



**HAL**  
open science

# Étude des propriétés CP du boson de Higgs dans le canal tau tau dans l'expérience CMS auprès du LHC

Guillaume Bourgatte

► **To cite this version:**

Guillaume Bourgatte. Étude des propriétés CP du boson de Higgs dans le canal tau tau dans l'expérience CMS auprès du LHC. Physique [physics]. Université de Strasbourg, 2020. Français. NNT : 2020STRAE029 . tel-03235975

**HAL Id: tel-03235975**

**<https://theses.hal.science/tel-03235975>**

Submitted on 26 May 2021

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

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

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE PHYSIQUE  
Institut Pluridisciplinaire Hubert Curien (IPHC), UMR 7178

## THÈSE

présentée par :

**Guillaume Bourgatte**

soutenue le : 10 décembre 2020

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/S spécialité : Physique des particules

**Étude des propriétés CP du boson de Higgs  
dans le canal tau tau dans l'expérience CMS  
auprès du LHC**

**THÈSE** dirigée par :

Mme Anne-Catherine LE BIHAN

Institut Pluridisciplinaire Hubert Curien

**RAPPORTEURS :**

M. David ROUSSEAU

Laboratoire de physique des deux infinis Irène Joliot-Curie

M. Jean-Baptiste DE VIVIE DE RÉGIE

Laboratoire de physique subatomique & cosmologie de Grenoble

**AUTRES MEMBRES DU JURY :**

M. Iouri BELIKOV

Institut Pluridisciplinaire Hubert Curien

(président)

Mme Elisabetta GALLO

Deutsches Elektronen-Synchrotron

M. Éric CHABERT

Institut Pluridisciplinaire Hubert Curien

(co-encadrant)



*A Odin.*



# Remerciements

J'aimerais tout d'abord remercier Anne-Catherine Le Bihan, ma directrice de thèse. Ta personnalité calme et rassurante a été en parfaite adéquation avec mon tempérament de stressé et pessimiste. C'est aussi grâce à toi que j'ai pu obtenir ce diplôme tant désiré. J'en profite pour remercier également Stéphanie Beauceron et Muriel Vander Donckt, mes tutrices de stage de M2, sans qui je n'aurais pu faire cette thèse. Merci également à tous les étudiants et professeurs que j'ai croisés durant mon cursus.

Je remercie ensuite le groupe CMS de l'IPHC. Vous avez tous vos personnalités mais vous avez un point commun : vous êtes cool. Ce fut un réel plaisir de travailler dans cet environnement. Un merci tout particulier à Éric Chabert et Jean-Laurent Agram avec qui j'ai travaillé sur l'étude des performances du trajectographe de l'expérience CMS.

Merci à toutes les autres personnes ayant travaillé avec moi durant ces trois années. Tout d'abord Vladimir Cherepanov, post-doctorant dans le bureau à côté du mien lors de mon arrivé. Tu m'as permis de commencer ma thèse dans de bonnes conditions et de m'apprendre à utiliser nos outils de travail. Ensuite, je remercie mon groupe d'analyse pour leur bienveillance et pour le fait de toujours avoir été disponible pour répondre à mes questions. Sans eux je n'aurais jamais pu faire partie de la publication de l'analyse.

Je souhaiterais également remercier tous les membres de mon jury de thèse pour avoir accepté d'en faire partie. Merci d'avoir suivi mes travaux et de m'avoir accordé ce diplôme. Un merci particulier à mes deux rapporteurs, David Rousseau et Jean-Baptiste de Vivie de Régie, pour avoir lu très attentivement mon manuscrit et pour m'avoir fait des retours pertinents.

Mes années de thèse m'ont permis de faire de très belles rencontres. Merci à tous les doctorants et post-doctorants que j'ai côtoyés durant ma thèse à l'IPHC, notamment Nicolas Tonon et Markéta Jansová qui m'ont tout de suite très bien intégré à leur groupe. Je pense aussi évidemment aux doctorants actuels qui vont beaucoup me manquer : Douja, Clément, Dylan, Mario, Alex, Julie, Robin, tous les nouveaux en première année, et j'en oublie forcément. Je vous souhaite à tous une excellente thèse. Nos jeux de cartes après manger et nos pauses café vont vraiment me manquer. Je souhaite à Mario Sessini, qui me succède, de réussir dans tous ses projets et qu'il vive une thèse très enrichissante sur tous les plans.

Enfin, je termine par Simon et Geoffrey, mes frères. Vous m'avez permis de passer des moments magiques, comme toujours. Cela m'a aussi aidé pendant les périodes de stress afin de décompresser et penser à autre chose. Je me rappelle notamment de ton chalet Simon, où j'ai du concilier pendant plusieurs jours travail sur mon ordinateur et fête. J'espère que ces moments avec vous ne sont pas près de s'arrêter.

# Sommaire

<b>Introduction</b>	<b>1</b>
<b>1 Le Modèle Standard de la physique des particules</b>	<b>3</b>
1.1 Particules	3
1.2 Interactions	4
1.2.1 Interaction forte	4
1.2.2 Interaction électromagnétique	6
1.2.3 Interaction faible	6
1.2.4 Interaction électrofaible	7
1.2.5 Mécanisme de Brout-Englert-Higgs et brisure de symétrie électrofaible	8
1.2.6 Couplages de Yukawa	10
1.3 Limites du Modèle Standard	10
1.4 Études du boson de Higgs	13
<b>2 L'expérience CMS</b>	<b>17</b>
2.1 Le Conseil européen pour la recherche nucléaire	17
2.2 Le Grand collisionneur de hadrons	18
2.3 L'expérience CMS	22
2.3.1 Système de coordonnées	22
2.3.2 Le trajectographe en silicium	23
2.3.3 Le calorimètre électromagnétique	25
2.3.4 Le calorimètre hadronique	26
2.3.5 L'aimant	28
2.3.6 Les chambres à muons	28
2.3.7 Le système de déclenchement	30
2.4 Le LHC à haute luminosité	31
<b>3 Reconstruction des particules et des objets</b>	<b>36</b>
3.1 L'algorithme du flux de particules	36
3.2 Objets physiques reconstruits	39
3.2.1 Muons	39
3.2.2 Électrons	40
3.2.3 Photons	42



3.2.4	Jets	43
3.2.5	Taus	45
3.2.5.1	Adaptation de l'algorithme de rejet des électrons pour la Phase 2	53
3.2.6	Énergie transverse manquante	60
<b>4</b>	<b>Le trajectographe en silicium et l'étude des particules hautement ionisantes</b>	<b>62</b>
4.1	Les capteurs en physique des particules	62
4.1.1	Les semi-conducteurs	62
4.1.2	Les capteurs à pistes en silicium de CMS	64
4.2	Particules hautement ionisantes	66
4.2.1	Contexte de l'étude menée	67
4.2.2	Études précédentes sur les HIP	67
4.2.3	Étude de l'inefficacité de coup dans le trajectographe en silicium avec les données 2018	71
4.3	Conclusion	77
<b>5</b>	<b>Mesure de l'état CP du boson de Higgs à <math>\sqrt{s} = 13</math> TeV</b>	<b>79</b>
5.1	Méthodes	79
5.1.1	Méthode du paramètre d'impact	79
5.1.2	Méthode du plan de désintégration	81
5.1.3	Méthode du vecteur polarimétrique	82
5.2	Échantillons	89
5.2.1	Données	89
5.2.2	Simulations	89
5.2.3	Échantillons embedded	90
5.3	Sélections des objets	91
5.3.1	Leptons	91
5.3.2	Jets	91
5.3.3	MET	91
5.3.4	Vertex	92
5.3.5	Système de déclenchement et filtres	98
5.4	Coupures	99
5.4.1	Filtres MET	99
5.4.2	Veto de lepton supplémentaire	99
5.4.3	Veto dilepton	100
5.4.4	Sélections	100
5.5	Corrections	101
5.5.1	Échelle d'énergie des $\tau_h$	101
5.5.2	Efficacité d'identification des $\tau_h$	101
5.5.3	Facteurs d'échelle du système de déclenchement du tau hadronique	102

## Sommaire

5.5.4	Mauvaise identification des électrons et muons en $\tau_h$ . . . . .	102
5.5.5	Corrections de recul MET . . . . .	102
5.5.6	Repondération de l'empilement . . . . .	103
5.5.7	Repondération du $p_T$ et de la masse du $Z$ . . . . .	103
5.5.8	Repondération du $p_T$ du top . . . . .	103
5.5.9	Poids prefiring . . . . .	105
5.6	Estimation du bruit de fond . . . . .	105
5.6.1	Méthode des fake factors . . . . .	105
5.6.1.1	Fake factors dans le canal $\tau_h\tau_h$ . . . . .	106
5.6.1.2	Traitement des incertitudes statistiques . . . . .	109
5.7	BDT de classification . . . . .	110
5.8	Lissage des modèles . . . . .	112
5.9	Incertitudes . . . . .	120
5.9.1	Incertitudes de normalisation . . . . .	120
5.9.2	Incertitudes de forme . . . . .	124
5.9.3	Corrélations entre les incertitudes MC et embedded . . . . .	127
5.10	Résultats . . . . .	127
5.10.1	Résultats dans le canal $a_1^{3pr} a_1^{3pr}$ . . . . .	128
5.11	Conclusion . . . . .	138
<b>Conclusion</b>		<b>140</b>
<b>A Tableaux des échantillons utilisés dans le canal <math>\tau_h\tau_h</math></b>		<b>142</b>
<b>B Extraction des corrections d'identification des <math>\tau_h</math></b>		<b>147</b>
<b>C Distributions avant ajustement de <math>\phi_{CP}</math> dans le canal <math>a_1^{3pr} a_1^{3pr}</math></b>		<b>152</b>
<b>Bibliographie</b>		<b>157</b>



# Introduction

Durant le XX<sup>e</sup> siècle, la physique entra dans une nouvelle ère avec l'essor de la cosmologie, l'élaboration de la relativité restreinte et la découverte des propriétés du monde quantique. Développé à partir des années 60, le Modèle Standard de la physique des particules, basé sur la théorie quantique des champs, décrit les constituants les plus élémentaires de l'Univers et leurs interactions. Ce dernier explique la majorité des résultats expérimentaux observés jusqu'à présent avec une précision inégalée. La dernière particule prédite et non découverte jusque là, le boson de Higgs, fut observée au Grand collisionneur de hadrons (LHC) en 2012.

Le [chapitre 1](#) décrit les particules fondamentales du Modèle Standard, notamment le boson de Higgs, et leurs interactions. Cependant, et malgré les nombreux succès du Modèle Standard, ce dernier n'explique pas tout. En effet, par exemple, il ne décrit pas l'interaction gravitationnelle ni n'explique l'asymétrie matière-antimatière de l'Univers. Un modèle plus général et englobant le Modèle Standard doit donc être trouvé. C'est pourquoi les prédictions du Modèle Standard sont testées avec une grande précision et notamment les propriétés du boson de Higgs. En effet, des observations expérimentales différant des prédictions seraient le signe d'une nouvelle physique au-delà du Modèle Standard.

Le [chapitre 2](#) présente d'abord le LHC, puis plus en détail les différents sous-détecteurs de l'expérience CMS (« Compact Muon Solenoid ») situé sur l'anneau du LHC et qui, conjointement avec l'expérience ATLAS (« A Toroidal LHC ApparatuS »), ont permis de découvrir expérimentalement le boson de Higgs. Les améliorations qui seront apportées à l'expérience CMS en vue de la phase à haute luminosité du LHC sont aussi décrites.

Le [chapitre 3](#) se concentre sur la reconstruction des objets physiques au sein de l'expérience CMS. En effet, l'algorithme du flux de particules (PF) permet de combiner toutes les informations des différents sous-détecteurs, comme des dépôts d'énergie dans les calorimètres ou des traces dans le trajectographe, afin de remonter aux particules initialement produites. Mon étude sur l'identification des leptons taus et plus précisément sur le rejet des électrons se faisant faussement reconstruire comme des taus hadroniques est présentée en [sous-sous-section 3.2.5.1](#).

## Introduction

Le [chapitre 4](#) détaille le fonctionnement des pistes en silicium du trajectographe de CMS avant de présenter mon étude sur les événements hautement ionisants (HIP) et leur effet sur l'efficacité de reconstruction des impacts. En effet, ces HIP interviennent lors d'interactions nucléaires inélastiques au niveau du volume sensible des capteurs. Les hautes énergies déposées saturent l'électronique frontale et induisent donc un temps mort.

Enfin, le [chapitre 5](#) expose l'analyse, à laquelle j'ai participé, de la collaboration CMS sur la mesure de l'état de charge-parité (de nature scalaire ou pseudoscalaire) du boson de Higgs dans ses désintégrations en paires de leptons taus. Mon travail a été d'étudier le canal  $a_1^{3pr} a_1^{3pr}$  et de mettre en œuvre la méthode dite du vecteur polarimétrique qui améliore la sensibilité dans ce canal. Le boson de Higgs étant prédit par le Modèle Standard comme une particule scalaire, une déviation dans les observations signifierait que ce boson de Higgs n'est pas celui prédit mais qu'il pourrait provenir d'un modèle au-delà du Modèle Standard.

# Chapitre 1

# Le Modèle Standard de la physique des particules

## 1.1 Particules

Le Modèle Standard (MS) de la physique des particules est le modèle actuel qui décrit les particules constituant notre univers et leurs interactions les plus fondamentales. Ce modèle, se basant sur la théorie quantique des champs et les symétries de jauge, fut développé au début des années 70. Deux grandes familles de particules élémentaires y sont présentes : les fermions de spin demi-entier et les bosons de spin entier.

Les premiers constituent la matière et sont représentés par des champs spinoriels. En font partie les leptons chargés électriquement (électrons, muons et taus), neutres électriquement (neutrinos) et les quarks classés en trois générations de masse croissante (voir Fig. 1.1). De plus, à chacune de ces particules est associée une antiparticule ayant des charges opposées.

La seconde famille contenant les bosons possède elle-même deux catégories de particules. La première contient les bosons de spin 1, vecteurs des interactions faibles (bosons W et Z), électromagnétiques (photons) et fortes (gluons) (voir Fig. 1.1). En effet, les interactions entre les particules sont décrites comme l'échange d'une particule médiatrice appelée boson de jauge représentée par un champ vectoriel. La seconde catégorie ne contient que le boson de Higgs de spin nul représenté par un champ scalaire. Plus de détails sur le boson de Higgs et son champ seront donnés plus bas.

Chaque particule précédemment citée est décrite par un champ associé à une probabilité de trouver la particule en question à un certain point de l'espace-temps. L'évolution de ce champ  $\phi$  est décrite par une action  $S$  qui est une densité lagrangienne  $\mathcal{L}$  intégrée sur un volume de l'espace-temps :

$$S \equiv \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x, \quad (1.1)$$

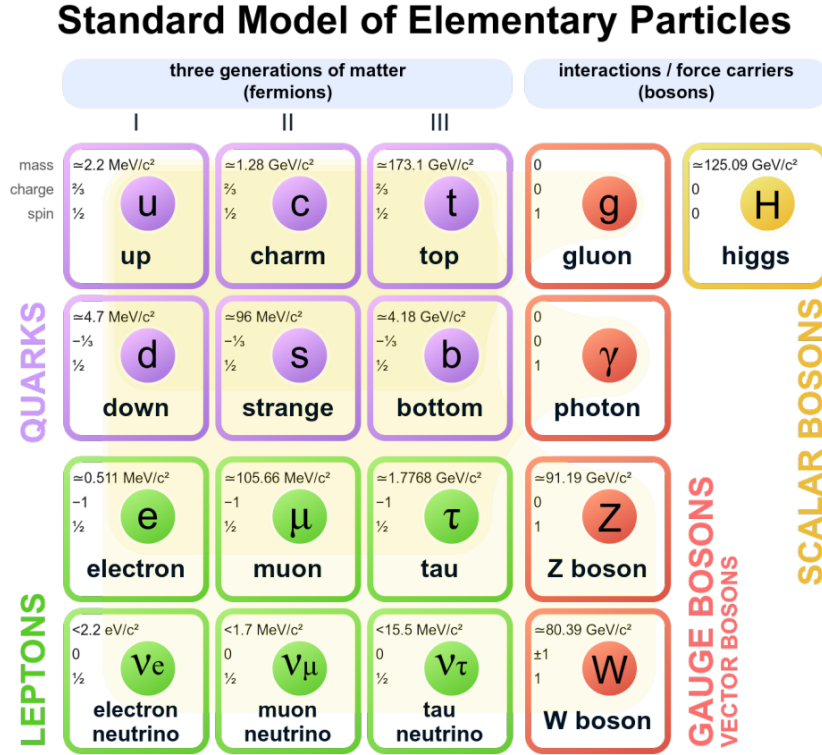


FIGURE 1.1 – Modèle Standard de la physique des particules.

avec  $\mathcal{L}$  contenant une partie cinétique  $T$  et potentielle  $V$  (cette dernière étant associée à la masse et aux interactions de la particule) :

$$\mathcal{L}(\phi, \partial_\mu \phi) = T(\phi, \partial_\mu \phi) - V(\phi, \partial_\mu \phi). \quad (1.2)$$

Le théorème de Noether veut qu'une correspondance soit faite entre un lagrangien invariant sous une symétrie et une quantité physique conservée. Par exemple, l'invariance du lagrangien par translation dans l'espace est associée à la conservation de l'impulsion, tandis que l'invariance dans le temps est associée à la conservation de l'énergie. Le MS est fondé sur la symétrie  $SU(3)_C \times SU(2)_L \times U(1)_Y$ .

## 1.2 Interactions

### 1.2.1 Interaction forte

La chromodynamique quantique (QCD) est la théorie de l'interaction forte. Elle est fondée sur le groupe de symétrie  $SU(3)_C$ . Ce dernier est non abélien : les gluons, vecteurs de l'interaction forte, portent donc eux-mêmes une charge dite de couleur qui est une grandeur conservée et interagissent donc aussi entre eux. Les quarks se combinent en singlets de couleurs et forment des mésons (spin entier avec nombre pair de quarks de

couleurs opposées) ou des baryons (spin demi-entier avec nombre impair de quarks de couleurs différentes). Le lagrangien QCD peut s'écrire comme :

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i (i\gamma_\mu D_{ij}^\mu - m\delta_{ij})\psi_j - \frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu}, \quad (1.3)$$

avec  $i$  et  $j$  les indices de couleurs,  $\psi$  le champ de Dirac du quark,  $m$  sa masse,  $\delta_{ij}$  le delta de Kronecker égal à 1 si  $i=j$  et égal à 0 autrement,  $\gamma^\mu$  les matrices de Dirac, la dérivée covariante :

$$D_{ij}^\mu = \partial^\mu \delta_{ij} + ig_s t_{ij}^a A_a^\mu, \quad (1.4)$$

et le tenseur chromodynamique :

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f_{abc} A_\mu^b A_\nu^c, \quad (1.5)$$

avec  $g_s$  la constante de couplage de l'interaction forte,  $t^a$  les huit générateurs de  $SU(3)_C$  pouvant s'exprimer en fonction des matrices de Gell-Mann  $\lambda^a$  comme  $t^a = \frac{1}{2}\lambda^a$ ,  $A_a^\mu$  le champ vectoriel des gluons de couleur  $a$  et  $f_{abc}$  les constantes de structure de  $SU(3)_C$ .

L'interaction forte possède une caractéristique non usuelle car à basse énergie, c'est-à-dire à grande distance, la force de couplage augmente et oblige les quarks et gluons à se confiner dans des hadrons. Tandis qu'à haute énergie, donc à petite distance, la valeur de la force de couplage diminue et les quarks sont libres (voir Fig. 1.2).

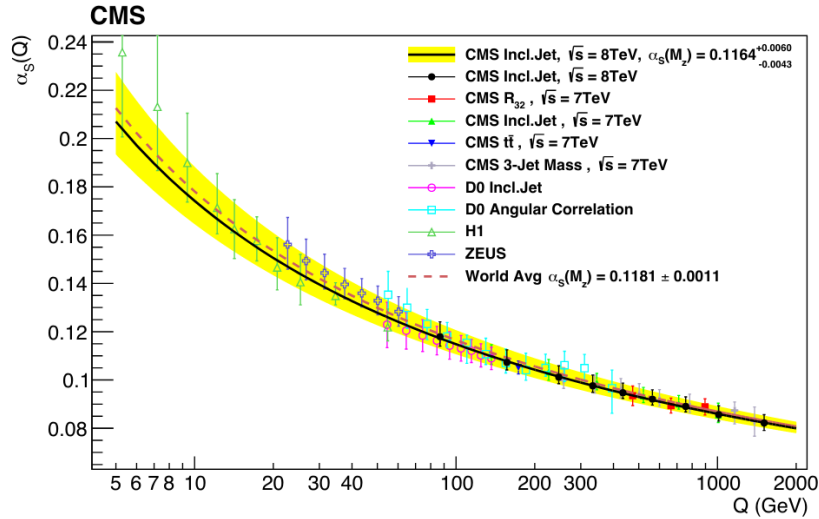


FIGURE 1.2 – Valeurs mesurées de la force de la constante de couplage QCD  $\alpha_s = \frac{g_s^2}{4\pi}$  en fonction de l'énergie  $Q$  [1].



## 1.2.2 Interaction électromagnétique

La théorie de l'électrodynamique quantique (QED) est fondée sur un lagrangien composé de trois parties :

$$\begin{aligned}
 \mathcal{L}_{\text{QED}} &= \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Maxwell}} + \mathcal{L}_{\text{int}} \\
 &= \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - e\bar{\psi}\gamma^\mu\psi A_\mu \\
 &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu},
 \end{aligned} \tag{1.6}$$

avec  $A_\mu$  le potentiel vecteur électromagnétique,  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$  la force du tenseur de champ et  $D_\mu = \partial_\mu + ieA_\mu$  la dérivée covariante permettant de préserver l'invariance de jauge locale. Le groupe de symétrie  $U(1)$  de la QED est abélien : le photon, vecteur de l'interaction électromagnétique, ne possède donc pas de charge électrique. Selon le théorème de Noether, la jauge de symétrie de l'action implique une conservation de la charge électrique.

## 1.2.3 Interaction faible

L'interaction faible fut d'abord décrite comme une interaction de contact par Fermi en 1933 puis devint une interaction à courte distance par la suite. Elle est fondée sur le groupe de symétrie  $SU(2)_L$  donnant trois bosons de jauge vecteurs de l'interaction que sont les bosons  $Z$ ,  $W^+$  et  $W^-$ . L'interaction faible est particulière car elle viole la symétrie de parité ( $P$ ) qui inverse les trois directions spatiales. Par exemple, une expérience fut réalisée par Mme C.S. Wu et ses collaborateurs en 1957 [2]. Ils alignèrent le spin des noyaux de cobalt 60 et comptèrent le nombre d'électrons provenant des désintégrations  $\beta$  des noyaux radioactifs dans le sens du spin nucléaire et dans le sens opposé. Leurs résultats montrèrent qu'il y avait bien plus d'électrons émis dans le sens opposé du spin nucléaire que dans le même sens. Or, selon la symétrie de parité, on devrait observer autant d'électrons dans un sens que dans l'autre, cette symétrie est donc violée.

La solution envisagée pour ce problème est l'introduction d'interactions différentes pour les fermions gauches et droits. Les interactions via un boson  $W^\pm$  ne contiennent que les champs fermioniques gauches et les interactions via un boson  $Z^0$  contiennent les deux champs fermioniques gauches et droits mais avec des couplages différents. En effet, les spineurs de Dirac représentant les champs des fermions, peuvent être décomposés en deux composantes chirales gauche et droite :  $\psi = \psi_L + \psi_R$ , avec  $\psi_L = P_L\psi$ ,  $P_L = \frac{(1-\gamma_5)}{2}$ ,  $\psi_R = P_R\psi$  et  $P_R = \frac{(1+\gamma_5)}{2}$ .

Par la suite on crut que la véritable symétrie était celle dite de charge-parité (CP) qui transforme un fermion gauche en un anti-fermion droit. Mais il fut prouvé expérimentalement que la symétrie CP était aussi violée (0.3% des interactions faibles), bien que moins souvent que la seule symétrie P. Une véritable symétrie, considérée comme étant une propriété fondamentale des lois physiques, dont on n'a jamais découvert de violation est celle dite CPT, où en plus de la symétrie CP, une inversion temporelle est appliquée. Une telle violation de symétrie impliquerait une violation de l'invariance de Lorentz. C'est pourquoi, afin de préserver cette symétrie, toute violation de symétrie combinée de deux de ses composantes (comme CP) implique une violation de la troisième composante (comme T), car mathématiquement équivalente.

### 1.2.4 Interaction électrofaible

Un des principaux objectifs en physique des particules est d'unifier toute les interactions en une seule. L'interaction électrofaible est un premier succès et combine les interactions électromagnétique et faible, en partant de quatre bosons sans masse,  $W^1$ ,  $W^2$ ,  $W^0$  et  $B^0$ . Les deux bosons neutres sont à l'origine des deux bosons neutres physiques :

$$\begin{pmatrix} Z^0 \\ \gamma \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W^0 \\ B^0 \end{pmatrix}, \quad (1.7)$$

avec  $\theta_W$  l'angle de mélange électrofaible dit de Weinberg qui relie les interactions électromagnétique et faible comme  $e = g \sin \theta_W = g' \cos \theta_W$ , avec  $e$  la charge électrique,  $g'$  la constante de couplage associée au champ de jauge B et  $g$  la constante de couplage associée aux champs de jauge W. Les deux bosons chargés physiques proviennent de  $W^1$  et  $W^2$  :

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2). \quad (1.8)$$

Le lagrangien électrofaible peut s'écrire comme :

$$\mathcal{L}_{\text{EWK}} = (i\bar{\psi}_L \gamma_\mu D_\mu^L \psi_L) + (i\bar{\psi}_R \gamma_\mu D_\mu^R \psi_R) - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu}^0 B_0^{\mu\nu}, \quad (1.9)$$

avec :

$$\begin{aligned} D_\mu^L &= \partial_\mu + i\frac{g}{2} W_\mu^a T_a + i\frac{g'}{2} Y B_\mu, \\ D_\mu^R &= \partial_\mu + i\frac{g'}{2} Y B_\mu, \\ W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon_{bc}^a W_\mu^b W_\nu^c, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned} \quad (1.10)$$

avec  $T_a$  les trois composantes de l'isospin faible, générateurs du groupe  $SU(2)_L$  et  $Y$  l'hypercharge faible, générateur du groupe  $U(1)_Y$ . Mais comme énoncé précédemment, les bosons  $W$  et  $Z$  devraient avoir une masse nulle, c'est là qu'intervient le boson de Higgs et le mécanisme de Brout-Englert-Higgs.

### 1.2.5 Mécanisme de Brout-Englert-Higgs et brisure de symétrie électrofaible

Une brisure de symétrie doit être introduite si l'on veut que les bosons  $W$  et  $Z$  aient une masse tout en gardant le photon sans masse. Dans le MS, la brisure de symétrie électrofaible se fait via le mécanisme de Brout-Englert-Higgs [3-5]. Ce mécanisme introduit un doublet complexe de champs scalaires :

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}. \quad (1.11)$$

Une contribution supplémentaire au lagrangien électrofaible est alors ajoutée :

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \quad (1.12)$$

avec  $D_\mu$  égal à  $D_\mu^L$  défini en Eq. 1.10 et le potentiel  $V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$ . Le terme  $\lambda$  décrit l'auto-couplage du champ scalaire et si l'on suppose la stabilité du vide, sa valeur doit être supérieure à 0. Le signe de  $\mu^2$  définit la forme du potentiel du champ de Higgs. En effet,  $\mu^2 < 0$  correspondrait à un puits de potentiel autour d'un minimum qu'est le vide  $\phi_0 = 0$ . Le lagrangien en Eq. 1.12 représenterait donc une particule scalaire de masse  $\mu$  et une auto-interaction en quatre points de couplage  $\lambda$ . En revanche,  $\mu^2 > 0$  impliquerait un potentiel en forme de chapeau mexicain (voir Fig. 1.3). Cette forme de potentiel permet une infinité d'états minimums satisfaisant :

$$\langle \phi \rangle = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}}, \quad (1.13)$$

avec  $v$  la « vacuum expectation value » (v.e.v.) ou valeur attendue dans le vide d'une valeur proche de 246 GeV. Ceci a pour conséquence de rendre le système instable à  $\phi = 0$  et le force à choisir une valeur minimale particulière, c'est ce qu'on appelle une brisure de symétrie spontanée. L'ensemble des solutions possibles partage la symétrie du lagrangien mais pas la solution choisie seule. On dit que la véritable symétrie du système est cachée par le choix d'un état fondamental spécifique qui peut s'écrire comme :

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.14)$$

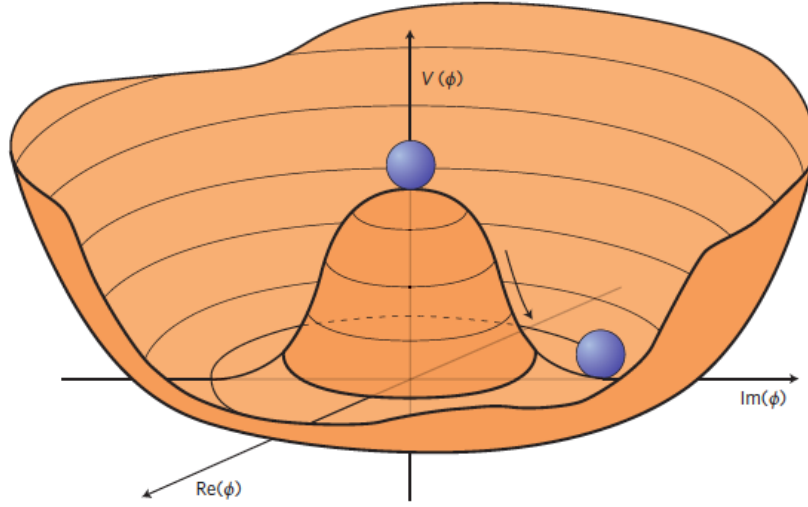


FIGURE 1.3 – Forme du potentiel du champ de Higgs, dit en chapeau mexicain, menant à une brisure spontanée de symétrie [6].

Le théorème de Goldstone dit qu'à chaque fois qu'une symétrie continue est spontanément brisée, chaque générateur brisé donne un boson scalaire sans masse dit boson de Goldstone. Si on développe le champ autour de l'état fondamental on obtient :

$$\phi(\mathbf{x}) = \frac{1}{\sqrt{2}} e^{\frac{i\xi_a(\mathbf{x})T^a}{v}} \begin{pmatrix} 0 \\ v + h(\mathbf{x}) \end{pmatrix}. \quad (1.15)$$

Les quatre degrés de liberté de l'Eq. 1.11 sont portés par les trois bosons de Goldstone  $\xi_a$  et le champ scalaire  $h$ . L'invariance du lagrangien sous les transformations  $SU(2)_L$  permet de choisir  $\xi_a(\mathbf{x}) = 0$ , appelé choix de jauge unitaire. Le champ de Higgs reste alors le seul champ physique présent. Les trois degrés de liberté des bosons de Goldstone sont transformés en trois degrés de liberté longitudinaux des états propres de masse des  $W^\pm$  et  $Z^0$  qui deviennent massifs. Dit autrement, en développant les termes cinétique et potentiel de l'Eq. 1.12 dans le choix de jauge unitaire, on génère des termes de masses pour les bosons  $W$  et  $Z$  ainsi que pour le boson de Higgs qui dépend de  $v$  tout en laissant le photon sans masse :

$$\begin{aligned} m_W &= \frac{gv}{2}, \\ m_Z &= \frac{v\sqrt{g^2 + g'^2}}{2}, \\ m_H &= \sqrt{2\lambda}v. \end{aligned} \quad (1.16)$$

### 1.2.6 Couplages de Yukawa

La brisure de symétrie électrofaible génère des termes de masse pour les bosons  $W$  et  $Z$ , mais le champ de Higgs donne aussi une masse aux fermions via les interactions de Yukawa [7] qui couplent un champ de Dirac  $\psi$  avec un champ scalaire  $\phi$ . Le terme de Yukawa dans le lagrangien est :

$$\mathcal{L}_{\text{Yukawa}} = -y_f(\bar{f}_L\phi f_R + \bar{f}_R\phi f_L), \quad (1.17)$$

avec  $f_L$  les doublets gauches et  $f_R$  les singlets droits des champs fermioniques, et  $y_f$  le couplage de Yukawa des fermions au champ de Higgs proportionnel à la masse des fermions :

$$y_f = -\frac{\sqrt{2}m_f}{v}. \quad (1.18)$$

## 1.3 Limites du Modèle Standard

Malgré la robustesse du MS qui décrit la grande majorité des observations faites en physique des particules, certaines et non des moindres restent inexplicées. Une liste non exhaustive des limites du MS est présentée ci-dessous.

La plus évidente des lacunes du MS est l'absence de la quatrième interaction fondamentale qu'est la gravitation, décrite dans la relativité générale par une distorsion de l'espace-temps. Ceci n'est pas handicapant pour la physique des particules expérimentale car à cette échelle la force de gravitation est négligeable. Cependant, être incapable de concilier la physique quantique avec la relativité générale pose un problème fondamental si on recherche une seule grande théorie expliquant tout ce qui nous entoure.

Les neutrinos sont considérés comme sans masse. Cependant, l'observation de l'oscillation des neutrinos, qui changent de saveur au cours de leur voyage, implique qu'ils aient une masse. De plus, leur seule chiralité gauche empêche l'introduction d'un terme de masse par des mécanismes similaires à ceux utilisés pour les autres fermions.

Plusieurs observations astronomiques indiqueraient l'existence d'une matière dite noire. Par exemple, la vitesse de rotation des étoiles situées aux périphéries des galaxies est bien plus importante que prévue si on calcule la masse de la galaxie avec la matière visible. Une particule de matière noire massive sensible à l'interaction gravitationnelle, mais interagissant très faiblement avec le MS, pourrait expliquer ce phénomène. Cependant, de telles particules n'ont pas encore été découvertes. La théorie supersymétrique (SUSY) associe une particule partenaire à chaque particule du MS (un boson à un fermion et inversement). De nombreux modèles supersymétriques prédisent que la particule supersymétrique la plus légère doit être stable et neutre électriquement. De plus, elle

doit faiblement interagir avec les particules du MS, ce qui en ferait une excellente candidate de matière noire.

Notre Univers est principalement constitué de matière. Or le Big Bang aurait dû créer autant de matière que d'antimatière qui se seraient mutuellement annihilées. Nos observations laissent donc penser qu'il a dû y avoir une baryogenèse dans l'Univers primordial, qui fit passer ce dernier d'un état où matière et antimatière sont en quantités égales à un état avec un excès de matière. L'asymétrie baryonique de l'Univers est définie comme :

$$\eta = \frac{n_B}{n_\gamma} = \frac{n_b - n_{\bar{b}}}{n_\gamma} \approx 6.10^{-10}, \quad (1.19)$$

avec  $n_\gamma$ ,  $n_b$  et  $n_{\bar{b}}$  les densités de photons, baryons et antibaryons respectivement.

Trois ans après la découverte de violation CP dans les désintégrations  $K_L \rightarrow 2\pi$ , Sakharov établit trois conditions permettant la baryogenèse : violation du nombre baryonique, violation C et CP, et rupture de l'équilibre thermique [8]. Cette dernière est indispensable car sans ça, la symétrie CPT compenserait les processus en augmentant et diminuant le nombre de baryons. Les violations CP découvertes jusqu'à ce jour sont insuffisantes pour expliquer l'excès de matière observé. C'est dans ce contexte que la recherche de violation CP dans le couplage de Yukawa pourrait jouer un rôle.

Un scénario possible est celui de la baryogenèse électrofaible se produisant avant ou pendant la brisure de symétrie électrofaible. En effet, cette symétrie étant restaurée à très haute température ( $T \geq 100$  GeV), le refroidissement de l'Univers a pour conséquence une transition de phase cosmologique où la symétrie est brisée et le champ de Higgs tombe dans un état du vide. L'expansion de l'Univers, suffisamment accélérée à cette température là, permettrait alors une rupture de l'équilibre thermique. Seule une transition de phase du premier ordre permettrait de sortir de l'équilibre [9]. Des bulles de champ de Higgs de différentes valeurs du vide seraient alors créées suffisamment rapidement par rapport aux taux d'interaction des particules. Les particules du plasma environnant interagissent avec les parois des bulles et peuvent générer une asymétrie C et CP [10] (baryogenèse électrofaible dite non locale, voir Fig. 1.4). Ces asymétries se diffusent dans la phase non brisée où les sphalerons, solutions statiques des équations du champ électrofaible du MS, vont produire plus de baryons que d'antibaryons. Une partie des baryons produits à l'extérieur des bulles se retrouvent à l'intérieur de ces dernières qui continuent de croître. À l'intérieur des bulles, les processus sphalerons sont très fortement supprimés et ne peuvent donc pas compenser le surplus de baryons. Ces bulles vont s'étendre jusqu'à qu'il ne reste plus que la phase brisée dans l'Univers. Des études évoquent alors que les leptons  $\tau$  interagissent moins fortement que les quarks top et qu'ils auraient un libre parcours moyen plus long, ce qui leur permettraient de

franchir les parois plus facilement.

Par conséquent, les leptons  $\tau$  dans la baryogenèse électrofaible pourraient donc jouer un rôle important [11, 12] en contribuant à l'augmentation de la densité fermionique gauche. Cependant, le boson de Higgs du MS est trop lourd pour avoir une transition de phase du premier ordre, mais des modèles SUSY pourraient résoudre ce problème en prédisant d'autres bosons de Higgs.

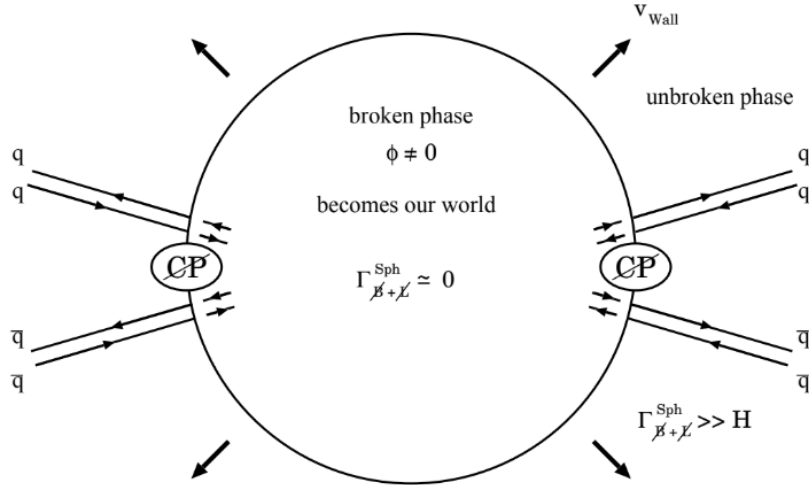


FIGURE 1.4 – Schéma d'une baryogenèse électrofaible non locale [10].

Un autre problème du MS bien connu est que les corrections à la masse du boson de Higgs du MS devrait en faire une particule bien plus lourde que la masse mesurée. En effet, le boson de Higgs se couplant aux fermions, bosons et à lui même, ces couplages impliquent des corrections radiatives à la masse du boson de Higgs. Cette dernière est un paramètre libre mais mesuré du MS. La correction totale dépend quadratiquement de l'échelle d'énergie  $\Lambda$ . On peut choisir cette dernière comme l'échelle de Planck si on considère que le MS est valide jusqu'à cette énergie :  $m_p \sim 2.4 \cdot 10^{18}$  GeV. Dans ce cas, la valeur de la correction de la masse devrait être bien supérieure à la masse elle même. Un ajustement des paramètres est donc nécessaire à une trentaine de décimales près. Cela pose problème car ne semblant pas naturel et une nouvelle physique plus générale et valide jusqu'à l'énergie de Planck pourrait exister, comme une théorie SUSY. En effet, les nouvelles particules prédites par la supersymétrie compenseraient la contribution à la masse du boson de Higgs des particules du MS. Un modèle prometteur est le « Next-to-Minimal Supersymmetric Standard Model » (NMSSM) qui contient un singlet de boson de Higgs de plus que le MSSM. Le NMSSM contiendrait donc trois bosons de Higgs CP-pairs  $H_1$ ,  $H_2$  et  $H_3$ , deux bosons de Higgs CP-impairs  $A_1$  et  $A_2$  et deux bosons de Higgs chargés  $H^+$  et  $H^-$ . Ce modèle est particulièrement intéressant car il autorise la violation CP à l'ordre de l'arbre dans le secteur du boson de Higgs, tandis que le MSSM la préserve, en particulier dans les désintégrations du boson de Higgs en leptons  $\tau$  [13].

## 1.4 Études du boson de Higgs

La plus récente et précise mesure de la masse du boson de Higgs [14] et de sa largeur [15] dans l'expérience CMS (voir section 2.3 pour plus de détails à propos de l'expérience CMS) sont  $125.38 \pm 0.14$  GeV pour un intervalle de confiance (CL) de 68%, et  $3.2_{-3.12}^{+5.96}$  MeV pour 95% CL respectivement. Le couplage du boson de Higgs aux particules élémentaires du MS est linéairement proportionnel à la masse des fermions et proportionnel à la masse au carré des bosons et a été mesuré (voir Fig. 1.5). Le couplage du boson de Higgs aux gluons et photons, qui sont sans masse, est induit par des boucles dans les diagrammes de Feynmann à l'ordre le plus bas. Les désintégrations du boson de Higgs en  $W^-W^+$  [16],  $ZZ$  [17],  $\gamma\gamma$  [18],  $b\bar{b}$  [19] et  $\tau^-\tau^+$  [20] ont été observées. De plus, la première évidence à  $3.0 \sigma$  du boson de Higgs se désintégrant en  $\mu^-\mu^+$  [21] a été annoncée en 2020. Le quark top quant à lui est trop lourd pour provenir de la désintégration d'un boson de Higgs mais la production d'un boson de Higgs en association avec une paire de quarks top a été observée [22]. L'auto-couplage du boson de Higgs peut très difficilement être mesuré dans la version actuelle du LHC à cause de la très faible section efficace estimée de production d'une paire de bosons de Higgs (39.6 fb à 14 TeV dans le centre de masse pour des bosons de Higgs de 125 GeV). Il faudra attendre le HL-LHC (voir section 2.4) pour espérer y être un tant soit peu sensible.

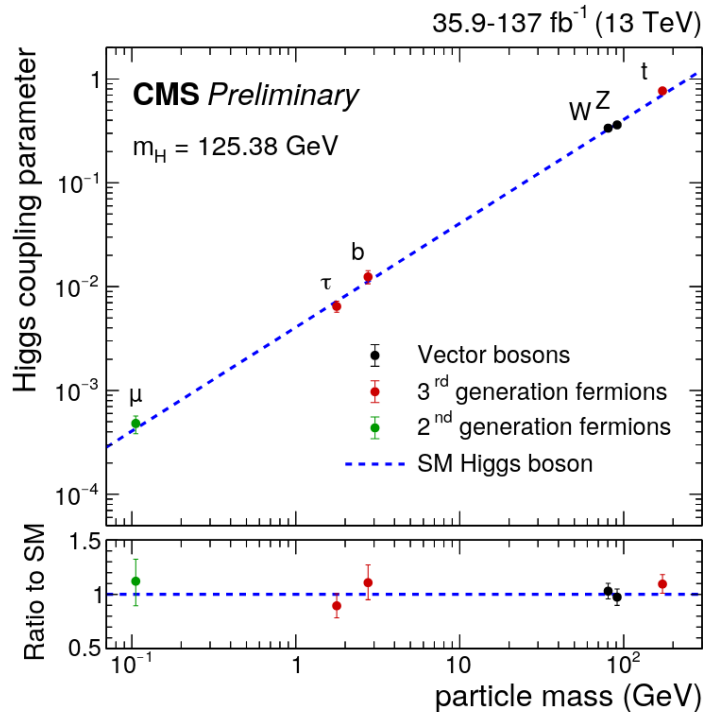


FIGURE 1.5 – Meilleure valeur d'ajustement des paramètres de couplage du boson de Higgs avec différents fermions et bosons. La ligne en pointillé représente la prédiction du MS [21].



Le boson de Higgs prédit par le MS a un état CP-pair ( $J^{\text{CP}} = 0^{++}$ ). Le nombre quantique CP du boson de Higgs est une observable de grand intérêt car une déviation par rapport à un état purement CP-pair (boson purement scalaire) serait le signe tangible d'une nouvelle physique.

Les collaborations ATLAS [23-28] et CMS [29-34] ont effectué de nombreuses analyses de couplages du boson de Higgs avec des bosons vecteurs, dont des tests de violation CP. Ces études ont exclu que le boson de Higgs soit un état purement CP-impair (pseudoscalaire,  $J^{\text{CP}} = 0^{+-}$ ).

Une violation CP dans les couplages de Yukawa aux fermions est possible et tous les couplages sont indépendants et doivent être étudiés. Ces couplages peuvent se décomposer comme :

$$\mathcal{L}_{\text{Yukawa}} = -\frac{m_f}{v} \kappa_f (\cos(\phi_{\text{ff}}) \bar{\psi}_f \psi_f + \sin(\phi_{\text{ff}}) \bar{\psi}_f i \gamma_5 \psi_f) h, \quad (1.20)$$

avec  $\tan(\phi_{\text{ff}}) = \frac{\tilde{\kappa}_f}{\kappa_f}$  l'angle de mélange, et  $\tilde{\kappa}_f$  et  $\kappa_f$  les forces de couplages CP-impair et CP-pair respectivement, avec dans le MS  $\tilde{\kappa}_f = 0$  et  $\kappa_f = 1$ . Pour  $\phi_{\text{ff}}=0$  ( $\frac{\pi}{2}$ ) le boson est scalaire (pseudoscalaire). Pour toute autre valeur de  $\phi_{\text{ff}}$  le boson est dans un état mixte et une violation CP est présente. Cette dernière est maximale quand l'angle de mélange a pour valeur  $\frac{\pi}{4}$ .

Une mesure dans la désintégration en  $b\bar{b}$  est difficile car elle serait affectée par l'hadronisation des quarks b. Le canal  $H \rightarrow c\bar{c}$  souffre du même problème et cette désintégration n'a encore jamais été observée. En revanche, l'état CP du boson de Higgs peut se mesurer dans le couplage de Yukawa au top lors de la production d'un boson de Higgs avec une paire de quarks top où le boson donnera, par exemple, une paire de photons [35]. Cette mesure a été effectuée en s'appuyant sur la définition de  $f_{\text{CP}}^{\text{Htt}}$  :

$$f_{\text{CP}}^{\text{Htt}} = \frac{|\tilde{\kappa}_t|^2}{|\kappa_t|^2 + |\tilde{\kappa}_t|^2} \text{signe}(\tilde{\kappa}_t/\kappa_t). \quad (1.21)$$

Cette analyse étudiant les quarks top utilise une observable,  $D_{0-}$ , qui sépare les états CP-pair et CP-impair. Un arbre de décision boosté (BDT) fut entraîné afin de séparer les deux états CP-pairs et CP-impairs et obtenir  $D_{0-}$ . Les variables d'entrée de l'entraînement sont des variables cinématiques des six premiers jets et du système di-photon (mais pas la masse invariante de la paire), le score d'identification des quarks b, et pour le canal leptonique, la multiplicité des leptons ainsi que des variables cinématiques du lepton de plus haute impulsion transverse. La structure CP mesurée du boson de Higgs a pour valeur  $f_{\text{CP}}^{\text{Htt}} = 0.00 \pm 0.33$  à 68% CL. L'état CP purement pseudoscalaire est rejeté dans le couplage au top à  $3.2 \sigma$  en analysant toutes les données du Run 2.

Initialement, les mesures de haute précision dans le secteur du boson de Higgs (masse, force de couplage et nature CP) étaient pensées pour se faire auprès des collisionneurs électron-positron. En effet, leur environnement très propre dû aux collisions de leptons et leurs détecteurs de haute précision en faisaient de bons candidats. Le plus connu d'entre eux, mais pas encore construit, est le collisionneur linéaire international (ILC) [36, 37] qui prévoit de faire se rencontrer des électrons et positrons polarisés à une énergie initiale de 250 GeV, possiblement augmentée plus tard à 500 GeV, puis potentiellement à 1 TeV. En particulier, la recherche de violation CP dans le couplage de Yukawa au lepton tau a été documentée en détails pour l'ILC. En effet, le référentiel au repos du boson de Higgs y est accessible car s'agissant de collisions  $e^+e^-$ . L'étude présentée dans cette thèse s'appuie en partie sur les études phénoménologiques développées pour l'ILC [38, 39]. Cependant, des études ont démontré que des mesures de violation CP dans les désintégrations du boson de Higgs en paires de leptons  $\tau$  au sein du LHC sont faisables [40-44].

L'objectif de ma thèse est de rechercher une possible violation CP du boson de Higgs dans le couplage de Yukawa du lepton tau, complémentaire aux mesures des moments dipolaires électriques (EDM) qui permettent également de contraindre le couplage CP-impair du boson de Higgs aux leptons taus [45] :

$$\begin{aligned} \frac{d_e}{e} &= 3.7 \cdot 10^{-29} \cdot \tilde{\kappa}_\tau \text{ cm}, \\ \frac{d_n}{e} &= (1.0 \pm 0.5) \cdot 22.3 \cdot 10^{-29} \cdot \tilde{\kappa}_\tau \text{ cm}, \end{aligned} \tag{1.22}$$

avec  $d_e$  l'EDM de l'électron,  $d_n$  l'EDM du neutron et  $e$  la charge élémentaire. Les dernières limites supérieures sur les mesures de l'EDM de l'électron et du neutron sont  $|\frac{d_e}{e}| < 1.1 \cdot 10^{-29} \text{ cm}$  [46] et  $|\frac{d_n}{e}| < 1.8 \cdot 10^{-26} \text{ cm}$  [47] respectivement. Ces mesures permettent, avec l'Eq. 1.22, d'obtenir  $\tilde{\kappa}_\tau < 0.3$  et  $\tilde{\kappa}_\tau < [53.8, 161.4]$ , la mesure de l'EDM de l'électron étant plus contraignante que celle du neutron.

Au LHC, les corrélations de spins des fermions sont exploitées. La désintégration d'un boson (pseudo)scalaire en deux fermions peut s'écrire comme [48] :

$$\Gamma(h \rightarrow f\bar{f}) \propto 1 - s_z \bar{s}_z \pm s_\perp \bar{s}_\perp, \tag{1.23}$$

avec  $\pm$  dépendant de l'état scalaire ou pseudoscalaire du boson,  $s$  et  $\bar{s}$  les vecteurs spin des fermions et anti-fermions respectivement, et  $z$  et  $\perp$  spécifiant les composantes longitudinales et transverses respectivement. La configuration de polarisation du tau affecte la distribution angulaire des désintégrations du système di-tau. Pour une désintégration en un pion chargé et un neutrino par exemple, les composantes transverses de l'impulsion des pions chargés sont anti-alignées pour un boson scalaire, et alignées pour un pseudoscalaire, comme représenté en Fig. 1.6. Par conséquent, l'angle entre les plans de désintégration des leptons taus est sensible à l'état CP du boson de Higgs. Cette propriété sera exploitée dans l'analyse présentée au chapitre 5.

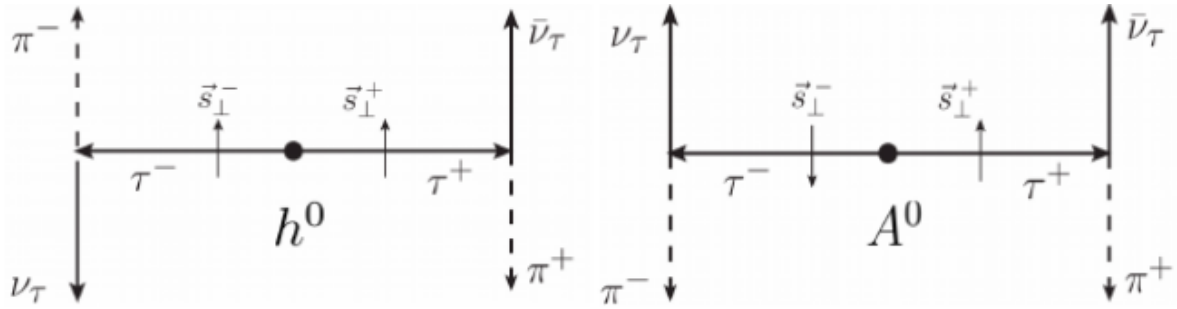


FIGURE 1.6 – Topologie d'une désintégration en un pion et un neutrino d'un boson de Higgs scalaire,  $h^0$ , avec des produits de désintégration anti-parallèles (gauche) et d'un boson de Higgs pseudoscalaire,  $A^0$ , avec des produits de désintégration parallèles [49].

## Chapitre 2

# L'expérience CMS

### 2.1 Le Conseil européen pour la recherche nucléaire

À la fin des années 40, un groupe de physiciens de plusieurs pays imagine la création d'un laboratoire européen de physique atomique. Le but était de partager les coûts croissants des installations de recherche nucléaire ainsi que de freiner la fuite des cerveaux vers les USA qui a commencé durant la Seconde Guerre mondiale et tenter d'unifier l'Europe d'après-guerre. La Suisse fut choisie pour la construction du Conseil européen pour la recherche nucléaire (CERN) grâce à sa localisation centrale en Europe, au fait qu'elle ait déjà accueilli un certain nombre d'organisations internationales et à sa neutralité durant la Seconde Guerre mondiale. La convention écrite en 1953 stipule que les activités du CERN n'auront aucun but militaire et que les résultats des travaux expérimentaux et théoriques seront publiés ou rendus accessibles.

Les travaux commencèrent en 1954 et le premier accélérateur de particules du site, le Synchro-Cyclotron (SC), fut construit en 1957 et accélérât un faisceau de protons à 600 MeV sur cible fixe. L'année suivante, sa première grande contribution fut la découverte de la désintégration rare d'un pion en électrons. En 1959, le Proton Synchrotron (PS) accélérât des protons à 28 GeV, battant le record mondial à 10 GeV tenu par les Russes.

Il fallait, dans les années 60, examiner des millions de photographies de chambres à bulles ou à étincelles. George Charpak développa en 1968 la « chambre proportionnelle multicâbles » qui est une boîte remplie de gaz avec un grand nombre de câbles détecteurs en parallèles connectés chacun à des amplificateurs individuels. Cette boîte était reliée à un ordinateur et pouvait atteindre un taux de comptage mille fois supérieur à ceux habituels.

Par la suite dans les années 70, le CERN utilisa le PS afin d'alimenter les nouveaux accélérateurs en particules et on collisionna pour la première fois au monde des protons entre eux en 1971 dans les anneaux de stockage à intersections (ISR) avec une énergie dans le centre de masse de 62 GeV au lieu d'utiliser une cible fixe. Cette nouvelle façon de faire permet, pour une même énergie de collision, de drastiquement diminuer l'énergie

des faisceaux par rapport à une collision sur cible fixe. En 1976, le Super Synchrotron à Protons (SPS), d'une circonférence de sept kilomètres, est le premier anneau géant souterrain du CERN et le premier accélérateur à traverser la frontière franco-suisse. Son énergie dans le centre de masse était de 400 GeV. En 1981 le SPS est modifié et collisionne des protons avec des antiprotons en permettant la découverte en 1983 des bosons W et Z.

En 1989, Tim Berners-Lee, un chercheur britannique, inventa le « World Wide Web » afin que les scientifiques d'instituts du monde entier puissent s'échanger des informations. Le premier site Web créé au monde fut au CERN et était destiné à ce projet. En 1993 le CERN mit ce logiciel dans le domaine public puis en fit une version sous licence libre.

Le Large Electron-Positron (LEP) entra en service en 1989 et devint l'accélérateur électron-positron le plus grand au monde avec ses 27 kilomètres de circonférence. Il commença avec une énergie dans le centre de masse de 91 GeV (la masse du boson Z) pour finir à 209 GeV en 2000 lors de son arrêt. Le LEP prouva, entre autre, qu'il n'y a que trois générations de fermions légers se couplant au boson Z.

## 2.2 Le Grand collisionneur de hadrons

En 2008, le Grand collisionneur de hadrons (LHC), collisionnant des protons, démarre et se situe dans le tunnel du LEP faisant moins de 4 m de diamètre entre 45 et 170 m sous terre. C'est le plus grand et le plus puissant des accélérateurs de particules conçu initialement pour découvrir le boson de Higgs, découvert en 2012 par les collaborations ATLAS [50] et CMS [51] (voir [section 1.4](#)). Mais il permet aussi de découvrir de potentielles nouvelles particules au-delà du MS grâce à son énergie dans le centre de masse de l'échelle du TeV. Les processus issus de ses collisions sont appelés des événements. Une part de son temps de fonctionnement est aussi dédiée à des collisions d'ions lourds comme des noyaux de plomb.

Une vue schématique du complexe des accélérateurs du CERN avec ses principales expériences est en [Fig. 2.1](#). Les protons sont obtenus en arrachant les électrons d'un gaz d'hydrogène, injecté dans un cylindre métallique appelé duoplasmatron, grâce à un champ électrique. Les protons sont ensuite envoyés dans une cavité appelée Quadrupole Radio Fréquence (RFQ) qui regroupe les protons en paquets, les accélère et les concentre. Une fois accélérés à 750 KeV, les protons entrent dans l'accélérateur linéaire appelé Linac 2 pour atteindre 50 MeV. Ensuite, ces derniers arrivent dans le booster du synchrotron à protons (PSB) de 150 m de circonférence pour atteindre 1.4 GeV. Puis ils atterrissent dans le PS pour enfin être dans le SPS avec une énergie de 26 GeV et 450 GeV respectivement. Finalement, les paquets de protons entrent dans le LHC pour être de nouveau accélérés par huit cavités radiofréquence (RF) par faisceau. Jusqu'à 2556

paquets de protons par faisceau sont présents avec environ 115 milliards de protons par paquet. Ces paquets sont espacés de 25 ns (50 ns pendant le Run 1) et font 11000 tours par seconde. Les paquets entrent en collision à une fréquence de 40 MHz quand la capacité des anneaux est au maximum. La vitesse atteinte par les protons n'est que de 3.1 m/s de moins que la vitesse de la lumière. Le LHC possède 9593 aimants afin de courber et concentrer les faisceaux. Ces derniers sont dans un vide ultra haut de  $10^{-10} - 10^{-11}$  mbar afin d'éviter toute collision avec du gaz.

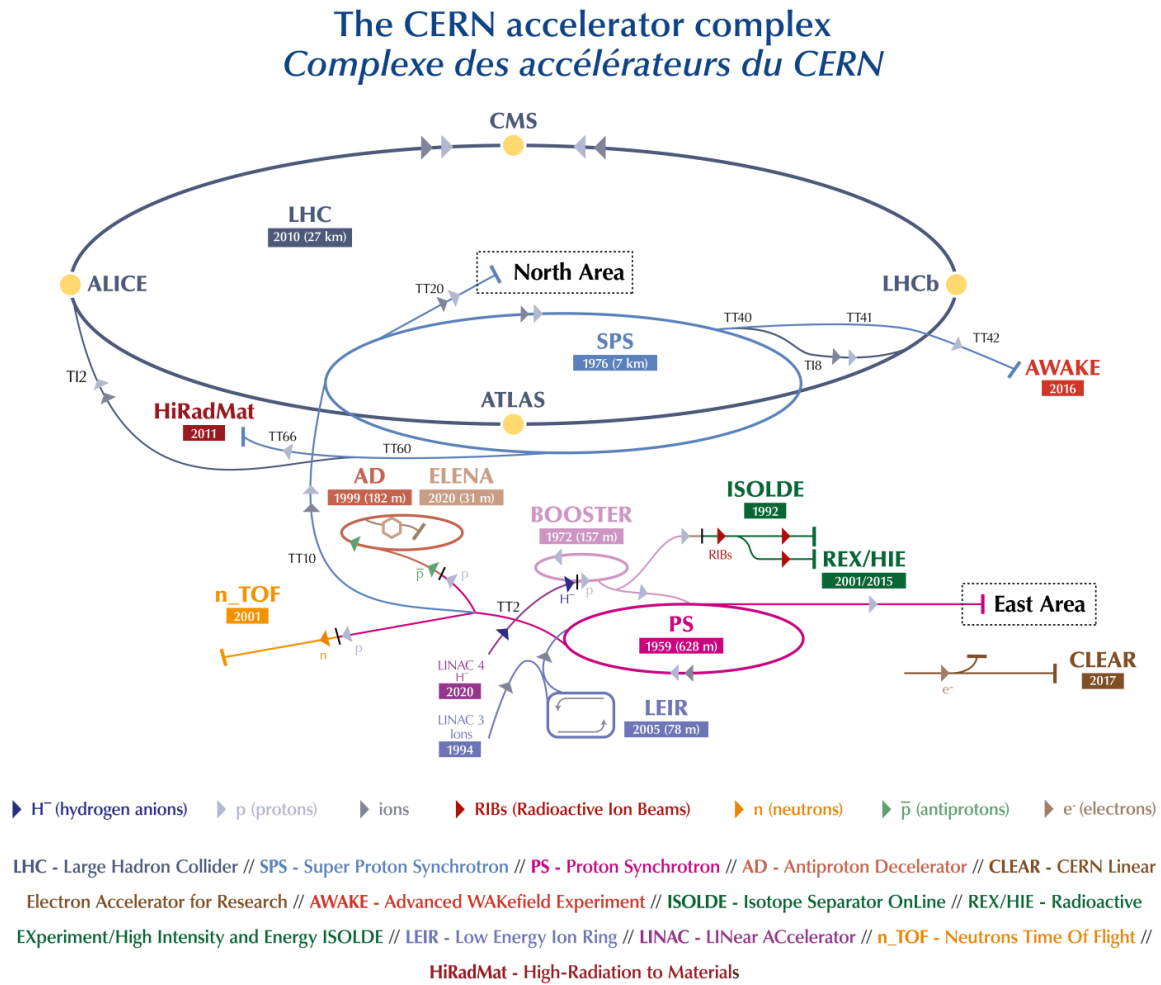


FIGURE 2.1 – Complexe des accélérateurs du CERN [52].

Les périodes de prise de données sont divisées en phases, elles-mêmes divisées en runs. Seulement une semaine après son démarrage en septembre 2018, le LHC a fait face à un accident technique dû à une connexion électrique défectueuse, ce qui décala la première prise de donnée de 14 mois et on prit la décision d'abaisser l'énergie des faisceaux [53, 54]. La première période dite Run 1, s'étendant de 2009 à 2013, a vu les faisceaux de protons du LHC atteindre une énergie dans le centre de masse de 7 TeV (2010, 2011)

et 8 TeV (2012), et une luminosité instantanée allant jusqu'à  $7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ . Le LHC fit ensuite une pause de deux ans afin d'améliorer les détecteurs et augmenter l'énergie de collision. Il redémarrera en 2015 pour le Run 2 avec une énergie dans le centre de masse de 13 TeV jusqu'à fin 2018. Initialement, le LHC était prévu pour délivrer une luminosité instantanée de  $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Mais elle atteignit  $2.06 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  durant le Run 2. L'accélérateur est de nouveau en pause jusqu'au Run 3 qui devrait débuter en 2022 et où l'énergie de collision devrait atteindre 14 TeV et la luminosité instantanée  $3.0 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Cette période permettra de doubler la quantité de données collectées jusqu'à présent. Ces trois Runs représentent ce que l'on appelle la Phase 1.

Quatre grandes expériences se situent sur les quatre points de l'anneau du LHC où les paquets de protons se rencontrent. ATLAS (« A Toroidal LHC ApparatuS ») [55] et CMS (« Compact Muon Solenoid ») [56] sont des détecteurs généralistes capables de servir à des mesures de précision comme à la découverte de nouvelles particules. Leurs sous-détecteurs utilisant des technologies différentes, ils sont complémentaires et permettent de corroborer les mesures et découvertes. En effet, le boson de Higgs fut indépendamment découvert par les deux expériences et la combinaison des résultats permit d'améliorer la confiance en cette observation.

La troisième principale expérience est « A Large Ion Collider Experiment » (ALICE) [57] qui étudie les collisions d'ions lourds. En effet, ce détecteur analyse le plasma de quarks et de gluons [58] d'une extrême densité et correspondant aux premiers instants de l'Univers. Cette phase de la matière ne voit plus les quarks et les gluons se lier en hadrons. Étudier ce plasma permet d'en savoir plus sur l'interaction forte.

Enfin, la dernière grande expérience est le « Large Hadron Collider beauty » (LHCb) [59] qui est spécialisée dans la physique du quark b. Elle sert à mesurer les paramètres de la matrice CKM, rechercher la violation CP, et étudier l'universalité leptonique.

Une quantité importante en physique des particules est la luminosité qui est liée au nombre d'évènements produits  $\frac{\partial N}{\partial t}$  :

$$\frac{\partial N}{\partial t} = \mathcal{L} \times \sigma, \quad (2.1)$$

avec  $\sigma$  la section efficace du processus et  $\mathcal{L}$  la luminosité instantanée définie comme :

$$\mathcal{L} = \frac{N_p^2 n_b f}{4\pi\sigma_x\sigma_y} F, \quad (2.2)$$

avec  $N_p$  le nombre de particules par paquet (environ  $10^{11}$ ),  $n_b$  le nombre de paquets par faisceau,  $f$  la fréquence de croisement des paquets (40 MHz),  $\sigma_{x,y}$  les tailles transverses de faisceau selon les axes x et y, et F le facteur de perte de luminosité dû à l'angle

de croisement au point d'interaction.

Par intégration de la luminosité instantanée dans le temps, on obtient la luminosité intégrée  $L$ . La Fig. 2.2 montre l'évolution cumulative de la luminosité intégrée pour chaque année de la Phase 1 délivrée à l'expérience CMS (voir section 2.3).

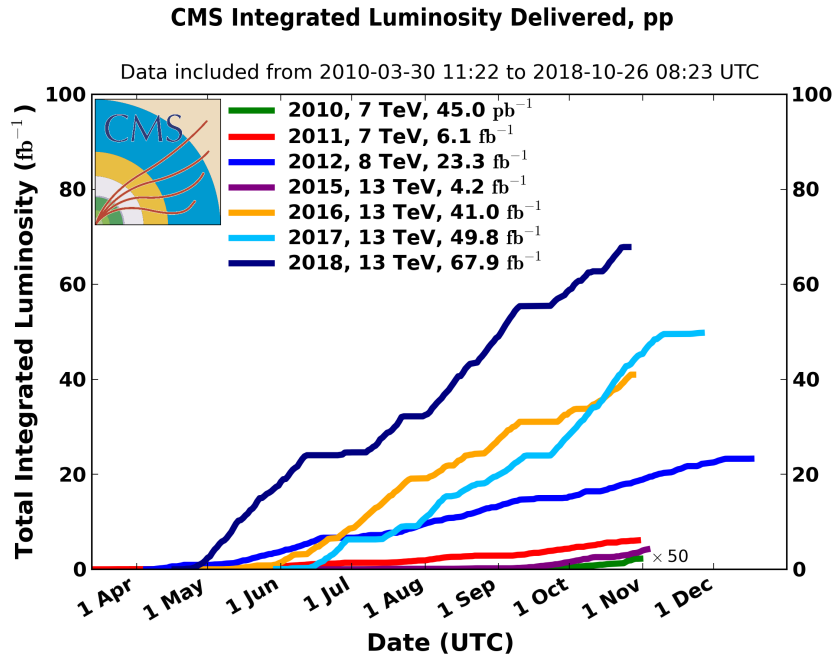


FIGURE 2.2 – Évolution cumulée de la luminosité intégrée délivrée à CMS au cours du temps pendant des faisceaux stables de collisions proton-proton pour chaque année de prise de données [60].

L'augmentation de la luminosité instantanée, outre le fait qu'elle permet la collecte d'une plus grande quantité de données indispensable à l'analyse de processus rares, a pour conséquence l'augmentation de l'empilement (PU). En effet, à chaque croisement de paquets, il peut y avoir de multiples interactions proton-proton, ce qui augmente l'occupation des sous-détecteurs. Or il y a très peu de chance qu'il y ait au moins deux interactions intéressantes pour un événement donné. Ces interactions additionnelles n'étant pas celles d'intérêt sont appelées interactions d'empilement. Le nombre moyen d'interactions par croisement de paquets pour chaque année de la Phase 1 est montré en Fig. 2.3. Cet empilement affecte directement la précision des mesures.

Ce genre d'empilement est appelé empilement « in-time », mais il existe un empilement « out-of-time » qui correspond aux interactions des croisements d'avant ou d'après. On distingue les interactions d'empilement « in-time » grâce à la position de chaque vertex d'interaction le long de l'axe du faisceau. Ces derniers sont moins énergétiques que le vertex dit primaire qui nous intéresse. Le temps de réponse des détecteurs est en général de plus de 25 ns, l'empilement « out-of-time » peut biaiser les mesures.



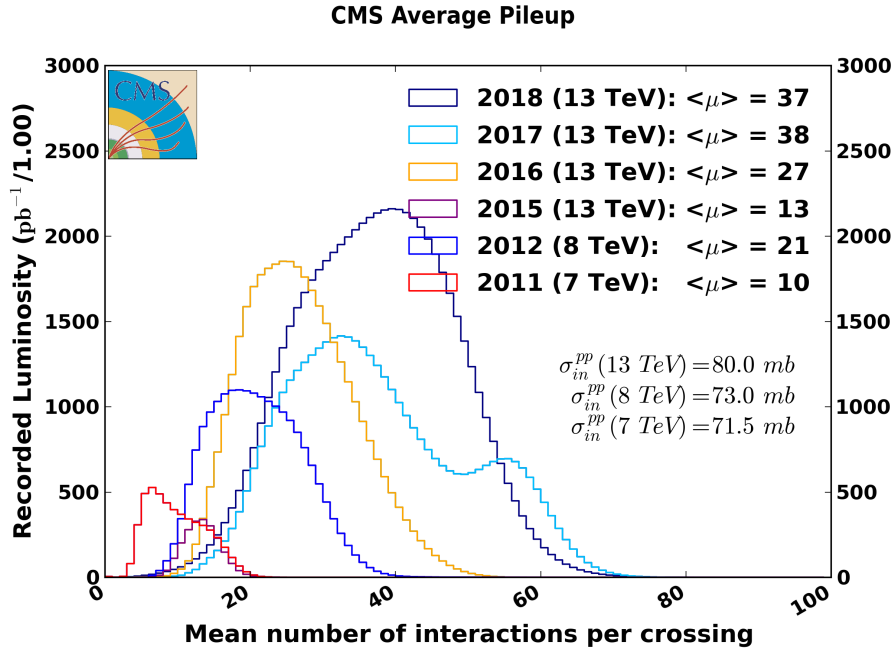


FIGURE 2.3 – Nombre moyen d’interactions par croisement de paquets pour chaque année de la Phase 1 [60].

Un détecteur de grande granularité et de bonne résolution temporelle situé à proximité du point d’interaction est donc indispensable.

Le détecteur de l’expérience CMS est décrit en détails ci-après.

## 2.3 L’expérience CMS

Le détecteur CMS, de forme cylindrique, mesure 22 m de long, 15 m de diamètre et pèse 14000 tonnes. Ses principaux composants sont représentés en Fig. 2.4.

### 2.3.1 Système de coordonnées

Le système de coordonnées (voir Fig. 2.5) est défini à partir du centre du détecteur. L’axe x pointe au centre du LHC, l’axe y en haut du détecteur et l’axe z le long du faisceau. L’angle azimutal  $\Phi$  est mesuré à partir de l’axe x dans le plan x-y et l’impulsion transverse est définie comme :

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (2.3)$$

Les partons dans les faisceaux de protons ont une impulsion négligeable dans le plan x-y et, par conservation de l’impulsion, la somme des impulsions de toutes les particules originaires de la même collision doit être nulle. Une impulsion transverse non nulle

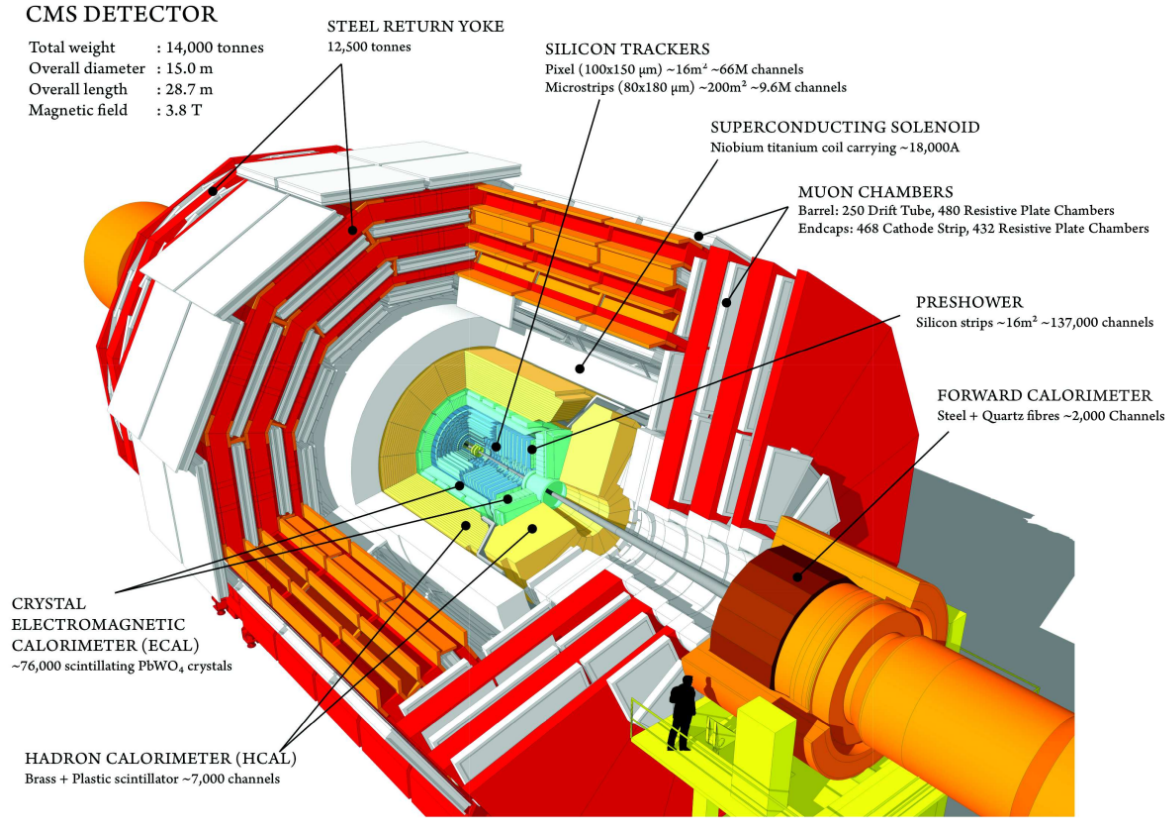


FIGURE 2.4 – Vue sectionnelle du détecteur CMS [61].

indique donc des particules non mesurées, comme des neutrinos ou des particules non découvertes. L'angle polaire  $\theta$  est quant à lui mesuré à partir de l'axe z. Cet angle permet de calculer une quantité physique très utile qu'est la pseudorapidité dont la différence entre deux particules est invariant de Lorentz :

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \quad (2.4)$$

Une pseudorapidité tendant vers  $\pm\infty$  correspond donc à une trajectoire parallèle à l'axe du faisceau. Une autre quantité importante est la séparation spatiale définie comme :

$$\Delta R = \sqrt{\Delta\Phi^2 + \Delta\eta^2}. \quad (2.5)$$

### 2.3.2 Le trajectographe en silicium

Le trajectographe en silicium fait 5.6 m de long et 2.2 m de diamètre, et les premières couches sont composées de pixels tandis que les plus extérieures sont faites de pistes. Il se situe au plus proche du point d'interaction et permet de reconstruire les traces des particules chargées, courbées par le fort champ magnétique, grâce à leur énergie déposée

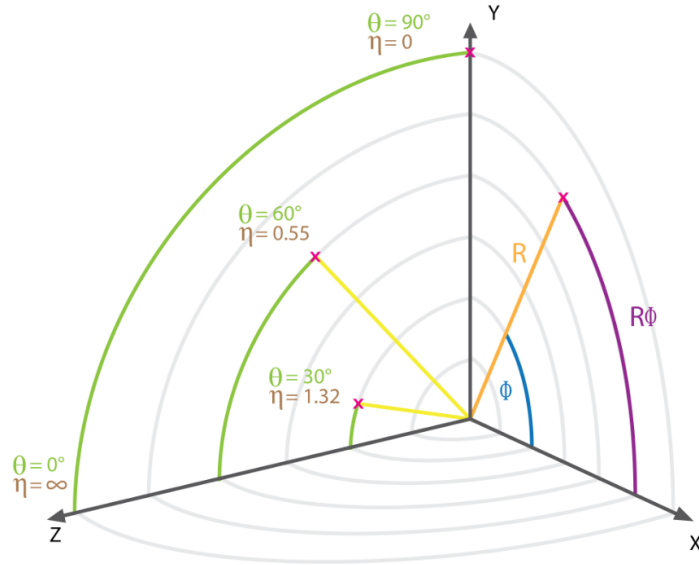


FIGURE 2.5 – Système de coordonnées de CMS [62].

dans les capteurs en silicium. La courbure permet de mesurer à la fois la charge électrique et l'impulsion de la particule chargée. Le fonctionnement physique des capteurs en silicium est expliqué en [chapitre 4](#). Un compromis a dû être trouvé entre la vitesse de lecture de l'électronique, la résolution spatiale, la quantité de matériel, le pouvoir de dissipation de la chaleur et la résistance aux radiations.

Durant l'arrêt hivernal en 2017, le trajectographe à pixels en silicium fut remplacé à cause des hautes radiations endommageant les capteurs. Ce nouveau sous-détecteur possède quatre couches dans le tonneau couvrant  $3 \leq |r| \leq 16$  cm et trois disques dans les bouchons couvrant  $29.1 \leq |z| \leq 51.6$  cm. La grande granularité du trajectographe à pixels permet de déterminer la position du point d'interaction (vertex primaire) avec une excellente résolution spatiale, indispensable afin de distinguer les interactions d'intérêt des interactions d'empilement. De plus, les hadrons beaux et charmés et les leptons taus ont un assez grand temps de vie pour voyager, on peut donc reconstruire un vertex secondaire où ils se désintègrent.

Le trajectographe à pistes en silicium possède quatre partitions, chacune étant composée de différentes couches de modules. Ces derniers sont soit mono, c'est-à-dire un seul côté de capteur, soit stéréo, c'est-à-dire deux côtés de capteurs. Les modules stéréo possèdent deux modules à pistes dos-à-dos et sont tournés de 100 mrad afin d'obtenir une mesure en 3D. Le tonneau intérieur (TIB) est composé de deux couches de modules stéréo puis de deux couches de modules mono. Le tonneau extérieur (TOB) entoure le TIB et sa partie la plus intérieure possède deux couches de modules stéréo et quatre mono. De chaque côté du tonneau se situent les disques intérieur (TID) et les bouchons

(TEC). Le TID possède trois roues (perpendiculaires à l'axe  $z$ ) ainsi que trois anneaux de modules (parallèles à l'axe  $z$ ). Le TEC quant à lui possède neuf roues avec quatre à sept anneaux de modules de chaque côté. Chaque roue est un mélange de modules mono et stéréo, mais chaque anneau est soit l'un soit l'autre. Les bouchons permettent d'étendre l'acceptance géométrique du trajectographe jusqu'à  $|\eta| < 2.5$ .

Une vue schématique est en Fig. 2.6. Les lignes simples représentent les modules mono, tandis que les doubles lignes représentent les modules stéréo.

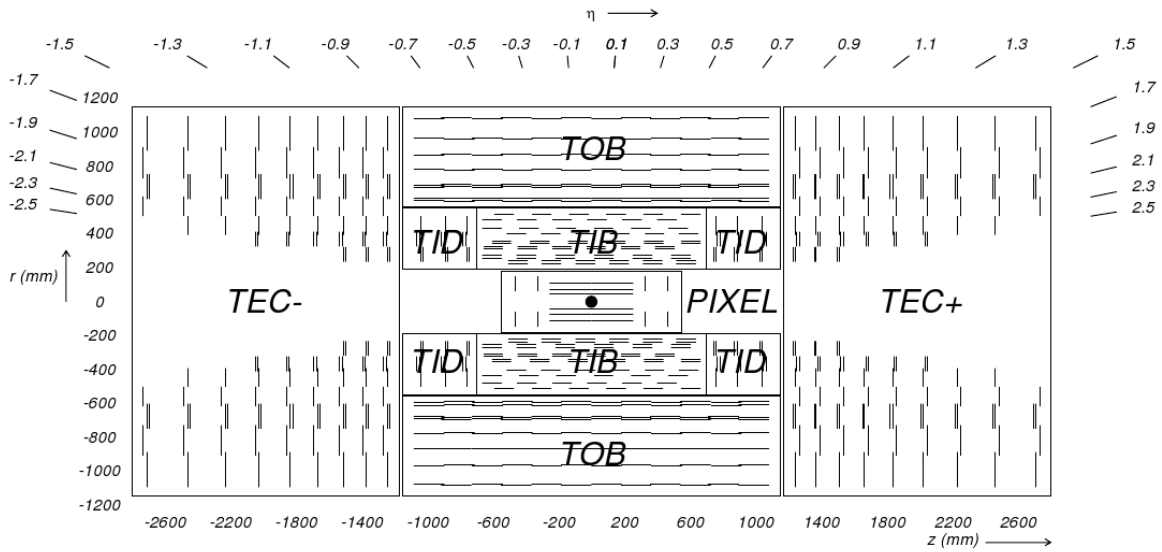


FIGURE 2.6 – Vue schématique du trajectographe de CMS où les positions en  $r$ ,  $z$  et  $\eta$  des différentes parties sont indiquées. Les lignes indiquent les modules de détection [63].

### 2.3.3 Le calorimètre électromagnétique

Le calorimètre électromagnétique (ECAL) est situé après le trajectographe et a pour but de mesurer précisément l'énergie des électrons et photons. En effet, ces derniers créent des gerbes électromagnétiques à cause des photons produisant des paires électron-positron et à cause des électrons et positrons rayonnant des photons. L'ECAL est composé de 75848 cristaux de plomb-tungstate ( $PbWO_4$ ). Ce matériau a comme propriétés intéressantes une certaine résistance aux radiations, une émission de lumière rapide (80% de la lumière est émise dans les 25 ns) et une bonne sensibilité. De plus, il possède une faible distance de radiation (0.89 cm) et un rayon de Molière (2.2 cm). Ces deux constantes, caractéristiques propres à un matériau, représentent les distances caractérisant la perte d'énergie d'une particule interagissant électromagnétiquement dans la direction longitudinale et transverse respectivement. La dimension des cristaux a été choisie afin de contenir entièrement les gerbes électromagnétiques.

Un schéma du ECAL est disponible en Fig. 2.7. Il possède un tonneau (EB) contenant deux moitiés (EB<sup>+</sup> et EB<sup>-</sup>) possédant chacune 18 supermodules et couvrant  $|\eta| < 1.479$ , ainsi que deux bouchons (EE) allant jusqu'à  $|\eta| = 3$ . Le sous-détecteur de « preshower » (ES) se situe devant les bouchons, améliore la résolution spatiale en  $1.65 < |\eta| < 2.6$  et possède des pistes en silicium de 2 mm de long qui aident à distinguer des photons seuls de hautes énergies des paires de photons rapprochés de basses énergies provenant de pions neutres.

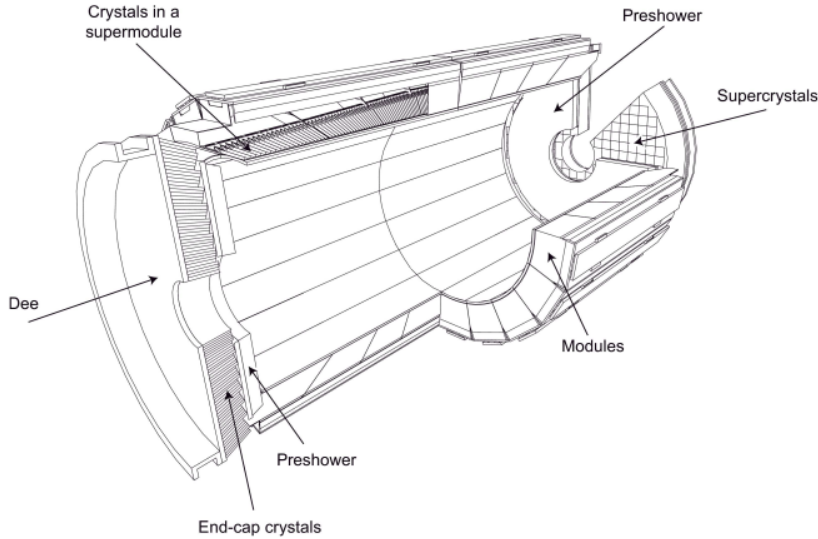


FIGURE 2.7 – Schéma du calorimètre électromagnétique de CMS [56].

Une caractéristique très importante du ECAL est sa résolution en énergie  $\sigma_E$  qui a été mesurée dans des données à 7 TeV [64] :

$$\frac{\sigma_E}{E} = \frac{0.027}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.005, \quad (2.6)$$

avec  $E$  l'énergie. Le premier terme est le terme stochastique, décrivant les fluctuations statistiques intrinsèques à la cascade électromagnétique, le second prend en compte le bruit et le troisième qui est constant couvre principalement la non uniformité, la fuite d'énergie et les mauvaises calibrations.

### 2.3.4 Le calorimètre hadronique

Le calorimètre hadronique (HCAL) mesure l'énergie des hadrons et des gerbes de particules. Sa taille est limitée par sa position entre le ECAL et l'aimant. Par conséquent, Le HCAL est constitué de quatre parties distinctes (voir Fig. 2.8). Tout d'abord le HCAL possède un tonneau (HB,  $|\eta| < 1.4$ ) et un bouchon (HE,  $1.4 < |\eta| < 3$ ) constitués d'absorbeurs en laiton (couches dites passives) alternés avec des scintillateurs en plastique (couches dites actives). En effet, les particules interagissent dans les couches passives et produisent des particules secondaires en cascade au fur et à mesure de leur

traversée. Ces différentes particules atteignent ensuite les couches actives en émettant une lumière bleue qui sera absorbée par des fibres plus petites qu'un millimètre. Cette lumière est ensuite décalée dans le vert et transportée par des câbles optiques vers des boîtes de lectures.

Cependant, les tonneaux du ECAL et du HCAL ne sont pas suffisants pour des particules fortement énergétiques pour  $|\eta| < 1.3$ . Par conséquent, une partie scintillateur (HO) est placée après l'aimant où ce dernier fait office d'absorbeur. Finalement, un sous-détecteur placé à l'avant (HF) est positionné après les chambres à muons à  $|z| = 11.2$  m et couvre la région  $2.9 < |\eta| < 5.2$ . Le HF étant très exposé aux radiations avec environ un tiers des particules produites lors de la collision p-p l'atteignant, il est constitué de fer comme absorbeur et de fibres de quartz comme volume actif.

La résolution en énergie hadronique combiné des tonneaux du ECAL et du HCAL a été mesurée en faisceau test avec des pions de 5 à 300 GeV [65] :

$$\frac{\sigma_E}{E} = \frac{0.847}{\sqrt{E}} \oplus 0.074. \quad (2.7)$$

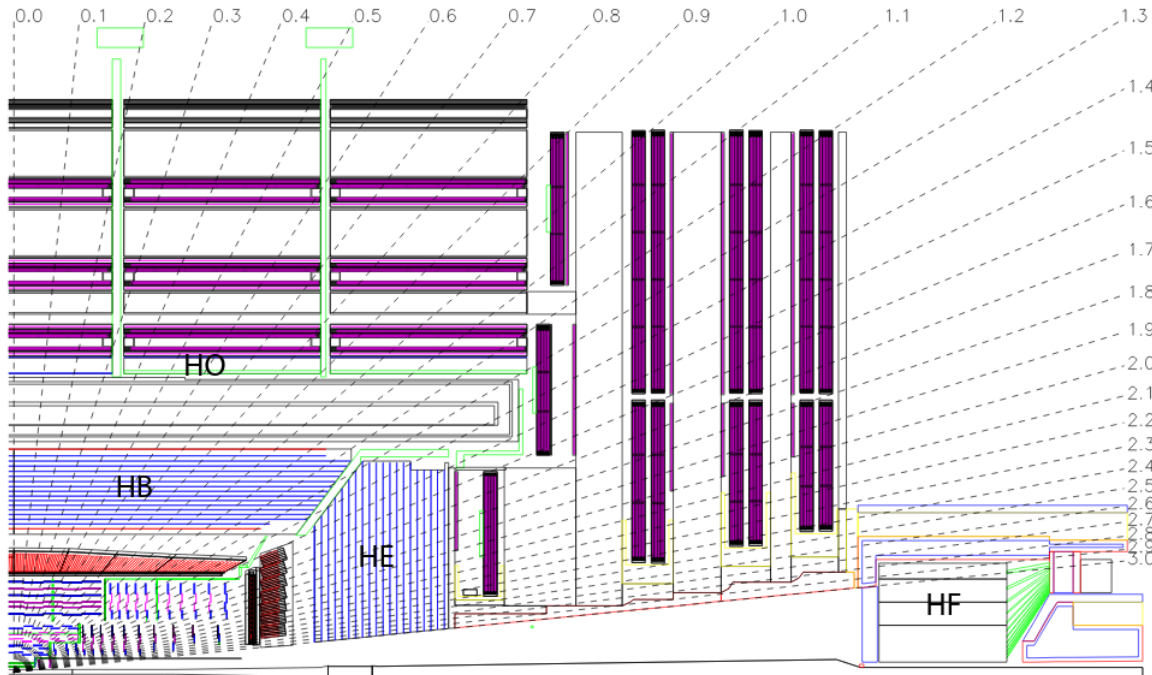


FIGURE 2.8 – Schéma du calorimètre hadronique de CMS [56].

### 2.3.5 L'aimant

L'expérience CMS est plongée dans un fort champ magnétique uniforme de 3.8 T grâce à un solénoïde supraconducteur. Il englobe le trajectographe et les deux calorimètres et permet de dévier les trajectoires des particules chargées afin de connaître le signe de leur charge électrique ainsi que leur impulsion. Cet aimant, le plus grand jamais construit au monde (6 m de diamètre, 12.5 m de long), était initialement prévu pour fournir un champ de 4 T, mais il fut abaissé à 3.8 T en raison de son risque de vieillissement prématuré [66, 67]. Afin d'exploiter ses propriétés supraconductrices, l'aimant est refroidi à 4.1 K avec de l'hélium liquide et placé dans une cuve à vide afin de l'isoler thermiquement. L'aimant est aussi placé et entouré d'un attelage cylindrique en fer. Cet attelage permet de canaliser le champ magnétique de retour à l'extérieur du solénoïde. En effet, le champ magnétique dans les chambres à muons, situées à l'extérieur du solénoïde, est de 2 T et est orienté dans le sens opposé à celui délivré dans la partie intérieure. L'attelage, large et dense, permet de filtrer les particules car seuls les muons et neutrinos peuvent le traverser.

### 2.3.6 Les chambres à muons

Le système à muons se situe à l'extérieur de l'aimant supraconducteur et a pour but d'identifier l'énergie d'ionisation des muons et d'améliorer la résolution sur la mesure de l'impulsion de ces derniers grâce à la combinaison de ses informations avec celles du trajectographe. De plus, certains muons comme les cosmiques ou issus de vertex très déplacés ne laissent pas de trace dans le trajectographe interne. Le système à muons étant situé loin du point d'interaction, jusqu'à 11 m, il doit couvrir une large surface. C'est pour quoi son spectromètre contient des détecteurs gazeux de trois types : les tubes à dérive (DT), les chambres à fils cathodiques (CSC) et les chambres à plaques résistives (RPC) (voir Fig. 2.9). Ces détecteurs gazeux contiennent 1400 chambres placées en couches successives afin de reconstruire la trajectoire des muons.

Les DT dans le tonneau couvrent la région  $|\eta| < 1.2$  et sont constitués de cinq sections transverses divisées en 12 secteurs de  $30^\circ$ . De plus, chaque secteur contient des chambres à dérives réparties en 4 stations, elles-mêmes possédant 3 couches de 4 tubes d'épaisseur. Les tubes de deux de ces couches sont orientés parallèlement à l'axe du faisceau afin de mesurer la coordonnée  $r-\phi$ . Dans la troisième couche, les tubes sont orientés perpendiculairement à l'axe du faisceau pour mesurer la coordonnée  $r-z$ . Lorsqu'un muon traverse un tube, le gaz (85% Ar et 15%  $\text{CO}_2$ ) est ionisé et des électrons vont migrer vers l'anode. Le temps mis par le signal pour traverser les tubes et la différence entre le nombre de charges collectées aux deux extrémités de l'anode donnent la position de la particule.

Les CSC sont dans les bouchons et couvrent la région  $0.9 < |\eta| < 2.4$ . Ils sont fait d'anodes sous forme de fils perpendiculaires au faisceau croisant des bandes de cathodes afin d'obtenir une position en 2D. Ces sous-détecteurs ont une réponse rapide, une fine segmentation et une résistance aux radiations (indispensable car proches de l'axe du faisceau).

Enfin, les RPC sont présents dans le tonneau et les bouchons afin de mieux identifier les paquets de protons d'où viennent les muons. Ils sont composés de plaques de bakélite parallèles de charges opposées séparées par du gaz. Les charges produites par la traversée d'un muon sont collectées par des pistes. La résolution spatiale des RPC est moins bonne que celle des DT et CSC mais les RPC possèdent un temps de réponse bien plus court et inférieur à 25 ns.

Lors de la pause hivernale de 2017, un nouveau sous-détecteur fut installé dans la région  $1.6 < |\eta| < 2.2$  afin d'améliorer le système à muons dans les bouchons et utilise la technologie « gas electron multiplier » (station GEM) [68, 69].

Ces sous-détecteurs sont utilisés dans le système de déclenchement pour les évènements contenant un candidat muon.

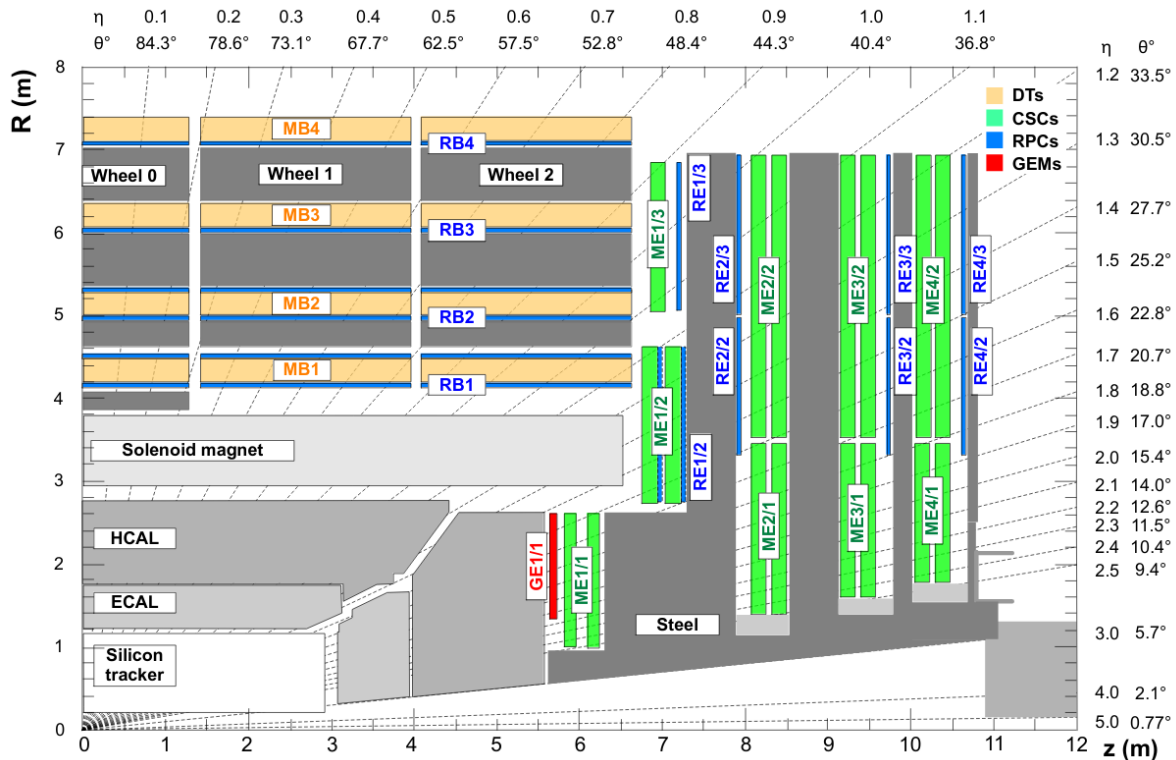


FIGURE 2.9 – Quadrant de la vue R-z du système à muons [70].



### 2.3.7 Le système de déclenchement

Les paquets de protons se croisent toutes les 25 ns et nécessitent un système de déclenchement [71] pour les millions d'événements produits chaque seconde. Or seule une infime fraction des événements produits sont intéressants pour une analyse de physique comme le montre la Fig. 2.10 qui représente la section efficace de production de différents processus en fonction de l'énergie dans le centre de masse par rapport à la section efficace totale. Un système de filtre des événements est donc nécessaire et un compromis a dû être trouvé entre la réponse en temps de l'électronique et les ressources informatiques disponibles.

Le premier niveau dit de bas niveau (L1) se base sur des informations hardware uniquement qui proviennent des calorimètres et des chambres à muons (l'algorithme de reconstruction des traces dans le trajectographe prendrait trop de temps). Une première identification des particules est alors faite. En effet, les électrons et photons sont identifiés par leurs dépôts d'énergie dans le ECAL, les muons par l'ensemble des informations du spectromètre à muons, les taus par les produits de désintégrations dans les calorimètres, les jets, cônes de hadrons et d'autres particules produites par l'hadronisation de quarks et de gluons, par les deux calorimètres et l'énergie transverse manquante en sommant l'énergie de tout les dépôts dans les calorimètres. L'électronique prend une décision en  $3.4 \mu s$  après la reconstruction de l'événement à partir d'un ensemble de conditions appelé menu L1. Le contenu de ce dernier a été étoffé lors de la première grande pause du LHC (LS1) en augmentant le nombre d'algorithmes disponibles de 128 à 256 et en les optimisant afin de mieux fonctionner dans un environnement à haute luminosité et empilement. Le L1 réduit la fréquence des événements de 40 MHz à 100 kHz.

Ensuite, le second niveau dit de haut niveau (HLT), ne traitant que les événements passant le L1, utilise les informations de tous les sous-détecteurs. Le niveau HLT permet de reconstruire plus précisément les objets grâce à une fréquence initialement réduite par le L1 qui permet aux algorithmes du flux de particules et de traçage itératif de fonctionner. Les dépôts d'énergie peuvent enfin être associés aux traces reconstruites permettant de connaître les charges électriques. De plus, le trajectographe permet d'identifier les gerbes provenant de quarks b car ils ont un vertex déplacé. Une liste de conditions est contenue dans le menu HLT afin de réduire la fréquence à 1 kHz, fréquence maximale à laquelle les données peuvent être traitées et enregistrées, grâce à l'exploitation de la puissance de calcul de la ferme informatique mise à disposition. Les événements passant le système de déclenchement sont ensuite enregistrés localement sur disques avant d'être transférés dans un centre informatique afin d'être enregistrés plus durablement et distribués.

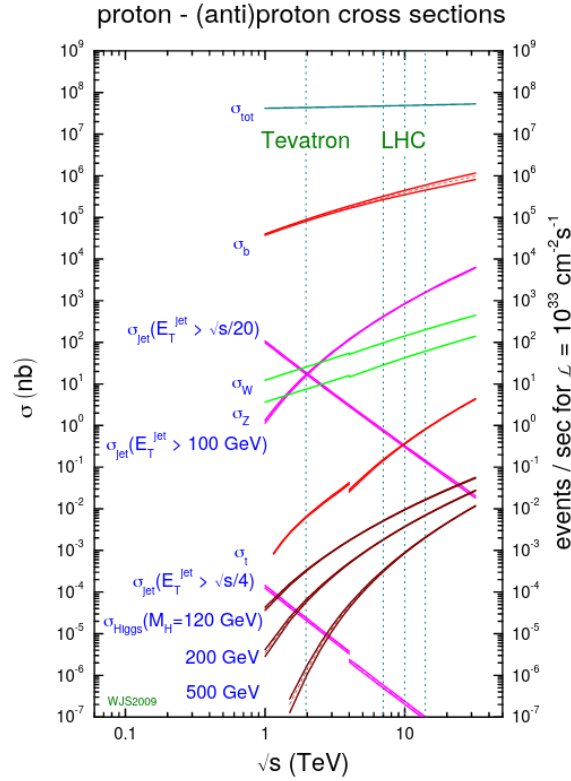


FIGURE 2.10 – Sections efficaces de production de différents signaux et bruits de fond dans des collisions proton-(anti)proton. La discontinuité des courbes est due au fait que le Tevatron utilisait un antiproton et un proton tandis que le LHC utilise deux protons. [72].

## 2.4 Le LHC à haute luminosité

Le LHC sera grandement amélioré à partir de 2025 afin de fonctionner à une plus haute luminosité instantanée durant la Phase 2, au moins à  $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , à partir de 2027 et sera renommé HL-LHC (voir Fig. 2.11). La luminosité pourra monter jusqu'à  $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , presque quatre fois la luminosité du Run 2, en augmentant l'intensité des faisceaux et en les compressant plus fortement aux points d'interaction. Ainsi, il y aura plus de collisions par croisement de paquets de protons, ce qui augmentera également le nombre moyen d'interactions d'empilements jusqu'à potentiellement 200 au lieu d'environ 60 lors du Run 2. De plus, l'énergie de collision devrait passer de 13 TeV à 14 TeV lors du Run 3. La quantité de données ainsi collectée devrait atteindre au minimum  $3000 \text{ fb}^{-1}$ , ce qui correspond à 10 fois celle de la Phase 1. Ceci devrait pouvoir, entre autres, nous permettre d'observer les premiers signes de l'auto-couplage du boson de Higgs. Cette augmentation en luminosité nous oblige donc à rendre les sous-détecteurs au moins aussi performants qu'actuellement dans cet environnement à grand empilement et fortes radiations. L'expérience CMS verra donc ses sous-détecteurs mis à niveau pour la Phase 2 [73].

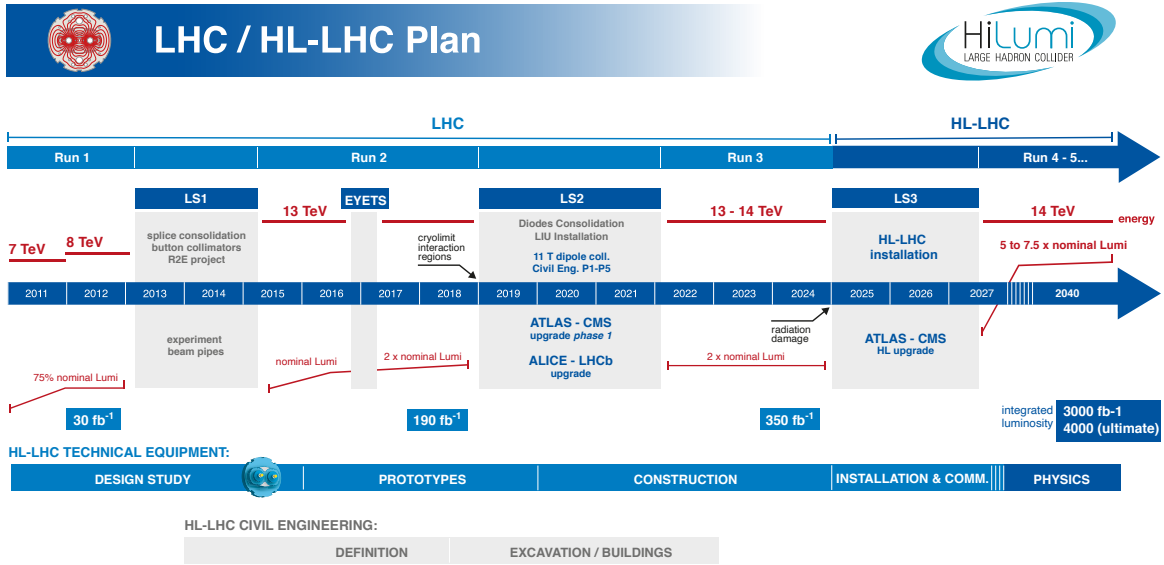


FIGURE 2.11 – Plan dans le temps du LHC et du HL-LHC. Les périodes de prises de données sont entrecoupées par des périodes de pause où l'accélérateur et les expériences sont améliorés. [74].

Le trajectographe sera entièrement remplacé par un nouveau détecteur plus léger (voir Fig. 2.12), plus granulaire et plus résistant aux radiations car l'actuel sera trop endommagé à la fin du Run 3 [75]. La partie intérieure sera faite de pixels et couvrira  $|\eta| < 4.0$  avec une granularité 6 fois plus importante comme montré en Fig. 2.13. La partie extérieure comportera des modules, dit  $p_T$ , de deux sortes et possédant chacun deux capteurs séparés de quelques mm : les modules 2S avec deux capteurs à micro-pistes comportant chacun deux colonnes de 1016 micro-pistes de taille  $5 \text{ cm} \times 90 \mu\text{m}$ , et les modules PS avec un capteur possédant deux colonnes de 960 micro-pistes de  $2.4 \text{ cm} \times 100 \mu\text{m}$  d'un côté et un capteur avec un macro-pixel de  $1.5 \text{ mm} \times 100 \mu\text{m}$  de l'autre côté. De plus, ces modules  $p_T$  contribueront au L1 du système de déclenchement à la fréquence de 40 MHz en rejetant les particules chargées ayant une basse impulsion transverse (voir Fig. 2.14).

L'électronique du tonneau des calorimètres sera remplacée car elle ne supportera pas la nouvelle latence et fréquence du nouveau L1 du système de déclenchement [77]. Les cartes dites très à l'avant (VFE) du tonneau du ECAL qui amplifient et mettent en forme le signal seront remplacées afin d'obtenir une meilleur résolution en temps et filtrage du bruit. Le L1 du système de déclenchement pourra utiliser l'information des cristaux seuls, au lieu d'un ensemble de  $5 \times 5$  cristaux pour la Phase 1, afin de faire correspondre précisément les gerbes électromagnétiques avec les traces. De plus, la température passera de  $18 \text{ }^\circ\text{C}$  à environ  $9 \text{ }^\circ\text{C}$ . Le remplacement des scintillateurs et fibres du tonneau

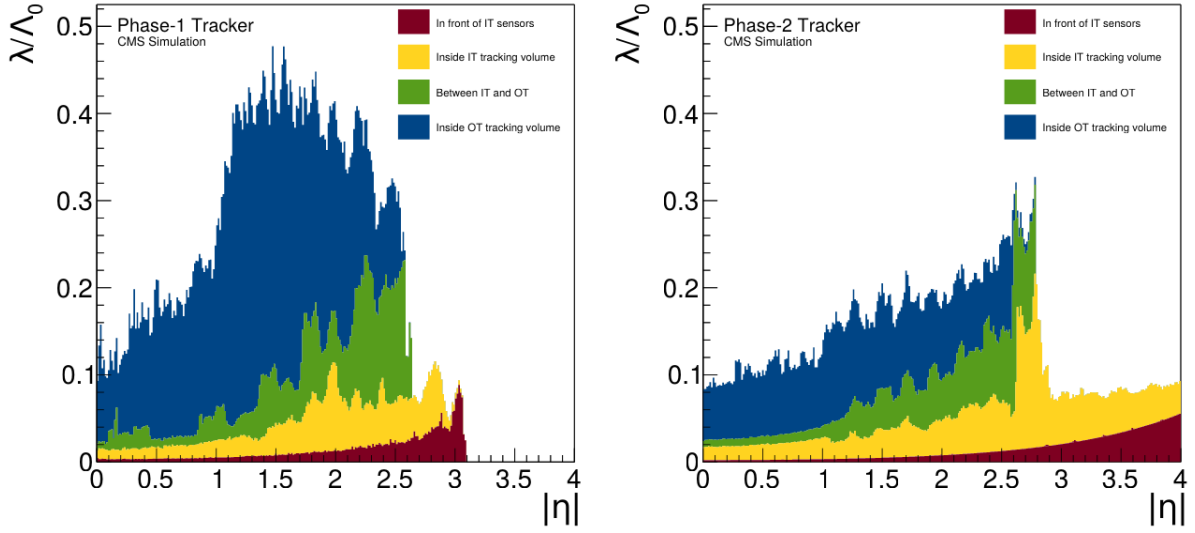


FIGURE 2.12 – Quantité de matière du trajectographe de CMS durant la Phase 1 (gauche) et la Phase 2 (droite) en unités de longueurs d'interaction nucléaire en fonction de la pseudorapidité pour les différents sous-détecteurs [75].

du HCAL ne sera pas nécessaire car les dommages de radiations sont négligeables durant la Phase 2. Cependant, les photodiodes hybrides (HPDs) seront remplacées par des photomultiplicateurs en silicium (SiPMs).

Les bouchons des calorimètres seront complètement remplacés car ils ne supporteront pas les très hauts niveaux de radiation [78]. À la place, un calorimètre à haute granularité (HGCAL) en silicium sera présent et couvrira  $1.5 < |\eta| < 3.0$  (voir Fig. 2.15). Il contiendra 28 couches d'absorbeurs en tungstène et en cuivre alternées avec des couches de capteurs en silicium dans la partie électromagnétique. La partie hadronique quant à elle contiendra 24 couches d'absorbeurs en fer inoxydable aussi alternées avec des

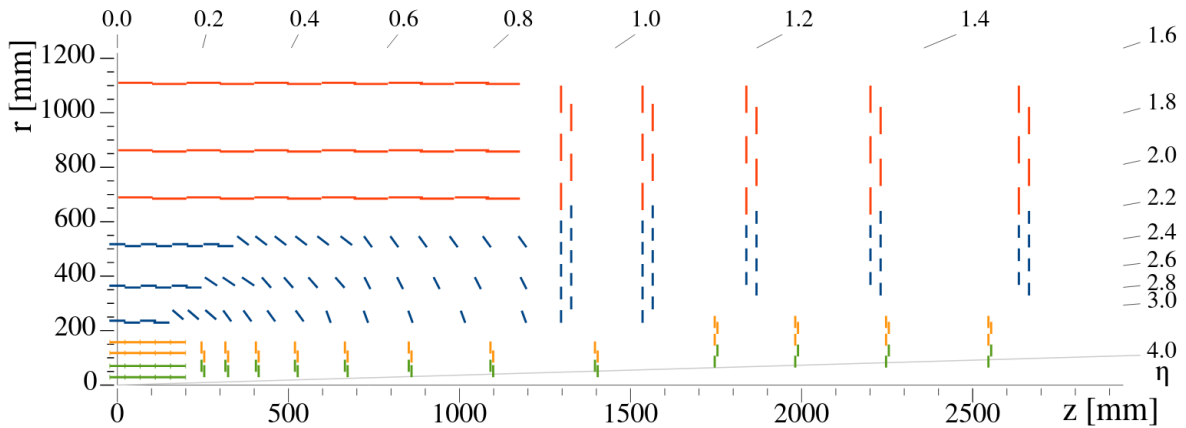


FIGURE 2.13 – Schéma du trajectographe de la Phase 2 dans le plan r-z [75].

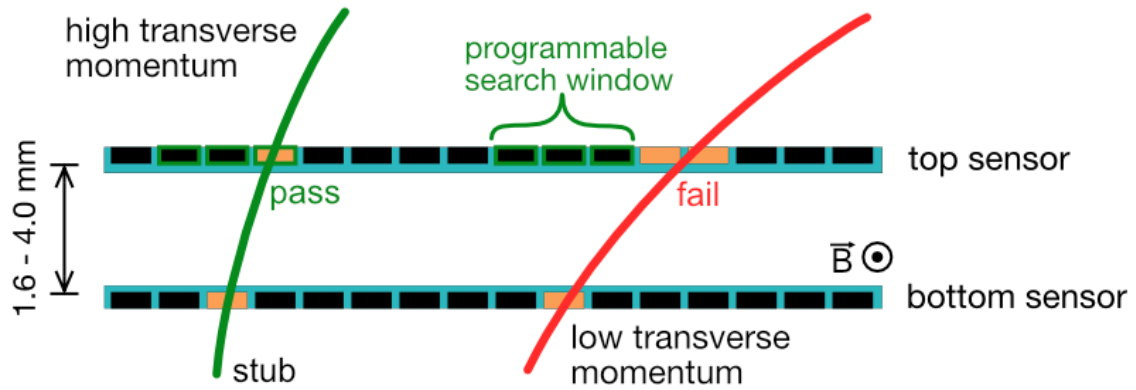


FIGURE 2.14 – Vue schématique d'un module  $p_T$  [76].

couches de capteurs en silicium et de scintillateurs en plastiques. La température sera de  $-30\text{ }^\circ\text{C}$  afin de baisser suffisamment le bruit électronique dû à l'augmentation du courant de fuite ainsi que de garantir l'efficacité de la collecte de charges après irradiation. Les capteurs en silicium fourniront une information en temps dont la résolution atteindra quelques dizaines de picosecondes par couche pour les gerbes électromagnétiques et plusieurs couches peuvent se combiner pour améliorer cette résolution. Le HGCALE permettra donc d'obtenir une imagerie des gerbes de particules en 5D (énergie, x, y, z et temps) ce qui améliorera grandement la réjection de l'empilement, la résolution en énergie et les résultats de l'algorithme du flux de particules.

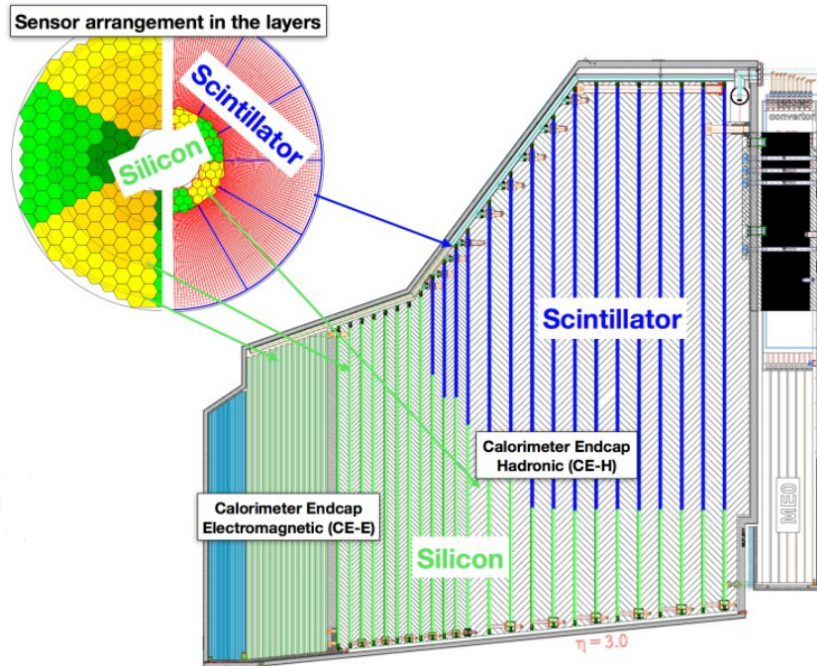


FIGURE 2.15 – Vue schématique du HGCALE.

L'électronique des DT, CSC et RPC du système à muon sera changée [79] afin de fonctionner avec le nouveau système de déclenchement. L'acceptance sera étendue jusqu'à  $|\eta| = 2.8$  afin de bénéficier de la nouvelle couverture spatiale du trajectographe et du HGCALE. De nouveaux RPC et stations GEM seront installés afin d'améliorer la reconstruction et le système de déclenchement pour  $1.6 < |\eta| < 2.4$ . De plus, un nouveau module de type GEM appelé ME0 utilisant 6 couches (au lieu de 2 pour les précédentes stations GEM) sera installé afin d'étendre l'acceptance jusqu'à  $|\eta| = 2.8$ .

Un nouveau sous-détecteur nommé « Minimum Ionizing Particles Timing Detector » (MIP timing detector ou MTD) [80] sera construit. C'est, par manque de place disponible, une fine couche entre le trajectographe et les calorimètres avec une partie tonneau pour  $|\eta| < 1.5$  et deux parties bouchons afin de couvrir jusqu'à  $|\eta| = 3.0$ . La partie tonneau sera constituée de cristaux et de photomultiplicateur en silicium (SiPMs) tandis que la partie bouchons aura des diodes à avalanche à faible gain (LGAD). Son but est de mesurer le temps de vol des particules chargées avec une résolution en temps de 30 à 40 ps. Une simulation 3D du sous-détecteur est en Fig. 2.16.

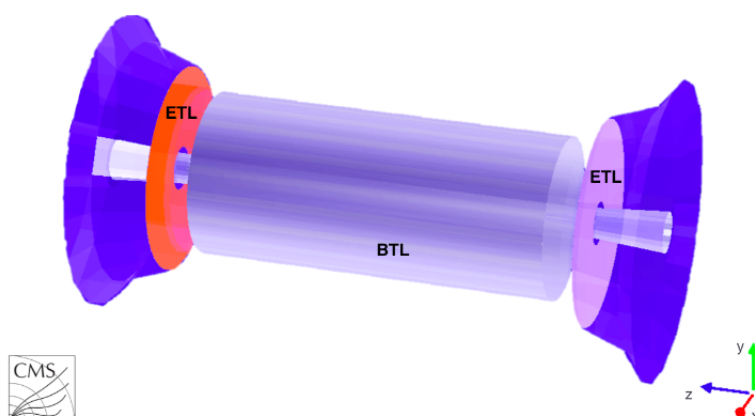


FIGURE 2.16 – Simulation 3D avec l'outil GEANT [81] du MTD [80]. La partie tonneau entre le trajectographe et le ECAL est en gris et les deux bouchons en face des bouchons des calorimètres sont en orange et violet clair.

Le L1 du système de déclenchement verra aussi son électronique changer [82]. En effet, la plus grande luminosité nous oblige à modifier le système de déclenchement. La fréquence de sortie du L1 passera de 100 kHz à 750 kHz et le niveau HLT sélectionnera des événements à une fréquence de 7.5 kHz. De plus, la latence, temps pendant lequel les informations sont gardées en mémoire avant d'être envoyées au HLT, passera de  $3.2 \mu\text{s}$  à  $12.5 \mu\text{s}$ . Les améliorations, présentées ci-dessus, des sous-détecteurs pour la Phase 2 permettront d'utiliser l'algorithme du flux de particules (voir section 3.1) dès le L1 du système de déclenchement. De plus, ceci permettra d'identifier la majorité des particules et d'appliquer la technique de l'empilement par particule (PUPPI) qui associe un poids aux particules reconstruites proportionnel à leur probabilité de provenir d'une interaction d'empilement (voir sous-section 3.2.6).

## Chapitre 3

# Reconstruction des particules et des objets

### 3.1 L'algorithme du flux de particules

Une particule donne, en général, plusieurs informations différentes dans les sous-détecteurs lors de sa traversée. Afin de reconstruire des événements, l'expérience CMS utilise l'algorithme du flux de particules (PF) [83] qui fait des liens entre les différentes signatures laissées par les particules que l'on peut voir en Fig 3.1.

Ceci est possible grâce aux spécificités uniques du détecteur CMS. En effet, le trajectographe étant très granulaire avec ses 125 millions de pixels et ses 9.3 millions de pistes, la mesure des impulsions des particules chargées en est facilitée. Mais ceci n'est possible que grâce au fort champ magnétique de 3.8 T du solénoïde qui courbe les trajectoires des particules chargées et les sépare des particules neutres. De plus l'aimant étant placé après les calorimètres, les interactions secondaires des particules avec la matière avant leur détection sont réduites. Le très segmenté calorimètre électromagnétique avec ses 75848 cristaux scintillants de tungstate de plomb permet de faire correspondre l'énergie déposée aux traces reconstruites. De plus, le HCAL, bien hermétique, complète les énergies déposées par celles des hadrons neutres et chargés. Pour terminer, les chambres à muons permettent une bonne reconstruction des traces des muons. Par exemple, un photon déposera son énergie dans le ECAL sans laisser de trace dans le trajectographe tandis qu'on observera la trace du muon, grâce au champ magnétique, qui traversera tout le détecteur pour atteindre les chambres à muons.

L'algorithme du flux de particules commence par utiliser des informations de bas niveau afin de reconstruire les électrons, muons, photons et hadrons. Puis dans un second temps des objets physiques plus complexes sont reconstruit comme les taus hadroniques ou les gerbes de particules. Pour faire cela, des éléments de base sont combinés : les traces et les amas d'énergie.

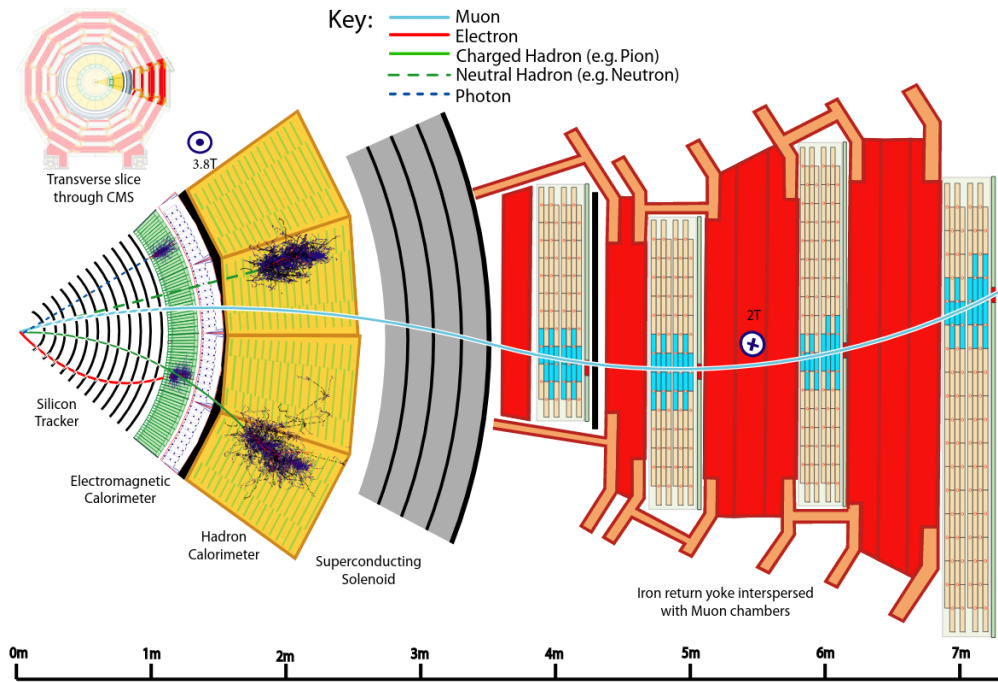


FIGURE 3.1 – Schéma d'une vue transverse de l'expérience CMS montrant les interactions des différentes particules avec les sous-détecteurs [83].

Les digits, qui sont les coups ADC après soustraction des piédestaux et de l'application du mode de suppression des zéros (voir chapitre 4 pour plus d'informations sur les capteurs du trajectographe), sont regroupés en impacts qui ont une position grossière et une erreur associée. Ensuite, la reconstruction des traces se fait en quatre étapes en utilisant un outil appelé « Combinatorial Track Finder » (CTF) [84] fondé sur des filtres de Kalman [85, 86] :

- pose de graines à partir de trois impacts reconstruits dans le trajectographe à pistes ou à pixels, ou à partir de deux impacts avec une contrainte sur la position du faisceau ;
- construction des traces qui a pour but de trouver tous les impacts venant d'une même particule. Pour cela, l'algorithme va démarrer par des graines et chercher les impacts de couche en couche en testant la compatibilité de l'impact avec la trace en cours de construction par un test  $\chi^2$  ;
- ajustement de la trajectoire qui va chercher à obtenir les meilleurs paramètres physiques de la trace en question et va ainsi recalculer la position des impacts avec une meilleure précision ;
- un filtre sur les traces est utilisé afin d'éliminer celles qui sont soit mal reconstruites soit fausses en utilisant un test  $\chi^2$ , le nombre de couches avec un impact associé à la trace et la probabilité de la trace d'avoir pour origine le vertex primaire.



On utilise une méthode de traçage itératif [87] qui applique ces étapes 12 fois, où les impacts associés aux traces reconstruites après chaque itération sont masqués afin de réduire les combinaisons et éviter de les compter plusieurs fois. Cette simplification combinatoire, grâce au masquage des impacts, permet de relâcher les coupures pour les prochaines itérations et de trouver des traces moins évidentes. Chaque itération recherche un type de trace précis. Les premières itérations commencent par chercher les traces les plus évidentes, comme celles semblant provenir d'une particule ayant la plus haute impulsion transverse ( $p_T$ ). On peut voir en Fig. 3.2 les efficacités et les taux de faux de l'algorithme CTF en fonction du  $p_T$  des traces de hadrons chargés dans des événements QCD sans empilement. La faible courbure des traces à haut  $p_T$  due au champ magnétique rend leurs propriétés plus difficile à mesurer. Les faibles performances à bas  $p_T$  sont quant à elles dues aux interactions nucléaires plus présentes à basse impulsion transverse ainsi qu'aux « loopers », particules ayant une trajectoire s'enroulant sur elle-même.

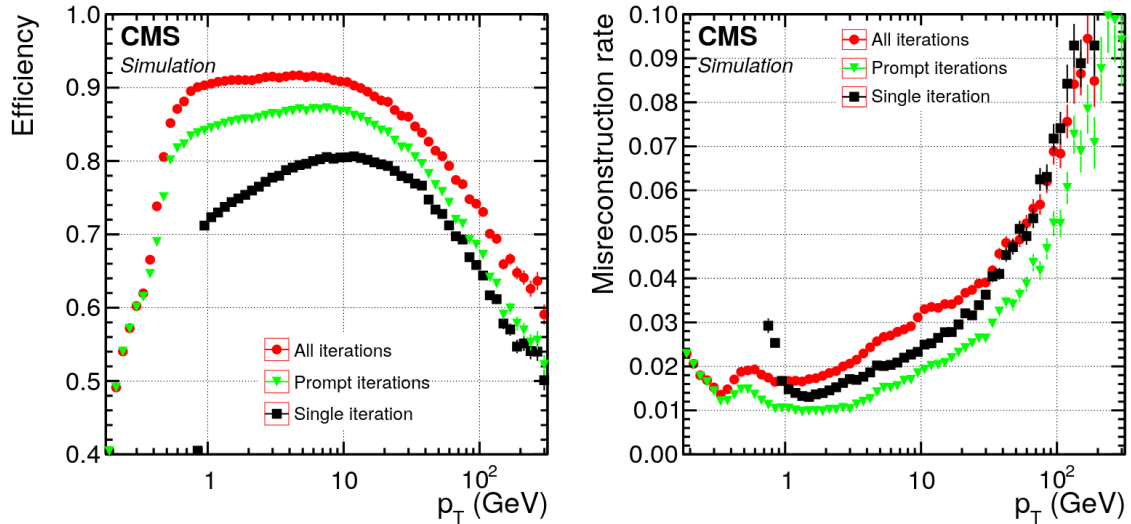


FIGURE 3.2 – Efficacité (gauche) et taux de faux (droite) de l'algorithme CTF (noir) en fonction du  $p_T$  des traces de hadrons chargés dans des événements purement QCD sans empilement. Le résultat est aussi montré pour la méthode de traçage itératif avec des itérations promptes basées sur des graines avec au moins un impact dans le détecteur à pixel (vert) et avec toutes les itérations dont celles avec des graines déplacées (rouge) [87].

L'algorithme de flux de particules utilise aussi les amas d'énergie reconstruits dans les ECAL et HCAL. D'une façon analogue aux traces, une graine est définie dans un maximum d'énergie local où les cellules voisines, avec un signal plus grand que deux fois le niveau de bruit, sont ajoutées pour former ce qu'on appelle un amas d'énergie topologique.

Une fois que ces deux types d'objets sont reconstruits, les traces sont propagées aux amas d'énergie du ECAL dans le but de savoir si les informations, avec des critères

géométriques basés sur leur distance dans le plan  $(\eta, \phi)$ , coïncident. Ensuite, on teste les amas d'énergies des deux calorimètres afin de vérifier qu'ils partagent la même position dans le plan précédemment cité. Si plusieurs amas différents sont possibles, les éléments les plus proches sont choisis par le test. L'algorithme de liens va finalement connecter les éléments PF des différents sous-détecteurs afin de reconstruire des objets physiques plus complexes comme des particules. Cet algorithme teste plusieurs paires d'éléments, restreintes aux plus proches voisins dans le plan  $(\eta, \phi)$ , afin que le temps de calcul ne croisse pas de façon quadratique avec le nombre de particules dans l'événement. Si deux éléments sont liés, une distance quantifiant la certitude du lien est définie entre les deux. L'algorithme crée alors des « blocs PF » des différents éléments liés. Mais si, par exemple, plusieurs amas sont liés à la même trace ou si plusieurs traces sont liées au même amas, seul le lien ayant la plus petite distance est gardé.

## 3.2 Objets physiques reconstruits

Les muons sont les premières particules reconstruites, et on supprime ensuite leurs traces et amas des blocs PF. Puis, vient le tour des électrons et des photons, où de la même manière, leurs traces et énergies déposées sont effacées des blocs PF après leur reconstruction. Ensuite les hadrons chargés et neutres sont identifiés. Pour terminer, une dernière étape est effectuée après que tous les blocs ont été traités. En effet, il existe une faible probabilité, mais non nulle, que des particules soient mal identifiées et reconstruites. Dans de rares cas une grande impulsion transverse manquante artificielle est reconstruite, souvent due à un muon de haut  $p_T$  mal reconstruit ou mal identifié. Les analyses recherchant de la nouvelle physique pourraient donc sélectionner de mauvais événements. Cette dernière étape consiste à sélectionner les particules qui mènent à ces cas. Puis, des corrélations sont faites avec le  $p_T$  et la direction de la particule, et l'amplitude et la direction de l'impulsion transverse manquante. Enfin, l'identification et la reconstruction des particules en question sont modifiées si cela réduit l'impulsion transverse manquante de plus de moitié.

### 3.2.1 Muons

Les muons traversent le détecteur CMS entier, ce qui permet d'exploiter les informations laissées par leurs traces à la fois dans le trajectographe et dans les chambres à muons situées à l'extérieur. La position des chambres à muons permet d'obtenir une haute pureté car les autres particules sont arrêtées plus tôt dans les calorimètres (sauf les neutrinos). Un muon reconstruit seulement avec les informations d'impacts dans les CSC, DT et RPC des chambres à muons est dit « standalone », tandis qu'un muon reconstruit seulement avec les impacts du trajectographe est dit « trace interne » (à l'opposé des chambres à muons qui sont très externes). On peut effectuer un ajustement global afin de combiner les informations des deux sous-détecteurs et ainsi obtenir une

meilleure estimation des propriétés de la particule. Cela peut se faire dans deux sens différents.

On peut d'abord partir d'un muon « standalone » où l'on va tenter de faire correspondre sa trace avec les traces du trajectographe en silicium. On appellera cet objet un muon global. Mais on peut aussi partir des traces internes avec  $p_T > 0.5$  GeV et une impulsion totale supérieure à 2.5 GeV, et extrapoler jusqu'aux chambres à muons afin de les faire correspondre à des impacts. On utilise pour cela les coordonnées (x,y) locales du système. La correspondance est vérifiée si la valeur absolue de la différence de position entre la trace et le segment est inférieure à 3 cm en x ou si le ratio de cette distance sur son incertitude est plus petite que 4. La trace en question du trajectographe en silicium est alors nommée trace de muon du trajectographe. Environ 99% des muons produits dans l'acceptance géométrique des chambres à muons sont reconstruits comme des muons globaux et/ou des muons du trajectographe. Un muon global et un muon du trajectographe qui partagent la même trace interne sont fusionnés en une seule particule. Une condition supplémentaire, d'isolement, est nécessaire afin de rejeter les gerbes de particules se faisant passer pour des muons ou des muons provenant de quarks lourds :

$$I_{\mu^i} = \sum_{\Delta R(i,j) < 0.4}^{\text{Ch-PV}} p_T^j + \max \left[ 0, \sum_{\Delta R(i,j) < 0.4}^N p_T^j + \sum_{\Delta R(i,j) < 0.4}^{\gamma} p_T^j - \frac{1}{2} \sum_{\Delta R(i,j) < 0.4}^{\text{Ch-PU}} p_T^j \right], \quad (3.1)$$

où chaque somme parcourt les particules j avec  $\Delta R < 0.4$  par rapport au muon. Ch-PV et Ch-PU sont les particules chargées associées au vertex primaire et à ceux d'empilement respectivement. N et  $\gamma$  correspondent aux hadrons neutres et photons reconstruits avec le PF respectivement. La soustraction de la moitié de la contribution Ch-PU correspond à la contamination de l'empilement. En effet, la moitié du flux de particules provenant de l'empilement est visible sous forme de traces de particules chargées, la part de particules neutres étant plus difficilement identifiable.

Les muons sont reconstruits et identifiés avec une efficacité supérieure à 96% et une résolution spatiale des impacts reconstruits de 50 à 300  $\mu\text{m}$  jusqu'à  $|\eta| = 2.4$ . La distribution de la masse invariante des paires de muons avec des données de 2018 peut être vue en Fig. 3.3.

### 3.2.2 Électrons

Les électrons laissent une trace dans le trajectographe en silicium ainsi qu'un dépôt d'énergie dans le ECAL. Mais avant d'atteindre ce dernier et à cause de la quantité de matière élevée du trajectographe, les électrons perdent une part importante de leur énergie par rayonnement continu de freinage sous forme de photons (autour de 30%

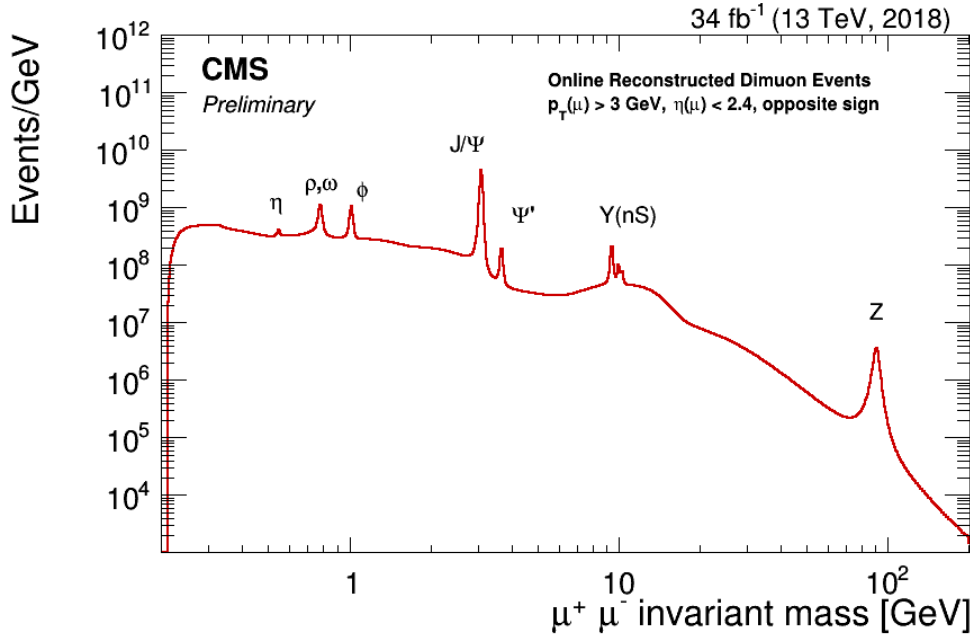


FIGURE 3.3 – Distribution de la masse invariante de paires de muons de charges opposées avec des données de 2018 au niveau HLT. Les résonances connues sont aussi indiquées [88].

d'énergie en moyenne pour une pseudorapidité d'environ 0, mais qui peut atteindre jusqu'à 86% à plus grande pseudorapidité). On a donc besoin d'estimer ce qui a été perdu afin de connaître l'impulsion totale de l'électron.

On emploie pour cela l'algorithme « Gaussian-sum-filter » (GSF), qui utilise des sommes pondérées de gaussiennes afin de modéliser les photons émis, plutôt que les mêmes algorithmes d'ajustement des autres particules chargées s'appuyant sur des filtres de Kalman. Les dépôts d'énergie des électrons et des photons dans le ECAL sont regroupés en super amas où les amas d'énergie initiaux des électrons sont étendus pour récupérer leurs éventuelles radiations dans une petite fenêtre en  $\eta$  et une plus grande en  $\phi$  (afin de prendre en compte la courbure des électrons dans le champ magnétique). L'algorithme GSF commence à partir d'une graine créée par un super amas dans le ECAL ou par deux ou trois impacts dans le trajectographe et la trace est alors extrapolée de la graine au trajectographe entier. Dans le dernier cas, le super amas doit correspondre à la trace de l'électron.

Comme les muons, une condition d'isolement est appliquée mais avec  $\Delta R < 0.3$  au lieu de 0.4. Les électrons sont reconstruit pour  $|\eta| < 2.5$  dans un large intervalle en  $p_T$ . L'efficacité, mesurée dans des données en  $Z \rightarrow e^+e^-$  ayant un intervalle en  $p_T$  entre 10 et 100 GeV, se situe entre 88 et 98% dans les tonneaux et entre 90 et 96% dans les bouchons. Un électron bien mesuré dans les tonneaux aura une résolution en impulsion

totale d'environ 1.7% tandis que ce sera environ 4.5% pour un électron mal mesuré dans les bouchons [89]. On peut voir en Fig. 3.4 la distribution de la masse invariante de paires d'électrons.

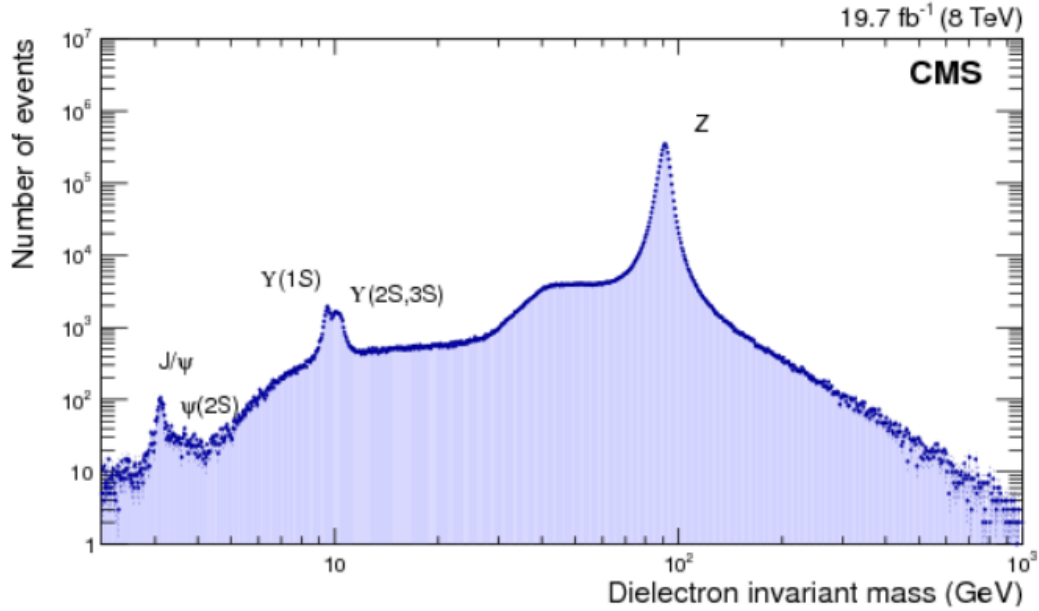


FIGURE 3.4 – Distribution de la masse invariante de paires d'électrons avec des données à 8 TeV en utilisant un système de déclenchement di-électrons [90].

### 3.2.3 Photons

Les photons ne laissent pas de trace dans le trajectographe mais seulement un dépôt d'énergie dans le ECAL. On les reconstruit à partir de graines posées dans des super amas du ECAL ayant une énergie transverse supérieure à 10 GeV et qui ne sont pas utilisés pour des traces GSF. Ces photons doivent être isolés des autres amas du ECAL ou de toutes traces extrapolées du trajectographe. La haute densité de matériaux du trajectographe fait qu'un photon peut donner une paire électron-positron avant d'atteindre le ECAL. Ces particules créées vont à leur tour parfois rayonner un photon. Ces émissions consécutives de particules sont désignées par « gerbes de particules ». Par conséquent, la distribution en énergie des cellules du ECAL et le ratio d'énergie entre le HCAL et le ECAL doit être compatible avec ce qui est attendu d'une gerbe de photons. Les désintégrations de mésons comme les  $\pi^0$ , donnant des paires de photons, sont un grand bruit de fond pour la reconstruction des photons.

### 3.2.4 Jets

Lorsque des gluons ou des quarks sont produits, ils rayonnent et s'hadronisent par interaction forte, ce qui ajoute beaucoup de particules supplémentaires. On reconstruit ces gerbes de particules, appelées jets, en utilisant les objets reconstruits avec le PF - après suppression des muons, électrons et photons de la collection - grâce à un algorithme anti-kt [91]. Ce dernier fonctionne récursivement en regroupant les particules PF dans des cônes en fonction de leur  $p_T$  et de leur distance, définis comme :

$$\begin{aligned} d_{ij} &= \min(p_{T_i}^{-2}, p_{T_j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \\ d_{iB} &= p_{T_i}^{-2}, \end{aligned} \quad (3.2)$$

avec  $p_{T_i}$  et  $p_{T_j}$  l'impulsion transverse de l'objet  $i$  et  $j$  respectivement,  $\Delta R_{ij}$  la séparation spatiale entre les objets  $i$  et  $j$ , et  $R$  le paramètre de distance choisi comme étant 0.4 pour le Run 2. Lors du Run 1, ce paramètre était de 0.5, mais ce changement fut décidé au vu de l'augmentation de l'énergie et de l'empilement entre les deux périodes de prise de données.  $d_{ij}$  et  $d_{iB}$  sont calculées pour tous les objets et la valeur minimale est gardée. Si  $d_{ij}$  est minimale, les objets  $i$  et  $j$  sont fusionnés en un seul objet. Si  $d_{iB}$  est minimale, l'objet  $i$  est considéré comme un jet et est supprimé de la liste des objets. Ces étapes sont répétées jusqu'à qu'il n'y ait plus d'objet.

Dans le cas d'objets physiques boostés donnant des partons, plusieurs jets peuvent être fusionnées en ce qu'on nomme des « fat jets ». Dans ces cas là, le paramètre de distance est augmenté à 0.8 ou 1.0. La couverture du trajectographe ne permet de reconstruire les particules chargées des jets que jusqu'à  $|\eta| = 2.4$ , tandis que les particules neutres sont reconstruites jusqu'à  $|\eta| = 5$ .

Il existe plusieurs vertex primaires par évènement mais seul celui qui a la plus grande somme quadratique du  $p_T$  de ses traces,  $\sum p_T^2$ , est considéré comme intéressant, les autres étant issus de l'empilement. Les hadrons chargés provenant de l'empilement sont supprimés de la liste des particules reconstruites, ce qui permet de réduire la contamination due à l'empilement. C'est la méthode de soustraction des hadrons chargés (CHS). De plus, un BDT est utilisé pour les jets restants. En effet, les interactions d'empilement génèrent souvent des jets avec une impulsion transverse de quelques GeV, or avec la superposition des jets, plusieurs de ces derniers sont fusionnés en un seul ayant un  $p_T$  de plusieurs dizaines de GeV. Le problème étant que ces jets reconstruits vont passer les sélections. Ce BDT utilise des variables comme la multiplicité de particules chargées et neutres afin de rejeter de tels jets en fonction de l'impulsion transverse et de la pseudo-rapidité.

Une autre correction est appliquée afin que l'énergie du jet reconstruit soit égale en moyenne à l'énergie du jet généré suivant différents intervalles en pseudorapidités reconstruites et impulsions transverses générées. Cette correction est mesurée en Fig. 3.5 dans des échantillons simulés à 13 TeV pour l'année 2018. Les corrections sont stables dans les tonneaux mais chutent pour  $p_T < 30$  GeV à cause de l'acceptance du HCAL. En revanche, le ratio dépend fortement du  $p_T$  dans les bouchons et dans le calorimètre à l'avant. La partie EC2 des bouchons est affectée par la dégradation du calorimètre au cours du temps. Les performances varient fortement pour  $3.0 < |\eta| < 3.2$  à cause d'une transition dans le détecteur et pour  $|\eta| > 4.5$  due à l'acceptance.

De plus, les jets sont corrigés d'une manière similaire, mais cette fois-ci par rapport aux données. On appelle toutes ces corrections en énergie : échelle d'énergie des jets (JES). Après ces corrections et d'une façon similaire, les différences entre les données et la simulation dans la résolution de l'énergie des jets (JER) sont corrigées en rendant moins parfaite la distribution en énergie de la simulation.

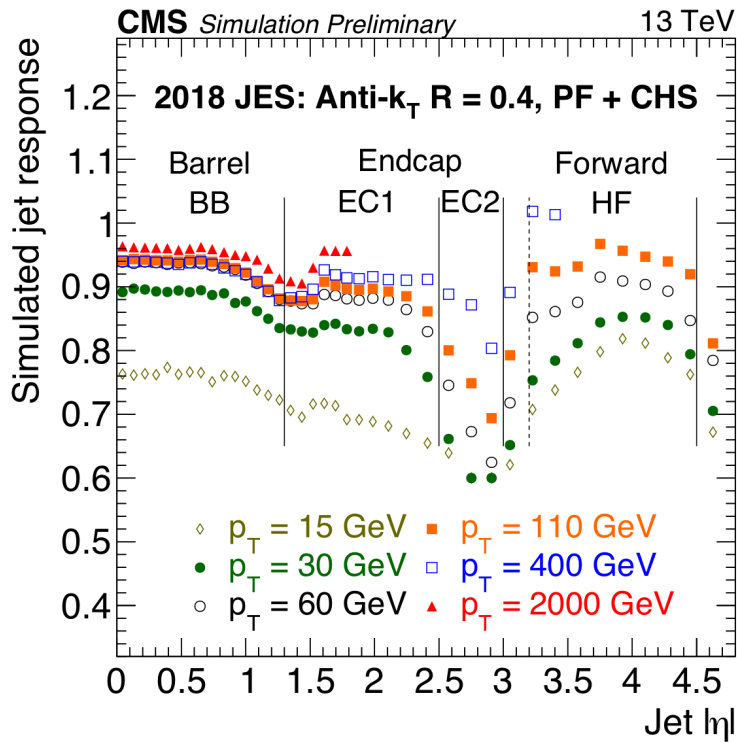


FIGURE 3.5 – Impulsion transverse moyenne reconstruite des jets divisée par l'impulsion transverse moyenne générée des jets en fonction de leurs pseudorapidités absolues (13 TeV pour 2018). Cette correction dépend de la pseudorapidité reconstruite et de l'impulsion transverse générée des jets [92].

### 3.2.5 Taus

Les leptons taus se désintègrent par interaction faible via un boson  $W$  virtuel (voir Fig. 3.6) et donnent dans 2/3 des cas des hadrons avec un neutrino (on parlera alors de taus hadroniques,  $\tau_h$ ), le tiers restant étant soit un électron soit un muon accompagné de deux neutrinos (voir tableau 3.1). C'est le seul lepton pouvant se désintégrer en hadrons. Le mode de désintégration donnant 3 pions chargés, un pion neutre et un neutrino, n'est pas étudié dans le chapitre 5 car il ne passe pas par une résonance en  $a_1$ .

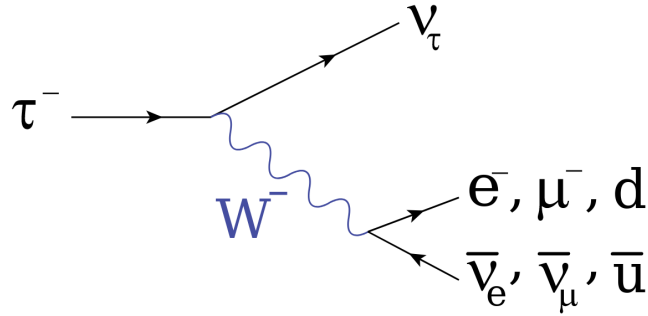


FIGURE 3.6 – Diagramme de Feynman de la désintégration du lepton tau au premier ordre.

Modes de désintégration	Résonance	BR [%]
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \rightarrow h^- \nu_\tau$		11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
$\tau^- \rightarrow h^- 2\pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8

TABLE 3.1 – Principaux modes de désintégration du tau avec leurs rapports d'embranchements (BR).  $h^-$  représente les pions, et les kaons avec une probabilité moindre. Cependant les résonances ne sont présentes que pour  $h^- = \pi^-$  seulement [93].

Sa détection directe dans CMS est impossible car son temps de vie n'est que de  $290 \times 10^{-15}$  s, ce qui ne lui permet pas d'atteindre les premières couches de détection. Ce sont donc ses produits de désintégration que l'expérience CMS va détecter. Dans le canal leptonique, les électrons et les muons seront détectés et il sera difficile de les différencier d'une production directe à cause du très court temps de vie du tau. Concernant le canal hadronique, les hadrons produits seront aussi détectés.

On utilise l'algorithme « hadron-plus-strip » (HPS) [94] afin de reconstruire les  $\tau_h$  et leurs modes de désintégration. Les hadrons chargés sont identifiés en partant des jets



reconstruits par l'algorithme PF. En effet, ces particules laisseront une trace dans le trajectographe qui sera reliée à un dépôt d'énergie dans les calorimètres. Leur énergie sera la moyenne pondérée de l'impulsion de la trace et du dépôt associé. Les hadrons chargés candidats sont collectés dans un cône de signal de rayon variable  $R = 3/p_T$  centré sur la direction du  $\tau_h$  et doivent avoir  $p_T > 0.5$  GeV (illustré en Fig. 3.7).

Les hadrons neutres sont principalement des pions neutres donnant une paire de photons qui eux mêmes peuvent donner une paire d'électrons par conversion. L'algorithme HPS reconstruit les pions neutres en utilisant des bandes calorimétriques, dans le plan  $(\eta, \phi)$ , contenant des photons et des électrons provenant de conversions avec  $p_T > 0.5$  GeV (illustré en Fig. 3.7).

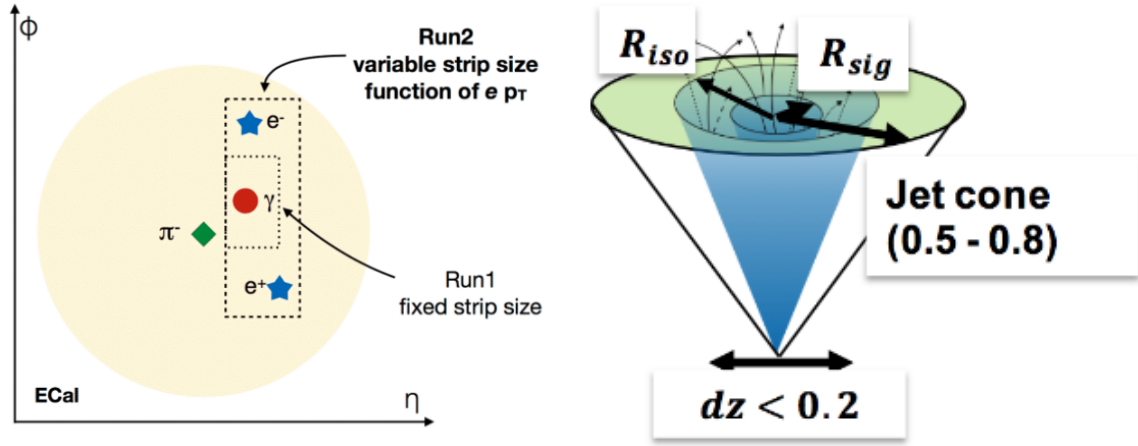


FIGURE 3.7 – Illustration des bandes calorimétriques employées pour reconstruire les hadrons neutres (gauche) et des cônes utilisés pour reconstruire les hadrons chargés (droite).

Pendant le Run 1, les bandes avaient une dimension fixe de  $0.05 \times 0.20$ . Mais pour le Run 2, cette taille est devenue variable [95] et dépend maintenant du  $p_T$  des électrons et photons. En effet, dorénavant l'électron ou le photon avec le plus haut  $p_T$  qui n'est pas encore dans une bande est utilisé pour en créer une nouvelle. Ensuite, le prochain électron ou photon étant dans ces intervalles  $\Delta\eta$  et  $\Delta\phi$ , centrés sur la bande, est ajouté à la bande calorimétrique :

$$\begin{aligned}\Delta\eta &= f(p_T^{e/\gamma}) + f(p_T^{\text{strip}}), \\ \Delta\phi &= g(p_T^{e/\gamma}) + g(p_T^{\text{strip}}),\end{aligned}\tag{3.3}$$

avec  $p_T^{e/\gamma}$  l'impulsion transverse de l'électron ou du photon candidat et  $p_T^{\text{strip}}$  l'impulsion transverse de la bande avant l'ajout du candidat. De plus, ces intervalles sont

bornés avec  $0.05 < \Delta\eta < 0.15$  et  $0.05 < \Delta\phi < 0.3$ . Les fonctions  $f(p_T)$  et  $g(p_T)$  sont définies comme :

$$\begin{aligned} f(p_T) &= 0.2 p_T^{-0.66}, \\ g(p_T) &= 0.35 p_T^{-0.71}. \end{aligned} \tag{3.4}$$

Puis, la position de la bande est recalculée comme la moyenne pondérée par le  $p_T$  de tous les électrons et photons de la bande. La construction de la bande se termine s'il n'y a plus d'électron ou de photon dans la fenêtre en  $\eta - \phi$ , et une nouvelle bande se construit à partir du prochain électron ou photon ayant le plus grand  $p_T$ . Les bandes contenant au moins un électron ou un photon avec une somme de  $p_T$  supérieure à 2.5 GeV sont gardées comme de potentiels pions neutres.

L'algorithme combine alors les bandes avec les particules chargées du jet PF pour créer les candidats  $\tau_h$ . Suivant le nombre d'objets reconstruits et avec une condition sur la masse invariante afin de reconstruire les résonances  $\rho(770)$  et  $a_1(1260)$ , chaque tau est assigné à un mode de désintégration (DM) :

- $h^\pm$  : une seule particule chargée sans bande (DM0) ;
- $h^\pm\pi^0$  : une particule chargée + une bande avec  $0.3 - \Delta m_\tau < m_\tau < 1.3 \cdot \sqrt{p_T/100} + \Delta m_\tau$  avec la limite supérieure entre 1.3 et 4.2 GeV (DM1) ;
- $h^\pm\pi^0\pi^0$  : une particule chargée + deux bandes avec  $0.4 - \Delta m_\tau < m_\tau < 1.2 \cdot \sqrt{p_T/100} + \Delta m_\tau$  avec la limite supérieure entre 1.2 et 4.0 GeV (DM2) ;
- $h^\pm h^\mp h^\pm$  : trois particules chargées sans bande avec  $0.8 - \Delta m_\tau < m_\tau < 1.5 + \Delta m_\tau$  avec les traces devant provenir du même vertex dans un intervalle  $d_z < 0.4$  cm, avec  $d_z$  le paramètre d'impact longitudinal du tau par rapport au vertex primaire (DM10) ;
- $h^\pm h^\mp h^\pm \pi^0$  : trois particules chargées avec une bande (DM11).

$\Delta m_\tau$  représente le changement de masse du  $\tau_h$  lors de l'ajout des électrons et photons aux bandes calorimétriques dynamiques.

Une nouvelle méthode d'identification des modes de désintégrations du tau hadronique a été développée en utilisant un BDT [96, 97]. En effet, dans le cadre de l'étude de l'état CP du boson de Higgs donnant une paire de taus hadroniques, il est important de bien reconstruire les différents modes de désintégrations qui sont différemment caractérisables vis-à-vis de l'état CP. Une très grande majorité des désintégrations donnant une résonance  $a_1$  puis  $\pi^\pm 2\pi^0 \nu$  sont reconstruits comme DM0 ou DM1 avec l'algorithme HPS car ce dernier ne distingue pas efficacement les deux pions neutres. La raison est que la taille dynamique des bandes calorimétriques permet de mieux récupérer les électrons et photons venant d'un tau hadronique, mais a le désavantage de moins bien distinguer les différents pions neutres. Le BDT considère donc les événements reconstruits comme

DM0, DM1 et DM2 et utilise des variables comme le mode de désintégration HPS, la masse de la résonance  $\rho$  et des pions neutres ou encore l'énergie des pions.

Le canal  $a_1 \rightarrow 3\pi^\pm$  est principalement contaminé par des événements donnant trois pions chargés et un pion neutre, ce qui diminue la sensibilité de ce canal à la détermination de l'état CP du Higgs. Une façon de faire, similaire aux événements avec un seul pion chargé est donc utilisée pour mieux séparer ces deux canaux. L'algorithme HPS reconstruit les bandes calorimétriques de chaque événement et la bande avec le plus haut  $p_T$  est gardée afin d'utiliser ses propriétés dans le BDT. Les variables utilisées sont la masse invariante des paires de pions chargés ou la fraction d'énergie portée par chaque pion chargé.

Le BDT assigne donc chaque tau à un nouveau mode de désintégration nommé MVA DM. Les matrices de confusion du MVA DM pour 2018 sont données en Fig. 3.8. De plus, la différence de pureté et d'efficacité entre les HPS DM et les MVA DM se trouvent en Fig. 3.9. Le MVA DM améliore la pureté pour tous les DM et améliore leur efficacité s'ils possèdent au moins un pion neutre. Pour le canal  $a_1 \rightarrow 3\pi^\pm$  par rapport à l'algorithme HPS par exemple, on perd 2% en efficacité mais le gain en pureté est de 15%.

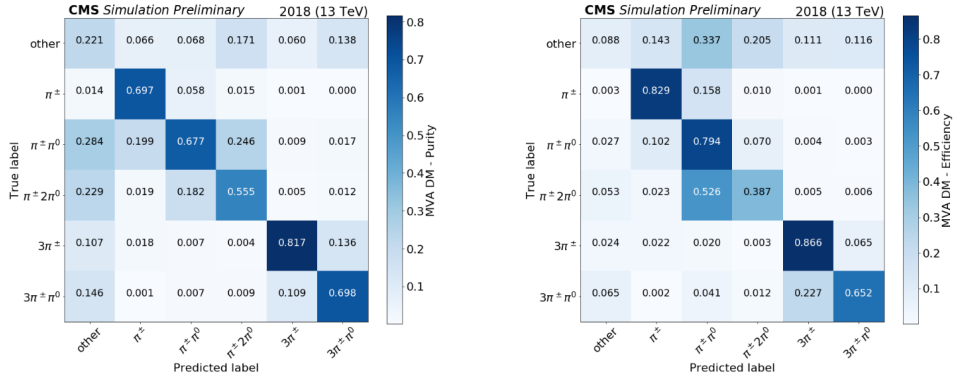


FIGURE 3.8 – Matrices de confusion du MVA DM où la diagonale de la figure de gauche donne la pureté des MVA DM tandis que celle de droite donne l'efficacité des MVA DM pour 2018 [97].

Cependant les interactions QCD donnent aussi des jets pouvant s'apparenter à un tau se désintégrant en différents hadrons. Les principales différences sont que les jets provenant de leptons taus ont une faible multiplicité de particules, déposent de l'énergie dans une région restreinte et sont bien isolés des autres particules de l'évènement (voir Fig. 3.10). Une sélection sur l'isolement du candidat  $\tau_h$  est requise afin de rejeter un maximum de jet QCD faussement reconstruit en  $\tau_h$ . Cet isolement est calculé comme [98] :

$$I_{\tau_h} = \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}) + \max(0, \sum p_T^\gamma - \Delta\beta \sum p_T^{\text{charged}}(d_z > 0.2 \text{ cm})), \quad (3.5)$$

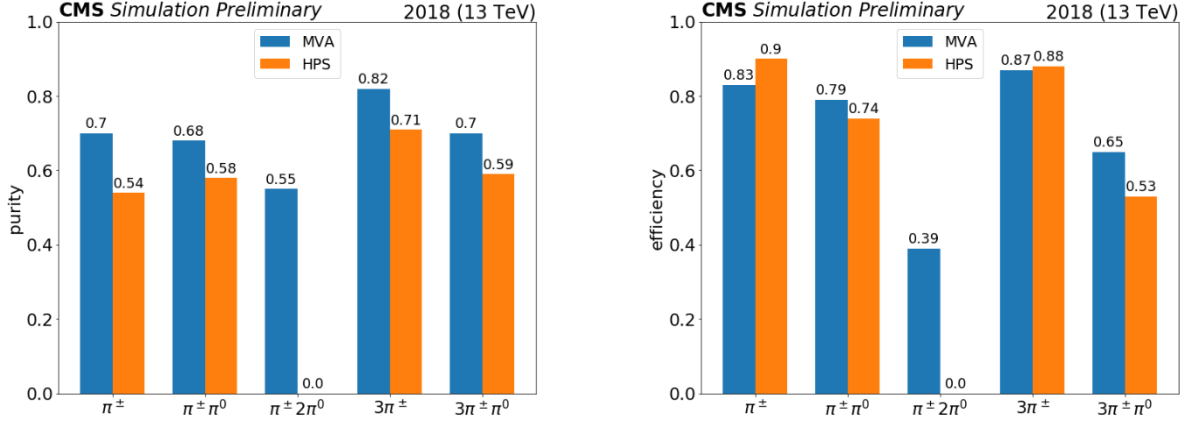


FIGURE 3.9 – Comparaison de la pureté (gauche) et de l’efficacité (droite) du MVA DM (bleu) et du HPS DM (orange) pour 2018 [97].

avec  $\sum p_T^{\text{charged}}$  et  $\sum p_T^\gamma$  la somme scalaire du  $p_T$  des particules chargées et des photons PF respectivement ne faisant pas partie du  $\tau_h$ , reconstruits dans un cône d’isolement de taille  $\Delta R = 0.5$  centré autour de la direction du  $\tau_h$ . La condition  $d_z < 0.2$  cm pour le calcul de  $\sum p_T^{\text{charged}}$  permet de supprimer les contributions venant de l’empilement. Concernant  $\sum p_T^\gamma$ , la suppression de cette contribution pour les photons est faite en soustrayant la part de  $\sum p_T^{\text{charged}}$  ne provenant pas du vertex de production du  $\tau_h$  ( $d_z > 0.2$  cm) et étant reconstruit dans un cône de taille  $\Delta R = 0.8$ . Ce terme est multiplié par  $\Delta\beta = 0.2$  qui prend en compte le ratio hadrons neutres sur hadrons chargés produits pendant l’hadronisation ainsi que les différentes tailles de cônes utilisées pour estimer l’empilement. Différents points de fonctionnement sont définis : Loose, Medium et Tight pour  $I_{\tau_h}$  inférieur à 2.5, 1.5 et 0.8 GeV respectivement.

Une condition d’isolement supplémentaire est appliquée concernant les bandes calorimétriques. En effet, un photon en dehors du cône de signal peut tout de même contribuer au signal.  $I_{\tau_h}$  diminue ce qui augmente la proportion de jets faussement reconstruits comme un  $\tau_h$ . Pour corriger cela, on utilise la somme scalaire du  $p_T$  des électrons et photons candidats au sein des bandes mais situés en dehors du cône de signal [98] :

$$p_T^{\text{strip,outer}} = \sum p_T^{e/\gamma}(\Delta R > R_{\text{sig}}). \quad (3.6)$$

Cette quantité doit être inférieure à 10% de l’impulsion transverse du tau hadronique, ce qui permet de réduire de 20% les identifications de jets en  $\tau_h$ .

Les jets ne sont pas les seuls objets à être reconstruits comme des  $\tau_h$ , c’est aussi le cas des muons et des électrons. En effet, un électron éjectant un photon par rayonnement continu de freinage pourrait déposer son énergie dans le ECAL et son photon pourrait s’apparenter à un produit de désintégration du pion neutre. Le tout ressemblerait à un

tau hadronique. Un muon peut aussi passer les sélections d'isolement si le signal dans les chambres à muons est près de la direction du  $\tau_h$ . C'est pour cela que des critères d'identification de réjection sont aussi appliqués aux deux leptons afin de rejeter les fausses reconstructions.

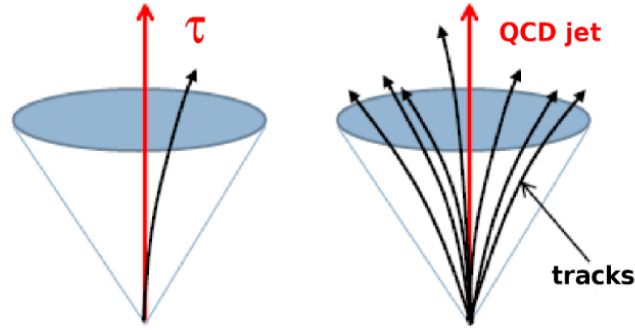


FIGURE 3.10 – Schémas représentant les différences entre un jet venant d'un tau et venant d'une interaction QCD.

Un BDT peut être utilisé pour rejeter les muons faussement reconstruits en taus. Les variables utilisées pour l'entraînement sont [94] :

1. Pseudorapidité du  $\tau_h$  ;
2. Énergie des calorimètres associée à la particule chargée de plus haut  $p_T$  du  $\tau_h$  en sommant séparément le ECAL et le HCAL ;
3. Énergie des calorimètres associée à n'importe quelle particule chargée ou photon reconstruits avec le PF faisant partie du  $\tau_h$  en sommant séparément le ECAL et le HCAL ;
4. Fraction du  $p_T$  de la particule chargée de plus haut  $p_T$  ;
5. Nombre de segments dans les chambres à muons reconstruits dans un cône de taille  $\Delta R = 0.5$  centré sur la direction du  $\tau_h$  ;
6. Nombre de stations de muons avec au moins un impact détecté dans un cône de taille  $\Delta R = 0.5$  centré sur la direction du  $\tau_h$  et calculé séparément pour les sous-détecteurs DT, CSC et RPC.

Ce BDT a été entraîné avec des échantillons  $Z/\gamma^* \rightarrow \tau\tau$ ,  $Z/\gamma^* \rightarrow \mu\mu$ ,  $W \rightarrow \tau\bar{\nu}_\tau$ ,  $W \rightarrow \mu\bar{\nu}_\mu$ ,  $t\bar{t}$ ,  $H \rightarrow \tau\tau$ ,  $Z' \rightarrow \tau\tau$ ,  $Z' \rightarrow \mu\mu$ ,  $W' \rightarrow \tau\bar{\nu}_\tau$  et  $W' \rightarrow \mu\bar{\nu}_\mu$ . Le candidat  $\tau_h$  reconstruit est comparé à un  $\tau_h$  ou à un muon généré dans un cône de rayon  $\Delta R < 0.3$ , afin de pouvoir le considérer comme du signal ou du bruit de fond.

Les électrons peuvent aussi être rejetés avec un BDT. Ce dernier est entraîné avec des grandeurs physiques classiques comme le  $p_T$ , la pseudorapidité du  $\tau_h$  et des traces GSF mais aussi avec ces variables [94, 98] :

1.  $E_{\text{ECAL}}/(E_{\text{ECAL}} + E_{\text{HCAL}})$ , fraction d'énergie déposée dans le ECAL ;

2.  $E_{\text{ECAL}}/p$  et  $E_{\text{HCAL}}/p$ , ratio d'énergie dans le ECAL et le HCAL sur l'impulsion de la trace ayant le plus haut  $p_T$  constituant le  $\tau_h$  ;
3.  $\sqrt{\sum(\Delta\eta)^2 p_T^\gamma}$  et  $\sqrt{\sum(\Delta\phi)^2 p_T^\gamma}$ , moyenne quadratique des distances en  $\eta$  et  $\phi$  entre les photons des bandes et la particule chargée de plus haut  $p_T$  pondéré par le  $p_T$  (en GeV) des photons ;
4.  $\sum E_\gamma/E_\tau$ , fraction d'énergie des photons par rapport à l'énergie du  $\tau_h$  ;
5.  $F_{\text{brem}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}}$ , différence en impulsion entre l'entrée et la sortie du trajectographe divisé par l'impulsion à l'entrée ;
6.  $(E_e + \sum E_\gamma)/p_{\text{in}}$ , énergie des amas venant des électrons et des photons du rayonnement continu de freinage divisé par l'impulsion en entrée du trajectographe ;
7.  $\sum E_\gamma/(p_{\text{in}} - p_{\text{out}})$ , ratio de l'énergie des photons du rayonnement continu de freinage mesurée dans le ECAL sur celle mesurée dans le trajectographe ;
8.  $m_{\tau_h}$ , masse du tau hadronique ;
9.  $(N_{\text{hits}}^{\text{GSF}} - N_{\text{hits}}^{\text{KF}})/(N_{\text{hits}}^{\text{GSF}} + N_{\text{hits}}^{\text{KF}})$ , avec  $N_{\text{hits}}^{\text{GSF}}$  et  $N_{\text{hits}}^{\text{KF}}$  le nombre d'impacts, dans les pixels et pistes, avec l'algorithme GSF et le filtre de Kalman respectivement ;
10.  $\chi^2$  par degré de liberté des traces GSF ;
11. Nombre de photons dans une bande associé au  $\tau_h$  ;
12. Moyenne quadratique de la distance en  $\eta$  et  $\phi$  entre les photons inclus dans des bandes et la trace de plus haut  $p_T$  du  $\tau_h$  pondéré par le  $p_T$  des photons.

L'entraînement est effectué sur des échantillons de  $H \rightarrow \tau\tau$ ,  $Z' \rightarrow ll$ ,  $W' \rightarrow l\nu$  et  $Z/\gamma^* \rightarrow ll$ , avec  $l = e$  ou  $\tau$ , où tout comme pour le rejet des muons, on regarde la correspondance du  $\tau_h$  reconstruit avec un  $\tau_h$  ou un électron généré avec un cône de taille  $\Delta R = 0.3$ .

Concernant mon analyse ([chapitre 5](#)), je n'ai pas utilisé ces BDT pour l'identification des leptons mais un DNN appelé DeepTau et développé dans CMS qui rejette les mauvaises identifications des  $\tau_h$  en jets, électrons ou muons [99]. Ce DNN combine une variable globale de l'évènement qui est la densité d'énergie moyenne déposée  $\rho$  avec des dizaines d'autres variables comme des informations sur les amas d'énergies dans les calorimètres ou sur la qualité des traces reconstruites. Les performances de ce DNN sont comparées avec celles des BDT utilisés dans la collaboration en Fig. 3.11 avec cette présélection :  $20 < p_T < 1000$  GeV,  $|\eta| < 2.3$  et  $|dz| < 0.2$  cm. L'efficacité d'identification du tau hadronique est estimée à partir d'échantillon  $H \rightarrow \tau\tau$  avec une correspondance du tau au niveau généré. Le taux de faux est estimé à partir d'échantillon  $t\bar{t}$  et Drell-Yan ( $Z/\gamma^* \rightarrow ll$ , DY) pour le rejet des jets, et des électrons et muons respectivement. Pour calculer ce taux lors du rejet des jets, le tau reconstruit ne doit pas correspondre à un électron ou muon ne venant pas d'un tau ni correspondre à un produit de désintégration d'un tau hadronique généré. Le taux pour le rejet des électrons et muons nécessite une correspondance entre le tau reconstruit et l'électron ou le muon généré. L'utilisation de DeepTau permet donc de grandement améliorer l'identification des leptons taus.

### Chapitre 3. Reconstruction des particules et des objets

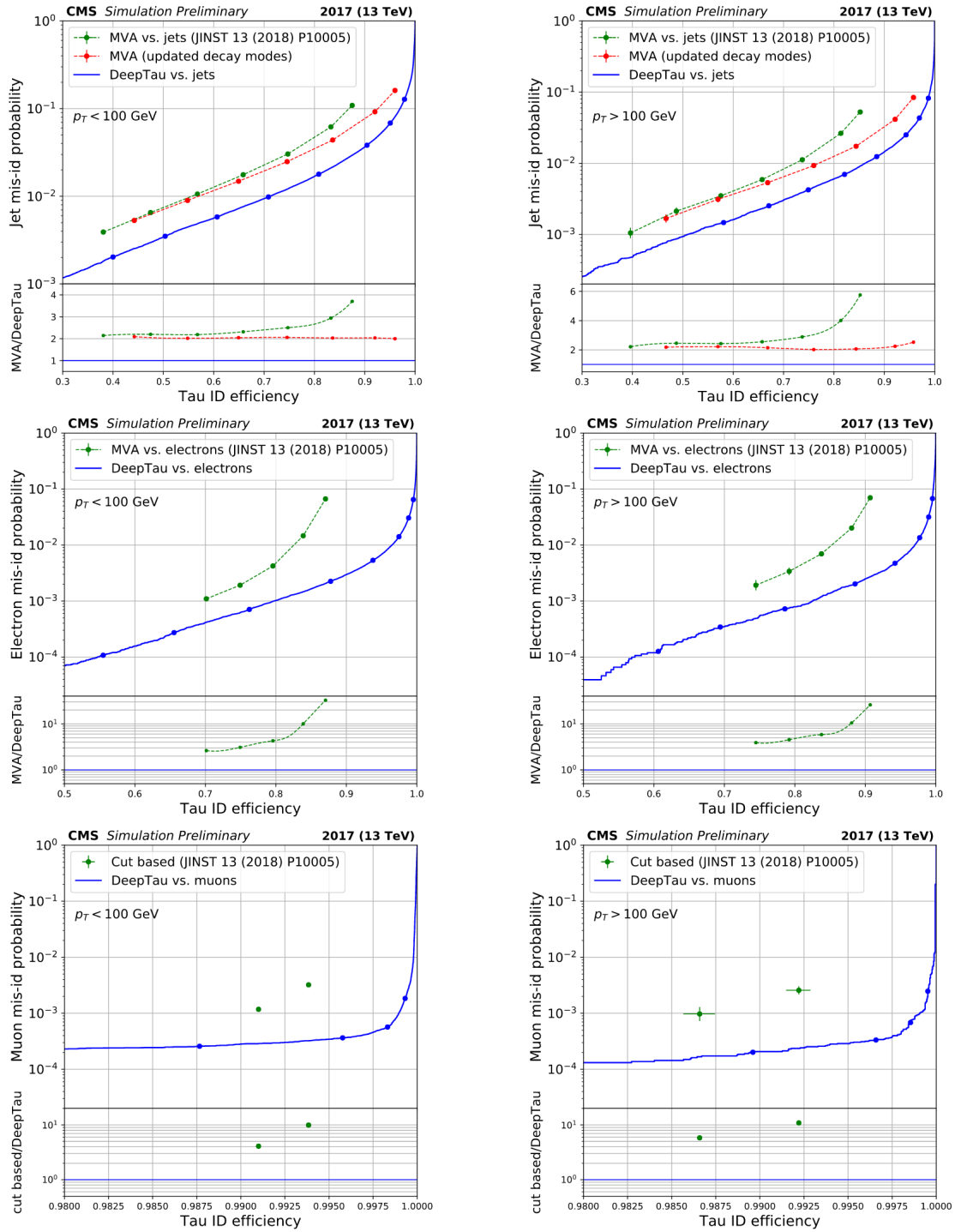


FIGURE 3.11 – Probabilités de mal reconstruire les objets en fonction de l’efficacité pour les jets (haut), les électrons (milieu) et les muons (bas) avec des échantillons Run 2 2017. Deux intervalles en  $p_T$  sont montrés,  $p_T < 100$  GeV (gauche) et  $p_T > 100$  GeV (droite). Les points sont les différents points de fonctionnement disponibles. Les performances sont comparées à celles obtenues avec le BDT classique [99].

### 3.2.5.1 Adaptation de l’algorithme de rejet des électrons pour la Phase 2

Afin de préparer la Phase 2, j’ai réoptimisé le BDT rejetant les électrons se faisant passer pour des taus hadroniques. En effet, le trajectographe aura moins de matière, une acceptation plus étendue, un calorimètre à très haute granularité sera ajouté dans la partie avant (HGCALE) ainsi qu’un détecteur en temps (MTD). Cependant, la luminosité et donc le nombre d’interactions d’empilement vont fortement augmenter lors de cette phase (140 à 200 interactions attendues par croisement de faisceaux au lieu d’une cinquantaine), ce qui rendra plus difficile l’identification des particules. Le développement d’un nouvel entraînement est donc nécessaire. J’ai ajouté de nouvelles catégories correspondants à trois régions spatiales différentes du détecteur CMS ( $0 < |\eta| < 1.4$  (tonneaux),  $1.6 < |\eta| < 2.4$  (HGCALE en silicium + scintillateurs) et  $2.4 < |\eta| < 3.0$  (HGCALE en silicium seulement)), avec une région de transition non utilisée entre les tonneaux et le HGCALE. Ces catégories impliquent donc l’entraînement de non pas un seul BDT, mais de douze au lieu de huit pour le Run 2 (voir tableau 3.2).

Région Spatiale	Trace GSF	Photons
$ \eta  < 1.4$		
$ \eta  < 1.4$	✓	
$ \eta  < 1.4$		✓
$ \eta  < 1.4$	✓	✓
$1.6 <  \eta  < 2.4$		
$1.6 <  \eta  < 2.4$	✓	
$1.6 <  \eta  < 2.4$		✓
$1.6 <  \eta  < 2.4$	✓	✓
$2.4 <  \eta  < 3.0$		
$2.4 <  \eta  < 3.0$	✓	
$2.4 <  \eta  < 3.0$		✓
$2.4 <  \eta  < 3.0$	✓	✓

TABLE 3.2 – Tableau des catégories utilisées dans le BDT rejetant les électrons pour la Phase 2. Chaque ligne correspond à une catégorie. Un ✓ signifie que la condition est vérifiée et son absence signifie qu’elle n’est pas remplie.

J’ai ajouté des variables utilisant des informations du nouveau sous-détecteur HGCALE et d’autres concernant la cascade électromagnétique afin d’entraîner les BDT. Des variables d’isolement ainsi que la densité d’énergie de l’évènement ont été testées mais la différence de performance était minime, j’ai donc choisi de les supprimer afin de ne pas complexifier inutilement les entraînements. La liste exhaustive des variables utilisées est en tableau 3.3. Certaines distributions de signal et de bruit de fond peuvent être vues en Fig. 3.12. Les BDT ont été entraînés sur environ 13 millions d’évènements de signal et de bruit de fond simulés à 14 TeV et un empilement de 200. Les processus utilisés pour l’entraînement sont des dibosons,  $t\bar{t}$ ,  $Z/\gamma^* \rightarrow ll$  et  $H \rightarrow \tau\tau$ . La distribution du score de



sortie du BDT pour le signal et le bruit de fond est montrée en Fig. 3.13 - 3.15 pour les trois régions spatiales ayant chacune quatre BDT différents. L'efficacité et le taux de faux en fonction de  $p_T$ ,  $\eta$  et  $\phi$  sont donnés en Fig. 3.16 pour des échantillons DY pour différents points de fonctionnement. On peut observer la région de transition pour  $\eta \approx 1.5$ . Les grandes barres d'erreur du taux de faux en fonction du  $p_T$  ont pour origine le manque de statistique à haute impulsion transverse. Les efficacités visées pour les différents points de fonctionnements sont 0.98 (Very Loose), 0.9 (Loose), 0.8 (Medium), 0.7 (Tight) et 0.6 (Very Tight) et les coupures sur les BDTs pour définir les points de fonctionnement sont estimées pour les intervalles en  $p_T$  suivant :  $[0,20]$ ,  $[20,30]$ ,  $[30,40]$ ,  $[40,60]$  et  $[60,\infty[$  GeV pour les 12 catégories citées plus haut.

Ces coupures et sélections sont le fruit d'une optimisation afin d'obtenir les meilleures performances possibles. Des comparaisons de courbes ROC (Receiver Operating Characteristic, taux de faux en fonction de l'efficacité) pour l'entraînement du Run 2 et celui de la Phase 2 sont montrées en Fig. 3.17 et 3.18. Tous les modes de désintégration HPS sont considérés et le tau hadronique reconstruit doit correspondre au tau hadronique généré ou à l'électron généré pour l'efficacité ou le taux de faux respectivement. La courbe verte correspond à l'entraînement du Run 2 appliqué à des échantillons DY du Run 2, la mauve au même entraînement mais appliqué à des échantillons DY de la Phase 2 et la bleue à mon entraînement pour la Phase 2 appliqué à des échantillons DY de la Phase 2. On observe que les BDT du Run 2 appliqués à des échantillons de la Phase 2 ne sont pas efficaces, ce qui prouve qu'une optimisation est nécessaire. Le but est d'obtenir des performances similaires au BDT du Run 2 appliqué à des échantillons du Run 2. Les nouveaux BDT ont des performances supérieures aux anciens du Run 2 pour les hautes efficacités (à partir de 90% pour les tonneaux et 80% pour les bouchons), mais restent inférieurs pour les plus basses. De plus, l'application des nouveaux BDT dans les régions des bouchons est plus efficace que dans les tonneaux. Ceci peut s'expliquer par le fait que des variables associées au HGCALE sont utilisées dans l'entraînement et ne sont disponibles que pour  $|\eta| > 1.6$ .

J'ai donc amélioré le BDT actuel qui rejette les électrons se faisant passer pour des taus hadroniques lors de la reconstruction afin qu'il soit utilisable lors de la Phase 2 du LHC. Les performances obtenues sont, pour les hautes efficacités, meilleures que pour la Phase 1, notamment pour la région du HGCALE. En effet, ce dernier permet d'obtenir de très bonnes performances grâce à l'algorithme PF qui en devient amélioré. L'entraînement a été intégré dans les codes utilisés par la collaboration CMS pour les futures analyses HL-LHC. La dégradation restante dans les tonneaux est due à l'empilement et plus particulièrement aux particules neutres de l'empilement. Afin d'encore améliorer les performances de ce BDT, on pourrait tirer parti du nouveau sous-détecteur MTD qui permettra d'obtenir une information temporelle. De plus, l'utilisation d'un réseau de neurone profond (DNN) pourrait potentiellement nous offrir de meilleures performances.

### Chapitre 3. Reconstruction des particules et des objets

Variables	Définitions
TauPt	$p_T^\gamma$
$\frac{\text{LeadChargedPFCandPt}}{\text{TauPt}}$	$\frac{p_T^{\text{lead h}}}{p_T^\gamma}$
EmFraction	$\frac{E_{\text{ECAL}}^{\text{lead h}}}{E_{\text{ECAL}}^{\text{lead h}} + E_{\text{HGCAL}}^{\text{lead h}}}$
LeadPFChargedHadrHoP	$\frac{E_{\text{HGCAL}}^{\text{lead h}}}{p_{\text{lead h}}}$
LeadPFChargedHadrEoP	$\frac{E_{\text{ECAL}}^{\text{lead h}}}{p_{\text{lead h}}}$
EtaAtEcalEntrance	$\eta^\gamma$ à l'entrée du ECAL
LeadChargedPFCandEtaAtEcalEntrance	$\eta^{\text{lead h}}$ à l'entrée du ECAL
VisMassIn	Masse invariante de la somme vectorielle des photons et hadrons
dCrackEta	Distance en $\eta$ de la plus proche zone de transition dans le tonneau ECAL
SignalPFGammaCandsIn	Nombre de photons dans le cône de signal
SignalPFGammaCandsOut	Nombre de photons en dehors du cône de signal
GammaEtaMomIn	$\sqrt{\sum_i^{\gamma \text{ in}} (p_T^\gamma \times (\eta^\gamma - \eta^{\text{lead h}})^2) \times \sqrt{p_T^\gamma}}$
GammaEtaMomOut	$\sqrt{\sum_i^{\gamma \text{ out}} (p_T^\gamma \times (\eta^\gamma - \eta^{\text{lead h}})^2) \times \sqrt{p_T^\gamma}}$
GammaPhiMomIn	$\sqrt{\sum_i^{\gamma \text{ in}} (p_T^\gamma \times (\phi^\gamma - \phi^{\text{lead h}})^2) \times \sqrt{p_T^\gamma}}$
GammaPhiMomOut	$\sqrt{\sum_i^{\gamma \text{ out}} (p_T^\gamma \times (\phi^\gamma - \phi^{\text{lead h}})^2) \times \sqrt{p_T^\gamma}}$
GammaEnFracIn	$\frac{\sum_i^{\gamma \text{ in}} p_T^\gamma}{p_T^\gamma}$
GammaEnFracOut	$\frac{\sum_i^{\gamma \text{ out}} p_T^\gamma}{p_T^\gamma}$
Chi2NormGSF	$\chi^2$ par degré de liberté des traces GSF
Chi2NormKF	$\chi^2$ par degré de liberté des traces KF
GSFNumHits	Nombre d'impacts avec l'algorithme GSF
$\frac{\text{GSFNumHits} - \text{KFNumHits}}{\text{GSFNumHits} + \text{KFNumHits}}$	Différence du nombre d'impacts entre GSF et le filtre de Kalman divisée par leur somme
GSFTrackResol	Résolution du $p_T$ de la trace GSF
GSFTracklnPt	Logarithme népérien du $p_T$ de la trace GSF
Eecal	Énergie déposée de l'électron dans le ECAL
$(\text{Pin} - \text{Pout})/\text{ElecPin}$	Fraction d'impulsion perdue de l'électron dans le trajectographe
Eecal/Pout	Énergie déposée dans le ECAL divisée par l'impulsion en sortie du trajectographe
SigmaEtaEta	$\sigma^{\eta\eta}$ de l'amas PF
HadEnergy	Énergie des amas associées aux hadrons PF
DeltaEta	$\Delta\eta$ entre le amas PF et la trace GSF
DeltaPhi	$\Delta\phi$ entre le amas PF et la trace GSF
SigmaIPhiPhi5x5	Largeur en $\eta$ du amas en 5x5
SigmaIEtaIEta5x5	Largeur en $\phi$ du amas en 5x5
ShowerCircularity	$1 - \frac{E_{1 \times 5}}{E_{5 \times 5}}$
R9	$\frac{E_{3 \times 3}}{E_{\text{superamas}}}$
eSeedClusterOverPout	Énergie de la graine du amas sur l'impulsion sortante
SuperClusterEtaWidth	Largeur en $\eta$ du super amas
SuperClusterPhiWidth	Largeur en $\phi$ du super amas
SigmaUU	Propagation perpendiculaire du amas dans le HGCAL
SigmaVV	Propagation radiale du amas dans le HGCAL
SigmaEE	Propagation en $\eta$ du amas dans le HGCAL
SigmaPP	Propagation en $\phi$ du amas dans le HGCAL
NLayers	Nombre de couches de détection du amas dans le HGCAL
FirstLayer	Première couche de détection du amas dans le HGCAL
LastLayer	Dernière couche de détection du amas dans le HGCAL
LayerEfrac10	Première couche ayant 10% de l'énergie dans le EE du HGCAL
LayerEfrac90	Première couche ayant 90% de l'énergie dans le EE du HGCAL
EcEnergyEE	Énergie dans le EE du HGCAL
EcEnergyFH	Énergie dans le FH du HGCAL
FHoverE	Ratio d'énergie entre le FH et le EE du HGCAL
MeasuredDepth	Profondeur mesurée de la cascade électromagnétique dans le HGCAL
ExpectedDepth	Profondeur prédite de la cascade électromagnétique dans le HGCAL
ExpectedSigma	Propagation prédite de la profondeur de la cascade électromagnétique dans le HGCAL
DepthCompatibility	$\frac{\text{ExpectedDepth} - \text{MeasuredDepth}}{\text{ExpectedSigma}}$

TABLE 3.3 – Tableau des variables utilisées dans l'entraînement du BDT anti-électron. « lead h » désigne le hadron ayant le plus haut  $p_T$  et «  $E_{m \times n}$  » l'énergie dans une matrice de cristaux de taille  $m$  par  $n$ . Les variables en rouges sont les nouvelles variables que j'ai ajoutées [77, 100].

### Chapitre 3. Reconstruction des particules et des objets

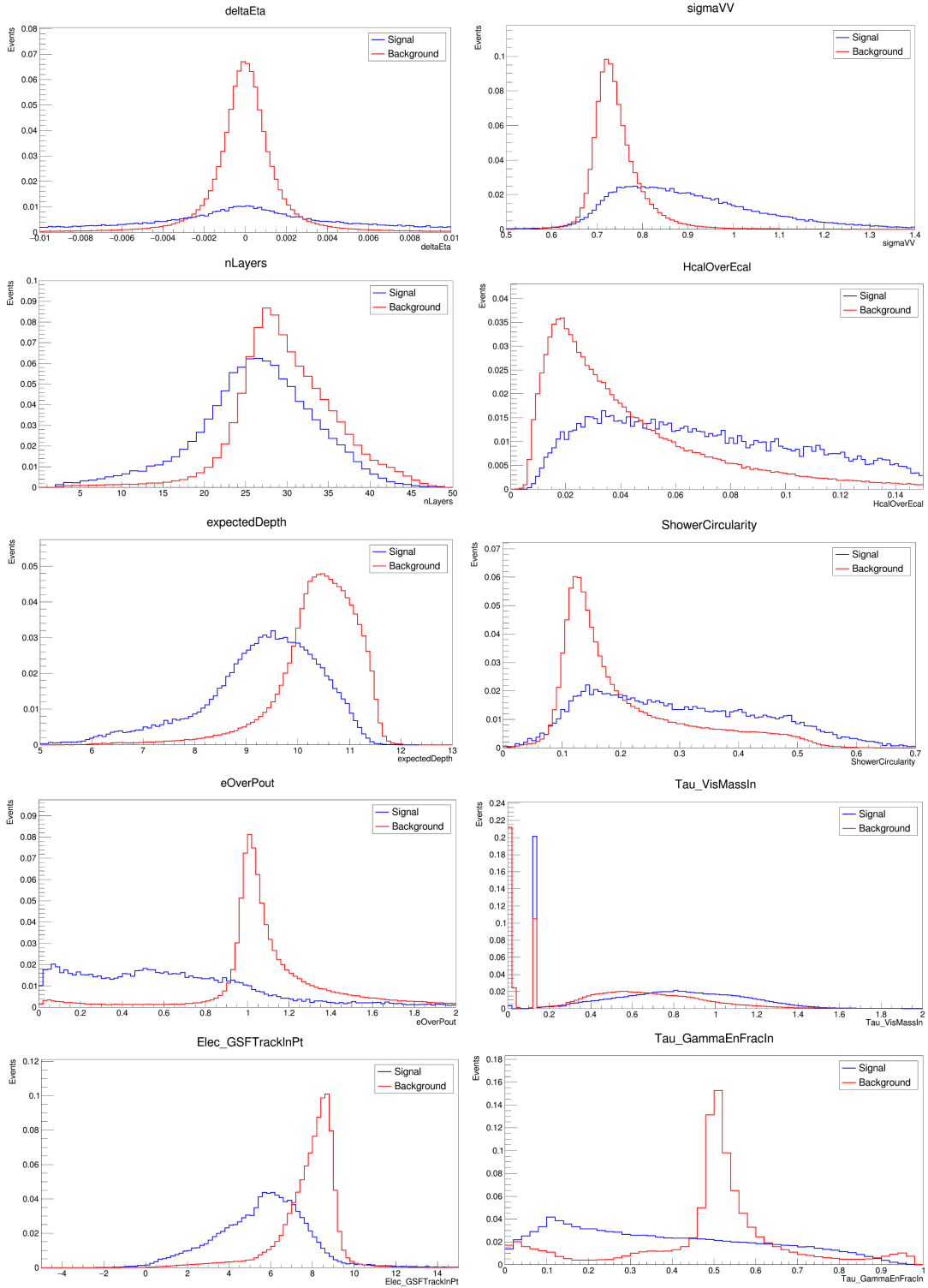


FIGURE 3.12 – Distributions de certaines variables utilisées dans les entraînements des BDT montrant la séparation du signal (bleu) et du bruit de fond (rouge) normalisés à l'unité.

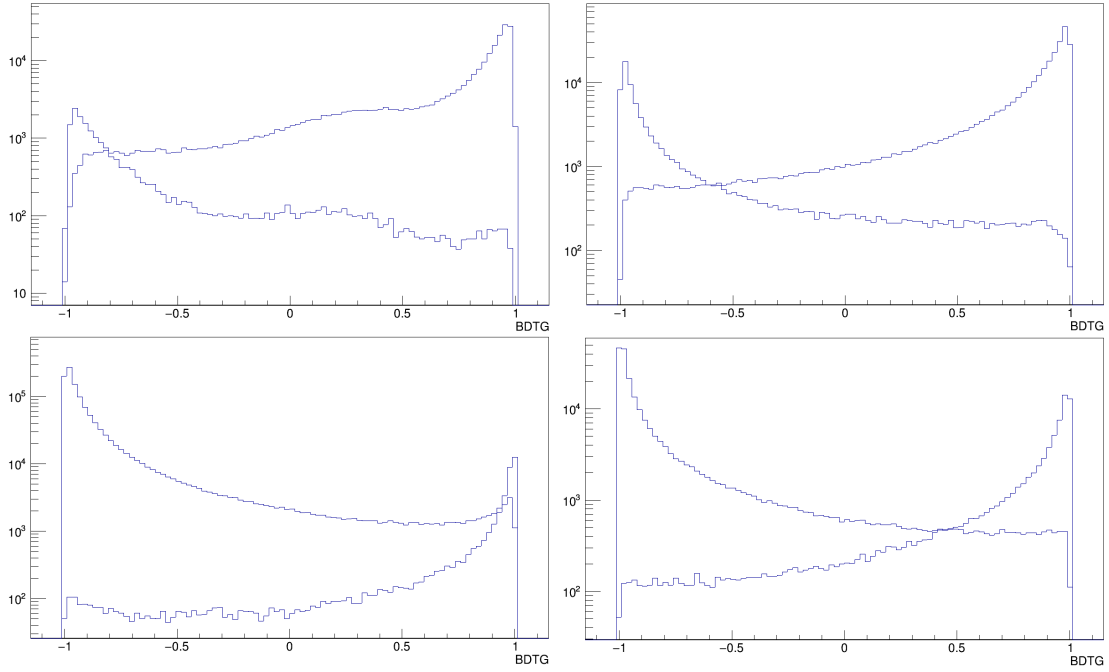


FIGURE 3.13 – Scores de sortie du BDT pour le signal et le bruit de fond dans la région  $0 < |\eta| < 1.4$  pour les quatre catégories : sans trace GSF ni photon (haut gauche), sans trace GSF mais avec photon (haut droite), avec trace GSF sans photon (bas gauche) et avec trace GSF plus photon (bas droite)

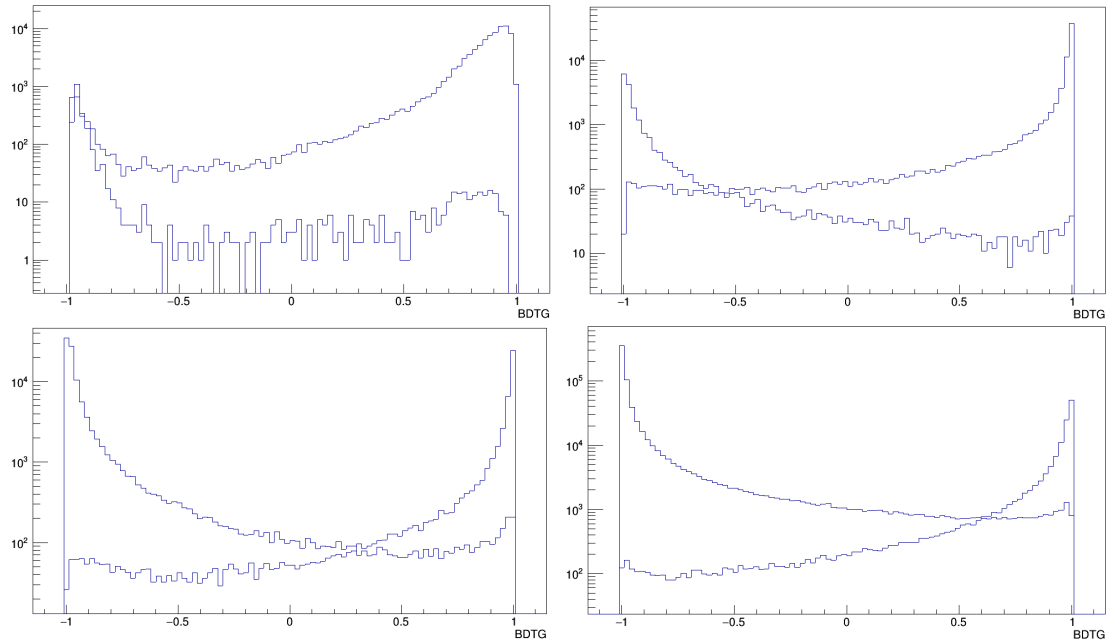


FIGURE 3.14 – Scores de sortie du BDT pour le signal et le bruit de fond dans la région  $1.6 < |\eta| < 2.4$  pour les quatre catégories : sans trace GSF ni photon (haut gauche), sans trace GSF mais avec photon (haut droite), avec trace GSF sans photon (bas gauche) et avec trace GSF plus photon (bas droite).

### Chapitre 3. Reconstruction des particules et des objets

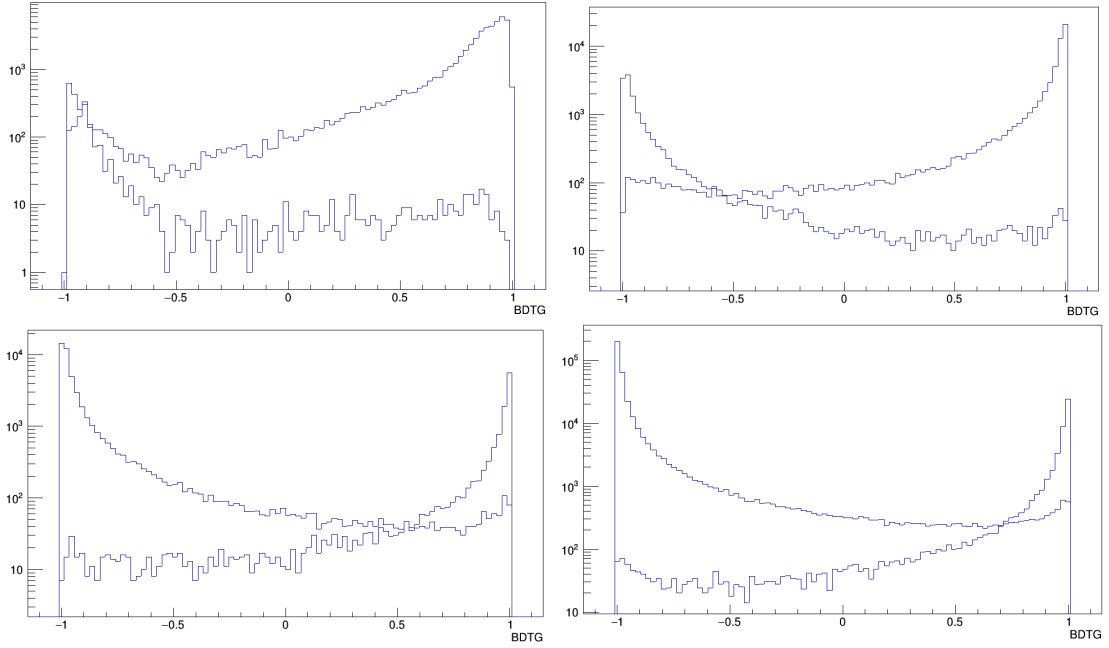


FIGURE 3.15 – Scores de sortie du BDT pour le signal et le bruit de fond dans la région  $2.4 < |\eta| < 3.0$  pour les quatre catégories : sans trace GSF ni photon (haut gauche), sans trace GSF mais avec photon (haut droite), avec trace GSF sans photon (bas gauche) et avec trace GSF plus photon (bas droite).

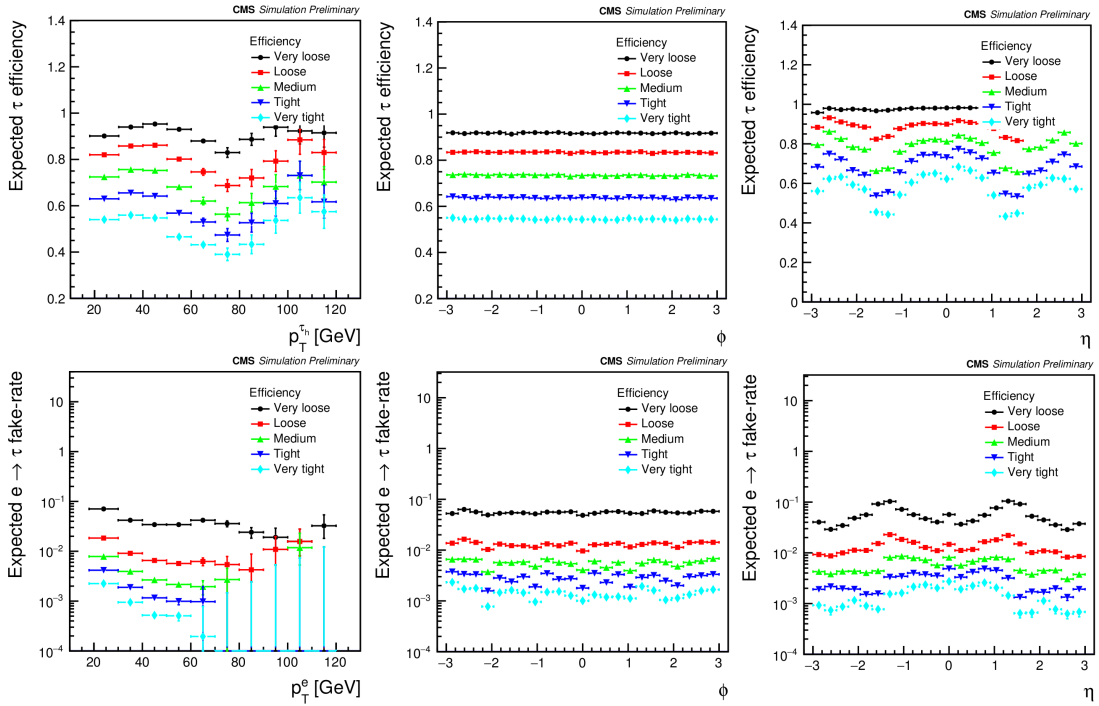


FIGURE 3.16 – Efficacité de reconstruction du tau hadronique (haut) en fonction du  $p_T$  (gauche),  $\phi$  (milieu) et  $\eta$  (droite) du tau et taux de faux (bas) en fonction du  $p_T$  (gauche),  $\phi$  (milieu) et  $\eta$  (droite) de l'électron appliqué au processus DY pour différents points de fonctionnement.

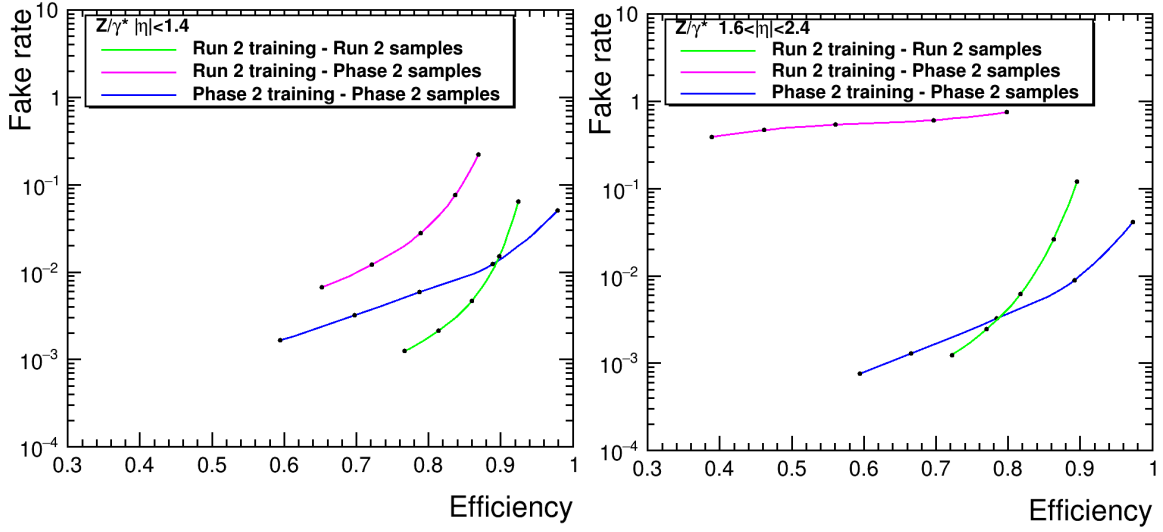


FIGURE 3.17 – Courbes ROC de l’entraînement du Run 2 appliqué à des échantillons DY du Run 2 (vert), appliqué à des échantillons DY de la Phase 2 (mauve) et entraînement pour la Phase 2 appliqué à des échantillons DY de la Phase 2 (bleu) dans les régions spatiales  $0 < |\eta| < 1.4$  (gauche) et  $1.6 < |\eta| < 2.4$  (droite).

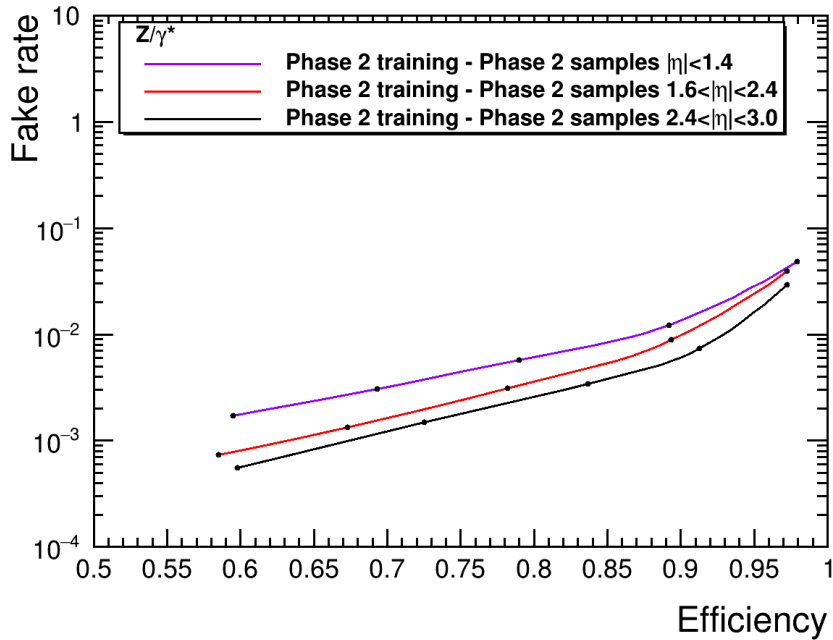


FIGURE 3.18 – Courbes ROC de l’entraînement de la Phase 2 appliqué à des échantillons DY de la Phase 2 suivant les régions spatiales  $0 < |\eta| < 1.4$  (violet),  $1.6 < |\eta| < 2.4$  (rouge) et  $2.4 < |\eta| < 3.0$  (noir).

### 3.2.6 Énergie transverse manquante

L'énergie transverse manquante ( $\vec{p}_T^{\text{miss}}$ ) est une grandeur physique très utile et est définie comme la projection sur le plan perpendiculaire au faisceau de la somme vectorielle négative de l'impulsion de toutes les particules reconstruites de l'évènement :

$$\vec{p}_T^{\text{miss}} = - \sum_i \vec{p}_T(i). \quad (3.7)$$

Sa magnitude est symbolisée par  $E_T^{\text{miss}}$ . Cette énergie représente les particules non détectées comme les neutrinos ou de potentielles particules au-delà du MS par conservation du quadrivecteur énergie-impulsion.

Cette somme vectorielle, dépendante de la reconstruction des autres particules, peut être mal estimée à cause de la non linéarité dans la réponse du calorimètre aux hadrons, du seuil d'énergie minimum dans les calorimètres et de l'inefficacité de reconstruction des traces [101]. On applique donc une correction dite de type I par la propagation des corrections sur les jets et la somme devient :

$$\vec{p}_T^{\text{miss}} = - \sum_i \vec{p}_T(i) - \sum_{\text{jets}} (\vec{p}_{T,\text{jets}}^{\text{corr}} - \vec{p}_{T,\text{jets}}), \quad (3.8)$$

avec  $\vec{p}_{T,\text{jets}}^{\text{corr}}$  et  $\vec{p}_{T,\text{jets}}$  le  $p_T$  corrigé et non corrigé des jets respectivement.

Une autre correction dite de type 0 est possible et fut mise en place pour les données 2012. La direction de l'énergie transverse manquante dépend de la direction de la somme vectorielle des particules neutres. Cette correction corrige donc la direction en prenant en compte la direction des particules chargées associées à l'empilement.

Cette énergie transverse manquante est dite PF car elle exploite les particules de l'algorithme du flux de particule. Mais il existe une autre façon d'obtenir cette grandeur physique en utilisant la méthode d'identification de l'empilement par particule (PUPPI). Cette méthode est construite sur l'algorithme CHS. En plus de cela, elle calcule la probabilité que chaque particule neutre soit originaire de l'empilement, en s'appuyant sur les informations des particules environnantes et leur association au vertex primaire, et multiplie son énergie en fonction de cette probabilité. Cette façon de faire permet de réduire encore plus la dépendance à l'empilement et est celle que j'utilise dans mon analyse.

La résolution de  $\vec{p}_T^{\text{miss}}$  est quantifiée avec la mesure de la résolution du recul hadronique, ce dernier étant défini comme la somme vectorielle des impulsions de tous les objets de l'évènement sauf le boson Z. La résolution du recul hadronique dans des évènements  $Z \rightarrow e^+e^-$  est donnée en Fig. 3.19. L'utilisation de  $p_T^{\text{miss}}$  PUPPI offre une meilleure résolution à partir de 10 vertex pour la composante longitudinale, et pour toutes valeurs pour la composante orthogonale.

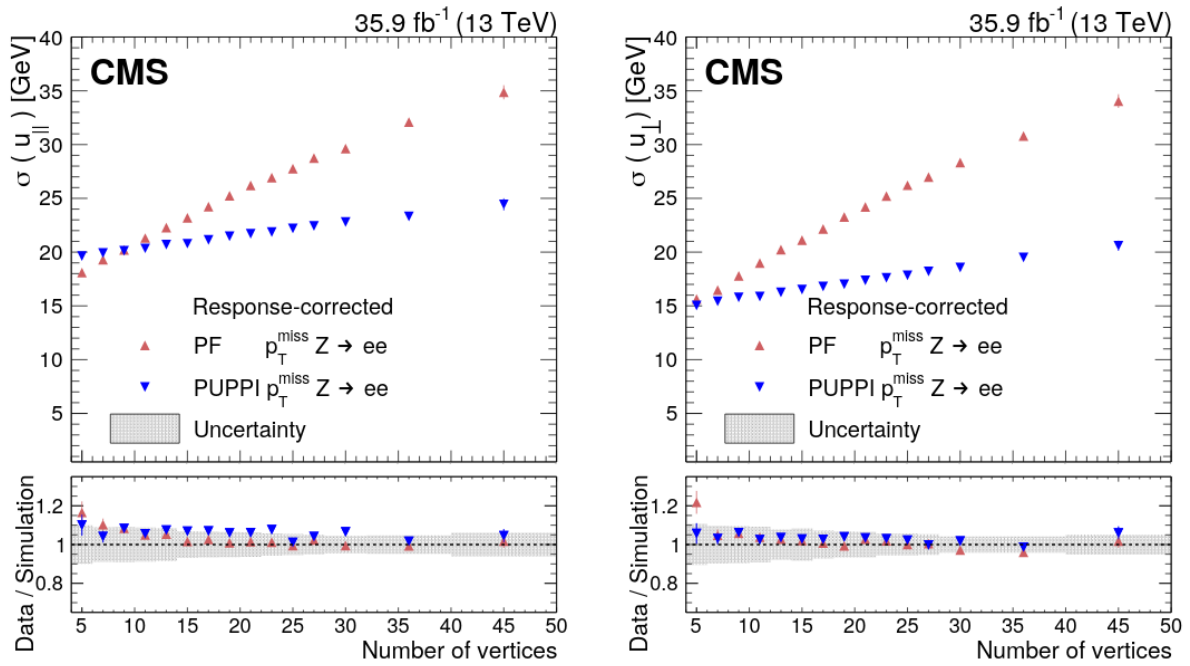


FIGURE 3.19 – Résolution des composantes  $u_{\parallel}$  (gauche) et  $u_{\perp}$  (droit) du recul hadronique pour  $\vec{p}_T^{\text{miss}}$  PUPPI et PF en fonction du nombre de vertex pour des évènements  $Z \rightarrow e^+e^-$  dans des données. Le ratio donnée/MC est donné en dessous des deux graphiques. La bande grise correspond aux incertitudes systématiques venant des variations de l'échelle et de la résolution de l'énergie des jets ainsi que de l'énergie non regroupée [102].



## Chapitre 4

# Le trajectographe en silicium et l'étude des particules hautement ionisantes

Ce chapitre expliquera tout d'abord le fonctionnement des semi-conducteurs et notamment leur utilisation dans le trajectographe en silicium de l'expérience CMS. Ensuite il sera question des particules hautement ionisantes (HIP) qui sont plus précisément des événements avec un fort dépôt d'énergie suite à une interaction nucléaire inélastique et qui affectent l'électronique de lecture induisant une inefficacité de reconstruction des impacts des particules. Je passerai en revue les études antérieures menées par CMS sur ce sujet. Pour terminer, je présenterai une étude que j'ai menée sur les données 2018 afin de savoir si l'inefficacité observée dans le trajectographe en silicium pouvait être expliquée à partir des HIP.

### 4.1 Les capteurs en physique des particules

#### 4.1.1 Les semi-conducteurs

En physique quantique des solides, la théorie des bandes modélise ces derniers avec des électrons qui ne peuvent prendre que des valeurs d'énergie comprises dans certains intervalles nommés bandes. Dans l'état au repos du système, les électrons se situent tous dans la bande de valence. De l'énergie peut être apportée par excitation thermique ou par ionisation due au passage d'une particule et faire passer des électrons de la bande de valence à la bande de conduction, ces dernières étant séparées par la bande interdite (band gap). La conductivité électrique d'un semi-conducteur dépend donc de sa structure électronique. Dans le cas d'une ionisation par une particule incidente, l'énergie déposée vient de la perte d'énergie de la particule traversant le matériau. La majorité des particules chargées détectées dans l'expérience CMS et provenant de collisions proton-proton sont faiblement énergétiques et sont appelées « particules au minimum d'ionisation » (MIP). La formule de Bethe-Bloch qui décrit l'énergie perdue  $-dE$  par

interaction électromagnétique, par distance traversée  $dx$  et par densité  $\rho$  du matériau [103] modélise bien la gamme d'énergie des MIP :

$$-\frac{1}{\rho} \frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{z^2}{\beta^2} \left[ \frac{1}{2} \ln \left( \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} \right) - \beta^2 - \frac{\delta}{2} - \frac{C}{Z} \right], \quad (4.1)$$

avec  $N_A$  le nombre d'Avogadro,  $r_e$  le rayon classique de l'électron,  $m_e$  la masse de l'électron,  $c$  la vitesse de la lumière,  $z$  la charge électrique de la particule incidente en unités de  $e$ ,  $\beta = \frac{v}{c}$  la vitesse,  $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ ,  $Z$  le numéro atomique,  $A$  la masse atomique,  $T_{\max}$  l'énergie maximale transférée en une interaction,  $I$  le potentiel moyen d'excitation du matériau qui dépend du numéro atomique,  $\delta$  le facteur de correction de la densité et  $\frac{C}{Z}$  la correction d'effet de couche qui prend en compte le fait qu'à basse énergie les électrons des couches les plus basses seront protégés.

Si suffisamment d'énergie est transférée, un électron peut traverser la bande d'énergie interdite et arriver dans la bande de conduction en laissant derrière lui un trou qui aura une charge positive par absence de cet électron. Ces mouvements d'électrons induisent un courant électrique dans le semi-conducteur.

Il y a dans un semi-conducteur dit intrinsèque, c'est-à-dire pur, autant d'électrons excités dans la bande de conduction que de trous dans la bande de valence. Une MIP traversant  $300\mu\text{m}$  ne déplacerait qu'environ  $2 \times 10^4$  porteurs de charge alors qu'un silicium intrinsèque possède à température ambiante environ  $10^9$  porteurs de charge libres. Ce signal serait donc perdu avec ce grand nombre de porteurs de charges libres et le bruit de fond associé. On pourrait y remédier en refroidissant jusqu'au zéro absolu le silicium mais ce n'est pas ce qu'il y a de plus pratique. Cependant le trajectographe est tout de même refroidi à environ  $-20^\circ\text{C}$ .

On préfère augmenter la conductivité d'un semi-conducteur en introduisant des impuretés, c'est ce qu'on appelle faire du dopage. Ces impuretés sont d'autres atomes que l'on va rajouter, et plus précisément cela ajoutera de nouveaux niveaux d'énergie autorisés. Le semi-conducteur sera impur et donc extrinsèque. On peut utiliser des atomes accepteurs, habituellement du bore, qui ont un électron de moins que ceux de silicium, et qui introduiront des trous (de type p), ou des atomes donneurs, habituellement du phosphore, qui à l'inverse ont un électron de plus et qui introduiront des électrons (de type n). Ceci aura pour effet de changer la structure électronique du matériau pour le rendre plus conducteur.

Si on joint les deux types de dopages, on crée une jonction p-n. Les électrons du côté n se déplacent vers le côté p et inversement pour les trous. Cela crée une zone de déplétion autour de l'interface de la jonction, région dépourvue de porteurs de charge

car ces derniers ont migré. Le côté n de la zone de déplétion est positivement chargé et le côté p négativement, ce qui crée un champ électrique qui s'oppose à la diffusion des charges jusqu'à atteindre un équilibre dynamique. Cette zone de déplétion peut être vue comme une barrière de potentiel empêchant le mouvement des charges libres. Or il serait préférable d'augmenter la taille de cette barrière de potentiel car des mouvements de charges sont autant de bruit de fond pour l'expérience.

Pour faire cela on utilise une polarisation inverse (reverse bias) en appliquant une tension externe entre les deux côtés avec l'anode connectée au côté p et la cathode au côté n. La zone de déplétion va alors s'agrandir avec un champ électrique plus important qui va éloigner encore plus les électrons et les trous. Cela permet de réduire le bruit de fond des capteurs même à température ambiante. Cette zone de déplétion est appelée « volume actif du détecteur » car les charges créées en dehors de cette zone se recombinent immédiatement avec leurs charges opposées et ne contribuent pas au signal. Tandis qu'au sein de la zone active, les paires électrons-trous, formées suite au passage d'une particule chargée, vont migrer vers les électrodes sous l'effet du champ électrique afin d'obtenir un signal. Il est donc important d'avoir une zone de déplétion aussi grande que le capteur. C'est le cas quand on dépasse la tension de déplétion entière (full depletion voltage,  $V_{FD}$ ) et le champ électrique peut être considéré comme constant. La tension de déplétion entière peut s'exprimer comme :

$$V_{FD} = \frac{d^2}{2 \cdot \epsilon_0 \cdot \epsilon_{Si} \cdot \mu_e \cdot \rho}, \quad (4.2)$$

avec  $d$  l'épaisseur du capteur,  $\epsilon_0$  la permittivité du vide,  $\epsilon_{Si}$  la permittivité du silicium,  $\mu_e$  la mobilité des électrons et  $\rho$  la résistivité du silicium.

### 4.1.2 Les capteurs à pistes en silicium de CMS

Le silicium est le principal matériau semi-conducteur utilisé dans les détecteurs en physique des particules. Les capteurs en silicium se sont principalement développés à la fin des années 70 [104, 105]. Leurs avantages sont un bas coût de production et des propriétés physiques très utiles. L'énergie requise pour créer une paire électron-trou dans du silicium est de 3.6 eV [106], ce qui est 10 fois inférieur à l'ionisation d'une molécule de gaz dans un détecteur à gaz. On peut encore améliorer ces capteurs avec des informations spatiales en segmentant finement la jonction p-n en pistes ou pixels avec la technique planaire [104].

Les pistes en silicium seront présentées dans la suite. Le silicium a plusieurs intérêts comme être à l'état solide dans des conditions normales de température et de pression, ce qui permet de se passer de gaz ou de liquides qui sont plus instables. En contrepartie l'état solide est sujet à la dégradation par les radiations qui vont modifier la structure interne du matériau. Un autre avantage est que l'industrie des matériaux en silicium

nous permet d'obtenir des capteurs très purs en silicium qui ont une efficacité proche de 100%, la fraction restante étant due à des régions non sensibles comme les bords ou à un phénomène de partage de charges entre pistes.

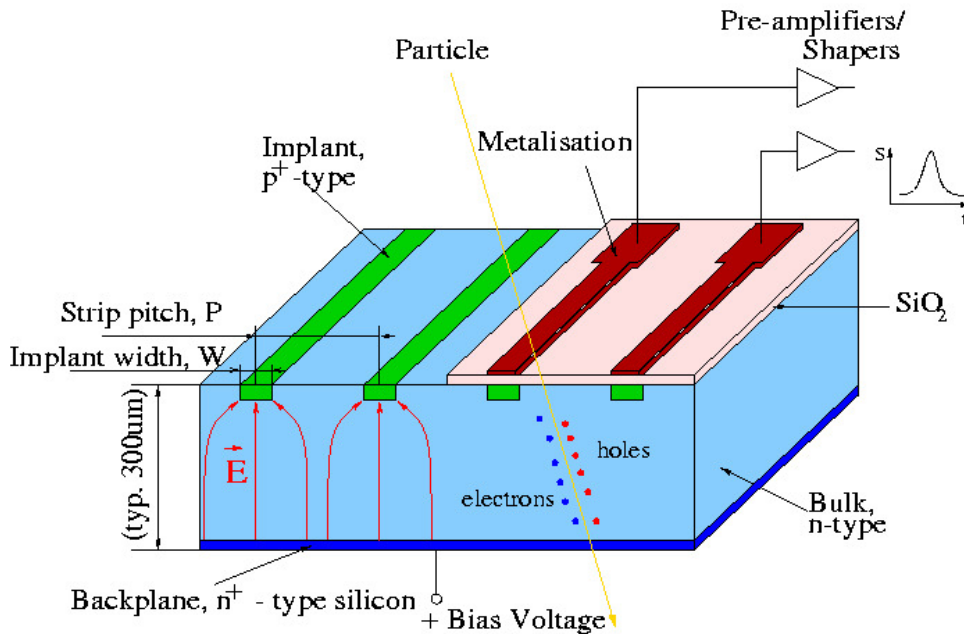


FIGURE 4.1 – Schéma d'un capteur en silicium de CMS [107].

Un schéma d'un capteur à pistes en silicium de CMS est montré en Fig. 4.1. Il est composé d'un volume de type n avec un fond  $n^+$  et de l'autre côté des pistes  $p^+$ . L'ajout d'un « + » démontre un fort dopage du semi-conducteur, tandis qu'un « - » serait le signe d'un faible dopage. Une tension de polarisation inverse est appliquée afin de complètement dépléter le volume (l'épaisseur des pistes  $p^+$  et du fond  $n^+$  est négligeable). Le signal est collecté par les pistes  $p^+$  et un courant circule alors dans les pistes en aluminium situées juste au dessus d'elles. Celles-ci sont séparées par une fine couche de dioxyde de silicium qui fait office de couplage capacitif. Cette couche est nécessaire car il existe un courant de fuite entre les côtés p et n qui rajoute du bruit. Afin de découpler ce courant, le signal est lu par couplage capacitif. Les pistes d'aluminium sont reliées à l'électronique de lecture Fig. 4.2. Ces dernières sont reliées par des groupes de 128 canaux par où transite le signal pour arriver dans des composants électroniques frontaux appelés APV25 (Analog Pipeline Voltage) que j'appellerai seulement APV dans la suite et qui vont l'amplifier, le mettre en forme et adapter l'impédance. Ces APV sont aussi équipées « d'inverter resistor » qui sont donc communs aux 128 canaux et permettent, entre autre, de stabiliser les variations entre les canaux.

Suite à la réception d'un signal de déclenchement, le signal analogue est envoyé par fibres optiques à des cartes électroniques de traitement appelées FED (Front End Driver) où il va être converti en signal numérique sur 10 bits dont l'unité sera en « Analog Digital

Counts » (ADC). Ce signal est appelé « raw digi ». Mais deux soustractions sont ensuite appliquées à ce dernier. La première est le piédestal qui est le bruit ou « signal moyen » obtenu en l'absence de signal physique. Plusieurs fois par an, pour chaque piste, le bruit et le piédestal sont mesurés durant des prises de données spéciales appelés « pedestal runs ». Ensuite le « bruit de mode commun » (CMN ou ligne de base) est le bruit de l'électronique et de l'alimentation. Il est commun à tous les canaux d'une APV car pris comme la médiane sur la population de pistes et diffère d'un évènement à un autre. Une fois les deux soustractions effectuées, les valeurs ADC négatives sont tronquées à 0 car les FED ne gèrent que les valeurs positives.

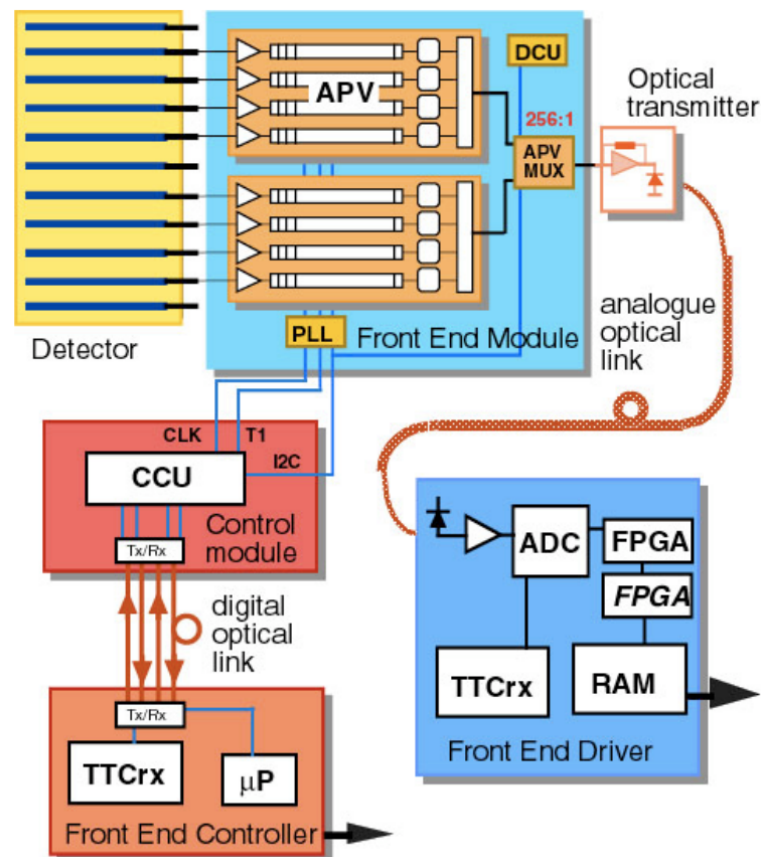


FIGURE 4.2 – Schéma de la chaîne de lecture du trajectographe [108].

## 4.2 Particules hautement ionisantes

Une haute efficacité de reconstruction des différents objets physiques dans l'expérience CMS est nécessaire. L'étude de l'efficacité de reconstruction des coups dans le trajectographe à piste en silicium dans le contexte des particules dites hautement ionisantes avec des données 2018 est présentée ci-après.

### 4.2.1 Contexte de l'étude menée

On observa, au début du Run 2, où le temps entre deux croisements de faisceau passa de 50 à 25 ns, un nombre de coups reconstruits bien plus faible que le nombre attendu. Cette forte inefficacité de reconstruction des traces était dépendante de la luminosité instantanée [109]. La cause était due à une saturation dans le pré-amplificateur de l'APV. En effet la vitesse de drainage des pré-amplificateurs faisait que la décharge des amplificateurs était trop lente. Cette vitesse fut augmentée durant le mois d'août 2016 en modifiant la paramétrisation des APV, ce qui corrigea le problème. Le graphique de gauche en Fig. 4.3 montre l'efficacité de coup, calculée comme le ratio du nombre de coups associés aux traces reconstruites sur le nombre de coups attendus, pour chaque couche du tonneau, des disques et des bouchons.

L'algorithme calculant l'efficacité de coups part des trajectoires reconstruites et vérifie si ces dernières traversent des couches avec un coup présent dit attendu. Ceci est mesuré en permettant qu'un coup soit manqué, car l'algorithme de trajectographie le permet, et en vérifiant que la couche en question est bien dans une surface utile du détecteur. Sur le graphique de droite est présentée l'efficacité de coup en fonction de la luminosité instantanée. La partition la plus touchée par l'inefficacité est le TOB et plus précisément sa première couche avec près de 8% de perte d'efficacité pour les plus hautes luminosités intégrées avec les cercles vides, ce qui correspond à l'ancienne paramétrisation des APV. Les cercles pleins rouges représentent la nouvelle paramétrisation.

Mais dans ce contexte, différentes hypothèses pouvant être la cause des problèmes observés furent étudiés et notamment un : les événements avec des particules hautement ionisantes (HIP). Cette appellation est un abus de langage et on devrait plutôt parler d'événements hautement ionisants. En effet, la perte d'énergie d'une particule dans le trajectographe en silicium se fait principalement via l'interaction électromagnétique par ionisation. Mais la particule peut aussi perdre de l'énergie dans les capteurs par interaction nucléaire élastique et inélastique avec les noyaux de silicium. La conséquence de cela est un recul nucléaire et à cela s'ajoute une fragmentation dans le cas d'une collision inélastique. Une assez grande énergie de recul peut déplacer les noyaux adjacents. La perte d'énergie de la particule incidente a pour conséquence un grand dépôt d'énergie dans les capteurs.

### 4.2.2 Études précédentes sur les HIP

Ces études ont commencé avant la construction du LHC. Des événements HIP interagissant avec un volume de silicium furent simulés avec un faisceau de pions de 200 MeV et 120 GeV [110]. Il fut montré que les interactions inélastiques dans le silicium génèrent approximativement à chaque fois un HIP pouvant être équivalent à 1000 MIP pour une épaisseur de 500  $\mu\text{m}$  de silicium. Mais plus aucun événement n'est détecté au

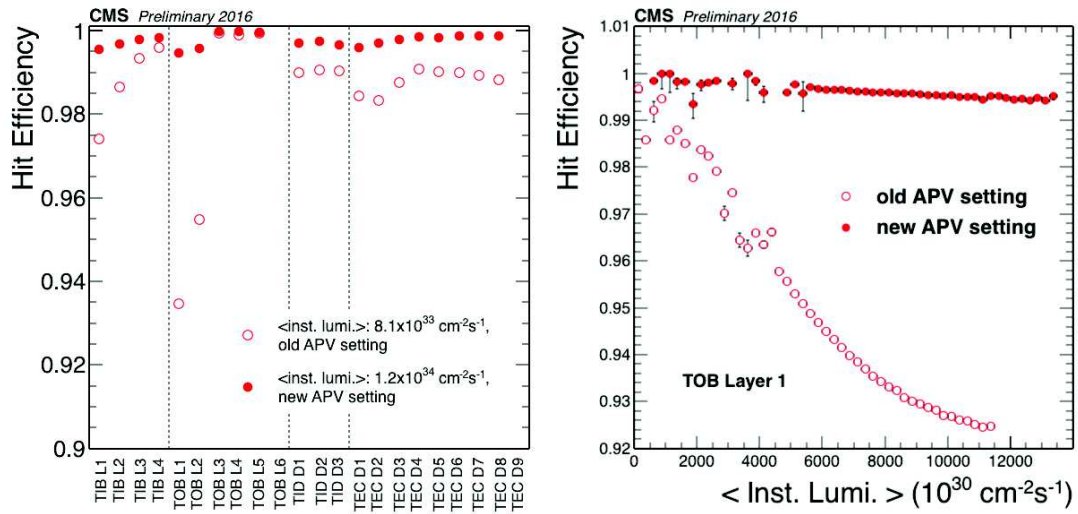


FIGURE 4.3 – (gauche) Efficacité de reconstruction des coups pour toutes les couches du tonneau, des disques et des bouchons. Les cercles vides représentent l'efficacité de coup avec l'ancienne paramétrisation de l'électronique frontale (APV) et les cercles pleins rouges représentent l'efficacité de coup avec la nouvelle paramétrisation des APV de début août 2016. La luminosité instantanée des deux runs est indiquée dans la légende. (droite) Efficacité de coups en fonction de la luminosité intégrée. Seule la première couche du TOB est montrée avec la même légende que la figure de gauche [109].

dessus de 300 MeV d'ionisation. Les fragments lourds dus aux interactions inélastiques atteignent une énergie de 20 MeV. À l'inverse, les interactions élastiques ne déposent pas assez d'énergie. Cela veut dire que les grands dépôts d'énergies laissés par les HIP sont toujours la conséquence de plusieurs particules et que les charges déposées sont partagées par plusieurs pistes.

Des données sur faisceaux test furent aussi collectées à l'Institut Paul Scherrer (PSI) et avec le faisceau X5 du CERN [111]. À PSI, des pions de 300 MeV, énergie estimée des pions dans l'expérience CMS, et un système de déclenchement aléatoire furent utilisés avec des capteurs en silicium sur des modules TIB (320  $\mu\text{m}$  d'épaisseur et 120  $\mu\text{m}$  de distance inter-pistes) ainsi que TOB (500  $\mu\text{m}$  d'épaisseur et 183  $\mu\text{m}$  de distance inter-pistes). La probabilité qu'un HIP se produise fut mesurée et se situe entre  $4 \times 10^{-4}$  et  $10 \times 10^{-4}$  pour une épaisseur de 320  $\mu\text{m}$ , et est bien d'environ  $\frac{5}{3}$  fois plus grande que pour une épaisseur de 500  $\mu\text{m}$  comme attendu.

Si les énergies déposées dépassent les capacités de l'électronique de lecture des modules du trajectographe, cela peut engendrer une saturation de l'électronique suivie d'un temps mort pour tous les canaux connectés à la même APV et donc une inefficacité de reconstruction des coups. Durant ce temps mort les signaux détectés ont une amplitude diminuée qui est insuffisante pour que le signal soit propagé dans le reste de la chaîne de

lecture. Le signal est donc en-dessous du seuil minimal pour la transmission par laser. La prise de données habituelle ne nous permet pas de détecter les HIP à cause du mode de suppression de zéro. En effet les canaux ayant des valeurs ADC négatives sont tronqués à 0, ce qui empêche l'étude des HIP qui ont pour conséquence ces valeurs négatives. Mais un autre mode de prise de données est possible et n'effectue aucune suppression ni soustraction, au prix d'une augmentation de la taille des événements : le mode « virgin raw » (données brutes, VR). Cela n'empêche pas d'effectuer les étapes habituelles de traitement par la suite. La ligne de base sera donc observée et une valeur négative sera le signe d'un temps mort. Les données de PSI furent analysées [112] et un HIP est considéré dans l'évènement si la ligne de base atteint -20 ADC ou moins.

En Fig. 4.4, tirée de l'analyse des faisceaux tests, est montrée l'évolution dans le temps de la réponse de l'APV à un évènement HIP pour six plans de modules équipant le TOB. Les ordonnées représentent la valeur ADC après soustraction des piédestaux et les abscisses les différents canaux formés par 4 APV. Sans HIP, les valeurs ADC sont autour d'une valeur nominale. Un HIP est présent dans le second module en partant du bas dans la seconde APV. À 300 ns, on y voit un large pic et un léger décalage des autres canaux de l'APV. À 350 ns, les autres canaux sont maintenant fortement décalés vers des valeurs d'ADC négatives. Les canaux ne commencent à retrouver leur capacités qu'à partir de 575 ns. Un MIP peut être observé en haut à droite dans les premiers modules en partant du haut dans la seconde APV et atteint même le dernier module à partir de 525 ns. Un second MIP apparaît à 575 ns. Tout les modules sont affectés sauf le module avec le HIP où l'APV touchée ne peut lire le signal. Ce temps pendant lequel aucun signal ne peut être lu sera appelé temps mort.

Le temps mort fut aussi mesuré avec les données PSI [113] et était calculé comme une différence entre l'efficacité de reconstruction des coups pour des APV subissant les conséquences d'un HIP ( $\epsilon_{\text{hit}}^{\text{HIP}}$ ) et des APV en fonctionnement optimal ( $\epsilon_{\text{hit}}^{\text{good}}$ ) avec  $\epsilon_{\text{hit}} = \frac{N_{\text{hit}}}{N_{\text{tracks}}}$  qui est le ratio du nombre de coups sur le nombre de traces. Le temps mort est considéré tant que l'APV n'est pas complètement opérationnelle ( $\epsilon_{\text{hit}}^{\text{HIP}} < \epsilon_{\text{hit}}^{\text{good}}$ ). Ces valeurs de temps mort sont présentées dans tableau 4.1 avec deux types de capteurs, TIB et TOB, ainsi que différentes valeurs d'« inverter resistor », 50 et 100  $\Omega$ . Différents résultats sont présents, le temps mort moyen et maximal, et le ratio entre le temps mort moyen à 50  $\Omega$  et 100  $\Omega$ . Ces mesures nous apprennent plusieurs choses. Tout d'abord le temps mort dans le TIB et dans le TOB n'est pas le même. Ce dernier a un temps mort moyen et maximal plus élevé que ce premier. Cela peut s'expliquer par la plus grande épaisseur des modules TOB qui augmente la quantité de charges collectées. On voit que la valeur de la résistance a une influence, en effet une valeur plus basse semble diminuer la durée du temps mort. En revanche cette baisse de résistance augmente le nombre de faux amas reconstruits avec une grande charge déposée et une grande largeur d'environ 10 canaux à cause d'une distorsion de la ligne de base devenant moins plate. Mais il fut



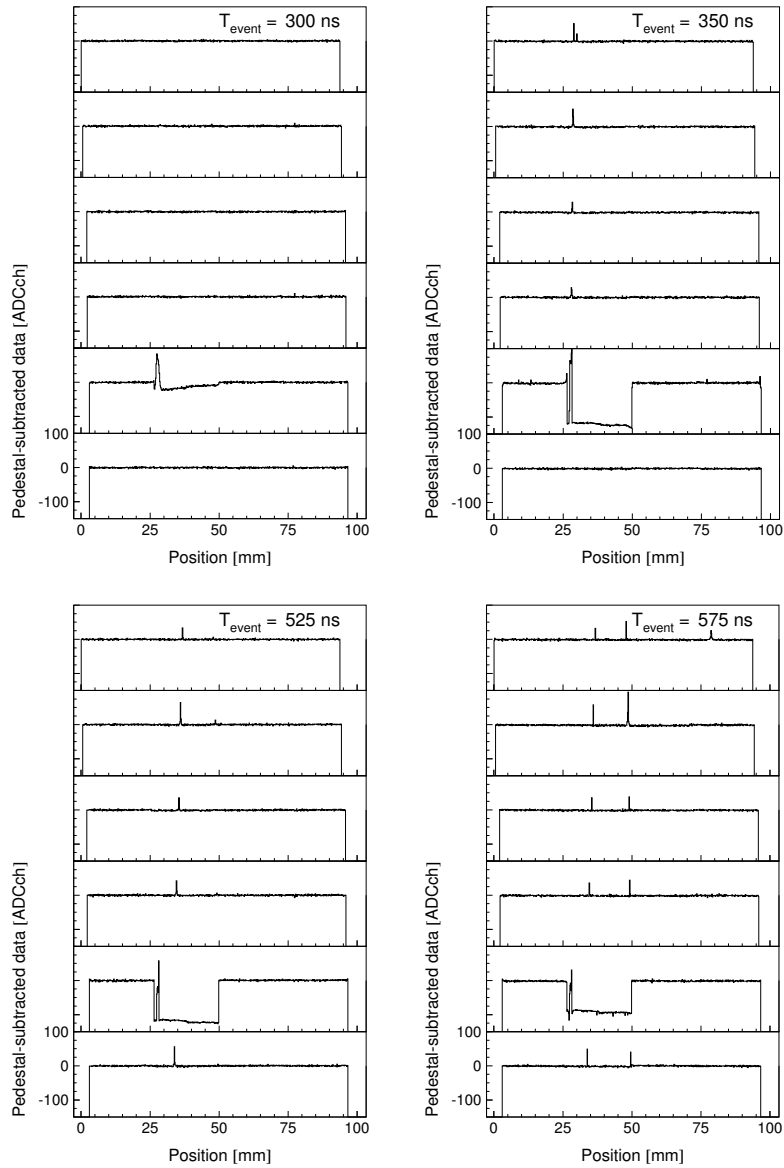


FIGURE 4.4 – Valeurs ADC après soustraction du piédestal en fonction des canaux. Quatre différents moments sont montrés (300 ns en haut à gauche, 350 ns en haut à droite, 525 ns en bas à gauche et 575 ns en bas à droite) pour six plans de modules correspondant à ceux équipant le TOB [113].

démonstré que cette augmentation était négligeable et l'utilisation de résistances à 50  $\Omega$  fut décidée afin de réduire le temps mort.

La probabilité d'obtenir un évènement HIP par trace a été mesurée à PSI :

$$P_{HIP} = \frac{N_{HIP}}{N_{tracks}}, \quad (4.3)$$

avec  $N_{HIP}$  le nombre d'évènements HIP sélectionnés avec une ligne de base en dessous de -20 ADC et  $N_{tracks}$  le nombre de traces traversant le capteur. Les résultats montrèrent

Sensor type and $R_{\text{inv}}$ [ $\Omega$ ]	$\Gamma_{\text{mean}}$ [ns]	$\Gamma_{\text{max}}$ [ns]
TIB 100 $\Omega$	$99.5 \pm 12.0$	$200 \pm 25$
TIB 50 $\Omega$	$69.6 \pm 9.4$	$250 \pm 25$
TOB 100 $\Omega$	$122.5 \pm 12.6$	$275 \pm 25$
TOB 50 $\Omega$	$100.5 \pm 3.6$	$275 \pm 25$
TIB $\Gamma_{\text{mean}}(50/100)$	$0.70 \pm 0.13$	
TOB $\Gamma_{\text{mean}}(50/100)$	$0.82 \pm 0.09$	

TABLE 4.1 – Temps mort moyen et maximal pour des modules TIB et TOB pour des résistances de 50 et 100  $\Omega$  tirés des faisceaux tests [113].

une probabilité d'environ  $10^{-3}$  pour le TOB et  $10^{-4}$  pour le TIB et que l'augmentation du flux n'augmentait pas forcément la probabilité par trace mais était due à l'épaisseur des modules qui était importante. Mais ce fut effectué avec un faisceau de pions de 300 MeV et bien que ce soit proche des conditions de l'expérience CMS, les particules et le spectre en énergie sont différents. De plus, seuls les modules TIB et TOB furent testés. C'est pourquoi une étude a été effectuée avec des données du Run 2 de CMS [112]. La probabilité fut calculée d'une manière équivalente à l'étude précédente et la mesure donne entre  $10^{-3}$  et  $10^{-2}$ , ce qui est un ordre de grandeur au dessus. Ce résultat peut s'expliquer par la mesure des traces qui se fait différemment. Une seconde façon de calculer la probabilité d'un HIP tout en étant indépendant de la reconstruction des traces a été utilisée :

$$P_{\text{HIP}}(\text{PU}) = \frac{N_{\text{HIP}}}{N_{\text{all}} \times \text{PU}}, \quad (4.4)$$

avec  $N_{\text{all}}$  le nombre total d'évènements et PU l'empilement maximal du remplissage. Ce nouveau calcul nous donne une probabilité située entre  $10^{-6}$  et  $10^{-5}$ . En Fig. 4.5 la probabilité d'apparition d'un HIP en pourcentage est donnée pour chaque couche de détection du trajectographe. Elle est donnée par interaction proton-proton, afin de ne pas être biaisée par les conditions du LHC, et par APV, car quand un HIP survient c'est l'APV entière qui en subit les conséquences. Ces probabilités furent calculées pendant une prise de données en 2018 en mode VR. On peut clairement voir une tendance pour le TOB où la probabilité diminue à mesure que l'on avance dans les couches, ce qui pourrait être relié au flux reçu.

### 4.2.3 Étude de l'inefficacité de coup dans le trajectographe en silicium avec les données 2018

La grande inefficacité observée en 2016 était trop importante pour être expliquée par des HIP et un changement du paramètre de vitesse du drainage des pré-amplificateurs des APV corrigea ce problème. Mais l'efficacité actuelle n'est toujours pas parfaite, j'ai

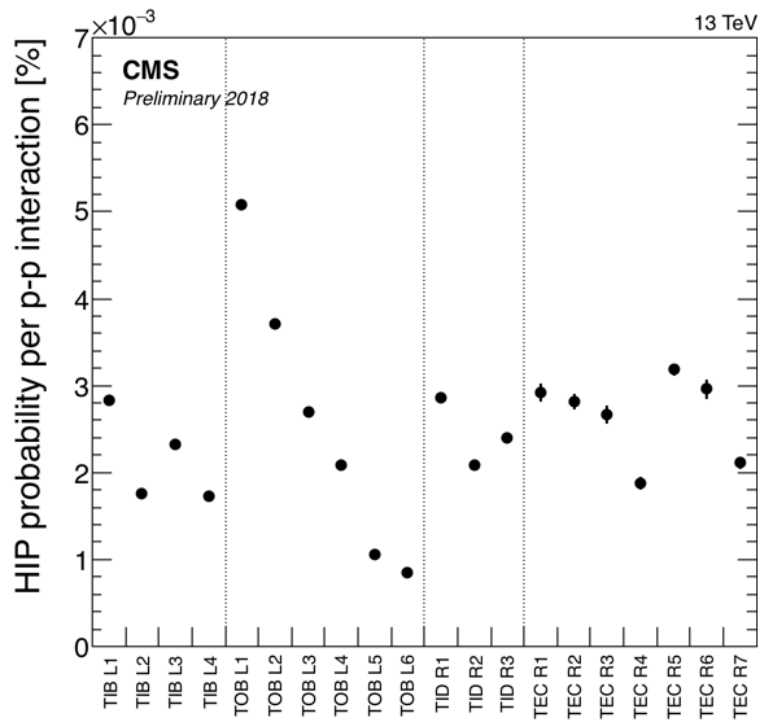


FIGURE 4.5 – Probabilité qu’un HIP survienne en pourcentage par interaction proton-proton et par APV pour chaque couche du trajectographe. Ces mesures ont été effectuées pendant une prise de données en 2018 en mode VR [114].

donc voulu vérifier si l’inefficacité restante était due aux HIP en me basant sur les probabilités estimées précédemment. L’efficacité de coup est calculée de cette manière :

- Les traces utilisées doivent passer des sélections permettant de s’assurer que ce sont de vraies traces [115] ;
- Les trajectoires passant près des bords des capteurs dans la couche en question sont exclues afin d’éviter de compter des signaux lus par deux APV, ce qui empêcherait d’avoir un calcul d’efficacité par APV. L’efficacité est mesurée comme le rapport du nombre de traces qui passent avec un coup reconstruit où qu’il soit (sauf dans les régions exclues) dans un intervalle de 15 pistes sur le nombre de traces attendues ;
- La trace doit commencer avant le module et doit se terminer après le module. Il n’y a donc pas de mesure dans les dernières couches que sont le TOBL6 et le TECD9 ;
- Les modules qui sont connus pour être défectueux et les modules ayant une efficacité très différente du reste des modules de la couche sont exclus afin d’éviter les biais ;
- Pour les couches doubles, l’efficacité moyenne des deux types de capteurs, mono et stéréo, est prise en compte.

J'ai étudié un remplissage représentatif de 2018 qui durait assez longtemps et avec un intervalle de luminosité instantanée assez important : j'ai donc choisi le fill 6714 où chaque faisceau contient 2556 paquets de protons. L'efficacité de coup des bons modules qui passent les précédentes sélections énoncées plus haut est montrée en Fig. 4.6 pour chaque couche de détection. On peut noter que l'efficacité est vraiment haute car entre 99% et plus de 99.9%. En Fig. 4.7 l'efficacité de coup en fonction de la luminosité instantanée et de l'empilement est montrée pour les différentes couches du TIB et du TOB. On peut y voir que l'efficacité de reconstruction des coups diminue linéairement avec la luminosité instantanée ou l'empilement. De plus, les différentes couches d'une même partition sont toutes décalées. En effet, les premières couches sont moins efficaces que les dernières. La différence de flux de particules peut expliquer ce comportement, les couches internes recevant plus de radiations que les externes, et un haut empilement ou une forte luminosité instantanée rend la reconstruction des coups moins efficace.

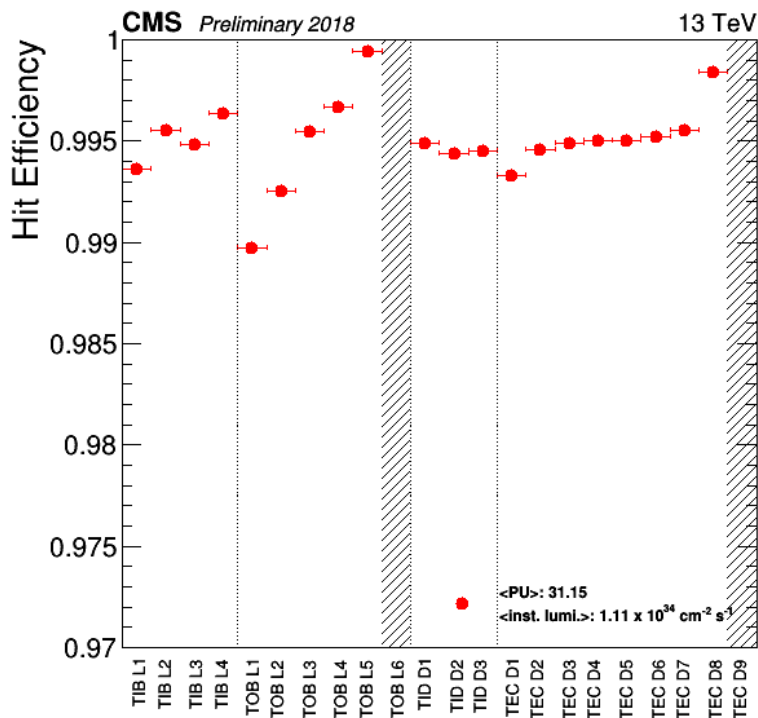


FIGURE 4.6 – Efficacité de reconstruction des coups pour le Fill 6714 pour chaque couche de détection en ne prenant en compte que les bons modules passant les critères de sélection.

Au vu de ces résultats, on a besoin de quantifier ces pentes. J'ai donc réalisé un ajustement linéaire des inefficacités en fonction de l'empilement pour chaque couche de détection (Fig. 4.8). En rouge l'ordonnée à l'origine, équivalente à un empilement nul, et en bleu la valeur de la pente. Premièrement on peut noter que les pentes et les ordonnées à l'origine suivent, globalement, les mêmes tendances, tout comme avec la Fig. 4.6. L'extrapolation à un empilement nul nous donne des valeurs d'inefficacités

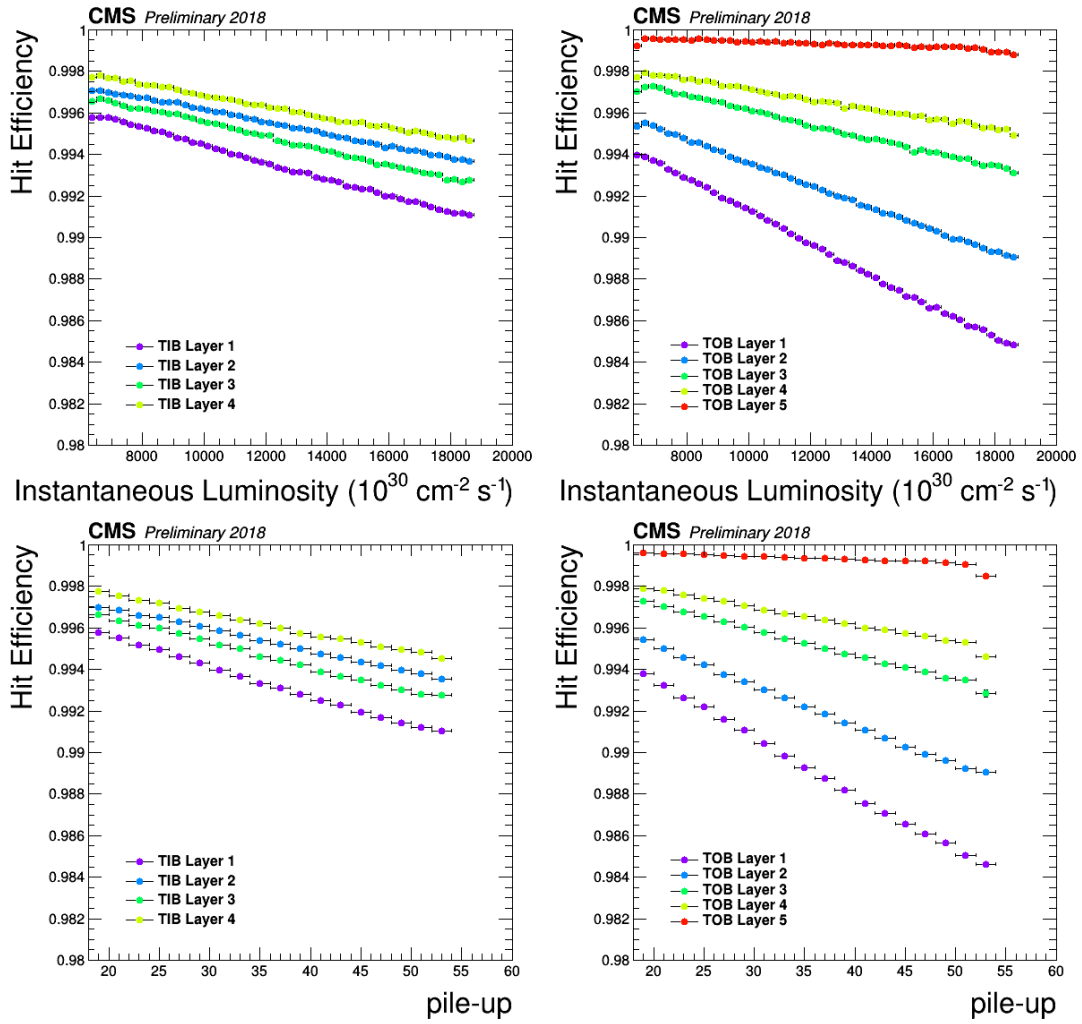


FIGURE 4.7 – Efficacité de coup en fonction de la luminosité instantanée (haut) et en fonction du PU (bas) avec les données 2018 du fill 6714. Les différentes couches du TIB (gauche) et du TOB (droite) sont montrées.

entre 0.2% et moins de 0.02%. On peut aussi voir une claire tendance dans le TOB où la pente diminue à mesure que l'on s'éloigne des couches externes. Il serait intéressant de comparer ces mesures avec les probabilités de la Fig. 4.5. Pour cela on a besoin de définir théoriquement l'inefficacité qui devrait être induite par les HIP en prenant en compte ces probabilités. La probabilité  $p(\geq 1\text{HIP})$  d'obtenir au moins un HIP par APV pour un empilement donné est :

$$\begin{aligned}
 p(\geq 1\text{HIP}) &= p(1\text{HIP}) + p(2\text{HIP}) + \dots \\
 &= C_1^N \times p \times (1 - p)^{N-1} + o(p^2) + \dots \\
 &\approx N \times p \\
 &\approx N \times p(\text{HIP}/\text{PU}) \times \text{PU},
 \end{aligned}
 \tag{4.5}$$

avec  $N$  le temps mort exprimé en nombre de croisement de paquets de protons (bx),  $p$  la probabilité d'avoir un HIP par APV pour un empilement donnée et  $PU$  l'empilement. La probabilité d'avoir au moins 2 HIP dans moins de  $N$  bx est négligée. On estime aussi que  $(1 - p)^{N-1} \approx 1$  au premier ordre car  $p$  est faible.  $p(\text{HIP}/PU)$  étant la probabilité indiquée en Fig. 4.5.

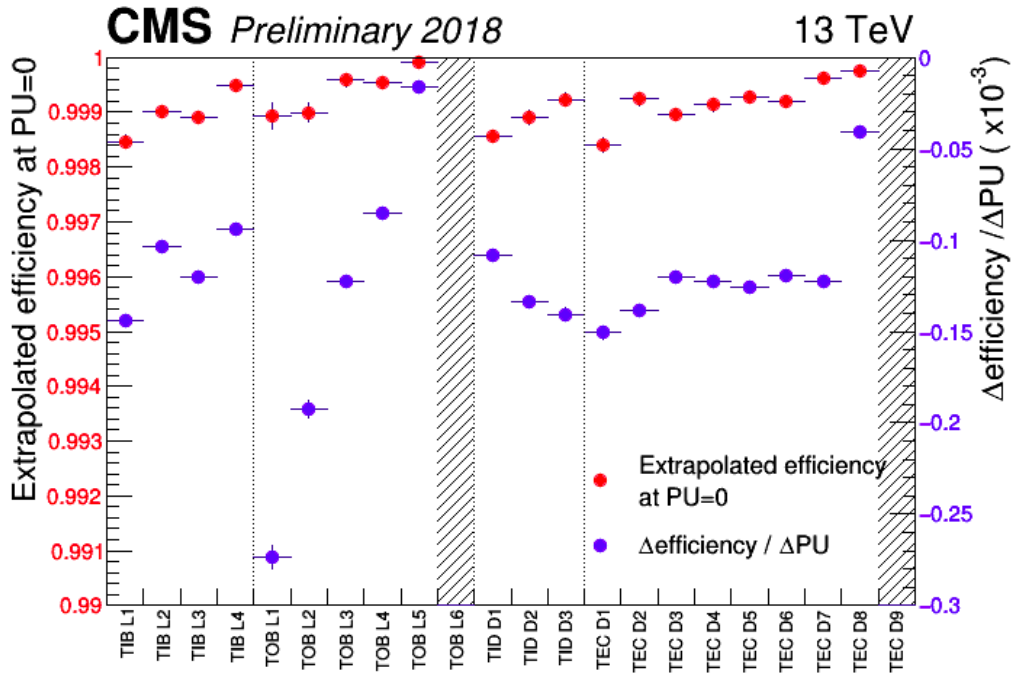


FIGURE 4.8 – Ordonnées à l'origine (rouge) et pentes (bleu) des ajustements linéaires effectués sur les efficacités de coup en fonction du PU dans les données 2018 avec le fill 6714. Les bandes grisées sont les couches où aucune mesure n'est faite comme expliqué dans les critères de sélection.

L'inefficacité observée dans les données,  $1 - \epsilon$ , pour un empilement donné est :

$$1 - \epsilon = 1 - s \times PU - d^{\text{ajust}}, \quad (4.6)$$

avec  $d^{\text{ajust}}$  l'ordonnée à l'origine et  $s$  la pente négative venant de la paramétrisation en Fig. 4.8. Dans l'hypothèse où l'inefficacité observée a pour seule cause les HIP, on peut établir cette relation :

$$\begin{aligned} 1 - \epsilon &= p(\geq 1\text{HIP}), \\ 1 - s \times PU - d^{\text{ajust}} &\approx N \times p(\text{HIP}/PU) \times PU. \end{aligned} \quad (4.7)$$

Or même en l'absence totale de HIP, l'efficacité n'est pas de 1, on peut donc définir un décalage. L'ordonnée à l'origine de l'ajustement linéaire peut être choisie et correspond à l'extrapolation des données pour un empilement nul ( $d^{\text{ajust}}$ ). Mais on peut aussi choisir

un décalage mesuré dans des données à très bas empilement ( $d^{\text{basPU}}$ ). J'ai donc essayé les deux possibilités et pour définir  $d^{\text{basPU}}$ , j'ai choisi l'efficacité des bons modules du fill 6891 qui a une luminosité instantanée maximale de  $0.006 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  et un empilement de 0.297. Les deux possibilités testées sont donc :

$$d^{\text{ajust}} - s \times \text{PU} - d^{\text{ajust}} = -s \times \text{PU}, \quad (4.8)$$

$$d^{\text{basPU}} - s \times \text{PU} - d^{\text{ajust}}. \quad (4.9)$$

Je compare en Fig. 4.9 les inefficacités observées à partir des données (marqueurs noirs) et les inefficacités prédites à partir des probabilités de HIP (marqueurs rouges) pour chaque couche du trajectographe. La figure de gauche utilise l'équation 4.8 et celle de droite l'équation 4.9. L'empilement est posé arbitrairement à 31.15 car c'est l'empilement moyen du fill 6714. Le temps mort  $N$  est un paramètre libre qui varie afin d'obtenir le meilleur ajustement possible. Mais afin de prendre en compte les différences entre les différentes partitions comme la géométrie, le paramètre  $N$  sera commun au sein d'une même partition sauf le TEC, qui possède deux épaisseurs de modules différentes, où  $N$  sera indépendant entre TEC R1-R4 et TEC R5-R6. Les différentes valeurs du temps mort mesuré sont en tableau 4.2 avec les deux types de décalages utilisés. Les partitions en bleu ont des modules de  $320 \mu\text{m}$  d'épaisseur tandis que celle en vert ont des modules de  $500 \mu\text{m}$  d'épaisseur. On peut tout d'abord dire qu'il est difficile de choisir entre l'un ou l'autre des décalages mais l'accord entre l'efficacité estimée à partir des probabilités de HIP et celle à partir des données de CMS sont en accord à 20% près. De plus les temps morts estimés pour une épaisseur de module donnée sont très proches et globalement ces mesures de temps mort sont du même ordre de grandeur que les précédentes études (voir tableau 4.1).

Partitions	Temps mort (bx)	
	$d^{\text{ajust}}$	$d^{\text{basPU}}$
TIB	4.69	6.26
TOB	4.15	5.22
TID	4.38	5.94
TEC R1-R4	4.49	5.91
TEC R5-R6	3.24	4.12

TABLE 4.2 – Tableau montrant le paramètre libre  $N$  en nombre de paquets de protons se croisant (bx) pour chaque partition et pour les deux décalages utilisés. Les partitions en bleu ont des modules d'une épaisseur de  $320 \mu\text{m}$  et les partitions en vert ont des modules d'une épaisseur de  $500 \mu\text{m}$ . Les erreurs de l'ajustement linéaire et de l'ajustement du paramètre libre ont été négligées.

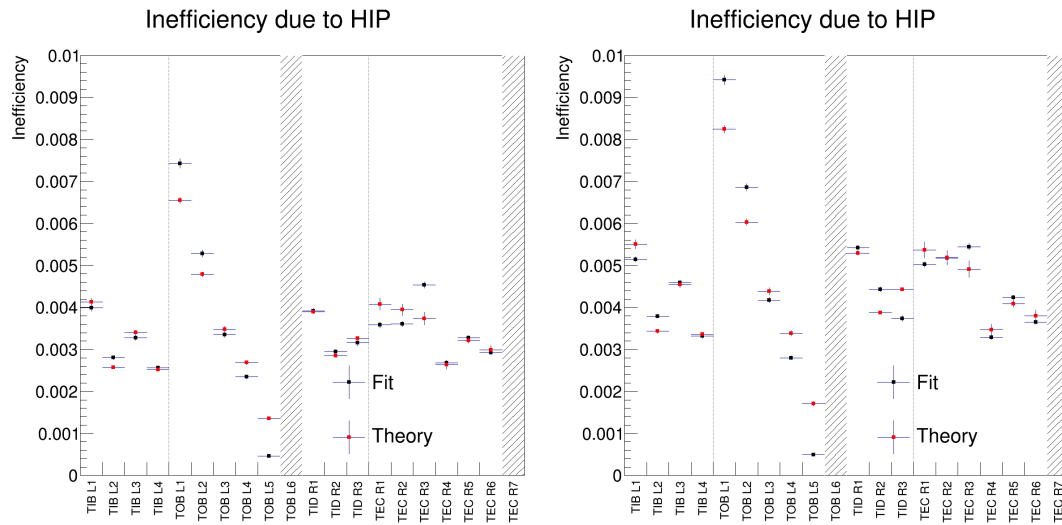


FIGURE 4.9 – Inefficacités dans le trajectographe en silicium mesurées en fonction des données (marqueurs noirs) et calculées à partir des probabilités (marqueurs rouges) pour chaque couche avec un décalage pris comme l'extrapolation à empilement nul de l'ajustement linéaire (gauche) et un décalage pris comme l'efficacité des bons modules du fill 6891 à très bas empilement.

### 4.3 Conclusion

Au début du Run 2 une inefficacité de reconstruction des coups importante, jusqu'à 8% pour la première couche du TOB, fut observée. Les particules hautement ionisantes (HIP) étaient une cause probable, or on découvrit que les inefficacités observées étaient trop importantes et que la raison était un choix non optimal des paramètres des APV. Mais même cette correction ne rend pas les modules parfaitement efficaces. Il était donc intéressant de connaître la cause des inefficacités restantes, comprises entre 0% à 2%, et vérifier si cela pouvait s'expliquer par les HIP.

Pour cela j'ai étudié le fill 6714 qui est représentatif des données 2018 en paramétrisant la dépendance de l'efficacité de coup dans les bons modules en fonction de l'empilement. On part du principe que les HIP sont la seule source d'inefficacité observée, or pour un empilement nul nous n'avons pas une efficacité de 1. J'ai donc eu besoin de rajouter un terme d'inefficacité sans empilement. J'en ai testé deux différents (Eq. 4.8 et Eq. 4.9), le premier étant l'ordonnée à l'origine de notre paramétrisation linéaire, le second étant l'efficacité dans les bons modules du fill 6891 car ayant un empilement de 0.297 donc presque nul.

J'ai ensuite comparé l'inefficacité dans les données avec l'inefficacité prédite par les probabilités qu'un HIP survienne par interaction proton-proton et par APV pour chaque couche de détection. L'empilement a été posé arbitrairement et le temps mort  $N$  a été



laissé comme paramètre libre pour chaque partition ayant une même épaisseur de module. Il en ressort un accord de l'ordre de 20% et des temps morts estimés en accord avec de précédentes études sur les HIP. C'est la première fois que l'on met en évidence le fait que les HIP sont la principale source d'inefficacité de coups restante dans le trajectographe en silicium.

Cependant les points de mesures ne se situent pas dans les barres d'erreurs et nous ne pouvons pas savoir quel décalage est le plus adapté entre  $d^{adjust}$  et  $d^{basPU}$  (voir Eq. 4.8 et 4.9) car le modèle utilisé est très simplifié. En effet, nous supposons que les HIP sont la seule source d'inefficacité observée, nous négligeons la probabilité qu'il y ait au moins 2 HIP par APV pendant un temps mort  $N$  et les différences observées peuvent également être dues aux épaisseurs et distances inter-pistes des modules ainsi qu'à l'angle d'incidence de la particule. On peut observer la dépendance des probabilités de HIP en fonction du volume sensible des modules en Fig. 4.10. Cette division permet d'effacer les différences observées induites par les différentes géométries des modules (épaisseurs et distances inter-pistes). La probabilité semble diminuer en fonction du rayon par rapport au point d'interaction. De plus le TIBL1 et le TIDR1 ont une probabilité semblable et sont presque alignés. Cette différence en fonction du rayon pourrait s'expliquer par le flux reçu qui est plus important pour les couches proches du point d'interaction que pour celles plus éloignées. Ce flux plus important a pour conséquence plus d'interactions et donc plus de chance d'avoir un HIP.

Une étude plus approfondie avec un modèle plus complexe permettrait de mieux quantifier cela, comme la prise en compte du remplissage de l'anneau du LHC. Plusieurs résultats que j'ai obtenus et qui sont présentés dans ce chapitre sont rassemblés dans une note (CMS-DP-2018-052) [116] et sont maintenant publics.

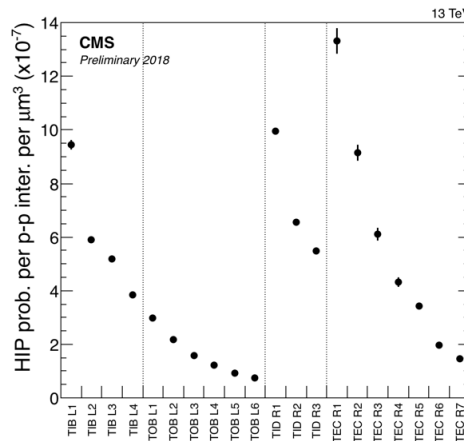


FIGURE 4.10 – Probabilité qu'un HIP survienne en pourcentage par interaction proton-proton, par APV et par volume sensible en  $\mu\text{m}^3$  pour chaque couche [114].

## Chapitre 5

# Mesure de l'état CP du boson de Higgs à $\sqrt{s} = 13$ TeV

Cette analyse a pour but de mesurer un éventuel état CP-mixte du boson de Higgs dans le couplage de Yukawa au lepton tau grâce à l'application d'une méthode originale et prometteuse appelée méthode du vecteur polarimétrique dans le cas où les deux taus donnent trois pions chargés via une résonance  $a_1^{3pr}$  (canal  $a_1^{3pr} a_1^{3pr}$ ). Elle s'insère dans le contexte de la première analyse CMS effectuée à l'été 2020 et dont les résultats ont été présentés dans un « Physics Analysis Summary » (PAS) [117]. Celui-ci s'appuie sur les données du Run 2 ( $137 \text{ fb}^{-1}$ ) et les canaux considérés sont  $\tau_\mu \tau_h$  et  $\tau_h \tau_h$ . Les désintégrations du  $\tau_h$  utilisées sont un pion chargé, une résonance  $\rho$  ou encore une résonance  $a_1$ . Mon travail de thèse consiste en l'étude du canal  $a_1^{3pr} a_1^{3pr}$  avec la méthode du vecteur polarimétrique qui remplacera la méthode du plan de désintégration présente dans le PAS pour le même canal dans une publication en cours.

Le chapitre présente dans un premier temps les méthodes habituellement utilisées pour mesurer l'angle de mélange CP du boson de Higgs pour ensuite présenter l'analyse menée dans le canal  $a_1^{3pr} a_1^{3pr}$ .

## 5.1 Méthodes

Plusieurs méthodes peuvent être utilisées pour différents modes de désintégrations en définissant différents plans sensibles aux composantes transverses des spins des taus afin de mesurer l'angle dit acoplanaire  $\phi_{CP}$  entre ces derniers. Cette observable fut initialement introduite pour des collisionneurs  $e^+e^-$  où l'impulsion du lepton tau peut être entièrement reconstruite.

### 5.1.1 Méthode du paramètre d'impact

La méthode du paramètre d'impact peut être utilisée pour les désintégrations en une particule chargée (1pr) en exploitant le temps de vie fini du tau. Ce paramètre  $\mathbf{j}^\pm$  est défini comme le vecteur entre le vertex primaire (vertex d'interaction proton-proton,

PV) et le point d'approche le plus près (PCA) mesuré entre le vertex primaire et la trace extrapolée de la particule chargée. Le plan de désintégration est alors défini par le vecteur du paramètre d'impact et la direction de l'impulsion de la particule chargée. Mais la sensibilité de l'observable, définie comme étant l'angle entre les deux plans (angle acoplanaire,  $\phi_{CP}$ ) illustré en Fig. 5.1, à l'état CP du boson de Higgs s'améliore si on se place dans le référentiel où la somme des impulsions des particules chargées provenant des désintégrations du boson de Higgs est nulle (ZMF, signalé avec \*). Pour cela, on construit les quadrivecteurs  $\lambda^\pm = (0, \hat{\mathbf{j}}^\pm)$  dans le référentiel du laboratoire avec  $\hat{\mathbf{j}}^\pm$  le vecteur paramètre d'impact unitaire. Ces quadrivecteurs sont boostés dans le ZMF tout comme l'impulsion du pion chargé qui deviennent  $\lambda^{*\pm}$  et  $q^{*\pm}$  respectivement.  $\hat{\lambda}_\perp^{*\pm}$  est ensuite défini comme la composante transverse de  $\lambda^{*\pm}$  par rapport à  $q^{*\pm}$ . On peut construire, grâce à ces vecteurs, les angles  $O^*$  et  $\phi^*$  comme :

$$\begin{aligned}\phi^* &= \arccos(\hat{\lambda}_\perp^{*+} \cdot \hat{\lambda}_\perp^{*-}), \\ O^* &= \hat{q}^{*-} \cdot (\hat{\lambda}_\perp^{*+} \times \hat{\lambda}_\perp^{*-}).\end{aligned}\tag{5.1}$$

À partir de  $O^*$  et  $\phi^*$  on reconstruit  $\phi_{CP}$  dans l'intervalle  $[0, 2\pi[$  comme :

$$\phi_{CP} = \begin{cases} \phi^* & \text{si } O^* \geq 0 \\ 2\pi - \phi^* & \text{si } O^* < 0 \end{cases}.\tag{5.2}$$

Cette méthode dépend particulièrement de la reconstruction du vertex primaire et est applicable à tous les modes de désintégrations reconstruits du tau :

- Pour  $\tau^\pm \rightarrow \pi^\pm \nu$ , la sensibilité est optimale ;
- Pour  $\tau^\pm \rightarrow \rho^\pm \nu \rightarrow \pi^\pm \pi^0 \nu$ , différentes régions de l'espace des phases de la désintégration contribuent avec un signe différent dans la corrélation spin-impulsion, ce qui mène à des interférences négatives. Cela peut se résoudre en rejetant les pions chargés ayant une impulsion transverse de moins de 40 GeV. Cependant les pions au dessus de ce seuil ont un pouvoir de résolution de spin dégradé par rapport à la désintégration en un seul pion chargé et la statistique est faible ;
- Pour  $\tau^\pm \rightarrow a_1^\pm \nu \rightarrow \pi^\pm \pi^0 \pi^0 \nu$  (avec une résonance notée  $a_1^{1pr}$  par la suite), le même problème est présent que pour la résonance  $\rho$  avec un pouvoir de résolution de spin encore plus faible ;
- Pour  $\tau^\pm \rightarrow a_1^\pm \nu \rightarrow \pi^\pm \pi^\pm \pi^\pm \nu$  (avec une résonance notée  $a_1^{3pr}$  par la suite), la méthode est aussi applicable malgré les trois pions chargés. Cependant, au lieu d'extrapoler une trace pour obtenir le vecteur paramètre d'impact, on le reconstruit comme la différence entre le vertex primaire et secondaire (point où le tau se désintègre, SV). De plus, l'impulsion du pion est remplacée par l'impulsion de la somme des pions chargés. En revanche, la contribution d'un  $a_1^{3pr}$  avec une polarisation différente rend ce canal quasi non sensible avec cette méthode ;

- Pour les désintégrations leptoniques, deux régions de l'espace des phases contribuent avec une phase différente. Une coupure similaire à celle citée plus haut est applicable à 20 GeV au lieu de 40 GeV. Mais le pouvoir de résolution de spin est aussi diminué. De plus, la région au dessus de 20 GeV a une phase qui change de signe par rapport à la désintégration en un seul pion chargé.

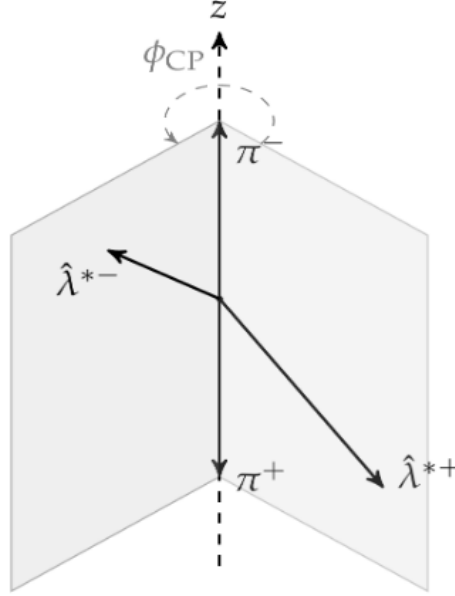


FIGURE 5.1 – Illustration de l'angle  $\phi_{CP}$  avec la méthode du paramètre d'impact avec des vecteurs 3D [118].

### 5.1.2 Méthode du plan de désintégration

Cette méthode peut s'appliquer aux modes de désintégrations ayant plus d'un hadron. Au lieu d'utiliser le paramètre d'impact, dans le cas d'un hadron chargé présent avec un ou plusieurs pions neutres, l'impulsion du pion neutre ou la somme des impulsions des deux pions neutres est utilisée pour estimer  $\lambda$ . La façon de calculer l'angle acoplaire est analogue à la méthode du paramètre d'impact et les pions sont boostés dans le ZMF. Cependant, afin d'éviter les interférences négatives venant d'états différemment polarisés, une nouvelle variable doit être définie :

$$\begin{aligned}
 y^{\tau^-} &= \frac{E_{\pi^-} - E_{\pi^0}}{E_{\pi^-} + E_{\pi^0}}, \\
 y^{\tau^+} &= \frac{E_{\pi^+} - E_{\pi^0}}{E_{\pi^+} + E_{\pi^0}}, \\
 y^{\tau} &= y^{\tau^-} y^{\tau^+},
 \end{aligned} \tag{5.3}$$

avec  $E_{\pi^\pm}$  et  $E_{\pi^0}$  l'énergie des pions chargés et neutres respectivement dans le référentiel au repos des  $\tau^\pm$  respectifs. Si  $y^\tau$  est négatif,  $\phi_{\text{CP}}$  est obtenu avec  $2\pi - \phi_{\text{CP}}$ . la méthode est illustrée en Fig. 5.2. Elle peut s'appliquer à toutes désintégrations avec une résonance  $\rho$  ou  $a_1$ . Contrairement à la méthode du paramètre d'impact, elle ne dépend pas de la reconstruction du vertex primaire. Tout comme la méthode du paramètre d'impact, la sensibilité dépend du mode de désintégration :

- Avec une résonance  $\rho$ , la sensibilité au niveau générateur est comparable à celle obtenue avec la méthode du paramètre d'impact dans le canal leptonique. Cependant, la dépendance en impulsion transverse des particules chargés produites dans l'état final est négligeable et donc aucune coupure n'est nécessaire ;
- Avec une résonance  $a_1^{1\text{pr}}$  donnant un pion chargé et deux neutres, les mêmes arguments que pour le  $\rho$  sont valables à l'exception près que le pouvoir de résolution est bien plus bas ;
- La méthode est aussi applicable pour la désintégration en 3 pions chargés via une résonance  $a_1^{3\text{pr}}$ . La paire de pions de charges électriques opposées ayant la masse invariante la plus proche du  $\rho$  neutre est sélectionnée (voir Fig. 5.3). Puis le pion de la paire de charge électrique opposée au tau est considéré comme un pion neutre et seul l'impulsion du pion restant de la paire est utilisée pour calculer le ZMF. La méthode du plan de désintégration pour une seule particule chargée est alors appliqué.

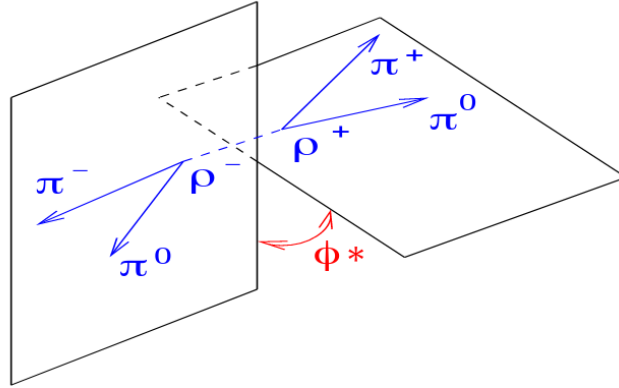
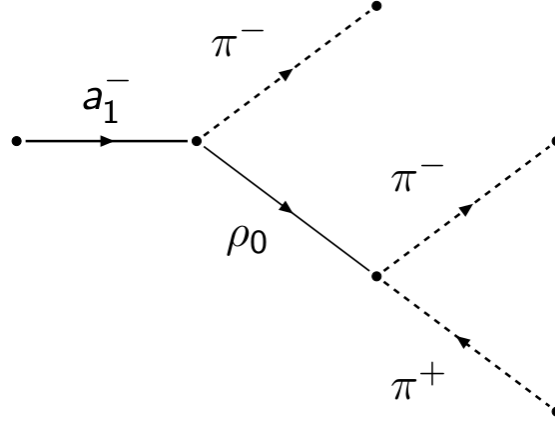


FIGURE 5.2 – Illustration de l'angle  $\phi^*$  avec la méthode du plan de désintégration [119].

### 5.1.3 Méthode du vecteur polarimétrique

Afin de caractériser les corrélations de spin transverses auxquelles l'angle CP du boson de Higgs est sensible, l'utilisation de ce qu'on appelle le vecteur polarimétrique est possible. Cette méthode a été proposée par Vladimir Cherepanov en 2017-2018 à l'Institut Pluridisciplinaire Hubert Curien (IPHC) et a fait l'objet d'une publication


 FIGURE 5.3 – Chaîne de désintégration du  $a_1^{3pr}$  donnant trois pions chargés.

de phénoménologie [120] ainsi que d'un code calculant ce vecteur polarimétrique. Le vecteur polarimétrique  $\vec{h}$  d'un tau est donné par les impulsions des hadrons et des taus produits et par le mode de désintégration :

$$\begin{aligned}\vec{h}^\pi &= -\vec{n}_\pi, \\ \vec{h}^\rho &= m_\tau \frac{2(qN)\vec{q} - q^2\vec{N}}{2(qN)(qP) - q^2(NP)},\end{aligned}\tag{5.4}$$

avec  $\vec{h}^\pi$  et  $\vec{h}^\rho$  les vecteurs polarimétriques pour une désintégration en un pion chargé et via une résonance  $\rho$  respectivement.  $\vec{n}_\pi$  représente un vecteur unitaire pointant dans la direction de vol du pion,  $m_\tau$  la masse du tau,  $q$  la différence des quadrivecteurs du pion chargé et du pion neutre,  $N$  le quadrivecteur du neutrino obtenu en soustrayant les quadrivecteurs des deux pions à celui du tau et  $P$  le quadrivecteur du tau. Le vecteur polarimétrique  $\vec{h}^{a_1^{3pr}}$  pour une désintégration via une résonance  $a_1^{3pr}$  et qui est utilisé dans mon analyse n'est pas donné ici car la modélisation de la résonance  $a_1^{3pr}$  est bien plus complexe [121]. Cette résonance donne un  $\rho$  et un pion chargé puis finalement 3 pions chargés. Deux combinaisons de  $\pi^+\pi^-$  sont donc possibles pour représenter la résonance intermédiaire  $\rho$  et ont des interférences constructives. De plus, le calcul du vecteur polarimétrique fait intervenir le courant hadronique qui dépend de l'impulsion des hadrons et de leur polarisation. Le modèle de la résonance  $a_1^{3pr}$  est pris de TAUOLA [122-124] et les paramètres de la collaboration CLEO [125].

Le vecteur polarimétrique peut se définir comme la direction la plus probable du spin du tau ( $\vec{s}$ ) dans le référentiel au repos de ce dernier. En effet, pour n'importe quel mode de désintégration du tau dans son référentiel au repos, sa largeur de désintégration différentielle dépend du produit scalaire du vecteur polarimétrique et du vecteur spin :  $d\Gamma \propto (1 + \vec{h} \cdot \vec{s})$ .  $d\Gamma$  est donc maximal pour  $\vec{h}$  orienté comme  $\vec{s}$ . On construit donc les

plans formés par les directions des taus et les vecteurs polarimétriques dans le référentiel du boson de Higgs au repos afin de mesurer l'angle  $\phi_{CP}$  illustré en Fig. 5.4.

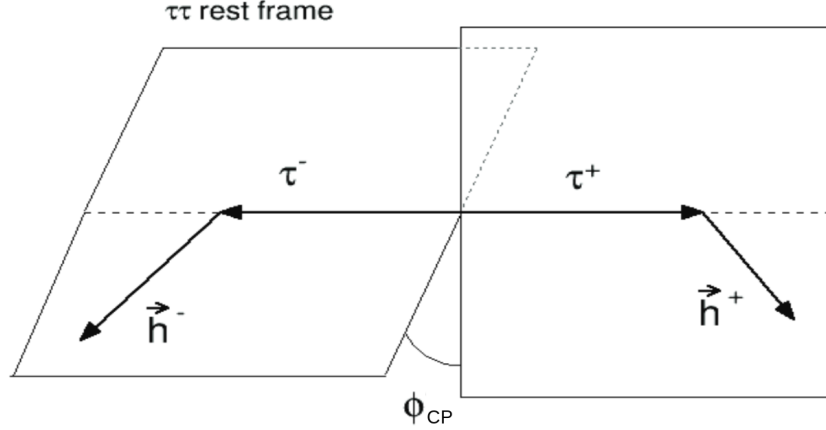


FIGURE 5.4 – Illustration de l'angle  $\phi_{CP}$  avec la méthode du vecteur polarimétrique.

La méthode du vecteur polarimétrique requérant la reconstruction des quadrivercteurs énergie-impulsion des taus, ces derniers doivent donc être estimés en prenant en compte les neutrinos. Cela est possible grâce à une méthode d'ajustement d'évènement global (GEF) [126] applicable dans des évènements contenant au moins un lepton tau se désintégrant en trois pions chargés via une résonance  $a_1^{3pr}$ . En effet, la méthode GEF reconstruit l'impulsion du tau en incluant le neutrino, et les trois pions permettent de reconstruire le point, appelé vertex secondaire (SV), d'où ils sont créés et qui, avec la connaissance du vertex primaire, permet d'obtenir la direction du tau. Ceci est faisable car le temps de vie du boson de Higgs est un milliard de fois plus petit que celui du lepton tau. Par conséquent, son temps de vol est négligeable et on considérera donc géométriquement que le tau émerge du vertex primaire. Ces deux vertex seront ainsi utilisés pour estimer la véritable (en prenant donc en compte le neutrino non détecté par l'expérience CMS) direction du quadrivercteur énergie-impulsion du tau. Dans cette analyse, seul le canal  $a_1^{3pr} a_1^{3pr}$  a été étudié avec la méthode du vecteur polarimétrique, ce qui permet d'utiliser une version simplifiée de la méthode GEF, suffisamment puissante pour ce canal de désintégration.

Afin de mesurer le vecteur polarimétrique  $\vec{h}$ , les impulsions des trois pions chargés et celle du tau estimée par la méthode GEF sont exigées. La direction du vecteur impulsion du tau est donnée par la position du vertex primaire et secondaire :  $\vec{SV} - \vec{PV}$ . Pour obtenir la valeur de l'impulsion du tau, on traite le canal de désintégration du lepton tau en trois pions chargés et un neutrino comme une désintégration à deux corps grâce à la résonance  $a_1^{3pr}$ . À partir de la conservation de l'énergie et de l'impulsion du système dans le référentiel du laboratoire et en connaissant la masse au repos du lepton tau, on

peut obtenir la valeur de l'impulsion du tau. Cette dernière dépend de l'impulsion du  $a_1^{3\text{pr}}$  et d'un angle  $\theta_{\text{GJ}}$  appelé Gottfried-Jackson. Cet angle est défini comme l'angle entre la direction de vol du tau et la direction du  $a_1^{3\text{pr}}$  [127] (voir Fig. 5.5) :

$$\begin{aligned} P_\tau &= P_{a_1} + P_\nu \\ \Leftrightarrow m_\tau^2 &= m_{a_1}^2 + 2E_{a_1} \sqrt{|\vec{p}_\tau|^2 + |\vec{p}_{a_1}|^2 - 2|\vec{p}_\tau||\vec{p}_{a_1}| \cos \theta_{\text{GJ}} - 2|\vec{p}_{a_1}||\vec{p}_\tau| \cos \theta_{\text{GJ}} + 2|\vec{p}_{a_1}|^2}, \end{aligned} \quad (5.5)$$

$$\begin{aligned} |\vec{p}_\tau| &= |\vec{p}_\tau|(m_{a_1}, |\vec{p}_{a_1}|, \theta_{\text{GJ}}) \\ &= \frac{(m_{a_1}^2 + m_\tau^2)|\vec{p}_{a_1}| \cos \theta_{\text{GJ}} \pm \sqrt{(m_{a_1}^2 + \vec{p}_{a_1}^2)((m_{a_1}^2 - m_\tau^2)^2 - 4m_\tau^2 \vec{p}_{a_1}^2 \sin^2 \theta_{\text{GJ}})}}{2(m_{a_1}^2 + \vec{p}_{a_1}^2 \sin^2 \theta_{\text{GJ}})}. \end{aligned} \quad (5.6)$$

Deux solutions sont alors possibles : le  $a_1^{3\text{pr}}$ , dans le référentiel au repos du tau, est émis dans la même direction que la direction de vol du tau dans le référentiel du laboratoire ou le  $a_1^{3\text{pr}}$  est émis dans la direction opposée. Une seule solution existe quand la racine carrée est nulle dans l'Eq. 5.6 et l'angle est alors à sa valeur maximale :

$$\begin{aligned} (m_{a_1}^2 + \vec{p}_{a_1}^2) ((m_{a_1}^2 - m_\tau^2)^2 - 4m_\tau^2 \vec{p}_{a_1}^2 \sin^2 \theta_{\text{GJ}}) &= 0 \\ \Leftrightarrow \theta_{\text{GJ}}^{\text{max}}(m_{a_1}, |\vec{p}_{a_1}|) &= \arcsin \left( \frac{m_\tau^2 - m_{a_1}^2}{2m_\tau |\vec{p}_{a_1}|} \right). \end{aligned} \quad (5.7)$$

Si la racine carrée est négative, alors les solutions ne sont pas physiques. En effet, l'incertitude de mesure sur la reconstruction des vertex primaires et secondaires peuvent mener à une solution non physique. Dans ce cas, on pose  $\theta_{\text{GJ}} = \theta_{\text{GJ}}^{\text{max}}$  (voir Fig. 5.5). Chaque tau possédant jusqu'à deux solutions possibles, l'évènement en contient donc quatre. L'ambiguïté est levée en prenant la paire de solution ayant la masse invariante la plus proche de celle du boson de Higgs. Pour un tau donné, la solution la plus proche de la vrai impulsion est choisie dans 75% des cas.

On définit alors pour chaque tau un plan formé par l'impulsion du tau en question et son  $\vec{h}$  associé. L'angle acoplaire  $\phi_{\text{CP}}$  est alors calculé comme :

$$\begin{aligned} \phi^* &= \arccos(\vec{k}_1 \cdot \vec{k}_2), \\ \phi_{\text{CP}} &= \begin{cases} \phi^*, & \text{si } (\vec{h}_1 \times \vec{h}_2) \cdot \vec{n}_1 \leq 0 \\ 2\pi - \phi^*, & \text{si } (\vec{h}_1 \times \vec{h}_2) \cdot \vec{n}_1 > 0 \end{cases}, \end{aligned} \quad (5.8)$$

avec  $\vec{k}_{1,2} = \frac{\vec{h}_{1,2} \times \vec{n}_{1,2}}{|\vec{h}_{1,2} \times \vec{n}_{1,2}|}$  et  $\vec{n}_{1,2}$  le vecteur unitaire de la direction du tau dans le référentiel au repos du boson de Higgs.



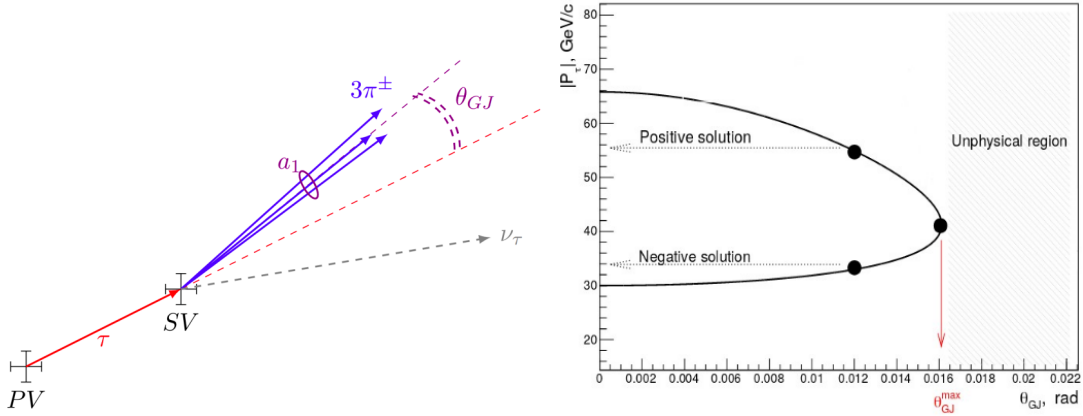


FIGURE 5.5 – Schéma représentant l'angle Gottfried-Jackson (gauche) et sa dépendance avec l'impulsion estimée (droite) [127].

Afin de reconstruire l'impulsion totale du tau, l'algorithme SVfit [128] aurait pu être utilisé à la place de la méthode GEF. L'algorithme SVfit reconstruit la masse invariante du boson de Higgs et les quadrivecteurs des deux leptons taus en combinant l'énergie transverse manquante mesurée avec des probabilités grâce à une méthode d'éléments de matrices. Cependant, SVfit n'utilise pas les vertex primaire et secondaire pour contraindre le quadrivecteur énergie-impulsion des taus mais seulement les quadrivecteurs visibles et l'énergie transverse manquante (MET) avec sa matrice de covariance. De ce fait, la direction de l'impulsion estimée est très proche de la direction visible, et la précision sur les angles est faible car ces derniers sont petits et car les contraintes utilisant la MET ne sont pas fortes. Par conséquent, la méthode GEF est préférable à SVfit car elle fournit de meilleures résolutions angulaires (voir Fig. 5.6 et 5.7).

Après avoir appliqué toutes les sélections usuelles de l'analyse, la distribution de l'angle  $\phi_{CP}$  pour du signal est montrée en Fig 5.8. La méthode GEF est comparée à l'utilisation de SVfit pour les différents états CP-pair et impair du boson de Higgs. L'ajustement suivant est appliqué :  $b + a \cdot \cos(x + c)$  avec  $a$  l'amplitude,  $b$  la moyenne et  $c$  une phase. Le ratio  $|a/b|$  quantifie le pouvoir séparateur et est donné dans la légende. On remarque que l'utilisation de SVfit donne de moins bons résultats que la méthode GEF, le ratio  $|a/b|$  est réduit de 55% à 70%. Une comparaison est aussi effectuée entre les méthodes du plan de désintégration et du vecteur polarimétrique en utilisant la méthode GEF pour cette dernière en Fig. 5.9 pour la période du Run 2. Dû à une convention différente, les extrema des deux méthodes sont décalés de  $\pi$ . La méthode du vecteur polarimétrique possède un meilleur pouvoir séparateur que la méthode du plan de désintégration expliquée en sous-section 5.1.2.

La liste des différentes méthodes utilisées suivant le mode de désintégration est en table 5.1.

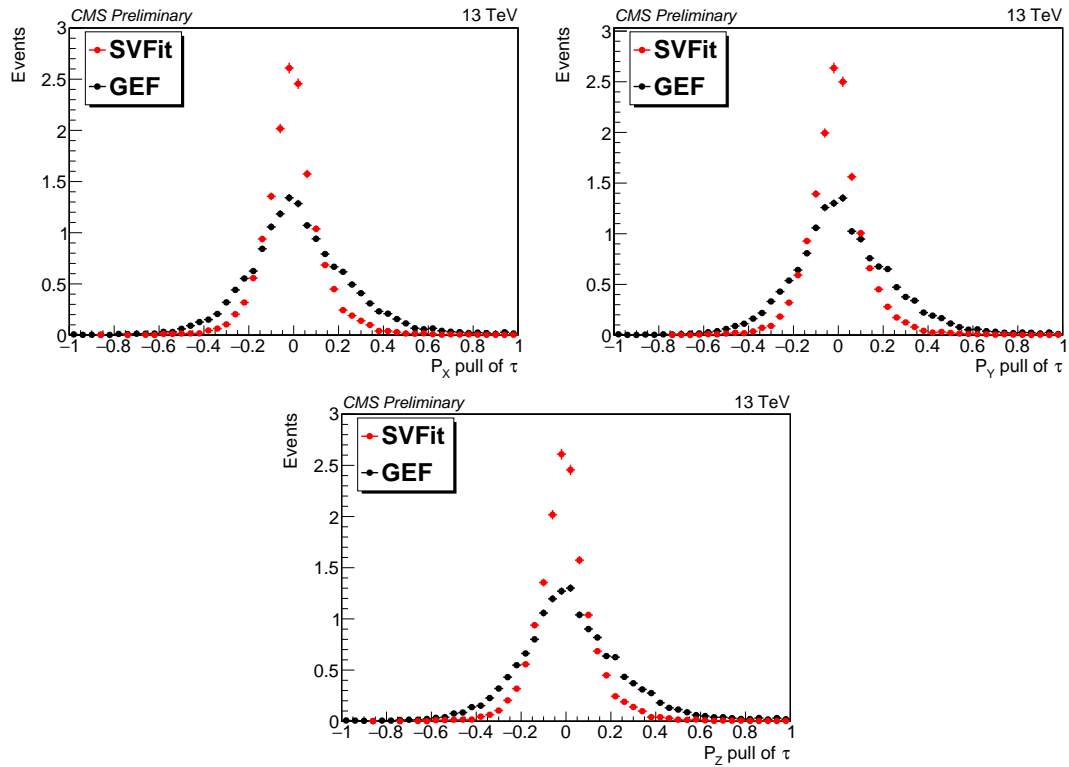


FIGURE 5.6 – Comparaison de la résolution des trois composantes du vecteur impulsion des taus entre SVfit (rouge) et la méthode GEF (noir) en utilisant des échantillons de signaux ggH, VBF et VH.

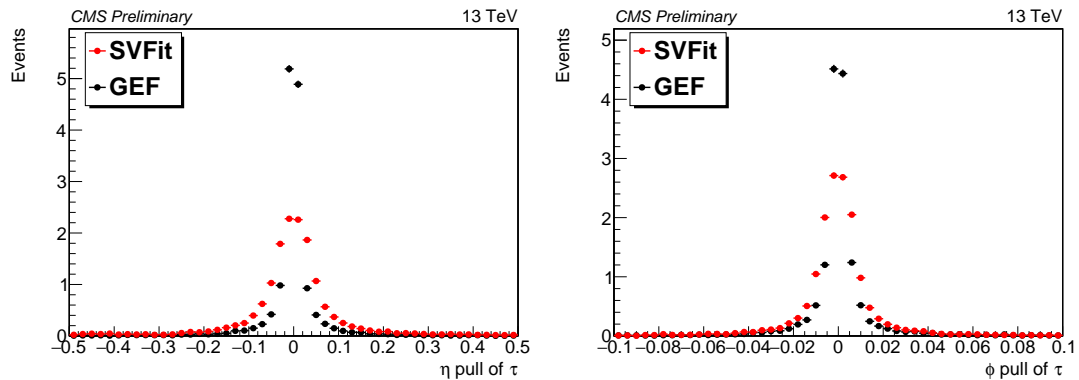


FIGURE 5.7 – Comparaison de la résolution de la pseudorapidité  $\eta$  et de l'angle azimutal  $\phi$  des taus entre SVfit (rouge) et la méthode GEF (noir) en utilisant des échantillons de signaux ggH, VBF et VH.

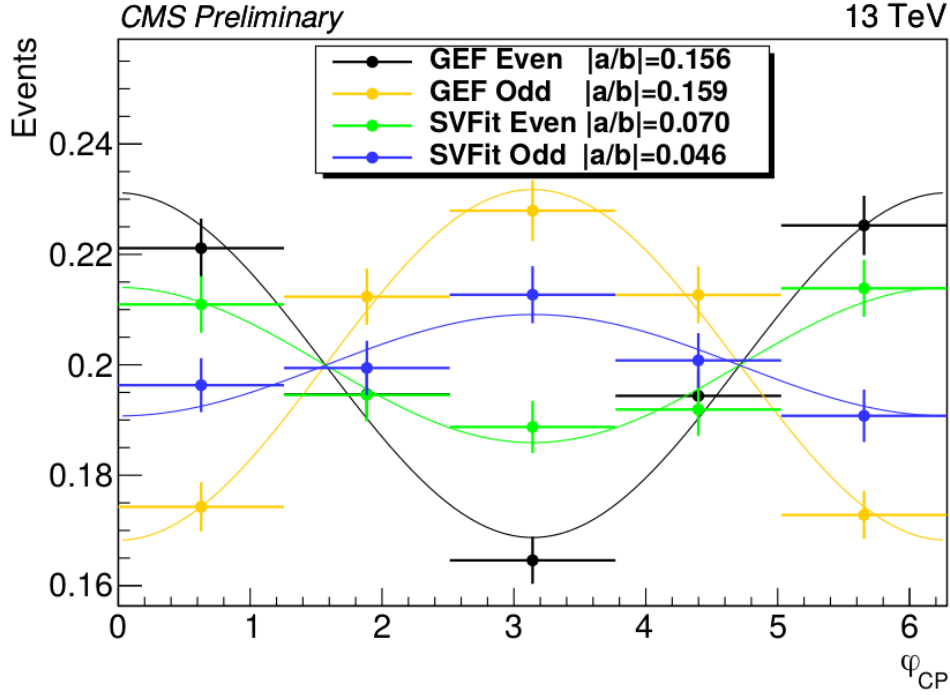


FIGURE 5.8 – Comparaison de la distribution de l'angle  $\phi_{CP}$  entre l'utilisation de la méthode GEF et SVfit pour un boson de Higgs CP-pair et CP-impair en utilisant des échantillons de signaux ggH, VBF et VH. Le ratio  $|a/b|$ , quantifiant le pouvoir séparateur, est aussi donné.

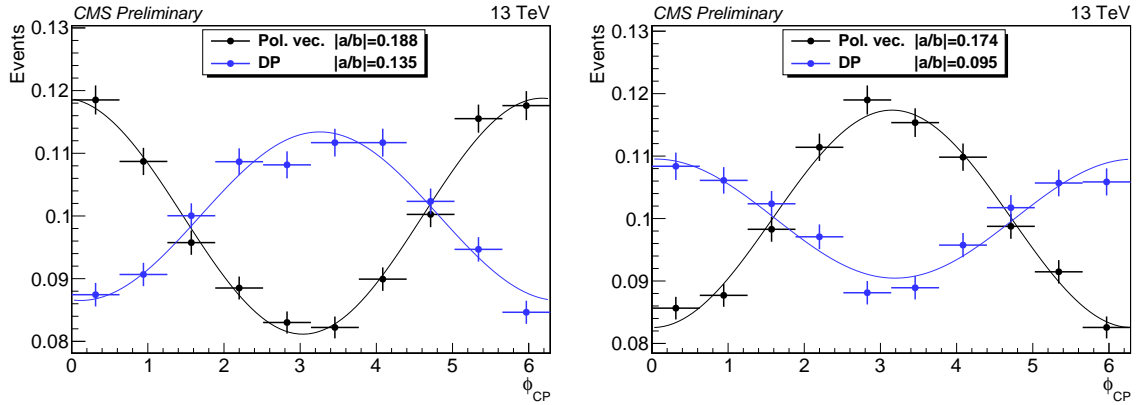


FIGURE 5.9 – Comparaison de la distribution de l'angle  $\phi_{CP}$  au Run 2 (2016+2017+2018) entre la méthode du vecteur polarimétrique (noir) et celle du plan de désintégration (bleu) pour un boson de Higgs CP-pair (gauche) et CP-impair (droite) en utilisant des échantillons de signaux de ggH, VBF et VH. Le ratio  $|a/b|$ , quantifiant le pouvoir séparateur, est aussi donné. La différence de phase est due à une convention différente entre les deux méthodes.

Canaux	Méthodes
$e\pi$	IP-IP
$e\rho$	IP-DP
$ea_1^{1\text{pr}}$	IP-DP
$ea_1^{3\text{pr}}$	IP-DP
$\mu\pi$	IP-IP
$\mu\rho$	IP-DP
$\mu a_1^{1\text{pr}}$	IP-DP
$\mu a_1^{3\text{pr}}$	IP-DP
$\pi\pi$	IP-IP
$\pi\rho$	IP-DP
$\pi a_1^{1\text{pr}}$	IP-DP
$\pi a_1^{3\text{pr}}$	IP-DP
$\rho\rho$	DP-DP
$\rho a_1^{1\text{pr}}$	DP-DP
$\rho a_1^{3\text{pr}}$	DP-DP
$a_1^{1\text{pr}} a_1^{1\text{pr}}$	DP-DP
$a_1^{1\text{pr}} a_1^{3\text{pr}}$	DP-DP
$a_1^{3\text{pr}} a_1^{3\text{pr}}$	DP-DP ou PV-PV

TABLE 5.1 – Liste des canaux et des méthodes associées utilisées selon le canal de désintégration. IP signifie « paramètre d'impact », DP « plan de désintégration » et PV « vecteur polarimétrique ».

## 5.2 Échantillons

### 5.2.1 Données

Cette analyse étudie toutes les données du Run 2, 2016 (legacy 94X), 2017 et 2018, collectées par l'expérience CMS. Ceci correspond à une énergie dans le centre de masse de 13 TeV et à une luminosité intégrée de 35.9 (2016) + 41.5 (2017) + 59.7 (2018)  $\text{fb}^{-1}$ , soit 137.1  $\text{fb}^{-1}$  au total. Une liste des données analysées dans le canal  $\tau_h\tau_h$  se trouve en [Appendice A](#) dans les tableaux [A.1](#) - [A.3](#).

### 5.2.2 Simulations

Le signal et le bruit de fond MC utilisés dans cette analyse sont résumés en [Appendice A](#) dans les tableaux [A.4](#) - [A.5](#) (2016), [A.6](#) - [A.7](#) (2017) et [A.8](#) - [A.9](#) (2018). Les échantillons de signaux sont produits en utilisant POWHEG avec l'outil tauspinner [129]. En effet, ces échantillons sont produits dépolarisés, ce qui permet de les polariser par la suite, avec une pondération des événements, en différents états CP du boson de Higgs. Une seule production d'échantillons est donc nécessaire par type d'évènement.

Seulement trois poids et distributions différentielles sont nécessaires pour modéliser tout les états CP possibles : le pur CP-pair ( $\phi_{\tau\tau} = 0^\circ$ ), le pure CP-impair ( $\phi_{\tau\tau} = 90^\circ$ ), et un état CP-mixte (choisi comme étant  $\phi_{\tau\tau} = 45^\circ$ ). Les distributions différentielles correspondant à ces trois états CP sont sommées suivant la formule :

$$\begin{aligned} \frac{d\sigma}{dx}(\phi_{\tau\tau}) = & (\cos^2 \phi_{\tau\tau} - \cos \phi_{\tau\tau} \sin \phi_{\tau\tau}) \cdot \frac{d\sigma^{\text{CP-pair}}}{dx} + (\sin^2 \phi_{\tau\tau} - \cos \phi_{\tau\tau} \sin \phi_{\tau\tau}) \cdot \frac{d\sigma^{\text{CP-impair}}}{dx} \\ & + 2 \cos \phi_{\tau\tau} \sin \phi_{\tau\tau} \cdot \frac{d\sigma^{\text{CP-mixte}}}{dx} . \end{aligned} \quad (5.9)$$

De plus, les échantillons de signaux sont filtrés au niveau générateur avec un filtre d'efficacité indiqué en [Appendice A](#) dans les tableaux [A.4](#) (2016), [A.6](#) (2017) et [A.8](#) (2018). Ce filtre supprime des évènements en se basant sur la pseudorapidité et l'impulsion transverse visible des leptons du système di-taus. Ceci est très utile pour cette analyse car beaucoup de sélections, comme le système de déclenchement, suppriment un grand nombre d'évènements. Les valeurs de coupures pour ce filtre sont choisies comme étant 10% (20%) plus faibles que celles choisies dans l'analyse pour les leptons (taus) et sont affichées en tableau [5.2](#). De cette façon, le filtre augmente le nombre d'évènements disponible pour l'analyse d'un facteur 4 et rend le processus de simulation plus efficace.

Channel	$p_T^{\text{lepton1}}$ (GeV)	$p_T^{\text{lepton2}}$ (GeV)	$ \eta _{\text{lepton1}}$	$ \eta _{\text{lepton2}}$
$\tau_\mu\tau_h$	> 19	> 16	< 2.5	< 2.7
$\tau_h\tau_h$	> 33	> 33	< 2.5	< 2.5

TABLE 5.2 – Sélection au niveau générateur sur des échantillons  $H\tau\tau$ .

Les échantillons de processus Drell-Yan, produits avec un nombre défini et non définis de jets, sont combinés en appliquant des poids d'évènement dépendants de leur luminosité intégrée associée. La même procédure est appliquée aux échantillons  $W \rightarrow L\nu$ .

### 5.2.3 Échantillons embedded

Les échantillons dit embedded [[130](#)] permettent une estimation du bruit de fond  $Z \rightarrow \tau_h\tau_h$  à partir de données. Ils sont produits en sélectionnant des évènements  $Z \rightarrow \mu\mu$  dans ces dernières. On enlève ensuite de l'évènement les traces des deux muons ainsi que les informations calorimétriques associées. Puis la cinématique du système di-muons est utilisée afin de simuler des évènements  $Z \rightarrow \tau_h\tau_h$ . L'avantage de cette technique est qu'on obtient une meilleure description des jets, de l'empilement, ainsi que du bruit du détecteur et des effets de résolution. De plus, moins de corrections sont nécessaires comparé à des données simulées (par exemple, pas besoin de correction sur l'énergie des

jets). Les listes des échantillons embedded utilisés dans l'analyse dans le canal  $\tau_h\tau_h$  sont données en [Appendice A](#) dans les tableaux [A.10](#) (2016), [A.11](#) (2017) et [A.12](#) (2018).

## 5.3 Sélections des objets

### 5.3.1 Leptons

Les électrons et muons doivent être reconstruits comme tels par le PF. Les désintégrations des taus hadroniques sont classifiées en utilisant le MVA DM (voir [sous-section 3.2.5](#)). Les hadrons chargés et neutres sont regroupés dans des combinaisons compatibles avec les modes de désintégrations des taus hadroniques et le quadrivecteur énergie-impulsion des candidats est calculé.

Le DNN DeepTau (voir [sous-section 3.2.5](#)) réduit le nombre de jets, électrons et muons se faisant passer pour un tau avec les points de fonctionnement (WP) Medium, VVLoose et VLoose respectivement pour le canal  $\tau_h\tau_h$  et correspondant à une efficacité de 70%, 99% et 99.95% respectivement.

### 5.3.2 Jets

Les jets sont construits à partir des particules candidates PF en utilisant l'algorithme anti-kt avec un paramètre de distance de taille  $R=0.4$ . L'algorithme CHS est appliqué afin de réduire le nombre de jets associés à un vertex d'empilement ainsi que les corrections sur les jets (voir [sous-section 3.2.4](#)). Afin d'exclure les candidats taus de la liste des jets, ces derniers doivent être séparés des candidats de  $\Delta R > 0.5$ . Seules les jets ayant un  $p_T$  corrigé supérieur à 30 GeV et une pseudorapacité absolue inférieure à 4.7 sont gardés. De plus, une grande quantité de bruit dans les jets reconstruits est présente en 2017 à cause d'un bruit élevé dans les bouchons du ECAL. Pour pallier cela, les jets en 2017 ayant une impulsion transverse non corrigée inférieure à 50 GeV et  $2.65 < |\eta| < 3.139$  sont rejetés.

### 5.3.3 MET

L'énergie transverse manquante est reconstruite en utilisant l'algorithme PUPPI (voir [sous-section 3.2.6](#)).

### 5.3.4 Vertex

Une détermination précise de la position du vertex d'interaction pp, dit vertex primaire (PV), est essentielle. La liste des traces reconstruites de l'évènement est donc utilisée afin d'appliquer un ajustement grâce à un ajusteur adaptatif de vertex [131] pour obtenir une meilleure estimation des paramètres du PV, dont sa position dans l'espace. Cet ajusteur fonctionne de la manière suivante. Tout d'abord, une première estimation de la position du vertex est obtenue en utilisant la méthode dite « fraction of sample mode with weights » (FSMW) [131] et une grande incertitude initiale est assignée à cette position. Les traces sont rangées suivant leur distance à cette position. Puis les traces sont ajoutées une par une au vertex et sa position est actualisée par des filtres de Kalman [85, 86]. Après tous ces ajouts, la nouvelle position du vertex est comparée à la première estimation. Si la position a grandement bougé, les traces sont pondérées suivant leur compatibilité avec la nouvelle position du vertex, calculée avec l'estimateur de compatibilité de l'algorithme de filtre de Kalman.

Ces étapes sont répétées avec la nouvelle liste de traces pondérées, en commençant de nouveau avec la première estimation de la position et ses incertitudes. Cela se répète jusqu'à que le déplacement de la position du vertex soit assez petit après plusieurs itérations ou qu'un nombre maximal d'itérations soit atteint. La position et les incertitudes obtenues à la dernière itération sont considérées comme finales et sont utilisées dans l'analyse.

Cet ajustement est amélioré pour cette analyse en imposant une contrainte dite du faisceau. En effet, le faisceau est la région où les protons collisionnent dans le détecteur CMS. Les paramètres de ce faisceau sont déterminés à partir d'une moyenne sur beaucoup d'évènements [132]. La contrainte du faisceau signifie que la position de ce dernier est utilisée comme première estimation de la position du vertex au lieu d'appliquer la méthode FSMW. De plus, la matrice de covariance du faisceau est utilisée comme première incertitude à la place d'une large incertitude. Le faisceau étant mesuré avec une grande précision, les incertitudes des paramètres de trace sont incapables de bouger significativement le PV. Par conséquent, l'incertitude totale sur la position du PV est dominée par l'incertitude de la position du faisceau, du moins dans le plan transverse. Cette contrainte améliore grandement la résolution de la position du PV en x et y, tandis que la position z est quasiment inchangée.

Les traces des désintégrations des taus peuvent être supprimées lors de l'ajustement afin d'éviter un biais du fait du temps de vie des  $\tau$ . Une comparaison de la résolution du vertex primaire nominal, ajusté avec et sans la contrainte du faisceau en excluant les traces provenant des taus et ajusté avec la contrainte du faisceau en gardant ces traces est montrée en Fig. 5.10. Ces résolutions sont obtenues avec des échantillons ggH, VBF et VH. La condition du faisceau améliore grandement la résolution en x et y, tandis

que l'exclusion ou non des traces provenant des taus change très peu la résolution. En revanche, la composante en z est très peu affectée par les ajustements. Le vertex primaire choisi pour l'analyse est celui ajusté avec la condition du faisceau et en excluant les traces des désintégrations des taus.

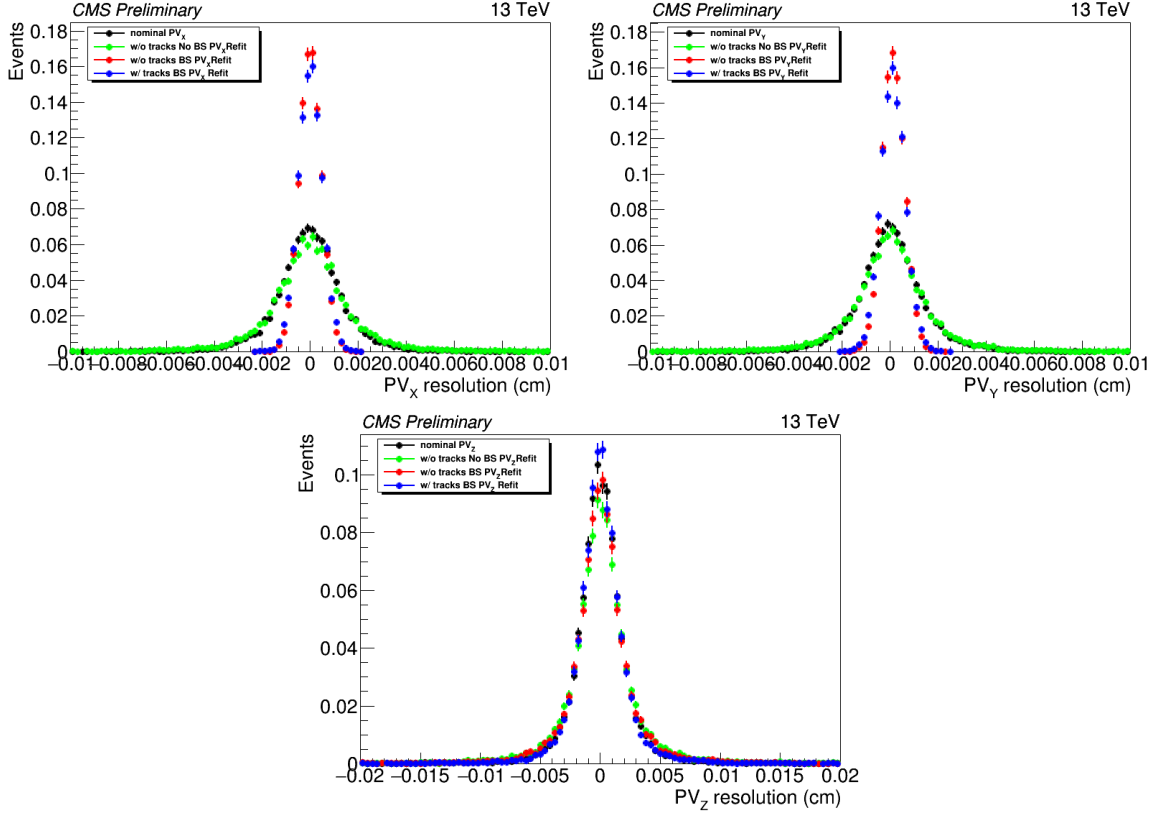


FIGURE 5.10 – Distributions des résolutions des composantes x (gauche), y (droite) et z (milieu) du vertex primaire nominal (noir), ajusté sans les traces provenant des taus avec la contrainte du faisceau (rouge) et sans cette contrainte (vert), ainsi qu'en gardant les traces provenant des taus avec la contrainte du faisceau (bleu) dans le canal  $a_1^{3pr} a_1^{3pr}$  avec des échantillons de signaux 2018 (ggH, VBF et VH). Le vertex primaire utilisé dans l'analyse est celui correspondant aux points rouges.

Le vertex secondaire est aussi ajusté en se basant sur les traces des pions chargés venant du tau. La bonne modélisation du vertex secondaire dans le canal  $a_1^{3pr} a_1^{3pr}$ , dominé par les faux taus, est vérifiée en Fig. 5.11 – 5.13 pour les trois composantes du vecteur formé par le vertex primaire et secondaire pour les années 2016, 2017 et 2018. Les mêmes distributions, mais dans le canal  $\tau_\mu \tau_h$  contenant principalement des vrais taus, sont en Fig. 5.14 pour l'année 2018. Seules les incertitudes statistiques sont affichées.



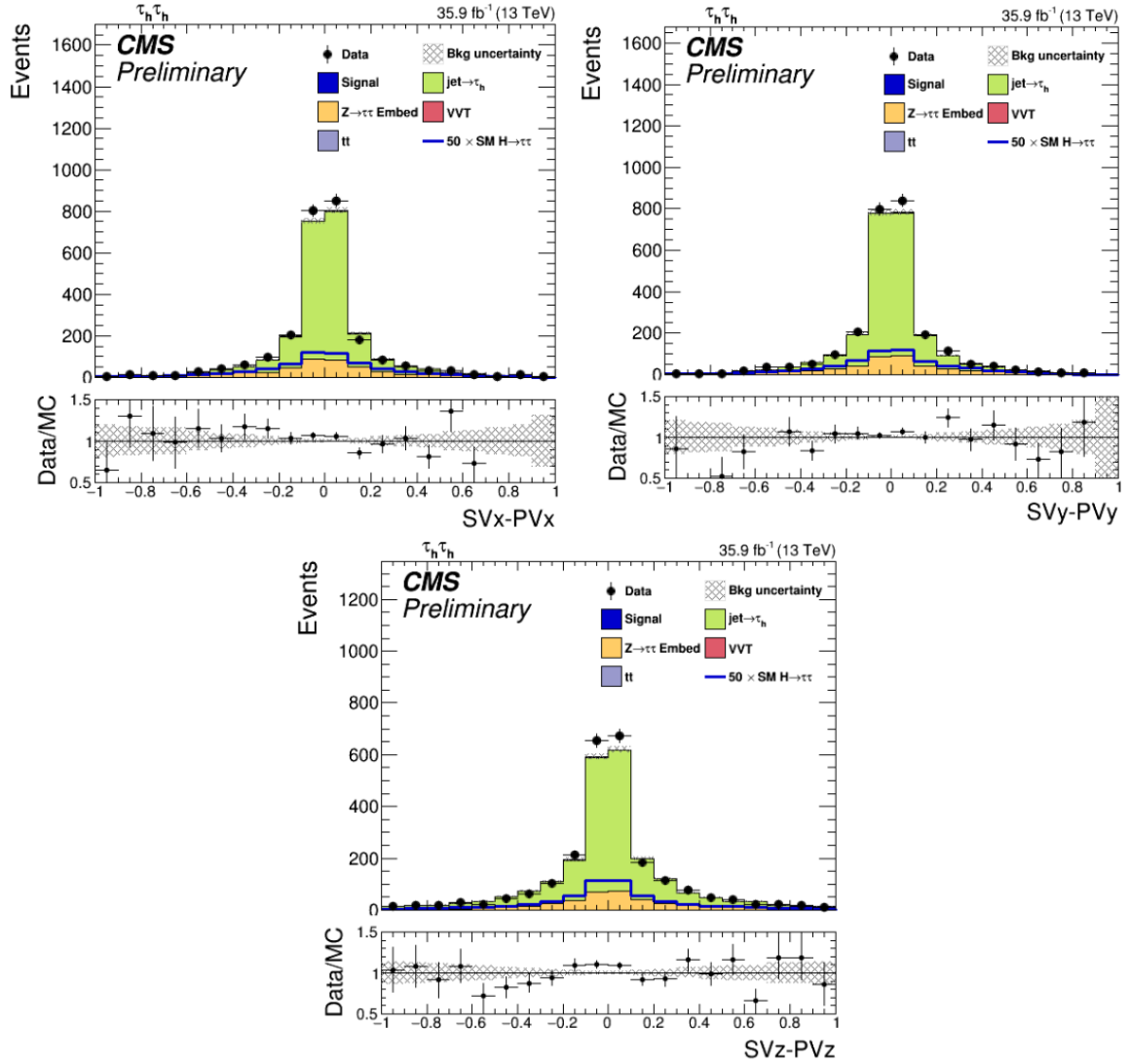


FIGURE 5.11 – Distributions des composantes x (gauche), y (droite) et z (milieu) du vecteur formé par les vertex primaire et secondaire dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  avec des échantillons 2016. Les bandes grises sont seulement les incertitudes statistiques.

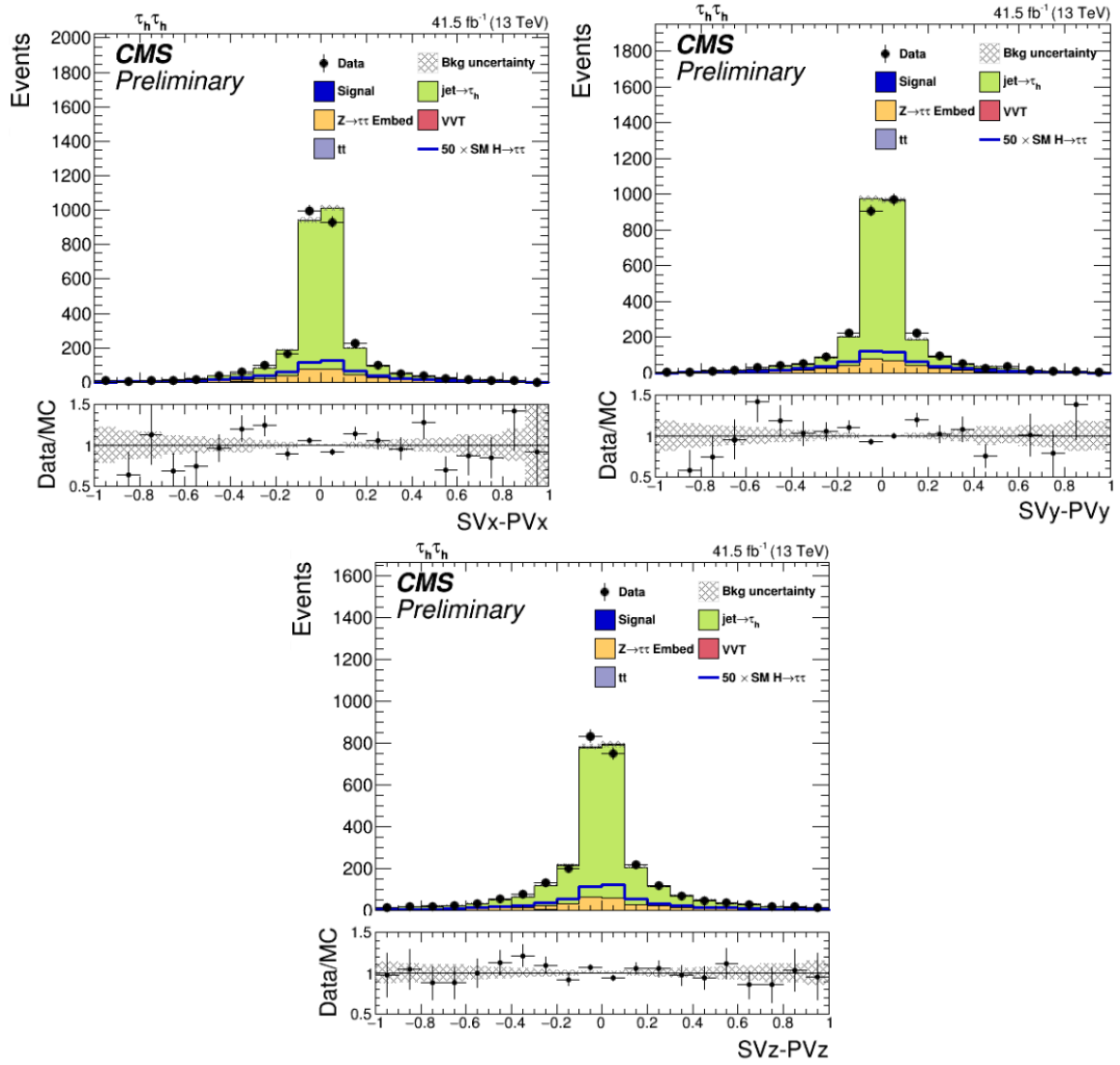


FIGURE 5.12 – Distributions des composantes x (gauche), y (droite) et z (milieu) du vecteur formé par les vertex primaires et secondaires dans le canal  $a_1^{3pr} a_1^{3pr}$  avec des échantillons 2017. Les bandes grises sont seulement les incertitudes statistiques.

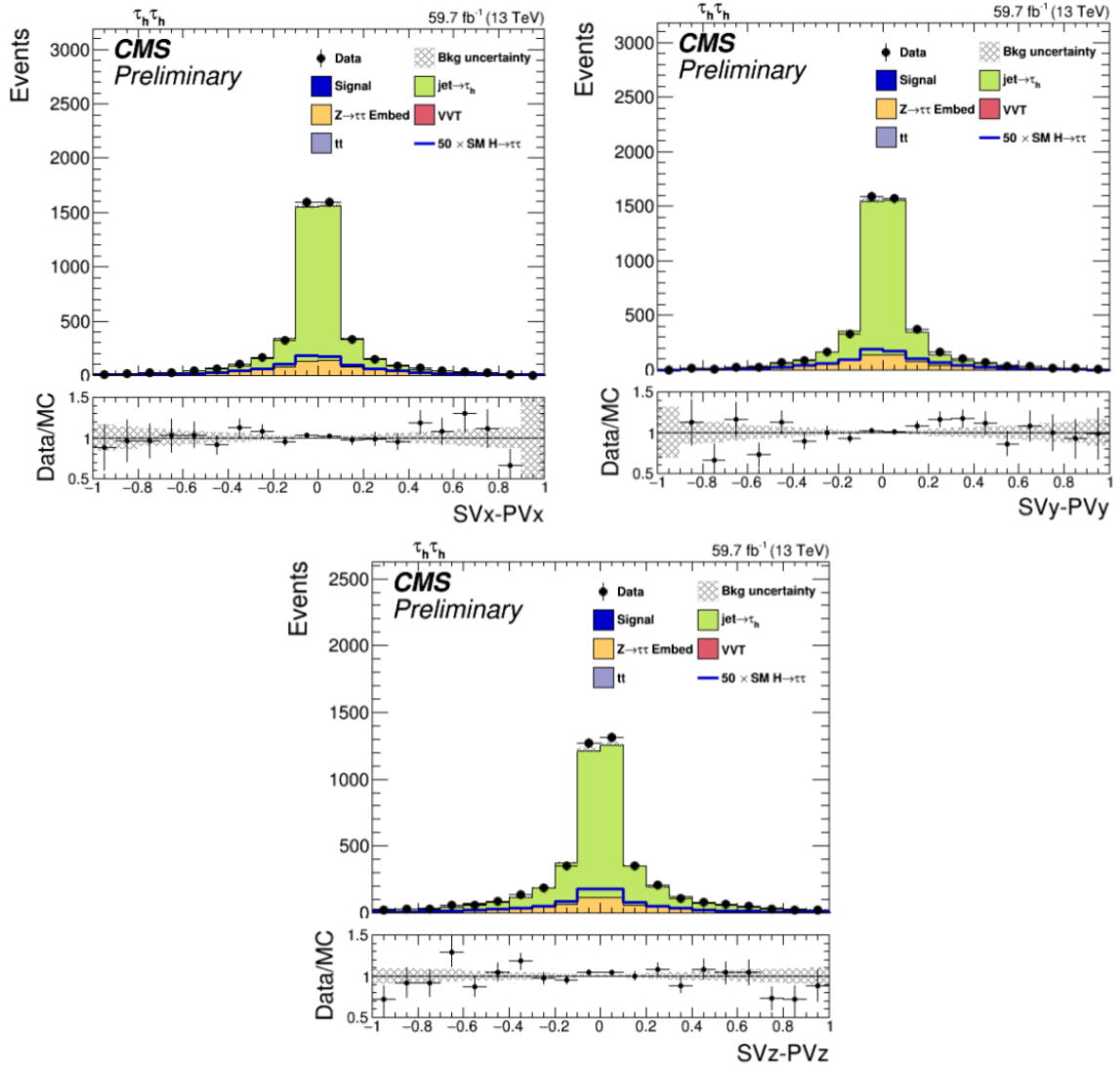


FIGURE 5.13 – Distributions des composantes x (gauche), y (droite) et z (milieu) du vecteur formé par les vertex primaires et secondaires dans le canal  $a_1^{3pr} a_1^{3pr}$  avec des échantillons 2018. Les bandes grises sont seulement les incertitudes statistiques.

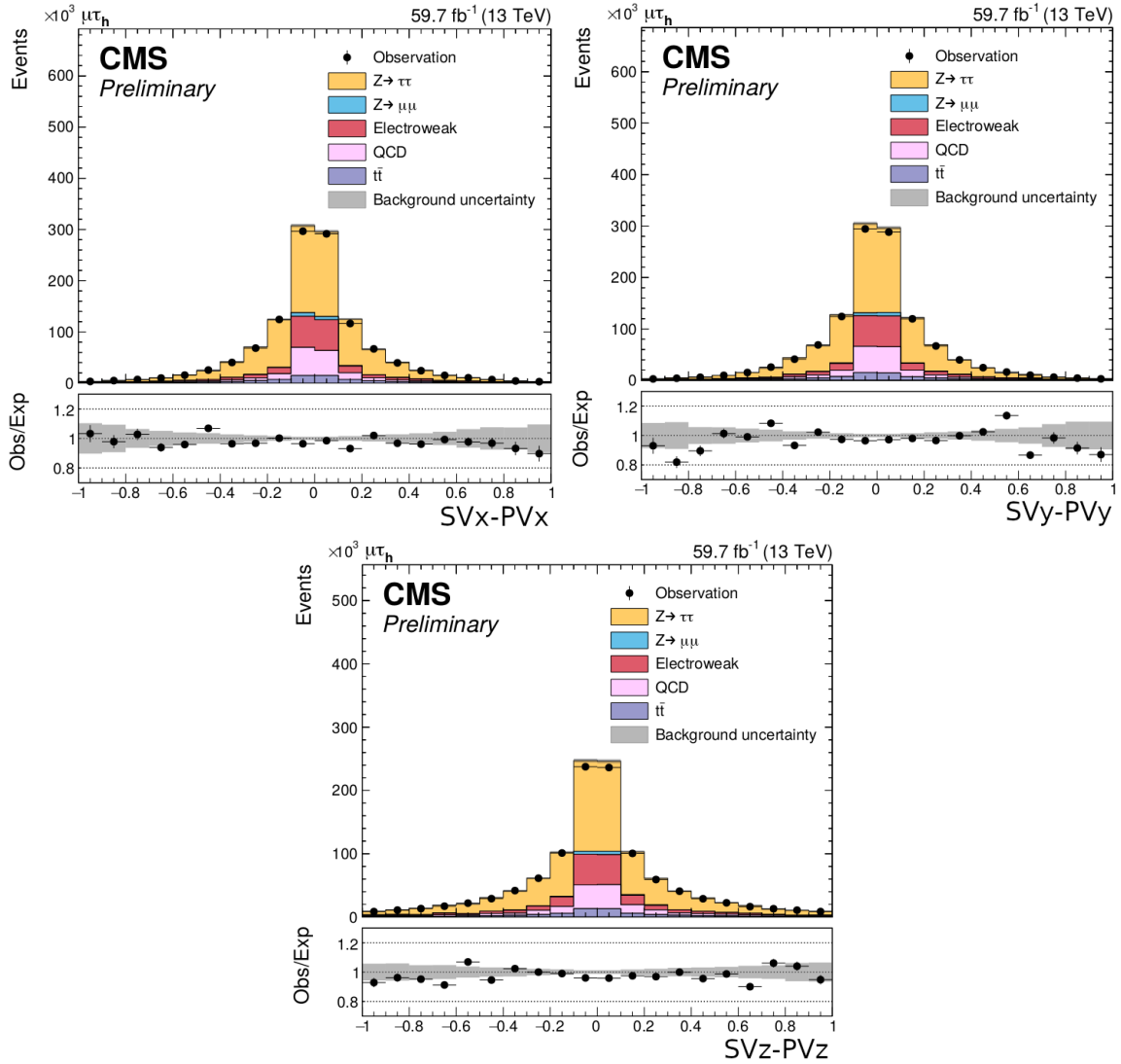


FIGURE 5.14 – Distributions des composantes x (gauche), y (droite) et z (milieu) du vecteur formé par les vertex primaires et secondaires dans le canal  $\tau_\mu \tau_h$  avec des échantillons 2018. Les bandes grises sont seulement les incertitudes statistiques [133].

### 5.3.5 Système de déclenchement et filtres

Les évènements sont collectés s'ils déclenchent un des systèmes de déclenchement listés dans les tableaux 5.3 (2016), 5.4 (2017) et 5.5 (2018) :

- **2016**

Les évènements passent avec un OU de deux chemins HLT, lesquels requièrent deux candidats  $\tau_h$  avec  $|\eta| < 2.1$  et  $p_T > 35$  GeV en ligne ;

- **2017**

Les évènements passent avec un OU de trois chemins HLT, lesquels requièrent différents seuils de  $p_T$  en ligne et différents isolements en ligne ;

- **2018**

Les chemins HLT dépendent du numéro de run car la reconstruction HPS au niveau HLT fut implémentée au milieu de la prise de données 2018. Le facteur d'échelle data/MC du système de déclenchement appliqué dans l'analyse prend en compte ce changement.

HLT Path	L1 seed	Tau filter to match
HLT_DoubleMediumIsoPFTau35 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>1</sup>	hltDoublePFTau35TrackPt1MediumIsolationDz02Reg
HLT_DoubleMediumCombinedIsoPFTau35 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>1</sup>	hltDoublePFTau35TrackPt1MediumCombinedIsolationDz02Reg

<sup>1</sup>L1\_DoubleIsoTau26er OR L1\_DoubleIsoTau27er OR L1\_DoubleIsoTau28er OR L1\_DoubleIsoTau30er OR  
L1\_DoubleIsoTau32er OR L1\_DoubleIsoTau39er

TABLE 5.3 – Systèmes de déclenchement utilisés pour sélectionner les évènements 2016 dans le canal  $\tau_h\tau_h$ .

HLT Path	L1 seed	Tau filter to match
HLT_DoubleTightChargedIsoPFTau40 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>1</sup>	hltDoublePFTau40TrackPt1TightChargedIsolationDz02Reg
HLT_DoubleMediumChargedIsoFTau40 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>1</sup>	hltDoublePFTau40TrackPt1MediumChargedIsolationAndTightOOSCPphotonsDz02Reg
HLT_DoubleTightChargedIsoPFTau35 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>1</sup>	hltDoublePFTau35TrackPt1TightChargedIsolationAndTightOOSCPphotonsDz02Reg

<sup>1</sup>L1\_DoubleIsoTau28er2p1 OR L1\_DoubleIsoTau30er2p1 OR L1\_DoubleIsoTau32er2p1 OR L1\_DoubleIsoTau33er2p1 OR  
L1\_DoubleIsoTau34er2p1 OR L1\_DoubleIsoTau35er2p1 OR L1\_DoubleIsoTau36er2p1 OR L1\_DoubleIsoTau38er2p1 OR  
L1\_DoubleTau70er2p1

TABLE 5.4 – Systèmes de déclenchement utilisés pour sélectionner les évènements 2017 dans le canal  $\tau_h\tau_h$ .

HLT Path	L1 seed	Tau filter to match
HLT_DoubleTightChargedIsoPFTau40 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>†</sup>	hltDoublePFTau40TrackPt1TightChargedIsolationDz02Reg
HLT_DoubleMediumChargedIsoPFTau40 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>†</sup>	hltDoublePFTau40TrackPt1MediumChargedIsolationAndTightOOSCPHOTONSdz02Reg
HLT_DoubleTightChargedIsoPFTau35 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>†</sup>	hltDoublePFTau35TrackPt1TightChargedIsolationAndTightOOSCPHOTONSdz02Reg
(MC, embedded or data after run 317509 only this) HLT_DoubleMediumChargedIsoPFTauHPS35 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTauOR <sup>†</sup>	hltHpsDoublePFTau35TrackPt1MediumChargedIsolationAndDz02Reg

<sup>†</sup>L1\_DoubleIsoTau28er2p1 OR L1\_DoubleIsoTau30er2p1 OR L1\_DoubleIsoTau32er2p1 OR L1\_DoubleIsoTau33er2p1 OR  
L1\_DoubleIsoTau34er2p1 OR L1\_DoubleIsoTau35er2p1 OR L1\_DoubleIsoTau36er2p1 OR L1\_DoubleIsoTau38er2p1 OR  
L1\_DoubleIsoTau70er2p1

TABLE 5.5 – Systèmes de déclenchement utilisés pour sélectionner les évènements 2018 dans le canal  $\tau_h \tau_h$ .

## 5.4 Coupures

### 5.4.1 Filtres MET

Les évènements doivent passer ces filtres de  $p_T^{\vec{p}_T^{\text{miss}}}$  :

- filtre du vertex primaire (données et MC) ;
- filtre très serré du halo de faisceau (données et MC) ;
- filtre de bruit du HCAL (données et MC) ;
- filtre de bruit du HCAL isolé (données et MC) ;
- filtre de bruit des mauvais super-cristaux du EE (données) ;
- filtre des cristaux masqués du ECAL (données et MC) ;
- filtre des mauvais muons (données et MC) ;
- filtre de mauvaise calibration du ECAL (données et MC) (2017 et 2018).

### 5.4.2 Veto de lepton supplémentaire

Ces évènements sont rejetés s'ils contiennent un troisième lepton séparé des deux autres taus sélectionnés par  $\Delta R > 0.5$  et passant ces critères :

**Électron :**

- $p_T > 10$  GeV et  $|\eta| < 2.5$  ;
- $d_z < 0.2$  cm et  $d_{xy} < 0.045$  cm ;
- passe le MVA d'identification de l'électron avec le WP à 90% d'efficacité sans isolement (entraînement Fall17-V2) ;
- passe le veto de conversion ;
- nombre d'impacts intérieurs manquants  $\leq 1$  ;
- isolement  $< 0.3 \times p_T^e$  avec un cône de taille 0.3.

**Muon :**

- $p_T > 10$  GeV et  $|\eta| < 2.4$ ;
- $d_z < 0.2$  cm, et  $d_{xy} < 0.045$  cm ;
- passe l'identification du muon en Medium ;
- isolement  $< 0.3 \times p_T^\mu$  avec un cône de taille 0.4.

### 5.4.3 Veto dilepton

Un veto est appliqué aux événements contenant une paire de leptons où ces derniers de signes opposés sont séparés de  $\Delta R > 0.15$  et passent ces conditions :

**Di-électrons :**

- $p_T > 15$  GeV ;
- $|\eta| < 2.5$  ;
- passe les coupures d'identification de l'électron (Fall17V2) ;
- $d_z < 0.2$  cm ;
- $d_{xy} < 0.045$  cm ;
- isolement  $< 0.3 \times p_T^e$  avec  $\rho$  corrigé et un cône de taille 0.3.

**Di-muons :**

- $p_T > 15$  GeV ;
- $|\eta| < 2.4$  ;
- le muon doit être un muon global ;
- le muon doit être un muon du trajectographe ;
- le muon doit être un muon PF ;
- $d_z < 0.2$  cm ;
- $d_{xy} < 0.045$  cm ;
- isolement  $< 0.3 \times p_T^\mu$  avec un cône de taille 0.4.

### 5.4.4 Sélections

Un événement est présélectionné si au moins une paire de taus hadroniques est trouvée. Ces  $\tau_h$  doivent être de charges électriques opposées et séparés de  $\Delta R > 0.5$ . Un candidat  $\tau_h$  est sélectionné s'il satisfait ces conditions :

- $p_T > 40$  GeV, et  $|\eta| < 2.1$  ;
- HPS DM 5 et 6 exclus (désintégration en deux particules chargées) ;
- DeepTau vs jets avec WP Medium ;

- DeepTau vs électrons avec WP VVLoose ;
- DeepTau vs muons avec WP VLoose ;
- Vertex des particules et faisceau séparés de  $d_z < 0.2$  cm.

Le tau de la paire ayant la plus grande impulsion transverse est nommé candidat 1 et l'autre tau candidat 2. Si plusieurs paires candidates sont présentes, un algorithme de tri est utilisé et la paire finale doit avoir une masse invariante visible  $> 40$  GeV :

- choisir la paire avec le candidat 1 le plus isolé, c'est à dire en utilisant la valeur de sortie de DeepTau vs jets ;
- si l'isolement est le même entre deux paires, celle avec le candidat 1 de plus haut  $p_T$  est choisie ;
- si le  $p_T$  est le même, la paire avec le candidat 2 le plus isolé est choisie ;
- si l'isolement est le même, alors la paire avec le candidat 2 de plus haut  $p_T$  est choisie.

De plus, après sélection de la paire finale, les objets sélectionnés hors lignes lors de l'application des chemins HLT doivent correspondre aux objets sélectionnés en ligne lors de l'utilisation du système de déclenchement ( $\Delta R < 0.5$ ).

## 5.5 Corrections

Des corrections sont appliquées afin de tenir compte des différences entre les données et la simulation. En plus des corrections détaillées en [section 3.2](#), d'autres sont expliquées ci-dessous.

### 5.5.1 Échelle d'énergie des $\tau_h$

L'énergie des taus hadroniques est corrigée à partir de mesures dans le canal  $\tau_\mu\tau_h$  en utilisant la masse du  $\tau_h$  et la masse visible du système  $\tau_\mu\tau_h$  comme observables. La correction en énergie (appelée TES) est obtenue en ajustant le maximum de vraisemblance pour différentes valeurs d'énergies simulées du  $\tau_h$ . Cette correction est alors propagée à la MET et à toutes les variables basées sur le quadrivecteur du  $\tau_h$  et sur la MET.

### 5.5.2 Efficacité d'identification des $\tau_h$

Des différences en efficacité pour les taus hadroniques existent entre les données et la simulation lors de l'étape de reconstruction et d'identification. On corrige donc ces différences avec des intervalles (ces intervalles dans les histogrammes seront appelés « bins » par la suite) en MVA DM au lieu de la correction habituelle en fonction du HPS DM. De plus, ces corrections sont mesurées pour des impulsions transverses inférieures et supérieures à 40 GeV, mais seule cette dernière est utilisée dans le canal  $\tau_h\tau_h$  car on sélectionne les taus hadroniques à partir de 40 GeV. L'extraction de ces corrections est détaillée en [Appendice B](#).



### 5.5.3 Facteurs d'échelle du système de déclenchement du tau hadronique

Les facteurs d'échelle du système de déclenchement du tau hadronique sont calculés pour chaque MVA DM dans le canal  $\tau_\mu\tau_h$ . Pour ces mesures, le bruit de fond est estimé avec les échantillons W+jet normalisés aux données en utilisant la région de haute  $m_T$  ( $m_T > 70$  GeV). Puis les conditions suivantes sont appliquées :

- le muon passe le système de déclenchement pour un muon seul ;
- isolation du muon  $< 0.15$  ;
- DeepTau vs jets, électrons et muons avec WP Medium, VVVLoose et Tight respectivement ;
- les évènements avec des leptons supplémentaires sont rejetés ;
- le muon a un  $p_T > 25$  GeV pour 2016 et un  $p_T > 29$  GeV pour 2017 et 2018 ;
- les évènements possédant des jets b sont rejetés ;
- $m_T < 40$  afin de diminuer le bruit de fond W+jets ;
- $50 \text{ GeV} < m_{\text{vis}} < 90 \text{ GeV}$ .

Les systèmes de déclenchement utilisés pour le  $\tau_h$  sont montrés en table 5.6 pour le canal  $\tau_h\tau_h$ . Afin de mesurer l'efficacité des systèmes de déclenchement, les différents bins en  $p_T$  du  $\tau_h$  utilisés sont [20,21,22,23,24,26,28,30,35,40,50,60,75,110,410] GeV et l'efficacité dans chaque bin est définie comme la fraction des évènements qui passe ce système. On utilise un processus gaussien pour ajuster les graphiques d'efficacité avec une fonction continue non paramétrique [134] (voir Fig. 5.15). Les facteurs d'échelle sont calculés en divisant la fonction ajustée des données par celle MC (ou embedded), et l'incertitude est estimée avec la propagation d'erreur standard.

### 5.5.4 Mauvaise identification des électrons et muons en $\tau_h$

L'efficacité du discriminant rejetant les électrons et muons faussement reconstruits en  $\tau_h$  n'est pas parfaitement modélisée dans la simulation. Des facteurs d'échelle sont donc calculés pour les régions en  $\eta$  des tonneaux et des bouchons, ainsi que pour chaque HPS DM. Ces valeurs de correction sont spécifiques à l'algorithme DeepTau et sont appliquées aux évènements avec  $l \rightarrow \tau_h$  dans la simulation. Cette correction n'est pas appliquée sur les échantillons embedded qui sont estimés à partir des données.

### 5.5.5 Corrections de recul MET

$\vec{p}_T^{\text{miss}}$  est mal modélisée dans les échantillons de simulation de DY, W+jet et boson de Higgs. Une correction est donc mesurée dans des évènements  $Z \rightarrow \mu\mu$  et est appliquée sur la différence vectorielle du  $\vec{p}_T^{\text{miss}}$  mesuré et de l'impulsion transverse totale des neutrinos venant des désintégrations de Z, W et boson de Higgs. Dans ce canal le recul leptonique

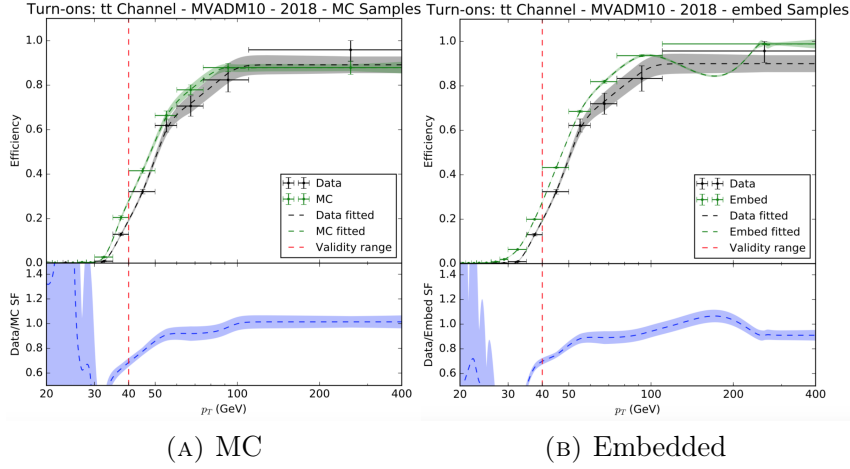


FIGURE 5.15 – Efficacités (haut) et facteurs de correction (bas) dans les données 2018, MC (gauche) et embedded (droite), facteurs d'échelle et ajustements correspondant au canal  $\tau_h\tau_h$  pour MVA DM = 10 [133].

ne contient pas de neutrino et le quadrivecteur du boson Z peut être précisément mesuré [135].

### 5.5.6 Repondération de l'empilement

Les événements MC sont pondérés afin de correspondre aux conditions d'empilement dans les données. Ces facteurs sont calculés comme un ratio, bin par bin, de la distribution de l'empilement entre la simulation et les données. Cette correction n'est pas appliquée sur les échantillons embedded.

### 5.5.7 Repondération du $p_T$ et de la masse du Z

On repondère les événements DY de simulation afin de corriger les distributions de  $p_T$  et de masse générés dans les échantillons dit leading order (LO). Les poids proviennent des différences entre la simulation et les données dans une région de contrôle pure en  $Z \rightarrow \mu\mu$ .

### 5.5.8 Repondération du $p_T$ du top

On a besoin de corriger l'impulsion transverse des quarks top des bruits de fond  $t\bar{t}$ . Ceci est fait grâce à une fonction dépendant du  $p_T$  des quarks (anti)top et est appliqué pour chacun de ces quarks :

$$w = \sqrt{\exp(a + b \cdot p_T + c \cdot p_T^2)}, \quad (5.10)$$

avec  $a=0.088$ ,  $b=-0.00087$ ,  $c=9.2 \cdot 10^{-7}$  et  $p_T$  en GeV.

Year	HLT path(s)	Filter(s) to match for data/MC	Filter(s) to match for em-bedding	Extra	Notes
2016	HLT_IsoMu19_eta2p1_MediumIsoPFTau32_Tk1_eta2p1_Reg_v	htOverlapFilterIsoMu19MediumIsoPFTau32Reg	hitPFTau32Reg	35 GeV HLT pr cut	–
2016	HLT_IsoMu19_eta2p1_MediumCombinedIsoPFTau32_Tk1_eta2p1_Reg_v	htOverlapFilterIsoMu19MediumCombinedIsoPFTau32Reg	hitPFTau32Reg	35 GeV HLT pr cut	–
2017	HLT_IsoMu24_eta2p1_TightChargedIsoPFTau35_Tk1_TightID_Reg_CrossL1	hitSelectedPFTau35TrackPt1TightChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	32 GeV (isolated) L1 pr cut	–
2017	HLT_IsoMu24_eta2p1_MediumChargedIsoPFTau35_Tk1_TightID_Reg_CrossL1	hitSelectedPFTau35TrackPt1TightChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	32 GeV (isolated) L1 pr cut	–
2017	HLT_IsoMu24_eta2p1_TightChargedIsoPFTau35_Tk1_eta2p1_Reg_CrossL1	hitSelectedPFTau35TrackPt1TightChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	40 GeV HLT pr cut & 32 GeV (isolated) L1 pr cut	–
2018	HLT_IsoMu24_eta2p1_TightChargedIsoPFTau35_Tk1_TightID_Reg_CrossL1	hitSelectedPFTau35TrackPt1TightChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	32 GeV (isolated) L1 pr cut	Not used in MC
2018	HLT_IsoMu24_eta2p1_MediumChargedIsoPFTau35_Tk1_TightID_Reg_CrossL1	hitSelectedPFTau35TrackPt1MediumChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	32 GeV (isolated) L1 pr cut	Not used in MC
2018	HLT_IsoMu24_eta2p1_TightChargedIsoPFTau35_Tk1_CrossL1	hitSelectedPFTau35TrackPt1TightChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	40 GeV HLT pr cut & 32 GeV (isolated) L1 pr cut	Not used in MC
2018	HLT_IsoMu24_eta2p1_MediumChargedIsoPFTau35_Tk1_eta2p1_Reg_CrossL1_v	hitHpsSelectedPFTau35TrackPt1MediumChargedIsolationAndTightOOSCPPho-tonsL1HLTMatchedReg	hitSingleL2IsoTau26eta2p2	32 GeV (isolated) L1 pr cut	–

TABLE 5.6 – Résumé des systèmes de déclenchement utilisés pour mesurer les efficacités des systèmes de déclenchement dans le canal  $\tau_h\tau_h$ .

### 5.5.9 Poids prefiring

Les données 2016 et 2017 ont un problème de prefiring du L1. En effet, durant ces années, un décalage progressif en temps dans le ECAL eu lieu et fut mal propagé aux primitives du système de déclenchement, ces dernières étant les informations utilisées par le L1 du système de déclenchement. Par conséquent, une fraction de ces primitives de déclenchement à grandes pseudorapidités dans le ECAL fut associée au croisement de paquet de protons précédent. Cet effet est pris en compte dans la simulation et on applique un poids d'évènement. Cette correction est par exemple autour de 0.95 pour des échantillons de signaux VBF [135]. Cette correction n'est pas appliquée sur les échantillons embedded.

## 5.6 Estimation du bruit de fond

Cette analyse considère plusieurs bruits de fond : Drell-Yan,  $t\bar{t}$ ,  $W$ +jets, QCD, di-boson, single-top, et boson  $W$  et  $Z$  produits par interaction électrofaible. Pour le canal  $\tau_h\tau_h$ , des échantillons embedded sont utilisés pour modéliser tous les processus produisant deux vrais leptons taus. De plus, les bruits de fond incluant des jets reconstruits en candidats  $\tau_h$  (écrit comme  $j \rightarrow \tau_h$  dans la suite) sont estimés à partir des données avec la méthode dite des fake factors (FF), le reste étant des échantillons de simulations.

### 5.6.1 Méthode des fake factors

Les bruits de fond ayant un jet reconstruit en  $\tau_h$  sont mal modélisés avec des échantillons simulés. De plus, la faible probabilité de ces mauvaises reconstructions implique la simulation d'énormément d'évènements. C'est pourquoi ces processus sont estimés à partir des données dans cette analyse grâce à l'utilisation de fake factors.

La méthode des fake factors modélise les bruits de fond  $j \rightarrow \tau_h$  à partir de régions dans les données où un candidat  $\tau_h$  ne passe pas le WP Medium de DeepTau, appliqué pour la région de signal (SR), mais passe le WP VVVLoose. Ces régions seront appelées régions d'application (AR). L'idée de cette méthode est de pondérer les évènements dans l'AR par un facteur :

$$FF = \frac{N(\text{Medium})}{N(\text{VVVLoose} \ \&\& \ !\text{Medium})}, \quad (5.11)$$

avec  $N(X)$  le nombre d'évènements vérifiant la condition  $X$ . Ces ratios FF sont mesurés dans des régions de contrôle enrichies en évènements  $j \rightarrow \tau_h$  nommées régions de détermination (DR). Ces facteurs dépendent du nombre de jets ( $N_{\text{jets}}$ ), du MVA DM, et la dépendance en  $p_T$  est prise en compte pour l'utilisation de fonctions ajustées. Les processus dominants dans le canal  $\tau_h\tau_h$  sont des processus QCD. Les ratios FF sont

corrigés pour tenir compte de l'extrapolation entre la DR et la SR, ainsi que pour la dépendance envers des variables non prévues (via des corrections dite non-closure).

### 5.6.1.1 Fake factors dans le canal $\tau_h\tau_h$

Les FF sont seulement mesurés pour le processus dominant QCD dans le canal  $\tau_h\tau_h$ . La DR est définie en inversant la condition voulant que les taus dans les paires soient de charges électriques opposés (OS), afin de ne sélectionner que les paires de mêmes signes (SS). Toutes les autres sélections restent inchangées comme dans la SR. On ne mesure les FF que pour le tau de plus haut  $p_T$  au sein de la paire, et pour trois bins en  $N_{\text{jets}}$  (0, 1,  $\geq 2$ ) et cinq bins en MVA DM (0, 1, 2, 10, 11). Dans chaque bin, la dépendance en  $p_T$  du  $\tau_h$  est ajustée en utilisant la superposition d'une Landau et d'une fonction constante. Quand la population statistique d'un bin est insuffisante, l'ajustement peut échouer ou donner un mauvais résultat. Dans ces cas là, la dépendance en  $p_T$  est cette fois-ci ajustée avec un polynôme du premier degré. Les ajustements sont effectués séparément pour les trois années. Ces ajustements sont montrés en Fig. 5.16 et 5.17 pour l'année 2018 et illustrent leurs dépendances en MVA DM et  $N_{\text{jets}}$ .

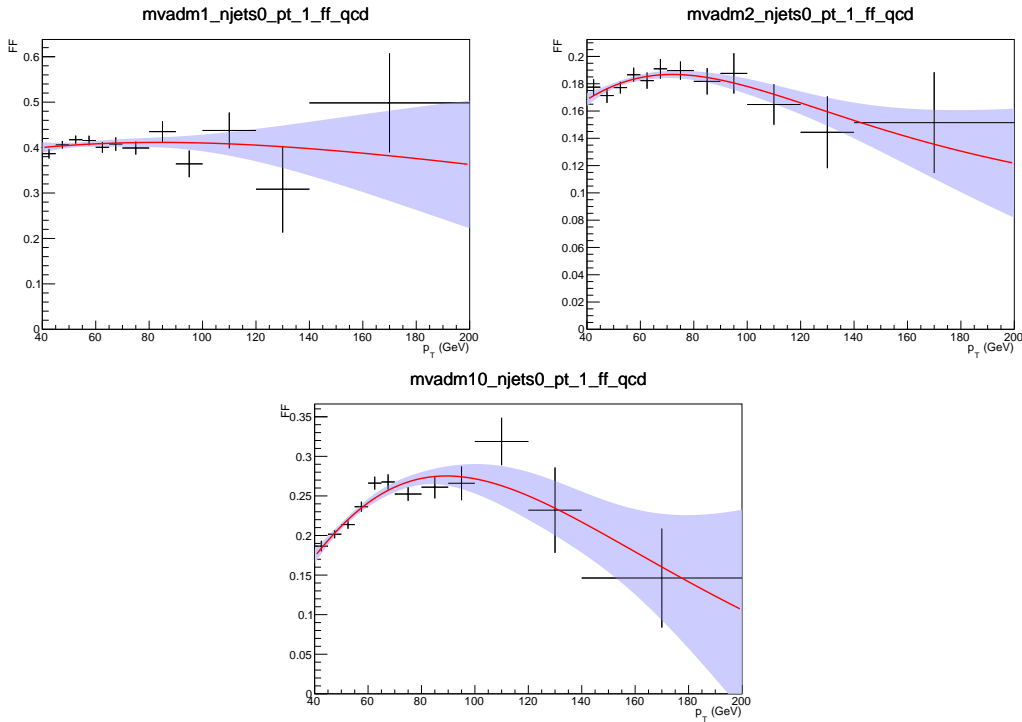


FIGURE 5.16 – Exemples d'ajustement des fake factors 2018 dans le canal  $\tau_h\tau_h$  en fonction du  $p_T$  illustrant la dépendance en MVA DM (0 (gauche), 1 (droite) et 10 (milieu)) [133].

Les FF mesurés sont ensuite corrigés pour tenir compte de potentielles dépendances non prises en compte et de différences entre DR et SR. On applique le FF aux événements

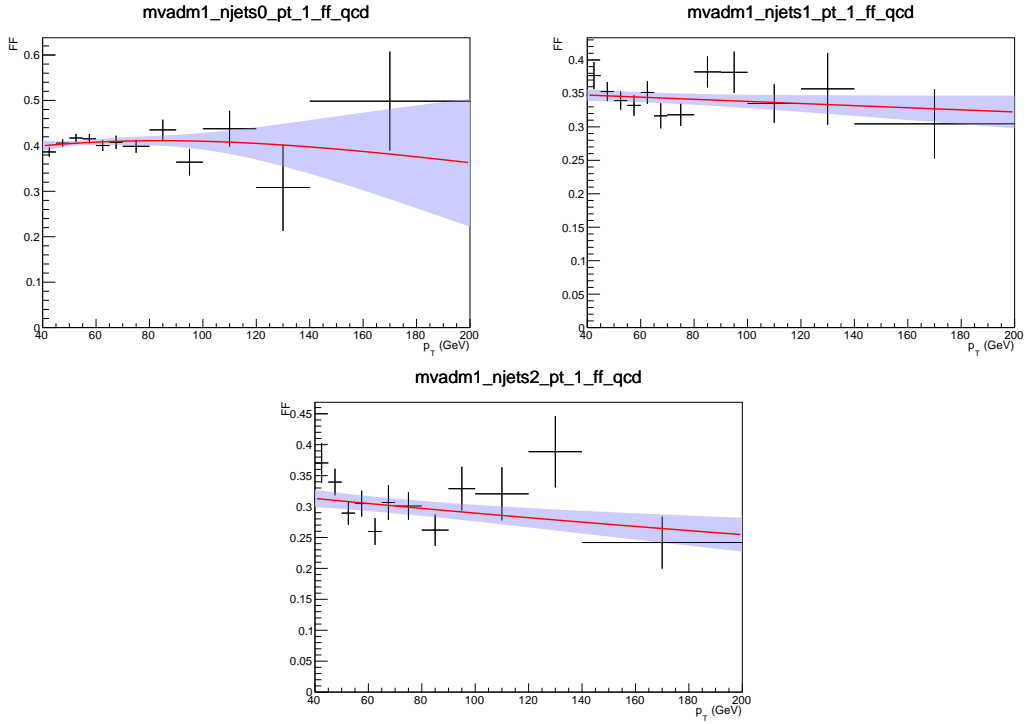


FIGURE 5.17 – Exemples d’ajustement des fake factors 2018 dans le canal  $\tau_h\tau_h$  en fonction du  $p_T$  illustrant la dépendance en  $N_{\text{jets}}$  (0 (gauche), 1 (droite) et  $\geq 2$  (milieu)) [133].

DR et on prend les différences résiduelles données/MC comme des corrections non-closure. D’importantes différences sont observées pour les variables liées à  $p_T^{\text{miss}}$  (sa magnitude  $E_T^{\text{miss}}$  et le  $\Delta\phi$  entre le tau et  $p_T^{\text{miss}}$ ) et sont les plus importantes pour  $N_{\text{jets}} = 0$ . La cause provient d’évènements où l’énergie des jets se faisant passer pour des  $\tau_h$  est mal mesurée, ce qui donne lieu à de fausses  $E_T^{\text{miss}}$  et change l’isolation du candidat  $\tau_h$ . Une variable C est définie dans le but de paramétrer cette correction :

$$C = \frac{E_T^{\text{miss}}}{p_T^{\tau_h}} \cos \Delta\phi(p_T^{\text{miss}}, \vec{\tau}_h), \quad (5.12)$$

avec  $\Delta\phi(p_T^{\text{miss}}, \vec{\tau}_h)$  la différence d’angle azimutal entre le  $\tau_h$  et  $p_T^{\text{miss}}$ . Cette variable représente la quantité de fausse  $E_T^{\text{miss}}$  alignée avec le  $\tau_h$ , par rapport au  $p_T$  du  $\tau_h$ . La correction est mesurée séparément pour les bins  $N_{\text{jets}} = 0, 1, \geq 2$  qui sont ajustés avec des polynômes. Cette correction est montrée en Fig. 5.18 pour l’année 2018.

Une correction supplémentaire est appliquée et prend en compte les différences entre la DR et AR/SR qui apparaissent s’il y a de légères différences de taux de faux dans les paires OS et SS. La cause est due à différentes compositions de saveurs de jets dans les évènements OS et SS, choses auxquelles les taux de faux sont dépendants. Ces compositions de saveurs ont pour conséquence une cinématique différente et donc

une correction OS/SS pouvant varier différemment. On sélectionne une seconde région orthogonale à la SR en inversant la condition d'isolement du tau candidat 2, c'est à dire qu'il passe le WP VVVLoose mais pas le VVLoose. Un ensemble de FF et de corrections non-closure sont dérivés pour les événements SS en utilisant exactement la même procédure citée ci-dessus. Ces FF sont ensuite appliqués aux événements OS anti-isolés et comparés aux événements OS isolés afin d'obtenir la correction. Cette dernière est calculée en bins de  $N_{\text{jets}}$ , est paramétrisée comme une fonction de  $\Delta R$  entre les deux taus candidats et est ajustée en utilisant un polynôme du premier degré. Ce choix est motivé par le fait que la composition en saveur du jet varie le plus fortement comme une fonction de ces variables, et plus significativement pour les événements avec  $N_{\text{jets}} = 0$  (les bins de plus haute multiplicité de jets ont une correction proche de l'unité). On pourrait en principe considérer d'autres variables pour cette paramétrisation, mais celle-ci prend en compte les plus grandes dépendances et donne de bons accords pour toutes les variables importantes.

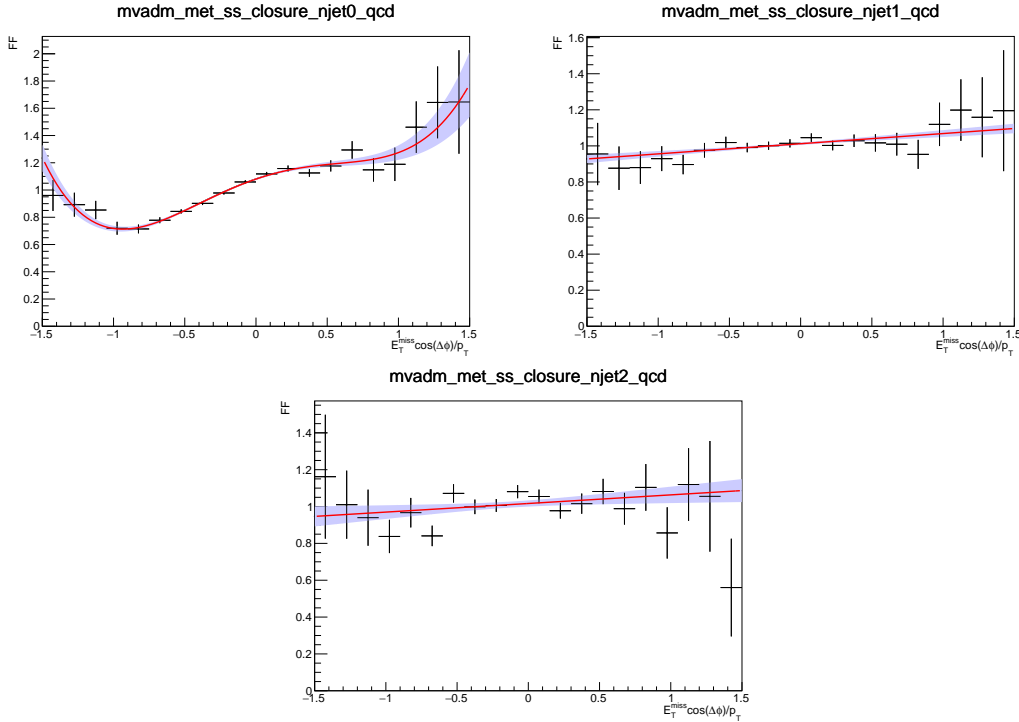


FIGURE 5.18 – Corrections non-closure pour les fake factors 2018 comme une fonction de  $C$  pour le canal  $\tau_h \tau_h$ . Cette correction est montrée pour 0 (gauche), 1 (droite) et  $\geq 2$  (milieu) jets [133].

L'AR est définie en exigeant que le tau de plus haut  $p_T$  passe le WP VVVLoose mais pas le Medium de DeepTau, les autres sélections restant identiques à la SR. Le bruit de fond  $j \rightarrow \tau_h$  dans la SR est alors estimé en appliquant les FF mesurés aux événements dans la AR. Les processus produisant de vrais  $\tau_h$  ou des leptons reconstruits en  $\tau_h$  sont soustraits en utilisant les échantillons simulés. Mais cette procédure seule mènerait à une sous-estimation des processus avec un seul jet reconstruit en  $\tau_h$ , où le tau candidat

2 est en réalité un jet comme dans les évènements  $W$ +jets. Ce type d'évènement est donc ajouté en utilisant les échantillons simulés.

### 5.6.1.2 Traitement des incertitudes statistiques

Les bandes d'incertitudes des fonctions ajustées pour les fake factors non corrigés sont utilisées pour estimer les incertitudes statistiques sur les fake factors mesurés. Si les valeurs les plus hautes et basses des bandes étaient prises comme une variation de  $\pm 1\sigma$ , les fake factors dans tout l'intervalle en  $p_T$  ne pourraient qu'augmenter ou diminuer simultanément pendant l'ajustement aux données. Or, on s'attend à ce que les fonctions ajustées aient au moins un point dans lequel l'incertitude change de signe et où les fake factors en dessous de ce point sont diminués tandis qu'ils augmentent au dessus. En principe, une incertitude par degré de liberté de la fonction ajustée pourrait être incluse pour réduire cet effet. En revanche, la forme fonctionnelle des fake factors ont typiquement quatre degrés de liberté, ce qui nécessiterait un très large ensemble de paramètres de nuisances à inclure (quatre par bin  $N_{\text{jets}}/MVA\text{-DM}$  par canal par processus par an).

Afin de réduire le nombre de paramètres de nuisances par environ un facteur deux, les incertitudes statistiques sont décomposées en deux contributions. La première affecte seulement la normalisation d'un bin en  $N_{\text{jets}}/MVA\text{-DM}$  (c'est-à-dire que l'incertitude relative est constante par rapport à  $p_T$ ). La seconde contribution change le signe à une valeur particulière de  $p_T$ , où en dessous de cette valeur le fake factor est diminué (augmenté) et au dessus augmenté (diminué) par le tirage vers le haut (bas) sur le paramètre de nuisance. Ces deux contributions doivent aussi s'additionner quadratiquement à l'incertitude totale.

La valeur en  $p_T$  où l'incertitude de forme change de signe,  $p_T^0$ , est déterminée par le point où la taille de l'incertitude relative totale est minimale. Cette taille en ce point détermine aussi l'incertitude de normalisation (car la seconde incertitude en ce point est nulle par construction). La magnitude de la seconde incertitude est alors déterminée comme une fonction de  $p_T$  en soustrayant de façon quadratique l'incertitude de normalisation ( $\sigma_1$ ) de l'incertitude totale ( $\sigma_{\text{tot}}$ ). L'incertitude de forme ( $\sigma_2$ ) est alors donnée par :

$$\sigma_2(p_T) = \begin{cases} -\sqrt{\sigma_{\text{tot}}^2 - \sigma_1^2}, & \text{si } p_T < p_T^0 \\ +\sqrt{\sigma_{\text{tot}}^2 - \sigma_1^2} & \text{autrement} \end{cases}. \quad (5.13)$$

Des exemples d'incertitudes décomposées sont montrés en Fig. 5.19. Le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  étudié dans mon analyse utilise ces mêmes fake factors. Cependant, leurs mesures n'a pas pris en compte la présence obligatoire d'un vertex secondaire par tau, requis pour la méthode du vecteur polarimétrique. Une vérification est donc effectuée sur la distribution du score BDT de la catégorie jet fakes (expliqué en section 5.7) avec des évènements



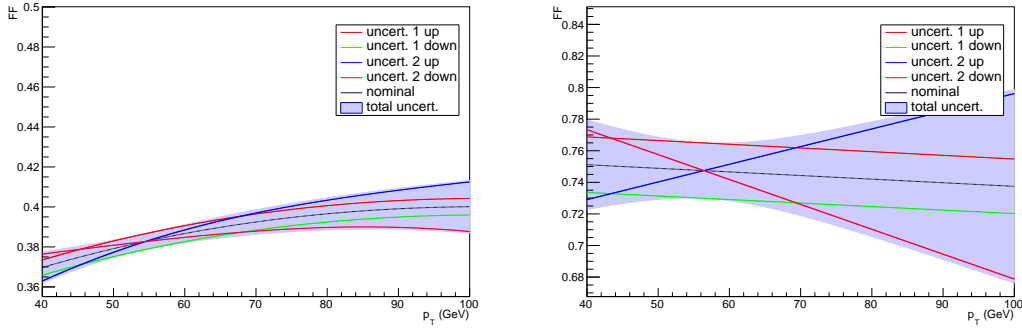


FIGURE 5.19 – Exemples de décompositions des bandes d’incertitudes statistiques en deux contributions. Une affecte seulement la normalisation notée « uncert. 1 » dans la légende, et la seconde affecte aussi la forme notée « uncert. 2 » dans la légende. Les exemples sont la forme fonctionnelle standard Landau+fonction constante (gauche) et un ajustement plus simple d’un polynôme du premier degré (droite) [133].

possédant une paire de taus de mêmes signes en Fig. 5.20. Seules les incertitudes statistiques sont affichées. Aucun problème particulier n’est présent dans l’emploi des fake factors avec l’ajout de la présence de vertex secondaires.

## 5.7 BDT de classification

La classification des évènements est faite en utilisant un BDT. Le canal  $\tau_h\tau_h$  est dominé par des bruits de fond contenant des contributions de vrais et faux taus, principalement des évènements  $Z \rightarrow \tau\tau$  et multijets. Le BDT est alors entraîné pour définir trois catégories :

- Higgs : combine les échantillons  $H \rightarrow \tau\tau$  incluant ggH, VBF et VH ;
- ztt embedded : processus contenant de vrais leptons  $\tau$  ;
- jets fakes : processus contenant de faux leptons  $\tau$ .

Les échantillons ggH, VBF et VH sont pondérés par leur section efficace avant de fusionner en une seule classe d’entraînement, puis les trois classes d’entraînement (Higgs, ztt embedded et jets fakes) sont normalisées pour tenir compte des différences de datasets grâce à une classe de poids qui pondère chaque catégorie d’entraînement afin qu’elles soient traitées de manière égale durant l’entraînement :

$$c_n = \frac{\sum_{i \in \text{events inclusively}} w_i}{\sum_{j \in \text{events} \in \text{category } n} w_j}, \quad (5.14)$$

avec  $i$  parcourant l’ensemble complet des données d’entraînement et  $j$  parcourant les évènements associés à une classe pour laquelle le poids est calculé.

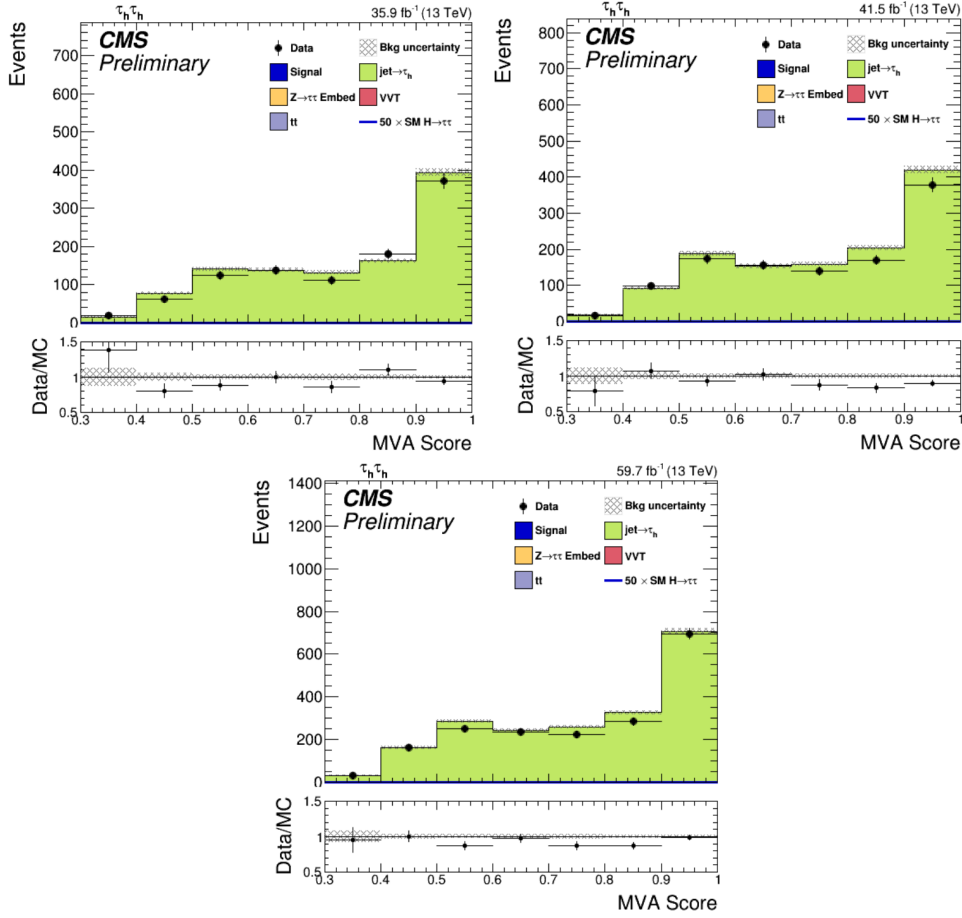


FIGURE 5.20 – Distribution du score BDT de la catégorie jet fakes avec des événements possédant une paire de taus de mêmes signes pour 2016 (gauche), 2017 (droite) et 2018 (milieu). La couleur verte représente les échantillons modélisant des jets se faisant passer pour des leptons taus. Les bandes grises sont seulement les incertitudes statistiques.

Des tests ont été effectués afin de montrer qu'il y a une différence négligeable comparée aux sensibilités attendues obtenues par entraînement sur des échantillons repondérés pour l'état pseudoscalaire du boson de Higgs. Ceci permet de montrer que l'entraînement est indépendant de l'état CP du boson de Higgs utilisé et que les variables d'entrée ne biaisent pas les résultats.

Pour l'entraînement, les lots de données sont séparés en numéros d'événements pair et impair qui à leur tour sont séparés en lots de données d'entraînement et de validation. Ce genre de validation croisée permet à des BDT séparés d'être entraînés sur blocs pair et impair et peuvent ainsi être testés en appliquant le BDT entraînés sur le bloc pair sur le bloc impair et inversement. Cela assure que toute la statistique disponible est utilisée pour l'entraînement et qu'il n'y a pas de biais introduit. Finalement, les sélections décrites en [section 5.4](#) sont appliquées à l'entraînement. Les variables suivantes

sont utilisées en entrée pour l'entraînement du BDT :

- $p_T$  du candidat  $\tau_h$  de plus haute impulsion transverse ;
- $p_T$  visible du système di- $\tau_h$  ;
- $p_T$  total du système di- $\tau_h + \text{MET}$  ;
- masse visible du système di- $\tau_h$  ;
- masse SVfit du système di- $\tau_h$  ;
- $p_T$  du jet de plus haute impulsion transverse ;
- MET ;
- nombre de jets ;
- masse invariante du système di-jets ;
- différence absolue entre les pseudorapidités des deux jets de plus haut  $p_T$ .

Si  $N_{\text{jets}} = 0, 1$ , les variables reliées au système di-jets ne sont pas utilisées. Si  $N_{\text{jets}} \geq 3$ , les deux jets de plus hautes impulsions transverses sont utilisés pour ces variables. Les distributions des variables d'entrée pour les trois années dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  sont montrées en Fig. 5.21 – 5.24. Le signal MS est aussi superposé et multiplié par 50. Le ratio données/MC est affiché en dessous de chaque distribution et les bandes grises d'incertitudes sont purement statistiques.

Dans le cadre de l'analyse [117], un entraînement séparé a été effectué pour chaque année. Le BDT fournit trois sorties : Higgs, ztt embedded et jets fakes. Chaque événement est donc assigné, suivant le plus haut score obtenu, à une de ces trois catégories. Les distributions de ces catégories pour le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  se trouvent en Fig. 5.25 – 5.27.

## 5.8 Lissage des modèles

Les modèles de signaux et de bruits de fond utilisés dans cette analyse peuvent souffrir d'une statistique limitée, notamment le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  étudié avec la méthode du vecteur polarimétrique. Appliquer un lissage à ces modèles permet de réduire l'impact des incertitudes dites intervalle-par-intervalle (bin-by-bin) sur les résultats. De plus, cela réduit aussi la possibilité que le résultat soit biaisé à cause des fluctuations statistiques.

Afin de minimiser ces impacts, les modèles sont lissés en exploitant les symétries connues des distributions. Ces symétries sont celles des modèles étant plats en fonction de  $\phi_{\text{CP}}$ , et des modèles (anti)symétriques par rapport à  $\phi_{\text{CP}} = \pi$ . Ces deux types seront nommés « aplatissement » et « symétrisation » respectivement.

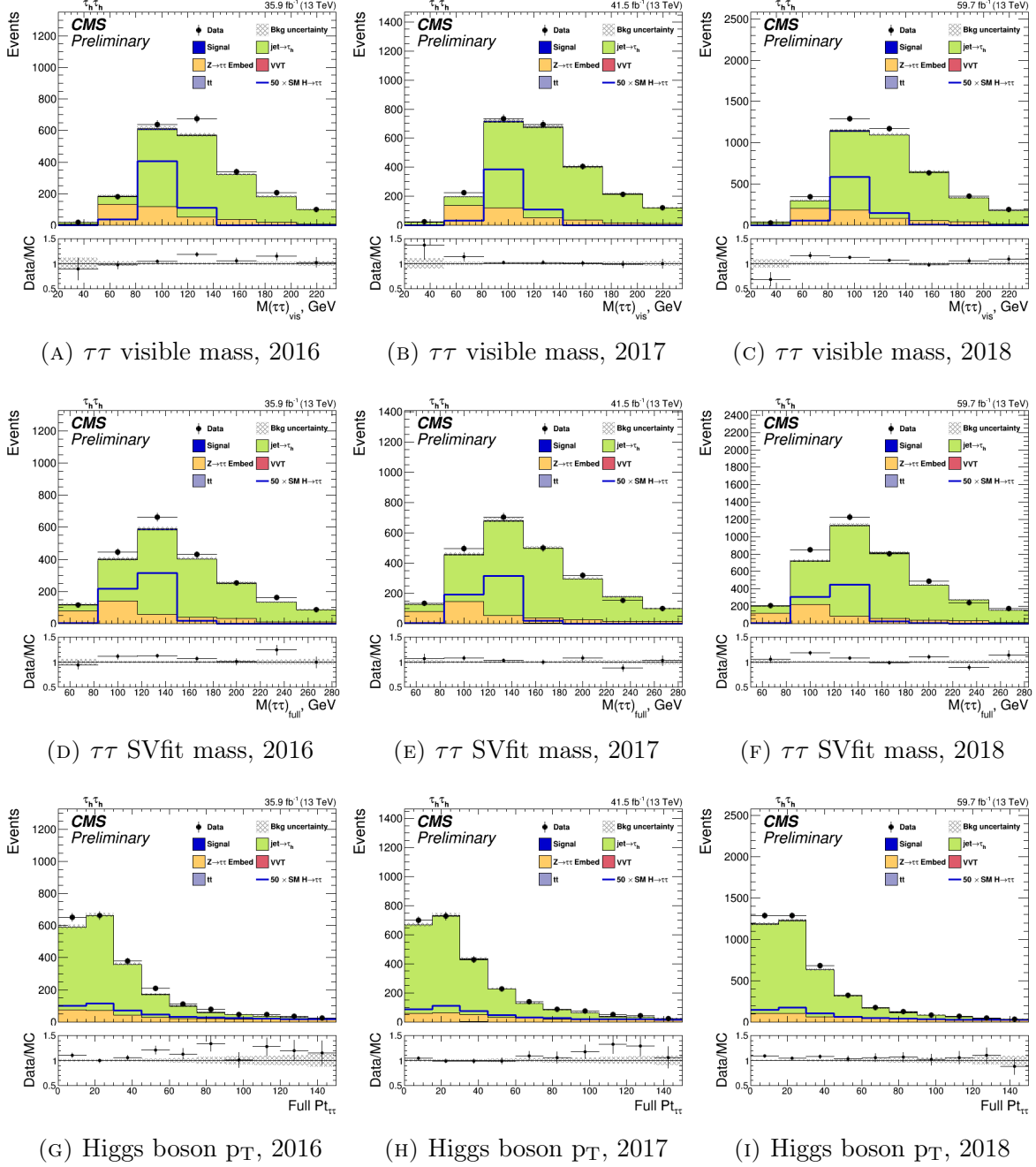


FIGURE 5.21 – Distribution avant ajustement final pour 2016, 2017 et 2018 des variables d'entrée du BDT de classification dans le canal  $a_1^{3pr} a_1^{3pr}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

Chapitre 5. Mesure de l'état CP du boson de Higgs à  $\sqrt{s} = 13$  TeV

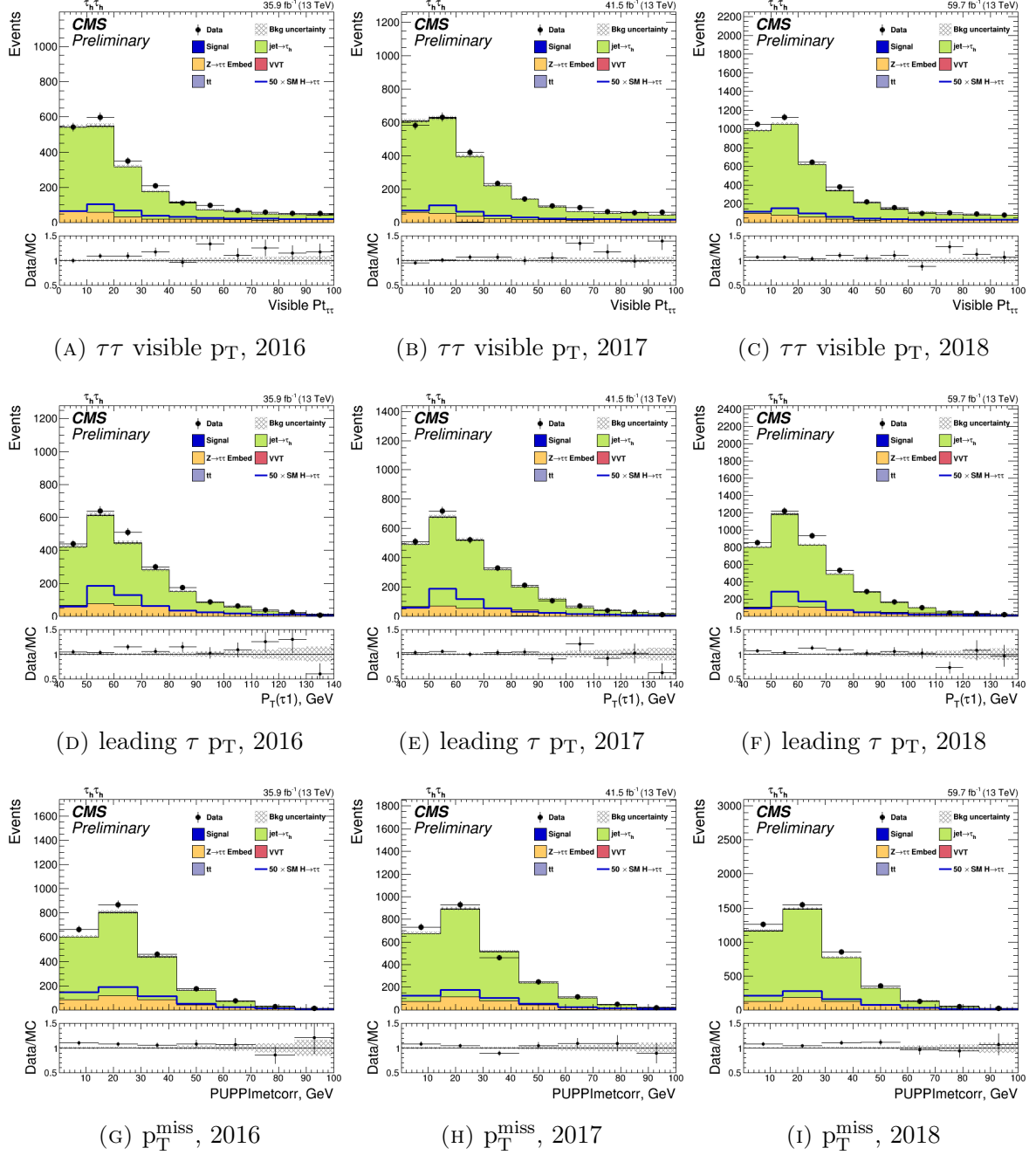


FIGURE 5.2 – Distribution avant ajustement final pour 2016, 2017 et 2018 des variables d'entrée du BDT de classification dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

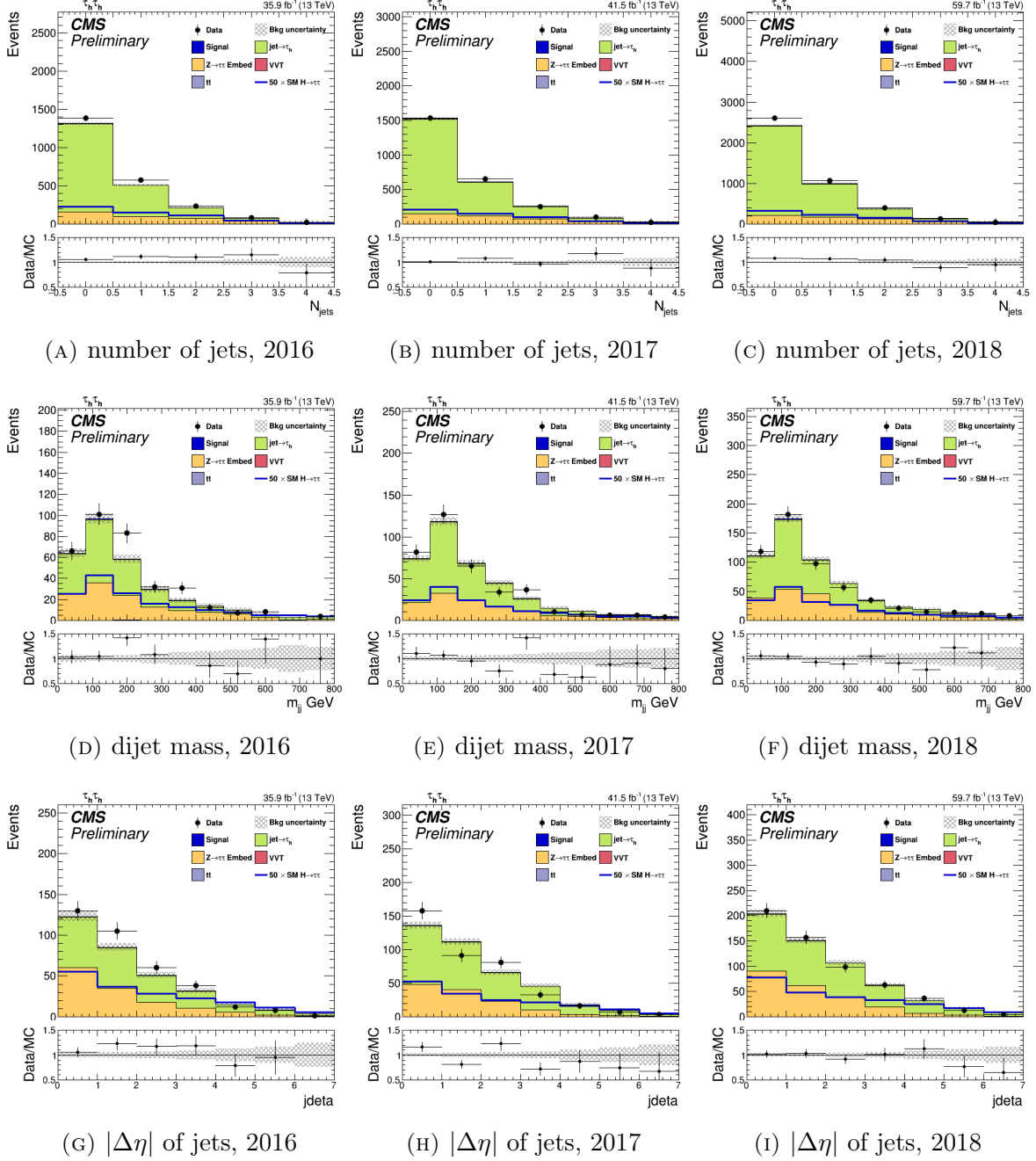


FIGURE 5.23 – Distribution avant ajustement final pour 2016, 2017 et 2018 des variables d'entrée du BDT de classification dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

Chapitre 5. Mesure de l'état CP du boson de Higgs à  $\sqrt{s} = 13$  TeV

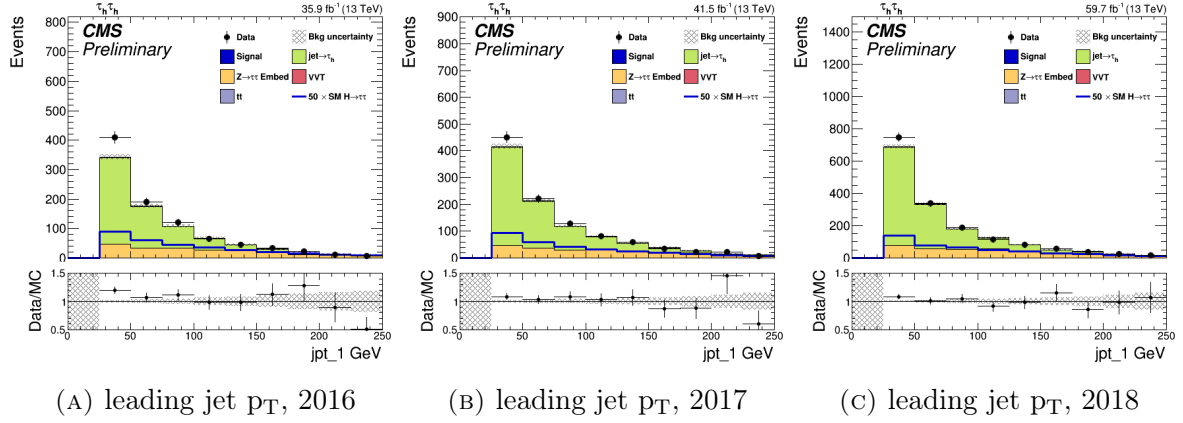


FIGURE 5.24 – Distribution avant ajustement final pour 2016, 2017 et 2018 des variables d'entrée du BDT de classification dans le canal  $a_1^{3pr} a_1^{3pr}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

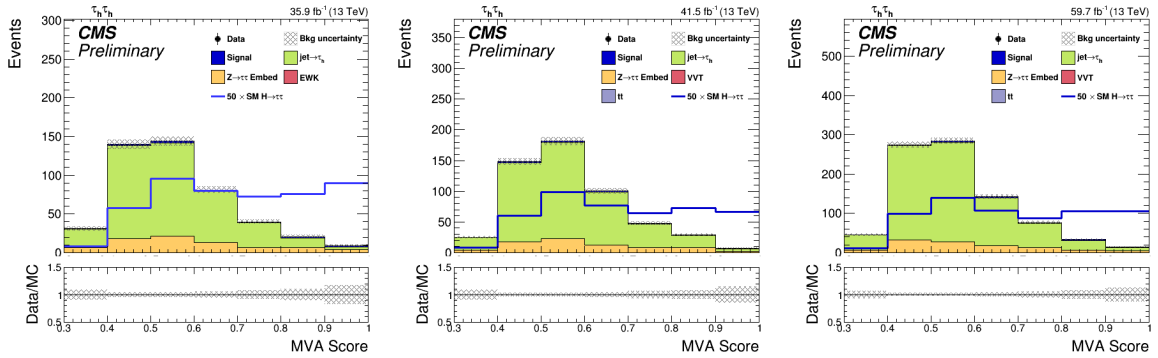


FIGURE 5.25 – Distribution des catégories Higgs avant ajustement final pour 2016, 2017 et 2018 dans le canal  $a_1^{3pr} a_1^{3pr}$  avec les données cachées. Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

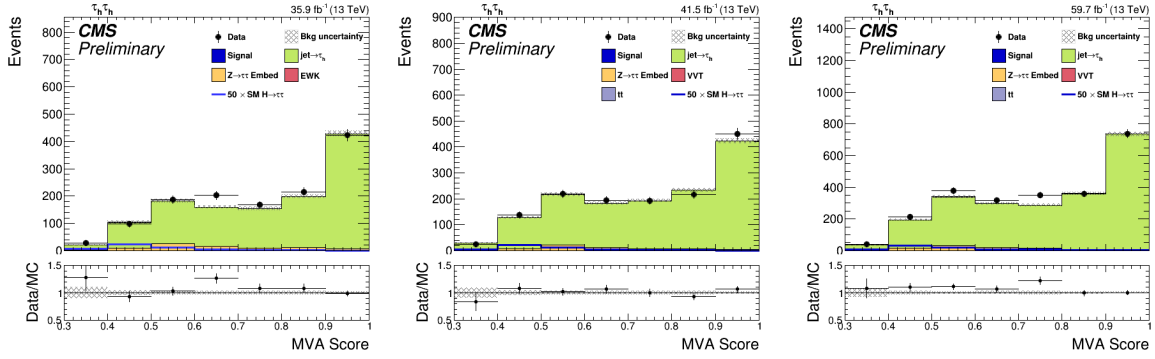


FIGURE 5.26 – Distribution des catégories jets fakes avant ajustement final pour 2016, 2017 et 2018 dans le canal  $a_1^{3pr} a_1^{3pr}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.

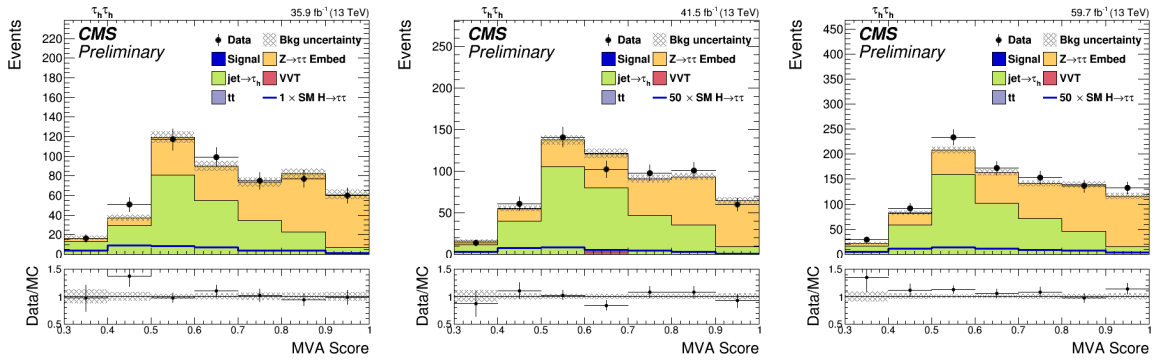


FIGURE 5.27 – Distribution des catégories ztt embedded avant ajustement final pour 2016, 2017 et 2018 dans le canal  $a_1^{3pr} a_1^{3pr}$ . Le signal est superposé et multiplié par 50. Les bandes d'incertitudes sont seulement statistiques.



En dehors des effets de reconstruction, tous les bruits de fond possédant de vrais leptons taus sont connus pour être plats en fonction de  $\phi_{\text{CP}}$ . De plus, un smearing indépendant des quadrivecteurs n'introduirait pas de biais vers des valeurs particulières de  $\phi_{\text{CP}}$ , car il n'y aurait pas de corrélation. Un exemple notable de biais induit par smearing est celui du smearing du PV qui décale les deux paramètres d'impact dans les canaux  $\pi + \pi$  et  $\mu + \pi$ . La conséquence est que  $\phi_{\text{CP}}$  est piqué en  $\phi_{\text{CP}} = 0, 2\pi$ . Cet effet est illustré en Fig. 5.28 A) qui montre au niveau générateur  $\phi_{\text{CP}}$  avec et sans smearing du PV. De mauvais alignements de détecteurs peuvent aussi introduire de tels biais mais ces effets seraient petits et l'analyse non sensible à ces derniers.

Dans le cas où la distribution ne serait pas plate, une seconde symétrie peut être exploitée et cette dernière ne devrait pas être détruite par des effets de reconstruction. Le signe de la variable  $O^*$  définie en Eq. 5.1) détermine si un évènement va dans l'intervalle  $\phi_{\text{CP}} < \pi$  ou  $\phi_{\text{CP}} > \pi$ . L'effet du smearing du PV dans ce cas n'affecte pas différemment les régions  $O^* < 0$  et  $O^* > 0$  comme illustré en Fig. 5.28 B) et donc ne détruit pas la symétrie autour de  $\phi_{\text{CP}} = \pi$ .

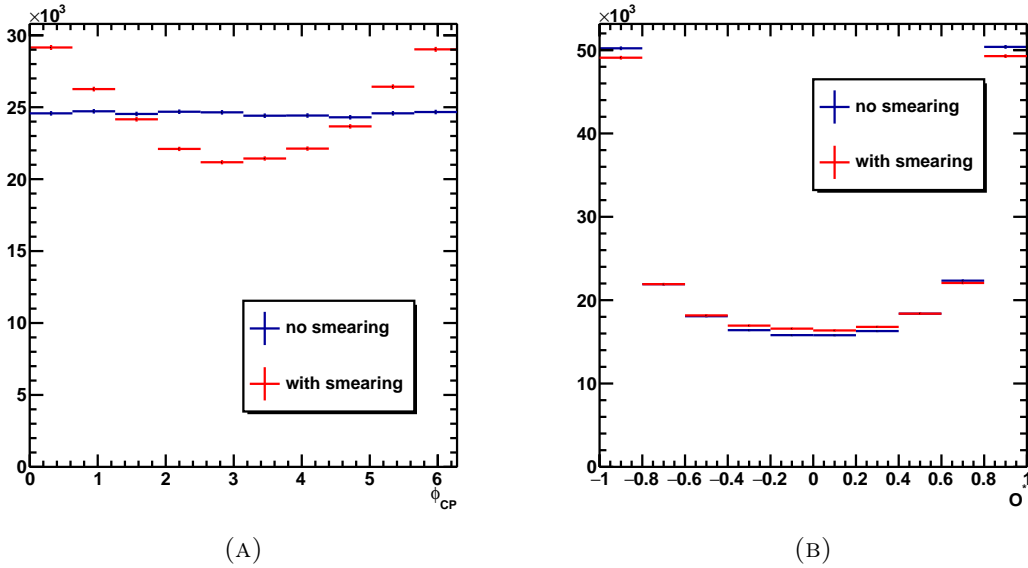


FIGURE 5.28 – Gauche : distribution au niveau générateur de  $\phi_{\text{CP}}$  pour des évènements  $\pi + \pi$  sans aucun smearing du PV (bleu) et après smearing du PV (rouge). Droite : distribution au niveau générateur de  $O^*$  (défini en Eq. 5.1) pour des évènements  $\pi + \pi$  sans aucun smearing du PV (bleu) et après smearing du PV (rouge) [133].

Pour toutes ces raisons, les bruits de fond possédant deux vrais leptons taus sont aplatis à l'exception des canaux  $\pi + \pi$  et  $\mu + \pi$  qui sont symétrisés. Les bruits de fond avec  $j \rightarrow \tau_h$  ont en général une forme symétrique mais non plate à cause d'effets cinématiques qui créent une préférence pour les valeurs  $\phi_{\text{CP}} = \frac{\pi}{2}$ . Tous les modèles  $j \rightarrow \tau_h$

dans toutes les catégories sont alors symétrisés plutôt qu'aplatés.

Concernant les plus petits bruits de fond restant qui incluent au moins un vrai  $\tau_h$  (le plus grand étant  $t\bar{t}$ ), les modèles semblent suivre la même tendance que les bruits de fond possédant deux vrais  $\tau_h$ . C'est-à-dire que les distributions sont plates dans toutes les catégories à part les canaux  $\pi + \pi$  et  $\mu + \pi$  où elles sont symétriques. Par conséquent, les mêmes procédures sont appliquées.

Les processus de signaux CP-pairs et impairs possèdent aussi une symétrie autour de  $\phi_{CP} = \pi$  qui peut être exploitée. Le scénario CP-mixte ( $\phi_{\tau\tau} = 45^\circ$ ) n'est pas symétrique autour de  $\phi_{CP} = \pi$ , mais la contribution de l'interférence au modèle est antisymétrique, c'est-à-dire qu'après soustraction des contributions CP-pair et impair au modèle, la valeur contenue dans les bins est égale mais leur signe est opposé. L'antisymétrie est alors exploitée en pondérant le processus de signal pour  $\phi_{\tau\tau} = -45^\circ$  (qui a le signe opposé de l'interférence) et en appairant les bins de la région  $\phi_{CP} < \pi$  du modèle  $\phi_{\tau\tau} = 45^\circ$  avec ceux de la région  $\phi_{CP} > \pi$  du modèle  $\phi_{\tau\tau} = -45^\circ$ . La procédure de symétrisation implique de poser le contenu,  $c$ , de deux bins symétriques ( $i$  et  $j$ ) égal à leur moyenne :

$$c'_i = c'_j = \frac{1}{2}(c_i + c_j), \quad (5.15)$$

avec ' signifiant la valeur après symétrisation. Les incertitudes statistiques,  $u$ , des bins sont posées comme :

$$u'_i = u'_j = \frac{1}{2}\sqrt{u_i^2 + u_j^2}. \quad (5.16)$$

Pour l'antisymétrisation, la même procédure est appliquée à l'exception que le bin  $i$  provient du modèle  $\phi_{\tau\tau} = 45^\circ$  tandis que le bin  $j$  provient du modèle  $\phi_{\tau\tau} = -45^\circ$ .

La procédure d'aplatissement implique de poser le contenu de  $N$   $\phi_{CP}$  bins égal à leur moyenne :

$$c'_j = \frac{1}{N} \sum_{i=1}^N c_i \text{ pour } 1 \leq j \leq N. \quad (5.17)$$

L'incertitude statistique dans ce cas est :

$$u'_j = \frac{1}{N} \sqrt{\sum_{i=1}^N u_i^2} \text{ pour } 1 \leq j \leq N. \quad (5.18)$$

Dans les deux cas, les incertitudes statistiques bin-by-bin sont correctement corrélées durant l'ajustement aux données et le nombre d'incertitudes indépendantes bin-by-bin est réduite d'un facteur  $\frac{1}{2}$  pour les modèles symétrisés et par  $\frac{1}{N}$  pour les modèles aplatés.

Les choix faits pour le lissage des modèles de bruits de fond sont validés de deux façons. D'abord, des tests  $\chi^2$  sont effectués et démontrent que les distributions lissées sont compatibles avec les distributions non lissées au vu des variations statistiques.

Ensuite, une étude dite toy est effectuée afin de vérifier si le résultat est biaisé par la procédure de lissage :

- un ensemble de données toy est généré selon les modèles non lissés en autorisant les fluctuations statistiques dans le contenu des bins (donc pas de variation d'incertitude systématique) ;
- un ensemble de données toy est généré selon les modèles lissés en autorisant aussi les fluctuations statistiques seulement ;
- ces deux ensembles sont ajustés en utilisant un modèle de bruit de fond lissé et des comparaisons sont faites des distributions de la meilleure valeur d'ajustement de  $\phi_{\tau\tau}$  et des profils de la fonction de vraisemblance  $-2\Delta \ln L = -2(\ln L(\phi_{\tau\tau} = X) - \ln(\hat{\phi}_{CP}))$ , où X correspond au scénario de l'angle de mélange  $\phi_{\tau\tau}$  supposé pour le signal lors de la génération des toys.

La procédure ci-dessus est appliquée pour des scénarios d'angle de mélange  $X = 0^\circ$ ,  $45^\circ$ , et  $90^\circ$ . L'idée derrière ces tests est de supposer que les données sont distribuées selon les modèles non lissés puis de vérifier si les choix de modèles lissés pour les signaux et bruits de fond biaisent les résultats en supposant que les données ne sont pas distribuées selon le modèle lissé.

Les distributions de  $\phi_{\tau\tau}$  et  $-2\Delta \ln L$  pour O(200) toys sont montrées en Fig. 5.29 et 5.30 pour chaque scénario d'angle de mélange. Les distributions pour les toys lissés et non lissés sont très similaires tout comme les valeurs moyennes des distributions indiquées par des flèches. Ceci démontre que le choix du modèle de bruit de fond lissé ne biaise pas les résultats.

## 5.9 Incertitudes

### 5.9.1 Incertitudes de normalisation

Les incertitudes n'affectant que la normalisation des bruits de fond et signaux dans le canal  $a_1^{3pr} a_1^{3pr}$  sont :

- **Luminosité**

L'incertitude de luminosité s'élève à 2.5%, 2.3% et 2.5% sur les données collectées en 2016, 2017 et 2018 respectivement. Elle est appliquée à tous les processus de bruits de fond et de signaux où le nombre d'événements est estimé à partir des MC. Cette incertitudes est partiellement corrélée entre les années comme recommandé en [136] ;

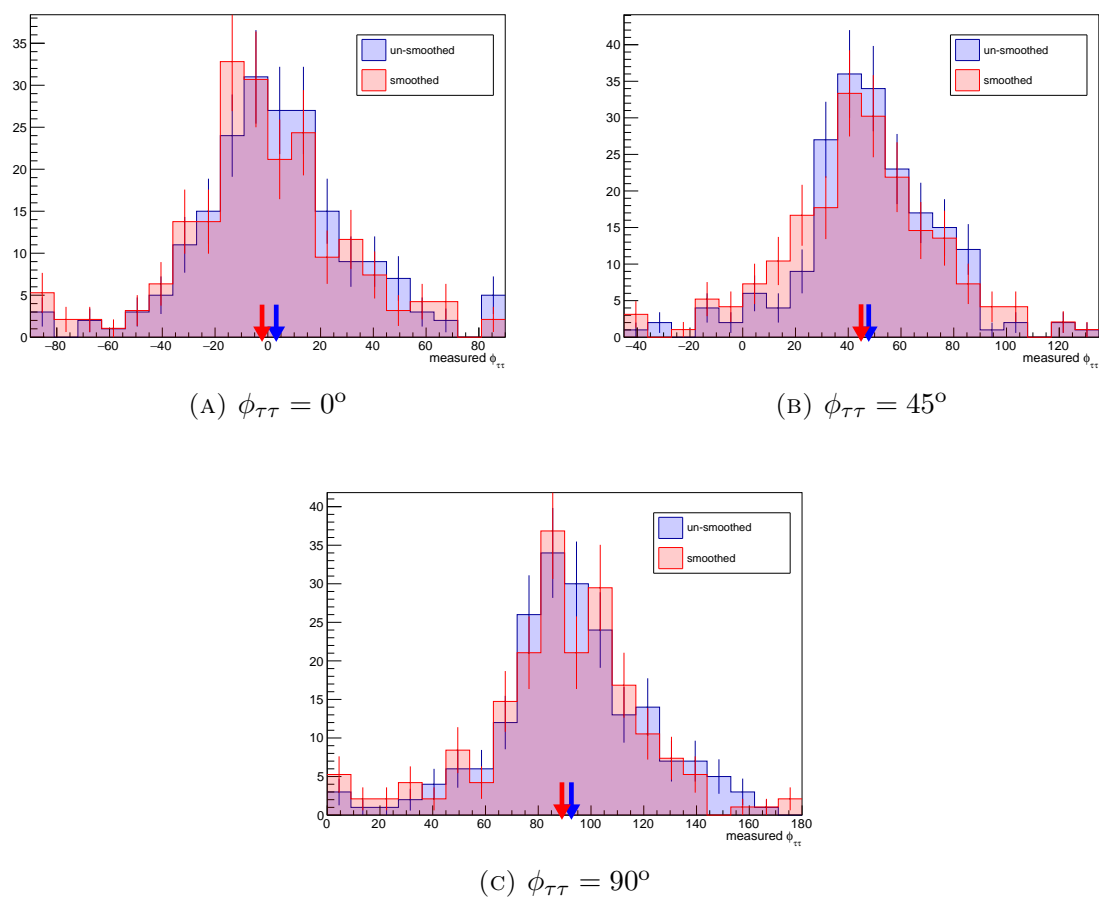


FIGURE 5.29 – Meilleures valeurs d’ajustement de  $\phi_{\tau\tau}$  pour des expériences toys générées selon un modèle de bruit de fond lissé (rouge) et non lissé (bleu). Les flèches indiquent la valeur moyenne des distributions. Ces dernières sont affichées pour trois scénarios différents de l’angle de mélange du signal  $\phi_{\tau\tau} = 0^\circ$  (A),  $45^\circ$  (B), et  $90^\circ$  (C) [133].

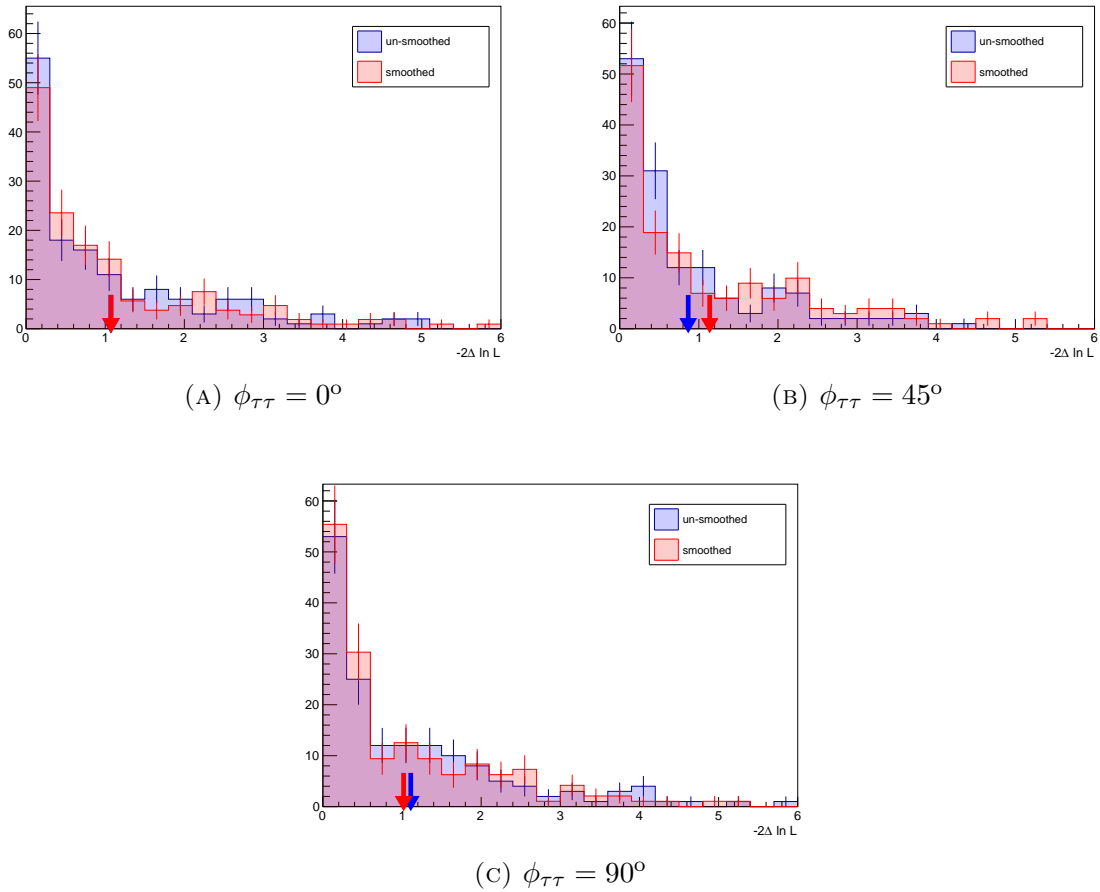


FIGURE 5.30 – Meilleures valeurs d'ajustement de  $-2\Delta \ln L = -2(\ln L(\phi_{\tau\tau} = X) - \ln(\hat{\phi}_{CP}))$  pour des expériences toys générées selon un modèle de bruit de fond lissé (rouge) et non lissé (bleu). Les flèches indiquent la valeur moyenne des distributions. Ces dernières sont affichées pour trois scénarios différents de l'angle de mélange du signal  $\phi_{\tau\tau} = 0^\circ$  (A),  $45^\circ$  (B), et  $90^\circ$  (C) [133].

- **Incertitudes de normalisation du bruit de fond**

4% d'incertitude est appliqué sur le nombre d'évènements des échantillons embedded. Cela provient de l'incertitude sur la mesure de l'identification du muon et des efficacités du système de déclenchement utilisées pour normaliser les échantillons embedded (2% par muon donc 4% au total). Ces incertitudes sont non corrélées entre les années ;

4.2% d'incertitude est appliqué aux processus  $t\bar{t}$  afin de tenir compte de l'incertitude sur la section efficace qui est entièrement corrélée entre les années ;

2% d'incertitude est appliqué aux processus  $Z \rightarrow ll$  afin de tenir compte de l'incertitude sur la section efficace Drell-Yan entièrement corrélée entre les années ;

5% d'incertitude est appliqué aux processus di-boson et single-top afin de tenir compte de l'incertitude sur la section efficace de production entièrement corrélée entre les années ;

4% d'incertitude est appliqué aux processus électrofaibles afin de tenir compte de l'incertitude sur la section efficace de production entièrement corrélée entre les années ;

- **Incertitudes de normalisation du signal**

Les incertitudes sur la section efficace de production ggH sont appliquées comme recommandé en [137]. Elles incluent une incertitude de 3.9% due aux variations des échelles QCD et une de 3.2% due aux variations des fonctions de distribution de partons et  $\alpha_s$ . Ces incertitudes sont entièrement corrélées entre les années ;

Les incertitudes sur la section efficace de production VBF sont appliquées comme recommandé en [137]. Elles incluent une incertitude de 0.4% due aux variations des échelles QCD et une de 2.1% due aux variations des fonctions de distribution de partons et  $\alpha_s$ . Ces incertitudes sont entièrement corrélées entre les années ;

Les incertitudes sur la section efficace de production WH (ZH) sont appliquées comme recommandé en [137]. Elles incluent une incertitude de 0.7% (3.8%) due aux variations des échelles QCD et une de 1.9% (1.6%) due aux variations des fonctions de distribution de partons et  $\alpha_s$ . Ces incertitudes sont entièrement corrélées entre les années ;

Les incertitudes recommandées sur les rapports d'embranchement  $H \rightarrow \tau\tau$  données en [137] sont appliquées à tous les signaux Higgs (ggH, VBF, et VH). Elles

Groupe	Membres du groupe	Corrélation
Absolute	AbsoluteMPFBias, AbsoluteScale, Fragmentation, PileUpDataMC, PileUpPtRef, RelativeFSR, SinglePionECAL, SinglePionHCAL	oui
Absolute_year	AbsoluteStat, RelativeStatFSR, TimePtEta	non
BBEC1	PileUpPtBB, PileUpPtEC1, RelativePtBB	oui
BBEC1_year	RelativeJEREC1, RelativePtEC1, RelativeStatEC	non
EC2	PileUpPtEC2	oui
EC2_year	RelativeJEREC2, RelativePtEC2	non
FlavorQCD	FlavorQCD	oui
HF	PileUpPtHF, RelativeJERHF, RelativePtHF	oui
HF_year	RelativeStatHF	non
RelativeBal	RelativeBal	oui
RelativeSample	RelativeSample	non

TABLE 5.7 – Groupes d'incertitudes JES utilisés dans l'analyse. La dernière colonne indique la corrélation entre les années.

incluent une incertitude de 1.7% due aux ordres supérieurs de corrections manquants, une de 0.99% due à la paramétrisation des masses des quarks et une de 0.62% due à la paramétrisation de  $\alpha_s$  ;

- **Vertex secondaire**

Une incertitude de 2% est appliquée aux échantillons embedded, jet  $\rightarrow \tau_h$  et MC pour tenir compte de la présence requise du vertex secondaire dans le canal  $a_1^{3pr} a_1^{3pr}$  avec la méthode du vecteur polarimétrique. Ces incertitudes sont décorréliées entre elles.

## 5.9.2 Incertitudes de forme

Certaines incertitudes de forme sont un mélange de forme et de normalisation. De tels cas sont indiqués dans les descriptions. Les incertitudes de formes utilisées dans le canal  $a_1^{3pr} a_1^{3pr}$  sont :

- **Efficacités de reconstruction/identification des taus**

Pour les candidats  $\tau_h$  correspondant aux  $\tau_h$  au niveau générateur, les corrections d'identification sont calculées en deux bins de  $p_T$  (20–40 et  $>40$  GeV, mais la sélection des  $\tau_h$  à 40 GeV rend le premier bin inutilisé), et en 5 bins de MVA DM (MVA DM = 0, 1, 2, 10, 11, mais le canal  $a_1^{3pr} a_1^{3pr}$  n'utilise que MVA DM = 10). L'incertitude dépend aussi du  $p_T$  et du MVA DM et provient des incertitudes des ajustements des corrections d'identification comme décrit en sous-section 5.5.2, et est de l'ordre de 3% et statistiquement dominée.

Des incertitudes de forme non corrélées sont considérées pour chaque bin nommé précédemment et sont non corrélées entre les années car elle sont principalement statistiques. Ces incertitudes couvrent de potentielles différences de nombre d'évènements dans chaque bin en MVA DM qui peuvent être dû à une mauvaise modélisation du HPS DM, des discriminants DeepTau, ou du discriminant MVA DM. En revanche, comme le pouvoir de séparation du discriminant  $\phi_{CP}$  dépend du véritable mode de désintégration du tau sélectionné dans chaque bin en MVA DM, une incertitude supplémentaire pourrait être considérée pour la migration des différents modes de désintégrations entre les bins MVA DM. Ces effets ont été étudiés en utilisant des valeurs exagérées d'incertitudes, mais ils furent négligeables comparé aux incertitudes statistiques des modèles de signaux.

De plus, une incertitude sur le nombre d'évènements de 3% est prise en compte pour les vrais  $\tau_h$  par rapport aux discriminants anti électron et anti muon ;

- **Efficacités du système de déclenchement des taus**

Pour les systèmes de déclenchement sur des candidats  $\tau_h$ , l'incertitude dépend du  $p_T$  et du MVA DM, et provient des ajustements des turn-on décrits en 5.5.3. Ces incertitudes sont non corrélées entre les MVA DM, sont principalement statistiques et sont donc non corrélées entre les années ;

- **Incertitudes de TES**

La TES des vrais  $\tau_h$  est corrigée par mode de désintégration. L'incertitude dépend du mode de désintégration et de l'année. Pour les événements MC cela varie entre 0.8 et 1.1%. Pour les événements embedded cela varie entre 0.2 et 0.5%. Les incertitudes sont non corrélées entre les modes de désintégrations et les années. Le décalage de l'énergie du  $\tau_h$  induit par la TES est propagé à la MET ;

- **Incertitudes de JES**

Il y a 27 sources d'incertitudes qui sont rassemblées en différents groupes en se basant sur les corrélations entre les incertitudes, principalement reliés aux régions du détecteur. Ces groupes sont détaillés en tableau 5.7, qui indique aussi comment les incertitudes sont corrélées entre les années. Pour les échantillons MC sans correction du recul appliquée, les incertitudes sont aussi propagées à la MET ;

- **Incertitudes de JER**

Les jets dans la simulation sont smearés afin de correspondre à la résolution en énergie des jets dans les données. Les incertitudes dans le smearing sont prise en compte comme une incertitude de forme supplémentaire et sont non corrélées entre les années ;

- **Incertitudes de l'énergie de la MET non groupées**

L'incertitude de l'énergie de la MET non groupée est appliquée à tous les MC qui n'ont pas de correction de recul appliquée ( $t\bar{t}$  et processus VV) [138]. L'incertitude est non corrélée entre les années ;



- **Incertitudes de la MET avec correction de recul**

La réponse hadronique et sa résolution sont modifiées pour tous les processus MC suivant une mesure effectuée durant le calcul de la correction de recul. L'incertitude est non corrélée entre les années ;

- **Contamination des échantillons embedded**

Les échantillons embedded sont une estimation de tous les processus produisant une paire de vrais leptons taus et contiennent donc une petite quantité d'évènements  $t\bar{t}$  et  $VV$ . On applique donc une incertitude de forme définie comme  $\pm 10\%$  de ces contributions appliquées à ces MC. Ceci couvre les biais introduits par la procédure d'embedding en modélisant ces évènements. L'incertitude est entièrement corrélée entre les années ;

- **Repondération du  $p_T$  du top**

Une incertitude entre aucune et deux fois la correction de la repondération du  $p_T$  du top est appliquée aux évènements  $t\bar{t}$ . L'incertitude est entièrement corrélée entre les années ;

- **Repondération du  $p_T$  du Z**

Une incertitude de  $\pm 10\%$  de la correction du  $p_T$  du Z est appliquée dans les évènements  $Z \rightarrow ll$ . L'incertitude est corrélée entre les années 2017 et 2018 et est non corrélée pour 2016 à cause de spectres en  $p_T$  différents au niveau générateur ;

- **Incertitudes des fake factors**

Les bruits de fond avec un jet mal identifié comme un  $\tau_h$  sont estimés à partir des données avec la méthode des fake factors (FF). Il y a plusieurs sources d'incertitudes reliées à cette méthode. Dans le canal  $\tau_h\tau_h$  ces incertitudes sont appliquées :

- Incertitudes statistiques sur la mesure des FF : ces incertitudes sont prises des bandes d'incertitudes des fonctions ajustées, séparées en deux incertitudes indépendantes en utilisant la procédure expliquée en [sous-sous-section 5.6.1.2](#). Elles sont aussi séparées en bins de  $N_{\text{jets}}$  et MVA DM et sont non corrélées entre les années ;
- Incertitude due à la correction dites non-closure de la MET : cette incertitude est définie en appliquant deux fois la correction pour la valeur haute et zéro fois pour la valeur basse. L'incertitude est non corrélée entre les classes  $N_{\text{jets}}$  dans lesquels elle est mesurée et est aussi non corrélée entre les années ;
- Incertitude due à la correction de l'extrapolation OS/SS : cette incertitude est définie en prenant la différence entre la correction en fonction de  $\Delta R$  et  $N_{\text{jets}}$  et la correction constante. Une incertitude supplémentaire de normalisation est ajoutée et est égale à la correction OS/SS constante. L'incertitude est non corrélée entre les bins en  $N_{\text{jets}}$  dans lesquels c'est mesuré et est entièrement non corrélée entre les années.

- **Incertitudes théoriques des signaux**

Les incertitudes théoriques dominantes sur l'acceptance du signal et la forme des

distributions sont considérées. Elles incluent les variations d'échelle  $\mu_R$  et  $\mu_F$  QCD et les variations d'échelles  $\mu_R$  de la gerbe de partons. Les autres sources que sont les fonctions de distribution de partons,  $\alpha_s$  et les accords d'évènements sous-jacent ont des effets négligeables en comparaison.

Pour l'incertitude d'échelle QCD, les variations hautes et basses sont données par  $\mu_R = \mu_F = 2$  et  $\mu_R = \mu_F = \frac{1}{2}$ . Ce choix permet aussi de déterminer les corrélations entre les différents bins et catégories d'évènements.

Pour l'incertitude sur la gerbe de partons, l'échelle  $\mu_R$  utilisée dans la génération de la gerbe de partons varie d'un facteur 2 et  $\frac{1}{2}$ . Cette variation est effectuée séparément pour chaque contribution ISR et FSR, ce qui mène à deux paramètres de nuisance indépendants. La variation de  $\mu_R$  est faite en utilisant des poids d'évènement calculés par PYTHIA 8 [139];

- **Incertitudes de prefiring**

En 2016 et 2017, les incertitudes sur la probabilité de prefiring sont prises comme le maximum des incertitudes statistiques sur la carte de prefire et 20%. L'incertitude est corrélée entre les années;

- **Statistique finie dans les modèles de bruits de fond**

Les limitations dans la statistique disponible des modèles sont comptées selon la méthode « Barlow-Beeston », qui assigne un paramètre de nuisance unique par bin par processus [140, 141]. Pour les modèles de bruits de fond qui ont été lissés comme décrit en section 5.8, les incertitudes bin-by-bin ont été corrélées de tel sorte qu'il n'y a qu'un seul paramètre de nuisance indépendant par N bins de  $\phi_{CP}$  pour les bruits de fond aplatis, et un paramètre pour chaque paire de bins symétrisés.

### 5.9.3 Corrélations entre les incertitudes MC et embedded

Pour les incertitudes communes aux échantillons MC et embedded, les incertitudes d'identification des  $\tau_h$  et de TES sont à 50% corrélées. Toutes les autres incertitudes sont non corrélées.

## 5.10 Résultats

Afin d'extraire l'angle de mélange CP  $\phi_{\tau\tau}$ , un ajustement simultané aux données est effectué en utilisant la fonction de vraisemblance  $L(\vec{\sigma}, \mathcal{B}^{\tau\tau}, \phi_{\tau\tau}, \vec{\theta})$ . Cette dernière dépend des sections efficaces de production du boson de Higgs  $\vec{\sigma} = (\sigma_{ggh}, \sigma_{VBF}, \sigma_{hW}, \sigma_{hZ})$ , du rapport d'embranchement des désintégrations  $h \rightarrow \tau\bar{\tau}$ , de l'angle de mélange CP  $\phi_{\tau\tau}$  et des paramètres de nuisance  $\vec{\theta}$  qui tiennent compte des incertitudes systématiques. La fonction de vraisemblance est définie comme un produit de probabilités conditionnelles

P sur des distributions d'observables discriminantes dans chaque catégorie d'évènement  $j$  :

$$L(\vec{\sigma}, \mathcal{B}^{\tau\tau}, \phi_{\tau\tau}, \vec{\theta}) = \prod_j^{N_{\text{categories}}} \prod_i^{N_{\text{bin}}} P(n_{i,j} | \mathcal{L} \cdot \vec{\sigma} \cdot \mathcal{B}^{\tau\tau} \cdot \vec{A}_{i,j}(\vec{\theta}, \phi_{\tau\tau}) + B_{i,j}(\vec{\theta})) \times \prod_m^{N_{\text{nuisance}}} C_m(\vec{\theta}), \quad (5.19)$$

avec les distributions de Poisson  $P$  correspondant à l'observation des évènements  $n_{i,j}$  dans chaque bin  $i$  de l'observable discriminante étant donné les attentes pour le bruit de fond  $B_{i,j}(\vec{\theta})$ , et pour le signal  $S_{i,j}(\vec{\theta}) = \mathcal{L} \cdot \vec{\sigma} \cdot \mathcal{B}^{\tau\tau} \cdot \vec{A}_{i,j}(\vec{\theta}, \phi_{\tau\tau})$ , où  $\mathcal{L}$  est la luminosité intégrée et  $\vec{A}_{i,j}(\vec{\theta}, \phi_{\tau\tau})$  est l'acceptance du signal dans chaque bin de production. Les contraintes sur les paramètres de nuisance, correspondant aux incertitudes systématiques décrites en [section 5.9](#), sont représentés par les fonctions  $C_m(\vec{\theta})$ . L'ajustement est réalisé en utilisant l'outil Combine [\[142\]](#).

Durant l'extraction de  $\phi_{\tau\tau}$ , deux paramètres de taux sont requis, un pour échelonner la section efficace ggH (multipliée par  $\mathcal{B}^{\tau\tau}$ ) et un autre pour la section efficace qqH d'une façon équivalente. Les deux sont laissées libres lors de l'ajustement comme paramètres de nuisance.

La fonction log-vraisemblance négative définie comme :

$$-2\Delta \ln L = -2 \cdot [\ln(L(\phi_{\tau\tau})) - \ln(L(\phi_{\tau\tau}^{\text{best fit}}))], \quad (5.20)$$

permet de trouver les intervalles de confiance (CL) à 68% et 95% quand  $-2\Delta \ln L = 0.99$  et  $-2\Delta \ln L = 3.84$  respectivement.

Les distributions  $\phi_{\text{CP}}$  sont données en entrée des ajustements des fonctions de vraisemblance réparties en bins de scores BDT (distributions dites « unrolled ») et pour chaque catégorie (Higgs, ztt et jet fakes) (voir [Appendice C](#)). Cela permet aux bruits de fond et incertitudes systématiques d'être plus fortement contraints et aide à améliorer l'ajustement.

### 5.10.1 Résultats dans le canal $a_1^{3\text{pr}} a_1^{3\text{pr}}$

À partir des résultats attendus sur un ensemble de données dites Asimov, c'est à dire des évènements MC simulant des données, on obtient une exclusion d'un boson de Higgs purement CP-impair à 0.17 déviation standard ( $\sigma$ ) pour le Run 2 avec l'ancienne méthode, et 0.28  $\sigma$  pour le Run 2 avec la nouvelle méthode du vecteur polarimétrique, ce qui correspond à une amélioration de la sensibilité dans ce canal de 60%.

Une procédure d'« unblinding » dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  a été effectuée afin de s'assurer de la stabilité de la démarche avant de dévoiler le résultat dans les données. La

première étape consista à obtenir la liste des sources d'incertitudes et de confirmer que l'ajustement n'a aucun biais statistique. La liste des 90 plus grandes systématiques dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  et leurs conséquences sur la mesure de l'angle  $\phi_{\tau\tau}$  sont en Fig. 5.31 – 5.33. Les plus importantes systématiques sont celles dites bin-by-bin à cause de la faible statistique du canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$ . Des tests de la qualité de l'ajustement (GOF) ont été effectués pour la catégorie de signal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  et les deux catégories de bruits de fond (contenant tous les modes de désintégrations) pour chaque année. Ces tests - dits saturés - car on fait correspondre le modèle aux données bin par bin, ce qui permet de définir une hypothèse nulle, consistent à estimer l'accord entre des valeurs observées et attendues pour un modèle statistique donné [143]. Les tests sont donnés en Fig. 5.34 – 5.37. Chaque entrée des histogrammes correspond à un test  $\chi^2$ . Ces tests GOF sont satisfaisants car les p-values, correspondant à la proportion de l'aire sous la courbe à droite des flèches bleues, sont autour de 0.5.

La seconde étape était d'obtenir les distributions « unrolled » sans cacher les données. Ces distributions avant ajustement sont montrées en Appendice C en Fig. C.1 – C.9 dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  pour les trois années et les trois catégories, correspondant aux trois sorties du BDT de classification. Les distributions « unrolled » après ajustement pour les trois années dans le même canal sont montrées en Fig. 5.38 – 5.40 pour la catégorie de signal, seule catégorie utilisée lors de l'ajustement final. On remarque que l'oscillation du signal s'inverse après ajustement, c'est-à-dire que le signal ressemble à un boson Higgs CP-impair.

La troisième et dernière étape de la procédure d'« unblinding » fut l'obtention des fonctions de vraisemblance avant et après ajustement final de l'angle de mélange  $\phi_{\tau\tau}$  qui donnent dans le canal étudié la sensibilité attendue et observée respectivement. Ces sensibilités sont montrées en Fig. 5.41. La valeur obtenue après ajustement,  $\phi_{\tau\tau} = 80.5^{\circ}_{-35.6^{\circ}}^{+9.5^{\circ}}$ , correspond à ce que l'on peut observer dans les distributions « unrolled », c'est-à-dire à un signal à caractère CP-impair. On peut voir à partir de la fonction de vraisemblance, après ajustement, du paramètre de taux commun  $\mu_{\text{tautau}}$ , paramètre échelonnant de la même manière tous les modes de production du signal, en Fig. 5.42, que sa valeur est élevée ( $2.24^{+0.94}_{-0.88}$ ) mais tout de même compatible avec les dernières analyses étudiant ces couplages [144].

Cette grande valeur de  $\mu_{\text{tautau}}$  et la fluctuation statistique du canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  expliquent cette observation. Un tel comportement a également été observé avec la méthode du plan de désintégration dans le même canal. De plus, le paramètre de taux commun devrait être mieux contraint lorsque l'ajustement final de ce canal avec la méthode du vecteur polarimétrique se fera avec tous les autres canaux.

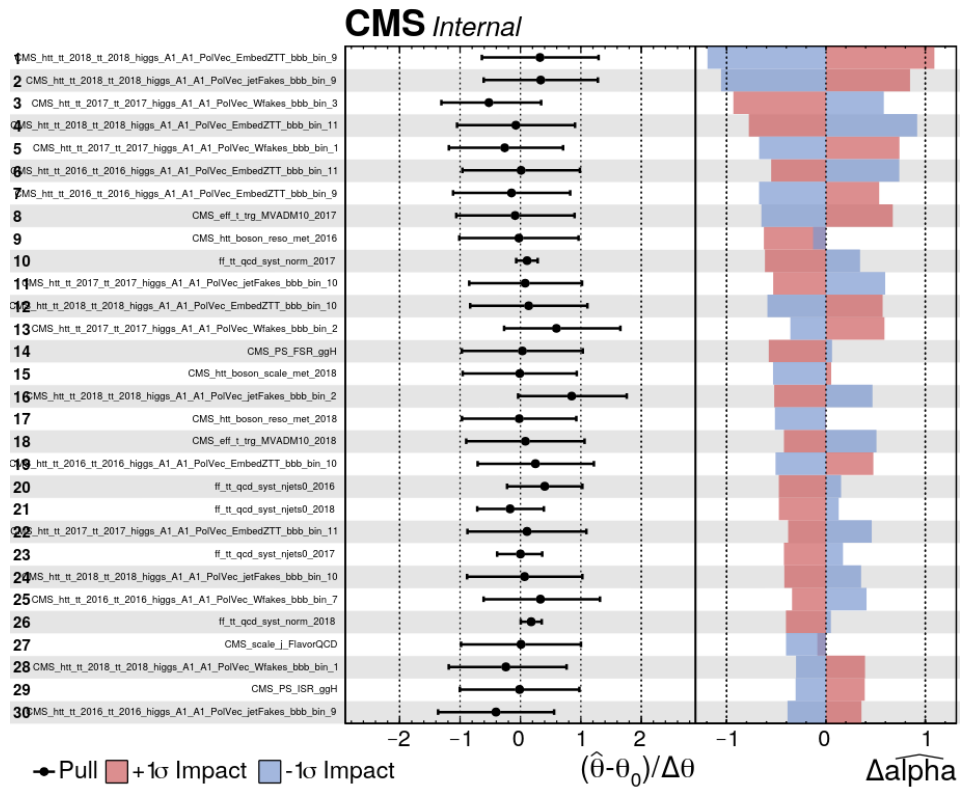


FIGURE 5.31 – Liste des 1 à 30 plus importantes systématiques.

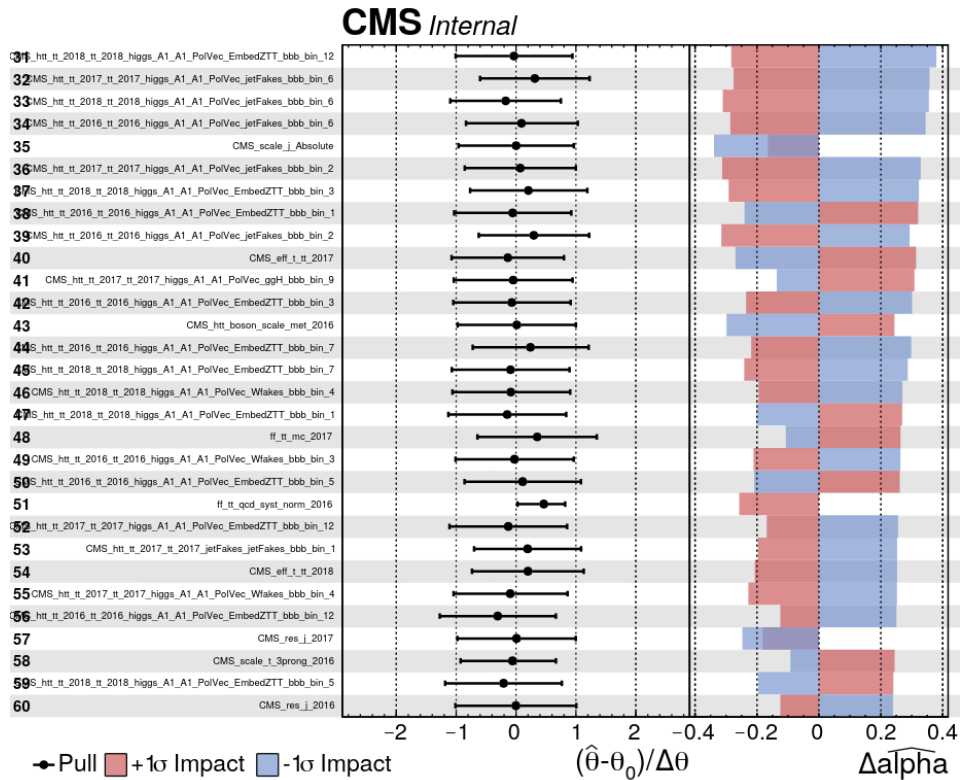


FIGURE 5.32 – Liste des 31 à 60 plus importantes systématiques.

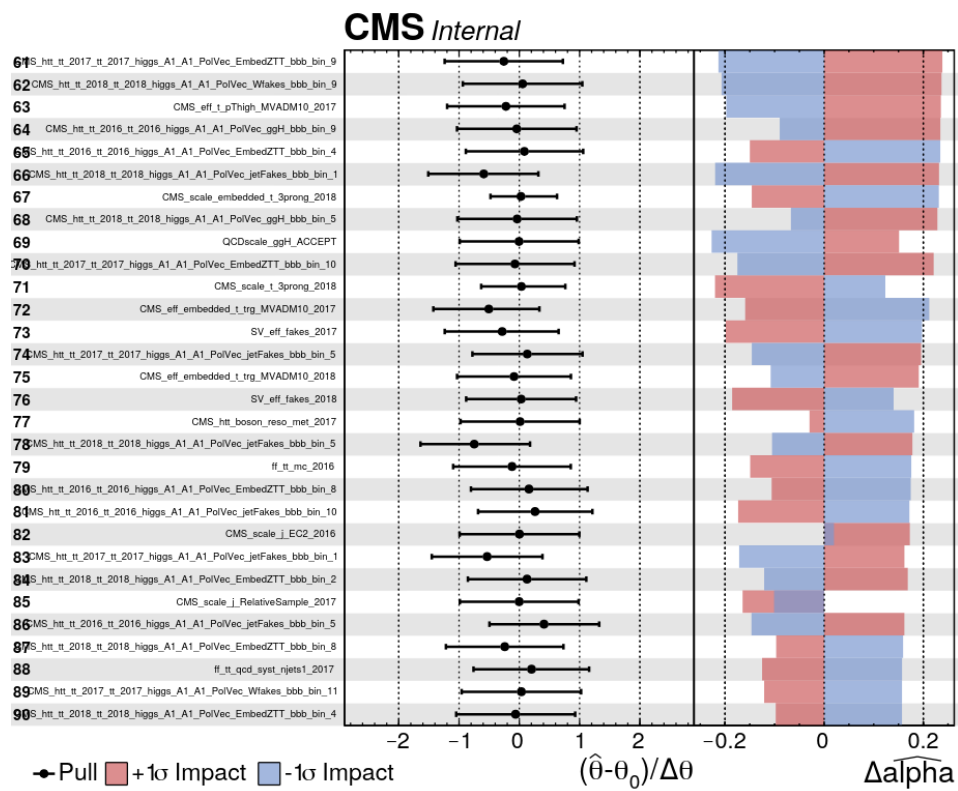


FIGURE 5.33 – Liste des 61 à 90 plus importantes systématiques.

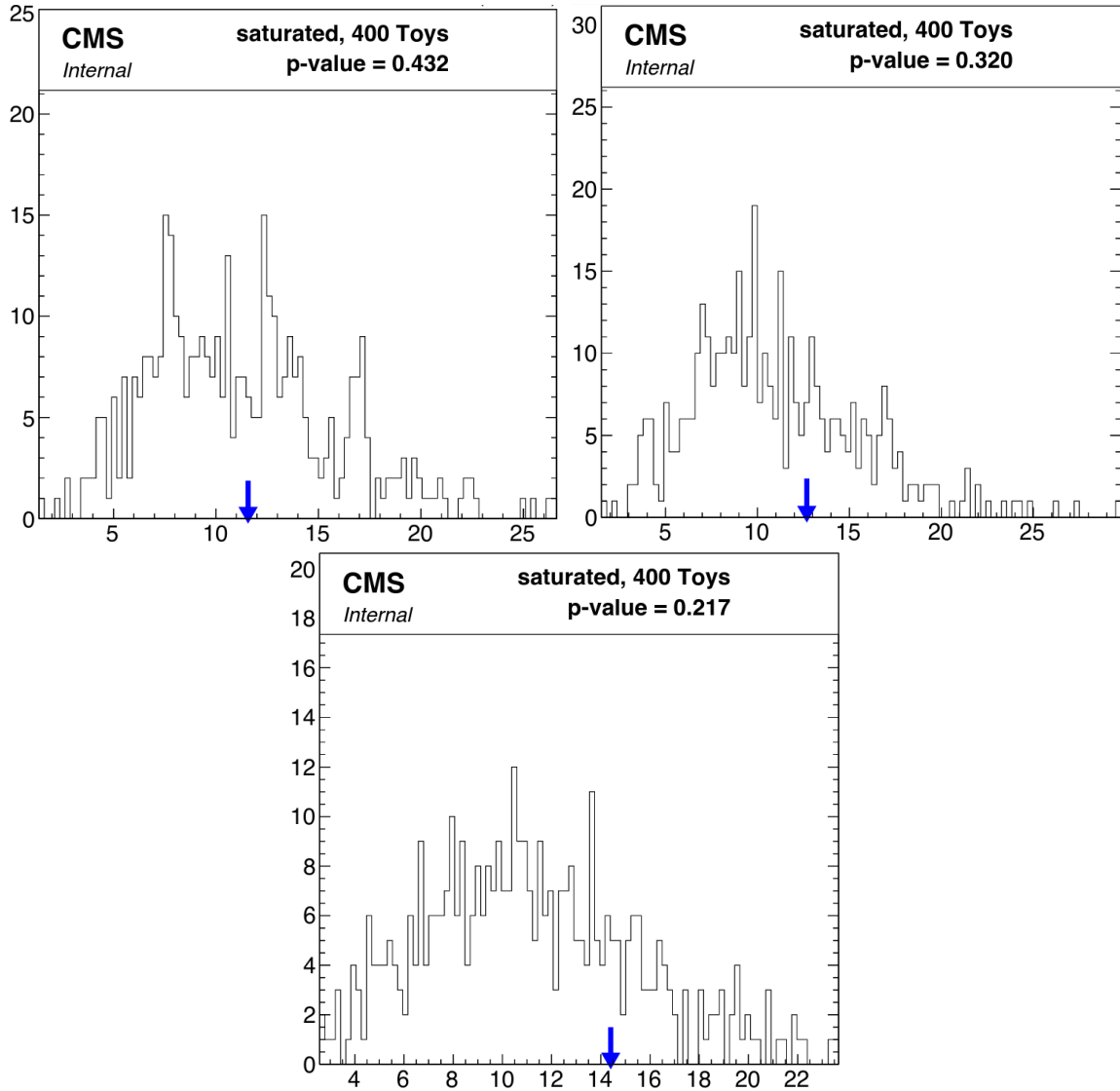


FIGURE 5.34 – Test de la qualité de l’ajustement avec un modèle saturé pour la catégorie de **signal**  $a_1^{3pr} a_1^{3pr}$  pour l’année 2016 (gauche), 2017 (droite) et 2018 (milieu). Chaque entrée de l’histogramme représente un test  $\chi^2$ .

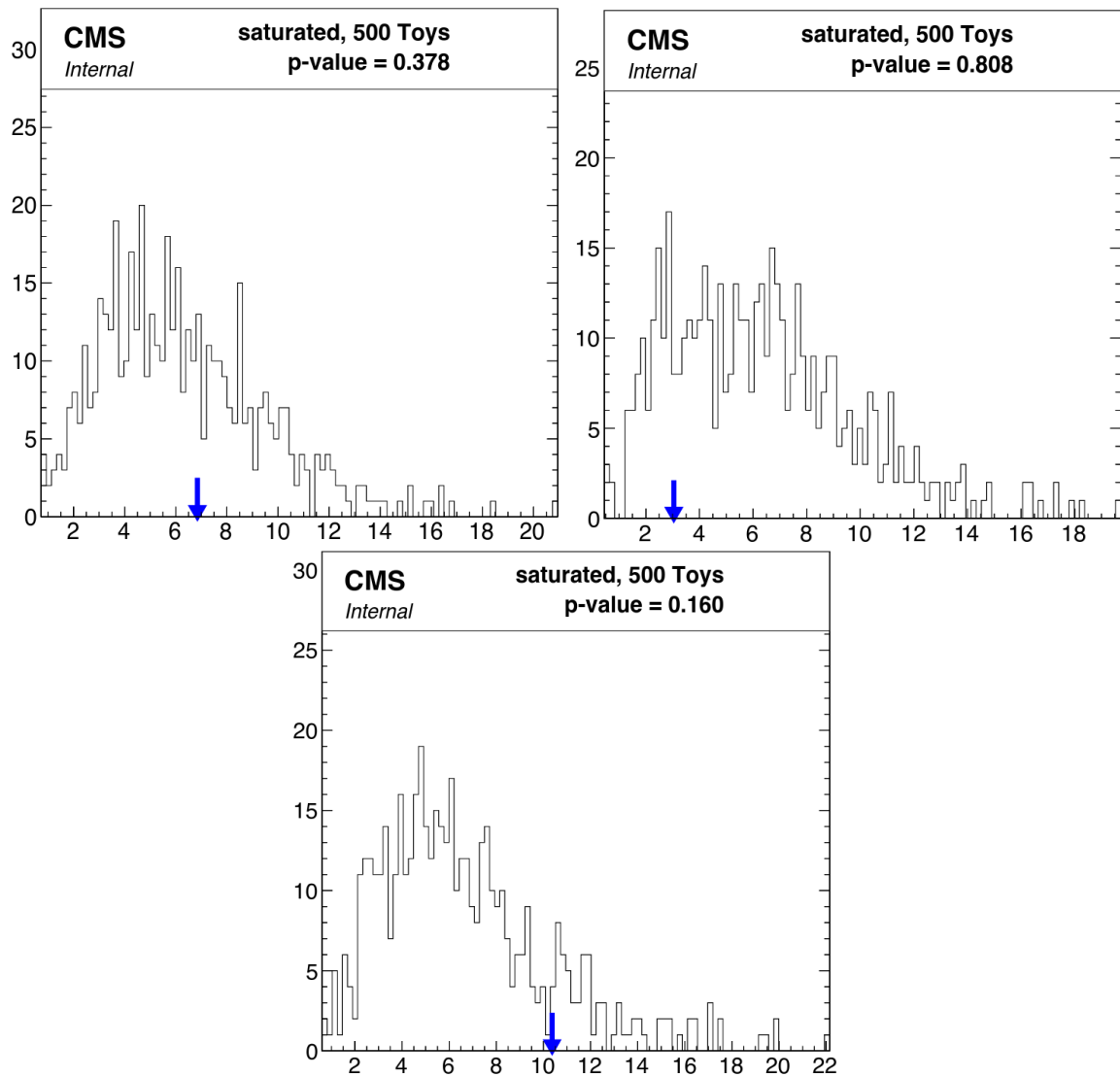


FIGURE 5.35 – Test de la qualité de l’ajustement avec un modèle saturé pour la catégorie **jets fakes** pour l’année 2016 (gauche), 2017 (droite) et 2018 (milieu). Chaque entrée de l’histogramme représente un test  $\chi^2$ .



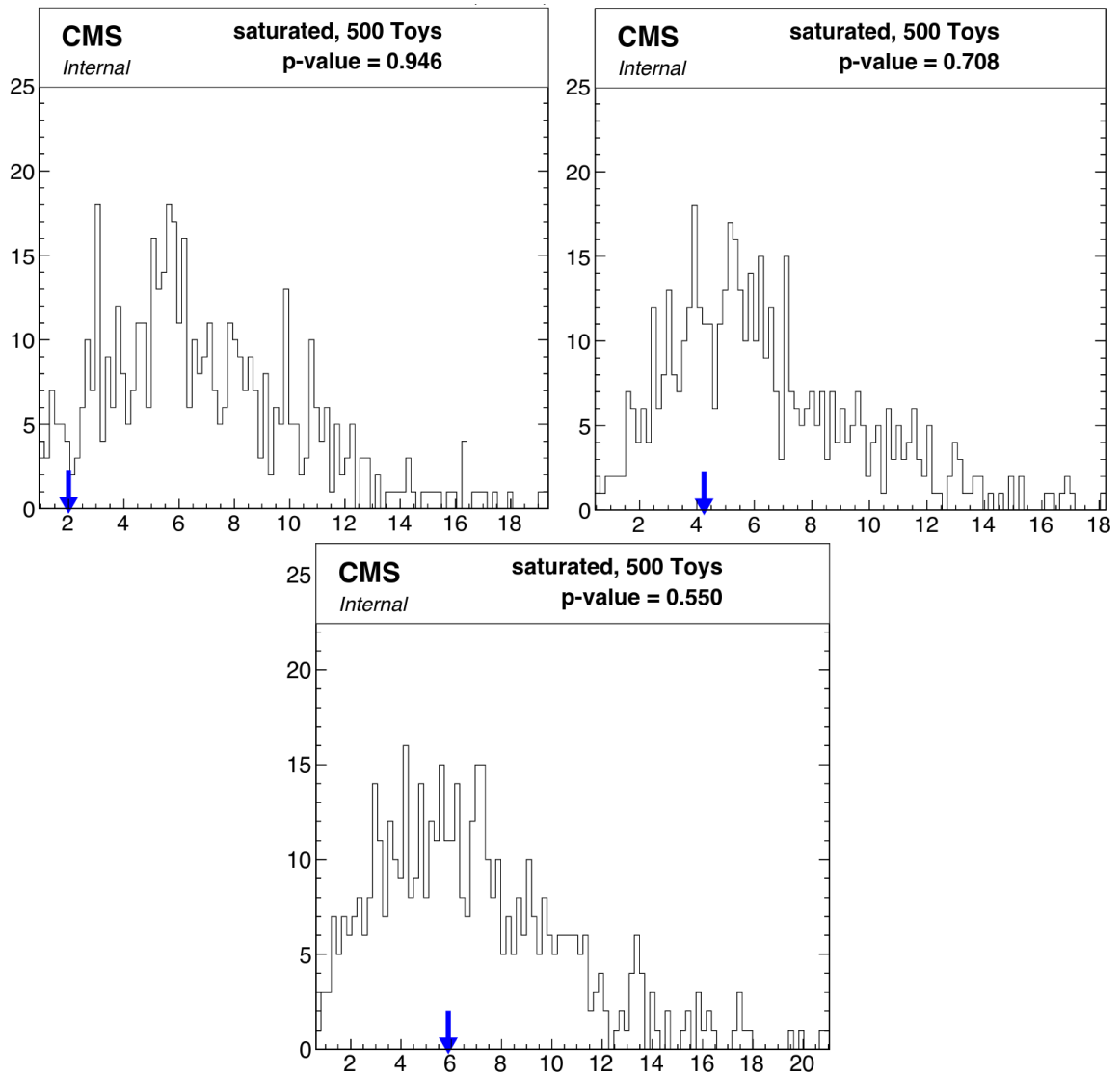


FIGURE 5.36 – Test de la qualité de l’ajustement avec un modèle saturé pour la catégorie **ztt embedded** pour l’année 2016 (gauche), 2017 (droite) et 2018 (milieu). Chaque entrée de l’histogramme représente un test  $\chi^2$ .

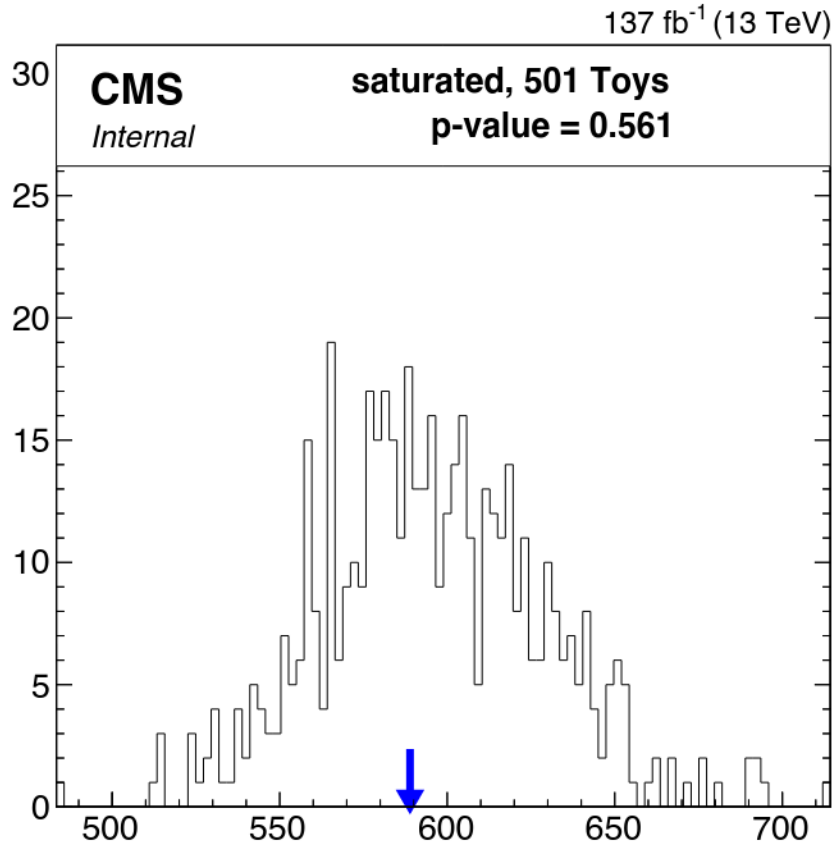


FIGURE 5.37 – Test de la qualité de l’ajustement avec un modèle saturé pour les catégories de **signal**  $a_1^{3pr} a_1^{3pr}$ , **jets fakes** et **ztt embedded** combinées pour le Run 2. Chaque entrée de l’histogramme représente un test  $\chi^2$ .

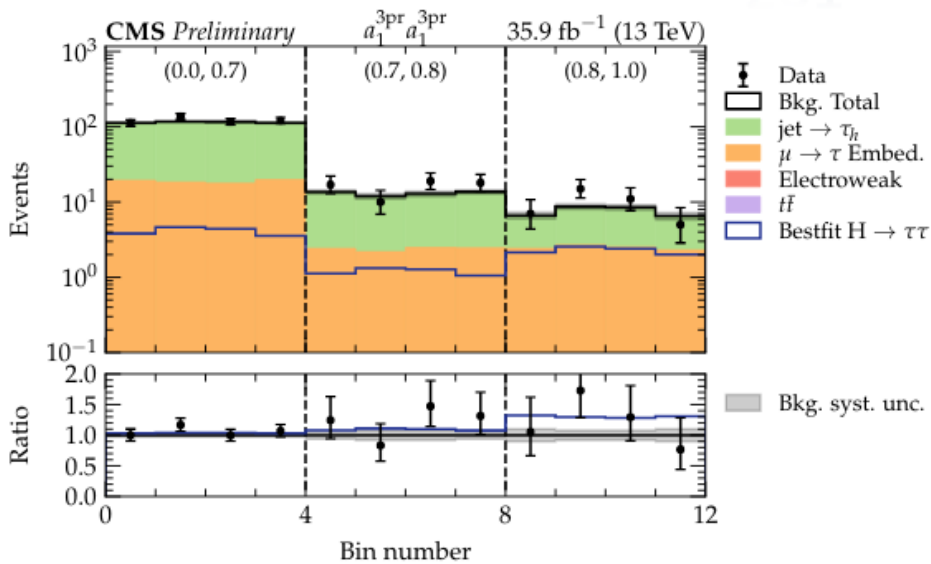


FIGURE 5.38 – Distribution 2016 après ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L’axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

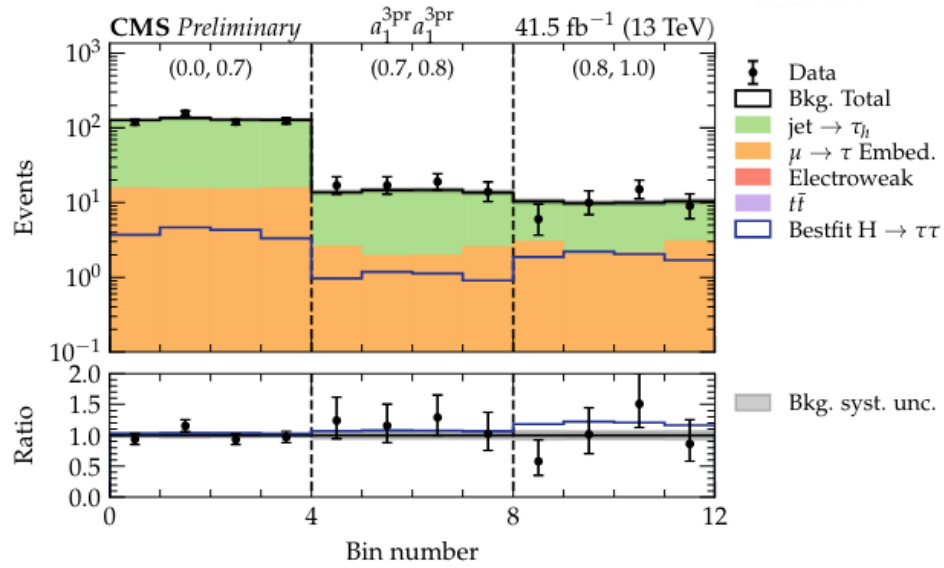


FIGURE 5.39 – Distribution 2017 après ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

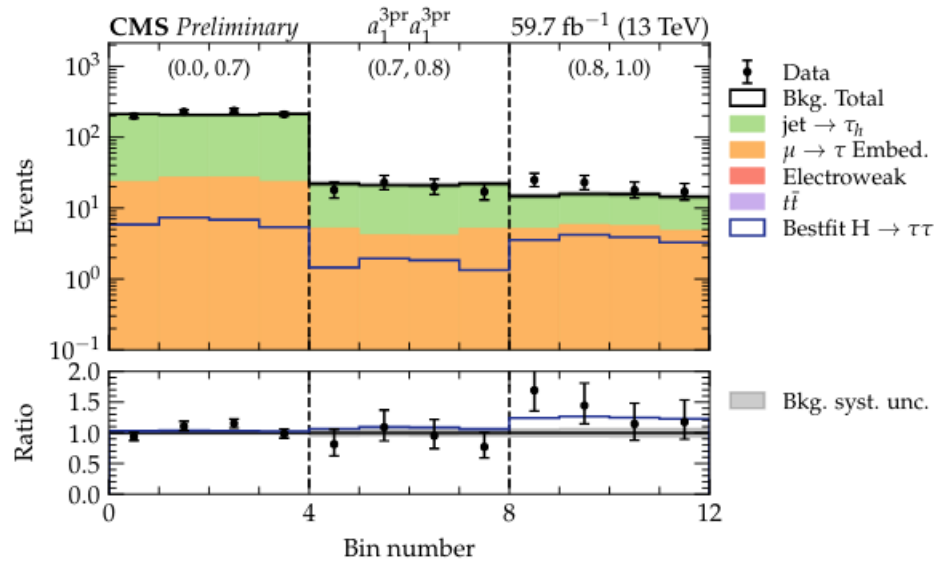


FIGURE 5.40 – Distribution 2018 après ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

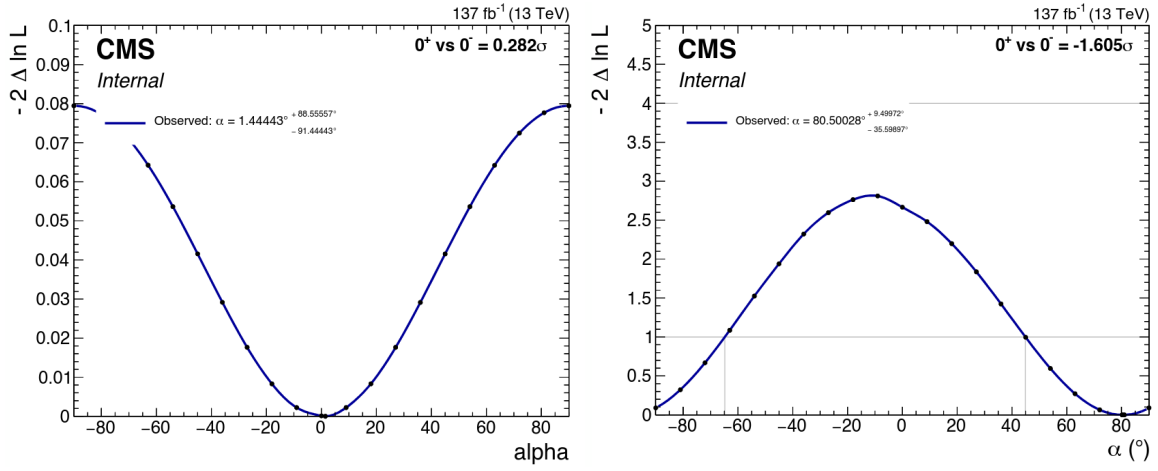


FIGURE 5.41 – Fonctions log-vraisemblance négatives de l'angle  $\phi_{\tau\tau}$  attendue (gauche) et observés (droite) pour les trois années combinées dans le canal  $a_1^{3pr} a_1^{3pr}$  avec la méthode du vecteur polarimétrique.

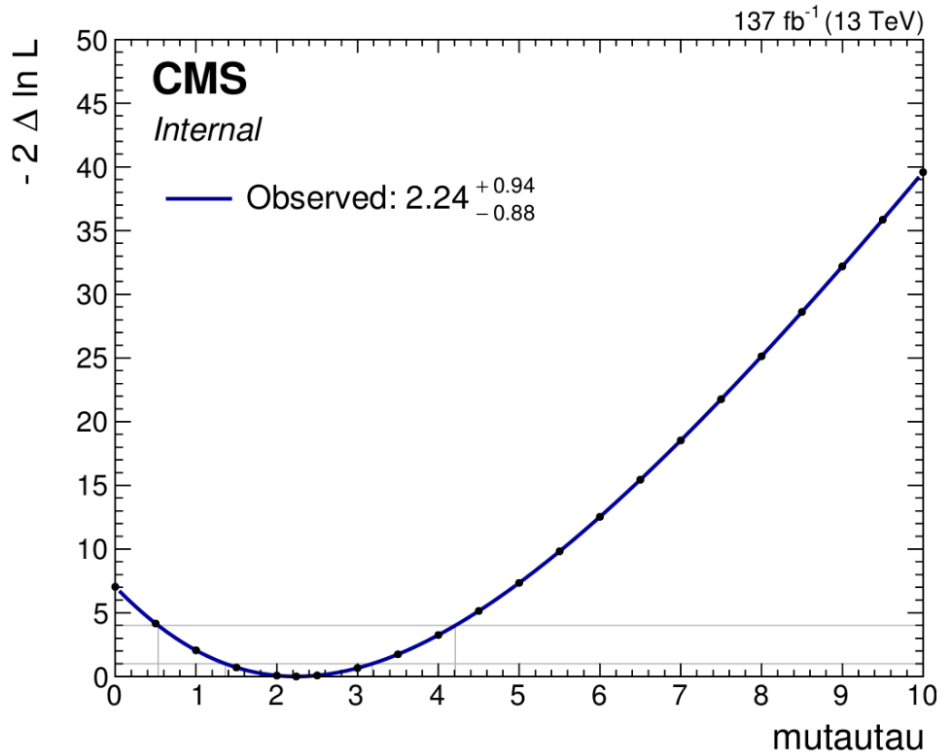


FIGURE 5.42 – Fonctions log-vraisemblance négatives du paramètre de taux commun  $\mu_{\text{tautau}}$  observés pour les trois années combinées dans le canal  $a_1^{3pr} a_1^{3pr}$  avec la méthode du vecteur polarimétrique.

## 5.11 Conclusion

L'analyse CMS combinant les canaux  $\tau_\mu\tau_h$  et  $\tau_h\tau_h$  a fait l'objet d'un « Physics Analysis Summaries » (PAS) [117], et le papier est en cours de rédaction. L'ajustement permettant de séparer les états CP-pair et impair du boson de Higgs donne une sensibilité observée (attendue) de  $3.2$  ( $2.3$ )  $\sigma$  pour les canaux  $\tau_\mu\tau_h$  et  $\tau_h\tau_h$  combinés avec toute la statistique disponible du Run 2. La valeur de l'angle  $\phi_{\tau\tau}$  observée (attendue) est de  $4 \pm 17^\circ$  ( $0 \pm 23^\circ$ ) à 68% CL et  $\pm 36^\circ$  ( $\pm 55^\circ$ ) à 95% CL.

L'incertitude peut être décomposée en différentes contributions : statistique, fluctuations bin-by-bin dans les modèles de bruit de fond, systématiques expérimentales et incertitudes théoriques. La valeur décomposée attendue de l'angle  $\phi_{\tau\tau}$  pour le Run 2 est donc de  $(4 \pm 17(\text{stat}) \pm 2(\text{bin} - \text{by} - \text{bin}) \pm 1(\text{syst}) \pm 1(\text{théorie}))^\circ$ . Ce résultat est compatible avec les prédictions du MS. La sensibilité attendue du canal  $\tau_h\tau_h$  seul est de  $1.8\sigma$  tandis que celle du canal  $\tau_\mu\tau_h$  est de  $1.5\sigma$ . Les incertitudes statistiques dans les modèles de bruit de fond sont la source dominante d'incertitude. Les bruits de fond dominants étant estimés à partir des données, la quantité de ces dernières est le facteur limitant. La différence entre les sensibilités attendues et observées s'explique par le fait que les données sont plus séparées de la prédiction CP-impair que celles-ci ne le sont de la prédiction CP-pair à cause des fluctuations statistiques. La figure 5.43 illustre ceci. On y voit le signal, comme les données après soustraction du bruit de fond, affiché avec les distributions d'un boson de Higgs scalaire (bleu) et pseudoscalaire (vert). Les points sont pondérés par  $A*S/(S+B)$  avec S le signal, B le bruit de fond et

$$A = \frac{1}{N_{\text{bins}}} \sum^{N_{\text{bins}}} \frac{|CP^{\text{even}} - CP^{\text{odd}}|}{CP^{\text{even}} + CP^{\text{odd}}} \text{ l'asymétrie moyenne.}$$

Durant ma thèse, j'ai mis à jour le code d'analyse et notamment inséré les différentes sélections, corrections, l'estimation du bruit de fond, les incertitudes systématiques ou encore la possibilité d'analyser les trois années de données. J'ai mis en œuvre la méthode du vecteur polarimétrique et ai étudié le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$ . L'utilisation de cette méthode en remplacement de la méthode du plan de désintégration utilisée dans le PAS permet d'augmenter la sensibilité dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  d'environ 60%. Par conséquent, cette méthode remplacera l'ancienne dans ce canal et sera intégrée à l'analyse CMS dont le papier est en cours de rédaction.

La méthode du vecteur polarimétrique est donc très prometteuse et est applicable à toutes désintégrations d'un boson de Higgs en deux leptons taus où ces derniers auront leur quadri-vecteur énergie-impulsion reconstruits avec une bonne précision angulaire. La reconstruction de l'impulsion des taus est difficile à cause de la présence d'énergie transverse manquante. En revanche, les désintégrations en une résonance  $a_1^{3\text{pr}}$ , donnant trois pions chargés, permet de reconstruire le vertex secondaire qui sera utilisé pour estimer la direction des quadri-vecteurs énergie-impulsion. En implémentant la méthode

GEF complète, elle devient alors utilisable au LHC dans les désintégrations ayant au moins un lepton tau donnant une résonance  $a_1^{3pr}$  ( $\mu a_1^{3pr}$ ,  $\pi a_1^{3pr}$ ,  $\rho a_1^{3pr}$ ...). La méthode du vecteur polarimétrique devrait permettre d'augmenter davantage la sensibilité en l'étendant à d'autres canaux ainsi qu'aux données prises lors du Run 3 et du HL-LHC.

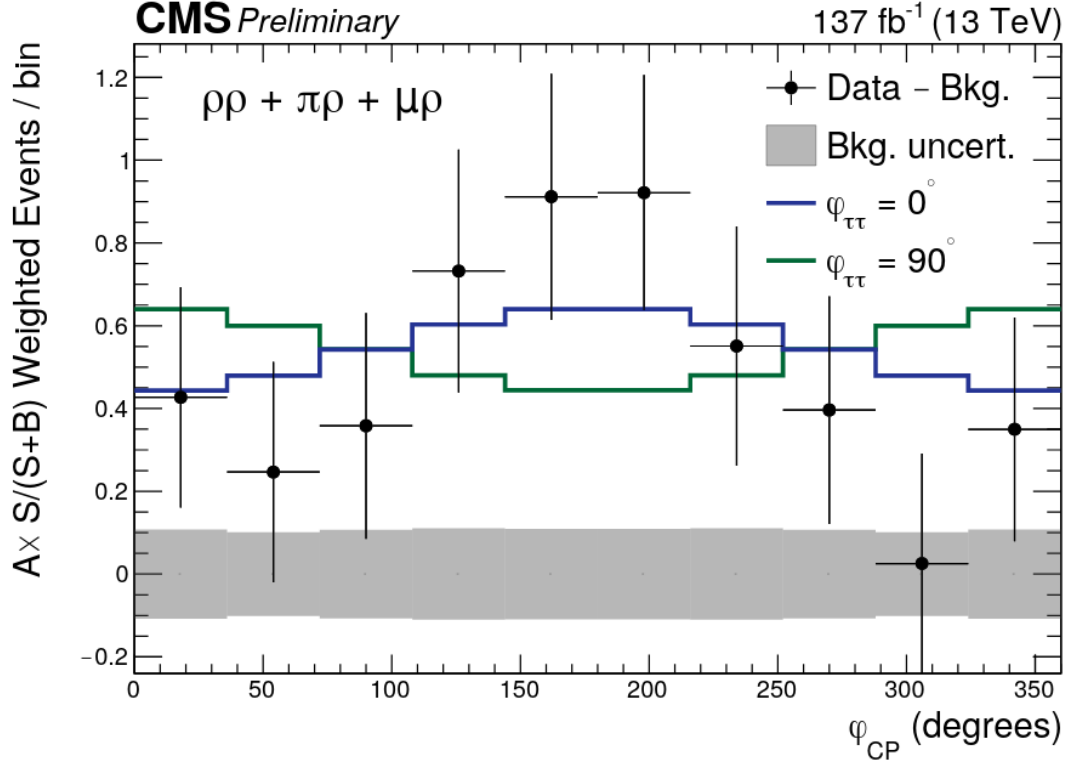


FIGURE 5.43 – Distributions de  $\phi_{CP}$  pour toutes les années et catégories pour les trois canaux les plus sensibles :  $\rho\rho + \pi\rho + \mu\rho$ . Les événements sont pondérés par  $A \cdot S/(S+B)$  avec S le signal, B le bruit de fond soustrait des données et  $A = \frac{1}{N_{bins}} \sum \frac{|CP^{even} - CP^{odd}|}{CP^{even} + CP^{odd}}$  l'asymétrie moyenne entre les distributions CP-paire (bleu) et CP-impair (vert) normalisées par le nombre de bins de l'histogramme. La bande d'incertitude grise indique l'incertitude sur le bruit de fond soustrait [117].

# Conclusion

Cette thèse présente trois études que j'ai effectuées. La première est l'amélioration pour la Phase 2 de l'arbre de décision boostée (BDT) rejetant les électrons se faisant passer pour des taus hadroniques ([sous-sous-section 3.2.5.1](#)). La seconde est l'étude des particules hautement ionisantes (HIP) qui induisent un temps mort de l'électronique dans le trajectographe ([section 4.2](#)). Enfin, la troisième est l'étude principale de ma thèse et consiste en la mesure de l'état CP du boson de Higgs dans ses désintégrations  $H \rightarrow \tau\tau \rightarrow a_1^{3\text{pr}} \bar{\nu}_\tau a_1^{3\text{pr}} \nu_\tau \rightarrow 3\pi \bar{\nu}_\tau 3\pi \nu_\tau$  grâce à la méthode dite du vecteur polarimétrique ([chapitre 5](#)).

Le BDT actuel permettant d'identifier correctement les leptons taus en les différenciant des électrons ne sera plus aussi performant pendant la phase à haute luminosité du LHC. En effet, la luminosité pourra être multipliée par presque 4 et l'empilement devrait atteindre 200. Dans un tel contexte, une nouvelle optimisation est nécessaire. J'ai donc ajouté de nouvelles variables tirant parti du nouveau sous-détecteur qu'est le HG-CAL. De plus, j'ai catégorisé l'espace en 3 régions suivant la valeur en pseudorapidité, ces dernières correspondant au tonneau, aux bouchons à l'avant et aux bouchons très à l'avant jusqu'à  $|\eta| = 3.0$ . Les performances obtenues sont comparables à celles du Run 2 et sont mêmes un peu meilleures dans la région avant. Ces performances pourraient être améliorées en tirant parti du nouveau sous-détecteur MTD qui permet d'avoir une information en temps, ainsi qu'en utilisant un apprentissage par réseaux de neurones profonds.

Les HIP réduisent l'efficacité de reconstruction des coups dans le trajectographe. En effet, la grande charge électrique déposée induit un temps mort dans l'électronique. Mon étude consistait en la vérification de l'hypothèse selon laquelle les HIP seraient responsables de l'inefficacité observée dans le trajectographe. J'ai démontré que les HIP étaient bien la principale source des inefficacités constatées, mais un modèle plus complexe, comme la prise en compte du schéma de remplissage de l'anneau du LHC, est nécessaire afin de mieux quantifier ces résultats.

Enfin, ma thèse a principalement consisté en la mesure de l'état CP du boson de Higgs dans ses désintégrations en paires de leptons taus dans le canal  $a_1^{3\text{pr}} a_1^{3\text{pr}}$  grâce à l'utilisation de la méthode dite du vecteur polarimétrique. Une publication de CMS sur la mesure de l'état CP du boson de Higgs donnant deux taus est en cours d'écriture et la

## Conclusion

valeur observée de l'angle  $\phi_{\tau\tau}$  est de  $4 \pm 17^\circ$  à 68% CL pour une sensibilité de  $3.2\sigma$ . Dans le canal  $a_1^{3pr} a_1^{3pr}$ , la méthode initialement appliquée est celle du plan de désintégration. Or, la méthode du vecteur polarimétrique permet d'améliorer la sensibilité de ce canal de 60%. Mes travaux vont donc remplacer la méthode du plan de désintégration dans ce canal afin d'être intégrés dans la publication de la collaboration CMS. Cette nouvelle méthode est prometteuse car elle est applicable à n'importe quel canal où les impulsions des deux taus sont très bien reconstruits. Ceci est difficile au LHC mais la méthode du vecteur polarimétrique est utilisable dans les canaux contenant au moins un  $a_1^{3pr}$ , car la reconstruction du vertex secondaire est possible, ce qui permet une reconstruction précise de la direction du tau. De plus, les principales incertitudes de l'analyse CMS étant statistiques, la collecte prochaine de données du Run 3 et pendant la Phase 2 permettra d'obtenir un résultat bien plus significatif.



## Annexe A

# Tableaux des échantillons utilisés dans le canal $\tau_h\tau_h$

Channel	Dataset	Run range
$\tau_h\tau_h$	/Tau/Run2016B-17Jul2018-ver2-v1/MINIAOD	272007–275376
$\tau_h\tau_h$	/Tau/Run2016C-17Jul2018-v1/MINIAOD	275657–276283
$\tau_h\tau_h$	/Tau/Run2016D-17Jul2018-v1/MINIAOD	276315–276811
$\tau_h\tau_h$	/Tau/Run2016E-17Jul2018-v1/MINIAOD	276831–277420
$\tau_h\tau_h$	/Tau/Run2016F-17Jul2018-v1/MINIAOD	277772–278808
$\tau_h\tau_h$	/Tau/Run2016G-17Jul2018-v1/MINIAOD	278820–280385
$\tau_h\tau_h$	/Tau/Run2016H-17Jul2018-v1/MINIAOD	280919–284044

TABLE A.1 – Liste des données 2016 (legacy 94X) utilisées dans l’analyse dans le canal  $\tau_h\tau_h$ .

Channel	Dataset	Run range
$\tau_h\tau_h$	/Tau/Run2017B-31Mar2018-v1/MINIAOD	297047–299329
$\tau_h\tau_h$	/Tau/Run2017C-31Mar2018-v1/MINIAOD	299368–302029
$\tau_h\tau_h$	/Tau/Run2017D-31Mar2018-v1/MINIAOD	302031–302663
$\tau_h\tau_h$	/Tau/Run2017E-31Mar2018-v1/MINIAOD	303824–304797
$\tau_h\tau_h$	/Tau/Run2017F-31Mar2018-v1/MINIAOD	305040–306460

TABLE A.2 – Liste des données 2017 utilisées dans l’analyse dans le canal  $\tau_h\tau_h$ .

Channel	Dataset	Run range
$\tau_h\tau_h$	/Tau/Run2018A-17Sep2018-v1/MINIAOD	315252–316995
$\tau_h\tau_h$	/Tau/Run2018B-17Sep2018-v1/MINIAOD	317080–319310
$\tau_h\tau_h$	/Tau/Run2018C-17Sep2018-v1/MINIAOD	319337–320065
$\tau_h\tau_h$	/Tau/Run2018D-PromptReco-v2/MINIAOD	320673–325175

TABLE A.3 – Liste des données 2018 utilisées dans l’analyse dans le canal  $\tau_h\tau_h$ .

Annexe A. Tableaux des échantillons utilisés dans le canal  $\tau_h\tau_h$

Signal Monte Carlo samples				
Process	Dataset Name	Cross-Section [pb]	Events	Filter Eff.
ggH $h \rightarrow \tau_h\tau_h$	/GluGluHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	3.046	20156550	0.2455
VBF $h \rightarrow \tau_h\tau_h$	/VBFHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.237	18671016	0.2727
$W^+h \rightarrow \tau_h\tau_h$	/WplusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0527	2735246	0.2425
$W^-h \rightarrow \tau_h\tau_h$	/WminusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0358	1884919	0.2596
Zh $\rightarrow \tau_h\tau_h$	/ZHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0594	3557541	0.2546

<sup>1</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3-v1/MINIAODSIM

TABLE A.4 – Liste des échantillons MC 2016 utilisés pour modéliser le signal  $pp \rightarrow h \rightarrow \tau_h\tau_h$ .

Background Monte Carlo samples		
Process	Dataset Name	Cross-Section [pb]
$t\bar{t} \rightarrow LL\nu\nu$	/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8 <sup>1</sup>	88.29
$t\bar{t} \rightarrow$ hadronic	/TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8 <sup>1</sup>	377.96
$t\bar{t} \rightarrow$ semi-leptonic	/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8 <sup>1</sup>	365.34
Single top	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 <sup>2</sup>	35.6
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 <sup>2</sup>	35.6
	/ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1 <sup>1</sup>	136.02
	/ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1 <sup>1</sup>	80.95
Z $\rightarrow$ LL + jets	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>3,5</sup>	6077.22 (NNLO)
Z $\rightarrow$ LL + 1jet	/DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>1</sup>	1012.5 (LO)
Z $\rightarrow$ LL + 2jets	/DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2</sup>	332.8 (LO)
Z $\rightarrow$ LL + 3jets	/DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2</sup>	101.8 (LO)
Z $\rightarrow$ LL + 4jets	/DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2</sup>	54.8 (LO)
Low mass Z $\rightarrow$ LL + jets	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2</sup>	18610
W $\rightarrow$ $L\nu$ + jets	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2,5</sup>	61526.7 (NNLO)
W $\rightarrow$ $L\nu$ + 1jet	/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>1</sup>	9644.5 (LO)
W $\rightarrow$ $L\nu$ + 2jets	/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2,3</sup>	3144.5 (LO)
W $\rightarrow$ $L\nu$ + 3jets	/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2,3</sup>	954.8 (LO)
W $\rightarrow$ $L\nu$ + 4jets	/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 <sup>2,3,4</sup>	485.6 (LO)
EWK $W^+ \rightarrow L^+\nu$ + 2jets	/EWKWPlus2Jets_WToLNu_M-50_13TeV-madgraphMLM-pythia8 <sup>2,3,5</sup>	25.62 (LO)
EWK $W^- \rightarrow L^-\nu$ + 2jets	/EWKMinus2Jets_WToLNu_M-50_13TeV-madgraphMLM-pythia8 <sup>2,3,5</sup>	20.25 (LO)
EWK Z $\rightarrow$ LL + 2jets	/EWKZ2Jets_ZToLL_M-50_13TeV-madgraphMLM-pythia8 <sup>2,3,5</sup>	3.987 (LO)
VV	/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin-pythia8 <sup>2,3</sup>	11.95
ZZ	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>1</sup>	3.22
ZZ	/ZZTo4L_13TeV_amcatnloFXFX-pythia8 <sup>3</sup>	1.212
WW	/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>2</sup>	49.997
WZ	/WZJToLLNu_TuneCUETP8M1_13TeV-amcnlo-pythia8 <sup>1</sup>	4.708
WZ	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>2</sup>	5.595
WZ	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin-pythia8 <sup>2</sup>	3.05
WZ	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>2</sup>	10.71

<sup>1</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3-v1

<sup>2</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3-v2

<sup>3</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3\_ext1-v2

<sup>4</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3\_ext2-v1

<sup>5</sup> /RunIISummer16MiniAODv3-PUMoriond17\_94X\_mcRun2\_asymptotic\_v3\_ext2-v2

TABLE A.5 – Liste des échantillons MC 2016 utilisés pour modéliser les bruits de fond  $Z/\gamma^* \rightarrow LL$ ,  $W \rightarrow L\nu$ ,  $t\bar{t}$ , single top and di-boson (WW, WZ, ZZ) avec  $L = e, \mu, \tau$ .

Signal Monte Carlo samples				
Process	Dataset Name	Cross-Section [pb]	Events	Filter Eff.
ggH $h \rightarrow \tau_h\tau_h$	/GluGluHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	3.046	19420205	0.2447
VBF $h \rightarrow \tau_h\tau_h$	/VBFHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.237	20397834	0.2697
$W^+h \rightarrow \tau_h\tau_h$	/WplusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0527	2640112	0.2394
$W^-h \rightarrow \tau_h\tau_h$	/WminusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0358	1986464	0.2567
Zh $\rightarrow \tau_h\tau_h$	/ZHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0594	2477724	0.2514

<sup>1</sup> /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v1/MINIAODSIM

TABLE A.6 – Liste des échantillons MC 2017 utilisés pour modéliser le signal  $pp \rightarrow h \rightarrow \tau_h\tau_h$ .

Annexe A. Tableaux des échantillons utilisés dans le canal  $\tau_h\tau_h$

Background Monte Carlo samples		
Process	Dataset Name	Cross-Section [pb]
$t\bar{t} \rightarrow LL\nu\nu$	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 <sup>8</sup>	88.29
$t\bar{t} \rightarrow$ hadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8 <sup>11</sup>	377.96
$t\bar{t} \rightarrow$ semi-leptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8 <sup>8</sup>	365.34
Single top	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 <sup>2</sup>	34.97
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 <sup>2</sup>	35.85
	/ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8 <sup>8</sup>	136.02
	/ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8 <sup>2</sup>	80.95
Z $\rightarrow$ LL + jets	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>5,6</sup>	6077.22 (NNLO)
Z $\rightarrow$ LL + 1jet	/DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>8,7</sup>	877.8 (LO)
Z $\rightarrow$ LL + 2jets	/DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1,9</sup>	304.4 (LO)
Z $\rightarrow$ LL + 3jets	/DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1,9</sup>	111.5 (LO)
Z $\rightarrow$ LL + 4jets	/DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>10</sup>	44.03 (LO)
Low mass Z $\rightarrow$ LL + jets	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1,1a</sup>	18610
W $\rightarrow$ L $\nu$ + jets	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2,2a</sup>	61526.7 (NNLO)
W $\rightarrow$ L $\nu$ + 1jet	/W1JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>12</sup>	9644.5 (LO)
W $\rightarrow$ L $\nu$ + 2jets	/W2JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>13</sup>	3144.5 (LO)
W $\rightarrow$ L $\nu$ + 3jets	/W3JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	954.8 (LO)
W $\rightarrow$ L $\nu$ + 4jets	/W4JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>11</sup>	485.6 (LO)
EWK W <sup>+</sup> $\rightarrow$ L <sup>+</sup> $\nu$ + 2jets	/EWKwPlus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>11</sup>	25.62 (LO)
EWK W <sup>-</sup> $\rightarrow$ L <sup>-</sup> $\nu$ + 2jets	/EWKwMinus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>11</sup>	20.25 (LO)
EWK Z $\rightarrow$ LL + 2jets	/EWKZ2Jets_ZToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>11</sup>	3.987 (LO)
ZZ	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 <sup>1</sup>	3.688
ZZ	/ZZTo2L2Nu_13TeV-powheg-pythia8 <sup>1</sup>	0.6008
ZZ	/ZZTo4L_13TeV_amcatnloFXFX-pythia8 <sup>8,14</sup>	1.325
WW	/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	11.08
WW	/WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8 <sup>8</sup>	45.99
WZ	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8_v2 <sup>1</sup>	6.331
WZ	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8_v2 <sup>1</sup>	3.293
WZ	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 <sup>2</sup>	11.66
WZ	/WZTo3LNu_13TeV_amcatnloFXFX_madspin_pythia8_v2 <sup>8</sup>	5.052

- 1 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v1/MINIAODSIM
- 1<sup>a</sup> /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14\_ext1-v1/MINIAODSIM
- 2 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v2/MINIAODSIM
- 2<sup>a</sup> /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14\_ext1-v2/MINIAODSIM
- 3 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v3/MINIAODSIM
- 4 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v4/MINIAODSIM
- 5 /RunIIFall17MiniAODv2-PU2017RECOSIMstep\_12Apr2018\_94X\_mc2017\_realistic\_v14-v1/MINIAODSIM
- 6 /RunIIFall17MiniAODv2-PU2017RECOSIMstep\_12Apr2018\_94X\_mc2017\_realistic\_v14\_ext1-v1/MINIAODSIM
- 7 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_v3\_94X\_mc2017\_realistic\_v14\_ext1-v2
- 8 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_new\_pmx\_94X\_mc2017\_realistic\_v14-v1
- 9 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_new\_pmx\_94X\_mc2017\_realistic\_v14\_ext1-v2
- 10 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_v2\_94X\_mc2017\_realistic\_v14-v2
- 11 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_new\_pmx\_94X\_mc2017\_realistic\_v14-v2
- 12 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v3
- 13 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14-v4
- 14 /RunIIFall17MiniAODv2-PU2017\_12Apr2018\_94X\_mc2017\_realistic\_v14\_ext2-v1

TABLE A.7 – Liste des échantillons MC 2017 utilisés pour modéliser les bruits de fond Z/ $\gamma^*$   $\rightarrow$  LL, W  $\rightarrow$  L $\nu$ ,  $t\bar{t}$ , single top and di-boson (WW, WZ, ZZ) avec L = e,  $\mu$ ,  $\tau$ .

Annexe A. Tableaux des échantillons utilisés dans le canal  $\tau_h\tau_h$

Signal Monte Carlo samples				
Process	Dataset Name	Cross-Section [pb]	Events	Filter Eff.
ggH $h \rightarrow \tau_h\tau_h$	/GluGluHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	3.046	20283379	0.2446
VBF $h \rightarrow \tau_h\tau_h$	/VBFHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.237	24065183	0.2695
W <sup>+</sup> h $\rightarrow \tau_h\tau_h$	/WplusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0527	2716402	0.2397
W <sup>-</sup> h $\rightarrow \tau_h\tau_h$	/WminusHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0358	1876352	0.2563
Zh $\rightarrow \tau_h\tau_h$	/ZHToTauTauUncorrelatedDecay_Filtered_M125_TuneCUETP8M1_13TeV-powheg-pythia8 <sup>1</sup>	0.0594	3049541	0.2513

<sup>1</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-v1/MINIAODSIM

TABLE A.8 – Liste des échantillons MC 2018 utilisés pour modéliser le signal  $pp \rightarrow h \rightarrow \tau_h\tau_h$ .

Background Monte Carlo samples		
Process	Dataset Name	Cross-Section [pb]
$t\bar{t} \rightarrow LL\nu\nu$	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	88.29
$t\bar{t} \rightarrow$ hadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	377.96
$t\bar{t} \rightarrow$ semi-leptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	365.34
Single top	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 <sup>3</sup>	34.97
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 <sup>3</sup>	35.85
	/ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8 <sup>1</sup>	136.02
	/ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8 <sup>1</sup>	80.95
Z $\rightarrow$ LL + jets	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	6077.22 (NNLO)
Z $\rightarrow$ LL + 1jet	/DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	877.8 (LO)
Z $\rightarrow$ LL + 2jets	/DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	304.4 (LO)
Z $\rightarrow$ LL + 3jets	/DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	111.5 (LO)
Z $\rightarrow$ LL + 4jets	/DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	44.03 (LO)
Low mass Z $\rightarrow$ LL + jets	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	18610
W $\rightarrow$ L $\nu$ + jets	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	61526.7 (NNLO)
W $\rightarrow$ L $\nu$ + 1jet	/W1JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	9644.5 (LO)
W $\rightarrow$ L $\nu$ + 2jets	/W2JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	3144.5 (LO)
W $\rightarrow$ L $\nu$ + 3jets	/W3JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	954.8 (LO)
W $\rightarrow$ L $\nu$ + 4jets	/W4JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>2</sup>	485.6 (LO)
EWK W <sup>+</sup> $\rightarrow$ L <sup>+</sup> $\nu$ + 2jets	/EWKWplus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	25.62 (LO)
EWK W <sup>-</sup> $\rightarrow$ L <sup>-</sup> $\nu$ + 2jets	/EWKWminus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	20.25 (LO)
EWK Z $\rightarrow$ LL + 2jets	/EWKZ2Jets_ZToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 <sup>1</sup>	3.987 (LO)
ZZ	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>1</sup>	3.688
ZZ	/ZZTo2L2Nu_13TeV_powheg_pythia8 <sup>4,5</sup>	0.6008
ZZ	/ZZTo4L_13TeV_amcatnloFXFX-pythia8 <sup>4,5</sup>	1.325
WW	/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	11.08
WW	/WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8 <sup>1</sup>	45.99
WZ	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin-pythia8_v2 <sup>1</sup>	6.331
WZ	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin-pythia8_v2 <sup>1</sup>	3.293
WZ	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>1</sup>	11.66
WZ	/WZTo3LNu_13TeV_amcatnloFXFX_madspin-pythia8_v2 <sup>1,4</sup>	5.052
WZ	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin-pythia8 <sup>1</sup>	11.66

<sup>1</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-v1/MINIAODSIM

<sup>2</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-v2/MINIAODSIM

<sup>3</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-ext1-v1/MINIAODSIM

<sup>4</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-ext1-v2/MINIAODSIM

<sup>5</sup> /RunIIAutumn18MiniAOD-102X\_upgrade2018\_realistic\_v15-ext2-v2/MINIAODSIM

TABLE A.9 – Liste des échantillons MC 2018 utilisés pour modéliser les bruits de fond  $Z/\gamma^* \rightarrow LL$ ,  $W \rightarrow L\nu$ ,  $t\bar{t}$ , single top and di-boson (WW, WZ, ZZ) avec  $L = e, \mu, \tau$ .

Annexe A. Tableaux des échantillons utilisés dans le canal  $\tau_h\tau_h$

Channel	Dataset Name
$\tau_h\tau_h$	/EmbeddingRun2016B/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016C/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016D/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016E/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016F/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016G/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2016H/TauTauFinalState <sup>3</sup>

<sup>1</sup>-inputSep16DoubleMu\_mirror\_miniaOD-v2/USER

<sup>2</sup>-inputPromptDoubleMu\_mirror\_miniaOD-v2/USER

<sup>3</sup>-inputPromptDoubleMu\_mirror\_miniaOD-v3/USER

TABLE A.10 – Liste des échantillons embedded 2016 pour le canal  $\tau_h\tau_h$ .

Channel	Dataset Name
$\tau_h\tau_h$	/EmbeddingRun2017B/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2017C/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2017D/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2017E/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2017F/TauTauFinalState <sup>1</sup>

<sup>1</sup>-inputDoubleMu\_94X\_miniaOD-v2/USER

TABLE A.11 – Liste des échantillons embedded 2017 pour le canal  $\tau_h\tau_h$ .

Channel	Dataset Name
$\tau_h\tau_h$	/EmbeddingRun2018A/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2018B/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2018C/TauTauFinalState <sup>1</sup>
$\tau_h\tau_h$	/EmbeddingRun2018D/TauTauFinalState <sup>1</sup>

<sup>1</sup>-inputDoubleMu\_102X\_miniaOD-v1/USER

TABLE A.12 – Liste des échantillons embedded 2018 pour le canal  $\tau_h\tau_h$ .

## Annexe B

# Extraction des corrections d'identification des $\tau_h$

Un échantillon enrichi en événements  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$  est sélectionné. Les sélections suivantes sont appliquées pour 2016 (2017/2018) :

- Le muon doit passer l'ID Medium,  $I_{\text{rel}} < 0.15$ ,  $p_T > 23(28)$  GeV et le système de déclenchement avec un muon seul ;
- Le tau hadronique doit passer  $p_T > 20$  GeV, l'algorithme HPS, le WP Medium de DeepTau contre les jets, le WP Tight de DeepTau contre les muons et le WP VVVLoose de DeepTau contre les électrons ;
- Le muon et le tau doivent avoir des charges électriques opposées et être séparés de  $\Delta R > 0.5$ . Si plus d'une paire vérifie ces conditions, alors la paire contenant les leptons les plus isolés est choisie ;
- La masse transverse  $m_T(\mu, \text{MET})$ , définie comme :

$$m_T(i, j) = \sqrt{2p_T^i p_T^j (1 - \cos \Delta\phi_{ij})} \quad (\text{B.1})$$

doit être inférieur à 40 GeV ;

- Les événements possédant des électrons, muons ou des jets b supplémentaires sont rejetés.

Un échantillon d'événements  $Z/\gamma^* \rightarrow \mu\mu$  est aussi utilisé afin de s'assurer que l'ajustement est capable de différencier les différences dans l'échantillon  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$  dus au facteur correctif tau ID (qui augmente ou diminue le nombre d'événements  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$ ) et d'autres effets comme un décalage dans la section efficace (qui décale simultanément  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$  et  $Z/\gamma^* \rightarrow \mu\mu$ ).

Les sélections d'événements  $Z/\gamma^* \rightarrow \mu\mu$  sont autant que possible similaires à ceux en  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$ . Les sélections suivantes sont appliquées pour 2016 (2017/2018) :

- Les événements contiennent deux muons passant l'ID Medium avec  $I_{\text{rel}} < 0.15$  et  $p_T > 10$  GeV ;

## Annexe B. Extraction des corrections d'identification des $\tau_h$

- Le muon de plus haut  $p_T$  doit de plus avoir  $p_T > 23$  (28) GeV et passer le système de déclenchement avec un muon seul ;
- Les muons de la paire doivent avoir une charge électrique opposée et être séparés de  $\Delta R > 0.5$ . Si plus d'une paire vérifie ces conditions, alors la plus isolée est choisie ;
- La masse invariante de la paire de muons doit être comprise entre 70 et 120 GeV ;
- Les événements avec des électrons, muons et jets b supplémentaires sont rejetés.

Dans la catégorie  $\tau_\mu\tau_h$ , la masse invariante du muon et du tau hadronique est utilisée comme variable discriminante. Seul le taux d'événements est utilisé dans la catégorie  $\mu\mu$  afin de contraindre le nombre d'événements DY. Un paramètre de taux commun est ajouté dans le but d'échelonner simultanément les événements  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$  et  $Z/\gamma^* \rightarrow \mu\mu$  dans chaque catégorie. Un paramètre de taux supplémentaire correspondant au facteur d'échelle tau ID et qui échelonne donc seulement le nombre d'événements  $Z/\gamma^* \rightarrow \tau_\mu\tau_h$  est aussi inclus. Des ajustements simultanés du maximum de vraisemblance des catégories  $\mu\mu$  et  $\tau_\mu\tau_h$  sont appliqués afin d'extraire les facteurs d'échelle tau ID.

Les bruits de fond suivant sont considérés lors des ajustements :  $t\bar{t}$ , di-boson, single-top, W+jets et QCD. Des facteurs d'échelle sont aussi calculés pour les échantillons embedded. Pour ces derniers, la petite fraction d'événements DY dans le canal  $\tau_\mu\tau_h$  qui n'est pas couvert par les échantillons embedded est aussi considérée comme un bruit de fond. Ce sont des événements DY où le tau est faussement reconstruit comme tel et est en réalité un muon ou un jet. Tous les processus de bruits de fond, mis à part la QCD, sont modélisés avec la simulation MC. Le bruit de fond multijets est estimé à partir de données avec deux candidats de même signe électrique échelonnées par un facteur constant égal à 1.12 afin de prendre en compte les différents taux dans la production QCD de mêmes signes et différents signes. Le nombre d'événements W+jets est aussi estimé en utilisant les données en normalisant le nombre d'événements aux données dans la région de haut  $m_T$  ( $m_T > 70$  GeV) qui est très pure en événements W+jets.

Les ajustements sont effectués séparément pour les échantillons MC et embedded. Pour ces derniers, le canal  $\mu\mu$  est modélisé avec un échantillon embedded d'événements  $Z \rightarrow \mu\mu$  produit avec la même procédure que pour  $Z \rightarrow \tau\tau$  sauf que les muons sélectionnés dans les données sont enlevés et remplacés par des muons simulés au lieu de taus.

De plus, les incertitudes systématiques suivantes affectant la normalisation et les formes des modèles de signaux et bruits de fond sont aussi incluses lors des ajustements :  
**incertitudes de normalisation :**

- **luminosité** : l'incertitude sur la luminosité est de 2.5% et est appliquée à tous les processus où le nombre d'événements est estimé à partir de la simulation MC ;
- **muon ID/efficacité d'isolement** : l'incertitude sur l'identification du muon et sur son isolement est de 1% par muon ;

- **taux de faux  $j \rightarrow \tau_h$**  : une incertitude de 20% est appliquée à tous les processus contenant un faux  $j \rightarrow \tau_h$  où le nombre d'évènements est estimé à partir de la simulation MC ;
- **taux de faux  $\mu \rightarrow \tau_h$**  : une incertitude de 20% est appliquée au processus  $Z \rightarrow \mu\mu$  dans les canaux  $\tau_\mu\tau_h$  où un des muons est faussement reconstruit comme un tau hadronique ;
- **efficacité du système de déclenchement** : l'incertitude sur l'efficacité du système de déclenchement avec un muon seul est de 2% ;
- **efficacité d'identification des jets b** : l'incertitude sur l'efficacité d'identification des jets b dû au rejet de ces derniers est appliquée sur les évènements diboson et  $t\bar{t}$  seulement et est d'environ 5% ;
- **incertitude de normalisation du bruit de fond** :
  - 4% d'incertitude sur les évènements embedded  $Z \rightarrow \mu\mu$  et  $Z \rightarrow \tau\tau$  ;
  - 2% d'incertitude sur les évènements MC  $Z \rightarrow \mu\mu$  et  $Z \rightarrow \tau\tau$  afin de tenir compte de l'incertitude sur la section efficace DY ;
  - 5% d'incertitude sur les évènements diboson et single-top afin de tenir compte des incertitudes sur les sections efficaces de production ;
  - 4.2% d'incertitude sur les évènements  $t\bar{t}$  afin de tenir compte des incertitudes sur la section efficace de production ;
  - 10% d'incertitude sur les évènements W+jets afin de tenir compte de l'extrapolation entre les régions de haute et de basse  $m_T$  lors de l'estimation du nombre d'évènements W ;
  - 10% d'incertitude sur les évènements QCD afin de tenir compte des incertitudes de mesures du ratio OS/SS ;
  - Une incertitude additionnelle, lors de l'ajustement des facteurs d'échelle embedded seulement, due aux facteurs d'échelle tau ID dans les MC, de 15% est appliquée. Elle n'affecte que les évènements MC non couverts par la procédure d'embedding, c'est à dire une fraction des évènements diboson et  $t\bar{t}$ .

**incertitudes de formes :**

- **échelle d'énergie des taus hadroniques** : quatre incertitudes non corrélées et variant entre  $\sim 1$  et 1.5% ;
- **échelle d'énergie du  $\mu$**  : l'incertitude de l'échelle d'énergie du muon est fonction de la pseudorapidité et varie entre 0.4 et 2.7% ;
- **incertitudes statistiques** : le nombre d'évènements limité dans les modèles de signaux et de bruits de fond est pris en compte par la méthode Barlow-Beeston [140, 141] ;
- **repondération du  $p_T$  du top** : l'incertitude est appliquée comme zéro correction et deux fois la correction de repondération du  $p_T$  du top sur des évènements  $t\bar{t}$  (voir sous-section 5.5.8) ;



- **repondération du  $p_T$  du Z DY** : l'incertitude est appliquée comme  $\pm 10\%$  de la correction du  $p_T$  du Z DY sur des événements MC  $Z \rightarrow \mu\mu$  et  $Z \rightarrow \tau\tau$  (voir sous-section 5.5.7);
- des incertitudes sur la résolution et l'échelle de la MET sont estimées en faisant varier la correction de recul.

Des exemples de comparaisons de données et bruits de fond après ajustements pour  $p_T > 40$  GeV en 2018 sont montrés en Fig. B.1 pour les mesures avec simulations MC et embedded avec MVA DM = 10 [117]. Un bon accord avec les données est observé. Les résultats des ajustements du maximum de vraisemblance des facteurs d'échelle pour différents MVA DM sont en Fig. B.2 [117].

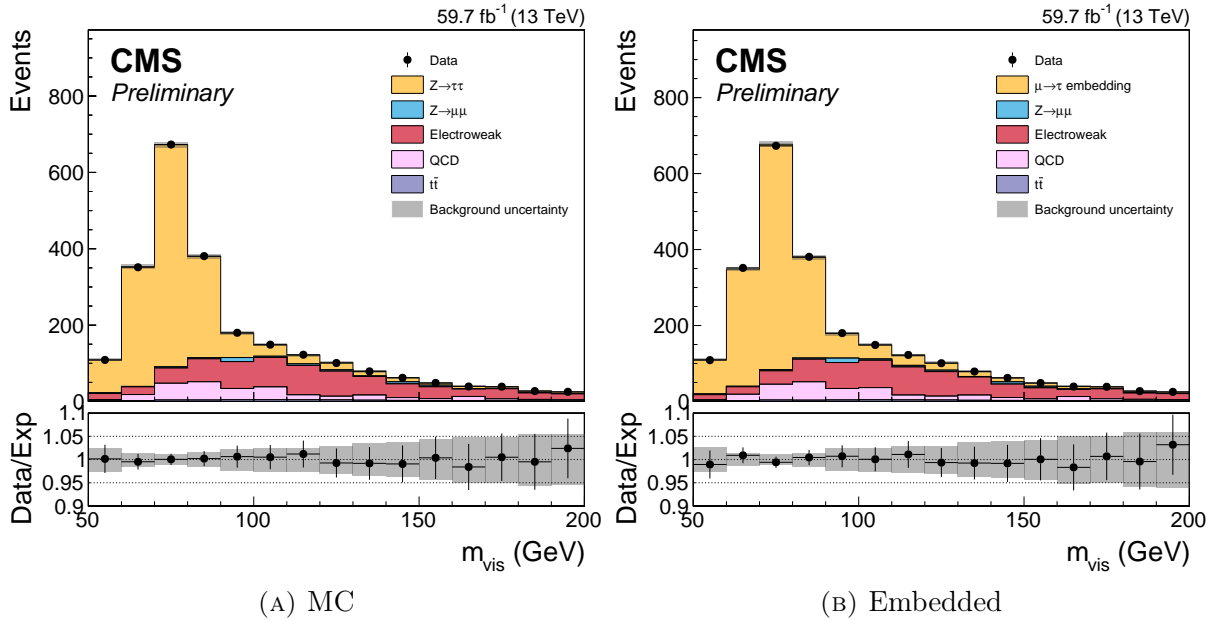


FIGURE B.1 – Distributions après ajustement montrant les accords données/bruits de fond pour  $p_T > 40$  GeV pour le MVA DM = 10 dans lesquels les facteurs d'échelle tau ID MC (gauche) et embedded (droite) sont mesurés. Les bandes d'incertitudes grises incluent les incertitudes statistiques et systématiques [117].

Annexe B. Extraction des corrections d'identification des  $\tau_h$

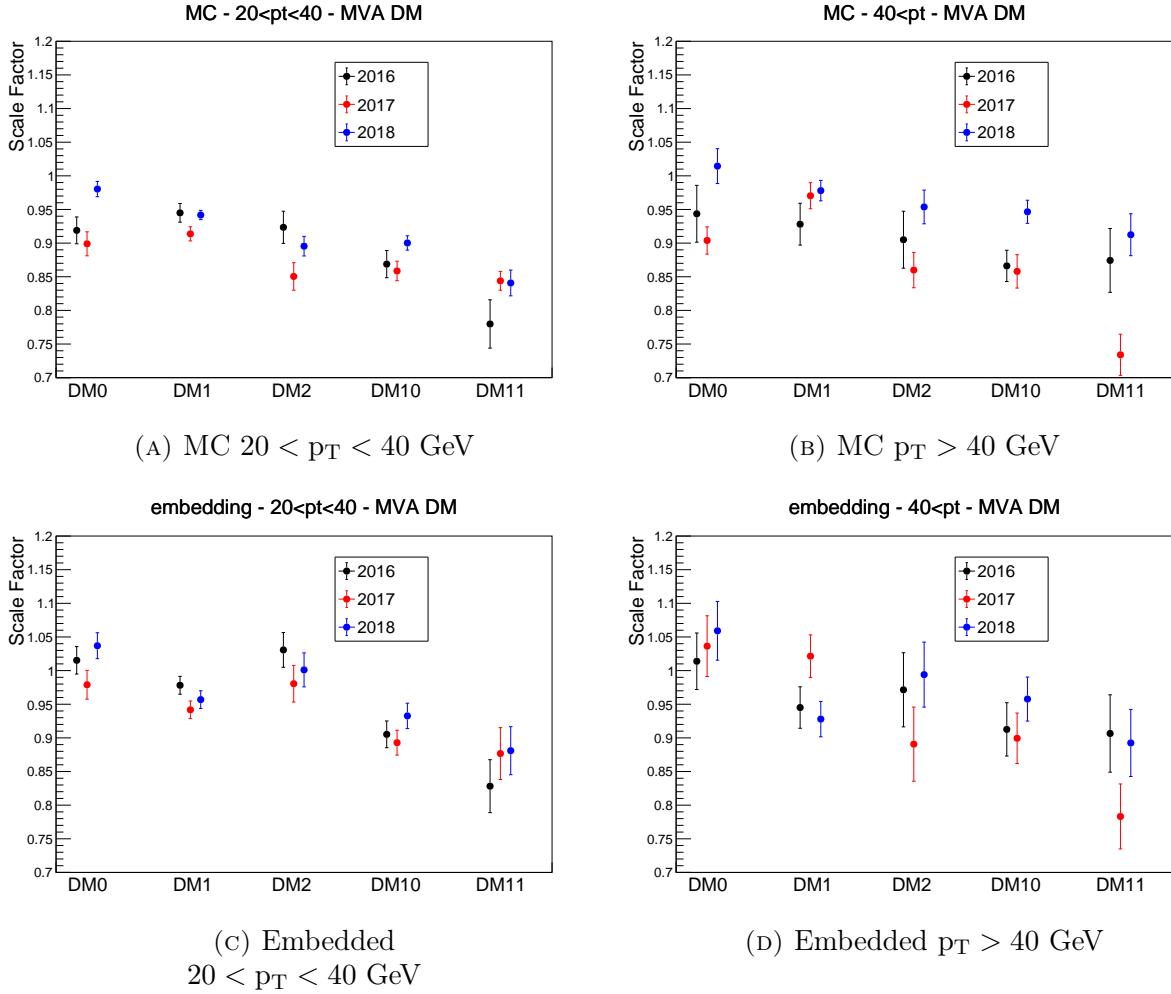


FIGURE B.2 – Valeurs du meilleur ajustement des facteurs d'échelle tau ID MC (haut) et embedded (bas) pour  $20 < p_T < 40$  GeV (gauche) et  $p_T > 40$  GeV (droite) [117].

## Annexe C

# Distributions avant ajustement de $\phi_{CP}$ dans le canal $a_1^{3pr} a_1^{3pr}$

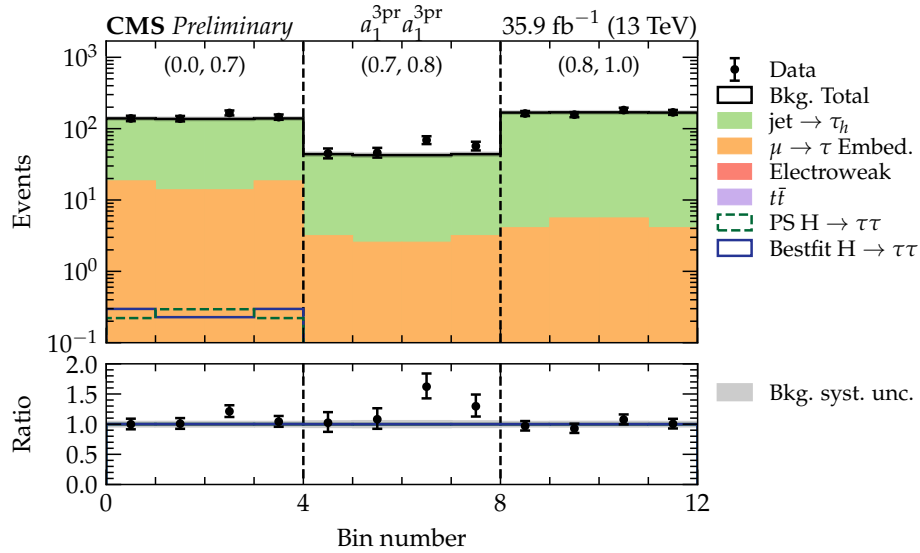


FIGURE C.1 – Distribution 2016 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **jets fakes**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

Annexe C. Distributions avant ajustement de  $\phi_{CP}$  dans le canal  $a_1^{3pr} a_1^{3pr}$

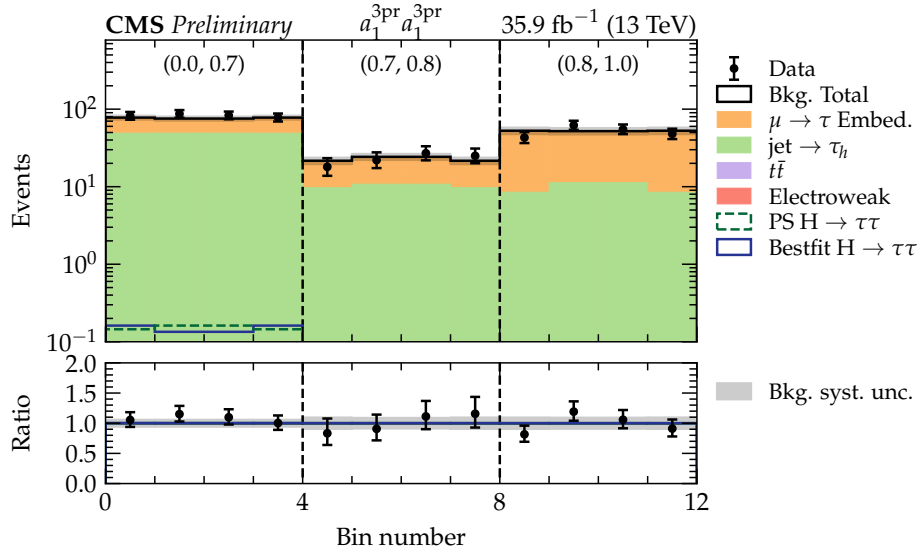


FIGURE C.2 – Distribution 2016 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **ZTT embedded**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

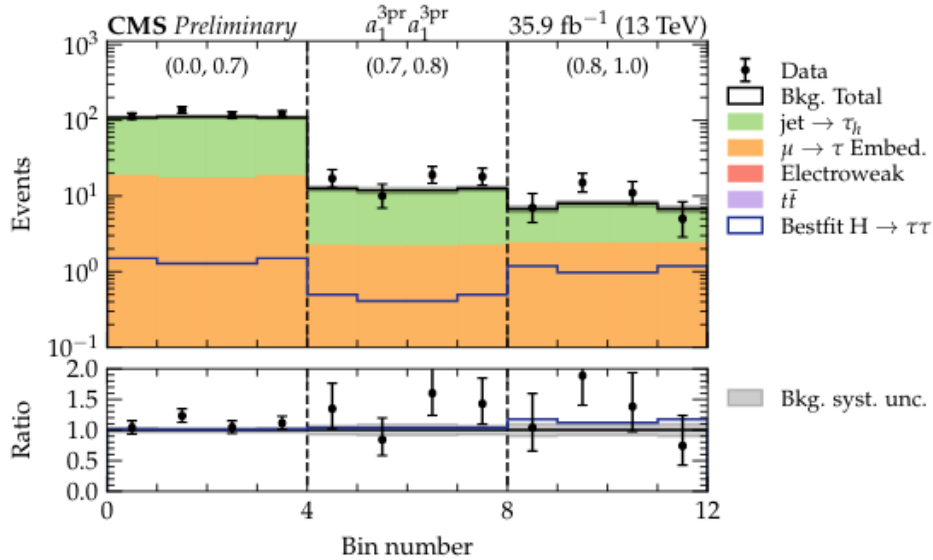


FIGURE C.3 – Distribution 2016 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

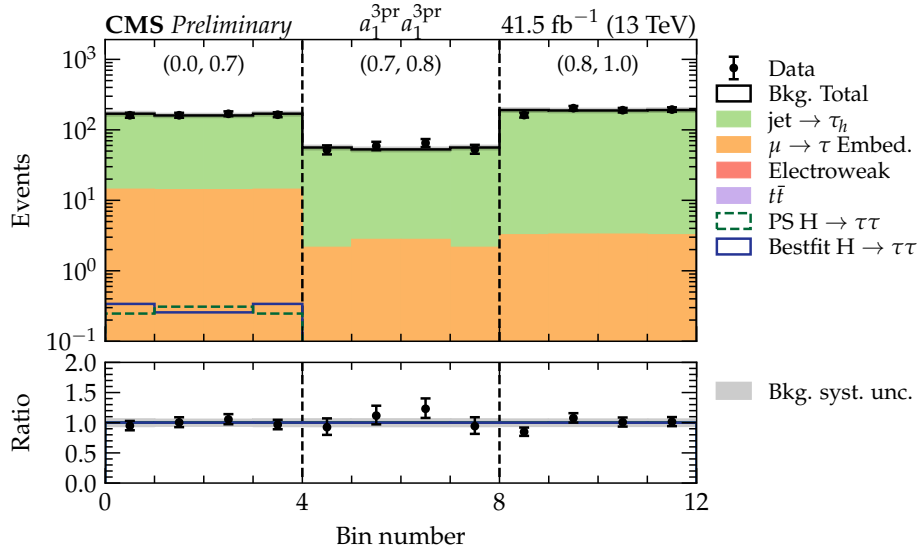


FIGURE C.4 – Distribution 2017 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **jets fakes**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

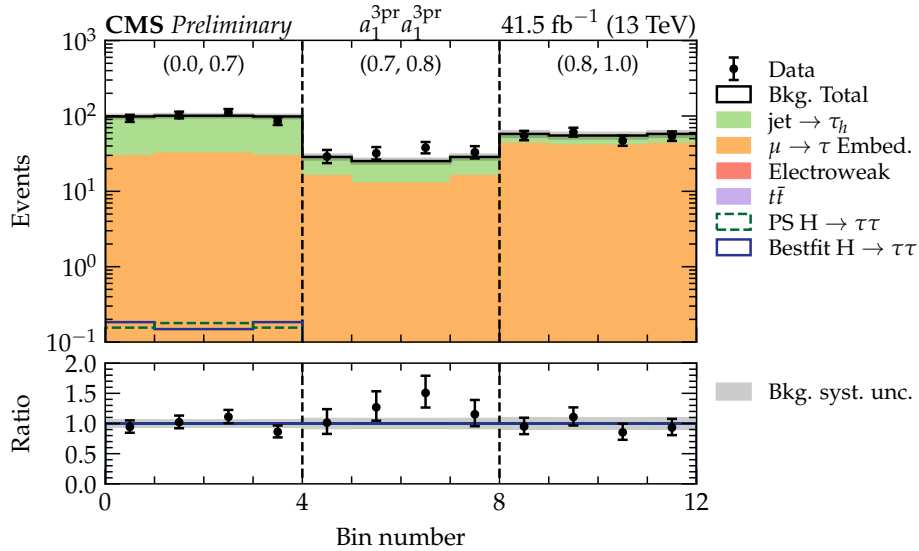


FIGURE C.5 – Distribution 2017 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **ZTT embedded**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

Annexe C. Distributions avant ajustement de  $\phi_{CP}$  dans le canal  $a_1^{3pr} a_1^{3pr}$

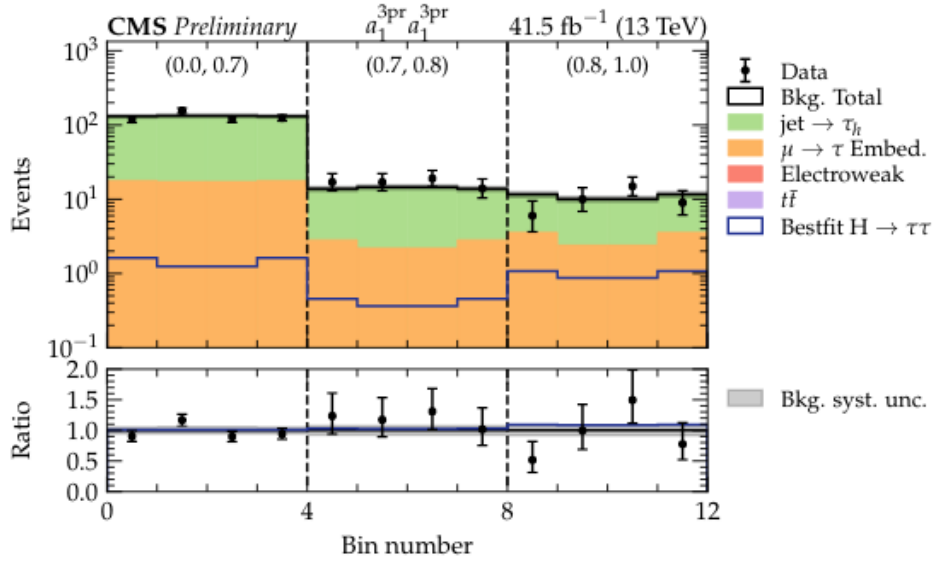


FIGURE C.6 – Distribution 2017 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

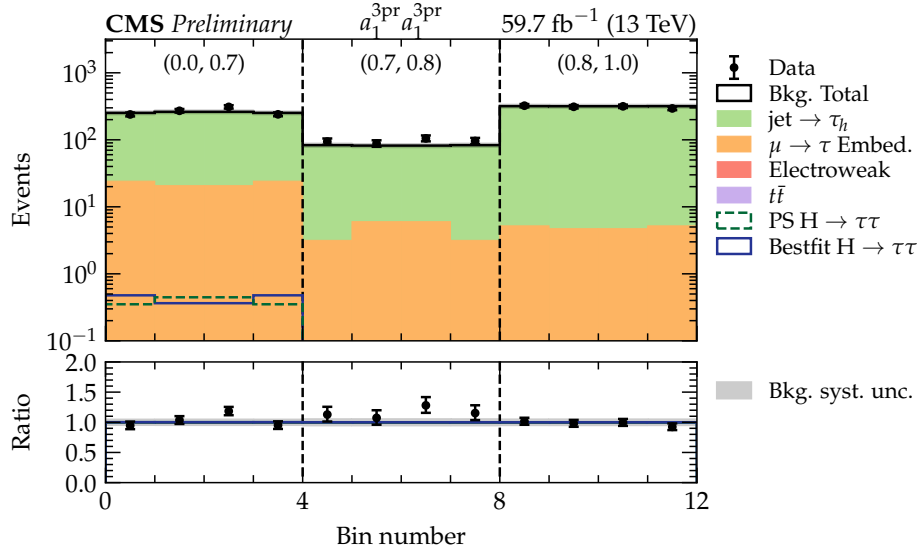


FIGURE C.7 – Distribution 2018 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **jets fakes**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{cp}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

Annexe C. Distributions avant ajustement de  $\phi_{CP}$  dans le canal  $a_1^{3pr} a_1^{3pr}$

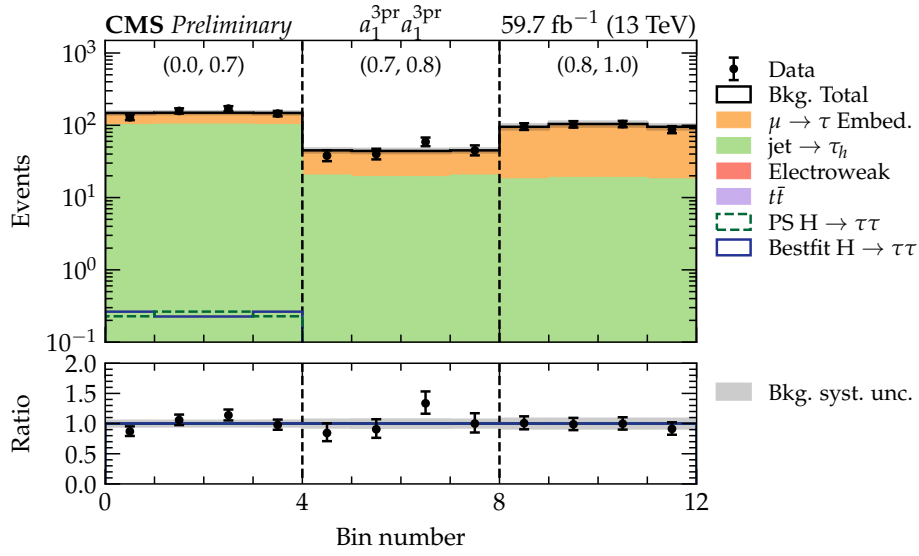


FIGURE C.8 – Distribution 2018 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie **ZTT embedded**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

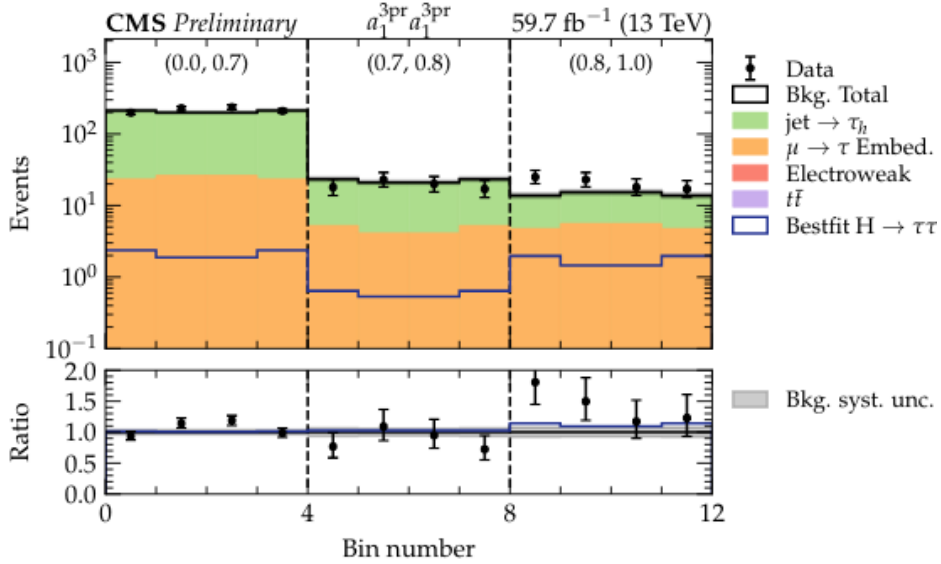


FIGURE C.9 – Distribution 2018 avant ajustement de  $\phi_{CP}$  du canal  $a_1^{3pr} a_1^{3pr}$  dans la catégorie de **signal**. L'axe des abscisses représente les valeurs cycliques de  $\phi_{CP}$  entre 0 et  $2\pi$ , réparties en intervalles de score BDT [0.0,0.7,0.8,1.0]. Le ratio donnée/MC est donné en bas de la distribution.

# Bibliographie

- [1] V. KHACHATRYAN et al. “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at  $\sqrt{s} = 8$  TeV and cross section ratios to 2.76 and 7 TeV”. In : *Journal of High Energy Physics* 2017.3 (2017). ISSN : 1029-8479. DOI : [10.1007/jhep03\(2017\)156](https://doi.org/10.1007/JHEP03(2017)156). URL : [http://dx.doi.org/10.1007/JHEP03\(2017\)156](http://dx.doi.org/10.1007/JHEP03(2017)156).
- [2] C.S. WU et al. “Experimental Test of Parity Conservation in  $\beta$  Decay”. In : *Phys. Rev.* 105 (1957), p. 1413-1414. DOI : [10.1103/PhysRev.105.1413](https://doi.org/10.1103/PhysRev.105.1413).
- [3] F. ENGLERT et R. BROUT. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In : *Phys. Rev. Lett.* 13 (9 1964), p. 321. DOI : [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321). URL : <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [4] Peter W. HIGGS. “Broken Symmetries and the Masses of Gauge Bosons”. In : *Phys. Rev. Lett.* 13 (16 1964), p. 508. DOI : [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508). URL : <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [5] G. S. GURALNIK, C. R. HAGEN et T. W. B. KIBBLE. “Global Conservation Laws and Massless Particles”. In : *Phys. Rev. Lett.* 13 (20 1964), p. 585. DOI : [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585). URL : <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [6] John ELLIS, Mary K GAILLARD et Dimitri V NANOPOULOS. “A Historical Profile of the Higgs Boson. An Updated Historical Profile of the Higgs Boson”. In : arXiv :1504.07217. KCL-PH-TH-2015-20. LCTS-2015-10. CERN-PH-TH-2015-098 (2015), 22 p. DOI : [10.1142/9789814733519\\_0014](https://doi.org/10.1142/9789814733519_0014). URL : <http://cds.cern.ch/record/2012465>.
- [7] Hideki YUKAWA. “On the Interaction of Elementary Particles. I”. In : *Progress of Theoretical Physics Supplement* 1 (jan. 1955), p. 1-10. ISSN : 0375-9687. DOI : [10.1143/PTPS.1.1](https://doi.org/10.1143/PTPS.1.1). eprint : <https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf>. URL : <https://doi.org/10.1143/PTPS.1.1>.
- [8] A.D. SAKHAROV. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In : *Sov. Phys. Usp.* 34.5 (1991), p. 392-393. DOI : [10.1070/PU1991v034n05ABEH002497](https://doi.org/10.1070/PU1991v034n05ABEH002497).



## Bibliographie

- [9] V. A. KUZMIN, V. A. RUBAKOV et M. E. SHAPOSHNIKOV. “On anomalous electroweak baryon-number non-conservation in the early universe”. In : *Physics Letters B* 155.1-2 (mai 1985), p. 36-42. DOI : [10.1016/0370-2693\(85\)91028-7](https://doi.org/10.1016/0370-2693(85)91028-7).
- [10] Werner BERNREUTHER. “CP violation and baryogenesis”. In : *Lect. Notes Phys.* 591 (2002), p. 237-293. arXiv : [hep-ph/0205279](https://arxiv.org/abs/hep-ph/0205279).
- [11] James M. CLINE, Kimmo KAINULAINEN et Axel P. VISCHER. “Dynamics of two-Higgs-doublet CP violation and baryogenesis at the electroweak phase transition”. In : *Physical Review D* 54.4 (1996), 2451–2472. ISSN : 1089-4918. DOI : [10.1103/PhysRevD.54.2451](https://doi.org/10.1103/PhysRevD.54.2451). URL : <http://dx.doi.org/10.1103/PhysRevD.54.2451>.
- [12] Michael JOYCE, Tomislav PROKOPEC et Neil TUROK. “Nonlocal electroweak baryogenesis. I. Thin wall regime”. In : *Physical Review D* 53.6 (1996), 2930–2957. ISSN : 1089-4918. DOI : [10.1103/PhysRevD.53.2930](https://doi.org/10.1103/PhysRevD.53.2930). URL : <http://dx.doi.org/10.1103/PhysRevD.53.2930>.
- [13] S.F. KING et al. “Exploring the CP-violating NMSSM : EDM constraints and phenomenology”. In : *Nuclear Physics B* 901 (2015), p. 526 -555. ISSN : 0550-3213. DOI : <https://doi.org/10.1016/j.nuclphysb.2015.11.003>. URL : <http://www.sciencedirect.com/science/article/pii/S0550321315003788>.
- [14] A.M. SIRUNYAN et al. “A measurement of the Higgs boson mass in the diphoton decay channel”. In : *Physics Letters B* 805 (2020), p. 135425. ISSN : 0370-2693. DOI : [10.1016/j.physletb.2020.135425](https://doi.org/10.1016/j.physletb.2020.135425). URL : <http://dx.doi.org/10.1016/j.physletb.2020.135425>.
- [15] A. M. SIRUNYAN et al. “Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state”. In : *Physical Review D* 99.11 (2019). ISSN : 2470-0029. DOI : [10.1103/PhysRevD.99.112003](https://doi.org/10.1103/PhysRevD.99.112003). URL : <http://dx.doi.org/10.1103/PhysRevD.99.112003>.
- [16] S. CHATRCHYAN et al. “Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states”. In : *Journal of High Energy Physics* 2014.1 (2014). ISSN : 1029-8479. DOI : [10.1007/jhep01\(2014\)096](https://doi.org/10.1007/jhep01(2014)096). URL : [http://dx.doi.org/10.1007/JHEP01\(2014\)096](http://dx.doi.org/10.1007/JHEP01(2014)096).
- [17] S. CHATRCHYAN et al. “Measurement of the properties of a Higgs boson in the four-lepton final state”. In : *Phys. Rev. D* 89 (9 2014), p. 092007. DOI : [10.1103/PhysRevD.89.092007](https://doi.org/10.1103/PhysRevD.89.092007). URL : <https://link.aps.org/doi/10.1103/PhysRevD.89.092007>.
- [18] V. KHACHATRYAN et al. “Observation of the diphoton decay of the Higgs boson and measurement of its properties”. In : *The European Physical Journal C* 74.10 (2014). ISSN : 1434-6052. DOI : [10.1140/epjc/s10052-014-3076-z](https://doi.org/10.1140/epjc/s10052-014-3076-z). URL : <http://dx.doi.org/10.1140/epjc/s10052-014-3076-z>.

## Bibliographie

- [19] S. CHATRCHYAN et al. “Search for the standard model Higgs boson produced in association with a  $W$  or a  $Z$  boson and decaying to bottom quarks”. In : *Phys. Rev. D* 89 (1 2014), p. 012003. DOI : 10.1103/PhysRevD.89.012003. URL : <https://link.aps.org/doi/10.1103/PhysRevD.89.012003>.
- [20] S. CHATRCHYAN et al. “Evidence for the 125 GeV Higgs boson decaying to a pair of  $\tau$  leptons”. In : *Journal of High Energy Physics* 2014.5 (2014). ISSN : 1029-8479. DOI : 10.1007/jhep05(2014)104. URL : [http://dx.doi.org/10.1007/JHEP05\(2014\)104](http://dx.doi.org/10.1007/JHEP05(2014)104).
- [21] *Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. Rapp. tech. CMS-PAS-HIG-19-006. Geneva : CERN, 2020. URL : <https://cds.cern.ch/record/2725423>.
- [22] A. M. SIRUNYAN et al. “Observation of  $t\bar{t}H$  Production”. In : *Phys. Rev. Lett.* 120 (23 2018), p. 231801. DOI : 10.1103/PhysRevLett.120.231801. URL : <https://link.aps.org/doi/10.1103/PhysRevLett.120.231801>.
- [23] Georges AAD et al. “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”. In : *Phys. Lett. B* 726 (2013), p. 120. DOI : 10.1016/j.physletb.2013.08.026. arXiv : 1307.1432 [hep-ex].
- [24] Georges AAD et al. “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”. In : *Eur. Phys. J. C* 75 (2015), p. 476. DOI : 10.1140/epjc/s10052-015-3685-1. arXiv : 1506.05669 [hep-ex].
- [25] Georges AAD et al. “Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector”. In : *Eur. Phys. J. C* 76 (2016), p. 658. DOI : 10.1140/epjc/s10052-016-4499-5. arXiv : 1602.04516 [hep-ex].
- [26] Morad AABOUD et al. “Measurement of inclusive and differential cross sections in the  $H \rightarrow ZZ^* \rightarrow 4\ell$  decay channel in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In : *JHEP* 10 (2017), p. 132. DOI : 10.1007/JHEP10(2017)132. arXiv : 1708.02810 [hep-ex].
- [27] Morad AABOUD et al. “Measurement of the Higgs boson coupling properties in the  $H \rightarrow ZZ^* \rightarrow 4\ell$  decay channel at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In : *JHEP* 03 (2018), p. 095. DOI : 10.1007/JHEP03(2018)095. arXiv : 1712.02304 [hep-ex].
- [28] Morad AABOUD et al. “Measurements of Higgs boson properties in the diphoton decay channel with  $36 \text{ fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In : *Phys. Rev. D* 98 (2018), p. 052005. DOI : 10.1103/PhysRevD.98.052005. arXiv : 1802.04146 [hep-ex].
- [29] Serguei CHATRCHYAN et al. “On the mass and spin-parity of the Higgs boson candidate via its decays to  $Z$  boson pairs”. In : *Phys. Rev. Lett.* 110 (2013), p. 081803. DOI : 10.1103/PhysRevLett.110.081803. arXiv : 1212.6639 [hep-ex].

## Bibliographie

- [30] Serguei CHATRCHYAN et al. “Measurement of the properties of a Higgs boson in the four-lepton final state”. In : *Phys. Rev. D* 89 (2014), p. 092007. DOI : [10.1103/PhysRevD.89.092007](https://doi.org/10.1103/PhysRevD.89.092007). arXiv : [1312.5353](https://arxiv.org/abs/1312.5353) [hep-ex].
- [31] Vardan KHACHATRYAN et al. “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”. In : *Phys. Rev. D* 92 (2015), p. 012004. DOI : [10.1103/PhysRevD.92.012004](https://doi.org/10.1103/PhysRevD.92.012004). arXiv : [1411.3441](https://arxiv.org/abs/1411.3441) [hep-ex].
- [32] Vardan KHACHATRYAN et al. “Limits on the Higgs boson lifetime and width from its decay to four charged leptons”. In : *Phys. Rev. D* 92 (2015), p. 072010. DOI : [10.1103/PhysRevD.92.072010](https://doi.org/10.1103/PhysRevD.92.072010). arXiv : [1507.06656](https://arxiv.org/abs/1507.06656) [hep-ex].
- [33] Vardan KHACHATRYAN et al. “Combined search for anomalous pseudoscalar HVV couplings in  $VH(H \rightarrow b\bar{b})$  production and  $H \rightarrow VV$  decay”. In : *Phys. Lett. B* 759 (2016), p. 672. DOI : [10.1016/j.physletb.2016.06.004](https://doi.org/10.1016/j.physletb.2016.06.004). arXiv : [1602.04305](https://arxiv.org/abs/1602.04305) [hep-ex].
- [34] Albert M SIRUNYAN et al. “Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state”. In : *Phys. Lett. B* 775 (2017), p. 1. DOI : [10.1016/j.physletb.2017.10.021](https://doi.org/10.1016/j.physletb.2017.10.021). arXiv : [1707.00541](https://arxiv.org/abs/1707.00541) [hep-ex].
- [35] CMS COLLABORATION. “Measurements of  $t\bar{t}H$  production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel”. In : *Phys. Rev. Lett.* 125. arXiv : [2003.10866](https://arxiv.org/abs/2003.10866). CMS-HIG-19-013-003. 6 (2020), 061801. 18 p. DOI : [10.1103/PhysRevLett.125.061801](https://doi.org/10.1103/PhysRevLett.125.061801). URL : <https://cds.cern.ch/record/2713705>.
- [36] Lyn EVANS et Shinichiro MICHIZONO. *The International Linear Collider Machine Staging Report 2017*. 2017. arXiv : [1711.00568](https://arxiv.org/abs/1711.00568) [physics.acc-ph].
- [37] Chris ADOLPHSEN et al. *The International Linear Collider Technical Design Report - Volume 3.II : Accelerator Baseline Design*. 2013. arXiv : [1306.6328](https://arxiv.org/abs/1306.6328) [physics.acc-ph].
- [38] K DESCH et al. “Probing the CP nature of the Higgs boson at linear colliders with  $\tau$  spin correlations; the case of mixed scalar-pseudoscalar couplings”. In : *Physics Letters B* 579.1-2 (2004), p. 157. ISSN : 0370-2693. DOI : [10.1016/j.physletb.2003.10.074](https://doi.org/10.1016/j.physletb.2003.10.074). arXiv : [0307331](https://arxiv.org/abs/0307331) [hep-ph]. URL : <http://dx.doi.org/10.1016/j.physletb.2003.10.074>.
- [39] G.R. BOWER et al. “Measuring the Higgs boson’s parity using  $\tau \rightarrow \rho\nu$ ”. In : *Physics Letters B* 543.3-4 (2002), 227–234. ISSN : 0370-2693. DOI : [10.1016/S0370-2693\(02\)02445-0](https://doi.org/10.1016/S0370-2693(02)02445-0). arXiv : [hep-ph/0204292](https://arxiv.org/abs/hep-ph/0204292) [hep-ph]. URL : [http://dx.doi.org/10.1016/S0370-2693\(02\)02445-0](http://dx.doi.org/10.1016/S0370-2693(02)02445-0).

## Bibliographie

- [40] S. BERGE et al. “How to pin down the CP quantum numbers of a Higgs boson in its  $\tau$  decays at the LHC”. In : *Physical Review D* 84.11 (2011). ISSN : 1550-2368. DOI : 10.1103/physrevd.84.116003. URL : <http://dx.doi.org/10.1103/PhysRevD.84.116003>.
- [41] Stefan BERGE, Werner BERNREUTHER et Sebastian KIRCHNER. “Determination of the Higgs CP-mixing angle in the tau decay channels at the LHC including the Drell–Yan background”. In : *The European Physical Journal C* 74.11 (2014), p. 3164. ISSN : 1434-6052. DOI : 10.1140/epjc/s10052-014-3164-0. arXiv : 1408.0798. URL : <https://doi.org/10.1140/epjc/s10052-014-3164-0>.
- [42] Stefan BERGE, Werner BERNREUTHER et Sebastian KIRCHNER. “Prospects of constraining the Higgs boson’s CP nature in the tau decay channel at the LHC”. In : *Physical Review D* 92.9 (2015). ISSN : 1550-2368. DOI : 10.1103/physrevd.92.096012. arXiv : 151003850 [hep-ph]. URL : <http://dx.doi.org/10.1103/PhysRevD.92.096012>.
- [43] Stefan BERGE, Werner BERNREUTHER et Sebastian KIRCHNER. “Determination of the Higgs CP-mixing angle in the tau decay channels”. In : *Nuclear and Particle Physics Proceedings* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), p. 841 -845. ISSN : 2405-6014. DOI : <https://doi.org/10.1016/j.nuclphysbps.2015.09.129>. URL : <http://www.sciencedirect.com/science/article/pii/S2405601415006185>.
- [44] *Probing the CP nature of the Higgs boson coupling to  $\tau$  leptons at HL-LHC*. Rapp. tech. ATL-PHYS-PUB-2019-008. Geneva : CERN, 2019. URL : <http://cds.cern.ch/record/2665667>.
- [45] Joachim BROD, Ulrich HAISCH et Jure ZUPAN. “Constraints on CP-violating Higgs couplings to the third generation”. In : *Journal of High Energy Physics* 2013.11 (2013). ISSN : 1029-8479. DOI : 10.1007/jhep11(2013)180. URL : [http://dx.doi.org/10.1007/JHEP11\(2013\)180](http://dx.doi.org/10.1007/JHEP11(2013)180).
- [46] *Particle Data Group*. 2020. URL : <https://pdg.lbl.gov/2020/tables/rpp2020-sum-leptons.pdf>.
- [47] C. ABEL et al. “Measurement of the Permanent Electric Dipole Moment of the Neutron”. In : *Phys. Rev. Lett.* 124 (8 2020), p. 081803. DOI : 10.1103/PhysRevLett.124.081803. URL : <https://link.aps.org/doi/10.1103/PhysRevLett.124.081803>.
- [48] M. KRÄMER et al. “Prospects of measuring the parity of Higgs particles”. In : *Zeitschrift für Physik C Particles and Fields* 64.1 (1994), 21–30. ISSN : 1434-6052. DOI : 10.1007/bf01557231. arXiv : hep-ph/9404280 [hep-ph]. URL : <http://dx.doi.org/10.1007/BF01557231>.
- [49] W. BERNREUTHER et S. BERGE. Figure used with permission of the authors.

## Bibliographie

- [50] G. AAD et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In : *Physics Letters B* 716.1 (2012), p. 1. ISSN : 0370-2693. DOI : 10.1016/j.physletb.2012.08.020. arXiv : 12077214 [hep-ph]. URL : <http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [51] S. CHATRCHYAN et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In : *Physics Letters B* 716.1 (2012), p. 30. ISSN : 0370-2693. DOI : 10.1016/j.physletb.2012.08.021. arXiv : 12077235 [hep-ph]. URL : <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [52] Esma MOBS. “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019”. In : (2019). General Photo. URL : <https://cds.cern.ch/record/2684277>.
- [53] G. BRUMFIEL. “Eight-month delay for LHC”. In : *Nature* 455 (2008), 1015–1015. DOI : 10.1038/4551015a.
- [54] CERN. *CERN releases analysis of LHC incident*. <https://home.cern/news/press-release/cern/cern-releases-analysis-lhc-incident>.
- [55] G AAD et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In : *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. 437 p. DOI : 10.1088/1748-0221/3/08/S08003. URL : <https://cds.cern.ch/record/1129811>.
- [56] S CHATRCHYAN et al. “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment”. In : *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. 361 p. DOI : 10.1088/1748-0221/3/08/S08004. URL : <https://cds.cern.ch/record/1129810>.
- [57] K. AAMODT et al. “The ALICE experiment at the CERN LHC”. In : *JINST* 3 (2008), S08002. DOI : 10.1088/1748-0221/3/08/S08002.
- [58] Helmut SATZ. “The Quark-Gluon Plasma – A Short Introduction”. In : *Nuclear Physics A* 862-863 (2011). The Sixth International Conference on Physics and Astrophysics of Quark Gluon Plasma (ICPAQGP-2010), p. 4 -12. ISSN : 0375-9474. DOI : <https://doi.org/10.1016/j.nuclphysa.2011.05.014>. URL : <http://www.sciencedirect.com/science/article/pii/S0375947411003435>.
- [59] A Augusto ALVES et al. “The LHCb Detector at the LHC”. In : *JINST* 3.LHCb-DP-2008-001. CERN-LHCb-DP-2008-001 (2008). Also published by CERN Geneva in 2010, S08005. DOI : 10.1088/1748-0221/3/08/S08005. URL : <https://cds.cern.ch/record/1129809>.
- [60] CMS. *Public CMS Luminosity Information*. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [61] *CMS detector design*. <https://cms.cern/news/cms-detector-design>.

## Bibliographie

- [62] Thomas LENZI. “Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”. Presented 28 May 2013. 2013. URL : <https://cds.cern.ch/record/1551944>.
- [63] D. SPRENGER et al. “Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap”. In : *JINST* 5.arXiv :1003.5460. CMS-NOTE-2010-002 (2010). Comments : 11 pages, 8 figures. CMS NOTE-2010/002, P06007. 11 p. DOI : 10.1088/1748-0221/5/06/P06007. URL : <http://cds.cern.ch/record/1256099>.
- [64] *Electromagnetic calorimeter calibration with 7 TeV data*. Rapp. tech. CMS-PAS-EGM-10-003. Geneva : CERN, 2010. URL : <http://cds.cern.ch/record/1279350>.
- [65] S. CHATRCHYAN et al. “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”. In : *Journal of Instrumentation* 5 (mar. 2010), T03012.
- [66] Sergey CHATRCHYAN et al. “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays”. In : *Journal of Instrumentation* 5 (mar. 2010), T03021.
- [67] Vyacheslav KLYUKHIN et al. “Measurement of the CMS Magnetic Field”. In : *Applied Superconductivity, IEEE Transactions on* 18 (juil. 2008), p. 395 -398. DOI : 10.1109/TASC.2008.921242.
- [68] V. KARJAVIN. “The upgrade of endcap CMS muon system”. In : *Physics of Particles and Nuclei* 49 (jan. 2018), p. 54-59. DOI : 10.1134/S1063779618010197.
- [69] A COLALEO et al. *CMS Technical Design Report for the Muon Endcap GEM Upgrade*. Rapp. tech. CERN-LHCC-2015-012. CMS-TDR-013. 2015. URL : <https://cds.cern.ch/record/2021453>.
- [70] A COLALEO et al. *CMS Technical Design Report for the Muon Endcap GEM Upgrade*. Rapp. tech. CERN-LHCC-2015-012. CMS-TDR-013. 2015. URL : <https://cds.cern.ch/record/2021453>.
- [71] Vardan KHACHATRYAN et al. “The CMS trigger system”. In : *JINST* 12.01 (2017), P01020. DOI : 10.1088/1748-0221/12/01/P01020. arXiv : 1609.02366 [physics.ins-det].
- [72] Tilman PLEHN. “Lectures on LHC Physics”. In : *Lecture Notes in Physics* (2012). ISSN : 1616-6361. DOI : 10.1007/978-3-642-24040-9. URL : <http://dx.doi.org/10.1007/978-3-642-24040-9>.

## Bibliographie

- [73] D CONTARDO et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Rapp. tech. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02. Upgrade Project Leader Deputies : Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts : Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch. Geneva, 2015. URL : <http://cds.cern.ch/record/2020886>.
- [74] HL-LHC PROJECT. *Project Schedule*. <https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>.
- [75] *The Phase-2 Upgrade of the CMS Tracker*. Rapp. tech. CERN-LHCC-2017-009. CMS-TDR-014. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2272264>.
- [76] *Phase 2 Outer Tracker*. [https://cms.desy.de/activities/detector\\_upgrade/phase\\_2\\_outer\\_tracker/](https://cms.desy.de/activities/detector_upgrade/phase_2_outer_tracker/).
- [77] *The Phase-2 Upgrade of the CMS Barrel Calorimeters*. Rapp. tech. CERN-LHCC-2017-011. CMS-TDR-015. This is the final version, approved by the LHCC. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2283187>.
- [78] *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Rapp. tech. CERN-LHCC-2017-023. CMS-TDR-019. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2293646>.
- [79] *The Phase-2 Upgrade of the CMS Muon Detectors*. Rapp. tech. CERN-LHCC-2017-012. CMS-TDR-016. This is the final version, approved by the LHCC. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2283189>.
- [80] *Technical Proposal for a MIP Timing Detector in the CMS experiment Phase 2 upgrade*. Rapp. tech. CERN-LHCC-2017-027. LHCC-P-009. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2296612>.
- [81] S. AGOSTINELLI et al. “Geant4—a simulation toolkit”. In : *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), p. 250 -303. ISSN : 0168-9002. DOI : [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL : <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [82] *The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report*. Rapp. tech. CERN-LHCC-2017-013. CMS-TDR-017. This is the CMS Interim TDR devoted to the upgrade of the CMS L1 trigger in view of the HL-LHC running, as approved by the LHCC. Geneva : CERN, 2017. URL : <https://cds.cern.ch/record/2283192>.
- [83] A. M. Sirunyan et al. [CMS COLLABORATION]. “Particle-flow reconstruction and global event description with the CMS detector”. In : *Journal of Instrumentation* 12.10 (2017), p10003. DOI : 10.1088/1748-0221/12/10/p10003. URL : <https://doi.org/10.1088/1748-0221/12/10/p10003>.

## Bibliographie

- [84] R. FRUHWIRTH. “Application of Kalman filtering to track and vertex fitting”. In : *Nucl. Instrum. Meth. A* 262 (1987), p. 444-450. DOI : [10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [85] Pierre BILLOIR. “Progressive track recognition with a Kalman like fitting procedure”. In : *Comput. Phys. Commun.* 57 (1989), p. 390-394. DOI : [10.1016/0010-4655\(89\)90249-X](https://doi.org/10.1016/0010-4655(89)90249-X).
- [86] Pierre BILLOIR et S. QIAN. “Simultaneous pattern recognition and track fitting by the Kalman filtering method”. In : *Nucl. Instrum. Meth. A* 294 (1990), p. 219-228. DOI : [10.1016/0168-9002\(90\)91835-Y](https://doi.org/10.1016/0168-9002(90)91835-Y).
- [87] CMS COLLABORATION. *Iterative Tracking*. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideIterativeTracking>.
- [88] “HLT dimuon invariant mass distributions in 2017 and 2018”. In : (2018). URL : <http://cds.cern.ch/record/2644377>.
- [89] CMS COLLABORATION. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s}=8$  TeV”. In : *Journal of Instrumentation* 10 (fév. 2015). DOI : [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [90] Vardan KHACHATRYAN et al. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”. In : *JINST* 10.06 (2015), P06005. DOI : [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005). arXiv : [1502.02701](https://arxiv.org/abs/1502.02701) [[physics.ins-det](https://arxiv.org/abs/1502.02701)].
- [91] Matteo CACCIARI, Gavin P. SALAM et Gregory SOYEZ. “The anti- $k_t$  jet clustering algorithm”. In : *JHEP* 04 (2008), p. 063. DOI : [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv : [0802.1189](https://arxiv.org/abs/0802.1189) [[hep-ph](https://arxiv.org/abs/0802.1189)].
- [92] “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018”. In : (2020). URL : <https://cds.cern.ch/record/2715872>.
- [93] “Review of Particle Physics : Particle Data Group”. In : *Physical Review D* 98 (août 2018).
- [94] “Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu\tau$  at CMS”. In : *Journal of Instrumentation* 11 (jan. 2016). DOI : [10.1088/1748-0221/11/01/P01019](https://doi.org/10.1088/1748-0221/11/01/P01019).
- [95] *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2*. Rapp. tech. CMS-PAS-TAU-16-002. Geneva : CERN, 2016. URL : <http://cds.cern.ch/record/2196972>.
- [96] Mohammad Hassan HASSANSHAHI. Thèse de doct. Imperial College, sous presse.
- [97] “Identification of hadronic tau decay channels using multivariate analysis (MVA decay mode)”. In : (2020). URL : <https://cds.cern.ch/record/2727092>.



## Bibliographie

- [98] CMS COLLABORATION. “Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV”. In : *JINST* 13.arXiv :1809.02816. CMS-TAU-16-003-003 (2018), P10005. 71 p. DOI : 10 . 1088/1748-0221/13/10/P10005. URL : <http://cds.cern.ch/record/2637646>.
- [99] “Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons”. In : (2019). URL : <https://cds.cern.ch/record/2694158>.
- [100] *CMS-HGCAL/EgammaTools github*. <https://github.com/CMS-HGCAL/EgammaTools/blob/master/VARIABLES.md>.
- [101] Albert M SIRUNYAN et al. “Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector”. In : *JINST* 14.07 (2019), P07004. DOI : 10.1088/1748-0221/14/07/P07004. arXiv : 1903.06078 [hep-ex].
- [102] CMS COLLABORATION. *Pileup mitigation at CMS in 13 TeV data*. 2020. arXiv : 2003.00503 [hep-ex].
- [103] Oliver POOTH. *The CMS Silicon Strip Tracker : Concept, Production and Commissioning*. Springer, 2010. ISBN : 9783834810038.
- [104] J. KEMMER. “Fabrication of low noise silicon radiation detectors by the planar process”. In : *Nucl. Instrum. Meth.* 169 (1980), p. 499-502. DOI : 10.1016/0029-554X(80)90948-9.
- [105] J. KEMMER. “Improvement of Detector Fabrication by the planar Process”. In : *Nucl. Instrum. Meth.* A226 (1984).
- [106] Paul Leroux JEFFREY PRINZIE Michiel Steyaert. *Radiation Hardened CMOS Integrated Circuits for Time-Based Signal Processing*. Springer, 2018. ISBN : 9783319786155.
- [107] High Energy Physics Department of FIRENZE. [http://hep.fi.infn.it/CMS/sensors/Silicon\\_Detector.gif](http://hep.fi.infn.it/CMS/sensors/Silicon_Detector.gif).
- [108] W. ADAM et al. “Performance studies of the CMS Strip Tracker before installation”. In : (jan. 2009). DOI : 10.1088/1748-0221/4/06/P06009.
- [109] CMS COLLABORATION. *Strips Offline Plots 2016*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StripsOfflinePlots2016>.
- [110] M. HUHTINEN. *Highly ionising events in silicon detectors*. CMS-NOTE-2002-011, CERN-CMS-NOTE-2002-011. 2002.
- [111] Ian R TOMALIN. “Test beam analysis of the effect of highly ionising particles on the silicon strip tracker”. In : (sept. 2003). URL : <http://hal.in2p3.fr/in2p3-00020330>.

## Bibliographie

- [112] Marketa JANSOVA. “Recherche du partenaire supersymétrique du quark top et mesure des propriétés des dépôts dans le trajectographe à pistes de silicium de l’expérience CMS au Run 2”. Thèse de doct. Université de Strasbourg, 2018. URL : <http://www.theses.fr/2018STRAE018>.
- [113] Robert John BAINBRIDGE. “Influence of highly ionising events on the CMS APV25 readout chip”. Thèse de doct. Imperial Coll., London, 2004.
- [114] CMS COLLABORATION. *Strips Offline Plots 2018*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StripsOfflinePlots2018>.
- [115] The CMS COLLABORATION. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In : *Journal of Instrumentation* 9.10 (2014), P10009–P10009. ISSN : 1748-0221. DOI : [10.1088/1748-0221/9/10/p10009](https://doi.org/10.1088/1748-0221/9/10/p10009). URL : <http://dx.doi.org/10.1088/1748-0221/9/10/P10009>.
- [116] “Silicon Strip Tracker Performance results 2018”. In : (2018). URL : <https://cds.cern.ch/record/2638062>.
- [117] CMS COLLABORATION. *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. Rapp. tech. CMS-PAS-HIG-20-006. Geneva : CERN, 2020. URL : <http://cds.cern.ch/record/2725571>.
- [118] Stefan BERGE, Werner BERNREUTHER et Sebastian KIRCHNER. “Prospects of constraining the Higgs CP nature in the tau decay channel at the LHC”. In : *Physical Review D* 92.9 (2015). ISSN : 1550-2368. DOI : [10.1103/physrevd.92.096012](https://doi.org/10.1103/physrevd.92.096012). URL : <http://dx.doi.org/10.1103/PhysRevD.92.096012>.
- [119] Malgorzata WOREK. *Higgs CP from H/A tau tau decay*. 2003. arXiv : [hep-ph/0305082](https://arxiv.org/abs/hep-ph/0305082) [hep-ph].
- [120] V. CHEREPANOV, E. RICHTER-WAS et Z. WAS. “Monte Carlo, fitting and Machine Learning for Tau leptons”. In : *15th International Workshop on Tau Lepton Physics*. T. 1. Amsterdam, Netherlands, sept. 2018, p. 018. DOI : [10.21468/SciPostPhysProc.1.018](https://doi.org/10.21468/SciPostPhysProc.1.018). URL : <https://hal.archives-ouvertes.fr/hal-01937914>.
- [121] A. STAHL et al. *Physics with Tau Leptons*. Springer Tracts in Modern Physics. U.S. Government Printing Office, 2000. ISBN : 9783540662679.
- [122] Stanisław JADACH, Johann H. KÜHN et Zbigniew WAS. “TAUOLA - a library of Monte Carlo programs to simulate decays of polarized  $\tau$  leptons”. In : *Computer Physics Communications* 64.2 (1991), p. 275 -299. ISSN : 0010-4655. DOI : [https://doi.org/10.1016/0010-4655\(91\)90038-M](https://doi.org/10.1016/0010-4655(91)90038-M). URL : <http://www.sciencedirect.com/science/article/pii/001046559190038M>.

## Bibliographie

- [123] M. JE.ABEK et al. “The  $\tau$  decay library TAUOLA, update with exact  $O(\alpha)$  QED corrections in  $\tau \rightarrow \mu(e) \nu\nu$  decay modes”. In : *Computer Physics Communications* 70.1 (1992), p. 69 -76. ISSN : 0010-4655. DOI : [https://doi.org/10.1016/0010-4655\(92\)90092-D](https://doi.org/10.1016/0010-4655(92)90092-D). URL : <http://www.sciencedirect.com/science/article/pii/001046559290092D>.
- [124] S. JADACH et al. “The tau decay library TAUOLA, version 2.4”. In : *Computer Physics Communications* 76.3 (1993), p. 361 -380. ISSN : 0010-4655. DOI : [https://doi.org/10.1016/0010-4655\(93\)90061-G](https://doi.org/10.1016/0010-4655(93)90061-G). URL : <http://www.sciencedirect.com/science/article/pii/001046559390061G>.
- [125] D. M. ASNER et al. “Hadronic structure in the decay  $\tau^- \rightarrow \nu_\tau \pi^- \pi^0 \pi^0$  and the sign of the tau neutrino helicity”. In : *Phys. Rev. D* 61 (1 1999), p. 012002. DOI : [10.1103/PhysRevD.61.012002](https://doi.org/10.1103/PhysRevD.61.012002). URL : <https://link.aps.org/doi/10.1103/PhysRevD.61.012002>.
- [126] Vladimir CHEREPANOV et Alexander ZOTZ. *Kinematic reconstruction of  $Z/H \rightarrow tt$  decay in proton-proton collisions*. 2018. arXiv : 1805.06988 [hep-ph].
- [127] Clément GRIMAUULT. Mém. de mast. Institut Pluridisciplinaire Hubert Curien, 2018.
- [128] Lorenzo BIANCHINI et al. “Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of  $\tau$  leptons using matrix element techniques”. In : *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment* 862 (2017), 54–84. ISSN : 0168-9002. DOI : [10.1016/j.nima.2017.05.001](https://doi.org/10.1016/j.nima.2017.05.001). URL : <http://dx.doi.org/10.1016/j.nima.2017.05.001>.
- [129] T. PRZEDZINSKI, E. RICHTER-WAS et Z. WAS. “Documentation of TauSpinner algorithms : program for simulating spin effects in tau-lepton production at LHC”. In : *The European Physical Journal C* 79.2 (2019). ISSN : 1434-6052. DOI : [10.1140/epjc/s10052-018-6527-0](https://doi.org/10.1140/epjc/s10052-018-6527-0). URL : <http://dx.doi.org/10.1140/epjc/s10052-018-6527-0>.
- [130] A.M. SIRUNYAN et al. “An embedding technique to determine  $\tau\tau$  backgrounds in proton-proton collision data”. In : *Journal of Instrumentation* 14.06 (2019), P06032–P06032. ISSN : 1748-0221. DOI : [10.1088/1748-0221/14/06/p06032](https://doi.org/10.1088/1748-0221/14/06/p06032). URL : <http://dx.doi.org/10.1088/1748-0221/14/06/p06032>.
- [131] R FRÜHWIRTH, Wolfgang WALTENBERGER et Pascal VANLAER. *Adaptive Vertex Fitting*. Rapp. tech. CMS-NOTE-2007-008. Geneva : CERN, 2007. URL : <https://cds.cern.ch/record/1027031>.
- [132] The CMS COLLABORATION. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In : *Journal of Instrumentation* 9.10 (2014), P10009–P10009. ISSN : 1748-0221. DOI : [10.1088/1748-0221/9/10/p10009](https://doi.org/10.1088/1748-0221/9/10/p10009). URL : <http://dx.doi.org/10.1088/1748-0221/9/10/p10009>.

## Bibliographie

- [133] MICHAL BLUJ, GUILLAUME BOURGATTE, ANDREA CARDINI, DAVID COLLING, ALBERT DOW ET AL. “Analysis of the Higgs CP state in  $\tau$  decays”. Analysis Note. 2020.
  - [134] Konstantin Androsov HALE SERT. *Tau trigger scale factors for DeepTau for Run 2 (update)*. URL : [https://indico.cern.ch/event/868279/contributions/3668554/attachments/1959280/3266317/2019-12-17\\_Tau\\_trigger\\_SF\\_for\\_DeepTau.pdf](https://indico.cern.ch/event/868279/contributions/3668554/attachments/1959280/3266317/2019-12-17_Tau_trigger_SF_for_DeepTau.pdf).
  - [135] ANDREW LOELIGER, CECILE CAILLOL, ABHI MALLAMPALLI, JITHIN MADHUSUDANAN, SRIDHARA DASU, TULIKA BOSE, DOYEONG KIM, TYLER MITCHELL, YURII MARAVIN, ABDOLLAH MOHAMMADI, KETI KAADZE, SENKA DURIC, ISOBEL OJALVO, SAM HIGGINBOTHAM, AND DAN MARLOW. “Measurement of Higgs boson production and decay to a pair of tau leptons on the full Run 2 data set using a cut-based approach”. In : *CMS Note 2019/019* (2019).
  - [136] CMS COLLABORATION. *Luminosity Physics Object Group (Lumi POG)*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM>.
  - [137] D. de FLORIAN et al. “Handbook of LHC Higgs Cross Sections : 4. Deciphering the Nature of the Higgs Sector”. In : (2016). DOI : [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). arXiv : [1610.07922](https://arxiv.org/abs/1610.07922) [hep-ph].
  - [138] CMS COLLABORATION. *MET Corrections and Uncertainties for Run-II*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections>.
  - [139] S. MRENNNA et P. SKANDS. “Automated Parton-Shower Variations in Pythia 8”. In : *Phys. Rev. D* 94.7 (2016), p. 074005. DOI : [10.1103/PhysRevD.94.074005](https://doi.org/10.1103/PhysRevD.94.074005). arXiv : [1605.08352](https://arxiv.org/abs/1605.08352) [hep-ph].
  - [140] Roger BARLOW et Christine BEESTON. “Fitting using finite Monte Carlo samples”. In : *Computer Physics Communications* 77 (1993), p. 219-228.
  - [141] J. S. CONWAY. “Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra”. To be published in a CERN Yellow Report. 2011.
  - [142] CMS COLLABORATION. URL : <https://cms-analysis.github.io/CombineHarvester/>.
  - [143] R. COUSINS. “Generalization of Chisquare Goodness-ofFit Test for Binned Data Using Saturated Models , with Application to Histograms”. In : 2013.
  - [144] *Measurement of Higgs boson production and decay to the  $\tau\tau$  final state*. Rapp. tech. CMS-PAS-HIG-18-032. Geneva : CERN, 2019. URL : <http://cds.cern.ch/record/2668685>.
-

# Étude des propriétés CP du boson de Higgs dans le canal tau tau dans l'expérience CMS auprès du LHC

## RÉSUMÉ

Cette thèse a pour sujet la recherche de violation CP dans les désintégrations tau tau du boson de Higgs ( $H \rightarrow \tau\tau \rightarrow a_1\nu a_1\nu \rightarrow 3\pi\nu 3\pi\nu$ ) dans l'expérience CMS, auprès de l'accélérateur du LHC. L'analyse a porté sur un échantillon de données correspondant à l'intégralité des données du Run 2 du LHC, avec une luminosité intégrée de  $137 \text{ fb}^{-1}$ . La méthode dite du vecteur polarimétrique a été employée pour définir l'observable caractérisant l'état CP du boson de Higgs.

Cette analyse s'accompagne de deux autres études. La première est l'amélioration des performances d'identification des leptons taus pour la phase de haute luminosité du LHC (HL-LHC), pour le rejet des électrons, par l'emploi d'arbres de décision boostés. Les électrons émettant des photons par radiation continue de freinage, ils constituent un bruit de fond conséquent à l'identification des leptons taus dans leurs modes de désintégrations hadroniques.

La seconde étude est celle de l'efficacité de reconstruction des impacts dans le trajectographe à pistes en silicium. L'origine des inefficacités observées a été évaluée dans le contexte de la présence de particules hautement ionisantes.

MOT-CLÉS : physique des particules, LHC, CMS, trajectographe, particules hautement ionisantes, lepton tau, boson de Higgs, violation CP, vecteur polarimétrique.

## ABSTRACT

The subject of this thesis is the search for a CP violation in tau tau decays of the Higgs boson ( $H \rightarrow \tau\tau \rightarrow a_1\nu a_1\nu \rightarrow 3\pi\nu 3\pi\nu$ ) in the CMS experiment at LHC. The analysis focused on a data sample corresponding to the entire LHC Run 2 data, with an integrated luminosity of  $137 \text{ fb}^{-1}$ . The polarimetric vector method was used to define the observable characterizing the CP state of the Higgs boson.

This analysis is accompanied by two other studies. The first one is the improvement of the identification of taus leptons for the high luminosity phase of the LHC (HL-LHC), for the rejection of electrons, by the use of boosted decision trees. The electrons emitting photons by bremsstrahlung radiation, they constitute a consequent background to the identification of taus leptons in their hadronic decay modes.

The second study is the hit efficiency in the silicon tracker. The origin of the observed inefficiencies was evaluated in the context of the presence of highly ionizing particles.

KEYWORDS : particle physics, LHC, CMS, tracker, highly ionizing particles, tau lepton, Higgs boson, CP violation, polarimetric vector.