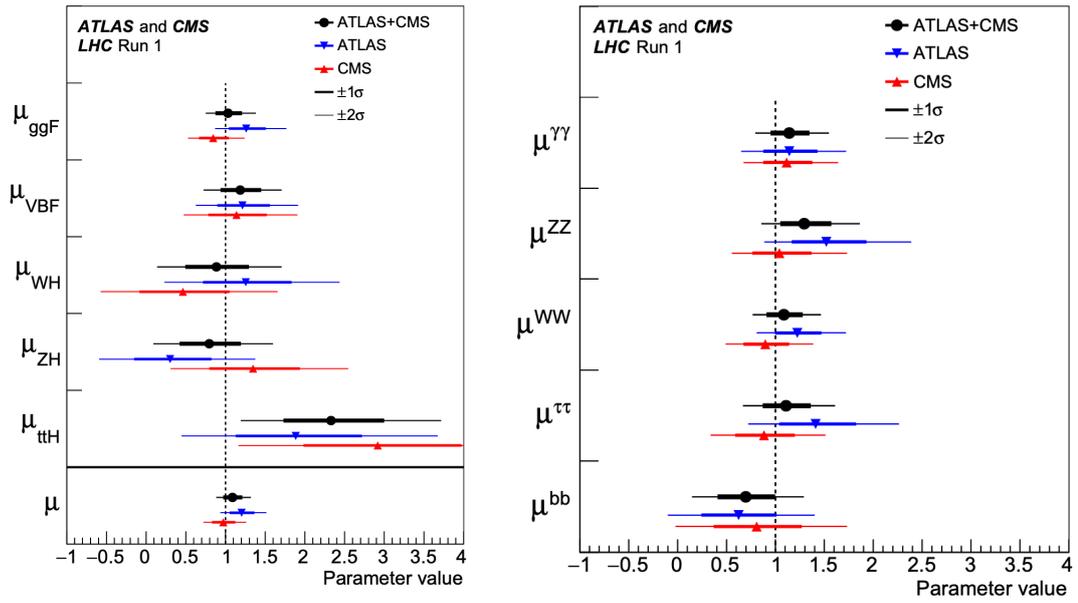


Figure I.5 – Value of the signal strength for several production modes (left) and decay channels (right) of the Higgs boson using data collected by the ATLAS and CMS experiments during the Run 1 of the LHC [18].

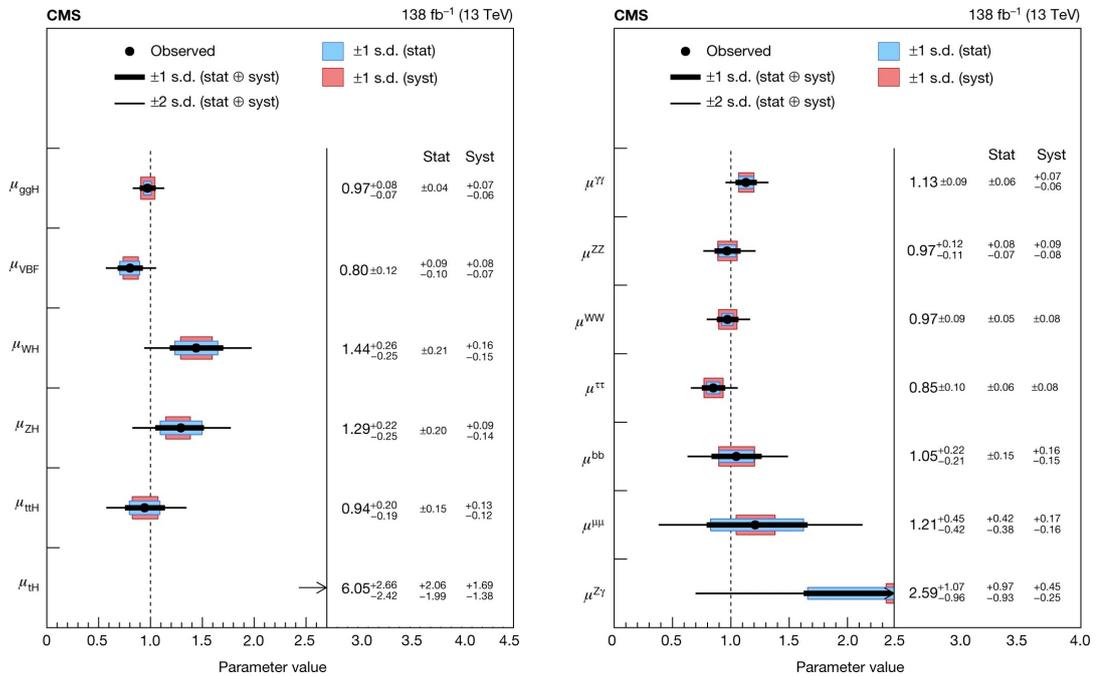


Figure I.6 – Value of the signal strength for several production modes (left) and decay channels (right) of the Higgs boson using data collected by the CMS experiment during the Run 2 of the LHC [16].

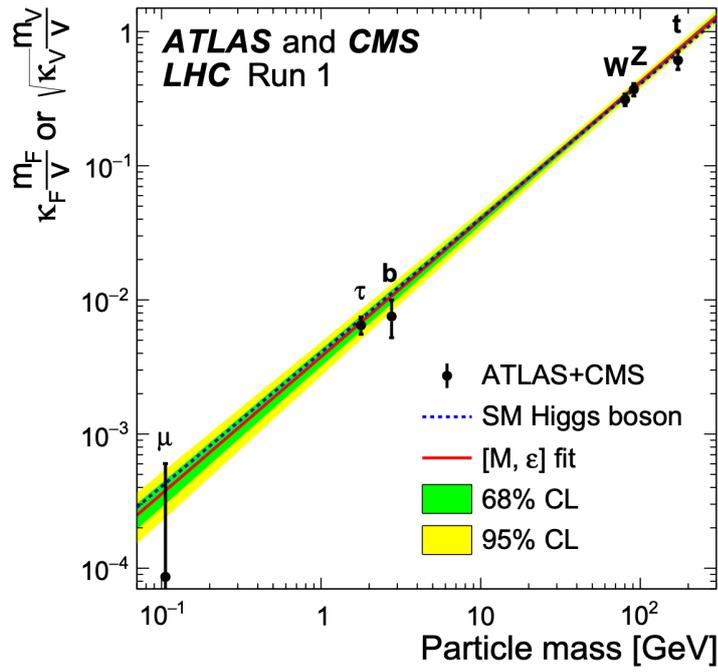


Figure I.7 – Coupling constants of the interaction of the Higgs boson with different particles represented as a function of the mass of these particles using data collected by the ATLAS and CMS experiments during the Run 1 of the LHC [18].

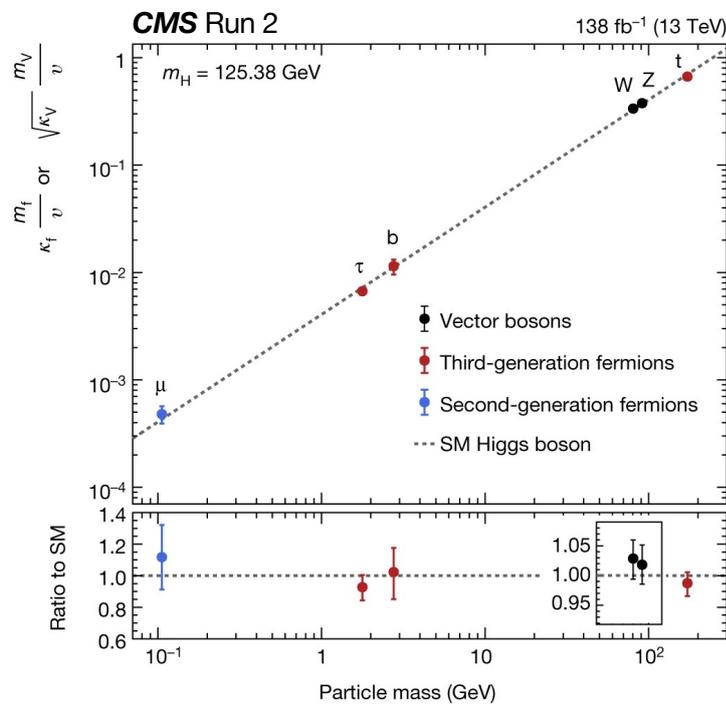


Figure I.8 – Coupling constants of the interaction of the Higgs boson with different particles represented as a function of the mass of these particles using data collected by the CMS experiment during the Run 2 of the LHC [16].

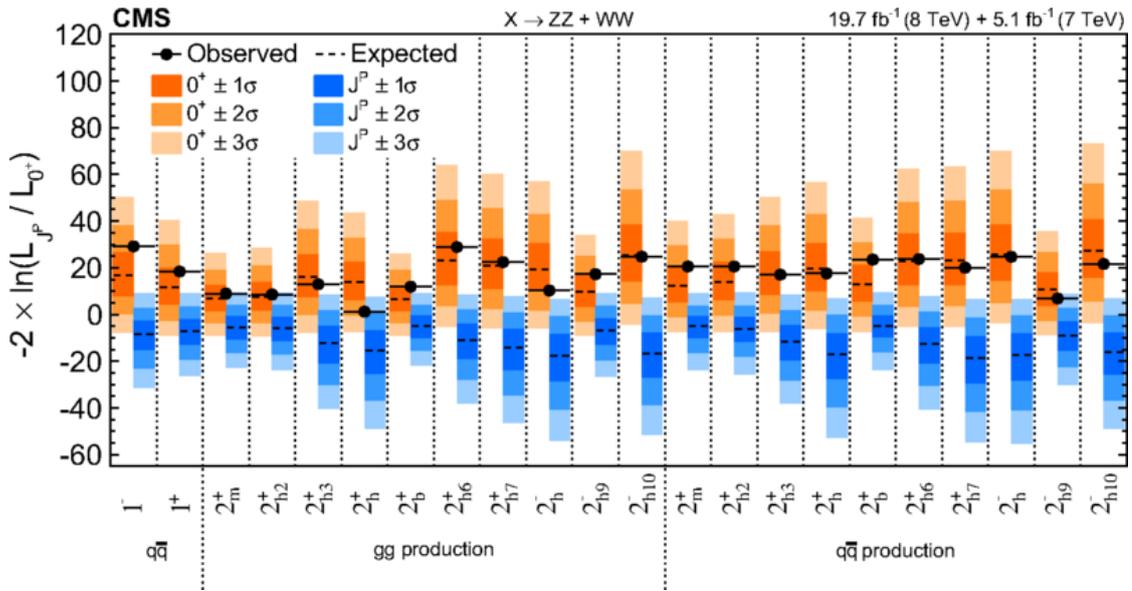


Figure I.9 – Distributions of the test statistics for alternative J^P models tested against the SM Higgs boson hypothesis in its ZZ and WW decay channels [22]. The expected median and the 68.3%, 95.3% and 99.7% CL regions are shown for the SM Higgs boson (orange) and for the alternative (blue) hypotheses. The observed q values are indicated by black dots.

$q = -2 \ln(L_{J^P}/L_{0^+})$ is defined, and the observed value of the test statistics is indicated by the black dot. As the decay of the Higgs boson has been observed in the diphoton decay channel, the $J = 1$ hypotheses are ruled out as a consequence of the Landau-Yang theorem [23, 24]. This is also confirmed in the ZZ and WW decay channels where the observed value of q excludes the $J = 1$ hypotheses at a greater than 99.999% confidence level (CL). For alternative $J = 2$ hypotheses, the same combination of decay channels is able to exclude them at a 99% CL or higher. Moreover, the same analysis excludes a pure pseudoscalar $J^P = 0^-$ hypothesis at a 99.98% CL. All of this indicates that the observed boson is consistent with a scalar SM-like Higgs boson. However, the constraints of the analysis are not enough to reject CP-violating couplings of the Higgs boson. This would be strong evidence of BSM physics.

Recent analyses of the Run 2 have tried to constrain further the possibility of CP-violating couplings of the Higgs boson. Both the interactions of the Higgs boson with gauge bosons and with fermions are studied to determine if they contain small contributions from CP-odd couplings called anomalous couplings (AC). Since the behaviour of the kinematic variables of the decay products depends on the CP-nature of the Higgs boson, different AC hypotheses can be evaluated. These analyses compare these AC hypotheses to the SM hypothesis (no CP-odd contribution) using Run 2 data and extract the most likely fractions of events resulting from these anomalous contributions. Results from an AC analysis in the Yukawa interactions are presented in Fig. I.10. This analysis observes the $t\bar{t}H$ production of the Higgs boson and its decay into two photons to determine if there is any anomalous contribution in the interaction between the Higgs boson and the top quark. Events are sorted among three bins depending on their kinematic variables, and results are reported in the scan of AC hypothe-

ses (f_{CP}^{Htt} , top right of Fig. I.10). The observed 68% CL region is $[-0.33, 0.33]$ for the value of f_{CP}^{Htt} . The anomalous couplings of the Higgs boson with the weak

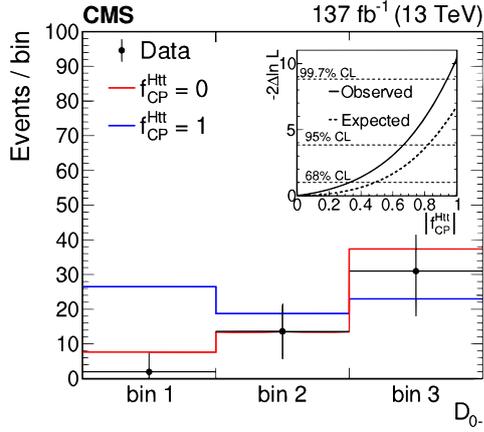


Figure I.10 – Distribution of $ttH(H \rightarrow \gamma\gamma)$ events in bins of a discriminant (D_{0-}) based on their kinematic variables [25]. The negative log-likelihood scan of the fraction of events originating from CP-odd couplings of the Higgs boson with t quarks (f_{CP}^{Htt}) is shown in the top right corner.

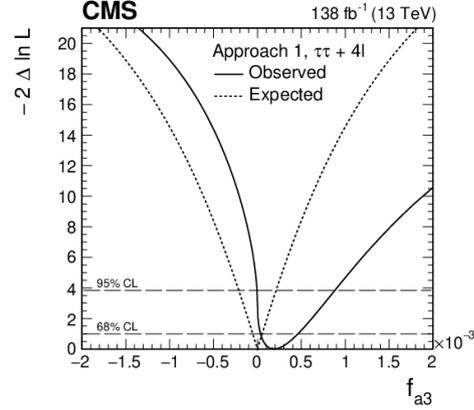


Figure I.11 – Negative log-likelihood scan of the fraction of events originating from CP-odd couplings of the Higgs boson with weak gauge bosons (f_{a3}) for the combination of $H \rightarrow ZZ$ and $H \rightarrow \tau\tau$ analyses [26].

bosons (Z, W) have been studied in its decay into four leptons [27] and into a pair of tau leptons [26], their combined results on the fraction of CP-odd Higgs boson events (f_{a3}) are shown in Fig. I.11. They report an observed 68% CL interval of $[-0.16 \times 10^{-3}, 0.26 \times 10^{-3}]$ for the value of f_{a3} . This thesis presents, in Chapter V, an additional way of constraining the value of f_{a3} by using the decay of the Higgs boson into a pair of photons.

I.3 Physics beyond the standard model

The SM offers a consistent theory depicting numerous physical effects. It has also been able to predict the properties of elementary particles with great accuracy. All the recent measurements performed at the LHC have shown results compatible with the expectations of the SM. Yet, a number of questions are left unanswered by the current formulation of the SM.

Some of these questions relate to the content of the SM. For instance, the scale of the electroweak symmetry breaking seems strangely fine-tuned, which is visible in the mass of the Higgs boson. In the SM, the physical mass observed experimentally is expected to arise from a bare mass getting very large quantum corrections [28]. These corrections are predicted to reach values of the order of the Planck scale $\approx 10^{19}$ GeV meaning that there is a compensation between two numbers over 10^{17} orders of magnitude to reach the observed mass of $m_H = 125$ GeV. It appears as an unlikely coincidence if not explained by an underlying effect. The Higgs boson mass is not the only one raising questions,

as neutrinos are considered massless in the SM. However, experiments such as Super-Kamiokande or the Sudbury Neutrino Observatory observed a phenomenon known as neutrino oscillation [29], indicating that each of the neutrino flavours has a nonzero mass. A recent measurement from Fermilab's Muon $g-2$ experiment confirmed the disagreement observed between SM predictions and the observed value of the $g-2$ anomalous magnetic moment of the muon. This value is computed from the loop corrections predicted by the SM to the scattering of muons on an external magnetic field. The latest value reported by the Muon $g-2$ experiment [30] shows a strong disagreement with respect to the latest theoretical previsions [31], hinting at effects from new physics.

The necessity to devise BSM theories also arises from observations of physics phenomena not described by the SM. Therefore, the SM can be considered as a solid foundation which should be included in a broader model. In particular, some cosmological observations are currently unexplained by the SM. A striking example is the gravitational interaction, which is the last fundamental force not encompassed by the SM. By observing the large-scale structure of the Universe, it was inferred that most of our universe consists of dark energy ($\approx 68\%$) and dark matter ($\approx 27\%$) [32]. Nevertheless, no particle of the SM constitutes a convincing candidate corresponding to dark energy or dark matter. The remaining 5% of the content of the Universe is composed of baryonic matter. This also raises a fundamental question: Why is there so much matter and so little anti-matter? Explanations for this asymmetry have been proposed by Sakharov [33], who described that for a universe out of thermal equilibrium, if the baryon quantum number and the CP symmetry are violated, then the Universe could cascade in a matter-dominated state. We know that the weak interaction is a source of CP violation, but this contribution is too small for the effects described by Sakharov. It means that the SM should contain other sources of CP violation. The interaction between the Higgs boson and other particles could be one of these sources, so it motivates measurement like the one presented in Chapter V.

All of these questions justify the necessity to design theories describing physics beyond the standard model. Therefore, one of the main objectives in the physics programme of the LHC experiments is now to determine where these theories could stem from by probing as many physical effects as possible and comparing the consistency of the results with the SM.

CHAPTER II

The CMS Experiment

II.1 The LHC and the CMS detector	24
II.1.1 The large hadron collider at CERN	24
II.1.2 A compact muon solenoid: the CMS detector	28
II.2 High-luminosity LHC and subdetectors adapted for timing in CMS phase II.	38
II.2.1 A luminosity upgrade for the LHC	38
II.2.2 CMS 2.0: probing a 4th dimension	40
II.3 DDMTD simulation for precise clock monitoring in CMS subdetectors	43
II.3.1 Clock distribution	43
II.3.2 Characterisation study of the lpGBT version 1	45
II.3.3 The DDMTD system	46
II.3.4 Simulating a DDMTD: <i>pyDDMTD</i>	49
II.3.5 Realistic jitter simulation	53
II.3.6 Evaluating corrections on the clock jitter	57
II.4 Conclusion	59

The large hadron collider (LHC) is a particle accelerator descending from a long list of machines and past experience probing the fundamental constituents of our Universe. It is operated by the European Organisation for Nuclear Research (CERN, from the French "Conseil Européen pour la Recherche Nucléaire") and located at the Swiss-French border. The specifications of the LHC, at the time of its design, were driven by its physics program: discover or rule out the existence of the Higgs boson predicted by the SM, push further the precision on the measurement of particle properties, probe sectors of the SM that would be sensitive to BSM physics.

Collisions at the LHC happen at four interaction points where four experiments record the subsequent production and decay of particles. One of these experiments is the compact muon solenoid (CMS), which provides the data and context for this thesis.

After approximately sixteen years of physics operations (which started in 2010), the LHC and its detectors will undergo significant upgrades to enter the high-luminosity phase of the LHC (HL-LHC). As there will be at least five times more collisions occurring each second, the CMS detector will have to deal with a harsher operating environment. One way to mitigate these effects in the CMS experiment and retain good precision on the relevant physics effects will be to rely on the time of arrival of particles in the subdetectors. Therefore, precise timing calibration and synchronisation will be needed over the entire CMS detector with the HL-LHC.

A description of the current operation of the LHC and CMS experiment is given in Section II.1 of this chapter. The upgrades planned for the HL-LHC and CMS are presented in Section II.2. Finally, my contribution to the timing calibration of CMS detectors in preparation for the HL-LHC is detailed in Section II.3.

II.1 The LHC and the CMS detector

II.1.1 The large hadron collider at CERN

Design

At this time, the LHC is the largest accelerator ever built, granting access to unprecedented levels of energy in the collisions of particles. This synchrotron is installed in the tunnel initially built to accommodate the large electron-positron collider (LEP). Thus, the particle beams of the LHC are accelerated through beam-lines of 27 km of circumference located 70 m to 140 m underground.

The LHC is primarily designed for proton-proton (pp) collisions. However, for about one month per year, heavy ions (lead ions) are accelerated and collide to probe the quark-gluon plasma, an exotic state of matter where quarks and gluons are deconfined. A one-month period of proton-lead collisions also took place beginning of 2013. In all cases, the acceleration path of the beams can be followed through the accelerator complex of CERN detailed in Fig. II.1.

The CERN accelerator complex Complexe des accélérateurs du CERN

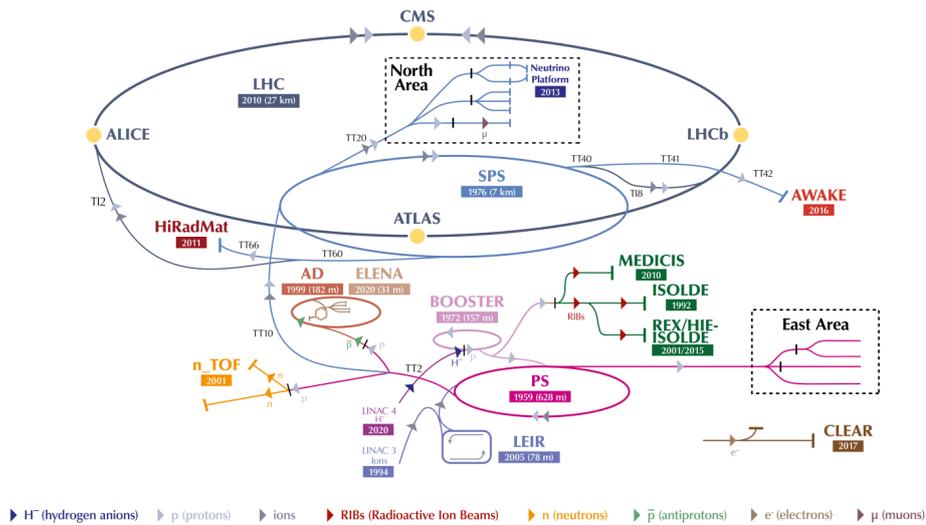


Figure II.1 – Illustration of the CERN accelerator complex [34].

The proton beam takes its source in a stock of negatively charged hydrogen ions (H^-) which are accelerated up to 160 MeV by the LINAC 4, a linear accelerator which replaced the LINAC 2 since 2020 to prepare for the HL-LHC. The beam then enters the proton synchrotron booster (BOOSTER in Fig. II.1 or PSB), the first circular accelerator of the chain. There, thanks to a strong magnetic field and a carbon stripping foil, more than 99% of the H^- ions are stripped of a pair of electrons, leaving a beam of proton boosted to 2 GeV. Then, the proton synchrotron (PS) accelerates the beam up to 26 GeV and passes it to the super proton synchrotron (SPS). The SPS is famous for allowing the discovery of the Z and W bosons in 1983 when it was colliding proton-antiproton beams. It is now in charge of accelerating the proton beam to 450 GeV before the LHC. The proton beam is split in two to circulate in opposite directions in the parallel beamlines of the LHC until collisions are started at the designated interaction points.

The proton beams of the LHC are accelerated with 16 radiofrequency (RF) cavities where electromagnetic fields modulated at a frequency of 400 MHz are synchronised with the passing of packets of protons (bunches). After the PS accelerator, bunches are 25 ns apart in time, corresponding to a frequency of 40 MHz, so the bunches are indeed synchronised with the fields of the RF cavities. In order to align their trajectory with the LHC ring and squeeze protons within each bunch, thousands of superconducting electromagnets are arranged along the beamlines. These magnets come in different shapes and sizes depending on their role. A total of 1232 15 m-long dipole magnets are used to bend the beams within the circular geometry of the LHC. Quadrupole magnets (392 in number and between 5 m to 7 m in length) focus the beams to increase the collision rate. Moreover, multipole magnets of higher orders correct imperfections in the spread of the protons around each bunch. All of these magnets are made of copper-clad niobium-titanium (Nb-Ti), so they are cooled down to 1.9 K (-271.3°C) using helium-4 to reach their superconducting state.

Operation

The maximal energy reached by the proton beams in the LHC increases with each data-taking period (called Run). Figure II.2 shows the evolution of the center-of-mass energy, starting at 7 TeV–8 TeV during Run 1 (2010–2013), reaching 13 TeV during Run 2 (2015–2018) and attaining now 13.6 TeV with Run 3 (2022–ongoing). In its high-luminosity phase, the LHC should reach its design energy of 14 TeV; however, as its name suggests, a focus is given to increasing the luminosity of the accelerator.

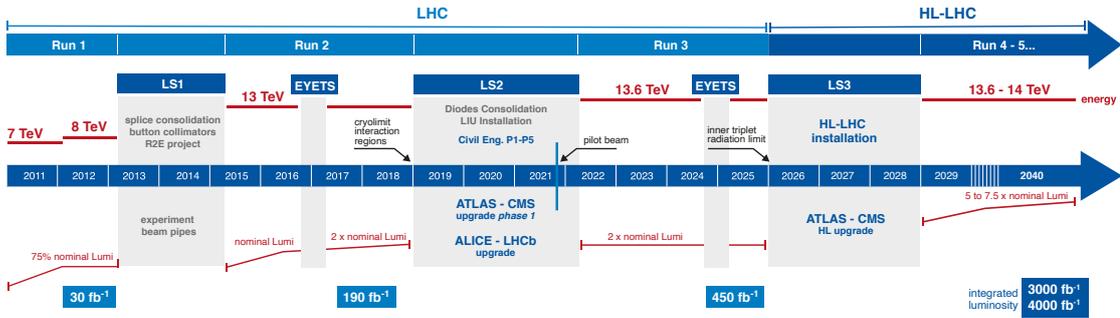


Figure II.2 – Summary of the LHC operations and projections for the HL-LHC phase. Each Run is separated by a long shutdown (LS) period to allow for accelerator and detector upgrades. They can also be interrupted during an extended year-end technical stop (EYETS).

Apart from the energy of the beam, an essential parameter of the beam is its luminosity. The luminosity is the representation of the rate of collisions happening between protons of the two beams. The luminosity is expressed in two ways, through the instantaneous luminosity \mathcal{L} expressed in $\text{cm}^{-2} \text{s}^{-1}$ or through the integrated luminosity $L = \int \mathcal{L} dt$ which is the instantaneous luminosity integrated throughout the experiment and expressed in fb^{-1} . The luminosity directly represents the expected number of times a given process should occur (such occurrence is called an event). Given a physical process with cross section σ_{proc} , the expected number of events is $N = \sigma_{\text{proc}} L$. Evolution of the instantaneous luminosity with each Run of the LHC is also presented in Fig. II.2 with a value of around $2 \times 10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$ for Run 3. The increase in total integrated luminosity recorded by the CMS experiment each year is summarised in Fig. II.3.

With a high rate of collisions, the average number of collisions occurring in the same bunch crossing (pileup, PU) increases. The average number of PU events is given by

$$\langle \mu \rangle = \frac{\mathcal{L} \sigma_{\text{in}}^{\text{pp}}}{n_b f_{\text{LHC}}} \quad (\text{II.1})$$

where σ_{pp} is the cross section of the inelastic pp collision process at 13 TeV, n_b is the number of bunches being accelerated in the LHC, and f_{LHC} is the revolution frequency. The nominal value for the number of bunches n_b is 2808, and the revolution frequency f_{LHC} is 11 245 Hz. With that in mind, the cross section of the inelastic pp process for different centre-of-mass energies is given in Fig. II.4, with the associated average PU separated year by year.

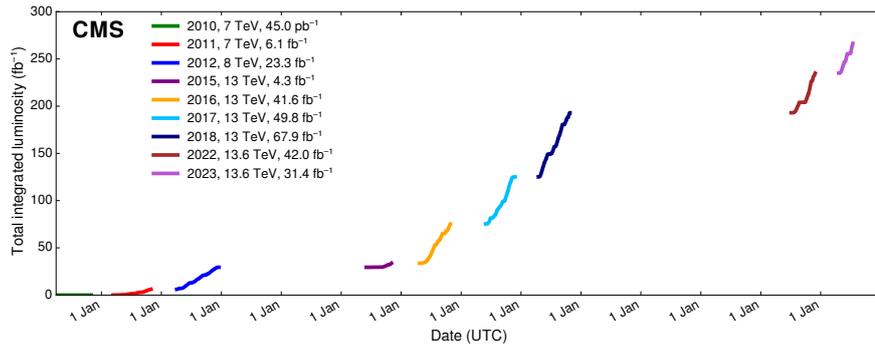


Figure II.3 – Luminosity of pp data delivered by the LHC as recorded by the CMS detector [35].

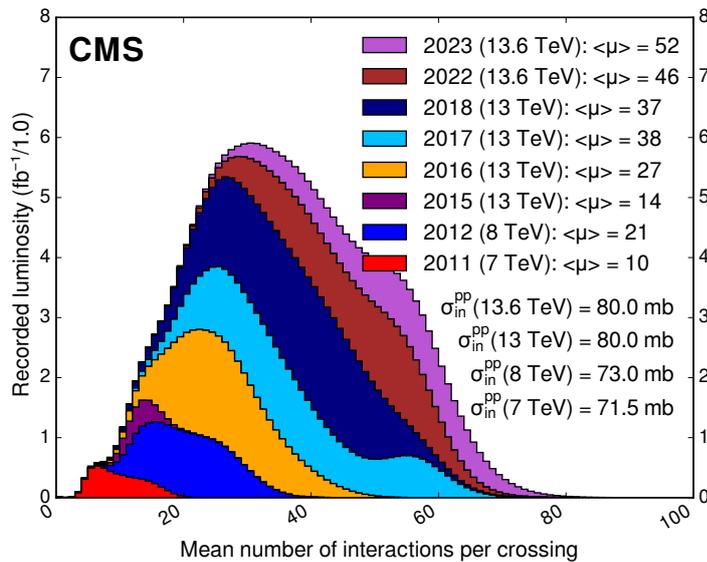


Figure II.4 – Distributions of the average number of interactions per bunch crossing separated by year [35].

Experiments

Four detectors collect collision data at four interaction points of the LHC, as described in Fig. II.1. The ATLAS (A Toroidal LHC Apparatus) and the CMS (Compact Muon Solenoid) are experiments with comprehensive physics programs utilising as much luminosity as possible to study rare processes. One of their primary goals is the study of the Higgs boson and its properties, and they reached a significant milestone in 2012 with the discovery of the Higgs boson. A detailed description of the CMS detector is given in Section II.1.2. The LHCb (LHC beauty) and ALICE (A Large Ion Collider Experiment) are the other two detectors operating at lower luminosity. LHCb specialises in heavy flavour quark physics, and some of its objectives are to study CP violation to search for BSM physics and observe rare hadrons of b and c quarks. Finally, ALICE relies on heavy ion collisions to focus on the strong interaction of the SM and probe the quark-gluon plasma, a state of matter that occurred at a very early stage of our Universe.

II.1.2 A compact muon solenoid: the CMS detector

The CMS experiment has a broad physics program. In addition to increasing the precision of measurements of observables described by the SM, one of its goals is to be able to observe any new physics phenomena produced from collisions at the LHC. The detector is cylindrical, with a central region called the barrel and both ends of the detectors forming the endcaps. Thus, it covers most of the solid angle around the interaction point. It consists of several concentric layers of subdetectors, each with a different specificity and dedicated to detecting different kinds of particles. All of the detector parts are presented in Fig II.5:

- **Tracking system:** its role is to identify the trajectory of charged particles (tracks) by reconstruction from the successive signal pulses (hits) triggered in the different layers of the tracker by the particles along their path. From these tracks, we can reconstruct the origin of the tracks (vertex).
- **Electromagnetic calorimeter (ECAL):** subdetector designed to collect the energy of electromagnetic showers. They originate from electrons or photons identified by gathering energy deposits as clusters in the ECAL.
- **Hadronic calorimeter (HCAL):** most hadrons leave only a small part of their energy in the ECAL; the HCAL is designed to collect the rest of the hadron energy.
- **Solenoid magnet:** it provides an intense magnetic field of 3.8 T within its volume, allowing to separate efficiently neutral particles from charged particles with a bent trajectory. By placing this superconducting magnet around the tracker, ECAL, and HCAL, less material impacts the particles, and it is easier to connect tracks in the tracker and energy deposits in the calorimeters.
- **Steel return yokes:** given the solenoid nature of the magnet, large iron panels are used to constrain, within the entire detector, the looping magnetic field outside of the solenoid. The magnetic field intensity around the return yokes is about 2 T. In addition to constraining the magnetic field, they also serve as absorber material for the muon chambers.
- **Muon chambers:** alternating with the steel return yokes are muon chambers. They are used for the tracking of the muons as these particles have little interaction with the rest of the subdetectors.

The constituents of the subdetectors are described in more detail in this section.

A coordinate system common to all the subdetectors of CMS is represented in Fig. II.6. Its origin is defined at the interaction point (IP) where the two beams collide, and both a cartesian or cylindrical right-handed system can be used. The cartesian system is defined with the x coordinate pointing towards the centre of the LHC and the z coordinate in the direction of the beamline pointing anti-clockwise when looking at the LHC from above. The kinematic properties of a particle within the detector are often described through the cylindrical coordinate system, with a preference for coordinates not dependent on the boost of

CMS DETECTOR

Total weight : 14,000 tonnes
 Overall diameter : 15.0 m
 Overall length : 28.7 m
 Magnetic field : 3.8 T

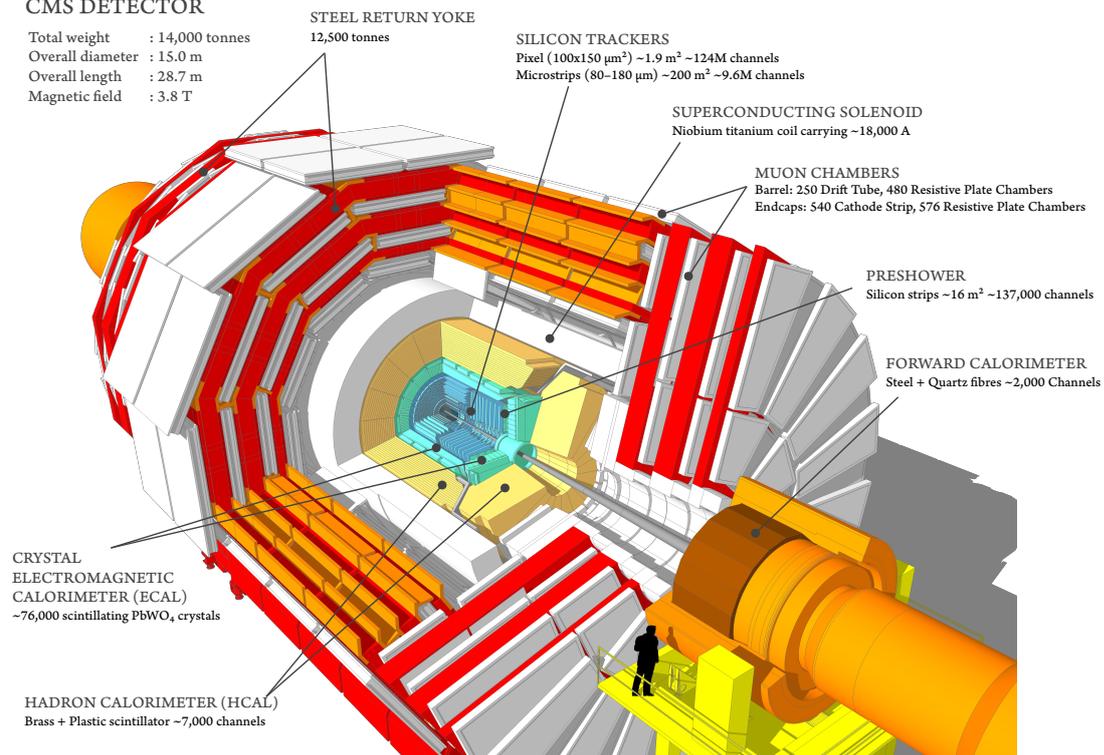


Figure II.5 – Schematic view of the CMS detector detailing all its subdetectors [36].

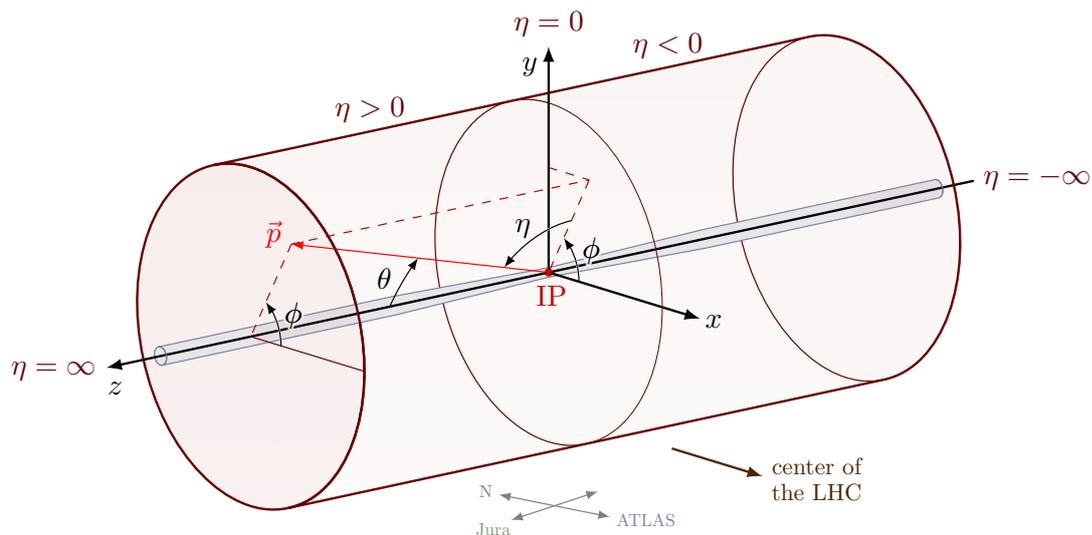


Figure II.6 – The CMS coordinates system [37].

acteristics of the silicon modules change depending on their distance to the IP. The PIXEL part of the detector is the closest to the IP ($29 \text{ mm} < r < 10 \text{ cm}$) and consists of pixel cells of size $100 \mu\text{m} \times 150 \mu\text{m}$. It grants a spatial resolution on the measurement of the vertex position of $15 \mu\text{m}$ to $20 \mu\text{m}$. Further, come the tracker inner barrel (TIB) and the tracker inner disks (TID) ($20 \text{ cm} < r < 55 \text{ cm}$) with silicon microstrips of typical size $10 \text{ cm} \times 80 \mu\text{m}$. At the edge of the barrel ($55 \text{ cm} < r < 120 \text{ cm}$) is the tracker outer barrel (TOB) with larger silicon strips of size $25 \text{ cm} \times 180 \mu\text{m}$. Finally, the tracker endcaps (TEC) cover the region $|z| > 118 \text{ cm}$ with silicon strips in the form of disks with a radial size between $100 \mu\text{m}$ and $180 \mu\text{m}$. Each layer of pixel cells or strips is shifted with respect to the previous one to reach the highest acceptance.

Electromagnetic calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) of the CMS experiment [41] encloses the tracking system and is dedicated to measuring the energy of photons and electrons. Consequently, the performance of the ECAL is directly impacting the performance of the $H \rightarrow \gamma\gamma$ analysis in the CMS experiment. The subdetector relies on lead-tungstate (PbWO_4) scintillating crystals, a single medium to act as absorber (*i.e.* force electrons and photons to deposit their energy) and active (collect these energy deposits) material. Photons and electrons leave their energy by producing electromagnetic showers when interacting with the lead nuclei of these dense crystals. By using a homogeneous calorimeter, the CMS detector has access to the full deposited energy and is less impacted by fluctuations due to non-measured showers. When luminescent centres [42] of the crystals are excited by these showers, they emit a signal in the form of light, which is collected on the crystal side opposite to the IP. This means that the crystals need to be as transparent as possible to ensure high efficiency in the light collection.

The longitudinal expansion of the electromagnetic showers depends on the radiation length X_0 of the absorber material. This radiation length is defined as the typical length after which an electron loses most of its energy ($\approx 63\%$ of lost energy). In a material with a small radiation length, electromagnetic showers are narrower, allowing for a better separation of electromagnetic candidates. With a high density (8.28 g cm^{-3}), small radiation length (8.9 mm), and fast signal production (80% of the light is produced in 25 ns in an ECAL crystal), the lead-tungstate crystals offers a compact and highly granular solution for the ECAL. However, these crystals have a relatively low light yield ($\approx 30\gamma$ produced per MeV), so their signal is amplified with silicon avalanche photodiodes (APD) in the barrel and vacuum phototriodes (VPT) in the endcaps.

The disposition of the ECAL crystals is presented in Fig. II.8. About 61 200 trapezoidal crystals of surface $22 \text{ mm} \times 22 \text{ mm}$ covers the barrel region $|\eta| < 1.479$. For each endcap $1.479 < |\eta| < 3.0$, a total of 7324 crystals of surface $28.6 \text{ mm} \times 28.6 \text{ mm}$ are used. The mean size of the crystals in the ECAL is $\approx 23 \text{ cm}$ corresponding to $\approx 26X_0$. The size and alignment of the crystal are carefully chosen. All the crystals point towards the nominal IP in a quasi-projective geometry, but with a 3° angle to avoid acceptance gaps between the crystals. Their surface is computed

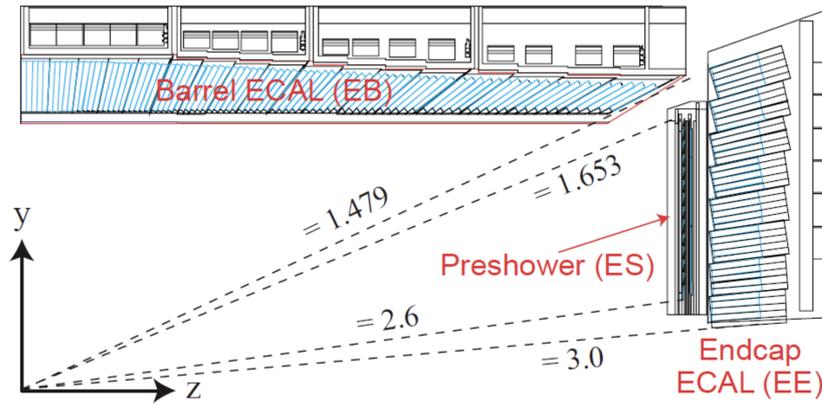


Figure II.8 – Illustration of a quadrant of the ECAL subdetector of CMS from Ref. [43]. The three main structures of the ECAL are represented: crystals in the barrel (EB) and endcap (EE) regions and the preshower lead/silicon-strips detector (ES).

to confine the longitudinal expansion of the electromagnetic shower produced by lead-tungstate. Therefore, the total volume of the crystal contains more than 98% of the electromagnetic deposit delivered by photons and electrons with energy up to 1 TeV. A preshower detector made of a lead absorber and silicon strips is located in front of each of the endcaps in the region $1.653 < |\eta| < 2.6$. The silicon sensors measure the x and y coordinates of electrons and photons before they shower in the ECAL crystals.

The evaluation of the performance of the ECAL is described in detail in Ref. [44]. The resolution of an electromagnetic calorimeter is parameterised as:

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

where N is the noise term due mostly to the readout electronics, S is the stochastic term arising from statistics fluctuations in the showers or from the energy lost in the absorber, and C is a constant term due to the channel-to-channel intercalibration as well as to the time variation of the channel response. For the ECAL of the CMS detector, they are measured to be $N = 0.12 \text{ GeV}$, $S = 0.028 \sqrt{\text{GeV}}$ and $C = 0.3\%$ from showers reconstructed in 3×3 crystals during electron beam tests [45]. The low value of the stochastic term is possible thanks to the homogeneous nature of the ECAL. Its resolution during Run 2 after calibration is shown in Fig. II.9 from the invariant mass distribution of electron pairs coming from the Z boson decay.

The high radiation level reaching the ECAL crystals damages the crystalline structures, creating coloured centres that reduce the crystal transparency. Thus, the crystal transparency is monitored during data-taking periods with a laser monitoring system [47]. Thanks to this system, corrections are applied to the crystal response to take into account the loss of transparency. These corrections must be applied every 40 min to keep the impact of the transparency loss on the constant term C below 0.2%.

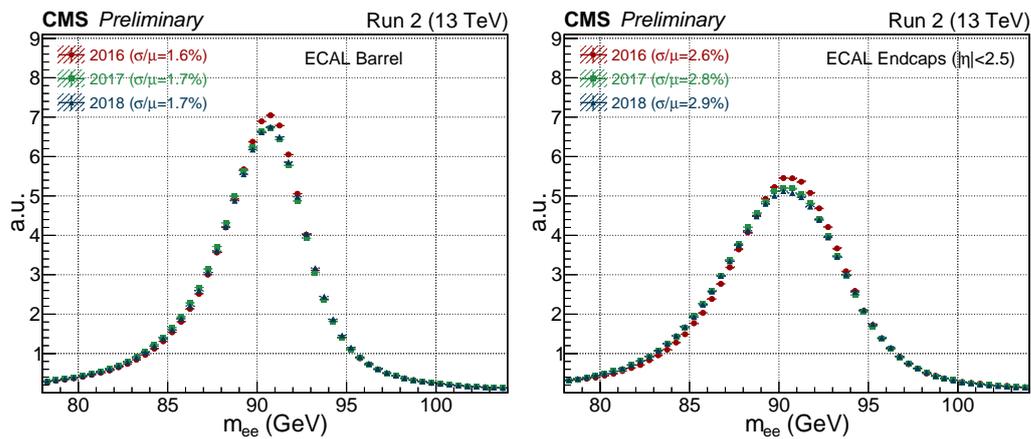


Figure II.9 – Resolution of the ECAL from the invariant mass distribution of electron pairs coming from the Z boson decay for each year of the Run 2 [46].

Hadronic calorimeter (HCAL)

The hadronic calorimeter (HCAL) of the CMS experiment [41] surrounds the ECAL and is the last subdetector with structures within the solenoid magnet. The HCAL is responsible for the detection of hadrons. Because of QCD colour confinement, quarks and gluons cannot exist as free states. They immediately hadronise after collisions to form hadrons, and instead of single partons, narrow jets made of hadrons and photons are observed in the detectors. To collect the energy of the jets, a dense absorber material is used to increase the probability of interaction of the hadrons with the medium, and a scintillating material gives information on the quantity of energy deposited. Similarly to the radiation length X_0 used for electrons and photons, a relevant length is defined for hadrons with the hadron interaction length λ_i . This time, λ_i represents the average distance a hadron can cross through a medium before interacting with a nucleus.

The nature of the absorber and scintillating materials depends on the targeted region of the detector. The different structures of the HCAL are illustrated in Fig. II.10 and summarised here:

- the barrel hadronic calorimeter (HB) and endcap hadronic calorimeters (HE) are located inside the solenoid magnet and cover respectively the $|\eta| < 1.4$ and $1.3 < |\eta| < 3$ regions. They rely on alternations of brass absorbers and plastic scintillators connected to hybrid photodiodes. The HB and HE are between $7\lambda_i$ and $10\lambda_i$ thick.
- the outer hadronic calorimeter (HO) is used to increase the longitudinal confinement of hadronic showers. It is located outside of the solenoid magnet and covers a $|\eta| < 1.4$ region. The scintillator is placed in the steel yoke, which plays the role of absorber material and is read out by silicon photomultipliers (SiPM).
- the forward hadronic calorimeters (HF) are located close to the beamline, covering the $3 < |\eta| < 5.2$ region. It uses steel absorbers and quartz fibres emitting Cherenkov light, which is collected by photomultiplier tubes (PMT).



Figure II.10 – Layout of a quadrant of the HCAL [48] and of its main structures: the barrel HCAL (HB), the endcap HCAL (HE), the outer HCAL (HO) and the forward HCAL (HF). The tracker system and ECAL are also illustrated in the bottom left corner.

Given that hadrons can already start showering in the ECAL, and some of the energy is not deposited in the HCAL, the global resolution of the ECAL+HCAL system on the energy measurement is reported. Performance is evaluated by using the resolution on the energy E of pions in the range 2 GeV to 350 GeV [49]:

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\% .$$

The first term is associated with stochastic effects (*e.g.* fluctuation in the shape and energy deposited by the hadronic showers), while the constant term accounts for effects independent of the energy scale (*e.g.* imperfect calibration of the calorimeters).

Muon chambers

The muon chambers of the CMS experiment [41] are the outermost subdetector. The muons targeted by physics analyses of the CMS collaboration have energy in the range of a few GeV to hundreds of GeV. Within this range, their energy loss through the multiple layers of the CMS detector is generally small. At the same time, almost all other particles are absorbed (apart from neutrinos), and muons are the only particles reaching the external layers of the detector.

The muon system is outside of the solenoid magnet, but it is intertwined with steel return yokes so the residual magnetic field still has an intensity of around 2 T, allowing the bending of the trajectory of muons and, thus, the measurement of their p_T and electric charge. However, the magnetic field behaviour within a muon chamber depends on its location. To cope with the differences in magnetic field and particle rates, three different strategies were designed for the muon system: drift tubes (DT) chambers, cathode strip chambers (CSC), and resistive plate chambers (RPC). All of these strategies rely on gas ionisation. Their location in a quadrant of the CMS detector is shown in Fig. II.11.

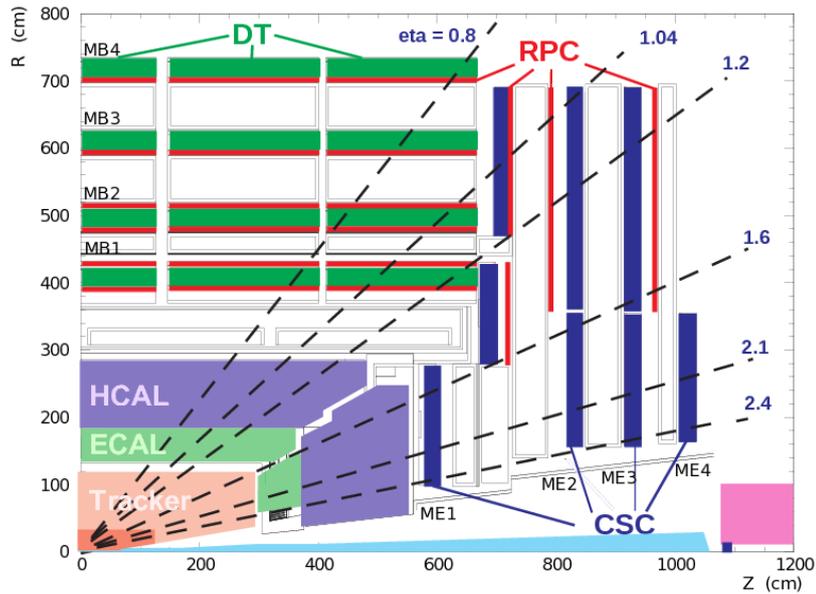


Figure II.11 – Depiction of a quadrant the CMS detector from Ref. [50] focusing on the three technologies employed for the muon system: drift tubes (DT) chambers, cathode strip chambers (CSC), and resistive plate chambers (RPC).

The DT chambers are in the barrel region with $|\eta| < 1.2$. They are arranged in 4 stations (MB1 to MB4 in Fig. II.11), each divided in 12 chambers (except for MB4 which has 14). They are filled with a mixture of argon (85%) and CO_2 (15%). Their spatial resolution is between $80 \mu\text{m}$ and $120 \mu\text{m}$ in the (r, ϕ) plane and between $200 \mu\text{m}$ and $300 \mu\text{m}$ in the z direction.

In the endcaps region ($0.9 < |\eta| < 2.4$), chambers endure a large muon rate and non-uniform magnetic field. Hence the choice of cathode-strips-based detectors (CSCs), which have a very fast response time. They are arranged in 4 stations (ME1 to ME4 in Fig. II.11) in the form of disks. These multi-wire chambers are filled with argon (40%), CO_2 (50%), and CF_4 (10%). They provide a spatial resolution between $30 \mu\text{m}$ and $150 \mu\text{m}$.

The RPCs overlap with other chambers in the barrel and endcap regions in order to add robustness and redundancy to the muon trigger. They have a moderate spatial resolution but an excellent time resolution of less than 3 ns , thus helping to measure the correct beam-crossing time. They consist of double-gap chambers filled with a mixture of $\text{C}_2\text{H}_2\text{F}_4$ (95.2%), $i\text{-C}_4\text{H}_{10}$ (4.5%) and SF_6 (0.3%), and operated in avalanche mode to ensure good operation at high rates.

Evaluation of the performance of the muon system using pp collisions at $\sqrt{s} = 7 \text{ TeV}$ is described in detail in Ref. [51].

Particle reconstruction and identification

The standard reconstruction of objects in the CMS detector is described in this section. A more detailed description of the object reconstruction and identification in the context of the $H \rightarrow \gamma\gamma$ analysis is given in Section V.2 of Chapter V.

This standard reconstruction relies on the particle flow (PF) algorithm [52]. Since each type of particle (photon, electron, muon, charged or neutral hadron) leaves a specific combination of signals in the subdetectors, as illustrated in Fig. II.12, the PF algorithm uses associations of information from all subdetectors to reconstruct the particles. This information is separated into two classes of basic elements: the tracks and the energy deposits in the calorimeters (clusters). The tracks are reconstructed from a series of hits in the tracker and are connected to a reconstructed vertex. Clusters are groups of crystals (in the ECAL) or scintillator segments (in the HCAL) that hold the energy of a single electromagnetic or hadronic shower.

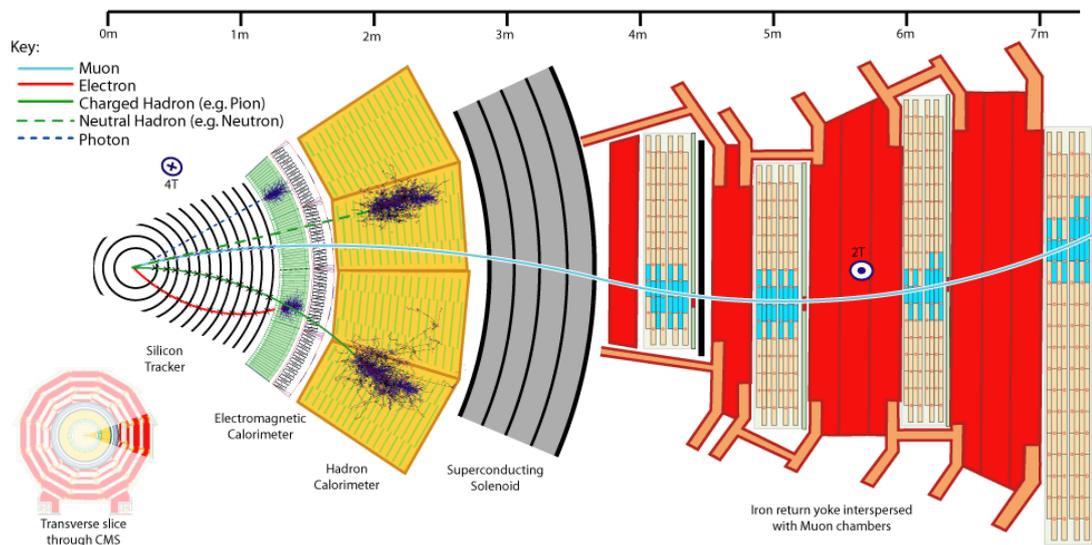


Figure II.12 – Slice of the CMS detector illustrating the expected interaction of each particle type with each subdetector [53].

Once the tracks and energy clusters of the event have been identified, the PF algorithm attempts a geometric association of the different elements. The strategy of the PF algorithm can be summarised as follows:

- First, if tracks reconstructed in the muon chambers and in the inner tracker are compatible, they form a muon candidate. These tracks and the potentially associated clusters along their way are removed from the following steps of the algorithm.
- The second step is dedicated to the reconstruction of electrons. Electrons have bent tracks in the inner tracker and electromagnetic deposits in the ECAL. A selection on the quality of electron candidates is defined from an additional algorithm dedicated to electron reconstruction [54] using the tracker and ECAL information. If the candidate passes the selection, its tracks and clusters are removed from the next steps.

- Then, the remaining tracks of the inner tracker should belong to charged hadrons depositing energy in the HCAL. The compatibility between the track momentum and the energy deposited in the HCAL is checked. Typical discrepancies between the two come from excesses of energy in the cluster. This indicates that overlapping neutral particles left energy within the same cluster in the HCAL. To account for this, a photon (or, if needed, a neutral hadron) is created, carrying the difference in energy.
- Finally, the remaining energy clusters in the ECAL are associated with photons and those of the HCAL are associated with neutral hadrons.

Trigger system

This section summarises the global trigger system in the CMS experiment [55]. The triggering of the CMS detector to collect $H \rightarrow \gamma\gamma$ events is described in Section V.2 of Chapter V.

Collecting the full information of all subdetectors of the CMS experiment for each collision occurring at 40 MHz would require an unattainable data throughput and a hardly-conceivable amount of storage. Moreover, most of these data are associated with uninteresting events (associated with well-known physics). That is why a trigger system ensures that the CMS experiment records only events of potential interest. The CMS trigger system performs an online events selection in two steps thanks to the level-1 (L1) trigger and the high-level trigger (HLT).

The goal of the L1 trigger is to limit the event rate to 100 kHz, the limit of the readout electronics. The L1 trigger is based on a simplified version of the information of the subdetectors. An L1 trigger menu of about 440 algorithms is used to look for specific signatures in the detector response (*e.g.* cluster in the ECAL with a certain energy level, ...) corresponding to an object then passed to the HLT. The L1 trigger works at a fixed latency of 4 μ s to decide on accepting or rejecting an event. This decision is based on information from calorimeters and muon detectors, and the triggering process is described in Fig. II.13. Trigger primitives are computed from a fast readout and a limited granularity on energy deposits in the ECAL and HCAL and on track segments and hit patterns in the muon chambers (DT, CSC, RPC). From the trigger primitives, coarse versions of the physics objects are built: muon candidates from the muon trigger and electron, photons, jet, or tau candidates from the calorimeter trigger. Finally, the combined information is used by the global trigger and depending on the trigger menu, the event is discarded or processed by the HLT.

The HLT is based on software close to offline reconstruction, identification and analysis and relies on the full readout of the detectors. For instance, the object reconstruction follows a simplified version of the PF algorithm. However, the HLT must trigger at a high rate, so all these software are whittled down to their elementary steps to gain computational efficiency. The HLT rate is, on average, around 1 kHz, and it delivers the complete raw detector data for permanent storage and offline steps. This high rate is also achievable thanks to a dedicated computing farm. With the start of Run 3, some graphics processing units (GPU)

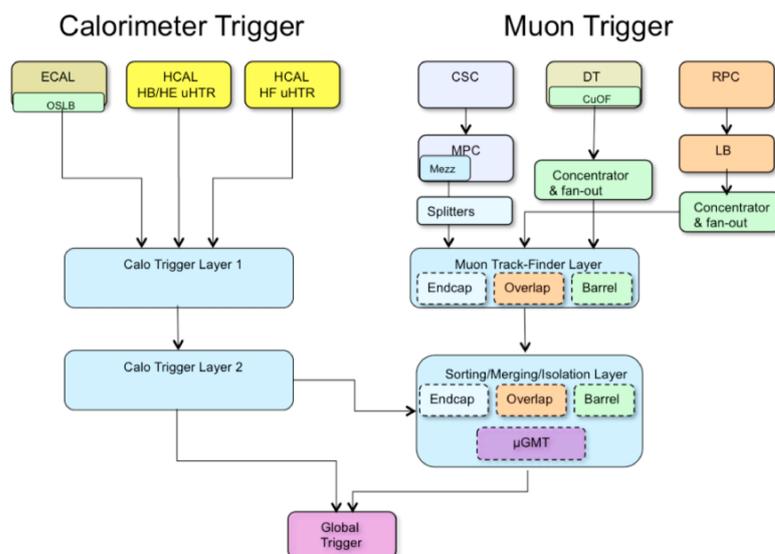


Figure II.13 – Diagram of the L1 trigger system for the CMS detector [56].

are used in addition to many central processing units (CPU) already present in previous runs. It amounts to a total of 25 600 CPU cores and 400 GPU cores. The GPU cores are particularly efficient for neural network implementations. With the development of advanced machine learning techniques for particle reconstruction or identification, some high-performing networks are now included in the HLT step. HLT candidates generally come from L1 trigger objects and are tested by multiple algorithms (HLT paths) to decide on the trigger. There are more than a hundred HLT paths, each corresponding to different usage of the events: some paths collect events for physics analyses and others for data quality monitoring, detector calibration, ...

II.2 High-luminosity LHC and subdetectors adapted for timing in CMS phase II

II.2.1 A luminosity upgrade for the LHC

With the end of Run 3 planned for the end of 2025, the latest previsions foresee a total integrated luminosity of $\approx 300 \text{ fb}^{-1}$ over sixteen years of physics operation. The LHC already achieved some of its goals in this period, with the discovery of the Higgs boson in 2012 and many measurements with record precision (all confirming the robustness of the SM so far), thanks to its four large-scale experiments. However, after this point, the physics benefits from collecting data at this rate will become minimal. That is why the LHC and its experiments will undergo significant upgrades during the long shutdown 3 to enter a high-luminosity phase. The goal for the HL-LHC is to reach an instantaneous luminosity between $5 \times 10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$ and $7.5 \times 10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$ during this phase and to provide an integrated luminosity of 250 fb^{-1} per year. At the end of the high-

luminosity phase (after a dozen years), it should amount to a total integrated luminosity between 3000 fb^{-1} and 4000 fb^{-1} .

To achieve such luminosity, an upgrade of the LHC is needed, equivalent to installing new parts of an accelerator over 1.2 km in various places of the current ring. The main modifications come from the improvement of quadrupole magnets and the addition of crab cavities around the ATLAS and CMS detectors. The quadrupole magnets are responsible for squeezing bunches. This is particularly important around the interaction points where more focused beams mean more chances of collisions. As mentioned in the previous section, the current magnets of the LHC are made of copper-clad niobium-titanium, forming a magnetic field of up to 9 T. To achieve even higher magnetic fields at the HL-LHC, quadrupole magnets around the interaction points of the LHC will be replaced by intermetallic niobium-tin (Nb_3Sn) magnets producing fields of up to 11 T [57]. In addition to the more focused beams, crab cavities are proposed as the baseline solution to increase the overlap between crossing bunches at the interaction point [58]. The crab cavities are additional RF cavities also creating a 400 MHz electric field. However, their role is not to accelerate bunches but to give them torque to introduce a rotation and align the colliding bunches. On the opposite beamline, a crab cavity gives a torque in the opposite direction to recover the initial alignment of the beam and ensure stability. This process is illustrated in Fig. II.14.

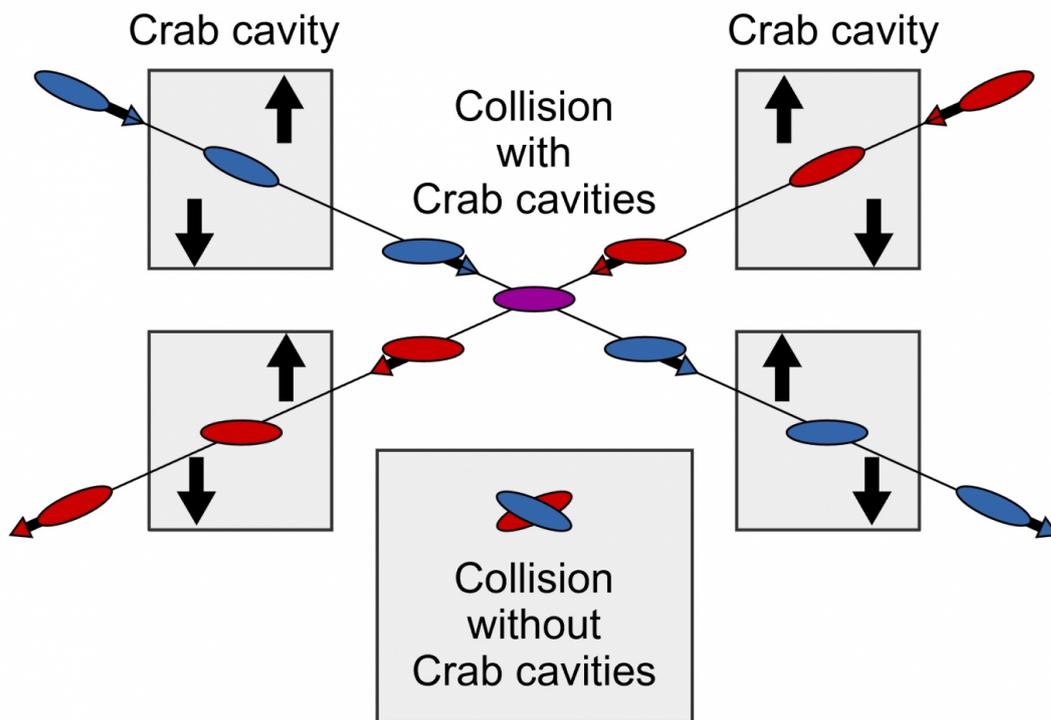


Figure II.14 - Illustration of the rotation introduced by the crab cavities. Bunches are flying sideways towards the interaction point, hence the name of the crab cavities.

II.2.2 CMS 2.0: probing a 4th dimension

During the HL-LHC phase, the goal of the CMS experiment is to maintain its excellent performance in terms of efficiency and resolution on object reconstruction and identification. However, as described by Eq. II.1, increasing the instantaneous luminosity means increasing the average PU. The PU level recorded by the CMS detector during the first three runs of the LHC is shown in Fig. II.4 and has a value of ≈ 40 in the nominal luminosity setting of the LHC ($\mathcal{L} = 10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$). By increasing the instantaneous luminosity by a factor 5, we expect an average PU of ≈ 200 . At this level of PU, it will become a significant challenge for the CMS experiment. Another challenge will be the radiation damage to the detector.

To overcome these challenges, a set of upgrades is studied for several subdetectors of the CMS experiment. The addition of a new detector system is planned: the minimum ionising particle (MIP) timing detector (MTD). A major idea to increase the PU rejection during the HL-LHC phase is to rely on the timing information provided by the CMS subdetectors. Currently, the spatial resolution of the tracking system is good enough to allow the PF algorithm to correctly connect particles of an event to their primary vertex with a low fake rate. However, it will not be enough under the harsher conditions of the high-luminosity phase. However, by identifying the time of arrival of particles, the PF reconstruction can recover performance close to what it achieved during Run 2. This is demonstrated in Fig. II.15. The resolution on the timing information of particles is determined

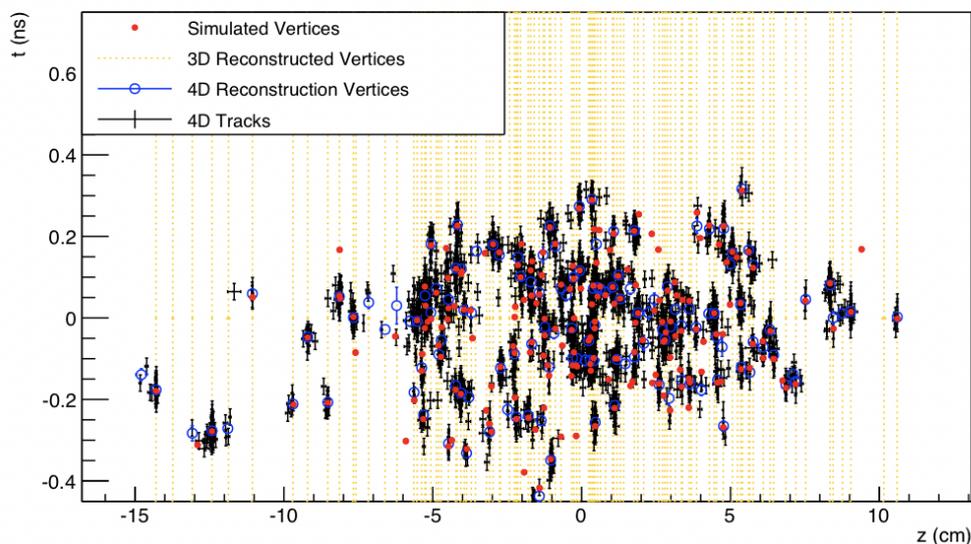


Figure II.15 – Evaluation of the vertex reconstruction performance when assuming timing detectors with ≈ 30 ps time resolution as presented in Ref. [59].

to get back to PU levels of 40–60. By considering the beam spot sliced into time exposures, these PU values are recovered with 30 ps to 40 ps time exposures.

A summary of the upgrades planned for phase 2 of the CMS detector is given in this section. The baseline upgrade plan is presented in detail in Ref. [60].

Tracker system

Because of significant radiation damages during its operation in phase 1 of the LHC, the tracker system will be replaced before the high-luminosity phase [61]. Moreover, modifications of the components of the tracker modules are also planned. The PIXEL system will be reinstalled with smaller pixels of $25\ \mu\text{m} \times 100\ \mu\text{m}$ or $50\ \mu\text{m} \times 50\ \mu\text{m}$. In the rest of the system, the silicon strips will be shortened by a factor 4 without changing their width. As the full detector will be about four times more granular than its current design, this upgrade will improve the separation between tracks close to each other, for instance, in boosted jets, and should compensate for the higher PU. Finally, with additional disks in the end-cap regions, the coverage of the tracker will be extended to $|\eta| < 4$.

Calorimeter endcaps

The electromagnetic and hadronic calorimeter endcaps will also be replaced during the long shutdown 3 because of their loss of transparency due to radiations. A new system is designed to serve as a replacement for these calorimeter endcaps: the high granularity calorimeter (HGCal) [61]. The layout of the HGCal is presented in Fig. II.16. The electromagnetic section (CE-E) consists of tungsten

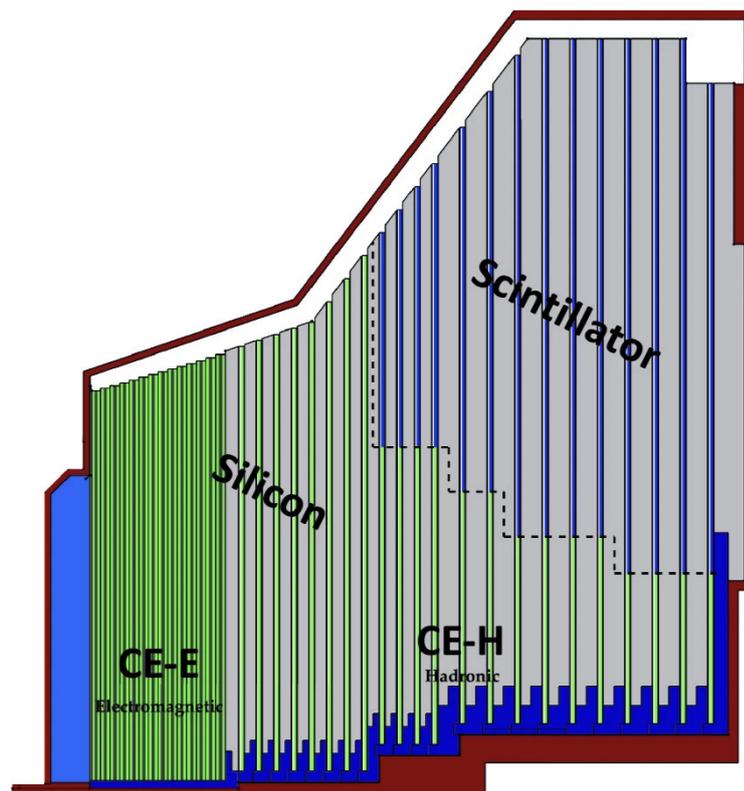


Figure II.16 – Arrangement of the electromagnetic (CE-E) and hadronic (CE-H) sections of the HGCal.

and copper plates interlaced with silicon sensors that are less than $1\ \text{cm}^2$. The depth of this section is about $25X_0$ or $1\lambda_i$. The hadronic section (CE-H) is an alternation of brass and copper plates and silicon sensors over a depth of around

$3.5\lambda_7$. Given the intrinsically fast response time of the silicon sensor, the HGAL will provide high-precision timing capabilities. The outermost part of the HGAL consists of the same design of brass plates and plastic scintillator currently used in the HCAL endcap. The total depth of the system is about $10\lambda_7$.

Muon chambers

The four stations of CSC in the high pseudorapidity region $1.6 < |\eta| < 2.4$ will be enhanced by adding redundancy. Two gas electron multiplier (GEM) chambers will be added to the two innermost stations. GEM chambers will be filled with a mixture of argon (70%) and CO_2 (30%) and will improve the momentum resolution for the muon trigger of the L1 trigger. The two outermost stations will include additional RPCs with intrinsically good timing capabilities, as is done currently for the DT chambers and CSCs at lower pseudorapidity.

Trigger system

The L1 trigger will be improved by using the tracking and high-granularity information [62]. This will be possible by upgrading the readout electronics of some subdetectors. To account for the additional time needed for the track reconstruction and matching with calorimeter information, the latency of the L1 trigger will increase from $4\ \mu\text{s}$ to $12.5\ \mu\text{s}$. From its expected performance with track information, its rate must increase from 100 kHz to 750 kHz to keep comparable performance with its current version despite the increase in PU. The current selection of events between the L1 and HLT output leads to 1 out of 100 events being saved. Assuming the same proportion during the high-luminosity phase, a 7.5 kHz rate is required for the HLT [63].

MIP timing detector (MTD)

A new detector dedicated to timing measurements will be added to the CMS experiment. The MTD [59] is designed to detect MIP deposition from charged particles with a high signal-to-noise ratio. It will consist of a thin layer located between the tracker and the ECAL. The barrel timing layer (BTL) covers the region $|\eta| < 1.5$ and will be equipped with LYSO scintillating crystals, read out by SiPMs. Crystals have a surface of $5.7\ \text{cm} \times 3\ \text{mm}$ and a thickness between 2.4 mm and 3.7 mm. The LYSO crystals are chosen because of their high light yield (40 000 photons per MeV), fast scintillation ($< 100\ \text{ps}$) and relatively short decay time ($\approx 40\ \text{ns}$). The endcap timing layer (ETL) extends over the region $1.6 < |\eta| < 3$ with two disks per endcap. Given the high radiation level, the ETL relies on planar silicon devices with internal gain: low gain avalanche detectors (LGAD). A total of about 4×10^6 LGAD sensors with size $1.3\ \text{mm} \times 1.3\ \text{mm}$ are used per endcap.

II.3 DDMTD simulation for precise clock monitoring in CMS subdetectors

II.3.1 Clock distribution

To provide precise timing information about the arrival of a particle, the detector response must be compared to a reference time. Within the detector system, the initial timing information of a time to digital converter (TDC) is the reference t_0 . The role of a TDC is to return a digital representation of the time of occurrence of a given signal, such as the detector pulse triggered by the arrival of a particle. This time of detection t_1 is compared to t_0 , and the difference between the two gives the precise time of arrival of the particle.

As mentioned in Section II.2.2, timing information in the subdetectors should have a 30 ps to 40 ps resolution to reduce the impact of the expected pile-up (200 events per bunch crossing) by improving the matching of tracks, vertices and energy deposits [59, 64]. It means that all components of the subdetectors need to have a precisely synchronised time of reference t_0 . This is done by distributing a common clock throughout the full subdetector. Distributing an accurate clock in the context of CMS subdetectors is a challenging task as it means sending the clock signal to thousands of readout units (ROU) in charge of reading the signal produced by the sensitive part of the detector. Moreover, each of these ROU is placed in different parts of the detector, separated by cables of hundreds of meters, and within significantly different radiation and temperature conditions. All these elements can alter the clock and introduce jitter in the signal.

A perfect digitised clock signal consists of a square wave oscillating between two amplitude levels corresponding to 0 and 1, with the rising and falling edges defining the unit of time as they follow the constant frequency of the clock. In reality, the square wave is imperfect and consists of fast-rising and falling edges (but with a small slope), and noise in the signal causes a shift in the time associated with an edge. These shifts of the clock edges are called jitter. This jitter can be decomposed into two components, the random jitter (RJ), where the shifts are randomly distributed. Since the random jitter originates from many independent physical effects, it creates shifts following a normal distribution around 0. The second component is the deterministic jitter (DJ), which produces periodic bounded shifts in the clock signal. Given the periodic nature of this noise, a frequency is associated with the DJ (f_{DJ}), and multiple DJ components can be defined if they impact the clock signal at different frequencies. This is an important element of jitter estimation and clock monitoring since our measurement or monitoring of a clock is done at a given sampling frequency f_{samp} . Then from the Nyquist-Shannon theorem, it means that our measurement is only sensitive to DJ components with frequency $f_{DJ} < f_{\text{samp}}/2$.

To study the jitter of a clock, we rely on the distribution of the shifts $\delta(t)$ between the clock we want to characterise and a clean clock of reference. Then these shifts can be studied in the time or frequency domain. By looking at the distribution of differences between the measured edges of a clock and the refer-

ence edge time, we measure a time interval error (TIE) histogram. The standard deviation or the root mean square (RMS) of the TIE histogram gives a figure of merit of the clock quality and of the jitter introduced by the distribution chain. In the frequency domain, an interesting figure of merit is the phase noise which shows at which frequencies the DJ is impacting the clock. The phase noise $\mathcal{L}(f)$ is defined as half the power spectral density (PSD) $S_{\delta}(f)$ of the shifts. The phase noise measurement is also useful to quantify the RMS jitter within a given frequency range $[f_1, f_2]$ as

$$\text{RMS}([f_1, f_2]) = \frac{\sqrt{2\Delta\phi^2}}{2\pi f_{\text{samp}}}, \text{ with } \Delta\phi^2 = \int_{f_1}^{f_2} \mathcal{L}(f) df. \quad (\text{II.2})$$

Examples of TIE histograms and phase noise plots are shown in Figs II.20 and II.21.

In order to achieve the required 30 ps to 40 ps resolution on the time information of particles, the clock distribution strategy should have less than 15 ps RMS jitter in all paths to the ROU [59, 64]. Three frequency ranges are defined with respect to their potential jitter contribution:

- *High frequencies* (100 kHz to 1 MHz): the clock signal can be embedded in a data stream at some stages of the distribution chain. The clock information is recovered thanks to phase-locked loops (PLLs). PLLs are expected to cancel most of the jitter in the clock signal in this range.
- *Intermediate frequencies* (10 Hz to 100 kHz): the jitter in this range should be monitored and cleaned if possible.
- *Low frequencies* (below 10 Hz): jitter in this range is qualified as wander as it causes the clock to slowly drift apart from its original phase. The TCLink system [65] was developed within the high precision timing distribution (HPTD) group of CERN to control this wander.

In this section, we describe a simulation framework of the digital dual mixer time difference (DDMTD) system to evaluate its correction potential of the jitter in the intermediate frequency range (10 Hz to 100 kHz). This study was performed with an application on the MTD and HGAL subdetectors in mind which have a similar clock distribution strategy [59, 64]. A simplified diagram of this clock distribution strategy is shown in Fig.II.17 and relies on a solution where the clock signal is embedded in the data acquisition (DAQ) path.

The LHC clock is derived directly from its radio frequency (RF) cavities, operating at a frequency of 400.788 MHz. After a division of its frequency by 10, the LHC clock is synchronised with the bunch crossings at a frequency of 40.079 MHz. This clock starts with a jitter specification of 9 ps RMS jitter [66] and should achieve better performance with improvements foreseen for the HL-LHC. The baseline suggestion for the clock distribution is to include the clock in the DAQ path.

First, the LHC clock is received by the new version of the trigger and clock distribution system (TCDS2) of the CMS detector. Then the TCDS2 transmits the clock or multiples of it (160 MHz, 320 MHz, ...) with the trigger and fast controls

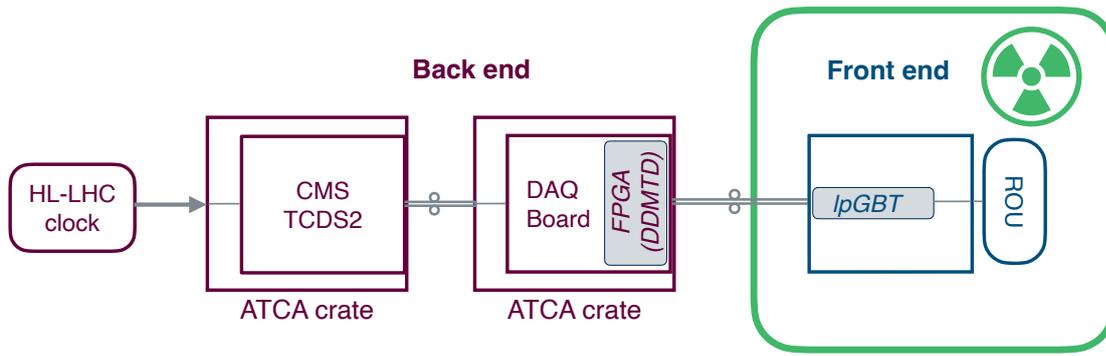


Figure II.17 – Diagram of the baseline clock distribution chain considered for the MT and HGAL subdetectors. A description of the distribution chain is provided in the text.

to the DAQ boards. Both the TCDS2 and DAQ boards are located within advanced telecommunications computing architecture (ATCA) crates and can rely on the high-speed lane of the ATCA backplane. There the field-programmable gate arrays (FPGAs) encode the clock to send it downlink to the front end at 2.5 Gb/s. At the front end level, low-power gigabit transceivers (lpGBTs) [67] are responsible for recovering the clock signal and propagating it to the ROUs. The lpGBT receives high-speed serial data streams sent without an explicit accompanying clock. But from the encoding of the clock information by means of a short recognisable periodic sequence in the stream, the PLL at the lpGBT level is able to lock on the phase of this sequence and recover a clock signal. Then this same clock is used as a reference to generate the uplink frame at 10 Gb/s. Even with this asymmetry in data throughput, the backend can recover the LHC clock. The monitoring (and cleaning) of the clock can be done thanks to systems implemented on the FPGAs using this clock loopback. For instance, a DDMTD system, as described in Section II.3.3, implemented on the FPGAs can compare the uplink clock to the downlink clock.

II.3.2 Characterisation study of the lpGBT version 1

To characterise the new version of the lpGBT (version 1), we perform an evaluation of its clock distribution performance. We used a distribution chain mimicking the back end and front end configuration of the detector chain. The test bench is summarised in Fig. II.18.

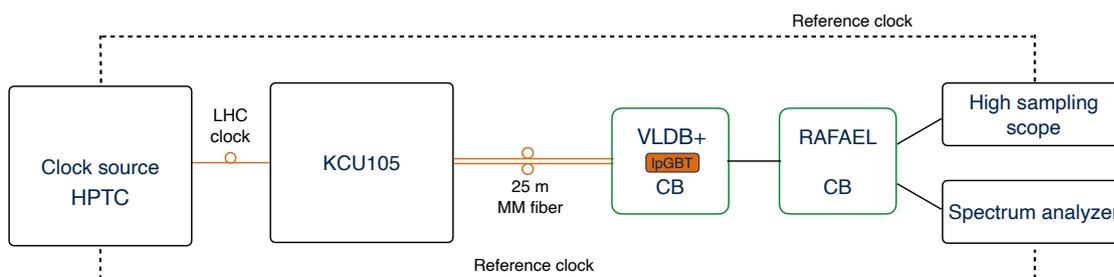


Figure II.18 – Diagram of the test bench for the characterisation of lpGBTv1.

First, a high-precision timing clock (HPTC) is generated and sent to the KCU105 equipped with an FPGA, which emulates the back end system. Its role is to encode the clock in the data stream and distribute it to the versatile link demo board (VLDB), which contains the lpGBTv1. The lpGBTv1 recovers the clock and propagate it through a RAFAEL ASIC (application-specific integrated circuit), which is dedicated to fan out the clock signal to all the sensor part of the detector. In this case, the clock does not reach the detector but is studied by a high sampling scope and a spectrum analyser. A picture of the test bench is shown in Fig. II.19.

The TIE histogram of Fig II.20 is measured from a clock with a frequency of 320 MHz. The output clock of the RAFAEL ASIC is compared to the input reference clock and shows a small RMS jitter contribution of 2.9 ps. Moreover, for the same configuration, the phase noise is measured, and results are shown in Fig II.21. From this measurement, we can compute the RMS jitter in the range 1 Hz–1 MHz and we find a low value of 0.97 ps.

These measurements show that the lpGBTv1 provides a clock distribution well within requirements for the phase 2 of the CMS detector. It can also be compared to the previous RMS jitter found for the lpGBTv0. Characterisation gave a value of 8.2 ps RMS jitter introduced by the lpGBTv0, so the improvements brought by version 1 are very promising.

II.3.3 The DDMTD system

As its name suggests, the digital DMTD (DDMTD) is a digital implementation of the dual mixer time difference system [68]. The DMTD was developed to transform the measurement of very fine phase shifts between two signals in a low-frequency domain with amplified shifts. It proves particularly useful when measuring differences in clock signals running at the same frequency. A digital implementation of the DMTD is proposed in Ref. [69]. This version of the DMTD was developed to run with digitised signals and can be adapted to many systems such as FPGAs. A simplified description of the DDMTD system is shown in Fig. II.22.

The system takes two clock signals u_1 and u_2 as input with the same frequency f and with a shift varying in time $\Delta t(t)$. An additional clock u_{dmtd} is derived from the input clocks and is slightly shifted in frequency with respect to the original signals thanks to a crucial parameter which is an integer commonly referred to as the N parameter of the DDMTD. This shift in frequency is defined as

$$f_{dmtd} = \frac{N}{N+1}f \quad (II.3)$$

where f_{dmtd} is the frequency of the u_{dmtd} clock. Both input signals pass through a gate called a D flip-flop [70] which uses as a reference clock the u_{dmtd} clock. In a D flip-flop, the signal (D in Fig. II.22) is effectively sampled by the reference clock (clk). In practice, it means that at each rising edge of the reference clock, the input signal is probed: if it is 1, the output signal is set to 1 until the next rising edge, and vice versa if the input is 0. The sampled clocks returned by the

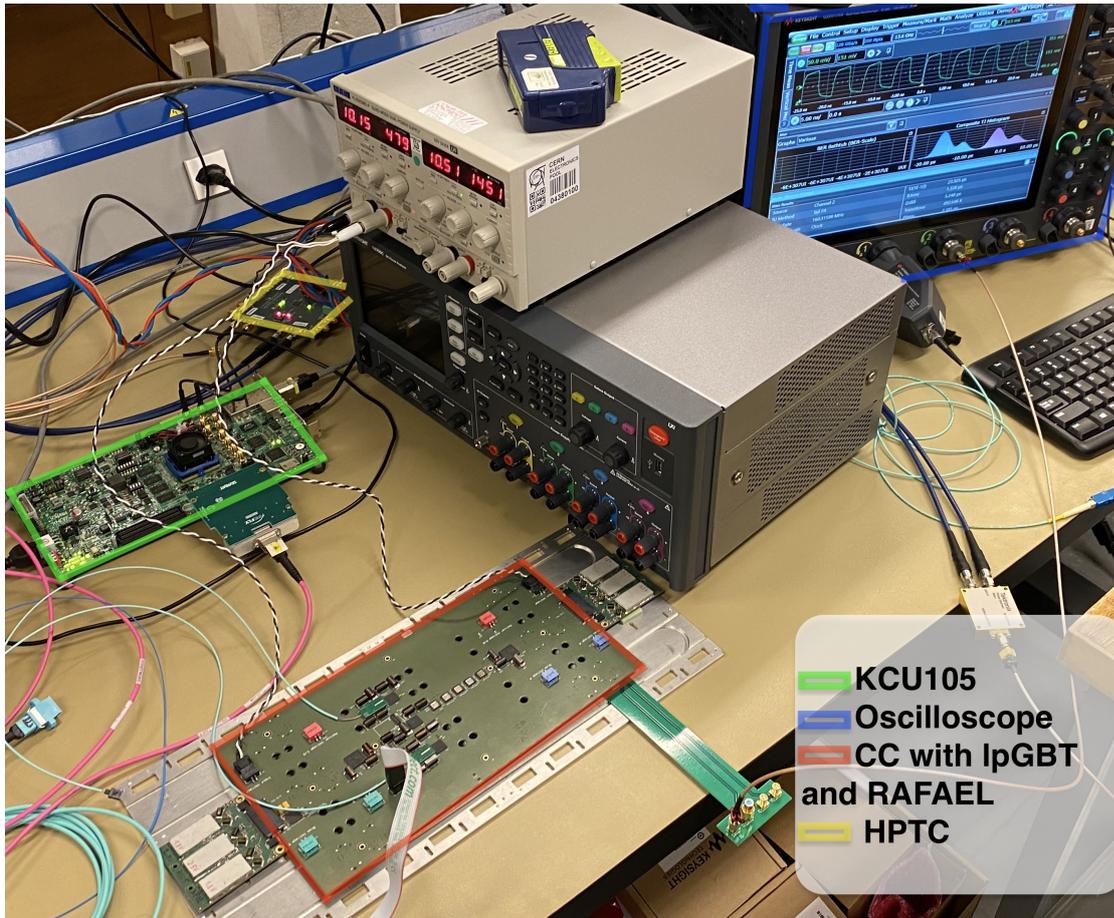


Figure II.19 – Picture of the test bench being validated on a IpGBT version 0. The board close to the table edge is the concentrator card (CC) with the IpGBT. It is replaced by the VLDB and IpGBTv1 in the actual characterisation study. Above are the KCU105 board (green) and clock generator (yellow).

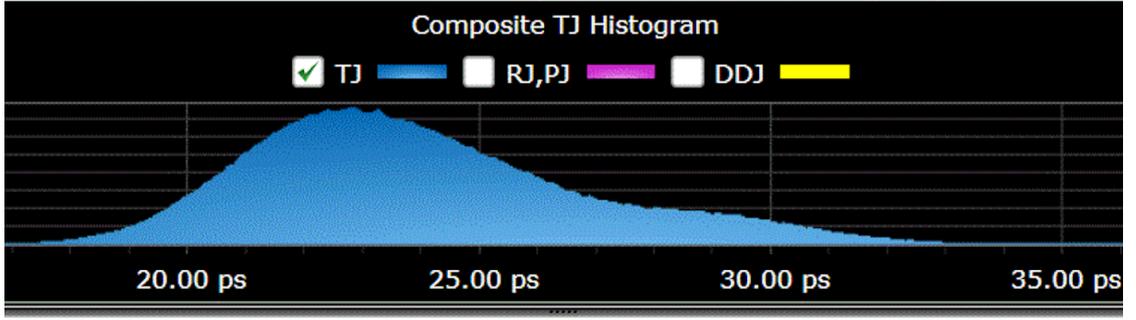


Figure II.20 – TIE histogram of the 320 MHz output clock of the lpGBTv1+RAFAEL system.

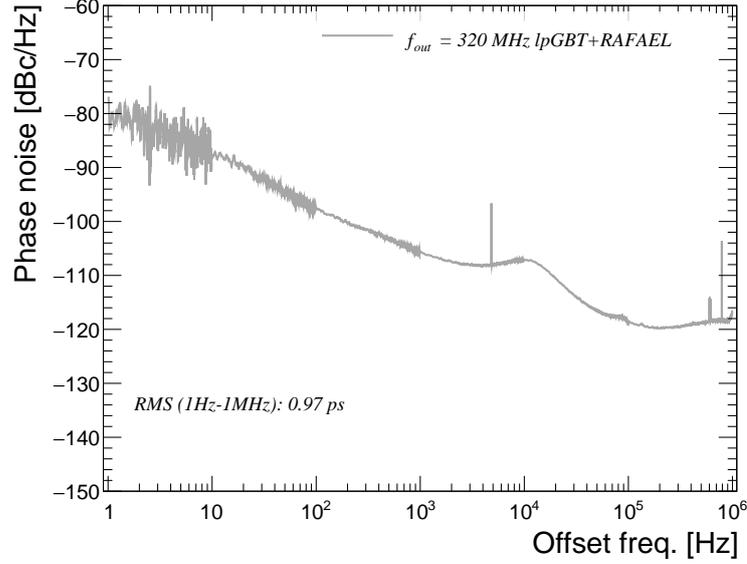


Figure II.21 – Phase noise of the 320 MHz output clock of the lpGBTv1+RAFAEL system.

D flip-flops, u_{beat1} and u_{beat2} , have a frequency

$$f_{\text{beat}} = f - f_{\text{dmt}} = \frac{f}{N + 1}. \quad (\text{II.4})$$

This process of creating the beat clocks is illustrated in Fig. II.23. If the input clocks u_1 and u_2 have a noisy signal, passage through the D flip-flops can create glitches where the beat clocks have several rising and falling edges at each transition before stabilising. Hence the necessity of a deglitching procedure for the beat clocks. Finally, the time difference Δt_{beat} between the rising or falling edges of u_{beat1} and u_{beat2} gives a proxy to access the actual time difference Δt between u_1 and u_2 . The conversion between the beat frequency domain and the original high frequency domain is given by

$$\Delta t = \Delta t_{\text{beat}} \frac{f_{\text{beat}}}{f} = \frac{\Delta t_{\text{beat}}}{N + 1}. \quad (\text{II.5})$$

The DDMTD system provides a way to measure narrow differences between clock signals but at the cost of not comparing as many rising edges since the comparison is done with the rising edges of the beat clocks, which run at a much

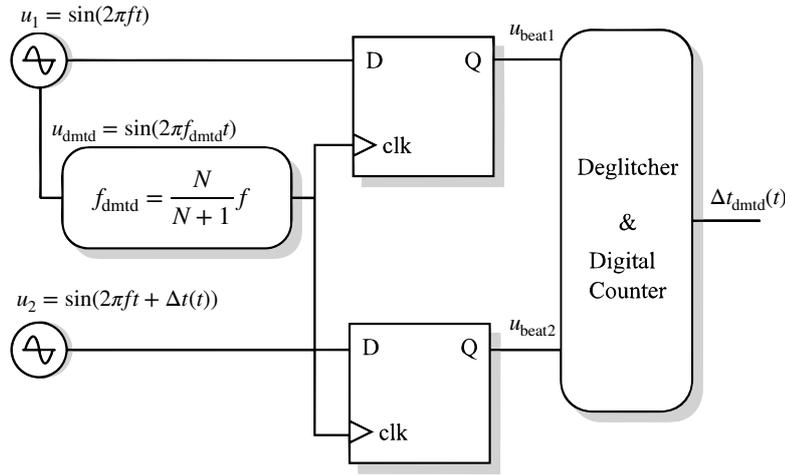


Figure II.22 – Schematic description of the DDMTD system adapted from Ref. [69].

slower frequency. It also imposes a restriction on the minimum Δt value that can be measured. If there is no visible shift in the beat domain, we cannot access the Δt information. To make sure there is a measurable difference between u_{beat1} and u_{beat2} , the Δt_{beat} value needs to be higher than the period of the $u_{\text{dmt d}}$ clock $\Delta t_{\text{beat}} > 1/f_{\text{dmt d}}$. From this requirement and using Eqs. II.3 and II.5, we can determine the lowest Δt accessible to the DDMTD:

$$\begin{aligned} \Delta t_{\text{beat}} > \frac{1}{f_{\text{dmt d}}} &\Leftrightarrow \Delta t_{\text{beat}} > \frac{N+1}{N \times f} \\ &\Leftrightarrow \Delta t > \frac{1}{N \times f} \end{aligned}$$

II.3.4 Simulating a DDMTD: *pyDDMTD*

The DDMTD system is a good candidate to monitor the clock jitter and potentially correct it to ensure conformity with the clock distribution requirements of CMS subdetectors during HL-LHC. Its hardware implementation has already been extensively tested in the CERN community, especially via the HPTD group [71]. However, in order to expand our understanding of the system and make future developments of monitoring or correction strategies easier, we developed a fast and accessible Python-based DDMTD simulation framework: *pyDDMTD*.

The simulation of the working principle of a DDMTD in a perfect clocks scenario is summarised in Figs. II.23 and II.24. First the input clocks u_1 and u_2 are defined with a frequency of 40 MHz and a constant offset $\Delta t = 4$ ns between the two. Then the $u_{\text{dmt d}}$ is defined with $N = 5$ in this example, and the dmt d clock is used to get the two beat clocks u_{beat1} and u_{beat2} . The sampling process is illustrated with the two red arrows of Fig. II.23. Then from the shifts of u_{beat1} and u_{beat2} the Δt_{beat} signal is built (see Fig. II.24) and can be converted to Δt using Eq. II.5.

Jitter can be added to the input clocks of the DDMTD simulation. To stay close to the typical operation expected in the HL-LHC, input clocks have a frequency $f = 40$ MHz, a shift of $\Delta t = 4$ ps is introduced between the two clocks and a parameter $N = 10000$ is taken for the DDMTD to observe fine shift due to the

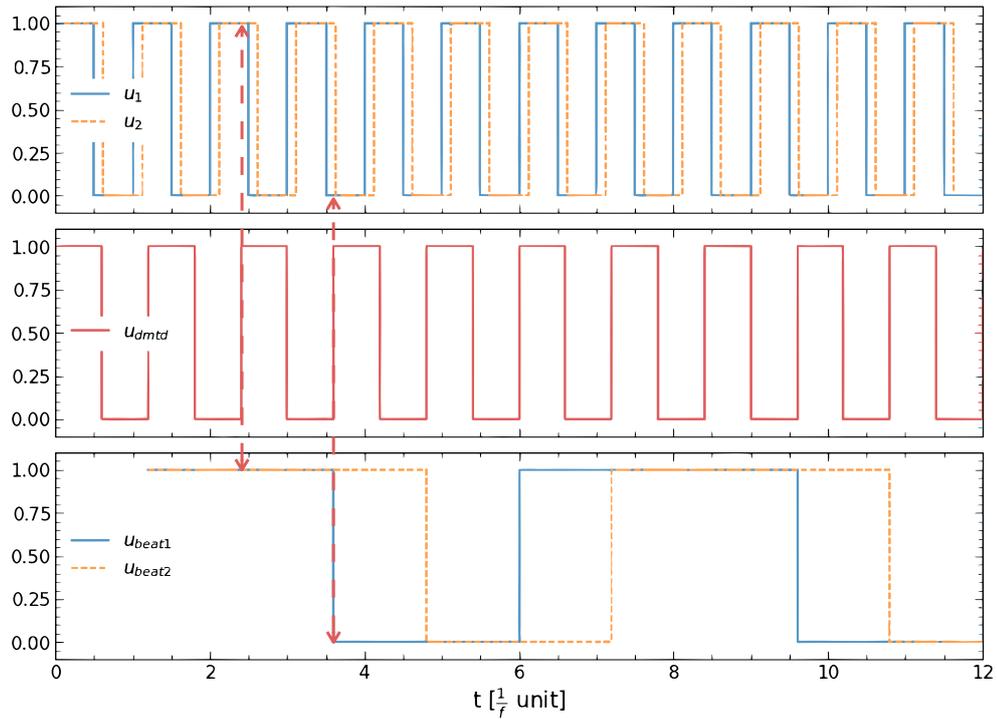


Figure II.23 – Clock signals are shown in units of the period of u_1 and u_2 , $1/f$, to avoid unnecessary large time values when drawing a small number of beat clocks cycles. *Top*: simulation of two clock signals u_1 and u_2 running at the same frequency and with a small offset. *Middle*: simulation of the DMTD clock u_{dmtd} with a frequency slightly lower than u_1 and u_2 . *Bottom*: beat clocks u_{beat1} and u_{beat2} sampled from u_1 and u_2 with u_{dmtd} . The two arrows illustrate the sampling process to create the two beat clocks.

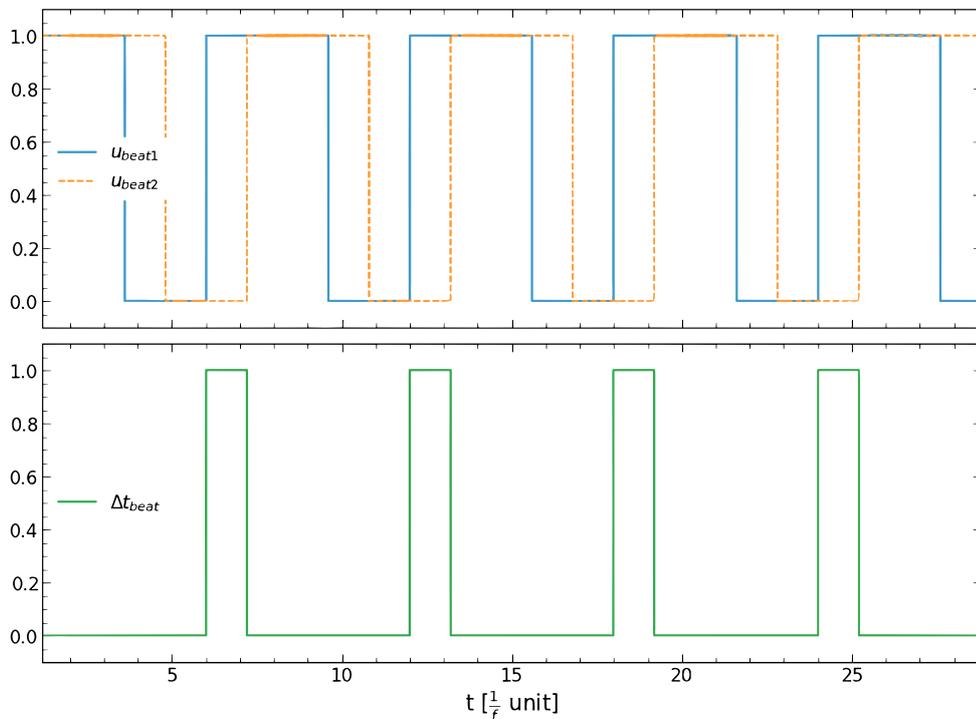


Figure II.24 – Extraction of the Δt_{beat} signal from the two beat clocks.

jitter of the input clocks. With such settings, the beat clocks run at frequency $f_{\text{beat}} = 4 \text{ kHz}$ and the DDMTD can probe shifts down to $\Delta t_{\text{min}} = 2.5 \text{ ps}$. Random jitter is added to u_1 and u_2 by shifting their rising and falling edges with values drawn from a Gaussian distribution with mean 0 and standard deviation 0.5 ns. As expected and as reported in Fig. II.25, glitches appear in the beat clocks and in the Δt_{beat} measurements in the form of packets of rising and falling edges instead of single edges at the truth value.

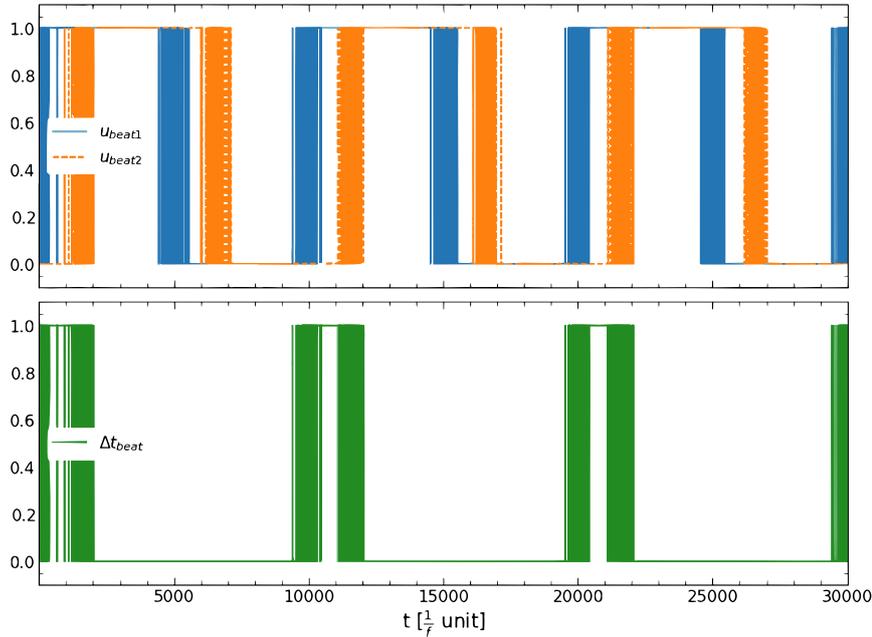


Figure II.25 – Impact of jitter on the u_{dmttd} clock. Glitches appear in the two beat clocks (top) and distort the Δt_{beat} signal measurement (bottom).

These glitches can be studied by looking at the distribution of rising edges when they are all folded in a clock cycle. The comparison of distributions with and without glitches is shown in Fig. II.26. In the clean clocks case, the edges are always falling at the same position in the clock cycle resulting in a peak with a number of edges equal to the number of simulated clock cycles (100 in the case of Fig. II.26). In the same way, the converted Δt measurements are at their truth value of 4 ns with no dispersion. However, in the case of noisy clock signals, rising edges follow a Gaussian distribution centred around the truth value but with a dispersion corresponding to the intensity of RJ included. Note also that the number of rising edges is significantly higher than the number of simulated clock cycles (100), this is a clear symptom of glitches in the beat clocks. The effects are then clearly visible in the Δt measurements. With the rapid alternation of rising and falling edges in the beat clocks, a lot of Δt_{beat} signals are observed (as shown in Fig. II.25) with duration close to 0 ps. Hence the peak at 0 ps in the final Δt measurements of Fig. II.26.

In following studies of the DDMTD performance [72], three deglitching procedures were proposed: the *first edge*, *mean edge*, and *zero count* algorithms. These methods resolve the glitches issue by respectively selecting the first, mean, or median edge of the glitch packet as the correct edge. We studied the effects

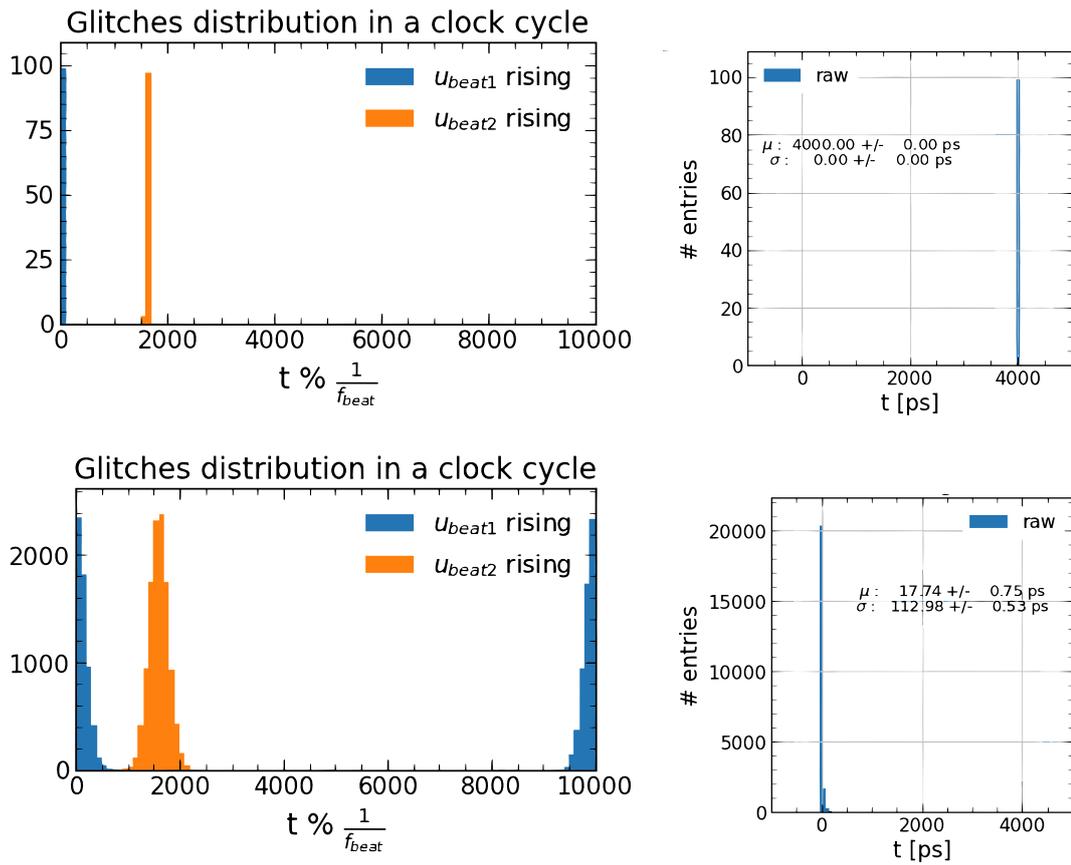


Figure II.26 – Left plots show the distribution of all rising edges of the two beat clocks folded into a single clock cycle and right plots show the distribution of Δt measurements. Two configurations are shown: without (above) and with (below) glitches due to jitter in the input clocks.

of the three methods on the distribution of edges and Δt measurements simulated with *pyDDMTD*, and the comparison is detailed in Fig. II.27. Since all of these procedures select a single edge, they all fix the issue of having mainly $\Delta t_{\text{beat}} \approx 0$ durations and recover the expected number of measured Δt_{beat} . However, their performance is not equivalent. The *first edge* algorithm is the simplest to implement in practice. It does not select edges close to their truth value but instead assumes that, for independent glitch packets, the minima of Gaussian-distributed edges keep the same distance as their mean. Thus, rising edges are lower than their truth value, but the mean of the Δt measurements is close to the correct value. In the case of the *mean edge* and *median edge* algorithms, selected edges are expected to be close to their truth value, and the Δt_{beat} value is recovered by construction. Finally, as presented in Fig. II.27, best results are obtained with the *median edge* method as it shows the lowest dispersion of DDMTD outputs with a standard deviation of $\approx 1\%$ around the mean. This result is consistent with the hardware implementation studied in Ref. [72].

To confirm these observations were not dependent on the amount of RJ introduced in the signal, the same study is performed with varying RJ in the u_1 and u_2 input clocks. Results are reported in Fig. II.28 and show the *median edge* algorithm should be favoured for any RJ hypothesis. Very high RJ values are tested, but less than 5 ps is expected on the input clocks in the context of HL-LHC.

II.3.5 Realistic jitter simulation

As mentioned already in Section II.3.1, the total jitter observed on a clock signal can be decomposed into two main components: the random jitter and the deterministic jitter. To offer more simulation possibilities, especially when considering realistic scenarios, both types of jitter can be simulated with *pyDDMTD*. The RJ is emulated by shifting the clock edges with values following a normal distribution $\mathcal{N}(0, \sigma_{\text{RJ}})$ where σ_{RJ} is the RJ intensity. The DJ corresponds to a periodical shift of the edges and is added through the sine wave $\sigma_{\text{DJ}}/2 \sin(2\pi f_{\text{DJ}} t)$ where σ_{DJ} is the DJ intensity impacting frequency f_{DJ} . As the DJ can impact multiple frequencies with different intensities, the condensed notation $\sigma_{\text{DJ}}@f_{\text{DJ}}$ is used in this section to clearly identify the components of DJ. Examples of jitter simulation are presented in Fig. II.29.

These examples are described through TIE histograms and, although this kind of distribution measure Δt shifts between two clocks, note that they represent a concept distinct from DDMTD measurements. The goal of a TIE histogram is to characterise a single clock by comparing it to a reference clock that is the perfectly clean equivalent of the measured clock. In practice, these TIE histograms are measured with high-frequency oscilloscopes and are sensitive to very fine shifts. On the other hand, the DDMTD output compares and characterises shifts between two (potentially noisy) input clocks. The sensitivity of the DDMTD depends on its N parameter. As such, in this section, TIE histograms serve as truth information to evaluate a clock and compare it to the DDMTD measurements.

With these additional handles to simulate jitter in clocks, we can reproduce the jitter expected in the actual clock distribution chain developed for subdetectors

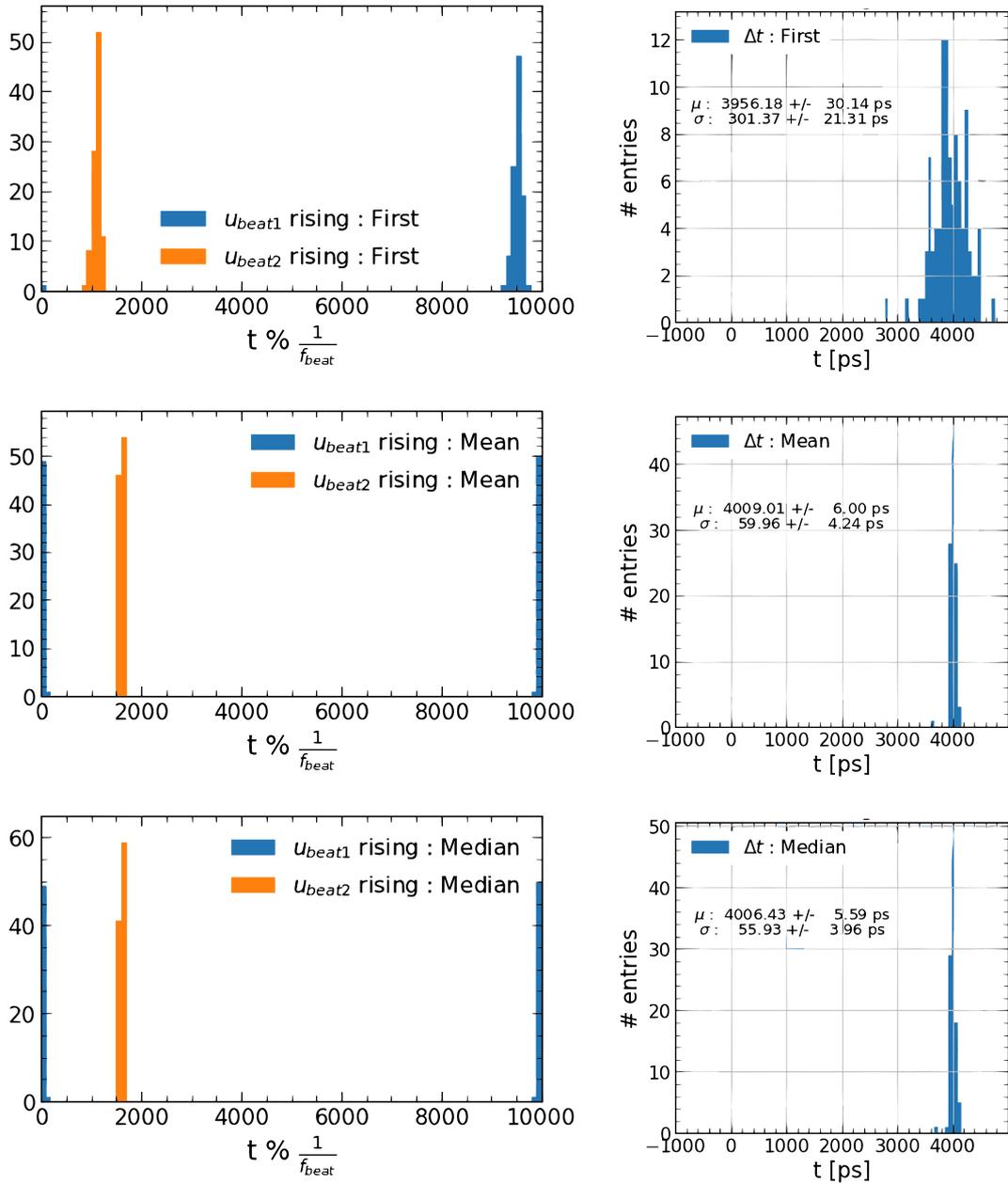


Figure II.27 – Left plots show the distribution of rising edges of the two beat clocks in a clock cycle, and right plots show the distribution of Δt measurements. Plots are showing the effects of the *first edge* (top), *mean edge* (middle), and *median edge* (bottom) deglitching methods.

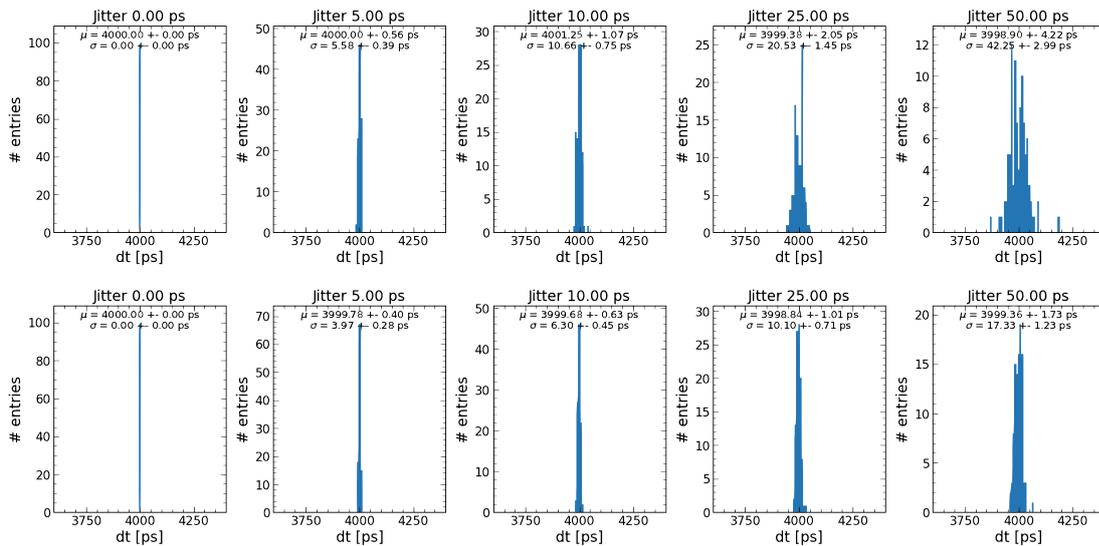


Figure II.28 – Distribution of Δt measurements for an increasing random jitter on the input clocks (left to right). The *first* deglitching method is applied in the top row, and the *median* method is applied in the bottom row.

of CMS phase II. To evaluate this expected jitter, we performed measurements on a test bench imitating the distribution chain and characterised the output clock of the lpGBT that corresponds to the clock guiding the front end systems. A clean clock with frequency 160 MHz is delivered to the distribution chain, and a TIE histogram of the output clock is performed with a high-sampling oscilloscope. The measured TIE histogram is reported in Fig. II.30. From this jitter profile, a satisfactory simulation is done with *pyDDMTD* using the following jitter characteristics:

- RJ: 2.8 ps
- DJ: 20 ps@0.1 Hz / 1 ps@100 kHz / 15.7 ps@40 MHz

This jitter profile is measured at the front end level, so it cannot be accessed during the normal operation of the CMS detector. Instead, what is observed by the DDMTD is the clock sent back to the back end (uplink). Then, it seems sensible to formulate a symmetry hypothesis by assuming the same contribution from downlink and uplink and to model the jitter at the front end as half of what is measured by the DDMTD. The relation between the jitter at the detector level (front end) and the jitter measured at the back end has been studied during a measurement campaign by teams from the HPTD group, Minnesota University and CEA Saclay. To study this relation, both the fibres and the front end have been put in a climate chamber alternately to assess the effects of temperature on jitter. Their results are presented in Fig. II.31. When the fibres are in the climate chamber, the downlink and uplink paths seem to add the same phase variation. This is visible on the left plot of Fig. II.31 where the symmetry hypothesis ($0.5 \cdot \text{DDMTD}$) matches what is observed in the output clocks of the lpGBT (ECLK0, ECLK1, PSCLK0). However, a significant difference is observed when only the front end is put in the climate chamber: the phase shift in the front end is approximately 7 times higher than the symmetry hypothesis. These results show

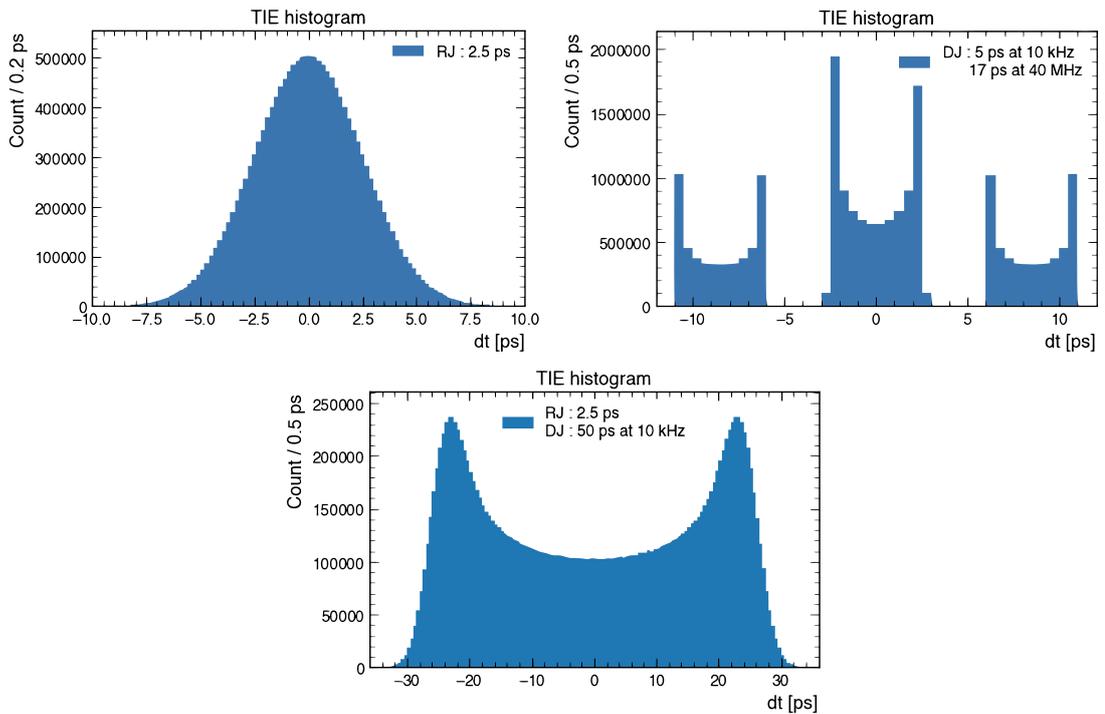


Figure II.29 – TIE histograms of clocks simulated with RJ (2.5 ps) in the top left, DJ (5 ps@10 kHz and 17 ps@40 MHz) in the top right and a combination of RJ (2.5 ps) and DJ (50 ps@10 kHz) at the bottom.

that in the feedback path to the DDMTD, the clock is not passing through some of the internal components of the lpGBT and on-detector modules.

A model of this asymmetry is implemented in *pyDDMTD* based on the result of this measurement campaign and on the following assumptions:

- external factors introduce slow jitter components (below 100 Hz) in the front end,
- we expect higher power and temperature fluctuations within one order of magnitude of the L1 trigger rate (750 kHz).

As such, DJ within these frequency ranges is reduced by a factor 7 at the DDMTD level. The symmetric part is also included by considering the same jitter contribution from downlink and uplink. The impact of this modelling on the back end measurement of a clock with the jitter profile described in Fig. II.30 is shown with the TIE histograms of Fig. II.32.

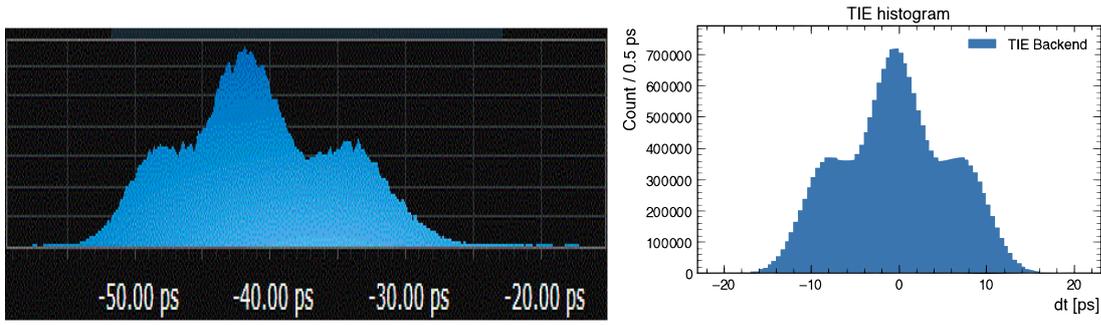


Figure II.30 – TIE histogram measured on the output clock of a lpGBT (v0) with frequency 160 MHz (right) and TIE histogram of a clock simulated with *pyDDMTD* at the same frequency with 2.8 ps RJ and 20 ps@0.1 Hz, 1 ps@100 kHz, and 15.7 ps@40 MHz DJ.

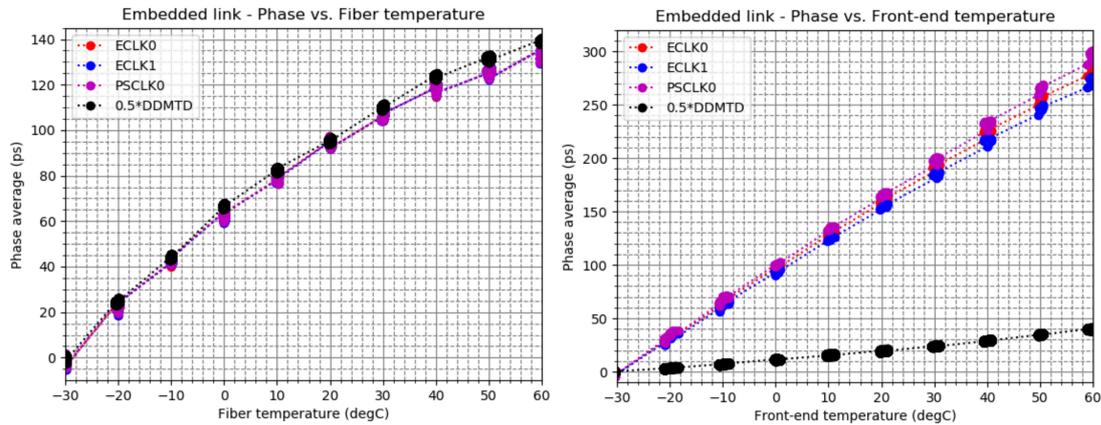


Figure II.31 – Measurements comparing the impact of temperature on the clock signal when the fibres (left) or the front end only (right) are put in a climate chamber. The plots show the compatibility between the phase shift in multiple clock links at the front end level (ECLK0, ECLK1, PSCLK0) and the assumption that the phase shift is half of the DDMTD measurement at the back end level (0.5*DDMTD) [71].

II.3.6 Evaluating corrections on the clock jitter

The goal of implementing DDMTDs on FPGAs of the back end is to monitor and correct the jitter introduced by the distribution of the clock to the detector and to ensure synchronisation among all parts of the subdetectors. These DDMTDs take as input the downlink clock sent to the front end and the uplink clock coming back. Thus, their output tracks down the jitter introduced along the clock path and brings correction possibilities.

In order to extract an accurate correction value, we need to evaluate the distribution of the DDMTD output *i.e.* we need to register the DDMTD information over several clock cycles. The number of clock cycles needed to compute a correction defines a monitoring window of the DDMTD output with a duration Δt_{window} . This duration sets the frequency of the correction on the clock with $f_{corr} = 1/\Delta t_{window}$, and it marks a limit on the maximal frequency of DJ we can correct as the correction does not impact DJ with $f_{DJ} > f_{corr}/2$. Examples of correction of noisy clocks using DDMTD measurements are shown in Fig. II.33. The DDMTD is simulated with a parameter $N = 8192$, and its measurements are

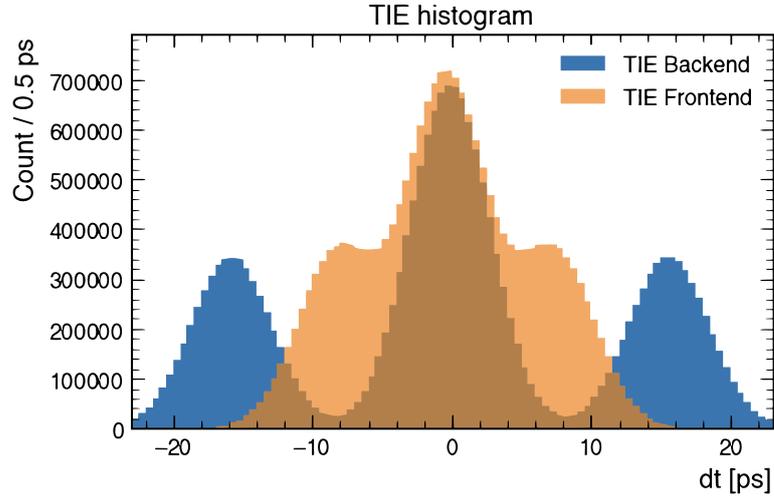


Figure II.32 – Simulation of the effect of additional jitter in the front end components (orange) on the measurement at back end level (blue).

collected during monitoring windows of 5 ms. The correction value is computed as the mean of the DDMTD output distribution acquired during that window. It corresponds to a correction frequency $f_{\text{corr}} = 200$ Hz. Two jitter configurations are tested in Fig. II.33, associated with the top row and bottom row examples. Both clocks have a frequency $f = 160$ MHz, a RJ of 2.5 ps, a high-frequency DJ component of 17 ps@40 MHz but they have different low-frequency DJ. The clock on the top row example has a low-frequency DJ of 50 ps@100 Hz. Thus, the distributions of the DDMTD output are shifted in each monitoring window since they collect measurements faster than the effects of the DJ. This can be seen in the leftmost plot of Fig. II.33, where each colour represents a different monitoring window. Then, each monitoring window is corrected by the mean of the DDMTD measurements, resulting in the central plots where all distributions are aligned. The effects on the full clock signal are visible from the TIE histograms on the right, where the standard deviation of the TIE distribution decreased from 6.8 ps to 6.4 ps. However, the bottom row example uses a clock with a low-frequency DJ components of 50 ps@1 kHz. Since the correction frequency is 200 Hz, it is not sensitive to the jitter components of this clock. Indeed, no effect is seen on the distribution of the DDMTD measurements, and the TIE histogram is unchanged before or after correction.

So while we need to accumulate several clock cycles to increase the precision on the correction value, we also need a fast correction process if we want to reduce the effects of DJ at high frequencies. The number of DDMTD measurements in a monitoring window is given by

$$n_{\text{ddmtd}} = \frac{f}{(N + 1)f_{\text{corr}}} \quad (\text{II.6})$$

As it is inversely proportional to f_{corr} , there is a direct trade-off between the number of measurements per window to compute an accurate correction and the maximum jitter frequency that can be corrected. A solution could be to decrease the value of N while increasing f_{corr} . But this cannot be done arbitrarily as N also imposes the precision of the DDMTD measurement through $\Delta t_{\text{min}} = 1/(Nf)$,

so smaller N also decreases the accuracy on the correction value. To evaluate the impact of N on the correction performance, we scanned the standard deviation of TIE histograms of clocks with varying low-frequency DJ (f_{DJ}) and using DDMTDs with varying N . The results of these scans are detailed in Fig. II.34. Given our asymmetry assumption, the clock measured at the back end shows a smaller jitter than expected at the front end. Effects of a low N are noticeable as the correction, in this case, mostly degrades the clock. However, we can also see that the correction is able to lower the jitter level in the front end up to 15% depending on the N setting. It means that a DDMTD only correction can already bring improvements in the clock distribution and allows some cleaning of the jitter in the frequency range above 10 Hz.

II.4 Conclusion

The first part of this chapter is dedicated to the description of the LHC at CERN, of the CMS experiment, and of the planned upgrades to adapt to a high-luminosity phase for the LHC.

A particular focus is given to the clock distribution chain within CMS subdetectors and the requirements for the detectors to uphold the quality of their physics measurements. To ensure a high level of synchronicity between the different components of the detectors, careful monitoring of the jitter introduced in their reference clock is needed. I present in this chapter a simulation framework to evaluate the monitoring and correction possibilities brought by a DDMTD system implemented at the back end level for jitter impacting frequency above 10 Hz. This framework gives a fast and accessible simulation of a DDMTD with tunable parameters and is applied on a customisable jitter profile. As a preliminary study, we study the correction possibilities of such DDMTD on an asymmetric jitter introduced at the front end level. Results show that with an appropriate selection of the parameters of the DDMTD, the system is able to reduce the jitter components with frequencies below half of the correction frequency.

This framework can now be used to test different correction strategies. The correction presented in this chapter is not the only possible one, and more advanced strategies can be considered, for instance by introducing multiple DDMTD (each with a different N parameter or with an offset between each other) and computing a correction based on a combination of their outputs. We can even think of machine-learning-based corrections where a model is trained to learn the asymmetry in jitter between the front end and the back end. In that case, *py-DDMTD* can be used to produce training samples for the machine-learning models.

CHAPTER III

Deep Learning

III.1 Machine learning and neural networks	62
III.1.1 Development of machine learning	62
III.1.2 Building an artificial neuron	63
III.1.3 Neural networks	64
III.2 Teaching a machine	66
III.2.1 Learning paradigms	66
III.2.2 Loss and metrics	67
III.2.3 Gradient descent and backpropagation	68
III.3 Deep neural networks	71
III.3.1 Overtraining	71
III.3.2 Vanishing and exploding gradients	73
III.3.3 Learning rate	74
III.3.4 An example of DNN: the convolutional neural network	76
III.4 Generative adversarial networks	78
III.5 Summary	81

Applications for machine learning are numerous as it is a rapidly growing field, and physicists in high energy physics (HEP) are prolific contributors. Given the high performance necessary from detectors and the precision required in the statistical analyses of data, machine learning is indeed a relevant tool that benefits from the large amount of data available. In that regard, one of the main aspects of my PhD topic is to apply advanced machine learning techniques to different aspects of physics analyses of the CMS experiment.

This chapter covers the developments in deep learning that are necessary for describing my work in Chapter IV and V. It is also a good way to transcribe my experience as the contact person between the machine learning group and the Higgs physics analysis group of CMS. It starts with a summary of machine learning and how to build artificial neural networks. Section III.2.1 describes the learning process of such networks. Finally, the last two sections detail deep neural networks and generative adversarial networks, respectively, as well as the common regularisation techniques needed for their training.

III.1 Machine learning and neural networks

III.1.1 Development of machine learning

Machine learning is a field studying how a computing machine can learn the optimal set of parameters needed to perform a given task. For instance, high-performing data-fitting methods such as the least squares methods or maximum likelihood estimation are building blocks of such parameter optimisation. But a shift in paradigm was brought in 1943 with the first mathematical model of the neuron by McCulloch and Pitts [73]. The 1950s saw the first learning neural network machines [74, 75] and, in particular, the first implementation of the McCulloch and Pitts neuron with the perceptron of Rosenblatt [76]. This interest continued throughout the 1960s with the parallel development of non-neural-network-based techniques such as support vector machines (SVM) [77] or the nearest neighbour algorithm [78]. This enthusiasm surrounding machine learning then came to a halt for more than a decade with the realisation that a single-layer neural network could only achieve optimal classification for linearly separable classes and that they were computationally limited (with the book *Perceptron* [79]). Only in the 1980s came a resurgence of the field with the development of neural networks with more complex architectures (precursors of the convolutional neural networks (CNNs) [80], first recurrent neural networks (RNNs) [81], ...) and the application of backpropagation to machine learning [82–84] for a great gain in computational performance (although backpropagation was already described since 1970 [85]). Outside of neural networks, this period saw the creation of additional techniques such as random forest algorithms [86] or boosted decision trees (BDT) [87]. Finally, in the last 10 to 20 years, the release of exhaustive databases (MNIST [88], ImageNet [89], ...), the implementation of high-level software libraries (Torch [90], Tensorflow [91], ...) and the upgrade of computing hardware (especially graphics processing units (GPU) and now artificial intelligence (AI) dedicated hardware [92]) enabled to study deeper and

deeper neural network models and to enter the deep learning era.

III.1.2 Building an artificial neuron

As for many technological progress, machine learning was shaped by observations of nature. In this case, the concepts of neural networks are directly inspired by the learning process in the animal kingdom. The support of this process is the neurons and their interconnections, forming intricate circuits throughout the whole body, especially in the brain. A healthy human brain is expected to contain around 90 billion neurons, each connected to thousands of others [93]. This figure can serve as a reference to compare with the size of neural networks presented in this thesis; for instance, the networks trained in Chapter IV have approximately 30 000 neurons each and the network described in Section V.4 of Chapter V contains approximately 80 000 neurons. All of these connections need to cover a very broad spectrum of tasks going from perception to complex reasoning but also taking care of memory, movement and other physiological processes (sleep, respiration, etc.).

Acting as a building block of the learning process, the neuron is the cell whose role is to convey information. A simple representation of a biological neuron is shown in Fig. III.1 with its different constituents:

- **Dendrites:** where the information is received from precedent neurons in the chain.
- **Cell body:** where one finds the usual components of a cell (nucleus, mitochondria, ...).
- **Axon:** its role is to propagate the information thanks to an electrical signal.
- **Presynaptic terminal:** which translates the information from an electrical signal to a biochemical one by releasing neurotransmitters outside of the neuron. These neurotransmitters cross the gap between two neurons (the synapse) before reaching the dendrites of the next neuron.

The most popular model behind the working principle of a neuron can be easily summarised by following the role of each of its constituents [95]. When dendrites receive neurotransmitters from other neurons, their contribution to the potential of the neuron's membrane is summed both in time and in multiplicity. Some of the dendrites have an inhibitory function, so they contribute negatively to this sum. If the potential reaches a certain threshold, the neuron activates, and an electric pulse is sent along its axon. This maximum of the pulse has a fixed value and is not dependent on the sum of the stimuli received by the dendrites: it is an all-or-nothing mechanism. Only the frequency of the pulses is affected by the input stimuli. Then the learning process emerges from the interconnection of these neurons. Each of these connections must be tuned (number of connected neurons, number of neurotransmitters released, frequency of the pulses, ...) to give the optimal neuronal chain for a given task. This tuning is

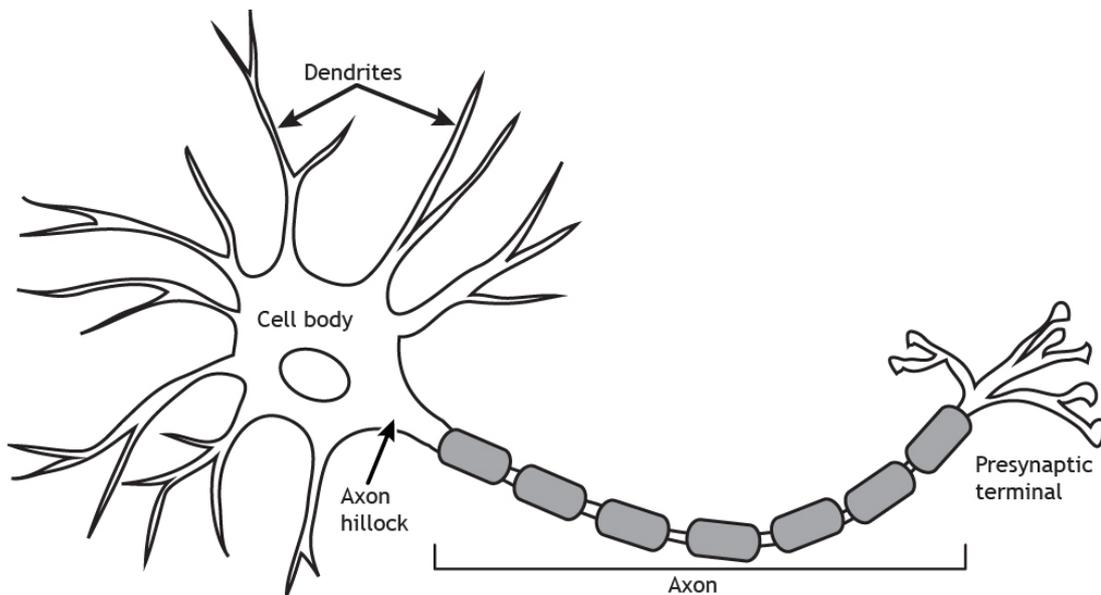


Figure III.1 – Simplified representation of a neuron from Ref. [94]

updated constantly through life by learning new tasks or training by repeating them, for instance, and is one of the main mechanisms behind neuroplasticity.

As mentioned above, translations of this biological neuron to an artificial one started in 1943 with McCulloch and Pitts [73] and were developed by Rosenblatt [76] to the model presented here. They realised that we can model a neuron as a basic mathematical function. It receives a number of input stimuli (a_i), which are weighted (w_i) by importance (some being even negative) and integrated over time ($\sum a_i w_i$). Then it outputs a certain number of pulses depending on this weighted sum of inputs if some threshold has been reached $f(\sum a_i w_i)$, where f is called the activation function. This activation function is of primary importance as it allows for a non-linear response of the neurons with respect to the inputs, increasing its capabilities in learning and reproducing complex functions. A bias (b) can also be added to shift the combination of inputs towards an optimal range in the activation function domain $f(\sum a_i w_i + b)$. From this, we can model a simple artificial neuron as described in Fig. III.2.

III.1.3 Neural networks

Once we have built this elementary block, we can assemble more complex structures by connecting artificial neurons together. The first emerging structure is a layer of neurons. A layer consists of parallel neurons that are not interconnected but that share the same input neurons and the same output neurons. Then, by concatenating several layers, we arrange a network of neurons or, in other terms, an artificial neural network. In its most basic form, a neural network consists of at least an input layer — where the neurons take as inputs the values evaluated by the network — and of an output layer — where the neurons output the value predicted by the network. Any layers between the input and output layers are called hidden layers.

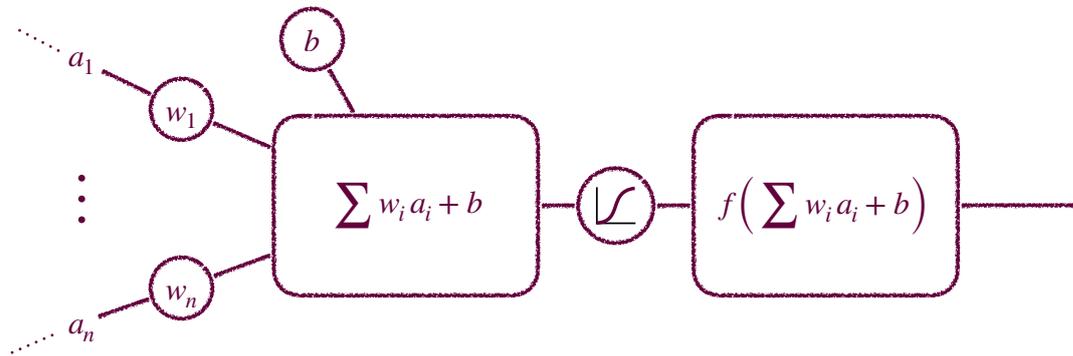


Figure III.2 – Model for an artificial neuron. a_i are the inputs of the neuron, w_i represents the importance given to each input, b is the bias and f is the activation function.

A simple architecture of such a neural network is presented in Fig. III.3 and is called a feedforward neural network. All outputs of one layer proceed forward to the next one; there is no loop. However, layers can take more elaborate structures. Some layers can take as input the output of anterior layers and not just the previous one (skip connection layers), and some can take as additional input an earlier state of the network in time (recurrent networks), or share the same weights between multiple neurons, etc. Actually, a layer represents a more general idea of a specific arrangement of neurons dedicated to a precise task, and the final output of the network is the transformation of the input values after going through each layer. It means that the output of a network $\hat{\mathbf{y}}$ can be decomposed as a combination of functions f_k applied to the input vector \mathbf{x} . So, for a network with L layers:

$$\hat{\mathbf{y}} = f_L \circ \dots \circ f_k \circ \dots \circ f_1 \left(\sum_j w_{ij}^1 x_j + b_i^1 \right) \quad \text{(III.1)}$$

and in the case of a feedforward network, all neurons of a layer are connected to all neurons of the previous layers so the pass through each layer can be expressed as a simple matrix multiplication:

$$\hat{\mathbf{y}} = f_L \left(\mathbf{W}^L f_{L-1} \left(\dots f_1 \left(\mathbf{W}^1 \mathbf{x} + \mathbf{b}^1 \right) \dots \right) \right) \quad \text{(III.2)}$$

where \mathbf{W}^l is the matrix representation of layer l meaning that each component of the matrix w_{ij}^l is the weight applied to input j of neuron i in layer l .

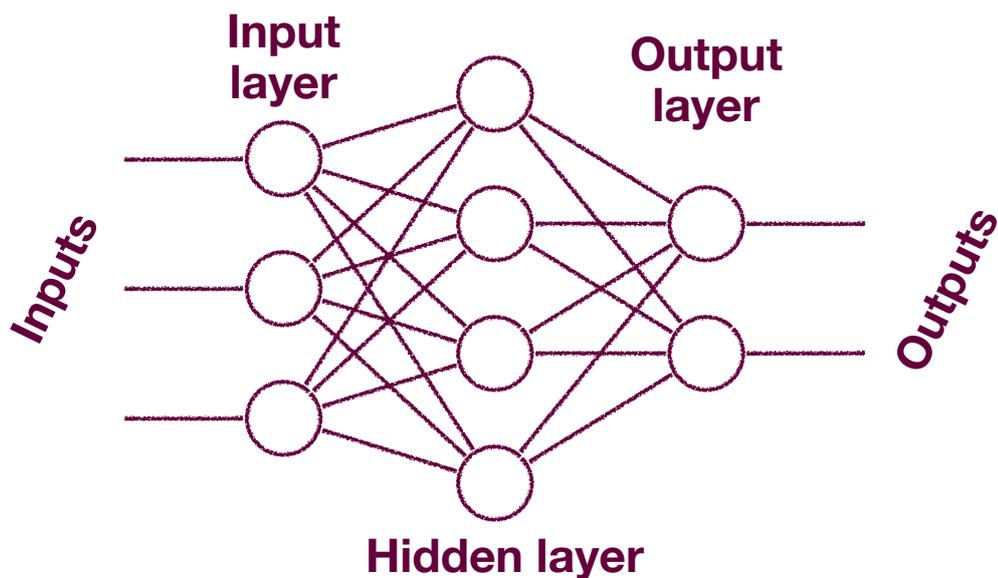


Figure III.3 – Model of a feedforward neural network.

III.2 Teaching a machine

III.2.1 Learning paradigms

Given a set of inputs, the output of a neural network is completely defined by the set of weights, biases and activation functions of each layer. While the activation functions are defined when building the network architecture, the weights and biases are the tunable parameters that must be optimised. This step of learning the optimal set of parameters from successive predictions over available data is called the training of the machine learning technique. These tunable weights are to be distinguished from other parameters of the training that can be adapted to improve the training process but which are fixed at training time such as the width of the network layers, the update strategy for the weights, the learning rate (see Section III.2.3), *etc.* The latter are called hyperparameters. The learning procedure requires first to define a precise model of the objective that the algorithm should reach through the learning process. This objective is highly dependent on the task assigned to the algorithm and on the nature of the training data. Machine learning methods are usually categorised between three main objectives representing different learning processes:

- **Supervised learning:** each training data point links a set of inputs to a label, and the machine learning algorithm should learn how to predict the correct label from a given input. An optimal algorithm can then infer labels even for inputs not seen in the training data. For instance, algorithms built for classification or regression are trained with supervised learning. As such, it is the preferred learning mode for physics analyses, whether it is for event tagging and categorisation or regression of event observables.

- **Unsupervised learning:** it corresponds to a case where the training data has no label or it is irrelevant. The objective of the algorithm while training is to reproduce the input data. It forces the algorithm to find the relevant features that constitute the minimal information from which one can extrapolate the rest of the data. In HEP, it is particularly useful for clustering or dimension-reduction tasks. It is also used in anomaly-detection-based searches or data quality monitoring.
- **Reinforcement learning:** in this context, the machine learning algorithm is considered as an agent belonging to a given environment, and its objective is to learn what is the correct next action to perform. This action impacts the environment, which brings a reward to the agent and sets it in a new state. It can be used when there is a clear set of possible actions to choose from, and it is typically how the algorithms controlling autonomous vehicles or the ones playing games are trained. It can be seen as a model of the reward system of the brain. Its applications are rare in the context of collider experiments, but it could be applied to the control of particle accelerator operations [96].

III.2.2 Loss and metrics

In the context of machine learning, the learning objective needs to be translated into a mathematical expression to give the algorithm an estimation of how it performs and how to improve. This expression is a function of the algorithm output and is commonly called a loss function. The definition of this loss function depends on the goal of the algorithm, but there are common losses for each context, especially when training neural networks. For instance, in classification tasks, the objective is to check the agreement between the network output and the class of the input data. Outputs of a classifier network are often bounded between 0 and 1 and can be interpreted as probabilities to belong to a given class. The loss is expressed as the cross entropy H between these two probability distributions:

$$H(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^n y_i \log \hat{y}_i \quad \text{(III.3)}$$

where n is the number of classes, $\hat{\mathbf{y}}$ is a vector encoding the network output, so each \hat{y}_i represents the probability of the input data to belonging to class i and \mathbf{y} is the vector encoding the true class, *i.e.* $y_i = \delta_{ij}$ for class j . Note that Eq. III.3 simplifies as $H(y, \hat{y}) = -(y \log \hat{y} + (1 - y) \log (1 - \hat{y}))$ for the binary case. As the network learns to predict the label value, the $\hat{y} \approx 0$ cases are not pathological since they correspond to an associated $y = 0$. For unsupervised learning or supervised regression, the goal of the network is to reproduce the input data or the continuous label of the input data. Then, an evaluation of the error of the network is preferred, *e.g.* with the mean squared error:

$$\text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{(III.4)}$$

where n is the number of output neurons and $\hat{\mathbf{y}}$ and \mathbf{y} the predicted and expected output respectively.

Even if the loss function usually gives crucial insight into the state of the training, it is useful, most of the time, to define additional metrics of the network performance. These metrics are independent of the learning process as they are not entering the update of the weights (so they are not affected by constraints on the continuity of their derivatives), but they act as indicators that the training is going in the correct direction and that the network is learning the proper task. For example, when training a network for a classification task, a powerful tool to compute its classification performance is the receiver operating characteristic curve (ROC curve), from which one can extract its area under the curve as a metric. In the context of applying machine and deep learning to Physics, this is typically where our intuition as physicists comes into play. Physicists are used to develop meaningful figures of merit to summarise information from complex contexts, and this can be applied directly to network training by using additional metrics to check for expected physical consequences or to rule out unphysical results.

III.2.3 Gradient descent and backpropagation

As described in Eq. III.1, the output of a network is entirely determined by its input and weights $\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{x})$. So any loss for a network is, in fact, a function of the inputs, the weights of the network \mathbf{w} and the expected output: $L_{\mathbf{w}}(\mathbf{y}, \hat{\mathbf{y}}(\mathbf{x}))$. This lays the foundations of the neural network training (a model of the learning process), as the goal of this phase is to learn the optimal distribution of the weights — the one that minimises the loss. In fact, it becomes a high-dimensional optimisation problem or, in other words, a fit over many parameters. Given the high number of trainable parameters (usually of the order of several millions) and sometimes of input variables, it is often impossible to know analytically the loss function over its full definition domain. Instead, the resolution of this problem is brought by numerical methods where we replace the full knowledge of the loss function with an estimation using the highest possible number of evaluations, as the final goal is to get a network with high predictive performance on unseen samples. This is also why neural network trainings require a high number of training data points.

One of the best-performing algorithms to solve optimisation problems is the gradient descent algorithm. The gradient descent algorithm starts by computing the gradient of the loss function for a given input point. Since we want to update the weights, the inputs and expected outputs of the network are considered fixed, and the gradient is computed with respect to each of the trainable weights of the network. By definition, the gradient indicates the direction where the function changes the most. It means that, by using the gradient information, we can update the weights towards the direction of the steepest descent, and the network gets closer to the minimal value of the loss. After the training step t ,

each weight can be updated as:

$$w_{t+1} = w_t - \gamma \frac{1}{n_{\mathcal{X}}} \sum_{\mathbf{x} \in \mathcal{X}} \nabla_{w_t} L_w(\mathbf{y}, \mathbf{x}) \quad (\text{III.5})$$

where \mathcal{X} is the full training sample, $n_{\mathcal{X}}$ is the size of the training sample and γ is called the learning rate and is used to decrease or increase the importance of the update following the direction opposite to the gradient. The gradient with respect to a given weight w_{ij}^l at training step t can be derived as follow using the chain rule:

$$\begin{aligned} \nabla_{w_{ij}^l} L_w(\mathbf{y}, \hat{\mathbf{y}}) &= \frac{\partial L}{\partial w_{ij}^l} = \frac{\partial L}{\partial \hat{\mathbf{y}}^l} \cdot \frac{\partial \hat{\mathbf{y}}^l}{\partial w_{ij}^l} \\ &= \frac{\partial L}{\partial \hat{\mathbf{y}}^l} \cdot \frac{df_L(\mathbf{s}^L)}{d\mathbf{s}^L} \cdot \frac{\partial \mathbf{s}^L}{\partial w_{ij}^l} \\ &= \frac{\partial L}{\partial \hat{\mathbf{y}}^l} \cdot f_L'(\mathbf{s}^L) \cdot \mathbf{W}^L \frac{\partial \hat{\mathbf{y}}^{L-1}}{\partial w_{ij}^l} \\ &= \frac{\partial L}{\partial \hat{\mathbf{y}}^l} \cdot f_L'(\mathbf{s}^L) \cdot \mathbf{W}^L \dots f_l'(\mathbf{s}^l) \cdot \frac{\partial \mathbf{W}^l}{\partial w_{ij}^l} \cdot \hat{\mathbf{y}}^{l-1} \end{aligned} \quad (\text{III.6})$$

where $\hat{\mathbf{y}}^l$ is the vector of outputs of layer l (so $\hat{\mathbf{y}} = \hat{\mathbf{y}}^L$) and $\mathbf{s}^l = \mathbf{W}^l \hat{\mathbf{y}}^{l-1} + \mathbf{b}^l$ is the weighted sum of the inputs of neurons at layer l . Although the method of gradient descent was already described by Cauchy in 1847 [97], its application to machine learning and neural networks was limited because of the high number of computation it requires. It was only with the rediscovery of backpropagation (presented in III.7) and its application to the training of neural networks in 1982 [82] that a new interest in neural networks emerged since it improved greatly the computational efficiency of gradient descent. Indeed, by computing the output and its derivatives for each neuron when evaluating the loss function and then storing them, one can extract the gradient of the loss with respect to a specific weight by simple matrix multiplication:

$$\begin{aligned} \nabla_{w_{ij}^L} L_w(\mathbf{y}, \hat{\mathbf{y}}) &= \frac{\partial L}{\partial \hat{\mathbf{y}}^L} \cdot f_L'(\mathbf{s}^L) \cdot \frac{\partial \mathbf{W}^L}{\partial w_{ij}^L} \cdot \hat{\mathbf{y}}^{L-1} \\ &= \delta^L \cdot \frac{\partial \mathbf{W}^L}{\partial w_{ij}^L} \cdot \hat{\mathbf{y}}^{L-1}, \\ \nabla_{w_{ij}^{L-1}} L_w(\mathbf{y}, \hat{\mathbf{y}}) &= \delta^L \cdot \mathbf{W}^L \cdot f_{L-1}'(\mathbf{s}^{L-1}) \cdot \frac{\partial \mathbf{W}^{L-1}}{\partial w_{ij}^{L-1}} \cdot \hat{\mathbf{y}}^{L-2} \\ &= \delta^{L-1} \cdot \frac{\partial \mathbf{W}^{L-1}}{\partial w_{ij}^{L-1}} \cdot \hat{\mathbf{y}}^{L-2}, \\ &\vdots \\ \nabla_{w_{ij}^l} L_w(\mathbf{y}, \hat{\mathbf{y}}) &= \delta^l \cdot \frac{\partial \mathbf{W}^l}{\partial w_{ij}^l} \cdot \hat{\mathbf{y}}^{l-1}. \end{aligned} \quad (\text{III.7})$$

Note that the computation starts with the final layers and goes decreasingly in their position, hence the backward propagation. This method corresponds to

the reverse version of automatic differentiation. The benefit with respect to a naive computation of the gradient for each weight is twofold: first, the repetition of calculations is avoided by computing each derivative only once, and second, by going backwards, it is only a vector and matrix multiplication instead of matrix and matrix multiplications if it was going forward. Hence the usefulness of GPUs which are particularly powerful when dealing with vectors and matrices operations. This method of backpropagation also imposes some conditions on the loss and activation functions of each layer. If some gradient-based optimisation is performed, such as gradient descent, the loss and activation functions should be differentiable in addition to being non-linear. For instance, the Heaviside step function — that would be the simplest representation of an all-or-nothing behaviour — is not differentiable in 0 and has a null derivative everywhere else, so it would not train with a standard gradient descent method.

With the development of a new method for updating weights to increase convergence capabilities and decrease the training time, the basic update strategy of Eq. III.5 has been superseded. A short description of some optimiser algorithms useful for this thesis will be given here. Firstly, most modern optimisers rely on a batch version of the gradient descent algorithm. Instead of computing the update as the average gradient over the full training sample X , the training sample is divided into smaller batches $\mathcal{B}(\mathcal{X})$ and the update of the weights is computed as the average gradient over these batches:

$$w_{t+1} = w_t - \gamma \frac{1}{n_{\mathcal{B}}} \sum_{\mathbf{x} \in \mathcal{B}} \nabla_{w_t} L_{\mathbf{w}}(\mathbf{y}, \mathbf{x}) \quad \text{(III.8)}$$

where $n_{\mathcal{B}}$ is the size of batch \mathcal{B} . The unit of training composed after going through all batches of the training sample once is called an epoch. This is a good compromise between the standard gradient descent computed over the full training sample, which is expensive regarding memory and computational power, and a stochastic gradient descent that uses only one batch as an epoch. In addition to the training by batch, some optimisers add a momentum term in the update, which includes information about the gradient at previous training steps. It is the case for the Adam optimiser [98], which is considered a standard optimiser given its good performance on a wide spectrum of training settings. The LAMB optimiser is also worth mentioning. It is a modified version of the Adam optimiser, which adds a layerwise normalisation factor to the weight update. This normalisation is particularly useful when training with a large batch size (*i.e.* smaller training time) as the direction of the gradient is preserved while mitigating the negative impact of too low or too high gradients.

III.3 Deep neural networks

As mentioned already in Section III.1.1, one of the reasons behind the loss of interest in neural networks was the realisation that a single-layer network is, in any case, limited to elementary tasks. One could always train multiple single-layer networks and perform an intricate non-linear combination of their output, but this is mostly impractical. A set of theorems called universal approximation theorems brought a solution to overcome this crippling shortcoming.

These theorems show that under certain conditions, a neural network can approximate any continuous function. The first proven universal approximation theorem states that a single-layer feedforward network is a universal approximator if it uses a non-linear activation function and that its hidden layer has an infinite width [99]. Then additional theorems confirmed that it is also the case for networks with an infinite number of fixed-width hidden layers [100] and that, fortunately, networks with a finite number of layers with finite width can approximate any function if they have a large enough number of neurons [101, 102]. These results also extend to other popular network architectures [103].

This set of theorems implies that by going towards deeper network architecture, *i.e.* adding more hidden layers, we can describe more intricate functions with neural networks and use them for more ambitious tasks. Moreover, the constant improvement of the computational capabilities of modern hardware made the idea of building and training models with a great number of neurons possible on reasonable timescales. These observations drove the rise of the deep learning field, which aims to train a class of networks with an extensive number of neurons, from simple deep feedforward neural networks to more convoluted architectures.

III.3.1 Overtraining

While the large number of neurons provided by deep architecture promises better performance, it also leads to additional challenges when training. Probably the most common one is the issue of overtraining. It is coming from the bias-variance trade-off arising when training a machine learning algorithm.

The bias of the algorithm describes its capacity to provide an accurate output close to the expected value, and the variance represents its capacity to return consistent results when the input is impacted by small variations due to noise. This problem is well known in optimisation procedures as fits or trainings reaching low bias and low variance are rarely possible and often limited by available data. Usually, an adjustment is needed between a network able to learn general rules from a training sample but reaching lower performance (underfitting) and a network giving the best performance on the training sample but thrown off when presented with unseen configuration (overtraining). This compromise depends on the number of times the network sees the training data. Networks underfit at the beginning of the training and overfit after a large number of epochs.

However, overtraining can be easily monitored by evaluating the loss value of

the network on a validation sample. This sample must be independent of the training sample and not seen by the network while training. It allows regular checks of the generalisation capabilities of the network. In this way, if the loss computed on the training sample keeps decreasing even though the loss from the validation sample starts to increase, we have a clear indication that the network starts to overtrain. This is also a good way to know when to stop the training. This technique is called early stopping and checks for an increase in the validation loss to stop the training of the network. An example of overtraining happening while training is shown in Fig. III.4. Additional techniques can be ap-

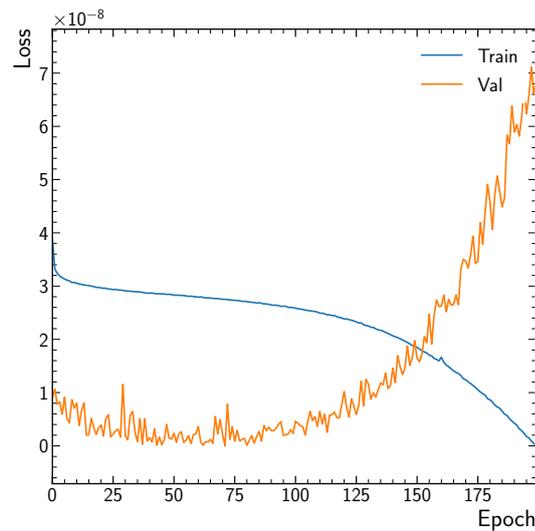


Figure III.4 – Example of overtraining seen when training a DNN. The loss of the training sample is dropping rapidly, but the loss from the validation sample is skyrocketing, indicating overtraining from the network.

plied to prevent such effects rather than just stopping the training when the network starts to overtrain. Most of them rely on limiting the complexity of the network by reducing the range of possible values for the weights and not letting them take arbitrarily high values. This can be achieved in several ways, with the most straightforward way consisting of clipping the weights within a small range [104]. A penalty term can also be added to the loss function of the network as the L_1 or L_2 norm of the vector of weights:

$$L_1(\mathbf{w}) = \sum |w_i| \tag{III.9}$$

$$L_2(\mathbf{w}) = \sum w_i^2 \tag{III.10}$$

as it pushes the weights toward smaller values while training.

Finally, a popular method to increase the robustness of the network against small variations of the input is to use dropout [105]. Dropout refers to the technique of randomly selecting a given fraction of the weights and setting their value to 0 after each training step — as if some of the neuron connections were dropped out (illustrated in Fig. III.5). It forces the network not to rely on a few very important weights (or series of weights) and to spread the computation over the full network.

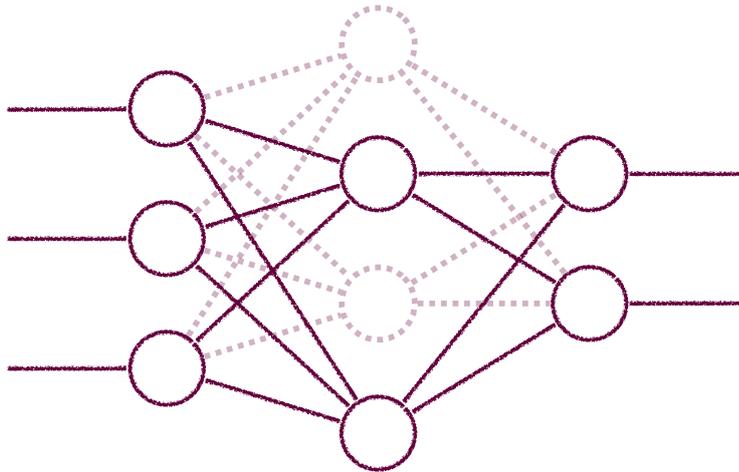


Figure III.5 – When using dropout, a random set of neuron connections are dropped at each training step. All connections are restored for evaluation.

III.3.2 Vanishing and exploding gradients

The choice of activation functions also plays a big role when training deep architectures, as their effect is applied after each layer. This is crucial when training a network with gradient descent and backpropagation. Indeed, as mentioned already in Section III.2.3, if the derivative of the activation functions is null, then the weights are not updated, and the network effectively stops learning. From my teaching experience, I would translate that as if your lecture lacks interesting twists, then students get bored and stop learning. And this can happen even without regions of null derivative in the activation functions. If the derivative has values in $[0, 1]$, then by using backpropagation and multiplying the derivatives of layer l by the gradients computed for layer $l + 1$, the gradients are actually exponentially decreasing, and the first layers of the network get very small updates. Conversely, some activation functions have derivatives taking very high values that will be enlarged during backpropagation leading to the inverse problem of exploding gradient. This could be translated as if you keep exhausting your students with a constant flow of difficult information, they will saturate and stop learning.

The most popular activation functions were shaped to avoid these problems. We already saw that the linear and Heaviside step functions were removed from the equation. Given their properties, the sigmoid (Eq. III.11) and hyperbolic tangent (Eq. III.12) functions are good candidates for activation functions. They are continuously derivable, bounded and, with their S shape, they are close to the description of an all-or-nothing mechanism. However, such S-shaped functions have a gradient tending to 0 when their input tends to $\pm\infty$, so they are particularly prone to bringing vanishing gradients. They are still good options for the activation of the output neurons. This led to the design of a set of activation

functions called rectified linear unit functions which are piecewise linear functions. In its basic form, the ReLU [106] (Eq. III.13) is linear for positive input and null otherwise. It allows the recovery from vanishing gradients in one direction at the cost of having a function non-differentiable in one point (but this is easily fixed by arbitrarily setting the gradient at 0 to 0 or 1). Several alternatives were developed from the original ReLU [107–109]; a popular one is the LeakyReLU (Eq. III.14) that also fixes the issue of vanishing gradients for negative inputs.

$$\sigma(x) = \frac{1}{1 + e^x} \quad \text{(III.11)}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{(III.12)}$$

$$\text{ReLU}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \quad \text{(III.13)}$$

$$\text{LeakyReLU}(\alpha, x) = \begin{cases} \alpha x & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \quad \text{(III.14)}$$

In addition to the choice of activation functions, some techniques were developed to avoid vanishing or exploding gradients. In the search for how to properly initialise the weights of the network before training, it was shown that this initialisation could prevent any vanishing or exploding gradients. For instance, in the case of a network with sigmoid-activated hidden layers, the problem of vanishing gradient can be fixed if the weights have a random initialisation following the popular Glorot uniform (Eq. III.15) or normal (Eq. III.16) [110] or by following the recent initialisation developed by Yilmaz and Poli [111] (Eq. III.17):

$$\mathbf{W}^l \sim \mathcal{U}\left(-\frac{\sqrt{6}}{\sqrt{n_l + n_{l+1}}}, \frac{\sqrt{6}}{\sqrt{n_l + n_{l+1}}}\right) \quad \text{(III.15)}$$

$$\mathbf{W}^l \sim \mathcal{N}\left(0, \frac{2}{n_l + n_{l+1}}\right) \quad \text{(III.16)}$$

$$\mathbf{W}^l \sim \mathcal{N}\left(\max\left(-1, -\frac{8}{n_l}\right), 0.01\right) \quad \text{(III.17)}$$

with \mathbf{W}^l the set of weights of layer l and n_l the number of weights of layer l . Similarly to the weight clipping used to mitigate overtraining, a gradient clipping can also be applied to avoid exploding gradient effects [112].

III.3.3 Learning rate

The landscape of the loss function can be quite complex as the dimensionality of its definition domain increases with the number of neurons. Since the gradient descent technique uses local information to find the direction of the minimum, it can start targeting a saddle point or a local minimum and get stuck around it, effectively stopping the training. To avoid such premature convergence while

training, the size of the step performed in the gradient descent (*i.e.* the size of the weights update) must be carefully selected. This is the role of the learning rate, denoted as γ in Eq. III.5.

If the learning rate is too big, then chances are that the gradient descent will miss the global minimum, and if the learning rate is too small, then the time of convergences of the networks increases, and the algorithm is more likely to find a local minimum as illustrated in Fig. III.6. That is why the learning rate is rarely constant and changes throughout the training by planning a learning rate schedule. A simple yet popular way of scheduling the evolution of the learning rate is to let it decrease exponentially with respect to the epochs or each training step. It allows the learning rate to start at a rather high value and then take smaller steps as the weights get closer to their optimal value. Alternatively, a strategy is to decrease the learning rate by a given factor each time the loss reaches a plateau. Some optimisers are also effectively adapting the learning rate as the training progresses. For example, in the Adam and LAMB optimisers, the moments of the gradient from the previous training steps effectively act as an adjustment to the learning rate. It is also worth mentioning the notion of

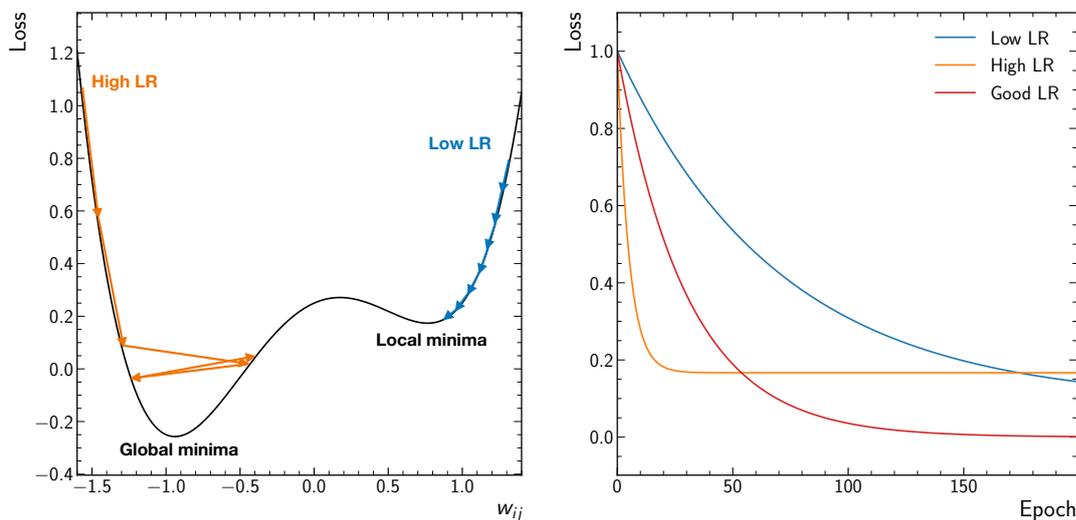


Figure III.6 – Illustration of the impact of the learning rate (LR) during the gradient descent when updating a given weight (left) and over the full training (right). If the learning rate is too high (orange), the training is suboptimal as the weights might not converge to the global minimum. If the learning rate is too low (blue), the training time increases unnecessarily, and the weights are more likely to get stuck in a local minimum.

warm restarts of the learning rate [113]. In trainings where the training is clearly getting stuck in a local minimum, regular and sudden increases in the learning rate can help to unblock the network and put it back on the way to the global minimum.

III.3.4 An example of DNN: the convolutional neural network

A short description of convolutional neural networks is given in this section as, first, they are a typical example of deep learning technique, and second, they are one of the building blocks of the method presented in Chapter IV. The convolutional neural network (CNN) is an attempt at extracting high-level information from organised arrays of data such as images, texts, spectrograms, and many more. Thus, it makes sense that their architecture was inspired by the operation of the visual cortex in biological brains.

When a sensory signal enters the visual cortex, it first goes through layers of *simple cells* that extract some level of information from these stimuli, for instance, particular lines, shapes and orientations at precise positions. The extracted information is then passed to layers of *complex cells* that build higher-level concepts, such as a movement in a specific direction and are usually unaffected by the absolute position of the stimuli in their perceptive field. This led to the creation of the precursor of CNNs in 1979 by Fukushima: the neocognitron [80], which was successfully trained to recognise Japanese characters and numbers. The prevalent architecture for CNNs and their training with end-to-end back-propagation of all the layers was developed in 1989 by LeCun et al. [84]. This standard architecture is composed of several layers alternating between convolutional layers and pooling layers:

- **Convolutional layers:** as their name suggests, they rely on the convolution operation, so the neurons are not fully connected to all neurons of the previous layer but rather to a smaller set of adjacent neurons of the previous layer. Since the goal of the layer is to learn a feature that could appear several times and anywhere on the output of the previous layer, all neurons covering the full input share the same set of weights. This is analogous to a small filter that would be convoluted to patches of an image and then shifted to repeat the operation over the full image (as described in Fig. III.7). A filter would learn to extract one type of feature, so convolutional layers usually consist of several filters learning a different type of feature each. The output of a filter over the full input is called a feature map, as it returns a mapping of the presence or absence of the feature in the input.

Note that, as for any layer, the convolutional layer can be represented by a matrix. It allows the definition of transposed convolutional layers that perform the inverse operation (sometimes called deconvolution) in the sense they can learn to extrapolate a higher dimensional representation from a feature map. This is useful to define networks symmetrical to CNNs that would take as input a vector of features and return the associated deconvoluted image, e.g. when using generative adversarial networks (see Section III.4).

- **Pooling layers:** the goal of pooling layers is to extract the prevailing aspects of each feature map while reducing the position dependency. This is done by looking at patches of the input feature maps and summarising

this patch with a single figure. For instance, it can be the average of the values over this patch or, most commonly, the maximum value of the patch. When applying a convolution layer, the feature maps usually have a size close to the input, so a network without pooling layers would learn to look for specific input regions to find a feature. However, the position of the feature relative to other features is usually more important than its absolute position; hence this downsampling is performed after each convolutional layer. Such a layer is entirely defined and has no trainable parameter.

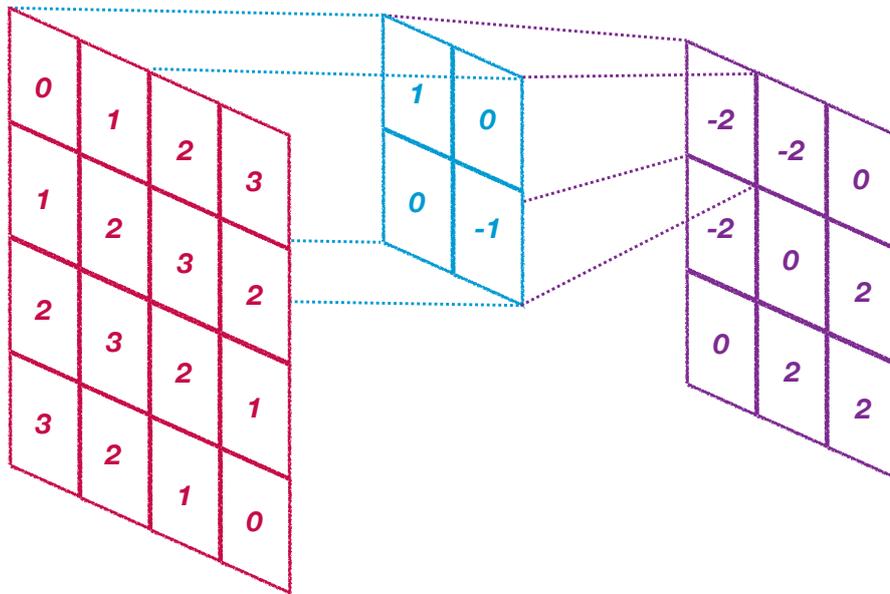


Figure III.7 – Illustration of the effect of one filter (middle) producing a feature map (right) from an input array (left). A convolutional layer can have several filters where the values of the filter are the trainable weights.

This alternation of convolutional and pooling layers makes CNN very robust against the vanishing and exploding gradient problems as the weight sharing acts as an efficient regularisation. Moreover, by combining the input through multiple applications of convolutions, CNNs learn to extract high-level features and the correlations between them. This is particularly useful in HEP. It can have intuitive applications on the response of the detector since they are already close to what is found in an image — by using the energy deposit in each pixel of the detector, for instance. Even the final reconstruction of an event can be represented as an image, with each pixel being one of the observables of the event.

III.4 Generative adversarial networks

A special field in artificial intelligence is the study of techniques with generative capabilities. The goal of such techniques is to generate samples that are possibly statistically independent from the training samples. Deep learning models are largely prevalent in this generation of all sorts of data, from text with large language models such as BERT [114] or CHATGPT and its recent GPT-4 architecture [115], to images with DALL-E [116] or MIDJOURNEY [117] or even molecules with ALPHAFOLD [118]. Along with the rise of generative pre-trained transformers (GPTs) in recent years, one of the most prominent frameworks for generation purposes is generative adversarial networks (GANs) [119, 120]. A GAN consists of two deep networks trained with competitive objectives. On one side, a generator network G is trained to map a latent vector of randomly generated inputs $\mathbf{z} \in \mathcal{Z}$ to meaningful and realistic data $G(\mathbf{z})$. On the other side, a discriminator network D is trained to distinguish between data from a sample of reference $\mathbf{x} \in \mathcal{X}_{ref}$ for which the expected output is $y(\mathbf{x}) = 1$ and generated data from the generator $\mathbf{x} \in G(\mathcal{Z})$ for which $y(\mathbf{x}) = y(G(\mathbf{z})) = 0$. It means that the generator is actually a forger trying to fool the discriminator, and the discriminator is a critic trying to spot any counterfeit (as represented in Fig. III.8).

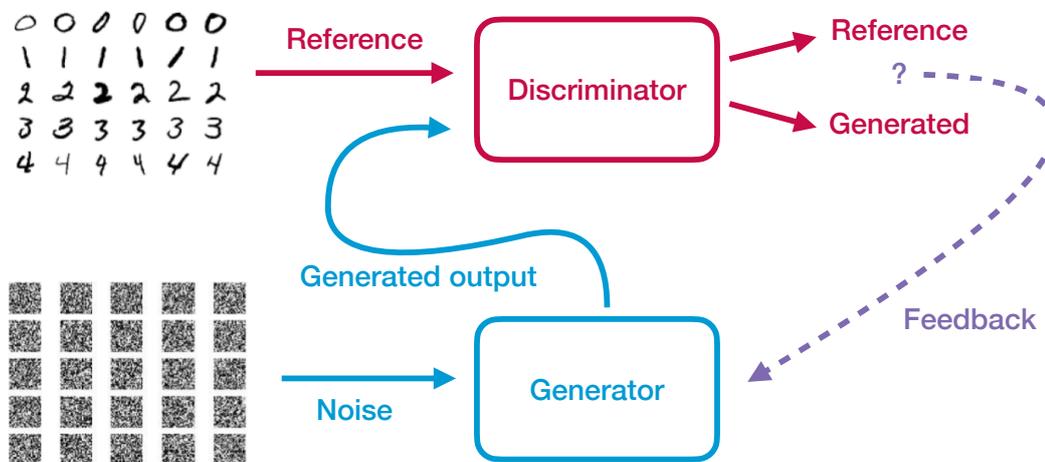


Figure III.8 – Simplified representation of a GAN.

So both networks have opposed objectives, which are translated mathematically by two different loss functions. The discriminator is classifying its inputs between reference $y(\mathbf{x}) = 1$ and generated $y(G(\mathbf{z})) = 0$ so its binary cross-

entropy loss writes as:

$$\begin{aligned}
 L_d &= \frac{1}{2 n_B} \left(\sum_{\mathbf{x} \in \mathcal{B}(\mathcal{X}_{ref}) \cup \mathcal{B}(G(\mathcal{Z}))} y(\mathbf{x}) \ln(D(\mathbf{x})) + (1 - y(\mathbf{x})) \ln(1 - D(\mathbf{x})) \right) \\
 &= \frac{1}{2 n_B} \left(\sum_{\mathbf{x} \in \mathcal{B}(\mathcal{X}_{ref})} \ln(D(\mathbf{x})) + \sum_{\mathbf{z} \in \mathcal{B}(\mathcal{Z})} \ln(1 - D(G(\mathbf{z}))) \right).
 \end{aligned} \tag{III.18}$$

Since the objective of the generator is to fool the discriminator, its loss is based on the output of the discriminator and the expected outputs are reversed, *i.e.* $y(G(\mathbf{z})) = 1$:

$$\begin{aligned}
 L_g &= \frac{1}{n_B} \left(\sum_{\mathbf{x} \in \mathcal{B}(G(\mathcal{Z}))} y(\mathbf{x}) \ln(D(\mathbf{x})) + (1 - y(\mathbf{x})) \ln(1 - D(\mathbf{x})) \right) \\
 &= \frac{1}{n_B} \sum_{\mathbf{z} \in \mathcal{B}(\mathcal{Z})} \ln D(G(\mathbf{z})).
 \end{aligned} \tag{III.19}$$

The networks are trained simultaneously as adversaries, and they compete in a zero-sum game — each gain for one of the networks is a loss for the other — and the final goal is to reach a Nash equilibrium [121] where both networks cannot improve anymore against the other.

Overall, GANs are known to be difficult models to train for several reasons. First, this zero-sum game brings the necessity for both networks to perform relatively well against the other in order to reach convergence. Indeed, if one of the networks starts to perform too well, then the other is stuck with a very high loss, no matter the direction of the weight update, meaning that its gradient gets close to 0, becoming a vanishing gradient problem. Yet, if the loss of each network should not increase or decrease too much, then a successful GAN training shows very stable losses during the training. It means that the evolution of the loss for each epoch does not give any information about the absolute performance of each network. This is unlike deep networks trained in non-adversarial settings where the loss should decrease during training and allows the direct comparison of different training frames, for instance, when looking for the best hyperparameters. Thus, the definition of additional metrics for the evaluation of the performance of the network is needed in the context of GANs. Second, the convergence of the networks toward a Nash equilibrium is not guaranteed in practice [122]. This is coming from the concurrent update of both networks, where they each try to minimise their loss through gradient descent and independently of the evolution of the other network. Since their loss is heavily dependent on the performance of the other network, each update can become a reaction to compensate for the effects of the previous update on the other network, leading easily to diverging oscillations in the evolution of the respective loss functions. Finally, an unfortunate pitfall of GANs appears when the generator needs to learn how to generate different classes of objects and generate only a few or even only one class of objects. This is called mode collapse or "the Helvetica scenario" [119] and is a direct example of what happens when the generator gets stuck in a local minima of the weights space.

Given the high popularity of GANs in the last 5 years, variations around this kind of deep model gave birth to a large variety of GAN-based techniques. A few notable architectures are presented here:

Convolutional GANs As their name suggests, convolutional GANs consist of GANs using convolutional neural networks. More precisely, the discriminator uses regular convolutional layers, but the generator is transforming a random latent vector to an organised array of data, so it usually uses deconvolutional layers. This is also coming from the fact that, in order to get networks of comparable potential in the GAN, it is easier to use networks that are comparable by construction. Thus, the discriminator and generator usually have symmetrical architectures. Convolutional GANs have a lot of applications in image generation, such as STYLEGAN [123]. STYLEGAN introduced training in a progressive fashion, beginning with a training to generate low-quality images, and after each training, a new layer is added to the generator and to the discriminator to upscale the image generation to higher quality images. STYLEGAN is famous for its generation of deceptively good portraits; it is indeed very difficult to distinguish which portrait was generated in Fig. III.9.

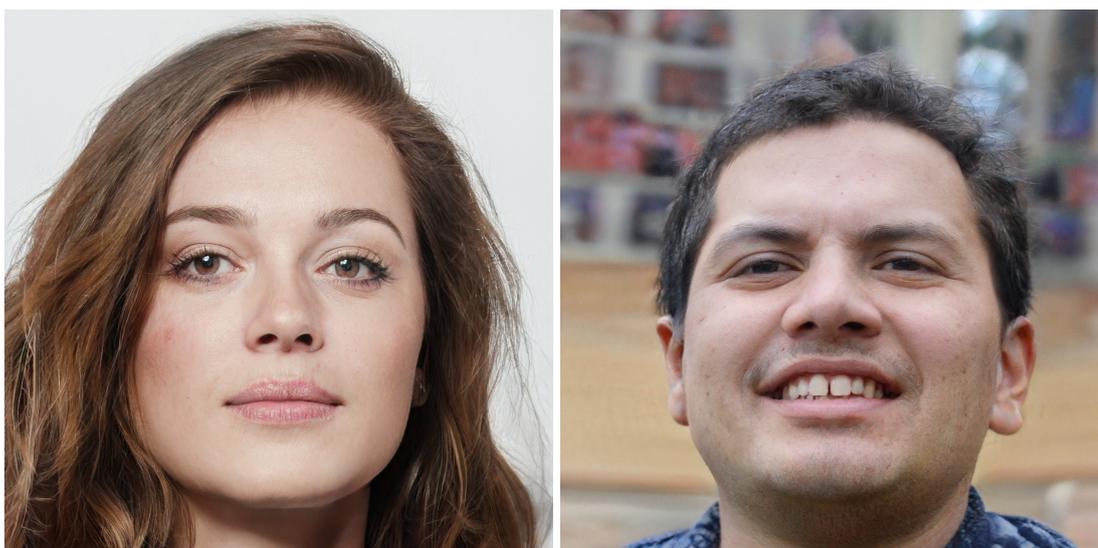


Figure III.9 – Examples of portrait generation with STYLEGAN. Both faces are actually generated and are not existing persons.

Conditional GANs With conditional GANs, some features are shown to the generator and discriminator. The GAN learns about correlations between the conditional features and the data that should be generated, hence conditioning the generator output to have specific features. This can be used, for example, to generate an image but keeping a background given as a conditional feature or to generate a random painting but in the style of a given painter.

Wasserstein GANs The concept behind this is to change its loss function rather than changing its architecture [104]. Instead of using the binary cross entropy to get the loss functions, the Wasserstein metric [124, 125] is used. It provides a meaningful loss that should allow for clear hyperparameter selection, contrary to the standard GAN. Wasserstein GANs were also developed to improve the training convergence and avoid mode collapse.

III.5 Summary

This thesis takes place in the context of the growing contribution from the machine learning field to high-energy physics research activities. For instance, the training of BDTs and simple binary DNNs is gaining momentum in collider experiments communities, and it is becoming a popular method for events classification, particle identification or object reconstruction. Reciprocally, HEP provides a favourable environment for the development and application of modern deep learning techniques, and covering them would go beyond the scope intended for this chapter. Such techniques include generative models for data sample generation using GANs (such as the study described in the next chapter) or normalising flows [126]. Other methods rely on the structure of data collected in HEP: features of an event can be seen as properties interconnected through a graph, and graph neural networks are, indeed, showing high performance on tasks such as particle identification [127, 128]. Finally, another example of active developments based on deep learning arises from the nature of research in Physics with models taking into account the uncertainty on their inputs [129–131].

This chapter summarises the working principle of neural networks and how they are trained. A particular focus is given to deep neural networks (DNN), generative adversarial networks (GAN), and regularisation techniques to improve their training. It gives a broad overview of all the considerations we take into account when training and optimising such models. Indeed, a crucial aspect of the data-driven estimation technique presented in Chapter IV is the careful optimisation of a conditional and convolutional GAN, which drives the quality of the generated background sample. In a similar way, a key element to the successful definition of phase space regions with high sensitivity for the analysis described in Chapter V is the training of a DNN model providing good discrimination of the events.

CHAPTER IV

Data Driven Estimation of Background Using GAN

IV.1 Motivations	84
IV.2 Data-driven estimation of the background in the $H \rightarrow \gamma\gamma$ analysis of CMS	85
IV.3 Generating photons with a GAN: a brilliant method?	90
IV.3.1 Methodology	90
IV.3.2 Model architecture and training	91
IV.3.3 Performance metrics	92
IV.3.4 Hyperparameters optimisation	95
IV.4 GAN performance	100
IV.5 Application to data	104
IV.5.1 Generating a $\gamma + \text{jets}$ sample with the GAN	105
IV.5.2 Generating a $\gamma + \text{jets}$ sample with a 1D PDF	106
IV.5.3 Comparison of the two methods	108
IV.6 Conclusion	112

IV.1 Motivations

In high energy physics (HEP), characterising a signal hypothesis requires distinguishing its signature from a large number of background processes with similar final states. Observables of physics objects are used to construct classification algorithms that can discriminate signal signatures from the background processes. An accurate description of these final states and their observables with the detailed modelling of the detector responses is crucial for the sensitivity of the analysis.

Monte Carlo (MC) techniques are widely used in HEP experiments to simulate a process from a physics model of interest (denoted as signal) and other standard model (SM) processes (denoted as background). Software libraries such as GEANT4 [132, 133] are used to model detailed descriptions of the modern colossal particle detectors and thereby provide an accurate simulation of the detector responses to these processes. Due to the intricate nature of these detectors, the large amount of data delivered by the colliders and the rarity of the signal, a substantial computational infrastructure in the form of grid processing power and a significantly large storage-disc volume are needed. This requirement usually constrains the simulation sample production to have a limited number of events in the tails of discriminating observables. Furthermore, inaccuracies in the underlying physics model and in the description of detector responses limit the use of MC simulations for background description.

Many leading background processes have signatures mimicking the signal due to one or more misidentified particles in their final states. For instance, the Higgs boson signature with two isolated photons in the final state has to be distinguished from other SM processes with a single photon and multiple jets in the final state where one of the jets is misidentified as a photon. A similar example can be given for the signal signatures where two b jets are expected. In this case, processes with a single b jet can have a second light flavour jet misclassified as b jet, thus populating the signal region. ML classification algorithms are widely used for optimal separation between signal and background processes. They can extract the higher-order relations between observables to provide a better classification performance [134] with respect to techniques treating observables sequentially. Therefore, the training samples should provide a good description of signal and background observables and their correlations. Modelling the misidentification of the physics objects is challenging as it might be subject to systematic effects creating discrepancies at the tails of distributions. Various data-driven techniques are used to mitigate the possible impacts of mis-modelling. For instance, a data-driven technique that requires two additional sub-sample spaces, known as the ABCD technique, is widely used in HEP, e.g. in Ref. [135]. Even though this technique can estimate the yields of the different processes, the shapes or the correlations of the observables in the high-purity signal region cannot be retained. Other techniques are also tailored explicitly for particular signatures. However, their generalisation cannot be assured, and they may still suffer from the aforementioned shortcomings.

This chapter describes my work on a novel data-driven technique using a condi-

tional generative adversarial network (GAN) to model backgrounds with misidentified particles. This GAN generates new observables for a particle that fails the identification criteria to mimic a misidentified object whose observables retain correlations with other event observables. I demonstrate the technique in the context of the CMS analysis of the Higgs boson decay into a pair of photons ($H \rightarrow \gamma\gamma$) as presented in Ref. [136, 137].

Between the beginning of this study and its publication, additional methods were introduced to tackle this objective of improving the description of background processes thanks to machine learning and generative techniques [138–140]. However, contrary to these methods, we train our generative model on a signal-enriched region in simulated MC samples and generate a complete event by making use of the large amount of discarded data events not passing the signal selection requirements. Furthermore, by using all features that may have correlations with the generated observables as conditional inputs to the GAN rather than considering only a single feature (such as the reconstructed mass of a particle, cf. [139]), the network is able to model linear and nonlinear correlations with both the generated object features and the conditional features.

This study was published in volume 83 of The European Physical Journal C [141] and presented at multiple conferences such as the 2023 Electroweak session of the 57th Rencontres de Moriond or the 26th International Conference on Computing for High Energy and Nuclear Physics (CHEP2023).

The chapter is organised as follows: a description of a typical data-driven estimation for the background of the $H \rightarrow \gamma\gamma$ analysis is given in Section IV.2. Section IV.3 describes the GAN architecture as well as its training and evaluation procedure, and the performance of the GAN is presented in Section IV.4. Finally, the impact of the method on the training of discriminants aiming at rejecting background is shown in Section IV.5.

IV.2 Data-driven estimation of the background in the $H \rightarrow \gamma\gamma$ analysis of CMS

After collecting data and reconstructing the physical objects in the event, a typical analysis flow in HEP experiments starts with identifying the physics objects. Multivariate analysis techniques are widely used to provide an identification (ID) score. For instance, a photon ID score is developed to discriminate real prompt photons γ (originating from the primary vertex) from jets reconstructed as photon $\not{\gamma}$ (named misidentified hereafter or sometimes fake in the literature) in the $H \rightarrow \gamma\gamma$ analysis. In this manuscript, the notations γ and $\not{\gamma}$ are used to identify prompt and misidentified photons, respectively. Misidentified photons mainly originate from the fragmentation of particles and the decay of neutral mesons, such as π^0 or η , into two collimated photons. During the second run of the LHC, the photon ID was computed from a boosted decision tree (BDT) trained on object properties with the highest discriminating power. These variables are mainly extracted from the reconstruction steps performed with the energy deposits measured in the ECAL. Indeed, the two boosted photons usually leave a

wider shape profile in the ECAL compared to a single prompt photon of comparable energy, as shown in Fig. IV.1.

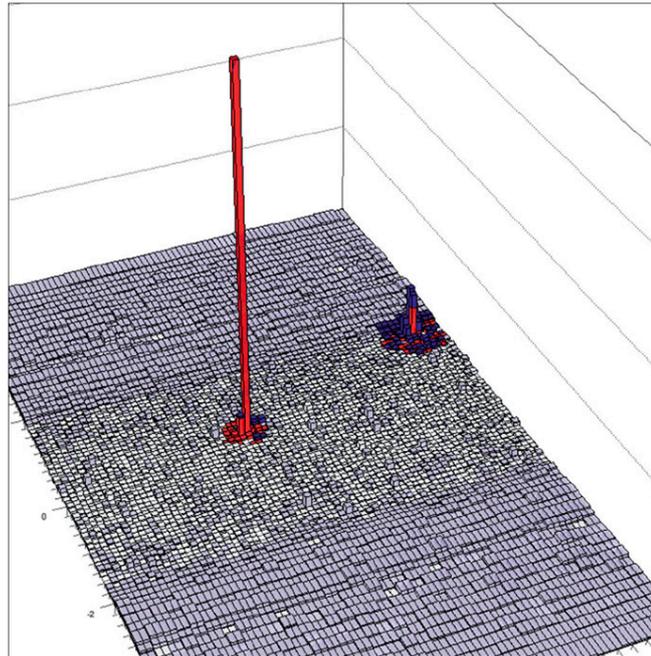


Figure IV.1 – Event display from Ref. [142] illustrating the differences in shower shape in the ECAL between prompt (narrow peak, on the left) and non-prompt (wide shape, on the right) photons.

Moreover, as these non-prompt photons emerge from the fragmentation of jets, they are often reconstructed close to additional objects. This leads to higher energy deposition in the surroundings of the photon candidate, as can be seen in Fig. IV.2. By defining the isolation as the sum of transverse energy measured in a cone surrounding a candidate particle, we can check for additional objects nearby. These two considerations drive the relevance of using shower shape and isolation variables for their discriminating potential.

The performance of the photon ID score is also improved thanks to the BDT compared to a cut-based ID, which allows a more distinct separation between prompt and misidentified photons as presented in Fig. IV.3. For a given signal efficiency, the background efficiency is lower on the BDT curve than the cut-based dots, meaning that more background is rejected. Examples of inputs used for the BDT include the isolation variables of Fig. IV.2, the sum of energy deposits and their distribution in neighbouring crystals or the η position of the energy deposit.

Object candidates passing a certain threshold on the ID score are identified as physics objects (e.g. signal photons). Additional selection criteria are applied to choose events with similar final states to the signal process of interest, creating a signal region named SR. Various background processes with similar signatures may pass these selection criteria. In the context of the $H \rightarrow \gamma\gamma$ analysis, photon candidates in the final state are selected by imposing photon ID criteria. This requirement on the ID of the two expected photons defines the SR

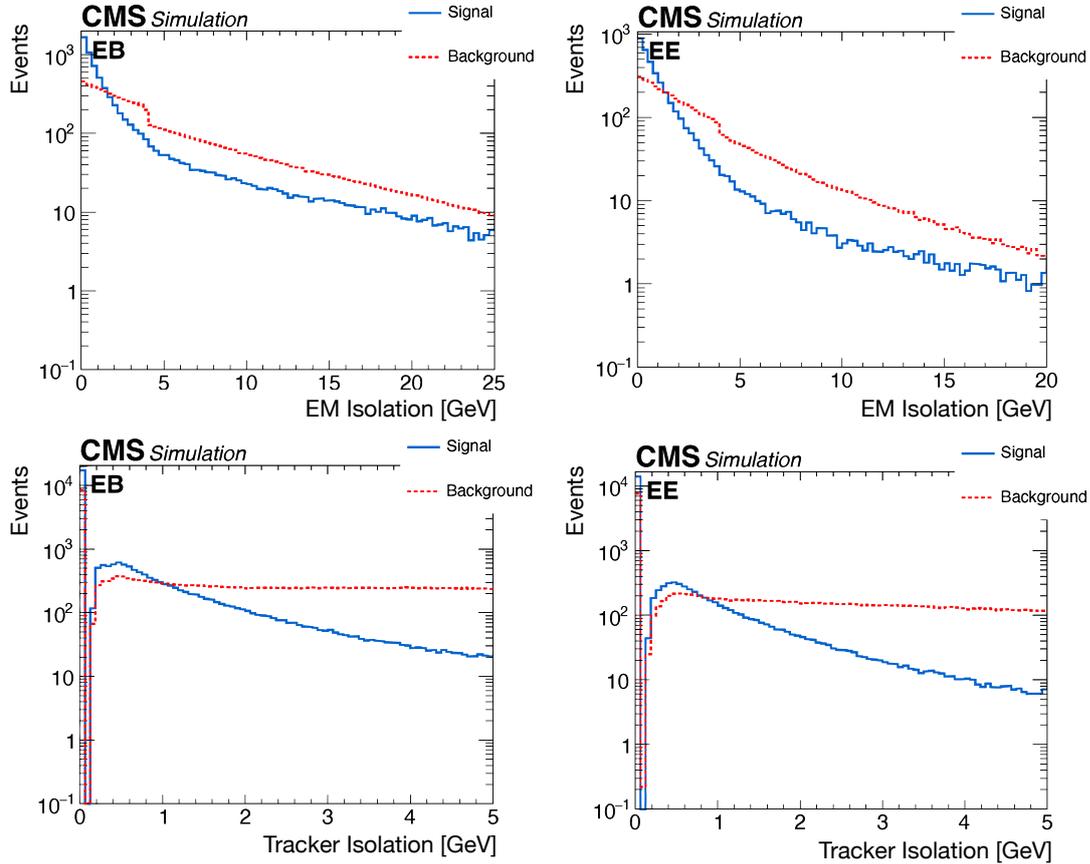


Figure IV.2 – The isolation is defined as the sum of transverse energy measured in a cone of $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the photon candidate as presented in Ref. [143]. Distributions are shown for photons reconstructed in the ECAL barrel (EB) or endcap (EE) and taking into account energy deposited by photons (EM isolation) or charged hadrons (Charged Isolation) according to the particle flow algorithm. These isolation variables allow us to discriminate between prompt photons (in blue) and misidentified photons (in red).

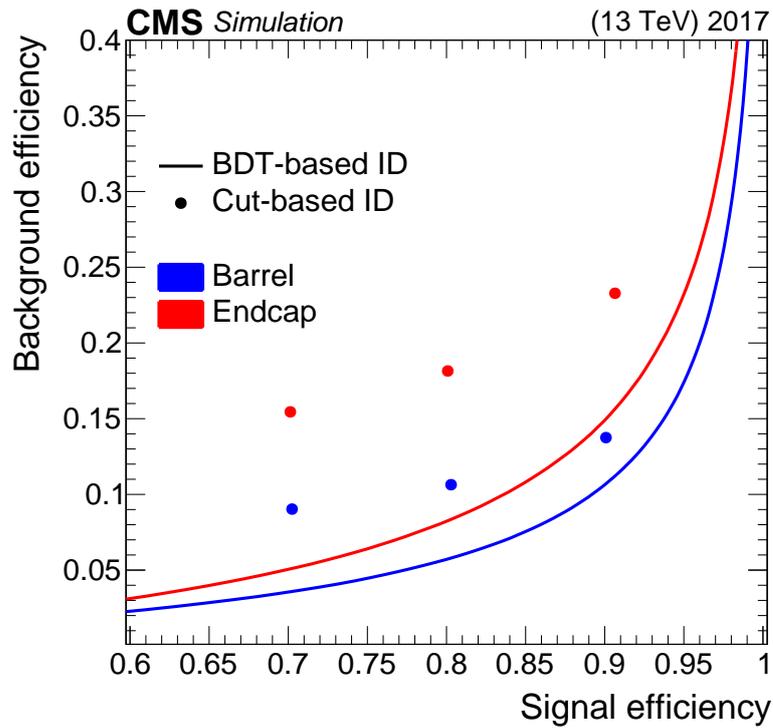


Figure IV.3 – Comparison from Ref. [144] of the identification of prompt photons when relying on the BDT score (line) or on loose, medium and tight cuts on the object features (dots).

in this work. A jet passing photon-ID-selection criteria is thus misidentified as a photon, contributing to the background in the SR. Indeed, processes with a photon and multiple jets ($\gamma + \text{jets}$) in the final state are major backgrounds in the $H \rightarrow \gamma\gamma$ analysis as well as processes solely composed of jets produced through the strong interaction where two of them are misidentified as photons (multijet events MJ).

The contribution of the background processes in data can be clearly identified in Fig. IV.4, where signal events in red are overwhelmed by several orders of magnitude of background events in blue. These histograms represent the distribution of the photon ID with the lowest score, and they illustrate the separation potential offered by the ID algorithm since the $\gamma + \text{jets}$ and MJ background processes have a lower minimal ID than processes with no misidentified photons (Fig. IV.4). Using features of the selected objects, a multivariate technique can further be used to increase the signal purity in the SR. They usually rely on the ID score of the photon, among other features, given its discrimination potential.

To build an efficient discrimination for background processes, the widespread strategy in HEP is to rely on a precise description of such background behaviour by means of simulation. Unfortunately, the MC simulations of the $\gamma + \text{jets}$ and MJ have a limited number of simulated events, and they show discrepancies when compared to data (due to reasons already mentioned in Section IV.1). This is particularly true when the analysis focuses on SRs with additional selection criteria, e.g. on the jet selection. These shortcomings led to an estimation of the background based on data outside of the SR in some $H \rightarrow \gamma\gamma$ studies such as

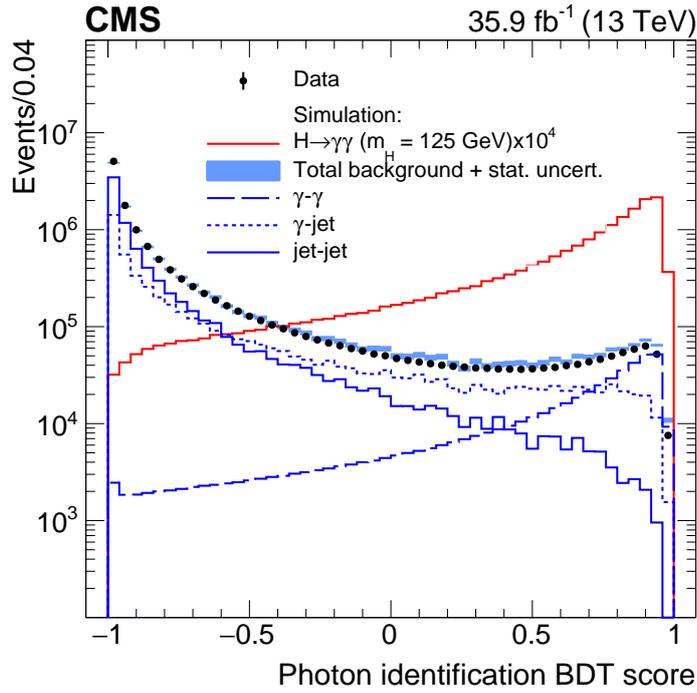


Figure IV.4 – Distribution of the minimal photon ID score for simulated background processes (in blue), signal processes (in red), and in data (black dots).

Ref. [145]. The method relies on selecting a control region (CR) using data events where one photon fails the ID selection criteria as illustrated in Fig. IV.5. Most data events with one object failing photon selection criteria, and the other one passing the selection, are $\gamma + \text{jets}$ or MJ events and constitute a CR with similar physics properties as the SR, especially in terms of the number of jets, jet identifications, jet kinematics, jet flavours, etc. This approach brings improvements in two ways. First, by using data events of the CR that would be discarded otherwise, we gain access to a sizeable amount of events, often offering a larger number of events than the SR of the MC simulations. Then, it grants a better modelling of the physics (even if a reweighting is sometimes needed to account for differences in the kinematics of the photons between the two regions).

Yet, by selecting events where one photon fails the ID requirement to build the CR the ID of the misidentified photon cannot be used any longer as it is falling outside the range of what is expected in the SR. Since the photon ID is one of the crucial features used to reject background, a workaround to retrieve the photon ID is to generate a new one. A first approach for generating a new ID is described in Ref. [25]. By extracting the 1D probability distribution function (PDF) of the minimal photon ID in the SR, a new ID can be randomly generated in the correct range. Nevertheless, a random generation using a 1D PDF is not reproducing the correlations of the photon ID with other observables of the misidentified photon (e.g. p_T , η , ...) or with other observables of the event (e.g. p_T of the other photon, number of jets, ...). Furthermore, there are kinematics differences between events in the SR and events in the CR. And even if a reweighting of the events is possible, there is some subjectivity in selecting the

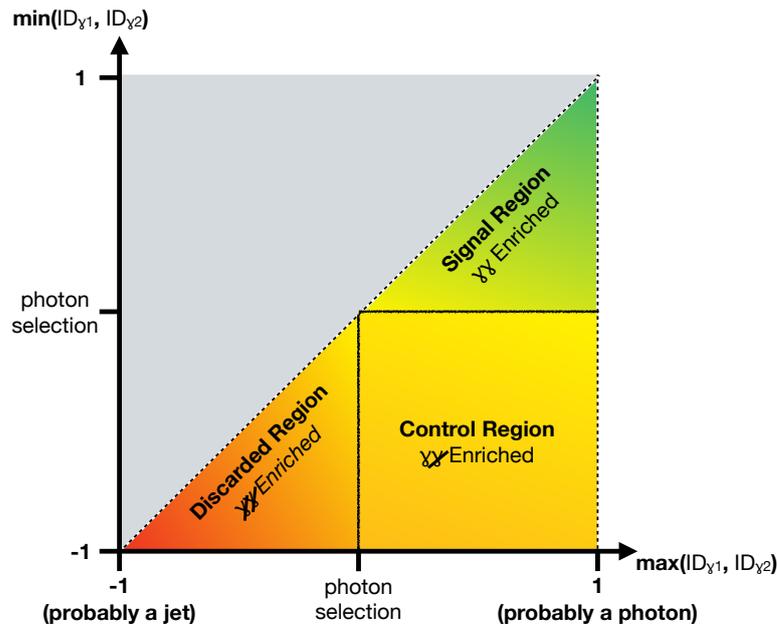


Figure IV.5 – Diagram of the signal and control region selection in the $H \rightarrow \gamma\gamma$ analysis.

final set of reweighting features, and it can lead to a suboptimal description of the background in the SR.

That is why we propose to improve the generation step thanks to a GAN by generating the full misidentified object. It resolves the issue of replacing the photon ID while keeping realistic correlations with other observables, and at the same time, it produces a misidentified photon whose kinematics match the expectations in the SR.

IV.3 Generating photons with a GAN: a brilliant method?

IV.3.1 Methodology

Here is a detailed description of the entire proposed strategy :

1. Selection of the SR of interest where the main background sources are processes with misidentified objects.
2. Definition of an orthogonal CR, based on the properties of a misidentified object (e.g. photon ID of a misidentified photon).
3. Training of the GAN in the SR using simulated background events and optimising the hyperparameters to reach the generator model with optimal performance. The misidentified object used in the training is selected with the MC truth information.
4. Using the generator network of the GAN, new objects are generated to replace misidentified objects in events of the CR of data. In data, the misiden-

tified object is taken as the one failing the SR criteria.

5. The produced sample can be used to define the analysis strategy (training of ML techniques for signal extraction, optimisation of selections, ...).

In the context of the $H \rightarrow \gamma\gamma$ analysis, the goal of the GAN is to generate a photon that replaces the photon failing the ID selection criteria. In that sense, it generates a new misidentified photon that passes the ID requirement, effectively converting events from the CR into events of the SR. Generating a photon means generating its associated observables: $p_{T,\gamma}$, η_γ , ϕ_γ , ID_γ , and then recomputing the related features of the event such as the diphoton pair p_T , the mass of the diphoton system, *etc.*. As the GAN needs to learn the observables and correlations of a misidentified photon (originating from a jet) from background events comparable to signal events, the GAN is trained with events from the SR of a $\gamma + \text{jets}$ simulation. This is done with a $\gamma + \text{jets}$ simulated dataset from CMS open data [146] containing around 2.7 million events where a specific filter ensures that they contain at least two reconstructed photons.

IV.3.2 Model architecture and training

In order for the GAN to learn the correlations between the generated outputs and the rest of the event, it is necessary to show the model the features of interest that it needs to consider while generating a misidentified photon. In addition, correlations are extracted by observing specific patterns in relations among multiple features, so it leads naturally to the choice of a convolutional architecture for both networks composing the GAN. Our GAN model is both conditional and convolutional, following what is described in Section III.4 of Chapter III. Actually, it is an extension of the deep convolutional GAN architecture described in Ref. [147] with the addition of conditional features as illustrated in Fig. IV.6. Unlike regular GANs, which only use randomly generated inputs for the generator, the conditional architecture allows the GAN to reproduce correlations. These conditional features are concatenated to the inputs of the generator and also to the inputs of the discriminator. The architecture of the GAN was studied as one of the hyperparameters of the training; a more detailed description of the networks' architecture is given in section IV.3.4 after the full optimisation.

Training and validation samples are taken from an MC simulation of the process of interest in the SR. For each event i , a set \mathbf{x}_i of n_{feat} features is defined. From this vector, a subset of conditional features $\mathbf{x}_{i,\text{cond}}$ of size n_{cond} is selected and concatenated to a vector of random latent features sampled from a normal distribution \mathbf{z}_i of size n_{rand} to form an input vector to the generator model $g : \mathbb{R}^{n_{\text{cond}}+n_{\text{rand}}} \rightarrow \mathbb{R}^{n_{\text{feat}}}$. The goal of the generator model is to produce a set of observables of size $n_{\text{out}} \equiv n_{\text{feat}} - n_{\text{cond}}$ describing a misidentified object in the SR (*e.g.* transverse momentum p_T , pseudorapidity η , ...). These generated observables are, in turn, concatenated to the set of conditional observables forming an output vector $g(\mathbf{x}_{i,\text{cond}}, \mathbf{z}_i)$. The output vectors are used and compared to the original observables \mathbf{x}_i to train the discriminator model $d : \mathbb{R}^{n_{\text{cond}}+n_{\text{out}}} \rightarrow [0, 1]$. To distinguish between the generated and original observables, the events are coupled to a discriminator label y_i with the choice of 0 for the generated ob-

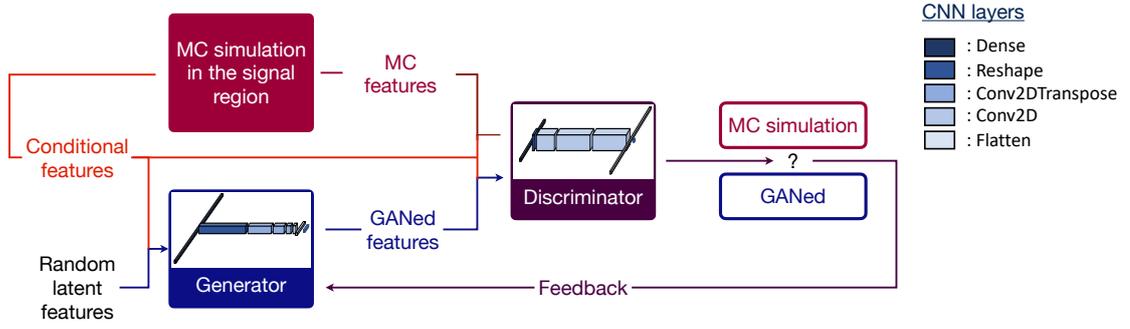


Figure IV.6 – Layout of the proposed conditional GAN architecture training. Conditional features and random latent features are given as inputs to the generator model. The generator model creates a new object with properties of a misidentified object of the SR. The discriminator model is expected to classify between MC-simulated events and GANed events.

servables ($\hat{\mathbf{y}}_i = g(\mathbf{x}_{i,\text{cond}}, \mathbf{z}_i)$, $l_i = 0$) and 1 for the original ones ($\hat{\mathbf{y}}_i = \mathbf{x}_i$, $l_i = 1$). Additional noise is added to the training labels as it has been shown to increase the chance of convergence for GAN [148] and is called label smoothing. The noise value is generated uniformly between 0 and a maximum ϵ , meaning $l_i \in [0, 0 + \epsilon] \cup [1 - \epsilon, 1]$. Finally, the objective for the discriminator model is to return a value as close as possible to the input label (*i.e.* $d(\hat{\mathbf{y}}_i) = l_i$).

Based on the binary cross-entropy, the following loss functions L_d and L_g are respectively defined for the discriminator and generator networks:

$$L_d = \frac{1}{n_B} \sum_{i=1}^{n_B} l_i \ln(d(\hat{\mathbf{y}}_i)) + (1 - l_i) \ln(1 - d(\hat{\mathbf{y}}_i)) \quad (\text{IV.1})$$

$$L_g = \frac{1}{n_B} \sum_{i=1}^{n_B} \ln d(g(\mathbf{x}_{i,\text{cond}}, \mathbf{z}_i)) \quad (\text{IV.2})$$

where n_B is the batch size (defined in Section III.2.3 of Chapter III). Trainable weights of the network are initialised following the Glorot uniform distribution defined in Eq. III.15 of the previous chapter. Then, the weights of events are updated after computing the loss for a given batch of events by using backpropagation and optimizers derived from the gradient descent algorithm. Three optimizer algorithms already described in Section III.2.3 of Chapter III are tested: the stochastic gradient descent algorithm (SGD) [149], the Adam optimizer [98] and the LAMB optimizer [150].

IV.3.3 Performance metrics

As the generator and discriminator models are trained adversarially in a zero-sum game, their loss functions do not reflect the absolute performance of each neural network. And because the losses of the two networks are balanced, an independent figure-of-merit is required to assess the performance of the generator. Once again, one of the primary objectives of this study is to make the generator model learn the correlations between the observables of the event. Therefore, we propose a performance score based on the negative log-likelihood of

the underlying probability distributions of the generated and original samples. We demonstrate that this score suffers from large fluctuations, intrinsic to the nature of GANs, and presents a way to stabilise it. This allows the selection of the best-performing model without introducing a significant computational overhead, an appreciated possibility for the optimisation of the hyperparameters of the training.

From an input sample $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ with N events and a sample of random vectors $\mathcal{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$, an output sample $\hat{\mathcal{Y}} = \{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_N\} = g(\mathcal{X}, \mathcal{Z})$ is generated with the GAN. Events are distributed in bins numbered from 1 to N_b , which span a multidimensional space chosen to reflect the physics one wants to reproduce (in our case p_T, η , ID score of the misidentified photon in the $H \rightarrow \gamma\gamma$ analysis, $p_{T\gamma\gamma}/m_{\gamma\gamma}$). We define a negative log-likelihood performance metric (NLL metric) as:

$$-2 \ln \Lambda(\mathcal{X}|\hat{\mathcal{Y}}) = -2 \sum_k^{N_b} m_k \ln p_{\hat{\mathcal{Y}}}(k) \quad (\text{IV.3})$$

where we use the frequency $p_{\hat{\mathcal{Y}}}(k) = n_k/N$ as an estimation of the probability of an event to fall in bin k estimated from sample $\hat{\mathcal{Y}}$ and $m_k(n_k)$ is the number of $\mathbf{x}_i(\hat{\mathbf{y}}_i)$ in bin number k . For each dimension in the log-likelihood, we transform the distribution of the corresponding variable to be uniform so we can use 10 bins by dimension, which is enough to capture the variable shape while retaining a sufficient amount of data per bin.

This metric is computed on the training sample and on an independent validation sample after each epoch to check for any overtraining effect. The optimal state of the GAN is then chosen as the set of weights giving the lowest $-2 \ln \Lambda$ value on the validation sample.

The balance needed between the performance of the two networks makes them prone to collapse towards suboptimal states, which produces a poor description of the event observables. Even when the GAN converges, large fluctuations are usually observed in the performance metrics. These fluctuations make the optimization of the network challenging. An averaging method was developed to overcome this limitation and better assess the performance of the GAN. The careful reader remembers that a crucial aspect of the proposed GAN architecture is that features of an event are given as input to the model in addition to the random latent space. It means that a GAN generator can produce different objects for the same event by using several random vectors for the same conditional variables (*i.e.* $g(\mathbf{x}_i, \mathbf{z}_i) \neq g(\mathbf{x}_i, \mathbf{z}'_i)$). An example is presented in Fig. IV.7 by generating ten thousand objects for a set of conditional features coming from one event of the SR. As the GAN learns the correlations between conditional features and properties of the misidentified photons, the generated distributions (in red) deviate from the original distributions of the SR (in blue) to form an excess around the actual values of the replaced photon (vertical lines). However, by generating only one photon per event when evaluating the performance of the GAN, we are not properly estimating the location of the excesses, *i.e.* its understanding of the correlations. In the context of performance metrics, by generating multiple objects, the GAN output is effectively averaged over the random latent space, giving a more accurate estimator $p_{\hat{\mathcal{Y}}}$ in Eq. IV.3. This effect is

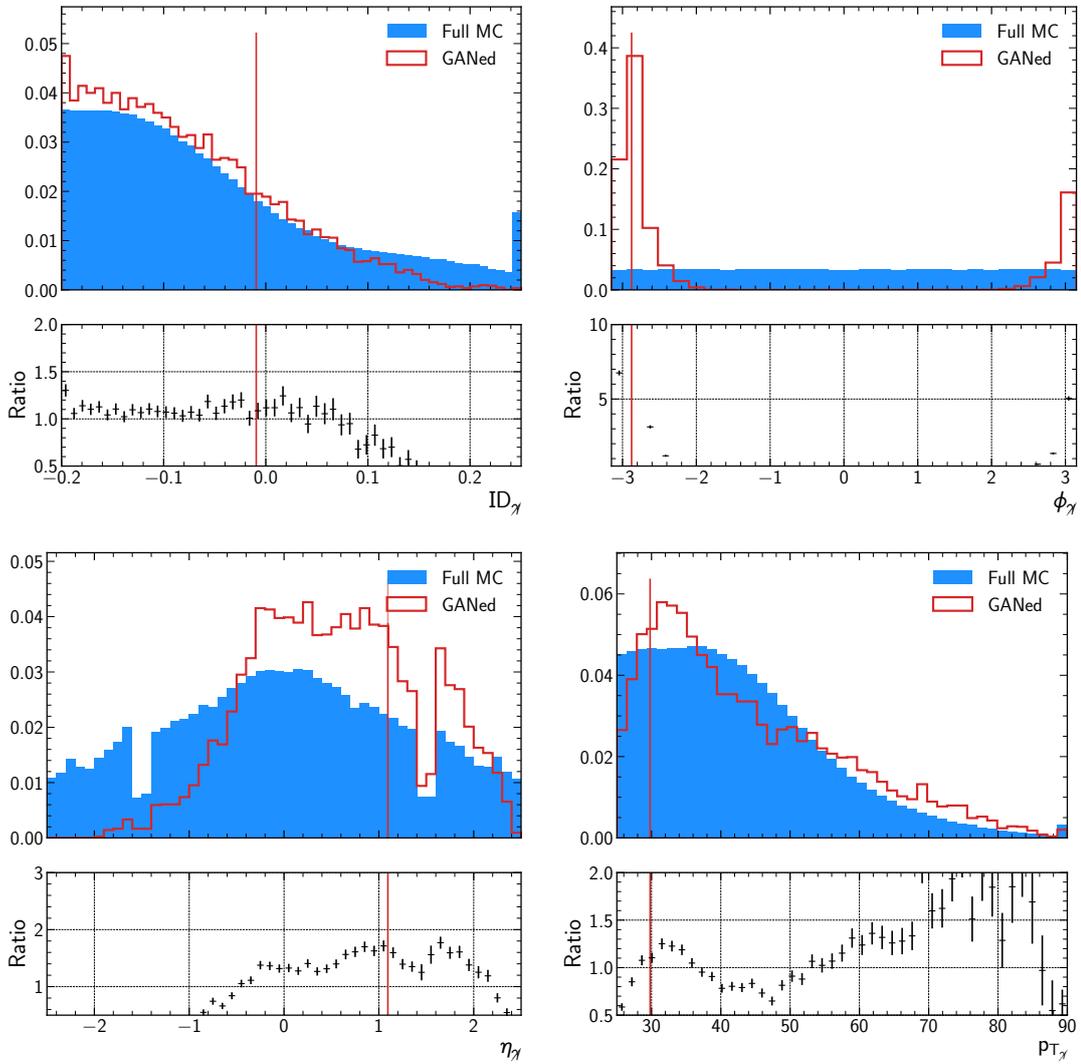


Figure IV.7 – Example of generation of 10000 objects with the GAN from one set of conditional variables (in red). The generated distributions are compared to the original distributions described in the SR of the MC sample ("Full MC" in blue), and the original properties of the misidentified photons are shown with vertical lines.

demonstrated in Fig. IV.8 by using three different configurations where we generate 1, 10 and 100 objects for each event. Fluctuations are decreasing when more objects are generated per event. As the evaluation time increases with the number of generated objects, a compromise of 100 objects per event was chosen in this work.

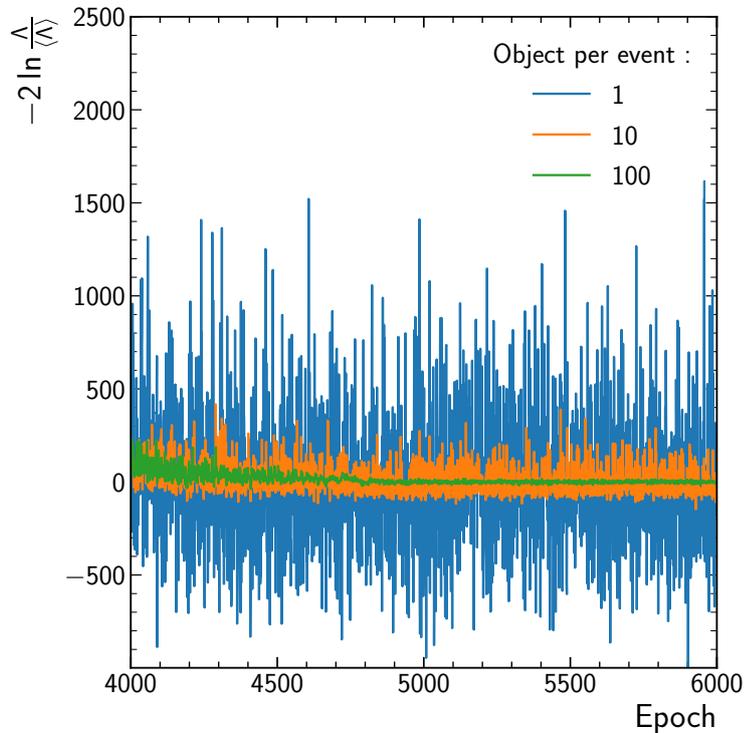


Figure IV.8 – Comparison of the $-2 \ln \Lambda$ metric evolution during training for different numbers of objects generated per event using different random latent spaces. Each metric is shifted by its average over the last thousand epochs ($\langle \Lambda \rangle = \prod_{i=5000}^{6000} \Lambda_i^{0.001}$, where i refers to the epoch number) to better illustrate the stabilisation of the metric.

IV.3.4 Hyperparameters optimisation

Thanks to the negative log-likelihood metric and the reduction of the fluctuations, it becomes possible to efficiently rank different training strategies, different models, etc. One of the first studies on improving the training concerned the preprocessing of the input features. The goal of the preprocessing step is to transform the original input vectors into a representation more suitable for the training of neural networks. This transformation needs to be bijective so that a transformed vector can be processed back to its original values in a unique way. It can mitigate the impact of rapidly falling or non-smooth distributions of the physics observables (e.g. due to detector effects), those being harder to learn for a network. Multiple preprocessing methods are tested from the Scikit-learn module [151], and the best performance is obtained with the quantile transformation to a uniform output. The quantile transformation relies on the method of inverse random sampling: for any random variable $X \in \mathbb{R}$ with cumulative distri-

bution function (CDF) F , the inverse CDF (or quantile function) F^{-1} applied on a random variable $U \sim \mathcal{U}(0, 1)$ gives a random variable following the distribution of X , *i.e.* $F^{-1}(U) \sim X$. This method is used to generate random numbers following any distribution. Knowing that a CDF applied to its associated random variable is also a random variable following a uniform distribution $F(X) \sim \mathcal{U}(0, 1)$, then a transformation from one distribution X to another one Y with CDF G can be derived using $G^{-1}(F(X)) \sim Y$. In our case, the quantile transformation is fitted to map the distribution of the conditional and generated features to follow $\mathcal{U}(-1, 1)$. The impact of the preprocessing step on the training is visible in Fig. IV.9, where we can clearly see that it helps the GAN to learn non-smooth distributions such as the gaps in the pseudorapidity distribution due to the junction between the barrel and the endcap of the detectors.

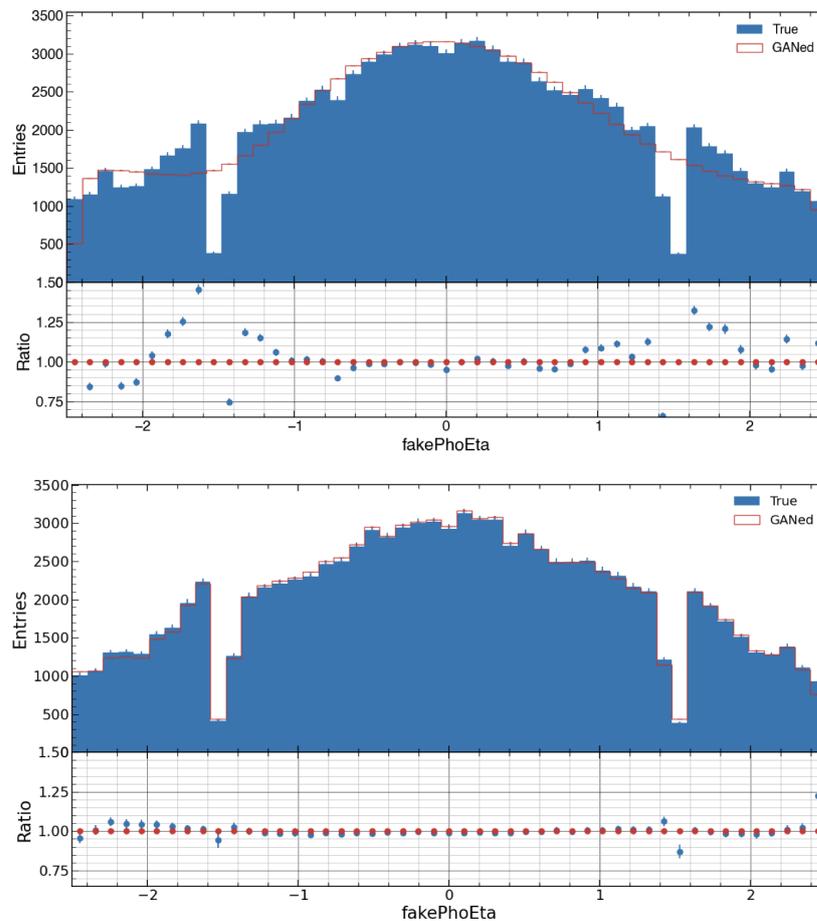


Figure IV.9 – Comparison of the distribution of η_γ when simulated (in blue) or generated by the GAN (in red) with (below) and without (above) quantile transformation before training.

Another major tuning was the identification of the best set of conditional features. Three different training strategies with respect to the feature sets are considered and compared. The first strategy follows the vanilla GAN application as described in Section IV.3, *i.e.* the latent space is purely composed of random variables. In the second strategy (partial set), we consider as conditional features the observables of the prompt object. These conditional features, together with

the random latent ones, are used as input to the GAN. The last strategy (full set) takes as input random latent features and an extended set of conditional features: the observables of the prompt object together with additional event observables, *i.e.* $ID_{\gamma}, p_{T\gamma}, \eta_{\gamma}, \phi_{\gamma}, N_{\text{jets}}, N_{\text{vtx}}$. We performed three trainings corresponding to these three strategies to test the impact of the conditional features. As shown in Fig. IV.10, the best-performing strategy is to use the full set of conditional features as the training loss reaches the lowest value of $-2 \ln \Lambda$. Besides

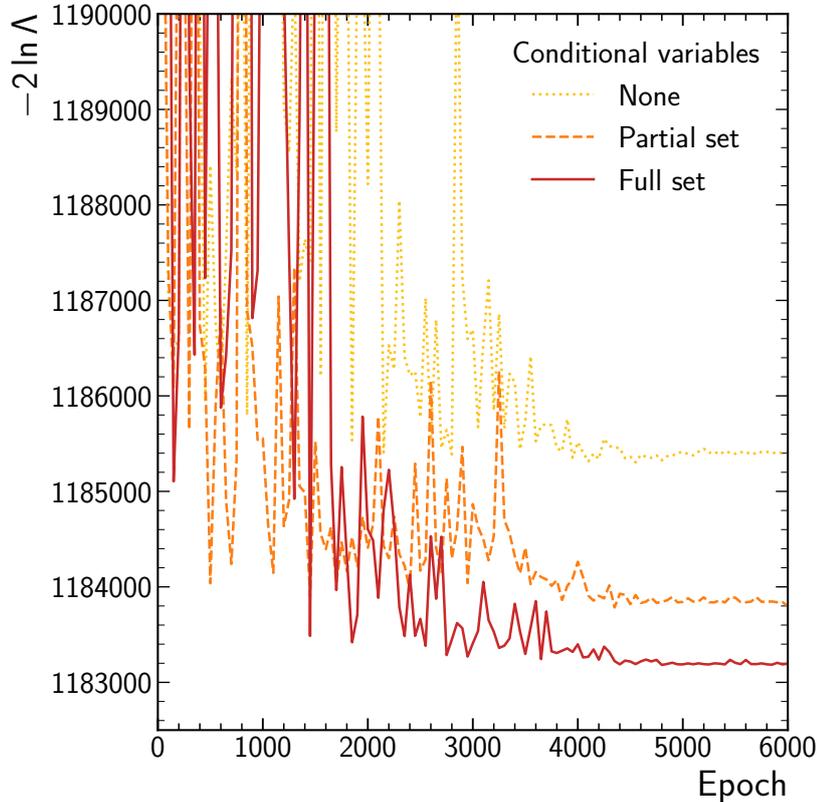


Figure IV.10 – Comparison of the $-2 \ln \Lambda$ figure of merit evolution during the training of the different strategies tested for the addition of conditional features to the random latent ones.

the values of the NLL metrics, additional distributions were drawn to check if the GAN is properly reproducing the correlations. The effects of conditional features on the GAN ability to learn correlation among features of the event are compelling from distributions such as Fig. IV.11. When no conditional features are shown (in yellow), the generated objects have features independent from one another or with respect to the rest of the event. This flat distribution is actually what one would get when using a 1D PDF for ID generation. With a partial set of conditional features (in orange), the GAN learns some correlation between features. Finally, with the full set (in red), the GAN is able to reproduce the correlations of the original sample (in blue).

We further investigate the impact of different parameters of the models in the GAN. For instance, we find that increasing the dimension of the latent space to more than 32 does not provide additional performance improvement, as demonstrated in Fig. IV.12.

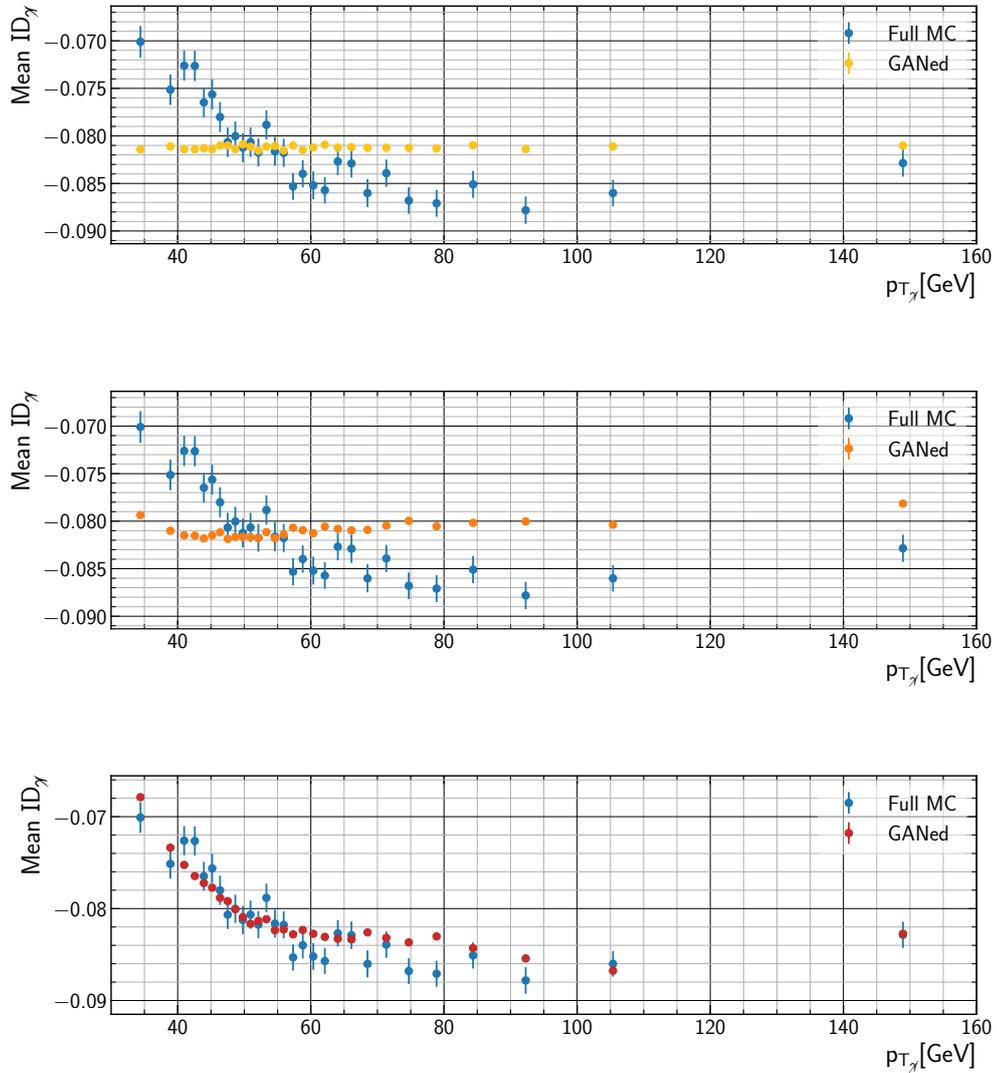


Figure IV.11 – Scatter plots showing the correlations between two properties of the misidentified photon γ . An average of its ID is computed for each p_T bin. The three different trainings with different sets of conditional variables (no conditional variables at the top, partial set in the middle, full set at the bottom) are compared to the SR of the MC simulation ("Full MC" in blue).

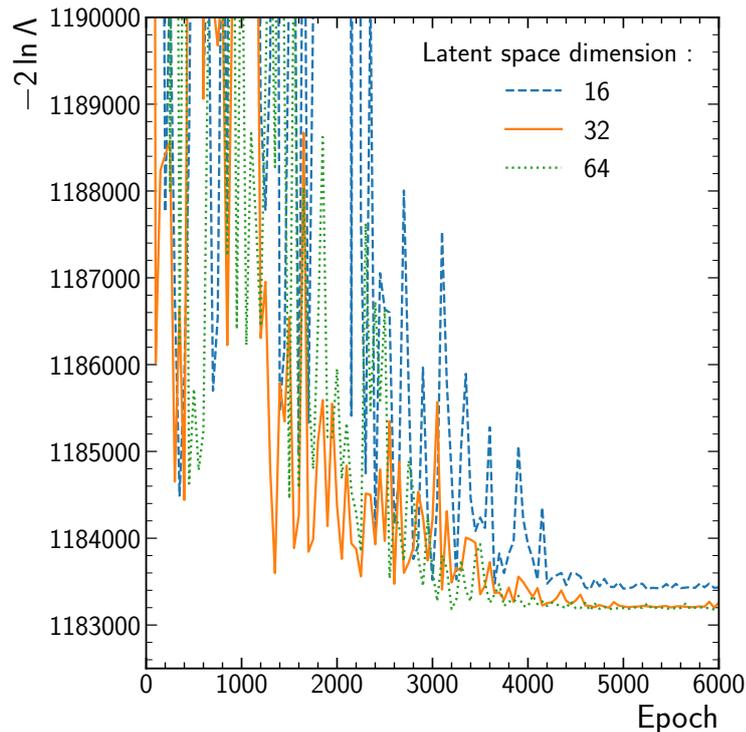


Figure IV.12 – Example of hyperparameter optimisation where different configurations are tested. Lines with different colours correspond to different numbers of dimensions of the random latent spaces given as input to the generator model.

Overall, the optimised parameters are found to be:

- *random latent space dimension* : 32
- *number of training events* : 100,000 events
- *gradient descent optimizer* : LAMB optimizer
- *learning rate* : cosine decay as described in [113] starting at 0.001 and reaching 0 after 5,000 training epochs
- *batch size* : 1024
- *noise on training labels* : 0.15
- *model architecture* :

generator : a dense input layer of 1024 nodes, three 2D deconvolution layers with 32/16/8 filters of size 4x4/2x2/2x2 respectively, one 2D convolution with 4 filters of size 3x3 and a dense layer with 4 outputs with hyperbolic tangent activation function

discriminator : a dense input layer of 256 nodes, three 2D convolution layers with 32/64/64 filters of size 2x2/2x2/4x4 respectively and with a LeakyReLU activation function [152], a dense layer with 1 output with sigmoid activation function. A dropout [105] of 20% is also implemented before the last layer of the discriminator.

IV.4 GAN performance

As described in the previous sections, we select an SR with two photons passing stringent photon ID criteria $ID_{\gamma_{1,2}} \geq -0.2$. A CR composed of events with one photon candidate passing the ID criteria and another one not passing them is formed. As the second object fails to pass the photon selection criteria, it is likely to originate from a jet. This latter object is replaced with a misidentified photon γ generated by the GAN model, thus with SR properties. The striking transformation capability of this technique is demonstrated in Fig. IV.13. The p_T distribution of the GANed γ matches the distribution of the same observable in the SR, while the MC-simulated misidentified object from the CR has different characteristics. This is a strong piece of evidence that the GAN adapts the kinematics of the misidentified object where a reweighting of the events would be needed with a generation of the ID only.

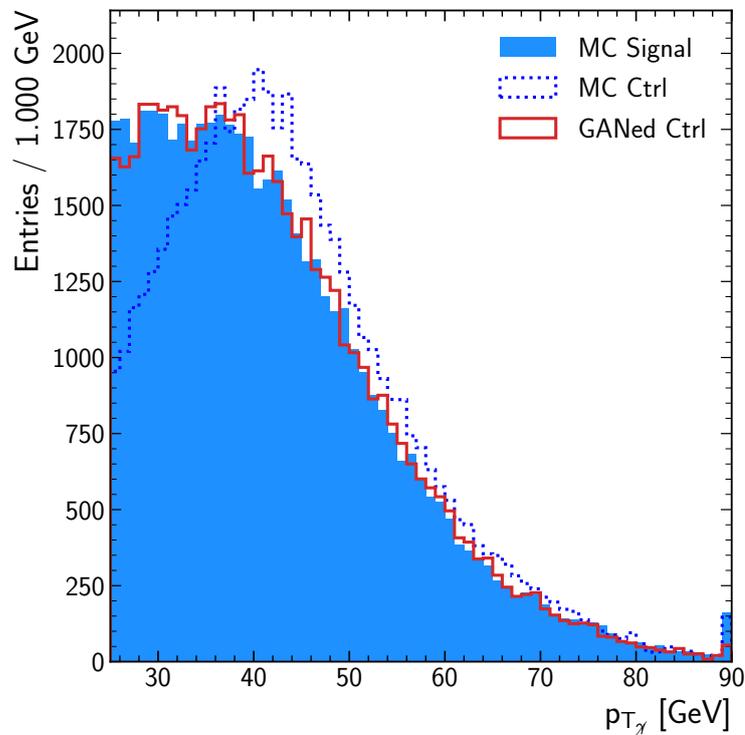


Figure IV.13 – Distribution of the misidentified photon p_T in the MC SR (MC Signal), MC CR (MC Ctrl) and a misidentified photon generated by the GAN using observables of the events in the CR (GANed Ctrl).

Figure IV.14 shows an excellent agreement of the GANed-object observable distributions (named GANed in Fig. IV.14) compared to the ones from actual misidentified photons of the SR (named Full MC in Fig. IV.14). Furthermore, the fact that the isopleths match between the two distributions indicates that the generator also reproduces the correlations between GANed and original Full MC observables. Additional scatter plots illustrating some of the correlations learnt by the GAN between the misidentified photon and the rest of the event are shown in Fig. IV.15.

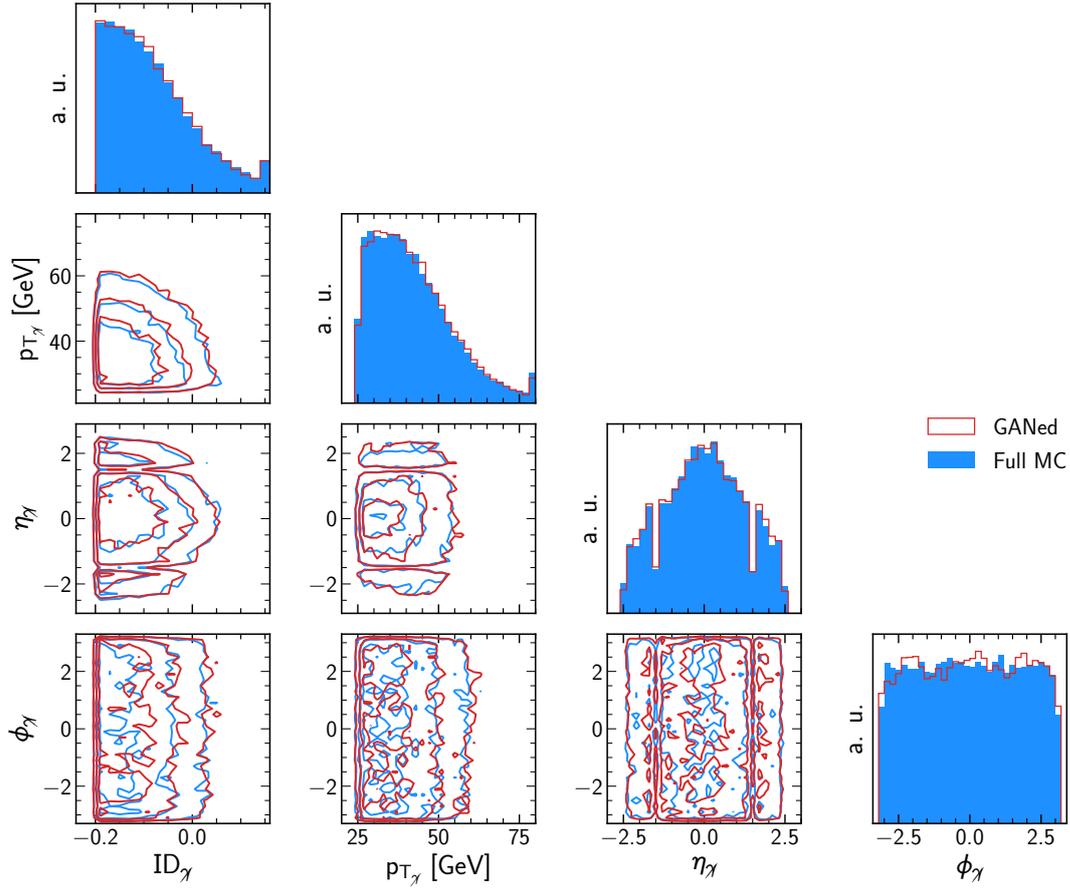


Figure IV.14 – Diagonal: Comparison of the misidentified photon-observable distributions from the MC SR (named Full MC) and from the application of the GAN to the MC CR (named GANed). Off diagonal: Contour lines (named isopleths) containing 25 %, 50 % and 75 % of the events and showing correlations between pairs of observables.

In order to assess the performance of the GAN in terms of reproduction of the correlations between observables, we use the distance correlation as defined in Ref. [153]. This metric allows us to quantify both the nonlinear and linear correlations between the observables of the event. The distance correlation coefficient goes from 0 to 1, where 0 indicates that two observables are independent and nonzero values mean that there exists some correlation (linear or nonlinear). We measure the correlations between the misidentified photon properties and other event observables (prompt photon properties, N_{jets} , N_{vtx} ...) for both the SR and the CR with a GANed misidentified object. These correlation matrices are shown in Fig. IV.16. We use a χ^2 to evaluate the difference between the 2 matrices defined as:

$$\chi \equiv \frac{1}{N_{\chi^2}} \sum_{i < j} \sqrt{\chi_{ij}^2} \quad \text{with} \quad \chi_{ij}^2 \equiv \frac{\left(d_{ij}^{\text{GAN}} - d_{ij}^{\text{SR}}\right)^2}{\sigma_{ij}^2},$$

where (i, j) is a pair of observables, d_{ij} the corresponding distance correlation and N_{χ^2} the number of couples with $i < j$. The uncertainty σ_{ij} on the computation of the distance correlation for both the MC and GANed samples is estimated

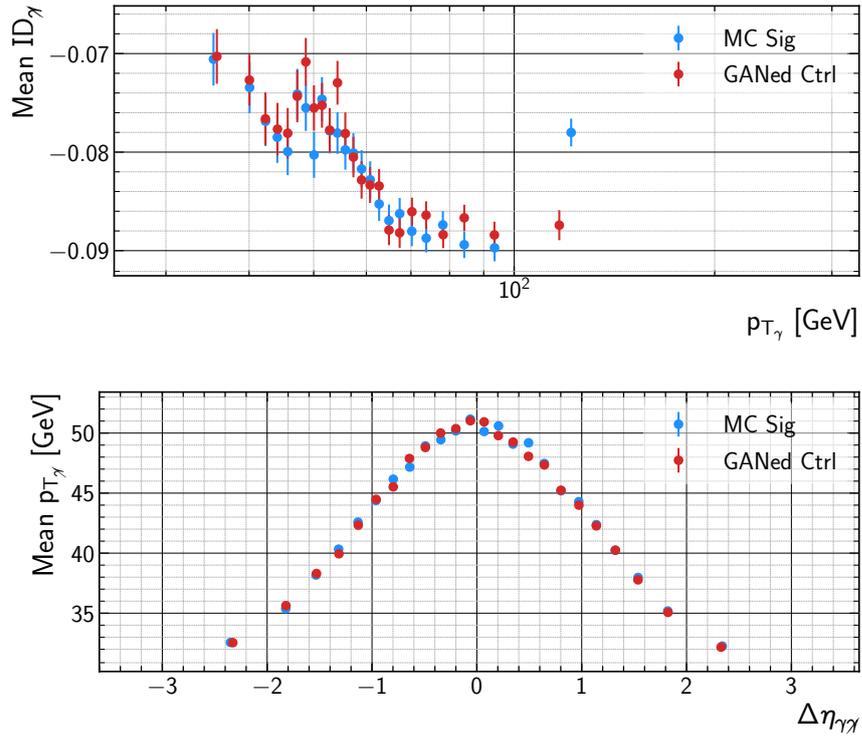


Figure IV.15 – Scatter plots constructed in a similar fashion as Fig. IV.11 and showing correlations between features of the generated object and other features of the event. Correlations generated with the GAN applied on the CR (GANed Ctrl in red) are compared with the target correlations from the MC SR (MC Sig in blue).

through two contributions. The first one comes from the finite size of the MC sample and is evaluated by splitting the sample, computing the distance correlation coefficients of each subsample and extracting their standard deviation over all the subsamples σ_{ij}^{MC} . The second contribution is introduced by the random latent space used as part of the GAN input. As described already in Section IV.3.3, we rely on the generation of multiple objects for the same event to assess the level of uncertainty. A hundred different samples are generated with the GAN from the same initial events, and the distance correlation coefficients are computed for each sample allowing the determination of their standard deviation σ_{ij}^{GAN} . Thus, the combined uncertainty is $\sigma_{ij}^2 = (\sigma_{ij}^{\text{MC}})^2 + (\sigma_{ij}^{\text{GAN}})^2$. As shown in Fig. IV.17, correlations are well reproduced and compatible with originating from statistical fluctuations as $\chi \approx 1.1$. While most of the individual $\sqrt{\chi_{ij}^2}$ are below 3, a few exceptions might be noted: the correlations between $p_{T_{\gamma\gamma}}$ and $m_{\gamma\gamma}$ are at the level of $\sqrt{\chi_{p_{T_{\gamma\gamma}} m_{\gamma\gamma}}^2} \approx 5.8$, denoting some degree of imperfection in the GAN generation.

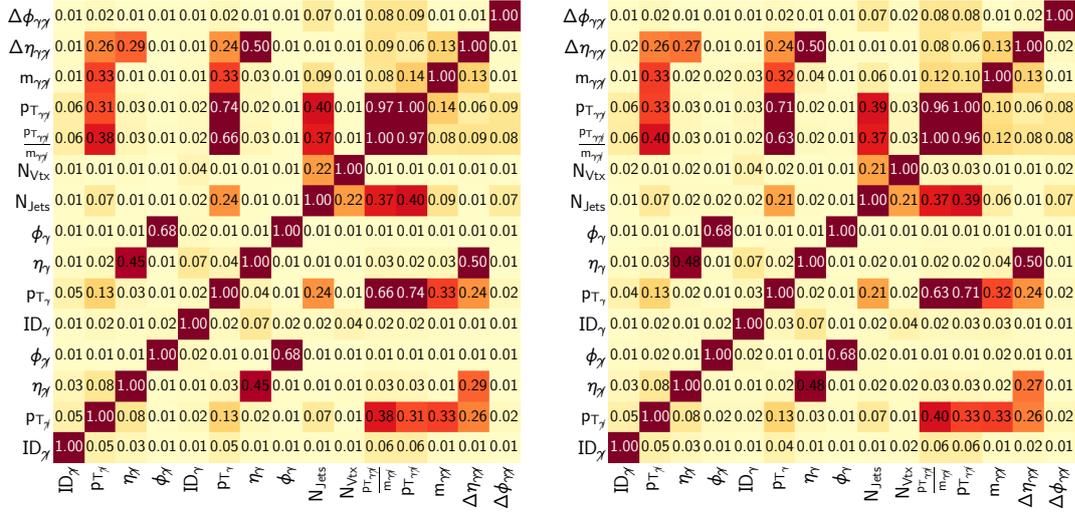


Figure IV.16 – Correlation matrices computed using the distance correlation on the MC SR (left) and on sample produced from the GAN application to the MC CR (right).

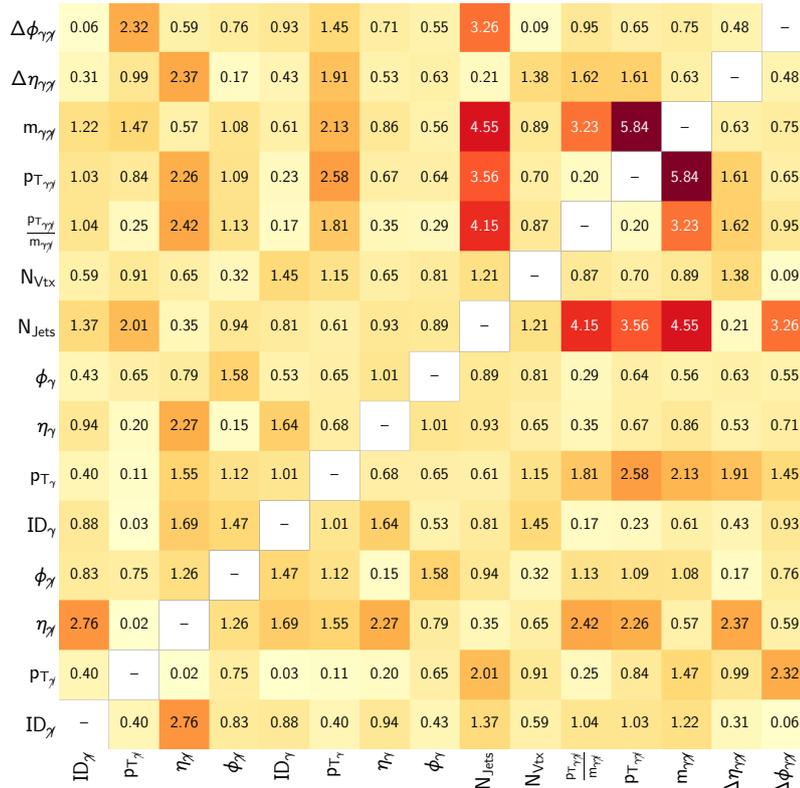


Figure IV.17 – Values of the $\sqrt{\chi_{ij}^2}$ estimating the agreement between generated and target correlation for each pair of features.

IV.5 Application to data

At collider experiments, the strategy of an analysis is usually built from MC simulations. By applying the analysis strategy to the simulated physical processes, we get an estimation of the sensitivity of the analysis. Then, the strategy can be tuned to improve the expected sensitivity of the analysis. In this regard, a good agreement between simulated processes and data is crucial.

A key step of the analysis strategy, after the reconstruction and selection of physical objects, is rejecting background events to base the measurement of physics observable on signal-enriched events. Thanks to their classification performance on problems with high dimensionality, machine learning techniques are one of the standard procedures for signal extraction/background rejection in collider analyses. It is the case in $H \rightarrow \gamma\gamma$ analyses where a BDT (the diphoton BDT) has been implemented to identify events where the diphoton pair is produced by the decay of a Higgs boson and not by any other SM process. The diphoton BDT score is computed using as inputs: kinematic variables of both photons, photonID BDT score, the resolution estimates on the diphoton mass and an estimation of the probability that the diphoton system is associated with the correct vertex. The output of the diphoton BDT is evaluated in Fig. IV.18 on different background (red) and signal (blue) processes, as well as on data (black dots).

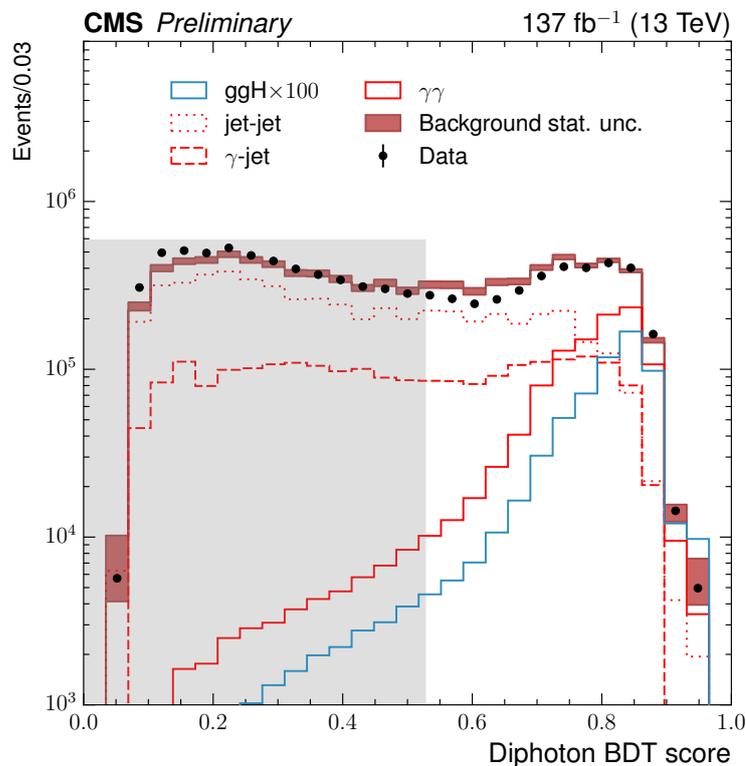


Figure IV.18 – Evaluation of the diphoton BDT on simulated samples of background (red) and signal (blue) processes, as well as on data (black dots). The figure comes from Ref. [145], and the shaded region represents the scores of events considered as background in the analysis.

To determine the level of improvements brought by the GAN, we propose to apply it to a CR from the data recorded by CMS in 2017 to generate an estimation of the $\gamma + \text{jets}$ process. This GANed $\gamma + \text{jets}$ sample is used to replace the MC simulation of the process in the training of a BDT with a role akin to the diphoton BDT, *i.e.* to single out events with a Higgs boson decaying to two photons. A second $\gamma + \text{jets}$ sample is produced by replacing the ID of the misidentified photon (photon with the lowest score) in the same CR of data with a generated ID following the estimated 1D PDF of the ID of the misidentified photon in the SR of the MC sample as described in Ref. [25]. This second sample is, in turn, used to train a diphoton BDT. The performance of the GANed sample is assessed from the improvement of the BDT to reject background. To get a fair comparison of the discrimination power of both networks, the background sample used for the final evaluation is taken from the SR using only the sidebands of data, defined as the region where $m_{\gamma\gamma} \notin [115 \text{ GeV}, 135 \text{ GeV}]$. The data sidebands exclude most of the Higgs-boson signal, ensuring that we evaluate networks on mostly $\gamma\gamma + \text{jets}$ and $\gamma + \text{jets}$ background.

IV.5.1 Generating a $\gamma + \text{jets}$ sample with the GAN

To investigate a concrete application of the GAN to a $H \rightarrow \gamma\gamma$ analysis, the GAN is trained with a $\gamma + \text{jets}$ sample generated with MADGRAPH with an additional filter enriching the sample with events with two photons. The SR is defined as the region where events have one photon with an ID score greater than 0 and the other photon with an ID score above -0.8 and the CR has a requirement reversed for the second photon (below -0.8). Standard selection criteria used in $H \rightarrow \gamma\gamma$ analyses are applied. A summary of all these criteria is provided in Table IV.1. Since we do not have information on the prompt or misidentified

SR	CR
$\text{ID}_{\gamma_{\max}} \geq 0.0$	$\text{ID}_{\gamma_{\max}} \geq 0.0$
$\text{ID}_{\gamma_{\min}} \geq -0.8$	$\text{ID}_{\gamma_{\min}} < -0.8$
$100 \text{ GeV} \leq m_{\gamma\gamma} \leq 180 \text{ GeV}$ $ \eta_{\gamma} \leq 2.5$ $p_{T\gamma_1} \geq 35 \text{ GeV}$ and $p_{T\gamma_2} \geq 25 \text{ GeV}$ $p_{T\gamma_1} \geq m_{\gamma\gamma}/3$ and $p_{T\gamma_2} \geq m_{\gamma\gamma}/4$	

Table IV.1 – Definition of the SR and CR to reproduce a $H \rightarrow \gamma\gamma$ analysis setting.

nature of photons in data, they are ordered by transverse momentum with γ_1 (γ_2) being the photon with leading (subleading) transverse momentum, *i.e.* $p_{T\gamma_1} \geq p_{T\gamma_2}$. They can be also ordered by ID score with γ_{\max} (γ_{\min}) being the photon with max (min) ID, *i.e.* $\text{ID}_{\gamma_{\max}} \geq \text{ID}_{\gamma_{\min}}$. These two sorting systems are independent, meaning that γ_{\max} does not necessarily have leading p_T and vice versa. The GAN generator is used to replace photons with the lowest ID score in events of the CR, and the features of the diphoton system are recomputed accordingly. Since the GANed photon is not guaranteed to have an ID below the ID of the unchanged

photon, photons are reordered after the generation, making sure that $p_{T\gamma_1} \geq p_{T\gamma_2}$ and $ID_{\gamma_{\max}} \geq ID_{\gamma_{\min}}$.

A reweighting of the events is needed to correct for kinematic differences between the CR and the SR. As demonstrated in Section IV.4, we planned to reweight the events based on the kinematic properties of the non-generated photon, given that the generated photons were shown to match the properties of the SR. However, we found that a non-negligible amount of MJ events were populating the CR in the data. Unfortunately, the current available MJ simulated samples by CMS do not allow the training of the GAN on a mixture of $\gamma + \text{jets}$ and MJ events. Therefore our final reweighting was performed in two dimensions along the transverse momentum of both photons. Contrarily to the CR, a major contribution to the background in the SR is coming from $\gamma\gamma + \text{jets}$ events. Before reweighting the generated $\gamma + \text{jets}$ sample to the SR in data, a careful treatment of the $\gamma\gamma + \text{jets}$ event is necessary. The proportion of $\gamma\gamma + \text{jets}$ and $\gamma + \text{jets}$ in the SR of data is determined by a maximum likelihood fit between a mixture of the MC simulation for these processes and the data sidebands. This fit is performed in 2D over the $ID_{\gamma_{\max}}$ and the $ID_{\gamma_{\min}}$. Once the proportion of $\gamma\gamma + \text{jets}$ events in the SR is measured, it can be subtracted from the data distribution in the $(p_{T\gamma_{\max}}, p_{T\gamma_{\min}})$ plane used for the reweighting. Examples of distributions before and after fit plus reweighting are shown in Fig. IV.19.

IV.5.2 Generating a $\gamma + \text{jets}$ sample with a 1D PDF

The generation of the estimated $\gamma + \text{jets}$ sample with a 1D PDF is similar to the generation from the GAN. The definitions of the SR, the CR, and the selection criteria are kept identical, and the method is taken from Ref. [25]. Before generating a photon ID, the PDF of the misidentified photon ID score is derived from the SR of the MC simulation of the $\gamma + \text{jets}$ sample (see Fig. IV.20). For all events, a value is randomly drawn with a probability following this PDF to replace the photon ID of the photon with the lowest ID. This new ID is generated in the range $[-0.8, ID_{\gamma_{\max}}]$ to ensure that $ID_{\gamma_{\min}} \leq ID_{\gamma_{\max}}$. In addition, a per-event weight w_{ID} is computed to correct the $ID_{\gamma_{\max}}$ distribution from its shape in the CR to its expected shape in the SR:

$$w_{ID} = \frac{\int_{-0.8}^{ID_{\gamma_{\max}}} \text{PDF}}{\int_{-1}^{-0.8} \text{PDF}}.$$

In this case, no need to reorder the photons in terms of ID or p_T , so the fit plus reweighting is directly performed, following the procedure described in Section IV.5.1. Examples of distributions before and after fit plus reweighting are shown in Fig. IV.21.

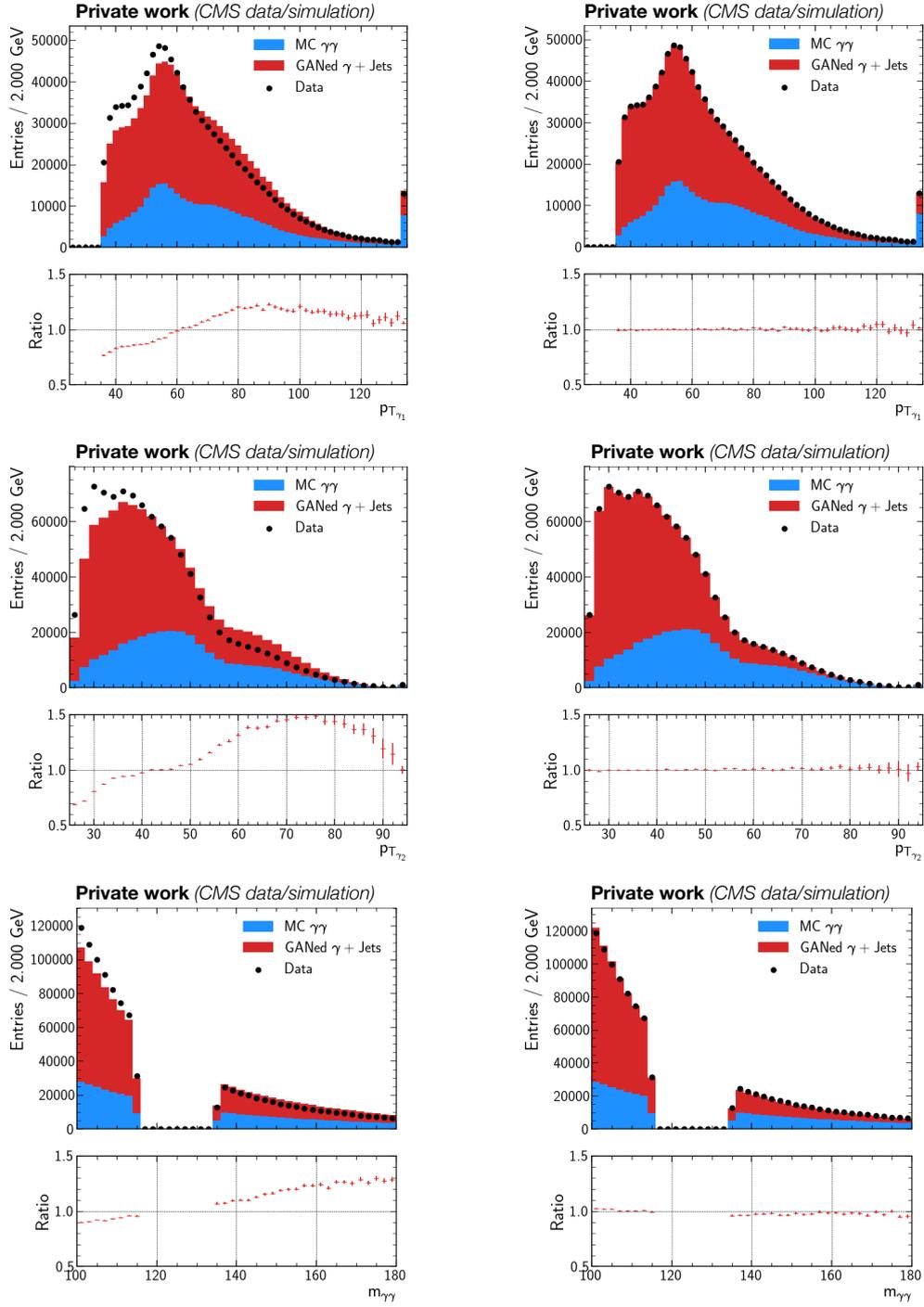


Figure IV.19 – Distribution of the $p_{T\gamma_1}$, $p_{T\gamma_2}$, and $m_{\gamma\gamma}$ using the GANed sample as estimation of the $\gamma + \text{jets}$ sample before (left column) and after (right column) fit plus reweighting to the data sidebands.

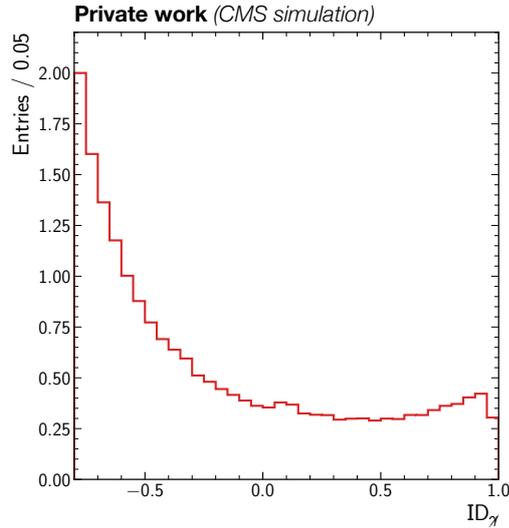


Figure IV.20 – Probability density function of the ID_γ evaluated in the SR from the γ +jets simulated sample.

IV.5.3 Comparison of the two methods

Two BDTs are trained to mimic the role of the diphoton BDT and to compare the quality of the two generated samples. Both BDTs are trained using MC samples of processes producing a Higgs boson (signal sample), and the background sample is the γ + jets sample generated with a PDF for one BDT (BDT_{PDF}) while the other (BDT_{GAN}) takes the γ +jets sample generated with the GAN as background.

As inputs to the BDTs, the kinematic features of the photons as well as their ID are used. The output represents the probability that the diphoton pair originates from a Higgs boson. While training, they are also evaluated on an independent validation sample to check for any overtraining and the training is stopped when the validation loss starts increasing. Evaluation of the two BDTs on test samples is shown in Fig. IV.22. To have a fair comparison of the background rejection capabilities of the two trainings, both are evaluated on the same sample from the data sidebands of the SR. These data sidebands are mainly composed of $\gamma\gamma$ + jets or γ + jets events. To evaluate the performance of a BDT on γ + jets events of the data sidebands, the BDT score distribution is computed from an MC $\gamma\gamma$ + jets sample and the expected fraction of $\gamma\gamma$ + jets events (computed from the fits of Sections IV.5.1 and IV.5.2) is subtracted from the data sidebands distribution.

The figure of merit used to quantify the discrimination capabilities of a classification technique is the receiver operating characteristics (ROC) curve. A decision on the type of an event (signal or background) can be made by setting a threshold on the BDT output. Above the threshold, events are considered as signal, and below, they are rejected as background. Therefore, we define the rate of true positive results as the number of signal events above the threshold divided by the total number of evaluated signal events and the rate of false positive results as the number of background events above the threshold divided by the total number of evaluated background events. The true positive and false positive

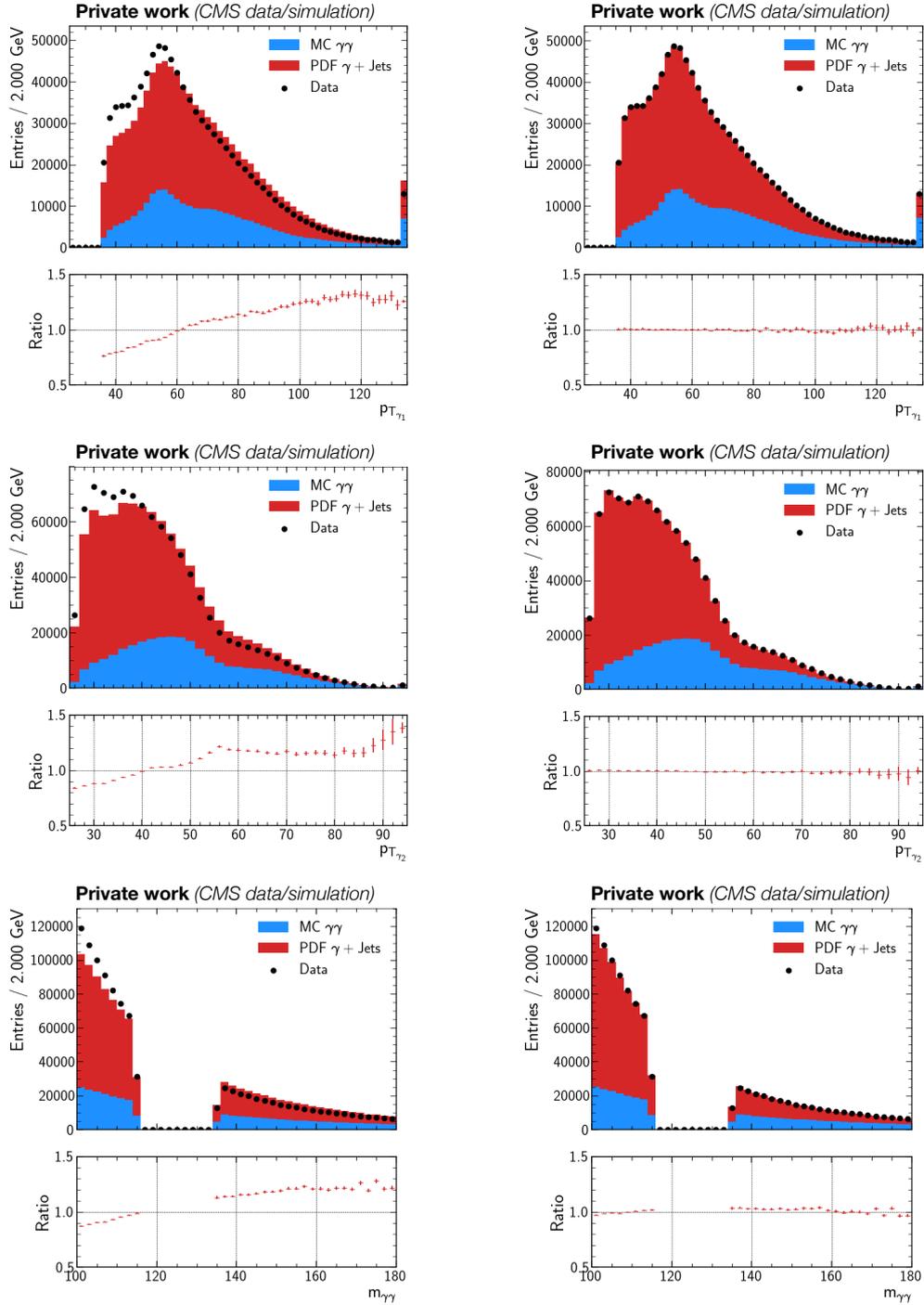


Figure IV.21 – Distribution of the p_{T,γ_1} , p_{T,γ_2} , and $m_{\gamma\gamma}$ using the sample where $ID_{\gamma_{\min}}$ is generated from a 1D PDF as estimation of the $\gamma + \text{jets}$ sample before (left column) and after (right column) fit plus reweighting to the data sidebands.

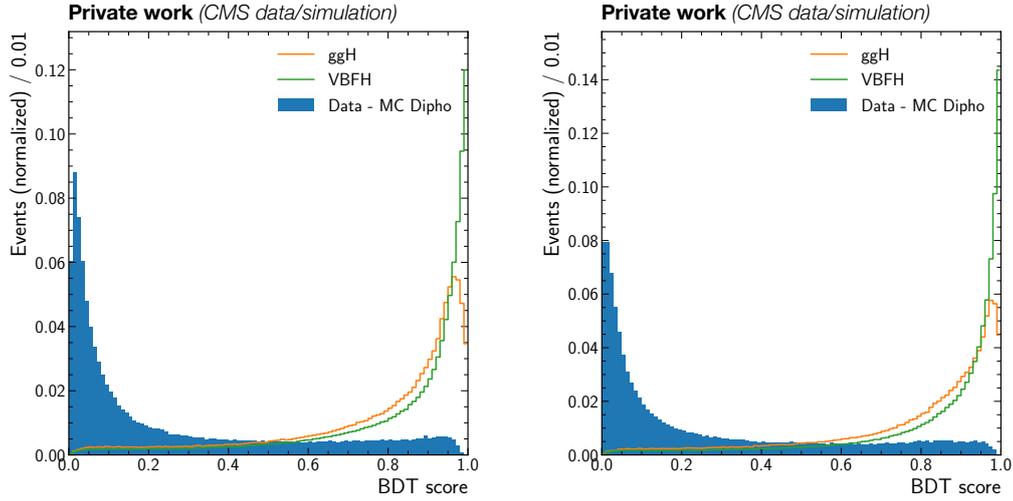


Figure IV.22 – Evaluation of the output of BDT_{PDF} (left) and BDT_{GAN} (right) on the background (data - MC dipho) and signal (ggH, VBFH) samples.

rates are also called signal efficiency and background efficiency, respectively. By scanning the threshold over the full BDT output range, we extract points in the (signal efficiency, background efficiency) plane forming the ROC curve. The area under the ROC curve (AUC) gives meaningful information on the discrimination power of the classification method as a perfect classifier would get an AUC of 1 and a random classifier an AUC of 0.5.

The ROC curves of the two BDTs of interest are compared in Fig. IV.23. As the highest AUC is obtained with BDT_{GAN} , it means that the sample generated with the GAN is improving the background rejection possibilities. A comparison of the impact on signal and background efficiencies for fixed working points is given through the ratio plots of Fig. IV.23. The ratio of signal efficiencies obtained from the two trainings (ratio sig) shows that BDT_{GAN} based selection offers a higher signal efficiency for a given background efficiency. In the same way, ratio bkg demonstrates that BDT_{GAN} gives a lower background efficiency at fixed signal efficiency levels. While their performance is similar at very high signal and background efficiency, BDT_{GAN} reduces background down to 88% of what could be achieved with BDT_{PDF} depending on the targeted signal efficiency level. By using S/\sqrt{B} , where S and B are, respectively, the signal and background yields, as an estimation of the sensitivity of a region in the phase space, reducing the level of background by 12% for the same signal efficiency means improving the sensitivity of this region by $\approx 7\%$.

To show the difference when applying selection requirements on the score of BDT_{PDF} or BDT_{GAN} , we report in Table IV.2 the number of events selected from each sample when targeting 80% signal efficiency. A zoom around this efficiency level is provided in Fig. IV.23. As expected, the selection from the output of both BDTs reduces mainly $\gamma + \text{jets}$ events (equivalent to Data - ($\gamma\gamma + \text{jets}$) events in Table IV.2); they are the main source of background when no cut is applied and $\gamma\gamma + \text{jets}$ events become dominant after the selection. However, at the same signal efficiency level, the selection on BDT_{GAN} rejects more background from both $\gamma\gamma + \text{jets}$ and $\gamma + \text{jets}$ events.

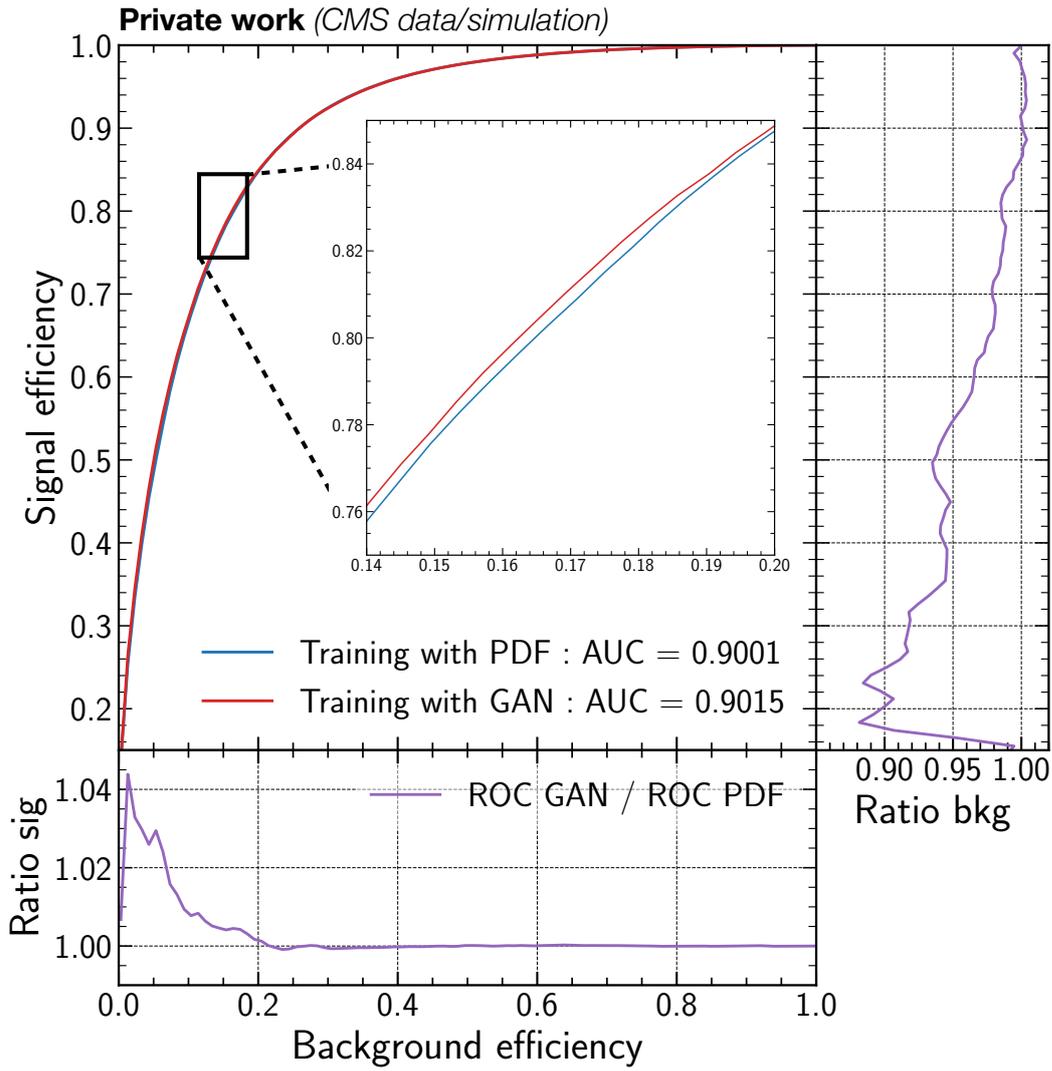


Figure IV.23 – Comparison of the ROC curve of BDT_{PDF} (blue) and BDT_{GAN} (red). Ratio sig is the ratio of signal efficiencies of both ROC curves at a given value of background efficiency. And vice versa, ratio bkg is the ratio of background efficiencies of both ROC curves at a given value of signal efficiency.

	Signal	$\gamma\gamma + \text{jets}$	Data – ($\gamma\gamma + \text{jets}$)
No cut	56	227 220	739 674
Cut BDT_{PDF} ($\epsilon_{\text{sig}} = 0.8$)	45 (80%)	170 534 (75%)	121 909 (17%)
Cut BDT_{GAN} ($\epsilon_{\text{sig}} = 0.8$)	45 (80%)	168 259 (74%)	118 756 (16%)

Table IV.2 – Estimated yield of each sample when applying cut on the score of BDT_{PDF} and BDT_{GAN} to target a signal efficiency (ϵ_{sig}) of 80%.

IV.6 Conclusion

In this chapter, I have presented a new data-driven technique to create background samples for HEP background processes with one misidentified object. A CR is defined by requiring an object to fail ID criteria. A generated misidentified object replaces this object to simulate an event in the SR. The technique is based on conditional generative adversarial networks (GANs), which are known to be challenging to train. To assess the generator performance, we developed a figure of merit based on a negative log-likelihood. Due to random fluctuations intrinsic to the latent space used in GANs, we introduced a multiple-sampling method to obtain more consistent results in the model performance evaluation.

The application and the performance of the technique are demonstrated on the $\gamma + \text{jets}$ background in the context of the $H \rightarrow \gamma\gamma$ analysis at the LHC. I have shown that the conditional GAN-based technique produces object observables that have excellent agreement with the signal-like object observables and non-linear correlations of these observables within themselves and with the properties of the rest of the event. Therefore, the samples generated by this technique can be used to improve the background description in the $H \rightarrow \gamma\gamma$ analysis. This is especially true in SR with specific constraints where the MC simulation might be suboptimal due to its low number of events and its inaccurate description of the $\gamma + \text{jets}$ background. This work is published as Ref. [141].

Finally, to quantify the improvements brought by the GAN in the estimation of the $\gamma + \text{jets}$ background, a BDT was trained to reject background thanks to the GANed sample. This BDT was compared to a second one, this time trained with a $\gamma + \text{jets}$ sample generated with the method of the ID_γ probability distribution function. Their receiver operating curve is drawn to confront their classification performance on data sidebands. From the area under the ROC curves, a final score arbitrates that the GAN offers better discrimination possibilities to improve the absolute sensitivity of the analysis.

It has to be noted that although this study was carried out in the context of the $H \rightarrow \gamma\gamma$ analysis, this method could be trained to generate other objects and help other analyses dealing with backgrounds coming from misidentified objects.

Constraints on the Anomalous Couplings of the Higgs Boson in the VH Production Mode and Diphoton Final State

V.1	Physics motivations and analysis strategy	114
V.1.1	Modeling couplings of the Higgs boson with electroweak gauge bosons.	114
V.1.2	Strategy	116
V.2	Trigger and objects definition	116
V.3	Highlighting $H \rightarrow \gamma\gamma$ events	121
V.3.1	Collected data and MC simulations	121
V.3.2	Reweighting of JHU samples	122
V.3.3	Phenomenology of $V_{\text{had}}H$ events and preselection.	124
V.4	Machine learning for events discrimination.	129
V.5	Optimisation of analysis categories	132
V.6	Statistical inference and results	136
V.6.1	Inference procedure	136
V.6.2	Background modelling	139
V.6.3	Signal modelling	140
V.6.4	Systematic uncertainties	142
V.6.5	Results.	152
V.7	Conclusion	155

V.1 Physics motivations and analysis strategy

This chapter lays out my involvement in the analysis of the H boson couplings with electroweak gauge bosons (W and Z bosons) using the H boson decay channel into a pair of photons. Ten years after the discovery of the H boson, many analyses aim at precise measurements of its properties thanks to the increasing amount of data collected by the CMS and ATLAS experiments. This is a way of checking the consistency of the H boson properties with the predictions of the standard model (SM). Any deviation from these predictions could hint at new physics effects that would require theories going beyond the standard model (BSM). Since the SM does not explain several fundamental questions (as mentioned already in Chapter I), major efforts at the LHC are carried out to identify which sectors of the SM could open the door to new physics.

The SM predicts the H boson to have spin-parity quantum numbers $J^{PC} = 0^{++}$, *i.e.* a spin 0 and a CP-even behaviour. Constraints consistent with the SM predictions have already been put by the CMS [22, 25, 27, 154] and ATLAS [155–158] experiments on the spin-parity nature of the H boson. However, these constraints are leaving room for small BSM couplings and some of them would indicate a potential CP-odd nature for the H boson. These small couplings are not expected in the SM, and in that sense, are regarded as *anomalous couplings*. These anomalous couplings were studied in different decay channels ($H \rightarrow ZZ$ [27] or $H \rightarrow \tau\tau$ [154]) and in the $H \rightarrow \gamma\gamma$ channel but in the particular case of the $t\bar{t}H$ production mode [25]. This chapter presents a complementary way of probing the CP nature of the H boson by analysing its production in the VBF and VH ($V = Z$ or W) modes and in the $H \rightarrow \gamma\gamma$ channel. Preliminary studies showed that, given a large enough data set, this analysis is sensitive to the couplings of the H boson with electroweak gauge bosons through the VVH coupling vertex in the vector boson fusion (VBF) and VH productions [143]. The analysis is shared between several CMS analysis teams with the VBF production analysed by the team of INFN Rome, the VH production where V decays to leptons ($V_{\text{lep}}H$) and V decays to neutrinos ($V_{\text{MET}}H$) is treated by teams of the University of Minnesota, the John Hopkins University and the National Taiwan University, while I am responsible of the VH production where V decays hadronically ($V_{\text{had}}H$). We analyse data collected by the CMS experiment during the three years (2016, 2017, 2018) of the second run of the LHC, corresponding to an integrated luminosity of 137.6 fb^{-1} .

V.1.1 Modeling couplings of the Higgs boson with electroweak gauge bosons

We rely on an effective description of the interaction between a spin-zero H boson and two spin-one gauge bosons. This is achieved by defining the scattering amplitude of this interaction $A(HVV)$ in its most general form and by measuring or constraining the parameters manifesting the intensity of each coupling term. It allows for a model-independent study of the HVV vertex, and specific theories can confront their model to the measured couplings. By taking the same

notation as previous CMS analyses [22], the scattering amplitude takes the form

$$A(\text{HVV}) \sim \left[a_1^{\text{VV}} + \frac{\kappa_1^{\text{VV}} q_1^2 + \kappa_2^{\text{VV}} q_2^2}{(\Lambda_1^{\text{VV}})^2} + \frac{\kappa_3^{\text{VV}} (q_1 + q_2)^2}{(\Lambda_Q^{\text{VV}})^2} \right] m_{V_1}^2 \epsilon_{V_1}^* \epsilon_{V_2}^* + a_2^{\text{VV}} f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + a_3^{\text{VV}} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu}, \quad (\text{V.1})$$

where m_{V_i} , q_i and ϵ_i are the mass, momentum and polarisation vector of vector boson i , allowing the definition of its field strength tensor $f^{(i),\mu\nu}$ and its dual $\tilde{f}^{(i),\mu\nu}$ as $f^{(i),\mu\nu} = \epsilon_i^\mu q_i^\nu - \epsilon_i^\nu q_i^\mu$ and $\tilde{f}^{(i),\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}^{(i)}$. The a_i are the coupling constants modifying the contribution of each term in the total amplitude, and Λ_1 and Λ_Q correspond to the energy scale of BSM effects. The VV pair stands for ZZ, WW, gg, $\gamma\gamma$ or $Z\gamma$ but through the VBF and VH production mode, we are only sensitive to the impact of the couplings of the H boson with ZZ, WW or $Z\gamma$ pairs. Some of these couplings are expected to have nonzero values by the SM. In particular, the a_1^{VV} parameter corresponds to the contribution from the tree diagram of a H boson with a pair of gauge bosons, so in the SM, $a_1^{\text{ZZ}} = a_1^{\text{WW}} = 1$ while $\kappa_1, \kappa_2, a_2, a_3 = 0$. The a_2^{VV} , $\kappa_i^{\text{VV}}/(\Lambda_1^{\text{VV}})^2$ and $\kappa_3^{\text{VV}}/(\Lambda_Q^{\text{VV}})^2$ couplings are part of CP-even terms. They may get loop-induced contribution from the SM at the level of ($\mathcal{O}(10^{-2}-10^{-3})$), not accessible with the current data. Tight constraints have already been set on Λ_Q^{VV} [159], so this parameter is not considered in the analysis. The a_3^{VV} term would indicate a pseudoscalar H boson (*i.e.* with CP-odd behaviour), so it is also considered as an anomalous coupling. Because of symmetry and gauge invariance arguments, constraints can be put on the following couplings $\kappa_1^{\text{ZZ}} = \kappa_2^{\text{ZZ}} = -\exp(i\phi_{\Lambda_1}^{\text{ZZ}})$, $\kappa_1^{\text{Z}\gamma} = 0$ and $\kappa_2^{\text{Z}\gamma} = -\exp(i\phi_{\Lambda_1}^{\text{Z}\gamma})$. We assume the following symmetry for all parameters $x^{\text{ZZ}} = x^{\text{WW}}$ where x is a_i , κ_i or Λ_1 , so the superscripts are dropped if they concern parameters of HZZ or HWW processes.

To represent the physical effects of these anomalous couplings, it is relevant to define their associated cross section fraction f_x and phases ϕ_x . By defining the total cross section as $\sigma^{\text{tot}} = |a_1|^2 \sigma_1 + |a_2|^2 \sigma_2 + |a_3|^2 \sigma_3 + \tilde{\sigma}_{\Lambda_1}/(\Lambda_1)^4 + \tilde{\sigma}_{\Lambda_1}^{\text{Z}\gamma}/(\Lambda_1^{\text{Z}\gamma})^4$ where $\tilde{\sigma}_{\Lambda_1}$ ($\tilde{\sigma}_{\Lambda_1}^{\text{Z}\gamma}$) is the effective cross section for the process with Λ_1 ($\Lambda_1^{\text{Z}\gamma}$) = 1 TeV and thus in units of fb TeV⁴, we have the final list of observables of interest for this analysis:

$$\begin{aligned} f_{a_2} &= \frac{|a_2|^2 \sigma_2}{\sigma^{\text{tot}}}, & \phi_{a_2} &= \arg\left(\frac{a_2}{a_1}\right), \\ f_{a_3} &= \frac{|a_3|^2 \sigma_3}{\sigma^{\text{tot}}}, & \phi_{a_3} &= \arg\left(\frac{a_3}{a_1}\right), \\ f_{\Lambda_1} &= \frac{\tilde{\sigma}_{\Lambda_1}/(\Lambda_1)^4}{\sigma^{\text{tot}}}, & \phi_{\Lambda_1} &= \arg(\kappa_1), \\ f_{\Lambda_1^{\text{Z}\gamma}} &= \frac{\tilde{\sigma}_{\Lambda_1}^{\text{Z}\gamma}/(\Lambda_1^{\text{Z}\gamma})^4}{\sigma^{\text{tot}}}, & \phi_{\Lambda_1^{\text{Z}\gamma}} &= \arg(\kappa_1^{\text{Z}\gamma}). \end{aligned}$$

As it is a ratio of cross sections, many systematic uncertainties are cancelling for these parameters, and they act as mixture coefficients between the different BSM hypotheses, so their range is bounded between 0 and 1.

V.1.2 Strategy

The outline of the analysis strategy is similar to what was proposed in previous $H \rightarrow \gamma\gamma$ analyses [145]:

- **Preselection:** After the physics objects reconstruction and correction, a preselection is applied based on the kinematic features of the event to enhance the signal purity.
- **Categorisation:** First, a machine-learning-based algorithm is trained to discriminate Higgs boson signals (SM and BSM) from SM backgrounds and SM Higgs boson production from BSM Higgs boson production. Based on the score of these classifiers, we define categories targeting regions of the phase space where we expect different behaviour from the different physics model hypotheses. It allows the final statistical analysis to probe the compatibility between hypotheses.
- **Statistical analysis:** Models for the signal and background distributions of the diphoton mass $m_{\gamma\gamma}$ are defined in each category. These models are fitted to data, and a log-likelihood ratio is defined as a statistical test to put constraints on a parameter of interest (like f_{a3}).

In this chapter, I will focus on my contribution to the design of the $V_{\text{had}}H$ part. After the physics objects reconstruction and correction (summarised in Section V.2), we study the specific behaviour of the $V_{\text{had}}H$ -event properties. Section V.3.3 compares the kinematics of the different production modes to clarify the preselection requirements. To enhance the discrimination power, a deep neural network (DNN) is trained from kinematics features to classify events between three classes: VH SM signal, VH BSM signal and background described in Section V.4. Since the outputs of the DNN provide a greater separation compared to kinematics variables only, categories are built from these outputs to create regions in phase space enriched either with SM or BSM signal. The optimisation of the categories is described in Section V.5 and is done by computing an estimate of the upper limit on f_{a3} at a 95% CL. Finally, the expected results are extracted using common statistical tools developed within CMS collaboration and are presented in Section V.6.5.

V.2 Trigger and objects definition

Triggers

As described in Chapter II, because of constraints on the data collection throughput and available storage, event information is stored for offline analyses only if they pass some trigger requirements. These requirements target specific signatures to enrich the recorded data with potentially interesting events in terms of physics. Two trigger levels must be satisfied based on the hardware response: the first-level trigger (L1 trigger) and the high-level trigger (HLT). The trigger requirements for $H \rightarrow \gamma\gamma$ analyses are detailed in this section.

The L1 trigger expects at least one electromagnetic candidate in the ECAL. These candidates are built by aggregating the energy deposits of adjacent crystals of the ECAL, thus forming relevant clusters of crystals. If two candidates are detected, then the requirements on the E_T of the candidates are lowered to 25 GeV and 14 GeV (23 GeV and 10 GeV) for 2017-2018 (2016) data.

From the clusters passing the L1 requirements, the HLT system builds further requirements to ensure the relevance of candidate events containing electrons or photons (e/γ). A version of the CMS-offline-reconstruction algorithm whittled down to its elementary steps is applied to the electromagnetic clusters to form e/γ candidates. With these steps, e/γ are defined through kinematic, isolation and shower shape variables:

- E_T : transverse energy of the photon;
- H/E : ratio between the energy collected in the HCAL (H) and energy deposited in the ECAL (E);
- R_9 : fraction of the supercluster total energy contained in a 3×3 crystal surrounding the supercluster seed;
- $\sigma_{i\eta i\eta}$: lateral extension of the electromagnetic shower in terms of crystal;
- $\text{Iso}_{\text{ph}}^{\text{HLT}}$: sum of the transverse energy of other particles identified as photon by the particle-flow algorithm [52] and falling in a cone with $R = 0.3$ around the photon candidate direction;
- $\text{Iso}_{\text{track}}^{\text{HLT}}$: sum of the transverse energy of tracks falling in a cone with $R = 0.3$ around the photon candidate direction but not falling in the inner cone with $R = 0.04$;

The set of HLT requirements dedicated to photon candidates in the $H \rightarrow \gamma\gamma$ analysis is summarised in Table V.1 and is applied on the single or pair of clusters of the event that passed the L1 criteria.

EB		EE	
$E_T > 30 \text{ GeV}$		$E_T > 30 \text{ GeV}$	
$ \eta < 2.5$		$ \eta < 2.5$	
$H/E < 0.12$		$H/E < 0.1$	
$R_9 > 0.5$		$R_9 > 0.8$	
$R_9 > 0.85$	$R_9 \leq 0.85$	$R_9 > 0.9$	$R_9 \leq 0.9$
—	$\sigma_{i\eta i\eta} < 0.015$	—	$\sigma_{i\eta i\eta} < 0.035$
—	$\text{Iso}_{\text{ph}}^{\text{HLT}} < 6.0 + 0.012 E_T$	—	$\text{Iso}_{\text{ph}}^{\text{HLT}} < 6.0 + 0.012 E_T$

Table V.1 – Requirements of the HLT to collect an event in the $H \rightarrow \gamma\gamma$ path.

Finally, if at least one of the clusters passes the L1 and HLT selections, then the entire ECAL is clustered, and a pair of γ is required (where at least one is the L1 candidate). A new set of selection criteria is applied to all the electromagnetic objects reconstructed as clusters in the ECAL detector. They are the same as

Table V.1 with the following substitution and additions : E_T of the photons above 22 GeV (18 GeV) in 2017-2018 (2016) data, $\text{Iso}_{\text{track}}^{\text{HLT}} < 6.0 + 0.002 E_T$, and $m_{\gamma\gamma} > 90$ GeV.

The performance of the triggers is studied by measuring the efficiency of the trigger selection using the tag and probe method on $Z \rightarrow ee$ events. Yet, because of the material located in front of the ECAL, electrons have different shower shapes compared to photons. Due to bremsstrahlung emission along the electron track path, the electron showers tend to be broader. The η distribution of photons coming from the Higgs boson decay also differs significantly from the one of electrons arising in the Z boson decay. To correct these effects, events are reweighted in 2D with $R_9 - |\eta|$ weight matrices computed from $H \rightarrow \gamma\gamma$ and Drell-Yan $Z \rightarrow ee$ simulated samples. Finally, the computed efficiencies for each year are binned in p_T , R_9 , and $|\eta|$ and are used to correct simulated samples. This correction reproduces the effects of the trigger on the expected background and signal yields in data.

Photons

As mentioned already in the description of the trigger, photons are reconstructed from clusters of energy deposits in the ECAL. Photons leave all their energy in the form of an electromagnetic shower in the ECAL, which can spread over multiple crystals. In some occurrences, interactions happen with the material upstream from the ECAL, and photons can convert into an electron-positron pair. Since the ECAL is within the high magnetic field of the CMS detector, electrons and positrons have bent trajectories that lead to energy deposits spread over an even larger number of crystals. In order to recover all possible information, these clusters are grouped into larger ensembles called superclusters, representing the complete shower stemming from a prompt photon, converted or not. Information on the pseudorapidity and azimuthal angle of the reconstructed photons is extracted from the supercluster position. However, the energy of the photon candidate is inferred from a multivariate approach to compensate for the energy loss due to upstream material or due to imperfections in the cluster aggregations. This method allows a more accurate measurement of the photon energy and also gives an estimation of the resolution of the energy measurement. By using $Z \rightarrow ee$ events, a correction on the energy scale of the regressed energy can be determined from the difference of the Z peak in MC simulations and in data. After applying the scale correction, a correction on the resolution is computed by applying Gaussian smearing centred around the corrected peak to MC simulations and using a maximum likelihood fit to data to find the correct modification needed on the energy resolution.

A description of the photon ID was already given in Section IV.2 of Chapter IV, particularly how it is derived from shower shape and isolation variables. Each reconstructed photon is given a score (the photon ID score) representing its likelihood to be a prompt photon, as opposed to a misidentified photon originating from the decay of neutral mesons. It means that discrepancies in the photon ID distributions between MC samples and data indicate that part of these shower shape and isolation variables are impacted by imperfect modelling of the CMS

detector simulation. To mitigate these effects, corrections can be applied to MC simulations. They are computed from a procedure based on chained quantile regression [145] that aim at shifting the input distributions of the photon ID BDT in the simulation until their cumulative distribution function matches the ones of data. Again, these corrections are derived from $Z \rightarrow ee$ events. It allows the recovery of shower shape and isolation distributions as observed in data, and it corrects the bias on the photon efficiency.

Then, from the reconstructed and corrected photons of the events, diphoton pairs are formed. In addition to the requirements of the triggers on photons, selection criteria are applied on the p_T of the photons to recover a smoothly falling background shape in the diphoton mass distribution: $p_{T\gamma_1}/m_{\gamma\gamma} > 1/3$ and $p_{T\gamma_2}/m_{\gamma\gamma} > 1/4$.

Vertices

To reach the highest precision in reconstructing the Higgs boson properties, all information on the diphoton system needs to be precisely reconstructed. In particular, any shift in the reconstructed vertex position is detrimental as it degrades the mass resolution of the diphoton system and would squander the excellent energy resolution of the ECAL. A first set of primary vertices is reconstructed for each event through a standard CMS algorithm [160]. Based on this set of reconstructed vertices, a vertex identification algorithm was developed as a common remedy for $H \rightarrow \gamma\gamma$ analyses [145].

Because of additional collisions occurring at the same time, known as pile-up collisions (PU), several primary vertices are reconstructed for a given event with a common primary vertex reconstruction algorithm within the CMS collaboration. Thus, a first BDT is trained to give a vertex-based score (the vertex identification score $vtxID$), indicating its likelihood to be the primary vertex associated with the diphoton system. The $vtxID$ BDT takes as input variables related to the diphoton topology in the detector as well as variables of the other tracks associated with the vertex. A signal sample is built from $H \rightarrow \gamma\gamma$ MC samples with the primary vertices closest to the true generated one, and the background sample consists of any other primary vertices. After computation of the $vtxID$ of each primary vertex in $H \rightarrow \gamma\gamma$ analyses, the final vertex chosen for a given event is the one with the highest $vtxID$ score.

A second BDT defines the per-event probability that the chosen vertex is the correct one (the vertex probability score $vtxprob$). The impact of the diphoton vertex displacement on the $m_{\gamma\gamma}$ reconstruction is negligible compared to the effects of the ECAL resolution if it is less than 10 mm away from its true position. Therefore, the $vtxprob$ BDT determines the probability that the selected diphoton vertex is within 10 mm of the true vertex, *i.e.* the fraction of events with $|\Delta z| = |z_{\text{selected}} - z_{\text{true}}| < 10$ mm where z_{selected} is the position of the chosen diphoton vertex and z_{true} its true position. It is trained using the same signal and background samples as the $vtxID$ BDT, and it uses as inputs the p_T of the diphoton pair, the number of primary vertices, the three highest $vtxID$, the distance between the chosen vertex and other candidates, and the number of converted

photons.

The performance of the full vertex identification algorithm is validated by using $Z \rightarrow \mu\mu$ events from data and MC samples. As the algorithm is built around photons, *i.e.* with no track information, it can be used on reconstructed muons by withholding their track information and comparing the results of the algorithm with the actual reconstructed vertex of the dimuon system. A reasonable agreement is observed between data and MC simulations (see Fig. V.1), and scale factors are computed to account for the remaining differences. These scale factors are computed as the ratio between the per-event probability computed in data and MC samples for $Z \rightarrow \mu\mu$ events, and they are applied to $H \rightarrow \gamma\gamma$ MC simulations as a function of the diphoton p_T .

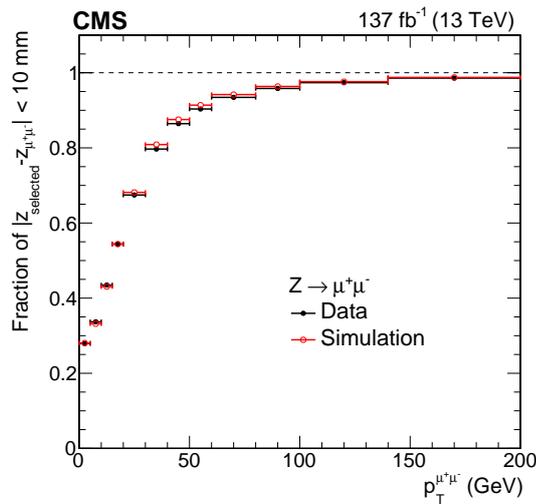


Figure V.1 – Validation of the vertex identification algorithm on $Z \rightarrow \mu\mu$ events from Ref. [145].

Jets

Jets are objects reconstructed from the imprint left in the detector by the shower of particles originating from quarks or gluons. Indeed, they immediately fragment and hadronise, so they cannot be observed directly. The reconstruction of all particles present in these showers is essential as it impacts the inferred properties of the initial parton. Jets are reconstructed from the set of particle candidates clustered by the PF algorithm using the anti- k_T algorithm [161] with a radius parameter set to 0.4. Charged particles are removed from this clustering step if they are not associated with the primary vertex identified with the vtxID. This is to mitigate the contribution from particles produced by PU happening within the same bunch crossing. Corrections to the jet energy are derived with respect to the jet type and its kinematic variables and are applied to jets in data and MC simulations. In addition, measurements show that the energy resolution of the jets in data is worse than in MC. This is taken into account by smearing the jets p_T in MC samples to match the jet energy resolution observed in the data.

V.3 Highlighting $H \rightarrow \gamma\gamma$ events

V.3.1 Collected data and MC simulations

Data samples

This analysis uses the full data set collected during the second data acquisition period of the LHC, spanning three years from 2016 to 2018 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The integrated luminosity collected for each year is as follows: 35.9 fb^{-1} in 2016, 41.5 fb^{-1} in 2017 and 59.8 fb^{-1} in 2018. This corresponds to a total integrated luminosity of 137.2 fb^{-1} . Data samples are blinded during the design of the analysis so as not to bias ourselves by masking events with a diphoton mass close to the mass of the H boson: $m_{\gamma\gamma} \notin [115 \text{ GeV}, 135 \text{ GeV}]$.

Simulated samples

Monte Carlo programs are used to simulate specific physics processes arising in the proton-proton collisions following matrix element calculations with a leading-order (LO) or next-to-leading-order (NLO) description of QCD effects, depending on the generator. The $pp \rightarrow \gamma + \text{jets}$ and $pp \rightarrow \gamma\gamma$ background processes are generated with the MADGRAPH5_AMC@NLO [162, 163] and SHERPA [164] generators respectively. The background samples are used to define the analysis strategy and train potential discriminates, but they are not used to do the final background estimation (as it is done directly from data sidebands). Samples describing processes with anomalous couplings of the H boson are generated with the JHUGEN [165–167] program for different values of the a_2 , a_3 , Λ_1 or $\Lambda_1^{Z\gamma}$ anomalous coupling parameters. The SM productions of the H boson from gluon fusion (ggH), vector boson fusion (VBF), or associated with a Z or W boson (VH), or a $t\bar{t}$ pair ($t\bar{t}H$) are described with the POWHEG [168–170] and MADGRAPH programs, with a NLO generation. MADGRAPH is also used to generate samples used in the signal modelling with alternative mass value hypotheses for the Higgs boson at 120 GeV and 130 GeV. The samples from POWHEG are favoured for the design of the analysis strategy and the training of MVA methods as they have the benefit of providing non-negative weights. It allows the comparison of NLO and LO descriptions of the processes (see Section V.3.2). The NNPDF 3.0 (3.1) [171] parton distribution functions are used for 2016 (2017 and 2018).

The PYTHIA [172] program is interfaced on all samples to include all particle decays, to describe parton showering and hadronisation, and also to overlay additional pileup events following the luminosity profile. All cross sections and branching ratios are taken from the recommendations of the LHC CrossSection-Working-Group [173]. These samples are processed through a GEANT4 [132, 133] description of the CMS detector to simulate detector effects. Finally, corrections described in Section V.2 are applied to correct MC simulations.

V.3.2 Reweighting of JHU samples

The anomalous productions of the H boson are simulated using the dedicated generator JHUGEN at LO QCD. To compare the deviations from an NLO generation, an SM sample is generated with JHUGEN by removing all anomalous contributions. The JHUGEN (LO) and POWHEG (NLO) simulations are explicitly compared after parton showering in the SM case, and discrepancies are found in kinematic observables as demonstrated in Fig. V.2. For instance, a large deviation is found

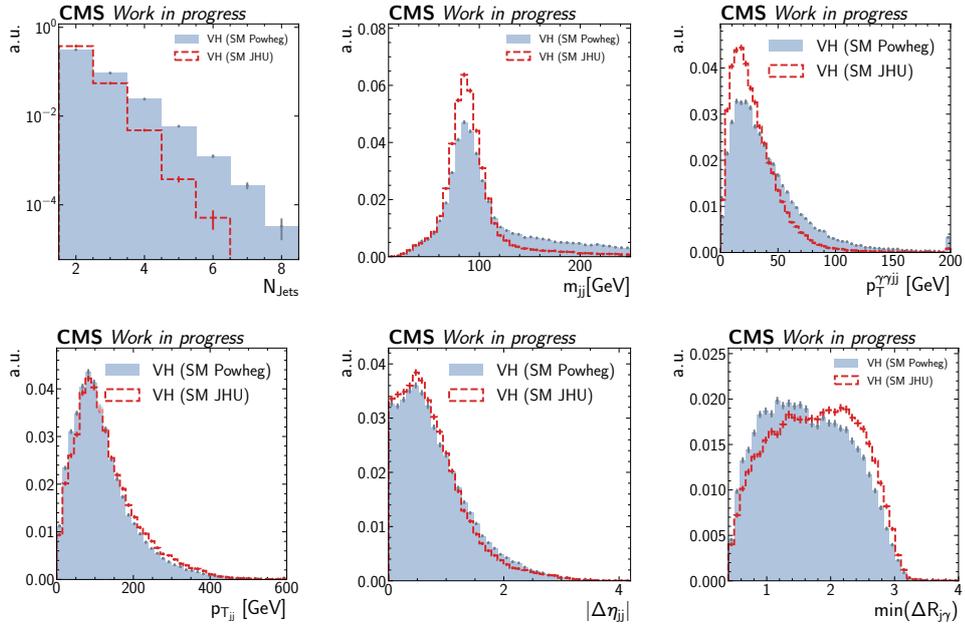


Figure V.2 – Comparison of the differences in event properties between the generation of an SM sample at LO with JHUGEN and at NLO with POWHEG.

in the number of jets N_{Jets} due to a lower number of jets produced at LO. This has a significant impact on the dijet mass distribution as the combinatorial background is reduced. The two jets coming from the decay of the V boson are identified more frequently, resulting in a thinner dijet mass distribution. Another difference is visible in the minimal angle between a jet and a photon $\min(\Delta R_{j\gamma})$. As less jets are generated, $\min(\Delta R_{j\gamma})$ is higher in the JHUGEN case. The p_T of the reconstructed intermediate vector boson ($p_T^{\gamma jj}$) is also softer in the LO case. Therefore, a reweighting of the JHUGEN samples is implemented to correct for these discrepancies.

To implement a reweighting efficient on all BSM samples generated with JHUGEN, the reweighting is computed from event properties that do not depend on the BSM hypothesis, such as the number of jets N_{jets} or the dijet mass m_{jj} , as shown in Fig. V.3. This reweighting is then applied to all JHUGEN samples to improve the description of BSM processes. After numerous attempts, no set of variables has proven capable of correcting all discrepancies between the JHUGEN and POWHEG descriptions of the SM distributions. The best reweighting we reached is offered by using the $(m_{jj}, N_{\text{jets}})$ pair of distributions in two dimensions. An illustration of some reweighting choices is shown in Fig. V.4 with examples both before (filled histogram) and after reweighting (dashed histograms).

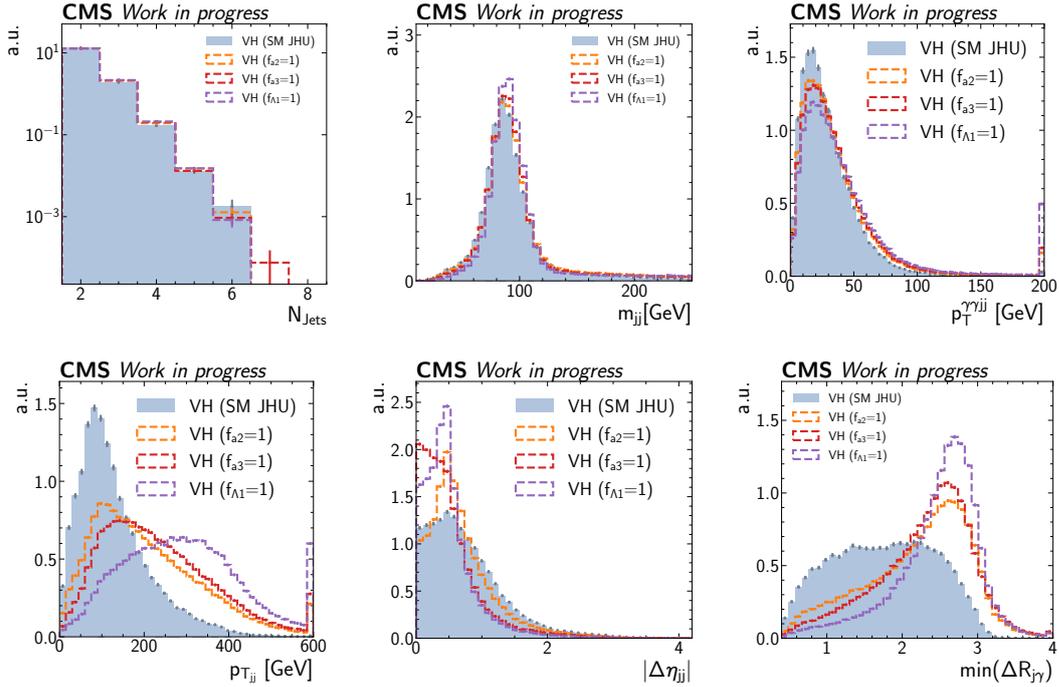


Figure V.3 – Comparison of different BSM hypotheses.

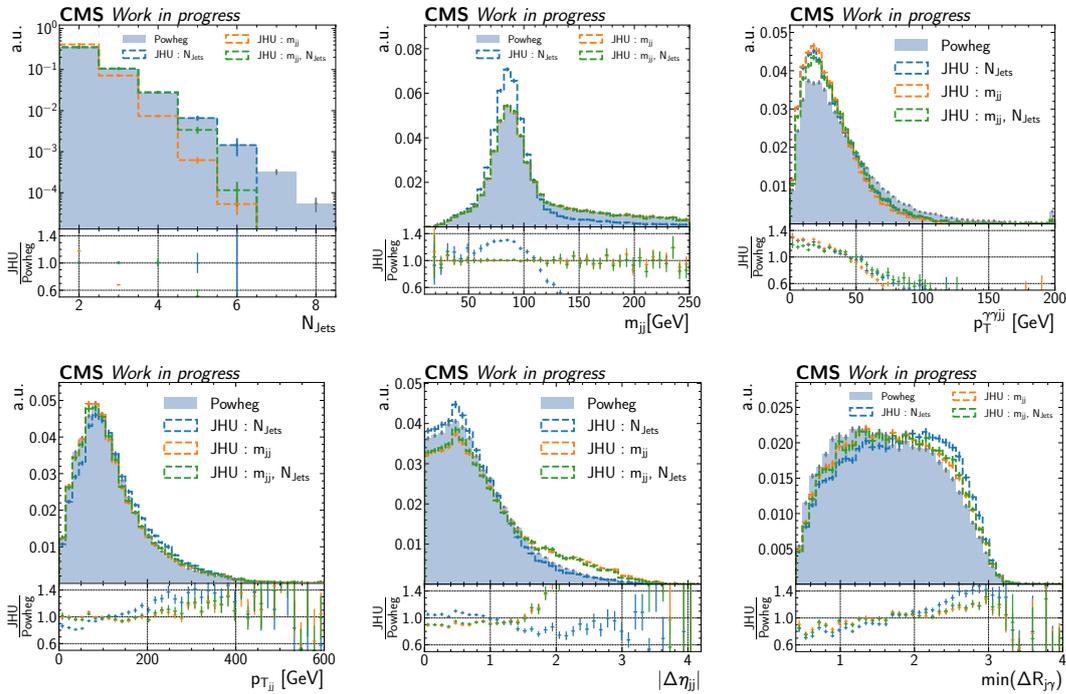


Figure V.4 – Example of kinematic distributions of the VH process in the SM with NLO QCD generation (POWHEG, filled histogram) and attempts at reweighting the SM sample with LO generation (JHUGEN, dashed histograms).

The goal of this reweighting is to get a better description of the BSM distributions to improve the definition of the categories defined in Section. V.5. These categories are defined from the outputs of a multiclassifier DNN trained to enhance the discrimination of the different physics processes. Thus, the reweighting is done on the DNN outputs directly to reach the optimal agreement between the DNN distributions of both SM samples (LO and NLO). More details on this reweighting are given in Section. V.5. In addition to the reweighting, a systematic uncertainty is derived to account for the effects of an imperfect reweighting and is described in Section. V.6.4. A similar procedure is applied in the VBF and $V_{\text{lep}}H$ parts of the analysis.

V.3.3 Phenomenology of $V_{\text{had}}H$ events and preselection

As mentioned in Section. V.1.1, the $H \rightarrow \gamma\gamma$ analysis is sensitive to the couplings of the Higgs boson with electroweak gauge bosons through the VH and VBF production modes of the Higgs boson. The HVV vertex can be clearly identified from their Feynman diagram presented in Fig. V.5.

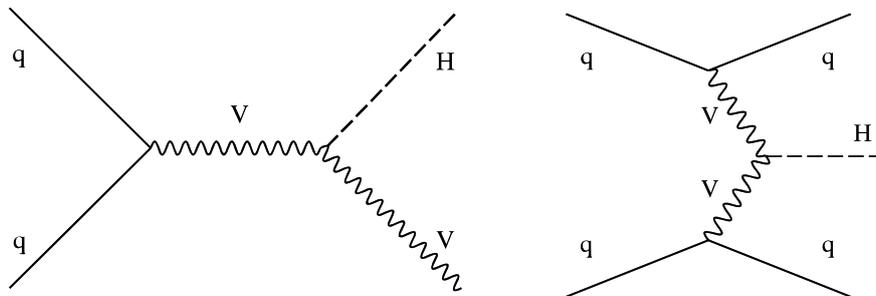


Figure V.5 – Feynman diagrams of the VH (left) and VBF (right) production modes of the H boson.

The analysis is divided into orthogonal phase spaces. The first one is dedicated to the VBF production, and two others depend on the final state of the VH production (where $V = Z$ or W). Indeed, with the production of a gauge boson in the VH case, we can expect the vector boson to decay in two hadrons ($\mathcal{B}(Z \rightarrow q\bar{q}) = 70\%$, $\mathcal{B}(W \rightarrow q\bar{q}) = 67\%$), in two leptons ($\mathcal{B}(Z \rightarrow \ell\ell) = 3.3\%$ with $\ell = e, \mu$ or τ), in one lepton and one neutrino ($\mathcal{B}(W \rightarrow \ell\nu) = 11\%$), or two neutrinos ($\mathcal{B}(Z \rightarrow \nu\nu) = 20\%$). Events with at least one lepton are gathered as $V_{\text{lep}}H$, events with a significant missing transverse momentum correspond to $V_{\text{MET}}H$, and other events with at least two jets belong to the $V_{\text{had}}H$ phase space.

Since we are responsible for optimising the sensitivity in the $V_{\text{had}}H$ case, we aim to define a way to extract $V_{\text{had}}H$ events from other processes. As such, during the design of $V_{\text{had}}H$ strategy, we consider other SM productions of the H boson as background, in addition to the QCD processes producing two photons in their final state ($\gamma\gamma + \text{jets}$) or with misidentified photons ($\gamma + \text{jets}$). This extraction is done through preselection of the events from their distinctive kinematic features and by training a DNN (described in Section V.4) to classify events. The preselection already provides a way to separate $V_{\text{had}}H$ events from other pro-

duction modes of the SM. The $V_{\text{lep}}H$ part of the analysis requires leptons in its final state, so we apply a veto on events with leptons. Contrary to the VBF phase space, we expect two hadrons originating from the decay of a W or Z boson, so the invariant mass of the dijet system m_{jj} should be close to m_W or m_Z . Thus, we focus on the region $m_{jj} < 250$ GeV while the VBF part focuses on the region where $m_{jj} > 250$ GeV. Additional requirements are introduced, following recommendations of physics groups of the CMS collaboration in charge of physics object reconstruction, and the final set of requirements is given in Table V.2.

Full $V_{\text{had}}H$ preselection
$100 \text{ GeV} < m_{\gamma\gamma} < 180 \text{ GeV}$
$p_{T\gamma_1}/m_{\gamma\gamma} > 1/3$
$p_{T\gamma_2}/m_{\gamma\gamma} > 1/4$
$ID_\gamma > 0$
$ \eta_\gamma < 2.5$
$p_{Tj} > 30 \text{ GeV}$
$ \eta_j < 2.5$
$m_{jj} < 250 \text{ GeV}$

Table V.2 – Presentation of the preselection behind the definition of the signal region for the $V_{\text{had}}H$ case.

Examples of distributions showing discrimination potential are shown in Fig. V.6. These discriminating features are particularly useful for the training of discriminants such as DNN. Most of the diphoton features allow the discrimination between Higgs and non-Higgs processes because the diphoton pair is more boosted when coming from the decay of a Higgs boson. This results in a harder p_T spectrum for the photons ($p_{T\gamma_1}$ and $p_{T\gamma_2}$) and tighter differences in azimuthal angles ($\Delta\phi_{\gamma\gamma}$) or pseudorapidity ($\Delta\eta_{\gamma\gamma}$). Other variables allow the direct separation of the $V_{\text{had}}H$ process from all background processes, such as the dijet mass m_{jj} (as already described) or the $\cos(\theta^*)$ property. The angle θ^* is defined as the angle between the direction of the diphoton pair in the diphoton-dijet centre-of-mass and the direction of the diphoton-dijet system in the lab frame. Given the common origin of the dijet and diphoton systems in the $V_{\text{had}}H$ case, the $\cos(\theta^*)$ is flat, while for other processes where the dijet and diphoton systems are not correlated, the distribution is peaked at $\cos(\theta^*) = -1$ or 1.

Distributions comparing the different BSM scenarios are presented in Fig. V.7. To allow for a fair comparison, the distributions from LO generation are compared to an NLO generation in Fig. V.8 to distinguish effects from the order of generation and effects due to different BSM hypotheses. A major difference between samples with a BSM scenario and SM distribution is that the BSM scenarios describe loop-induced processes. It results in a much harder p_T spectrum and more boosted reconstructed objects.

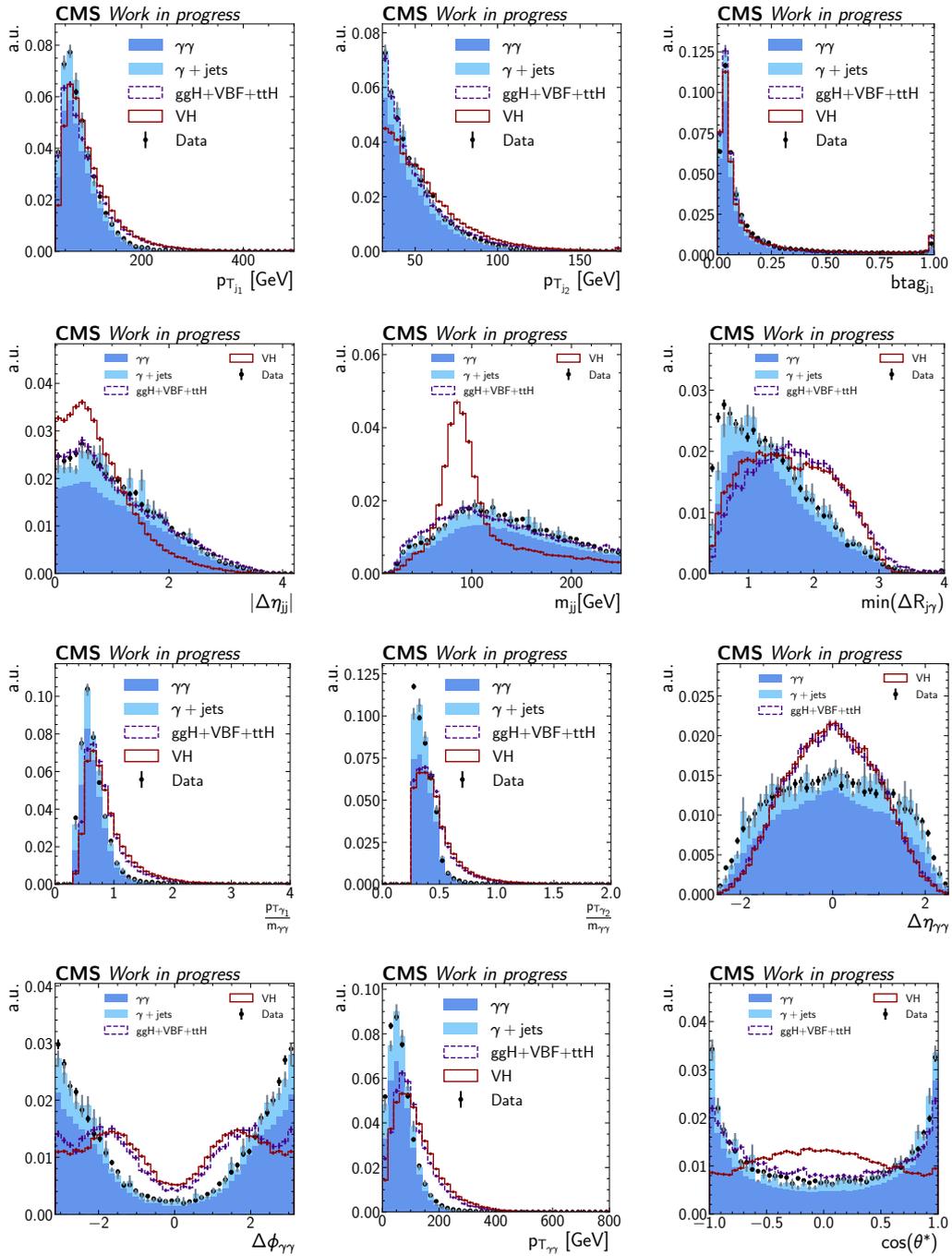


Figure V.6 – Distributions of event variables for SM samples. They are normalised to the VH sample to better illustrate differences in shape. The $\gamma\gamma$ and $\gamma + \text{jets}$ distributions are first stacked and then normalised.

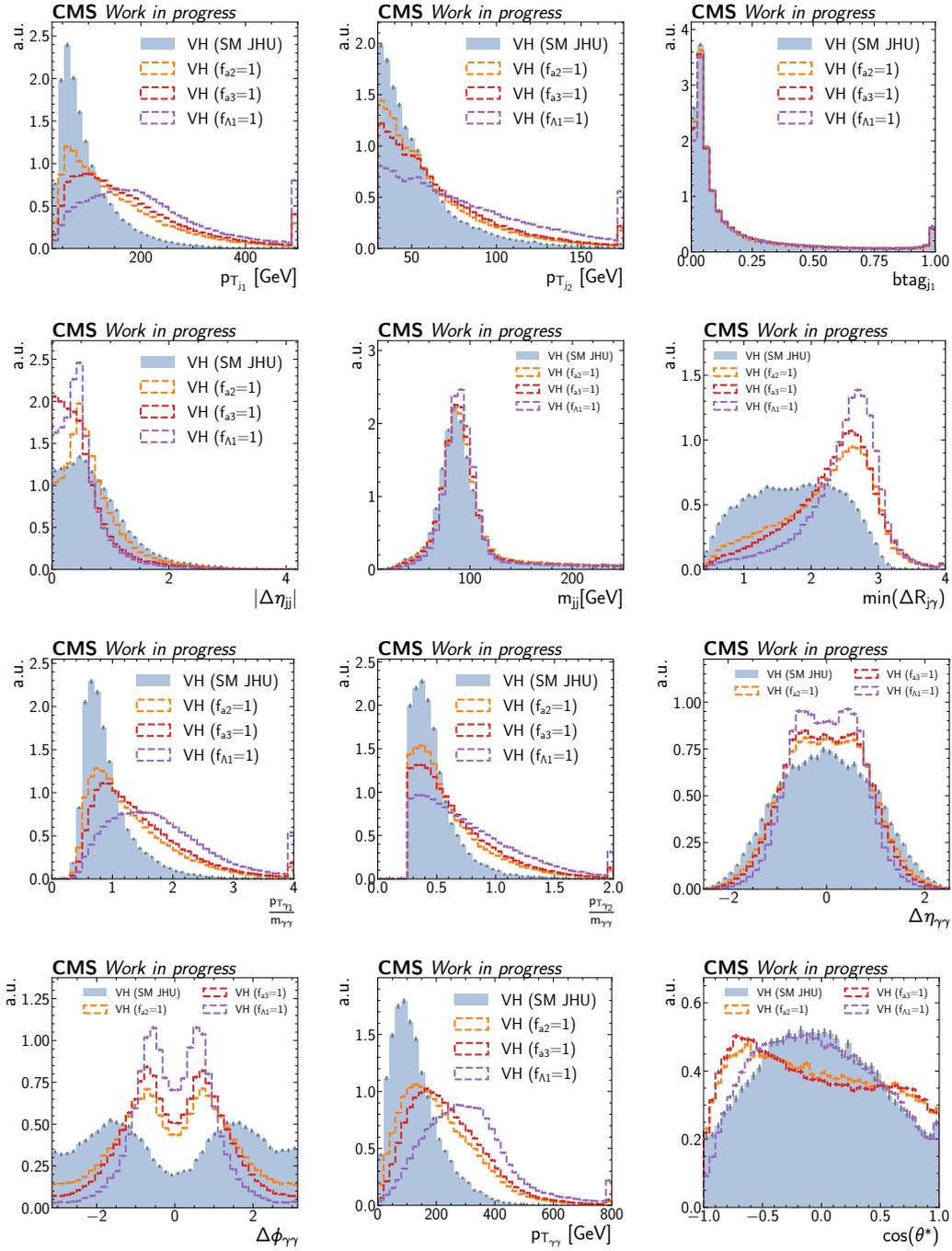


Figure V.7 – Distributions of event variables of the VH process in different BSM scenarios. They are normalised to the SM VH sample to better illustrate differences in shape.

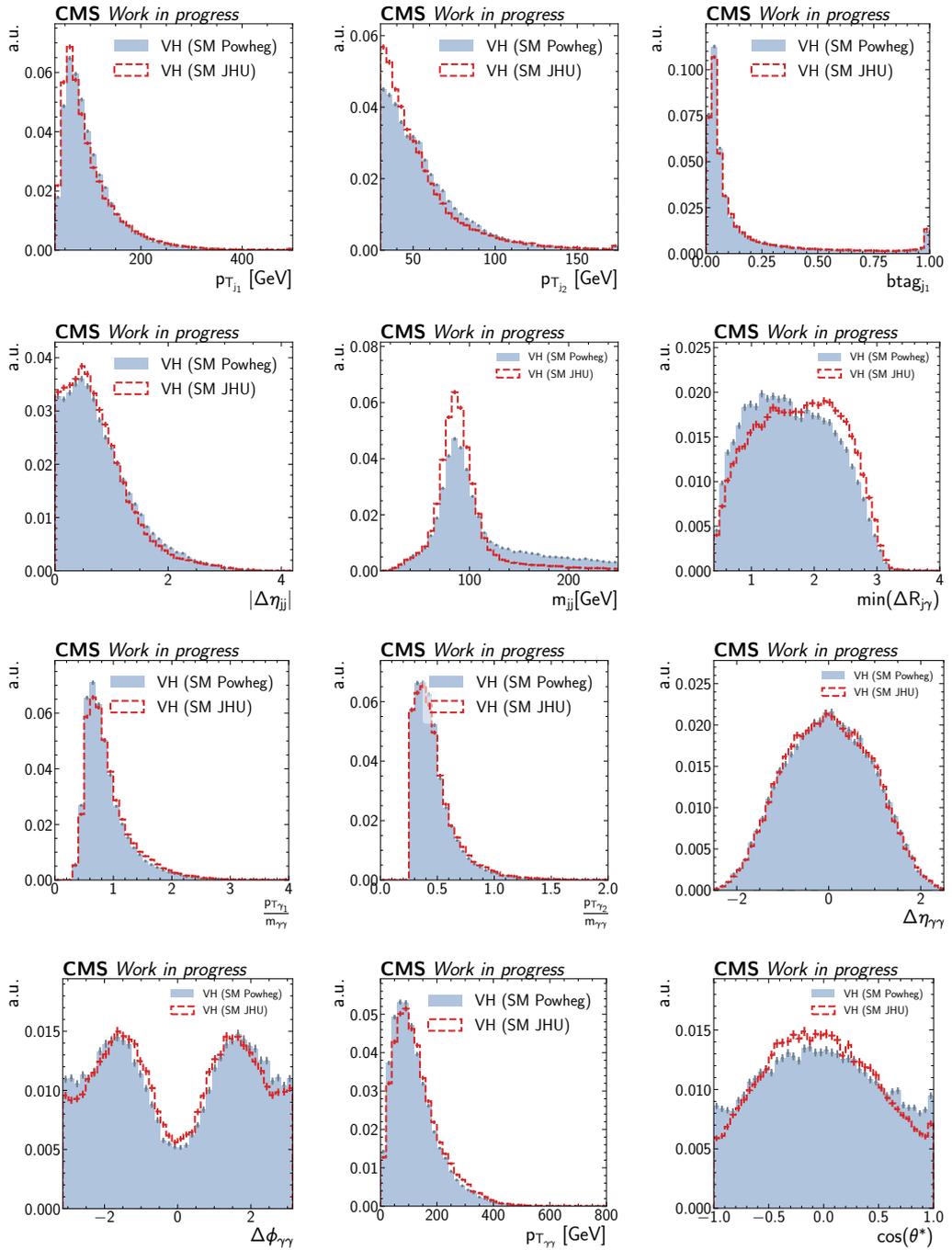


Figure V.8 – Distributions of event variables of the VH process in the SM case to compare the effect of using LO (JHUGEN) or NLO (POWHEG) computations to generate the simulated sample.

V.4 Machine learning for events discrimination

In addition to the dedicated preselection, a multiclassifier DNN is specifically trained to identify events within the $V_{\text{had}}H$ phase space. Multiclassifiers are neural networks with multiple outputs representing the probability for an event to belong to a given class. In our case, its role is to separate events into three classes:

- **$V_{\text{had}}H$ background:** backgrounds to the VH hadronic production mode include resonant background from other SM production modes of the Higgs boson or non-resonant background coming from $pp \rightarrow \gamma\gamma$ or $pp \rightarrow \gamma + \text{jets}$ processes. For the resonant background, the ggH sample is used, and each background component (ggH, $\gamma\gamma + \text{jets}$, $\gamma + \text{jets}$) is then mixed with the same proportion (33% of the total background weights).
- **$V_{\text{had}}H$ SM-like signal:** $V_{\text{had}}H$ events with couplings as expected in the SM.
- **$V_{\text{had}}H$ BSM-like signal:** $V_{\text{had}}H$ events displaying deviations from the expected couplings of the SM.

The three associated output probabilities given by the DNN are respectively DNN_{bkg} , DNN_{vh} , and DNN_{bsm} following the relation $DNN_{\text{bkg}} + DNN_{\text{vh}} + DNN_{\text{bsm}} = 1$. As inputs to the network, we use the kinematic properties of the two photons and the two jets with leading p_T and features with high discrimination power between the three considered classes. We base our selection of input features on the $V_{\text{had}}H$ BDT developed for the STXS analysis [145] and extend them to the following list: $p_{T\gamma_1}/m_{\gamma\gamma}$, η_{γ_1} , ID_{γ_1} , $p_{T\gamma_2}/m_{\gamma\gamma}$, η_{γ_2} , ID_{γ_2} , $|\Delta\eta_{\gamma\gamma}|$, $|\Delta\phi_{\gamma\gamma}|$, p_{Tj_1} , η_{j_1} , ϕ_{j_1} , btag_{j_1} , p_{Tj_2} , η_{j_2} , ϕ_{j_2} , btag_{j_2} , m_{jj} , $|\Delta\eta_{jj}|$, $\min(\Delta R_{j\gamma})$, $\cos\theta^*$. Some of the most discriminating features in the list of inputs are shown in Figs. V.6, V.8 and V.7.

The DNN is trained using Keras and its TensorFlow backend [91]. The training is done with MC samples taking into account the productions $\gamma + \text{jets}$, $\gamma\gamma$, ggH, VH 0+ and VH 0-. During the study of the reweighting procedure for JHUGEN samples, we found that using a reweighting based on $(p_{T\gamma\gamma}, N_{\text{jets}})$ before the training of the DNN helped the network to improve its performance. The weights of the events are used during the training to weight the categorical cross-entropy loss accordingly, and the sum of the weights in each class is normalised to bring the same contribution to the loss computation. A proportion of 25% of the total number of events is kept for validation, *i.e.* to check for any overfitting while training.

Multiple training setups are tested through the Hyperband algorithm [174] of the Keras tuner [175] to determine which provides the best-performing network. The Hyperband algorithm is an extension of the random search for the optimal set of hyperparameters. It considers the finite amount of resources available (*e.g.* training time, number of epochs, *etc.*) and relies on a tournament-based search. The algorithm starts by providing a small number of resources, for instance, a few training epochs, to a large number of trainings with randomly drawn hyperparameters configurations. After these few epochs, the algorithm discards two-thirds of the worst-performing trainings based on the result of their validation loss and continues training the best-performing one. Finally, only

one training configuration remains and is designated as the winner of the initial bracket. Five brackets are formed, and the model with the lowest validation loss among all brackets identifies the optimal set of hyperparameters. The Hyperband algorithm optimises hyperparameters among the number of neurons for each layer of the network, the percentage of dropout while training, the slope of the leakyReLU activation function, the batch size, and the learning rate. The optimal values are found to be :

- **Number of neurons:** 512 (first layer), 128 (second layer), 32 (third layer)
- **Dropout:** 0.25
- **Slope of leakyReLU:** 0.0
- **Batch size:** 128
- **Learning rate:** 10^{-3}

The performance of the optimised network is shown in Fig. V.9 with the evaluation of the three DNN outputs on samples belonging to the three different classes. The GJet, Diphoton and ggH samples corresponds respectively to the $\gamma + \text{jets}$, $\gamma\gamma + \text{jets}$ and ggH processes belonging to the background class. The VH sample represents the $V_{\text{had}}H$ SM-like signal class. Finally, the VH0M sample is describing events in the $f_{a3} = 1$ hypothesis which constitute the $V_{\text{had}}H$ BSM-like signal class. It illustrates the discrimination possibilities brought by the three DNN scores. Figure V.10 shows the same evaluation from a different point of view with the distributions of the three scores sample by sample and the 2D histograms hinting at correlations between the three possible pairs of DNN output. Furthermore, ROC curves are computed (see Fig. V.11) to probe the ability of each DNN score to distinguish events coming from two different classes.

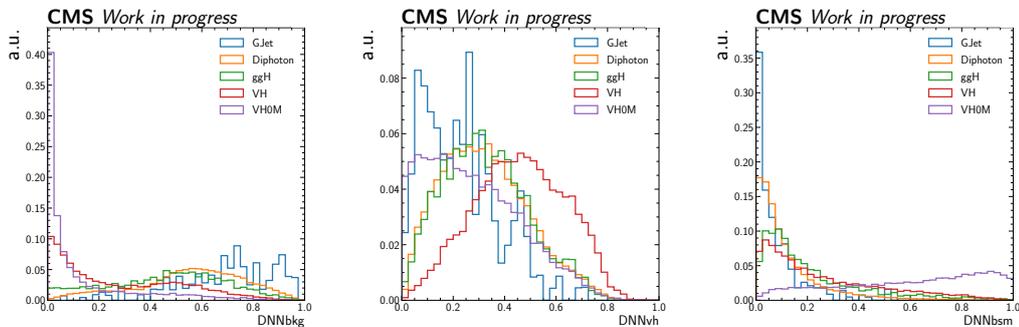


Figure V.9 – Evaluation of the DNN outputs on samples from the three classes.

From several figures of merits inspecting the performance of the DNN (especially Fig. V.10), it appears that the outputs of the DNN are strongly correlated. Thus, rather than relying on a single score for each event, a better strategy to build categories enriched in SM-like or BSM-like signal events is to rely on 2D cuts on two of the three DNN outputs. Since their sum is one, constraints on two of the DNN outputs are actually fixing the third one as well. We can see from Fig. V.10 that events are already localised in distinct areas based on their class in the plane $(DNN_{\text{bkg}}, DNN_{\text{bsm}})$ in the 3rd column, offering good categorisation possibilities. Fig. V.11 confirms this by showing that the background and BSM

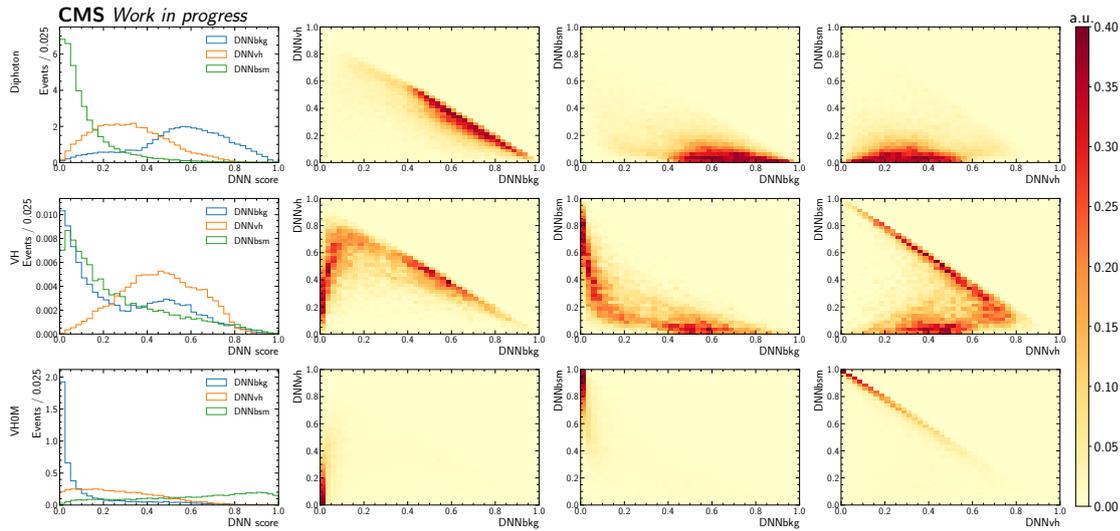


Figure V.10 – Evaluation of the DNN outputs sample by sample (one sample per row). The first column shows the distributions of the three DNN scores for a given sample. From the second to the fourth column, the distributions of the events of a given sample in the three possible 2D planes formed by pairs of the DNN scores.

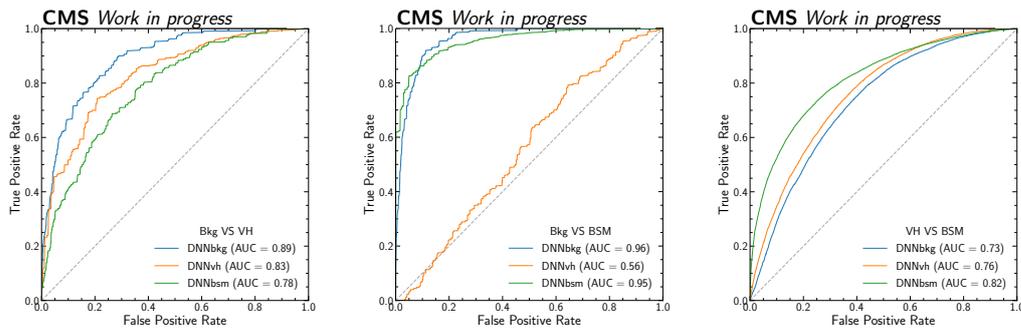


Figure V.11 – ROC curves evaluating the performance of each DNN score in identifying events belonging to the three possible pairs of classes.

scores of the DNN provide suitable identification for any pair of classes as the maximal AUC is always from the ROC curve of the background or BSM score.

As a comparison, two DNN classifiers with a single output each were trained, the first one to extract SM-like signal events and the second one to extract BSM-like events. Their respective output score was used to build categories and optimise them as described in the next section. This method showed 4% worst results on the figure of merit described in Section V.5 (upper limit of f_{a3}) after optimising categories. In addition to having worst results, this method requires training two DNNs instead of one multiclassifier, so the method is less relevant.

An additional test was performed by including the quark-gluon tag (qgtag) information of jets as inputs of the DNN multiclassifier. This score represents the likelihood for the jet to originate from a quark (score close to 1) or from a gluon (score close to 0). Since we target VH events where the gauge boson decays in two quarks (and not in leptons), the qgtag of both jets is peaked at 1 as opposed to the inclusive $H + 2\text{jets}$ production where the 2 jets tend to originate from gluon

radiations. Thus, it provides additional discrimination information to the DNN. And indeed, the area under the ROC curve computed over all classes, when training the DNN with the qqtag, improved by 2%. Unfortunately, for this iteration of the analysis, we abandoned this additional discrimination because the systematic uncertainty on the quark-gluon tagger is not yet determined. However, we recommend including it, if available, for the next iterations of DNN trainings in the $V_{\text{had}}H$ case.

V.5 Optimisation of analysis categories

The final results are extracted using a statistical test (described in Section V.6.1) which scans values of the observables of interest and tests the compatibility of data with the corresponding physics model. Hence the necessity to define categories targeting regions of the phase space where any difference between different models would be striking. Constructing categories allows the evaluation of the separation of the different hypotheses by the statistical test, and then, by improving the definition of the categories, we enhance the sensitivity of the analysis. The regions sensitive to signal are built from the DNN outputs, but before optimising these categories, the reweighting of the DNN outputs mentioned in Section V.3.2 is applied. The effects of this reweighting are demonstrated in Fig. V.12 by comparing it to a reweighting using physics variables. As expected, the DNN distributions are recovered for the SM distributions, where the reweighting on physics still shows significant deviations.

To increase our sensitivity to processes including anomalous couplings of the Higgs boson to gauge boson, we define categories enriched in SM- and BSM-like signal events in the $V_{\text{had}}H$ preselected phase space. The total number and definition of these categories are optimised to minimise the expected upper limit on f_{a3} at 95% confidence level. This upper limit on the value of f_{a3} is computed with a likelihood ratio as a test statistic and following prescriptions from previous Higgs boson searches at the LHC [176] and Cowan *et al.* [177]. When constraining only the f_{a3} parameter, the number of expected signal events in category k is parametrised as

$$\mu_{a_3} s_k = \mu_{a_3} [(1 - f_{a3}) s_k^{\text{SM}} + f_{a3} s_k^{a_3}] \quad (\text{V.2})$$

with s_k^{SM} the expected number of SM-like signal events, $s_k^{a_3}$ the number of events in the hypothesis $a_3 = 1$, $a_1 = 0$, and μ_{a_3} is the signal strength associated with a given value of f_{a3} . Under this decomposition, we can write the Poisson probability for n_k events to fall in category k , given the expected number of signal s_k and background b_k events :

$$\mathcal{P}(n_k | \mu_{a_3} s_k, b_k, f_{a3}) = \frac{(\mu_{a_3} s_k + b_k)^{n_k}}{n_k!} e^{-\mu_{a_3} s_k - b_k} . \quad (\text{V.3})$$

Then, the negative log-likelihood (NLL) is defined, up to a normalisation factor, as

$$-\ln \mathcal{L}(\mathbf{n} | \mu_{a_3} \mathbf{s}, \mathbf{b}, f_{a3}) = \sum_k^{\text{N}_{\text{cat}}} (\mu_{a_3} s_k + b_k) - n_k \ln (\mu_{a_3} s_k + b_k) . \quad (\text{V.4})$$

The signal strength μ_{a_3} is not known a priori, but its maximum likelihood estimator $\hat{\mu}_{f_{a3}}$ can be profiled from the NLL for each value of f_{a3} . The estimator $\hat{\mu}_{a_3}$

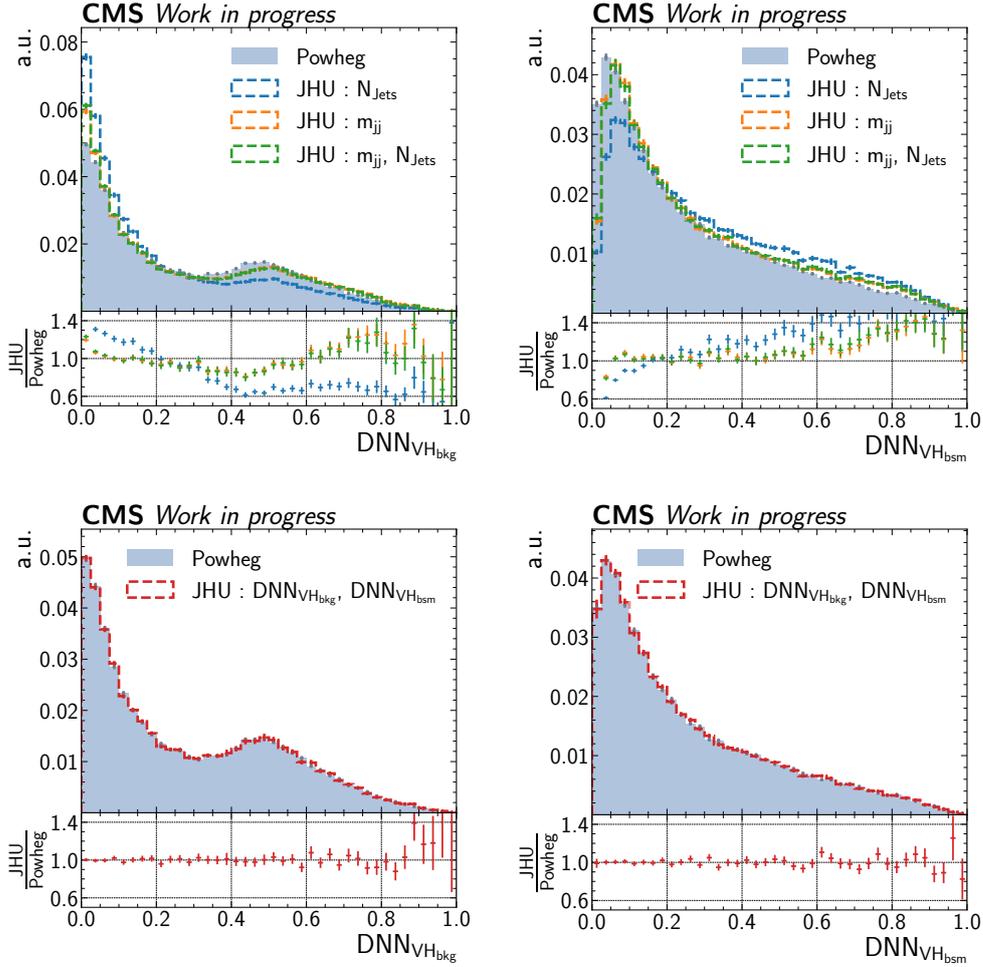


Figure V.12 – Comparison of the reweightings using physics variables (top row) and the DNN outputs (bottom row) on the shapes of the DNN outputs for the background (left column) and BSM classes (right column).

is the value of μ_{a_3} for which the NLL is minimal, so it can be extracted from the derivative of the NLL:

$$\left. \frac{d \ln \mathcal{L}}{d\mu_{a_3}} \right|_{\mu_{a_3} = \hat{\mu}_{a_3}} = 0 \Rightarrow \sum_k^{N_{\text{cat}}} \left(s_k - \frac{n_k s_k}{\hat{\mu}_{a_3} s_k + b_k} \right) = 0. \quad (\text{V.5})$$

Thus, for any value of f_{a_3} , we can use a test statistic $q_{f_{a_3}}$ entirely defined through

$$q_{f_{a_3}}(\mathbf{n}) = -2 \ln \frac{\mathcal{L}(\mathbf{n} | \hat{\mu}_{a_3} \mathbf{s}, \mathbf{b}, f_{a_3})}{\mathcal{L}(\mathbf{n} | \hat{\mu}_0 \mathbf{s}_{\text{SM}}, \mathbf{b}, 0)}. \quad (\text{V.6})$$

which test the compatibility between a given f_{a_3} hypothesis and the SM-only hypothesis ($a_3 = 0$ and thus $f_{a_3} = 0$). To estimate the expected upper limit on f_{a_3} offered by a given configuration of categories, we build an Asimov data set where the yield in each category assumes no contribution from BSM processes, *i.e.* $n_k^A = \mu^* s_k^{\text{SM}} + b_k$, where μ^* is an ad hoc normalisation to compensate for the low number of expected events. We use a value of $\mu^* = 5$ and keep it identical to compare all categories configurations. As such, the value obtained for the expected upper limit of f_{a_3} is not the final expected upper limit provided through

our analysis strategy but rather a relevant relative value for the classification of configurations. Finally, the expected upper limit at 95% confidence level is computed as the value of f_{a3} for which

$$\sqrt{q_{f_{a3}}(\mathbf{n}^A)} = \Phi^{-1}(1 - 0.95), \quad (\text{V.7})$$

where Φ is the cumulative distribution function of the normal distribution [177]. The number of expected signal events, \mathbf{s}^{SM} and \mathbf{s}^{a3} , corresponding to SM and BSM processes respectively, is taken as the number of events in the range $m_{\gamma\gamma} \in [122.5 \text{ GeV}, 127.5 \text{ GeV}]$ from their associated MC sample.

The expected number of background events, \mathbf{b} , is computed from a maximum likelihood fit to the $m_{\gamma\gamma}$ distribution in data sidebands, *i.e.* $m_{\gamma\gamma} \notin [115 \text{ GeV}, 135 \text{ GeV}]$ using a power-law model as illustrated in Fig. V.13. Then, the fitted function is integrated over the range $m_{\gamma\gamma} \in [122.5 \text{ GeV}, 127.5 \text{ GeV}]$ as well to estimate the number of background events. In addition to this number of events from the non-resonant background, we take into account events from the ggH process.

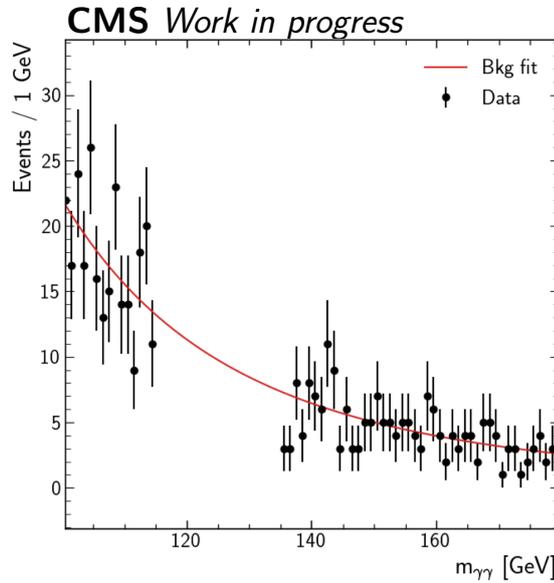


Figure V.13 – Example of background fit (red line) performed on the data sidebands (black dots) entering a given category.

To find the optimal number and shape of categories to build from the DNN_{bkg} and DNN_{bsm} scores, the $(\ln(\text{DNN}_{\text{bkg}}), \text{DNN}_{\text{bsm}})$ plane is filled with categories until the improvement on the expected upper limit of f_{a3} becomes negligible. It results in eight categories, whose borders have been optimised to minimise the upper limit on f_{a3} , resulting in an upper limit of 0.253 for $\mu^* = 5$. These categories are presented in Fig. V.14.

As all categories do not bring the same sensitivity to the estimation of f_{a3} , we remove each category one by one, compute the new upper limit (with a missing category) and compare it to the original value (with eight categories) to estimate the impact of each category on the optimisation of the upper limit of f_{a3} . Results of this study are summarized in Table V.3.

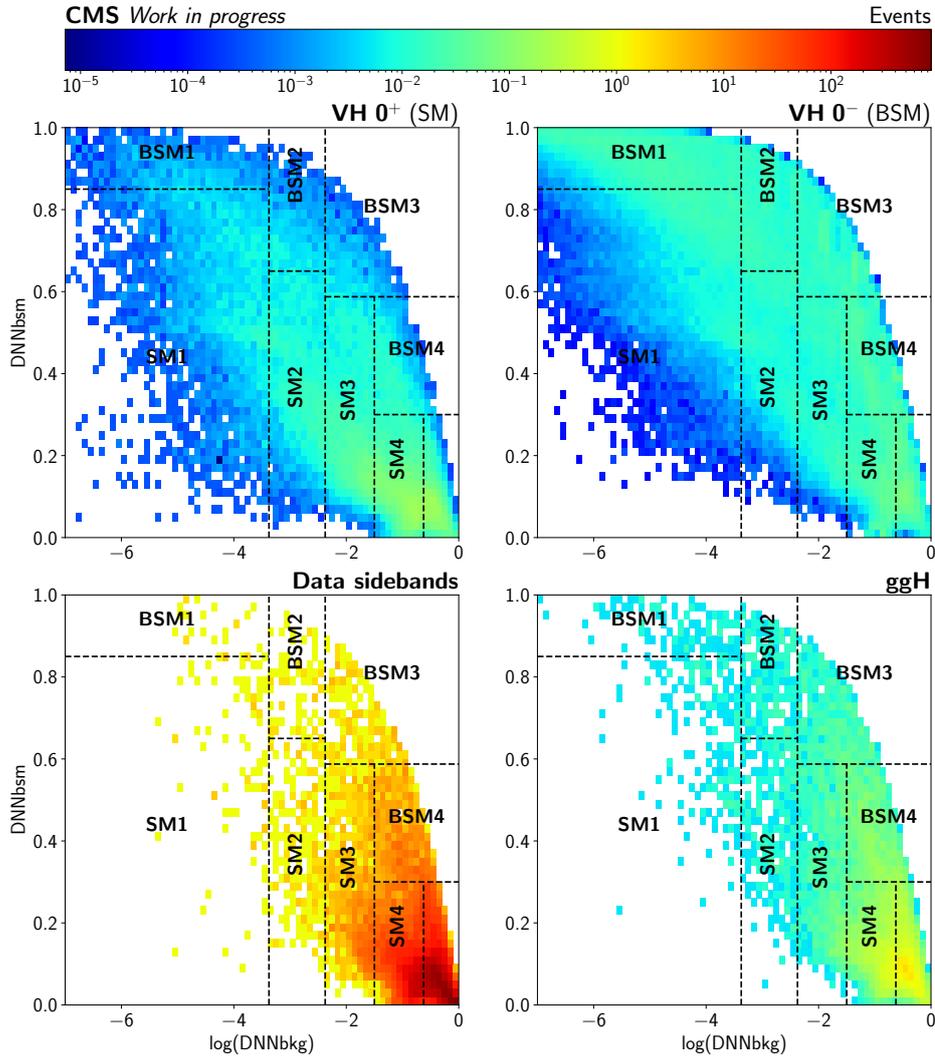


Figure V.14 – The categories are overlaid to the 2D distribution of the events as a function of their DNN scores. The 2D distributions represent different classes of the DNN (from left to right, top to bottom: SM-like, BSM-like, background from data sidebands and background from ggH process).

Removed	BSM1	BSM2	BSM3	BSM4	SM1	SM2	SM3	SM4
Impact on upper limit	+129%	+3%	+2%	< +1%	+3%	+15%	+13%	+8%

Table V.3 – Impact of each category on the estimation of the upper limit of f_{a3} .

From the impact values, we discard the BSM4 category. The SM1 category is sensitive because it mainly comprises BSM events, so the top of the SM1 category is merged with the BSM2 and BSM3 categories. It creates a category collecting events around the very sensitive BSM1 category. The bottom of category SM1 is merged with SM2, and categories SM3 and SM4 are kept in the same position.

The optimal categorisation is presented through Figs. V.15 and V.16. Five categories are left: two enriched in BSM-like signal events (called BSM1 and BSM2) and three enriched in SM-like events (SM1, SM2 and SM3). As some of the categories are chosen to be adjacent, nine boundaries are needed to build the five categories, creating a 9D optimisation problem. The optimal set of selection criteria gives the lowest expected upper limit on f_{a3} from the procedure described above. Figure V.15 shows the values of the expected limit of f_{a3} when scanning the possible boundaries for the definition of the SM1 (left) and SM2 categories (right). The optimal categories are overlaid to the 2D distributions of events from

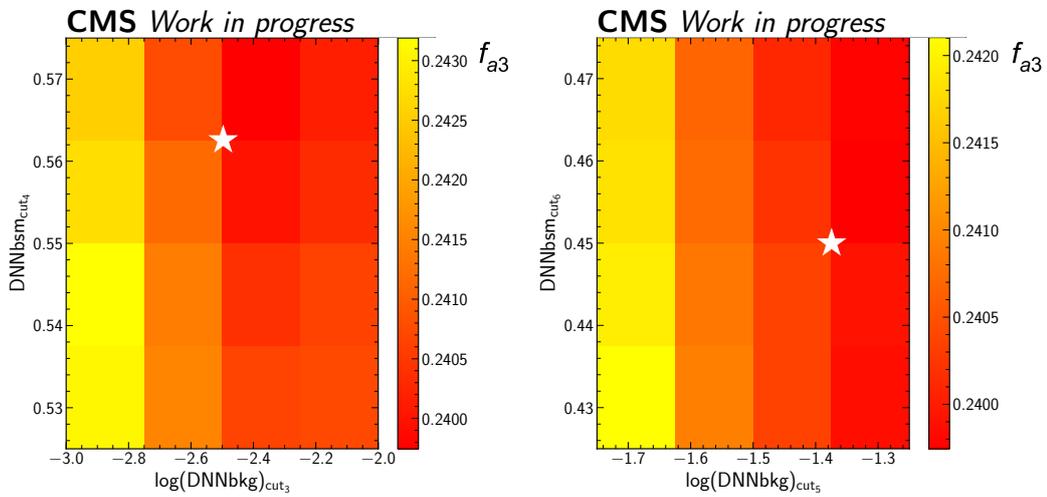


Figure V.15 – Projection of the 9D optimisation in 2D planes containing the lowest upper limit on f_{a3} (white star). The heatmaps show the values of f_{a3} when scanning the requirements on the DNN scores for category SM1 (cut3 and cut4 on the left) and for category SM2 (cut5 and cut6 on the right).

the three different classes in Fig. V.16. The lowest upper limit on f_{a3} at 95% confidence level is 0.227 for $\mu^* = 5$, and the expected yield in each category is summarised in Table V.4.

V.6 Statistical inference and results

V.6.1 Inference procedure

The statistical methods used to extract the expected and observed results of the analysis are widespread in analyses of the ATLAS and CMS experiments [176]. These methods rely on the definition of a test statistic, which acts as the figure of merit assessing the compatibility between two hypotheses. A hypothesis represents a physics model, and in our particular context, the two hypotheses (or

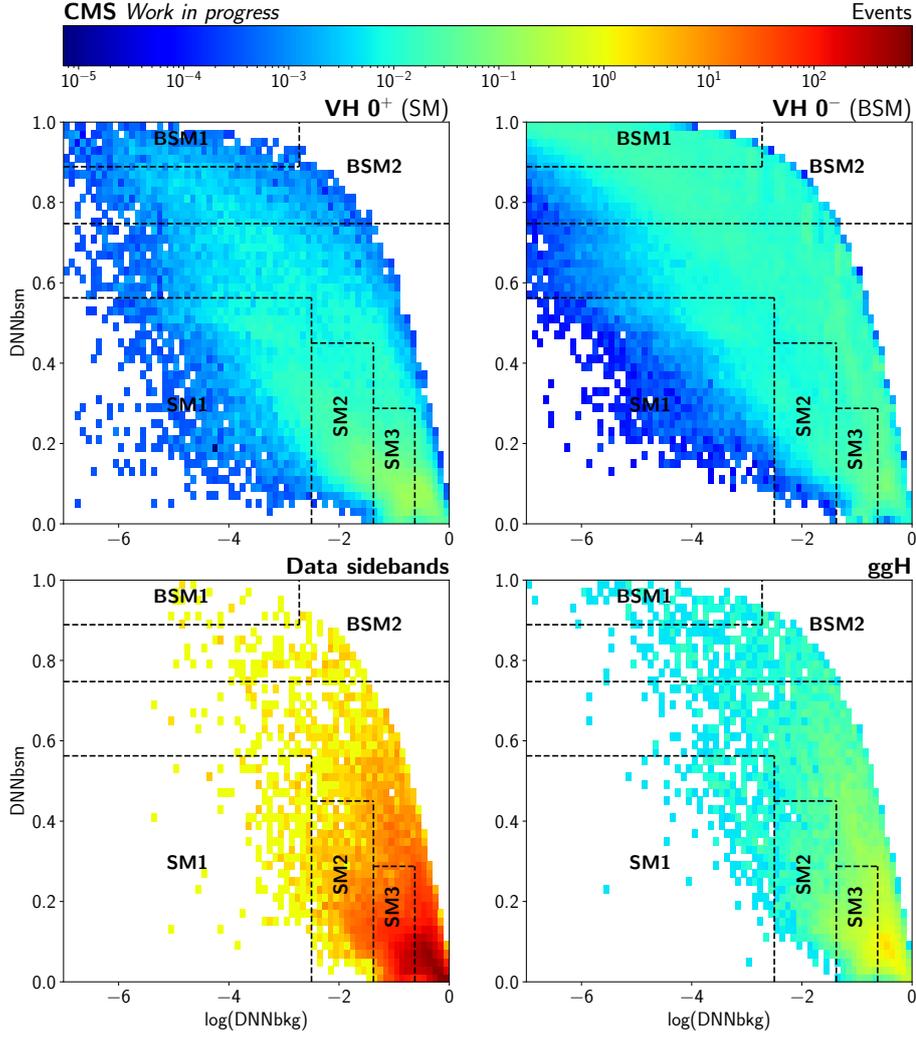


Figure V.16 – Final categories after the optimisation procedure. The categories are overlaid to the 2D distribution of the events as a function of their DNN scores. The 2D distributions represent different classes of the DNN (from left to right, top to bottom: SM-like, BSM-like, background from data sidebands and background from ggH process).

$V_{\text{had}}H$ category	SM-like	BSM-like	Background/GeV
BSM1	0.18	3.26	0.33
BSM2	0.62	3.57	1.69
SM1	2.10	1.57	2.37
SM2	3.67	1.62	16.44
SM3	4.91	1.92	83.59

Table V.4 – Estimated yield from each DNN class in the three optimised categories defined with the DNN scores. Values for the background are given in number of events per GeV within the 5 GeV mass region of interest.

models) we want to compare are the standard model and the model described in Section V.1, where any of the couplings (apart from a_1) is nonzero. As described in Section V.5, categories are built from the kinematics of the events (indirectly as they are transformed to more discriminating features with the DNN) since it differs significantly between the two hypotheses.

The test statistic used to extract the final value and uncertainties of the f_{a_i} quantities is defined in a similar fashion to what is described in Section V.5. Each f_{a_i} is scanned one by one by fixing all the others to 0. However, by taking into account the potential interference between the scanned fraction terms ($\mathbf{f} = (f_{a_2}, f_{a_3}, f_{\Lambda_1}, f_{\Lambda_1}^{Z\gamma})$) and the a_1 term, as described in Eq. V.1, the expected number of signal events in each category includes an additional term compared to Eq. V.2:

$$s_k = (1 - f_{a_i})s_k^{\text{SM}} + f_{a_i}s_k^{a_i} + 2\sqrt{f_{a_i}(1 - f_{a_i})}s_k^{\text{int}}\text{sgn}(a_1 \cdot a_i) \quad (\text{V.8})$$

with s_k^{SM} ($s_k^{a_i}$) the expected number of SM-like (BSM-like) signal events, and s_k^{int} is the effective number of signal event emerging from the interaction of the amplitudes related to the a_1 and a_i terms. Systematic uncertainties are added as nuisance parameters θ that potentially impact the event yields in each of the categories: $s_k \rightarrow s_k(\theta)$, $b_k \rightarrow b_k(\theta)$. This effect propagates to the probability of data events to fall in a given category $\mathcal{P}(n_k | \mu s_k(\theta), b_k(\theta), \mathbf{f})$ as defined in Eq. V.3. Constraints can be applied to the effects of systematic uncertainties by including a probability distribution function of the uncertainty $p(\theta_i)$ to the likelihood

$$\mathcal{L}(\mathbf{n} | \mu \mathbf{s}(\theta), \mathbf{b}(\theta), \mathbf{f}) = \prod_k^{\text{N}_{\text{cat}}} \mathcal{P}(n_k | \mu s_k(\theta), b_k(\theta), \mathbf{f}) \prod_i^{\text{N}_{\text{nuis}}} p(\theta_i). \quad (\text{V.9})$$

Thus, the test statistic, defined as a likelihood ratio, takes the form

$$q_{\mathbf{f}}(\mathbf{n}) = -2\Delta \ln \mathcal{L} = -2 \ln \frac{\mathcal{L}(\mathbf{n} | \mu \mathbf{s}(\theta), \mathbf{b}(\theta), \mathbf{f})}{\mathcal{L}(\mathbf{n} | \hat{\mu} \mathbf{s}(\hat{\theta}), \mathbf{b}(\hat{\theta}), \hat{\mathbf{f}})} \quad (\text{V.10})$$

where the hat symbol denotes the maximum likelihood estimator of the corresponding variables, *i.e.* their value is the most compatible one with data, and the denominator is actually the global maximum of the likelihood function.

In the diphoton decay channel, we expect a peaked signal in the diphoton mass distribution at the H boson mass (m_{H}) over a smoothly falling background. Therefore, the signal and background contributions are parameterised in each category as a function of the diphoton mass. The modelling of background and signal distributions are described in the following Sections V.6.2 and V.6.3, respectively. These modelling techniques were developed over previous H $\rightarrow \gamma\gamma$ analyses such as Ref. [145], so a short description of the methods is given before describing their application to the $V_{\text{had}}\text{H}$ categories. Once models are designed in each category, the likelihood ratio evaluates the compatibility of the data distribution over the diphoton mass with the combination of the background and signal models simultaneously over all categories. The likelihood ratio is scanned along each f_{a_i} parameter one by one (fixing other BSM contributions to 0), and for each value of the f_{a_i} , the signal strength, m_{H} and systematic uncertainties are profiled and taken as their maximum likelihood estimator. In order to get the

expected results of the analysis, the statistical test is performed on an Asimov data set which consists of a simulation where the maximum likelihood estimator of each parameter is equal to its expected value. In our case, it means that in the Asimov data set, the signal strength is set to 1, the H boson mass to its most precise value of 125.38 GeV [17], systematic uncertainties to their expected value and all the BSM f_{ai} are set to 0.

V.6.2 Background modelling

Standard procedure for H $\rightarrow \gamma\gamma$ analyses at CMS experiment

The modelling of the background distribution as a function of $m_{\gamma\gamma}$ relies on the data sidebands and is not based on MC simulations. Since it is an effective model based on empirical fits, data from all years are merged. The background model fit is performed with different families of functions to find the optimal description of the distribution. These families and their parameters p_i (and q_i when needed) are listed below for a given order N :

- Exponentials: $f_N(x) = \sum_{i=1}^N p_i e^{q_i x}$
- Power-laws: $f_N(x) = \sum_{i=1}^N p_i x^{-q_i}$
- Bernstein polynomials: $f_N(x) = \sum_{i=0}^N p_i \binom{N}{i} x^i (1-x)^{N-i}$
- Laurent polynomials: $f_N(x) = \sum_{i=1}^N p_i x^{-4+\sum_{j=1}^i (-1)^j (j-1)}$

To check the goodness-of-fit obtained with each function, an F-test [178] is computed from each fit and only functions passing a loose requirement on this test are kept as candidates for the final fit procedure. Since these functions can fit any distribution as their order N increases, *i.e.* there is a risk of overfitting data with high-order functions. To prevent this, functions in each family are fitted in increasing order, and a negative log-likelihood is computed at each order, NLL_{fit}^N . It is then compared to the negative log-likelihood of the next order as $2\Delta NLL_{\text{fit}}^{N+1} = 2(NLL_{\text{fit}}^{N+1} - NLL_{\text{fit}}^N)$ should have a χ^2 distribution with m degrees of freedom, where m is the difference in degrees of freedom of functions of order $N+1$ and N . In the tested families of functions, $m = 1$ or 2 as they either have N or $2N$ parameters at order N . A p-value is computed to compare which hypothesis (order N or order $N+1$) data are describing better

$$\text{p-value} = \mathcal{P} \left(2NLL_{\text{fit}}^N > 2\Delta NLL_{\text{fit}}^{N+1} \mid \chi_m^2 \right),$$

where \mathcal{P} is the probability density function associated with a χ^2 distribution with m degrees of freedom. For a p-value lower than 0.1, the procedure continues and tests following orders but for a p-value above 0.1, the higher order function is considered overfitting the data sidebands.

To account for the uncertainty on the final choice of background model among the remaining candidates, they are added as a discrete nuisance parameter to the final fit [179]. Then when scanning the likelihood ratio in the final fit to data,

the likelihood is computed for each of the background-model-candidates functions with the order determined from the modelling in each category and the minimal value is retained. It means that the final likelihood is actually an envelope of all profiles obtained for each background model. Thus, the shape of the final likelihood is broader than with a single fit, so it is a conservative way of taking into account the uncertainty of the background model.

Application to $V_{\text{had}}H$ categories

The above procedure is applied to all $V_{\text{had}}H$ categories described in Section V.5 and is shown in Fig. V.17 for data events from the mass sidebands. Between 5 and 7 background models are selected in each category, and the order of each function is fixed for the final fit. All of these functions are fitted in the final likelihood ratio scan, and the best fit is combined with the signal model defined in the next section.

V.6.3 Signal modelling

Standard procedure for $H \rightarrow \gamma\gamma$ analyses at CMS experiment

A signal model is built from MC samples as a function of $m_{\gamma\gamma}$ in each category, for each year, and separately for events where the right vertex (RV) or the wrong vertex (WV) is selected. This decomposition allows the computation of a combined signal model mixing the contribution from each process depending on the physics scenario (typically, in our case, how the f_{ai} observables are mixing each BSM contribution). The last division, depending on the correctness of the selected diphoton vertex, allows the computation of the fraction of events with correctly identified primary vertex (RV fraction) and to include an uncertainty on its value (in Table V.6).

These numerous signal models are fitted with a sum of one to five Gaussian functions. An F-test [178] is performed to identify the optimal number of Gaussian functions and their parameters to fit the signal distribution. The models representing RV and WV are summed, and the mixture between the two is fitted in each category from MC simulation to get the fraction of RV association.

The dependency of these signal models on m_H is determined by parameterising the models as functions of important parameters, namely: the total normalisation N_{exp} , the cross section σ , the branching ratio \mathcal{B} , the product of selection efficiency and acceptance $\epsilon \times A$, the RV fraction. The dependence on m_H of the cross section and branching ratio is taken from Ref. [173] while for every other parameter, a polynomial dependence is assumed. To obtain the latter, the signal models are derived for three different m_H samples ($m_H = 120$ GeV, 125 GeV, and 130 GeV) and the relevant parameters are fitted with an order 2 polynomial as a function of m_H . This allows interpolating the signal model for any value of m_H . This way, m_H remains floating around the best estimate during the final fit, and no assumption is made on its true value. Since the MC simulations describing different H boson mass are available only for SM signals, the dependency of the signal model on m_H is kept identical for BSM processes.

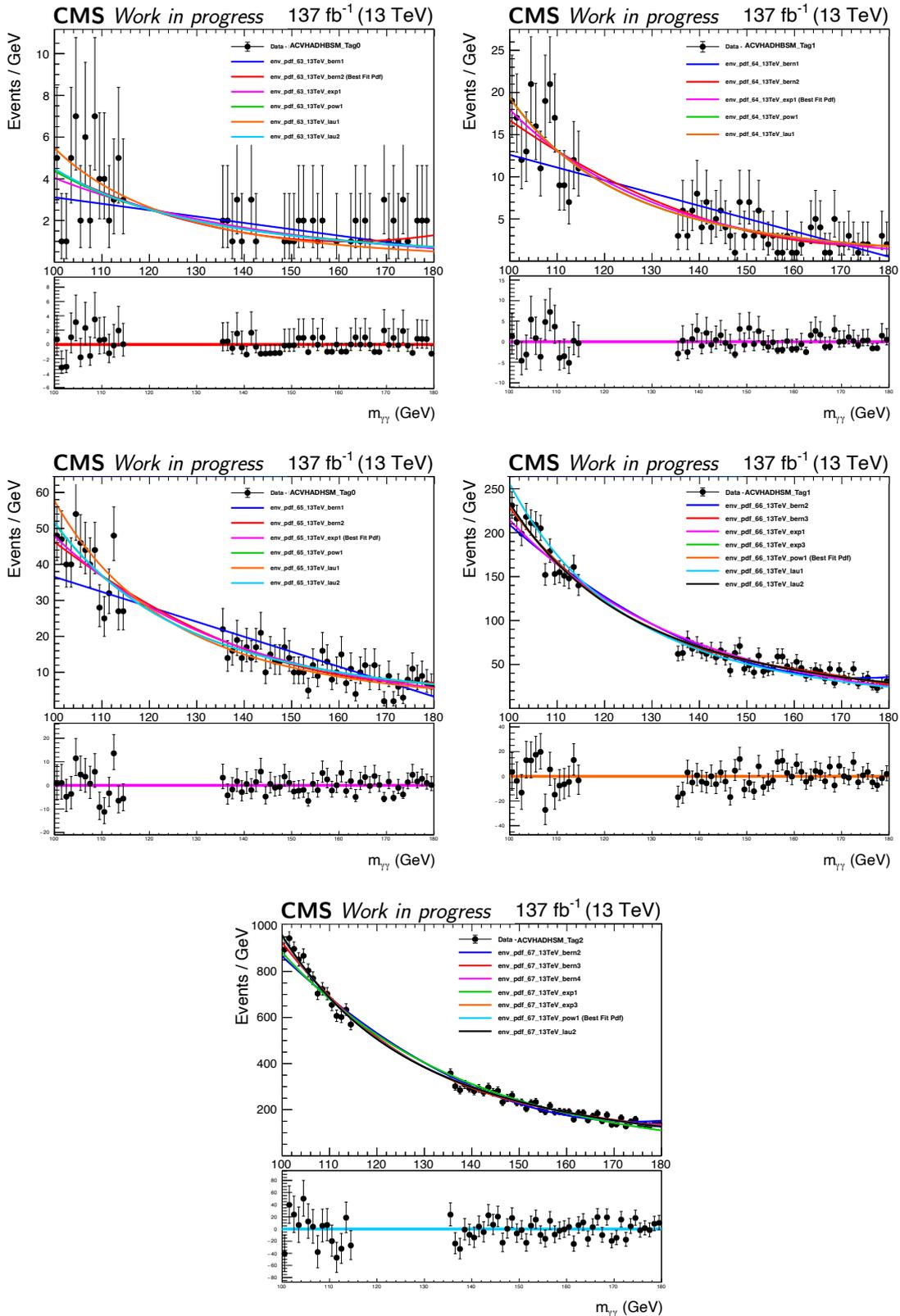


Figure V.17 – Representation of the background model candidates selected for each $V_{had}H$ category. The best fit in each category is indicated in the legend and is the one shown in the ratio below each plot. The association with categories of Section V.5 is (from left to right, top to bottom): BSM1, BSM2, SM1, SM2, SM3.

Finally, signal models are summed over the three data taking years weighted by their recorded luminosity to give the signal model of each signal process in each category that is used in the global likelihood ratio scan.

Application to $V_{\text{had}}H$ categories

Results of the identification of the optimal number of Gaussian functions to model some of the signal processes in the $V_{\text{had}}H$ categories are shown in Figs. V.18 and V.19. Figure V.18 shows the difference in signal distribution for the same process in the same category but with different vertex identification scenarios. The top row corresponds to events where the wrong vertex is associated with the diphoton pair. As mentioned already, it results in a worse mass resolution with more spread Gaussian functions compared to the right vertex identification shown on the bottom row. The number of Gaussian functions and the contribution of each Gaussian function to the signal model is shown in Fig. V.19 for some signal processes targeted by the different categories. The maximum number of Gaussian functions allowed in the fit is limited to 5 to avoid overfitting, but it is rarely a constraint as a good description of the signal ($\chi^2/n(\text{dof}) \approx 1$) is already achieved with fewer Gaussian functions.

The parameterisation of the signal models as a function of m_H are illustrated in Fig. V.20 for the SM WH process in the SM3 category. Three fits are performed with the signal model corresponding on the $m_H = 120$ GeV, 125 GeV, or 130 GeV hypotheses, allowing the determination of the evolution of the N_{exp} , σ , \mathcal{B} , $\epsilon \times A$, and RV fraction parameters as function of m_H (see left plot of Fig. V.20). From this evolution, a smooth interpolation of the signal models is derived over the potential range of the H boson mass (right plot of Fig. V.20).

The combination of the signal models of all signal processes for the three data-taking years is shown in Fig. V.21 for each of the $V_{\text{had}}H$ categories. A resolution on the $m_{\gamma\gamma}$ peak of approximately 1.3% is achieved. The signal models are also combined per year for all signal processes in all categories, as reported in Fig. V.22. The resolution per year is closer to 1.6% and is consistent for the three years.

V.6.4 Systematic uncertainties

As mentioned already in Section V.6.2, the systematic uncertainty associated with the background modelling on data sidebands is taken into account by the discrete profiling method [179]. Systematic uncertainties impacting the signal model are addressed differently if they affect the shape of the $m_{\gamma\gamma}$ distribution or just the event yield. In the latter case, the uncertainty is added as a log-normal deviation on the event yield. However, if an uncertainty is modifying the $m_{\gamma\gamma}$ distribution shape, it is included in the signal model as a nuisance parameter, potentially changing the parameters of the model and causing migration of events between categories.

These systematic uncertainties were derived for previous $H \rightarrow \gamma\gamma$ analyses [145], and their implementation in this analysis is mostly unchanged. Two sources of uncertainty are expected to bring most of the impact in $V_{\text{had}}H$ categories: the un-

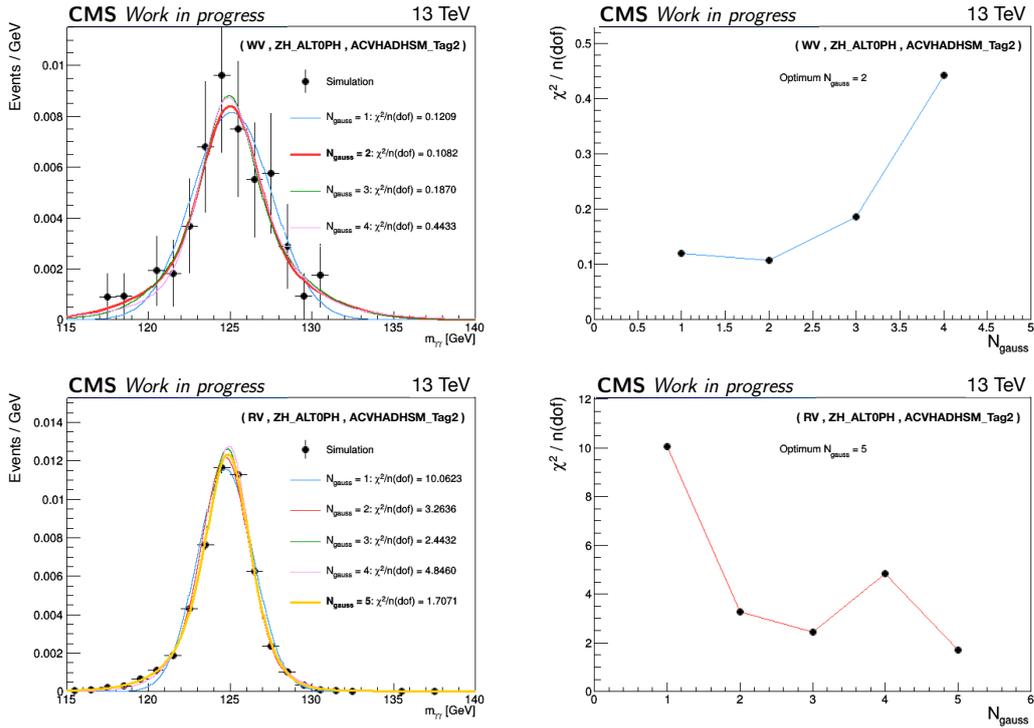


Figure V.18 – Illustration of the signal model building. The sum of Gaussian function fits is presented in the left column, and the evolution of the goodness-of-fit with an increasing number of Gaussian functions is on the right. Two vertex identification situations are shown for the same signal process in the same category. In the top row, the wrong vertex is associated with the diphoton pair, and in the bottom row, it is correctly identified.

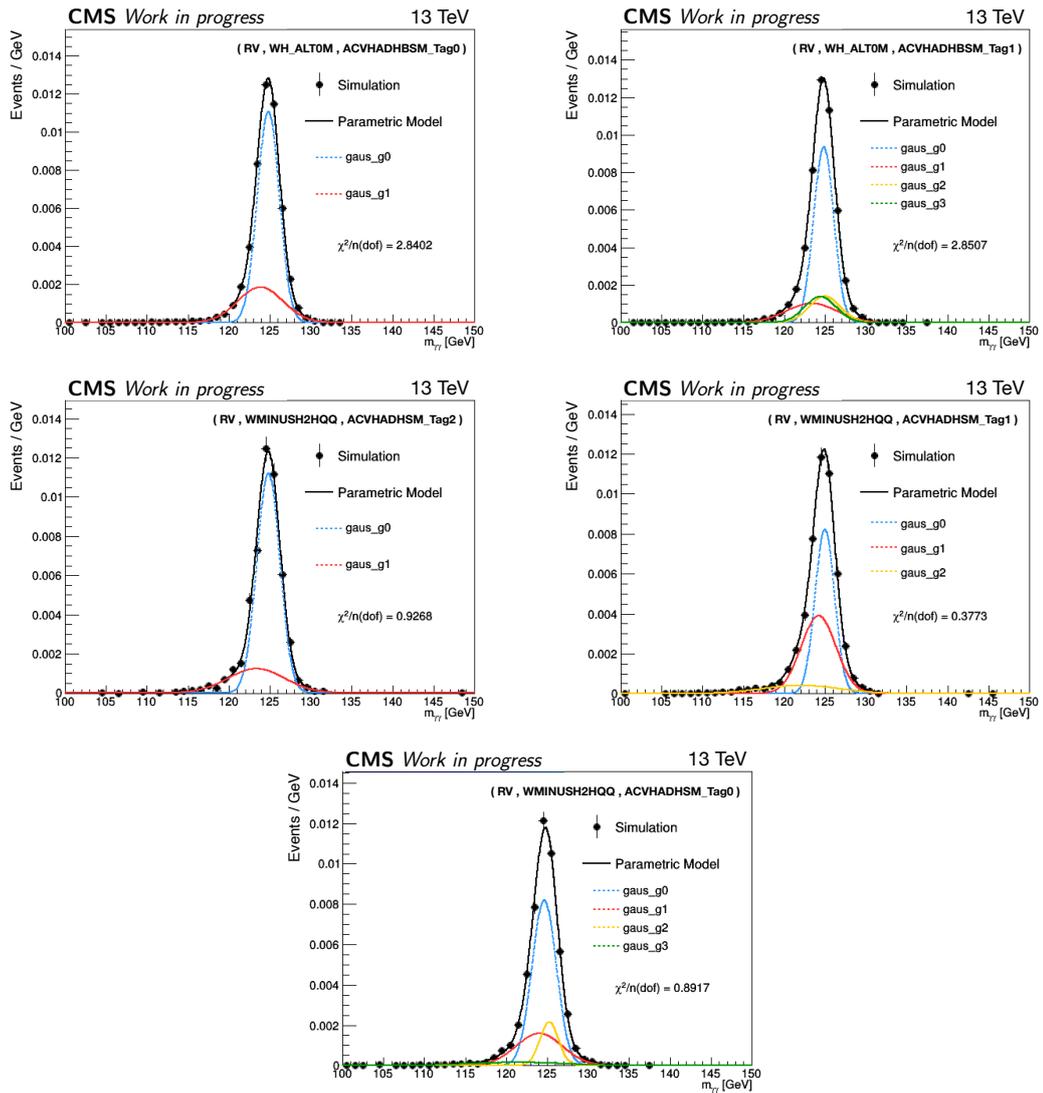


Figure V.19 – Contribution of each Gaussian function to the total parametric signal model for processes expected in the $V_{\text{had}}H$ categories. The association with categories of Section V.5 is (from left to right, top to bottom): BSM1, BSM2, SM1, SM2, SM3.

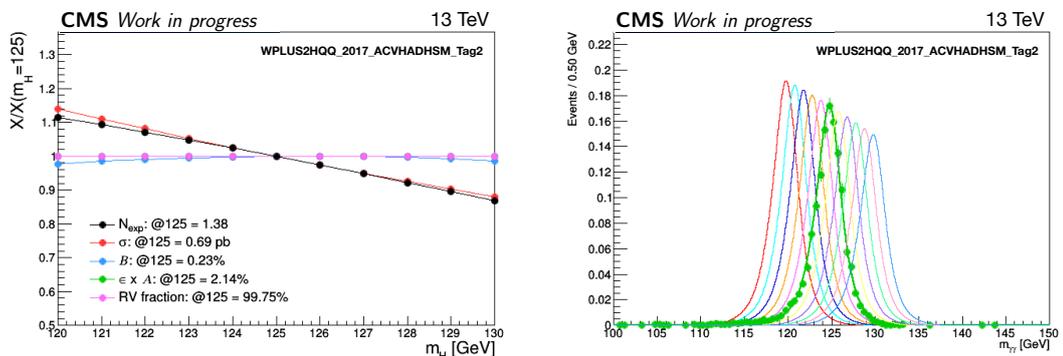


Figure V.20 – Left: Evolution of physical parameters extracted from the signal model as a function of m_H . Three mass hypotheses are fitted $m_H = 120$ GeV, 125 GeV, or 130 GeV. Right: Interpolation of the signal model with respect to m_H based on the parameters fitted for three different masses.

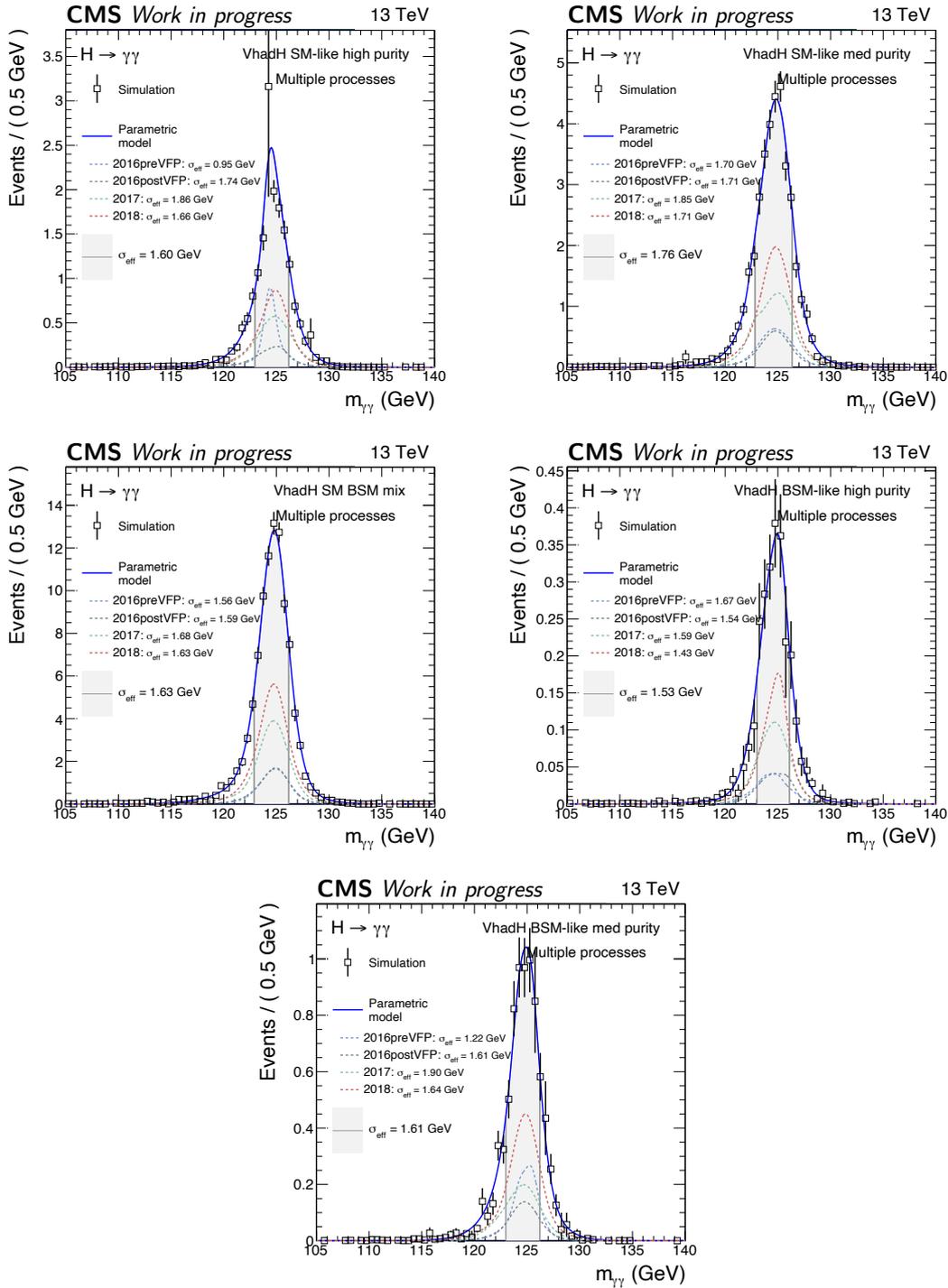


Figure V.21 – All signal models within a category are combined. The association with categories of Section V.5 is (from left to right, top to bottom): BSM1, BSM2, SM1, SM2, SM3.

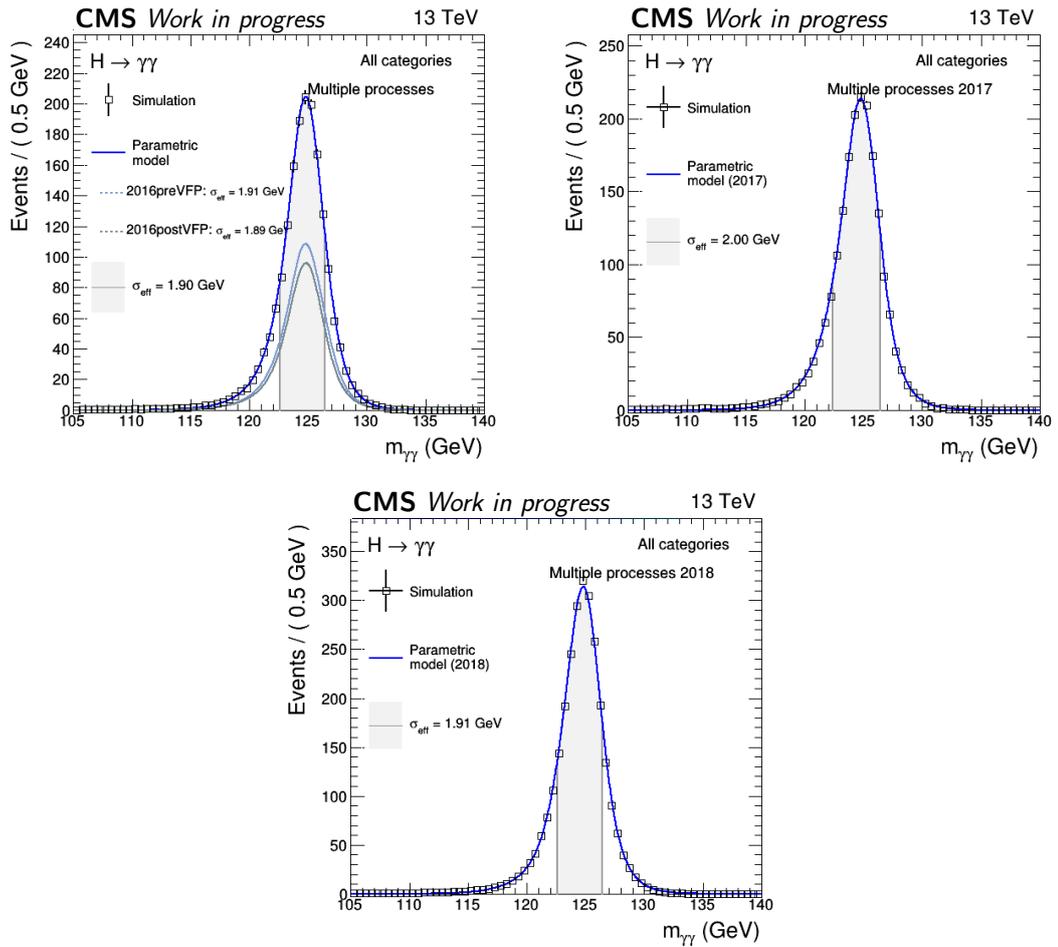


Figure V.22 – Combination of signal models per year.

certainty on the jet energy scale and smearing corrections and the uncertainty coming from the reweighting of JHUGEN samples. The first one is evaluated by using the p_T balance of jets with Z boson and photons in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ and γ +jets events, as well as the p_T balance between jets in dijet and multijet events. When propagating the uncertainty on the jet energy scale to the final result, an impact on the category yield of up to 22% can be observed. The uncertainty on the reweighting has an impact between 7% and 25% depending on the category and is detailed in a dedicated section after the description of the common systematic uncertainties for the $H \rightarrow \gamma\gamma$ analyses of the CMS collaboration.

Theoretical uncertainties

Theoretical predictions are needed to use the correct cross section, branching ratio, efficiency and acceptance in each category. Thus, uncertainties in these predictions lead to variations in the global normalisation of the event yield of the categories and also to the migration of events. A summary of theoretical uncertainties and their impact is reported in Table V.5.

Experimental uncertainties

Experimental uncertainties can impact the energy scale of individual photons and, thus, on the shape of the reconstructed mass of the diphoton system. A summary of experimental uncertainties impacting the shape of the $m_{\gamma\gamma}$ distribution is given in Table V.6. If the uncertainty affects only the event yield, it is described in Table V.7

Study of the uncertainty on the reweighting of JHUGEN BSM samples

The uncertainty on the reweighting of the JHUGEN BSM samples is implemented to account for the imperfection in the reweighting since, for a perfect match with an NLO sample, the generation of additional jets would be needed. To estimate this uncertainty, we compare the final reweighting of JHUGEN samples to alternative reweighting with other sets of variables. The final reweighting is done from the outputs of the DNN since the kinematic categories are built from these outputs. Several other reweightings were tested, and this final reweighting is compared to the best reweighting of the DNN output distributions we would reach using a set of physics variables.

To make sure that we are not also including variations due to the expected difference in the kinematics of SM and BSM samples, the reweighting from physics variables is selected from variables with distributions that are model-independent (the comparison is shown in Fig. V.7). Thus, we use a 2D reweighting based on the physics variables $(m_{jj}, N_{\text{jets}})$ as mentioned in Section V.3.2. The derivation procedure of this systematic uncertainties is summarised in Fig. V.23 and is done in two steps:

1. the comparison of the reweightings is done by computing the ratios, R_{SM} and R_{BSM} , of the DNN_x distributions obtained after reweighting $(m_{jj}, N_{\text{jets}})$ and reweighting $(\text{DNN}_{\text{bkg}}, \text{DNN}_{\text{bsm}})$ applied on the SM and BSM samples.

Uncertainty	Impact
QCD scale uncertainty: arises from variations of the renormalisation and factorisation scales when computing the expected SM cross section. Following recommendations from LHC Cross Section Working Group [173].	Overall normalisation varies between 0.5% and 15%. Migration of events around 1%.
ggH contamination: theoretical predictions for gluon fusion are less reliable in a regime where the Higgs boson is produced in association with jets.	Impact on the signal strength by about 2%.
qqH signal fraction: similar to ggH contamination but for VBF production. The uncertainty reflects also the migration of events to the region $m_{jj} < 250$ GeV.	Uncertainty of at most 8% in the SM prediction for cross section measurement.
Parton density functions: accounts for the uncertainty on which partons are most likely to initiate high energy events from proton-proton collisions [173, 180].	Overall normalisation uncertainty between 1% and 5%. Migration of events less than 1%.
QCD coupling constant: propagation of the uncertainty on the value of the strong interaction coupling α_S [180].	Impact on overall normalisation below 2.6%.
H $\rightarrow \gamma\gamma$ branching fraction: uncertainty on the value of the H $\rightarrow \gamma\gamma$ branching fraction [173].	Uncertainty on the SM prediction of 2%.
Underlying event and parton shower uncertainty: propagation of the uncertainty on the parameters used for MC generation.	Normalisation uncertainty around 5% (up to 30% for high- p_{TH} categories). Migration between 1% and 16%.

Table V.5 – Theoretical uncertainties

Uncertainty	Impact
Photon energy scale and resolution: uncertainty on the correction of the photon energy scale described in the photon definition of Section V.2.	Uncertainty on the energy scale between 0.05% and 3%.
Non-linearity of photon energy scale: uncertainty taking into account differences in the linearity of the photon energy scale between data and simulation.	Uncertainty on the energy scale between 0.2%.
Shower shape corrections: uncertainty on the correction of the shower shape described in the photon definition of Section V.2.	Uncertainty on the energy scale from 0.01% to 0.15%.
Non-uniformity of light collection: uncertainty from the model of light collection as a function of emission depth for a given ECAL crystal.	Uncertainty on the energy scale between 0.07% and 0.25%.
Modelling of material in front of the ECAL: covers the imperfect modelling of electromagnetic showers arising from the material upstream of the ECAL.	Uncertainty on the energy scale between 0.02% and 0.05% for central photons, less than 0.24% in the end-cap.
Vertex assignment: incorrect modelling of the underlying event leads to an uncertainty on the fraction of primary vertex correctly assigned.	Fraction of RV and WV in categories can vary by $\pm 2\%$.

Table V.6 – Experimental uncertainties impacting the shape of the $m_{\gamma\gamma}$ distribution.

Uncertainty	Impact
Integrated luminosity: uncertainty observed by the CMS luminosity monitoring.	Uncertainties of 2.5%, 2.3%, and 2.5% on the integrated luminosity for 2016, 2017, and 2018.
PhotonID BDT score: propagation of the uncertainty of the inputs on the training of the quantile regression correction described in Section V.2.	Impact on category yield below 3%.
Per-photon energy resolution: uncertainty on the per-photon resolution coming from the energy regression step.	Impact on category yield below 5%.
Trigger efficiency: uncertainty on the efficiency of the trigger selection measured with the tag-and-probe method.	Impact on category yield below 1.4%.
Photon preselection: account for the difference between preselection efficiency in data and in simulation.	Impact on category yield below 1%.
Pileup jet identification: uncertainty on the PU jet classification output score.	Impact on category yield around 1%.

Table V.7 – Experimental uncertainties impacting the event yields.

- the difference observed in the BSM case is compensated by the one observed in the SM case by computing the ratio $R_{\text{BSM}}/R_{\text{SM}}$. For a perfect reweighting, this ratio should be one. The deviation of this ratio to one is taken as systematic uncertainty ($+1\sigma$). This systematic uncertainty is also symmetrised around one (-1σ).

This uncertainty brings a shift between 7% and 25% on the expected number of BSM signal events, so it is one of the dominating systematic uncertainties in the $V_{\text{had}}H$ categories. While being included in the final analysis, the reweighting of the JHU samples based on DNN outputs and its associated uncertainty is not included at the current stage of preliminary results presented in this section.

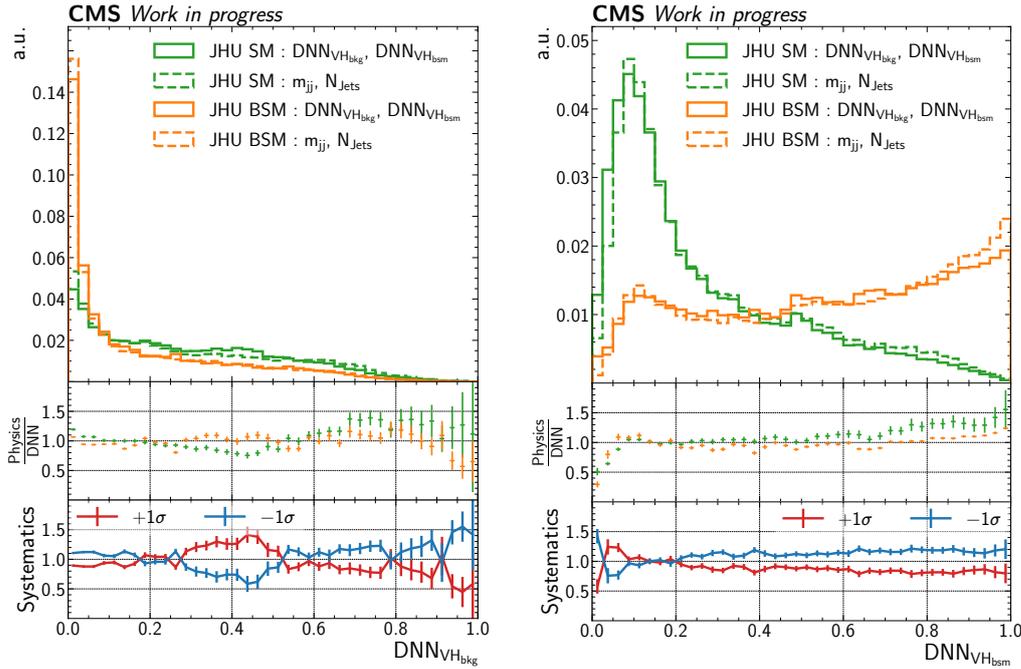


Figure V.23 – Decomposition of the derivation of the systematic uncertainty on reweighting of JHUGEN BSM samples. The top distributions show the effects of reweightings computed from the DNN outputs (solid histogram) or physics variables (dashed histogram) on the shape of the background (left) and BSM (right) outputs of the DNN. These reweightings are applied to SM (green) and BSM (orange) simulated events. The ratios below highlight the difference between the difference of the two reweightings, *i.e.* they are the ratio between the dashed histograms and the filled ones. Finally, the uncertainty is presented at the bottom as the $+1\sigma$ curve and is determined from the difference of reweightings on BSM samples compensated by the difference of reweightings on SM samples, *i.e.* orange over green from the ratio plot just above.

V.6.5 Results

Results of the analysis are extracted by scanning the binned likelihood ratio $-2\Delta \ln \mathcal{L}$ defined in Section V.6.1 for each parameter of interest ($f_{a2}, f_{a3}, f_{\Lambda 1}, f_{\Lambda 1}^{Z\gamma}$). The value of the likelihood function is computed from a fit of the combination of the background and signal models defined through Sections V.6.2 and V.6.3 and including the systematic uncertainties described in Section V.6.4 as nuisance parameters. The combination of signal and background models for the $V_{\text{had}}H$ categories over the three years is shown in Fig. V.24. The total signal and background models for all categories are also included in the figure. These combinations give a visual representation of the relative contribution of expected signal events compared to the background received in each category. While the expected signal yield in the $V_{\text{had}}H$ categories is relatively small (compared to inclusive categories, for instance), Fig. V.24 show that some of these categories have a promising signal purity.

The fit is performed with a common framework of the CMS experiment and is fitting simultaneously all categories of all production modes with the three years of data. Through the precise modelling of each signal process, this analysis is sensitive to the effective cross section fractions emerging from anomalous couplings of the H boson with gauge bosons $f_{a2}, f_{a3}, f_{\Lambda 1}, f_{\Lambda 1}^{Z\gamma}$. This thesis focuses on the expected results for the f_{a3} parameter. These results are extracted by fitting the likelihood ratio to an Asimov data set assuming the SM-only hypothesis (signal strengths are set to 1, m_H to 125.38 GeV, systematic uncertainties to their expected value and f_{a_i} are set to 0).

The results presented here are preliminary and may be subject to change as, for instance, the $V_{\text{lep}}H$ and $V_{\text{MET}}H$ categories are being integrated into the common analysis. To increase the sensitivity of the analysis, at least with respect to the SM background, production modes where categories were not optimised have default categories defined as in the STXS analysis [145].

Expected yields

The expected yields in categories optimised for this analysis are reported in Tables V.8 and V.9. Categories with the "AC" prefix are optimised and integrated to target a good sensitivity of the global analysis to anomalous couplings. Categories without this prefix are the STXS categories optimised on the SM-only signal yield versus background and thus provide some sensitivity to the effects of anomalous couplings. Table V.8 details yields for categories targeting the VH production mode while Table V.9 focuses on the VBF production. These tables include only the SM signals, hence the low purity in categories expecting contribution from BSM signals.

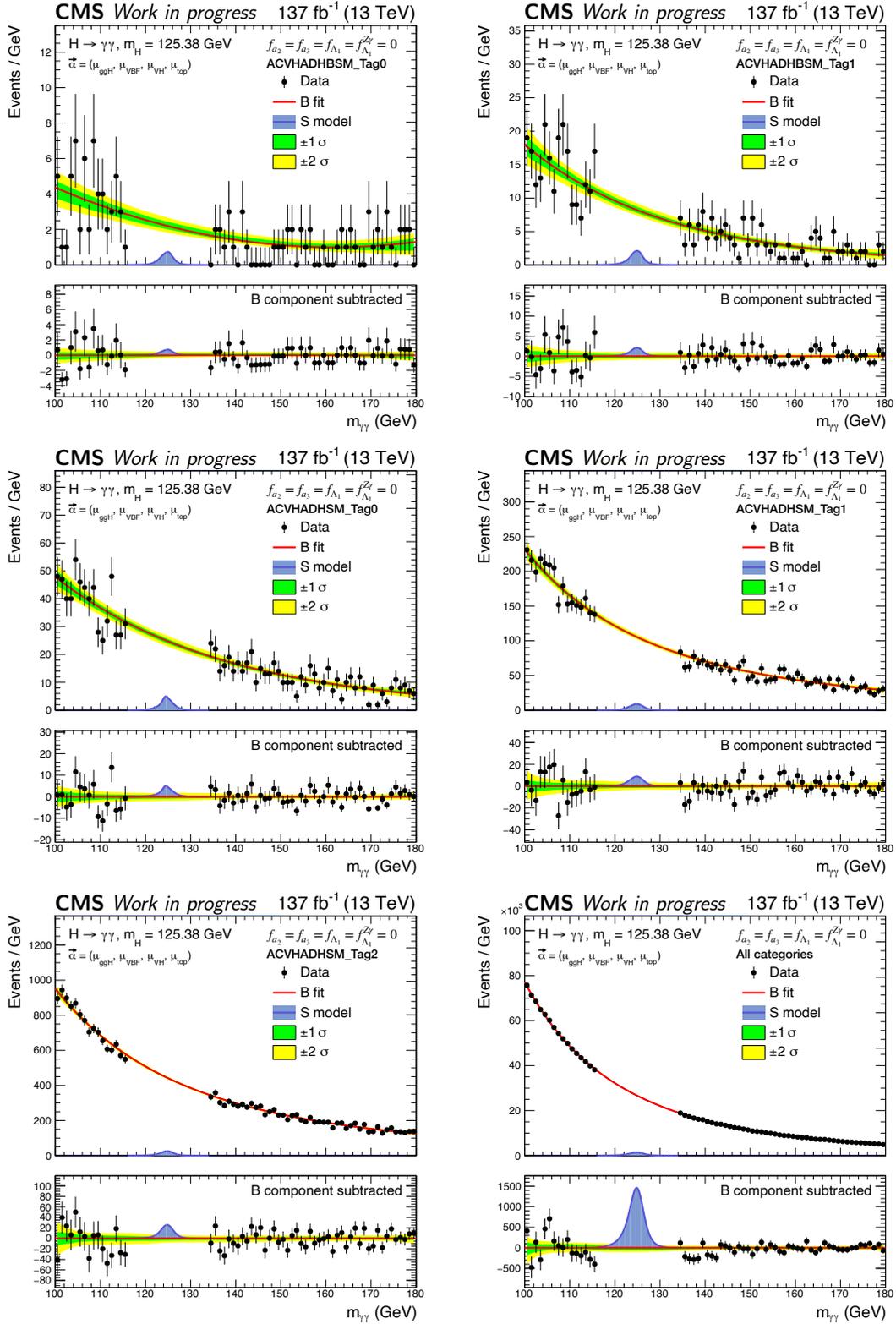


Figure V.24 – Combination of signal plus background models in categories of the $V_{\text{had}}H$ production mode and also shown when combining all categories (bottom right). The fit is performed on the blinded data, and the 68% (green) and 95% (yellow) bands are determined using toy MC simulations.

Analysis categories	SM 125 GeV Higgs boson expected signal			
	Total	Production Mode Fractions VH	σ_{eff} (GeV)	S/S+B
AC_VHADH_BSM_Tag0	2.7	27.0%	1.57	0.21
AC_VHADH_BSM_Tag1	7.9	31.6%	1.65	0.17
AC_VHADH_SM_Tag0	17.0	50.8%	1.63	0.13
AC_VHADH_SM_Tag1	37.0	36.4%	1.77	0.06
AC_VHADH_SM_Tag2	98.5	17.2%	1.64	0.04
VH_MET_Tag0	1.9	98.4%	2.00	0.46
VH_MET_Tag1	3.0	95.4%	2.03	0.37
VH_MET_Tag2	5.1	85.3%	1.98	0.19
WH_LEP_PTV_0_75_Tag0	1.8	99.2%	1.80	0.45
WH_LEP_PTV_0_75_Tag1	4.1	98.0%	1.92	0.19

Table V.8 – Expected event yields for categories targeting the VH production mode. Results are separated by: total expected event yield in the category (Total), fraction of signal events coming from the VH SM production, resolution on the $m_{\gamma\gamma}$ peak (σ_{eff}), and estimation of the purity of the category as S/S+B where S and B are the number of expected SM signal and background events within $\pm 1\sigma_{\text{eff}}$ of m_H .

Analysis categories	SM 125 GeV Higgs boson expected signal			
	Total	Production Mode Fractions VBF	σ_{eff} (GeV)	S/S+B
AC_GGH_Tag0	117.6	42.9%	1.89	0.06
AC_GGH_Tag1	54.2	27.3%	1.79	0.04
AC_VBF_BSM_Tag0	5.3	28.1%	1.49	0.35
AC_VBF_BSM_Tag1	26.1	70.6%	1.71	0.40
AC_VBF_SM_Tag0	77.0	78.0%	1.79	0.36

Table V.9 – Expected event yields for categories targeting the VBF production mode. Results are separated by: total expected event yield in the category (Total), fraction of signal events coming from the VBF SM production, resolution on the $m_{\gamma\gamma}$ peak (σ_{eff}), and estimation of the purity of the category as S/S+B where S and B are the number of expected SM signal and background events within $\pm 1\sigma_{\text{eff}}$ of m_H .

Anomalous couplings

Expected results on f_{a3} are obtained by scanning values of the f_{a3} parameter and setting other f_{a_i} to 0 as a first approach, as a simultaneous scan can also be considered. The likelihood profile is provided in Fig. V.25, and the expected constraints on f_{a3} are reported in Table V.10. Figure V.25 describes three likelihood scans performed over three different sets of categories. The green one does not include the $V_{\text{had}}H$ and VBF categories; only the categories for other production modes are taken into account, so it corresponds to the categories targeting the ggH , $t\bar{t}H$, $V_{\text{lep}}H$, and $V_{\text{MET}}H$ production modes. For these production modes, the definition of the categories of the STXS is used. These categories are defined to probe the complete phase space of production of the Higgs boson. In particular, some of these categories target events with a high reconstructed p_T for the H boson, and they already bring some sensitivity to the measurement of f_{a3} . The blue curve is the likelihood scan, including VBF categories. They are optimised to enhance the sensitivity to anomalous couplings, so they bring a significant contribution in constraining f_{a3} . Finally, the additional improvement in sensitivity brought by the $V_{\text{had}}H$ categories is demonstrated through the black curve. We defined and optimised these categories to discern effects from the anomalous contribution as well, and the high purity of some of the categories helps constrain further possible values of f_{a3} . The constraints at 68% and 95% confidence level (CL) are extracted for $-2\Delta \ln \mathcal{L}$ values reaching 0.99 or 3.84 respectively. Thus, values outside of the dashed lines on Fig. V.25 correspond to the ranges expected to be excluded at the 95% CL. From these projections, if the unblinded data are compatible with an SM-only hypothesis, the analysis will be able to rule out the pure CP-odd hypothesis ($|f_{a3}| = 1$) and set tight constraints on the possible values of f_{a3} .

The sensitivity of this analysis can be compared to anomalous coupling analyses already performed for other decay channels of the H boson. In particular, two analyses studied the full data set collected by the CMS experiment during the three years of the Run2 in the $H \rightarrow ZZ$ [27] and $H \rightarrow \tau\tau$ [154] decay channels. The $H \rightarrow \tau\tau$ analysis also includes combined results with the $H \rightarrow ZZ$ channel. The three sets of expected results are extracted from Refs. [27, 154] and reported in Table V.11. As the study of anomalous couplings through the $H \rightarrow \gamma\gamma$ decay channel shows comparable sensitivity, its combination with the results of these two analyses could tighten the constraints set on the values of anomalous couplings.

V.7 Conclusion

This last chapter describes the strategy and expected results for the analysis of the Higgs boson anomalous couplings with electroweak gauge bosons using the Higgs boson decay in two photons and the full Run2 data set recorded by the CMS experiment. In particular, a focus is given to my involvement in the design of the analysis strategy targeting the $V_{\text{had}}H$ production mode of the Higgs boson. While following a common global approach of preselecting events, producing a discriminant, and optimising categories, the strategy relies on non-standard

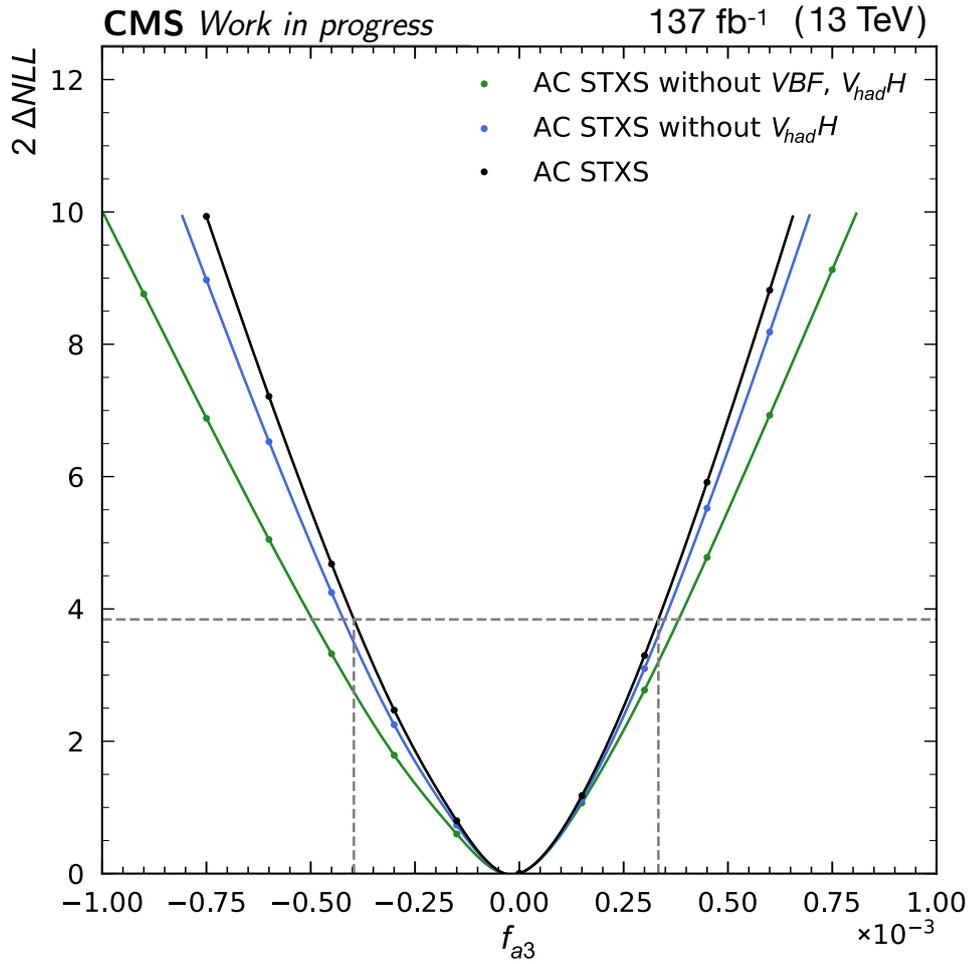


Figure V.25 – Likelihood scan of the f_{a3} parameter. Different configurations are shown to compare improvements brought by the optimised categories for the VBF and $V_{had}H$ production modes. In green, the likelihood is computed without the VBF and $V_{had}H$ categories. In blue, the VBF categories are included. The final likelihood scan using all categories (including $V_{had}H$ ones) is represented by the black curve.

Parameter	Expected constraints	
	68% CL	95% CL
f_{a3}	$[-1.7, 1.3] \times 10^{-4}$	$[-4.0, 3.3] \times 10^{-4}$

Table V.10 – Expected constraints on f_{a3} at 68% and 95% CL.

Analysis	Expected intervals for f_{a3}	
	68% CL	95% CL
H \rightarrow ZZ	$[-8.1, 8.1] \times 10^{-4}$	$[-41.2, 41.2] \times 10^{-4}$
H $\rightarrow \tau\tau$	$[-0.6, 0.6] \times 10^{-4}$	$[-2.3, 2.3] \times 10^{-4}$
Combined	$[-0.5, 0.5] \times 10^{-4}$	$[-2.1, 2.1] \times 10^{-4}$

Table V.11 – Expected constraints on f_{a3} at 68% and 95% CL for anomalous couplings analyses in the H \rightarrow ZZ [27] and H $\rightarrow \tau\tau$ [154] decay channels.

steps. I developed and trained a multiclassifier DNN to classify events between SM-like, BSM-like, and background-like classes. I also demonstrated that using a multiclassifier improved the sensitivity to f_{a3} while training only one DNN, compared to a more standard approach of training n-1 binary classifiers to separate events between n classes. Then, from the output scores of the DNN, I implemented a way to extract an estimator on the upper limit of f_{a3} . It allows the optimisation of the number, shape and definition of categories to enhance the final sensitivity of the analysis.

Finally, after combining the $V_{\text{had}}H$ categories to the common framework, I extracted the expected results on f_{a3} by applying common statistical inference steps and using the optimised categories for the $V_{\text{had}}H$ process. The expected results are computed with different sets of categories, and they show that the $V_{\text{had}}H$ categories bring additional constraints to the expected results of f_{a3} . The sensitivity of the analysis is comparable to anomalous couplings analyses targeting other decay channels, and thus, promising results with tight constraints on the anomalous couplings are expected from the unblinded fit to data.

Conclusion

With this thesis, I have presented my contribution to enhancing the $H \rightarrow \gamma\gamma$ analysis of the CMS experiment, notably by proposing deep learning methods at different steps of the global analysis framework. I have also described my participation in the timing calibration of the CMS detector in the context of its future upgrades through the development of a simulation framework to investigate monitoring and correction possibilities of the high-precision clock distribution.

The general theoretical context regarding the SM and the current experimental status of the Higgs boson properties are briefly covered by the first chapter of this thesis, allowing an introduction of the motivations for particle collider experiments and how they deepen our knowledge of fundamental laws of the Universe. The particular experimental context of this thesis, relying on data of proton-proton collisions produced by the LHC and recorded by the CMS experiment, is described in Chapter II. This description also covers upgrades planned for the CMS detector to adapt to a high-luminosity phase of the LHC. In particular, the CMS detector will rely on the timing information of particles to mitigate the effects of a harsher data-taking environment. In this context, the simulation framework described in Section II.3, *pyDDMTD*, provides a description of a DDMTD based monitoring of the clock to control the jitter introduced in the distribution chain and ensure a precise synchronisation of all detector components. From the results of simple simulation studies, a basic DDMTD based solution could already bring some corrections of the jitter with frequency components above 10 Hz. The *pyDDMTD* simulation tool could now be used to investigate and compare different correction strategies based on DDMTD systems.

My contributions to the $H \rightarrow \gamma\gamma$ analysis of the CMS experiment are presented in Chapters IV and V. Since they employ uncommon machine learning methods, a general introduction to deep learning is provided in Chapter III. Indeed, Chapter IV introduces a method based on advanced deep learning models: GANs. This GAN is implemented to generate misidentified photons from events of a control region in data to produce an accurate background sample for the $H \rightarrow \gamma\gamma$ analysis. An evaluation strategy was defined to assess the performance of the GAN, and the GAN is shown to learn how to generate convincing misidentified photons whose properties are correctly correlated with the rest of the event. A comparison study between the GAN-based method and the method used in

the last published $H \rightarrow \gamma\gamma$ analyses shows that BDTs trained with the GAN sample perform better in separating background and signal events for the $H \rightarrow \gamma\gamma$ analysis.

Finally, this thesis presents preliminary results of an $H \rightarrow \gamma\gamma$ analysis studying the couplings of the Higgs boson with the weak gauge bosons (Z and W) to determine if there is any contribution from anomalous couplings not expected by the SM and in particular CP-odd couplings. Small anomalous couplings could indicate a new source of CP-violation in the SM and signify that a fraction (f_{a3}) of the observed signal events have a CP-odd behaviour. The complete Run 2 dataset is analysed, corresponding to 137 fb^{-1} of data collected with the CMS detector in proton-proton collisions with a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. My contribution to this analysis concerns the design of the analysis strategy targeting events with a production of the Higgs boson in association with a Z or W boson which decays hadronically. This strategy includes a DNN multiclassifier, which relies on the properties of the events to class them as background, SM-like signal or anomalous-like signal. The preliminary results are reported, with expected constraints on f_{a3} of $[-0.17 \times 10^{-3}, 0.13 \times 10^{-3}]$ at the 68% confidence level. Therefore, tight constraints can be expected for the final results.

Bibliography

- [1] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Boulder, CO: Westview, 1995.
- [2] Carsten Burgard. *Standard model of physics: TikZ example*. url: <https://texample.net/tikz/examples/model-physics/> (visited on Aug. 8, 2023).
- [3] Particle Data Group, R. L. Workman, et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. doi: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [4] Sheldon L. Glashow. "The renormalizability of vector meson interactions". In: *Nuclear Physics* 10 (1959), pp. 107–117. issn: 0029-5582. doi: [10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8).
- [5] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19.21 (Nov. 1967), pp. 1264–1266. doi: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [6] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. doi: [10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).
- [7] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96.1 (Oct. 1954), pp. 191–195. doi: [10.1103/PhysRev.96.191](https://doi.org/10.1103/PhysRev.96.191).
- [8] Enrico Fermi. "Tentativo di una Teoria Dei Raggi β ". In: *Il Nuovo Cimento (1924-1942)* 11.1 (Jan. 1, 1934), pp. 1–19. issn: 1827-6121. doi: [10.1007/BF02959820](https://doi.org/10.1007/BF02959820).
- [9] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13.9 (Aug. 31, 1964), pp. 321–323. doi: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [10] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (Oct. 19, 1964), pp. 508–509. doi: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [11] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12.2 (Sept. 15, 1964), pp. 132–133. issn: 0031-9163. doi: [10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [12] John Ellis. "Higgs Physics". In: (2015), pp. 117–168. doi: [10.5170/CERN-2015-004.117](https://doi.org/10.5170/CERN-2015-004.117).
- [13] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 17, 2012), pp. 1–29. issn: 0370-2693. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).

- [14] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 17, 2012), pp. 30–61. issn: 0370-2693. doi: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [15] The CMS collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *Journal of High Energy Physics* 2013.6 (June 20, 2013), p. 81. issn: 1029-8479. doi: [10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081).
- [16] The CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607.7917 (July 2022), pp. 60–68. issn: 1476-4687. doi: [10.1038/s41586-022-04892-x](https://doi.org/10.1038/s41586-022-04892-x).
- [17] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (June 10, 2020), p. 135425. issn: 0370-2693. doi: [10.1016/j.physletb.2020.135425](https://doi.org/10.1016/j.physletb.2020.135425).
- [18] The ATLAS collaboration. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". In: *Journal of High Energy Physics* 2016.8 (Aug. 5, 2016), p. 45. issn: 1029-8479. doi: [10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045).
- [19] CMS Collaboration. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". In: *Phys. Lett. B* 779 (2018), pp. 283–316. doi: [10.1016/j.physletb.2018.02.004](https://doi.org/10.1016/j.physletb.2018.02.004).
- [20] CMS Collaboration. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. doi: [10.1103/PhysRevLett.121.121801](https://doi.org/10.1103/PhysRevLett.121.121801).
- [21] CMS Collaboration. "Evidence for Higgs boson decay to a pair of muons. Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 2101 (2021), p. 148. doi: [10.1007/JHEP01\(2021\)148](https://doi.org/10.1007/JHEP01(2021)148).
- [22] 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 13, 2015), p. 012004. doi: [10.1103/PhysRevD.92.012004](https://doi.org/10.1103/PhysRevD.92.012004).
- [23] Lev Davidovich Landau. "On the angular momentum of a system of two photons". In: *Dokl. Akad. Nauk SSSR* 60 (1948), p. 207.
- [24] Chen Ning Yang. "Selection rules for the dematerialization of a particle into two photons". In: *Phys. Rev.* 77.2 (1950), pp. 242–245.
- [25] CMS Collaboration. "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel". In: *Physical Review Letters* 125.6 (Aug. 5, 2020), p. 061801. doi: [10.1103/PhysRevLett.125.061801](https://doi.org/10.1103/PhysRevLett.125.061801).
- [26] CMS Collaboration. "Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs". In: *Physical Review D* 100.11 (Dec. 4, 2019), p. 112002. doi: [10.1103/PhysRevD.100.112002](https://doi.org/10.1103/PhysRevD.100.112002).
- [27] 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. 17, 2021), p. 052004. doi: [10.1103/PhysRevD.104.052004](https://doi.org/10.1103/PhysRevD.104.052004).

- [28] Ian Aitchison. *Supersymmetry in Particle Physics: An Elementary Introduction*. 1st ed. Cambridge University Press, Sept. 20, 2007. isbn: 978-0-521-88023-7. doi: [10.1017/CB09780511619250](https://doi.org/10.1017/CB09780511619250).
- [29] Andrea Taroni. “Nobel Prize 2015: Kajita and McDonald”. In: *Nature Physics* 11.11 (Nov. 2015), pp. 891–891. issn: 1745-2481. doi: [10.1038/nphys3543](https://doi.org/10.1038/nphys3543).
- [30] Muon g-2 Collaboration. “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm”. In: (2023). url: <https://muon-g-2.fnal.gov/result2023.pdf>.
- [31] Muon g-2 Collaboration. “The anomalous magnetic moment of the muon in the Standard Model”. In: *Physics Reports*. The anomalous magnetic moment of the muon in the Standard Model 887 (Dec. 3, 2020), pp. 1–166. issn: 0370-1573. doi: [10.1016/j.physrep.2020.07.006](https://doi.org/10.1016/j.physrep.2020.07.006).
- [32] John R. Ellis. “Dark Matter and Dark Energy: Summary and Future Directions”. In: *Philos. Trans. R. Soc. Lond. A* 361 (2003), p. 2607. doi: [10.1098/rsta.2003.1297](https://doi.org/10.1098/rsta.2003.1297).
- [33] Andrei D. Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (May 31, 1991), p. 392. issn: 0038-5670. doi: [10.1070/PU1991v034n05ABEH002497](https://doi.org/10.1070/PU1991v034n05ABEH002497).
- [34] Ewa Lopienska. “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”. In: (2022). url: <https://cds.cern.ch/record/2800984>.
- [35] CMS Collaboration. *LumiPublicResults*. url: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [36] Tai Sakuma. “Cutaway diagrams of CMS detector”. In: (2019). url: <https://cds.cern.ch/record/2665537>.
- [37] *University of Zurich: CMS Wiki Pages*. url: <https://wiki.physik.uzh.ch/cms/latex:tikz> (visited on Aug. 2, 2023).
- [38] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. 1997. url: <https://cds.cern.ch/record/368412>.
- [39] *The CMS tracker: addendum to the Technical Design Report*. 2000. url: <https://cds.cern.ch/record/490194>.
- [40] CMS Collaboration. “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment”. In: *JINST* 3 (2008), S08004. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [41] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997.
- [42] A. A Annenkov, M. V Korzhik, and P Lecoq. “Lead tungstate scintillation material”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 490.1 (Sept. 1, 2002), pp. 30–50. issn: 0168-9002. doi: [10.1016/S0168-9002\(02\)00916-6](https://doi.org/10.1016/S0168-9002(02)00916-6).
- [43] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. Geneva: CERN, 2006. url: <https://cds.cern.ch/record/922757>.

- [44] CMS Collaboration. "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV". In: *Journal of Instrumentation* 8.9 (Sept. 2013), P09009. issn: 1748-0221. doi: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009).
- [45] CMS Collaboration. *Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter*. Geneva: CERN, 2007. doi: [10.1088/1748-0221/2/04/P04004](https://doi.org/10.1088/1748-0221/2/04/P04004).
- [46] CMS Collaboration. "ECAL 2016 refined calibration and Run2 summary plots". In: (2020). url: <https://cds.cern.ch/record/2717925>.
- [47] M. Anfreville et al. "Laser monitoring system for the CMS lead tungstate crystal calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 594.2 (Sept. 1, 2008), pp. 292–320. issn: 0168-9002. doi: [10.1016/j.nima.2008.01.104](https://doi.org/10.1016/j.nima.2008.01.104).
- [48] Bannaje Sripathi Acharya et al. *The CMS Outer Hadron Calorimeter*. Geneva: CERN, 2006. url: <http://cds.cern.ch/record/973131>.
- [49] CMS HCAL/ECAL Collaborations. "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c". In: *The European Physical Journal C* 60.3 (Apr. 1, 2009), pp. 359–373. issn: 1434-6052. doi: [10.1140/epjc/s10052-009-0959-5](https://doi.org/10.1140/epjc/s10052-009-0959-5).
- [50] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. doi: [10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002).
- [51] The CMS collaboration. "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *Journal of Instrumentation* 8.11 (Nov. 2013), P11002. issn: 1748-0221. doi: [10.1088/1748-0221/8/11/P11002](https://doi.org/10.1088/1748-0221/8/11/P11002).
- [52] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. issn: 1748-0221. doi: [10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [53] Siona Ruth Davis. "Interactive Slice of the CMS detector". In: (2016). url: <http://cds.cern.ch/record/2205172>.
- [54] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10 (2015), P06005. doi: [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [55] CMS Collaboration. "The CMS trigger system". In: *Journal of Instrumentation* 12.1 (Jan. 2017), P01020. issn: 1748-0221. doi: [10.1088/1748-0221/12/01/P01020](https://doi.org/10.1088/1748-0221/12/01/P01020).
- [56] A Tapper, Darin Acosta, and A Tapper. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. 2013. url: <https://cds.cern.ch/record/1556311>.
- [57] Ezio Todesco et al. "Progress on high luminosity LHC Nb3Sn magnets". In: *IEEE Trans. Appl. Supercond.* 28.4 (2018), p. 4008809. doi: [10.1109/TASC.2018.2830703](https://doi.org/10.1109/TASC.2018.2830703).
- [58] Rama Calaga. "Crab Cavities for the High-luminosity LHC". In: (2018), THXA03. doi: [10.18429/JACoW-SRF2017-THXA03](https://doi.org/10.18429/JACoW-SRF2017-THXA03).
- [59] Collaboration CMS. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. 2019. url: <https://cds.cern.ch/record/2667167>.
- [60] D Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Geneva, 2015. doi: [10.17181/CERN.VU8I.D59J](https://doi.org/10.17181/CERN.VU8I.D59J).

- [61] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Geneva: CERN, 2017. doi: [10.17181/CERN.IV8M.1JY2](https://doi.org/10.17181/CERN.IV8M.1JY2).
- [62] CMS Collaboration. *The Phase-2 Upgrade of the CMS Level-1 Trigger*. Geneva: CERN, 2020. url: <https://cds.cern.ch/record/2714892>.
- [63] CMS Collaboration. *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger*. Geneva: CERN, 2021. url: <https://cds.cern.ch/record/2759072>.
- [64] Collaboration CMS. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. 2017. doi: [10.17181/CERN.IV8M.1JY2](https://doi.org/10.17181/CERN.IV8M.1JY2).
- [65] Eduardo Mendes et al. "TCLink: A Fully Integrated Open Core for Timing Compensation in FPGA-Based High-Speed Links". In: *IEEE Transactions on Nuclear Science* 70.2 (Feb. 2023), pp. 156–163. issn: 1558-1578. doi: [10.1109/TNS.2023.3240539](https://doi.org/10.1109/TNS.2023.3240539).
- [66] S. Baron et al. "Jitter impact on clock distribution in LHC experiments". In: *Journal of Instrumentation* 7.12 (Dec. 2012), p. C12023. issn: 1748-0221. doi: [10.1088/1748-0221/7/12/C12023](https://doi.org/10.1088/1748-0221/7/12/C12023).
- [67] Stefan Biereigel et al. "The lpGBT PLL and CDR Architecture, Performance and SEE Robustness". In: *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2019)*. Vol. 370. 2020, p. 034. doi: [10.22323/1.370.0034](https://doi.org/10.22323/1.370.0034).
- [68] D.W. Allan and H. Daams. "Picosecond Time Difference Measurement System". In: *29th Annual Symposium on Frequency Control*. 29th Annual Symposium on Frequency Control. May 1975, pp. 404–411. doi: [10.1109/FREQ.1975.200112](https://doi.org/10.1109/FREQ.1975.200112).
- [69] Pedro Moreira et al. "Digital dual mixer time difference for sub-nanosecond time synchronization in Ethernet". In: *2010 IEEE International Frequency Control Symposium*. 2010 IEEE International Frequency Control Symposium. June 2010, pp. 449–453. doi: [10.1109/FREQ.2010.5556289](https://doi.org/10.1109/FREQ.2010.5556289).
- [70] Montgomery Phister. *Logical Design of Digital Computers*. Wiley, 1958. 432 pp. isbn: 978-0-608-10265-8.
- [71] *High Precision Timing Distribution project*. url: https://espace.cern.ch/HighPrecisionTiming/_layouts/15/start.aspx#/SitePages/Home.aspx (visited on July 27, 2023).
- [72] Pedro Moreira et al. "Sub-nanosecond digital phase shifter for clock synchronization applications". In: *2012 IEEE International Frequency Control Symposium Proceedings*. 2012 IEEE International Frequency Control Symposium Proceedings. May 2012, pp. 1–6. doi: [10.1109/FCS.2012.6243715](https://doi.org/10.1109/FCS.2012.6243715).
- [73] Warren S. McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1, 1943), pp. 115–133. issn: 1522-9602. doi: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259).
- [74] N. Rochester et al. "Tests on a cell assembly theory of the action of the brain, using a large digital computer". In: *IRE Transactions on Information Theory* 2.3 (Sept. 1956), pp. 80–93. issn: 2168-2712. doi: [10.1109/TIT.1956.1056810](https://doi.org/10.1109/TIT.1956.1056810).
- [75] B. Farley and W. Clark. "Simulation of self-organizing systems by digital computer". In: *Transactions of the IRE Professional Group on Information Theory* 4.4 (Sept. 1954), pp. 76–84. issn: 2168-2704. doi: [10.1109/TIT.1954.1057468](https://doi.org/10.1109/TIT.1954.1057468).

- [76] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Psychological Review* 65.6 (1958), pp. 386–408. issn: 1939-1471. doi: [10.1037/h0042519](https://doi.org/10.1037/h0042519).
- [77] V. N. Vapnik and A. Ya. Chervonenkis. "A class of algorithms for pattern recognition learning". In: *Avtomatika i Telemekhanika* 25.6 (1964), pp. 937–945.
- [78] T. Cover and P. Hart. "Nearest neighbor pattern classification". In: *IEEE Transactions on Information Theory* 13.1 (Jan. 1967), pp. 21–27. issn: 1557-9654. doi: [10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964).
- [79] Marvin Minsky and Seymour A. Papert. *Perceptrons: an introduction to computational geometry*. 2. print. with corr. Cambridge/Mass.: The MIT Press, 1972. 258 pp. isbn: 978-0-262-63022-1.
- [80] Kunihiko Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4 (Apr. 1, 1980), pp. 193–202. issn: 1432-0770. doi: [10.1007/BF00344251](https://doi.org/10.1007/BF00344251).
- [81] J J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the National Academy of Sciences* 79.8 (Apr. 1982), pp. 2554–2558. doi: [10.1073/pnas.79.8.2554](https://doi.org/10.1073/pnas.79.8.2554).
- [82] Paul J. Werbos. "Applications of advances in nonlinear sensitivity analysis". In: *System Modeling and Optimization*. Ed. by R. F. Drenick and F. Kozin. Lecture Notes in Control and Information Sciences. Berlin, Heidelberg: Springer, 1982, pp. 762–770. isbn: 978-3-540-39459-4. doi: [10.1007/BFb0006203](https://doi.org/10.1007/BFb0006203).
- [83] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. issn: 1476-4687. doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [84] Y. LeCun et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1.4 (Dec. 1989), pp. 541–551. issn: 0899-7667. doi: [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541).
- [85] Seppo Linnainmaa. "Taylor expansion of the accumulated rounding error". In: *BIT Numerical Mathematics* 16.2 (June 1, 1976), pp. 146–160. issn: 1572-9125. doi: [10.1007/BF01931367](https://doi.org/10.1007/BF01931367).
- [86] Tin Kam Ho. "Random decision forests". In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Proceedings of 3rd International Conference on Document Analysis and Recognition. Vol. 1. Aug. 1995, 278–282 vol.1. doi: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994).
- [87] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (Aug. 1, 1997), pp. 119–139. issn: 0022-0000. doi: [10.1006/jcss.1997.1504](https://doi.org/10.1006/jcss.1997.1504).
- [88] Y. Lecun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (Nov. 1998), pp. 2278–2324. issn: 1558-2256. doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).

- [89] Jia Deng et al. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009 IEEE Conference on Computer Vision and Pattern Recognition. June 2009, pp. 248–255. doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- [90] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (Oct. 28, 2017).
- [91] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015.
- [92] Md Syadus Sefat et al. "Accelerating HotSpots in Deep Neural Networks on a CAPI-Based FPGA". In: *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. Aug. 2019, pp. 248–256. doi: [10.1109/HPCC/SmartCity/DSS.2019.00048](https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00048).
- [93] Frederico A.C. Azevedo et al. "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain". In: *Journal of Comparative Neurology* 513.5 (2009), pp. 532–541. issn: 1096-9861. doi: [10.1002/cne.21974](https://doi.org/10.1002/cne.21974).
- [94] Casey Henley. *Foundations of Neuroscience*. Michigan State University Libraries, Jan. 1, 2021. isbn: 978-1-62610-109-8.
- [95] Wulfram Gerstner and Werner M. Kistler. *Spiking neuron models: single neurons, populations, plasticity*. Cambridge, U.K.: Cambridge University Press, 2002. 480 pp. isbn: 978-0-511-07817-0.
- [96] Leander Grech et al. "Application of reinforcement learning in the LHC tune feedback". In: *Frontiers in Physics* 10 (2022). issn: 2296-424X.
- [97] Augustin-Louis Cauchy. *Oeuvres complètes d'Augustin Cauchy*. Vol. 10. 1. 1882.
- [98] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. doi: [10.48550/ARXIV.1412.6980](https://doi.org/10.48550/ARXIV.1412.6980).
- [99] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals and Systems* 2.4 (Dec. 1, 1989), pp. 303–314. issn: 1435-568X. doi: [10.1007/BF02551274](https://doi.org/10.1007/BF02551274).
- [100] G. Gripenberg. "Approximation by neural networks with a bounded number of nodes at each level". In: *Journal of Approximation Theory* 122.2 (June 1, 2003), pp. 260–266. issn: 0021-9045. doi: [10.1016/S0021-9045\(03\)00078-9](https://doi.org/10.1016/S0021-9045(03)00078-9).
- [101] Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2 (Jan. 1, 1991), pp. 251–257. issn: 0893-6080. doi: [10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T).
- [102] Vitaly Maiorov and Allan Pinkus. "Lower bounds for approximation by MLP neural networks". In: *Neurocomputing* 25.1 (Apr. 1, 1999), pp. 81–91. issn: 0925-2312. doi: [10.1016/S0925-2312\(98\)00111-8](https://doi.org/10.1016/S0925-2312(98)00111-8).
- [103] Ding-Xuan Zhou. "Universality of deep convolutional neural networks". In: *Applied and Computational Harmonic Analysis* 48.2 (Mar. 1, 2020), pp. 787–794. issn: 1063-5203. doi: [10.1016/j.acha.2019.06.004](https://doi.org/10.1016/j.acha.2019.06.004).

- [104] Martin Arjovsky, Soumith Chintala, and Léon Bottou. “Wasserstein Generative Adversarial Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, July 17, 2017, pp. 214–223.
- [105] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. issn: 1532-4435.
- [106] Kunihiko Fukushima. “Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements”. In: *IEEE Transactions on Systems Science and Cybernetics* 5.4 (Oct. 1969), pp. 322–333. issn: 2168-2887. doi: [10.1109/TSSC.1969.300225](https://doi.org/10.1109/TSSC.1969.300225).
- [107] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. “Rectifier Nonlinearities Improve Neural Network Acoustic Models”. In: ICML Workshop on Deep Learning for Audio, Speech, and Language Processing (WDLASL 2013). 2013.
- [108] Kaiming He et al. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. Feb. 6, 2015. doi: [10.48550/arXiv.1502.01852](https://doi.org/10.48550/arXiv.1502.01852). arXiv: [1502.01852](https://arxiv.org/abs/1502.01852).
- [109] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. Feb. 22, 2016. doi: [10.48550/arXiv.1511.07289](https://doi.org/10.48550/arXiv.1511.07289). arXiv: [1511.07289](https://arxiv.org/abs/1511.07289).
- [110] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feed-forward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, Mar. 31, 2010, pp. 249–256.
- [111] Ahmet Yilmaz and Riccardo Poli. “Successfully and efficiently training deep multi-layer perceptrons with logistic activation function simply requires initializing the weights with an appropriate negative mean”. In: *Neural Networks* 153 (Sept. 1, 2022), pp. 87–103. issn: 0893-6080. doi: [10.1016/j.neunet.2022.05.030](https://doi.org/10.1016/j.neunet.2022.05.030).
- [112] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. “On the difficulty of training recurrent neural networks”. In: *Proceedings of the 30th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, May 26, 2013, pp. 1310–1318.
- [113] Ilya Loshchilov and Frank Hutter. *SGDR: Stochastic Gradient Descent with Warm Restarts*. 2016. doi: [10.48550/ARXIV.1608.03983](https://doi.org/10.48550/ARXIV.1608.03983).
- [114] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. May 24, 2019. doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805). arXiv: [1810.04805](https://arxiv.org/abs/1810.04805).
- [115] *GPT-4*. url: <https://openai.com/gpt-4> (visited on July 12, 2023).
- [116] *DALL·E 2*. url: <https://openai.com/dall-e-2> (visited on July 12, 2023).
- [117] *Midjourney*. url: <https://www.midjourney.com/home/> (visited on July 12, 2023).
- [118] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. issn: 1476-4687. doi: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2).

- [119] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. doi: [10.48550/ARXIV.1406.2661](https://doi.org/10.48550/ARXIV.1406.2661).
- [120] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [121] John Nash. "Non-Cooperative Games". In: *Annals of Mathematics* 54.2 (1951), pp. 286–295. issn: 0003-486X. doi: [10.2307/1969529](https://doi.org/10.2307/1969529).
- [122] Farzan Farnia and Asuman Ozdaglar. "Do GANs always have Nash equilibria?" In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by Hal Daumé III and Aarti Singh. Vol. 119. Proceedings of Machine Learning Research. PMLR, July 13, 2020, pp. 3029–3039.
- [123] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*. Mar. 29, 2019. doi: [10.48550/arXiv.1812.04948](https://doi.org/10.48550/arXiv.1812.04948). arXiv: [1812.04948](https://arxiv.org/abs/1812.04948).
- [124] L. N. Vaserstein. "Markovian processes on countable space product describing large systems of automata". In: *Problemy Peredachi Informatsii* 5.3 (1969), pp. 64–72. issn: 0555-2923.
- [125] L. V. Kantorovich. "Mathematical Methods of Organizing and Planning Production". In: *Management Science* (July 1, 1960). doi: [10.1287/mnsc.6.4.366](https://doi.org/10.1287/mnsc.6.4.366).
- [126] Danilo Jimenez Rezende and Shakir Mohamed. *Variational Inference with Normalizing Flows*. June 14, 2016. doi: [10.48550/arXiv.1505.05770](https://doi.org/10.48550/arXiv.1505.05770). arXiv: [1505.05770](https://arxiv.org/abs/1505.05770).
- [127] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. "Graph neural networks in particle physics". In: *Machine Learning: Science and Technology* 2.2 (Dec. 2020), p. 021001. issn: 2632-2153. doi: [10.1088/2632-2153/abbf9a](https://doi.org/10.1088/2632-2153/abbf9a).
- [128] Huilin Qu and Loukas Gouskos. "Jet tagging via particle clouds". In: *Physical Review D* 101.5 (Mar. 26, 2020), p. 056019. doi: [10.1103/PhysRevD.101.056019](https://doi.org/10.1103/PhysRevD.101.056019).
- [129] Pablo de Castro and Tommaso Dorigo. "INFERNO: Inference-Aware Neural Optimisation". In: *Computer Physics Communications* 244 (Nov. 1, 2019), pp. 170–179. issn: 0010-4655. doi: [10.1016/j.cpc.2019.06.007](https://doi.org/10.1016/j.cpc.2019.06.007).
- [130] Aishik Ghosh, Benjamin Nachman, and Daniel Whiteson. "Uncertainty-aware machine learning for high energy physics". In: *Physical Review D* 104.5 (Sept. 28, 2021), p. 056026. doi: [10.1103/PhysRevD.104.056026](https://doi.org/10.1103/PhysRevD.104.056026).
- [131] Nathan Simpson and Lukas Heinrich. "neos: End-to-End-Optimised Summary Statistics for High Energy Physics". In: *Journal of Physics: Conference Series* 2438.1 (Feb. 2023), p. 012105. issn: 1742-6596. doi: [10.1088/1742-6596/2438/1/012105](https://doi.org/10.1088/1742-6596/2438/1/012105).
- [132] S. Agostinelli et al. "Geant4, a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. issn: 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [133] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. issn: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2016.06.125>.

- [134] Alexander Radovic et al. “Machine learning at the energy and intensity frontiers of particle physics”. In: *Nature* 560.7716 (2018), pp. 41–48. doi: [10.1038/s41586-018-0361-2](https://doi.org/10.1038/s41586-018-0361-2).
- [135] The CDF Collaboration. “Measurement of $\sigma B(W \rightarrow e\nu)$ and $\sigma B(Z^0 \rightarrow e^+e^-)$ in $\bar{p}p$ collisions at $\sqrt{s} = 1800$ GeV”. In: *Phys. Rev. D* 44.1 (July 1991), pp. 29–52. doi: [10.1103/PhysRevD.44.29](https://doi.org/10.1103/PhysRevD.44.29).
- [136] The CMS collaboration. “Observation of the diphoton decay of the Higgs boson and measurement of its properties”. In: *The European Physical Journal C* 74 (2014), p. 3076. doi: [10.1140/epjc/s10052-014-3076-z](https://doi.org/10.1140/epjc/s10052-014-3076-z).
- [137] The CMS collaboration. “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* (2018), p. 185. doi: [10.1007/JHEP11\(2018\)185](https://doi.org/10.1007/JHEP11(2018)185).
- [138] Joshua Lin, Wahid Bhimji, and Benjamin Nachman. “Machine learning templates for QCD factorization in the search for physics beyond the standard model”. In: *Journal of High Energy Physics* 2019.5 (2019), p. 181. doi: [10.1007/JHEP05\(2019\)181](https://doi.org/10.1007/JHEP05(2019)181).
- [139] Andrew Chisholm et al. “Non-parametric data-driven background modelling using conditional probabilities”. In: *Journal of High Energy Physics* 2022.10 (2022), p. 1. doi: [10.1007/JHEP10\(2022\)001](https://doi.org/10.1007/JHEP10(2022)001).
- [140] Anna Hallin et al. “Classifying anomalies through outer density estimation”. In: *Phys. Rev. D* 106.5 (Sept. 2022), p. 055006. doi: [10.1103/PhysRevD.106.055006](https://doi.org/10.1103/PhysRevD.106.055006).
- [141] Victor Lohezic et al. “Data driven background estimation in HEP using generative adversarial networks”. In: *The European Physical Journal C* 83.3 (Mar. 1, 2023), p. 256. issn: 1434-6052. doi: [10.1140/epjc/s10052-023-11347-8](https://doi.org/10.1140/epjc/s10052-023-11347-8).
- [142] Inna Kucher. “Search for the Higgs boson decaying to two photons and produced in association with a pair of top quarks in the CMS experiment”. PhD thesis. Université Paris Saclay (COMUE), July 17, 2017.
- [143] Martina Machet. “Higgs boson production in the diphoton decay channel with CMS at the LHC : first measurement of the inclusive cross section in 13 TeV pp collisions, and study of the Higgs coupling to electroweak vector bosons”. PhD thesis. Université Paris Saclay (COMUE), Sept. 26, 2016.
- [144] CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 16.5 (May 2021), P05014. issn: 1748-0221. doi: [10.1088/1748-0221/16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014).
- [145] CMS Collaboration. “Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2021.7 (July 2021), p. 27. issn: 1029-8479. doi: [10.1007/JHEP07\(2021\)027](https://doi.org/10.1007/JHEP07(2021)027). arXiv: [2103.06956](https://arxiv.org/abs/2103.06956).
- [146] The CMS collaboration. *Simulated dataset Gjet in AODSIM format for 2012 collision data*. 2017. doi: [10.7483/OPENDATA.CMS.2W51.W8AT](https://doi.org/10.7483/OPENDATA.CMS.2W51.W8AT). (Visited on July 12, 2023).
- [147] Alec Radford, Luke Metz, and Soumith Chintala. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”. 2015. doi: [10.48550/ARXIV.1511.06434](https://doi.org/10.48550/ARXIV.1511.06434).

- [148] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2818–2826. doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- [149] J. Kiefer and J. Wolfowitz. "Stochastic Estimation of the Maximum of a Regression Function". In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 462–466.
- [150] Yang You et al. *Large Batch Optimization for Deep Learning: Training BERT in 76 minutes*. 2019. doi: [10.48550/ARXIV.1904.00962](https://doi.org/10.48550/ARXIV.1904.00962).
- [151] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [152] Bing Xu et al. "Empirical Evaluation of Rectified Activations in Convolutional Network". 2015. doi: [10.48550/ARXIV.1505.00853](https://doi.org/10.48550/ARXIV.1505.00853).
- [153] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. "Measuring and testing dependence by correlation of distances". In: *The Annals of Statistics* 35.6 (2007), pp. 2769–2794. doi: [10.1214/009053607000000505](https://doi.org/10.1214/009053607000000505).
- [154] CMS Collaboration. *Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the $\tau\tau$ final state*. May 10, 2022. doi: [10.48550/arXiv.2205.05120](https://doi.org/10.48550/arXiv.2205.05120). arXiv: 2205.05120.
- [155] ATLAS Collaboration. "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector". In: *The European Physical Journal C* 75.10 (Oct. 6, 2015), p. 476. issn: 1434-6052. doi: [10.1140/epjc/s10052-015-3685-1](https://doi.org/10.1140/epjc/s10052-015-3685-1).
- [156] ATLAS Collaboration. "Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector". In: *The European Physical Journal C* 76.12 (Nov. 29, 2016), p. 658. issn: 1434-6052. doi: [10.1140/epjc/s10052-016-4499-5](https://doi.org/10.1140/epjc/s10052-016-4499-5).
- [157] ATLAS collaboration. "Measurement of the Higgs boson coupling properties in the $H \rightarrow ZZ^* \rightarrow 4l$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2018.3 (Mar. 15, 2018), p. 95. issn: 1029-8479. doi: [10.1007/JHEP03\(2018\)095](https://doi.org/10.1007/JHEP03(2018)095).
- [158] ATLAS Collaboration. "CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \rightarrow \gamma\gamma$ with the ATLAS Detector". In: *Physical Review Letters* 125.6 (Aug. 5, 2020), p. 061802. doi: [10.1103/PhysRevLett.125.061802](https://doi.org/10.1103/PhysRevLett.125.061802).
- [159] CMS Collaboration. "Limits on the Higgs boson lifetime and width from its decay to four charged leptons". In: *Physical Review D* 92.7 (Oct. 22, 2015), p. 072010. doi: [10.1103/PhysRevD.92.072010](https://doi.org/10.1103/PhysRevD.92.072010).
- [160] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. issn: 1748-0221. doi: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009).
- [161] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.4 (Apr. 2008), p. 063. issn: 1126-6708. doi: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).

- [162] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 17, 2014), p. 79. issn: 1029-8479. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [163] R. Frederix et al. "The automation of next-to-leading order electroweak calculations". In: *Journal of High Energy Physics* 2018.7 (July 31, 2018), p. 185. issn: 1029-8479. doi: [10.1007/JHEP07\(2018\)185](https://doi.org/10.1007/JHEP07(2018)185).
- [164] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Physics* 7.3 (Sept. 18, 2019), p. 034. issn: 2542-4653. doi: [10.21468/SciPostPhys.7.3.034](https://doi.org/10.21468/SciPostPhys.7.3.034).
- [165] Sara Bolognesi et al. "Spin and parity of a single-produced resonance at the LHC". In: *Physical Review D* 86.9 (Nov. 29, 2012), p. 095031. doi: [10.1103/PhysRevD.86.095031](https://doi.org/10.1103/PhysRevD.86.095031).
- [166] Ian Anderson et al. "Constraining anomalous HVV interactions at proton and lepton colliders". In: *Physical Review D* 89.3 (Feb. 19, 2014), p. 035007. doi: [10.1103/PhysRevD.89.035007](https://doi.org/10.1103/PhysRevD.89.035007).
- [167] 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. 28, 2020), p. 056022. doi: [10.1103/PhysRevD.102.056022](https://doi.org/10.1103/PhysRevD.102.056022).
- [168] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *Journal of High Energy Physics* 2004.11 (Dec. 2004), p. 040. issn: 1126-6708. doi: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040).
- [169] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), p. 070. issn: 1126-6708. doi: [10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070).
- [170] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 9, 2010), p. 43. issn: 1029-8479. doi: [10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043).
- [171] Richard D. Ball et al. "Parton distributions for the LHC run II". In: *Journal of High Energy Physics* 2015.4 (Apr. 8, 2015), p. 40. issn: 1029-8479. doi: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040).
- [172] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (June 1, 2008), pp. 852–867. issn: 0010-4655. doi: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036).
- [173] Daniel de Florian et al., eds. *Handbook of LHC Higgs Cross Sections: report of the LHC Higgs Cross Section Working Group. 4: Deciphering the nature of the Higgs sector*. CERN yellow reports Monographs volume 2. Geneva: CERN, 2017. 849 pp. isbn: 978-92-9083-442-7.
- [174] Lisha Li et al. "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization". In: *Journal of Machine Learning Research* 18.185 (2018), pp. 1–52. issn: 1533-7928.
- [175] Tom O'Malley et al. *KerasTuner*. <https://github.com/keras-team/keras-tuner>. 2019.

- [176] ATLAS and CMS Collaborations, and LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11. 2011. url: <https://cds.cern.ch/record/1379837>.
- [177] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 9, 2011), p. 1554. issn: 1434-6052. doi: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
- [178] R. A. Fisher. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P". In: *Journal of the Royal Statistical Society* 85.1 (1922), pp. 87–94. issn: 0952-8385. doi: [10.2307/2340521](https://doi.org/10.2307/2340521).
- [179] P. D. Dauncey et al. "Handling uncertainties in background shapes: the discrete profiling method". In: *Journal of Instrumentation* 10.4 (Apr. 2015), P04015. issn: 1748-0221. doi: [10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015).
- [180] Sergey Alekhin et al. *The PDF4LHC Working Group Interim Report*. Jan. 3, 2011. doi: [10.48550/arXiv.1101.0536](https://doi.org/10.48550/arXiv.1101.0536). arXiv: [1101.0536](https://arxiv.org/abs/1101.0536).