UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE

Institut pluridisciplinaire Hubert Curien (IPHC)

THÈSE présentée par

Océane PONCET

soutenue le: **12 Septembre 2025**

pour obtenir le grade de: **Docteur de l'Université de Strasbourg**
Discipline/Spécialité: Physique des particules

Search for charge-parity violation in the tau Yukawa coupling with CMS Run 3 data and preparation of the tracker irradiation tests for the high luminosity phase of the LHC

THÈSE dirigée par :

Mme Anne-Catherine LE BIHAN	Directrice de recherche, Institut pluridisciplinaire Hubert Curien (IPHC)

RAPPORTEURS:

M. Klaus DESCH	Professeur, Universität Bonn, Physikalisches Institut
Mme Justine SERRANO	Directrice de recherche, Centre de physique des particules de Marseille (CPPM)

AUTRES MEMBRES DU JURY:

Mme Sandrine COURTIN	Professeure, Institut pluridisciplinaire Hubert Curien (IPHC), Université de Strasbourg
Mme Saskia FALKE	Chargée de recherche, Institut pluridisciplinaire Hubert Curien (IPHC)
M. Daniel WINTERBOTTOM	Research Associate, Imperial College London

À mes parents

"How vain it is to sit down to write when you have not stood up to live."

Henry David Thoreau

Remerciements

« Le bonheur n'est réel que lorsqu'il est partagé », écrivait Christopher McCandless. Le bonheur d'achever cette thèse n'échappe pas à cette vérité : il est l'aboutissement d'un long chemin semé d'efforts, mais illuminé par la présence et le soutien de celles et ceux qui l'ont rendu possible. C'est à cet entourage que j'adresse ces quelques lignes de gratitude.

Comme chacun sait, l'exercice des remerciements est toujours délicat : pour beaucoup, la lecture de cette thèse s'arrêtera ici, avec un soupir de soulagement ou un léger pincement d'amertume de ne pas avoir aperçu leur nom. C'est pourquoi je commence par remercier toutes les personnes que j'aurais pu oublier, mais qui m'ont pourtant offert leur aide, leur soutien, ou simplement un café salvateur au moment opportun.

Mes premiers remerciements vont naturellement à mes directrices de thèse, Anne-Catherine et Saskia. Merci de m'avoir offert cette opportunité unique, de m'avoir fait confiance et de m'avoir accompagnée de manière complémentaire tout au long de cette aventure. Saskia, merci pour tes précieux conseils, qui ont continué à résonner en moi même après ton départ, et que j'ai fini par comprendre quelques années plus tard. Je resterai à la fois ta meilleure et ta pire doctorante, mais j'espère que tu te souviendras de moi comme de la meilleure. Anne-Catherine, merci d'avoir toujours pris le temps de me conseiller, de tout relire et de me guider, avec ce calme rassurant qui m'a tant aidée. Tes conseils m'ont permis d'avancer, de retrouver confiance en moi et de rester motivée. Merci à vous deux de m'avoir inspirée et guidée au cours de ces dernières années.

Je tiens également à remercier les membres de l'équipe CMS pour leur accueil et leurs conseils. Jamais aucun d'entre vous n'a rechigné à commenter mes présentations, pour mon plus grand désespoir... mais aussi pour mon plus grand bénéfice.

Un merci tout particulier à Jérémy, qui a été comme un troisième directeur de thèse et le meilleur pour les citations cultes. Merci également à Jean-Laurent,

Clément, Laurent et à toute l'équipe de Cyrcé pour leur aide précieuse sur la partie irradiation. Merci à Jean-Laurent pour son accompagnement et son expertise, et à Clément pour tous les câbles passés et sa grande gentillesse. Thanks to Stefano for the summer student and for the many discussions during my PhD, and to Marijus and Mauro for their help with PH2 ACF.

Merci à Éric pour nos discussions, qu'elles soient sportives ou sur l'avenir de la physique au-delà du Modèle Standard. Merci à Caroline de m'avoir fait découvrir CMS dès ma première année de fac. Merci à Daniel pour prendre si bien soin de Paul. Merci à Pierre pour ton sourire et tes chemises toujours colorées. Merci à Ulrich pour ses cours de master et ses histoires passionnantes racontées au CERN.

Thanks to Gourab for all your help and for the moments we shared, coding, discussing physics, failing, and trying again. It was a real pleasure working with you, and I wish you all the best in physics and in all your future projects.

Thanks also to the Tau CP group, and to my colleagues at Imperial College (Danny, Lucas, Klitos) and at DESY (Stepan, Aliaksei, Jacopo) for their expertise and for the time spent coordinating with us on the $\tau\tau$ channel and ColumnFlow.

I am also grateful to the members of the Tau POG/CQM (Konstantin, Michal, Dennis, Izaak, Andrea) for their availability and guidance, and especially to Andrea and Stepan for welcoming me at DESY.

Je tiens également à remercier les tuteurs de mes stages en physique des particules. En premier lieu, Antonin, qui m'a accueillie chez ALICE et m'a confortée dans l'idée que je voulais faire de la physique des particules. Sans ta pédagogie, ton accompagnement et ta gentillesse, je n'aurais peut-être pas eu autant envie de continuer, alors merci. Merci aussi à Boris pour le pseudo-stage sur Thermus, mais surtout pour ton suivi pendant le M2 et même après. Merci également à Jérôme et à toute l'équipe de Belle II pour le stage et pour l'accueil dans le couloir. Vous avez tous confirmé ma volonté de faire partie de cette belle communauté, alors merci.

Évidemment, je tiens à remercier particulièrement mes copains docteurs (ou presque). En commençant par ma première rencontre, Dylan, de prof d'informatique à CMS Week à Schuman : tes expressions lorraines et ta belle chevelure m'ont bien manqué en fin de thèse, mais ce fut un plaisir de la commencer avec toi. Merci pour ça. Mario, mon mentor τ, un autre chevalier de l'apocalypse parti trop τ mais qui m'a offert l'un des meilleurs débuts de thèse possibles. Raph, le yes man, toujours un bonheur d'être avec toi. Je t'ai vu te trouver et glow up au cours de ta thèse, pour mon plus grand plaisir. J'arrive bientôt pour tout CASSER au CERN, merci pour tous ces fous rires. Paul, mon frère d'aventure, merci de m'avoir acceptée dans ton bureau. Trois ans ou six semaines, ça passe vite ! Merci d'avoir partagé avec moi la douleur de l'analyse, de l'écriture, des slides et des corrections, mais aussi toutes ces pauses, de m'avoir initiée au ping-pong et de partager ta passion

pour la musique discutable. On se retrouve bientôt, sur une plage bretonne et au CERN. T'es un gros crack. Gael, merci : ton arrivée dans l'équipe a élevé le niveau de folie et de blagues. Je crois n'avoir jamais autant ri que dans ce bureau. Je te souhaite énormément de succès dans tes projets et de trouver ta voie, que tu sauras gravir avec brio. Merci à Martha, ce petit rayon de soleil, auquel je souhaite aussi tout le bonheur. Mathilde, merci de nous avoir rejoints, c'était un vrai plaisir d'avoir ta compagnie (girl poweeerrr !). Cyril, bon courage pour la suite de ta thèse et merci pour ta bonne humeur. Merci à Thomas, qui défie Paul au ping-pong et partage les pauses de midi. Merci à Corentin pour ses lectures qualitatives et ses fun (weird) facts, bisous à Doudou. Petros, it was a pleasure to share these last years of university with you. I really learned a lot during our first internship. It was great to meet a friend like you, to share our love of series and cats. Thank you for all the food you bring from Greece. I wish you the best, because you are really great.

Thanks to my friends and colleagues met at CERN events: Anna, Martina, Samuel, Tzvi, Yao, Federico, Elise, David, Benjamin, Marion and Line.

Si la thèse s'est aussi bien passée, c'est aussi parce que j'ai été entourée d'amour et de bienveillance sur le plan personnel. Merci à mes parents, à qui je dois tout, qui m'ont toujours soutenue pour accomplir mes objectifs, quels qu'ils soient, et qui sont un pilier dans ma vie. Merci de m'avoir transmis des valeurs qui me guident, et d'avoir tout donné pour nous. Merci à ma famille, et tout particulièrement à mon inspirant frère, qui n'a rien à envier en termes de réussite, et qui partage mon goût pour démarrer mille projets. Merci également à sa petite famille, Rose et Sophie, pour le bonheur que vous apportez. Merci enfin à mes oncles et tantes d'avoir fait le déplacement pour l'occasion, et pour tout l'encouragement et l'amour que vous m'avez témoigné.

Merci à mes amis, toujours présents pour embellir ma vie. À mes amis d'enfance : Diana, Julie, Ben, Jérémy, Rémi, Clémiche, Charlotte, que je vois peu mais qui ont contribué à façonner la personne que je suis. Aux vieilles destinées : Martina, Théo, Lucas, Eliot, pour nos premières années étudiantes. Aux Belettes : Rémi, Claudia, Céline, Daphné, Guillaume, David et Louis, pour tous les souvenirs partagés et pour enrichir mon quotidien avec chaque jour de nouvelles histoires. Merci d'être là et bravo au Dr Guillaume ! Merci à Lilie, mon amie de longue date, à qui je ne souhaite que du bonheur. Merci à Mathilde et Anthony, et un merci tout particulier à Mathilde d'avoir pris le temps de tout relire : tu es brillante et une personne merveilleuse.

Enfin, merci à Tom, mon amour, de rendre ma vie plus belle, plus simple et plus drôle. Merci de me soutenir dans mes mille projets par minute et de m'aider à les réaliser. La vie est dure, mais pas la nôtre.

Résumé

Le concept de *particules élémentaires* désigne des particules qui sont ponctuelles et ne possèdent pas de structure interne. Le modèle standard de la physique des particules (MS) décrit ces constituants fondamentaux ainsi que leurs interactions. En 2012, la découverte du boson de Higgs par les expériences ATLAS et CMS a complété le MS en confirmant le mécanisme de Brout–Englert–Higgs, introduit en 1964, qui explique la génération des masses via la brisure spontanée de la symétrie électrofaible.

Cependant, le MS ne permet pas d'expliquer l'asymétrie matière–antimatière observée dans l'Univers. La violation de symétrie charge-parité (CP) mesurée expérimentalement jusqu'à présent est trop faible, suggérant l'existence de nouvelles sources, potentiellement dans le secteur de Higgs. Si les mesures actuelles des couplages du Higgs aux bosons vecteurs indiquent une nature scalaire, des effets violant la symétrie CP pourraient émerger dans les couplages de Yukawa via un mélange scalaire–pseudoscalaire.

Cette thèse étudie la violation de CP dans la désintégration du boson de Higgs en une paire de leptons τ hadroniques τ_h, où les corrélations de spin des produits de désintégration permettent de sonder la structure CP du couplage. L'analyse bénéficiera de la statistique accrue du Run 3, offrant une sensibilité améliorée par rapport aux résultats de Run 2.

Une calibration précise des τ_h est essentielle. Ce travail propose une nouvelle méthode d'ajustement combiné permettant l'extraction simultanée de l'échelle en énergie et de l'efficacité d'identification des τ_h, en tenant compte de leurs corrélations.

Enfin, dans le cadre des préparatifs à la phase de haute luminosité du LHC (HL-LHC), cette thèse présente une étude dédiée à la préparation de la campagne d'irradiation à l'IPHC, des modules du futur trajectographe interne de CMS. Cette étude comprend la mise en place du dispositif expérimental pour les tests à froid, ainsi que la validation de ce dispositif. Par ailleurs, des tests ont été réalisés sur

le premier module irradié avec la nouvelle ligne d'irradiation du cyclotron CYRCé de l'IPHC.

Contexte théorique

Cette section présente le cadre théorique de la thèse, introduisant le MS, la violation de la symétrie CP et les observables associées, ainsi que les méthodes de reconstruction utilisées dans l'analyse.

Le modèle standard de la physique des particules

Le SM [1–3] repose sur le groupe de jauge:

$$SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}},$$

qui décrit respectivement les interactions forte, faible et électromagnétique [4–6]. Il contient douze fermions (et anti-fermions) élémentaires (six quarks, six leptons) et les bosons vecteurs médiateurs. L'interaction forte est décrite par la QCD via $SU(3)_{\mathrm{C}}$ [7], tandis que l'interaction électrofaible unifie $SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ [4, 5].

Le mécanisme de Brout–Englert–Higgs [8, 9] introduit un doublet scalaire Φ dont le potentiel entraîne une brisure spontanée de symétrie avec $\langle\Phi\rangle = \frac{v}{\sqrt{2}}$, $v \approx 246\,\mathrm{GeV}$. Cela confère une masse aux bosons W et Z, tandis que le photon reste sans masse. Les fermions obtiennent leur masse par couplages de Yukawa proportionnels à v [10], la valeur attendue dans le vide du champ de Higgs.

Mesures du boson de Higgs

Depuis sa découverte par ATLAS et CMS en 2012 [11, 12], le boson de Higgs fait l'objet d'études précises. Au LHC, il est majoritairement produit par la fusion gluon-gluon (ggF), puis par fusion de bosons vecteurs (VBF), production associée à un boson vectoriel (VH) ou à des quarks top ($t\bar{t}H$, tH) [13–15]. Ses désintégrations dominantes sont $H \to b\bar{b}$, WW^*, gg et $\tau^+\tau^-$ [15].

Les couplages au boson Higgs, proportionnels aux masses des fermions et à la racine de la masse des bosons, sont testés via les facteurs de modification du couplage κ, où $\kappa_i = 1$ correspond au SM [16]. Aucune déviation significative par rapport aux prédictions du MS n'a été observée à ce jour.

Recherche de violation de CP dans le secteur de Higgs

La violation de CP, insuffisante dans le MS pour expliquer l'asymétrie matière-antimatière, pourrait apparaître dans le secteur de Higgs via des couplages non

standards, notamment dans la désintégration $H \to \tau^+\tau^-$. Cette thèse étend les analyses du Run 2 de CMS [17] en exploitant les données du Run 3.

La structure CP du couplage Higgs–τ est étudiée via l'observable ϕ_{CP}, sensible aux corrélations de spin des produits de désintégration. L'analyse combine les canaux $e\tau_h$, $\mu\tau_h$ et $\tau_h\tau_h$, couvrant près de 70% des désintégrations $\tau\tau$. L'angle de mélange CP est extrait par ajustement, et a donné pour le Run 2 $\alpha_{\tau\tau} = -1 \pm 19°$ (68.3% CL), excluant un Higgs purement CP-impair à 3.0σ, en accord avec le SM mais limité par les incertitudes statistiques.

Observables sensibles à la violation de CP dans $H \to \tau\tau$

La sensibilité à la violation de CP dans les désintégrations du boson de Higgs en paires de τ repose sur l'analyse des corrélations angulaires entre les produits de désintégration des leptons τ. En particulier, les distributions des angles entre les plans de désintégration permettent de sonder la structure CP du couplage de Yukawa. Pour reconstruire ces plans, trois méthodes sont utilisées en fonction du canal de désintégration du τ :

- **Méthode du plan de désintégration (Decay Plane, DP)** : utilise les directions des produits visibles dans le référentiel du laboratoire pour approximer le plan de désintégration. Cette méthode est particulièrement adaptée à la désintégration $\tau \to \rho\nu \to \pi^\pm\pi^0\nu_\tau$, permettant une reconstruction géométrique directe à partir des impulsions des pions neutres et chargés.

- **Méthode du vecteur polarimétrique (Polarimetric Vector, PV)** : repose sur la reconstruction du vecteur polarimétrique des τ, calculé à partir des moments des hadrons dans le référentiel propre du τ. Elle est applicable à la désintégration $\tau \to a_1\nu_\tau$, mais requiert une estimation fiable de l'impulsion totale du τ (y compris la contribution des neutrinos) afin de reconstruire le vecteur polarimétrique ainsi que le plan défini par ce vecteur et l'impulsion totale du τ.

- **Méthode de l'impact parameter (IP)** : utilise le vecteur d'impact transverse de la trace du τ chargé pour définir le plan de désintégration. Cette approche est principalement utilisée pour les canaux leptoniques ($\tau \to \ell\nu\nu$) ou les désintégrations hadroniques à une seule particule visible (par exemple, $\tau \to \pi\nu$), où l'information directionnelle est limitée.

Les méthodes décrites s'expriment via une notation commune [18] fondée sur la définition du plan de désintégration de chaque τ à partir de deux vecteurs : le moment d'un produit chargé, $\vec{P}_\pm$ et un vecteur de référence $\vec{R}_\pm$, non parallèle à $\vec{P}_\pm$.

Après boost dans le référentiel à quantité de mouvement nulle (ZMF) des deux produits chargés, vérifiant

$$\vec{P}_+ + \vec{P}_- = 0,$$

l'angle d'acoplanarité ϕ^* est défini par

$$\phi^* = \begin{cases} \arccos\left(\frac{\vec{R}_+^\perp \cdot \vec{R}_-^\perp}{|\vec{R}_+^\perp||\vec{R}_-^\perp|}\right), & O \geq 0, \\ 2\pi - \arccos\left(\frac{\vec{R}_+^\perp \cdot \vec{R}_-^\perp}{|\vec{R}_+^\perp||\vec{R}_-^\perp|}\right), & O < 0, \end{cases}$$

avec

$$O = (\vec{R}_+^\perp \times \vec{R}_-^\perp) \cdot \vec{P}_-, \quad \vec{R}_\pm^\perp = \vec{R}_\pm - \frac{(\vec{R}_\pm \cdot \vec{P}_\pm)}{|\vec{P}_\pm|^2}\vec{P}_\pm.$$

Méthode	$\vec{P}_-$	$\vec{R}_-$	$\vec{P}_+$	$\vec{R}_+$	Décalage de phase π
Paramètre d'impact	$\vec{p}_p$	$\vec{IP}$	$\vec{p}_p$	$\vec{IP}$	-
Pion neutre	$\vec{p}_\pi$	$\vec{p}_{\pi^0}$	$\vec{p}_\pi$	$\vec{p}_{\pi^0}$	$y_+ \cdot y_- < 0$
Vecteur polarimétrique	$\vec{p}_\tau$	$\vec{h}$	$\vec{p}_\tau$	$\vec{h}$	-

Table 1: Résumé des méthodes et vecteurs utilisés pour la reconstruction des observables CP.

Du détecteur CMS au LHC au détecteur Phase-2 pour le HL-LHC

Cette section présente le contexte expérimental, du détecteur CMS actuel à sa mise à niveau appelée Phase-2 pour la phase de haute luminosité du LHC (HL-LHC).

Le LHC

Le Grand collisionneur de hadrons (LHC) au CERN, avec ses 27 km de circonférence, est le plus puissant accélérateur de particules au monde [19, 20]. Il produit des collisions proton-proton jusqu'à 13.6 TeV en énergie centre-masse. Le LHC alterne phases d'exploitation et arrêts techniques pour mises à niveau. Après la Phase-1 (2011–2026), le HL-LHC et la mise à niveau Phase-2 du détecteur CMS sont prévus pour 2030.

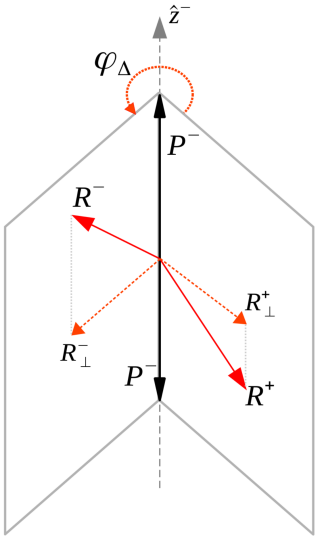


Figure 1: Définition schématique de l'angle d'acoplanarité ϕ^* entre les plans de désintégration des $\tau^\pm$ selon la notation commune. Les vecteurs $\vec{R}^\perp_\pm$ sont orthogonaux aux axes $\vec{P}_\pm$ dans les plans respectifs [18].

L'expérience CMS

Le détecteur CMS est un détecteur à usage général installé au Point 5 du LHC. Il est composé de plusieurs sous-systèmes organisés de manière cylindrique autour du point d'interaction, chacun étant optimisé pour la détection et la mesure de différents types de particules. Une vue schématique illustrant l'ensemble des sous-détecteurs est présentée à la Figure 2.

Le trajectographe interne, constitué de détecteurs à pixels et à microbandes en silicium, couvre la région pseudorapide $|\eta| < 2.5$ et assure la reconstruction précise des trajectoires des particules chargées. Il est suivi par le calorimètre électromagnétique (ECAL), conçu pour mesurer l'énergie des électrons et des photons. Le calorimètre hadronique (HCAL) entoure l'ECAL et permet la mesure de l'énergie des hadrons. Ces trois sous-détecteurs sont contenus à l'intérieur d'un solénoïde supraconducteur générant un champ magnétique de 3.8 T, indispensable pour la courbure des trajectoires des particules chargées. Enfin, le système de détection des muons, situé à l'extérieur du solénoïde et intégré dans le retour du flux magnétique, utilise différents types de chambres gazeuses (DT, CSC, RPC) pour identifier et mesurer les muons jusqu'à $|\eta| \sim 2.4$.

Au LHC, le taux de croisement de faisceaux (40 MHz) empêche l'enregistrement intégral des événements. Le déclenchement de CMS, en deux niveaux, sélectionne en temps réel les événements d'intérêt. Le L1, basé sur des FPGA et utilisant calorimètres et muons, réduit le flux à 100 kHz. Le High-Level Trigger (HLT),

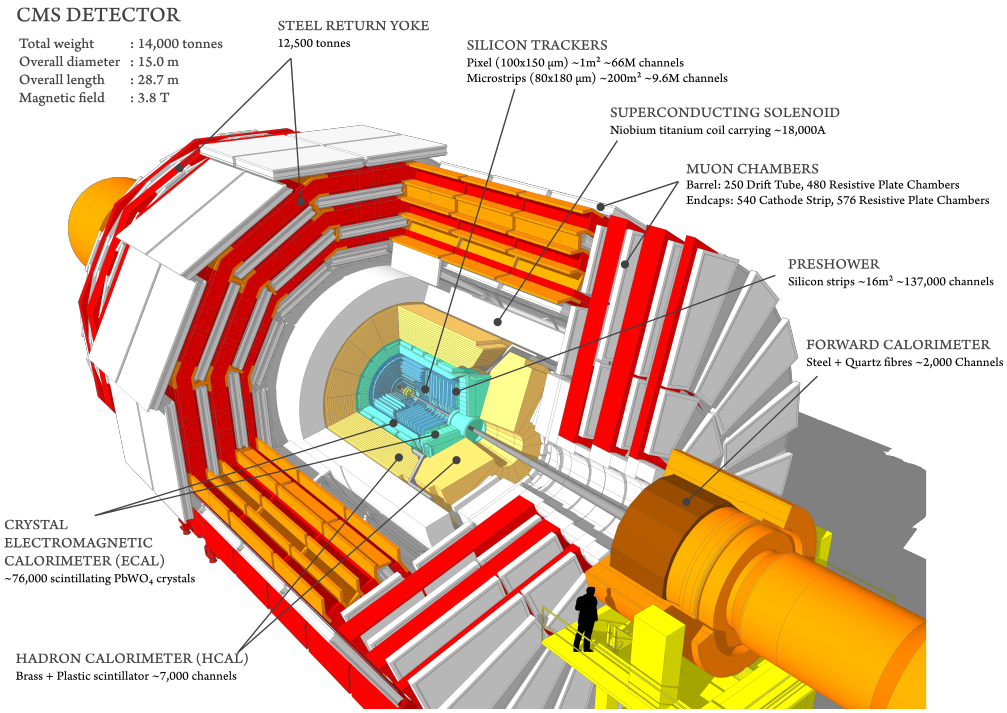


Figure 2: Vue shecmatique du detecteru CMS [21].

sur une ferme de processeurs, applique des algorithmes avancés pour aboutir à un taux final d'enregistrement d'environ 1 kHz, conciliant sélection fine et contraintes techniques.

Trajectographe en silicium

Les trajectographes en silicium sont essentiels en physique des hautes énergies pour reconstruire avec précision les trajectoires des particules chargées. Basés sur la création de paires électron-trou dans du silicium dopé, ces détecteurs collectent les charges via un champ électrique, offrant une granularité fine. On distingue deux types : les détecteurs à pixels, fournissant une localisation précise par pixel, et les détecteurs à microbandes (*strips*), qui offrent une bonne résolution directionnelle avec moins de canaux.

Ces détecteurs, placés près du point d'interaction, doivent aussi satisfaire à des exigences strictes de résistance aux radiations, de rapidité et de stabilité thermique, en particulier pour le HL-LHC.

Mise à niveau pour la Phase-2

Pour faire face aux conditions extrêmes du HL-LHC, le trajectographe interne de CMS sera entièrement rénové lors de la Phase-2. Le nouveau *Inner Tracker* sera un détecteur à pixels, proche du point d'interaction avec une granularité fine.

La conception inclut également une meilleure résistance aux radiations, grâce à des capteurs silicium avancés, des électroniques durcies et une gestion thermique optimisée, garantissant des performances stables durant plus de dix ans sous des doses jusqu'à 10^{16} $\mathrm{n_{eq}/cm^2}$.

Développement et validation du dispositif d'irradiation pour les modules pixel de la Phase 2 à CYRCé

Dans le cadre de la mise à niveau Phase-2 de l'Inner Tracker, les modules pixels doivent être qualifiés avant et après irradiation afin de garantir leur bon fonctionnement sous les doses extrêmes prévues au HL-LHC, atteignant jusqu'à $\phi_{\mathrm{eq}} = 3{,}5 \times 10^{16}$ $\mathrm{n_{eq}/cm^2}$ pour une luminosité intégrée de 4000 $\mathrm{fb^{-1}}$. Plusieurs campagnes d'irradiation sur faisceau ont permis de valider les performances des capteurs plans et 3D après exposition [22, 23]. La phase de production en cours impose également des irradiations systématiques des modules afin de vérifier leur conformité aux exigences en conditions extrêmes.

Le groupe CMS de l'IPHC a apporté sa contribution en mettant en place une ligne d'irradiation au cyclotron CYRCé, dédiée à la qualification des modules. Cette section décrit la mise en service et les résultats de la première campagne d'irradiation, réalisée en 2024.

Dans le cadre de cette thèse, un banc de test a été développé pour permettre la caractérisation des modules dans des conditions environnementales contrôlées (température, humidité). Un protocole de test a été établi et testé afin de garantir la reproductibilité des mesures et de démontrer l'opérabilité de l'installation. Ce développement assure que l'IPHC est techniquement prêt à mener une campagne d'irradiation à plus grande échelle, incluant la qualification systématique d'une dizaine de modules.

Dispositif expérimental

Le CYRCé (Cyclotron pour la Recherche et l'Enseignement) est un cyclotron compact TR24, en fonctionnement depuis 2013 à l'IPHC Strasbourg. Il fournit des faisceaux de protons jusqu'à 25 MeV avec des intensités comprises entre 100 attoampères et 10 nanoampères. Initialement dédié à la radiobiologie et à la production d'isotopes, il accueille désormais une ligne d'irradiation spécifiquement

conçue pour la qualification des capteurs silicium du tracker CMS, permettant d'atteindre des fluences allant jusqu'à 10^{16} n_{eq}/cm^2. Cette installation positionne l'IPHC parmi les acteurs européens majeurs des campagnes d'irradiation, aux côtés d'infrastructures telles que le KIT [24] et DESY [25].

Le module est irradié dans une *boîte froide* garantissant un environnement stable en température et humidité, grâce à un refroidissement par air sec, un dessiccateur intégré, ainsi que des capteurs environnementaux. La position du module peut être ajustée à l'aide d'une table motorisée, permettant un balayage uniforme du faisceau. La boîte froide assure également une isolation de l'air intérieur afin de limiter les activations radioactives.

Lors de la première campagne, un module RD53B CROC-V1 [26] équipé d'un capteur 3D a été irradié de manière homogène par balayage, avec un pas de 1 mm sur une durée de 1,5 heure, atteignant une fluence de 5×10^{15} n_{eq}/cm^2. La fluence a été mesurée par deux méthodes indépendantes, le courant de faisceau et l'activation d'un film de nickel, avec un accord meilleur que 15%, confirmant la fiabilité des mesures. L'homogénéité du faisceau a été vérifiée à l'aide de films Gafchromic [27].

Prototypage d'une boîte de test froide

Après irradiation, les modules sont transférés vers une *boîte de test froide* dédiée à leur calibration et caractérisation dans des conditions contrôlées, libérant ainsi la ligne d'irradiation pour d'autres utilisateurs. Cette section décrit les exigences, la conception et la validation de cette enceinte.

Motivations et exigences

Les tests post-irradiation requièrent un contrôle thermique strict pour éviter l'emballement thermique lié à l'augmentation du courant de fuite et limiter les effets de recuit modifiant les dommages radiatifs. La boîte froide doit maintenir une température stable proche des conditions CMS (~ -35 °C) via un refroidissement efficace.

Le contrôle de l'humidité est crucial : la ventilation à l'air sec empêche la condensation, nuisible à l'électronique des modules non hermétiques, en garantissant $T > T_{\text{rosée}}$, ce dernier calculé via la formule de Buck [28] en fonction de la température et de l'humidité relative (HR). Par ailleurs, même sans condensation, une humidité relative supérieure à 40% peut provoquer des ruptures électriques dans les modules irradiés [29], justifiant une humidité minimale pour préserver leur intégrité électrique.

Conception de la boîte froide

La boîte froide, installée dans une zone supervisée proche de la plateforme CYRCé, minimise les contraintes liées à l'activation résiduelle tout en facilitant l'accès aux tests. Le dispositif repose sur un congélateur commercial modifié, capable d'atteindre −35 °C grâce à un mode "super" pour un refroidissement rapide. De petits ventilateurs internes assurent une circulation d'air homogène, essentielle lors de la dissipation thermique des modules alimentés. Une ouverture latérale permet le passage des câbles et l'arrivée d'air sec, contrôlée par un débitmètre visible sur la Figure 3. Plusieurs capteurs de température et d'humidité, reliés à un Raspberry Pi, assurent une acquisition continue et un archivage des données. Des résistances et plaques chauffantes peuvent simuler la charge thermique des modules, permettant de tester la stabilité en conditions réalistes. L'ensemble utilise des matériaux antistatiques (tapis ESD, câbles à la terre) et la lecture des modules s'effectue via une carte FPGA refroidie, connectée à un PC dédié à l'acquisition.

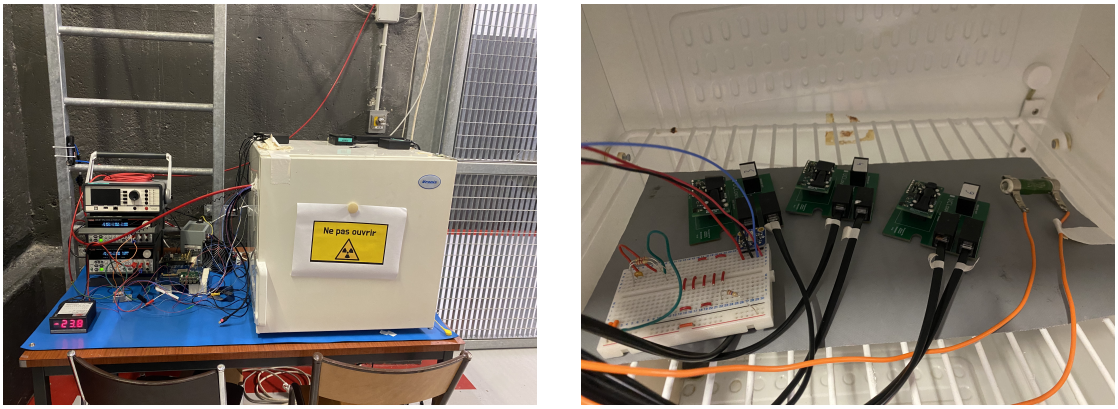


Figure 3: (Gauche) Vue extérieure de la boîte froide, montrant le débitmètre et les câbles d'alimentation. (Droite) Intérieur de la boîte avec capteurs de température et humidité, ainsi qu'une résistance chauffante simulant une charge thermique.

Campagne de validation et protocole de test

La validation de la boîte froide a suivi une procédure standard : injection d'air sec à un débit élevé (50 L/min) pour réduire l'humidité, suivie de la mise en marche du congélateur. Le débit d'air a été ajusté pour optimiser le contrôle de l'humidité tout en minimisant le temps de refroidissement. Durant le réchauffage, le flux d'air sec est augmenté pour éviter toute condensation.

Les mesures effectuées par trois capteurs répartis montrent une bonne cohérence des données, avec des écarts de l'ordre de ±1°C pour la température

et ±4°C pour le point de rosée. Celui-ci reste systématiquement inférieur à la température, assurant l'absence de condensation.

En mode "normal", le congélateur atteint environ −20 °C, insuffisant sous une charge thermique simulée par des résistances dissipant jusqu'à 9 W. Le mode "super" permet d'atteindre −35 °C, avec un débit d'air sec ajusté entre 8 et 50 L/min, maintenant une humidité relative basse.

L'ajout de ventilateurs internes améliore la dissipation thermique et la stabilité, permettant de maintenir environ −29 °C sous une charge thermique de 7 W, avec une humidité relative inférieure à 10% (Figure 4).

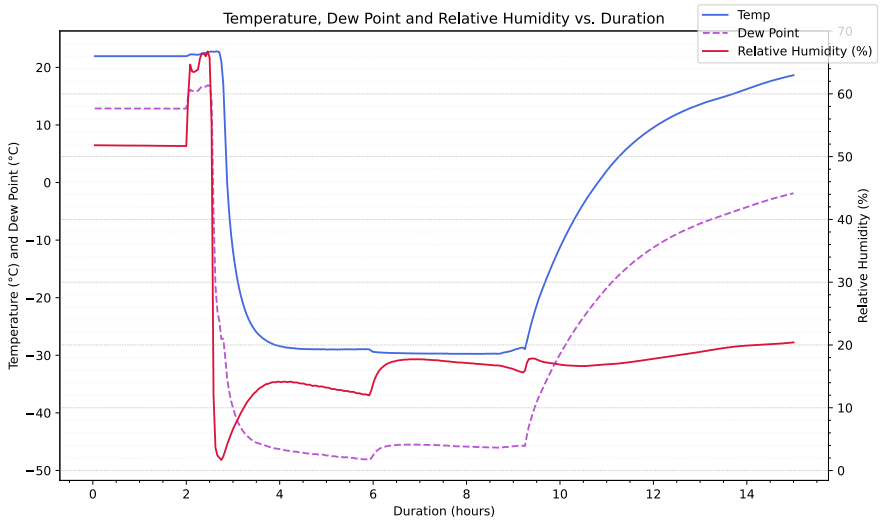


Figure 4: Température, humidité relative et point de rosée en mode "super" avec charge thermique simulée (7 W) et ventilation interne. La stabilité thermique et le contrôle de l'humidité sont assurés.

Performances avec modules irradiés

Un premier test avec un module CROC-V1 irradié a révélé qu'un démarrage prématuré à −35 °C sans ventilation suffisante entraîne un emballement thermique et une chute brutale de la tension d'alimentation.

Une seconde expérience, avec retrait du capuchon protecteur et ajout d'un ventilateur, a permis de stabiliser la température à −29 °C, avec un point de rosée bien inférieur à la température, garantissant une exploitation sans condensation (Figure 5). Le module a pu être alimenté et calibré avec succès, la montée en température restant maîtrisée (∼ +1°C).

Un suivi sur plusieurs jours a confirmé la stabilité à long terme, avec température maintenue sous −30 °C et humidité relative inférieure à 7%.

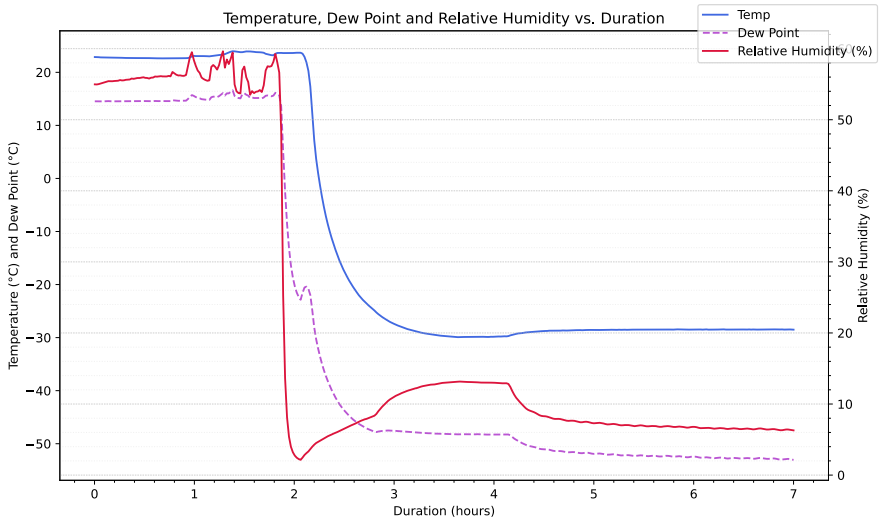


Figure 5: Température, humidité relative et point de rosée lors de l'opération stable d'un module CROC-V1 irradié en mode "super" avec ventilation active.

Résumé et perspectives

La campagne de qualification démontre que la boîte froide fournit un environnement stable en température et humidité, garantissant la sécurité des modules irradiés. Le mode "super", combiné à un flux d'air sec et une ventilation locale, permet de maintenir les modules à -30 °C. Le protocole standardisé de refroidissement et de réchauffage, associé à une surveillance continue, réduit les risques d'emballement thermique et d'annealing non contrôlé.

Les améliorations futures viseront à limiter la contamination particulaire et à optimiser l'ergonomie et la sécurité, en vue d'accueillir des modules plus complexes lors des prochaines campagnes.

Tests du CROC-V1 avant et après irradiation

Une puce CROC-V1, bump-bondée à un capteur 3D, a été irradiée à une fluence de 5×10^{15} n_{eq}/cm^2 dans la nouvelle installation d'irradiation du CYRCé. Des tests fonctionnels ont été réalisés avant et après irradiation afin d'évaluer la performance de la puce et de valider la chaîne de lecture à basse température comme présentée dans la section précédente.

Cette étude se décompose en trois étapes principales. Tout d'abord, la chaîne complète de lecture a été mise en service et validée, incluant la communication bas niveau, le réglage des seuils (trimming) et le comportement à la mise sous tension. Ensuite, une caractérisation avant irradiation a été réalisée, comprenant

l'analyse de la réponse des pixels, l'ajustement des seuils et les calibrations. Enfin, après irradiation, des mesures dédiées ont été menées pour déterminer la tension de déplétion du capteur, étudier des anomalies observées dans les premières analyses post-irradiation et calibrer les zones fonctionnelles de la puce.

Installation de test et configuration de la lecture

Le module CROC-V1 avec capteur 3D a été testé via l'infrastructure DAQ standard du pixel Phase-2, pilotée par le logiciel `Ph2_ACF` [30]. La communication passait par une carte KSU-FMC [31] connectée aux FPGA et alimentations externes.

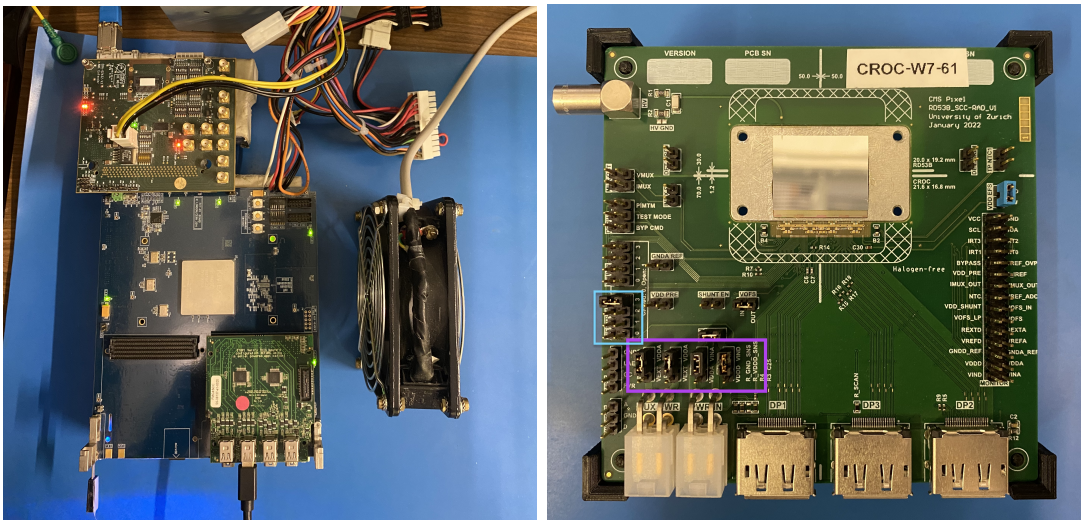


Figure 6: Banc de test du CROC-V1 : (gauche) chaîne de lecture complète avec FC7, KSU-FMC et alimentations ; (droite) module CROC-V1 sur carte porteuse, avec capteur 3D au centre et cavaliers LDO (violet) et trimming (bleu).

Le CROC-V1 était relié à la FMC par câble DisplayPort, avec la haute tension via câble LEMO et basse tension (1,8 V, 1,5 A max) fournie par une alimentation dédiée. La carte FC7, générant horloges et triggers, était alimentée séparément. Le module fonctionnait en mode LDO, avec la configuration suivant le manuel RD53 [26] (cf. Figure 6). Un ventilateur assurait le refroidissement du FPGA, et les précautions anti-ESD (tapis, bracelet) étaient systématiques.

Caractérisation avant irradiation

Avant irradiation, le module CROC-V1 a fait l'objet d'une calibration complète visant à valider sa fonctionnalité et uniformiser la réponse pixel par pixel. Les réglages des seuils, la calibration Time-over-Threshold (ToT) et le masquage des

pixels défectueux ont permis d'obtenir une matrice quasi intégralement active, sans zones mortes ni pixels bruyants (cf. Figure 7, gauche). La distribution de la S-curve montre une transition relativement nette entre 0% et 100% d'efficacité, témoignant d'une homogénéité des seuils (Figure 7, droite). Ces résultats constituent une référence fiable avant irradiation.

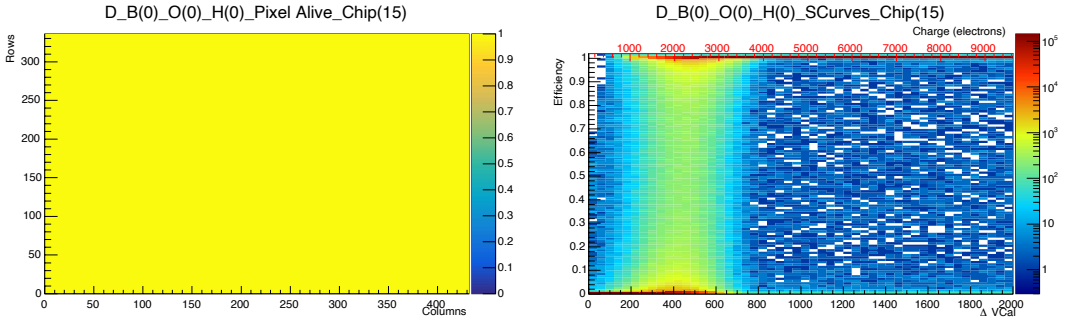


Figure 7: (Gauche) Carte d'occupation des pixels avant irradiation, indiquant une quasi-totalité des pixels actifs sans défaut majeur. (Droite) Distribution de la S-curve avant irradiation, illustrant une faible dispersion des seuils et un faible bruit.

Caractérisation après irradiation

Après irradiation et environ deux semaines de recuit à température ambiante, des mesures courant-tension (IV) ont été réalisées à différentes températures dans la boîte froide afin d'estimer la tension de pleine déplétion du capteur 3D. La courbe IV montre une plage de déplétion comprise entre 25 et 45 V, en accord avec les valeurs attendues après irradiation à fluence modérée. Il est important de noter que cette estimation reste approximative, la mesure capacitance-tension (CV) étant plus adaptée à une détermination précise.

Les mesures initiales ont révélé une zone en forme de "W" présentant un fort bruit et des pixels défectueux. Un scan de cross-talk (Figure 8) a permis d'identifier de nombreux bump-bonds déconnectés dans cette région, suggérant un défaut mécanique localisé.

Après masquage des pixels défectueux, la calibration a permis une égalisation efficace des seuils et une réponse homogène dans la région fonctionnelle. La comparaison des courbes S-curve montre une nette amélioration, avec une distribution plus étroite centrée autour du seuil cible (Figure 9).

En résumé, malgré les dommages localisés induits par l'irradiation, la majorité des pixels actifs conserve une fonctionnalité satisfaisante, avec un bruit et un seuil calibrés dans les plages attendues, attestant de la robustesse du capteur et de l'électronique associée après irradiation.

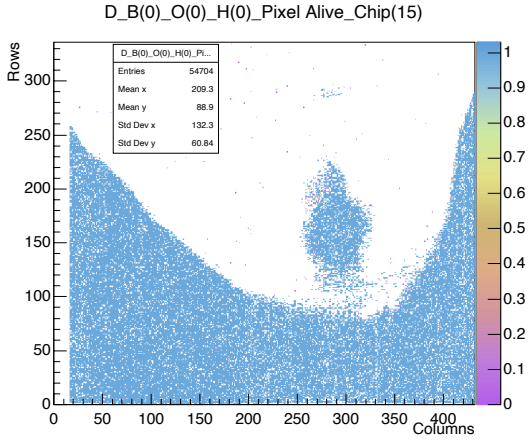


Figure 8: Scan de cross-talk post-irradiation montrant la région “W” avec des pixels déconnectés (absence de réponse dans les pixels voisins).

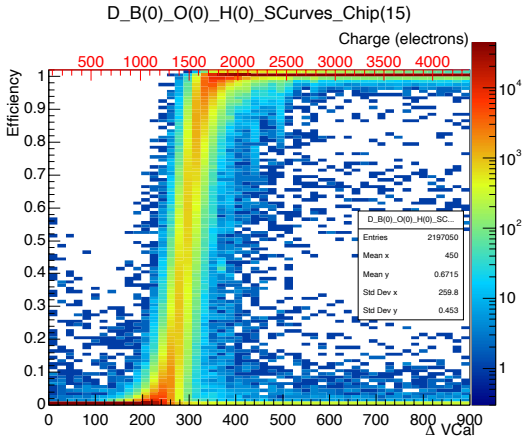


Figure 9: Courbe S-curve post-calibration des pixels fonctionnels, illustrant une réponse homogène et un bruit maîtrisé.

Conclusion

La première campagne d’irradiation a été menée avec succès, marquant la mise en service de la nouvelle ligne d’irradiation au CRYCé. Un module CROC-V1 a été irradié à froid à une fluence de 5×10^{15} n_{eq}/cm^2 selon un protocole et un dispositif de test dédiés. Les performances initiales après irradiation confirment que la puce reste fonctionnelle.

Un comportement non uniforme a été observé, lié à une zone localisée de bump-bonds déconnectés, vraisemblablement due à une faiblesse mécanique lors du col-

lage. Le masquage de cette région a permis de démontrer que le reste du module présente une réponse homogène, et que l'irradiation était uniforme.

Cette étude valide la procédure d'irradiation ainsi que le dispositif de test à froid. Avec l'infrastructure désormais opérationnelle, l'IPHC est prêt à poursuivre la campagne complète d'irradiation sur plusieurs modules.

Reconstruction et identification des τ hadroniques

La reconstruction des leptons τ hadroniques dans CMS repose sur l'algorithme *particle-flow* (PF) [32], qui combine les informations des sous-détecteurs pour reconstruire les particules individuelles. Cette approche est essentielle pour identifier correctement les désintégrations hadroniques du τ, qui produisent typiquement un ou plusieurs hadrons chargés et des photons issus de π^0.

L'algorithme *Hadron Plus Strips* (HPS) [33] reconstruit les τ_{had} à partir des candidats PF en regroupant les photons dans des « strips » pour tenir compte des conversions, et en classifiant les désintégrations selon leur *mode de désintégration* (Decay Mode, DM). Les principaux DMs incluent : DM0 ($h^{\pm}$), DM1 ($h^{\pm}\pi^0$), DM10 ($h^{\pm}h^{\mp}h^{\pm}$), et DM11 ($h^{\pm}h^{\mp}h^{\pm}\pi^0$).

Mesure combinée des facteurs d'échelle d'énergie et l'efficacité d'identification des τ_h

Cette section présente le développement d'une nouvelle méthode pour la mesure combinée des facteurs d'échelle d'énergie (TES) et l'efficacité d'identification (ID SF) des τ_h. L'approche proposée permet une extraction simultanée de ces deux paramètres à partir des données, tout en tenant compte de leurs corrélations. Elle vise à améliorer la précision de ces calibrations essentielles pour les analyses de physique impliquant des leptons τ hadroniques.

Technique de mesure

Les facteurs TES et ID SFs sont extraits simultanément via une méthode *tag-and-probe* appliquée aux événements $Z \to \tau_\mu \tau_h$, sélectionnés avec un muon isolé (tag) et un candidat τ_h (probe). L'analyse utilise les données pp à $\sqrt{s} = 13\,\text{TeV}$ (UL2018, $59.8\,\text{fb}^{-1}$) et $\sqrt{s} = 13.6\,\text{TeV}$ (Run 3 2022, $7.98\,\text{fb}^{-1}$ pré-EE et $26.67\,\text{fb}^{-1}$ post-EE).

Le TES ajuste l'énergie reconstruite du τ_h en simulation pour correspondre à celle observée en données. L'ID SF correspond au rapport d'efficacité d'identification

en données sur celle en simulation. Ces deux paramètres sont extraits par un ajustement du maximum de vraisemblance sur la masse visible :

$$m_{\text{vis}} = \sqrt{(E_{\tau_h} + E_\mu)^2 - |\vec{p}_{\tau_h} + \vec{p}_\mu|^2}.$$

L'ajustement est réalisé dans des régions définies par le mode de désintégration (DM) ou par DM et $p_T(\tau_h)$ (DM-p_T).

Contrairement aux analyses précédentes qui ajustaient séparément TES et ID SF, cette méthode effectue un ajustement combiné tenant compte des corrélations et incertitudes systématiques. Dans chaque région r, la vraisemblance est définie par :

$$L_r(TES_r, IDSF_r, \theta_r) = p(\text{données}|\text{POIs}, \theta_r) \prod_i p(\overline{\theta_i}|\theta_{i,r}),$$

où les POIs sont TES et ID SF, et θ_r l'ensemble des paramètres nuisances.

Le TES affecte la forme de m_{vis}, tandis que l'ID SF en modifie la normalisation. Des distributions de référence $Z \to \tau_h\mu$ (ZTT) sont générées avec une variation du TES de $\pm 3\%$ (DM) et $\pm 5\%$ (DM-p_T), puis interpolées par morphing horizontal [34]. La méthode, validée sur les données UL2018, est appliquée aux données de 2022, avec les régions d'ajustement listées dans le Tableau 2.

Table 2: Régions ajustées pour chaque jeu de données. Deux ajustements distincts ont été réalisés pour UL2018.

Jeu de données	Régions ajustées	Nombre de régions
UL2018	DM ; DM-p_T : $p_T = [20, 40], [40, 200]$ GeV	4 ; 8
2022 pré-EE	DM	4
2022 post-EE	DM	4

Développement de la méthode sur les données UL2018

La méthode combinée de mesure simultanée du TES et de l'ID SF a été initialement développée et validée sur les données de 2018 à $\sqrt{s} = 13$ TeV, correspondant à une luminosité intégrée de $59.8\,\text{fb}^{-1}$. Cette étape a permis d'optimiser la stratégie d'analyse avant son application aux données de Run 3.

Deux ajustements indépendants ont été réalisés : l'un dans quatre régions définies par le mode de désintégration (*Decay Modes*) du τ_h (DM0, DM1, DM10, DM11), et l'autre dans huit régions définies par la combinaison DM-p_T, avec deux intervalles de $p_T(\tau_h) = [20, 40]$ et $[40, 200]$ GeV pour chaque DM.

Les résultats obtenus sur UL2018 sont cohérents avec les analyses antérieures et montrent une bonne stabilité de la méthode. Toutefois, la segmentation en régions DM-p_T, bien qu'utile pour étudier une dépendance éventuelle du TES à l'impulsion transverse, reste limitée par la statistique disponible dans certaines régions, en particulier pour les hautes valeurs de p_T. En conséquence, le découpage en régions DM-p_T a été limité à 8 régions. Ces résultats sont présentés dans la Figure 10.

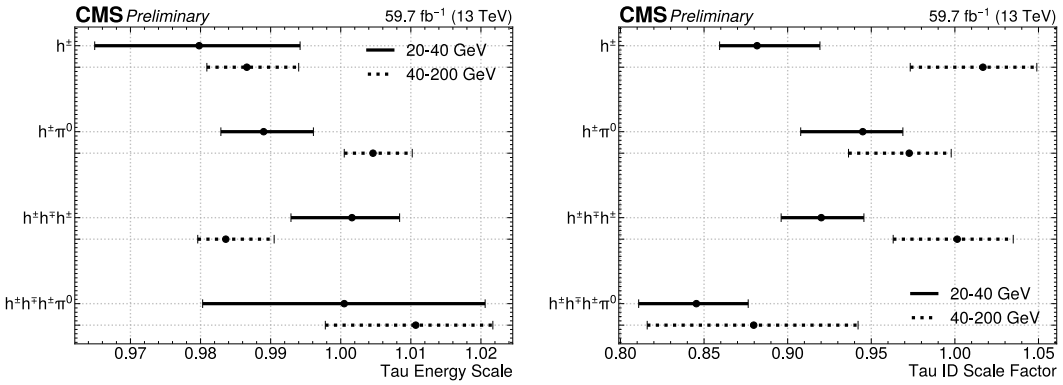


Figure 10: Synthèse des TES (gauche) et ID SF (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la log-vraisemblance négative (NLL).

Mesures combinées dans les jeux de données Run 3 2022

La méthode combinée développée pour UL2018 est appliquée aux données 2022 (preEE et postEE), en utilisant une sélection plus stricte sur la masse transverse ($m_T < 40\,\text{GeV}$). La faible statistique disponible, notamment en preEE, limite les mesures à des régions par DM, avec une plage TES élargie à $\pm 5\%$.

Les résultats pour preEE (Figure 11) montrent des TES compatibles avec l'unité pour tous les modes. Une amélioration marquée est observée pour DM0 par rapport aux données initiales du fait d'un bug dans la modélisation du spin du τ. Les ID SFs sont également proches de l'unité à 2σ près, sauf pour DM0, qui présente un déficit. Cette anomalie pourrait s'expliquer par une migration de vrais candidats DM0 vers DM1, renforcée par l'ID SF$>$1 observé pour DM1.

Les résultats postEE (Figure 12) confirment la tendance observée pour les données UL2018 : TES proches de l'unité et ID SF décroissant avec le nombre de particules chargées dans la désintégration du τ_h. Les distributions post-ajustement reproduisent bien les données, et la corrélation entre TES et ID SF reste faible, permettant de mesurer TES et les ID SFs avec la méthode de mesure séparée.

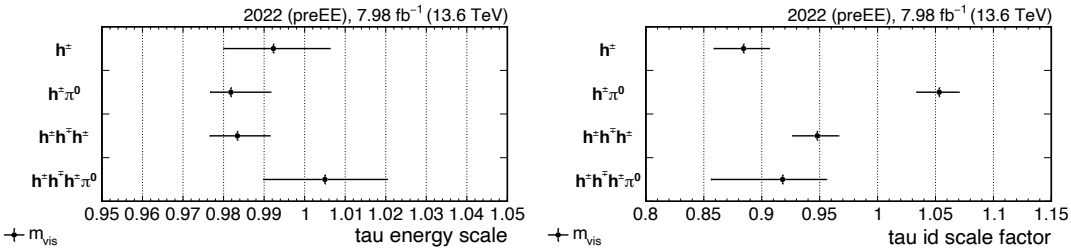


Figure 11: Synthèse des TES (gauche) et ID SFs (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la NLL.

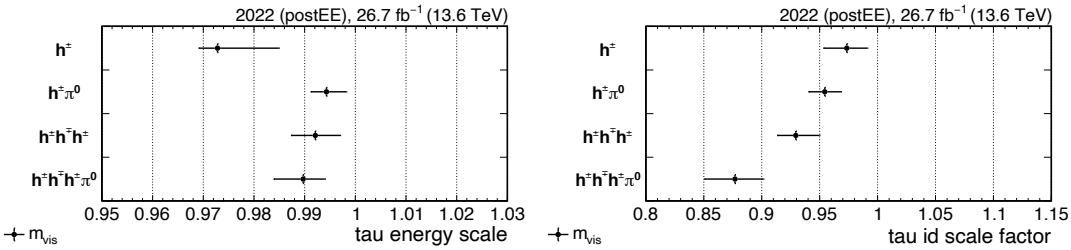


Figure 12: Synthèse des TES (gauche) et ID SFs (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la NLL.

Conclusion

Cette section a présenté la mesure combinée des TES et ID SFs. Développée sur les données UL2018, la méthode a démontré sa robustesse et a été adaptée aux données 2022, après correction d'un bug affectant la simulation DY.

Malgré son exigence statistique, l'ajustement combiné permet une extraction cohérente des paramètres en exploitant à la fois la forme et la normalisation du signal. Il s'avère particulièrement adapté aux échantillons de données avec beaucoup d'événements, tandis que l'ajustement individuel reste préférable dans des cas à faible statistique ou avec forte dépendance cinématique.

Les résultats obtenus sur les données 2022 confirment la validité et la précision de la méthode, en accord avec les approches indépendantes. Cette calibration combinée constitue une base fiable pour les analyses Run 3 impliquant des τ_h, et a été approuvée par la Collaboration pour être présentée dans le résultat public CMS-PAS-TAU-24-001 [35] et sera utilisée pour 2024.

Mesure de la violation de CP dans les désintégrations $H \to \tau_h \tau_h$

Cette analyse vise à mesurer une éventuelle violation de CP dans les désintégrations $H \to \tau\tau$, en se concentrant sur l'état final $\tau_h\tau_h$ dans les données Run 3 (2022–2023) collectées à $\sqrt{s} = 13{,}6$ TeV. Malgré un rapport d'embranchement important (42% des désintégrations $H \to \tau\tau$), ce canal est dominé par les bruits de fond Drell–Yan ($Z/\gamma^* \to \tau\tau$) et les processus *QCD multijets*, c'est-à-dire des événements contenant plusieurs jets issus de quarks et gluons produits par l'interaction forte, ainsi que par une résolution modérée de la masse du boson de Higgs due à la présence de neutrinos.

L'observable sensible à la violation de CP, ϕ_{CP}, est reconstruite à partir de la topologie des désintégrations τ. Sa résolution dépend fortement du mode de désintégration : les modes à trois prongs offrent plus d'information, tandis que des méthodes alternatives sont requises pour les désintégrations plus simples. Des techniques multivariées et une catégorisation fine des événements sont utilisées pour améliorer la sensibilité, en exploitant les différences topologiques entre le signal et le bruit de fond.

Mes contributions portent sur l'estimation des bruits de fond jets $\to \tau_h$ via la méthode des Fake Factors, ainsi que sur le développement de techniques de reconstruction des moments des τ basées sur l'apprentissage automatique, spécifiques à la recherche de violation de CP. J'ai également participé à la mise en œuvre de la sélection, des calibrations et de la catégorisation dans un framework avec un structure de données colonnaire.

Cette étude s'inscrit dans la continuité de l'analyse Run 2 [17], qui a exclu un état CP-impair pur à 95% CL et contraint l'angle de mélange CP à $\phi_{\tau\tau} = -1° \pm 18°$ à 68% CL. Les progrès apportés dans le cadre Run 3 visent à améliorer cette sensibilité malgré les défis statistiques et expérimentaux du canal $\tau_h\tau_h$.

Technologies d'analyse : `ColumnFlow`

L'analyse Run 3 exploite `ColumnFlow`, un cadre basé sur un modèle de données colonnaire, qui remplace les formats classiques `ROOT` [36], peu adaptés aux grands volumes et aux calculs vectorisés requis par les méthodes modernes, notamment l'apprentissage automatique. Le stockage par variable permet des opérations plus rapides et une meilleure gestion mémoire, facilitant aussi le parallélisme et la scalabilité. `ColumnFlow` s'appuie sur des outils comme `AwkwardArray` [37], `NumPy` [38] et le format `Parquet` [39] pour optimiser la lecture des données. La gestion automatisée du workflow repose sur le package *law* [40], assurant traçabilité, modularité et reproductibilité via une orchestration fine des tâches, de la calibration jusqu'à

la production d'histogrammes.

Estimation du taux de faux τ_h

Un bruit de fond important dans les états finaux incluant des τ_h provient de la mauvaise identification de jets initiés par quarks ou gluons en candidats τ_h, dominé par les processus QCD multijets. La simulation modélise imparfaitement ce bruit de fond, principalement en raison du nombre limité d'événements générés.

La méthode des facteurs de faux (*Fake Factor*, F_F) est une technique *data-driven*, c'est-à-dire basée directement sur les données expérimentales plutôt que sur la simulation, reposant sur une méthode ABCD utilisant des critères orthogonaux définis à partir des valeurs des discriminants d'identification des τ_h et de la charge des paires $\tau_h\tau_h$.

Les régions sont définies comme suit :

- **Région A (AR)** : paire $\tau_h\tau_h$ opposée en charge (OS), τ_h principal échoue VTight mais passe VVVLoose DeepTau,
- **Région B (DR)** : paire même charge (SS), τ_h principal échoue VTight, passe VVVLoose,
- **Région C (DR)** : paire SS, τ_h principal passe VTight,
- **Région D (SR)** : paire OS, les deux τ_h passent VTight.

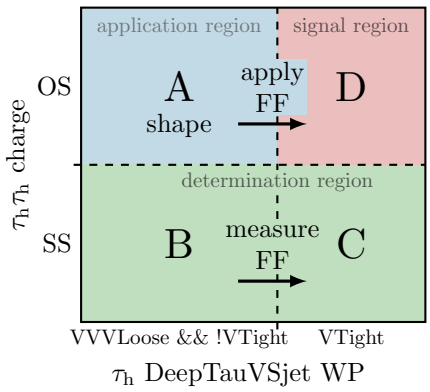


Figure 13: Schématisation de la méthode ABCD utilisée pour estimer le bruit de fond $j \to \tau_h$.

Le facteur de faux est défini par :

$$F_F = \frac{N_{\text{pass DR}}}{N_{\text{fail DR}}}, \tag{1}$$

avec $N_{\text{pass DR}}$ et $N_{\text{fail DR}}$ les événements du bruit de fond soustraits aux données dans les régions C et B. L'estimation du bruit de fond dans la région signal (SR) s'obtient par :

$$N_{\text{SR}} = N_A \cdot F_F, \tag{2}$$

où N_A est le nombre d'événements du bruit de fond soustrait aux données en région A.

Les fake factors sont déterminés selon la multiplicité en jets ($N_{\text{jet}} = 0, 1, \geq 2$) et le mode de désintégration (DM0,1,2,10) du τ_h principal, paramétrés en fonction de p_T via un polynôme du second ordre, constant au-delà de 100 GeV.

Les distributions des variables clés de l'analyse sont présentées en Figure 14 après la détermination et la mise en œuvre des fake factors.

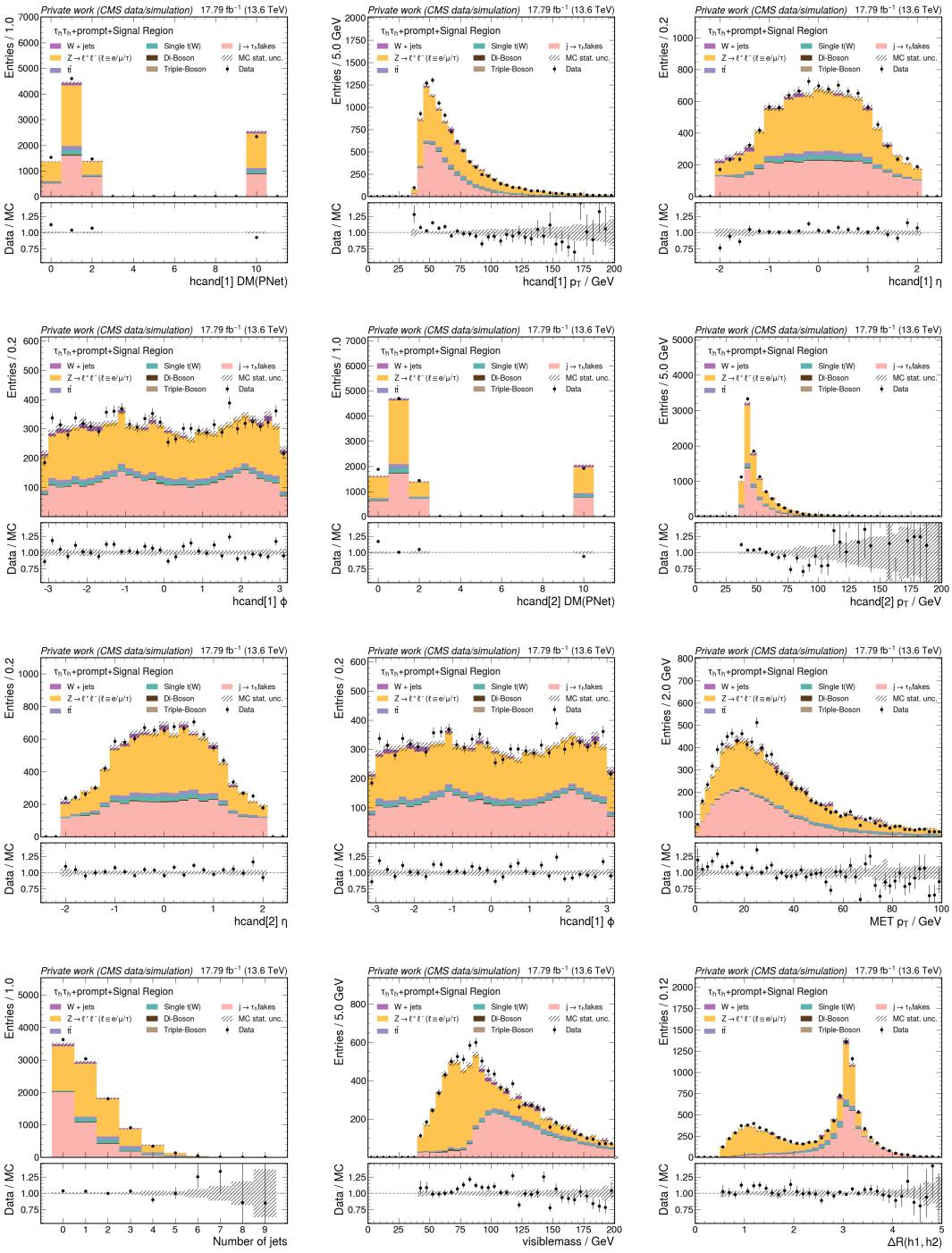


Figure 14: Distributions dans la région signal pour l'ère BPix 2023 avec le bruit de fond multijet modélisé par la méthode des fake factors.

Reconstruction du moment τ et observable CP

SVFit et FastMTT

La reconstruction de la masse invariante des paires $\tau\tau$ est essentielle pour distinguer le signal $H \to \tau\tau$ du fond dominant $Z \to \tau\tau$. La présence de neutrinos dans la désintégration des τ entraîne une perte d'énergie non détectée, dégradant la résolution lorsque l'on utilise uniquement les produits visibles.

L'algorithme *SVFit* [41] combine les produits visibles avec l'énergie transverse manquante (MET) et sa covariance via une méthode de vraisemblance dynamique intégrant les variables inconnues de désintégration $(\theta_{\text{inv}}, \phi_{\text{inv}}, m_{\nu\nu})$. Il calcule la probabilité de la masse hypothétique $m_{\tau\tau}$ par intégration sur ces paramètres, améliorant la résolution de masse d'environ 30%, équivalent à un gain de 70% en luminosité effective.

Pour réduire la lourdeur numérique, *FastMTT* [41] simplifie en supposant que les neutrinos sont colinéaires aux produits visibles, permettant une solution analytique. La Figure 15 compare leur résolution sur simulation Drell–Yan : FastMTT (écart-type 0.16) est comparable à SVFit (0.17).

Durant cette thèse, les deux méthodes ont été implémentées et adaptées en Python pour faciliter leur usage.

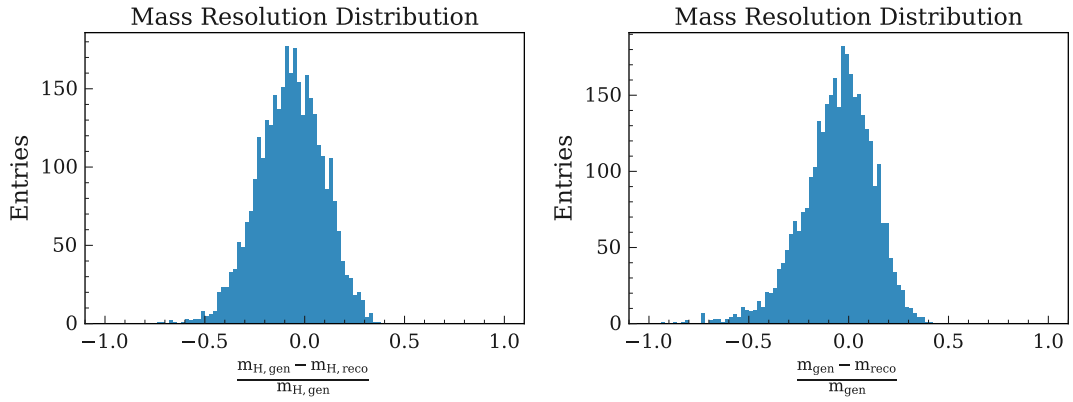


Figure 15: Résolution de la masse di-τ sur DY simulé avec FastMTT (gauche) et SVFit (droite).

Étude machine learning dans le canal $\rho\rho$

Le canal $\rho\rho$, représentant 25,9% des désintégrations hadroniques τ_h. Au niveau générateur, la méthode du vecteur polarimétrique (PV) offre une meilleure sensibilité à l'angle CP ϕ_{CP} que la méthode du plan de désintégration (DP), avec une

asymétrie CP $A_{\text{gen}} = 0{,}418$ contre 0,332, où

$$A = \frac{1}{N_{\text{bins}}} \sum_i \left| \frac{N_i^{\text{odd}} - N_i^{\text{even}}}{N_i^{\text{odd}} + N_i^{\text{even}}} \right|,$$

et $N_i^{\text{odd,even}}$ sont les nombres d'événements pondérés pour les hypothèses CP-impair et CP-pair.

Toutefois, au niveau reconstruit, la sensibilité PV chute ($A_{\text{reco}} = 0{,}010$) en raison de la perte d'information sur les neutrinos, tandis que DP reste plus performante ($A_{\text{reco}} = 0{,}265$), expliquant le choix de l'utilisation de la méthode DP pour $\rho\rho$ en Run 2.

Le vecteur polarimétrique est défini dans le référentiel τ_h par

$$\vec{h} = \mathcal{N}\big[2(q \cdot N)\vec{q} - q^2\vec{N}\big], \quad \text{avec} \quad \vec{q} = \vec{p}_{\pi^\pm} - \vec{p}_{\pi^0}, \quad \vec{p}_\nu = \vec{p}_\tau - (\vec{p}_{\pi^\pm} + \vec{p}_{\pi^0}).$$

Les résultats montrent que la régression complète des quadrivecteurs via un modèle LBN ne sépare pas efficacement les hypothèses CP avec la méthode PV. En revanche, la combinaison d'une régression partielle sur p_T avec des informations reconstruites atteint une sensibilité comparable à la méthode DP classique (cf. Figure 16). L'utilisation du p_T générateur n'améliore pas significativement la sensibilité, suggérant que la régression prédit correctement p_T, mais que les progrès doivent venir d'une meilleure estimation des angles η et ϕ.

Les développements futurs viseront donc à améliorer la reconstruction angulaire, notamment par des modèles hybrides intégrant apprentissage supervisé et contraintes cinématiques, pour accroître la sensibilité CP dans ce canal.

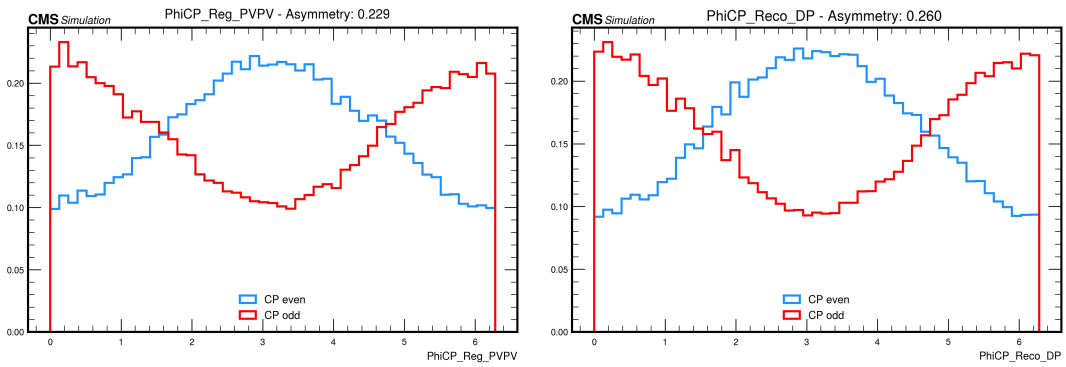


Figure 16: Comparaison des distributions de ϕ^*_{CP} obtenues avec : p_T regréssés combinée aux informations reconstruites (PVPV, à gauche), et méthode classique du plan de désintégration (DP, à droite) avec les informations reconstruites.

Catégorisation et discrimination du signal

Les événements sont classés pour exploiter les différences topologiques comme la multiplicité en jets ($N_{\text{jet}} = 0, 1, 2, \geq 3$) et les DMs.

Dans chaque catégorie, un classificateur fondé sur des arbres de décision boostés (BDT) distingue trois classes : signal Higgs, bruit de fonds avec vrais τ ($Z \to \tau\tau$) et bruit de fonds avec jets ou leptons mal identifiés comme τ_h. La probabilité la plus élevée sert de discriminant final pour l'ajustement de vraisemblance.

Un ajustement simultané de vraisemblance maximale, réalisé avec COMBINE [42], permet d'extraire l'angle de mélange CP $\alpha_{H\tau\tau}$ via la distribution paramétrée de l'observable ϕ_{CP}, combinant hypothèses CP-pair et CP-impair, avec $\alpha_{H\tau\tau}$ libre dans $[-90°, +90°]$.

Un scan sur le jeu de données Asimov PreEE 2022 ($26.671\,\text{fb}^{-1}$) limité aux seules incertitudes statistiques donne une sensibilité de 0.85σ avec le seul canal $\tau_h\tau_h$. Par extrapolation à la statistique attendue du Run 3, cette sensibilité pourrait atteindre environ 3.0σ. À titre de comparaison, l'analyse du Run 2, combinant tous les canaux et incluant les incertitudes systématiques, avait atteint une sensibilité équivalente [17]. Ces résultats confirment la robustesse de la stratégie d'analyse et mettent en évidence un potentiel d'amélioration significatif avec l'ensemble des données du Run 3.

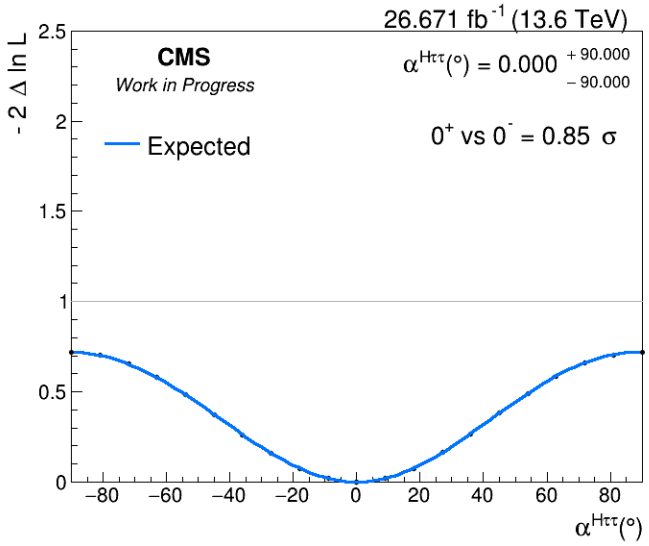


Figure 17: Scan du profil de vraisemblance de l'angle $\alpha_{H\tau\tau}$ extrait d'un fit simultané sur toutes les catégories avec le jeu de données Asimov PreEE 2022 [43] ($26.671\,\text{fb}^{-1}$), incertitudes statistiques uniquement.

Conclusion de l'analyse

Cette section a présenté la stratégie d'analyse du canal $H \to \tau_h\tau_h$ avec les premières données Run 3 à $\sqrt{s} = 13{,}6$ TeV. Ce canal, malgré un fort taux d'embranchement, reste difficile à analyser en raison d'une résolution de masse limitée et de bruit de fonds conséquents Drell–Yan et multijets.

J'ai contribué au développement et à l'intégration complète de l'analyse dans un framework colonnaire, notamment la sélection, la catégorisation, la calibration, et la validation de la méthode Fake Factor adaptée aux conditions Run 3. Pour améliorer la reconstruction des observables CP, un modèle basé sur les Lorentz Boost Networks a été entraîné sur l'état final $\rho\rho$.

Des architectures plus avancées, comme les Graph Neural Networks, seront explorées pour mieux modéliser les corrélations angulaires. La catégorisation sera étendue à différents modes de production du Higgs afin d'optimiser la sensibilité.

Les scans préliminaires sur les premières données Run 3 donnent des contraintes prometteuses sur l'angle de mélange CP, comparables au Run 2 malgré une moindre luminosité. À court terme, une étude complète de sensibilité sur Run 3 sera réalisée. L'objectif final est de mesurer la violation de CP dans $H \to \tau\tau$ avec Run 3, puis de combiner avec Run 2 pour atteindre la contrainte la plus précise sur l'angle de mélange CP Higgs–τ.

Conclusion

Le modèle standard décrit efficacement les particules et leurs interactions, et le mécanisme CKM permet en principe la violation de CP dans le secteur des quarks. Cependant, les mesures expérimentales montrent que cette violation est insuffisante pour expliquer l'asymétrie matière-antimatière observée, motivant la recherche de sources supplémentaires de violation de CP. La découverte du boson de Higgs en 2012 offre une voie pour étudier ces effets via ses couplages aux leptons τ. Cette thèse se concentre sur la recherche de violation de CP dans $H \to \tau_h\tau_h$ à partir des données Run 3 du CMS, utilisant un observable construit à partir des plans définis par les produits de désintégration des τ, sensibles aux corrélations de spin, bien qu'aucune observation expérimentale de violation de CP dans ce couplage n'ait encore été rapportée.

La première étude a porté sur la qualification des modules pixel Phase-2 via une campagne d'irradiation à l'IPHC, avec la mise en place d'un environnement froid et sec et un protocole précis. L'irradiation d'un module CROC-V1 a confirmé la robustesse du dispositif, isolant un défaut mécanique sans lien avec l'irradiation, validant ainsi la méthodologie pour la campagne d'irradiation des modules à l'IPHC.

La deuxième étude concerne la mesure combinée des facteurs d'échelle d'énergie

et l'efficacité d'identification des τ_h, cruciale pour les analyses CP-sensibles. Une méthode d'ajustement simultané, développée et validée sur les données Run 2 puis appliquée au Run 3, a permis d'extraire ces facteurs avec une bonne stabilité. L'ajustement combiné est privilégié pour les futures analyses, bien que la mesure différenciée en p_T reste limitée par la statistique de données disponible.

Enfin, la mesure de la violation de CP dans $H \to \tau_h\tau_h$ a été réalisée sur les premières données Run 3. J'ai contribué à l'insertion de l'analyse dans un framework colonnaire, incluant sélection, catégorisation, calibration, et estimation in-situ du bruit de fond QCD. Pour améliorer la reconstruction CP-sensible, un Lorentz Boost Network a été développé pour reconstruire le quadrivecteur complet des τ_h dans la catégorie $\rho\rho$. Cette approche permet d'exploiter la méthode polarimétrique, plus sensible, bien que la résolution angulaire reste encore à optimiser. Des architectures avancées telles que les Graph Neural Networks offrent des pistes prometteuses.

Les résultats préliminaires sur Run 3 sont encourageants, montrant une bonne sensibilité sur l'angle de mélange CP $\alpha_{H\tau\tau}$. Les perspectives incluent une extension de la catégorisation aux différents modes de production du Higgs et la combinaison avec les résultats Run 2, pour fournir une contrainte encore plus précise sur la violation de CP dans le couplage Higgs–τ.

Contents

Remerciements

Résumé **1**

Introduction **1**

1 Theoretical Background **3**

I. The Standard Model of Particle Physics 4

I.1. Symmetries 4

I.2. The Fundamental Particles and Forces 5

I.3. The Brout–Englert–Higgs Mechanism 7

II. Higgs boson Measurements 9

II.1. Production Modes 9

II.2. Decay modes 11

II.3. Couplings 12

III. Search of CP Violation in the Higgs Sector 13

III.1. Motivation for the Search for CP Violation 13

III.2. Anomalous Higgs Couplings to Gauge Bosons 15

III.3. CP Violation in Yukawa Couplings 17

IV. CP-Sensitive Observables in $H \to \tau\tau$ 21

IV.1. Polarimetric Method 24

IV.2. Impact Parameter Method 25

IV.3. Decay Plane Method 26

IV.4. Common Notation 28

2 From the CMS Phase-1 Detector at LHC to the Phase-2 Upgrade at HL-LHC **31**

I. The LHC 33

I.1. Accelerator Complex . . . 33
I.2. LHC Design and Specifications . . . 33
I.3. Schedule of Operations . . . 38
II. The CMS Experiment . . . 39
II.1. Coordinate System . . . 40
II.2. Anatomy of the Detector . . . 42
II.3. Trigger System . . . 53
III. Silicon Tracking Detectors for High Energy Physics . . . 53
III.1. Silicon Detectors Technology . . . 54
III.2. Signal Generation . . . 58
III.3. Radiation Damage . . . 60
IV. The CMS Tracker Phase-2 Upgrade . . . 63
IV.1. The High Luminosity LHC Upgrade . . . 63
IV.2. CMS Phase-2 Upgrade . . . 66
IV.3. Phase-2 Upgrade of the CMS Tracker . . . 67
IV.4. Phase-2 Upgrade of the CMS Inner Tracker . . . 69

3 Development and Validation of the Irradiation Setup for Phase-2 Pixel Modules at CYRCé 77
I. Experimental Setup for Irradiation . . . 78
I.1. The CYRCé Cyclotron at IPHC . . . 78
I.2. Irradiation Line and Irradiation Cold Box . . . 80
II. Prototyping a Cold Test Box . . . 82
II.1. Motivation and Requirements . . . 82
II.2. Design of the Cold Test Box . . . 83
II.3. Test Protocol and Validation Campaign . . . 84
II.4. Performance with Irradiated Modules . . . 89
II.5. Summary and Outlook . . . 90
III. Testing of CROC-V1 Before and After Irradiation . . . 91
III.1. Test Setup and Readout Configuration . . . 91
III.2. Pre-Irradiation Characterization . . . 92
III.3. Post-Irradiation Measurements . . . 96
IV. Conclusion . . . 102

4 Hadronic τ Reconstruction and Identification 103
I. Object Identification and Reconstruction in CMS . . . 104
I.1. Tracks and Vertices . . . 104
I.2. Particle-Flow Event Reconstruction . . . 105
I.3. Muons . . . 107
I.4. Electrons . . . 109
I.5. Jets . . . 110

I.6. Missing Transverse Energy (MET) 112
II. Hadronic τ Reconstruction and Identification 113
II.1. Properties of Hadronic τ Decays 113
II.2. Reconstruction with the Hadron Plus Strips (HPS) Algorithm 114
II.3. Identification with the Deep Neural Network DeepTau . . . 116

5 Combined Measurement of the τ_h Identification Scale Factor and Energy Scale 123
I. Baseline Event Selection with $Z \to \tau\tau$ events 124
I.1. Trigger Selections 124
I.2. Baseline Selection for the $\mu\tau_h$ Channel 125
I.3. Baseline Selection for the $\mu\mu$ Channel 126
II. Correction 127
II.1. Pileup Corrections 127
II.2. Electrons and Muons Efficiencies 127
II.3. Z p_T Corrections 128
III. Background Estimation 128
III.1. Background from QCD Multijet Processes 128
III.2. Normalization of the W + jets Background 129
IV. Combined Measurement of τ_h Identification and Energy Scale Factors 129
IV.1. Measurement Technique 129
IV.2. Fit Configuration 131
IV.3. Method Validation with the UL2018 Dataset 131
IV.4. Combined Measurements in the 2022 Run 3 Datasets 147
IV.5. Comparison with Individual Fit Results 153
V. Conclusion 153

6 CP measurement in $H \to \tau_h\tau_h$ decays 155
I. Analysis Technologies 157
I.1. Columnflow 157
I.2. Deep Learning and Neural Networks 161
II. Event Samples 164
II.1. Data Event Samples 164
II.2. Simulated Event Samples 164
II.3. Signal Modeling 166
III. Selection 167
III.1. Trigger and Preselection 167
III.2. Final Selection for $\tau_h\tau_h$ Events 167
III.3. Candidate Pair Selection and Extra Lepton Veto 169
IV. Corrections 169
IV.1. General Corrections 169

IV.2. Impact Parameter Vector Calibration 171
V. Fake Rate Estimation for τ_h 172
V.1. Background Overview 172
V.2. Methodology 172
V.3. Systematic Uncertainties 177
VI. τ_h Momentum Reconstruction and CP Observable 178
VI.1. SVFit and FastMTT 178
VI.2. Machine Learning Studies in the $\rho\rho$ Channel 179
VII. Event Categorization and Signal Discrimination 186
VIII. Sensitivity Scan of ϕ_{CP} 186
IX. Conclusion 187

Conclusion **189**

A Combined Measurement of τ_h Identification and Energy Scale Factor **193**
I. Combined fit of the TES and the ID SF by decay modes and $p_T(\tau_h)$ 194
II. Bug in Pythia 8 196
III. First Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets 197
IV. Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets 198
IV.1. PreEE 2022 198
IV.2. PostEE 2022 199
IV.3. Individual fit results 199

B CP measurement in $H \to \tau_h\tau_h$ decays **205**
I. Simulated Samples 205
I.1. Simulated Background Samples 205
I.2. Simulated Signal Samples 205
II. Performance of the Network 207

Bibliography **211**

List of Figures **229**

List of Tables **241**

Introduction

The Standard Model (SM) of particle physics describes the fundamental constituents of matter and their interactions. The Brout–Englert–Higgs mechanism, proposed in 1964, explains the origin of masses of the particles and predicts the existence of a scalar boson, the Higgs boson. A particle compatible with the scalar Higgs boson was discovered in 2012 by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC), completing the SM framework and allowing the precise study of the electroweak symmetry breaking.

While the SM accurately describes most observed phenomena, it does not explain the matter-antimatter asymmetry in the universe. Charge-Parity (CP) violation is a necessary condition for this asymmetry, but the CP violation present in the SM is insufficient to account for the observed imbalance. The Higgs sector is therefore a promising place to search for new sources of CP violation beyond the SM.

Current measurements of the Higgs boson couplings to vector bosons strongly favor a purely scalar nature, consistent with the SM prediction, and exclude a pure pseudoscalar component. However, CP violation may still arise in the Yukawa couplings through the mixing of scalar and pseudoscalar components.

This thesis focuses on the search for CP violation in the Higgs Yukawa coupling to hadronic τ leptons ($H \rightarrow \tau_h \tau_h$). Since the τ lepton decay products retain spin correlations that are sensitive to the CP properties of the interaction, this decay channel provides a powerful probe. While previous analyses based on Run 2 data have been limited by their statistical precision, the increased dataset from Run 3 offers an opportunity to significantly enhance the sensitivity to potential CP-violating effects.

Accurate reconstruction and calibration of τ_h decays are essential for this analysis, as well as for any physics study involving such particles. Precise knowledge of the τ_h energy scale and identification efficiency is particularly important for reducing systematic uncertainties and improving the sensitivity of measure-

ments that rely on τ_h final states. This thesis presents a dedicated study aimed at developing a novel technique for the simultaneous extraction of the τ_h energy scale and identification scale factors. By combining both into a combined fit strategy in order to handle the correlation between these two scale factors.

On the other hand, the High-Luminosity LHC (HL-LHC), scheduled to start operations in 2030, will deliver proton–proton collisions at unprecedented instantaneous luminosities. This upgrade will increase the accumulated dataset by an order of magnitude compared to Run 2 and Run 3, enabling more precise studies of the Higgs boson and rare processes. However, the extreme conditions at the HL-LHC also pose significant challenges for the CMS detector, necessitating substantial upgrades.

In preparation for HL-LHC, the CMS detector will undergo a major Phase-2 upgrade. The HL-LHC environment requires radiation-hard silicon pixel detectors and efficient cooling systems to ensure stable and reliable operation. This thesis also presents work related to the preparation of irradiation campaigns at IPHC's CYRCé cyclotron, including studies on the design and validation of a cold test box used for the characterization of irradiated modules. In addition, the first irradiation of a module was performed on the new irradiation line at CYRCé, with characterization measurements conducted both before and after irradiation.

The thesis is organized as follows. Chapter 1 introduces the theoretical background, including the SM, the Higgs mechanism, and CP violation studies in CMS. Chapter 2 describes the CMS detector and the Phase-2 tracker upgrade. Chapter 3 presents the irradiation campaign and cold box development. Chapter 4 details the reconstruction of hadronic τ leptons. Chapter 5 discusses the development and validation of the combined fit for τ_h calibration. Chapter 6 covers the search for CP violation in $H \to \tau_h\tau_h$ decays, including analysis strategy; estimation of the τ_h quark gluon fake rate, and machine learning studies for the τ_h momentum reconstruction.

Chapter 1

Theoretical Background

Content

I. The Standard Model of Particle Physics 4

I.1. Symmetries 4

I.2. The Fundamental Particles and Forces 5

I.3. The Brout–Englert–Higgs Mechanism 7

II. Higgs boson Measurements 9

II.1. Production Modes 9

II.2. Decay modes 11

II.3. Couplings 12

III. Search of CP Violation in the Higgs Sector 13

III.1. Motivation for the Search for CP Violation 13

III.2. Anomalous Higgs Couplings to Gauge Bosons 15

III.2.a. Anomalous Higgs to Electroweak Vector Boson Couplings 16

III.2.b. Higgs to Gluon Effective Couplings 17

III.3. CP Violation in Yukawa Couplings 17

III.3.a. CP Violation in the Top Yukawa Coupling . . 18

III.3.b. CP Violation in the τ Yukawa Coupling 19

IV. CP-Sensitive Observables in $H \to \tau\tau$ 21

IV.1. Polarimetric Method 24

IV.2. Impact Parameter Method 25

IV.3. Decay Plane Method 26
IV.4. Common Notation 28

This chapter provides an introduction to the key theoretical concepts relevant to understand and motivate the analysis presented in this thesis. It first introduces the Standard Model of Particle Physics (SM) which is the theory describing fundamental particles and the forces acting between them. It is followed by the motivation behind the searches of Charge-Parity (CP) violation, along with a summary of the experimental status of the CP searches in the Higgs sector at LHC. The chapter concludes with the description of the different techniques used to reconstruct the CP-sensitive observables in the $H \to \tau\tau$ analysis.

I. The Standard Model of Particle Physics

The SM [1–3] is a model that represents the best of our current understanding of the subatomic world. It describes the electromagnetic, weak, and strong nuclear forces, as well as the interactions of fundamental particles via these forces. For completeness and further details beyond what is described in this section, the reader is referred to Refs. [44–46].

I.1. Symmetries

Symmetries are fundamental to our understanding of the laws of nature. According to Noether's theorem, every continuous symmetry of a physical system implies a conserved quantity [47]. In the SM, the mathematical description of particles and their interactions is encoded in a Lagrangian, whose form is dictated by these symmetries.

The SM is constructed within the framework of special relativity [48], and as such, it is invariant under the global Poincaré group [49], which includes spacetime translations and Lorentz transformations. These symmetries lead to the conservation of energy, momentum, and angular momentum.

In addition to spacetime symmetries, the SM Lagrangian is invariant under local gauge transformations associated with the symmetry group

$$SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}.$$

The strong interaction is described by the non-Abelian gauge theory of quantum chromodynamics (QCD) based on the color symmetry $SU(3)_{\mathrm{C}}$ [7]. The electroweak interaction unifies the weak and electromagnetic forces via the $SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ symmetry, as introduced by Glashow [4], Weinberg [5], and Salam [6]. After electroweak symmetry breaking via the Brout–Englert–Higgs mechanism [8, 9], electric charge emerges as the conserved quantity for electromagnetic processes.

The SM also includes discrete symmetries such as parity (P), charge conjugation (C), and time reversal (T). While these are symmetries of many interactions, the weak interaction violates P and C maximally [50], and also violates the CP symmetry, as first observed in neutral kaon decays [51]. The combined CPT symmetry, however, remains an exact symmetry of any local, Lorentz-invariant quantum field theory [52, 53].

The search for CP violation is of major interest for explaining the matter–antimatter asymmetry of the universe and is further discussed in Section III.

I.2. The Fundamental Particles and Forces

The concept of *elementary particle* refers to particles that, to the best experimental knowledge, are point-like and not composed of smaller constituents. This concept has evolved as experimental resolution improved, revealing substructures over time. In the SM, elementary particles are divided into two categories, shown in Figure 1.1. *Fermions* are spin-$\frac{1}{2}$ particles composing matter (and antimatter), while *bosons* are spin-1 particles acting as force carriers.

There are twelve fundamental fermions: six quarks, which participate in all three fundamental interactions and carry color charge, and six leptons, which interact only via the electromagnetic or weak interactions, unified as the *electroweak* interaction. Each fermion has a corresponding antiparticle with opposite quantum numbers. These fermions are organized into three generations with identical gauge quantum numbers but differing in masses. Neutrino masses and mixing are not included in the SM but require extensions beyond it [54, 55].

The strong interaction is mediated by eight gluons, while the electromagnetic (EM) force is mediated by the photon (γ), and the weak force by the $W^{\pm}$ and Z bosons. Quarks and gluons both carry color charge, and the strong interaction is described by Quantum Chromodynamics (QCD) [7, 57], governed by an $SU(3)_{\mathrm{C}}$ local gauge symmetry. QCD exhibits two key properties: confinement, meaning quarks are never observed as free particles but only in color-singlet hadronic bound states, the *hadrons*; and asymptotic freedom, wherein the strong coupling decreases at high energy or short distances. The strong coupling constant α_s is defined as

$$\alpha_s(Q^2) = \frac{g_s^2(Q^2)}{4\pi}, \tag{1.1}$$

where g_s is the QCD gauge coupling evaluated at the scale Q [15].

The electroweak sector unifies the electromagnetic and weak interactions, governed by the $SU(2)_L \times U(1)_Y$ symmetry [4–6]. The weak isospin generators $I_i = \frac{1}{2}\sigma_i$, where σ_i are the Pauli matrices [58], generate $SU(2)_L$, and the weak hypercharge Y generates $U(1)_Y$. Electric charge Q is related to these quantum

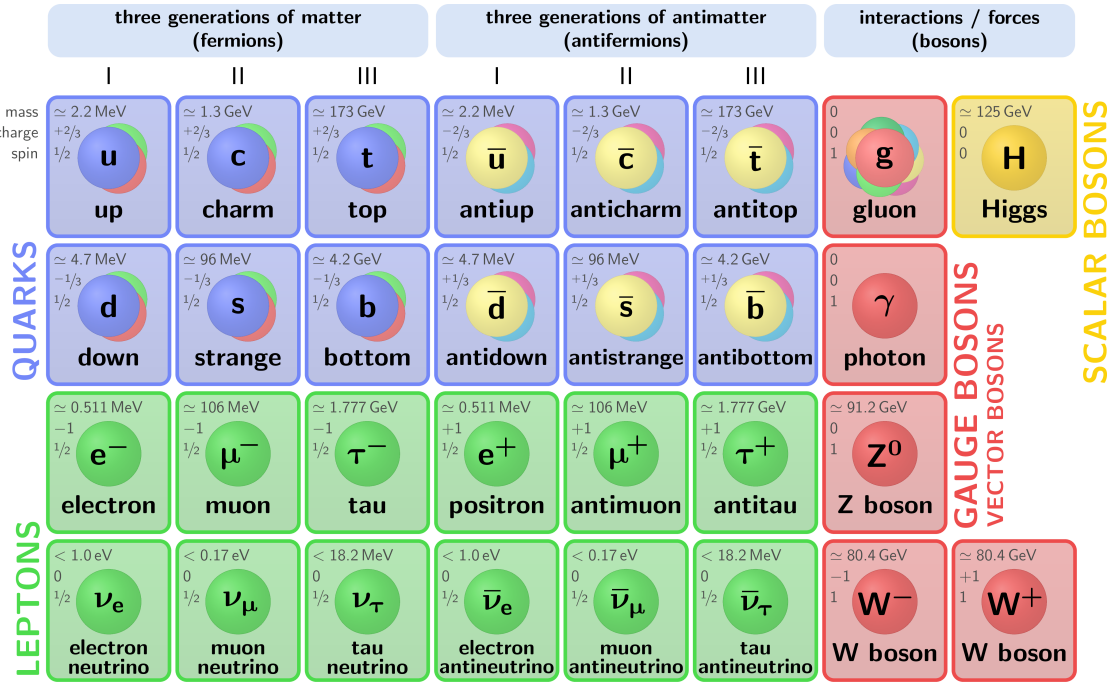


Figure 1.1: Summary of all SM elementary particles and their corresponding antiparticles. The three generations of quarks/anti-quarks (blue) and leptons/anti-leptons (green) are shown in the first six columns. The seventh column shows the force carriers of the SM (red), while the Higgs boson (yellow) is shown on the right as the only spin-0 particle of the SM. Figure adapted from [56].

numbers via the Gell-Mann–Nishijima relation [59, 60]:

$$Q = I_3 + \frac{Y}{2}. \tag{1.2}$$

The associated gauge fields W_μ^i and B_μ mix to form the physical bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{1.3}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \tag{1.4}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \tag{1.5}$$

where A_μ and Z_μ are the photon and Z boson fields, and θ_W is the weak mixing angle [5], defined by

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}, \quad \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \tag{1.6}$$

with g and g' the $SU(2)_L$ and $U(1)_Y$ gauge couplings respectively.

The charged weak interaction couples only to left-handed fermions, leading to maximal parity violation. Mass terms for gauge bosons and fermions explicitly break gauge invariance. This is resolved in the SM by spontaneous electroweak symmetry breaking via the Brout–Englert–Higgs mechanism [8, 9] (see Section I.3.) while preserving gauge invariance:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \tag{1.7}$$

Neutral weak currents were experimentally discovered in 1973 in the Gargamelle experiment [61], followed by the observation of the $W^{\pm}$ and Z bosons at CERN in the early 1980s [62, 63], confirming the electroweak unification.

I.3. The Brout–Englert–Higgs Mechanism

The Brout–Englert–Higgs mechanism, proposed in the 1960s [8, 9], explains the masses of the $W^{\pm}$ and Z^0 bosons without violating gauge invariance. Initially, gauge theories predicted massless gauge bosons, such as the photon. However, the short-range nature of the weak force requires its mediators to be massive. Direct mass terms for gauge bosons violate gauge invariance, rendering the theory non-renormalizable. The Higgs mechanism resolves this by introducing a scalar field, the Higgs field, which spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry and generates masses for the weak bosons while leaving the photon massless.

The SM is extended by introducing a scalar field Φ, known as the Higgs field, which transforms as a doublet under the electroweak gauge group $SU(2)_L \times U(1)_Y$:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad \Phi \sim (2, +\tfrac{1}{2}), \tag{1.8}$$

where $(2, +\frac{1}{2})$ denotes the representation under $SU(2)_L$ and the hypercharge $Y = +\frac{1}{2}$ under $U(1)_Y$. The Higgs field is a singlet under the strong interaction group $SU(3)_C$.

The scalar potential associated with Φ is chosen as

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

with $\mu^2 > 0$ and $\lambda > 0$. These conditions ensure the potential exhibits the characteristic "Mexican hat" shape necessary for spontaneous symmetry breaking (SSB). In particular, the negative quadratic term $-\mu^2 \Phi^\dagger \Phi$ drives the Higgs field to acquire a non-zero vacuum expectation value (VEV), minimizing the potential away from zero.

The Higgs field thus develops a VEV:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v = \sqrt{\frac{\mu^2}{\lambda}} \approx 246\,\text{GeV}, \tag{1.10}$$

which spontaneously breaks the electroweak symmetry as

$$SU(2)_L \times U(1)_Y \to U(1)_{\text{EM}},$$

where $U(1)_{\text{EM}}$ corresponds to the unbroken electromagnetic gauge symmetry.

Importantly, the VEV v is not simply a theoretical parameter but is precisely measured from the Fermi constant G_F, extracted from muon decay experiments [15]:

$$v = \left(\sqrt{2} G_F\right)^{-\frac{1}{2}} \approx 246\,\text{GeV}. \tag{1.11}$$

This relation directly connects the electroweak scale to experimentally measured quantities.

Expanding around this vacuum,

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

introduces the physical Higgs boson $h(x)$ as the quantum fluctuation about the vacuum.

Through the Higgs kinetic term and covariant derivatives, mass terms for the $W^\pm$ and Z bosons arise naturally:

$$m_W = \frac{gv}{2}, \quad m_Z = \frac{v}{2}\sqrt{g^2 + g'^2}, \quad m_\gamma = 0, \tag{1.13}$$

where the photon remains massless, consistent with the unbroken $U(1)_{\text{EM}}$.

Fermion masses originate from Yukawa interactions with the Higgs field,

$$\mathcal{L}_{\text{Yukawa}} = -y_f \overline{\psi}_L \Phi \psi_R + \text{h.c.}, \tag{1.14}$$

leading to fermion masses proportional to their Yukawa couplings [10]:

$$m_f = \frac{y_f v}{\sqrt{2}}. \tag{1.15}$$

Finally, the mass of the Higgs boson itself is determined by the curvature of the scalar potential at its minimum:

$$m_h = \sqrt{2\lambda} v. \tag{1.16}$$

II. Higgs boson Measurements

On the fourth of July 2012, the ATLAS and CMS collaborations announced the discovery of the Higgs boson [11, 12]. This discovery completed the SM but also opened a new field of exploration. Precise measurements of Higgs properties, such as differential production cross sections, branching ratios, and couplings, are particularly important, as deviations from SM predictions could indicate new physics beyond the SM (BSM). Moreover, characterizing other properties of the Higgs boson, such as its mass, is essential to fully understand this newly discovered particle.

One of the primary properties of the Higgs boson is its mass. Since Run 1, the ATLAS and CMS experiments have significantly improved the precision of Higgs mass measurements using Run 2 data. The most precise determination from the ATLAS Collaboration combines results from the $H \to \gamma\gamma$ and $H \to 4\ell$ decay channels [64], yielding

$$m_H = 125.11 \pm 0.09_{\text{stat}} \pm 0.06_{\text{syst}}\,\text{GeV}.$$

The CMS experiment reaches a similar precision in the $H \to ZZ^*$ decay channel [65].

The total width of the Higgs boson has been measured by both experiments using on-shell and off-shell Higgs production in the $H \to ZZ^*$ decay channel. The CMS result,

$$\Gamma_H = 3.0^{+2.0}_{-1.5}\,\text{MeV},$$

is consistent with the SM expectation of approximately 4.1 MeV [66].

This section will discuss the main production and decay modes of the Higgs boson, as well as the couplings that are especially sensitive to CP violation and potential BSM effects.

II.1. Production Modes

The main Higgs boson production mechanisms at leading order (LO) are illustrated in Figure 1.2, and their cross sections as a function of the center-of-mass energy are shown in Figure 1.3. The dominant production mode at the LHC is gluon-gluon fusion (ggF), which proceeds via a top-quark loop at next-leading order (NLO) in perturbation theory [13, 14]. The large ggF cross section arises in part from the high gluon density at small *parton* momentum fractions within the proton.

In perturbative QCD, the proton is described as a sea of partons (quarks and gluons). The probability of finding a parton of flavor i carrying a momentum fraction x at a given factorization scale Q^2 is encoded in the parton distribution functions (PDFs), $f_i(x, Q^2)$ [67]. The variable x, known as Bjorken-x, denotes the

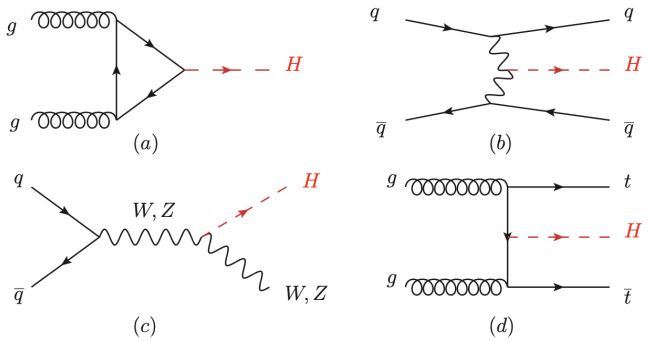


Figure 1.2: Feynman diagram of the Higgs production at tree level. In (a) gluon gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung or associated production with a W/Z boson and (d) $t\bar{t}$ associated production [15].

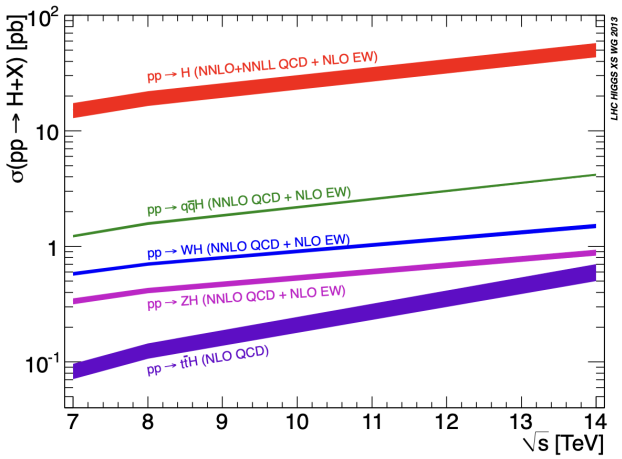


Figure 1.3: SM Higgs boson production cross section as a function of the center of mass energy $\sqrt{s}$ of the pp collision [15].

longitudinal momentum fraction of the proton taken by a parton in the hard scattering process [68]. At LHC energies, Higgs production via ggF typically involves gluons with $x \sim 10^{-3}$–10^{-2}, where the gluon PDFs are particularly large due to their steep rise at low x [67, 69]. The total proton–proton cross section for Higgs production is obtained by convoluting the partonic cross sections $\hat{\sigma}_{ij\to H}$ with the PDFs:

$$\sigma(pp \to H) = \sum_{i,j} \int dx_1\, dx_2\, f_i(x_1, Q^2)\, f_j(x_2, Q^2)\, \hat{\sigma}_{ij\to H}(x_1 x_2\, s, Q^2). \tag{1.17}$$

Here, x_1 and x_2 are the momentum fractions of the incoming partons, and $\sqrt{s}$ is

the proton–proton center-of-mass energy. The PDFs themselves are not calculable from first principles but are extracted from global fits to experimental data, including deep inelastic scattering, Drell–Yan production, and jet measurements.

The second most significant production mode is vector boson fusion (VBF), where the Higgs boson is produced via the fusion of two electroweak bosons (W or Z) radiated from incoming quarks. This process leads to a characteristic topology with two high-p_T jets in the forward region, a large pseudorapidity gap between them, and little hadronic activity in the central detector.

Other important modes include associated production with a vector boson (VH) and with top quarks, either a top or antitop pair ($t\bar{t}H$) or a single top quark (tH). Although these channels have smaller cross sections, they offer unique sensitivity to the Higgs couplings to gauge bosons and to the top quark. All production mechanisms are essential to rigorously test the consistency of the Higgs sector with the SM and to probe potential deviations arising from BSM physics.

II.2. Decay modes

The relative occurrence of a Higgs boson decay channel is quantified by its branching ratio, defined as $\mathrm{BR}(H \to X) = \Gamma_{H\to X}/\Gamma_H$, where $\Gamma_{H\to X}$ is the partial decay width into the final state X, with the total width given by $\Gamma_H = \sum_X \Gamma_{H\to X}$.

For a Higgs boson mass close to 125 GeV, the partial widths span several orders of magnitude. Figure 1.4 shows the different branching ratio for the Higgs decay. The largest branching fraction corresponds to the $H \to b\bar{b}$ channel, accounting for roughly 58% of all decays, followed by $H \to WW^*$, $H \to gg$, and $H \to \tau^+\tau^-$, with approximate branching ratios of 21%, 8.2%, and 6.3%, respectively. Subdominant channels such as $H \to c\bar{c}$, $H \to \gamma\gamma$, and $H \to Z\gamma$, while having significantly smaller branching fractions, are of particular theoretical and experimental interest. Their loop-induced nature provides sensitivity to virtual contributions from heavy charged particles, such as the W boson and top quark, making them valuable indirect probes of BSM physics.

Experimentally, the $\gamma\gamma$ and $ZZ^* \to 4\ell$ final states are favored for their clean signatures, excellent mass resolution, and low backgrounds, despite their low rates. In contrast, modes such as $b\bar{b}$ and WW^* benefit from higher production rates but are experimentally more challenging due to larger backgrounds and poorer mass resolution. A comprehensive understanding of the Higgs sector therefore requires combining results from multiple production and decay channels, balancing statistics with experimental sensitivity to provide a robust test of Higgs boson properties and to search for possible deviations from SM predictions.

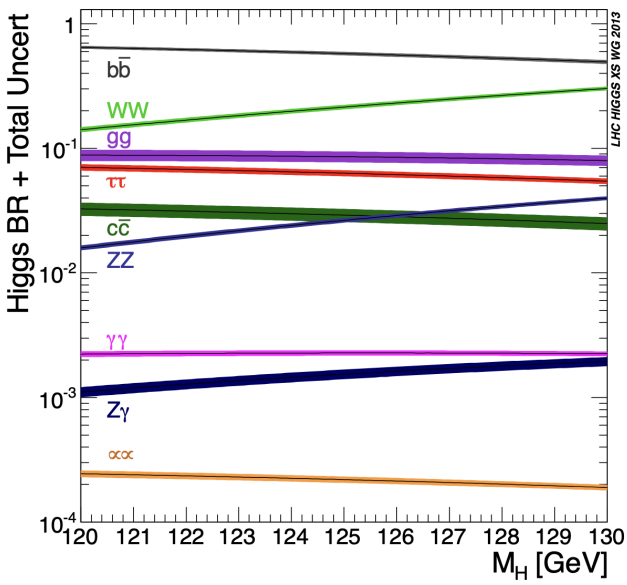


Figure 1.4: SM Higgs boson branching ratios as a function of the Higgs boson mass [15].

II.3. Couplings

The couplings of the Higgs boson to other particles are fundamental properties of the Higgs boson. At tree level, the SM predicts that these couplings are proportional to the masses of the fermions involved, and proportional to the square of the mass for the bosons. Any deviation from this trend could indicate BSM physics. To test for such deviations in a model-independent way, early LHC measurements adopted the so-called κ-framework, in which each Higgs coupling is rescaled by a modifier κ_i, defined such that production cross sections σ_i, partial widths Γ_i, and the total width Γ_H scale as κ_i^2 relative to their SM expectations. In the SM limit, all $\kappa_i = 1$ .While this approach cannot determine the detailed nature of the Higgs interactions, such as whether the couplings have a CP-violating component, it still serves as a simple and effective way to evaluate how much the observed rates differ from predictions.

Figure 1.5 shows the measured coupling modifiers of the Higgs boson to fermions (κ_f) and vector bosons (κ_V) as a function of the particle mass it decays to. For gauge bosons, the square root of the coupling modifier is shown to preserve the linear mass dependence predicted by the SM. The most precise determinations currently come from the W and Z boson couplings, which are known to better than 10%. Fermionic couplings such as κ_b and κ_μ remain less constrained, with uncertainties of around 20% or more. All measured values are consistent with the

SM within current uncertainties.

With the increasing statistical power available in Run 2 and beyond, Higgs boson studies have progressively moved beyond inclusive measurements of cross sections and branching ratios toward more differential observables. For example, the *Simplified Template Cross Section* (STXS) framework [70] splits the phase space of the Higgs production into a set of exclusive kinematic regions. These regions are defined to be theoretically well-controlled and experimentally accessible, facilitating a more precise comparison between data and theoretical predictions. The STXS framework retains sensitivity to specific production mechanisms (e.g. ggF, VBF, $t\bar{t}H$) while minimizing extrapolation uncertainties. It is organized in stages (e.g. Stage 0, Stage 1, Stage 2), with increasing granularity in variables such as jet multiplicity, the Higgs transverse momentum (p_T^H), or the invariant mass of associated jets.

In parallel, the *Effective Field Theory* (EFT) approach [71] provides a systematic and model-independent framework to probe for deviations from the SM. In EFT, the SM Lagrangian is extended with higher-dimensional operators constructed from SM fields, which are suppressed by powers of an energy scale Λ representing the mass scale of new physics. The coefficients of these operators encapsulate the strength of possible new interactions. For instance, dimension-six operators can induce modifications to Higgs couplings, introduce CP-violating terms, or generate anomalous interactions with gauge bosons.

Unlike the κ-framework, which assumes SM-like tensor structures and only rescales the couplings, EFT enables a broader characterization of new physics effects, including modifications to angular and kinematic distributions. Combining STXS measurements with EFT interpretations thus provides a powerful, high-resolution tool to test the Higgs sector and constrain or discover deviations from SM expectations.

III. Search of CP Violation in the Higgs Sector

III.1. Motivation for the Search for CP Violation

One major phenomenon that the SM does not explain is the matter-antimatter asymmetry in the universe, which is observed from the cosmic microwave background and the abundances of light primordial elements. This baryon asymmetry must have occurred during the early universe, in a phase called *baryogenesis*. Baryogenesis can only occur if the three Sakharov conditions are satisfied: baryon number violation, departure from thermal equilibrium, and CP violation [72]. While the Standard Model (SM) provides a source of CP violation via the complex phase in the Cabibbo–Kobayashi–Maskawa (CKM) matrix, it is insufficient by sev-

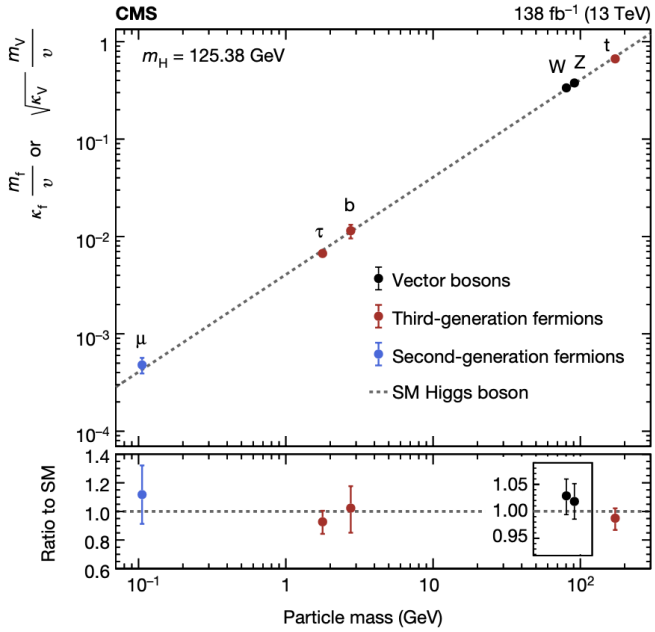


Figure 1.5: Coupling modifier of the Higgs bosons to fermions (κ_f) and heavy gauge bosons (κ_V) as a function of the fermion or gauge boson mass, where v is the vacuum expectation value of the Brout-Englert-Higgs field. For gauge bosons, the square root of the coupling modifier is plotted, to keep a linear proportionality to the mass [16].

eral orders of magnitude to explain the observed baryon asymmetry [73], thereby motivating searches for new sources of CP violation. Such searches are pursued both in the quark sector, through precision measurements of meson decays [15], and in the lepton sector, for example via electric dipole moment measurements and studies of neutrino oscillations [74]. Such new sources are predicted in many theoretical extensions of the SM, especially those involving an extended Higgs sector, where the Higgs boson plays a central role in the CP dynamics of the electroweak phase transition [75].

The conservation of the baryon number (B) is only an approximate symmetry in the SM, and it can be violated at high temperature through non-perturbative electroweak processes known as *sphalerons* [76]. These transitions, efficient in the early universe, violate $B + L$ and can therefore convert baryon and lepton number. Because sphalerons create baryons and antibaryons in equal amounts, an additional source of CP violation is required to generate a net baryon excess. CP violation thus becomes a necessary ingredient, together with baryon-number violation and a departure from thermal equilibrium, for any viable baryogenesis mechanism, as stipulated by the three Sakharov conditions.

A well-studied scenario, *electroweak baryogenesis*, posits that these conditions are satisfied during the electroweak phase transition. As the universe cools, the Higgs field acquires a non-zero vacuum expectation value, driving a transition from the symmetric to the broken phase. If this transition is strongly of first-order, it proceeds via nucleation of bubbles of broken phase within the symmetric plasma. CP-violating interactions at the bubble walls create an excess of left-handed fermions. In the symmetric phase, where sphalerons are active, part of this chiral excess is converted into a net baryon number. Once inside the bubbles, i.e. in the broken phase, sphaleron processes are suppressed, effectively "freezing in" the generated baryon asymmetry.

However, with a 125 GeV Higgs boson, the electroweak phase transition is not first-order but a smooth crossover, which prevents electroweak baryogenesis. Nevertheless, the actual measurement does not exclude the existence of an electroweak-scale singlet scalar field that would couple to the Higgs boson. This new scalar field would not only modify the electroweak phase transition but also provide a new source of CP violation, which could occur in the Higgs couplings. In particular, this would also generate non-zero electric dipole moments (EDMs). Actual measurements of the EDMs combined with LHC ones constrain these models [77, 78] without deviation from the SM [79]. At the LHC, the CP structure of the Higgs boson can be probed through its couplings to both gauge bosons and fermions. The following sections details the CMS results for CP measurements in the Higgs sector.

III.2. Anomalous Higgs Couplings to Gauge Bosons

CP violation in the gauge sector can be probed through the interactions of the Higgs boson with electroweak vector bosons and gluons. In the SM, such interactions are CP-even at tree level, and any CP-violating effects must therefore originate from loop-level contributions or from new physics encoded in higher-dimensional operators.

The interaction of the spin-zero Higgs boson with two spin-one gauge bosons VV, where $VV = WW, ZZ, Z\gamma, \gamma\gamma$, or gg, can be parametrized by the scattering amplitude [80]:

$$\mathcal{A}(HVV) \approx \left[a_1 + \frac{\kappa_1^{VV} q_1^2 + \kappa_2^{VV} q_2^2}{\Lambda^2}\right] m_{V1}^2\, \epsilon_1^* \cdot \epsilon_2^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2)\mu\nu}, \tag{1.18}$$

where a_1 is the SM-like coupling, $\kappa_{1,2}$ and a_2 correspond to CP-even anomalous terms, and a_3 is a CP-odd anomalous coupling. The tensors $f_{\mu\nu}^{(i)} = \epsilon_\mu^{(i)} q_\nu^{(i)} - \epsilon_\nu^{(i)} q_\mu^{(i)}$ and $\tilde{f}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}$ denote the field strength and its dual.

Experimentally, constraints on these couplings are reported in terms of cross-section fractions,

$$f_{a_i} = \frac{|a_i|^2\sigma_i}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\kappa_1} + |\kappa_2|^2\sigma_{\kappa_2}}, \tag{1.19}$$

which describe the relative contribution of each tensor structure to the total rate.

III.2.a. Anomalous Higgs to Electroweak Vector Boson Couplings

Results from CMS Run 1 set stringent limits on a wide range of anomalous Higgs boson interactions with electroweak vector bosons. Specifically, a broad class of spin-two models is excluded at the 99% confidence level or higher, independent of assumptions on the production mechanism. Any mixed-parity spin-one state is excluded in the ZZ and WW channels at a confidence level exceeding 99.999%. By combining measurements from the $H \to ZZ$ and $H \to WW$ channels, even tighter constraints on anomalous HVV interactions have been achieved. All observations remain consistent with the SM expectation of a CP-even scalar with quantum numbers $J^{PC} = 0^{++}$ [81].

During Run 2, CMS has extended these studies using larger datasets and multiple decay channels, each offering different advantages and challenges. The four-lepton (4ℓ) final state provides a clean experimental signature with high mass resolution, enabling precise reconstruction of the Higgs boson and its decay kinematics, which enhances the sensitivity to anomalous couplings [82]. The $\tau\tau$ decay channel benefits from a relatively large branching fraction and sensitivity to different Higgs production modes, but suffers from more challenging event reconstruction and higher backgrounds [83]. The WW channel offers high statistics due to the large branching ratio and is probed in both production and decay, although the presence of neutrinos in the final state reduces the mass resolution and complicates the kinematic reconstruction [84]. These analyses uniformly employ matrix element likelihood approaches (MELA) [85], which optimally exploit the full kinematic information of each event to distinguish between the SM and various anomalous coupling hypotheses, thus maximizing the sensitivity to CP violation and other non-standard effects in Higgs boson interactions.

Combining the 4ℓ and $\tau\tau$ channels improves the overall sensitivity by exploiting their complementary features and correlations in systematic uncertainties. The newest results in the WW channel further tighten constraints on anomalous couplings. All measurements consistently align with the SM expectations, and constraints are set on the fractional contributions of anomalous couplings to the Higgs boson production cross section.

III.2.b. Higgs to Gluon Effective Couplings

Anomalous Higgs couplings to gluons (Hgg), probed via the ggF production mode, arise through top-quark loop processes in the SM. Extensions beyond the SM may introduce CP-violating components via higher-dimensional operators, which would be sign of BSM physics or CP violation in the top Yukawa coupling.

The first limits on such anomalous Hgg couplings were derived in Run 1 indirectly through the $H \to ZZ^* \to 4\ell$ and $H \to \tau\tau$ channels by analyzing angular and kinematic observables sensitive to CP violation [81]. Run 2 measurements have expanded this approach across the 4ℓ, $\tau\tau$, and WW final states, employing the same analyses used for anomalous HVV couplings, but additionally constraining the effective Higgs–gluon vertex. As for HVV couplings, these analyses utilize MELA to optimally separate CP-even and CP-odd contributions by exploiting the full event kinematic [82, 83, 85].

A particularly sensitive probe arises in ggF events with two forward jets, resembling the VBF topology, where the azimuthal angle difference $\Delta\phi_{jj}$ between the jets exhibits characteristic modulations sensitive to interference between CP-even and CP-odd amplitudes. This observable provides an independent cross-check complementary to the usage of the MELA discriminants.

Results combining $H \to 4\ell$ and $H \to \tau\tau$ channels using both the MELA and $\Delta\phi_{jj}$ observables exclude a pure CP-odd Higgs–gluon coupling at an observed significance of 2.4σ. The WW channel contributes complementary but with less sensitive information [84]. The current combined measurements show no significant deviation from the SM expectation from a CP-even Hgg coupling and set stringent limits on possible CP-odd admixtures.

III.3. CP Violation in Yukawa Couplings

The most general, renormalizable Lagrangian for Higgs couplings to fermions can include both scalar (CP-even) and pseudoscalar (CP-odd) terms. It is parametrized as:

$$\mathcal{L}_{Hff} = -\frac{m_f}{v} H\, \bar{\psi}_f \left(\kappa_f + i\gamma_5 \tilde{\kappa}_f\right) \psi_f, \tag{1.20}$$

where m_f is the fermion mass, v is the vacuum expectation value of the Higgs field, and κ_f and $\tilde{\kappa}_f$ represent the CP-even and CP-odd coupling modifiers, respectively. In the SM, $\kappa_f = 1$ and $\tilde{\kappa}_f = 0$, while deviations from these values indicate new physics.

Experimentally, CP violation in the Yukawa sector can be characterized either by the CP-odd cross section fraction or by a CP-mixing angle. The former is defined as:

$$f_{\mathrm{CP}}^{Hff} = \frac{|\tilde{\kappa}_f|^2}{|\kappa_f|^2 + |\tilde{\kappa}_f|^2}, \tag{1.21}$$

which expresses the fraction of the total cross section arising from the CP-odd contribution. Alternatively, one can express the mixing angle α_{Hff} between the two components as:

$$\alpha_{Hff} = \tan^{-1}\left(\frac{\tilde{\kappa}_f}{\kappa_f}\right), \tag{1.22}$$

where $\alpha_{Hff} = 0$ corresponds to a pure scalar (SM-like) coupling and $\alpha_{Hff} = \pi/2$ to a pure pseudoscalar.

These quantities can be constrained using decay and production observables sensitive to the spin correlations and angular distributions of final-state fermions notably in $H \to \tau\tau$ decays and in the $t\bar{t}H$ production.

III.3.a. CP Violation in the Top Yukawa Coupling

The top Yukawa coupling can be accessed through several production mechanisms, most notably ggH and the Higgs production in association with top quarks ($t\bar{t}H$ and tH). A CP-violating admixture in the $H-t$ interaction modifies both the inclusive rates and the angular correlations of final-state objects, yielding experimentally accessible signatures [86].

In the $t\bar{t}H$ process, CP information is preserved through the polarization of the top quarks and the spin correlations between the t and $\bar{t}$ [86]. These correlations are reflected in the decay kinematics, allowing MELA and multivariate discriminants to separate CP-even and CP-odd components with high efficiency [82, 83, 87, 88]. The associated tH mode offers a complementary probe: its cross section is linearly sensitive to the sign and CP phase of the top Yukawa coupling because of interference between diagrams involving the Htt and HWW vertices. When combined with $t\bar{t}H$, this sensitivity enhances overall constraints.

Anomalous $H-t$ couplings are investigated in several clean final states. The $H \to \gamma\gamma$ and $H \to 4\ell$ channels provide excellent mass resolution and low backgrounds, and are exploited to study both $t\bar{t}H$ and tH production [82, 83, 87]. In the $H \to 4\ell$ analysis, dedicated MELA discriminators target the associated-production topologies, while the $H \to \gamma\gamma$ channel employs boosted-decision-tree classifiers [88].

To maximize the sensitivity, the $H \to 4\ell$ analysis combines $t\bar{t}H$ with ggH events under the assumption of the top-quark dominance in the ggH loop. The $t\bar{t}H$ cross section rises with the CP-odd fraction, whereas the ggH contribution falls, yielding strong discriminatory power. The combined result disfavors a purely CP-odd $H-t$ coupling at an observed significance of 3.7σ [87].

Figure 1.6 (left) shows the two-dimensional likelihood scan in the $(\kappa_t, \tilde{\kappa}_t)$ plane, excluding the pure CP-odd scenario. Figure 1.6 (right) presents the combined likelihood as a function of the CP-odd cross-section fraction f_{CP}^{Htt}, obtained from the $H \to \gamma\gamma$, $H \to 4\ell$, and multilepton channels. No significant deviation from

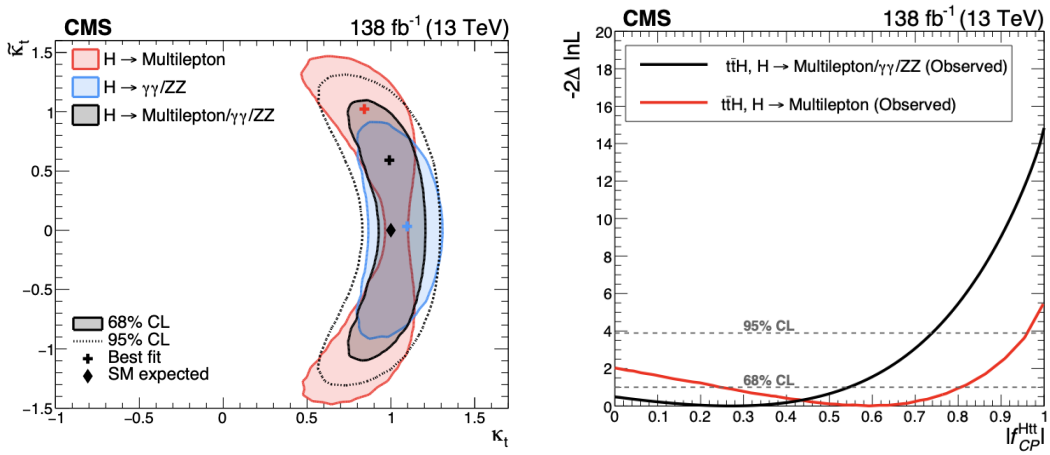


Figure 1.6: (Left) Two-dimensional likelihood scan in the $(\kappa_t, \tilde{\kappa}_t)$ plane from the combination of $t\bar{t}H$, tH, and ggH sensitive analyses in $H \to \gamma\gamma$, $H \to 4\ell$, and multilepton final states. (Right) Observed and expected negative log-likelihood scan as a function of the CP-odd cross section fraction f_{CP}^{Htt}. Figures from [87].

the SM expectation has been seen, although moderate CP-odd admixtures are still allowed.

III.3.b. CP Violation in the τ Yukawa Coupling

The search for CP violation in $H \to \tau\tau$ has already been performed by CMS using Run 2 data [17, 89]. This thesis focuses on developments using Run 3 data.

The CP structure of the Higgs boson coupling to τ leptons is directly probed in the $H \to \tau\tau$ decay channel using the CP-sensitive observable ϕ_{CP}, constructed from the spin correlations of the τ decay products, as described in Section IV..

In Run 2, results were obtained by combining several final states: $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$, where the notation indicates that one τ decays leptonically ($\tau \to e\nu_e\nu_\tau$ or $\tau \to \mu\nu_\mu\nu_\tau$) and the other hadronically. The decay modes and their branching fractions are summarized in Table 1.1. The analysis covers approximately 70% of the τ-pair final states in terms of total branching ratio.

In each channel, a neural-network-based discriminator is used to categorize events into multivariate analysis (MVA) bins, enhancing signal-to-background separation. The CP-mixing angle is extracted from a simultaneous fit to all channels. The measured value is $\alpha_{\tau\tau} = -1 \pm 19°$ at 68.3% confidence level (CL), and a purely CP-odd scenario is excluded at 3.0σ, as shown in Figure 1.7. The analysis is currently limited by the statistical precision of the dataset, and the result is consistent with the SM prediction.

Mode	$e^{\pm}\nu_e\nu_\tau$	$\mu^{\pm}\nu_\mu\nu_\tau$	$h^{\pm}\nu_\tau$	$h^{\pm}\pi^0\nu_\tau$	$h^{\pm}\pi^0\pi^0\nu_\tau$	$h^{\pm}h^{\mp}h^{\pm}\nu_\tau$
Type	τ_e	τ_μ	τ_h	τ_h	τ_h	τ_h
BR (%)	17.8	17.4	11.5	25.9	9.5	9.8
Resonance	–	–	–	$\rho(770)$	$a_1(1260)$	$a_1(1260)$
Symbol	e	μ	π	ρ	a_1^{pr}	a_3^{pr}

Table 1.1: Summary of main τ decay modes with their type, branching ratio (BR), and resonance label. The charged hadrons denoted as $h^{\pm}$ are assumed to be $\pi^{\pm}$ since the CMS detector is not able to distinguish between the different types of $h^{\pm}$ [17].

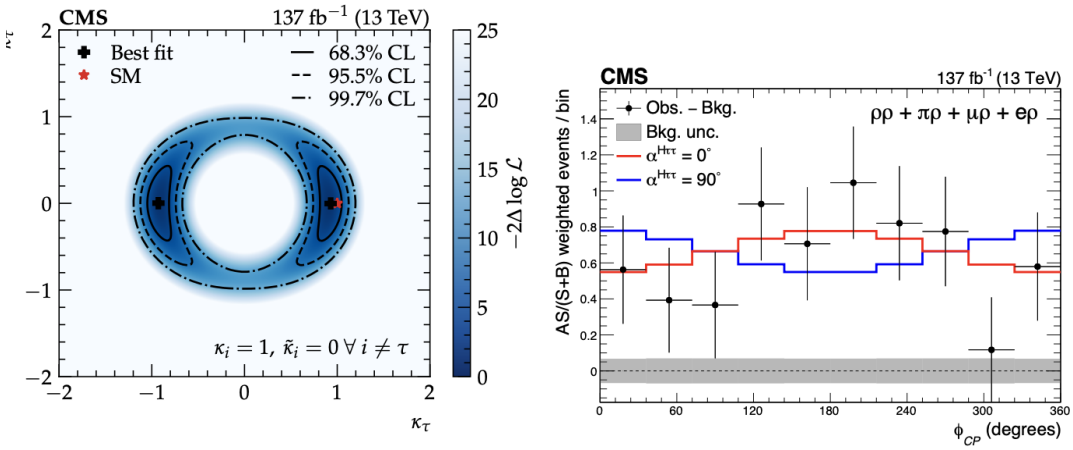


Figure 1.7: (Left) Two-dimensional likelihood scan in the $(\kappa_\tau, \tilde{\kappa}_\tau)$ plane. (Right) Weighted ϕ_{CP} distributions in most sensitive hadronic and leptonic τ decay channels, with background subtraction in all MVA bins. Each MVA score bin is weighted by $A \cdot S/(S+B)$ where S and B are the signal and background rates, and A is the average symmetry of the scalar and pseudoscalar distribution. The scalar hypothesis is shown in red, while the pseudoscalar hypothesis is shown in blue. The predictions are based on the best fit values of the rate parameters. The grey uncertainty band indicates the uncertainty in the background subtraction. A phase shift of 180° is applied to the leptonic channels due to the phase difference with the hadronic channels. Figures from [17].

Further details on the methodology, including background modeling, MVA construction, and fit strategy, are discussed in Chapter 6.

IV. CP-Sensitive Observables in $H \to \tau\tau$

This section gives the main clues to understand how the CP nature of the τ Yukawa coupling is probed by exploiting the correlation of the transverse spin of the two τ leptons and the propagation of the spin information in their decay. From [90–93], a CP-sensitive observable is built with different methods, the so-called *acoplanarity* angle, ϕ_{CP}.

As mentioned previously, the effective Lagrangian for the Higgs coupling to τ leptons can be written as:

$$\mathcal{L}_{H\tau\tau} = -\frac{m_\tau}{v}\bar{\tau}\left(\kappa_\tau + i\tilde{\kappa}_\tau\gamma_5\right)\tau H \tag{1.23}$$

where m_τ is the τ mass, v is the vacuum expectation value of the Higgs field, and κ_τ and $\tilde{\kappa}_\tau$ are the reduced CP-even and CP-odd Yukawa couplings, respectively. The CP mixing angle $\alpha_{\tau\tau}$ is then defined as:

$$\alpha_{\tau\tau} = \arctan\left(\frac{\tilde{\kappa}_\tau}{\kappa_\tau}\right). \tag{1.24}$$

As shown in [91], the partial decay width of the Higgs boson into τ leptons can be expressed in terms of the τ spin polarization vectors s^{τ^+} and s^{τ^-} as:

$$\Gamma(H_{\text{mix}} \to \tau^+\tau^-) = \Gamma_{\text{unpol}}\left(1 - s_\parallel^{\tau^+}\cdot s_\parallel^{\tau^-} + s_\perp^{\tau^+}R(2\alpha_{\tau\tau})s_\perp^{\tau^-}\right) \tag{1.25}$$

where Γ_{unpol} is the polarization-averaged decay width, and $s_\parallel^{\tau^\pm}$ and $s_\perp^{\tau^\pm}$ represent the components of the spin polarization vectors along and transverse to the direction of motion, respectively. The operator $R(2\alpha_{\tau\tau})$ corresponds to a rotation by an angle $2\alpha_{\tau\tau}$ around the longitudinal axis, acting on the transverse spin components.

In the SM hypothesis, which corresponds to the purely CP-even scenario, the CP-mixing angle is $\alpha_{\tau\tau} = 0°$, while a purely CP-odd coupling corresponds to $\alpha_{\tau\tau} = -90°$. In these two limiting cases, the rotation operator $R(2\alpha_{\tau\tau})$ reduces to the identity and Equation 1.25 simplifies as:

$$\Gamma(H_{\text{mix}} \to \tau^+\tau^-) = \Gamma_{\text{unpol}}\left(1 - s_\parallel^{\tau^+}\cdot s_\parallel^{\tau^-} \pm s_\perp^{\tau^+}\cdot s_\perp^{\tau^-}\right), \tag{1.26}$$

where the sign depends on whether the Yukawa coupling is CP-even (positive sign) or CP-odd (negative sign). As the longitudinal spin components of the τ^+ and τ^- are always collinear in the Higgs rest frame (due to angular momentum conservation and the scalar nature of the Higgs), their dot product simplifies to the product of their magnitudes:

$$s_\parallel^{\tau^+}\cdot s_\parallel^{\tau^-} = \|s_\parallel^{\tau^+}\|\cdot\|s_\parallel^{\tau^-}\| = \|s_\parallel^{\tau}\| \tag{1.27}$$

and, using the normalization condition $|s|^2 = 1$, the transverse components satisfy

$$\|s_\perp^\tau\|^2 = \|s_\perp^{\tau^+}\|^2 \cdot \|s_\perp^{\tau^-}\|^2 = 1 - \|s_\parallel^\tau\|^2. \tag{1.28}$$

Therefore, Equation 1.26 can be rewritten as:

$$\Gamma(H_{\text{mix}} \to \tau^+\tau^-) \propto 1 - \|s_\parallel^\tau\| \pm \sqrt{1 - \|s_\parallel^\tau\|^2} \cos\phi \tag{1.29}$$

where ϕ is the angle between the transverse components of the τ spin. It is straightforward to see that the decay rate depends only on the angle ϕ and of $\|s_\parallel^\tau\|$. The decay rates for scalar and pseudoscalar Higgs bosons (referred to as A) into fermion-antifermion pairs, $\Gamma(H \to f\bar{f})$ and $\Gamma(A \to f\bar{f})$, are shown as a function of the longitudinal spin component $\|s_\parallel^\tau\|$ and the angle ϕ between the transverse spin vectors in Figure 1.8. In the case of a scalar Higgs, the decay rate is maximum for $\phi = 180°$, meaning that in the SM scenario, the transverse components of the spin tend to be anti-aligned, while in the pseudoscalar scenario, the transverse spins tend to be aligned. This also shows that the CP nature is only observed in the transverse spin correlations of the two τ leptons.

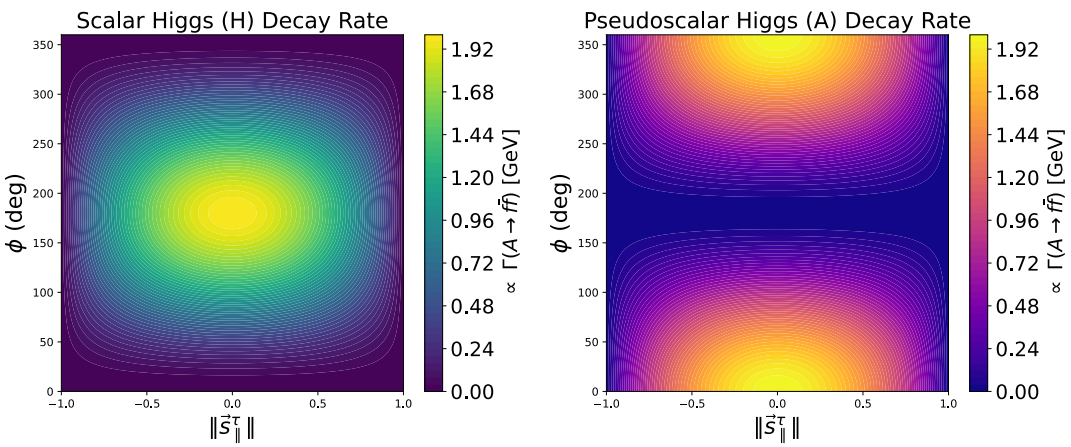


Figure 1.8: Spin correlation–dependent decay rates (see Equation 1.29) of the Higgs boson into $\tau^+\tau^-$ pairs as a function of the longitudinal spin component $\|s_\parallel^\tau\|$ and the angle ϕ between the transverse spin vectors. The scalar scenario is on the left and the pseudoscalar scenario on the right.

As the angle ϕ between the transverse spins of the τ leptons is not directly accessible, this analysis exploits the propagation of spin information to the τ decay products. The decay mode $h \to \tau^-\tau^+$, with subsequent decays:

$$h \to \tau^-\tau^+ \to a^- a'^+ + X, \tag{1.30}$$

is considered, where $a^{\pm}, a'^{\pm} \in \{e^{\pm}, \mu^{\pm}, \pi^{\pm}, \rho, a_1\}$, and X represents undetected particles such as neutrinos.

As shown in [93], the decay width for the Higgs decay into $\tau^+\tau^-$ with polarized τ leptons decaying into $a^{\pm}$ and X is given by:

$$d\Gamma(H \to \tau^+\tau^-) \sim 1 - b(E^+)b(E^-)\frac{\pi^2}{16}\cos(\phi_{CP} - 2\alpha_{H\tau\tau}), \tag{1.31}$$

where $E^{\pm}$ are the energies of the outgoing charged particles in their respective τ rest frames, and $b(E^{\pm})$ are the spectral functions that encapsulate the correlation between the τ spin and the momentum of the outgoing charged particle.

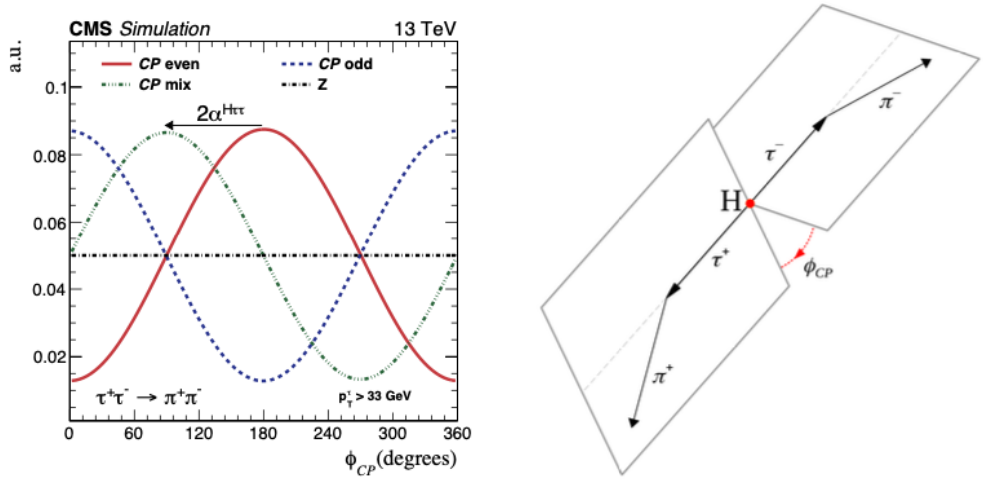


Figure 1.9: On the left, the normalized distribution of the acoplanarity angle ϕ_{CP} between the τ lepton decay planes in the Higgs rest frame at generator level, for both τ leptons decaying to a charged pion and a neutrino. Distributions are shown for a scalar (CP-even, solid red), pseudoscalar (CP-odd, dashed blue), maximal CP-mixing angle of 45° (CP-mix, dash-dot-dot green), and a Z boson (black dash-dot). On the right, an illustration of the decay planes of two τ leptons decaying to single charged pions. The ϕ_{CP} angle is defined as the angle between these decay planes in the Higgs rest frame. Figures from [17].

The ϕ_{CP} angle is the acoplanarity angle defined as the angle between the decay planes of the two τ leptons in the rest frame of the Higgs boson. The distribution of the acoplanarity angle ϕ_{CP} is shown in Fig. 1.9. Different scenarios (CP-even, CP-odd, maximal mixing) are shown, and the shift from the CP-even case in blue corresponds to $2\alpha_{\tau\tau}$. An illustration of the decay planes and the corresponding ϕ_{CP} distribution is shown in Figure 1.9.

The reconstruction of the ϕ_{CP} observable is based on the reconstruction of the

decay planes of the τ leptons. Different methods are used depending on the τ decay mode. These methods are detailed in the following sections.

IV.1. Polarimetric Method

The *polarimetric* vector, sometimes referred to as the *polarimeter* vector, is an optimal observable for the measurement of the τ spin [94]. It exploits the spin propagation information transmitted to the τ decay products. The polarimetric vector is computed in the τ rest frame and is thus sometimes referred to as a three-vector rather than a four-vector. It can be understood as an estimator of the most probable spin direction $\vec{s}$ of the τ lepton.

The decay rate of a polarized τ lepton can be expressed as [95]:

$$\frac{d\Gamma}{d\text{Lips}} \propto (1 + h^\mu s_\mu), \tag{1.32}$$

where h^μ is the polarimetric four-vector and s_μ is the τ spin four-vector and dLips denotes the Lorentz-invariant phase space element. In the rest frame of the τ, $s^0 = 0$ and $h^0 = 0$, so the spin-dependent term reduces to $-\vec{h} \cdot \vec{s}$. Thus, the decay probability is maximized when the spin and polarimetric vectors are aligned.

In principle, the polarimetric vector can be estimated for any hadronic τ decay. Since the acoplanarity angle ϕ^* is defined by the planes spanned by the τ directions and the corresponding polarimetric vectors in the Higgs boson rest frame, this method requires an accurate reconstruction of the full τ momenta. This is particularly challenging due to the presence of the undetected neutrinos.

For the one-prong decay, the polarimetric vector is $\vec{h} = \vec{n}_\pi$, where $\vec{n}_\pi$ is the unit vector of the charged pion originating from the τ decay. In the case of the τ decay into a ρ resonance, the polarimetric vector takes the form [96]:

$$\vec{h} = \mathcal{N}\left(2(q \cdot N)\vec{q} - q^2\vec{N}\right), \tag{1.33}$$

with $\mathcal{N}$ a normalization factor, $q = (q^0, \vec{q})$ the difference between the charged and neutral pion four-momenta, and $N = (N^0, \vec{N})$ the neutrino four-momentum, all defined in the τ rest frame.

For the $\tau \to a_1\nu_\tau$ decay, the polarimetric vector depends on the four-momenta of the three charged pions, and the full formula is given in [97].

To extract the CP-sensitive angle ϕ_{CP}, the polarimetric vector method constructs the acoplanarity angle ϕ^* between the planes defined by the directions of the two taus, denoted $\hat{n}_1$ and $\hat{n}_2$ (corresponding to the first and second τ leptons, respectively), and their corresponding normalized polarimetric vectors $\hat{h}_1$ and $\hat{h}_2$

in the Higgs boson rest frame. A vector $\vec{k}_{1,2}$ is defined as:

$$\vec{k}_{1,2} = \frac{\vec{h}_{1,2} \times \vec{n}_{1,2}}{|\vec{h}_{1,2} \times \vec{n}_{1,2}|}. \tag{1.34}$$

The angle ϕ^* is computed via:

$$\phi^* = \arccos(\vec{k}_1 \cdot \vec{k}_2), \tag{1.35}$$

and

$$\phi_{CP} = \begin{cases} \phi^*, & \text{if } (\vec{h}_1 \times \vec{h}_2) \cdot \vec{n}_1 \leq 0, \\ 2\pi - \phi^*, & \text{otherwise.} \end{cases} \tag{1.36}$$

IV.2. Impact Parameter Method

The impact parameter method was developed for τ leptons decaying into a single charged particle, such as $\tau^\pm \to \pi^\pm \nu_\tau$ or $\tau^\pm \to l^\pm \nu_\tau \nu_l$. It exploits the finite lifetime of the τ lepton by defining the impact parameter vector $\vec{IP}$ as the vector from the PV to the point of closest approach of the extrapolated charged particle track as shown in Figure 1.10

For each τ lepton, a decay plane is defined using the charged-particle momentum vector and the corresponding impact parameter vector. These planes are initially constructed in the laboratory frame. However, since the laboratory frame does not generally coincide with the Higgs boson rest frame, the method uses an approximation by boosting the relevant four-vectors into the zero-momentum frame (ZMF) of the two visible decay products of the τ leptons. This ZMF is defined by the sum of the four-momenta of the two charged particles from the τ decays. Impact parameter 4-vectors are then constructed as:

$$IP^\pm = (0, \vec{IP}^\pm), \tag{1.37}$$

and subsequently boosted into the ZMF, denoted by $IP^{*\pm}$. The four-momenta of the charged decay products, $p_p^\pm$, are also boosted into the ZMF and labeled $p_p^{*\pm}$.

The transverse components of the boosted impact parameter vectors, $IP_\perp^{*\pm}$, are defined with respect to the direction of $p_p^{*\pm}$. These transverse vectors are normalized to obtain unit vectors $\hat{IP}_\perp^{*\pm}$.The angle between the decay planes is then given by

$$\phi^* = \arccos\left(\hat{IP}_\perp^{*+} \cdot \hat{IP}_\perp^{*-}\right). \tag{1.38}$$

To resolve the orientation ambiguity, the following quantity is computed:

$$O^* = \hat{p}_p^{*-} \cdot \left(\hat{IP}_\perp^{*+} \times \hat{IP}_\perp^{*-}\right). \tag{1.39}$$

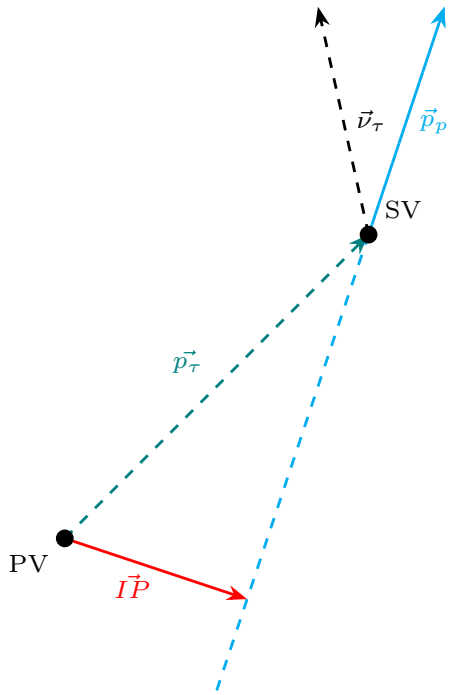


Figure 1.10: Schematic representation of the impact parameter vector $\vec{IP}$ of a charged particle of momemtum $\vec{p_p}$ from a secondary τ decay vertex (SV) relative to the primary vertex (PV).

The acoplanarity angle ϕ_{CP} is then defined in the interval $[0, 2\pi]$ as:

$$\phi_{CP} = \begin{cases} \phi^* & \text{if } O^* \geq 0, \\ 2\pi - \phi^* & \text{if } O^* < 0. \end{cases} \tag{1.40}$$

It is important to note that the spectral functions $b(E)$, which describe the spin analyzing power of the τ decay products as functions of their energy E, have opposite signs for single-pion and leptonic decay modes. This difference reflects the fact that the angular distributions of the decay products, sensitive to the τ polarization, exhibit different behavior depending on the decay channel [18].

IV.3. Decay Plane Method

The assumption that the neutrino momentum lies within the decay plane defined by the visible decay products is an approximation that simplifies the reconstruction of the polarimetric vector. While this assumption neglects the true three-dimensional kinematics of the τ decay, it has proven to provide a reasonably good

estimator of the spin direction for the dominant hadronic modes such as $\tau \to \rho\nu_\tau$ and $\tau \to a_1\nu_\tau$ [17]. The accuracy of this approximation depends on the specific decay mode and the kinematic regime; in particular, it holds better when the invariant mass of the visible system is close to the resonance mass and the missing neutrino momentum is small transverse to the visible system.

In the case of $\rho^\pm \to \pi^\pm\pi^0\nu_\tau$ decays, the decay plane is defined by the momentum vectors of the charged and neutral pions. The four-momentum of the π^0 is estimated by summing the energies of all electron or photon candidates identified by the HPS algorithm ([33]). The direction of the π^0 is taken to be that of the leading candidate, typically a photon, and is determined by extrapolating the position of the associated ECAL cluster to the primary vertex. The invariant mass of the π^0 is fixed to its known value. The CP observable ϕ_{CP} is then calculated analogously to the impact parameter method, using the plane spanned by the $\pi^\pm$ and π^0 momentum vectors in the zero-momentum frame (ZMF) of the $\tau\tau$ system.

To resolve the twofold ambiguity due to the symmetry of the decay planes, the variable

$$y^\pm = \frac{E_{\pi^\pm} - E_{\pi^0}}{E_{\pi^\pm} + E_{\pi^0}}, \quad y^\tau = y^- - y^+, \tag{1.41}$$

is used. If $y^\tau < 0$, the CP angle is modified by a phase shift: $\phi_{\mathrm{CP}} \to 360^\circ - \phi_{\mathrm{CP}}$.

The method is similarly extended to decays involving the $a_1(1260)$ meson. For 1-prong decays with two neutral pions, the π^0 candidates are treated collectively as a single system due to the limited experimental resolution. For 3-prong decays, the pair of oppositely charged pions with invariant mass closest to that of the ρ meson is selected. The pion with charge opposite to that of the τ_h is treated as a pseudo-neutral pion, and the same-sign pion is used to construct the ZMF. The decay plane method is then applied as in the 1-prong case.

In decays such as $\tau^\pm \to \pi^\pm\pi^0\nu_\tau$, the polarimetric vector in the τ rest frame can be approximated as:

$$\vec{h} = \mathcal{N}\left(2(q\cdot N)\vec{q} - q^2\vec{N}\right) \simeq \mathcal{N}\left(2m_\tau(E_{\pi^\pm} - E_{\pi^0})\vec{q} - q^2\vec{N}\right), \tag{1.42}$$

where $q = (q^0, \vec{q})$ is the four-momentum difference between the charged and neutral pions, and N is the neutrino four-momentum. The neutrino momentum is neglected, and the approximation introduces a sign flip in $\vec{h}$ depending on the energy difference $E_{\pi^\pm} - E_{\pi^0}$. Since both the polarimetric vector and τ momentum lie in the decay plane, their cross-product is approximately parallel to the vector orthogonal to the pion momenta:

$$\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0} \parallel \vec{h} \times \vec{p}_\tau = \vec{n}_\perp \tag{1.43}$$

Thus the ϕ_{CP} can be reconstructed as:

$$\phi^* = \arccos\left(\hat{n}_\perp^+ \cdot \hat{n}_\perp^-\right). \tag{1.44}$$

To resolve the orientation ambiguity, the following quantity is computed:

$$O^* = \hat{p}_{\pi^\pm}^- \cdot \left(\hat{n}_\perp^+ \times \hat{n}_\perp^-\right). \tag{1.45}$$

The acoplanarity angle ϕ_{CP} is then defined in the interval $[0, 2\pi]$ as:

$$\phi_{CP} = \begin{cases} \phi^* & \text{if } O^* \geq 0, \\ 2\pi - \phi^* & \text{if } O^* < 0. \end{cases} \tag{1.46}$$

IV.4. Common Notation

The methods discussed in the previous section can be expressed in a common notation [18], which is as follows.

All methods reviewed rely on defining the decay plane of each τ lepton using two vectors:

1. the momentum of a charged decay product, denoted as $\vec{P}_\pm$,
2. a reference vector $\vec{R}_\pm$, which must not be parallel to $\vec{P}_\pm$.

The Figure 1.11 shows an illustrative scheme of the acoplanarity angle. The acoplanarity angle is defined by boosting all vectors into the zero-momentum frame (ZMF) of the two charged decay products, where the total momentum satisfies:

$$\vec{P}_+ + \vec{P}_- = 0. \tag{1.47}$$

In this frame, the boosted momenta of the charged decay products are still denoted by $\vec{P}_\pm$, and the corresponding boosted reference vectors by $\vec{R}_\pm$. The acoplanarity angle is then computed as a function of these four vectors.

$$\phi^* = \begin{cases} \arccos\left(\frac{\vec{R}_+^\perp \cdot \vec{R}_-^\perp}{|\vec{R}_+^\perp||\vec{R}_-^\perp|}\right), & \text{if } O \geq 0, \\ 2\pi - \arccos\left(\frac{\vec{R}_+^\perp \cdot \vec{R}_-^\perp}{|\vec{R}_+^\perp||\vec{R}_-^\perp|}\right), & \text{if } O < 0. \end{cases} \tag{1.48}$$

where

$$O = (\vec{R}_+^\perp \times \vec{R}_-^\perp) \cdot \vec{P}_-, \tag{1.49}$$

and

$$\vec{R}_\pm^\perp = \vec{R}_\pm - \left(\frac{\vec{R}_\pm \cdot \vec{P}_\pm}{|\vec{P}_\pm|^2}\right)\vec{P}_\pm. \tag{1.50}$$

Method	P^-	R^-	P^+	R^+	π **Phase-Shift**
Impact parameter	$\vec{p}_p$	$\vec{IP}$	$\vec{p}_p$	$\vec{IP}$	-
Neutral pion	$\vec{p}_\pi$	$\vec{p}_{\pi^0}$	$\vec{p}_\pi$	$\vec{p}_{\pi^0}$	$y_+ \cdot y_- < 0$
Polarimetric vector	$\vec{p}_\tau$	$\vec{h}$	$\vec{p}_\tau$	$\vec{h}$	-

Table 1.2: Summary of methods and corresponding vectors used.

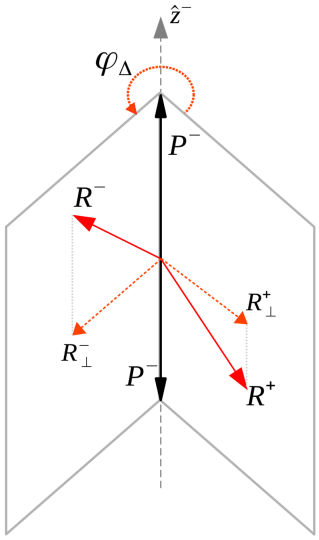


Figure 1.11: Schematic representation of the acoplanarity angle definition, following the notation introduced in this paper. The vectors $\vec{R}^\pm_\perp$ lie in the respective $\tau^\pm$ decay planes and are orthogonal to the axes defined by the vectors $\vec{P}^\pm$. The acoplanarity angle ϕ_Δ is defined as the counter-clockwise angle from the τ^+ decay plane to the τ^- decay plane, measured around the $\hat{z}^-$ axis. [18].

Chapter 2

From the CMS Phase-1 Detector at LHC to the Phase-2 Upgrade at HL-LHC

Content

I. The LHC 33

I.1. Accelerator Complex 33

I.2. LHC Design and Specifications 33

I.3. Schedule of Operations 38

II. The CMS Experiment 39

II.1. Coordinate System 40

II.2. Anatomy of the Detector 42

II.2.a. Superconducting Magnet 42

II.2.b. Silicon Tracker 43

II.2.c. Calorimeters 47

II.2.d. Muon Chamber 50

II.3. Trigger System 53

III. Silicon Tracking Detectors for High Energy Physics . . 53

III.1. Silicon Detectors Technology 54

III.1.a. Introduction to Semiconductors 54

III.1.b. p-n Junctions and their Role in Silicon Detectors 56

III.2. Signal Generation 58

III.2.a. Interactions of Charged Particles with Matter . 58

III.2.b. Signal Generation in a Silicon Tracker 59
III.3. Radiation Damage 60
III.3.a. Non-Ionizing Effects 60
III.3.b. Ionizing Effects 63
IV. The CMS Tracker Phase-2 Upgrade 63
IV.1. The High Luminosity LHC Upgrade 63
IV.1.a. Physics Motivation 63
IV.1.b. HL-LHC Upgrade Project 65
IV.2. CMS Phase-2 Upgrade 66
IV.3. Phase-2 Upgrade of the CMS Tracker 67
IV.4. Phase-2 Upgrade of the CMS Inner Tracker 69
IV.4.a. Overview 70
IV.4.b. The Silicon Pixel Module 70
IV.4.c. Pixel Module Readout Chip 72

High-energy physics experiments rely on particle colliders, which accelerate particles to high energies to collide them, and on particle detectors, which record the resulting interactions. The studies presented in this thesis were conducted with the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC).

Following the current operation of the Phase-1 CMS detector, which ends with Run 3, the upcoming high-luminosity LHC (HL-LHC) phase requires significant upgrades to maintain and extend the CMS performance under high-luminosity conditions. In this context, the CMS Phase-2 Upgrade is planned to prepare the detector for an operation at high instantaneous luminosity and radiation levels.

This chapter first introduces the LHC and its operating conditions, including the concepts of luminosity and pileup. The CMS detector and its key subsystems are then described, along with an overview of its coordinate and trigger systems. Particular focus is given to particle detection in silicon trackers, as this thesis work is related to such detectors. Finally, the motivations, design choices, and technical aspects of the CMS Tracker Phase-2 Upgrade are presented, forming the basis for the work discussed in Chapter 3.

I. The LHC

Located on the French–Swiss border near Geneva, the Large Hadron Collider (LHC) [19] at CERN [1] [20] is the world's largest and most powerful particle accelerator. Proposed in the early 1980s, the project was approved by the CERN Council in December 1994, while the Large Electron-Positron (LEP) collider was actively searching for the Higgs boson. In November 2000, LEP was officially shut down to allow the construction of the LHC. Following a decade of construction, the LHC began its first data-taking period (referred to as a *Run*) in 2008 and is currently in Run 3. At the end of Run 3, the LHC will be upgraded to achieve its High-Luminosity (HL-LHC) specifications, with operations expected to continue into the early 2040s. A detailed description of the accelerator complex and its operations is provided in the following sections.

I.1. Accelerator Complex

The LHC is a 27 km long circular collider designed to collide protons at a center-of-mass energy ($\sqrt{s}$) of 14 TeV. However, the highest collision energy recorded to date is 13.6 TeV, achieved in 2022. When not stated, the following description refers to protons, but the LHC is also capable of colliding heavy ions with energies up to 5.5 TeV.

The proton beams originate from dihydrogen (H_2) gas, where hydrogen atoms are stripped of their electrons using an electric field. Each beam is composed of bunches that contain approximately 1.15×10^{11} protons.

To reach high energies, protons are sequentially accelerated through a series of pre-existing accelerators that were upgraded and adapted for LHC operation. They are first accelerated to 50 MeV in the Linear Accelerator (Linac 4 [98]), to 1.4 GeV in the Proton Synchrotron Booster (PSB [99]), to 26 GeV in the Proton Synchrotron (PS [100]), and finally to 450 GeV in the Super Proton Synchrotron (SPS [101]), before being injected into the LHC ring.

A schematic view of the CERN accelerator complex is shown in Figure 2.1, where the four major detectors, ALICE [102], ATLAS [103], CMS [104], and LHCb [105], are located at each of the four interaction points (IP). Further details on the accelerator chain can be found in [19].

I.2. LHC Design and Specifications

Charged particles emit electromagnetic radiation, known as *synchrotron radiation*, when they are accelerated along a curved trajectory. The radiated power scales

[1] Conseil Européen pour la Recherche Nucléaire

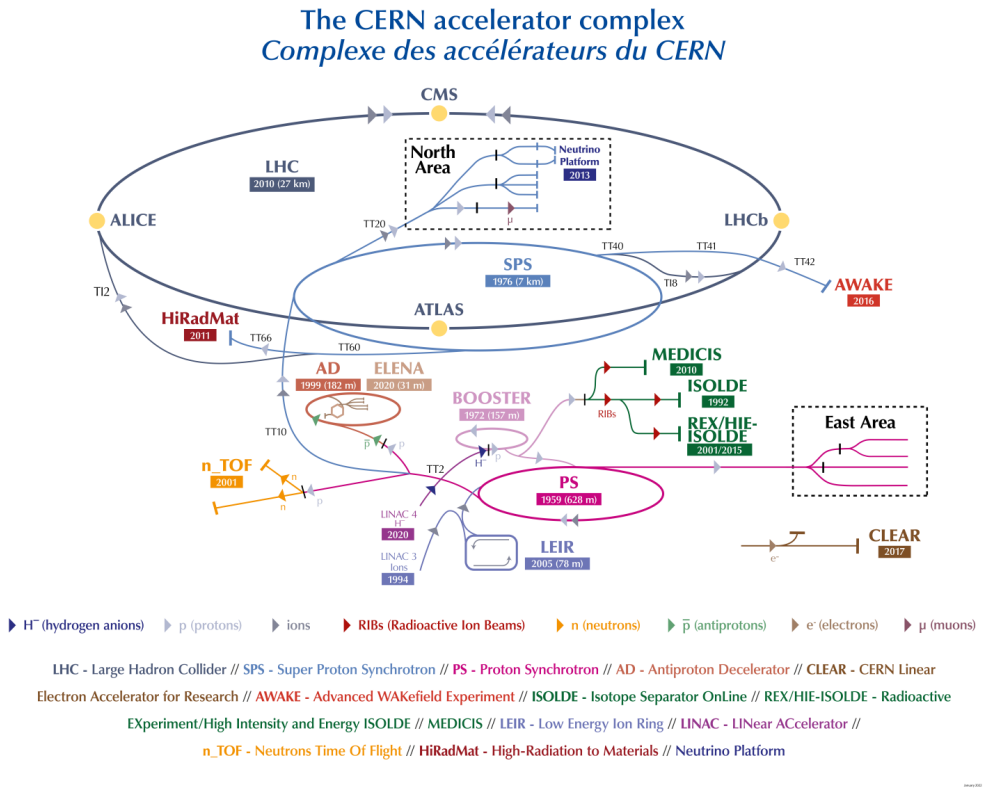


Figure 2.1: The CERN accelerator complex, showing the chain of accelerators leading to the LHC. The LHC is represented by the final dark blue ring, with the four major experiments indicated by yellow dots [106].

inversely with the fourth power of the particle mass m^{-4}, making the effect particularly significant for light particles such as electrons. Consequently, the LEP accelerator, an electron collider, was replaced by the LHC. As the larger mass of the proton reduces synchrotron radiation by a factor of approximately 10^{13}, it enables the LHC to reach higher energy, and probe new processes such as those including the discovery of the Higgs boson. However, protons are composite particles, which introduce more complex collision dynamics and a larger background than electron colliders.

In addition to the beam energy, a key parameter that impacts the collision rate of the LHC is the instantaneous luminosity, $\mathcal{L}$. The LHC operates with *bunched beams*, meaning that the proton beams are not continuous but consist of discrete bunches. Typically, up to 2808 bunches per beam are circulated simultaneously, with a bunch spacing of 25 ns during Run 2 operations. Some bunch slots are intentionally left empty to reduce beam loading effects, to provide machine protection

windows (such as abort gaps), or to facilitate background and noise measurements. The arrangement of filled and empty bunches along the beam is referred to as the *bunch filling scheme*, which is optimized to balance luminosity performance, beam stability, and detector requirements.

The instantaneous luminosity depends on several beam parameters and can be expressed as:

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

where N_b indicates the number of protons in each of the n_b bunches per beam, which circulate in the tunnel at a revolution frequency $f_{\text{rev}} \simeq 11.245\,\text{kHz}$. The collision frequency is therefore given by $n_b \times f_{\text{rev}}$, modulo the bunch filling scheme. The Lorentz factor of the protons is denoted by γ. The normalized transverse emittance ϵ_n characterizes the spread of the proton distribution in the transverse plane. The parameter β^* corresponds to the value of the optical β function at the IP. The β function, $\beta(s)$, describes the variation of the transverse beam size along the accelerator as a function of the longitudinal position s. It characterizes the envelope of the particle trajectories, which oscillate around the design orbit due to the focusing provided by the magnetic field. A smaller value of β implies a tighter focusing and hence a smaller beam size. A detailed description of the β function and its role in beam optics can be found in standard accelerator physics references, such as Ref. [107].

The factor F accounts for the reduction in instantaneous luminosity due to the crossing geometry of the beams. It depends on the crossing angle θ_c at the IP, as well as on the longitudinal and transverse root mean square (r.m.s.) bunch sizes, denoted σ_z and σ_{xy} respectively, at the point of collision. By definition, $F \leq 1$, and its general expression is:

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

The LHC parameters, including their nominal values, are summarized in Table 2.1. The instantaneous luminosity $\mathcal{L}$ is directly related to the number of *events* N produced per unit time, where an event corresponds to a single proton-proton collision resulting in the production of detectable final-state particles, as expressed by

$$\frac{\partial N}{\partial t} = \mathcal{L} \cdot \sigma \tag{2.3}$$

where σ is the total cross section for proton-proton interactions. This highlights that a large instantaneous luminosity is crucial to maximize the number of

Parameter	Description	Value
$\sqrt{s}$	Center-of-mass energy	14 TeV
N_b	Particles per bunch	1.15×10^{11}
n_b	Number of bunches per beam	2808
f_{rev}	Revolution frequency	11.2 kHz
ϵ_n	Transverse beam emittance	3.75 μm
β^*	Beta function at CMS IP (focal length)	0.55 m
Δt_b	Bunch spacing	25 ns
θ_c	Collision angle	285 μrad
σ_z	Bunch longitudinal r.m.s at CMS IP	7.55 cm
σ_{xy}	Bunch transverse r.m.s at CMS IP	16.7 μm

Table 2.1: Summary of LHC beam parameters at nominal conditions [19].

collisions produced. By integrating the instantaneous luminosity over time, the integrated luminosity, $\mathcal{L}_{\text{int}} = \int \mathcal{L}\,\mathrm{d}t$, is obtained, which evaluates the total number of collisions accumulated. The instantaneous luminosity is measured in $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, while the integrated luminosity is typically expressed in units of inverse femtobarns (fb^{-1}), with $1\,\mathrm{fb}^{-1} = 10^{39}\,\mathrm{cm}^{-2}$ and $1\,\mathrm{cm}^2 = 10^{24}\,\mathrm{b}$.

In the LHC, protons are grouped into discrete bunches that circulate in opposite directions in the two beams. A *bunch crossing* occurs when the two beams pass through each other at the interaction points (IPs), resulting in proton-proton collisions. However, due to the high luminosity at which the LHC operates, multiple proton-proton interactions can take place within the same bunch crossing window. This phenomenon, where more than one interaction happens simultaneously, is referred to as *pileup* (PU).

PU significantly affects the detector's capability to accurately reconstruct individual proton-proton interactions. Beyond the simple overlap of signals in the detector, a major challenge lies in correctly associating tracks and calorimeter clusters to their corresponding primary vertices. Moreover, PU can be categorized into *in-time* PU, arising from multiple interactions within the same bunch crossing, and *out-of-time* PU, originating from interactions in adjacent bunch crossings but still contributing signals due to the finite detector response time. The average PU, denoted as $\langle PU \rangle$, can be estimated from the inelastic proton-proton cross section, σ_{inel}, and the instantaneous luminosity using the formula:

$$\langle PU \rangle = \frac{\mathcal{L} \cdot \sigma_{\text{inel}}}{f_{\text{rev}}} \tag{2.4}$$

The inelastic proton-proton scattering cross section at a center-of-mass energy of 13 TeV is measured to be: $\sigma_{\text{inel}} = 68.6 \pm 0.5(\text{syst}) \pm 1.6(\text{lumi})\, mb$ [108]. This value leads to an average PU of approximately $\langle PU \rangle \approx 22$ under the LHC's nominal design conditions. However, as the LHC's luminosity has steadily increased over time, its operational conditions have evolved, and the instantaneous luminosity has now exceeded its nominal design value by a factor of 2 to 2.5. This has resulted in a significantly higher average PU of around 52, with peak PU reaching values over 80. Figure 2.2 shows the cumulative integrated luminosity delivered during different data-taking years, illustrating the steadily increasing dataset collected by CMS [109]. This increase in luminosity directly correlates with the rise in PU levels observed over the years.

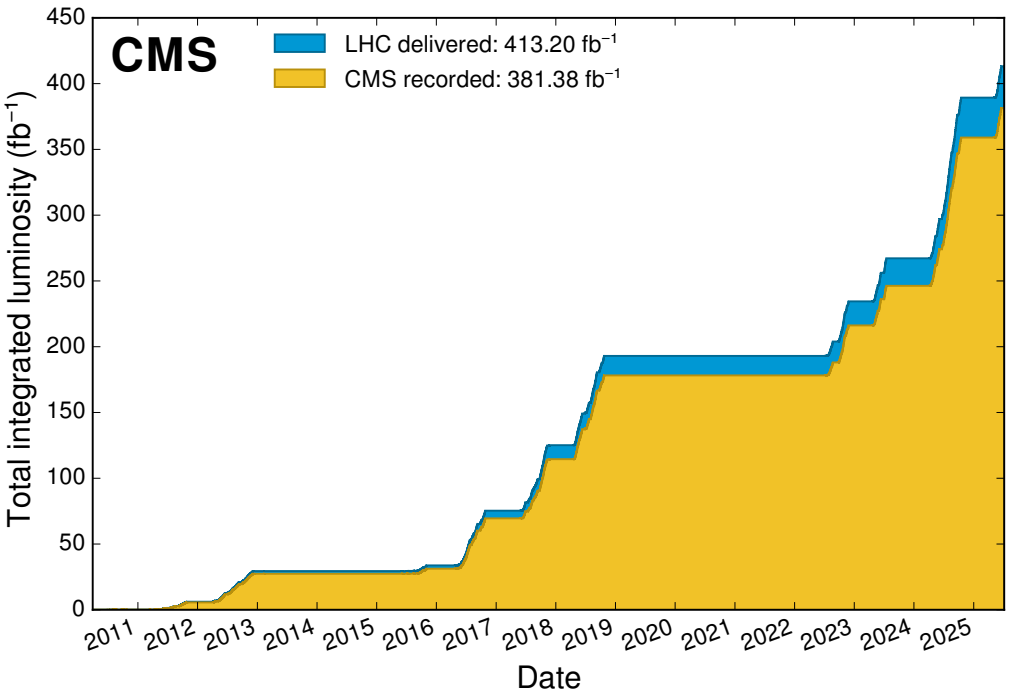


Figure 2.2: Cumulative integrated luminosity delivered by the LHC in proton-proton collisions over different data-taking years [110].

I.3. Schedule of Operations

The LHC's operational timeline is divided into two primary phases: Phase-1 (2011–2026) and Phase-2 (2029–2041). Figure 2.3 illustrates both the schedule and luminosity of LHC operations. Phase-1 will be followed by an upgrade to the High-Luminosity LHC (HL-LHC).

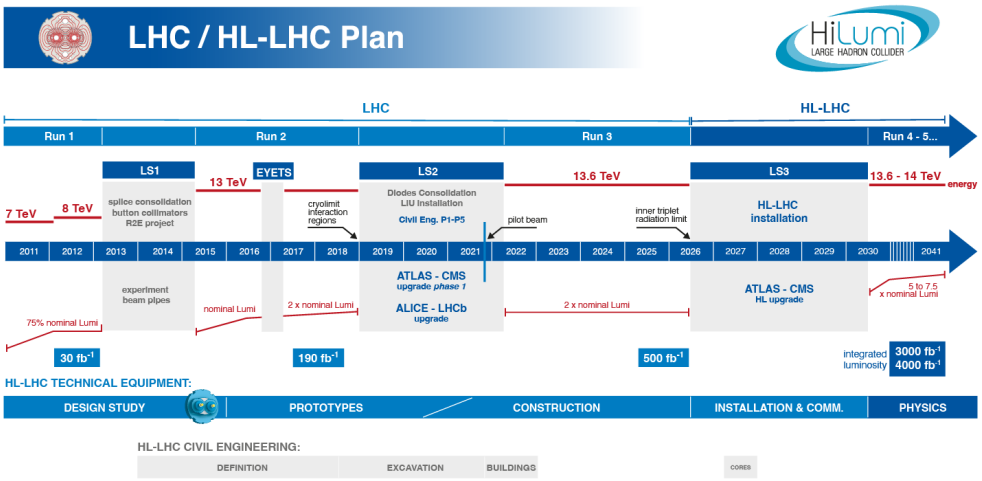


Figure 2.3: LHC operational phases and luminosity schedule, as planned for the HL-LHC upgrade [111].

Run 1, which started in 2009 and ended in 2013, marked the first completed operational phase of the LHC. During this period, the LHC recorded 6.1 fb^{-1} at 7 TeV and 23.3 fb^{-1} at 8 TeV, with a bunch spacing of 50 ns. These data were crucial for the historical discovery of the Higgs boson by the CMS and ATLAS Collaborations [11, 12], announced on July 4th, 2012. After Run 1, the LHC entered a two-year shutdown period, during which the beam energy was increased to 6.5 TeV by reinforcing the superconducting magnets and detector systems to support higher luminosities.

The second data-taking period, Run 2, began in 2015 and ended in 2018. During this period, the LHC operated at $\sqrt{s} = 13$ TeV. Over the four years of Run 2, the LHC delivered 4.2 fb^{-1}, 41.0 fb^{-1}, 49.8 fb^{-1}, and 67.9 fb^{-1} of integrated luminosity. The first year of Run 2 focused on commissioning the LHC in its upgraded configuration, while the subsequent years enabled significant discoveries, including the first observation of the Higgs boson's direct coupling to the τ lepton [112] and evidence of its coupling to muons [113]. At the end of Run 2, the LHC entered its Second Long Shutdown (LS2), lasting three years. Similar to LS1, LS2 in-

volved upgrades to enhance the subdetectors performances and initiated upgrades for Phase-2.

The latest operational phase of the LHC, Run 3, started in 2022 and is scheduled to last until 2026. This phase achieved a new record with a center-of-mass energy of $\sqrt{s} = 13.6$ TeV and an instantaneous luminosity of $2.6 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with peak PU exceeding 80. In its first three years, the LHC delivered integrated luminosities of 41.5 fb^{-1}, 32.7 fb^{-1}, and 122.2 fb^{-1}. This additional data will significantly enhance studies of rare processes and reduce uncertainties in existing measurements.

The conclusion of Run 3 will mark the end of Phase-1 and the beginning of Phase-2 with the transition to the HL-LHC. The Third Long Shutdown (LS3) will begin in 2026 and last three years. During this period, the LHC will undergo major upgrades to meet the HL-LHC specifications [114]. The instantaneous luminosity is projected to reach $5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, potentially reaching $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, over seven times the LHC's original design value, while restoring the center-of-mass energy to 14 TeV. Over its anticipated 12 years of operation, the HL-LHC is expected to deliver a total of 3000–4000 fb^{-1}.

II. The CMS Experiment

The site of CMS experiment is situated at one of the IP of the LHC, also known as Point 5, in the French village of Cessy. The CMS detector [115] is a general-purpose detector, designed to study a wide range of physics phenomena. The acronym CMS stands for *Compact Muon Solenoid*, each term reflecting a defining characteristic of the detector. While the term "compact" may appear counter-intuitive upon first encountering the detector, given its dimensions of 21 m in length, 15 m in diameter, and a total mass of approximately 14,000 tons, it is justified when compared to the ATLAS detector, which occupies nearly six times the volume while being roughly half as massive. The term "muon" highlights the extensive muon detection system, which occupies about 80% of the total detector volume and plays a crucial role in identifying and measuring high-momentum muons. Lastly, "solenoid" refers to the superconducting solenoid magnet, which produces a magnetic field of 3.8 T. This magnetic field bends the trajectories of charged particles to accurately measure their momentum and electric charge. As a general-purpose detector, CMS is designed to identify and reconstruct particles produced in collisions, along with their decay products. This is made possible by its cylindrical geometry, consisting of concentric layers of subdetectors, which are detailed in the following sections.

II.1. Coordinate System

Before any consideration, it is important to define the coordinate system used in this analysis, which is shown in Figure 2.4. The origin of the CMS coordinate system is defined at the interaction point of the beams. It is a right-handed Cartesian system, where the x-axis points toward the center of the LHC ring, and the y-axis points toward the surface, both forming a plane perpendicular to the proton beam, known as the transverse plane. The z-axis represents the longitudinal coordinate, pointing in the direction of the anticlockwise proton beam.

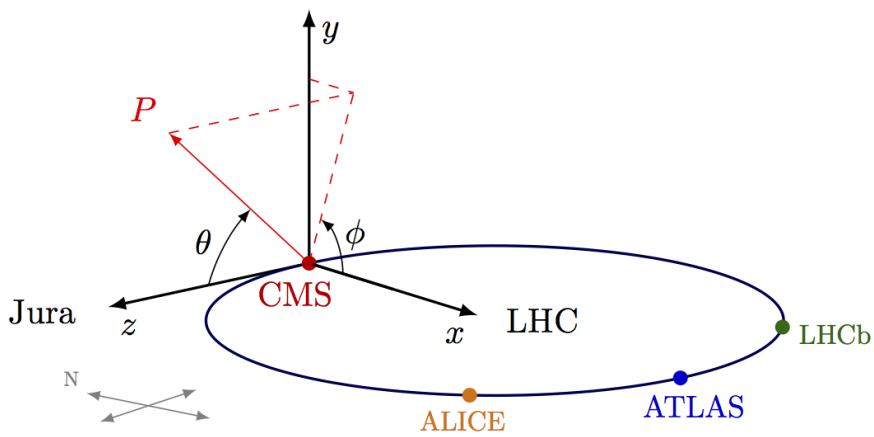


Figure 2.4: A schematic view of the standard coordinate system at the CMS detector, shown relative to the LHC and the other three main LHC experiments [56].

Given the cylindrical geometry of the detector, the polar coordinate system is also employed (see Figure 2.4). For any momentum vector $\vec{p}$, the transverse momentum p_T is defined as the component of the momentum in the transverse plane:

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

The azimuthal angle $\phi \in [0, 2\pi]$ represents the angle between p_T and the x-axis, defined as:

$$\phi = \tan^{-1}\left(\frac{p_y}{p_x}\right).$$

The polar angle $\theta \in [0, \pi]$ is the angle formed by $\vec{p}$ and the z-axis, given by:

$$\theta = \tan^{-1}\left(\frac{p_T}{p_z}\right).$$

as shown in Figure 2.4. Although the polar coordinate system is very useful, it is insufficient to fully describe the dynamics of hadronic collisions. This limitation arises because, in high-energy collisions, it is the individual constituents of the protons that interact. Indeed, the exact fraction of the proton's momentum carried by these constituents at the moment of the collision is unknown, so is the longitudinal boost of the collision's rest frame along the z-axis. To address this, the pseudorapidity is introduced because differences in pseudorapidity are invariant under longitudinal Lorentz boosts. It is defined as:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{2.5}$$

This variable ranges over $\eta \in [-\infty, \infty]$ and takes a value of zero for particles moving perpendicular to the beam (z) axis. The relation between θ and η is shown in Figure 2.5.

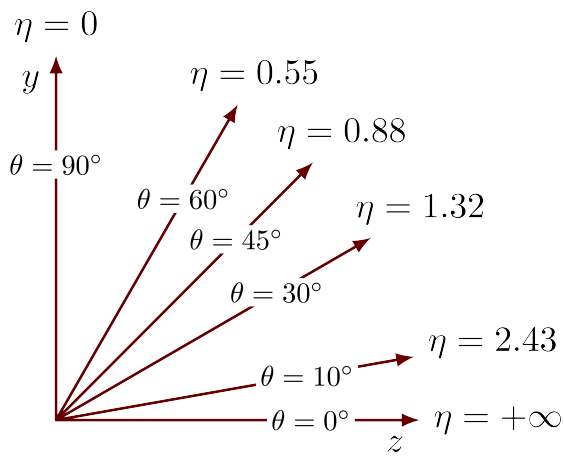


Figure 2.5: Relation between the pseudorapidity η and the polar coordinate θ from the standard coordinate system of CMS [56].

Having defined the pseudorapidity, one can also define the spatial separation between two particles:

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

As mentioned before, the total momentum along the z-axis is not known due to the unknown momentum fractions of the partons involved in the hard scatter. However, the transverse component of the parton momentum is typically negligible, and by conservation of momentum in the transverse plane, the sum of the transverse momenta of all final-state particles should be zero.

The CMS detector is designed to measure the energy and momentum of hadrons and charged leptons. However, some particles, such as neutrinos (or BSM particles), escape detection as they do not interact with the detector material. Their

presence can therefore only be inferred indirectly. To account for their contribution, the *missing transverse momentum* is defined as:

$$\vec{p}_T^{\text{miss}} = -\sum_i \vec{p}_T^{\,i}. \tag{2.7}$$

In the high-energy regime, the particle mass is typically negligible compared to its momentum, and hence the missing transverse energy, commonly referred to as MET (E_T^{miss}), is given by:

$$E_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}|. \tag{2.8}$$

II.2. Anatomy of the Detector

The CMS detector is composed of different subdetectors arranged in layers around the beamline, each with a specific role in the reconstruction of collision events. A schematic view of the CMS detector and it's subsystems is shown in Figure 2.6. As CMS has a cylindrical shape, it is composed of a cylindrical central region, called the *barrel*, and two *endcaps* that extend coverage at higher pseudorapidity.

Closest to the interaction point, the silicon pixel and strip detectors measure the trajectories of charged particles. Their curvature in the magnetic field generated by the solenoid magnet is used to estimate the particles' transverse momentum. The magnet provides a field strength of 3.8 T and encloses the inner tracker and the calorimeters.

Surrounding the tracker, the electromagnetic and hadronic calorimeters measure the energy of electrons, photons, and hadrons. Most particles stop in these layers and deposit their full energy. Neutrinos and other weakly interacting particles exit the detector without depositing energy.

Outside the calorimeters, muons are detected using dedicated chambers placed in the steel return yoke of the magnet. The trigger system selects a subset of collision events in real time for further analysis. Each of these systems will be described in more detail in the following subsections.

II.2.a. Superconducting Magnet

The superconducting magnet is a central component of the CMS detector. It generates a quasi-uniform magnetic field of 3.8 T along the z-axis to bend the momentum of charged particles. The curvature of a particle's trajectory, governed by the Lorentz force $\vec{F} = q(\vec{v} \times \vec{B})$, is used to infer its momentum. Inside the magnet, the uniform magnetic field bends particles with a lower momentum more sharply. Outside the magnet, the weaker field leads to reduced bending, making momentum measurements more challenging. The field's variation map is shown

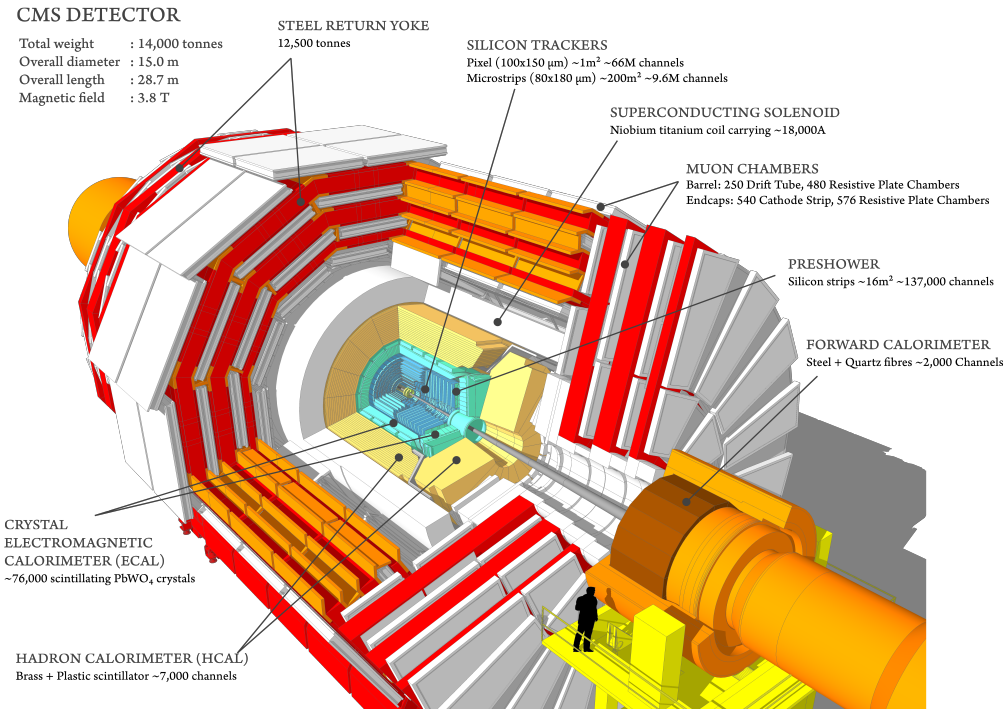


Figure 2.6: Schematic view of CMS detectors showing all the subdetectors [21].

on Figure 2.7 with an accuracy of less than 0.1% of the mapping, which is used for precise momentum reconstruction.

Regarding the magnet itself, it is a NbTi solenoidal magnet [117] with a diameter of 6 m, a length of 12.5 m, and a weight of 220 tonnes. This superconducting magnet requires cooling with a liquid He-4 cryostat system to maintain its operational temperature at 4.5 K ($-268.3°$C). Surrounding the magnet are the iron return yokes, which ensure a quasi-uniform 2 T magnetic field in the volume of the muon detectors, with a direction that varies spatially depending on the detector region.

II.2.b. Silicon Tracker

The silicon tracker depicted in Figure 2.8 is the innermost part of the detector [118, 119]. It relies on silicon sensor technology, which provides excellent spatial resolution and radiation hardness, along with a low material budget to minimize multiple scattering of charged particles. The readout electronics are synchronized with the 25 ns LHC bunch crossing clock, allowing hits to be associated with the corresponding bunch crossing, although the intrinsic time resolution of the sensors does not enable precise separation of closely spaced interactions in time. It is located between 4 cm and 110 cm in radius of the interaction point and up to 280 cm

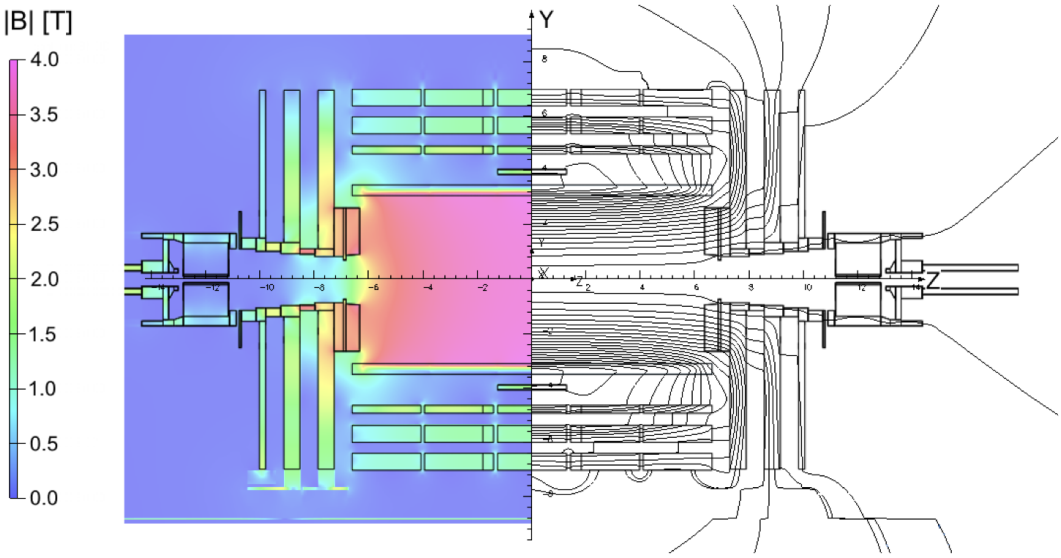


Figure 2.7: The value of the magnetic field strength $|B|$ (left) and the corresponding field lines (right) are shown for a longitudinal section of the CMS detector. This is based on the underground model, with a central magnetic flux density of 3.8 T. Each field line represents an increment of 6 Weber (Wb) in magnetic flux [116].

along the beam axis, covering a pseudo-rapidity range up to $|\eta| < 2.5$. It consists of 18 concentric barrel layers and 15 discs at both ends, covering a total surface of 210 m^2. The silicon tracker is composed of two tracking systems. The innermost part, referred as pixel tracker, consists of highly granular pixel detectors while the outer part, the strip tracker, of silicon strip detectors. Both detectors have independent cooling, powering, and read-out schemes. The pixel and strip trackers are cooled down at a temperature of about -10°C and -15°C, respectively to reduce the effects of radiation-induced damage and to minimize leakage current induced by the thermal noise.

Pixel Tracker: The pixel tracker receives the highest particle flux as it is the closest detector to the interaction point. For example, at a distance of 80 mm, a flux of around 10 million particles per square centimeter per second is expected. Hence, the detector requires a very good radiation tolerance, as it was designed to be operational for the three Runs. However, since the integrated luminosity exceeded the initially planned conditions, the pixel detector was upgraded during the 2016-2017 Run 2 Year-End Technical Stop (YETS) [121]. The specifications detailed here correspond to the pixel detector after the so-called Phase-1 upgrade.

The pixel tracker must also cope with high particle multiplicity and ensure high

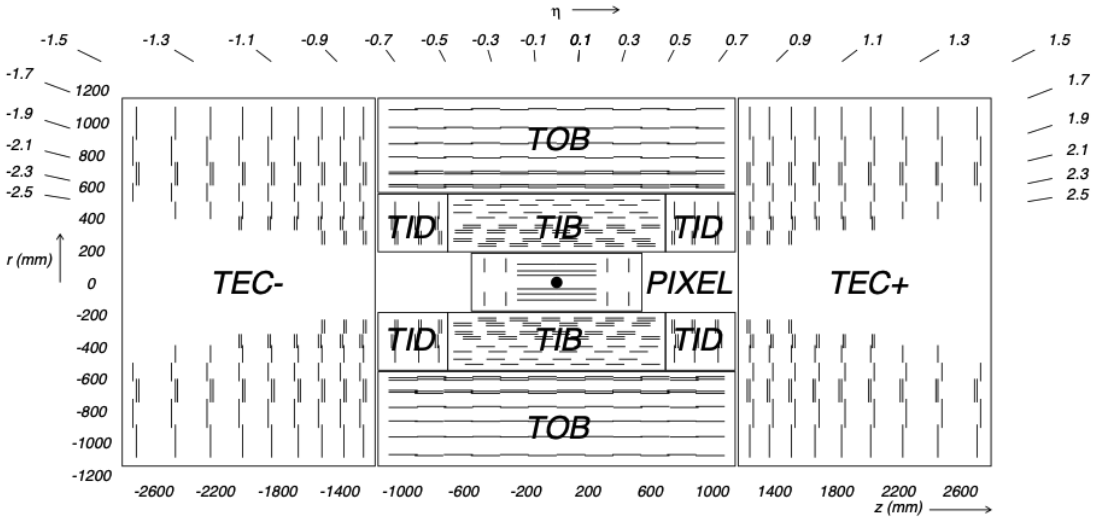


Figure 2.8: Schematic overview of the CMS tracker. Each line represents a detector module. Double lines indicate double-sided modules [120].

granularity. This is handled by the finely segmented pixel detector equipped with fast readout. This allows for an accurate measurement of the particles' charge and momentum by tracking their positions across multiple silicon detector layers. This information is then used for the reconstruction of the primary vertex, which is the initial hard-scatter interaction point, and to discriminate it against PU. Moreover, it enables the identification of in-flight decays, such as those of B-hadrons or τ leptons, allowing for the reconstruction of secondary vertices.

The pixel tracker consists of a barrel section made up of four barrel layers (BPIX) of 54 cm in length, located at radii of 2.9, 6.8, 10.9, and 16.0 cm, and two end-cap sections composed of three end-cap disks (FPIX) extending in radius from 4.5 to 16.1 cm, as shown in Figure 2.9. In total, the pixel tracker contains 79 million pixel sensors of size 100×150 μm^2. This ensures a spatial resolution on the reconstructed hits of 10 μm in the BPIX and 15 μm in the FPIX, where a *hit* refers to the signal produced when a particle interacts with the detector material, providing its position at a given moment.

The high segmentation of the detector generates a significant demand for readout electronics and power supply components, which introduces additional material budget, especially with the extra layer added during the Phase-1 upgrade. Therefore, the Phase-1 upgrade also introduced a new design for the pixel supports, a shift from on-board to off-board electronics, and the implementation of a CO_2 cooling system for heat dissipation. This led to a weight reduction of 40% and 80% for the barrel and endcap pixel detectors, respectively, with a substantial reduction in dead material, as shown in Figure 2.10.

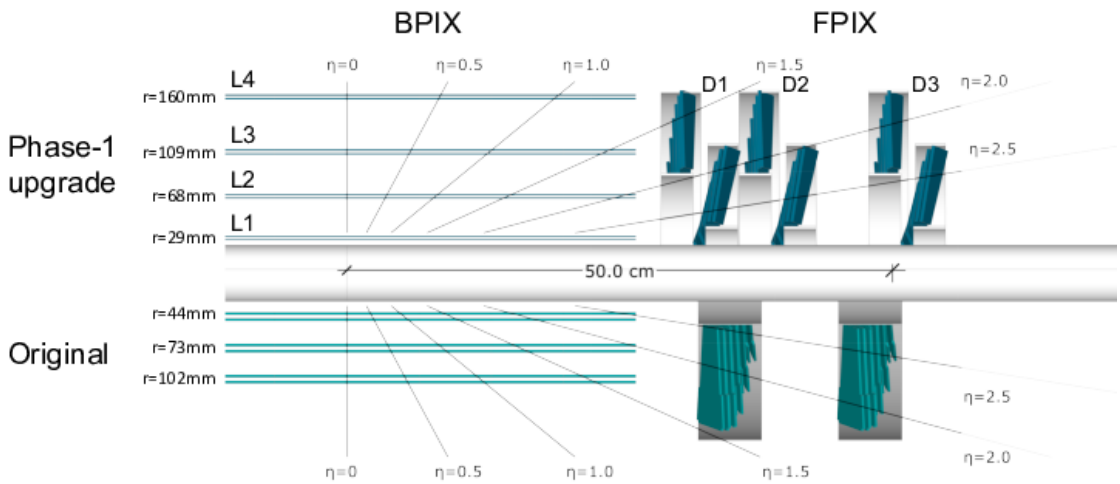
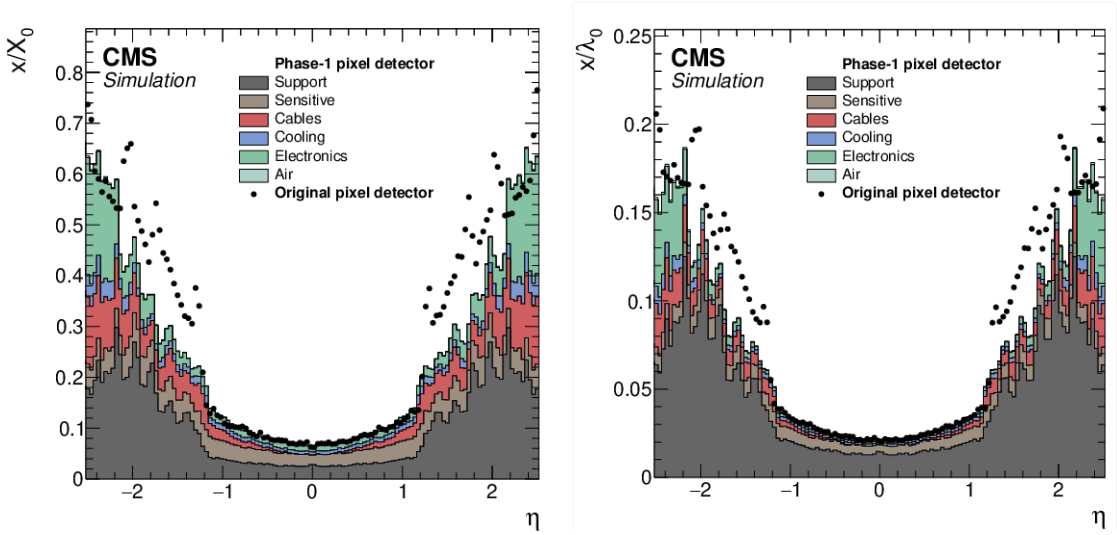


Figure 2.9: Layout of the pixel tracker compared to the original detector layout, in longitudinal view [121].

Strip Tracker: As the radial distance r from the interaction point increases, the particles spread out, covering a larger area. Consequently, at larger radii, the particle flux decreases. Hence, the fraction of the detector pixels that register hits from particles, referred to as *occupancy*, decreases with r^{-2}. As a result, a lower granularity detector can be used, and the strip tracker, which surrounds the pixel detector, is composed of silicon micro-strip modules. The strip tracker extends its volume up to a 110 cm radius and 540 cm length, covering a pseudorapidity range of $|\eta| < 2.5$. As shown in Figure 2.8, it is divided into four parts.

The Tracker Inner Barrel (TIB) is the innermost part, composed of four layers and covering the region $20 < r < 55$ cm. It is closed by the two endcaps of the Tracker Inner Disks (TID), consisting of three disks located at $58 < |z| < 124$ cm with radii up to 55 cm. These are surrounded by the six layers of the Tracker Outer Barrel (TOB), covering the region $55 < r < 116$ cm, and completed by the Tracker Endcap (TEC), with nine disks located at $124 < |z| < 282$ cm.

The strip modules of the TIB, TID, and the four innermost TEC rings have



(a) Radiation length X_0 (b) Hadronic interaction length λ_0

Figure 2.10: Simulated material budget of the CMS pixel tracker as a function of pseudorapidity η, expressed in units of radiation lengths (left) and hadronic interaction lengths (right). The original detector design is shown as black points, while the Phase-1 upgraded detector is represented with stacked histograms detailing contributions from individual components [121].

sensors with a thickness of 320 μm, while the outer modules have a thickness of 500 μm, with the inter-strip distance (strip pitch) varying from 80 μm to 205 μm for the different modules. The resolution for the TIB and TID strip modules in the $r - \phi$ plane is approximately 23–35 μm, depending on the radial position. For the outer TOB and TEC modules, the resolution in the $r - \phi$ plane ranges from 35 μm to 53 μm, with a larger uncertainty in the longitudinal direction due to the increased radial distance. Finally, to measure the endcap's radial coordinate and the barrel's longitudinal coordinate, several disks and layers are complemented by a second micro-strip sensor mounted back-to-back with the main one, tilted with a 100 mrad angle. This setup enhances the resolution of the system, especially in the longitudinal direction.

II.2.c. Calorimeters

The CMS calorimeter system is composed of an Electromagnetic Calorimeter (ECAL) [122] and a Hadronic Calorimeter (HCAL) [123], both designed to measure the energy of incident particles. They consist of high-density materials, the *active material* which force the incoming particle to decay in a cascade of secondary particles of lower energy, referred to as a *particle shower*. The two calorimeters

are optimized for different types of particle showers.

Electromagnetic showers result from electrons or photons losing energy through interactions with matter. This energy loss occurs through processes such as *Bremsstrahlung* (where electrons emit photons) and *electron-positron pair production* (where photons are converted into electron-positron pairs). As the shower progresses, these interactions produce additional secondary particles, leading to the creation of low-energy photons. The total detected energy corresponds to the energy of the incoming particle.

Hadronic showers are primarily driven by the strong nuclear force and result from hadron interactions, leading to the creation and decay of secondary hadrons. These hadrons ionize the active layers of the calorimeter, generating photons whose energy is collected by the scintillators. Additionally, some neutral hadrons, such as π^0, decay directly into photons, producing electromagnetic sub-showers, which are measured in the same way as those in the ECAL.

To accurately measure both the energy and trajectory of incoming particles, the CMS calorimeters are segmented in the transverse plane. This helps in differentiating between multiple particles and jets while also matching with particle trajectories recorded in the tracker.

Electromagnetic Calorimeter: The ECAL [122] is located just after the tracker in a cylindrical geometry. It is a homogeneous calorimeter formed by lead tungstate ($PbWO_4$) scintillating crystals. As previously explained, the measurement of energy relies on the electromagnetic shower initiated by the crystal material and resulting in the emission of low-energy photons. The choice of $PbWO_4$ was driven by its high density (8.28 g/cm^3), which ensures a small electromagnetic radiation length of $X_0 = 0.89\,\text{cm}$ and a low scintillation time, allowing approximately 80% of the scintillating light to be collected within the 25 ns time between two consecutive bunch crossings. Additionally, the material has a small Molière radius of $R = 2.19\,\text{cm}$, which contains 90% of the energy and allows the measurement of the incoming particle's direction of flight. This material is also characterized by good radiation resistance, enabling the calorimeter to operate with limited deterioration over several years. However, this comes at the cost of light yield, which is limited to a few photons per MeV, necessitating photodetectors with internal amplification, managed by the avalanche photo-diodes (APDs) in the barrel and vacuum photo-triodes (VPTs) in the endcaps.

As shown in the schematic view of the ECAL given in Figure 2.11, the ECAL is divided into three sections: the Preshower (ES) with pseudo-rapidity coverage $|\eta| < 0.9$ and $1.65 < |\eta| < 2.61$, one ECAL Barrel (EB) covering the pseudorapidity region up to $|\eta| < 1.479$, and two ECAL Endcaps (EE) closing the cylindrical design and covering the pseudorapidity region with $1.479 < |\eta| < 2.6$.

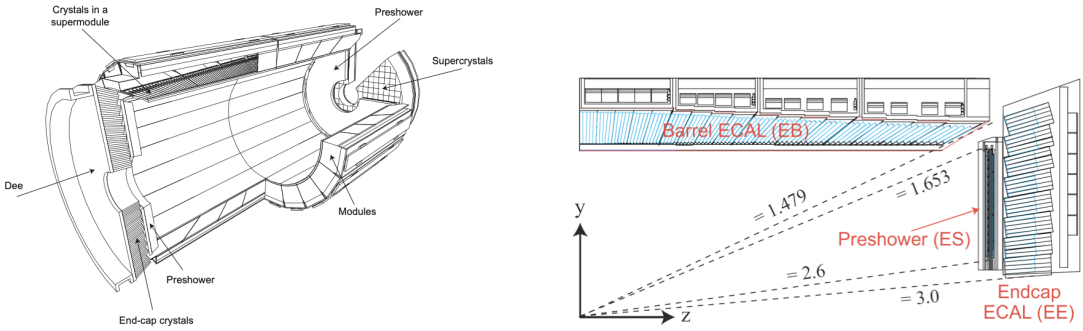


Figure 2.11: Schematic layout of the CMS electromagnetic calorimeter, presenting the arrangement of barrel, endcaps and the preshower in front (left). Geometric view of one quarter of the ECAL (right) [124].

A total of 68,524 crystals compose the barrel and endcap, with 61,200 in the barrel (length $25.8X_0$) and 7,324 in the endcap (length $27X_0$). The granularity for each crystal is $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$. To prevent photons from entering the small gap between adjacent crystals and escaping detection, the crystals are oriented with a slight tilt with respect to the nominal interaction point. Closing the barrel detectors on the sides, the EE comprises a total of 7,324 crystals, each 24.7 X_0 long. The high granularity of the EE allows differentiation between a single photon and two photons originating from the decay of a neutral pion when it is highly boosted.

The energy resolution of the ECAL, σ_{ECAL}, has been measured in 7 TeV data as [122]:

$$\sigma^2_{ECAL} = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2 \tag{2.9}$$

with E being the measured energy, where the first term is a stochastic term arising from statistical fluctuations in the energy of electromagnetic showers, the second is related to noise and PU, and the third is a term accounting for detector imperfections and energy leakage.

Hadronic Calorimeter: After the ECAL, the HCAL [123] is dedicated to the energy measurement of hadrons. Similarly to the tracker, the HCAL detector underwent a substantial upgrade, starting during LS1 and concluding during LS2, referred to as the HCAL Phase-1 upgrade. The HCAL covers the region $|\eta| < 5$. It is a sampling calorimeter consisting of alternating layers of passive absorbers and active detectors. As shown in Figure 2.12, the HCAL is divided into five distinct sections:

- **Barrel Hadronic Calorimeter (HB)**: The main part of HCAL, covering the $|\eta| < 1.4$ region. It is a sampling calorimeter consisting of interspersed brass absorbers and plastic scintillators. The HB is made up of 2304 so-called projective towers, which are small angular sections. Each projective tower covers a region $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, corresponding to a 5×5 array of ECAL crystals and ensuring high granularity in both the η and ϕ directions. Energy readings are performed using Silicon Photomultipliers (SiPMs).

- **Endcap Hadronic Calorimeter (HE)**: A sampling calorimeter made of brass absorbers and plastic scintillators. The HE covers the region $1.3 < |\eta| < 3$ and is segmented into 2304 towers, with lower granularity of $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$. Similar to the HB, energy readings are performed using SiPMs.

- **Outer Hadronic Calorimeter (HO)**: Located between the solenoid magnet and the muon chambers, it consists of plastic scintillators with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ and scintillating tiles located between the magnet's iron flux return yokes, which serve as absorber material. The readout of the HO is collected by wavelength-shifter fibers (WLF) coupled with SiPMs.

- **Forward Hadronic Calorimeter (HF)**: Located beyond the endcaps, the HF covers the region up to $|\eta| < 5$. Positioned at $|z| = \pm 11.1$ m away from the interaction point in the forward region of the detector, where the particle flux can reach $6 \times 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, it is composed of steel absorbers alternating with quartz fibers as active material. Incoming particles release energy in the form of Čerenkov radiation, which is read out by Photo-Multiplier Tubes (PMTs).

The energy resolution for the hadronic calorimeters in the barrel region and in the endcaps is [123]:

$$\frac{\sigma_E}{E} = \frac{1.72}{\sqrt{E\,(\mathrm{GeV})}} + 0.045. \tag{2.10}$$

And for the HF, the energy resolution is:

$$\frac{\sigma_E}{E} = \frac{1.72}{\sqrt{E\,(\mathrm{GeV})}} + 0.09. \tag{2.11}$$

II.2.d. Muon Chamber

The outermost part of the CMS detector is the muon-tracking system [126, 127], whose purpose is to detect muons and reconstruct their tracks. The muon system,

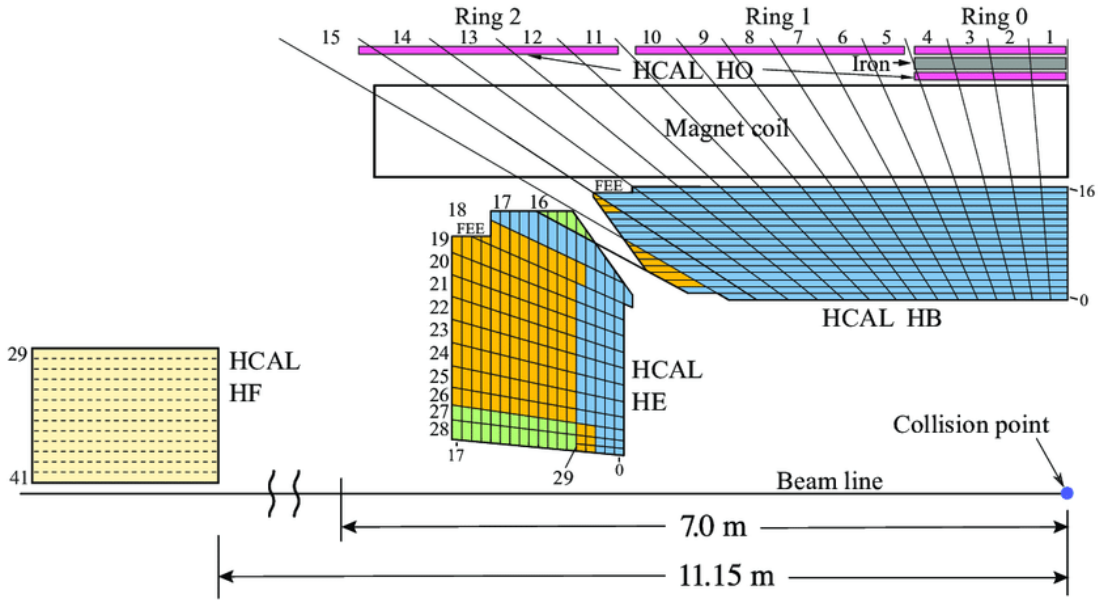


Figure 2.12: A schematic representation of one quarter of the CMS HCAL during the 2016 LHC operation, illustrates the locations of its four main components: the Hadron Barrel (HB), the Hadron Endcap (HE), the Hadron Outer (HO), and the Hadron Forward (HF) calorimeters. The layers marked in blue are grouped as depth = 1, while the layers in yellow, green, and magenta are assigned to depths 2, 3, and 4, respectively [125].

or the so-called muon chambers, consists gaseous detectors. With an average lifetime of $2.2\mu s$, the muons' decay length is much larger than the size of the CMS detector. Additionally, they are close to Minimum Ionizing Particles (MIPs), with energy ranging from a few to hundreds of GeV. Thus, they easily escape detection as they lose minimal energy while traversing the detector. Therefore, the muon detector is located outside the magnet where a 2 T magnetic field ensures a measurement of the muons' charge and momentum, complementary to that of the tracker. As shown in Figure 2.13, the muon chambers are divided in three regions with different technologies:

- Drift Tubes (DT): cover the central region of the detector with pseudo-rapidity coverage of $|\eta| < 1.2$. DTs are organized into four cylindrical concentric regions. The three innermost cylinders contain 60 drift chambers, while the outermost layer is made out of 70 drift chambers. Each tube has a section of 1.3×4.2 cm and is filled with gas mixture containing 85% Ar and 15% CO_2. The spatial resolution of the DTs is 100 μm.

- Cathode Strip Chambers (CSC): located in the forward region of CMS and

cover the region $0.9 < |\eta| < 2.4$ where 540 CSC chambers (270 per endcap) are implemented. These chambers are more granular to cope with high radiation exposure and the non-uniform magnetic field in which they are immersed. They are made with cathodes segmented into strips and with gas mixture of 40% Ar, 50% CO2 and 10% CF_4.

- Resistive Plate Chambers (RPC): consist of two parallel plates of phenolic resin with a separation of a few millimeters filled with a gaseous mixture. An electric field is applied between the RPC plates which act as electrodes, and the collected charges are read out by aluminum strips on the outside of the plates. The RPC's spatial resolution is poorer but it has excellent time resolution of 2-3 ns. This makes them a powerful trigger system.

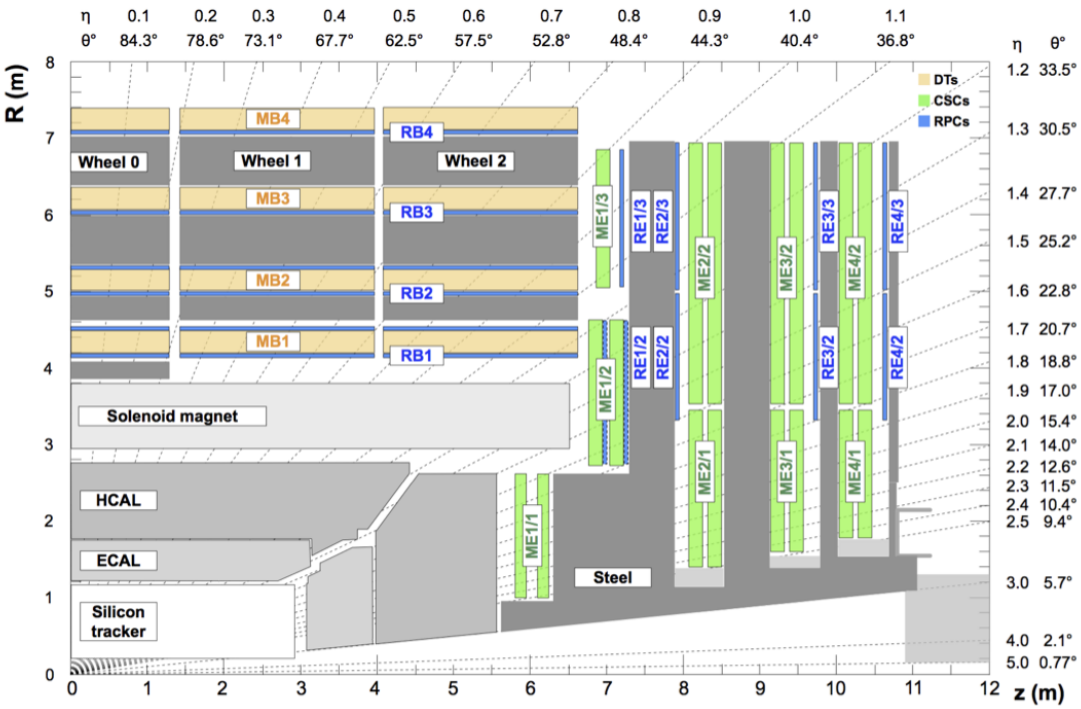


Figure 2.13: A schematic longitudinal view of one quarter of the muon detection system layout. The Drift Tubes (DT, yellow), Cathode Strip Chambers (CSC, green), and Resistive Plate Chambers (RPCs, blue) are displayed, with dashed lines indicating the pseudorapidity values [127].

II.3. Trigger System

The proton bunch crossing occurs every 25 ns (40 MHz) which corresponds to 40 million collisions per second. Given that each event produces around 1 MB of data, the total data size would reach 40 TB/s. This data volume is well beyond the capacity of the CMS Data Acquisition System (DAQ), which is capable of handling only about 100 MB/s for permanent storage. Furthermore, a significant fraction of these collisions consists of low-energy proton-proton interactions, that are not relevant to the CMS physics program, focused on hard scattering processes. Only a small subset of these events, referred to as *minimum bias*, is retained for detector calibration and luminosity measurements.

The CMS trigger system is used to reduce the data volume by filtering out the most interesting events. The trigger system operates in two stages: the Level-1 Trigger (L1) and the High-Level Trigger (HLT).

The L1 is implemented using dedicated hardware that processes the information from the calorimeters and muon systems with a reduced granularity. It has a maximum processing latency of 3.8 μs and selects the most significant events at a rate of up to 100 kHz. Events passing the L1 trigger are then transferred to the HLT for further filtering.

The HLT is a software-based selection system that uses the full detector granularity and operates on CPU and GPU processors. The HLT has a latency of 200 ms and further refines event selection, reducing the rate to approximately 1 kHz. Its primary task is to identify high-energy electrons, muons, photons, τ or jets while ensuring an efficient rejection of minimum bias events.

The selected events are then processed by the DAQ system and stored permanently in the CERN Tier-0 facility, the central hub of the computing grid. Overall, the trigger system achieves a real-time data reduction by a factor of 4×10^4, ensuring that only the most relevant events are retained while optimizing storage and processing efficiency.

III. Silicon Tracking Detectors for High Energy Physics

Silicon tracking detectors are widely used in high-energy physics experiments, especially for tracking and vertexing. They are mainly positionned at the innermost region of the detectors, close to the interaction point, where precise position measurements are essential.

One of their main advantages is that even with very thin sensors, typically only a few hundred microns thick, silicon can generate sufficiently strong signals due to its ionization properties. This allows for a minimal material budget, which

helps to reduce unwanted effects such as multiple scattering and secondary interactions, thereby improving tracking performance. Moreover, silicon sensors provide a prompt and localized electronic signal, which facilitates fast signal collection and allows for compact and efficient front-end electronics integration. This contributes to reducing the overall material in the tracker volume. Finally, silicon detectors exhibit good radiation hardness, enabling long-term operation under the harsh radiation environment close to the beamline. This section provides an introduction to silicon detectors, more details can be found in [128–130].

III.1. Silicon Detectors Technology

III.1.a. Introduction to Semiconductors

All solid materials are made up of an ensemble of atoms that define their electrical properties. Materials are classified into three classes, according to their electrical conductivity σ: insulators, semiconductors, or conductors. The notion of *energy bands* can be introduced to better understand the concept of conductivity and this classification.

The electrons of an atom have specific energy levels called atomic orbitals. When a large number of atoms are grouped together, they form a crystal lattice, resulting in solid material. In the lattice, the atomic orbitals of nearby atoms overlap, forming a continuum of energy levels known as the *energy band*. The filling of energy states with electrons follows the Fermi-Dirac distribution, expressed as:

$$f(E) = \frac{1}{1 + e^{(E-\mu)/(kT)}},$$

where k is the Boltzmann constant, T is the temperature, E is the energy and μ is the Fermi level. The Fermi level μ refers to the energy at which, at $T = 0$ K, all energy states are filled. The last band that is completely filled with electrons is known as the *valence band*, while the band above it, which may only be partially filled, is called the *conduction band*. The energy gap between the valence band and the conduction band is known as the *band gap*. The smaller the band gap, the higher the conductivity. The three different types of materials and the corresponding band energy levels are illustrated in Figure 2.14.

A conductor is a material that easily conducts electrical current. This is due to the fact that its valence band and conduction band overlap because its Fermi level lies in the conduction band. Thus, electrons can freely move in the material when a small electric field is applied.

In contrast, for insulators and semiconductors, the Fermi level lies within the band gap, meaning that additional energy is required for electrons to jump from the valence band to the conduction band. In the extreme case of a insulator, the

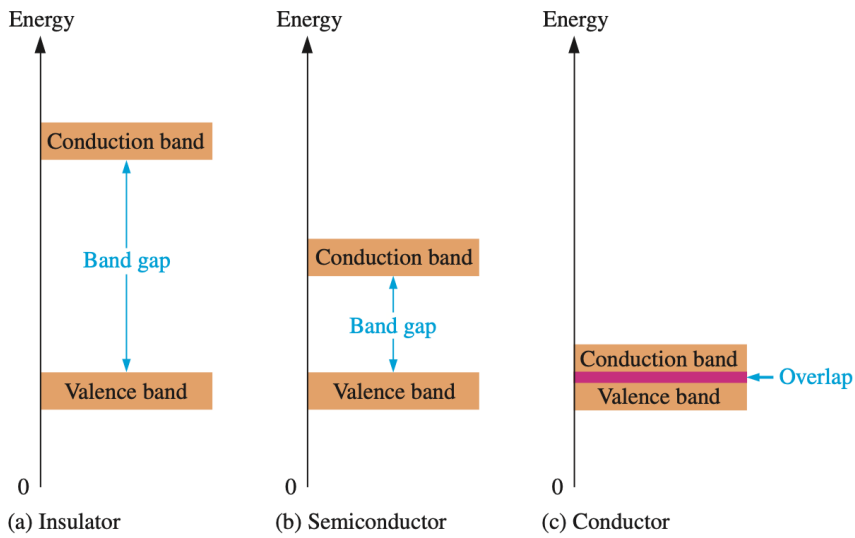


Figure 2.14: The three types of solid-state materials and their energy band diagrams: conductors, semiconductors, and insulators [131].

energy gap is so high that the material does not conduct an electrical current, except when a very high electric field applied.

In the case of semiconductors, the gap is much smaller compared to insulators. For example, in silicon at room temperature, the energy gap is typically $E_g = 1.12\,\text{eV}$ and the thermal energy kT can still excite an electron to jump from the valence band to the conduction band. When the electron leaves the valence band, this creates a hole that can be filled by the neighboring electron, producing an intrinsic current. For this reason, semiconductor materials are called *intrinsic semiconductors* and thermally excited electrons and holes are the intrinsic charge carriers. The electrons in an intrinsic semiconductor can also be excited by ionizing radiation, enabling them to act as detectors. Because intrinsic semiconductors are very sensitive to temperature, the noise level might be considerably higher than the signal produced by a charged particle traversing the silicon lattice, making them very poor detectors.

One way to change the conduction properties of a semiconductor is by introducing impurities into the silicon crystal. This process illustrated in Figure 2.15 is referred to as *doping* and semiconductors that have been doped are known as *extrinsic semiconductors*.

There are two types of doping. If the doping element belongs to group-V of the periodic table, which contains five valence electrons, a bound electron is added to the valence band. The free electron can jump into the conduction band, making the semiconductor's conductivity dominated by electrons. This type is known as

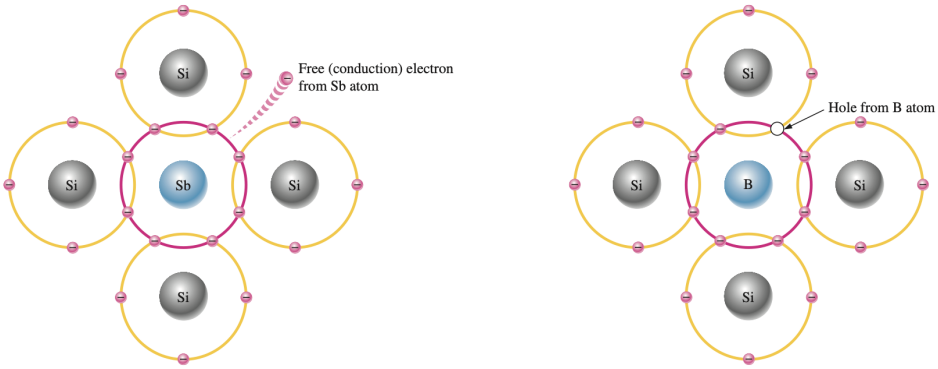


Figure 2.15: On the left, a pentavalent impurity atom of antimony (Sb) introduces an extra electron, creating an n-type semiconductor. On the right, a trivalent impurity atom of boron (B) creates a hole, forming a p-type semiconductor [131].

n-type semiconductor, and the dopant atoms are called *donors*.

On the other hand, if the impurity element belongs to group-III of the periodic table, a vacancy is created in the valence band, increasing the number of holes in the semiconductor. This type is known as *p-type* semiconductor, where the dopant atoms are called *acceptors*. The process of doping silicon causes a shift in the Fermi level and is used to produce one of the basic structures of silicon detectors, the *p-n junction*.

III.1.b. p-n Junctions and their Role in Silicon Detectors

As mentioned in the previous section, the noise in intrinsic silicon arising from thermal vibration dominates the signal produced by the ionization from passing particles. This makes intrinsic semiconductors not well-suited for particle detection. This thermal noise can be reduced by forming a *p-n junction*, created from p-doped and n-doped silicon materials.

A *p-n junction* is formed at the interface between p-type and n-type semiconductors, as illustrated in Figure 2.16. In the n-doped region, electrons are the majority carriers, and the Fermi level lies close to the conduction band. In contrast, in the p-doped region, holes dominate, and the Fermi level is near the valence band. Upon contact, electrons from the n-doped side diffuse into the p-doped side and recombine with holes, and vice versa. This movement, driven by the strong carrier concentration gradient, generates a diffusion current. As electrons and holes recombine at the interface, they leave behind ionized donor and acceptor atoms, leading to a region depleted of mobile charge carriers, known as the *depletion region*.

Although the depletion region is free of mobile carriers, the electric potential

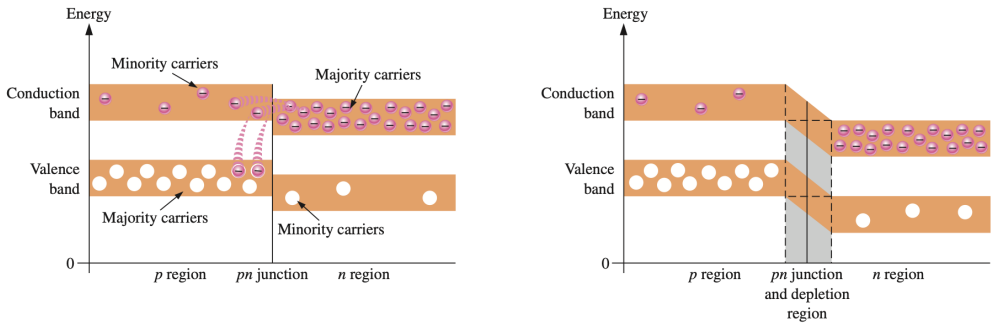


Figure 2.16: A schematic representation of the p–n junction at the instant of junction formation (left) and at equilibrium (right) [131].

difference induces an internal electric field pointing from the n-side to the p-side. The width of the depletion region can be controlled by applying an external voltage, called the *bias voltage*.

In *forward biasing*, an external electric field is applied with opposite polarization to the internal field, narrowing the depletion region and allowing the current to flow. In contrast, in *reverse biasing*, the typical operational mode for silicon detectors, the external electric field polarization is the same as the internal one, widening the depletion region and reducing conduction. The voltage at which the depletion region has the entire thickness of the sensor is called the *full depletion voltage*. Even in this state, a small *leakage current* persists, consisting of a diffusion component from minority carriers and a generation component due to thermally produced electron-hole pairs.

As the bias voltage increases much beyond the full depletion voltage, an *avalanche breakdown* may occur. This is characterized by a high increase in leakage current due to the multiplication of charge carriers. Figure 2.17 illustrates the current-voltage (IV) characteristics of the sensor, showing the onset of avalanche breakdown. This behavior needs to be carefully managed in detector design to avoid damage or loss of sensitivity, especially at high bias voltages.

The formation of the p–n junction creates an active volume sensitive to passing particles. Additionally, full depletion minimizes the noise and enhances charge collection by increasing the electric field, which improves carrier mobility and reduces the integration time.

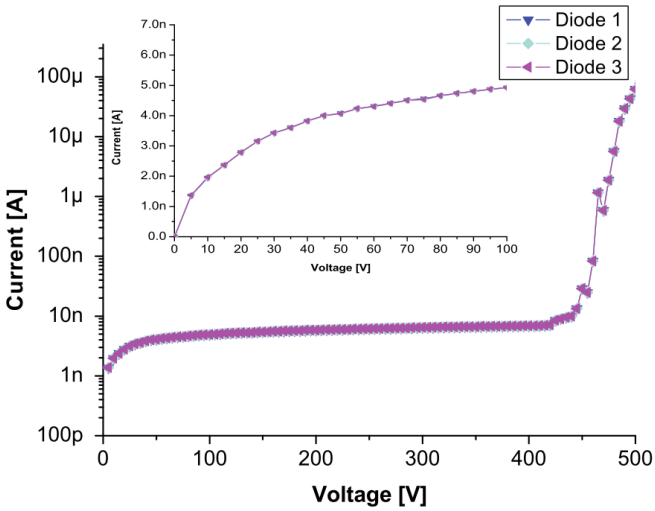


Figure 2.17: The current–voltage relation (I-V) for a silicon diode in the reverse bias direction is depicted. The expanded view shows the proportionality between I and $\sqrt{\mathrm{V}}$ and the full scan shows the breakdown at higher voltages [132].

III.2. Signal Generation

III.2.a. Interactions of Charged Particles with Matter

The principle of particle detection is based on measuring of the energy loss of a particle as it passes through the detectors. Through a semiconductor, an incident particle loses some of its kinetic energy, creating an electron-hole pair. The particles lose energy in matter either by ionization of atoms along their path or by excitation of electrons. Ionization occurs when an incident particle transfers its energy to an atom, producing a free electron. The incident particle energy can also excite an electron in an atom to elevate its energy to higher states. When this electron returns to its energy level, it emits a low-energy photon.

The average ionization and excitation energy loss per unit length for a relativistic charged particle in a medium is given by the Bethe-Bloch formula:

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c^2\beta^2\gamma^2 T_{\max}}{I^2}\right) - \beta^2 - \frac{\delta}{2}\right], \tag{2.12}$$

where N_A is the Avogadro's number, r_e and m_e the radius and mass of the electron, c the speed of light, z is the charge of the incident particles, Z and A are the atomic number and mass, $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$, T_{max} is the maximum kinetic energy that can be transferred to a free electron in a single collision, I the mean excitation energy and δ the density effect correction.

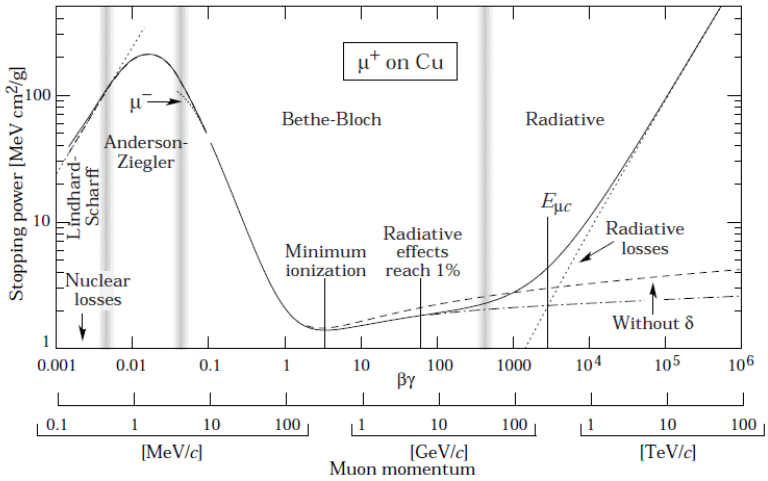


Figure 2.18: Mean energy loss in copper for traversing muons [15].

This formula is valid for particles with $0.1 < \beta\gamma < 100$, else, the full function of the average energy loss per unit length is shown in Figure 2.18. The minimum of this curve corresponds to the *minimum ionizing particles* (MIP). Since most of the particles of LHC are in this regime, detectors must be able to keep their noise well below this energy.

As the number of collisions and the energy transfer vary, the total energy loss is sensitive to statistical fluctuations. The statistics of the number of collisions can be described by a Poisson distribution, while the energy transfer in silicon is well described by a Landau distribution. This distribution is characterized by an asymmetric long tail toward high energy due to the high energy γ-electron that ionizes the material. As a result, the mean energy transfer is not equal to the most probable energy transfer, as illustrated in Figure 2.19. For example, in silicon, the mean ionization energy is 3.6 eV, about three times greater than the band gap of 1.12 eV. Thus, a minimum ionizing particle (MIP) traversing silicon typically produces about 76 electron–hole pairs per μm, with an average energy loss corresponding to approximately 108 pairs per μm in a 300μm thick sensor.

III.2.b. Signal Generation in a Silicon Tracker

At the LHC, silicon tracker detectors like those used in CMS rely on fully depleted p–n junctions. They are composed of a doped silicon base (p-type for the CMS Phase-2 upgrade) and strips or pixels of inverse type (n-type for the CMS Phase-2 upgrade). When a charged particle passes through the sensors, pairs of electron-

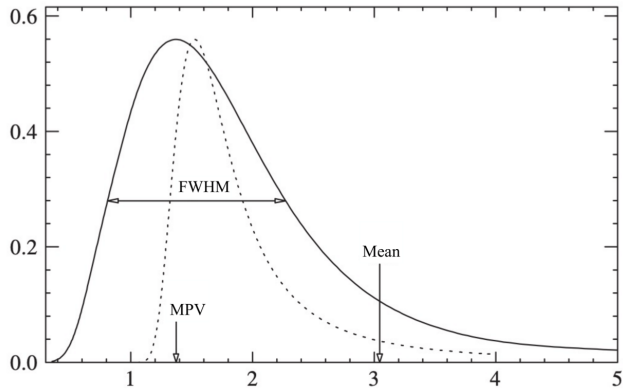


Figure 2.19: Landau distribution in solid line with arbitrary scales. The Full-Width-Half-Maximum (FWHM), the mean value and the Most Probable Value (MPV) have been added to the Figure [133] as illustration.

holes are created. These charge carriers drift into the depletion region and are finally collected by the electrode. According to the Shockley-Ramo theorem [134], the movement of a charge in the vicinity of an electrode induces a current at the electrode, creating a signal. This signal is then amplified and treated by the front-end electronics.

III.3. Radiation Damage

The cumulative effects of radiation can gradually damage silicon sensors. The damage can be induced by ionizing effects referred to as surface effects or by non-ionizing effects, referred to as bulk effects.

III.3.a. Non-Ionizing Effects

Non-ionizing effects are the main cause of bulk damage in silicon sensors. Also referred to as displacement effects, they are caused by heavy particles such as neutrons, protons, and pions. When passing through the material, these incident particles undergo elastic collisions with the atomic nuclei in the crystal lattice. Depending on the energy transfer, this can either knock atoms out of their lattice sites, creating point defects, or cause displacement cascades creating cluster defects.

When an atom moves from its initial position, it creates a silicon interstitial and a vacancy called a *Frenkel effect*, illustrated in Figure 2.20. Over time, the thermal energy can move the interstitial to its initial position (beneficial annealing) or to another stable point defect.

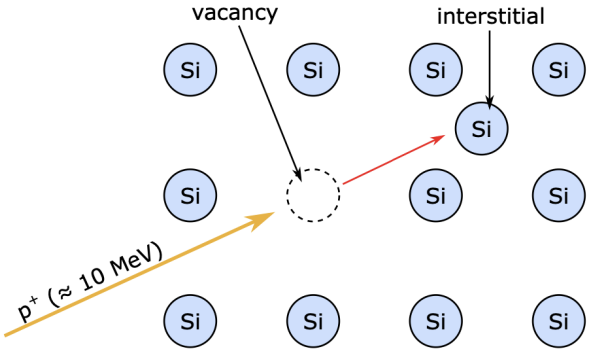


Figure 2.20: Vacancy and interstitial in the silicon lattice known as "Frenkel effect" [135].

The displacement damage is expressed in terms of the particle fluence ϕ, but can be quantified based on the effects on the materials using the *non-ionizing energy loss* (NIEL) hypothesis. In this hypothesis, any radiation damage induced by a displacement damage is linearly proportional to the non-ionizing energy loss of the traversing particles. As a result, the displacement damage depends on the nature of the particles. It is common to express the displacement damage by scaling the fluence ϕ to the equivalent fluence ϕ_{eq} of 1 MeV neutrons that would induce the same damage with the formula:

$$\phi_{eq} = \kappa\phi \tag{2.13}$$

where κ is the hardness factor describing the severity of the damage induced by the nature of the incident particles.

In the silicon sensor, displacement defects create additional energy levels in the band gap, as illustrated in Figure 2.21. If these defects are close to the edges of the band gap, they change the effective doping concentration as they act as donors or acceptors. Defects close to the center of the band gap act as traps.

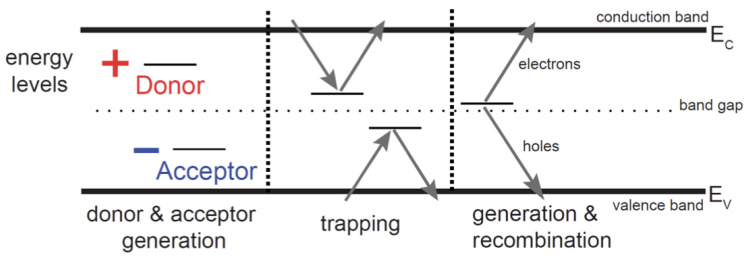


Figure 2.21: Radiation-induced defects in the silicon band gap [135].

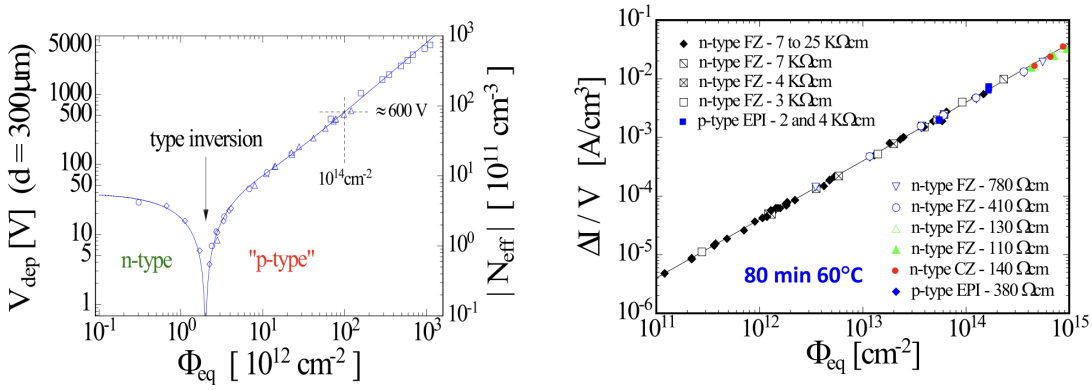


Figure 2.22: On the left, full depletion voltage and effective doping concentration as a function of the equivalent fluence ϕ_{eq}. On the right, leakage current increase caused by irradiation ΔI as a function of the equivalent fluence ϕ_{eq} for various silicon detectors [136].

Irradiation introduces defects that can act as effective donors or acceptors, depending on their energy levels relative to the band gap. In n-type silicon, these defects often act as acceptors, gradually compensating the original donor doping. This reduces the effective space charge and lowers the full depletion voltage, until the material undergoes *type inversion*, where it becomes a p-type. Beyond this point, the accumulation of acceptor-like defects increases the effective doping again, leading to a rise in the full depletion voltage, as illustrated in Figure 2.22.

In contrast, p-type silicon does not undergo type inversion at standard fluences. Instead, irradiation increases the effective acceptor concentration, which expands the depletion region and requires a higher bias voltage to achieve full depletion after significant damage. This increase in voltage leads to a higher leakage current, which in turn raises the power dissipated in the sensor. If the heat is not efficiently removed, this can cause a *thermal runaway*, where the increased temperature leads to a higher leakage current, which further increases the temperature. Therefore, efficient cooling becomes essential in maintaining sensor stability after high radiation exposure.

On the other hand, the traps induced by defects close to the center of the band gap affect the charge collection efficiency and increase the sensor leakage current. The leakage current increase δI caused by irradiation is linearly proportional to the particle fluence:

$$\delta I = \alpha \phi_{eq} V \tag{2.14}$$

with α the current-related damage rate and V the voltage. Figure 2.22 illus-

trates this relation.

III.3.b. Ionizing Effects

Surface effects are mainly caused by photons and charged particles that ionize matter as they pass through it, generating electron-hole pairs. The energy deposited in the material as the ionization energy is the dose measured in Gray (Gy).

When ionization occurs in the semiconductor, the electron-hole pairs recombine or drift toward the electrodes without causing any damage. However, ionization on the surface of the detector, is more problematic because the electron-hole pairs created in this material do not recombine in a very short time. This effect depends on the electric field, since the electric charges drift because of it. Electrons can leave the oxide because they have higher mobility, but holes are trapped in defect centers in the oxide, creating or activating defects at the silicon-oxide interface. More details on this can be found in [136].

IV. The CMS Tracker Phase-2 Upgrade

IV.1. The High Luminosity LHC Upgrade

The HL-LHC [114] is an upgrade of the existing LHC which aims to significantly increase its luminosity and enable precision measurements. The target is to reach an integrated luminosity of approximately 300 fb^{-1} per year, resulting in a total of up to 3000 fb^{-1} over the projected ten-year operational lifetime of the HL-LHC. The HL-LHC project, approved by the CERN Council in June 2016, is recognized as a top priority by the European Strategy for Particle Physics (ESPP) [137] and the US Particle Physics Project Prioritization Panel (P5) [138]. It is also included in the list of the European Strategy Forum on Research Infrastructures (ESFRI) "Roadmap 2016" strategy report on large research infrastructures [139].

IV.1.a. Physics Motivation

The HL-LHC is expected to operate at a center-of-mass energy of 14 TeV and with instantaneous luminosity five times higher than LHC, allowing for a significant enhancement in the precision of measurements. In particular, it will enable to measure rare and statistically limited SM processes and to probe BSM ones.

In the Higgs sector, the increased dataset and improved detector performance at the HL-LHC are expected to significantly improve the precision of coupling measurements, particularly for those processes that are currently limited by statistics [140]. For example, sub-percent precision is projected for couplings to vector bosons, while percent-level or better sensitivity will be achieved for couplings to

third-generation fermions. Moreover, the HL-LHC will allow the first meaningful constraints on rare decays such as $H \to \mu^+\mu^-$ and precise studies of the Higgs boson self-coupling, as illustrated in Figure 2.23. Regarding the search for CP violation in the Higgs to τ Yukawa coupling, the mixing angle $\alpha_{H\tau\tau}$ could be constrained between the scalar and pseudoscalar components of the Higgs boson to a precision of approximately 5° as shown in Figure 2.24. Additionally, precision measurements of the Higgs boson self-coupling via the observation of Higgs boson pair production are expected as well as the measurement of the trilinear Higgs boson coupling.

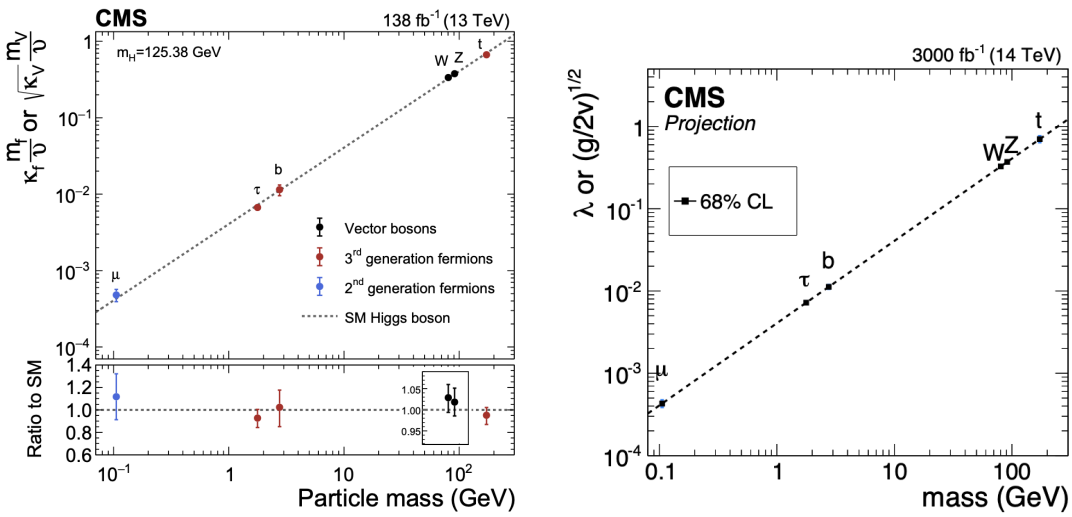


Figure 2.23: Higgs boson couplings to fermions (Yukawa couplings, λ_f) and to vector bosons (parametrized as $(g_V/2v)^{1/2}$, where v is the SM Higgs vacuum expectation value) as measured using Run 2 data [16] (left), and projected uncertainties for an integrated luminosity of 3000 fb^{-1}, assuming Standard Model couplings [141] (right).

The HL-LHC will significantly enhance the sensitivity to BSM physics, primarily due to the increased integrated luminosity and upgraded detector capabilities. Direct searches will extend the mass reach for supersymmetric particles (e.g. gluinos up to ∼3 TeV and electroweakinos up to ∼1 TeV), heavy resonances (Z', W'), and signatures of extra dimensions. In parallel, the larger dataset will improve sensitivity to rare or suppressed processes that may reveal indirect effects of BSM physics, such as rare Higgs decays, flavor-violating transitions, or deviations in electroweak couplings. These indirect probes offer complementary discovery potential in the absence of direct production [142].

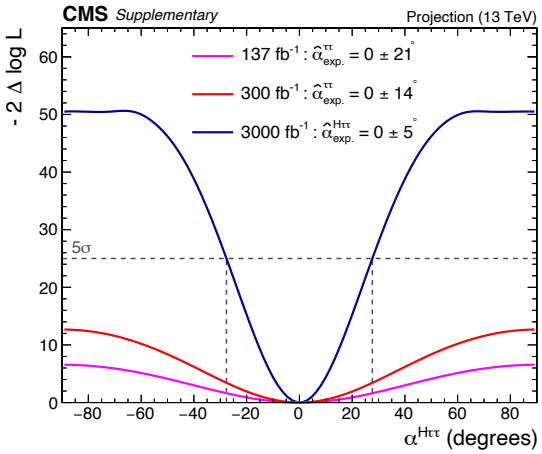


Figure 2.24: Expected negative log-likelihood scan for the combination of the $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels, including predictions for LHC Run-3 and Phase-2 integrated luminosities. The extrapolations have been made assuming systematic uncertainties as derived for the main analysis, while the statistical uncertainties were adjusted to different luminosity scenarios. The expected sensitivity to distinguish between the scalar and pseudo-scalar hypotheses, defined at $\alpha_{H\tau\tau} = 0$ and $\pm 90°$, respectively, exceeds 5 standard deviations for the Phase-2 scenario [17].

IV.1.b. HL-LHC Upgrade Project

The HL-LHC upgrade will start during the LS3 from 2026 to 2028. In order to reduce the beam size and reach targeted instantaneous luminosity, several enhancements to the accelerator are planned [114]. One of the major improvements is the replacement of the inner triplet quadrupole magnets in the insertion regions with more radiation-tolerant ones based on Nb_3Sn technology. These new quadrupole magnets will provide a higher magnetic field strength, up to 12 T, and feature a larger aperture. This larger aperture accommodates higher beam intensities and tighter focusing optics, enabling a reduced beam beta function (β^*) at the interaction point, which directly increases the luminosity.

As a consequence of a lower β^*, the crossing angle of the beam will increase and the geometrical luminosity reduction factor will be smaller. To counteract this, superconducting radiofrequency crab cavities will be installed as they will rotate the bunches before collisions, ensuring a head-on interaction. As a result, it will increase the overlap of the colliding bunches and improve the luminosity.

The increase in instantaneous luminosity will enable CMS to collect a total of up to 3000 fb^{-1} over the lifetime of the HL-LHC. If the ultimate instantaneous luminosity scenario of 7.5×10^{34} cm^{-2}s^{-1} is achieved, the total integrated luminosity could extend to 4000 fb^{-1}. However, the increase in luminosity will also result

in a significant increase of PU events. While the maximum PU in 2016 reached 53 events, the HL-LHC is expected to experience an average of 140 PU events at 5.0×10^{34} cm^{-2}s^{-1}, potentially increasing to 200 events under the ultimate luminosity scenario. To mitigate the effects of high PU and to maintain stable detector and data-taking conditions, luminosity leveling will be employed. This technique regulates the instantaneous luminosity during each LHC fill to keep it approximately constant, preventing large fluctuations in PU and ensuring optimal detector performance.

In preparation for the HL-LHC, several key infrastructure upgrades have already been completed. During Long Shutdown 2 (LS2), a new cryogenic system and more powerful bending magnets were installed to support higher beam intensities. Additionally, the LHC Injectors Upgrade (LIU) project, aimed at enhancing the injector complex performance, was successfully implemented. These improvements lay the groundwork for the HL-LHC era, with further upgrades planned during LS3.

IV.2. CMS Phase-2 Upgrade

The CMS detector needs to be substantially upgraded during LS3 in order to fully exploit the increase in luminosity provided by the HL-LHC. This upgrade is referred to as the CMS Phase-2 Upgrade. The increase in radiation levels requires improved radiation hardness, while the larger PU and associated increase in particle density demand higher detector granularity to reduce occupancy, increased bandwidth to accommodate higher data rates, and improved trigger capability to keep the trigger rate at an acceptable level while not compromising the physics potential.

The entire silicon tracking system, presently consisting of pixel and strip detectors, will be replaced. The new tracker will feature increased forward acceptance, increased radiation hardness, higher granularity, and compatibility with higher data rates and a longer trigger latency. Additionally, the tracker will provide tracking information (for tracks above a configurable transverse momentum threshold) to the L1 trigger, which is currently only available at the HLT. This will allow the trigger rates to be kept at a sustainable level without sacrificing the physics potential. A detailed discussion of the Silicon Tracker upgrade will be presented in the next section.

The MIP Timing Detector (MTD) [143] will be added between the Tracker and ECAL to enhance the timing resolution for MIP to 30 ps. The barrel section of the detector will use Silicon Photomultipliers and scintillating crystals, while Low Gain Avalanche Detectors will be used in the end-cap region.

In the barrel, the electromagnetic calorimeter uses lead-tungstate crystals read out with avalanche photodiodes. The crystals will be cooled to lower temperatures

than those currently used, and the front-end electronics will be improved to cope with the trigger latency and bandwidth requirements. The new front-end boards will allow the exploitation of information from single crystals in the L1 trigger, whereas the present system integrates the same information only in groups of 5×5 crystals. The hadronic calorimeter in the barrel region will be replaced with silicon photomultipliers before LS3. During LS3, the scintillator tiles close to the beamline will also be replaced.

The electromagnetic and hadronic endcap calorimeters will be replaced with a new combined electromagnetic and hadronic sampling calorimeter based primarily on silicon pad sensors. Plastic scintillator tiles, read out by SiPMs, will be used at large distances from the beamline in the hadronic section. This detector will provide high transverse and longitudinal granularity, leading to improved PU rejection and identification of electrons, photons, τ and jets.

Although the muon chambers are expected to cope with the increased particle rates, the front-end electronics for the drift tube chambers and cathode strip chambers will be replaced to increase radiation tolerance, readout speed, and performance. In the forward region, the muon system will be enhanced with improved resistive plate chambers and new chambers based on the gas electron multiplier (GEM) technique. The new chambers will improve triggering, reconstruction performance and increased acceptance in the forward detector region.

CMS trigger rates will be limited to 750 kHz at L1 and 7.5 kHz at HLT. The L1 trigger latency will be about 12.5 µs. The CMS DAQ and control system will be enhanced with extended bandwidth capabilities. Additionally, various improvements and upgrades will be made to data reconstruction, storage, analysis, access algorithms, luminosity measurement systems, and beam radiation protection.

IV.3. Phase-2 Upgrade of the CMS Tracker

The Phase-2 tracker [144] will consist of an Inner Tracker (IT) based on silicon pixel modules and an Outer Tracker (OT) made of silicon modules with strip and macro-pixel sensors. The Outer Tracker will consist of p_T-modules capable of rejecting low-p_T particles at the module level. These modules come in two varieties: *2S modules*, consisting of two strip sensors, and *PS modules*, which pair a strip sensor with a macro-pixel sensor. It will be composed of pixel modules, a detailed description of the IT is given in the next section. The key requirements for both trackers' upgrades can be summarized as follows:

- **Radiation tolerance:** The upgraded tracker must remain fully operational up to an integrated luminosity of 3000 fb^{-1}, with a safety margin to ensure longevity. The Inner Tracker, in particular, must withstand fluences exceeding 10^{16} $\mathrm{n_{eq}/cm^2}$ and doses above 10 MGy. This necessitates the use

of radiation-hard silicon sensors and advanced readout electronics capable of sustaining high doses without performance degradation. A map of the estimated fluence is shown on Figure 2.25, with high dependence on the fluence with the radial distance the beam r. The maximum fluence levels expected for various parts of the new tracker are summarized in Table 2.2.

Region or component	**Max. fluence [n_{eq}/cm^2]**	**r [mm]**	**z [mm]**
IT barrel layer 1	2.3×10^{16}	28	0
IT barrel layer 2	5.0×10^{15}	69	0
IT barrel layer 4	1.5×10^{15}	156	89
IT forward, ring 1	1.0×10^{16}	51	252
IT service cylinder	1.3×10^{15}	170	260
OT PS modules	9.6×10^{14}	218	129
OT 2S modules	3.0×10^{14}	676	2644

Table 2.2: Maximum fluence values for different regions/components along with corresponding r and z values [144].

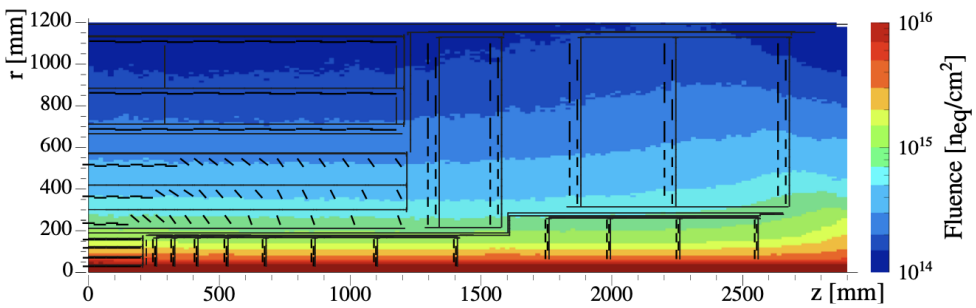


Figure 2.25: Integrated particle fluence in 1 MeV neutron equivalent per cm^2, for the Phase-2 tracker. The estimates shown correspond to a total integrated luminosity of 3000 fb^{-1} of pp collisions at $\sqrt{s} = 14$ TeV, and have been obtained with the CMS FLUKA geometry version [145].

- **Increased granularity:** Due to the high track density expected in the HL-LHC environment, finer segmentation is essential. A higher channel density will allow better spatial resolution, reducing the probability of hit ambiguities and improving the precision of track reconstruction. The pixel detector in

the Inner Tracker will have small pixel sizes (of the order of $25 \times 100\ \mu\text{m}^2$ or smaller) to optimize the tracking efficiency. Additionally, a new clustering algorithm will be used to improve two-track separation.

- **Low material budget:** Minimizing the amount of material in the tracker reduces multiple scattering and Bremsstrahlung, improving both momentum resolution and the overall detector performance. This requires lighter support structures, low-mass electrical and cooling systems, and the use of high thermal conductivity materials such as carbon-based composites to improve heat dissipation without adding excess material.

- **Robust pattern recognition:** High PU conditions demand advanced pattern recognition algorithms that can efficiently handle large numbers of tracks. The new tracker design will facilitate an improved hit association and track reconstruction, leveraging both hardware-based and software-based approaches, such as machine learning techniques and fast tracking algorithms.

- **Contribution to Level-1 trigger:** The upgraded tracker must provide fast tracking information to the L1 trigger system to improve event selection. The Outer Tracker will include modules capable of identifying track stubs, which can be used to reconstruct high-momentum tracks in real time, significantly enhancing the trigger's ability to reject background events while preserving physics signals of interest.

- **Extended tracking acceptance:** The upgraded tracker will extend the tracking coverage up to $|\eta| = 4.0$, providing better tracking performance in the forward region. This is particularly important for physics processes involving forward jets, vector boson fusion, and heavy flavor tagging in boosted topologies.

These requirements guide the technological choices for sensor development, readout electronics, mechanical design, and integration into the CMS detector. A detailed description of the IT is given in the next section.

IV.4. Phase-2 Upgrade of the CMS Inner Tracker

The Phase-2 Inner Tracker (IT) is a key component of the CMS detector upgrade, designed to maintain and improve the tracking and vertexing capabilities under the high PU conditions of the HL-LHC. The IT will be subjected to an ionizing radiation dose of up to 1.2 Grad and a hadron fluence of up to 2.3×10^{16} neq/cm^2 for 3000 fb^{-1}. The upgrade is necessary to handle the expected PU of up to 140 (or even 200) collisions per bunch crossing, with hit rates approaching 3 GHz/cm^2

in the inner layers for a PU of 200. A detailed understanding of the IT design is essential in the context of this thesis, as a dedicated chapter is devoted to the irradiation studies of its modules (see Chapter 3). This section provides an overview of the key characteristics of the IT; for further technical details, the reader is referred to [145].

IV.4.a. Overview

The IT will extend from -265 cm to +265 cm in z direction and occupy a region with radius < 260 mm. It will be composed of several parts, as shown in its layout in Figures 2.26 and 2.27. The closest part to the IP is the Tracker Barrel Pixel Detector (TBPX), followed by the Forward (TFPX) and Endcaps (TEPX) Pixel Detectors. The modules (detailed in Section IV.4.b.) in the barrel are arranged in *ladders* with staggered modules to overlap in $r-\phi$. The TBPX consists of 4 layers with 4 or 5 modules per ladder. In TFPX and TEPX the modules are arranged in half discs referred to as *Dees*. The TFPX is made of 8 small double-discs with 4 rings of modules per disc and the TEPX made of 4 large double-discs with 5 rings of modules per disc. The design also supports a modular structure that allows easy installation and potential replacement of degraded components during an Extended Technical Stop. A serial powering, support structures and cooling pipes shared among modules will minimize the material budget. The cooling will be provided by a CO_2-based system, to keep the sensors below $-20°C$.

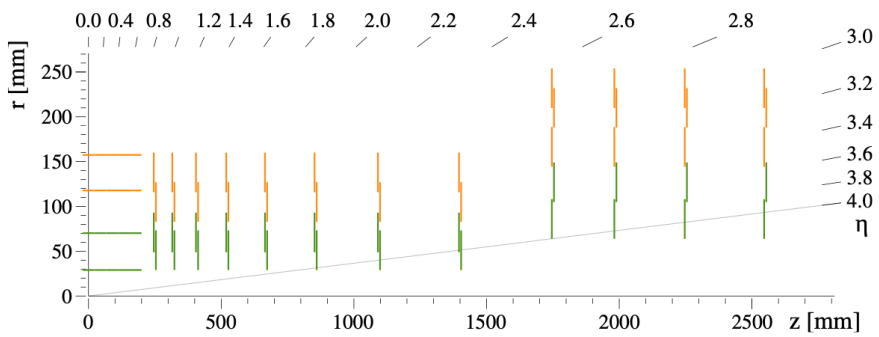


Figure 2.26: Sketch of the Phase-2 Inner Tracker layout, showing the barrel and endcap detector regions. The modules are arranged in layers and rings, and the pixel sensors are illustrated as colored elements [145].

IV.4.b. The Silicon Pixel Module

The IT modules are hybrid structures, where the sensors are bump-bonded to either two or four *Read-Out Chips* (ROCs). These ROCs are wire-bonded to

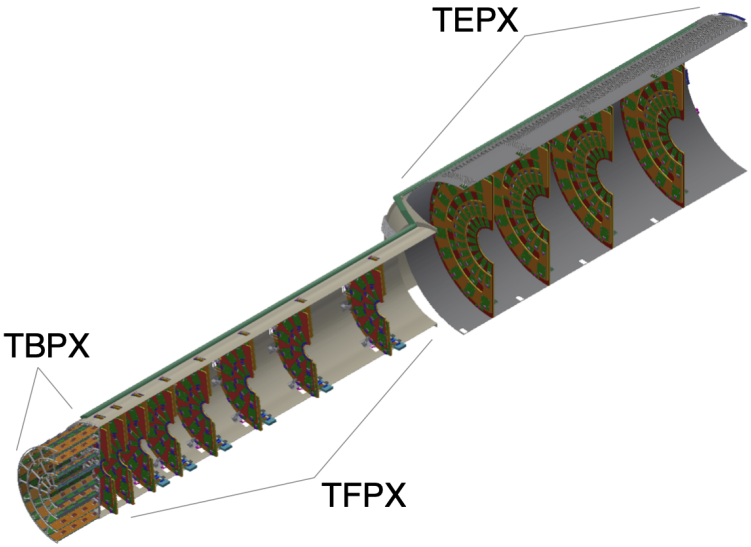


Figure 2.27: A perspective view of one quarter of the Inner Tracker, illustrating the TBPX ladders along with the TFPX and TEPX dees within the supporting structures. The pixel modules appear as orange elements in the TBPX and green elements in the TFPX and TBPX, while the dees are represented by red and orange surfaces [145].

a printed circuit board known as the *High-Density Interconnect* (HDI), which distributes low voltage to power the ROCs, high voltage to bias the sensor, provides clock and control signals, and routes the data output from the chips. The modules are connected electrically via high-speed links, known as e-links, to optoelectronic interface boards called *Portcards*. A detailed description on ROCs is given in Section IV.4.c.

The design of the modules is based on previous iterations used in the Phase-0 [115] and Phase-1 [120] pixel detectors. The two types of modules that will be used in the IT are shown in Figure 2.28. The modules differ in their sensor sizes and the number of readout chips they contain. The first type, known as the 1×2 module, incorporates two readout chips, and 50×50 μm^2 pixels. The second, known as the 2×2 module, consists of four chips arranged in a 2×2 layout and of $25\times$ 100 μm^2 pixels.

A summary of the number of modules, sensors, readout chips, and pixels in each Inner Tracker substructure, along with the corresponding active silicon area, is provided in Table 2.3.

As regard to the sensors, a planar n-in-p sensor technology is chosen as a double-sided processing is not needed, making them easier to produce and cheaper. For the inner layers, an alternative option using 3D sensors is also considered. They

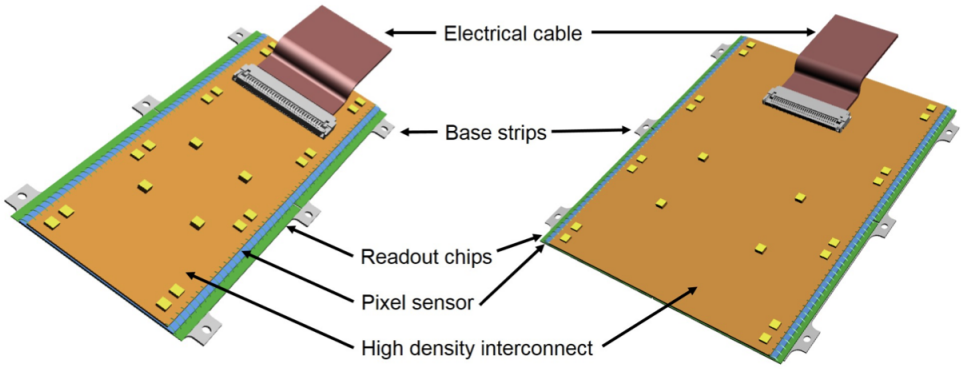


Figure 2.28: Illustration of the 1×2 (left) and 2×2 (right) pixel modules, with approximate dimensions of 1.8 × 4.4 cm^2 and 3.7 × 4.4 cm^2, respectively. Yellow elements indicate passive electrical components [145].

Table 2.3: Summary of the number of modules, sensors, readout chips, and pixels in each Inner Tracker substructure, along with the total active silicon area [145].

	TBPX	**TFPX**	**TEPX**	**Total**
Modules (type)	360 (1×2)	832 (1×2)	768 (1×2)	1960 (1×2)
	504 (2×2)	896 (2×2)	992 (2×2)	2392 (2×2)
Sensors	864	1728	1760	4352
Pixel chips	2736	5248	5504	13488
Pixels [×10^6]	395	757	794	1947
Silicon area [m^2]	0.99	1.89	1.99	4.87

have a better radiation resistance and require lower bias voltage, but with a larger pixel capacitance, are more expensive and more complex to produce. Both planar and 3D sensors have been studied in test beams [146] confirming their excellent performances.

IV.4.c. Pixel Module Readout Chip

The IT ROCs are critical components of the IT as they must ensure a high-speed data acquisition, tolerate a 1.2 Grad of total ionizing dose and a maximum hit rate of 3.2 GHz/cm^2. In the context of the Phase-2 upgrade, a custom ROC has been developed in a joint effort between ATLAS [147, 148] and CMS [149–151],

called the *RD53* collaboration. The latest generation of ROCs, called *RD53B*, differs between ATLAS and CMS as it uses different *Analog Front-Ends* (AFE). The RD53B [26] of CMS, also known as CMS Readout Chip version 1 (CROC-V1 or CROC) is using a Linear AFE. Its characteristics are summarized in Table 2.4.

Table 2.4: Design requirements and specifications of the CROC-V1 chip [26].

Parameter	**Value**
Maximum hit rate	3 GHz/cm^2
Trigger rate	750 kHz
Trigger latency	12.5 μs
Bump bonding pitch	50×50 μm^2
Pixel array	432×336 pixels
Chip dimensions	21.6×18.6 mm^2
Minimum threshold	600 e^-
Radiation tolerance	1 Grad over 10 years
Power delivery	Serial Powering
Power consumption	< 1 W/cm^2
Clock & Data Recovery	Clock and Command/Trigger stream: 160 MHz
	Internal clocks: 160 MHz, 64 MHz, 640 MHz, 1.28 GHz

Architecture and Data Flow: The CROC-V1 consists of two main components, the *pixel matrix* and the *chip bottom*. The pixel matrix is made up of pixels, each containing an AFE that processes the signal of the sensor. In the CROC, the pixels are grouped into different substructures, such as the *core*, shown in Figure 2.29. The concept behind these core regions is to segment the pixel matrix into manageable units that can efficiently share digital and analog resources. The chip bottom (or periphery) is the region beneath the pixel matrix that contains the digital logic and control circuits needed to bias, configure, monitor, and read out the chip. The chip bottom is divided in the Analog Chip Bottom (ACB) containing the analog circuitry and the Digital Chip Bottom (DCB) which implements the input, output, global control/monitoring and readout.

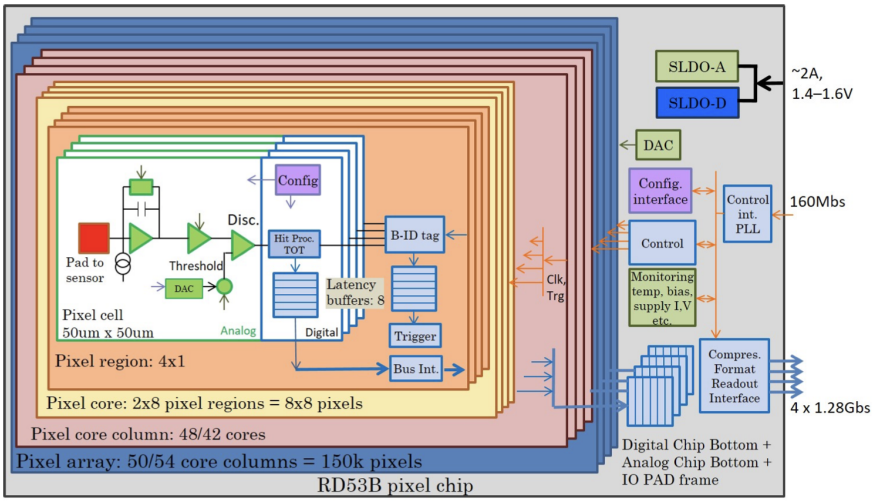


Figure 2.29: Data flow architecture of the RD53B readout chip [26].

Analogue Front-End and Hit Capture: The AFE is responsible for processing the initial analog signal generated by the interaction of a particle with the sensor. A simplified scheme of the CROC-V1 AFE is shown in Figure 2.30. First, the charge is induced in the pixel sensors (represented as reverse biased diodes) and transferred to the corresponding pixel chip via the bond bounding. This signal is then integrated by a pre–amplifier stage (PA) which converts the charge into a voltage signal via a charge sensitive amplifier (CSA) with *Krummenacher feedback*. The Krummenacher feedback generates a controlled current, called the Krummenacher current I_{Krum} providing a constant, linear discharge of the CSA capacitance. As shown on Figure 2.31 the signal is characterized by a fast-rising edge and by a slowly decreasing falling edge which depend on I_{Krum}. Once the signal is amplified, it is turned into a digital hit by a comparator, which decides if the signal is above a certain threshold V_{th}. This threshold is determined by the *Global threshold*, but each pixel can fine-tune it with a small 4-bit adjustment to correct for small differences between pixels, this is the *local threshold*.

In the CROC, the threshold V_{th} is commonly set around 1000 electrons. This ensures a high detection efficiency for a MIP with a charge of the order of 10000 electrons and a noise of 1000 electrons.

In order to measure the charge Q induced in a single pixel, the so-called *time-over-threshold* (ToT) technique is used. The ToT corresponds to the duration during which the signal remains above a predefined threshold. It is determined by counting the number of 25 ns clock cycles for which the output of the comparator remains high. The ToT value also depends on the Krummenacher current I_{Krum}, which controls the discharge rate of the preamplifier. As illustrated in Figure 2.31,

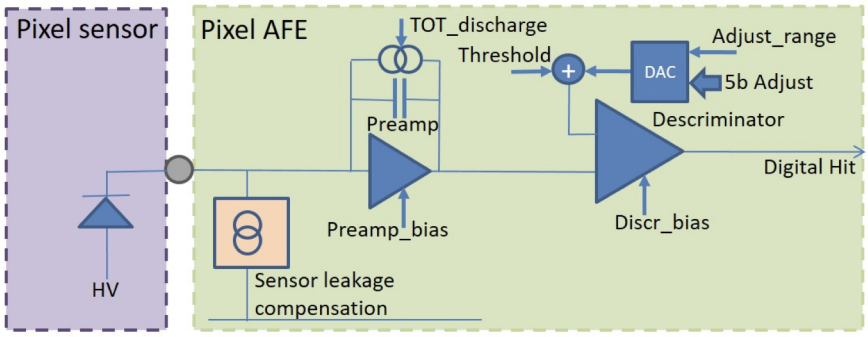


Figure 2.30: Block diagram of a generic Analog Front-End (AFE) used in pixel detectors. The diagram shows the pixel sensor, which converts the energy deposited by a particle into charge. This charge is then processed by the AFE, which includes a preamplifier for signal integration, a shaping stage to prepare the signal for digitization, a programmable current source (Krummernacher current) for signal discharge, and a discriminator to generate a digital hit signal when the input exceeds the threshold [26].

increasing I_{Krum} leads to a steeper return of the signal to the baseline, thereby reducing the measured ToT.

Front-End Registers: In the AFE, the *Digital to Analog Converter* (DAC) converts the digital settings (such as register values) into precise analog voltages or currents. One of the most important settings is the comparator threshold V_{Thr}, which determines the minimum signal level to register a hit. It is defined as:

$$V_{Thr} = V_{GDAC} - V_{GLDAC} \times TDAC \tag{2.15}$$

where V_{GDAC} is the Global Threshold, V_{GLDAC} the Local Threshold of the pixels and TDAC defined later in the `LDAC_LIN` register. The main registers are:

- **GDAC LIN**: It configures the Global Threshold of the AFE. The threshold is generated using a specific unit called a δV_{cal} unit. With a good approximation, one δV_{cal} corresponds to the injected charge $Q = 5e^-$.

- **LDAC_LIN:** The value of V_{LDAC} is set using the global register `LDAC_LIN`. Although V_{GDAC} is applied uniformly across all channels, variations in local ground reference levels between channels can introduce threshold dispersion. To mitigate this, an additional correction term $V_{\text{LDAC}} \times \text{TDAC}$ is included in Eq. 2.15. Here, `TDAC` is a 6-bit value (ranging from 0 to 32) specific to each channel and determined during a calibration procedure.

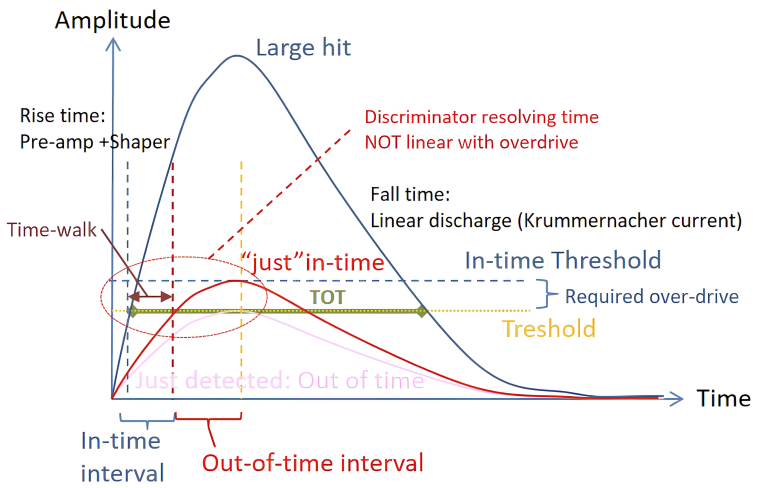


Figure 2.31: Illustration of charge measurement using the ToT method and the Krummernacher current scheme. The ToT method translates the amount of deposited charge into a time pulse width, providing an indirect but precise measure of the energy deposited by the particle. The Krummernacher current defines the discharge slope of the pulse, ensuring a consistent and linear correlation between the charge and time duration [26].

- **VCAL_MED and VCAL_HIGH:** These registers define the voltages used for charge injection during the calibration of the AFE. Both `VCAL_MED` and `VCAL_HIGH` are expressed in *Vcal units*, which correspond to the voltage seen across the feedback capacitance plates when they are charged with a known charge (typically equivalent to $\sim 5e^-$). The difference between the two, denoted as ΔVcal = VCAL_HIGH − VCAL_MED, determines the amplitude of the injected calibration pulse. It is important to emphasize that ΔVcal refers to the difference between two registered values, both in units of Vcal; it is not a physical unit itself.

The CROC-V1 calibration procedure is detailed in the Section III., where a study was conducted before and after the irradiation of a CROC-V1.

Chapter 3

Development and Validation of the Irradiation Setup for Phase-2 Pixel Modules at CYRCé

Content

I. Experimental Setup for Irradiation 78

I.1. The CYRCé Cyclotron at IPHC 78

I.2. Irradiation Line and Irradiation Cold Box 80

II. Prototyping a Cold Test Box 82

II.1. Motivation and Requirements 82

II.2. Design of the Cold Test Box 83

II.3. Test Protocol and Validation Campaign 84

II.4. Performance with Irradiated Modules 89

II.5. Summary and Outlook 90

III. Testing of CROC-V1 Before and After Irradiation . . 91

III.1. Test Setup and Readout Configuration 91

III.2. Pre-Irradiation Characterization 92

III.3. Post-Irradiation Measurements 96

IV. Conclusion 102

In the context of the Inner Tracker (IT) Phase-2 upgrade presented in Section IV.2., pixel modules are tested before and after irradiation to ensure that both the sensor and the readout chip can withstand the radiation levels expected at HL-LHC. Under the most optimistic scenario, the integrated luminosity is expected to reach 4000 fb^{-1}, corresponding to a maximum non-ionizing energy loss of $\phi_{\mathrm{eq}} = 3.5 \times 10^{16}$ $\mathrm{n_{eq}/cm^2}$ in the innermost regions of the IT. Several test beam campaigns have already been carried out to validate the design and performance of both planar and 3D sensors after irradiation [22, 23].

As the pixel module production phase has commenced, the IPHC CMS group joined the irradiation effort to contribute to the qualification of these new modules. This became possible thanks to the construction of a new irradiation line at the CYRCé accelerator, detailed in Section I.1. The study presented in this chapter focuses on the preparation and commissioning of the first irradiation of an IT module at IPHC.

The first part of the work consists of setting up a test setup to test the module before and after irradiation. This requires the development of a stable, dry and cold environment, suitable for the operation of irradiated modules, as well as the definition of a detailed and reproducible testing protocol to ensure reliable measurements.

The second part of the study presents the test results obtained before and after irradiation. One of the key motivations behind this work is to demonstrate the technical capacity and expertise required for IPHC to be recognized as an irradiation center within the CMS collaboration. The ultimate goal is for IPHC to serve as a reliable irradiation center, capable of receiving modules from various laboratories, performing standardized irradiation campaigns, and returning the modules for post-irradiation characterization.

My contribution to this project was focused on preparing and launching the first irradiation campaign, with the aim of establishing a robust infrastructure and workflow to support future module irradiation at IPHC.

I. Experimental Setup for Irradiation

I.1. The CYRCé Cyclotron at IPHC

The CYRCé facility (Cyclotron pour la Recherche et l'Enseignement) is a compact TR24 cyclotron located at IPHC in Strasbourg, shown in Figure 3.1. Operational since 2013, it delivers proton beams with energies up to 25 MeV and beam currents ranging from 100 attoamperes to 10 nanoamperes.

Originally developed for radiobiology, medical applications, and isotope production, with the capacity to produce over 20 different isotopes, CYRCé supports

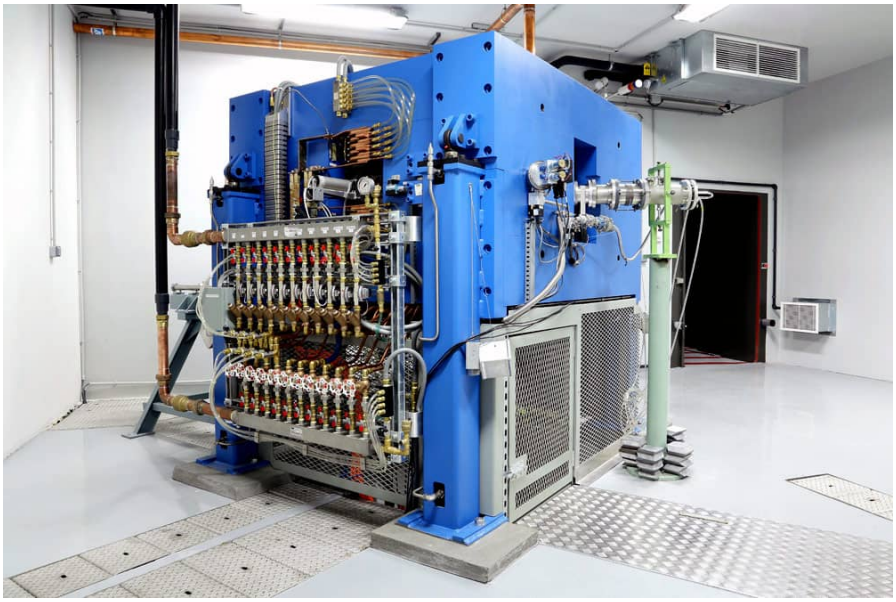


Figure 3.1: Pictures of CYRCé in the vault [152].

a wide range of research activities [152]. The PRECy project (Platform for Radiobiological Experiments from CYRCé) enables studies aimed at improving the understanding of relative biological effectiveness.

In parallel with PRECy, the facility has been extended to particle physics applications by developing a dedicated beamline specifically designed for the irradiation and qualification of silicon sensors, as illustrated in Figure 3.2. This beamline is primarily used for testing modules of the CMS Outer Tracker.

For this study, a new dedicated beamline was commissioned inside the cyclotron shielding casemate to overcome the dose limitations of the previous external irradiation setup. This infrastructure allows controlled irradiation with fluences up to 10^{16} n_{eq}/cm^2, corresponding to the integrated doses expected for tracker modules operated during HL-LHC conditions.

Given the limited availability of irradiation facilities in Europe and the high demand associated with the HL-LHC upgrade, CYRCé enables IPHC to actively participate in the irradiation campaign for the CMS tracker upgrade. The dedicated proton beam, with energies up to 25 MeV and adjustable flux, allows efficient irradiation to HL-LHC equivalent doses within reduced timescales. Furthermore, the beamline's location inside a radiation-shielded casemate ensures compliance with safety regulations for high-dose exposures. This facility complements existing irradiation and beam test sites, such as KIT [24] and DESY [25], thereby supporting ongoing irradiation campaigns.

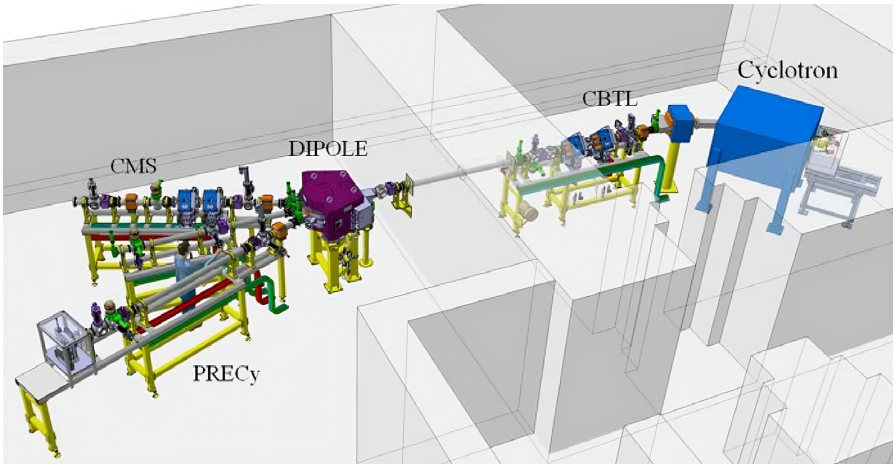


Figure 3.2: Layout of the CYRCé facility. The experimental area includes the PRECy platform for radiobiology studies and the CMS dedicated beamline for silicon sensor qualification. The new irradiation beamline, located inside the radiation-shielded vault, is designed for high-fluence irradiation of IT modules [152].

I.2. Irradiation Line and Irradiation Cold Box

The new irradiation line for this study was commissioned in 2024 at CYRCé to irradiate silicon modules up to fluences of 10^{16} n_{eq}/cm^2. The beamline is installed inside the shielding casemate, offering a dedicated and safe environment for high-dose irradiation. A picture of the beamline is shown in Figure 3.3.

The line consists of a series of beamline elements configured to transport and focus the 25 MeV proton beam onto the device under test. A dipole magnet is used to bend the beam from the cyclotron extraction point onto the irradiation axis. Two steerers provide fine control over the beam position in both transverse directions. The beam is then focused by two quadrupole magnets, allowing adjustment of the beam spot size at the irradiation plane. Beam monitoring is ensured through the use of two Faraday cups, installed upstream and downstream of the target position, and a retractable ionization chamber.

At the end of the line, the irradiation box, also referred to as the cold box, houses the module. The box is designed to maintain stable and controlled environmental conditions throughout the irradiation process. It includes an air-cooling system to remove heat from the irradiated modules, along with a dry-air inlet and desiccant granulates to suppress humidity. Internal temperature and humidity sensors allow continuous environmental monitoring. The device position can be adjusted using a motorized XY translation stage, enabling uniform irradiation. A webcam provides real-time visual feedback for remote supervision. Radiation pro-

Figure 3.3: Photograph of the new proton irradiation line installed in the CYRCé vault. The setup includes beamline elements such as the dipole, steerers, quadrupoles, and the cold box used for silicon module.

tection is also integrated into the design: the air inside the box is isolated during activation and released only after sufficient decay of short-lived isotopes.

The commissioning of the new line began in mid-2024. Initial beam tests were performed with a temporary setup, including limited power availability for the quadrupoles and a single operational Faraday cup. Full operational capability is expected to be reached for the irradiation campaigns planned in 2025.

As part of the first irradiation campaign, a RD53B CROC-V1 module (see Section IV.4.c.) equipped with a 3D sensor was irradiated under homogeneous conditions. The irradiation lasted approximately 1.5 hours, during which the beam was scanned across the module in 1 mm steps. A total of around 60 scans was performed, delivering an integrated fluence of 5×10^{15} n_{eq}/cm^2.

The fluence was determined using two independent methods. First, the beam current was measured with a Faraday cup, and the equivalent fluence was calculated using the appropriate hardness factor [153]. Second, a nickel activation foil was placed just behind the module to provide an independent measurement of the proton fluence based on gamma spectrometry. When protons interact with the nickel atoms in the foil, they produce specific radioactive isotopes whose decay can be measured. By identifying these isotopes and their relative abundances, in

particular, the ratio between Ni-57 and Co-55, it is possible to infer the energy of the incoming protons. The results showed that the protons had lost energy while passing through the sensor and the metallic support layer behind it, arriving at the foil with an estimated energy of about 16 MeV. Using known cross sections for this energy, the fluence measured from the activation analysis was calculated. The result differed by only about 15% from the value obtained using the Faraday cup and hardness factor method [152], which is within the expected uncertainty of the hardness factor itself (typically around 20%). This good agreement confirms the reliability of both techniques. Moreover, the nickel foil activation method used here is expected to achieve an accuracy of approximately 10% [154], offering a complementary and independent cross-check of the delivered fluence.

The homogeneity of the irradiation was assessed using Gafchromic films [27]. These films were exposed under the same scanning conditions and analyzed to produce a dose map across the module area. A dose gradient (or "nuancier") was also created by modulating the beam exposure across different regions of the film, enabling calibration and visualization of the dose uniformity.

II. Prototyping a Cold Test Box

After irradiation, modules are transferred from the beamline irradiation box to a separate enclosure referred to as the *cold test box*, dedicated to calibration and testing of the CROC. This second box ensures the availability of the irradiation infrastructure for other users while providing a controlled environment for systematic post-irradiation studies. This section presents the specifications and the design of the cold test box, as well as the validation tests performed to ensure its suitability for post-irradiation module studies.

II.1. Motivation and Requirements

Pre- and post-irradiation tests of the modules are part of the qualification process. After irradiation, the modules exhibit increased leakage current and power dissipation, making temperature control important. If the thermal environment is unstable, the module can reach thermal runaway, where the increase in current leads to further heating and a feedback loop. In parallel, annealing effects during the time between irradiation and testing can change the radiation-induced damage and affect the measurement. Thermal runaway (see Section III.3.) needs to be avoided to keep the module in a testable state, while annealing should be kept under control or monitored.

To address these points, tests are performed at low temperatures close to the CMS operating conditions, where the detector is cooled with CO_2 at approximately

−35 °C. The setup should provide a stable thermal environment, especially when the module is powered.

Humidity control is also needed. The modules are not sealed, and condensation can damage the electronics. To prevent this, the test box is ventilated with dry air.

The *dew point* is defined as the temperature at which the air becomes saturated with water vapor, such that any further cooling or increase in humidity causes condensation or even frost formation. It depends on both the ambient temperature and the relative humidity.

In this setup, the dew point is estimated from the measured temperature and relative humidity using the empirical formula described in [28]:

$$f(T) = \ln\left(\frac{\text{humidity}}{100}\right) + \left(23.036 - \frac{T}{333.7}\right) \cdot \frac{T}{279.82 + T}, \tag{3.1}$$

$$\text{Dew Point} = \frac{279.82 \cdot f(T)}{23.036 - f(T)}, \tag{3.2}$$

where T is the air temperature in degrees Celsius, and humidity is the relative humidity in percentage. This allows to control that the air temperature stays below the dew point. However, it is important to note that even in the absence of condensation, high relative humidity can lead to electrical breakdown in irradiated modules. In particular, values of relative humidity exceeding 40% have been observed to induce early breakdown behavior [29]. Maintaining low humidity is therefore essential, not only to avoid condensation but also to preserve the electrical integrity of the irradiated detectors.

Since the irradiation cold box needs to be available for other uses, the modules are moved after irradiation into a separate cold test box. The transfer time is kept as short as possible to limit uncontrolled annealing. The box is designed to cool the module after transfer and to manage temperature changes during removal.

II.2. Design of the Cold Test Box

The cold test box is located in a supervised area near the CYRCé facility. As the modules may be slightly activated after irradiation, this location avoids additional safety constraints while allowing easy access for testing.

The setup is kept simple, aiming to verify basic module functionality under controlled temperature and humidity. It consists of a modified commercial freezer, shown in Figure 3.4, capable of reaching temperatures down to approximately −35 °C. The freezer includes a "super" mode to support fast cooling. Small internal fans are added to enhance air circulation and distribute the thermal load from powered modules.

To allow the passage of cables and dry air lines, an opening was made in the side wall of the freezer and sealed with insulating foam. A flow-meter, visible in Figure 3.4, regulates the injection of dry air to maintain low humidity and avoid condensation. Multiple temperature and humidity sensors are placed inside the box, as seen in Figure 3.4, and are read out using a Raspberry Pi. This system provides continuous monitoring and data logging.

To simulate the thermal load of an operating module, resistive heaters and heat plates can be added. This helps evaluate the cooling performance and temperature stability under realistic conditions.

The test setup for the electronic readout of the module includes ESD-safe materials such as antistatic carpets and properly grounded cables. The module is read out using an FPGA board with active cooling, connected to a dedicated PC used to control the setup and collect data during the tests.

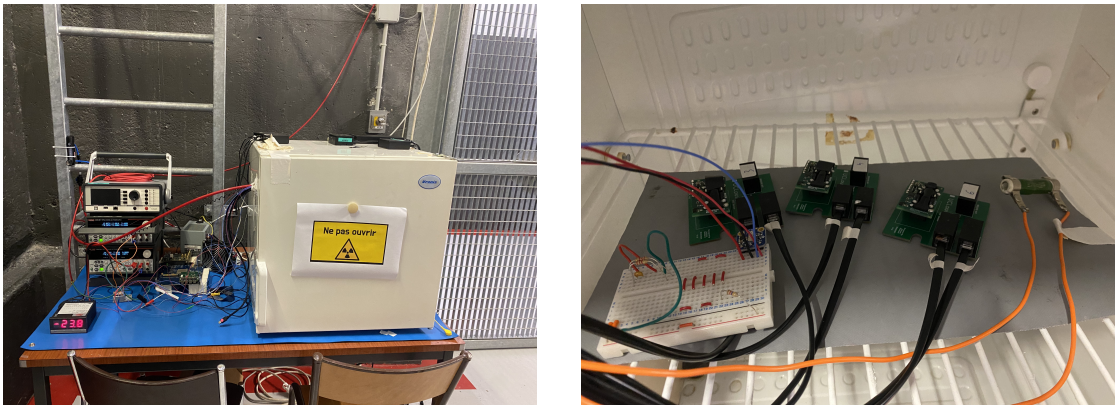


Figure 3.4: (Left) Full view of the cold box setup. The flow meter is visible with a tube supplying air inside the freezer. The power supply for the module and the sensors is also visible. (Right) Interior of the cold box, including a set of temperature and relative humidity (HR) sensors, along with a resistor.

II.3. Test Protocol and Validation Campaign

Following the requirements described earlier, a series of tests was performed to evaluate the cold box's performance under different conditions and to define a reliable procedure for cooling modules after irradiation. The main goals were to check the cooling system's stability, guaranty low humidity inside the box, and confirm that the setup is suitable for operating irradiated modules safely.

Common Procedure: All tests used the same basic procedure to study the system's temperature behavior and humidity control. First, dry air was injected

into the cold box at a high flow rate (usually 50 L/min unless stated otherwise) to lower the relative humidity. When the humidity reached a stable low level (generally below 10%), the freezer was switched on to start cooling. Once cooling began, the dry air flow was adjusted according to the specific requirements of each test. The airflow rate was chosen to balance maintaining low relative humidity while avoiding excessive cooling delays, since the dry air itself was not pre-cooled. During warming, the freezer was turned off and the dry air flow increased back to 50 L/min to keep the dew point below the air temperature and prevent condensation.

Normal Mode Tests: The first series of tests was conducted with the cold box operating in "normal" freezer mode. To estimate the uncertainty of the measurement system, three sensors were positioned at different locations inside the cold box. As shown in Figure 3.5, temperature readings agreed within ±1°C, while dew point values varied within ±4°C. These variations are attributed to the sensors' intrinsic characteristics and their spatial placement within the box. For all three sensors, the dew point consistently remained below the air temperature, indicating that condensation on the module is unlikely under these conditions. However, the minimum temperature achieved in this mode was around −20 °C, which is at the upper limit of the required range and could become insufficient when heat is dissipated by an operating module.

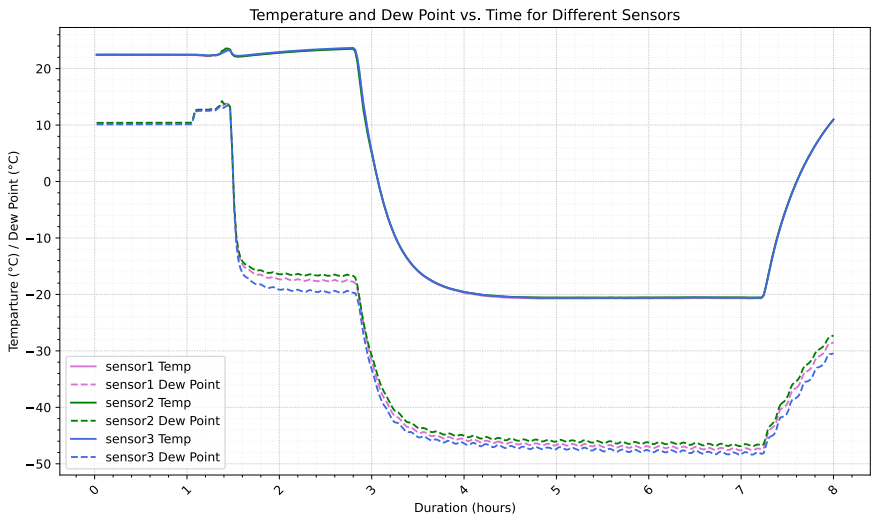


Figure 3.5: Temperature and dew point measurements from three sensors placed at different locations inside the cold box during operation in normal freezer mode.

In a subsequent test, two resistive heaters, dissipating 3.6 W and 9.3 W re-

spectively, were used to emulate the thermal load of the module during operation. As shown in Figure 3.6, the internal heating increased the air temperature by approximately 1°C maximum. Under this configuration, the system failed to reach temperatures below −20 °C, indicating that the "normal" cooling mode may not be suitable for operation with irradiated modules at full thermal load. Nevertheless, this mode remains useful for passive cooling phases where power dissipation is lower.

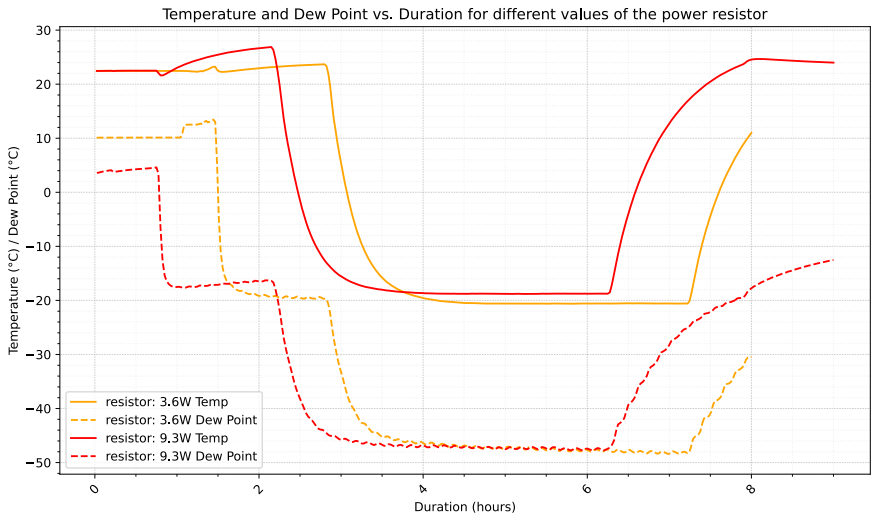


Figure 3.6: Temperature and dew point for two different resistor values.

"Super" Mode Tests: To improve thermal performance, the system was operated in "super" mode, which allows reaching lower temperatures. First, a test was conducted where the dry air was initially set to 50 L/min, reducing the relative humidity to approximately 1%. The flow was then decreased to 8 L/min while the freezer remained in "super" mode. The dew point remained safely below the air temperature, but the relative humidity increased, indicating that 8 L/min is near the lower limit for effective humidity control. Additionally, a brief increase in dew point was observed when dry airflow was reduced before the cooling cycle began, highlighting the importance of starting cooling prior to reducing airflow. In this mode, the system reached temperatures as low as −35 °C, which is appropriate for module operation (Figure 3.7).

In another test, a heating plate was used to simulate the thermal load corresponding to module power dissipation. This approach has the advantage of providing a flat and uniform surface, allowing temperature sensors to be directly attached, and more closely mimicking the thermal behavior of an actual module.

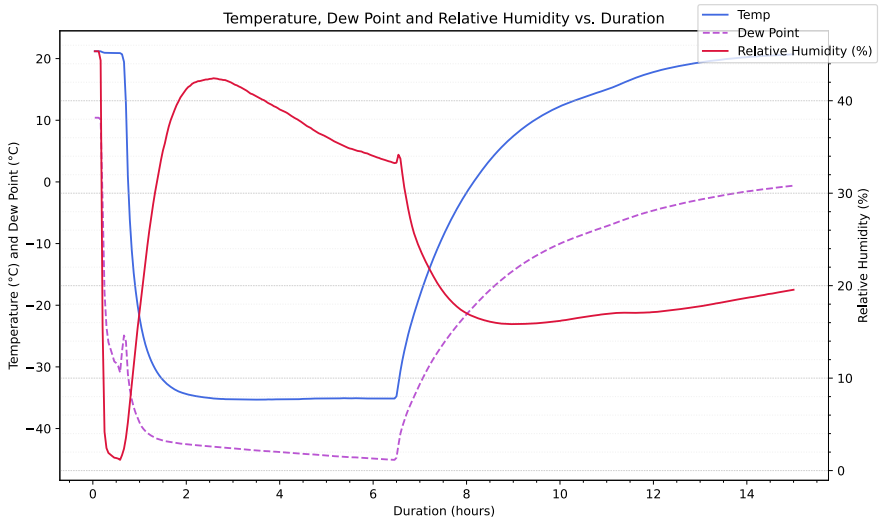


Figure 3.7: Temperature, relative humidity and dew point for "super" mode.

A temperature sensor was placed on the plate, although no automatic data logging was available for this measurement. With a dry air flow of 50 L/min and the freezer operating in “super” mode, the plate was initially powered at 7.5 W. When the heater was activated before reaching low temperatures, a transient increase in temperature was observed. After reducing the power to 3.2 W, the plate temperature stabilized at approximately −14.5 °C. While this test confirmed that the ambient air temperature could still be maintained below −30 °C (see Figure 3.8), it also demonstrated that the heating should only be activated after the cold box has reached its target low temperature. Activating the heater too early can compromise the module cooling.

In the final test, two small fans were added to improve heat dissipation. The freezer was operated in "super" mode, with dry air initially at 50 L/min, and the fans were switched on at the beginning of the run. The heating was activated only after reaching −30 °C. With this setup, the system stabilized at approximately −29 °C for a heating power of 7 W. The air remained cold and dry, with relative humidity below 10%. The test confirmed that this configuration is suitable for operating irradiated modules under stable conditions (Figure 3.9).

Conclusion: These tests demonstrate that the cold box provides a controlled environment in terms of temperature and humidity. The use of dry air, combined with the freezer and controlled heating, allows operation near the required conditions. The addition of fans helps dissipate the heat on the sensor (in this case, the heating plate), and the dew point remains safely below the module temperature

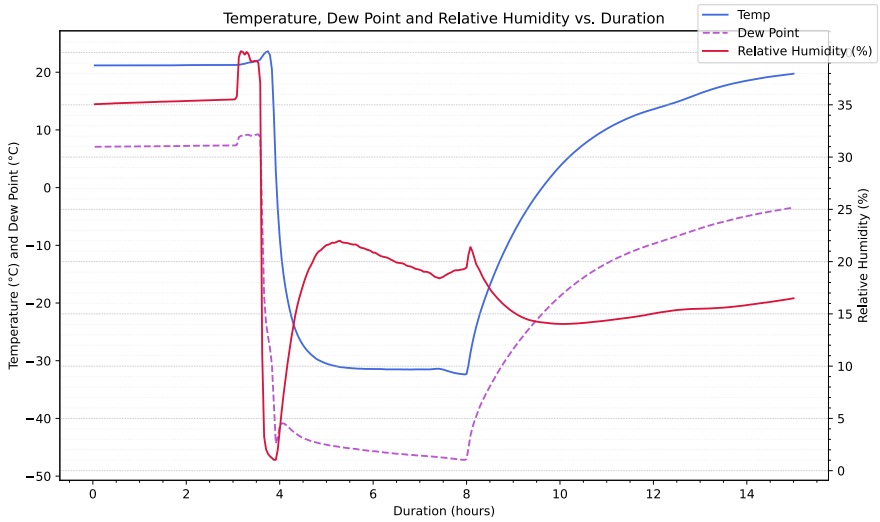


Figure 3.8: Temperature, relative humidity and dew point for "super" mode with heating plate.

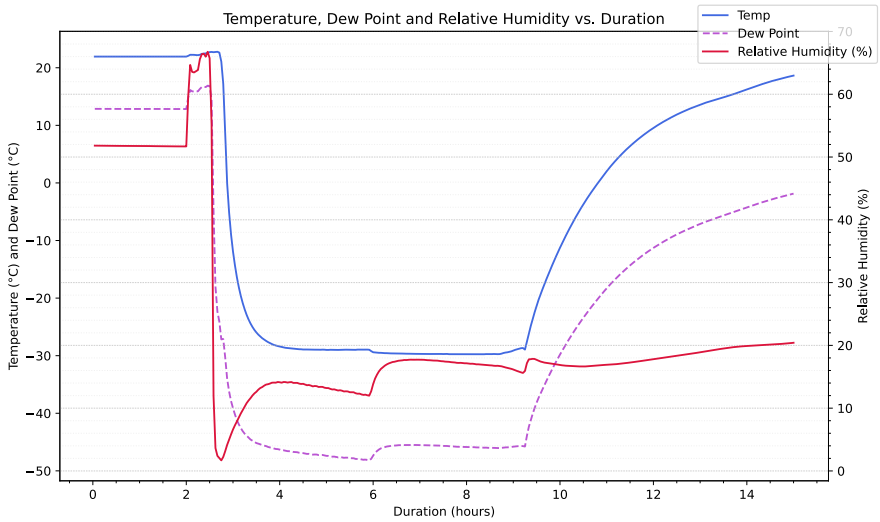


Figure 3.9: Temperature, relative humidity and dew point for "super" mode with heating plate and fans.

in all tested configurations.

II.4. Performance with Irradiated Modules

Following the irradiation of a CROC-V1 module, a first attempt to power and communicate with the chip was conducted. Due to time constraints, the module was placed in the cold box without removing the protective cap and without any fan for heat dissipation. The freezer was activated, and the chip was powered on before the system reached thermal equilibrium at −35 °C. As a result, after completing the first scan, the chip experienced thermal runaway, leading to a sudden drop in the supply voltage.

A second test was subsequently performed under improved thermal conditions. The protective cap was removed, and a small fan was added to enhance heat dissipation. As shown in Figure 3.10, the air temperature stabilized around −29 °C, with the dew point remaining well below the temperature, ensuring safe operation without condensation. Upon powering on the chip, a temperature increase of approximately +1°C was observed, but this remained within acceptable limits. Under these conditions, the module functioned correctly and scan calibrations were successfully carried out.

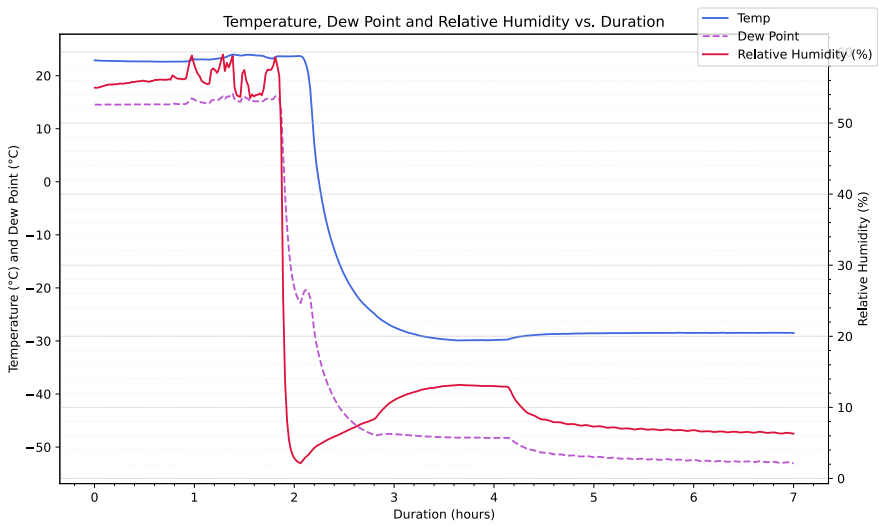


Figure 3.10: Temperature, relative humidity and dew point during operation of the irradiated CROC-V1 module in "super" mode with active cooling and a fan. The setup ensures stable conditions for safe operation.

A follow-up measurement was conducted after the module had remained in the cold box for several days under "super" mode conditions. As shown in Figure 3.11, the air temperature remained below −30 °C, and the relative humidity stabilized below 7%, confirming the long-term stability of the system for extended irradiation studies and the possibility to operate at lower temperature if needed.

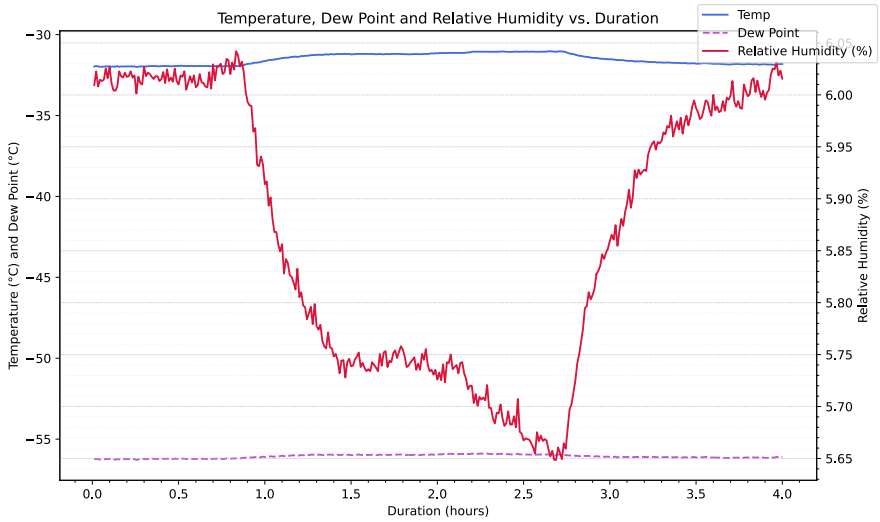


Figure 3.11: Long-term stability of temperature, relative humidity, and dew point in "super" mode with an irradiated module and fan, several days after initial installation.

II.5. Summary and Outlook

A dedicated test campaign was carried out to qualify the performance of the cold box system for the safe operation of irradiated pixel modules. The setup demonstrated reliable control over temperature and humidity, both of which are critical to preventing condensation and ensuring chip thermal stability. A standardized procedure was developed for cooling and warming cycles, supported by sensor cross-validation and continuous humidity monitoring.

Operation in "super" mode, combined with active dry air injection and localized fan cooling, allowed the system to reach and maintain temperatures around −30 °C, even under moderate power dissipation. This configuration was successfully validated using an irradiated CROC-V1 module. The system enabled stable operation, calibration scans, and extended storage with minimal risk of annealing or moisture-related degradation.

Some limitations remain, notably the presence of dust in the setup, which could be mitigated with improved environmental control. Since the initial tests, the system has been relocated and upgraded to improve safety and available workspace, enabling it to host new modules in a more permanent and controlled environment.

The current configuration is now considered operational and ready for upcoming irradiation campaigns. Given the successful performance with the CROC-V1 module, it is expected to be compatible with more complex assemblies, such as quad or dual-chip modules, provided that adequate thermal preconditioning is

applied.

III. Testing of CROC-V1 Before and After Irradiation

A single CROC-V1 chip bump-bonded to a 3D sensor was irradiated to a fluence of 5×10^{15} n_{eq}/cm^2 at the new irradiation facility at CYRCé. Functional tests were performed before and after irradiation to evaluate the performance of the chip and to validate the cold box setup described in the previous section.

This study is structured into three main phases. First, the full readout chain was commissioned and validated, including low-level communication, trimming procedures, and power-up behavior. Second, a comprehensive pre-irradiation characterization was carried out, covering pixel response, threshold tuning, and calibration scans. Finally, after irradiation, a dedicated set of measurements was performed to determine the sensor depletion voltage, investigate anomalies observed in initial post-irradiation scans, and calibrate the functional regions of the chip.

III.1. Test Setup and Readout Configuration

The CROC-V1 module, equipped with a 3D sensor, was tested using the standard Phase-2 pixel DAQ infrastructure. The test system operated with the `Ph2_ACF` software [30] providing the control and data acquisition functionalities. Communication with the module was ensured via a KSU-FMC board [31], which interfaces both the front-end (FE) and external clock/trigger sources.

The connection from the CROC-V1 to the FMC board was established through a DisplayPort cable. High voltage was supplied via a LEMO cable with adapter, while the low-voltage input, set to 1.8 V and 1.5 A (max), was provided using a dedicated power supply. The FC7 board, responsible for generating clock and trigger signals, was powered independently and connected to the FMC. The module was operated in LDO (Low Dropout) powering mode, with jumper configuration adapted accordingly, following the recommendations [26]. For thermal management, a small fan was positioned next to the FPGA to dissipate heat during continuous operation. All handling and measurement procedures were conducted under proper ESD (Electrostatic Discharge) precautions, including the use of an antistatic mat and grounding wrist strap.

Figure 3.12 shows a photograph of the complete test bench. Visible in the image are the CROC-V1 single-chip module mounted on its carrier card, the KSU-FMC board connected via a DisplayPort cable, and the FC7 FPGA board used for DAQ and control.

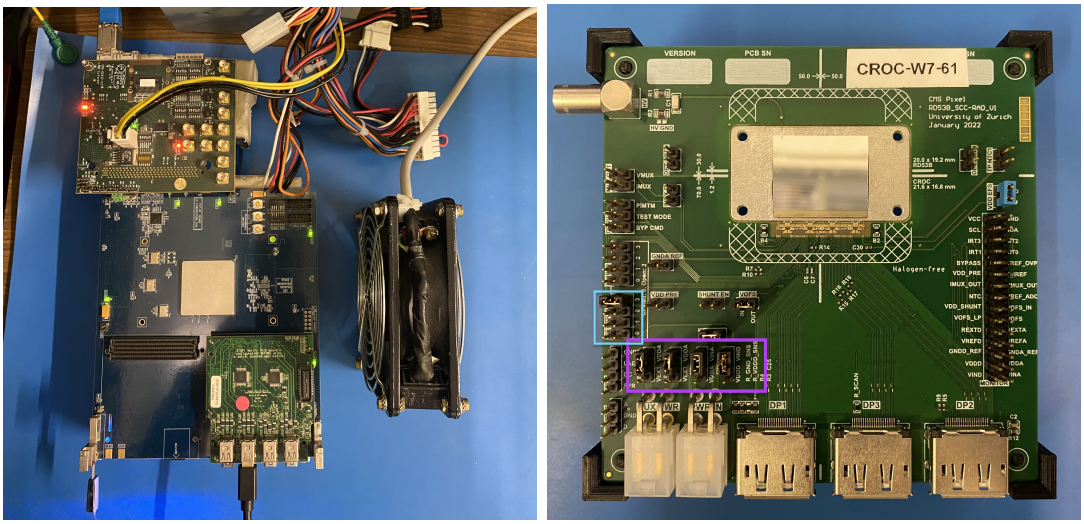


Figure 3.12: Photographs of the test setup for the CROC-V1 module. (Left) Full readout chain, including the FC7 board, KSU-FMC interface, and power connections. (Right) Close-up of the CROC-V1 module mounted on its carrier card, showing the 3D sensor at the center. The jumpers for the SLDO powering configuration are highlighted in purple, and the jumper used for trimming is highlighted in blue.

At power-on, the chip's current consumption depends on its default configuration. Trimming consists of fine-tuning the internal regulator voltages (V_{DDA} and V_{DDD}) and the main reference current (I_{ref}) to ensure correct operation. This involves measuring the voltages at specific pins and adjusting jumpers (cf. Figure 3.12) as needed. The goal is to set the analog and digital power domains to their nominal current values of approximately 630 mA for analog and 400 mA for digital, thus ensuring stable and optimal chip performance.

III.2. Pre-Irradiation Characterization

Before irradiation, a full calibration sequence was performed to validate the functionality of the CROC-V1 module and ensure uniformity in pixel response. The calibration process includes threshold tuning, Time-over-Threshold (ToT) calibration, and masking of noisy or dead pixels (see Section IV.4.c.). Threshold tuning adjusts the comparator levels so that each pixel triggers at the same input charge, while ToT calibration ensures that the pulse width correlates accurately with the deposited charge. These scans are essential for optimizing the performance and for evaluating the impact of radiation damage in subsequent steps.

Pixel Occupancy: Pixel occupancy (also referred to as pixel efficiency) quantifies how frequently a pixel registers an internal hit during a scan. It is computed by dividing the number of times a pixel registers such hits by the total number of trigger events. This measurement reflects the pixel's response to the internal test signals and does not involve real particle interactions. It is used to identify malfunctioning (dead) or excessively active (noisy) pixels.

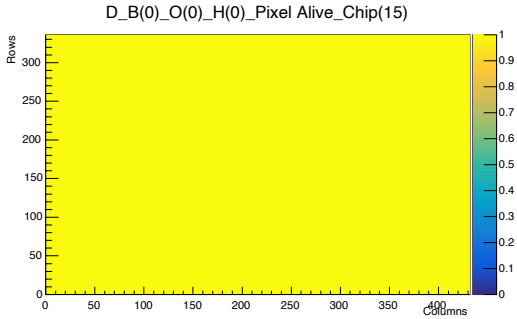


Figure 3.13: Pixel Alive map before irradiation. The z-axis corresponds to pixel occupancy.

Figure 3.13 shows the Pixel Alive scan before irradiation, with nearly all pixels exhibiting full efficiency, indicating no dead pixels or disconnected regions.

Time-over-Threshold (ToT): ToT is the time duration a signal remains above threshold and is correlated with the input charge. The ToT distribution and pixel map are shown in Figures 3.14. Before irradiation, the distribution is centered around a value of 7–8, consistent with nominal charge collection and amplification response. However, some strip-like patterns are visible on the ToT map. These patterns are not related to the physical properties of the chip but are artifacts introduced by the DAQ. Specifically, during the Pixel Alive scan, a predefined pattern of pixels is injected. If the DAQ encounters issues reading out a particular pattern, all pixels associated with that pattern may appear distorted, leading to these artificial structures. Such patterns are observed in nearly all scans but do not indicate any malfunction or degraded performance of the chip itself.

Threshold Tuning (TDAC): The threshold for each pixel is set to 2000 electrons before irradiation. The TDAC (Threshold DAC, see Section IV.4.c.) value is tuned per pixel to match this target. Figures 3.15 show the TDAC distribution and pixel map. Before irradiation, the TDAC values are uniformly distributed, with no edge effects or misbehaving areas.

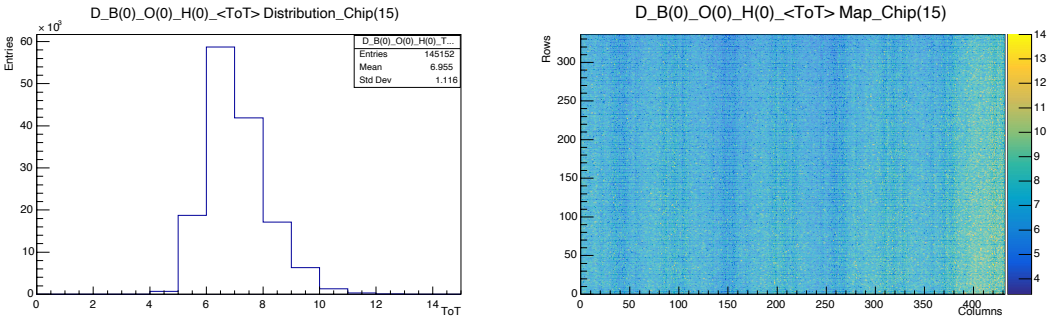


Figure 3.14: ToT distribution and map before irradiation.

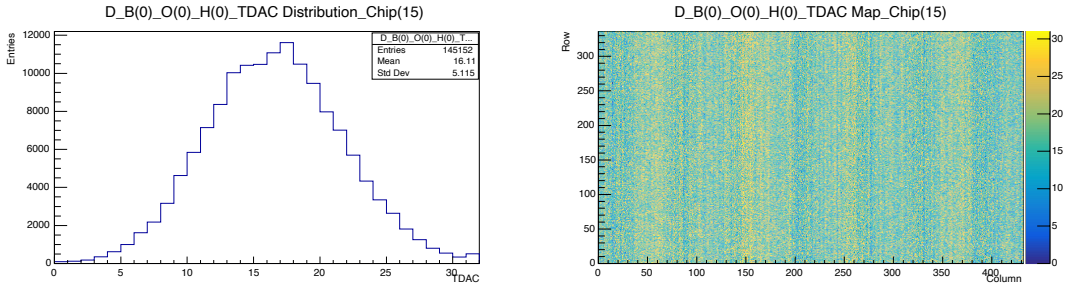


Figure 3.15: TDAC distribution before irradiation and map.

S-Curve Measurement: The S-curve scan measures the detection efficiency as a function of injected charge and provides information on the threshold sharpness and noise level of each pixel. An ideal S-curve shows a steep transition from 0% to 100% efficiency, indicating uniform and low-noise behavior. Figure 3.16 shows the S-curve distribution before irradiation, where the majority of pixels exhibit a sharp turn-on around the threshold. It should be noted, however, that the observed S-curve is relatively broad. This is expected, as the purpose of this scan was not to perform a full threshold calibration but rather to verify that the pixel array was functioning correctly. Consequently, the spread in the turn-on region reflects the preliminary nature of this measurement rather than intrinsic performance limitations.

Figures 3.17 and 3.18 show the threshold and noise distributions along with their corresponding pixel maps. The threshold distribution is narrow and centered around the target value, indicating good tuning with minimal pixel-to-pixel variation. Similarly, the noise distribution is uniform across the matrix, confirming the stability of the front-end electronics.

Conclusion: All calibration scans confirm that the chip was fully functional before irradiation, with uniform pixel response, minimal noise, and stable power

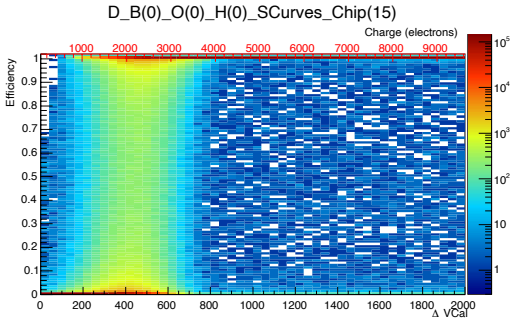


Figure 3.16: S-curve distribution before irradiation. The steep slope reflects a narrow threshold dispersion and low noise.

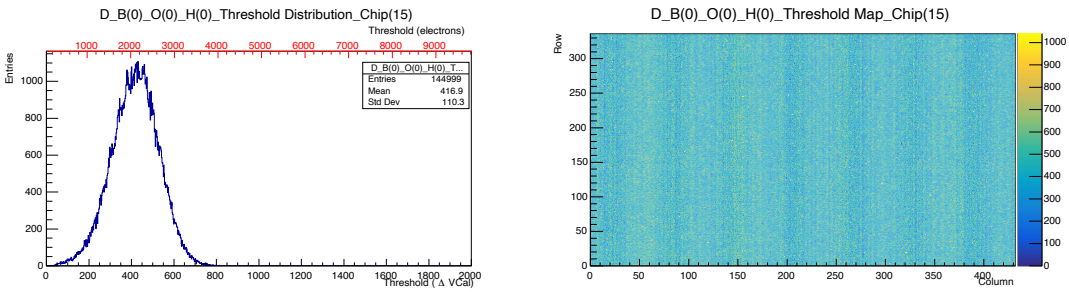


Figure 3.17: Threshold distribution (left) and threshold pixel map (right) before irradiation. The threshold is uniformly distributed across the pixel matrix.

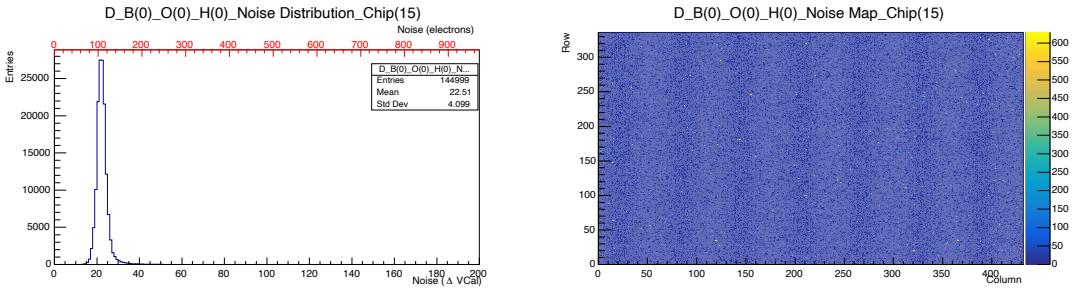


Figure 3.18: Noise distribution (left) and noise pixel map (right) before irradiation. The uniform noise response confirms stable analog front-end performance.

consumption. This provides a reliable baseline for evaluating radiation-induced degradation.

III.3. Post-Irradiation Measurements

Full Depletion Voltage Scan: After irradiation and approximately two weeks of annealing at room temperature, current-voltage (IV) measurements were performed at various ambient temperatures inside the cold box. The purpose of this scan was to estimate the full depletion voltage of the 3D sensor. Figure 3.19 shows the IV characteristics measured by gradually increasing the bias voltage and monitoring the corresponding leakage current.

It should be noted that determining the exact full depletion voltage from IV curves alone is challenging, as capacitance-voltage (CV) measurements are typically required for a more precise evaluation. Here, the aim is rather to establish a reasonable depletion voltage range and to verify that the sensor operates below the onset of thermal runaway. From the IV curve, the depletion voltage is estimated to lie between 25 and 45 V, based on the observed current behavior, which is consistent with expectations for 3D sensors after moderate fluence irradiation.

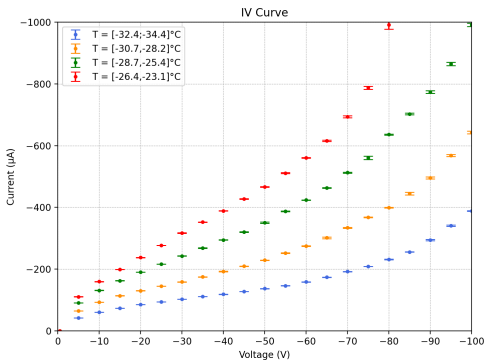


Figure 3.19: IV curve of the irradiated CROC-V1 module measured after approximately two weeks of annealing. The curve indicates a full depletion voltage in the range of 25–45 V.

Initial Post-Irradiation Scan: Following irradiation, a series of initial scans were performed to evaluate the functionality of the CROC-V1 chip. These included noise scans, pixel occupancy measurements, and threshold tuning.

As shown in Figure 3.20, the pixel occupancy map revealed a characteristic "W-shaped" pattern of noisy pixels. This observation was unexpected, given the homogeneous nature of the irradiation. Subsequent tests aimed to investigate whether this was the only affected region and to better understand the origin of this anomalous behavior. For all further measurements, the pixels within this W-shaped region were masked. Additionally, one column was manually masked due

to persistent readout errors, as indicated in the same figure. Aside from these masked regions, the average occupancy of the remaining pixels remained close to 1, indicating good behavior of the unmasked and untuned pixels.

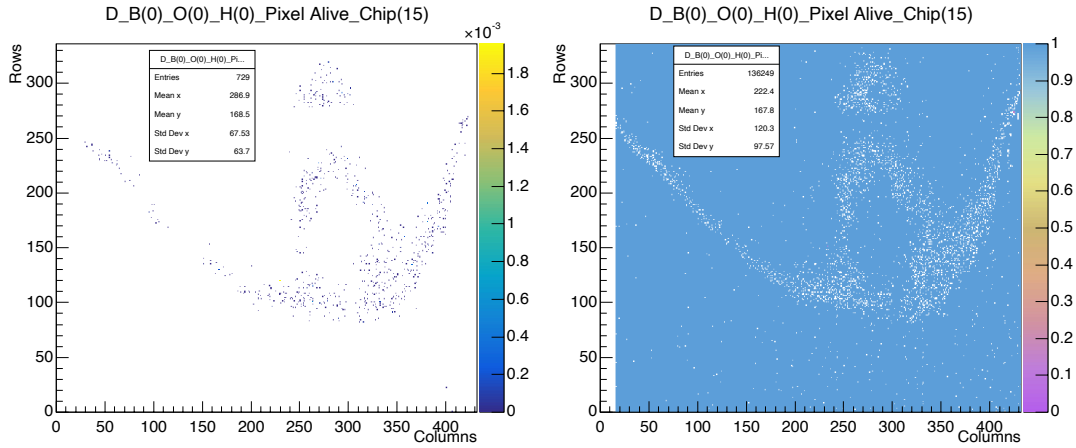


Figure 3.20: Noise scan (left) and pixel alive scan (right) following irradiation. The "W-shaped" noisy region and manually masked column are visible.

Threshold tuning was then performed in order to equalize the response of individual pixels. In an ideally tuned system, the occupancy distribution should peak around 50%, corresponding to the threshold-crossing point. However, as shown in Figure 3.21, a significant number of pixels exhibit occupancy values close to either 0 or 1, forming long tails in the distribution. Pixels with near-unity occupancy are either intrinsically noisy or have retained low thresholds due to imperfect tuning. These pixels often appear in vertical or horizontal band-like structures in the 2D map.

This behavior can be mitigated through per-pixel threshold tuning using the TDAC, as well as global threshold optimization via the LDAC (see Section IV.4.c.). Nevertheless, Figure 3.22 shows a residual tail in the TDAC distribution, indicating local variations not fully compensated by the tuning procedure. Moreover, the tuning behavior appears to differ between the regions above and below the W-shaped area, suggesting different local response characteristics. This discrepancy complicates the calibration process and raises questions about the underlying origin of the effect.

Investigation of the "W-shape": The origin of the W-shaped noisy region was extensively investigated through repeated scans performed under varying parameters and environmental conditions. Several hypotheses were evaluated. Non-uniform irradiation was ruled out based on post-irradiation Gafchromic film analysis,

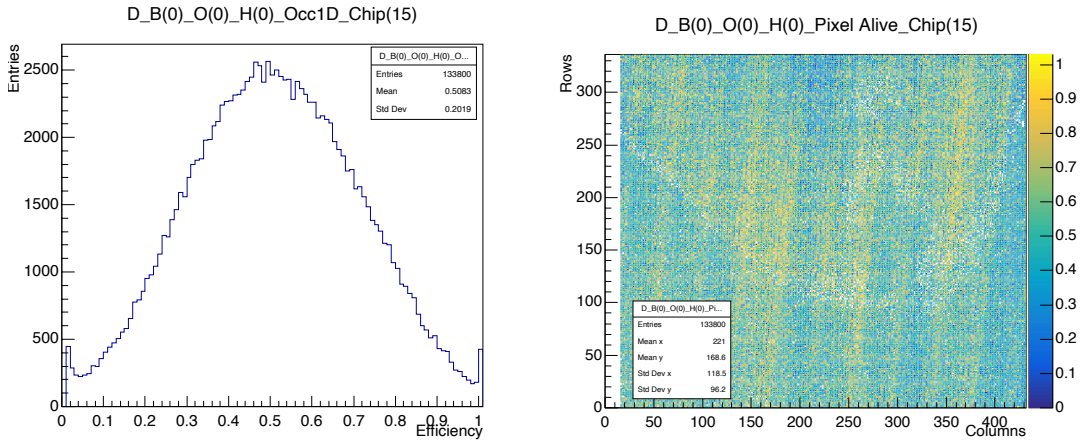


Figure 3.21: Occupancy distribution: 1D projection (left) and 2D map (right) after initial tuning.

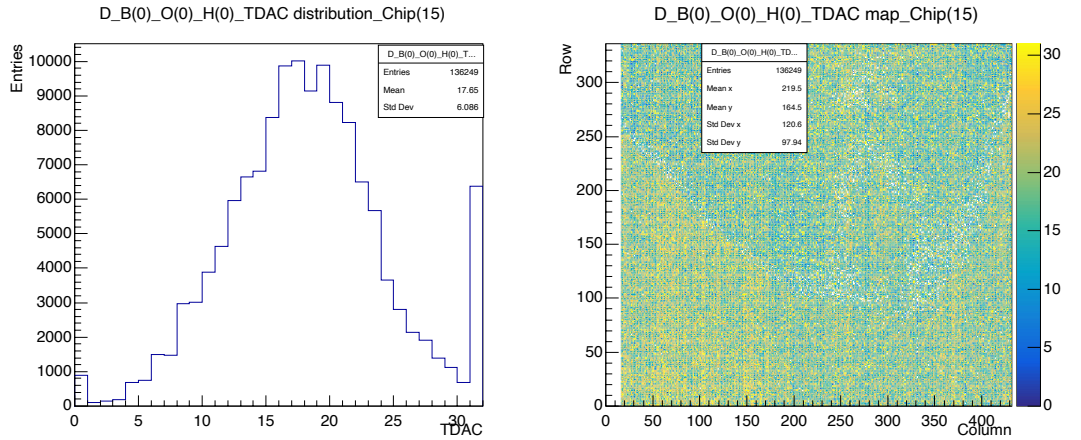


Figure 3.22: TDAC tuning distribution: 1D histogram (left) and 2D map (right).

which confirmed a homogeneous dose distribution. Incorrect trimming was also excluded, as re-trimming the chip after irradiation did not alter the observed behavior. Humidity or condensation was considered unlikely: multiple warm-up cycles produced no variation in the noisy region, and the environmental monitoring discussed in Section II. supports confidence in a controlled and dry environment. An electronic malfunction was deemed less probable, given the irregular and highly localized nature of the effect.

To further investigate, a *cross-talk scan* was performed. This scan takes advantage of the charge-sharing effect, where injected charge can propagate to neighboring pixels due to the design of the bump-bond layout and pixel proximity. A high

charge is injected into selected pixels, which are masked from the readout. The response of neighboring pixels is then recorded to evaluate charge propagation.

This method allows identification of disconnected bumps, pixels where the connection between sensor and readout chip is broken. The cross-talk scan results, shown in Figure 3.23, indicate that the region above the W-shape contains pixels with no neighboring response, suggesting a significant number of disconnected bump-bonds in this area.

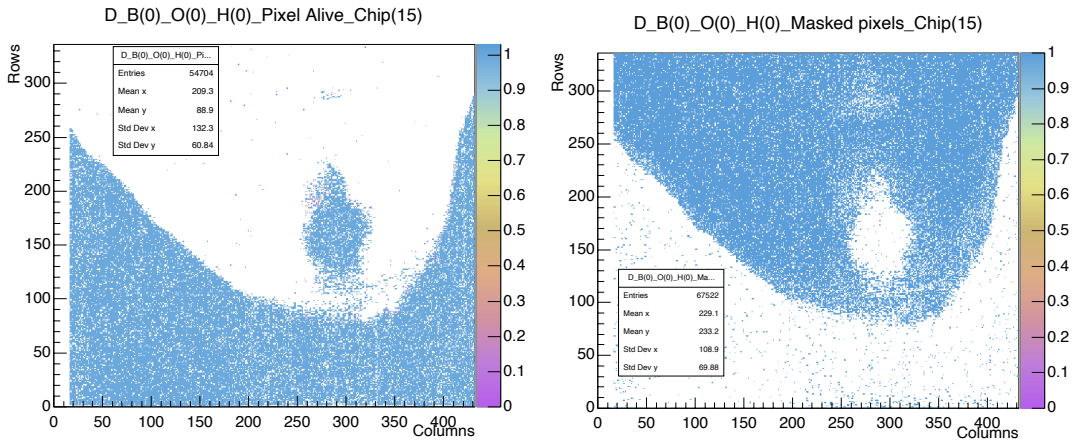


Figure 3.23: Cross-talk scan: pixel alive map (left) and masked pixels map (right).

The cause of these disconnected regions could be related to mechanical strain during handling or thermal cycling. However, given that the chip was manipulated with care, this behavior may point to pre-existing weaknesses in the bump-bonding. Further investigation, such as inspection via X-ray imaging, would be necessary to confirm the disconnection of the bump-bonds.

Calibration of the Functional Pixel Region: The chip was calibrated following the standard procedure, masking the disconnected bump-bond regions. This was done to verify the uniformity and tunability of the functional part of the chip after irradiation.

Threshold adjustment was performed resulting in a significantly more homogeneous pixel response. As shown in Figure 3.24, both the one-dimensional and two-dimensional occupancy distributions indicate a largely uniform activity, with residual strip-like patterns still visible.

The threshold equalization, illustrated in Figure 3.25, confirms that most pixels were successfully tuned. The TDAC distribution shows a peak with a suppressed tail compared to pre-tuning, again with the remaining tail attributed to the aforementioned DAQ artifacts. This demonstrates that, in the absence of bump-bond

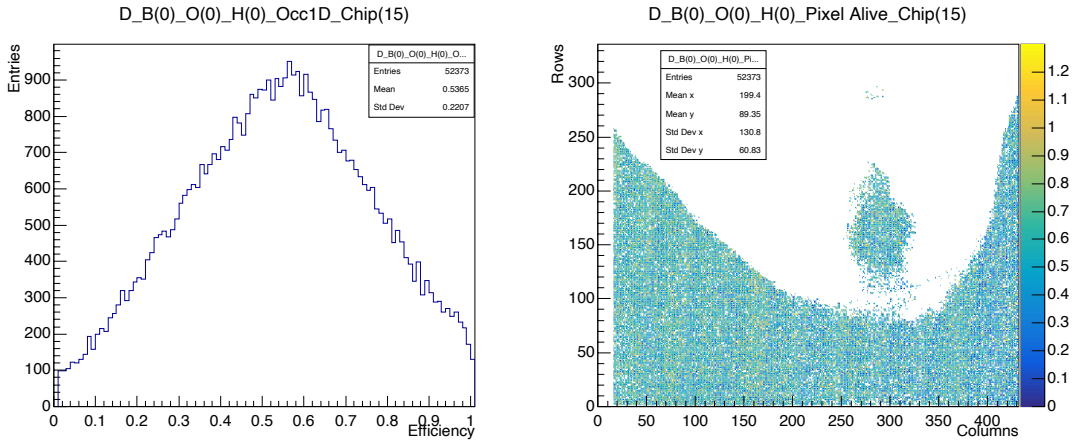


Figure 3.24: Occupancy distribution after tuning: 1D (left) and 2D (right).

failures, the chip is capable of achieving a stable and uniform threshold tuning.

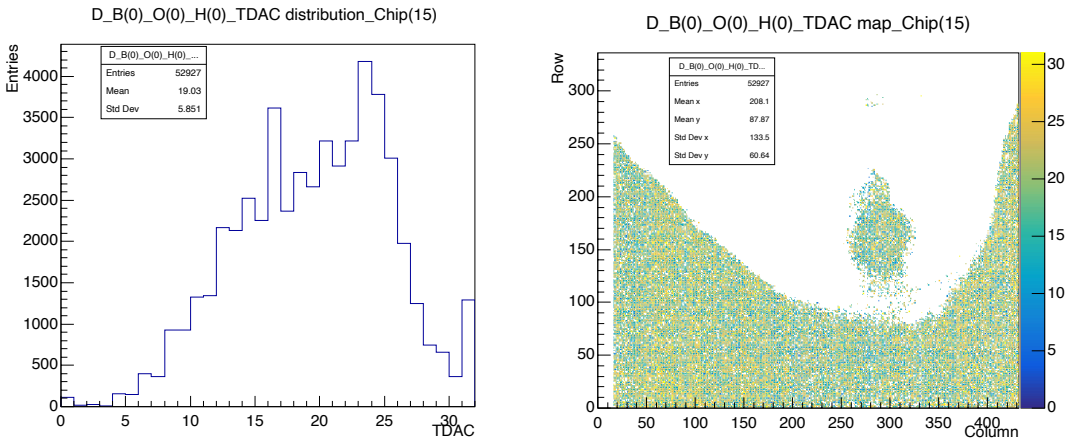


Figure 3.25: TDAC distribution: 1D (left) and 2D (right).

Noise measurements reveal the impact of irradiation more directly. In theory, the noise distribution should be centered close to zero since no signal is injected. However, as seen in Figure 3.26, a significant number of pixels, about 17% of the previously identified "healthy" region, were masked due to elevated noise levels. The comparison between S-curve fits for only the functioning pixels (left) and the full matrix including faulty pixels (right) shows that excluding the faulty areas leads to a narrower and cleaner distribution centered around the target value of 1500 electrons.

Figures 3.27 and 3.28 further detail the distributions of threshold and noise,

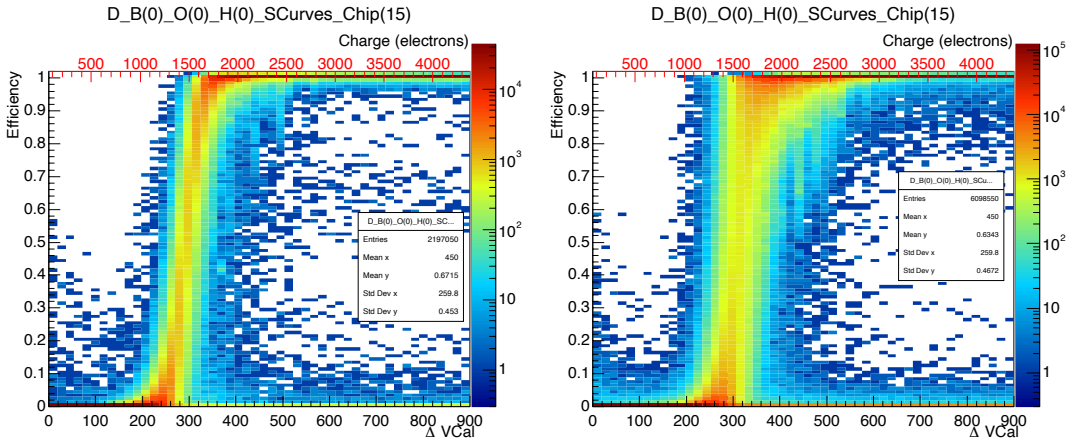


Figure 3.26: S-curve comparison between functional pixels only (left) and all pixels including faulty ones (right). Both tuned to 1500 electrons.

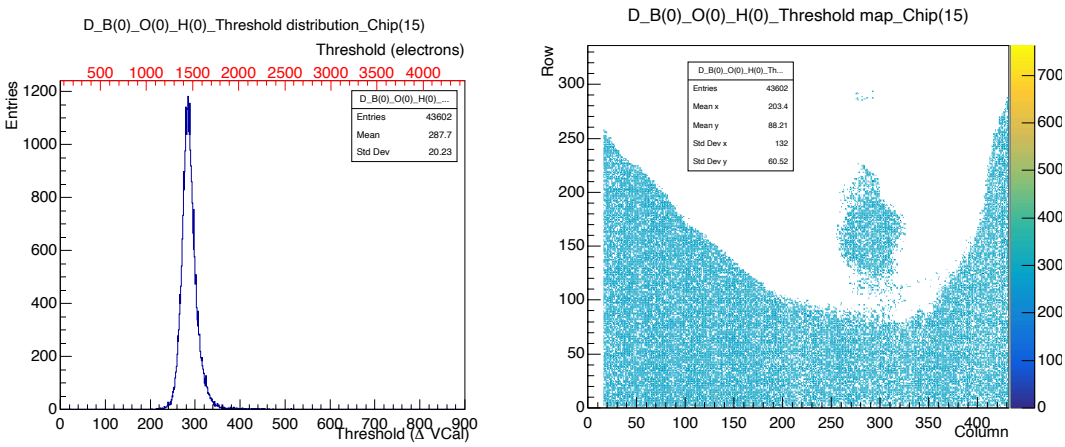


Figure 3.27: Threshold distribution in 1D (left) and 2D map (right).

respectively. The threshold map (1D and 2D) exhibits a single narrow peak, reflecting the chip's tuning uniformity across the functioning area. In contrast, the noise map shows an elevated level of noise across a wider region, consistent with post-irradiation expectations. As a result, approximately 30% of the initially considered functional pixels had to be masked due to excessive noise.

Altogether, this calibration confirms the remaining active pixels on the chip can still be tuned effectively, exhibiting a uniform threshold response and acceptable noise characteristics.

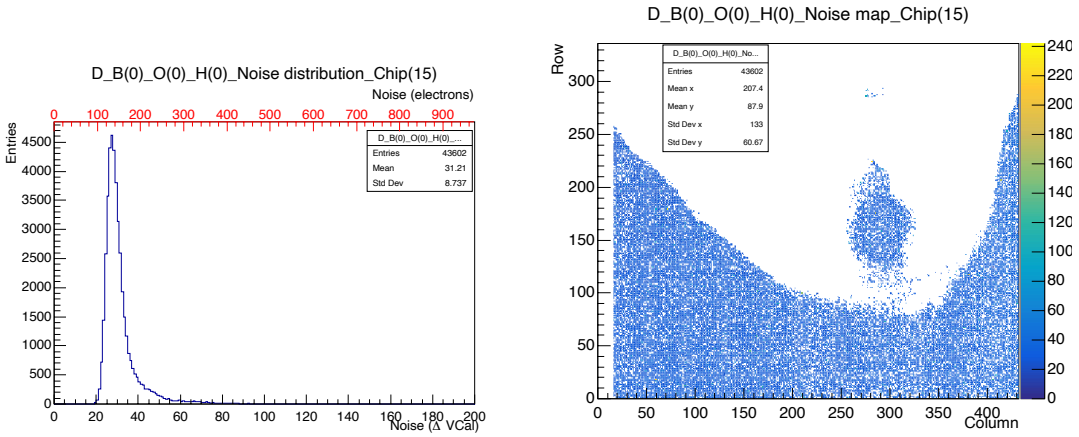


Figure 3.28: Noise distribution in 1D (left) and 2D map (right).

IV. Conclusion

The first irradiation campaign was successfully conducted, marking the commissioning of the new irradiation line at CRYCé. A CROC-V1 module was irradiated under cold conditions to a fluence of 5×10^{15} n_{eq}/cm^2, for which a dedicated setup and testing protocol were developed. The initial performance of the irradiated chip confirmed expectations, demonstrating that the chip remains functional after irradiation.

During post-irradiation testing, a non-uniform response was observed, originating from a localized region of disconnected bump-bonds. This effect prompted a detailed investigation, which ultimately pointed to mechanical weakness in the bump-bonding process as the most probable cause. By masking the affected region, it was confirmed that the rest of the chip behaves uniformly and that the irradiation itself was homogeneous.

This study validates both the irradiation procedure and the cold testing setup. With the infrastructure now operational and the methodology established, IPHC is prepared to proceed with the full irradiation campaign, including the irradiation of a series of different modules.

Chapter 4

Hadronic τ Reconstruction and Identification

Content

I. Object Identification and Reconstruction in CMS . . . 104

I.1. Tracks and Vertices . . . 104

I.2. Particle-Flow Event Reconstruction . . . 105

I.3. Muons . . . 107

I.4. Electrons . . . 109

I.5. Jets . . . 110

I.6. Missing Transverse Energy (MET) . . . 112

II. Hadronic τ Reconstruction and Identification . . . 113

II.1. Properties of Hadronic τ Decays . . . 113

II.2. Reconstruction with the Hadron Plus Strips (HPS) Algorithm . . . 114

II.3. Identification with the Deep Neural Network DeepTau . 116

II.3.a. Inputs . . . 117

II.3.b. Architecture and Loss Function . . . 118

II.3.c. Domain Adaptation by Backpropagation . . . 119

II.3.d. Expected Performance . . . 120

The τ leptons originating from a Z boson or a Higgs boson typically have energies on the order of several tens of GeV and decay before reaching the innermost layers of the pixel detector. As a consequence, τ leptons are not observed

directly in the detector and must be reconstructed and identified through their decay products. In CMS, the reconstruction of τ leptons relies on the particle-flow (PF) algorithm, which combines information from all sub-detectors to reconstruct individual particles such as charged hadrons, neutral hadrons, photons, electrons, and muons. These PF candidates form the basis for higher-level objects such as jets and hadronic τ decays. Therefore, understanding the reconstruction of these basic physics objects is a necessary step before addressing the τ lepton reconstruction and identification. This chapter first presents the PF reconstruction objects and then describes in detail the algorithms used for the τ lepton reconstruction and identification.

I. Object Identification and Reconstruction in CMS

I.1. Tracks and Vertices

The reconstruction of charged-particle tracks and primary-interaction vertices is performed using the Combinatorial Track Finder (CTF) algorithm [155]. This algorithm is based on an iterative Kalman Filter (KF) [156], which determines the momentum and trajectory of charged particles as they pass through their successive layers of the silicon tracker.

The track reconstruction starts by forming *seeds* from hits in the inner layers of the tracker. These seeds provide initial estimates of the trajectory and uncertainties. The seeds are then propagated outward, layer by layer, using the KF, which updates the estimated parameters at each step by incorporating compatible hits. The filter combines the measured hit positions with the predicted ones to refine the track while accounting for detector resolution and material interactions. After each fit, spurious hits are removed, and the trajectory is refitted until convergence. Tracks are retained only if they pass quality criteria designed to reject mis-reconstructed or fake tracks, such as a minimum number of hits, limits on the normalized χ^2, and impact parameter requirements with respect to the beam spot.

To efficiently handle the high hit density in LHC events, the reconstruction is performed in multiple iterations. The first iterations target tracks that are easier to find, such as those with high transverse momentum and originating near the primary interaction point. In the subsequent iterations, previously found hits are removed, and the algorithm focuses on more complex cases, such as tracks with lower p_T or large impact parameters, which are typical of displaced tracks from long-lived particles. For example, the third and fourth iterations are specifically designed to recover such displaced or low-momentum tracks. On average, about ten iterations are used in standard CMS reconstruction, depending on the event complexity and pileup conditions.

Once all tracks are reconstructed, interaction vertices are identified, including the primary vertex (PV) and pileup vertices. Tracks compatible with the *beam spot* are clustered in z using a deterministic annealing algorithm [157]. The beam spot (BS) is defined as the center of the luminous region where the LHC beams collide at Point 5. Its position is estimated by averaging the location of reconstructed primary vertices over many consecutive bunch crossings. As it reflects the overlap region of the two beams, the BS is approximately stable during an LHC fill, though it can vary between fills due to changes in beam conditions and focusing. Once the BS is determined, each vertex candidate is fitted using an adaptive vertex fitter (AVF) [158], which refines the vertex position and assigns weights to tracks according to their compatibility with the vertex. The primary vertex is defined as the one with the highest scalar sum of squared transverse momenta ($\sum p_T^2$) of the associated physics objects, such as jets. Secondary vertices (SVs) are also reconstructed and associated with the decay of long-lived particles like b-hadrons or τ leptons.

The performance of the tracking and vertexing algorithms was evaluated using data from Run 2 proton-proton collisions at $\sqrt{s} = 13$ TeV and early Run 3 data at $\sqrt{s} = 13.6$ TeV. The track reconstruction efficiency for promptly produced charged particles with transverse momentum $p_T > 0.9\,\text{GeV}$ has been measured to be approximately 99.9% for tracks associated with muons, uniformly across the full muon pseudorapidity acceptance in both Run 2 and early Run 3 data [159]. For primary vertices with a scalar sum of track transverse momenta exceeding 100 GeV, the transverse position resolution is better than $14\,\mu$m, while the longitudinal resolution is better than $19\,\mu$m [160].

I.2. Particle-Flow Event Reconstruction

The PF algorithm [32, 161, 162] is used to reconstruct all stable particles in an event, including electrons, photons, muons, and hadrons. This is achieved by combining information from all of the CMS sub-detectors shown in Figure 4.1 in order to reconstruct particle energies, momenta, types and trajectories. The output of the PF algorithm is a list of PF particles that can be used to reconstruct higher-level objects, such as jets, hadronically decaying taus (τ_h), and MET.

The PF algorithm reconstructs individual particles by combining information from the tracker and calorimeters. It begins with inputs from reconstructed tracks and calorimeter energy deposits. In the ECAL and HCAL barrel and endcap regions, clusters are formed by grouping neighboring cells with significant energy, whereas in the coarser granularity hadron forward (HF) calorimeter, each cell is treated as an individual cluster.

Neutral particles, such as photons and neutral hadron are identified as isolated clusters without associated tracks. Charged hadrons, on the other hand,

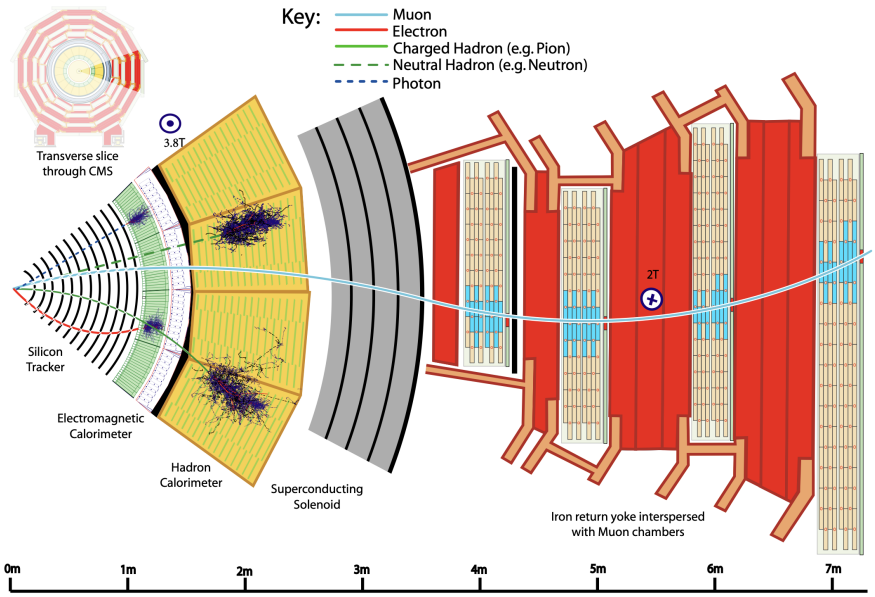


Figure 4.1: A scheme of example of particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector. The muon and the charged pion are positively charged, and the electron is negatively charged [162].

are characterized by a combination of tracks and corresponding energy deposits. A dedicated linking algorithm associates tracks to clusters based on geometric and topological compatibility. Tracks are extrapolated to the calorimeters, and matches are identified when the projected impact point falls within a cluster's spatial extent. This association enables robust particle identification, particularly in complex environments like jets.

To account for electromagnetic radiation, the algorithm also identifies Bremsstrahlung photons by searching for ECAL clusters aligned with the tangent direction of electron tracks. This feature enhances the reconstruction of electrons undergoing significant energy loss via photon emission.

In the CMS detector, this approach is particularly effective due to the detector geometry: the entire tracking volume and calorimetry system are located inside the superconducting solenoid. As a result, particles do not traverse any magnetic field between the tracker and the calorimeters, minimizing trajectory distortions and improving the spatial correlation between tracks and clusters. This is in contrast to the ATLAS detector design, where the magnetic field is situated between the tracker and calorimeters, complicating the track-cluster association.

Once the tracks and clusters are linked, the PF algorithm assigns them to spe-

cific particles. The algorithm follows a sequence to identify particles, starting with muons. Muons are identified by using the hits in the muon chambers and silicon tracking system, and by matching them with energy deposits in the calorimeters. After a track is assigned to a muon, it is removed from further reconstruction steps.

Next, the algorithm identifies electrons by pairing tracks in the silicon tracker with corresponding energy clusters in the ECAL. If the pair passes a multivariate discriminant, the object is identified as a PF electron, and both the track and cluster are removed.

The remaining tracks are identified as charged hadrons, with momentum given by the track momentum and mass set to that of a charged pion. The momentum is redefined if the calorimetric energy is consistent with the track momentum within uncertainties. If the calorimetric energy does not match the track momentum, the excess energy is attributed to overlapping photons or neutral hadrons. If the excess is larger than the ECAL energy, a PF photon is created with energy equal to the ECAL energy, and a PF neutral hadron is created with the remaining energy. If the excess is smaller, only a PF photon is created with the excess energy.

Finally, the PF algorithm estimates the total transverse energy of the event. In a proton-proton collision, the total transverse momentum must sum to zero. If this is not the case, the imbalance in the transverse plane is attributed to missing transverse energy. This MET could indicate the presence of invisible particles, such as neutrinos, or detector inefficiencies. An accurate estimation of the MET is especially crucial for the reconstruction of the τ_h full momenta.

I.3. Muons

The CMS detector is optimized for muon reconstruction [163, 164], taking advantage of the clean and distinct signatures muons leave in the outer muon chambers. Muon reconstruction relies partially on the PF algorithm and uses the tracks independently reconstructed in the inner tracker and in the muon chambers. Tracks in the pixel tracker are built using the PF approach, and are referred to as *tracker tracks*, while the *standalone tracks* use a standalone procedure to reconstruct tracks in the muon chambers. Muon candidates are built by combining these two different types of tracks, leading to three different muon types:

- **Standalone muons:** Reconstructed using only information from the muon chambers. Tracks are formed from hits in DT, CSC, or RPC systems, and are used to seed the track reconstruction via a KF, which accounts for energy loss in the detector material (see Section I.1.). Due to their susceptibility to contamination from cosmic rays and punch-through hadrons, pure standalone muons are typically not used directly in physics analyses.

- **Tracker muons:** Reconstructed using an inside-out approach. Tracks built in the inner silicon tracker with $p_T > 0.5$ GeV and $p > 2.5$ GeV are extrapolated outward to the muon chambers. If the extrapolated track matches at least one muon segment in the local (x, y) plane, either satisfying $\Delta x < 3$ cm or $\Delta x/\sigma_{\Delta x} < 4$, the track is classified as a tracker muon. This method ensures high efficiency, especially for low-momentum muons that may not reach deeper muon stations.

- **Global muons:** Reconstructed using an outside-in approach. Standalone muon tracks are matched to tracker tracks based on the compatibility of their extrapolated parameters at a common surface. The two tracks are then refitted together using a KF to form a global muon. Global muons are especially valuable for high-momentum tracks ($p_T > 200$ GeV), where combining both sub-detectors improves the momentum resolution.

Muons that share the same tracker track can be merged into a single candidate. Overall, CMS achieves a muon reconstruction efficiency greater than 99% for muons identified as either tracker or global muons. For muons with $p_T < 200$ GeV, the momentum assignment is based solely on the inner tracker, while for muons with $p_T > 200$ GeV, it is derived from the global fit.

The muon identification proceeds by applying a series of selection criteria based on variables related to the muon reconstruction, such as the number of hits in the inner tracker and in the muon chambers, the quality of the track fit (quantified by the χ^2 per degree of freedom), and the geometrical compatibility between the standalone muon track and the corresponding tracker track. This compatibility is evaluated using matching variables such as the angular separation and the consistency of the track parameters. These criteria are combined into several identification working points (WPs), commonly labeled as *Loose*, *Medium*, and *Tight*, to provide increasing levels of background rejection. Muon identification efficiencies for the Tight WP as a function of muon p_T are shown in Figure 4.2.

In addition, muon isolation is an important variable to discriminate prompt muons from those originating from in-flight decays of hadrons or from hadronization processes inside jets. The PF relative isolation variable is defined as:

$$I^{\mu}_{\text{rel}} = \frac{\sum p_T^{\text{charged}} + \max\left(0, \sum E_T^{\text{neutral}} + \sum E_T^{\gamma} - 0.5 \sum p_T^{\text{charged, PU}}\right)}{p_T^{\mu}},$$

where all quantities in the numerator are computed from PF candidates within a cone of radius $\Delta R < 0.4$ around the muon direction. The term

$$\sum E_T^{\text{neutral}} + \sum E_T^{\gamma} - 0.5 \sum p_T^{\text{charged, PU}}$$

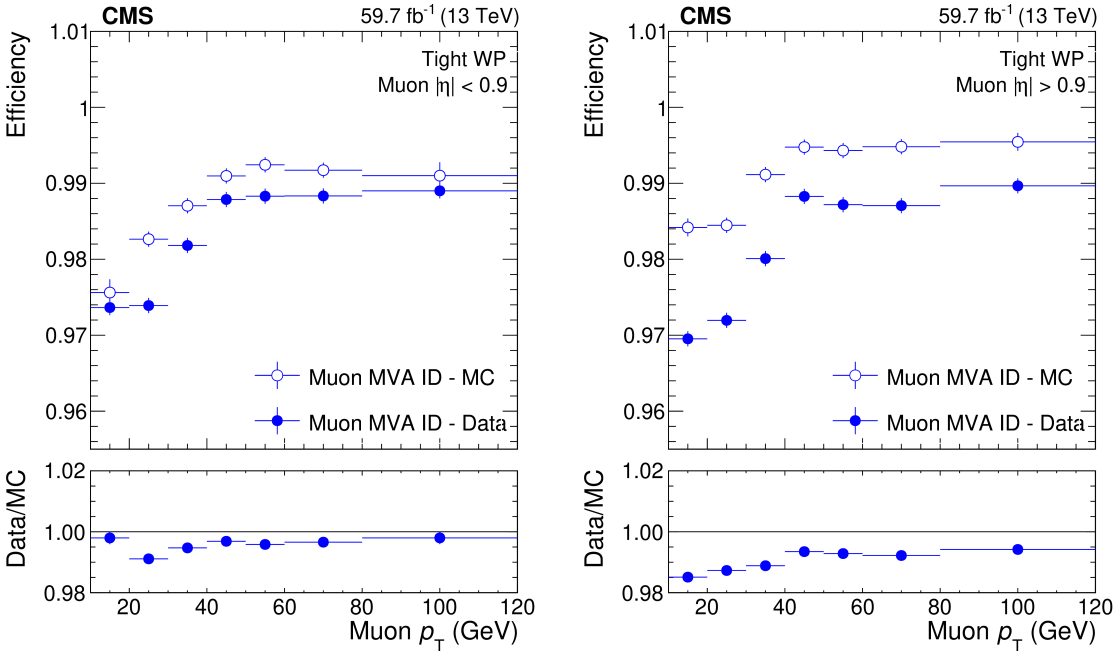


Figure 4.2: Muon identification efficiency for the Tight working points as a function of the muons p_T for muons with $|\eta| < 0.9$ (left) and $|\eta| > 0.9$ (right) [165].

corrects the neutral energy deposits for PU contributions. This correction relies on the approximate proportionality between neutral and charged particle production in PU interactions. It also accounts for the fact that charged particles originating from PU vertices can emit photons via Bremsstrahlung, which contribute to the neutral energy measured in the calorimeters. To prevent an overcorrection that would lead to negative values, the corrected neutral energy is set to be non-negative by applying a $\max(0, \cdot)$ function.

Muons play a crucial role in the $\tau\tau$ analysis. In particular, they are used as tag objects in the $Z \to \mu\tau_h$ events, which are employed to calibrate the hadronic τ, as detailed in Chapter 5.

I.4. Electrons

Unlike muons, which are detected by a dedicated subdetector, electrons are reconstructed by matching tracks from the tracker with energy clusters in the ECAL [166]. As electrons pass through the detector, they emit Bremsstrahlung photons, which lead to a significant energy loss. At $\eta \approx 0$, this energy loss represents about 33% of the total electron energy, increasing up to 86% at $\eta \approx 1.4$, where the material thickness is maximal.

The Bremsstrahlung effect causes the electron energy to be spread out, mainly

along the ϕ direction due to the high magnetic field. To reconstruct the electron's energy, clusters are combined into a *supercluster* (SC) around the seed PF cluster with the highest energy in a given $\eta \times \phi$ window. Two algorithms are used for SC reconstruction: the "hybrid" algorithm in the barrel and the "multi-5×5" algorithm in the endcaps. The hybrid algorithm uses a small η window and an extended ϕ window, exploiting the electromagnetic shower shaped by the magnetic field, while the multi-5×5 algorithm works with the endcap crystals. The total energy of the SC corresponds to the sum of the energies of all its clusters.

In the tracker, Bremsstrahlung also affects the electron track curvature, and the KF algorithm may fail to collect all the hits. Therefore, the KF is used for track seeding, while a Gaussian sum filter (GSF) algorithm is applied to handle sudden changes in the track path and energy loss at the material surface, ensuring compatibility with the SCs. Electron candidates are constructed by associating a GSF track with an ECAL supercluster.

To improve background rejection, a multivariate analysis (MVA) is used to identify electrons based on several discriminating variables, including the shower shape, HCAL energy fraction, track quality, and isolation. The isolation discriminant is defined as:

$$I_{\text{rel}} = \frac{1}{p_T^e} \left(\sum p_T^{h^\pm} + \max\left(0, \sum p_T^{h^0} + \sum p_T^\gamma - \rho A_{\text{eff}}\right)\right), \tag{4.1}$$

where p_T^e is the electron transverse momentum, ρ is the pileup energy density, and A_{eff} is an η-dependent effective area. The sum runs over charged hadrons ($h^\pm$), neutral hadrons (h^0), and photons (γ) within a cone of $\Delta R = 0.3$.

Electron identification is defined through working points based on the MVA score thresholds. The *Loose* working point achieves $\sim 90\%$ efficiency with ~ 2–5% misidentification rate, while the *Tight* working point provides a higher background rejection with $\sim 80\%$ efficiency. The selection strategy depends on the analysis-specific trade-off between signal efficiency and background suppression.

The reconstruction and identification efficiencies are validated using a tag-and-probe method in $Z/\gamma^* \to e^+e^-$ events. For the 2016 data, the tracking efficiency exceeds 95% in the range $25 < p_T < 500\,\text{GeV}$ and $|\eta| < 2.1$, demonstrating a robust performance across detector regions and electron kinematics [166].

I.5. Jets

Collimated sprays of hadrons, referred to as *jets*, originate from the hadronization and fragmentation of quarks and gluons produced in the hard scattering process [167]. These jets offer an effective way to probe the kinematic properties of the originating partons. From a theoretical point of view, this relation statement stays valid when the jet clustering algorithm satisfies the conditions of infrared

and collinear (IRC) safety, ensuring stability against additional soft radiation or collinear splittings.

The CMS experiment employs the anti-k_T clustering algorithm [168], which is IRC-safe and favors clustering around hard particles, yielding jets with a nearly conical geometry in the (η, ϕ) plane. Jets are reconstructed from PF candidates using the implementation provided by the `FastJet` package [169]. To mitigate the impact of PU, charged PF candidates not associated with the primary vertex are excluded using the charged hadron subtraction technique [170].

The anti-k_T algorithm defines a distance measure between objects i and j, or between object i and the beam axis, given by:

$$d_{ij} = \begin{cases} \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, & i \neq j \\ p_{T,i}^{2p}, & i = j \end{cases} \tag{4.2}$$

where $p = -1$ for anti-k_T, $\Delta R_{ij} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the distance in the η-ϕ plane, and R is the radius parameter, set to 0.4. The algorithm proceeds iteratively by computing all d_{ij} and d_{iB} values, merging the pair with the smallest d_{ij} if $i \neq j$, or declaring i a jet if d_{iB} is the minimum. This ordering ensures that hard, collimated particles are clustered first, producing jets with well-defined shapes and axes.

Jets considered in the analysis, are required to pass the loose identification criteria designed to suppress noise and pileup jets. In particular, for the problematic region $2.65 < |\eta| < 3.14$, where the ECAL endcap is affected by increased noise levels, a dedicated PU jet identification discriminant is employed [171]. This discriminant combines vertex association and jet shape observables, with the latter providing the full discrimination beyond $|\eta| > 2.5$, where tracking is unavailable.

To ensure an accurate measurement of the jet energy and direction, a sequence of jet energy corrections (JECs) is applied [172]:

- **Pileup offset correction:** estimated using dijet simulations with and without PU, and validated using the random cone (RC) method in zero-bias data events.
- **Relative and absolute response corrections:** derived as a function of η and p_T using multijet and $Z/\gamma^* \to \ell^+\ell^- +$ jets events.
- **Residual data/MC corrections:** applied to account for small discrepancies in the detector response.
- **Jet flavor corrections:** obtained by comparing simulations using different hadronization models (PYTHIA [173] and HERWIG++ [174]).
- **b-jet energy corrections:** validated with $Z/\gamma^* + b$ events in data.

The jet energy scale (JES) and resolution (JER) are determined from both simulation and in situ techniques. The JES uncertainty for central jets with $p_T >$ 30 GeV is below 2%, and is primarily limited by jet flavor dependence. The JES uncertainty remains relatively flat for $|\eta| < 2.5$ at approximately 2.5%, increasing in more forward regions due to the growing importance of PU and methodological uncertainties [175].

Jet reconstruction and identification efficiencies were measured using $2.3\,\mathrm{fb}^{-1}$ of collision data at $\sqrt{s} = 13$ TeV collected in 2015. The jet ID efficiency was found to be nearly 100% [176], confirming the robustness of the reconstruction algorithm and the selection criteria.

I.6. Missing Transverse Energy (MET)

Stable or long-lived weakly interacting particles, such as neutrinos or potential BSM candidates, traverse the CMS detector without interacting and cannot be directly reconstructed. Their presence is inferred from an imbalance in the transverse momentum of visible particles. The missing transverse momentum vector, $\vec{p}_T^{\,\mathrm{miss}}$, is defined as the negative vectorial sum of the transverse momenta of all reconstructed particles using the PF algorithm:

$$\vec{p}_T^{\,\mathrm{miss}} = -\sum_i \vec{p}_{T,i}, \qquad E_T^{\mathrm{miss}} = |\vec{p}_T^{\,\mathrm{miss}}|. \tag{4.3}$$

The MET measurement relies on the CMS detector's hermetic calorimetry and excellent muon reconstruction performance. Accurate reconstruction is essential, particularly in analyses with τ decays that contains neutrinos. Since the longitudinal momentum of the partons is not known, the MET is computed only in the transverse plane.

The accuracy of the MET measurement is affected by several factors, including nonlinearities in the calorimeter response, imperfect reconstruction of final-state particles. To mitigate such effects, jet energy corrections (JECs) are propagated to MET as follows:

$$\vec{E}_T^{\mathrm{miss,corr}} = \vec{E}_T^{\mathrm{miss,raw}} - \sum_{\mathrm{jets}} (\vec{p}_{T,\mathrm{jet}}^{\,\mathrm{corr}} - \vec{p}_{T,\mathrm{jet}}^{\,\mathrm{raw}}), \tag{4.4}$$

where the sum includes all jets with $p_T > 15\,\mathrm{GeV}$ and less than 90% of their energies in the ECAL [177].

To improve the data–simulation agreement, recoil corrections are applied to simulated samples. The hadronic recoil vector $\vec{U}_T$ is defined by subtracting the neutrino momenta from the reconstructed MET:

$$\vec{U}_T = \vec{E}_T^{\mathrm{miss}} - \sum \vec{p}_T^{\,\nu}, \tag{4.5}$$

and in $Z \to \mu\mu$ events as:

$$\vec{U}_T = -\vec{H}_T - \vec{p}_T^{\,Z}, \tag{4.6}$$

where $\vec{H}_T$ is the vector sum of jet momenta. The components of $\vec{U}_T$ parallel and perpendicular to $\vec{p}_T^{\,Z}$ are fitted in data and simulation using a Gaussian model, and correction factors are applied to match the resolution and scale:

$$U'_{\parallel} = \langle U_{\parallel} \rangle_{\text{data}} + (U_{\parallel} - \langle U_{\parallel} \rangle_{\text{MC}}) \frac{\sigma_{\text{data}}(U_{\parallel})}{\sigma_{\text{MC}}(U_{\parallel})}, \tag{4.7}$$

$$U'_{\perp} = U_{\perp} \cdot \frac{\sigma_{\text{data}}(U_{\perp})}{\sigma_{\text{MC}}(U_{\perp})}. \tag{4.8}$$

The MET is then recalculated using the corrected recoil. The MET performance has been evaluated in $Z/\gamma^* \to e^+e^-/\mu^+\mu^- +$jets events, which do not contain intrinsic MET. Good agreement between data and simulation was observed, with resolutions of the transverse and longitudinal components around 23 GeV at $p_T^Z \sim 100$ GeV and a small degradation with increasing boson p_T [175, 176].

II. Hadronic τ Reconstruction and Identification

II.1. Properties of Hadronic τ Decays

The τ lepton is the heaviest lepton, with a mass of $m_\tau = 1.776$ GeV. Its short lifetime, $\tau = 2.90 \times 10^{-13}$ s (or 0.290 ps), corresponds to a decay length of approximately 1 mm for a 20 GeV τ. The τ decays via the weak interaction, as illustrated by the tree-level Feynman diagram in Fig. 4.3.

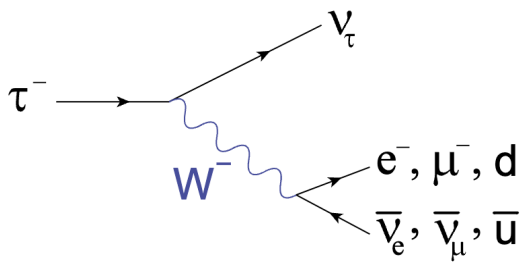


Figure 4.3: Tree-level Feynman diagram of τ decay via the weak interaction.

The τ decays are classified into leptonic and hadronic modes. Leptonic decays occur in 35.2% of cases and produce a lighter lepton (electron or muon) along with two neutrinos: $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ or $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$. The remaining 64.8% of decays are hadronic (τ_h), in which the τ decays into a ν_τ and a hadronic system, mainly composed of pions.

Hadronic decays are further classified by the number of charged particles in the final state, referred to as *prongs*. Most τ_h decays are either one-prong or three-prong. They are also categorized by the intermediate mesonic resonance involved:

- π mode: $\tau^- \to \pi^- \nu_\tau$
- ρ mode: $\tau^- \to \pi^- \pi^0 \nu_\tau$
- a_1 one-prong: $\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$
- a_1 three-prong: $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$

The branching ratios for these channels are shown in Figure 4.4.

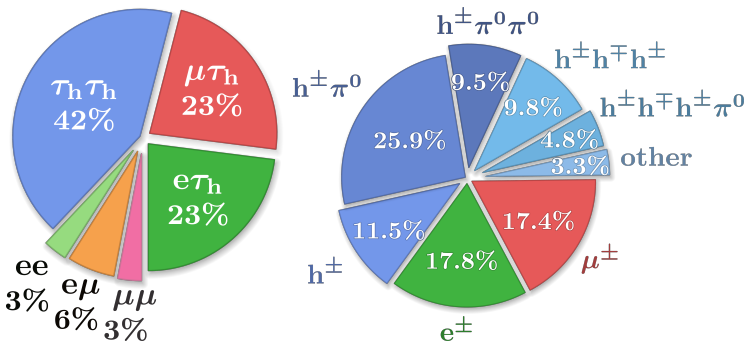


Figure 4.4: Right: Branching fractions of τ lepton decays into leptonic and hadronic final states. Left: Composition of the $\tau\tau$ decay final [56].

In $\tau^+\tau^-$ pair production, final states containing at least one hadronic decay dominate. The $\tau_h\tau_h$ and $\ell\tau_h$ (with $\ell = e, \mu$) final states account for approximately 42% and 46% of decays, respectively, while fully leptonic channels represent only 12%.

II.2. Reconstruction with the Hadron Plus Strips (HPS) Algorithm

The reconstruction of τ_h candidates in CMS is performed using the Hadron Plus Strips (HPS) algorithm [33]. The algorithm is designed to reconstruct the main intermediate resonances in hadronic τ decays and to classify them into distinct decay modes (DMs). Charged hadrons are reconstructed using the PF algorithm, while neutral pions are identified through the clustering of photons in the ECAL into so-called *strips*, which account for photon conversions and Bremsstrahlung. The reconstructed decay products are required to match the topology of one-prong or three-prong τ decays with up to two neutral pions. The HPS algorithm proceeds as follows:

1. **Jet Seed:** PF jet candidates identified by the anti-k_T algorithm with $\Delta R = 0.4$ are used as seeds.

2. **Identification of Candidates:**

   - *Charged Hadrons:* Selected from PF charged hadrons within the jet cone and used to form the charged prongs of the decay.
   - *Photons and Electrons:* These are used to reconstruct neutral pions via the strip reconstruction algorithm, described below.

3. **Strip Reconstruction:** Neutral pions from τ decays often produce photons that can convert into e^+e^- pairs. In the presence of the magnetic field, these particles are deflected predominantly in the ϕ direction, creating ECAL energy deposits that are narrow in η but extended in ϕ. These are clustered into so-called *strips*.

   A dynamic strip reconstruction algorithm was introduced in Run 2, replacing the fixed-size window of Run 1 ($\Delta\eta \times \Delta\phi = 0.05 \times 0.20$). It adapts to the p_T of the objects, improving performance for boosted topologies. The algorithm proceeds as follows:

   (a) A strip is seeded by the leading photon or electron found in the jet that has not yet been assigned to a strip.

   (b) The position and transverse momentum of the leading e/γ are assigned to the strip.

   (c) The next highest p_T e/γ candidate is added to the strip if the distance between its position and the strip center satisfies the following conditions:

   $$\Delta\eta(e/\gamma - \text{strip}) < \max\left(f(p_T^{e/\gamma}) + f(p_T^{\text{strip}}), 0.15\right)$$

   $$\Delta\phi(e/\gamma - \text{strip}) < \max\left(g(p_T^{e/\gamma}) + g(p_T^{\text{strip}}), 0.30\right)$$

   where the functions $f(p_T)$ and $g(p_T)$ are defined as:

   $$f(p_T) = 0.20 \cdot p_T^{-0.66}, \quad g(p_T) = 0.35 \cdot p_T^{-0.71}$$

   (d) If the e/γ candidate is included in the strip, the properties of the strip are recalculated, and the process continues until no new e/γ candidates are found within the strip window.

4. **Combination and Decay Mode Assignment:** The highest-p_T charged hadrons and up to six strips are combined to form τ_h candidates. The visible

four-momentum is reconstructed from these constituents, and the invariant mass m_{τ_h} is used to classify the candidate into a DM.

The decay mode is determined based on the number of charged hadrons (*prongs*), the number of strips (interpreted as neutral pions), and the invariant mass of the system. The modes are assigned as follows:

- **DM 0**: $h^{\pm}$ (1-prong, no strip)
- **DM 1**: $h^{\pm}\pi^0$ (1-prong, 1 strip)
- **DM 2**: $h^{\pm}\pi^0\pi^0$ (1-prong, 2 strips)
- **DM 10**: $h^{\pm}h^{\pm}h^{\mp}$ (3-prong, no strip)
- **DM 11**: $h^{\pm}h^{\pm}h^{\mp}\pi^0$ (3-prong, 1 strip)

The decay mode index is computed as:

$$\mathrm{DM} = 5 \times (n_{\mathrm{prongs}} - 1) + n_{\mathrm{strips}},$$

where n_{prongs} is the number of charged hadrons and n_{strips} is the number of strips.

5. **Final Selection:** Although the HPS algorithm defines multiple decay modes, only DM 0, 1, 2, 10, and 11 are typically used in physics analyses.

II.3. Identification with the Deep Neural Network DeepTau

Hadronic τ candidates reconstructed with the HPS algorithm can be misidentified by other objects, commonly referred to as *fakes*. The dominant sources are quarks and gluons jets ($j \to \tau_h$) and isolated electrons or muons ($\ell \to \tau_h$), especially in one-prong decay modes. To suppress these backgrounds, CMS employs the *DeepTau* discriminator, a deep convolutional neural network that performs multiclass classification to distinguish genuine τ_h decays from jets, electrons, and muons. The output is a set of class probabilities, used to define working points (WP) for analysis selections.

The version used in Run 3, DeepTau v2.5 [35], improves upon v2.1 [178] by introducing domain adaptation via adversarial training to reduce data–simulation discrepancies. It is trained on a balanced sample of simulated events and 2018 data at $\sqrt{s} = 13$ TeV. This version benefits from improved input representations, hyperparameter tuning, and a larger training dataset, resulting in significantly better background rejection for similar τ_h efficiency, particularly in $j \to \tau_h$ and $\ell \to \tau_h$ categories.

II.3.a. Inputs

The DeepTau algorithm exploits a combination of low- and high-level inputs to characterize reconstructed τ_h candidates and their surroundings. Particle-level inputs are arranged in two overlapping grids in η–ϕ space, centered on the τ_h axis (see Figure 4.5). The inner grid, covering the signal cone ($\Delta R = 0.1$), consists of 11×11 cells of size 0.02×0.02. The outer grid, covering the isolation region ($\Delta R = 0.5$), consists of 21×21 cells with a coarser resolution of 0.05×0.05 per cell.

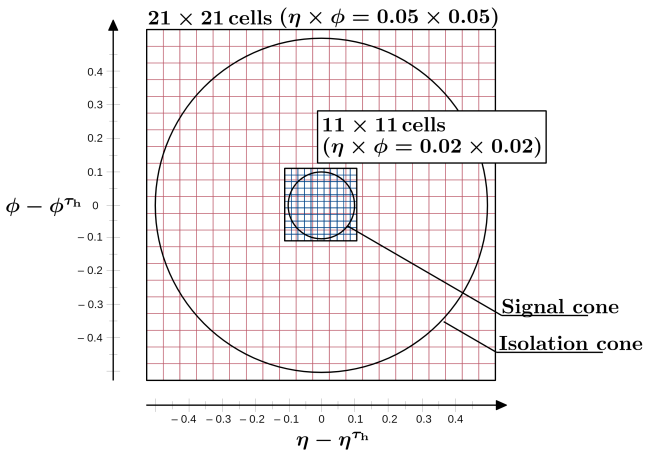


Figure 4.5: Layout of the inner and outer grids in η–ϕ space used as particle-level input to DeepTau. The inner grid covers the signal cone with radius 0.1, composed of 11×11 cells sized 0.02×0.02. The outer grid corresponds to the isolation cone with radius 0.5, composed of 21×21 cells sized 0.05×0.05 [178].

Each grid cell encodes up to seven particle types: five reconstructed by the PF algorithm (electrons, muons, photons, charged hadrons, neutral hadrons), plus standalone electrons and muons reconstructed with dedicated algorithms. Per-particle features include transverse momentum, angular distances ($\Delta\eta$, $\Delta\phi$) to the τ_h axis, electric charge, track quality variables, vertex compatibility (primary or secondary), and calorimeter energy deposits (ECAL, HCAL, muon systems). The PUPPI pileup identification score [179] is included to mitigate pileup effects.

In addition to these low-level inputs, DeepTau uses 43 high-level features. These comprise the τ_h candidate's kinematics (p_T, η, energy), charge, charged and neutral hadron multiplicities from the HPS decay mode, energy deposits from different particle types within the isolation cone, π^0 strip energy shapes in η and ϕ, pileup density observables, and electron rejection variables. Variables with known modeling discrepancies, such as the τ_h ϕ and absolute impact parameter, were

removed.

Each τ_h candidate is labeled according to its closest matching particle at generator level within a cone of radius $\Delta R < 0.2$. Candidates matched to electrons or muons with $p_T > 8$ GeV are labeled as e or μ, respectively (including leptonic tau decays). Those matched to generated τ_h with visible $p_T > 15$ GeV are labeled as genuine τ_h. All other candidates are classified as jets. Matching considers only particles from the primary proton–proton vertex.

Training uses approximately 100 million reconstructed τ_h candidates passing loose preselection: $20 < p_T < 1000$ GeV, $|\eta| < 2.5$, and longitudinal impact parameter $|dz| < 0.2$ cm. The training samples include simulated Drell–Yan (Z/γ^*+jets), $t\bar{t}$, W+jets, QCD multijets, $H \to \tau\tau$, τ-gun, and $Z' \to ee$ processes.

II.3.b. Architecture and Loss Function

DeepTau processes three inputs separately: high-level features, the inner grid (signal cone), and the outer grid (isolation cone), as shown in Figure 4.6.

Each input passes through fully connected or convolutional layers for feature extraction and dimensionality reduction (see Section I.2. for details on these layers). The grid outputs are reduced to 1×1 and concatenated with the high-level features. This combined vector is fed through fully connected layers, ending with a softmax (Section I.2.) that outputs classification probabilities:

$$\vec{y}_{\text{pred}} = (y_\tau, y_{\text{jet}}, y_e, y_\mu),$$

corresponding to genuine τ_h, jet, electron, and muon classes.

The training loss combines two components:

- **Categorical cross-entropy loss** measures how different the predicted probabilities are from the true class. It gives a higher penalty when the model is confident but wrong [180].
- **Focal loss** is a modified version of cross-entropy that reduces the impact of easy, well-classified examples. This helps the model focus more on difficult cases and improves performance when classes are imbalanced [181].

Physics discriminators are defined as

$$D_\alpha(\vec{y}) = \frac{y_\tau}{y_\tau + y_\alpha}, \quad \alpha \in \{\text{jet}, e, \mu\}.$$

Version 2.5 adds domain adaptation via a domain classifier trained alongside the main network to reduce differences between simulation and data, improving modeling in the high-score regions of D_{jet}. Details are given in the next section.

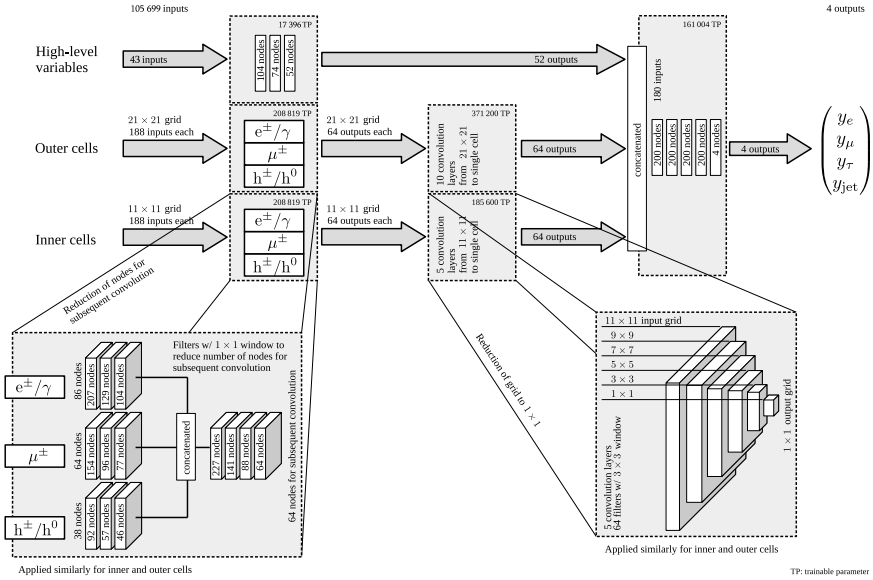


Figure 4.6: DeepTau network architecture: three input branches for high-level features, inner grid, and outer grid; merged features pass through fully connected layers for classification [35].

II.3.c. Domain Adaptation by Backpropagation

DeepTau v2.5 introduces domain adaptation [182] to reduce differences between simulated and real data. A small subnetwork shown in Figure 4.7, called a *domain classifier*, is added to the main network. It learns to identify whether a τ_h candidate comes from simulation or from real collision data. This helps the main network learn features that behave similarly in both domains, improving performance on data.

The domain classifier ends with a *sigmoid activation function* (see Section I.2.), which outputs a value $y_{\text{adv}} \in [0, 1]$, representing the likelihood that the input comes from real data (close to 1) or from simulation (close to 0). It is trained using a *binary cross-entropy loss*, defined as:

$$L_{\text{adv}} = -\left[y_{\text{true}} \log(y_{\text{adv}}) + (1 - y_{\text{true}}) \log(1 - y_{\text{adv}})\right],$$

where y_{true} is the true label (1 for data, 0 for simulation), and y_{adv} is the output of the domain classifier.

To make the shared layers learn domain-invariant features, the training uses a technique called *gradient reversal*. During backpropagation, the gradient from the adversarial loss is multiplied by -1 before being passed to the shared layers. This means the domain classifier tries to distinguish data from simulation, while

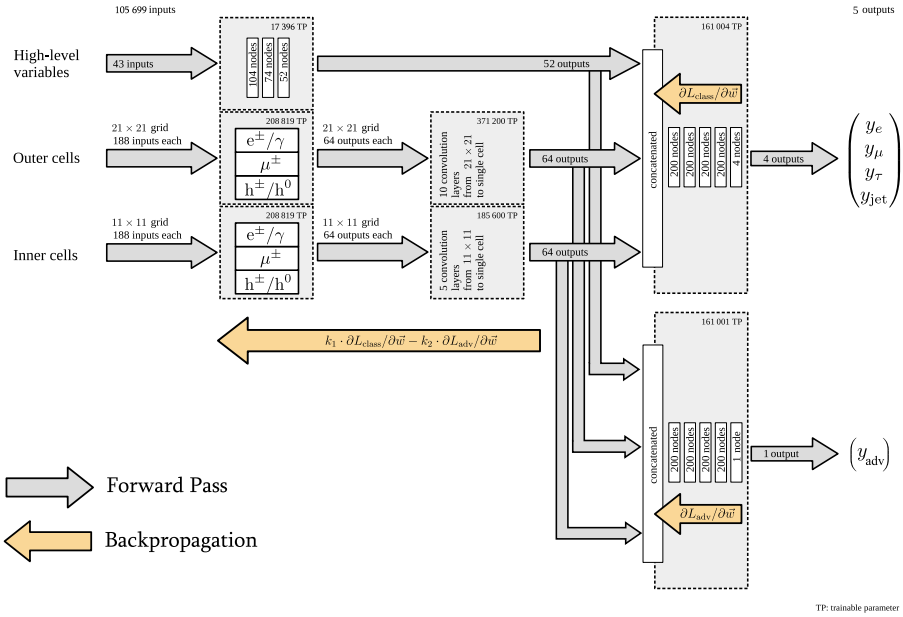


Figure 4.7: The DeepTau architecture with domain adaptation. A domain classifier is added to distinguish between data and simulation, with gradient reversal applied to the common layers during the training to promote domain-invariant features [35].

the main network learns to make them indistinguishable, a strategy known as *adversarial training*.

This method helps to reduce mismodeling effects, particularly in the high-score region of the D_{jet} distribution.

II.3.d. Expected Performance

DeepTau v2.5 demonstrates an improved performance over its predecessor across all misidentification channels while maintaining a similar or better τ_h identification efficiency. The performance was first evaluated using 2018 simulated datasets in the central pseudorapidity region ($|\eta| < 2.3$), with dedicated samples for each background source. The identification efficiency is extracted from reconstructed τ_h candidates matched to generator-level τ leptons from $H \to \tau\tau$ decays. Misidentification probabilities are evaluated using τ_h candidates that are not matched to prompt leptons or τ decay products.

A consistent improvement is observed, as shown in Figures 4.8, 4.9, and 4.10, with a notable reduction in the misidentification rates across the different background types. Specifically, the jet misidentification rate is reduced by approxi-

mately 50% across all working points. The electron rejection improves significantly, with the tightest working points showing a twofold reduction in the misidentification rate. Muon rejection remains comparable between DeepTau v2.1 and v2.5.

The relative misidentification probabilities across selected working points are summarized in Table 4.1.

Table 4.1: Target τ_h identification efficiencies for the different working points defined for the three discriminators. The target efficiencies are evaluated with the $H \to \tau\tau$ event sample for τ_h with $p_T \in [30, 70]$ GeV [35].

	VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
D_e	60%	70%	80%	90%	95%	98%	99%	99.5%
D_μ			99.5%	99.8%	99.9%	99.95%		
D_{jet}	40%	50%	60%	70%	80%	90%	95%	98%

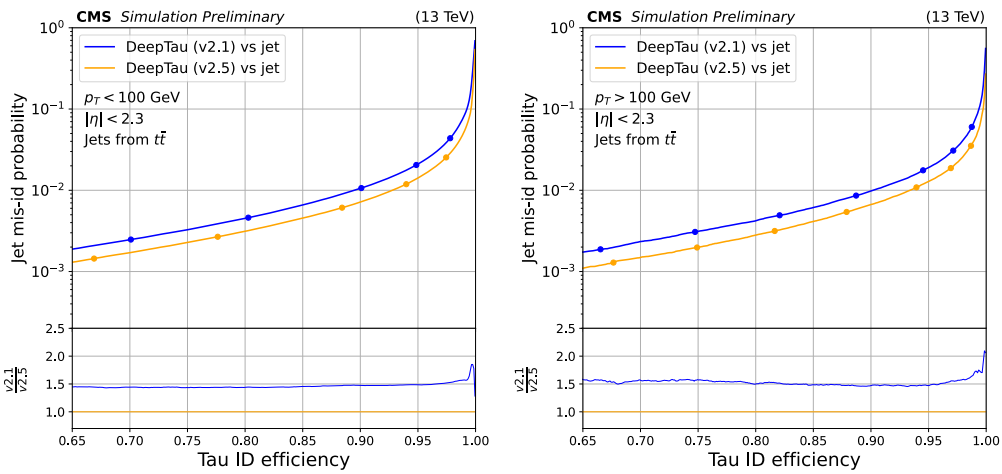


Figure 4.8: Jet misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35].

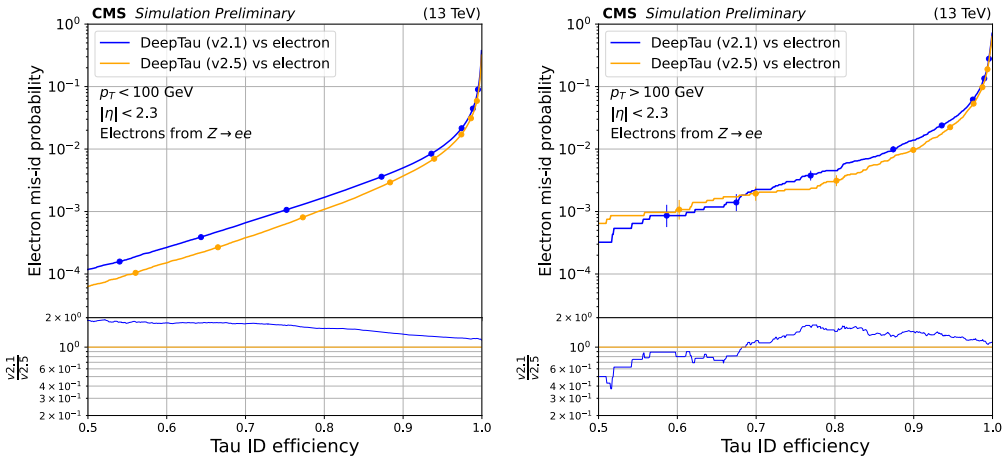


Figure 4.9: Electron misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35].

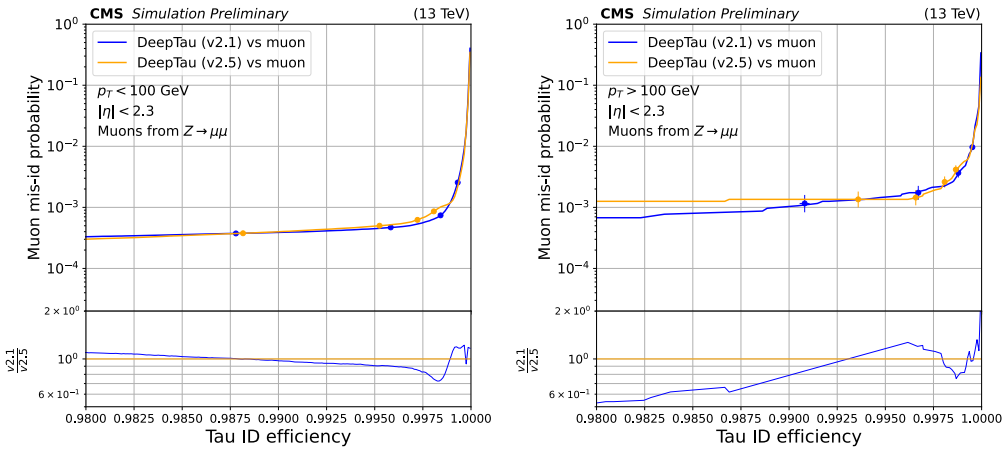


Figure 4.10: Muon misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35].

Chapter 5

Combined Measurement of the τ_h Identification Scale Factor and Energy Scale

Content

I. Baseline Event Selection with $Z \to \tau\tau$ events 124

I.1. Trigger Selections 124

I.2. Baseline Selection for the $\mu\tau_h$ Channel 125

I.3. Baseline Selection for the $\mu\mu$ Channel 126

II. Correction 127

II.1. Pileup Corrections 127

II.2. Electrons and Muons Efficiencies 127

II.3. Z p_T Corrections 128

III. Background Estimation 128

III.1. Background from QCD Multijet Processes 128

III.2. Normalization of the $W +$ jets Background 129

IV. Combined Measurement of τ_h Identification and Energy Scale Factors 129

IV.1. Measurement Technique 129

IV.2. Fit Configuration 131

IV.3. Method Validation with the UL2018 Dataset 131

IV.3.a. Simultaneous fit of the TES and the ID SF per decay modes 132

IV.3.b. Combined fit of the TES and the ID SF by decay modes and $p_T(\tau_h)$ 139
IV.4. Combined Measurements in the 2022 Run 3 Datasets . . 147
IV.4.a. 2022 preEE Measurements 147
IV.4.b. 2022 postEE Measurements 151
IV.5. Comparison with Individual Fit Results 153
V. Conclusion 153

This chapter details the combined measurement of the τ_h Identification Scale Factor (ID SF) and the τ_h Energy Scale (TES), a novel technique developed within this thesis. The TES is defined as the factor applied to the reconstructed τ_h energy in simulation to bring it into agreement with the τ_h energy observed in data. The τ_h identification efficiency scale factor is defined as the ratio of the number of τ_h candidates passing the selection criteria in data to that in simulation. Accurate determination of these scale factors (SFs) is essential for the proper calibration of both τ_h identification efficiency and energy reconstruction. The method is first validated using data from Run 2, followed by its application to the 2022 Run 3 datasets. Finally, the performance of the combined measurement technique is compared to that of the standalone method, where the scale factors are extracted from separate fits [183].

I. Baseline Event Selection with $Z \to \tau\tau$ events

This section describes the baseline event selection employed for the extraction of the SFs in $Z \to \tau\tau$ events. The measurement utilizes different final states: the $\mu\tau_h$ channel is used to determine the τ_h energy scale and identification efficiency, while the $\mu\mu$ channel serves as a control region to constrain the $Z/\gamma^* \to \mu\mu$ background.

I.1. Trigger Selections

Events are selected using High-Level Trigger (HLT) paths, specifically single-muon triggers in both the $\mu\tau_h$ and $\mu\mu$ final states. These triggers target events containing isolated, well-identified muons. The relevant HLT paths and their corresponding offline transverse momentum (p_T) thresholds for Run 2 and Run 3 datasets are summarized in Table 5.1.

Since the trigger efficiency exhibits a gradual turn-on behavior around the nominal threshold (the trigger turn-on region), a slightly higher offline p_T requirement is applied to ensure that selected events lie within the efficiency plateau, where the trigger response is well-modeled. For each event, the offline p_T threshold is

matched to the lowest-threshold trigger that fired. Trigger efficiency corrections are applied to account for differences between data and simulation.

The details of the HLT paths and their performance are described in the CMS Trigger documentation [184, 185].

Table 5.1: Single-muon HLT paths and corresponding offline p_T selection thresholds for Run 2 and Run 3 datasets.

Channel	**HLT Path**	**Offline p_T cut [GeV]**
$\mu\tau_h$ and $\mu\mu$	`HLT_IsoMu24`	25
	`HLT_IsoMu27`	28

I.2. Baseline Selection for the $\mu\tau_h$ Channel

The selection of $Z \to \mu\tau_h$ events requires exactly one muon with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.4$, passing a relative isolation requirement of $I_{\text{rel}} < 0.15$. Standard impact parameter selections, $|d_{xy}| < 0.045\,\text{cm}$ and $|d_z| < 0.2\,\text{cm}$, are applied. One τ_h candidate is required with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$, reconstructed in decay modes corresponding to DM0, DM1, DM10, or DM11.

The muon and τ_h candidate must be separated by $\Delta R > 0.5$ and have opposite-sign (OS) electric charges. In events with multiple τ_h candidates, the one with the highest p_T is selected. To suppress backgrounds from Z+jets, $t\bar{t}$, and diboson processes, events containing additional isolated leptons are rejected. Specifically, events with extra electrons or muons that satisfy loose identification, isolation, and standard impact parameter requirements, without overlapping with previously selected leptons or τ_h candidates, are vetoed if these leptons have $p_T > 10\,\text{GeV}$ for electrons and $p_T > 15\,\text{GeV}$ for muons. Events containing an OS dielectron or dimuon pair with $\Delta R > 0.15$ are also rejected. These criteria are summarized in Table 5.2.

To further reduce contamination from $Z \to ee$ and $Z \to \mu\mu$ events, DeepTau discriminators are applied. The Very-Very-Loose (VVLoose) working point (WP) of the electron discriminator (D_e) and the Tight WP of the muon discriminator (D_μ) are used.

The muon, electron, and hadronic tau reconstruction and identification criteria used for Run 2 correspond to those described in Refs. [33, 186, 187]. For Run 3, updated performance results and object definitions follow the latest CMS public results [188, 189].

Table 5.2: Summary of selection criteria for extra lepton vetoes applied to suppress background contamination.

Criteria	Extra electron	Extra muon	Dimuon
p_T	> 10 GeV	> 10 GeV	> 15 GeV
$\|\eta\|$	< 2.5	< 2.4	
$\|d_{xy}\|$	< 0.045 cm	< 0.045 cm	
$\|d_z\|$	< 0.2 cm	< 0.2 cm	
Isolation	$I_e < 0.10$ (ρ-corrected)	$I_\mu < 0.15$ ($\Delta\beta$-corrected)	
Identification	`MVAFall2017V2noIso` 90% WP	Medium ID	Global, tracker and PF muon
Other	Conversion veto ≥ 1 missing inner hits		Opposite-sign $\Delta R(\mu\mu) > 0.15$

I.3. Baseline Selection for the $\mu\mu$ Channel

The $\mu\mu$ control region is primarily used to constrain the Drell–Yan ($Z/\gamma^* \to \mu\mu$) background normalization. Events are required to contain exactly two muons satisfying the following criteria: the leading muon must have a transverse momentum $p_T > 25\,\text{GeV}$ and pseudorapidity $|\eta| < 2.4$, while the sub-leading muon must satisfy $p_T > 15\,\text{GeV}$ and $|\eta| < 2.3$. Both muons are required to be isolated, with relative isolation $I_{\text{rel}} < 0.15$, and to pass standard impact parameter selections to suppress non-prompt backgrounds [186].

Only the leading muon is required to match the HLT trigger object, ensuring trigger efficiency modeling consistency. The muon object definitions are kept consistent with those used in the $\mu\tau_h$ channel. The muon pair must be of opposite electric charge and satisfy a spatial separation requirement of $\Delta R > 0.5$. To reduce contamination from $t\bar{t}$, Z+jets, and diboson processes, events with additional isolated leptons are vetoed following the criteria detailed in Section I.2.

II. Correction

Accurate simulation of high-energy physics processes requires applying corrections for known differences between data and MC predictions. These discrepancies can arise from imperfect modeling of pileup interactions, detector response, object reconstruction and identification efficiencies, as well as inaccuracies in kinematic distributions. To mitigate these effects, data-driven SFs are derived and applied to the simulated events.

This section summarizes the principal corrections applied to simulated samples used in the extraction of the τ_h SFs. These include PU reweighting to match the observed number of simultaneous interactions per bunch crossing and additional object- or event-level SFs for muons, electrons, jets, and missing transverse energy.

II.1. Pileup Corrections

PU reweighting is applied to simulated events to ensure that the distribution of the number of PVs per event matches that observed in data. This correction improves the agreement in particle multiplicities and energy deposits across sub-detectors between simulation and real data.

The reweighting is performed using the true pileup reweighting method, which relies on the true number of interactions per bunch crossing. In simulation, this is given by the mean of the Poisson distribution used to sample the pileup. The reweighting is performed in 100 bins over the range of 0 to 100, aligning the simulated pileup distribution to that of the data.

In Run 3, higher instantaneous luminosities and changes in detector conditions have required refined pileup mitigation strategies compared to Run 2, Run 3 features updated calibrations and improved pileup mitigation techniques, such as the optimized *PUPPI* (PileUp Per Particle Identification) algorithm [179], which enhances jet energy resolution and the performance of missing transverse momentum reconstruction in high-pileup environments.

These pileup correction procedures and mitigation strategies are detailed in dedicated CMS performance studies, including Refs. [190, 191].

II.2. Electrons and Muons Efficiencies

Lepton identification and isolation efficiencies are derived from data and simulation using a tag-and-probe method in $Z \to \ell^+\ell^-$ events [192, 193]. The tag lepton must fulfill kinematic, identification, and isolation criteria and match a trigger object, while the probe lepton is required to satisfy basic kinematic selections. The efficiency is defined as the fraction of probes passing the identification and isolation requirements, extracted via a fit to the dilepton invariant mass around

the Z peak. SFs, defined as the ratio $\epsilon_{\text{data}}/\epsilon_{\text{sim}}$, are computed in bins of lepton p_T and $\|\eta\|$, and are applied as event weights.

II.3. Z p_T Corrections

The transverse momentum (p_T) spectrum of the dilepton system in Drell–Yan ($Z \to \ell^+\ell^-$) events is known to be imperfectly modeled in simulation, especially at low p_T. To improve the agreement between data and simulation, a reweighting procedure is applied based on the Z boson p_T distribution.

The reweighting factors are derived in a dedicated control region enriched in $Z \to \mu^+\mu^-$ events, selected by requiring two oppositely charged muons with $p_T >$ 15 GeV that pass medium identification and isolation criteria. The leading muon must satisfy a p_T threshold above the trigger requirement, and events containing additional leptons are vetoed to reduce backgrounds.

Weights are computed as a function of generator-level dilepton p_T and invariant mass ($m_{\mu\mu}$) by comparing simulation to data after subtracting background contributions. The resulting weights are normalized to preserve the overall Drell–Yan yield, ensuring the correction affects only the shape of the p_T spectrum. This procedure is performed separately for each data-taking year.

The performance of this correction is validated by comparing the post-reweighting distributions of $p_T^{\mu\mu}$ in simulation to data, demonstrating improved modeling across the relevant kinematic range.

III. Background Estimation

The estimation of background contributions relies primarily on simulated samples for all SM processes, except for the QCD multijet background, which is evaluated using data-driven techniques, as described in Section III.1. The normalization of the W+jets background is derived from a control region enriched in W+jets events, defined by a high transverse mass ($m_T(MET, \mu)$) requirement. Details of this sideband method are provided in Section III.2.

III.1. Background from QCD Multijet Processes

The contribution from QCD multijet processes is estimated using a data-driven technique, as simulation of such events is computationally expensive and less reliable due to challenges in modeling hadronization and jet fragmentation.

This analysis employs the same-sign (SS) method, which exploits the charge symmetry of QCD processes. Events with two same-sign leptons are selected to construct a QCD-enriched control region, in contrast to the opposite-sign (OS)

selection used in the signal region. The QCD yield in the OS region is then extrapolated from the SS region using the relation:

$$N_{\mathrm{OS}}^{\mathrm{QCD}} = F \times \left(N_{\mathrm{SS}}^{\mathrm{data}} - N_{\mathrm{SS}}^{\mathrm{MC}}\right), \tag{5.1}$$

where $N_{\mathrm{SS}}^{\mathrm{MC}}$ represents the non-QCD background contributions in the SS region, estimated from simulation, and F is a transfer factor correcting for the OS/SS asymmetry in QCD events.

The factor F is determined in a sideband of lepton isolation, specifically in the range $0.15 < I_\ell < 0.50$, and is found to be $F = 1.10$ in semileptonic final states. The QCD estimate is derived bin-by-bin for each observable, with uncertainties propagated from the statistical errors in data and simulation.

III.2. Normalization of the $W+$jets Background

The contribution from $W+$jets events is estimated using simulation, but its normalization is derived from data via a control region enriched in $W+$jets. This region is defined by inverting the transverse mass requirement on the lepton and the missing transverse energy. The transverse mass is computed as:

$$m_T(MET, \mu) = m_T = \sqrt{2\, p_T^\ell\, E_T^{\mathrm{miss}}\, (1 - \cos\Delta\phi)}, \tag{5.2}$$

where p_T^ℓ is the transverse momentum of the lepton, E_T^{miss} is the missing transverse energy, and $\Delta\phi$ is the azimuthal angle between the two.

After the baseline event selection described in Section I., a sideband region is defined by requiring $m_T > 80$ GeV. The normalization of the $W+$jets simulation is corrected by a SF derived in this sideband such that the yield in simulation matches the number of observed events in data. This correction accounts for potential mismodeling of the $W+$jets background normalization in the signal region.

IV. Combined Measurement of τ_h Identification and Energy Scale Factors

IV.1. Measurement Technique

In this new measurement approach, the TES and ID SF are determined simultaneously using a tag-and-probe method in $Z \to \tau_\mu \tau_h$ events. Events are selected with one well-identified and isolated muon (the "tag") and one τ_h candidate (the "probe"), following the baseline preselection described in Section I. The dataset used for this measurement includes proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$

collected during Run 2 (2018) and $\sqrt{s} = 13.6\,\text{TeV}$ data from Run 3 (2022), corresponding to integrated luminosities of $59.8\,\text{fb}^{-1}$ (UL2018), $7.98\,\text{fb}^{-1}$ (2022 preEE), and $26.67\,\text{fb}^{-1}$ (2022 postEE). The Run 3 2022 dataset is split into two periods, referred to as *preEE* and *postEE*, corresponding to data collected before and after the replacement of the CMS electromagnetic calorimeter (ECAL) endcap modules.

The TES and ID SF are extracted from a maximum likelihood fit to the visible mass distribution:

$$m_{\text{vis}} = \sqrt{(E_{\tau_h} + E_\mu)^2 - |\vec{p}_{\tau_h} + \vec{p}_\mu|^2}, \tag{5.3}$$

where E_{τ_h} and E_μ are the reconstructed visible energies, and $\vec{p}_{\tau_h}$, $\vec{p}_\mu$ are the momenta of the τ_h and muon, respectively. The fit is performed in different regions, defined either by the τ_h decay mode (DM) or by a combination of DM and the τ_h transverse momentum (p_T), referred to as DM-p_T regions.

In previous analyses, TES and ID SF were fitted separately on the m_{vis} distribution in the signal region, with one scale factor treated as the parameter of interest (POI) and the other as a nuisance parameter (NP).

The new technique performs a combined fit of TES and ID SF simultaneously on the same distribution, taking into account correlations and systematic uncertainties. The likelihood for each region r can be expressed as:

$$L_r(TES_r, IDSF_r, \theta_r) = p(\text{data}|TES_r, IDSF_r, \theta_r) \prod_{i=1}^{n} p(\overline{\theta_i}|\theta_{i,r}), \tag{5.4}$$

where $\theta_r = \{\theta_{i,r}\}_{i=1}^n$ is the set of *nuisance parameters* (NP) describing systematic uncertainties, constrained by auxiliary measurements $p(\overline{\theta_i}|\theta_{i,r})$. The index r denotes the fitted regions described in Table 5.3.

The combined fit is initially performed independently for each region r. Alternatively, a simultaneous fit across all regions can be carried out by defining TES and ID SF in all regions as *parameters of interest* (POIs). POIs are the parameters that the fit aims to measure directly, representing the physical quantities of interest, in this case, the TES and ID SF. All other parameters, such as systematic uncertainties, are treated as NPs and are profiled or constrained during the fit. A comparison of a fit across all region and separately is presented in Section IV.3.a.

While both scale factors are POIs, they affect the m_{vis} distribution differently: the ID SF modifies the normalization of the $Z \to \tau\tau$ (ZTT) signal, whereas the TES changes the shape of the distribution. Templates of the m_{vis} distribution for ZTT events are produced with TES varied in discrete steps of 0.1% over a range of -3% to 3% for fits performed by DM, and -5% to 5% for DM-p_T fits. This range covers the $\pm 1\sigma$ TES uncertainty for the UL2018 and 2022 datasets. Horizontal integral morphing [34] is used to interpolate between TES-varied templates; some instabilities of this method are discussed in Section IV.3.a.

This technique was developed and validated using the UL2018 dataset, and applied to the 2022 preEE and postEE datasets. The fitted regions per dataset are summarized in Table 5.3.

Table 5.3: Fitted regions for each dataset. For UL2018, two different fits have been performed.

Dataset	Fitted regions	Number of regions
UL2018	DM	4
	DM-p_T: $p_T(\tau_h) = [20, 40], [40, 200]$ GeV	8
2022 preEE	DM	4
2022 postEE	DM	4

IV.2. Fit Configuration

The fit includes two POIs: the TES and the ID SF and Systematic uncertainties are incorporated as NPs, either affecting the normalization (log-normal, lnN) or the shape of the m_{vis} distribution. Shape uncertainties are modeled using vertical template morphing between nominal, upward, and downward variations. Log-normal uncertainties are implemented using probability density functions following a log-normal distribution. A summary of the fit parameters and their associated uncertainties is given in Table 5.4.

IV.3. Method Validation with the UL2018 Dataset

This section presents a validation of the combined fit method using the UL2018 dataset, one of the largest and most reliable datasets from Run 2. The goal is to test and optimize the method under controlled conditions.

This study focuses on the stability of the fit, the optimization of selection criteria, and the identification of limitations inherent to the method. In particular, the approach is first validated in τ_h DM regions only (see Section IV.3.a.). Motivated by the fact that ID SFs are typically measured as a function of p_T, the analysis is later extended to DM–p_T regions (see Section IV.3.b.). This more granular binning reveals limitations of the method, which are investigated in detail to understand their origin and propose potential improvements.

In this section, the ID SF and the TES are determined for the following working points: D_{jet} Medium, D_e VVLoose, and D_μ Tight. Additionally, a cut on

Table 5.4: Summary of fit parameters and associated uncertainties.

Nuisance Parameter	**Type**	**Uncertainty**	**Processes Affected**
Integrated Luminosity	lnN	2.5%	All except QCD
Muon Efficiencies	lnN	2.0%	All except QCD
$t\bar{t}$ Cross Section	lnN	6.0%	TT
VV and ST Cross Sections	lnN	5.0%	VV, ST
QCD Normalization	lnN	10.0%	QCD
$j \to \tau_h$ Fake Rate	lnN	15.0%	W, QCD, ZJ, TTJ, STJ
$j \to \tau_h$ Energy Scale	Shape	5.0%	ZJ, W, TTJ
$\ell \to \tau_h$ Fake Rate	Shape	1 s.d.	ZL, TTL
$\mu \to \tau_h$ Energy Scale	Shape	2.0%	ZL, TTL
Z p_T Reweighting	Shape	10.0%	DY
Bin-by-bin Stats	Shape	–	All
DY Normalization	Rate Param	–	ZTT, ZL, ZJ
W+jets Normalization	Rate Param	–	W
ID SF	POI	30.0%	ZTT
TES	POI	5.0%	ZTT

$m_T(\text{MET}, \mu) < 65$ GeV is applied to enhance the significance of the $S/\sqrt{S+B}$ ratio.

IV.3.a. Simultaneous fit of the TES and the ID SF per decay modes

The combined measurement of TES and ID SF is performed separately for each DM. The $-2\Delta \ln L$ profiles of the TES are shown in Figure 5.3, with the minimum TES value subtracted. To mitigate fluctuations observed in some profiles, asymmetric parabolae (in red or black) are fitted to the likelihood values below 5, using the function:

$$f(\text{tes}) = \begin{cases} a_L(\text{tes} - \text{tes}_{\text{min}})^2 + c & \text{for tes} < \text{tes}_{\text{min}}, \\ a_R(\text{tes} - \text{tes}_{\text{min}})^2 + c & \text{for tes} \geq \text{tes}_{\text{min}}. \end{cases} \tag{5.5}$$

The fit has four parameters: left and right curvatures a_L and a_R, the TES best fit tes_{min}, and a vertical offset c, which is kept small. The best-fit TES is extracted at the parabola minimum, and the $\pm 1\sigma$ uncertainty is defined by the TES values satisfying $f(\text{tes}) = 1 + c$. The same procedure is applied to the ID SF. The results are summarized in Figure 5.5.

The results show that the TES values are consistent with unity across all DM regions within uncertainties. In contrast, the ID SF exhibits a clear trend as a function of DM. Specifically, DMs corresponding to higher track multiplicities (e.g., three-prong decays) tend to have ID SF values further from unity. This behavior can be attributed to a decrease in the reconstruction and identification efficiency as the number of decay products increases, likely due to increased complexity in track reconstruction and neutral pion identification in the hadronic τ decay.

Pre-fit and post-fit distributions of the visible mass (m_{vis}) for each DM are presented in Figures 5.1 and 5.2. The post-fit distributions are obtained by scaling the MC templates according to the measured TES, ID SF, and log-normal (lnN) nuisance parameters. The good agreement observed between data and the post-fit MC predictions demonstrates the reliability of the measurements and validates the fit model across all decay mode regions.

To ensure the robustness and reliability of the TES and ID SF measurements, a series of comprehensive stability tests have been performed in the $\mu\tau$ final state. These tests evaluate the sensitivity of the results to key analysis choices and systematic uncertainties, thereby probing the potential biases and limitations of the fitting procedure. Specifically, the stability checks address:

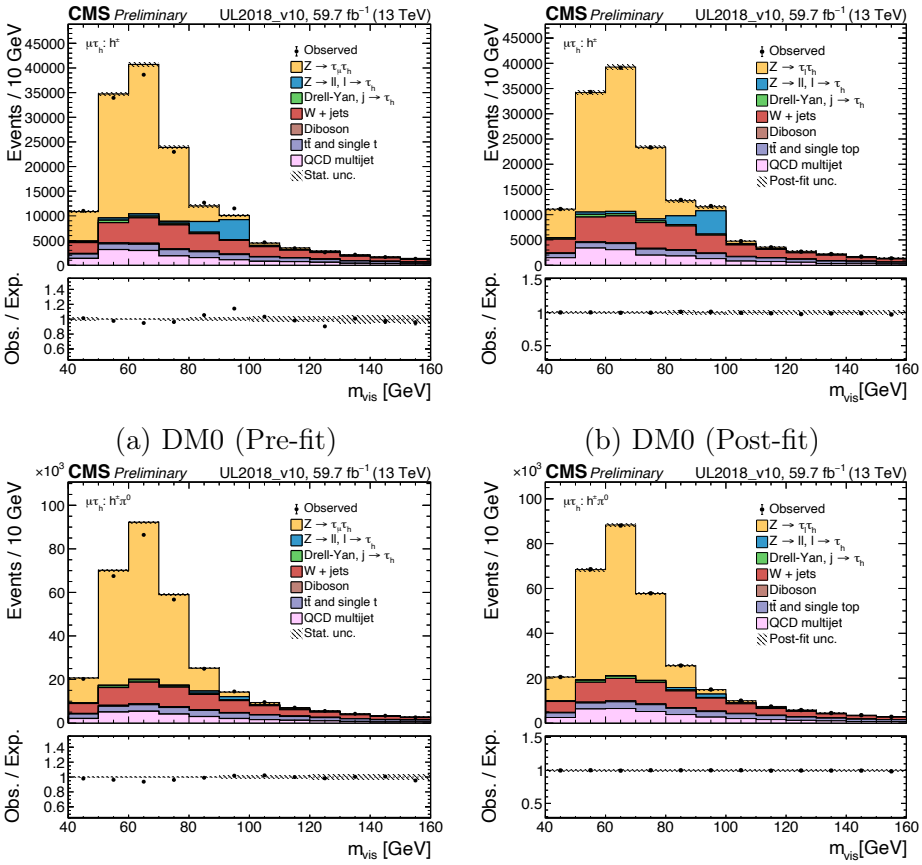
- **Impact of the DY control region (CR):** The Drell–Yan (DY) control region provides crucial constraints on the DY background normalization and shape. To evaluate its influence, the DY CR is removed from the fit, and instead a systematic uncertainty on the DY cross section is introduced. The TES measurement is then repeated, resulting in variations within 0.3%, indicating a limited dependence on the DY CR.

- **Effect of binning choices:** The visible mass (m_{vis}) is the primary discriminant in the fit. To assess the impact of binning granularity, three binning schemes (5, 10, and 15 bins) are tested over the range [50, 150] GeV. The observed TES variations remain within 1%, demonstrating the stability of the fit against binning-related statistical fluctuations. Coarser binning reduces statistical fluctuations but slightly increases the fit uncertainty by about 0.02%.

- **Variation of the fitting range:** The fit range is extended by 10 and 40 GeV at both lower and upper boundaries to check for potential biases induced by the choice of range. The resulting TES values vary by less than 0.5%,

confirming that the measurement is stable against reasonable modifications of the fit window.

- **Simultaneous fit across all decay modes:** To exploit correlations and common systematic uncertainties across different τ_h decay modes, a simultaneous fit defining separate TES and ID SF parameters per decay mode is performed. Common nuisance parameters (e.g., DY and QCD normalizations, VV and $t\bar{t}$ cross sections, muon efficiency, luminosity, and Z $p_T(\tau_h)$ reweighting) are shared across regions, while others related to fake rates and energy scales are treated separately. This approach tests the consistency of TES and ID SF across modes and leverages additional constraints from combined fits. The results obtained from the simultaneous fit across all the regions are found to be compatible with those from other methods. Notably, the simultaneous fit of the DMs has a minimal impact on the ID SF, with changes of less than 0.2% observed. However, this method is not employed as the default approach due to the instability issues encountered in the fit when profiling the TES. The TES profiling relies on morphing, which has exhibited certain instabilities during this study. Therefore, while the simultaneous fit approach is viable and yields compatible results, it is not the default method due to the identified instabilities during the TES profiling.



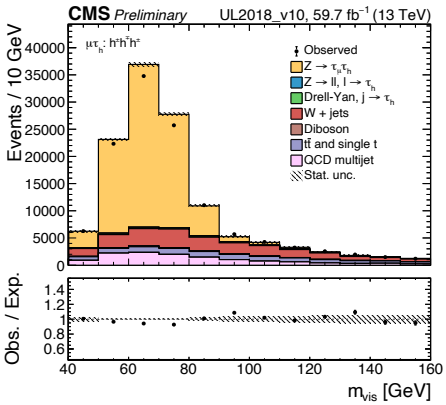
(a) DM0 (Pre-fit)

(b) DM0 (Post-fit)

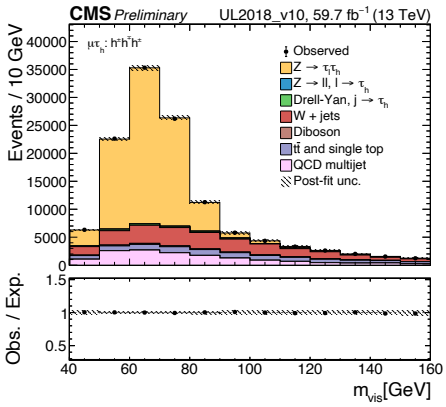
(c) DM1 (Pre-fit)

(d) DM1 (Post-fit)

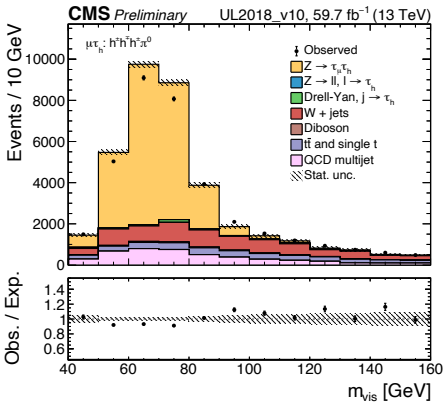
Figure 5.1: Pre-fit (left) and post-fit (right) distribution of m_{vis} for UL2018 for DM0 and DM1.



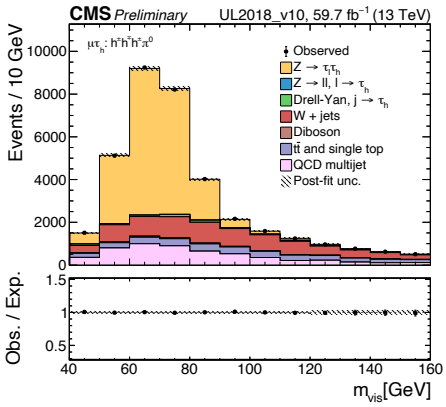
(a) DM10 (Pre-fit)



(b) DM10 (Post-fit)



(c) DM11 (Pre-fit)



(d) DM11 (Post-fit)

Figure 5.2: Pre-fit (left) and post-fit (right) distribution of m_{vis} for UL2018 for DM10 and DM11.

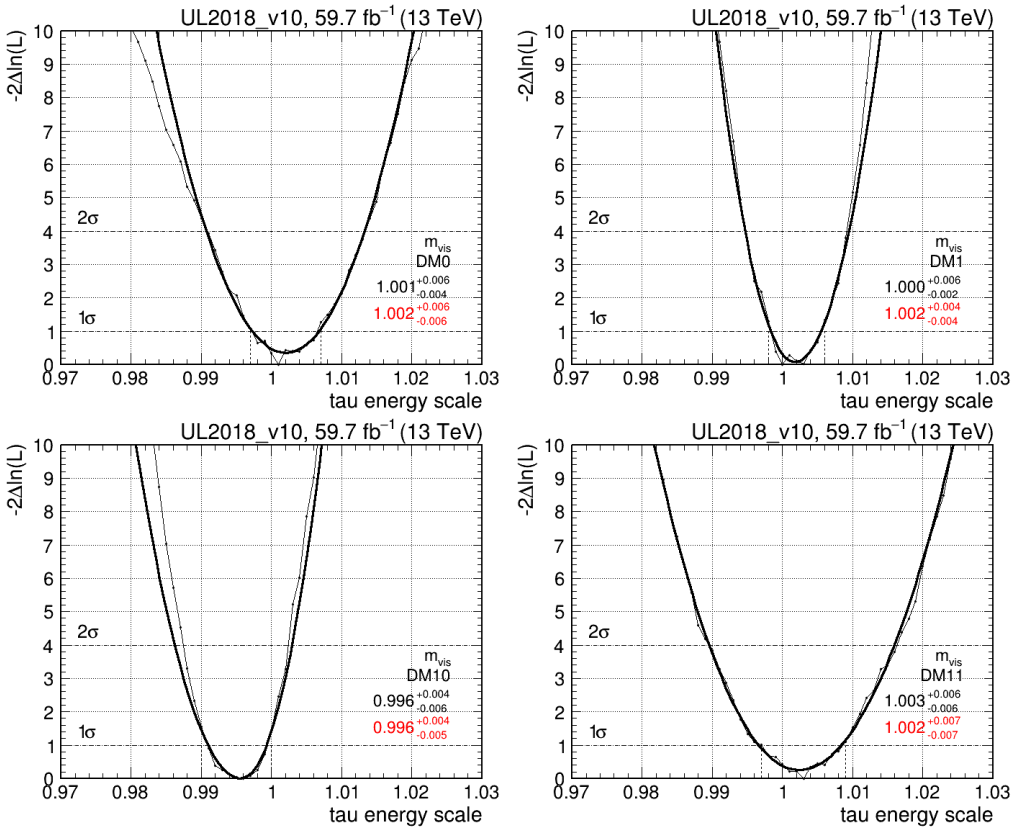


Figure 5.3: Negative log-likelihood profiled as a function of the TES, measured from the m_{vis} observable. Asymmetric parabolas (black line) are fitted to the $-2\Delta(\ln L)$ values below 5. The plots are arranged as follows: top-left for DM0, top-right for DM1, middle-left for DM10, and middle-right for DM11.

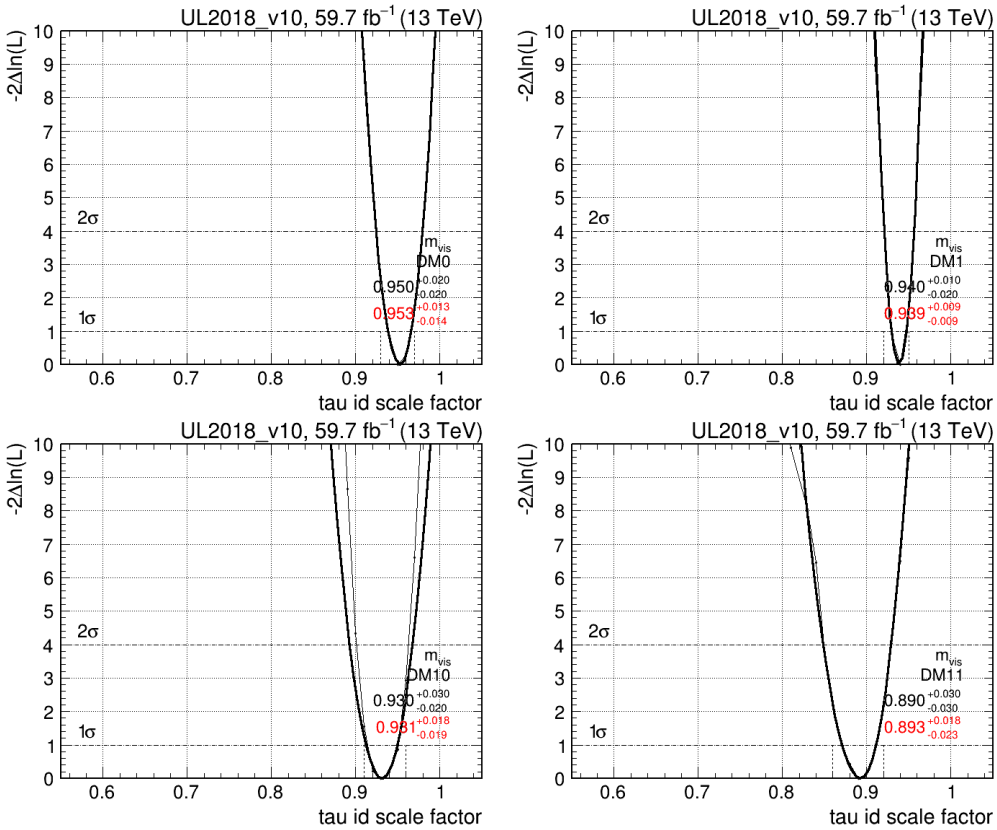


Figure 5.4: Negative log-likelihood profiled as a function of the ID scale factor (ID SF), measured from the m_{vis} observable. Asymmetric parabolas (black line) are fitted to the $-2\Delta(\ln L)$ values below 5. The plots are arranged as follows: top-left for DM0, top-right for DM1, middle-left for DM10, and middle-right for DM11.

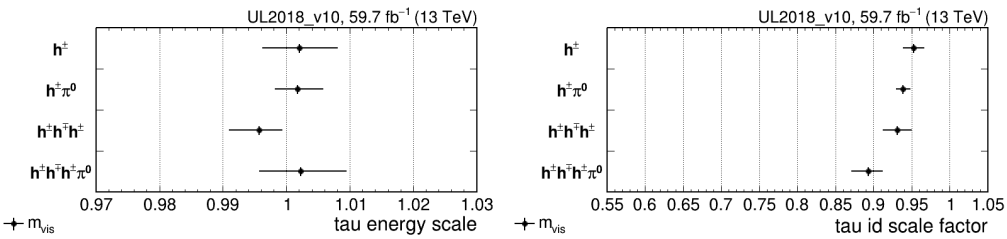


Figure 5.5: Summary of TES (left) and ID SF (right) per DM region for UL2018, obtained from parabolic fits.

IV.3.b. Combined fit of the TES and the ID SF by decay modes and $p_T(\tau_h)$

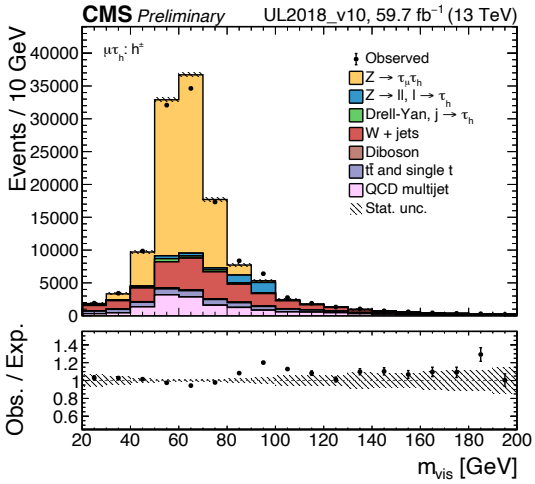
To account for variations in TES and ID SF across different τ_h decay modes and transverse momentum $p_T(\tau_h)$, a combined fit is performed considering both variables simultaneously. Pre-fit m_{vis} distributions per DM–p_T region are shown in Figures 5.6, 5.7, 5.8, and 5.9.

The $p_T(\tau_h)$ distributions for each DM, displayed in Figure A.1 of the appendix, are used to determine appropriate $p_T(\tau_h)$ regions. After testing various binning schemes, a division into [20, 40] GeV and [40, 200] GeV regions is chosen, providing sufficient statistics for a stable fit while minimizing fluctuations. Although a more refined binning, similar to that used in the individual ID SF fit [183], is technically feasible and has been tested, limited statistics lead to significant instabilities in the TES fit. Therefore, this analysis presents results based on the two-region $p_T(\tau_h)$ division.

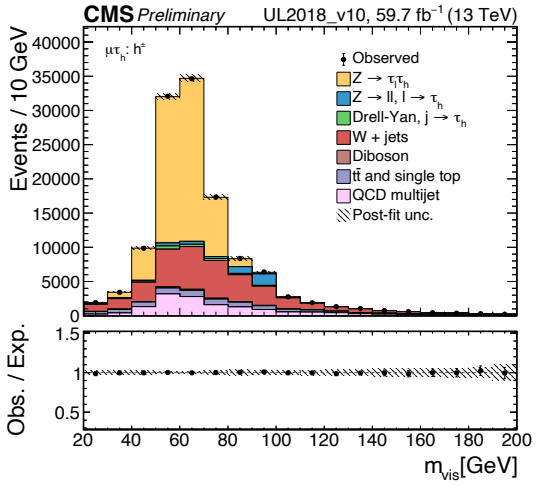
The scanned scale factors are shown in Figure 5.11, while the TES results are presented in Figure 5.10. A summary plot of the results as a function of $p_T(\tau_h)$ is provided in Figure 5.12. For the DM0 in the high $p_T(\tau_h)$ region, the NLL profile exhibits fluctuations and underestimates the uncertainties. To address this, the uncertainties have been symmetrized conservatively to the larger upper value.

The TES and ID SF fluctuations primarily arise from the limited size of the simulated samples. These fluctuations may be reduced in future analyses with larger datasets. The presence of fluctuations underscores the necessity of applying the parabolic fit.

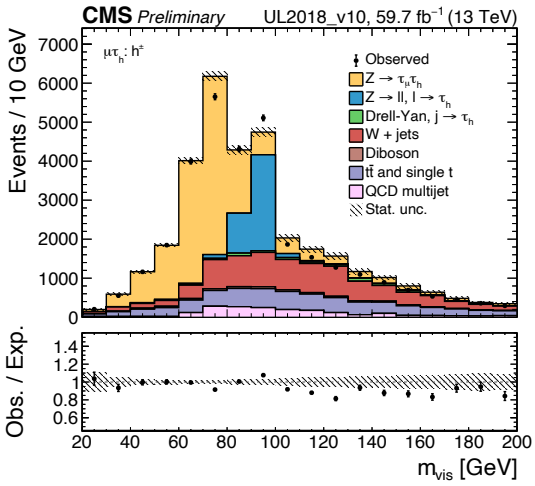
The dependency between the scale factors is illustrated in Figure A.2. At low $p_T(\tau_h)$ (20–40 GeV), the ID SF and TES show low dependency, while at high $p_T(\tau_h)$ (>40 GeV) they exhibit strong dependency. This arises from bin migration effects as TES varies, impacting normalization and creating dependency between the TES and ID SF. This study highlights the importance of performing a simultaneous fit of both scale factors. However, the limited statistics constrain the number of $p_T(\tau_h)$ bins, a limitation that could be overcome with increased event sample size.



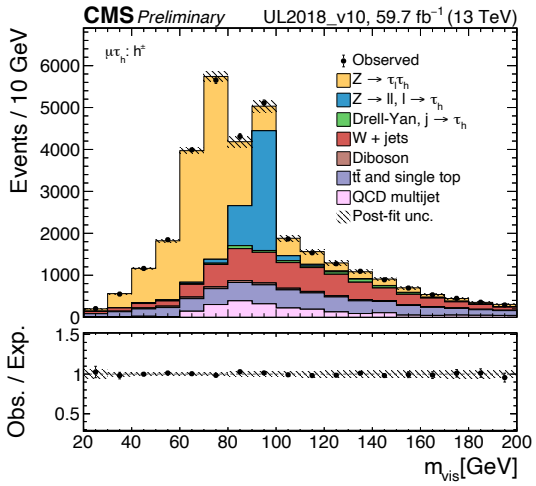
(a) DM0 $20 < p_T(\tau_h) \leq 40$ GeV (Pre-fit)



(b) DM0 $20 < p_T(\tau_h) \leq 40$ GeV (Post-fit)

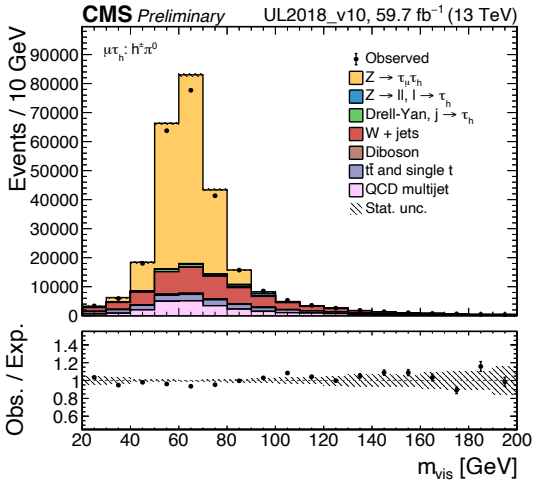


(c) DM0 $40 < p_T(\tau_h) \leq 200$ GeV (Pre-fit)

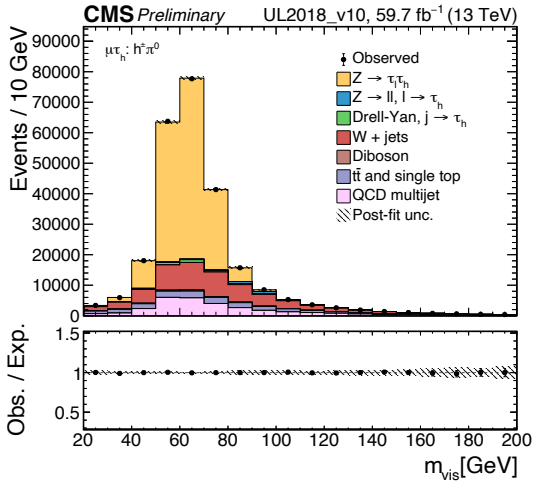


(d) DM0 $40 < p_T(\tau_h) \leq 200$ GeV (Post-fit)

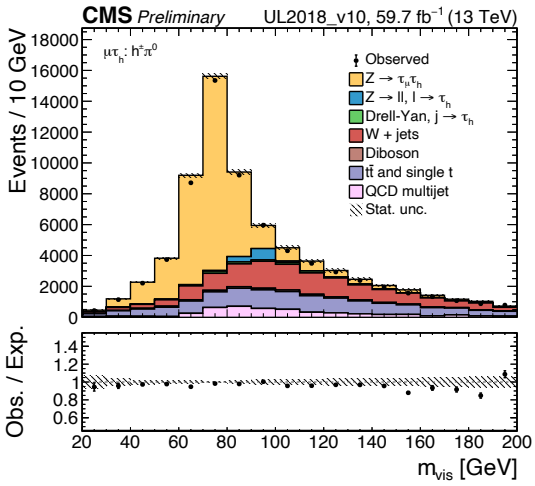
Figure 5.6: Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM0-p_T regions for UL2018.



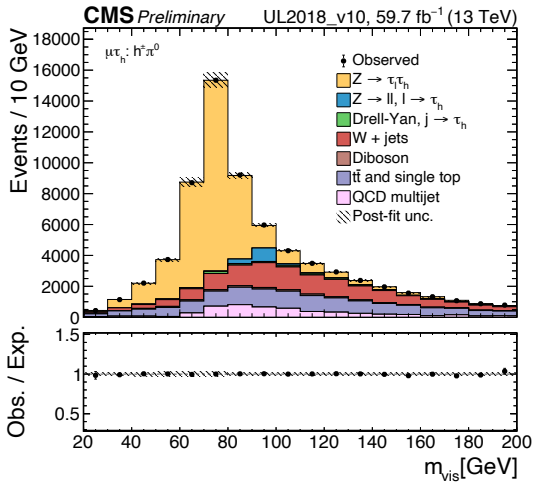
(a) DM1 $20 < p_T(\tau_h) \leq 40$ GeV (Pre-fit)



(b) DM1 $20 < p_T(\tau_h) \leq 40$ GeV (Post-fit)

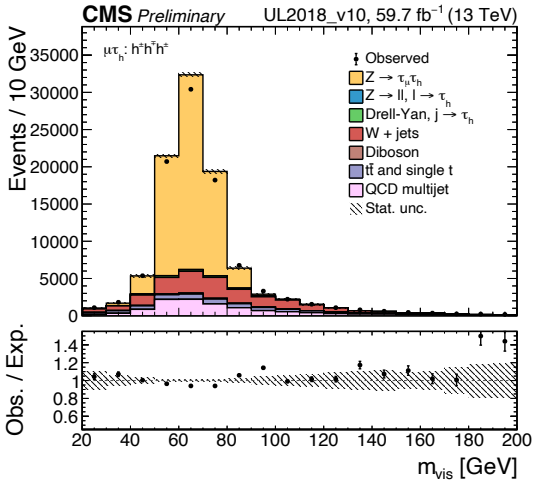


(c) DM1 $40 < p_T(\tau_h) \leq 200$ GeV (Pre-fit)

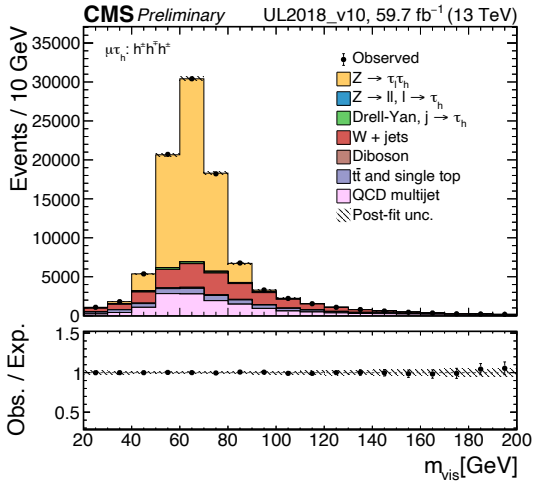


(d) DM1 $40 < p_T(\tau_h) \leq 200$ GeV (Post-fit)

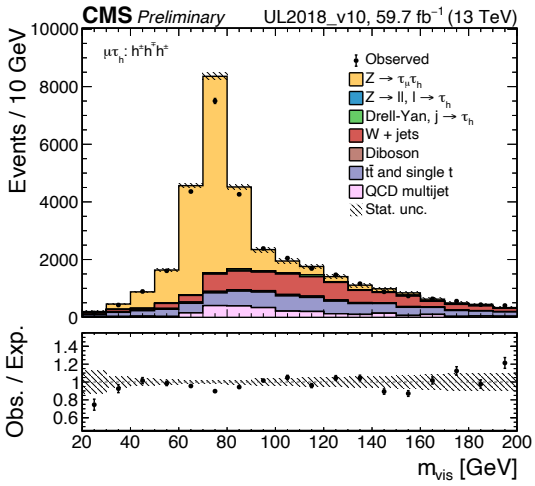
Figure 5.7: Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM1-p_T regions for UL2018.



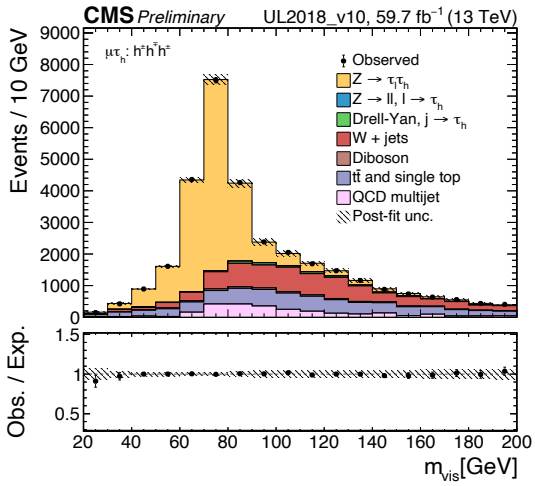
(a) DM10 $20 < p_T(\tau_h) \leq 40$ GeV (Pre-fit)



(b) DM10 $20 < p_T(\tau_h) \leq 40$ GeV (Post-fit)

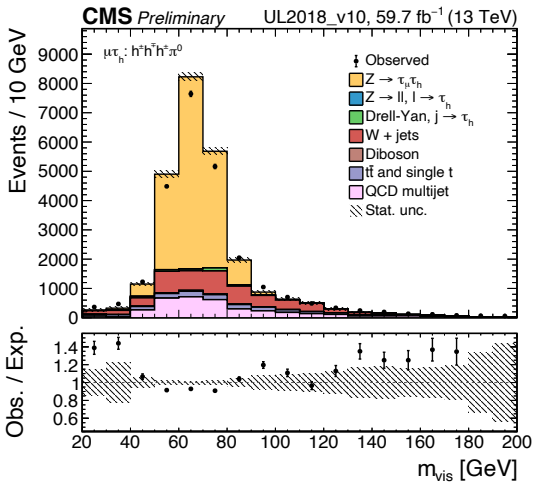


(c) DM10 $40 < p_T(\tau_h) \leq 200$ GeV (Pre-fit)

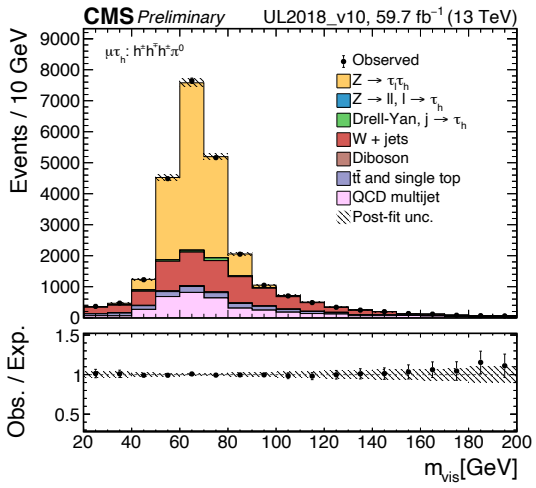


(d) DM10 $40 < p_T(\tau_h) \leq 200$ GeV (Post-fit)

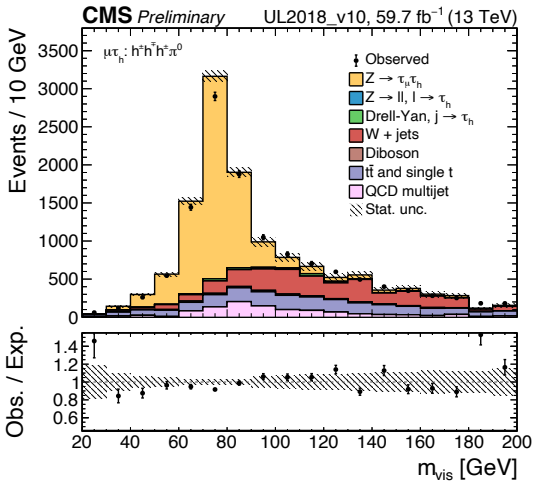
Figure 5.8: Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM10-p_T regions for UL2018.



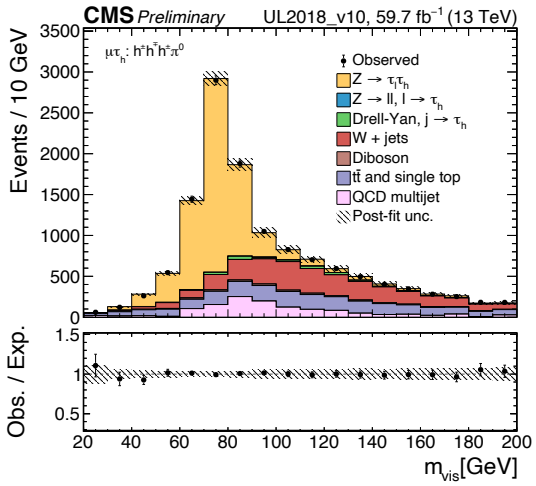
(a) DM11 $20 < p_T(\tau_h) \leq 40$ GeV (Pre-fit)



(b) DM11 $20 < p_T(\tau_h) \leq 40$ GeV (Post-fit)



(c) DM11 $40 < p_T(\tau_h) \leq 200$ GeV (Pre-fit)



(d) DM11 $40 < p_T(\tau_h) \leq 200$ GeV (Post-fit)

Figure 5.9: Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM11-p_T regions for UL2018.

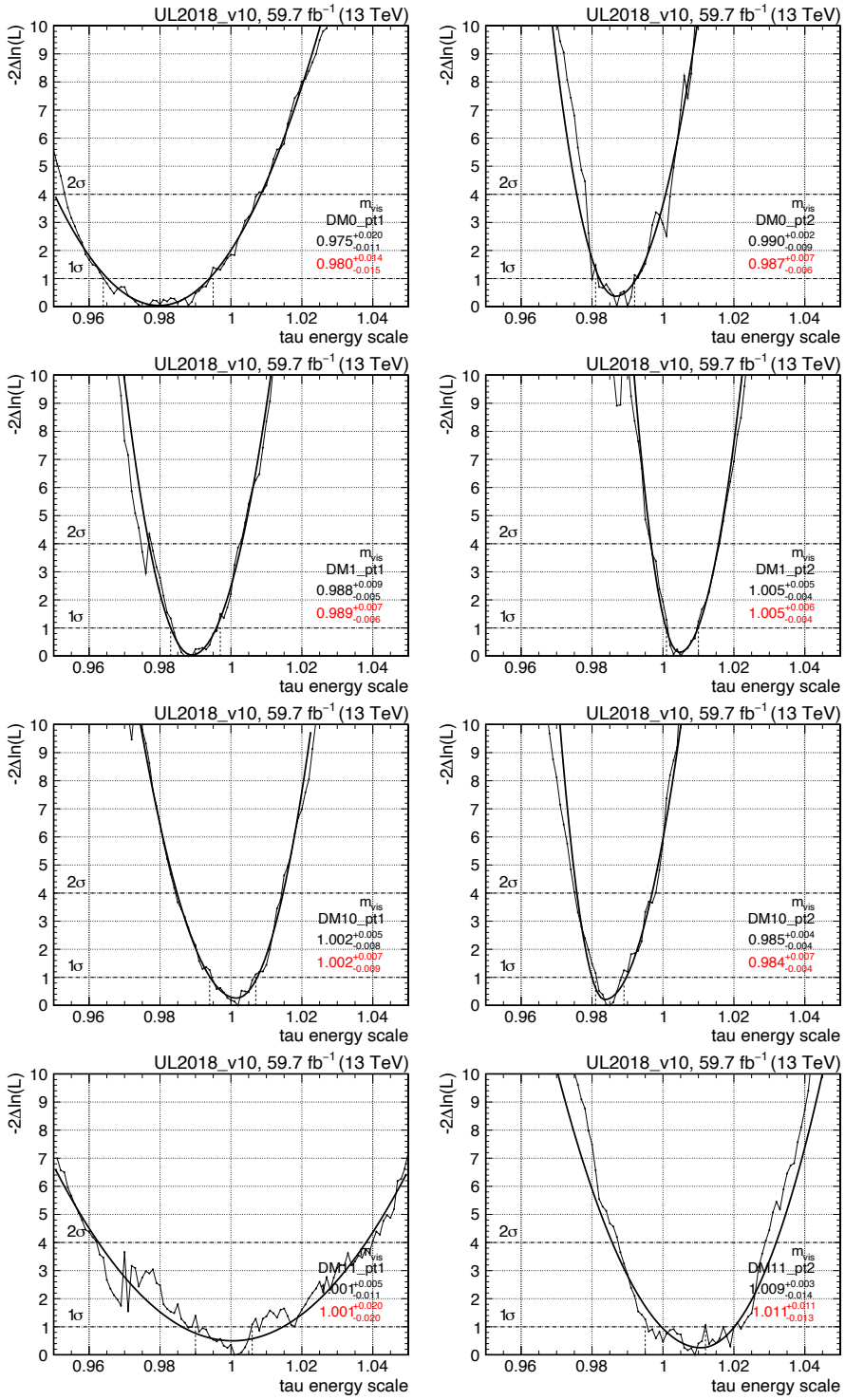


Figure 5.10: Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 5.

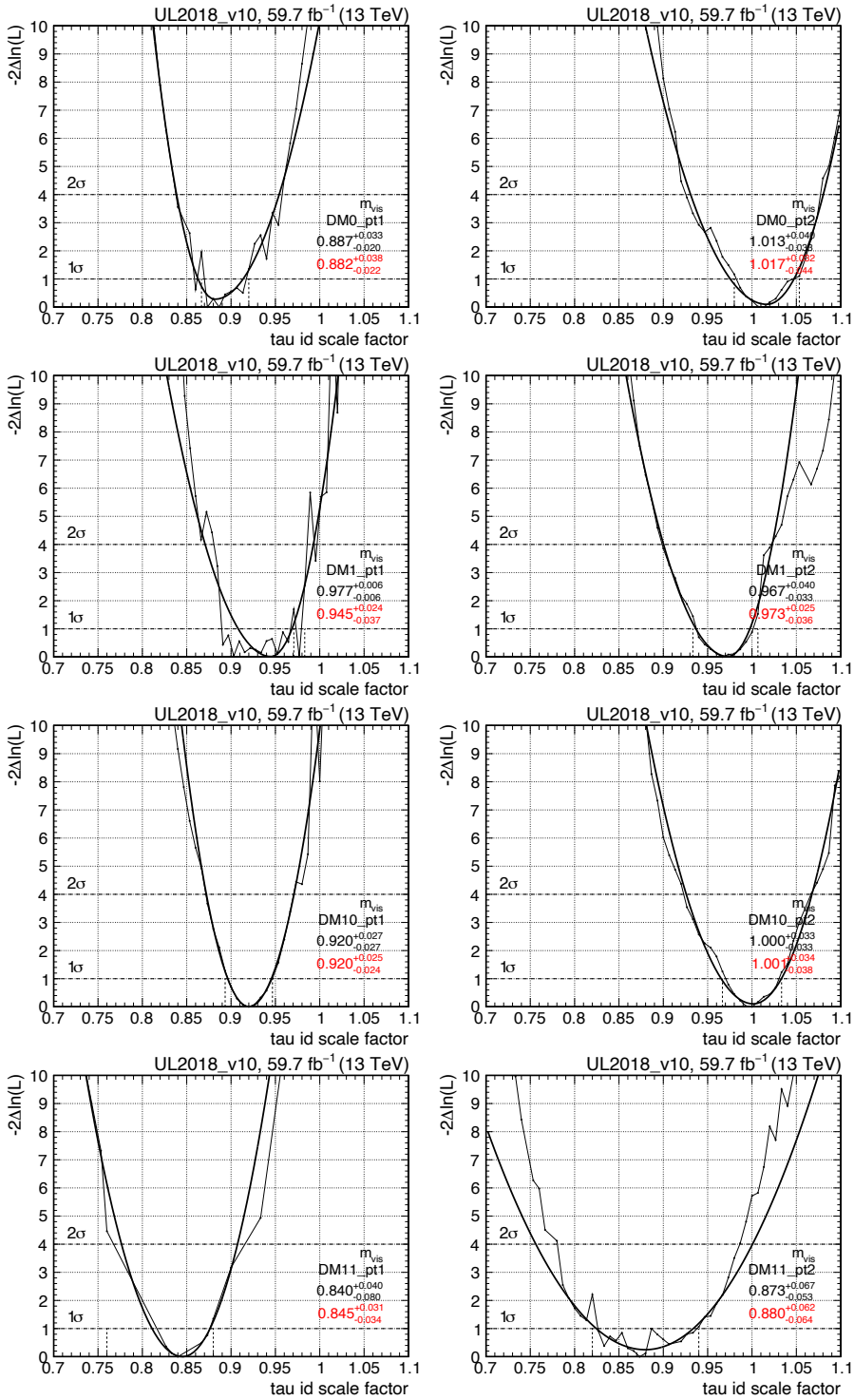


Figure 5.11: Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 5.

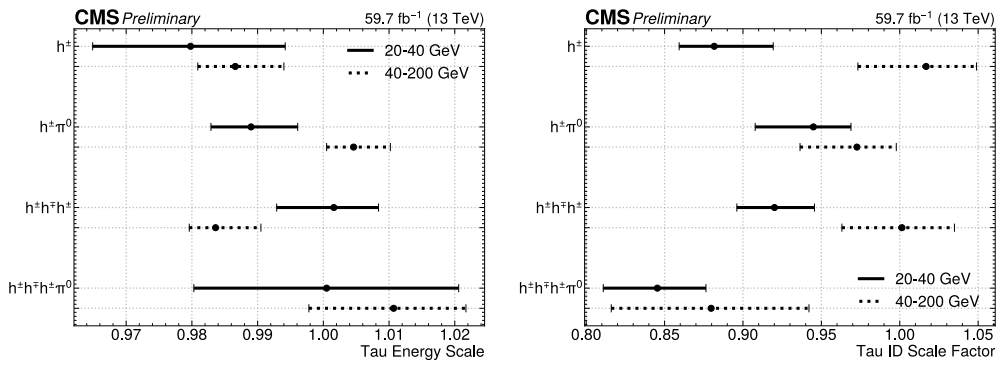


Figure 5.12: Summary of TES (left) and ID SF (right) per DM and $p_T(\tau_h)$ regions for UL2018, obtained from parabolic fits.

IV.4. Combined Measurements in the 2022 Run 3 Datasets

The combined–fit strategy developed for the Run 2 is re-used for Run 3. A first iteration of the measurement (see Appendix A Section III.) revealed an important data–simulation discrepancy, most pronounced in DM0. To improve the signal purity and reduce the dominant W+jets background, the transverse-mass requirement was tightened from $m_T < 65$GeV to $m_T < 40$GeV. For consistency, this selection is kept in the results presented here.

A second iteration became necessary after a bug was identified in the initial 2022 samples which was affecting the simulation in the τ_h polarization (details in Appendix A Section II.). The updated analysis presented here uses POWHEG–generated DY samples with lower statistics; consequently, the fit range is restricted to the DM region for both 2022 preEE and postEE datasets.

For the Run 3 analysis, PUPPI-based MET is used instead of PF MET, following the Run 3 recommendations. In addition, due to the absence of a dedicated measurement of the muon-to-τ_h fake rate for 2022, this background is treated as a yield-only parameter in the fit. In contrast, in the 2018 TES extraction, the same contribution was treated as a shape-parameterized nuisance. The SFs are measured using the following discriminator working points: D_{jet} Medium, D_e VVLoose, and D_μ Tight.

IV.4.a. 2022 preEE Measurements

Given the limited event statistics in the preEE dataset, a wider TES variation range of -5% to $+5\%$ is employed to cover the 1σ uncertainty interval. The profile likelihood scans for the TES extraction are shown in Figure 5.13. Due to the low statistics in the DM11 region, the profile NLL exhibits significant fluctuations. As a result, the final TES values are obtained from a parabolic fit to the NLL profile, shown in black. The ID SF is also profiled, and its NLL scans are displayed in Figure 5.14. The summary results for both TES and ID SF are shown in Figure 5.15.

Compared to the results obtained prior to the correction of the DY sample bug, the TES extracted for the DM0 decay mode is now significantly closer to unity, indicating a improved agreement between data and simulation. The TES values for the other decay modes (DM1, DM10, and DM11) remain consistent with the first results, which were less impacted by the bug, confirming the reliability of the measurement. Furthermore, the new samples enable a simultaneous scan of both the Tau ID scale factor and TES, which was not feasible in with the bugged samples.

Regarding the identification scale factors, all measured values are consistent with unity within 2σ uncertainties. The only notable deviation is observed for

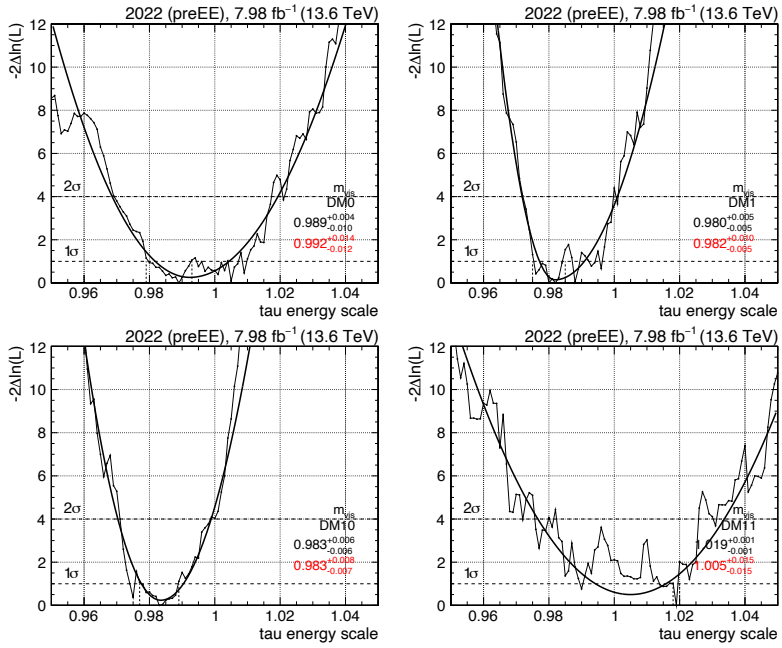


Figure 5.13: Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11.

DM0, which yields the lowest ID scale factor among the decay modes. Further investigation is required to clarify this discrepancy; one possible explanation is the misidentification of true DM0 candidates as DM1, due to reconstruction ambiguities. This hypothesis is supported by the observation that the DM1 identification scale factor is the only one significantly above unity, suggesting an overestimation of the DM1 yield and a corresponding underestimation of the DM0 yield.

The pre-fit versus post-fit distributions are shown in Figure 5.16, demonstrating good closure between the model and the data. The dependency plots between the TES and the ID scale factors for each decay mode are presented in Figure A.6. Similarly to the results obtained in the UL2018 dataset, the TES and ID scale factors are found to be only weakly correlated. Moreover, the extracted values are compatible with those obtained using the individual fit strategy described in Appendix A Section IV.3.

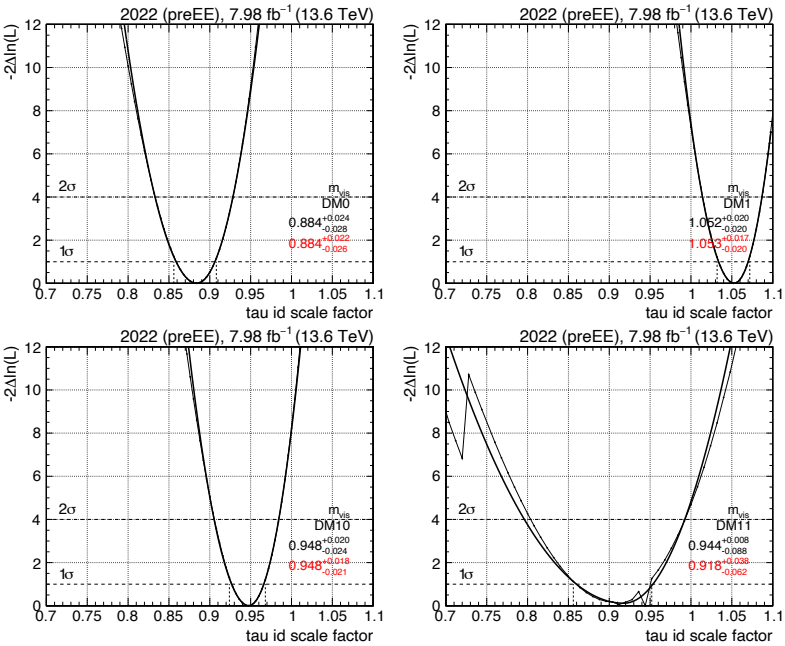


Figure 5.14: Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11.

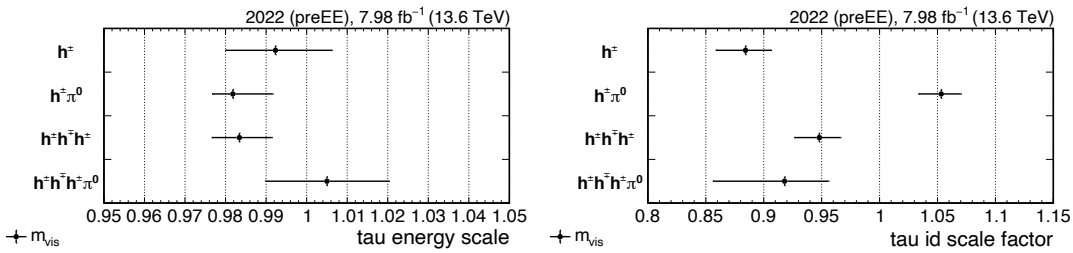
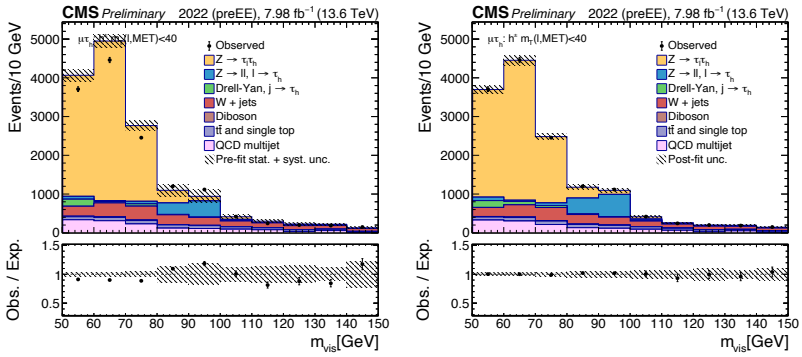
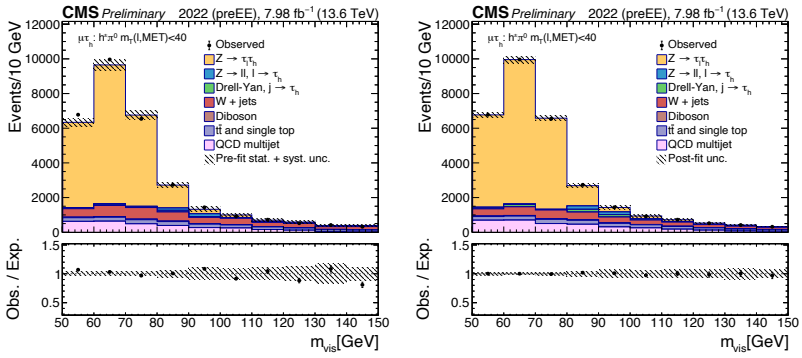


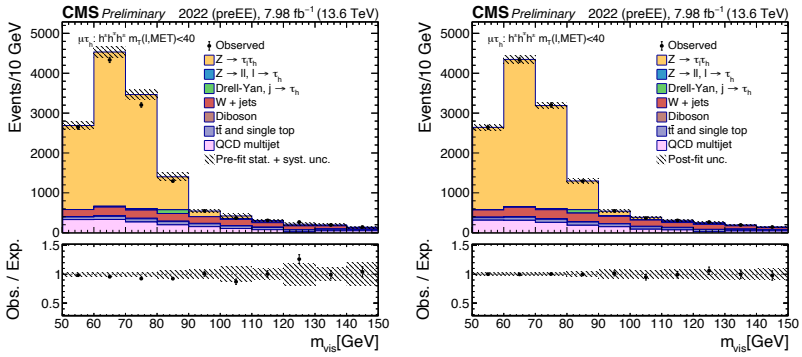
Figure 5.15: Summary of TES (left) and ID SF (right) per DM region for 2022 preEE, obtained from parabolic fits.



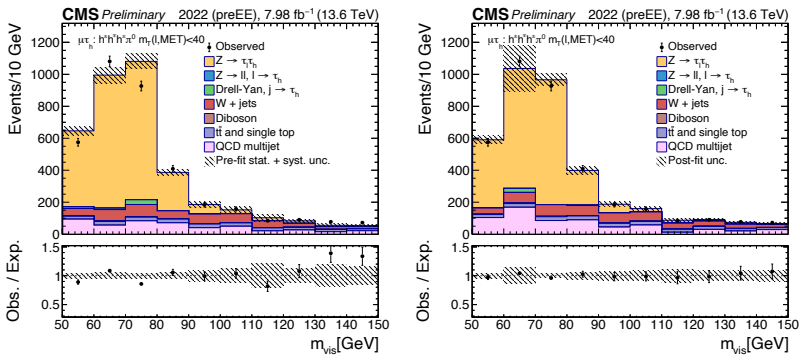
(a) DM0 (Pre-fit) (b) DM0 (Post-fit)



(c) DM1 (Pre-fit) (d) DM1 (Post-fit)



(e) DM10 (Pre-fit) (f) DM10 (Post-fit)



(g) DM11 (Pre-fit) (h) DM11 (Post-fit)

Figure 5.16: Pre-fit (left) and post-fit (right) distributions of the visible mass m_{vis} for 2022 preEE data in the DM0, DM1, DM10, and DM11 decay modes.

IV.4.b. 2022 postEE Measurements

Similarly to the preEE measurement, the postEE results are split by DM. Due to observed fluctuations, the results are extracted from a parabolic fit of the likelihood profiles (see Figure A.7). The summary of the results is provided in Figure 5.17.

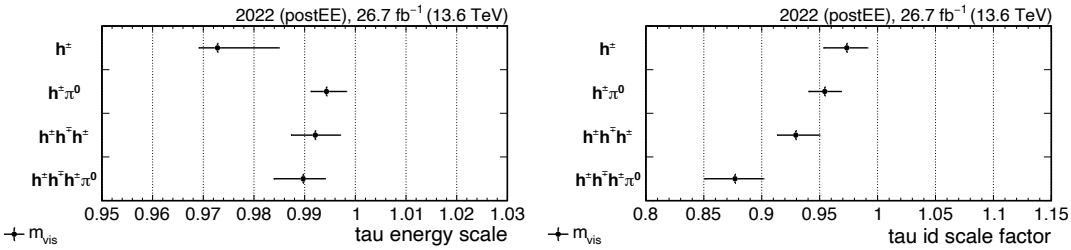
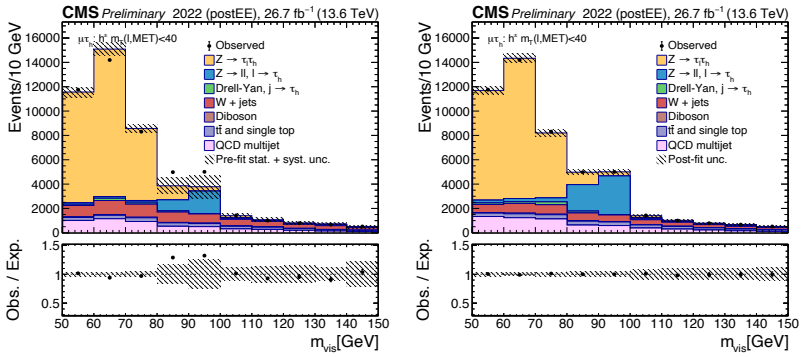


Figure 5.17: Summary of TES (left) and ID SF (right) per DM region for 2022 postEE, obtained from parabolic fits.

Once again, the TES values are consistent with unity within two standard deviations. However, the ID SF deviates further from unity as the number of charged particles in the τ_h decay increases, consistent with the trend observed in the 2018 analysis.

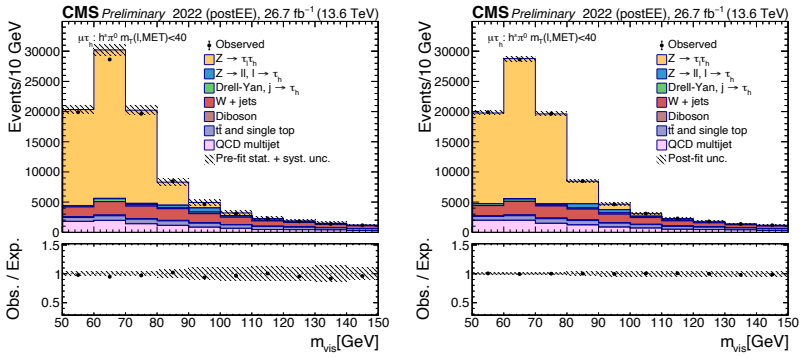
These results are consistent with the first set of measurements affected by the bug, except for DM0. A weak dependency is observed between the scale factors, which justifies performing the individual fit procedure as described in Appendix A Section IV.3.

Finally, the pre-fit versus post-fit shown in Figures 5.18 comparisons demonstrate good closure, confirming the robustness of the fit procedure.



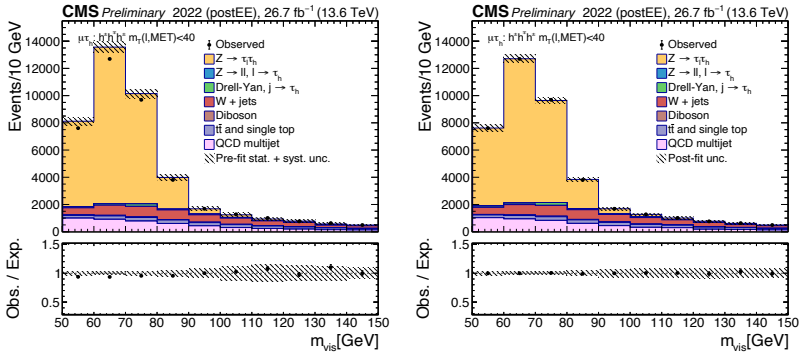
(a) DM0 (Pre-fit)

(b) DM0 (Post-fit)



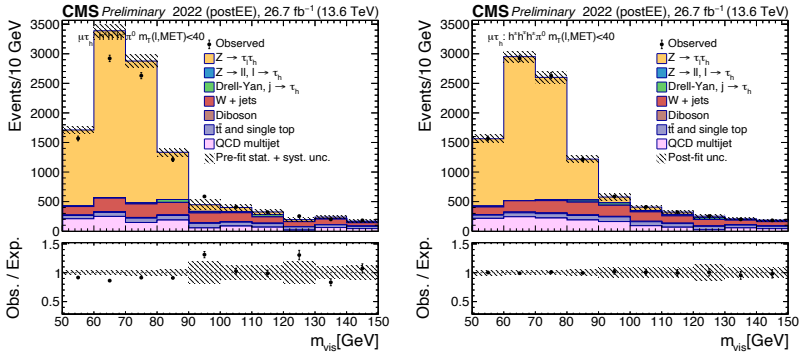
(c) DM1 (Pre-fit)

(d) DM1 (Post-fit)



(e) DM10 (Pre-fit)

(f) DM10 (Post-fit)



(g) DM11 (Pre-fit)

(h) DM11 (Post-fit)

Figure 5.18: Pre-fit (left) and post-fit (right) distributions of the visible mass m_{vis} for 2022 postEE data in the DM0, DM1, DM10, and DM11 decay modes.

IV.5. Comparison with Individual Fit Results

The choice between the combined fit and separate fit methods for deriving scale factors involves balancing the benefits and challenges of each approach. The combined fit method allows for simultaneous determination of the τ_h energy scale and identification efficiency scale factors, accounting for potential correlations and offering a comprehensive treatment of uncertainties. However, because it requires fitting not only the normalization of the m_{vis} distribution but also the shape, it demands a large dataset to achieve a stable fit. Additionally, statistical limitations prevent dividing the data into finer p_T regions, which poses a challenge for analyses sensitive to the p_T dependence of τ_h. In contrast, the separate fit method provides enhanced flexibility by allowing independent determination of each scale factor, making it well-suited for analyzing different kinematic regions. The separate fit does, however, assume low correlation between the factors, which must be verified for validity, and may offer a slightly less comprehensive uncertainty treatment if correlations are overlooked. To address these challenges, an alternative method for handling $\tau_h$energy scale uncertainties was proposed by externalizing them from the $\tau_h$identification scale factor determination, as described in the previous section. In the end, the selection of the method relies on statistical constraints of the analysis and on the specific requirements, particularly in regard to kinematic sensitivity.

V. Conclusion

In this chapter, the combined measurement of the τ_h identification scale factor and energy scale was presented, using a simultaneous fit approach that accounts for correlations and provides a thorough treatment of uncertainties. This method was developed through an extensive study on Run 2 data, where I established its performance and limitations. During the early Run 3 measurements, a bug in the Pythia simulation required significant effort to stabilize the fit. The latest results, free from this issue, now allow the full potential of the combined fit technique to be realized.

While statistically demanding, the combined fit method enables a consistent extraction of scale factors using both normalization and shape information. Comparison with the individual fit method showed that the combined fit is preferred when sufficient statistics are available and when strong sensitivity to $p_T(\tau_h)$ is not needed. On the other hand, the individual fit offers more flexibility in limited data or kinematic-dependent studies.

The results demonstrate good agreement between the two methods within uncertainties, confirming the reliability of the combined fit approach. This work provides a solid foundation for the precise calibration of τ_h objects in Run 3 analyses,

which is crucial for improving the sensitivity of physics measurements involving hadronic tau decays.

Looking ahead, the combined fit strategy and its calibrations will be used in future CMS Run 3 analyses, ensuring consistent and accurate treatment of τ_h energy and identification scale factors throughout the data-taking period.

This work is summarized in the CMS Physics Analysis Summary CMS-PAS-TAU-24-001 [35].

Chapter 6

CP measurement in $H \to \tau_h \tau_h$ decays

Content

I. Analysis Technologies 157

I.1. Columnflow 157

I.1.a. Columnar Analysis Framework 157

I.1.b. Analysis Workflow Management 158

I.2. Deep Learning and Neural Networks 161

I.2.a. Mathematical Foundations of Neural Networks 161

I.2.b. Neural Network Training 163

II. Event Samples 164

II.1. Data Event Samples 164

II.2. Simulated Event Samples 164

II.3. Signal Modeling 166

III. Selection 167

III.1. Trigger and Preselection 167

III.2. Final Selection for $\tau_h \tau_h$ Events 167

III.3. Candidate Pair Selection and Extra Lepton Veto 169

IV. Corrections 169

IV.1. General Corrections 169

IV.2. Impact Parameter Vector Calibration 171

V. Fake Rate Estimation for τ_h 172

V.1. Background Overview 172

V.2. Methodology 172
V.3. Systematic Uncertainties 177
VI. τ_h Momentum Reconstruction and CP Observable . . 178
VI.1. SVFit and FastMTT 178
VI.2. Machine Learning Studies in the $\rho\rho$ Channel 179
VI.2.a. Motivation 179
VI.2.b. Lorentz Boost Network 182
VI.2.c. Performance of the Network 183
VII. Event Categorization and Signal Discrimination 186
VIII. Sensitivity Scan of $\phi_{\mathbf{CP}}$ 186
IX. Conclusion 187

This chapter presents the analysis targeting the measurement of CP violation in $H \to \tau\tau$ decay, using Run 3 proton-proton collision data collected by the CMS experiment at a center-of-mass energy of 13.6 TeV during 2022 and 2023. The physical motivations are detailed in Chapter 1 Section III. The focus is placed on the fully hadronic final state ($\tau_h\tau_h$), which has the largest branching fraction among the $\tau\tau$ decay modes [15]. Despite this advantage, the final state imposes distinct challenges that strongly influence the overall analysis strategy.

At $\sqrt{s} = 13.6$ TeV, the inclusive Higgs boson production cross section is approximately $\sigma_H \approx 55$ pb, predominantly via gluon fusion [194]. The branching fraction for $H \to \tau\tau$ is approximately 6.3% [15], and the $\tau_h\tau_h$ final state accounts for about 42% of these decays [15]. As a result, the effective signal rate is relatively small, requiring high signal selection efficiency and sophisticated background suppression techniques.

The dominant background arises from Drell–Yan $Z/\gamma^* \to \tau\tau$ production, which occurs at rates significantly larger than the signal. In addition, QCD multijet events, where jets are misidentified as τ_h, contribute substantially to the background, particularly in signal-like regions of phase space. Accurate modeling of these backgrounds necessitates robust, data-driven estimation techniques.

The τ_h decays always involve at least one neutrino, leading to missing transverse energy in the event. This complicates the reconstruction of both the di-τ invariant mass and the CP-sensitive observable ϕ_{CP}. In τ decay modes with more visible particles (e.g., three-prong decays), the neutrino kinematics can be better inferred, yielding improved resolution on ϕ_{CP}. In contrast, for one-prong or less-informative decay modes, alternative reconstruction strategies are adopted to infer ϕ_{CP}. These approaches allow the analysis to retain sensitivity across all relevant τ decay channels (see Section IV.).

Given the moderate mass resolution and the low signal-to-background ratio after baseline selection, multivariate techniques and event categorization are employed to enhance the sensitivity. These methods exploit differences in event topology, kinematics, and τ decay structure between signal and background.

My contributions to this analysis are focused on the implementation of the Fake Factor method for estimating the jet $\rightarrow \tau_h$ background, and on the development of machine-learning-based τ momentum reconstruction techniques tailored to the CP-sensitive final states. Additionally, I participated in the implementation of the selection, calibration, and categorization steps within a columnar-based analysis framework.

The overall strategy of the analysis follows the template established by the Run 2 analysis [17] summarized in Section III.3.b. To recall the context, the combined Run 2 CMS analysis across all $\tau\tau$ channels achieved an observed (expected) significance of 3.0σ (2.6σ) in discriminating between pure CP-even and pure CP-odd Higgs hypotheses. The corresponding constraint on the CP-mixing angle was:

$$\phi_{\tau\tau} = -1^\circ \pm 19^\circ \, (0^\circ \pm 21^\circ) \quad \text{at 68\% CL}, \quad \phi_{\tau\tau} = -1^\circ \pm 39^\circ \, (0^\circ \pm 47^\circ) \quad \text{at 95\% CL}.$$

I. Analysis Technologies

I.1. Columnflow

In Run 2, most analyses relied on `ROOT` Trees [36], which store data on an event-by-event (row-based). This worked well for smaller datasets and sequential processing. However, Run 3 presents new challenges: the dataset size has grown significantly, data formats [195] are now more compact, and analyses increasingly use machine learning methods that require fast, flexible data access.

Row-based data formats are inefficient for such tasks because they require reading entire events even when only a few variables are needed. Moreover, they do not naturally support vectorized computations or parallel processing, which are essential for speeding up analysis workflows. These limitations motivated the adoption of a columnar data processing model with `ColumnFlow`. `ColumnFlow` [196, 197] is a framework for columnar-based data analysis which is also providing a fully automated workflow.

I.1.a. Columnar Analysis Framework

In the columnar model, data are stored by variable (column) rather than by event (row). This means that all values of a given variable are stored contiguously, allowing operations to be performed on entire arrays of values simultaneously. Such vectorized operations are faster and more memory-efficient.

Columnar data also enables straightforward parallel processing. Different columns or event subsets can be processed independently on multiple CPU cores or distributed computing nodes, greatly improving analysis throughput and scalability.

The `ColumnFlow` framework combines powerful tools like `AwkwardArray` [37] (to handle complex, irregular particle physics data), `NumPy` [38] (for fast numerical computations), and `Parquet` [39] (an efficient columnar file format) to support this approach. By selectively reading only the needed variables, `ColumnFlow` reduces input/output overhead and memory usage, which is especially beneficial for workflows involving systematic uncertainty evaluations and machine learning applications. The *Coffea* [198] package is also used for four-vector computations. Figure 6.1 illustrates how data are organized differently in row-based and columnar formats.

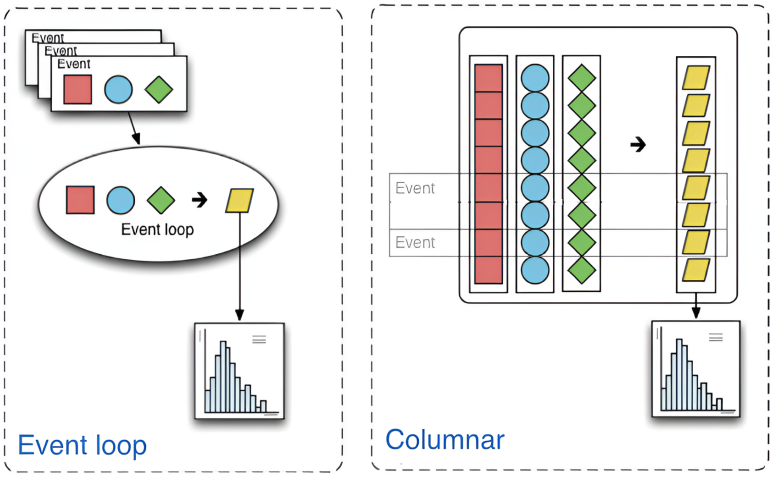


Figure 6.1: Illustration of row-based vs. columnar data layouts [199].

I.1.b. Analysis Workflow Management

Columnflow provides a fully automated workflow management system based on the Luigi Analysis Workflow (law) package [40], which uses the Luigi engine [200]. Each step in the analysis is implemented as a task with defined inputs, outputs, and dependencies. When executed, tasks automatically resolve these dependencies, running only what is needed and reusing existing results. Task parameters, including versioning, are encoded in output paths to ensure reproducibility and customization.

The workflow follows a typical high-energy physics analysis sequence and the key tasks include fetching datasets, applying calibrations, event selection, merging reduced datasets, reconstruction, machine learning evaluation, and histogram

creation. Each task can specify its own software environment using sandboxing, supporting shell scripts, Python virtual environments, or containers. The overall workflow is illustrated in Figure 6.2.

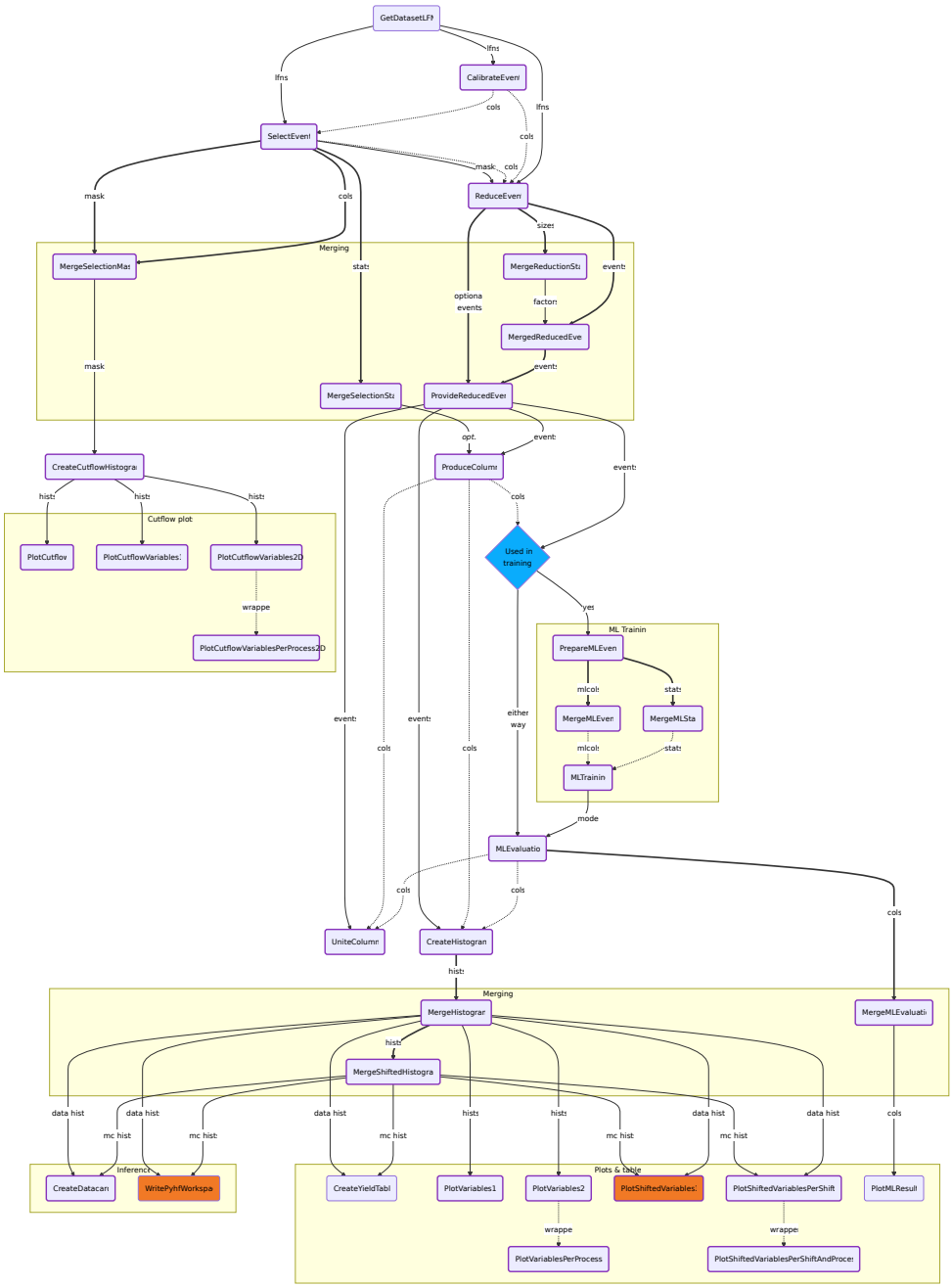


Figure 6.2: Schematic of the analysis workflow managed with columnflow [196].

I.2. Deep Learning and Neural Networks

Deep learning is a machine learning (ML) technique based on deep neural networks (DNNs). DNNs are layered structures of interconnected neurons designed to model complex, non-linear patterns in data. In high-energy physics, such networks are widely used for tasks like particle identification and event classification. This section outlines the key principles of multivariate analysis (MVA) with deep neural networks, focusing only on the components relevant to this analysis, such as network structure, training procedures, and generalization techniques. For a more detailed introduction to deep learning, the reader is referred to Ref. [201].

I.2.a. Mathematical Foundations of Neural Networks

A DNN is a parameterized function that maps an input vector of *features* $\vec{x}$ to an output *prediction* $\vec{y}$. The model is composed of multiple sequential *layers*, each consisting of *units* (also called *neurons*) connected by weights and biases. These connections define the model's learnable parameters. Figure 6.3 shows a typical DNN with an input layer ($l = 0$), $L-1$ *hidden layers*, and an output layer ($l = L$). The output of each layer l is computed by applying a linear transformation to the

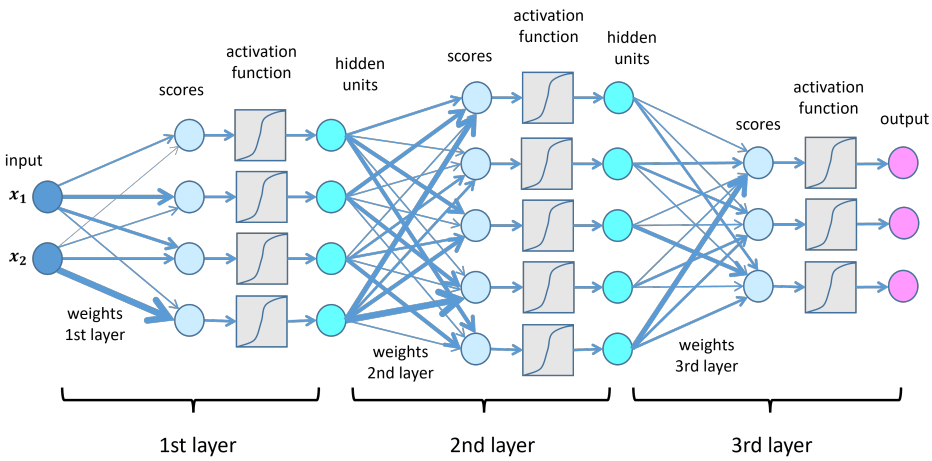


Figure 6.3: Schematic of a typical DNN [202].

outputs of the previous layer, followed by a non-linear activation [203]:

$$\vec{z}^{(l)} = \mathbf{W}^{(l-1)} \cdot \vec{a}^{(l-1)} + \vec{b}^{(l-1)}, \quad \vec{a}^{(l)} = \sigma^{(l)}(\vec{z}^{(l)}),$$

where $\mathbf{W}^{(l)}$ and $\vec{b}^{(l)}$ are the weights and biases, $\vec{a}^{(l)}$ is the layer output (also called *activation*), and $\sigma^{(l)}$ is a non-linear activation function applied element-wise. The

bias terms $\vec{b}$ allow each unit to shift the activation function output, increasing the model's flexibility.

The non-linearity introduced by the activation function is essential; without it, the network would reduce to a single linear transformation, regardless of its depth, limiting its ability to model complex patterns. Common choices for σ include ReLU, sigmoid, Leaky ReLU or tanh, with an illustration shown in Figure 6.4. The choice of activation function can also affect the training stability and final model performance.

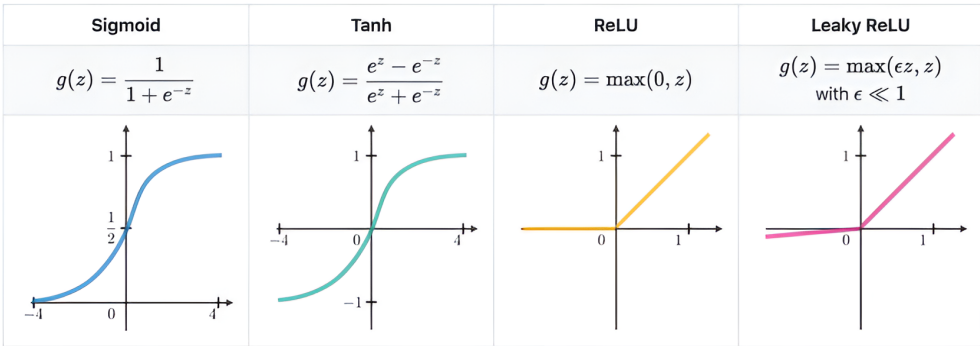


Figure 6.4: Activation function [204].

The full network function is then expressed as a composition of all layers:

$$\vec{y} = f(\vec{x} \mid \omega) = a^{(L)} \circ a^{(L-1)} \circ \cdots \circ a^{(1)}(\vec{x}),$$

where $\omega = \{\mathbf{W}^{(l)}, \vec{b}^{(l)}\}$ denotes the complete set of trainable parameters, which can number in the millions depending on the network size.

To determine these parameters, the model undergoes a *training* phase, during which it minimizes a task-specific *loss function* that measures the difference between the prediction $\vec{y}$ and the true target value. This optimization is performed iteratively.

The number of layers, number of units per layer, choice of activation functions, and other architectural decisions are collectively referred to as *hyperparameters*. These are not learned during training and must be defined beforehand, typically via cross-validation.

For multi-class classification tasks, the output layer uses a *softmax* activation function to ensure that the output values represent a normalized probability distribution:

$$\sigma_{\text{softmax}}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}.$$

Further details on the training algorithm and loss function used in this work are discussed in the following sections.

I.2.b. Neural Network Training

Training a neural network consists in finding the set of weights and biases ω that minimize a task-specific *loss function*, which measures the difference between the predicted output $\vec{y}_n(\vec{x})$ and the true label $\hat{\vec{y}}_n$. For regression, a common choice is the mean squared error (MSE), while classification tasks typically use the cross-entropy (CE) loss:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N} \sum_n \left(\hat{y}_n - y_n(\vec{x})\right)^2 , \tag{6.1}$$

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N} \sum_n \hat{\vec{y}}_n \cdot \log \vec{y}_n(\vec{x}), \tag{6.2}$$

where N is the batch size and the logarithm is applied element-wise.

The weights are updated iteratively using gradient descent:

$$\omega_{t+1} = \omega_t - \eta \cdot \nabla_\omega \mathcal{L}, \tag{6.3}$$

where η is the *learning rate*, and $\nabla_\omega \mathcal{L}$ is the gradient of the loss with respect to the model parameters.

The gradient is computed efficiently using the *backpropagation* algorithm, which recursively propagates errors from the output layer backward through the network. The partial derivatives of the loss with respect to the weights and biases are given by:

$$\frac{\partial \mathcal{L}}{\partial W_{ij}^{(l)}} = a_j^{(l-1)} \cdot \delta_i^{(l)}, \quad \frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = \delta_i^{(l)}, \tag{6.4}$$

where $\delta^{(l)}$ is the error in layer l, defined recursively based on the gradient at the output and the derivatives of the activation functions.

The training loop proceeds as:

1. Forward pass: compute network output $\vec{y}$ and loss $\mathcal{L}$.

2. Backward pass: compute gradients via backpropagation.

3. Update parameters using the gradient descent rule.

To avoid overfitting, the loss is also monitored on an independent *validation set*. If the validation loss stops improving or diverges from the training loss, training is stopped. A third, independent *test set* is used for final performance evaluation after training.

II. Event Samples

This section introduce the samples of measured and simulated events used for this analysis. The data events samples for early Run 3 are described, as well as simulated *Monte Carlo* (MC) [15] event samples for background and signal.

II.1. Data Event Samples

The Run 3 data-taking period used in this analysis includes both the 2022 and 2023 datasets [205], corresponding to an integrated luminosity of 38.5 fb^{-1} and 58.6 fb^{-1}, respectively. The 2022 data is split into 2022 preEE and 2022 postEE, corresponding to data taking before and after a water leak in the ECAL endcap. The affected crystals were masked, and the issue was resolved during the year-end technical stop. The 2023 data is split into 2023 preBPix and 2023 postBPix, corresponding to data taking before and after the failure of 27 BPix modules following Technical Stop 1. The resulting gap in pixel coverage affected tracking and jet energy measurements. Dedicated corrections and simulations are used to model the impact. The datasets used in each era are categorized by trigger object: `EGamma` for electrons, `Muon`, `SingleMuon`, and `DoubleMuon` for muons, `Tau` for τ_h candidates, and `MuonEG` for electron-muon final states.

II.2. Simulated Event Samples

Simulated MC samples are used to model the signal and background processes contributing to the analysis. The primary backgrounds include Drell–Yan ($Z/\gamma^* \to \ell\ell$), W+jets, top quark pair production ($t\bar{t}$), diboson (WW, WZ, ZZ), triboson (WWW, WWZ, WZZ, ZZZ), and single top production. Drell–Yan and W+jets processes are generated in exclusive bins of jet multiplicity and dilepton transverse momentum $p_T^{\ell\ell}$, to enhance modeling in boosted or high recoil topologies relevant for CP-sensitive observables.

All samples are generated with MadGraph5_aMC@NLO (v2.2.2) [206], interfaced with Pythia8 (v8.200) [207] for parton showering and hadronization. The top quark and Higgs boson masses are set to 172.5 GeV and 125.0 GeV, respectively [15]. The simulation is normalized to the corresponding cross sections listed in Table 6.1, which include higher-order corrections when available. A details list of samples is given in Appendix B Section I.

A known issue affects the `Pythia 8` Drell–Yan samples, specifically impacting the modeling of the DM0 (see Appendix A Section II.). The results shown in this thesis are based on the affected samples. A corrected set of samples (in production at the time of thesis writing) is expected to be used in the final analysis.

Table 6.1: Simulated signal and background samples used in this analysis. The *Label* column indicates groupings used in plots; the *Color* column refers to the consistent color coding used throughout this chapter.

Dataset	**Channel / Phase Space**	**Generator**	**Label**	**Color**
$Z/\gamma^* \to \ell\ell$	inclusive +1 jet +2 jets +3 jets low mass	MG5_aMC@NLO + Pythia8	$Z \to \ell^+\ell^-$	
$Z/\gamma^* \to \ell\ell$ +1 jet	40–100 GeV 100–200 GeV 200–400 GeV 400–600 GeV >600 GeV	MG5_aMC@NLO + Pythia8	$Z \to \ell^+\ell^-$	
$Z/\gamma^* \to \ell\ell$ +2 jets	40–100 GeV 100–200 GeV 200–400 GeV 400–600 GeV >600 GeV	MG5_aMC@NLO + Pythia8	$Z \to \ell^+\ell^-$	
$t\bar{t}$	inclusive	POWHEG + Pythia8	ttbar	
W + jets	inclusive +1 jet 200–400 GeV 400–600 GeV >600 GeV	MG5_aMC@NLO + Pythia8	W + jets	
Diboson (WW, WZ, ZZ)	inclusive	Pythia8	Di-boson	
Single top (all channels)	inclusive	POWHEG + Pythia8	SingleTop	
Triboson (WWW, WWZ, ...)	inclusive	Pythia8	Triple-boson	

II.3. Signal Modeling

The signal in this analysis corresponds to Higgs boson production via ggH, VBF, and associated production with a W or Z boson (WH, ZH, or collectively VH).

All signal samples are generated with POWHEG. The gluon fusion process with two extra jets ($gg \to H + 2\,\text{jets}$) uses the MiNLO method [208], while VBF and VH samples are produced using the MiNNLO approach.

For the ggH channel, dedicated MC samples are created to study different CP properties in both production and decay. Three separate samples represent:

- Pure CP-even (Standard Model–like) Higgs production,
- Pure CP-odd Higgs production,
- A maximally CP-mixed Higgs state.

The Higgs boson decay to τ_h leptons is simulated independently with PYTHIA 8.2. To include spin correlations and CP-sensitive effects in the decay, the TAUSPINNER tool [209] is applied after generation. This reweighs events for different CP mixing angles:

- Pure CP-even ($\phi_{\tau_h\tau_h} = 0°$),
- Pure CP-odd ($\phi_{\tau_h\tau_h} = 90°$),
- CP-mixed ($\phi_{\tau_h\tau_h} = 45°$).

Using these weights, the differential cross section for any CP mixing angle $\phi_{\tau_h\tau_h}$ can be expressed as:

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}x}(\phi_{\tau_h\tau_h}) = &(\cos^2\phi_{\tau_h\tau_h} - \cos\phi_{\tau_h\tau_h}\sin\phi_{\tau_h\tau_h}) \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}x}_{\text{CP-even}} \\ &+ (\sin^2\phi_{\tau_h\tau_h} - \cos\phi_{\tau_h\tau_h}\sin\phi_{\tau_h\tau_h}) \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}x}_{\text{CP-odd}} \\ &+ 2\cos\phi_{\tau_h\tau_h}\sin\phi_{\tau_h\tau_h} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}x}_{\text{CP-mix}}\end{aligned} \tag{6.5}$$

To improve the signal generation efficiency, a generator-level filter selects events where the visible p_T and $|\eta|$ of the τ_h decay products are slightly looser (about 10–20% lower) than the offline analysis requirements. This increases the fraction of simulated events that pass the final selection. The applied cuts for the $\tau_h\tau_h$ channel is $p_{\mathrm{T}}^{\tau_h{}^1} > 20$ GeV, $p_{\mathrm{T}}^{\tau_h{}^2} > 20$, $|\eta|^{\tau_h{}^1} < 3.0$ and $|\eta|^{\tau_h{}^2} < 3.0$.

III. Selection

The event selection aims to isolate a phase space region enriched in Higgs boson decays to τ_h while suppressing dominant backgrounds such as $Z/\gamma^* \to \tau\tau$, QCD multijet, and events with jets misidentified as τ_h. Selection criteria are chosen to maximize the expected sensitivity to the CP-mixing angle by balancing signal efficiency and background rejection. Additionally, the selection avoids regions where simulation models show significant discrepancies from data, ensuring reliable background modeling.

III.1. Trigger and Preselection

The $H \to \tau_h\tau_h$ final state is selected using dedicated HLT paths optimized for hadronic τ pairs. In particular, the following triggers are employed:

- `HLT_DoubleMediumDeepTauPFTauHPS35_L2NN_eta2p1`: requires two hadronic τ candidates with $p_T > 40$ GeV and $|\eta| < 2.1$,
- `HLT_DoubleMediumDeepTauPFTauHPS30_L2NN_eta2p1_PFJet60`: requires two hadronic τ candidates with $p_T > 35$ GeV, $|\eta| < 2.1$, and at least one AK4 jet with $p_T > 60$ GeV.

The overlap between these trigger paths is explicitly considered to enhance signal acceptance.

Preselection is applied to remove detector noise and non-physical events. All events are required to pass the full set of MET filters as recommended by the Jet-MET POG, including the primary vertex filter, beam halo filter, ECAL TP filter, bad PF muon filters, HF noisy hits filter, and the ee badSC noise filter. Additional event cleaning is performed for run numbers between 362433 and 367144 to mitigate the impact of a known bad ECAL crystal, which introduces high EM-fraction jets and spurious MET. Events in this period are rejected if $p_T^{\text{miss}} > 100$ GeV and there exists an AK4 jet with $p_T > 50$ GeV, $-0.5 < \eta < -0.1$, $-2.1 < \phi < -1.8$, a neutral or charged EM fraction greater than 0.9, and $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}}) > 2.9$.

The baseline object preselections are summarized in Table 6.2.

III.2. Final Selection for $\tau_h\tau_h$ Events

Events are retained if at least one pair of oppositely charged hadronic τ candidates is found, with $\Delta R > 0.5$ between the two τ_h. The final selection thresholds are optimized based on the trigger requirements. Each τ_h must satisfy $p_T > 35$ GeV and $|\eta| < 2.1$. They must pass the VVTight anti-jet, VVLoose anti-muon, and Tight anti-electron discriminators. Additional decay-mode-dependent selections

Table 6.2: Summary of event preselection and baseline object selection criteria.

Category	**Selection Criteria**
Electron Selection	
Kinematics	$p_T > 10$ GeV, $\|\eta\| < 2.5$
ID	MVA ID (90% WP) with isolation
Impact parameters	$\|d_z\| < 0.2$ cm, $\|d_{xy}\| < 0.045$ cm
Muon Selection	
Kinematics	$p_T > 10$ GeV, $\|\eta\| < 2.4$
ID	Medium ID
Impact parameters	$\|d_z\| < 0.2$ cm, $\|d_{xy}\| < 0.045$ cm
τ_h *Selection*	
Kinematics	$p_T > 20$ GeV, $\|\eta\| < 2.5$, $\|d_z\| < 0.2$ cm
Decay modes	ParticleNet DM = 0,1 (with HPS DM = 1)
	2 (HPS DM = 1), 10, 11; excluding HPS DM = 5, 6
Discriminators	Anti-jet: VVLoose WP
	Anti-muon: VVLoose WP
	Anti-electron: $\geq$ VLoose WP

are imposed: for ParticleNet DM = 0, the impact parameter significance must be greater than 1.25; for DM = 1 or 2, the energy asymmetry between the charged and neutral pion must exceed 0.2; and for DM = 10, a reconstructed secondary vertex is required.

III.3. Candidate Pair Selection and Extra Lepton Veto

In events with multiple valid $\tau_h\tau_h$ pairs, the pair selection prioritizes isolation and transverse momentum. The pair with the most isolated leading candidate is chosen, where isolation is defined by the DeepTau raw score. If the isolation is equal between pairs, the one with higher p_T is selected. If both leading candidates are identical, the same logic is applied to the subleading τ_h. Events containing additional loosely selected leptons (electrons or muons) with $\Delta R > 0.5$ from the τ_h candidates are vetoed. These leptons must pass minimal identification and isolation criteria, ensuring suppression of W+jets and $t\bar{t}$ backgrounds.

IV. Corrections

Residual differences between data and simulation are corrected using SFs applied to object efficiencies, energy scales, and resolutions. These corrections include τ_h identification and trigger efficiency SFs derived from dedicated control samples.

Corrections are applied either on an event-by-event basis or through reweighting techniques that adjust distributions of key observables. This ensures that simulation accurately reflects the detector response and data-taking conditions.

IV.1. General Corrections

A summary of the corrections applied in this analysis is provided in Table 6.3.

Table 6.3: Summary of standard corrections applied in the analysis.

Object	**Corrections**
Muon	ID SFs
	Isolation SF
	SingleMuon trigger SF ($p_T > 26$ GeV)
	CrossMuon trigger SF ("Orthogonal OR", $p_T < 25$ GeV)
	Energy scale correction (pre-selection)
Electron	ID SFs
	Isolation SF
	SingleElectron trigger SF ($p_T \geq 31$ GeV)
	CrossElectron trigger SF ("Orthogonal OR", $p_T < 30$ GeV)
	Scale and smearing corrections (pre-selection)
τ_h	ID SFs
	Lepton fake SF
	Trigger SF (Inclusive OR, jet $p_T > 60$ GeV)
	τ_h energy scale correction (pre-selection)
Jet	Jet energy correction (JEC)
	Jet energy resolution (JER)
	Propagation of JEC to MET
	TEC to MET
General	IP vector correction
	SF for IP significance
	MET recoil corrections
	Jet-to-τ_h fake factor
	Pileup weighting (PU-weight)
	Z p_T reweighting for DY
	τ_h spinner weights for signal
	$t\bar{t}$ shape uncertainty from p_T reweighting
	MET ϕ correction
	ggH quark mass effect

IV.2. Impact Parameter Vector Calibration

An essential component in the reconstruction of CP-sensitive observables in $H \to \tau\tau$ decays is the impact parameter ($\vec{IP}$) vector of charged tracks associated with τ lepton decays with is used in the $\tau \to \pi\nu_\tau$ decay. Accurate modeling of this vector improves the precision of decay plane reconstruction, particularly in $\tau_h\tau_h$ final state. As the $\vec{IP}$ vector resolution and shape are not perfectly described in MC simulation, a dedicated calibration procedure is employed to align simulation with data.

The correction is derived using $Z \to \mu\mu$ and $Z \to ee$ control samples, where the true $\vec{IP}$ distribution is well understood and dominated by detector resolution effects. The procedure targets the Cartesian components of the $\vec{IP}$ vector (x, y, and z) and applies a quantile mapping technique to transform the simulated distributions to match those observed in data.

For each $\vec{IP}$ component $i \in \{x, y, z\}$, the corrected value is computed via:

$$\mathrm{IP}_i^{\mathrm{corr}} = F_{\mathrm{data},i}^{-1}\left(F_{\mathrm{MC},i}\left(\mathrm{IP}_i\right)\right), \tag{6.6}$$

where $F_{\mathrm{MC},i}$ is the cumulative distribution function (CDF) of the simulated distribution and $F_{\mathrm{data},i}^{-1}$ is the quantile function (inverse CDF) derived from data. The correction is performed separately in bins of track pseudorapidity (η) to account for η-dependent resolution effects.

The η binning used for the $Z \to \mu\mu$ and $Z \to ee$ control samples is defined as follows:

- $Z \to \mu\mu$: $|\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, $2.1 < |\eta| < 2.4$,
- $Z \to ee$: $|\eta| < 0.9$, $0.9 < |\eta| < 1.48$, $1.48 < |\eta| < 1.65$, $1.65 < |\eta| < 2.1$, $2.1 < |\eta| < 2.5$.

Within each bin, the cumulative and quantile functions are derived and used to map the uncalibrated simulated IP distribution to the data reference. This correction is applied only to simulated events.

To preserve the physical displacement induced by genuine particle lifetimes, such as those from τ decays, the correction is applied to the difference between the reconstructed and generator-level IP vectors. The final corrected $\vec{IP}$ vector is then constructed as:

$$\mathrm{IP}_i^{\mathrm{corr}} = \mathrm{IP}_i^{\mathrm{gen}} + F_{\mathrm{data},i}^{-1}\left(F_{\mathrm{MC},i}\left(\mathrm{IP}_i - \mathrm{IP}_i^{\mathrm{gen}}\right)\right), \tag{6.7}$$

where $\mathrm{IP}_i^{\mathrm{gen}}$ is the generator-level value of the i-th $\vec{IP}$ component. This procedure ensures that detector resolution mismodeling is corrected without biasing the physical decay lengths of particles in the simulation.

V. Fake Rate Estimation for τ_h

V.1. Background Overview

A significant background in final states involving τ_h arises from the misidentification of quark- or gluon-initiated jets as τ_h candidates. This effect is especially pronounced in the fully hadronic final state, and is predominantly contributed by QCD multijet processes.

The modeling of this background in simulation is limited by the poor description of jet-to-τ_h fake rates and the computational cost of generating large enough samples to probe the rare misidentification process. To circumvent these limitations, a data-driven technique, the Fake Factor (F_F) method, is employed to estimate the contribution of jet-induced τ_h candidates in the signal region (SR).

V.2. Methodology

The F_F method is based on the ABCD technique, wherein orthogonal selection criteria are used to define control and application regions. These include combinations of the τ_h identification working points and the charge correlation between the τ_h pair. The key idea is to measure the misidentification rate (i.e., the fake factor) in dedicated derivation regions (DRs), and apply it to a sideband, or application region (AR), to predict the contribution of fake τ_h candidates in the SR.

In the fully hadronic ($\tau_h\tau_h$) channel, the regions are defined as follows:

- Region A (AR): OS $\tau_h\tau_h$ pair, leading τ_h fails VTight but passes VVVLoose DeepTau ID.
- Region B (DR): SS $\tau_h\tau_h$ pair, leading τ_h fails VTight but passes VVVLoose ID.
- Region C (DR): SS $\tau_h\tau_h$ pair, leading τ_h passes VTight ID.
- Region D (SR): OS $\tau_h\tau_h$ pair, both τ_h candidates pass VTight ID.

The schematic layout of this ABCD configuration is shown in Figure 6.5.

The fake factor is defined as:

$$F_F = \frac{N_{\text{pass DR}}}{N_{\text{fail DR}}}, \tag{6.8}$$

where $N_{\text{pass DR}}$ and $N_{\text{fail DR}}$ are the background-subtracted event counts in Region C and B, respectively. The estimate for the number of fake background events in the SR is then given by:

$$N_{\text{SR}} = N_A \cdot F_F, \tag{6.9}$$

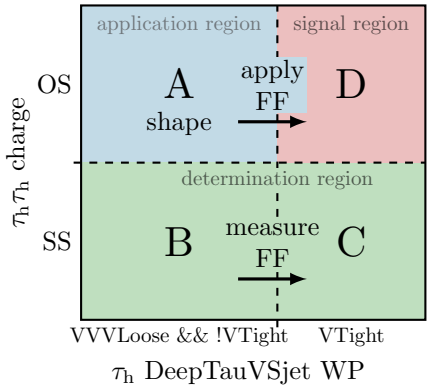


Figure 6.5: Schematic representation of the ABCD method used to estimate the $j \to \tau_h$ fake background. Region D represents the signal region (SR), A the application region (AR), and B/C the derivation regions (DRs). The fake factor is measured in the DRs and applied to the AR to predict the SR yield.

with N_A denoting the background-subtracted event yield in Region A.

To account for variations in fake rates across kinematic regimes, fake factors are derived in bins of jet multiplicity ($N_{\text{jet}} = 0$, 1, and ≥ 2) and by decay mode (DM0, DM1, DM2, and DM10) of the leading τ_h candidate. In each bin, the fake factor is parametrized as a function of the leading τ_h transverse momentum, p_{T}, using a third-order polynomial fit. For $p_{\text{T}} > 100$ GeV, the fake factor is fixed to a constant value due to limited statistics.

Figure 6.6 presents the fitted fake factors for the 2023 dataset, split between pre-BPix and post-BPix data-taking periods.

A closure test is performed by applying the fake factor to Region A (AR) and comparing the resulting background prediction with the data observed in that region. This serves as a validation of the F_F method. The distributions of key kinematic observables from this closure test are shown in Figure 6.7. In addition, the so-called "control plots" (Figure 6.8) show the predicted background obtained by applying the fake factors to the inclusive signal region, providing an additional cross-check of the method.

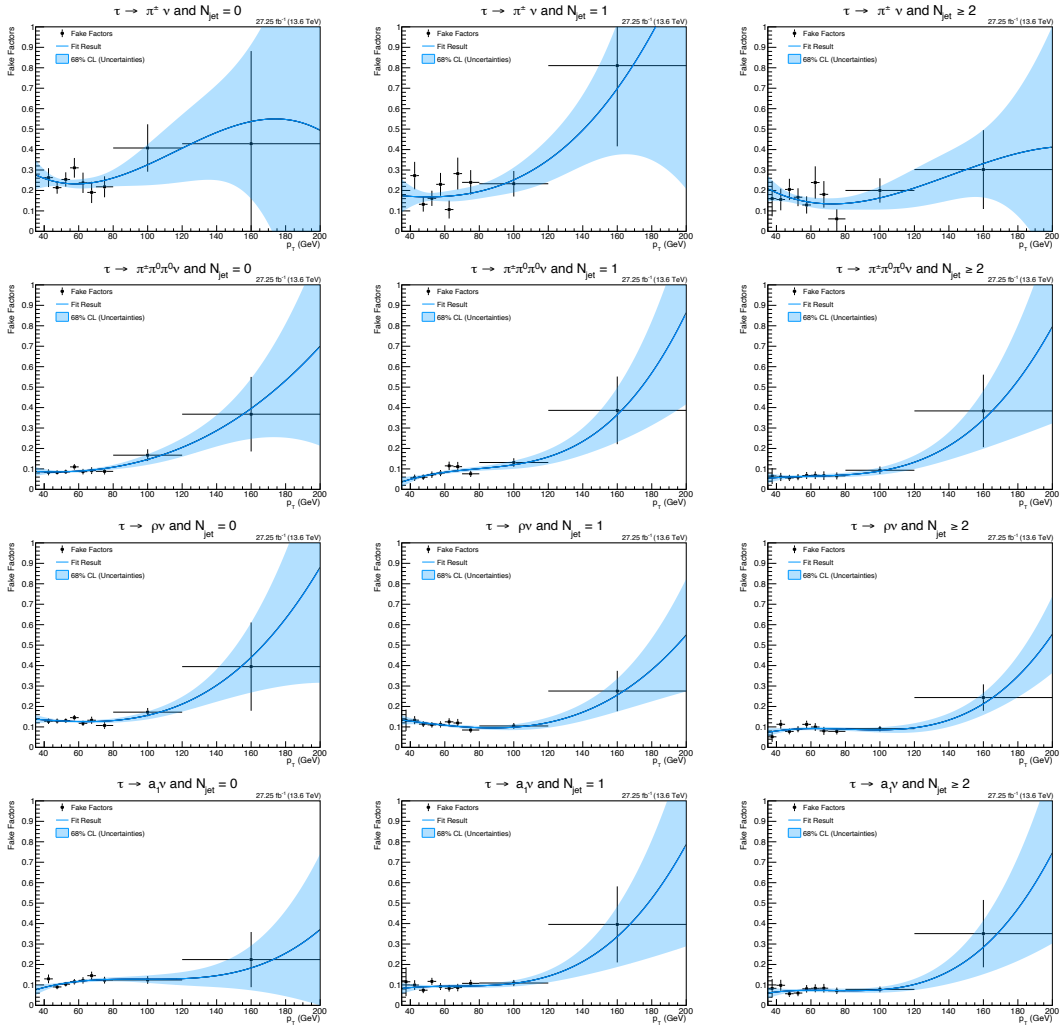


Figure 6.6: Fake factors measured in the combination of 2023 preBPix and 2023 postBPix data for the different N_{jet} multiplicities and decay modes of the leading τ_h candidate. The FFs are fitted with a polynomial function of the order 2, shown as the solid line, and the 68% CL contour is shown as the shaded area.

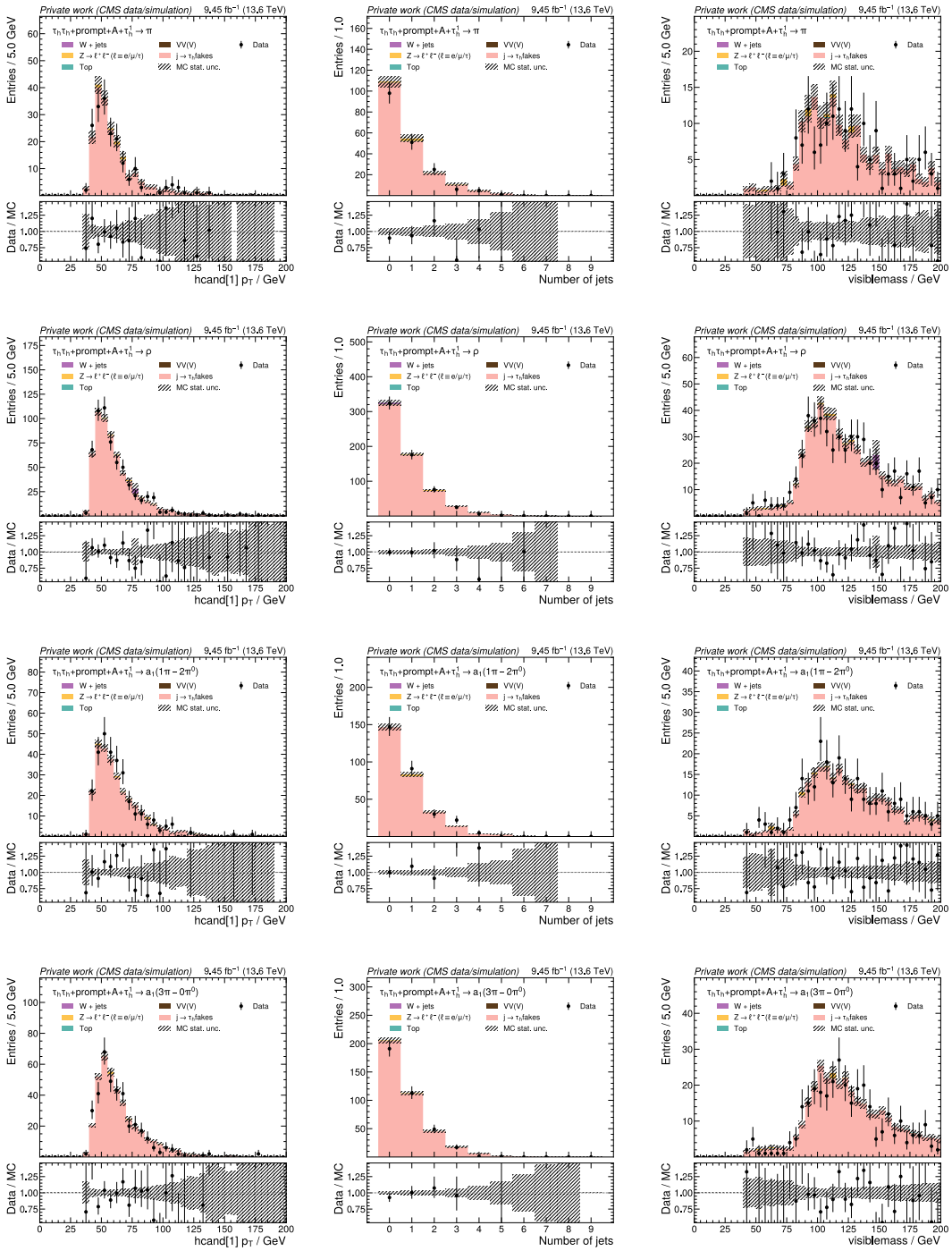


Figure 6.7: Control plots as a closure test for 2023 BPix era.

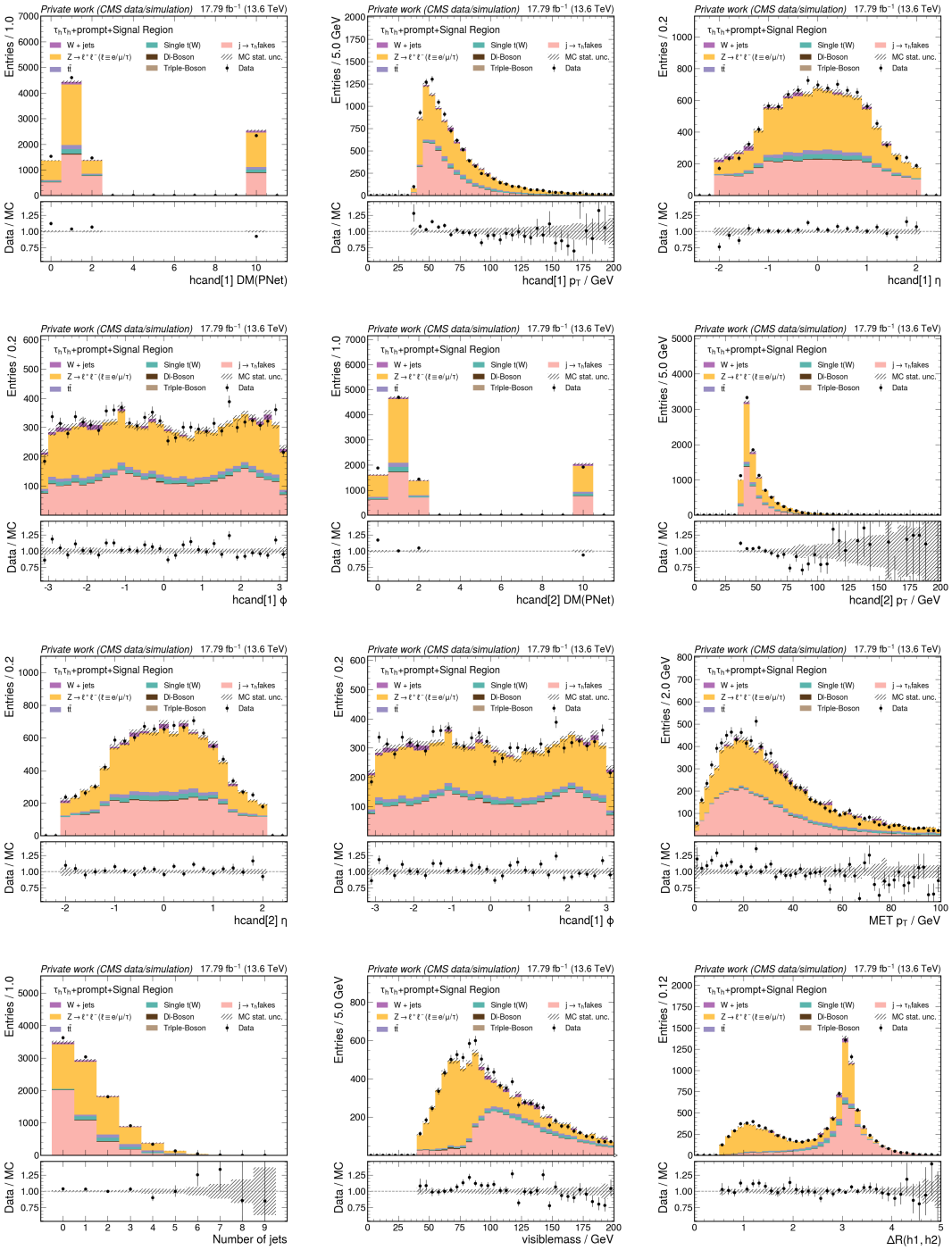


Figure 6.8: Control plots of the Signal Region for 2023 BPix era.

V.3. Systematic Uncertainties

The systematic uncertainty on the FF is estimated using the 68% confidence band obtained from the polynomial fit to the measured fake rates as a function of the leading τ_h transverse momentum.

Since a third-order polynomial function is used to parametrize the dependence,

$$F(p_T; \vec{p}) = p_0 + p_1 p_T + p_2 p_T^2 + p_3 p_T^3, \tag{6.10}$$

the uncertainty envelope around the fitted function is computed by propagating the uncertainties and correlations of the fit parameters $\vec{p} = (p_0, p_1, p_2, p_3)$.

This is achieved by evaluating the variance of the fit function at each point p_T using the covariance matrix $\mathbf{C}$ of the fit parameters:

$$\sigma_F^2(p_T) = \mathbf{J}^T(p_T)\,\mathbf{C}\,\mathbf{J}(p_T), \tag{6.11}$$

where $\mathbf{J}(p_T) = (1, p_T, p_T^2, p_T^3)^T$ is the Jacobian vector of partial derivatives of F with respect to the parameters.

To properly account for correlations between parameters and to obtain statistically rigorous uncertainty bands, the covariance matrix is diagonalized as

$$\mathbf{C} = \mathbf{U}\,\mathbf{D}\,\mathbf{U}^T, \tag{6.12}$$

where $\mathbf{D}$ is a diagonal matrix containing eigenvalues λ_i, and $\mathbf{U}$ is an orthogonal matrix whose columns are the eigenvectors.

Transforming the Jacobian vector into the eigenbasis,

$$\tilde{\mathbf{J}}(p_T) = \mathbf{U}^T \mathbf{J}(p_T), \tag{6.13}$$

the variance can be expressed as a sum of independent contributions:

$$\sigma_F^2(p_T) = \sum_{i=0}^{3} \lambda_i \tilde{J}_i^2(p_T), \tag{6.14}$$

where $\tilde{J}_i(p_T)$ are the components of the transformed Jacobian vector.

The uncertainty envelope is thus given by

$$F(p_T) \pm \sigma_F(p_T), \tag{6.15}$$

which defines the 68% confidence interval around the nominal fit.

This approach ensures a proper treatment of parameter correlations and yields a reliable estimate of the systematic uncertainty on the fake factor parametrization.

VI. τ_h Momentum Reconstruction and CP Observable

VI.1. SVFit and FastMTT

Reconstructing the invariant mass of τ lepton pairs is essential for distinguishing the $H \to \tau\tau$ signal from dominant backgrounds such as $Z \to \tau\tau$. Due to the presence of neutrinos in τ decays, a significant fraction of the decay energy is undetected, degrading the mass resolution when using only the visible decay products.

The *SVFit* algorithm (*Secondary Vertex Fit*) [41] is a likelihood-based method developed by CMS to reconstruct the full di-τ four-momentum. It uses the visible decay products, the MET, and the MET covariance matrix. A Dynamical Likelihood Method is employed to estimate the probability of the observed event kinematics given a hypothesized value of the invariant mass $m_{\tau\tau}$, by integrating over the unmeasured decay parameters.

For each τ decay, the kinematics in the τ rest frame are parametrized by:

- θ_{inv}: the polar angle between the visible and invisible momenta,
- ϕ_{inv}: the azimuthal angle of the invisible momentum in the transverse plane,
- $m_{\nu\nu}$: the invariant mass of the neutrinos (only for leptonic decays).

Hadronic decays require two integration variables, while leptonic decays require three. The likelihood is defined as:

$$\frac{dL(M_{\tau\tau})}{dM_{\tau\tau}} = \int \frac{df(\vec{x}_u|\vec{x}_m)}{d\vec{x}_u}\,\delta\left(M_{\tau\tau} - M_{\tau\tau}(\vec{x}_u, \vec{x}_m)\right), \tag{6.16}$$

where $\vec{x}_u$ are the unknown parameters and $\vec{x}_m$ the measured observables.

SVFit improves the mass resolution and separation between the Higgs and Z/γ^* peaks. This translates into a $\sim 30\%$ gain in analysis sensitivity, equivalent to a 70% increase in effective luminosity [41]. However, the computational cost of the multidimensional integration is significant.

To address this, a faster approximation, *FastMTT* [41], has been developed. It assumes the neutrino momenta are collinear with the visible τ decay products, enabling an analytical solution. Figure 6.9 shows the di-τ mass resolution obtained with both methods on simulated Drell–Yan events. SVFit and FastMTT achieve comparable performance, with standard deviations of 0.17 and 0.16 respectively.

During this PhD thesis, SVFit and FastMTT have both been implemented. An initial version of SVFit was integrated in the analysis framework, and the original C++ implementation of FastMTT was wrapped in Python. A standalone Python implementation of FastMTT has also been developed and is used.

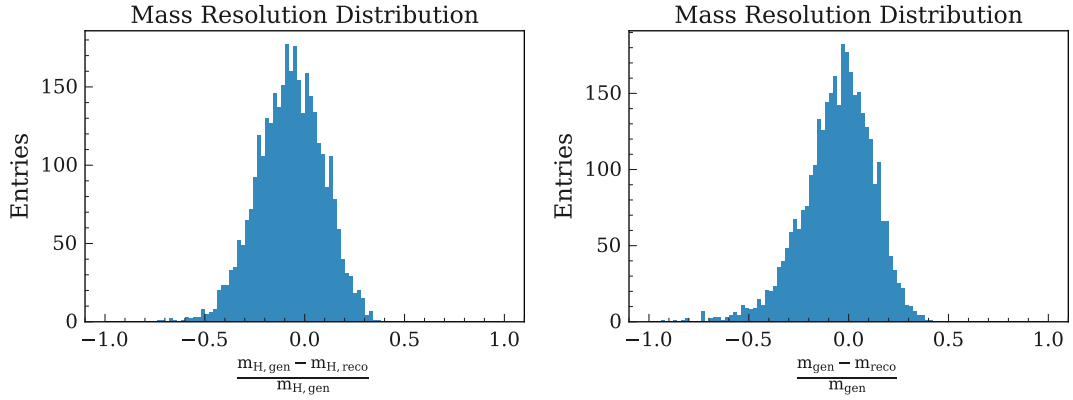


Figure 6.9: Di-τ mass resolution from FastMTT (left) and SVFit (right) on simulated DY events in columnflow. The standard deviation is 0.16 for FastMTT and 0.17 for SVFit.

VI.2. Machine Learning Studies in the $\rho\rho$ Channel

VI.2.a. Motivation

Among all hadronic decay modes of τ_h leptons, the $\rho\rho$ final state is particularly promising for probing CP violation in $H \to \tau\tau$ decays. It benefits from a relatively large branching ratio (25.9% of the τ_h decays), and constitutes one of the statistically dominant hadronic decay modes.

In addition, the $\rho\rho$ channel provides enhanced sensitivity to the CP-violating phase ϕ_{CP} when employing the polarimetric vector method (PV), in contrast to decay plane method (DP) based solely on the decay plane of the neutral pions (see Section IV.). At generator level, the CP-sensitive angular asymmetry reaches $A_{\mathrm{gen}} = 0.418$ for the PV-PV method, compared to $A_{\mathrm{gen}} = 0.332$ using the DP-DP method, as shown in Figure 6.10.

To quantify the CP sensitivity, the asymmetry is defined as:

$$A = \frac{1}{N_{\mathrm{bins}}} \sum_{i=1}^{N_{\mathrm{bins}}} \left| \frac{N_i^{\mathrm{odd}} - N_i^{\mathrm{even}}}{N_i^{\mathrm{odd}} + N_i^{\mathrm{even}}} \right|, \tag{6.17}$$

where N_i^{odd} and N_i^{even} represent the weighted event yields in bin i of a CP-sensitive angular observable, using event weights corresponding to the pure CP-odd ($\phi_\tau = \pi/2$) and pure CP-even ($\phi_\tau = 0$) hypotheses, respectively. This asymmetry captures the bin-by-bin shape difference between the angular distributions under the two extreme CP scenarios and serves as a proxy for the observable's sensitivity to ϕ_{CP}.

However, this gain in sensitivity is substantially degraded at reconstruction

level using the τ visible four-momentum. While the DP method preserves a significant asymmetry ($A_{\text{reco}} = 0.265$), the PV method exhibits a nearly flat distribution ($A_{\text{reco}} = 0.010$), as illustrated in Figure 6.11. This reduction is primarily driven by the loss of information associated with the unmeasured neutrino momentum, which is crucial for computing the polarimetric observables. Because of this, in Run 2 analyses, the DP–DP method was used for ρ–ρ decays, while the PV–PV method was used only for decays involving the a_1 resonance [17].

In the ρ decay mode, the polarimetric vector is defined as:

$$\vec{h} = \mathcal{N}\left(2(q \cdot N)\vec{q} - q^2\vec{N}\right), \tag{6.18}$$

where the hadronic vector is

$$\vec{q} = \vec{p}_{\pi^\pm} - \vec{p}_{\pi^0}, \tag{6.19}$$

and the neutrino momentum is obtained as

$$\vec{p}_\nu = \vec{p}_\tau - (\vec{p}_{\pi^\pm} + \vec{p}_{\pi^0}), \tag{6.20}$$

all quantities being defined in the τ_h rest frame. The normalization factor $\mathcal{N}$ ensures unit norm for the polarimetric vector.

At reconstruction level, the full four-momentum of the τ_h is not directly measurable due to undetected neutrinos in the decay. As a result, reconstructing the τ_h rest frame, and thus the polarimetric vector, requires estimates of both the τ_h energy and its flight direction. These are typically inferred using algorithms like FastMTT, which combine the visible decay products with the MET. However FastMTT assumes the collinear approximation, where the neutrino is aligned with the visible decay products which is the same hypothesis used in the DP method.

To recover some of the lost sensitivity, a ML approach is explored. The idea is to train regression models on simulated events to learn a mapping from reconstructed observables (such as visible τ_h decay products, MET, and their covariances) to τ_h kinematic quantities important for CP measurements, such as the energy, direction, or momentum. The goal is to improve the separation between CP-even and CP-odd hypotheses and to recover the generator-level sensitivity.

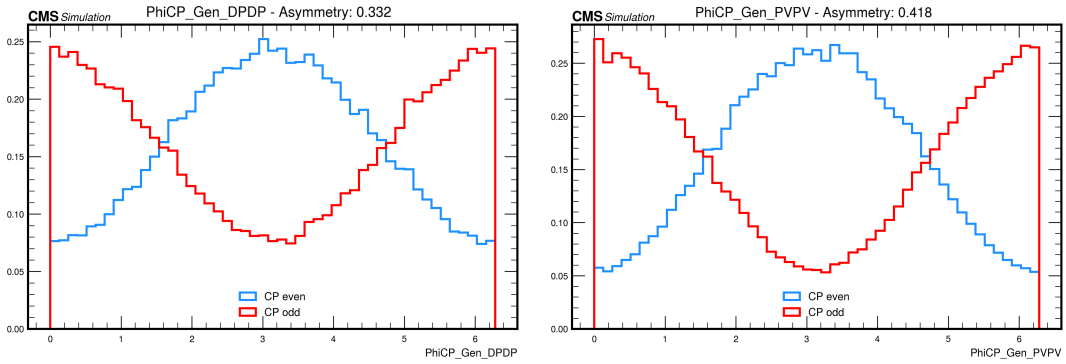


Figure 6.10: Decay Plan method (left) and Polarimetric vector method (right) using generator level information.

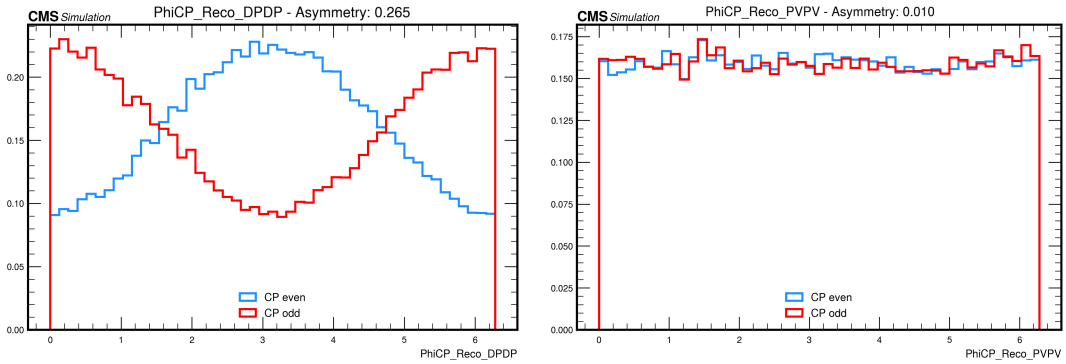


Figure 6.11: Decay Plan method (left) and Polarimetric vector method (right) using reconstructed level information (τ_h visible four-momentum).

VI.2.b. Lorentz Boost Network

Motivations: Conventional deep networks often struggle to learn the underlying physics when dealing with four-vectors, due to their lack of knowledge of Lorentz invariance and frame-dependent transformations. The Lorentz Boost Network (LBN) [210] addresses this by explicitly constructing composite systems and boosting to their respective rest frames. This enables the model to extract physically meaningful observables, such as decay angles and invariant masses, thereby improving the representation of kinematic information relevant for processes like $H \to \tau\tau$.

Architecture and Operation: The model's architecture detailed in the following was chosen after testing and tuning different hyperparameters. The goal was to find a setup that works well without being too complex or overfitting.

- **Inputs:** The LBN receives 7 four-vectors as input: the visible decay products of the two taus (τ_1, τ_2), four leading jets $(j_1 \dots j_4)$, and the missing transverse energy (MET), each expressed as (E, p_x, p_y, p_z). Additional auxiliary variables (e.g., event-level kinematic quantities or category labels) are appended after the LBN layer.

- **LBN Layer:** As illustrated in Figure 6.12, the LBN core layer constructs $N = 10$ trainable linear combinations of the input four-vectors. These represent intermediate candidate systems (e.g., τ_h-pair, visible Higgs decay, recoil system). For each combination, a Lorentz boost is applied to transform the event into its rest frame. Several tests indicate that $N = 10$ is near optimal for the number of LBN components in this study.

- **Feature Construction:** In these boosted frames, the LBN extracts physics-inspired features, including:

  - invariant masses of candidate systems,
  - angles between decay products (e.g. $\cos\theta^*$, azimuthal separations),
  - energy ratios between constituents and parents.

  These features form a compact and interpretable latent representation of the event's kinematics.

- **Dense Network:** The extracted features are concatenated with auxiliary inputs and passed to a fully connected neural network with the architecture:

  `[512 → 256 → 64]` nodes per layer, using `tanh` activation and a dropout rate of 0.2.

The final output is an 8-dimensional vector, corresponding to the predicted four-momenta of the two τ leptons: $(E, p_x, p_y, p_z)_{\tau_1}, (E, p_x, p_y, p_z)_{\tau_2}$.

- **Training:** The model is trained using a custom hybrid loss function that combines:
  - a MSE loss for four-momentum regression,
  - a histogram-based loss to preserve the statistical distribution of key observables such as the reconstructed ditau mass.

  Optimization is performed with the Adam optimizer using a learning rate of 10^{-2} and weight decay of 10^{-4}.

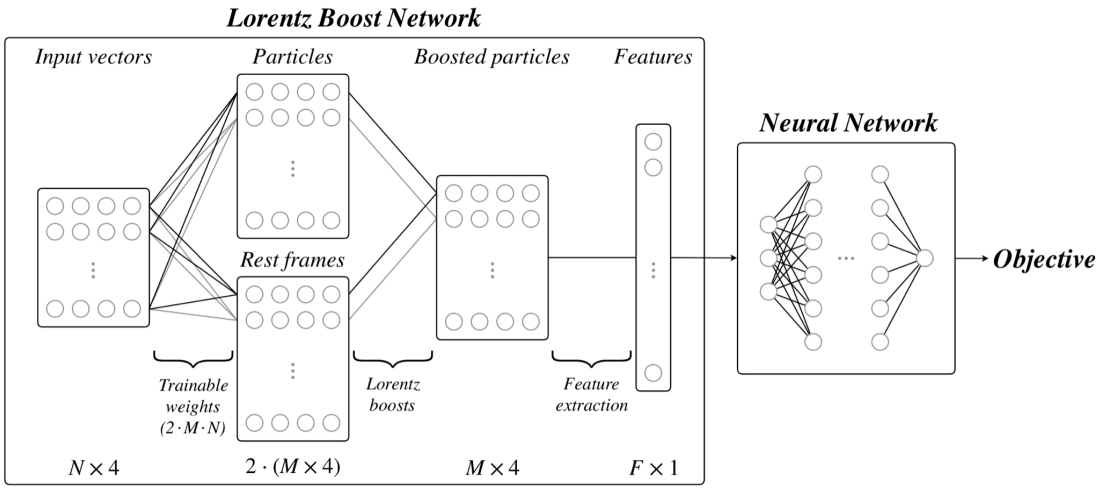


Figure 6.12: Schematic representation of the Lorentz Boost Network (LBN) architecture, including the trainable four-vector combination, boost to rest frames, and physics feature extraction.

VI.2.c. Performance of the Network

τ_h and Pion Target Regression: To evaluate the performance of the LBN in reconstructing τ_h decay products, the model is first trained to simultaneously regress the full four-momenta of the τ_h lepton and its visible decay products (pions), so 24 distributions to regress: p_x, p_y, p_z and E for each τ_h, neutral and charged pion.

As shown in Figure B.1 in Appendix B Section II., the regression for the τ_h leptons performs well across all components, especially for the p_x, p_y, p_z components. However, the pion-level regression shows reduced accuracy, likely due to

the highly non-Gaussian and sharply peaked distribution of pions, making it more difficult for the network to generalize. The resulting CP sensitivity is comparable to that obtained using reconstructed-level information alone.

τ_h-only Regression: Focusing the regression target solely on the τ_h leptons p_x, p_y, p_z and E improves the performance, since the network learns from a more uniform distribution with consistent trends. This allows the model to train more effectively and improves overall performance. Figure B.2 in Appendix B Section II. shows improved agreement with the truth-level τ kinematics. However, the reconstruction of the τ energy still has room for improvement.

Additional Inputs and Constraints: To enhance regression performance further, the following constraints are added:

- angular information between the ρ meson and the parent τ lepton,
- constraints on the invariant mass of the Higgs system implemented as a penalty term in the loss function to enforce the consistency with the known Higgs mass.

The regressed distributions are summarized in Figure B.3 in Appendix B Section II. The τ_h energy regression improves with the additional constraints. In particular, the invariant mass constraint helps the model learn physically meaningful correlations by penalizing deviations from the expected ρ mass. However, the angle is not well regressed because its distribution lacks the sharp features that the LBN can effectively learn, representing a limitation of the method.

Results on ϕ_{CP} Sensitivity: Figure 6.13 summarizes the CP sensitivity achieved with different approaches. The fully regressed LBN model (top) shows no significant separation between CP-even and CP-odd scenarios, indicating limited resolution in the regressed η and ϕ variables. However, combining partial regression (e.g., transverse momentum) with directions reconstructed from detector observables achieves sensitivity comparable to the standard DPDP method. Studies show that using generator-level p_T for PVPV and all reconstructed information does not significantly improve the PVPV method. This implies the regressed p_T is already well predicted by the ML algorithm, and further gains in sensitivity must come from improving the regression of the η and ϕ distribution.

These angular parameters directly affect the construction of polarimetric vectors and decay planes, dominating the CP observable resolution. Therefore, future model development will focus on improving the reconstruction of angular directions, potentially via hybrid models that combine learned regression with measured directions and constraints from τ_h decay kinematics.

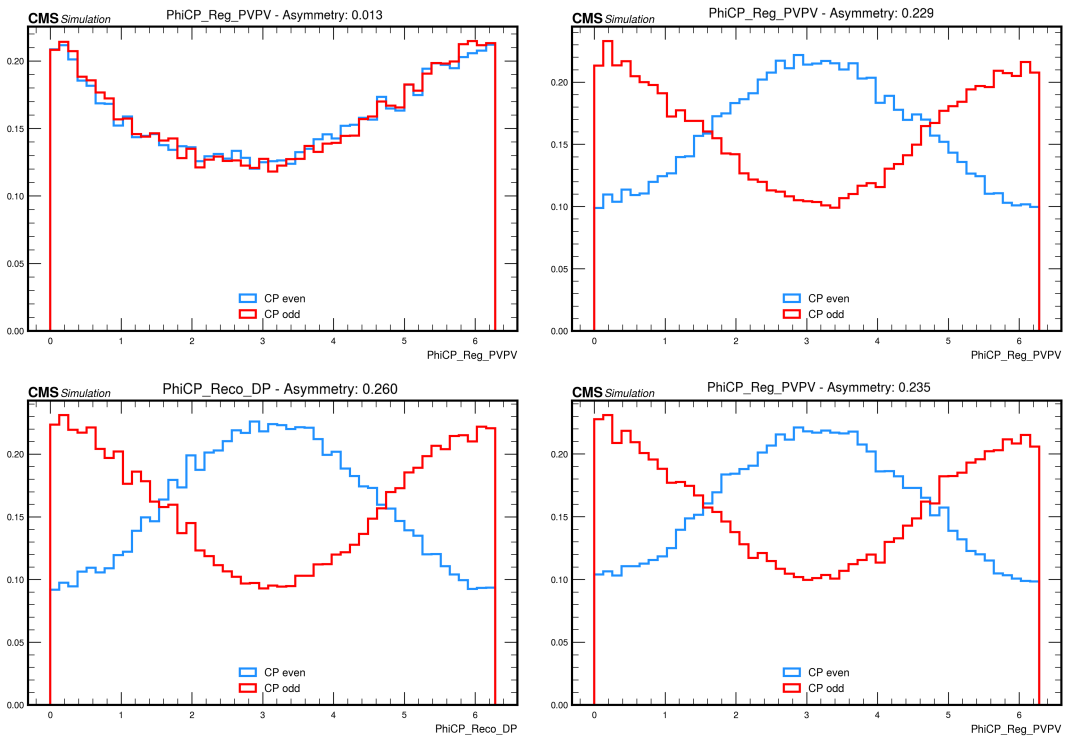


Figure 6.13: Comparison of the ϕ_{CP} distributions: (Top left) full four-momentum regression using the PVPV method; (Top right) partial regression (only p_T) combined with reconstructed information (visible τ_h; (Bottom left) reconstruction-based Decay Plane (DPDP) method using reconstructed information (visible τ_h); (Bottom right) PVPV method using generator-level p_T and all other reconstructed information.

VII. Event Categorization and Signal Discrimination

To exploit differences in event topology, events are further split according to PNet DMs [211]. Within each category, a machine learning classifier based on boosted decision trees is used to enhance separation of signal from background. The classifier is trained to distinguish between three classes:

1. Higgs signal events,
2. backgrounds with genuine tau leptons (mainly $Z \to \tau\tau$), and
3. backgrounds where a jet or lepton is misidentified as a τ_h.

Each event is assigned class probabilities, and events are labeled according to the class with the highest score. The associated probability is then used as the final discriminant. This approach allows the construction of robust, class-specific templates for the likelihood fit.

VIII. Sensitivity Scan of ϕ_{CP}

The event categorization strategy described in Section VII. enables the extraction of the CP-mixing angle $\alpha_{H\tau\tau}$ from data via a maximum likelihood fit. A simultaneous fit across all signal and background categories is performed using the Combine tool [42].

The distribution of the CP-sensitive observable ϕ_{CP} can be parametrized for any arbitrary value of $\alpha_{H\tau\tau}$ by a linear combination of the distributions corresponding to the pure CP-even and CP-odd hypotheses (see Equation 6.5). The parameter $\alpha_{H\tau\tau}$ is treated as the parameter of interest (POI) in the fit and is allowed to float within the range $[-90°, +90°]$.

An preliminary expected sensitivity scan is conducted using the so-called Higgs class, comprising all decay mode categories with Asimov dataset [43] including all signal and background samples. Only statistical uncertainties are considered in this preliminary study. The fit is applied to the 2022 postEE datasets, corresponding to an integrated luminosity of 26.671 fb^{-1}, and the resulting scan of the likelihood profile is shown in Figure 6.14. Only one category is considered in the Higgs class.

This preliminary result, which needs further validation, confirms the robustness of the analysis implementation and validates the event categorization strategies. They also highlight the promising sensitivity expected from the full Run 3 dataset, which is anticipated to yield significantly improved constraints on CP violation in Higgs–τ interactions.

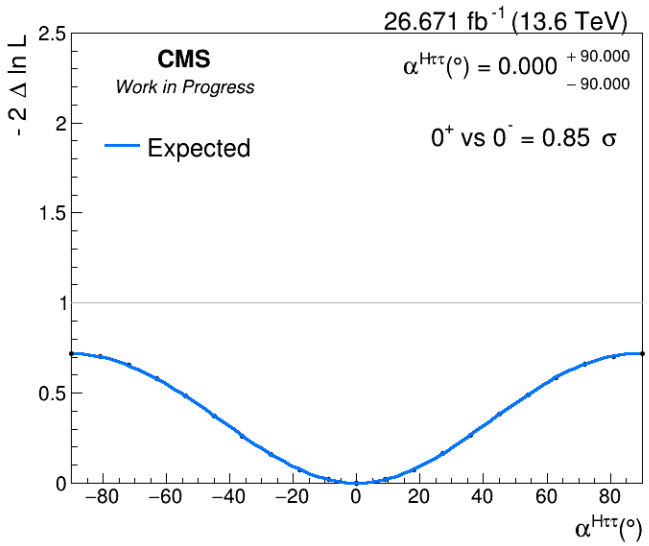


Figure 6.14: Profile likelihood scan of the CP-mixing angle $\alpha_{H\tau\tau}$ obtained from a simultaneous fit across all DM categories in the 2022 postEE Asimov dataset, corresponding to an integrated luminosity of 26.671 fb^{-1}. Only statistical uncertainties are included.

IX. Conclusion

This chapter presented the $H \to \tau_h\tau_h$ analysis strategy using early Run 3 CMS data collected at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV, focusing on the fully hadronic final state which benefits from a large branching ratio.

As part of this effort, I contributed to the development and integration of the full analysis pipeline within a columnar analysis framework. This included the implementation of the selection logic, categorization scheme, and object calibration. In particular, I implemented and validated the Fake Factor method on Run 3 data, adapting it to the specifics of the updated detector conditions and pileup environment.

To improve the reconstruction of CP-sensitive observables, I also explored the use of machine learning techniques. A regression model based on Lorentz Boost Networks (LBNs) was trained to reconstruct the full τ four-momentum in the $\rho\rho$ final state. While improvements in transverse momentum resolution were observed, the limited angular resolution, especially in η and ϕ, remained a key limitation to enhancing the resolution of the CP-sensitive observable ϕ^*_{CP}. To address this, additional physical constraints were incorporated into the training loss, including terms sensitive to invariant mass and angular correlations between decay products. Moderate gains were achieved, though further refinement is necessary.

Beyond the LBN approach, future work will explore more expressive architectures such as Graph Neural Networks (GNNs), which are well-suited to capturing the complex angular correlations between τ decay products. These models offer promising avenues to overcome the limitations of current architectures. On the event categorization front, the analysis will be extended to include categories targeting different Higgs production modes, enabling sensitivity optimization across multiple signal regions.

Preliminary sensitivity scans performed on early Run 3 data indicate promising constraints on the CP-mixing angle $\alpha_{H\tau\tau}$, with the potential to reach sensitivities comparable to or better than those achieved in Run 2, despite the lower integrated luminosity. These early results confirm the efficacy of the analysis approach, in particular the τ_h reconstruction and identification techniques.

In the short term, updated Fake Factors will be produced using corrected simulation samples. The previously identified bug affecting the DM0 category is not expected to significantly affect the $\rho\rho$ final state targeted in the ML studies. The immediate outlook includes producing the full early Run 3 sensitivity study.

The final goal of this analysis is to measure CP violation in $H \to \tau\tau$ using the full Run 3 dataset and to combine the results with Run 2 to maximize sensitivity. This will enable the most stringent constraints to date on the CP-mixing angle in the Higgs–τ coupling.

Conclusion

The SM of particle physics provides a remarkably successful framework for describing fundamental particles and their interactions. However, it does not fully account for the observed matter–antimatter asymmetry in the universe, which motivates the search for new sources of CP violation beyond the well-established quark sector. The discovery of the Higgs boson in 2012 opened a promising window into such effects, particularly through its couplings to fermions like the τ lepton. While no deviation from the SM predictions has yet been observed, this thesis focuses on probing CP violation in the Higgs boson decay to hadronic τ pairs ($H \to \tau_h \tau_h$) with CMS Run 3 data. The key CP-sensitive observable is constructed from the plane defined by the decay products of the τ leptons, which carry information about the τ spin correlations and are therefore sensitive to the CP nature of the Yukawa coupling between the Higgs and τ leptons.

The first study presented in this thesis concerns the development and validation of the irradiation setup for Phase-2 pixel modules at the CYRCé facility. This work was carried out in the context of the CMS Inner Tracker Phase-2 upgrade, where pixel modules must be qualified to withstand the unprecedented radiation levels expected at the HL-LHC. My contribution focused on preparing and commissioning the first irradiation campaign at IPHC, including the design and implementation of a stable, cold, and dry environment for module testing, as well as the development of a detailed and reproducible measurement protocol.

The successful irradiation of a CROC-V1 module under cold conditions demonstrated the functionality of both the readout chip and sensor after exposure to the target fluence. A localized non-uniform response detected during post-irradiation testing was investigated and attributed to mechanical issues in the bump-bonding process, rather than to irradiation effects. By isolating this region, the uniformity of the irradiation was confirmed across the remaining detector area.

This work validates the irradiation procedure and the cold testing setup, establishing a robust infrastructure and methodology at IPHC. With this foundation,

the group is now equipped to perform systematic irradiation campaigns on various pixel modules, playing an essential role in the qualification program for the CMS IT Phase-2 upgrade.

The second study presented in this thesis focuses on the combined measurement of the τ_h identification scale factor and energy scale, developed in the context of analyses sensitive to CP violation and other precision measurements relying on τ_h object reconstruction. Accurate calibration of the τ_h identification efficiency and energy response is essential to reduce systematic uncertainties and enhance the sensitivity of these analyses.

My contribution involved the development and validation of a simultaneous fitting technique that extracts both scale factors. This method was developed and tested on Run 2 data before being applied to early Run 3 datasets, where challenges such as statistics limitations were overcome to stabilize the fit and realize the full potential of the combined approach.

The results demonstrate that the combined fit technique provides results compatible with the individual fit method. Limitations arise when the fit is divided into too many fitted regions due to limited statistics. For instance, it is not currently feasible to measure the scale factors as a function of p_T, which is not problematic for analyses not sensitive to such dependencies. Nevertheless, the combined fit method is being adopted for deriving the scale factors for the upcoming Run 3 analyses which has larger event sample size.

The final study presented in this thesis concerns the measurement of CP violation in Higgs boson decays to hadronically decaying τ leptons, $H \to \tau_h\tau_h$, using early Run 3 data collected by the CMS experiment at $\sqrt{s} = 13.6$ TeV. This analysis targets the fully hadronic final state ($\tau_h\tau_h$), which benefits from the largest branching fraction among the $\tau\tau$ decay modes. Like all $H \to \tau\tau$ final states, it presents notable challenges, including a limited mass resolution and substantial backgrounds from Drell–Yan and QCD multijet processes.

My contributions to this work include the development of the analysis within a columnar framework, encompassing event selection, categorization, object calibration, and the implementation of CP sensitive observables in the different channels. Additionally, I measured the misidentification rate of quark and gluon jets as τ_h candidates using a data-driven method, which is essential for accurately estimating reducible backgrounds in the analysis. To improve the reconstruction of CP-sensitive observables, I investigated machine-learning-based techniques, focusing on a Lorentz Boost Network (LBN) for τ_h four-momentum reconstruction. The motivation was to enable the use of the polarimetric vector method, rather than the traditional decay plane method, in the $\rho\rho$ channel, which offers significantly better sensitivity at generator level. This approach is challenging, as it requires reconstructing the full τ momentum including the missing neutrino information.

The results demonstrate that the LBN achieves sufficient performance in reconstructing the transverse momentum component to recover the sensitivity of the decay plane method when replaced by the polarimetric method. However, further improvements in the angular resolution (η and ϕ) are necessary to enhance the overall sensitivity. These gains could be realized either by exploring more complex architectures such as Graph Neural Networks (GNNs) or by incorporating additional physical constraints directly into the LBN training.

Preliminary sensitivity studies performed on early Run 3 data show encouraging constraints on the CP-mixing angle $\alpha_{H\tau\tau}$, confirming the effectiveness of the event categorization and analysis framework. Near-term plans include a full early Run 3 sensitivity study, with potential improvements such as event categorization targeting different Higgs production modes, for example VBF or ggF. The ultimate goal is to combine Run 3 results with those from Run 2 to provide the most stringent constraints on CP violation in the Higgs–τ coupling to date, a critical probe of physics beyond the Standard Model.

Appendix A

Combined Measurement of τ_h Identification and Energy Scale Factor

Content

I. Combined fit of the TES and the ID SF by decay modes and $p_T(\tau_h)$ . . . 194

II. Bug in Pythia 8 . . . 196

III. First Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets . . . 197

IV. Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets . . . 198

IV.1. PreEE 2022 . . . 198

IV.2. PostEE 2022 . . . 199

IV.3. Individual fit results . . . 199

This appendix provides complementary plots and detailed explanations supporting the results presented in Chapter 5.

I. Combined fit of the TES and the ID SF by decay modes and $p_T(\tau_h)$

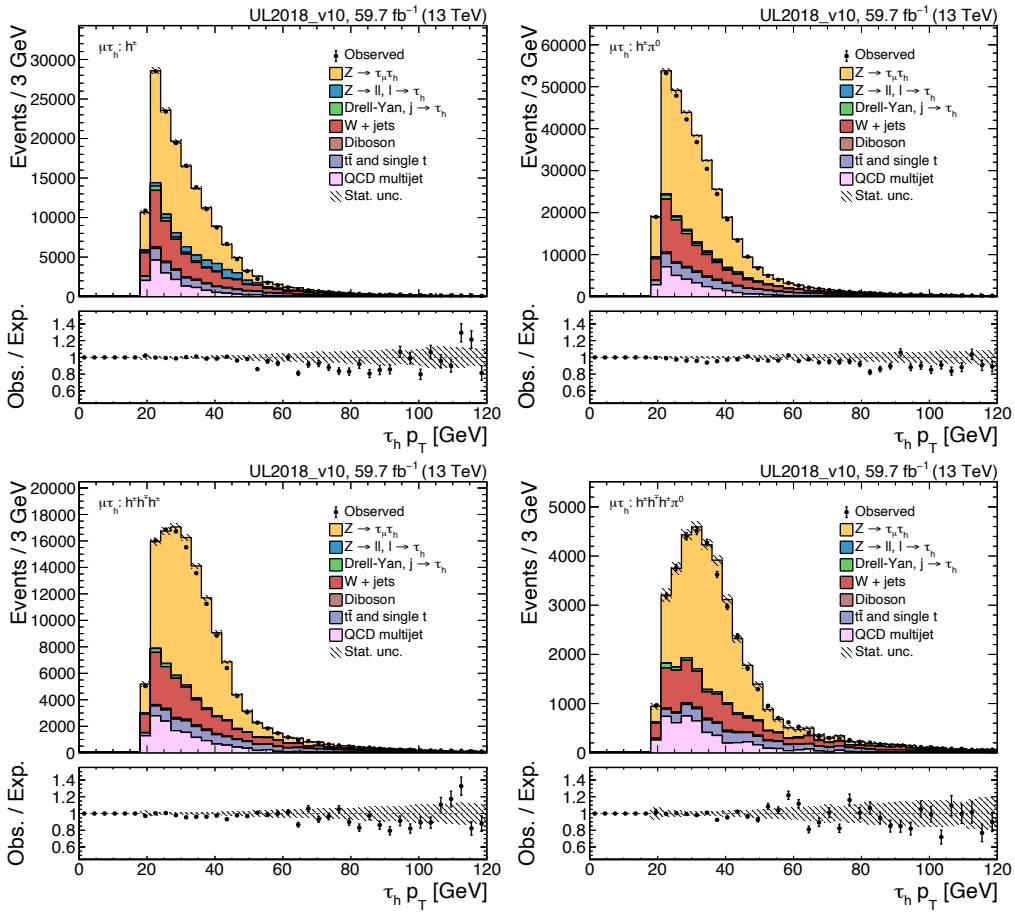
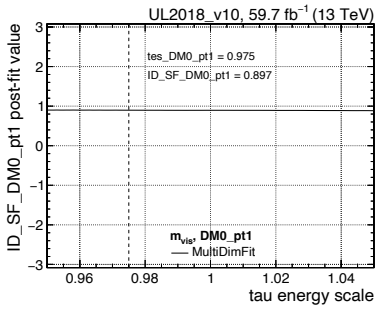
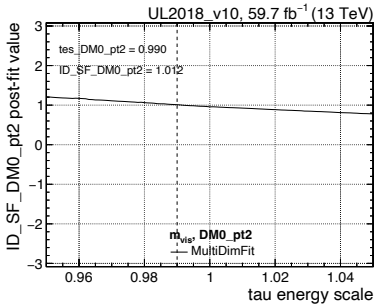


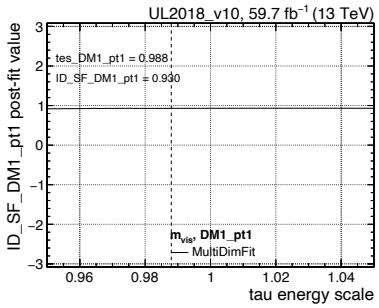
Figure A.1: Pre-fit $p_T(\tau_h)$ distributions for UL2018 across each DM region. The top-left panel shows DM0, top-right shows DM1, middle-left shows DM10, and middle-right shows DM11. A selection cut of $m_T(\mu, \tau) < 65$ GeV on the transverse mass of the muon and tau has been applied.



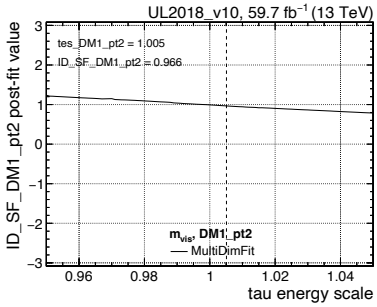
(a) DM0, $20 < p_T(\tau_h) \leq 40$ GeV



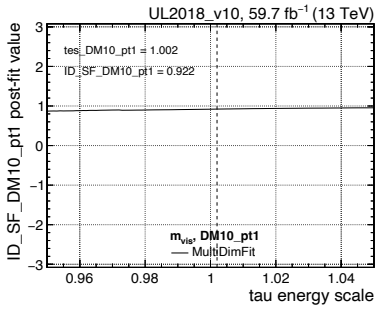
(b) DM0, $p_T(\tau_h) > 40$ GeV



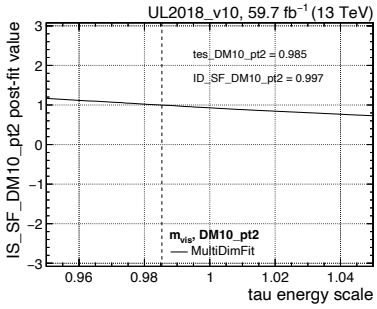
(c) DM1, $20 < p_T(\tau_h) \leq 40$ GeV



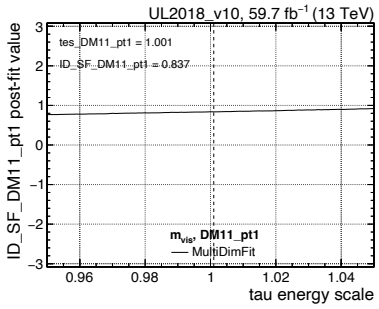
(d) DM1, $p_T(\tau_h) > 40$ GeV



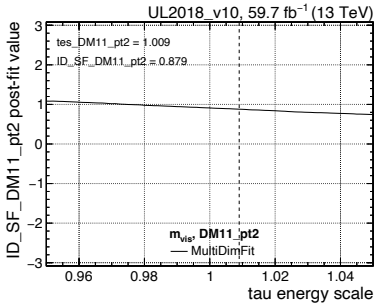
(e) DM10, $20 < p_T(\tau_h) \leq 40$ GeV



(f) DM10, $p_T(\tau_h) > 40$ GeV



(g) DM11, $20 < p_T(\tau_h) \leq 40$ GeV



(h) DM11, $p_T(\tau_h) > 40$ GeV

Figure A.2: Correlation plots between the TES and the ID SF. The scale factors for $20 < p_T(\tau_h) \leq 40$ GeV on the left-hand side and $p_T(\tau_h) > 40$ GeV on the right-hand side.

II. Bug in Pythia 8

An important issue has been discovered in the modeling of τ_h decays in early Run 3 Drell–Yan (DY) samples generated with MadGraph5_aMC@NLO [206] interfaced to Pythia 8 [207]. The problem arises from how τ_h spin correlations and polarization are treated, which impacts the visible transverse momentum (p_T) and energy distribution of the τ_h decay products.

Pythia 8 provides three options for the `TauDecays:externalMode` parameter that control tau spin correlation:

- **Option 0**: Treats τ_h pairs as uncorrelated, using only longitudinal polarization.
- **Option 1** (default for most Run 3 samples): Uses the mother particle's spin to correlate τ_h decays but ignores some spin correlations between incoming and outgoing particles, and uses a slightly different Weinberg angle.
- **Option 2** (used in Run 2): Uses Pythia's internal spin correlation model.

By comparing visible mass and energy distributions of τ_h decays, it was found that the spin treatment in these early Run 3 samples is incorrect. This causes:

- A shift in the mass peak and an overestimation of events in the $\mu\tau_h$ channel (especially for $\tau_h \to$ DM0 decays).
- A large underestimation of events in the $\tau_h\tau_h$ channel for DM0–DM0 and DM0–DM1 decay modes.

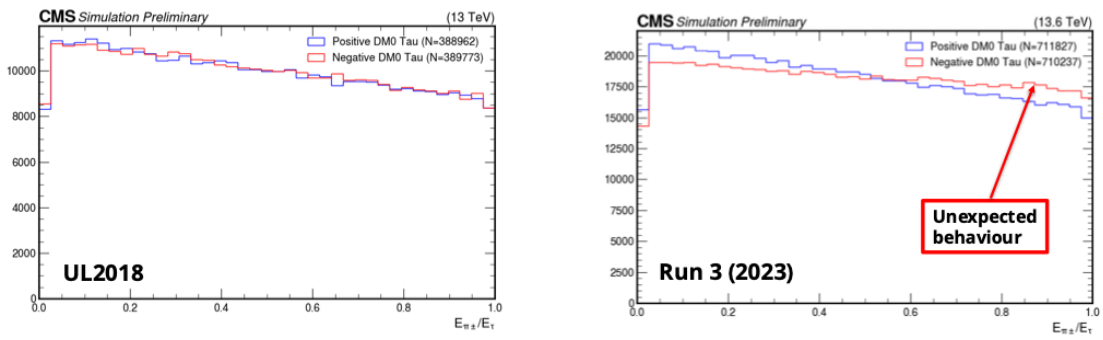


Figure A.3: Comparison of energy fraction carried by visible τ_h between early Run 3 (2023) and UL18 samples using GenPart in NanoAOD.

These issues originate from incorrect spin correlation and polarization modeling in the MadGraph/amcatNLO + Pythia 8 samples, which undermines their

accuracy. As a result, any scale factors or corrections derived from these samples (such as the first set of SF measurements discussed in Appendix III.) are affected.

To address this, a new set of measurements has been performed using samples generated with POWHEG [212], which does not suffer from this bug.

III. First Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets

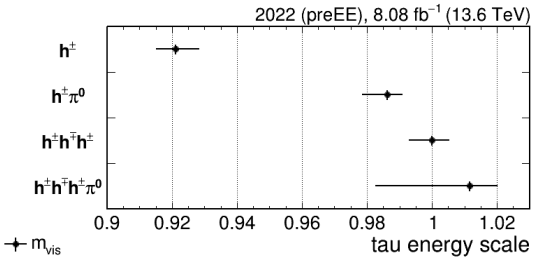


Figure A.4: Obtained TES per DM to be applied on 2022 preEE simulated genuine τ_h.

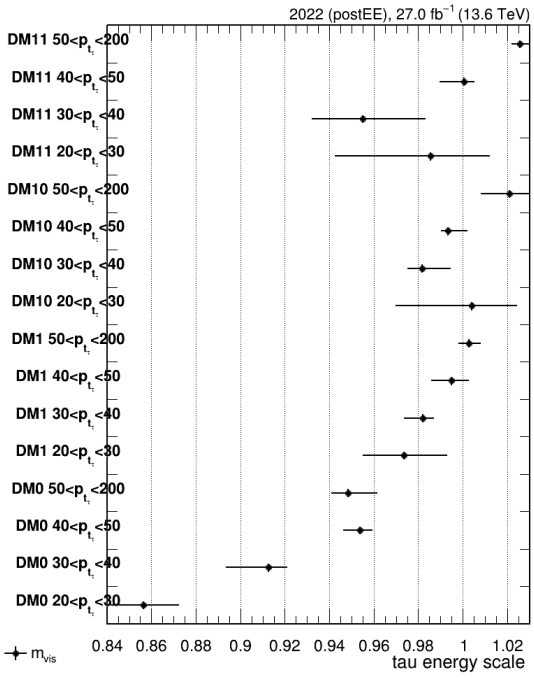


Figure A.5: Obtained TES per DM-p_T to be applied on 2022 postEE simulated genuine τ_h.

IV. Measurements of τ_h Energy Scale in the 2022 Run 3 Datasets

IV.1. PreEE 2022

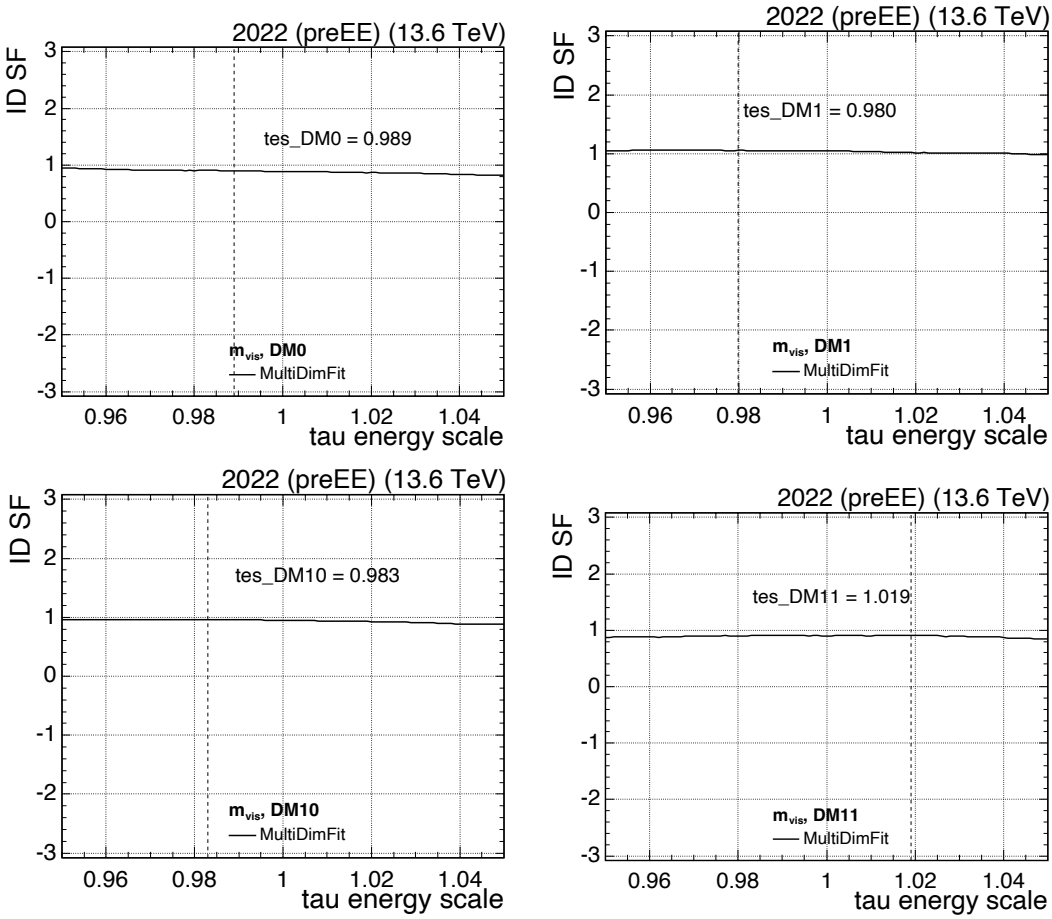


Figure A.6: The ID SF as a function of the TES during the NLL scan of the TES showing the dependency between the SFs.

IV.2. PostEE 2022

IV.3. Individual fit results

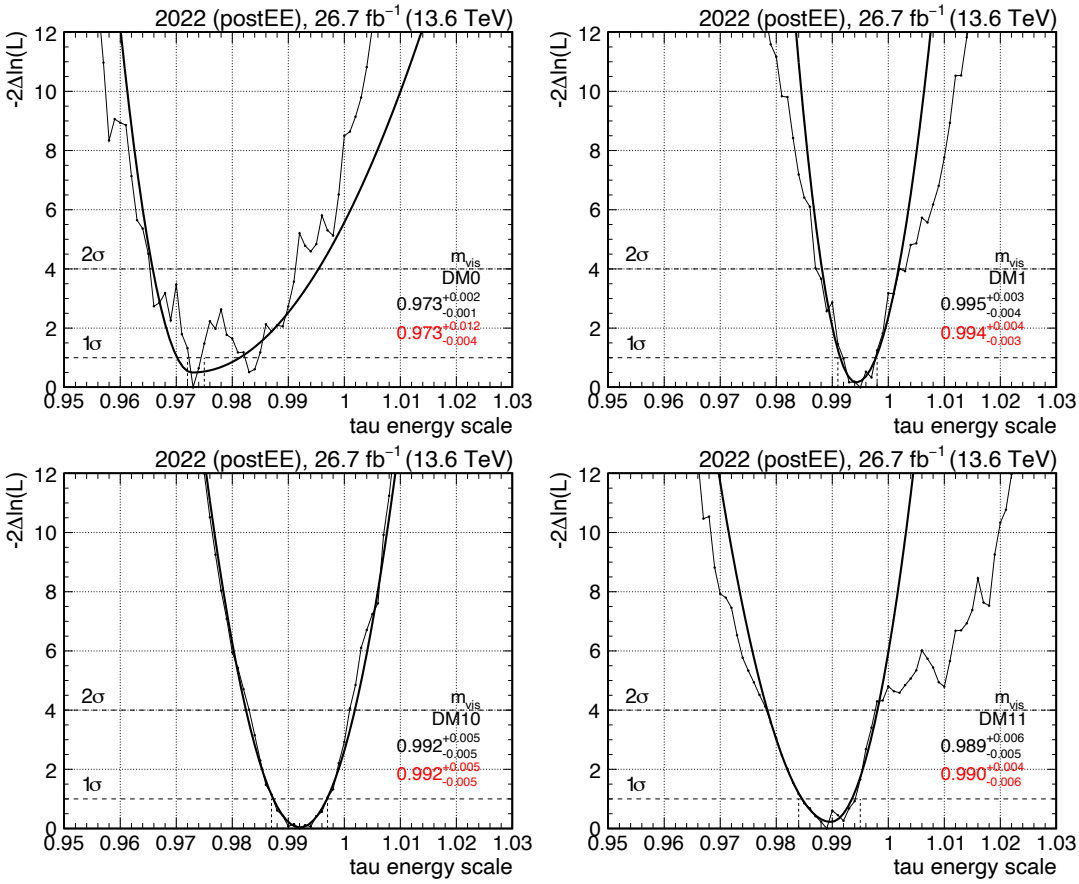


Figure A.7: Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11.

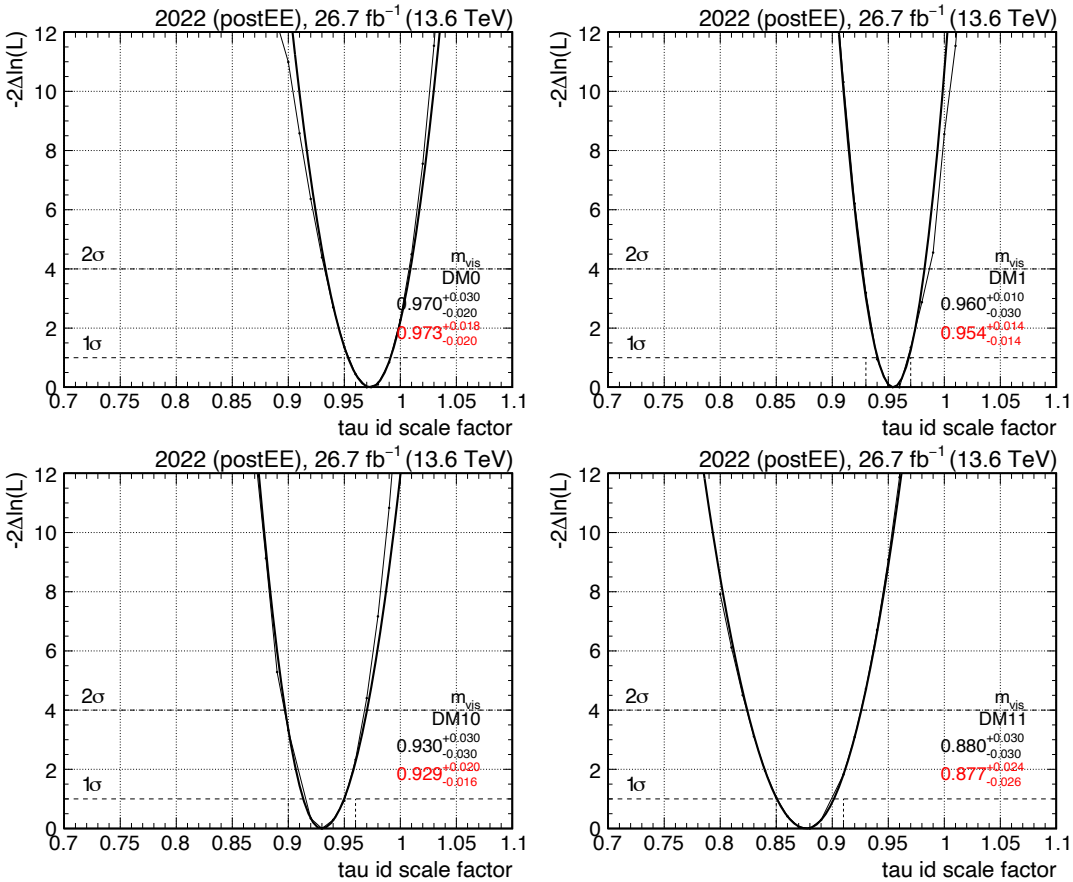


Figure A.8: Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11.

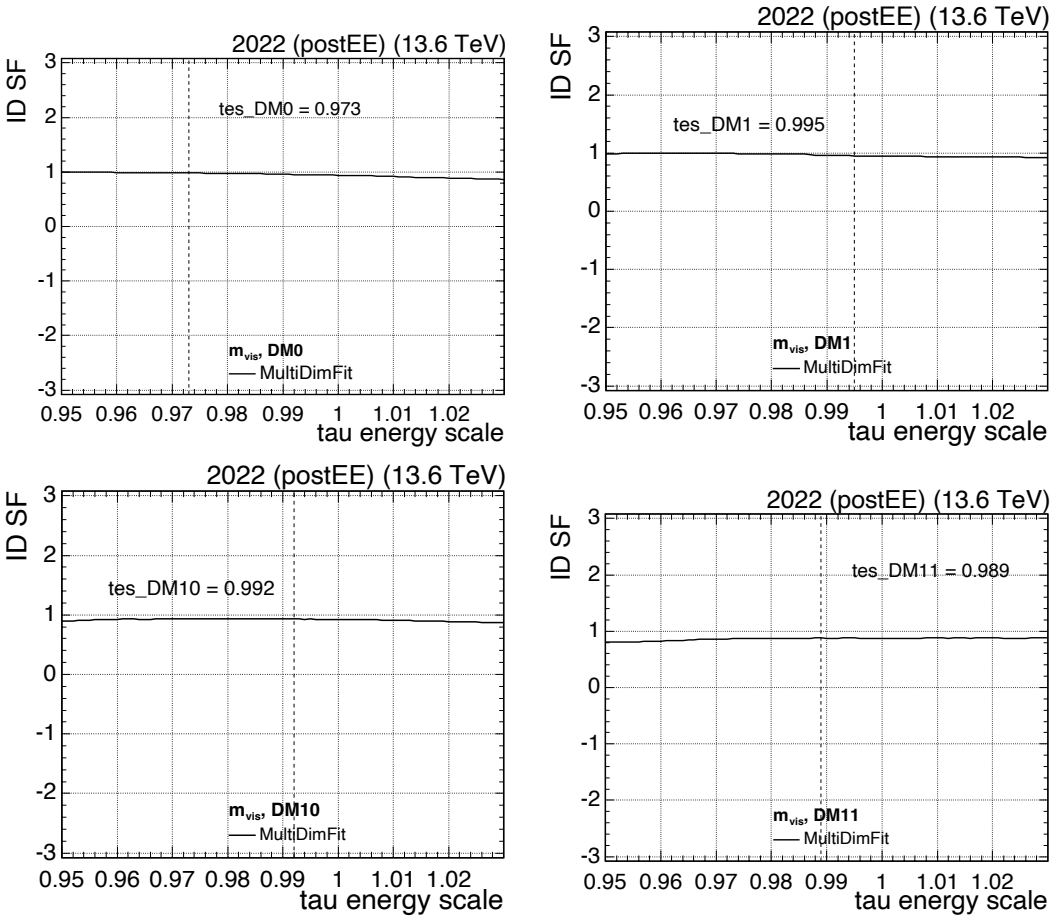


Figure A.9: The ID SF as a function of the TES during the NLL scan of the TES showing the dependency between the SFs.

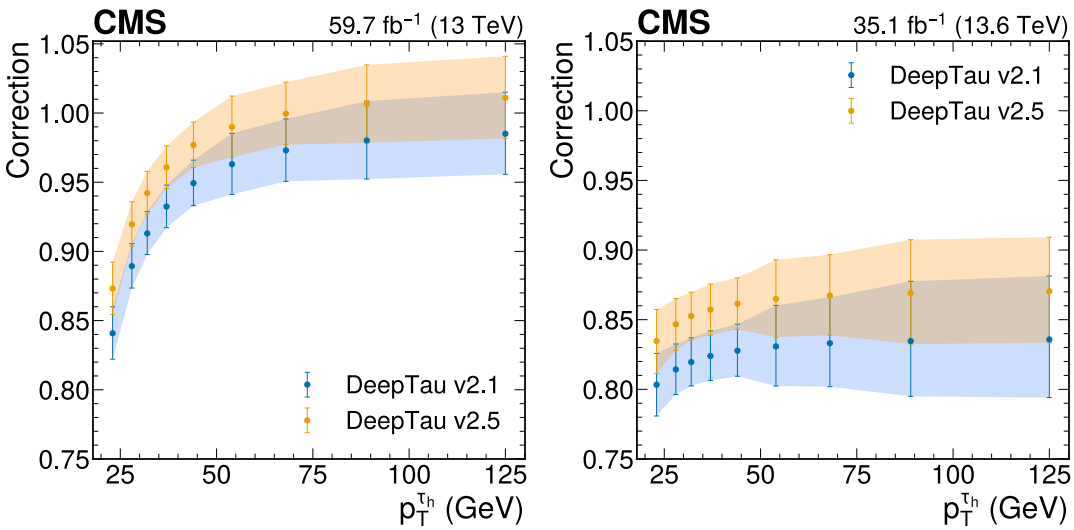


Figure A.10: The data-to-simulation scale factors of the τ_h identification efficiency as functions of p_T in the 2018 (left) and 2022 (right) data-taking periods, including all τ_h decay modes, and requiring the D_{jet} Medium working point and $m_T(\mu, p_T^{\text{miss}}) < 65$GeV. The vertical bars correspond to the combined statistical and systematic uncertainties in the individual scale factors.

Appendix B

CP measurement in $H \to \tau_h \tau_h$ decays

Content

I. Simulated Samples 205

I.1. Simulated Background Samples 205

I.2. Simulated Signal Samples 205

II. Performance of the Network 207

This appendix provides complementary plots and detailed explanations supporting the results presented in Chapter 6.

I. Simulated Samples

I.1. Simulated Background Samples

Monte Carlo (MC) simulations used for background modelling in this search are shown in Tables B.1. The simulated samples produced to model the signal processes are given in Tables B.2 and their production is further discussed in Section II.2.

I.2. Simulated Signal Samples

Dataset Description	XS [pb] × BR
Z → LL	6282.6 (NNLO)
Z + 1 jet	5364.0
Z + 2 jets	1019.0
Z + 3 jets	375.3
Low mass Z → LL	20950.0
Z + 1 jet, $p_T^{\ell\ell}$ 40–100 GeV	475.3
Z + 1 jet, $p_T^{\ell\ell}$ 100–200 GeV	45.42
Z + 1 jet, $p_T^{\ell\ell}$ 200–400 GeV	3.382
Z + 1 jet, $p_T^{\ell\ell}$ 400–600 GeV	0.1162
Z + 1 jet, $p_T^{\ell\ell} > 600$ GeV	0.01392
Z + 2 jets, $p_T^{\ell\ell}$ 40–100 GeV	179.3
Z + 2 jets, $p_T^{\ell\ell}$ 100–200 GeV	51.68
Z + 2 jets, $p_T^{\ell\ell}$ 200–400 GeV	7.159
Z + 2 jets, $p_T^{\ell\ell}$ 400–600 GeV	0.4157
Z + 2 jets, $p_T^{\ell\ell} > 600$ GeV	0.07019
$t\bar{t}$ (2L2ν)	98.04
$t\bar{t}$ (Lν2q)	405.75
$t\bar{t}$ (4q)	419.81
W+jets	63425.1 (NNLO+NLO_EW)
W+1 jet	9128.0
W+2 jets	2922.0
W+3 jets	861.3
W+4 jets	415.4
WW	122.27
WZ	41.15
ZZ	19.43
single top (t-channel, top)	123.8
single top (t-channel, anti-top)	75.47
single top (TW→2L2ν)	3.8
single top ($\overline{\text{T}}$W→2L2ν)	3.8
single top (TW→Lν2q)	15.8
single top ($\overline{\text{T}}$W→Lν2q)	15.9
WWW	0.22
WWZ	0.19
WZZ	0.062
ZZZ	0.016

Table B.1: List of simulated samples for Run 3 runs used for background modeling in the analysis.

Signal Process	XS [pb] $\times$ BR
ggH $h \to \tau\tau$ (CP-odd, uncorrelated)	3.2759
ggH $h \to \tau\tau$ (CP-mix, uncorrelated)	3.2759
ggH $h \to \tau\tau$ (CP-even, uncorrelated)	3.2759
VBF $h \to \tau\tau$ (uncorrelated)	0.2558
ZH $h \to \tau\tau$ (uncorrelated)	0.0592
W^+H $h \to \tau\tau$ (uncorrelated)	0.05575
W^-H $h \to \tau\tau$ (uncorrelated)	0.03561

Table B.2: List of signal samples for Run 3, with filter efficiencies and cross sections.

II. Performance of the Network

Tau and Pion Target Regression:

Tau-only Regression:

Additional Inputs and Constraints:

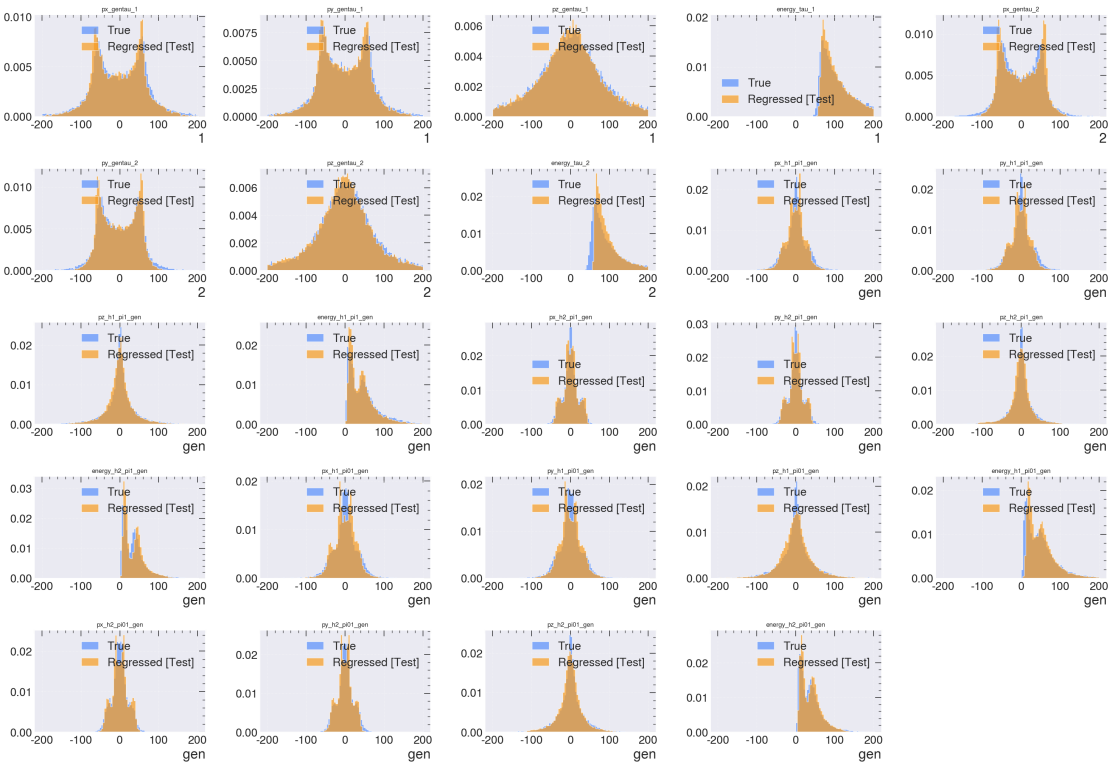


Figure B.1: Regressed distribution vs generator level distribution of the τ_h and pions.

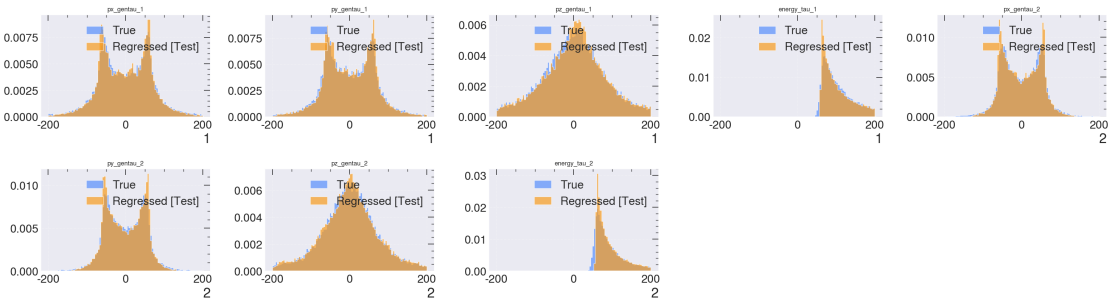


Figure B.2: Regressed distribution vs generator level distribution of the τ_h.

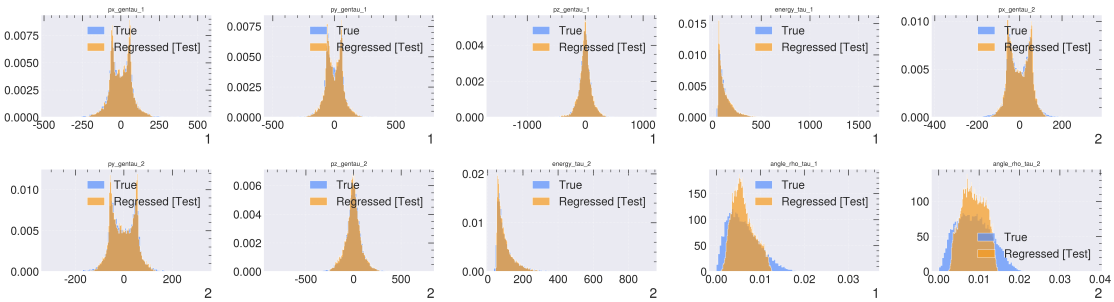


Figure B.3: Regressed distribution vs generator level distribution of the τ_h and the angle between τ_h and ρ resonance.

Bibliography

[1] Sheldon L. Glashow. Partial symmetries of weak interactions. *Nuclear Physics*, 22:579–588, 1961.

[2] Steven Weinberg. A model of leptons. *Physical Review Letters*, 19:1264–1266, 1967.

[3] Abdus Salam. Weak and electromagnetic interactions. In *Conf. Proc. C*, volume 680519, pages 367–377, 1968.

[4] S. L. Glashow. Partial symmetries of weak interactions. *Nucl. Phys.*, 22:579–588, 1961.

[5] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[6] A. Salam. Weak and electromagnetic interactions. In N. Svartholm, editor, *Elementary Particle Theory: Relativistic Groups and Analyticity*, pages 367–377. Almqvist and Wiksell, 1968.

[7] D. J. Gross, F. Wilczek, and H. D. Politzer. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973.

[8] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[9] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[10] Abdelhak Djouadi. The anatomy of electroweak symmetry breaking. *Physics Reports*, 457(1–4):1–216, February 2008.

[11] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.

[12] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.

[13] John R. Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. A Phenomenological Profile of the Higgs Boson. *Nucl. Phys. B*, 106:292, 1976.

[14] Michael Spira. Higgs boson production at the LHC. In *Ringberg Workshop: The Higgs Puzzle - What can We Learn from LEP2, LHC, NLC, and FMC?*, pages 122–130, 3 1997.

[15] Particle Data Group. Particle data group. http://pdg.lbl.gov/. Accessed: 2025-04-14.

[16] The CMS Collaboration. A portrait of the higgs boson by the cms experiment ten years after the discovery. *Nature*, 607(7917):60–68, July 2022.

[17] The CMS Collaboration. Analysis of the cp structure of the yukawa coupling between the higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} =$ 13 tev. *Journal of High Energy Physics*, 2022, 06 2022.

[18] Andrea Cardini. Methodologies to measure the cp structure of the higgs yukawa coupling to tau leptons. *Universe*, 8(5), 2022.

[19] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008.

[20] CERN. CERN Home Page. https://home.cern, 2024. Accessed: April 3, 2025.

[21] Tai Sakuma. Cutaway diagrams of CMS detector. 2019.

[22] Rudy Ceccarelli. Results on 3d pixel sensors for the cms upgrade at the hl-lhc. page 046, 03 2023.

[23] Massimiliano Antonello. Test beam results of irradiated modules for the CMS Phase-2 upgrade equipped with HPK planar pixel sensors and RD53B-CMS readout chips. *Nucl. Instrum. Meth. A*, 1064:169379, 2024.

[24] Institute of Experimental Particle Physics, KIT. Proton Irradiation at ETP, 2024. Accessed: 2025-06-13.

[25] R. Diener et al. The DESY II Test Beam Facility. *Nucl. Instrum. Meth. A*, 922:265–286, 2019.

[26] RD53B users guide. Technical report, CERN, Geneva, 2020.

[27] Ryszard Dąbrowski, Izabela Drozdyk, and Paweł Kukołowicz. High accuracy dosimetry with small pieces of gafchromic films. *Reports of Practical Oncology & Radiotherapy*, 23(2):114–120, 2018.

[28] Arden L. Buck. New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology and Climatology*, 20(12):1527–1532, 1981.

[29] V. Fadeyev, A. Affolder, A. Dowling, J. Fernandez-Tejero, J. Gunnell, N. Kang, F. Martinez-Mckinney, Q. Paddock, M. Ullan, Y. Unno, and J. Yarwick. Long-term humidity exposure of atlas18 itk strip sensors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1064:169430, 2024.

[30] Mauro E. Dinardo. The data acquisition system to test and characterize the pixel detector modules of the cms inner tracker for the high luminosity upgrade of lhc. In *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–7, 2021.

[31] R. Taylor. Ksu-fmc adapter manual for hardware version 1.04, 2021. Kansas State University.

[32] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical report, CERN, Geneva, 2009.

[33] Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. Technical report, CERN, Geneva, 2016.

[34] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. In *PHYSTAT 2011*, pages 115–120, 2011.

[35] Identification of tau leptons using a convolutional neural network with domain adaptation in the CMS experiment. 2025.

[36] Rene Brun and Fons Rademakers. ROOT — An Object Oriented Data Analysis Framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997. Proceedings AIHENP'96 Workshop, Lausanne, September 1996.

[37] Jim Pivarski, Isaac Osborne, Ionut Ifrim, Holger Schreiner, Alexander Hollands, Anish Biswas, Pritam Das, Shilpi Roy Choudhury, Nicholas Smith, and Manan Goyal. Awkward array, 2018. https://awkward-array.org.

[38] Charles R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020.

[39] Neal Richardson, Ian Cook, Neal Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragos Moldovan-Grünfeld, Jeroen Ooms, and Jonathan Wujciak-Jens. arrow: Integration to 'apache' 'arrow'. https://github.com/apache/arrow/, 2025. R package version 19.0.0.

[40] M. Rieger. End-to-end analysis automation over distributed resources with luigi analysis workflows. *EPJ Web of Conferences*, 295:05012, 2024.

[41] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques. *J. Phys. Conf. Ser.*, 513:022035, 2014.

[42] Aram Hayrapetyan et al. The CMS Statistical Analysis and Combination Tool: Combine. *Comput. Softw. Big Sci.*, 8(1):19, 2024.

[43] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[44] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[45] Brian Robert Martin and Graham Shaw. *Particle physics*. 2008.

[46] Steven Weinberg. *The Quantum Theory of Fields, Volume I: Foundations*. Cambridge University Press, 1995.

[47] E. Noether. Invariant variation problems. *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse*, pages 235–257, 1918. English translation: Transp. Theory Statist. Phys. 1 (1971) 186.

[48] A. Einstein. Zur elektrodynamik bewegter körper. *Annalen der Physik*, 322(10):891–921, 1905.

[49] E. Wigner. On unitary representations of the inhomogeneous lorentz group. *Annals of Mathematics*, 40:149–204, 1939.

[50] C. S. Wu et al. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, 1957.

[51] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.*, 13:138–140, 1964.

[52] G. Lüders. On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories. *Kong. Dan. Vid. Sel. Mat. Fys. Med.*, 28(5), 1954.

[53] W. Pauli. Exclusion principle, lorentz group and reflection of space-time and charge. *Nobel Lecture*, 1955.

[54] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.

[55] The Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81(8):1562–1567, August 1998.

[56] I. Neutelings. Cms wiki pages, how to draw diagrams in latex with tikz, 2023.

[57] H. D. Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.

[58] W. Pauli. Zur quantenmechanik des magnetischen elektrons. *Zeitschrift für Physik*, 43:601–623, 1927.

[59] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Nuovo Cim.*, 4 (Suppl 2):848–866, 1956.

[60] Kazuhiko Nishijima. Charge Independence Theory of V Particles. *Prog. Theor. Phys.*, 13(3):285–304, 1955.

[61] F. J. Hasert et al. Observation of Neutrino Like Interactions without Muon or Electron in the Gargamelle Neutrino Experiment. *Nucl. Phys. B*, 73:1–22, 1974.

[62] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983.

[63] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[64] Georges Aad et al. Combined Measurement of the Higgs Boson Mass from the $H\to\gamma\gamma$ and $H\to ZZ^*\to 4\ell$ Decay Channels with the ATLAS Detector Using s=7, 8, and 13 TeV pp Collision Data. *Phys. Rev. Lett.*, 131(25):251802, 2023.

[65] Aram Hayrapetyan et al. Measurement of the Higgs boson mass and width using the four-lepton final state in proton-proton collisions at s=13 TeV. *Phys. Rev. D*, 111(9):092014, 2025.

[66] Measurement of the Higgs boson mass and width using the four leptons final state. 2023.

[67] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, Emanuele R. Nocera, Juan Rojo, Luca Rottoli, Emma Slade, and Maria Ubiali. Parton distributions from high-precision collider data: Nnpdf collaboration. *The European Physical Journal C*, 77(10), October 2017.

[68] J. D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.*, 179:1547–1553, 1969.

[69] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the lhc era: Mmht 2014 pdfs. *The European Physical Journal C*, 75(5), May 2015.

[70] Nicolas Berger, Claudia Bertella, Matteo Bonanomi, Nihal Brahimi, Thomas P. Calvet, Milene Calvetti, Valerio Dao, Marco Delmastro, Michael Duehrssen-Debling, Paolo Francavilla, Yacine Haddad, Sarah Heim, Jelena Jovicevic, Oleh Kivernyk, Maria Moreno Llacer, Jonathon M. Langford, Changqiao Li, Giovanni Marchiori, Josh A. McFayden, Johannes K. L. Michel, Predrag Milenovic, Carlo E. Pandini, Edward Scott, Frank J. Tackmann, Kerstin Tackmann, Lorenzo Viliani, Meng Xiao, and Hongtao Yang. Simplified template cross sections – stage 1.1 and 1.2, 2025.

[71] Adam Falkowski and Adam Falkowski. Higgs Basis: Proposal for an EFT basis choice for LHC HXSWG. 2015.

[72] A. D. Sakharov. Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. [JETP Lett. 5 (1967) 24].

[73] M. B. Gavela et al. Standard Model CP violation and baryon asymmetry. *Mod. Phys. Lett. A*, 9:795–810, 1994.

[74] C. Hagedorn, R. N. Mohapatra, E. Molinaro, C. C. Nishi, and S. T. Petcov. CP Violation in the Lepton Sector and Implications for Leptogenesis. *Int. J. Mod. Phys. A*, 33(05n06):1842006, 2018.

[75] Mark Trodden. Electroweak baryogenesis. *Reviews of Modern Physics*, 71(5):1463–1500, October 1999.

[76] Frans R. Klinkhamer and N. S. Manton. A Saddle Point Solution in the Weinberg-Salam Theory. *Phys. Rev. D*, 30:2212, 1984.

[77] Masashi Aiko, Motoi Endo, Shinya Kanemura, and Yushi Mura. Electroweak baryogenesis in 2hdm without edm cancellation, 2025.

[78] Jordy de Vries, Marieke Postma, Jorinde van de Vis, and Graham White. Electroweak baryogenesis and the standard model effective field theory. *Journal of High Energy Physics*, 2018(1), January 2018.

[79] Henning Bahl, Elina Fuchs, Sven Heinemeyer, Judith Katzy, Marco Menen, Krisztian Peters, Matthias Saimpert, and Georg Weiglein. Constraining the cp structure of higgs-fermion couplings with a global lhc fit, the electron edm and baryogenesis. *The European Physical Journal C*, 82(7), July 2022.

[80] Ian Anderson et al. Constraining Anomalous HVV Interactions at Proton and Lepton Colliders. *Phys. Rev. D*, 89(3):035007, 2014.

[81] The CMS Collaboration. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev. D*, 92(1):012004, 2015.

[82] The CMS Collaboration. Constraints on anomalous higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state. *Physical Review D*, 104(5), September 2021.

[83] The 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. *Phys. Rev. D*, 108:032013, Aug 2023.

[84] Aram Hayrapetyan et al. Constraints on anomalous Higgs boson couplings from its production and decay using the WW channel in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 84(8):779, 2024.

[85] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, and Nhan V. Tran. Spin Determination of Single-Produced Resonances at Hadron Colliders. *Phys. Rev. D*, 81:075022, 2010.

[86] Fawzi Boudjema, Rohini M. Godbole, Diego Guadagnoli, and Kirtimaan A. Mohan. Lab-frame observables for probing the top-Higgs interaction. *Phys. Rev. D*, 92(1):015019, 2015.

[87] The CMS Collaboration. Search for cp violation in $t\bar{t}h$ and th production in multilepton channels in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, 2023(7), July 2023.

[88] Albert M Sirunyan et al. 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. *Phys. Rev. Lett.*, 125(6):061801, 2020.

[89] Georges Aad et al. Measurement of the CP properties of Higgs boson interactions with τ-leptons with the ATLAS detector. *Eur. Phys. J. C*, 83(7):563, 2023.

[90] K Desch, A Imhof, Z Was, and M Worek. Probing the cp nature of the higgs boson at linear colliders with τ spin correlations; the case of mixed scalar–pseudoscalar couplings. *Physics Letters B*, 579(1–2):157–164, January 2004.

[91] M. Krämer, J. Kühn, M. L. Stong, and P. M. Zerwas. Prospects of measuring the parity of higgs particles. *Zeitschrift für Physik C Particles and Fields*, 64(1):21–30, March 1994.

[92] V. Barger, K. Cheung, A. Djouadi, B. A. Kniehl, and P. M. Zerwas. Higgs bosons: Intermediate mass range at e^+e^- colliders. *Physical Review D*, 49(1):79–90, January 1994.

[93] Stefan Berge, Werner Bernreuther, and Sebastian Kirchner. Determination of the higgs cp-mixing angle in the tau decay channels at the lhc including the drell–yan background. *The European Physical Journal C*, 74(11), November 2014.

[94] B. K. Bullock, Kaoru Hagiwara, and Alan D. Martin. Tau polarization and its correlations as a probe of new physics. *Nucl. Phys. B*, 395:499–533, 1993.

[95] Yung-Su Tsai. Decay Correlations of Heavy Leptons in e+ e- —> Lepton+ Lepton-. *Phys. Rev. D*, 4:2821, 1971. [Erratum: Phys.Rev.D 13, 771 (1976)].

[96] Stanislaw Jadach, J H Kühn, and Zbigniew Was. TAUOLA: a library of Monte Carlo programs to simulate decays of polarized τ leptons. *Comput. Phys. Commun.*, 64:275–299, 1991.

[97] Vladimir Cherepanov and Christian Veelken. The polarimeter vector for $\tau \to 3\pi\nu_\tau$ decays, 2024.

[98] R. Garoby et al. Commissioning and First Operation of Linac4, the New LHC Injector. *Phys. Rev. Accel. Beams*, 2020.

[99] K. H. Reich. The CERN Proton Synchrotron Booster. *IEEE Trans. Nucl. Sci.*, 16:959–961, 1969.

[100] Simone Gilardoni and Django Manglunki, editors. *Fifty years of the CERN Proton Synchrotron: Volume 1*. CERN Yellow Reports: Monographs. 2011.

[101] The Super Proton Synchrotron. 2012.

[102] ALICE Collaboration. The alice experiment at the cern large hadron collider. *J. Instrum.*, 3(S08002), 2008.

[103] ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *J. Instrum.*, 3(S08003), 2008.

[104] CMS Collaboration. The cms experiment at the cern lhc. *J. Instrum.*, 3(S08004), 2008.

[105] LHCb Collaboration. The lhcb detector at the cern large hadron collider. *J. Instrum.*, 3(S08005), 2008.

[106] Ewa Lopienska. The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022. 2022. General Photo.

[107] Helmut Wiedemann. *Particle Accelerator Physics*. Springer, 4th edition, 2015.

[108] CMS Collaboration. Measurement of the inelastic proton-proton cross section at 13 tev. *Journal of High Energy Physics*, 2018(7):161, 2018.

[109] CMS Collaboration. Public luminosity results 2023. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2023. Last visited: August 1st, 2023.

[110] CMS Collaboration. CMS Luminosity Public Results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2025. Accessed: 2025-06-25.

[111] Markus Zerlauth and Oliver Bruning. Status and prospects of the hl-lhc project. page 615, 01 2024.

[112] CMS Collaboration. Observation of the higgs boson decay to a pair of τ leptons with the cms detector. *Physics Letters B*, 779:283–316, April 2018.

[113] CMS Collaboration. Evidence for higgs boson decay to a pair of muons. *Journal of High Energy Physics*, 2021(1):148, January 2021.

[114] G Apollinari, I Béjar Alonso, O Brüning, M Lamont, and L Rossi. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2015.

[115] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[116] CMS Collaboration. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021, mar 2010.

[117] *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[118] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[119] *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. CERN, Geneva, 2000.

[120] A Dominguez, D Abbaneo, K Arndt, N Bacchetta, A Ball, E Bartz, W Bertl, G M Bilei, G Bolla, H W K Cheung, M Chertok, S Costa, N Demaria, Daniel Dominguez Vazquez, K Ecklund, W Erdmann, K Gill, G Hall, K Harder, F Hartmann, R Horisberger, W Johns, H C Kaestli, K Klein, D Kotlinski, S Kwan, M Pesaresi, H Postema, T Rohe, C Schäfer, A Starodumov, S Streuli, A Tricomi, P Tropea, J Troska, F Vasey, and W Zeuner. CMS Technical Design Report for thePixel Detector Upgrade. Technical report, 2012.

[121] The CMS Collaboration. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021.

[122] P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.

[123] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[124] Andrea Benaglia. The CMS ECAL performance with examples. Technical report, CERN, Geneva, 2014.

[125] Albert M Sirunyan et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *JINST*, 15(05):P05002, 2020.

[126] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[127] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV. *JINST*, 13(06):P06015, 2018.

[128] Gerhard Lutz. *Semiconductor Radiation Detectors*. Springer, 1999.

[129] Salim Reza. *Semiconductor Radiation Detectors: Technology and Applications*. Taylor & Francis Group, Boca Raton, FL, 2017.

[130] Karl W. Böer and Udo W. Pohl. *Semiconductor Physics*. Springer, Cham, Switzerland, 1st edition, 2018.

[131] Thomas L. Floyd. *Introduction to Semiconductors*. Pearson, 2017. Accessed: 2025-04-11.

[132] Author(s) of the book. *Evolution of Silicon Sensor Technology in Particle Physics: Basics and Applications*. Publisher Name, Publisher's Location, Year of Publication.

[133] H. Bichsel. A method to improve tracking and particle identification in tpcs and silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 562:154–197, 2006.

[134] S. Ramo. Currents to conductors induced by a moving point charge. *Journal of Applied Physics*, 9(10):635–636, 1938.

[135] Viktoria Hinger. *Silicon Sensor Process Quality Control for the CMS Phase-2 Upgrade*. PhD thesis, Vienna University of Technology, 2021. Presented 24 Feb 2021.

[136] Michael Moll. Displacement damage in silicon detectors for high energy physics. *IEEE Transactions on Nuclear Science*, 65(8):1561–1582, 2018.

[137] 2020 Update of the European Strategy for Particle Physics (Brochure). Technical report, Geneva, 2020.

[138] Steve Ritz et al. Building for Discovery: Strategic Plan for U.S. Particle Physics in the Global Context. 5 2014.

[139] Adam Smoliński. *Strategy Report on research infrastructure*. 03 2016.

[140] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:221–584, 2019.

[141] D Contardo, M Klute, J Mans, L Silvestris, and J Butler. Technical Proposal for the Phase-II Upgrade of the CMS Detector. Technical report, Geneva, 2015. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[142] Xabier Cid Vidal et al. Report from Working Group 3: Beyond the Standard Model physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:585–865, 2019.

[143] Irene Dutta. Precision Timing with the CMS MIP detector. In *13th Conference on the Intersections of Particle and Nuclear Physics*, 9 2018.

[144] Armen Tumasyan et al. The Phase-2 Upgrade of the CMS Tracker. 6 2017.

[145] The CMS Collaboration. The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva, 2017.

[146] Irene Zoi, M Boscardin, G. F Dalla Betta, Mauro Dinardo, G Giacomini, Dario Menasce, R Mendicino, Marco Meschini, Alberto Messineo, Luigi Moroni, S Ronchin, D.M.S Sultan, Lorenzo Uplegger, Caterina Vernieri, Lorenzo Viliani, and Davide Zuolo. Beam Test Results of Thin n-in-p 3D and Planar Pixel Sensors for the High Luminosity LHC Tracker Upgrade at CMS. Technical report, CERN, Geneva, 2018.

[147] ATLAS Collaboration. ATLAS ITk Pixel Detector for Phase-2 Upgrade. In *ACES 2020*, 2020.

[148] ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical Report CERN-LHCC-2017-021, ATLAS-TDR-030, CERN, 2017.

[149] CMS Collaboration. CMS Pixel Detector Overview. In *ACES 2020*, 2020.

[150] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Upgrade. Technical Report CERN-LHCC-2012-016, CMS-TDR-11, CERN, 2012.

[151] E. Orfanelli and CMS Collaboration. The CMS Inner Tracker for the High Luminosity LHC. In *International Workshop on Future Colliders (Hiroshima 2019)*, 2019.

[152] E. Bouquerel, E. Traykov, C. Maazouzi, M. Rousseau, M. Pellicioli, T. Adam, P. Graehling, C. Mathieu, G. Heitz, M. Krauth, D. Oster, T. Foehrenbacher,

C. Ruescas, J. Schuler, U. Goerlach, and C. Haas. Design and commissioning of the first two cyrcé extension beamlines. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1024:166034, February 2022.

[153] P. Allport, F. Bögelspacher, K. Bruce, R. Canavan, A. Dierlamm, L. Gonella, P. Knights, I. Mateu, M. Moll, K. Nikolopoulos, B. Phoenix, T. Price, L. Ram, F. Ravotti, C. Simpson-Allsop, and C. Wood. Experimental determination of proton hardness factors at several irradiation facilities. *Journal of Instrumentation*, 14(12):P12004, dec 2019.

[154] Wenlin Li, Qifan Dong, Hantao Jing, Li Ou, Zhixin Tan, Sixuan Zhuang, and Qingbiao Wu. Measurements of proton beam flux and energy of apep using foil activation technique. *Nuclear Engineering and Technology*, 56(1):328–334, 2024.

[155] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, October 2014.

[156] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987.

[157] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998.

[158] R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical report, CERN, Geneva, 2007.

[159] Walaa Elmetenawee. CMS tracking performance in Run 2 and early Run 3. *PoS*, VERTEX2023:074, 2024.

[160] Erica Brondolin. CMS detector tracking performance in Run-II. *PoS*, ICHEP2016:1167, 2017.

[161] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical report, 2010.

[162] The CMS Collaboration. Particle-flow reconstruction and global event description with the cms detector. *Journal of Instrumentation*, 12(10):P10003, oct 2017.

[163] The CMS collaboration. Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$ tev. *Journal of Instrumentation*, 7(10):P10002, oct 2012.

[164] The CMS Collaboration. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 tev. *Journal of Instrumentation*, 13(06):P06015, jun 2018.

[165] The CMS Collaboration. Muon identification using multivariate techniques in the cms experiment in proton-proton collisions at sqrt(s)=13tev. *Journal of Instrumentation*, 19(02):P02031, February 2024.

[166] *Journal of Instrumentation*, 10(06):P06005–P06005, June 2015.

[167] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993.

[168] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[169] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[170] CMS Collaboration. Jet energy scale and resolution performances with 8 tev data. CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014.

[171] CMS Collaboration. Jet energy scale and resolution performance in proton-proton collisions at 13 tev. CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2016.

[172] V. Khachatryan et al. Observation of the higgs boson decay to two photons. *Phys. Lett. B*, 789:1–31, 2016.

[173] T. Sjostrand et al. Pythia 6.4: Physics and manual. *JHEP*, 05:026, 2006.

[174] M. Bahr et al. Herwig++ 2.5: A monte carlo event generator for high-energy hadron collisions. *Eur. Phys. J. C*, 58:639–657, 2008.

[175] CMS Collaboration. Cms performance for jets and missing transverse energy. CMS Detector Performance, Document CMS-DP-2016-020, 2016. CERN, Geneva.

[176] CMS Collaboration. Performance of the cms jet reconstruction with the 2015 data. CMS Detector Performance, Document CMS-DP-2015-020, 2015. CERN, Geneva.

[177] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical report, CERN, Geneva, 2016.

[178] The CMS collaboration. Identification of hadronic tau lepton decays using a deep neural network. *Journal of Instrumentation*, 17(07):P07023, jul 2022.

[179] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), October 2014.

[180] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[181] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2018.

[182] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation, 2015.

[183] Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons. 2019.

[184] Somnath Choudhury. Performance of the High-Level Trigger System at CMS in LHC Run-2. *IEEE Trans. Nucl. Sci.*, 68(8):2035–2042, 2021.

[185] Sanu Varghese. CMS Trigger Performance in Run 3. *Springer Proc. Phys.*, 304:961–964, 2024.

[186] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[187] Livia Soffi. CMS electron and photon performance at Run 2 and prospects for Run 3. *PoS*, ICHEP2020:780, 2021.

[188] Archana Sharma. Performance of the CMS Muon System with Early Run 3 Data. *PoS*, EPS-HEP2023:519, 2024.

[189] Electron and photon reconstruction and identification performance at CMS in 2022 and 2023. 2024.

[190] CMS Collaboration. Pileup mitigation at cms in 13 tev data. *JINST*, 15:P09018, 2020.

[191] Anna Benecke. Jet performance and pileup mitigation in Run 3 in CMS. Technical report, CERN, Geneva, 2024.

[192] The CMS collaboration. Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$ tev. *Journal of Instrumentation*, 7(10):P10002–P10002, October 2012.

[193] The CMS Collaboration. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 tev. *Journal of Instrumentation*, 13(06):P06015–P06015, June 2018.

[194] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. *CERN Yellow Rep. Monogr.*, 2:1–869, 2017.

[195] Marco Peruzzi, Giovanni Petrucciani, Andrea Rizzi, and for the CMS Collaboration. The nanoaod event data format in cms. *Journal of Physics: Conference Series*, 1525(1):012038, apr 2020.

[196] ColumnFlow Developers. Columnflow: A columnar data processing framework for hep, 2024. Accessed: 2025-07-05.

[197] Mathis Frahm. Orchestrated columnar-based analysis with columnflow. Technical report, CERN, Geneva, 2025.

[198] Nicholas Smith, Lindsey Gray, Matteo Cremonesi, Bo Jayatilaka, Oliver Gutsche, Allison Hall, Kevin Pedro, Maria Acosta, Andrew Melo, Stefano Belforte, and Jim Pivarski. Coffea columnar object framework for effective analysis. *EPJ Web of Conferences*, 245:06012, 2020.

[199] Nicholas Smith, Lindsey Gray, Matteo Cremonesi, Bo Jayatilaka, Oliver Gutsche, Allison Hall, Kevin Pedro, Maria Acosta, Alessandra Melo, Stefano Belforte, and Jim Pivarski. Coffea – columnar object framework for effective analysis, 08 2020.

[200] E. Bernhardsson and E. Freider. Luigi - a python package for building complex pipelines of batch jobs. https://github.com/spotify/luigi.

[201] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[202] Lamarr Institute. Deep neural networks – the backbone of modern ai, 2023. Accessed: 2025-07-05.

[203] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. Received 01 May 1986, Accepted 31 July 1986, Issue Date 09 October 1986.

[204] Karen Loaiza. *Deep learning for decision support in dermatology*. PhD thesis, 09 2020.

[205] CMS luminosity measurement for the 2022 data taking period. https://cms-results.web.cern.ch/cms-results/public-results/publications/LUM-22-001/, 2023.

[206] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[207] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. 45 pages.

[208] Pier Francesco Monni, Paolo Nason, Emanuele Re, Marius Wiesemann, and Giulia Zanderighi. MiNNLO$_{PS}$: a new method to match NNLO QCD to parton showers. *JHEP*, 05:143, 2020. [Erratum: JHEP 02, 031 (2022)].

[209] T. Przedzinski, E. Richter-Was, and Z. Was. TauSpinner: a tool for simulating CP effects in $H \to \tau\tau$ decays at LHC. *Eur. Phys. J. C*, 74(11):3177, 2014.

[210] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger. Physics inspired feature engineering with Lorentz Boost Networks. *J. Phys. Conf. Ser.*, 1525(1):012107, 2020.

[211] Huilin Qu and Loukas Gouskos. ParticleNet: Jet Tagging via Particle Clouds. *Phys. Rev. D*, 101(5):056019, 2020.

[212] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

List of Figures

1 Définition schématique de l'angle d'acoplanarité ϕ^* entre les plans de désintégration des $\tau^\pm$ selon la notation commune. Les vecteurs $\vec{R}^\perp_\pm$ sont orthogonaux aux axes $\vec{P}_\pm$ dans les plans respectifs [18]. . . 5

2 Vue shecmatique du detecteru CMS [21]. 6

3 (Gauche) Vue extérieure de la boîte froide, montrant le débitmètre et les câbles d'alimentation. (Droite) Intérieur de la boîte avec capteurs de température et humidité, ainsi qu'une résistance chauffante simulant une charge thermique. 9

4 Température, humidité relative et point de rosée en mode "super" avec charge thermique simulée (7 W) et ventilation interne. La stabilité thermique et le contrôle de l'humidité sont assurés. 10

5 Température, humidité relative et point de rosée lors de l'opération stable d'un module CROC-V1 irradié en mode "super" avec ventilation active. 11

6 Banc de test du CROC-V1 : (gauche) chaîne de lecture complète avec FC7, KSU-FMC et alimentations ; (droite) module CROC-V1 sur carte porteuse, avec capteur 3D au centre et cavaliers LDO (violet) et trimming (bleu). 12

7 (Gauche) Carte d'occupation des pixels avant irradiation, indiquant une quasi-totalité des pixels actifs sans défaut majeur. (Droite) Distribution de la S-curve avant irradiation, illustrant une faible dispersion des seuils et un faible bruit. 13

8 Scan de cross-talk post-irradiation montrant la région "W" avec des pixels déconnectés (absence de réponse dans les pixels voisins). . . . 14

9 Courbe S-curve post-calibration des pixels fonctionnels, illustrant une réponse homogène et un bruit maîtrisé. 14

10 Synthèse des TES (gauche) et ID SF (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la log-vraisemblance négative (NLL). 17
11 Synthèse des TES (gauche) et ID SFs (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la NLL. . . . 18
12 Synthèse des TES (gauche) et ID SFs (droite) par région DM pour UL2018, obtenus via des ajustements paraboliques de la NLL. . . . 18
13 Schématisation de la méthode ABCD utilisée pour estimer le bruit de fond $j \to \tau_h$. 20
14 Distributions dans la région signal pour l'ère BPix 2023 avec le bruit de fond multijet modélisé par la méthode des fake factors. 22
15 Résolution de la masse di-τ sur DY simulé avec FastMTT (gauche) et SVFit (droite). 23
16 Comparaison des distributions de ϕ^*_{CP} obtenues avec : p_T regréssés combinée aux informations reconstruites (PVPV, à gauche), et méthode classique du plan de désintégration (DP, à droite) avec les informations reconstruites. 24
17 Scan du profil de vraisemblance de l'angle $\alpha_{H\tau\tau}$ extrait d'un fit simultané sur toutes les catégories avec le jeu de données Asimov PreEE 2022 [43] (26.671 fb^{-1}), incertitudes statistiques uniquement. 25

1.1 Summary of all SM elementary particles and their corresponding antiparticles. The three generations of quarks/anti-quarks (blue) and leptons/anti-leptons (green) are shown in the first six columns. The seventh column shows the force carriers of the SM (red), while the Higgs boson (yellow) is shown on the right as the only spin-0 particle of the SM. Figure adapted from [56]. 6
1.2 Feynman diagram of the Higgs production at tree level. In (a) gluon gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung or associated production with a W/Z boson and (d) $t\bar{t}$ associated production [15]. 10
1.3 SM Higgs boson production cross section as a function of the center of mass energy $\sqrt{s}$ of the pp collision [15]. 10
1.4 SM Higgs boson branching ratios as a function of the Higgs boson mass [15]. 12
1.5 Coupling modifier of the Higgs bosons to fermions (κ_f) and heavy gauge bosons (κ_V) as a function of the fermion or gauge boson mass, where v is the vacuum expectation value of the Brout-Englert-Higgs field. For gauge bosons, the square root of the coupling modifier is plotted, to keep a linear proportionality to the mass [16]. 14

1.6 (Left) Two-dimensional likelihood scan in the $(\kappa_t, \tilde{\kappa}_t)$ plane from the combination of $t\bar{t}H$, tH, and ggH sensitive analyses in $H \to \gamma\gamma$, $H \to 4\ell$, and multilepton final states. (Right) Observed and expected negative log-likelihood scan as a function of the CP-odd cross section fraction f_{CP}^{Htt}. Figures from [87]. 19

1.7 (Left) Two-dimensional likelihood scan in the $(\kappa_\tau, \tilde{\kappa}_\tau)$ plane. (Right) Weighted ϕ_{CP} distributions in most sensitive hadronic and leptonic τ decay channels, with background subtraction in all MVA bins. Each MVA score bin is weighted by $A \cdot S/(S+B)$ where S and B are the signal and background rates, and A is the average symmetry of the scalar and pseudoscalar distribution. The scalar hypothesis is shown in red, while the pseudoscalar hypothesis is shown in blue. The predictions are based on the best fit values of the rate parameters. The grey uncertainty band indicates the uncertainty in the background subtraction. A phase shift of 180° is applied to the leptonic channels due to the phase difference with the hadronic channels. Figures from [17]. 20

1.8 Spin correlation–dependent decay rates (see Equation 1.29) of the Higgs boson into $\tau^+\tau^-$ pairs as a function of the longitudinal spin component $\|s_{\|}^{\tau}\|$ and the angle ϕ between the transverse spin vectors. The scalar scenario is on the left and the pseudoscalar scenario on the right. 22

1.9 On the left, the normalized distribution of the acoplanarity angle ϕ_{CP} between the τ lepton decay planes in the Higgs rest frame at generator level, for both τ leptons decaying to a charged pion and a neutrino. Distributions are shown for a scalar (CP-even, solid red), pseudoscalar (CP-odd, dashed blue), maximal CP-mixing angle of 45° (CP-mix, dash-dot-dot green), and a Z boson (black dash-dot). On the right, an illustration of the decay planes of two τ leptons decaying to single charged pions. The ϕ_{CP} angle is defined as the angle between these decay planes in the Higgs rest frame. Figures from [17]. 23

1.10 Schematic representation of the impact parameter vector $\vec{IP}$ of a charged particle of momemtum $\vec{p_p}$ from a secondary τ decay vertex (SV) relative to the primary vertex (PV). 26

1.11 Schematic representation of the acoplanarity angle definition, following the notation introduced in this paper. The vectors $\vec{R}_{\perp}^{\pm}$ lie in the respective $\tau^{\pm}$ decay planes and are orthogonal to the axes defined by the vectors $\vec{P}^{\pm}$. The acoplanarity angle ϕ_{Δ} is defined as the counter-clockwise angle from the τ^{+} decay plane to the τ^{-} decay plane, measured around the $\hat{z}^{-}$ axis. [18]. 29

2.1 The CERN accelerator complex, showing the chain of accelerators leading to the LHC. The LHC is represented by the final dark blue ring, with the four major experiments indicated by yellow dots [106]. 34

2.2 Cumulative integrated luminosity delivered by the LHC in proton-proton collisions over different data-taking years [110]. 37

2.3 LHC operational phases and luminosity schedule, as planned for the HL-LHC upgrade [111]. 38

2.4 A schematic view of the standard coordinate system at the CMS detector, shown relative to the LHC and the other three main LHC experiments [56]. 40

2.5 Relation between the pseudorapidity η and the polar coordinate θ from the standard coordinate system of CMS [56]. 41

2.6 Schematic view of CMS detectors showing all the subdetectors [21]. 43

2.7 The value of the magnetic field strength $|B|$ (left) and the corresponding field lines (right) are shown for a longitudinal section of the CMS detector. This is based on the underground model, with a central magnetic flux density of 3.8 T. Each field line represents an increment of 6 Weber (Wb) in magnetic flux [116]. 44

2.8 Schematic overview of the CMS tracker. Each line represents a detector module. Double lines indicate double-sided modules [120]. 45

2.9 Layout of the pixel tracker compared to the original detector layout, in longitudinal view [121]. 46

2.10 Simulated material budget of the CMS pixel tracker as a function of pseudorapidity η, expressed in units of radiation lengths (left) and hadronic interaction lengths (right). The original detector design is shown as black points, while the Phase-1 upgraded detector is represented with stacked histograms detailing contributions from individual components [121]. 47

2.11 Schematic layout of the CMS electromagnetic calorimeter, presenting the arrangement of barrel, endcaps and the preshower in front (left). Geometric view of one quarter of the ECAL (right) [124]. . . 49

2.12 A schematic representation of one quarter of the CMS HCAL during the 2016 LHC operation, illustrates the locations of its four main components: the Hadron Barrel (HB), the Hadron Endcap (HE), the Hadron Outer (HO), and the Hadron Forward (HF) calorimeters. The layers marked in blue are grouped as depth = 1, while the layers in yellow, green, and magenta are assigned to depths 2, 3, and 4, respectively [125]. 51
2.13 A schematic longitudinal view of one quarter of the muon detection system layout. The Drift Tubes (DT, yellow), Cathode Strip Chambers (CSC, green), and Resistive Plate Chambers (RPCs, blue) are displayed, with dashed lines indicating the pseudorapidity values [127]. 52
2.14 The three types of solid-state materials and their energy band diagrams: conductors, semiconductors, and insulators [131]. 55
2.15 On the left, a pentavalent impurity atom of antimony (Sb) introduces an extra electron, creating an n-type semiconductor. On the right, a trivalent impurity atom of boron (B) creates a hole, forming a p-type semiconductor [131]. 56
2.16 A schematic representation of the p–n junction at the instant of junction formation (left) and at equilibrium (right) [131]. 57
2.17 The current–voltage relation (I-V) for a silicon diode in the reverse bias direction is depicted. The expanded view shows the proportionality between I and $\sqrt{\mathrm{V}}$ and the full scan shows the breakdown at higher voltages [132]. 58
2.18 Mean energy loss in copper for traversing muons [15]. 59
2.19 Landau distribution in solid line with arbitrary scales. The Full-Width-Half-Maximum (FWHM), the mean value and the Most Probable Value (MPV) have been added to the Figure [133] as illustration. 60
2.20 Vacancy and interstitial in the silicon lattice known as "Frenkel effect" [135]. 61
2.21 Radiation-induced defects in the silicon band gap [135]. 61
2.22 On the left, full depletion voltage and effective doping concentration as a function of the equivalent fluence ϕ_{eq}. On the right, leakage current increase caused by irradiation ΔI as a function of the equivalent fluence ϕ_{eq} for various silicon detectors [136]. 62
2.23 Higgs boson couplings to fermions (Yukawa couplings, λ_f) and to vector bosons (parametrized as $(g_V/2v)^{1/2}$, where v is the SM Higgs vacuum expectation value) as measured using Run 2 data [16] (left), and projected uncertainties for an integrated luminosity of 3000 fb^{-1}, assuming Standard Model couplings [141] (right). 64

2.24 Expected negative log-likelihood scan for the combination of the $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels, including predictions for LHC Run-3 and Phase-2 integrated luminosities. The extrapolations have been made assuming systematic uncertainties as derived for the main analysis, while the statistical uncertainties were adjusted to different luminosity scenarios. The expected sensitivity to distinguish between the scalar and pseudo-scalar hypotheses, defined at $\alpha_{H\tau\tau} = 0$ and $\pm 90°$, respectively, exceeds 5 standard deviations for the Phase-2 scenario [17]. 65
2.25 Integrated particle fluence in 1 MeV neutron equivalent per cm^2, for the Phase-2 tracker. The estimates shown correspond to a total integrated luminosity of 3000 fb^{-1} of pp collisions at $\sqrt{s} = 14$ TeV, and have been obtained with the CMS FLUKA geometry version [145]. 68
2.26 Sketch of the Phase-2 Inner Tracker layout, showing the barrel and endcap detector regions. The modules are arranged in layers and rings, and the pixel sensors are illustrated as colored elements [145]. 70
2.27 A perspective view of one quarter of the Inner Tracker, illustrating the TBPX ladders along with the TFPX and TEPX dees within the supporting structures. The pixel modules appear as orange elements in the TBPX and green elements in the TFPX and TBPX, while the dees are represented by red and orange surfaces [145]. 71
2.28 Illustration of the 1×2 (left) and 2×2 (right) pixel modules, with approximate dimensions of 1.8×4.4 cm^2 and 3.7×4.4 cm^2, respectively. Yellow elements indicate passive electrical components [145]. 72
2.29 Data flow architecture of the RD53B readout chip [26]. 74
2.30 Block diagram of a generic Analog Front-End (AFE) used in pixel detectors. The diagram shows the pixel sensor, which converts the energy deposited by a particle into charge. This charge is then processed by the AFE, which includes a preamplifier for signal integration, a shaping stage to prepare the signal for digitization, a programmable current source (Krummernacher current) for signal discharge, and a discriminator to generate a digital hit signal when the input exceeds the threshold [26]. 75
2.31 Illustration of charge measurement using the ToT method and the Krummernacher current scheme. The ToT method translates the amount of deposited charge into a time pulse width, providing an indirect but precise measure of the energy deposited by the particle. The Krummernacher current defines the discharge slope of the pulse, ensuring a consistent and linear correlation between the charge and time duration [26]. 76

3.1 Pictures of CYRCé in the vault [152]. 79
3.2 Layout of the CYRCé facility. The experimental area includes the PRECy platform for radiobiology studies and the CMS dedicated beamline for silicon sensor qualification. The new irradiation beamline, located inside the radiation-shielded vault, is designed for high-fluence irradiation of IT modules [152]. 80
3.3 Photograph of the new proton irradiation line installed in the CYRCé vault. The setup includes beamline elements such as the dipole, steerers, quadrupoles, and the cold box used for silicon module. . . 81
3.4 (Left) Full view of the cold box setup. The flow meter is visible with a tube supplying air inside the freezer. The power supply for the module and the sensors is also visible. (Right) Interior of the cold box, including a set of temperature and relative humidity (HR) sensors, along with a resistor. 84
3.5 Temperature and dew point measurements from three sensors placed at different locations inside the cold box during operation in normal freezer mode. 85
3.6 Temperature and dew point for two different resistor values. 86
3.7 Temperature, relative humidity and dew point for "super" mode. . . 87
3.8 Temperature, relative humidity and dew point for "super" mode with heating plate. 88
3.9 Temperature, relative humidity and dew point for "super" mode with heating plate and fans. 88
3.10 Temperature, relative humidity and dew point during operation of the irradiated CROC-V1 module in "super" mode with active cooling and a fan. The setup ensures stable conditions for safe operation. 89
3.11 Long-term stability of temperature, relative humidity, and dew point in "super" mode with an irradiated module and fan, several days after initial installation. 90
3.12 Photographs of the test setup for the CROC-V1 module. (Left) Full readout chain, including the FC7 board, KSU-FMC interface, and power connections. (Right) Close-up of the CROC-V1 module mounted on its carrier card, showing the 3D sensor at the center. The jumpers for the SLDO powering configuration are highlighted in purple, and the jumper used for trimming is highlighted in blue. 92
3.13 Pixel Alive map before irradiation. The z-axis corresponds to pixel occupancy. 93
3.14 ToT distribution and map before irradiation. 94
3.15 TDAC distribution before irradiation and map. 94

3.16 S-curve distribution before irradiation. The steep slope reflects a narrow threshold dispersion and low noise. 95
3.17 Threshold distribution (left) and threshold pixel map (right) before irradiation. The threshold is uniformly distributed across the pixel matrix. 95
3.18 Noise distribution (left) and noise pixel map (right) before irradiation. The uniform noise response confirms stable analog front-end performance. 95
3.19 IV curve of the irradiated CROC-V1 module measured after approximately two weeks of annealing. The curve indicates a full depletion voltage in the range of 25–45 V. 96
3.20 Noise scan (left) and pixel alive scan (right) following irradiation. The "W-shaped" noisy region and manually masked column are visible. 97
3.21 Occupancy distribution: 1D projection (left) and 2D map (right) after initial tuning. 98
3.22 TDAC tuning distribution: 1D histogram (left) and 2D map (right). 98
3.23 Cross-talk scan: pixel alive map (left) and masked pixels map (right). 99
3.24 Occupancy distribution after tuning: 1D (left) and 2D (right). . . . 100
3.25 TDAC distribution: 1D (left) and 2D (right). 100
3.26 S-curve comparison between functional pixels only (left) and all pixels including faulty ones (right). Both tuned to 1500 electrons. . 101
3.27 Threshold distribution in 1D (left) and 2D map (right). 101
3.28 Noise distribution in 1D (left) and 2D map (right). 102

4.1 A scheme of example of particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector. The muon and the charged pion are positively charged, and the electron is negatively charged [162]. 106
4.2 Muon identification efficiency for the Tight working points as a function of the muons p_T for muons with $|\eta| < 0.9$ (left) and $|\eta| > 0.9$ (right) [165]. 109
4.3 Tree-level Feynman diagram of τ decay via the weak interaction. . . 113
4.4 Right: Branching fractions of τ lepton decays into leptonic and hadronic final states. Left: Composition of the $\tau\tau$ decay final [56]. . 114
4.5 Layout of the inner and outer grids in η–ϕ space used as particle-level input to DeepTau. The inner grid covers the signal cone with radius 0.1, composed of 11×11 cells sized 0.02×0.02. The outer grid corresponds to the isolation cone with radius 0.5, composed of 21×21 cells sized 0.05×0.05 [178]. 117

4.6 DeepTau network architecture: three input branches for high-level features, inner grid, and outer grid; merged features pass through fully connected layers for classification [35]. 119
4.7 The DeepTau architecture with domain adaptation. A domain classifier is added to distinguish between data and simulation, with gradient reversal applied to the common layers during the training to promote domain-invariant features [35]. 120
4.8 Jet misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35]. 121
4.9 Electron misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35]. . . . 122
4.10 Muon misidentification probability versus τ_h identification efficiency for DeepTau v2.1 and v2.5 in simulated 2018 data [35]. 122

5.1 Pre-fit (left) and post-fit (right) distribution of m_{vis} for UL2018 for DM0 and DM1. 135
5.2 Pre-fit (left) and post-fit (right) distribution of m_{vis} for UL2018 for DM10 and DM11. 136
5.3 Negative log-likelihood profiled as a function of the TES, measured from the m_{vis} observable. Asymmetric parabolas (black line) are fitted to the $-2\Delta(\ln L)$ values below 5. The plots are arranged as follows: top-left for DM0, top-right for DM1, middle-left for DM10, and middle-right for DM11. 137
5.4 Negative log-likelihood profiled as a function of the ID scale factor (ID SF), measured from the m_{vis} observable. Asymmetric parabolas (black line) are fitted to the $-2\Delta(\ln L)$ values below 5. The plots are arranged as follows: top-left for DM0, top-right for DM1, middle-left for DM10, and middle-right for DM11. 138
5.5 Summary of TES (left) and ID SF (right) per DM region for UL2018, obtained from parabolic fits. 138
5.6 Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM0-p_T regions for UL2018. 140
5.7 Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM1-p_T regions for UL2018. 141
5.8 Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM10-p_T regions for UL2018. 142
5.9 Pre-fit (left) and post-fit (right) distribution of m_{vis} for DM11-p_T regions for UL2018. 143
5.10 Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 5. 144

5.11 Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 5. 145
5.12 Summary of TES (left) and ID SF (right) per DM and $p_T(\tau_h)$ regions for UL2018, obtained from parabolic fits. 146
5.13 Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11. 148
5.14 Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11. 149
5.15 Summary of TES (left) and ID SF (right) per DM region for 2022 preEE, obtained from parabolic fits. 149
5.16 Pre-fit (left) and post-fit (right) distributions of the visible mass m_{vis} for 2022 preEE data in the DM0, DM1, DM10, and DM11 decay modes. 150
5.17 Summary of TES (left) and ID SF (right) per DM region for 2022 postEE, obtained from parabolic fits. 151
5.18 Pre-fit (left) and post-fit (right) distributions of the visible mass m_{vis} for 2022 postEE data in the DM0, DM1, DM10, and DM11 decay modes. 152

6.1 Illustration of row-based vs. columnar data layouts [199]. 158
6.2 Schematic of the analysis workflow managed with columnflow [196]. 160
6.3 Schematic of a typical DNN [202]. 161
6.4 Activation function [204]. 162
6.5 Schematic representation of the ABCD method used to estimate the $j \to \tau_h$ fake background. Region D represents the signal region (SR), A the application region (AR), and B/C the derivation regions (DRs). The fake factor is measured in the DRs and applied to the AR to predict the SR yield. 173
6.6 Fake factors measured in the combination of 2023 preBPix and 2023 postBPix data for the different N_{jet} multiplicities and decay modes of the leading τ_h candidate. The FFs are fitted with a polynomial function of the order 2, shown as the solid line, and the 68% CL contour is shown as the shaded area. 174
6.7 Control plots as a closure test for 2023 BPix era. 175
6.8 Control plots of the Signal Region for 2023 BPix era. 176

6.9 Di-τ mass resolution from FastMTT (left) and SVFit (right) on simulated DY events in columnflow. The standard deviation is 0.16 for FastMTT and 0.17 for SVFit. 179
6.10 Decay Plan method (left) and Polarimetric vector method (right) using generator level information. 181
6.11 Decay Plan method (left) and Polarimetric vector method (right) using reconstructed level information (τ_h visible four-momentum). . 181
6.12 Schematic representation of the Lorentz Boost Network (LBN) architecture, including the trainable four-vector combination, boost to rest frames, and physics feature extraction. 183
6.13 Comparison of the ϕ_{CP} distributions: (Top left) full four-momentum regression using the PVPV method; (Top right) partial regression (only p_T) combined with reconstructed information (visible τ_h; (Bottom left) reconstruction-based Decay Plane (DPDP) method using reconstructed information (visible τ_h); (Bottom right) PVPV method using generator-level p_T and all other reconstructed information. 185
6.14 Profile likelihood scan of the CP-mixing angle $\alpha_{H\tau\tau}$ obtained from a simultaneous fit across all DM categories in the 2022 postEE Asimov dataset, corresponding to an integrated luminosity of 26.671 fb^{-1}. Only statistical uncertainties are included. 187

A.1 Pre-fit $p_T(\tau_h)$ distributions for UL2018 across each DM region. The top-left panel shows DM0, top-right shows DM1, middle-left shows DM10, and middle-right shows DM11. A selection cut of $m_T(\mu, \tau) <$ 65 GeV on the transverse mass of the muon and tau has been applied. 194
A.2 Correlation plots between the TES and the ID SF. The scale factors for $20 < p_T(\tau_h) \leq 40$ GeV on the left-hand side and $p_T(\tau_h) > 40$ GeV on the right-hand side. 195
A.3 Comparison of energy fraction carried by visible τ_h between early Run 3 (2023) and UL18 samples using GenPart in NanoAOD. . . . 196
A.4 Obtained TES per DM to be applied on 2022 preEE simulated genuine τ_h. 197
A.5 Obtained TES per DM-p_T to be applied on 2022 postEE simulated genuine τ_h. 198
A.6 The ID SF as a function of the TES during the NLL scan of the TES showing the dependency between the SFs. 199
A.7 Negative log-likelihood profiled as a function of the TES, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11. 200

A.8 Negative log-likelihood profiled as a function of the ID SF, as measured from the m_{vis} observable. Asymmetric parabolae are fitted (black) to the values of $-2\Delta(lnL)$ below 4. Top left DM0, top right DM1, middle left DM10, middle right DM11. 201
A.9 The ID SF as a function of the TES during the NLL scan of the TES showing the dependency between the SFs. 202
A.10 The data-to-simulation scale factors of the τ_h identification efficiency as functions of p_T in the 2018 (left) and 2022 (right) data-taking periods, including all τ_h decay modes, and requiring the D_{jet} Medium working point and $m_T(\mu, p_T^{\text{miss}}) < 65$GeV. The vertical bars correspond to the combined statistical and systematic uncertainties in the individual scale factors. 203

B.1 Regressed distribution vs generator level distribution of the τ_h and pions. 208
B.2 Regressed distribution vs generator level distribution of the τ_h. . . . 208
B.3 Regressed distribution vs generator level distribution of the τ_h and the angle between τ_h and ρ resonance. 209

List of Tables

1 Résumé des méthodes et vecteurs utilisés pour la reconstruction des observables CP. 4
2 Régions ajustées pour chaque jeu de données. Deux ajustements distincts ont été réalisés pour UL2018. 16

1.1 Summary of main τ decay modes with their type, branching ratio (BR), and resonance label. The charged hadrons denoted as $h^{\pm}$ are assumed to be $\pi^{\pm}$ since the CMS detector is not able to distinguish between the different types of $h^{\pm}$ [17]. 20
1.2 Summary of methods and corresponding vectors used. 29

2.1 Summary of LHC beam parameters at nominal conditions [19]. . . . 36
2.2 Maximum fluence values for different regions/components along with corresponding r and z values [144]. 68
2.3 Summary of the number of modules, sensors, readout chips, and pixels in each Inner Tracker substructure, along with the total active silicon area [145]. 72
2.4 Design requirements and specifications of the CROC-V1 chip [26]. . 73

4.1 Target τ_h identification efficiencies for the different working points defined for the three discriminators. The target efficiencies are evaluated with the $H \to \tau\tau$ event sample for τ_h with $p_T \in [30, 70]$ GeV [35].121

5.1 Single-muon HLT paths and corresponding offline p_T selection thresholds for Run 2 and Run 3 datasets. 125
5.2 Summary of selection criteria for extra lepton vetoes applied to suppress background contamination. 126
5.3 Fitted regions for each dataset. For UL2018, two different fits have been performed. 131
5.4 Summary of fit parameters and associated uncertainties. 132

6.1 Simulated signal and background samples used in this analysis. The *Label* column indicates groupings used in plots; the *Color* column refers to the consistent color coding used throughout this chapter. . 165
6.2 Summary of event preselection and baseline object selection criteria. 168
6.3 Summary of standard corrections applied in the analysis. 170

B.1 List of simulated samples for Run 3 runs used for background modeling in the analysis. 206
B.2 List of signal samples for Run 3, with filter efficiencies and cross sections. 207

Université de Strasbourg

Océane PONCET

IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Search for charge-parity violation in the tau Yukawa coupling with CMS Run 3 data and preparation of the tracker irradiation tests for the high luminosity phase of the LHC

Résumé

Cette thèse présente une recherche de violation de charge-parité (CP) dans les désintégrations du boson de Higgs en paires de leptons τ se désintégrant hadroniquement (τ_h), utilisant les premières données Run 3 de l'expérience CMS à $\sqrt{s} = 13,6$ TeV. Une observable sensible à la violation de CP est construite à partir des plans de désintégration du système $\tau_h\tau_h$, exploitant les corrélations de spin. Le taux de mauvaise identification des jets de quarks et de gluons en candidats τ_h est mesuré avec une méthode fondée sur les données. Des techniques d'apprentissage automatique, notamment un Lorentz Boost Network, sont explorées pour améliorer la reconstruction du quadrivecteur des τ_h. Une méthode d'ajustement simultané est développée pour extraire les facteurs correctifs d'identification et d'échelle en énergie des τ_h pour le Run 3. Parallèlement, la mise à niveau du trajectographe pour la phase de haute luminosité du LHC (HL-LHC) est soutenue par la mise en service d'une ligne d'irradiation auprès du cyclotron CYRCé de l'IPHC. La première irradiation d'un module à pixels a validé le dispositif expérimental et les procédures, confirmant la capacité de l'IPHC à conduire un programme d'irradiation de modules.
Mots-clés : violation de charge-parité, boson de Higgs, leptons tau, apprentissage automatique, Lorentz Boost Network, calibration, irradiation, trajectographe, CMS, HL-LHC, cyclotron CYRCé

Abstract

This thesis presents a search for charge-parity (CP) violation in Higgs boson decays to hadronic τ lepton decays (τ_h) using early Run 3 CMS data at $\sqrt{s} = 13.6$ TeV. A CP-sensitive observable is built from the decay planes of the $\tau_h\tau_h$ system, exploiting spin correlations. The misidentification rate of quark and gluon-jets as τ_h candidates is measured with a data-driven method. Machine learning techniques, notably a Lorentz Boost Network, are explored to improve the τ_h four-momentum reconstruction. A simultaneous fit method is developed to extract the τ_h identification and energy scale factors for Run 3. In parallel, the high luminosity LHC (HL-LHC) tracker upgrade is supported by the commissioning of an irradiation beamline at IPHC's cyclotron CYRCé. The first irradiation of a CROC-V1 pixel prototype module has validated the experimental setup and procedure, confirming the module's irradiation uniformity and IPHC's readiness to conduct a module irradiation program.
Keywords: CP violation, Higgs boson, tau leptons, machine learning, Lorentz Boost Network, calibration, irradiation, Tracker, CMS, HL-LHC, CYRCé cyclotron.