



# CERTIFICATO DI FIRMA DIGITALE

Si certifica che questo documento informatico

**phd\_unisi\_076630.pdf**

composto da n239pagine

È stato firmato digitalmente in data odierna con Firma Elettronica Qualificata (FEQ), avente l'efficacia e gli effetti giuridici equivalenti a quelli di una firma autografa, ai sensi dell'art. 2702 del Codice Civile e dell'art. 25 del Regolamento UE n. 910/2014 eIDAS (electronic IDentification Authentication and Signature).

## PROCESSI INFORMATICI COMPLETATI

- **Apposizione di Firma Elettronica Qualificata Remota** emessa da Intesi Group S.p.A. in qualità di prestatore di servizi fiduciari qualificati autorizzato da AgID, per garantire con certezza l'autenticità, l'integrità, il non ripudio e l'immodificabilità del documento informatico e la sua riconducibilità in maniera manifesta e inequivoca all'autore, ai sensi dell'art. 20 comma 2 del CAD - D.lgs 82/2005.
- **Apposizione di Marca Temporale Qualificata** emessa da Intesi Group S.p.A. in qualità di prestatore di servizi fiduciari qualificati autorizzato da AgID, per attribuire una data e un orario opponibile a terzi, ai sensi dell'art. 20 comma 3 del CAD - D.lgs 82/2005 e per far sì che la Firma Elettronica Qualificata apposta su questo documento informatico, risulti comunque valida per i prossimi 20 anni a partire dalla data odierna, anche nel caso in cui il relativo certificato risultasse scaduto, sospeso o revocato.
- **Apposizione di Contrassegno Elettronico**, l'unica soluzione tecnologica che permette di prorogare la validità giuridica di un documento informatico sottoscritto con firma digitale e/o marcato temporalmente, rendendolo inalterabile, certo e non falsificabile, una volta stampato su supporto cartaceo, ai sensi dell'art. 23 del CAD - D.lgs 82/2005.



Per risalire all'originale informatico è necessario scansionare il Contrassegno Elettronico, utilizzando l'applicazione HONOS, disponibile per dispositivi Android e iOS.



# UNIVERSITÀ DI SIENA

1240

Department of Physical Sciences, Earth and Environment

PhD in Experimental Physics

XXXIII Cycle

Coordinator: Prof. Riccardo Paoletti

## **A new neural network to select the b-jet pair from $H \rightarrow b\bar{b}$ in $HH \rightarrow b\bar{b}\tau^+\tau^-$ searches at CMS**

Disciplinary Scientific Sector: FIS/01

PhD Student:

Maria Rosaria Di Domenico

Supervisor:

Prof.ssa Maria Agnese Ciocci

---

XXXIII CICLO

*A mi familia: La mejor mamá del mundo, Giacomo y Biscotto.*

## ABSTRACT

This thesis presents the search for non-resonant double Higgs boson (HH) production via gluon-gluon fusion (GGF) and vector boson fusion (VBF) mechanisms. The HH production gives access to two fundamental parameters of the SM: the Higgs boson trilinear self-coupling ( $\lambda_{\text{HHH}}$ ), and the coupling between two Higgs bosons and two vector bosons ( $\lambda_{2V}$ ), via the measurement of the GGF and VBF cross sections. These parameters are sensitive to the presence of physics beyond the Standard Model. The investigated final state has one of the Higgs bosons decaying into two b-quarks and the other decaying into two  $\tau$  leptons ( $\text{HH} \rightarrow \text{b}\bar{\text{b}}\tau^+\tau^-$ ). This process is studied through the examination of the three decay modes of the  $\tau^+\tau^-$  system:  $\tau_h\tau_h$ , where  $\tau_h$  means a  $\tau$  which decays into hadrons plus a  $\nu_\tau$ , and  $\tau_\ell\tau_\ell$  (with  $\ell$  denoting either an electron or a muon). The search uses data from proton-proton collisions at a center-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$  recorded with the Compact Muon Solenoid (CMS) detector at the LHC, corresponding to an integrated luminosity of  $137.1 \text{ fb}^{-1}$ .

The new techniques employed in the analysis allowed for a major improvement of the signal sensitivity compared to the previously published CMS results. Among these new techniques, a novel neural network, HH-Btag, is used to improve the selection of the b-jet pair forming the  $\text{HH} \rightarrow \text{b}\bar{\text{b}}$  with respect to the previous methods. HH-Btag has been designed and optimized for the searches of resonant and non-resonant production of the HH in the aforementioned final state. The performance of the algorithm with respect to their predecessors are evaluated under different hypotheses of mass and of spin in the case of resonant production, and as a function of couplings modifiers in the case of non-resonant production, in order to cover a large phase space of the possible scenario.

The internal CMS collaboration review of the analysis started in April 2021 and is still ongoing, thus the statistical analysis of the results is carried out without the inclusion of the observed data (blind analysis), and only expected results from standard model (SM) predictions are reported. The 95% Confidence Level upper limit on the total (GGF + VBF) cross section times branching fraction expected from SM is  $10.88 \text{ fb}$  (equivalent to 4.55 times the SM value). The study of the VBF process, done for the first time in this analysis, resulted in an exclusion limit on the production cross section expected via VBF of  $237.3 \text{ fb}$  (137.5 times the SM value).



# Contents

Introduction . . . . .	9
<b>1 Theoretical context of the double Higgs boson production</b>	<b>11</b>
1.1 The Standard Model of particle physics . . . . .	12
1.1.1 Fields and particle content . . . . .	12
1.1.2 Electroweak symmetry breaking and the Higgs Mechanism . . . . .	20
1.1.3 The Higgs mechanism . . . . .	21
1.2 Higgs boson phenomenology and experimental status . . . . .	27
1.3 Higgs boson pair production . . . . .	32
1.3.1 HH production mechanisms at LHC . . . . .	32
1.3.2 Double Higgs production beyond the SM . . . . .	35
1.4 Current searches of HH production at the LHC . . . . .	43
1.4.1 HH decay channels . . . . .	44
1.4.2 Status of experimental HH searches . . . . .	46
<b>2 The CMS experiment at LHC</b>	<b>51</b>
2.1 The Large Hadron Collider . . . . .	51
2.2 The Compact Muon Solenoid experiment at LHC . . . . .	57
2.2.1 The CMS coordinate system and useful variables . . . . .	57
2.2.2 Detector . . . . .	59
2.2.3 The trigger system . . . . .	73
<b>3 Event reconstruction and physical objects</b>	<b>77</b>
3.1 High level reconstruction objects . . . . .	77
3.1.1 The track reconstruction . . . . .	77
3.1.2 Primary vertex . . . . .	82
3.1.3 The particle-flow algorithm . . . . .	83
3.2 Physics object reconstruction . . . . .	87
3.2.1 Muon and electron isolation . . . . .	87
3.2.2 Muons . . . . .	88
3.2.3 Electrons . . . . .	90

---

3.2.4	Taus . . . . .	94
3.2.5	Jets . . . . .	99
3.2.6	Missing transverse energy . . . . .	104
<b>4</b>	<b>The <math>HH \rightarrow bb\tau\tau</math> analysis</b>	<b>105</b>
4.1	The $HH \rightarrow bb\tau\tau$ decay channel . . . . .	105
4.1.1	Signal definition . . . . .	105
4.1.2	Background definition . . . . .	106
4.2	Data samples . . . . .	108
4.3	Trigger selections . . . . .	109
4.4	Analysis flow . . . . .	112
4.4.1	$H \rightarrow \tau\tau$ selection . . . . .	113
4.4.2	$H \rightarrow bb$ and VBF-jets selection . . . . .	119
4.4.3	$HH$ signal region . . . . .	122
4.4.4	Event categorization and signal extraction . . . . .	124
4.5	Sample composition modeling . . . . .	131
4.5.1	Signal modeling . . . . .	131
4.5.2	Background modeling . . . . .	136
4.6	Systematic uncertainties . . . . .	142
4.6.1	Normalization uncertainties . . . . .	142
4.6.2	Shape uncertainties . . . . .	143
4.7	Statistical interpretation . . . . .	145
4.8	Results . . . . .	147
4.8.1	Inclusive GGF+VBF $HH$ production . . . . .	147
4.8.2	VBF $HH$ production . . . . .	153
<b>5</b>	<b>Introduction to neural networks</b>	<b>158</b>
5.1	Neural networks . . . . .	158
5.1.1	Weights . . . . .	159
5.1.2	Loss function and optimizers . . . . .	160
5.2	Recurrent neutral networks . . . . .	163
<b>6</b>	<b><math>HH</math>-BTag algorithm to select b-jets in <math>HH \rightarrow bb\tau\tau</math> searches</b>	<b>166</b>
6.1	Motivation for a new algorithm . . . . .	166
6.2	Signal samples for $HH$ production . . . . .	168
6.2.1	Signal samples for non-resonant SM and BSM $HH$ production . . . . .	168
6.2.2	Signal samples for resonant $HH$ production (BSM . . . . .	168
6.3	Event Selection . . . . .	178
6.4	Ground Truth definition . . . . .	178

6.5	Features and inputs . . . . .	181
6.6	HH-BTag algorithm . . . . .	183
6.6.1	Optimization of the NN hyperparameters . . . . .	185
6.6.2	Theory behind Bayesian optimization . . . . .	186
6.6.3	Guided Optimization . . . . .	189
6.6.4	Final architecture . . . . .	192
6.6.5	Training . . . . .	192
6.7	HH-BTag performance . . . . .	195
6.7.1	Gluon fusion non-resonant . . . . .	197
6.7.2	VBF non-resonant . . . . .	200
6.7.3	Gluon fusion resonant . . . . .	203
6.7.4	VBF resonant . . . . .	209
6.7.5	Comparison with DeepFlavour algorithm . . . . .	213
6.8	HH-BTag Results . . . . .	215
	Conclusions . . . . .	216
<b>A</b>	<b>MC Corrections</b>	<b>217</b>
A.1	Pile-up reweighting . . . . .	217
A.2	Pile-up jet identification scale factor . . . . .	217
A.3	Jet energy smearing . . . . .	217
A.4	L1 ECAL prefiring weight . . . . .	218
A.5	Trigger scale factors . . . . .	218
A.6	Trigger prescale weight . . . . .	219
A.7	b-tag scale factors . . . . .	219
A.8	DeepTau scale factors for hadronic taus . . . . .	220
A.9	Electron and muon scale factors . . . . .	220
A.10	Tau Energy Scale corrections . . . . .	221
A.11	Custom SF for the fully hadronic channel in 2017 . . . . .	221



# Introduction

The discovery of the Higgs boson (H) with a mass around 125 GeV<sup>1</sup>, in 2012 by the ATLAS and CMS collaborations [1, 2] was a first step in the exploration of the electroweak symmetry breaking (EWSB). So far, the particle found agrees with properties of the Higgs boson predicted by the SM (like the spin, charge and mass). However, now the efforts are focused on corroborate all the still-missing properties of the Higgs boson, to understand if its properties correspond exactly to ones predicted by the SM. After EWSB, the Higgs boson potential gives rise to cubic and quartic terms in the Higgs boson field, introducing two self-couplings. Their measurement will provide an independent test of the SM, allowing a direct measurements of the scalar sector properties.

In particular, the precise measurement of the trilinear self-coupling of the Higgs boson ( $\lambda_{\text{HHH}}$ ) represents a crucial test of the validity of the SM predictions, since in the SM the value of the trilinear self-coupling is defined as  $\lambda_{SM} = m_{\text{H}}^2/2v$ , and both the Higgs mass ( $m_{\text{H}}$ ) and the vacuum-expectation value ( $v$ ) are measured. Any deviation from the prediction would imply the existence of physics beyond the Standard Model (BSM).

The double Higgs (HH) production gives direct access to  $\lambda_{\text{HHH}}$ , since its value can be extracted from the measurement of the HH production cross section. Due to the direct relation of the HH production to the scalar potential, this process is very sensitive to the presence of BSM contributions, that could manifest either directly as new states decaying to a HH pair (resonant production), or as contributions in the quantum loops that would modify its cross section and kinematic properties (non-resonant production).

At the LHC, the HH production is extremely rare. The cross section of this process and the statistics already collected by ATLAS and CMS experiments, during Run II and in the near future in Run III, makes the SM HH process undetectable on these data: will be necessary to wait the High Lumi (HL) phase of LHC to measure the trilinear self-coupling predicted by the SM, if no new physics manifest itself at the LHC energy scale. During Run III (and analysis of legacy Run II data) it is still possible to explore other parts of the phenomenological parameter space that predict enhanced non-resonant double Higgs production, as well as search for the resonances, predicted by different physics models

---

<sup>1</sup>In this thesis the system of natural units is used with  $\hbar = c = 1$ , where  $\hbar = h/2\pi = 6.582\,119\,569(26) \times 10^{-22}$  MeV s

---

(like MSSM, warped extra dimensions) which decay in pair of Higgs bosons. The ATLAS and CMS collaborations reported their results for several decay modes of the Higgs pair production in the Ref. [3, 4].

The search of HH production explored in this thesis consists in the investigation of the non-resonant gluon-gluon fusion (GGF) and vector boson fusion (VBF) production mechanisms, using proton-proton collisions at  $\sqrt{s} = 13$  TeV. The final state where one Higgs boson decays to  $b\bar{b}$  and the other decays to  $\tau^+\tau^-$  ( $HH \rightarrow b\bar{b}\tau\tau$ <sup>2</sup>), characterized by a decay branching fraction of 7.3% for a Higgs boson mass of 125 GeV, is studied. The branching fraction combined with the high purity of the  $\tau\tau$  pair selection, makes this final state one of the most sensitive to Higgs boson pair production. The three decay modes of the  $\tau\tau$  system are considered:  $\tau_\mu\tau_h$ ,  $\tau_e\tau_h$  and  $\tau_h\tau_h$ , accounting for 87.6% of the full  $\tau\tau$  decay modes. This analysis is part of the so-called Run II legacy analysis, thus it uses the total LHC Run II integrated luminosity, amounting to  $137.1 \text{ fb}^{-1}$  collected by the CMS experiment during the 2016, 2017 and 2018 data taking periods.

Given the rarity of the process, the optimization of the signal selection is essential. Because of that, in this work is presented a new neural network (NN) to identify the reconstructed jets of the  $H \rightarrow b\bar{b}$  decay candidates, for the  $HH \rightarrow b\bar{b}\tau\tau$  final state, denoted HH-BTag. This is my major personal contribution to the analysis. The algorithm is designed and optimized to improve the selection of the pair of b-jets forming the  $H \rightarrow b\bar{b}$  in HH, for the resonant and non-resonant searches in the aforementioned final states. The performance with respect to the previous methods is evaluated using simulated samples, under different hypotheses of mass and of spin of the resonance decaying in HH, in the case of resonant production, and as a function of couplings modifiers in the case of non-resonant production, in order to cover a large phase space of the possible scenarios, and where both mechanism of production GGF and VBF are considered.

This thesis is structured as follows: The Standard Model theory, with a focus on the Higgs mechanism, providing a theoretical context of the HH production, along with the experimental status of the HH production, is presented in Chapter 1. In Chapter 2 is described the experimental apparatus, providing a description of the LHC machine, and the most important features of the CMS detector. The reconstruction and identification algorithms used to reconstruct the final states of the analysis will be described in the Chapter 3. In the Chapter 4 the analysis and their results are illustrated. Chapter 5 provides an introduction to the machine learning techniques, whose basic concepts will be useful for the Chapter 6, where the design, optimization and the performance of the HH-BTag algorithm is presented.

---

<sup>2</sup>For simplicity throughout this thesis it is not specified the charge of the leptons and the particle-antiparticle nature of the quarks:  $\tau\tau$  corresponds to  $\tau^+\tau^-$ , for leptons  $\ell\ell$  is equal to  $\ell^+\ell^-$ ,  $b\bar{b}$  stays for  $b\bar{b}$ , and in general  $qq$  stays for  $q\bar{q}$

---

# Chapter 1

## Theoretical context of the double Higgs boson production

The Standard Model of particle physics (SM) is a renormalizable quantum field theory that describes all the known fundamental particles and how they interact at subnuclear scale. It describes the elementary particles that compose matter and three of the four fundamental interactions: the strong, weak, and electromagnetic forces. The SM is so far the best rigorous theory that provides a unified description of particle physics, well corroborated in the last decades with the experimental discoveries of all the SM parameters, finalized with the observation of the missing piece of the puzzle, the Higgs boson, announced in 2012 by the ATLAS and CMS experiments at the LHC [1, 2].

Despite the high level of agreement with the SM predictions, due to the latest experimental results, the SM is still in contrast with some observations ranging from subnuclear to astrophysical scale and others theoretical considerations, leaving room for the presence of physics beyond the Standard Model (BSM). Extensions of the SM scalar sector can provide intriguing scenarios for new physics, and the Higgs boson discovery opens new ways to its exploration. The study of the double Higgs production is strongly related to the nature of the scalar sector, and can give precious information in this context, since the measurement of the cross section of the process determines the Higgs boson self-interaction, and provides a path to search for new physics BSM.

This chapter describes the theoretical framework and the importance of the study of the double Higgs production, the Section 1.1 will introduce the principles of the SM and the spontaneous symmetry breaking. The phenomenology of the Higgs production, and its experimental status is presented in Section 1.2, and in Section 1.3 the phenomenology of the double Higgs boson production is presented, in the context of searches for SM and BSM physics. Finally, a summary of the status of the experimental searches is presented in Section 1.4

---

## 1.1 The Standard Model of particle physics

### 1.1.1 Fields and particle content

The Standard Model [5, 6] describes with high precision all known phenomena in particle physics. A detailed description of SM can be found in References [7, 5].

The SM is a renormalizable quantum field theory that describes the phenomena at the subnuclear scales. It provides a unified description of the strong and electroweak forces, and contains a scalar sector responsible for the Brout-Englert-Higgs (BEH) mechanism that breaks the electroweak symmetry and generates the masses of the known elementary constituents, the W and Z bosons, and a new spin-0 boson. The SM formulation does not include the gravitational interaction, which is negligible at the subnuclear scales. Its mathematical formulation is based on the construction of a local gauge invariant Lagrangian [5] under the non-abelian group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ , where the subscript stands for the conserved charges: the strong charge or color  $C$ , the weak isospin  $T$  (or better, its third component  $T_3$ ) and the hypercharge  $Y$ . These quantities are connected to the electric charge  $Q$  (conserved too) through the Gell-Mann–Nishijima relation:  $Q = Y/2 + T_3$ .

Strong interactions are described by the quantum chromodynamics (QCD), a gauge field theory invariant under the colour symmetry gauge group  $SU(3)_C$ . The  $SU(3)_C$  invariance results in the existence of eight different mediators of the strong interaction, the gluons. The electroweak interactions are described by a gauge field theory invariant under the weak isospin and hypercharge symmetry gauge group  $SU(2)_L \times U(1)_Y$ , unifying weak and electromagnetic interactions, mediated by spin-1 bosons, the  $W^\pm$  and Z bosons and the photon ( $\gamma$ ), respectively.

In the SM, the matter is described by twelve fundamental spin-1/2 fermion fields: six quarks, which do not exist as free states, and six leptons. Fermions are organized in three generations according to their mass, starting from the lightest and most stable particles. Each generation consists of one “up-type” and one “down-type” fermion of electric charge<sup>1</sup>  $+2/3$  and  $-1/3$  and two leptons of electric charge  $-1$  and  $0$ ,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}.$$

In addition three generations with the corresponding anti-particles, with identical mass and opposite quantum numbers exist. All visible stable matter in the Universe is made of first generation fermions. When produced, the particles of the second and third generation, excluding the neutrinos, decay directly or indirectly into the particles of the

---

<sup>1</sup>The electric charge is expressed in terms of the elementary charge,  $e \approx 1.602 \times 10^{-19}C$ .

first generation.

The known universe is made by up- and down-quarks, electrons, and neutrinos. The remaining particles described in the SM are produced in accelerators, or in energetic cosmic rays collisions in the atmosphere, existed in a early stage of the universe, but since their lifetime is very short, they play no known role in the universe today. All the elementary particles in the SM, their spin, charge and approximate mass, are shown in Fig. 1.1 .

As already mentioned, the SM describes three fundamental interactions. The electromagnetic interaction is mediated by the photon and it occurs between all electrically charged particles. The strong interaction is mediated by the gluons and it occurs between all the particles that carry the colour charge, namely the quarks and gluons. Finally, the weak interaction is mediated by the  $W^\pm$  and  $Z$  bosons. All elementary fermions carry the charge of the weak interaction the weak isospin, and therefore are subject to it.

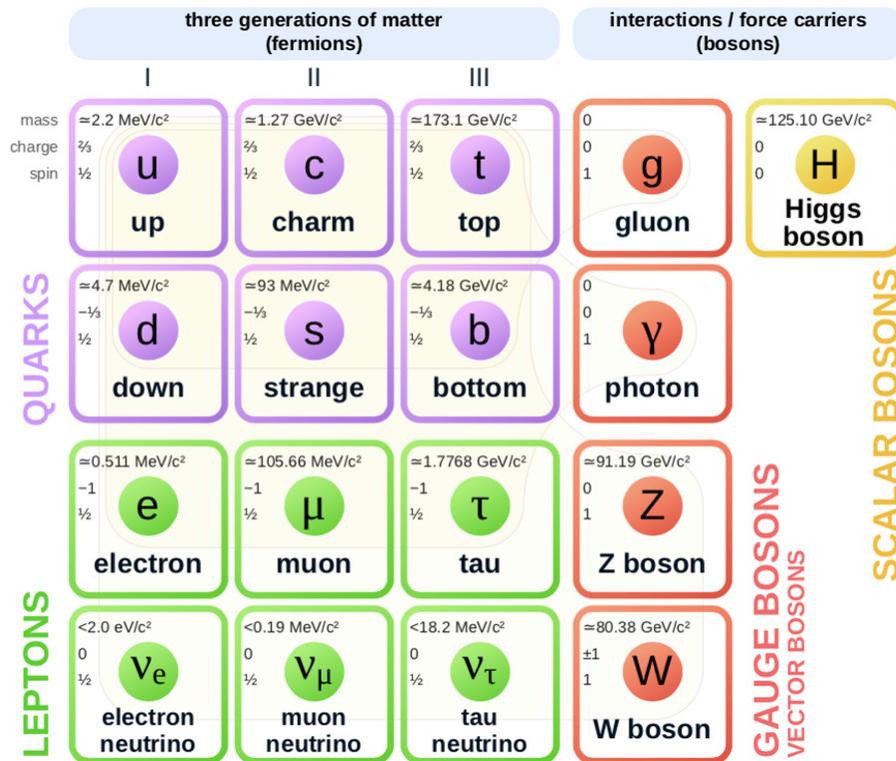


Figure 1.1: Particle content of the Standard Model [8].

## Quarks

The quarks can be of six flavours, are massive, have a fractional electric charge and interact through strong and electroweak interactions. For historical reasons the quarks are called up and down, for first generation, charm and strange, for second generation,

top and bottom for the third one. The up-type quarks of each family have an electric charge  $Q = 2/3$ , while down-quarks have  $Q = -1/3$ . In addition to the other quantum numbers, quarks carry a “color” charge, and three distinct color charge are allowed, “red” (R), “green” (G) and “blue” (B).

The quarks do not exist in free states, which means that they have never been observed in nature, i.e. quarks have to bind into color neutral states called hadrons, in a way that the color charge is confined, this mechanism is known as the color confinement. In more detail, thanks to this mechanism the quarks and anti-quarks are kept permanently inside hadrons, and this phenomenon is the reason why color-charged particles, such as quarks and gluons, can not be directly observed in normal conditions, and thus no isolated colour-charge particles can exist, only colourless singlet particles are observed, in form of baryons (bound states of three quarks) and mesons (bound states of a quark–antiquark pair). At distances larger than  $1/\Lambda$ , with  $\Lambda$  the energy scale, the quarks and gluons do not propagate as free particles, instead at smaller distances the perturbative expansion works, and the theory behaves as any other field theory, thanks to the asymptotic freedom property [9, 10, 11]. Thus, quarks do not exist as free states but can be experimentally observed only as bound states of quarks. As a consequence, the high energy experiments (HEP) can only detect hadrons.

The creation of hadrons from a single quark produced in a HEP experiment is a complex process that takes the name of “hadronization”. The hadronization, or showering, process occurs when highly energetic quark-antiquark pair, gluons, or in a more general way a parton<sup>2</sup>, is produced after an interaction. The potential between the produced partons, due to gluons exchange, tries to keep them together until the strength reaches a breaking point where further quark-antiquark pairs are created and finally bound together with the original parton. Finally, the parton generated during the interaction, can be observed as a collimated “jet” of particles approximately in the direction of the original parton.

The timescales of the hadronization, related to the QCD time scales, are of the order of  $10^{-24}$  s, thus the hard scatter and hadronization processes can be treated separately thanks to a factorization of their effects. The top-quark represents an exception since its lifetime is so short ( $\approx 0.5 \times 10^{-24}$  s) that it decays before bound states can be formed.

The quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, where interactions involving the W boson, (flavor changing-current) allow for mixing of the quark generations. This mixing, which is a generalisation of the Cabibbo hypothesis to all three quark generations, is parameterized in the SM by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which contains the information on the strength of

---

<sup>2</sup>The word parton originates from Feynman who called the constituent of the proton parton, so it refers to valence quarks, to both quarks and gluons.

---

the weak decays that are involved in the change of the quarks flavour.

## Leptons

The leptons are also arranged into three families, summing twelve fermions, counting anti-particles families. These particles have been experimentally observed and are included in the SM. The electron ( $e$ ), muon ( $\mu$ ) and tau ( $\tau$ ) have electric charge  $Q = -1$ , and are massive. Neutrinos are up-type fermions and they are electrically neutral particles. In the SM, neutrinos are massless, although this contradicts experimental evidence of their oscillation [12], which requires non-zero masses.

To each lepton down-type corresponds its own neutrino, denoted as  $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$ . Leptons have no colour charge, and only interact through the electromagnetic and weak forces. The electron is a stable particle with a mass of 551 keV, the muon has a mass of 105.7 MeV, and a lifetime of 2.2  $\mu$ s, which is enough time to pass through all of the subdetectors of the LCH experiments without decaying into more stable particles, so it is considered as a stable particle in the context. The tau lepton has an approximately mass of 1.8 GeV and a lifetime too short ( $2.9 \times 10^{-13}$  s) to be detected at the LHC experiments directly, however it can be observed through its decay products [8].

Since neutrinos have a neutral electric charge they interact with matter only through the weak force, which makes them undetectable at collider experiments. Little is known about their masses, but the observation of their flavour oscillations prove that their masses are not zero ( $m_\nu < 2$  eV) [8]. Massive neutrinos, can be considered in the SM introducing the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix for neutrino mixing, equivalent to the CKM matrix in the quark sector [13].

Type	1st gen.	2st gen.	3st gen.	$I_3$	$Y$	$Q$
Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix}$	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$
	$u_R$	$c_R$	$t_R$	0	4/3	2/3
	$d_R$	$s_R$	$b_R$	0	-2/3	-1/3
Leptons	$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
	$e_R$	$\mu_R$	$\tau_R$	0	-2	-1

Table 1.1: SM fermion fields in their  $SU(2)_L$  representation, with their main quantum numbers: weak isospin  $T_3$  ( $I_3$ ), hypercharge  $Y$  and electric charge  $Q$ . The  $L$  and  $R$  indexes denotes the left and right chiralities, respectively. [8].

## Strong interaction

Within the SM, the theory of the strong interaction between quarks and gluons is known as quantum chromodynamics (QCD). A complete, formal discussion of QCD, and in particular of the concepts of asymptotic freedom and confinement, is outside the scope of this thesis, for which the reader may be referred to standard textbooks [7, 5]. One of its most striking success is the possibility to interpret interactions of hadrons observed at high energy particle colliders, such as the LHC, in terms of elementary processes involving quarks and gluons, collectively indicated as partons.

The quarks in the  $SU(3)_C$  symmetry group have a fundamental representation as triplets of colour, hence the ‘‘C’’ in the subscript denotes that the members of each triplet differ only by their colour charge ( $R, G, B$ ). Leptons are singlets of  $SU(3)_C$ , which means they do not carry a colour charge, thus do not interact via the strong force.

The QCD Lagrangian is constructed starting from the free Lagrangian, for free Dirac massless fermions with spin-1/2 [14]:

$$\mathcal{L}_f = \sum_f \bar{q}_f (i\gamma^\mu \partial_\mu) q_f, \quad \text{where } \gamma^\mu \text{ are the Dirac matrices.} \quad (1.1)$$

The quark field of flavour  $f$  is a colour triplet of the  $SU(3)$  group:

$$q_f = \begin{pmatrix} q_f^R \\ q_f^G \\ q_f^B \end{pmatrix}.$$

In order to conserve the color charge, the density Lagrangian must be invariant under global and local transformations. The fermion field transforms in the following way under the  $SU(3)_C$  group global transformations:

$$q_f(x) \xrightarrow{SU(3)} q'_f(x) = e^{i\frac{\lambda^a}{2}\theta_a(x)} q_f(x), \quad (1.2)$$

where  $\lambda^a$  ( $a = 1, 2, \dots, 8$ ) are the eight Gell-Mann matrices, the generators of the  $SU(3)_C$  group, and  $\theta^a$  are functions that depends on space-time coordinates. Since  $SU(3)$  is a non-abelian group, the group algebra is defined by the commutation relation:

$$\left[ \frac{\lambda^a}{2}, \frac{\lambda^b}{2} \right] = if^{abc} \frac{\lambda^c}{2}, \quad (1.3)$$

where  $f^{abc}$  denotes the  $SU(3)_C$  structure constants.

Nonetheless, the derivatives  $\partial^\mu q_f$  do not transform in the same way, and it is needed to be redefined as covariant derivatives in order to maintain the Lagrangian density invariance under the transformation in Eq. 1.2. The covariant derivative is defined as:

$$D^\mu = \partial^\mu + ig_s \frac{\lambda^a}{2} G_\mu^a(x), \quad (1.4)$$

where eight gauge fields  $G_\mu^a$  (gluon), are introduced, responsible for the color charge interaction between fermions. The coupling constant of the strong interaction is labelled as  $g_s$ <sup>3</sup>.

The introduction of the vector fields  $G_\mu^a$  ensures the invariance, under the local gauge transformation of the Lagrangian. After adding the kinetic term for the gauge fields, the QCD Lagrangian is

$$\mathcal{L}_{\text{QCD}} = \sum_f \bar{q}_f (i\gamma^\mu D_\mu) q_f - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \quad (1.5)$$

where  $G_{\mu\nu}^a = \partial^\mu G_\nu^a - \partial^\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c$ , is the gluon field tensor. The last term in Eq. 1.5 is required as a consequence of the non-abelian nature of the  $SU(3)$  group, and it implies the existence of triple and quartic gluon self-interactions. Rewriting the last equation, using the definition of  $D_\mu$ , the Lagrangian density of QCD is:

$$\mathcal{L}_{\text{QCD}} = \bar{q}_f (i\gamma^\mu \partial_\mu) q_f - g_s \bar{q}_f^i \gamma^\mu \frac{\lambda_a}{2} q_f^j G_\mu^a - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a. \quad (1.6)$$

The first term represents the original Lagrangian density from Eq. 1.1, standing for the free-field propagation of the quarks. The second term comes from the covariant derivative and represents the interaction of the quarks with the vector field  $G_\mu$ , i.e. the interaction between the quarks and gluons, through the coupling constant  $g_s$ . Thanks to the second term it is possible the change of the colour of the quarks from  $i$  to  $j$ . The last term is required as a consequence of the non-abelian nature of the  $SU(3)$  group, and it implies the existence of triple and quartic gluon self-interactions.

In the QCD Lagrangian all the interactions are given in terms of a single universal coupling, the strong coupling constant  $g_s$ . The existence of self-interactions among the gauge fields is a new feature, unique for this interaction, and somehow these interactions could explain properties like the asymptotic freedom, in which strong interactions become weaker at short distances, and color confinement, where the strong forces increase at large distances. These properties are characteristic of QCD and do not appear in other theories [14].

## Electroweak interaction

The electroweak interaction (EW) is described by a single gauge theory, with the symmetry group  $SU(2)_L \times U(1)_Y$ , and includes the electromagnetic and weak interactions;

---

<sup>3</sup>The strength of the interaction is parametrized by the constant  $g_s$  usually redefined by the strong coupling constant  $\alpha_s = g_s^2/4\pi$ .

---

the unification of these interactions started at the end of the 1960s theorized by Glashow, Weinberg and Salam [15, 8, 16]. The EW interactions are explained in the SM with the same local gauge invariance mechanism as strong interactions by imposing a symmetry under the  $SU(2)_L \times U(1)_Y$  group.

In order to account for the observed parity violation in the weak interactions [9], Dirac spinors are decomposed into “left-handed” (L) and “right-handed” (R) chiral components. The left and right chiral components of a field are defined from the  $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$  used to define left and right chirality projection operators, as

$$P_{L,R} = \left[ \frac{1 \mp \gamma^5}{2} \right], \quad P_L\psi = \psi_L, \quad P_R\psi = \psi_R. \quad (1.7)$$

In the limit of a massless particle, the chirality corresponds to the helicity, that is defined as the normalized projection of the spin vector onto the spatial momentum vector. Fields therefore are represented as one doublet and two singlets for each family (Tab. 1.1), only exception are neutrinos, since as Goldhaber [17] has experimentally proved, neutrinos with positive chirality eigenvalues do not exist and the right-handed fermions in the SM, ought to be singlets for  $SU(2)_L$ . For each family, the fields are represented by:

$$\Psi_L \equiv P_L \begin{pmatrix} \psi \\ \psi' \end{pmatrix} = \begin{pmatrix} \psi_L \\ \psi'_L \end{pmatrix}; \quad \psi_R \equiv P_R\psi; \quad \psi'_R \equiv P_R\psi'; \quad (1.8)$$

where  $\psi$  and  $\psi'$  represent either the neutrino and charged lepton fields or the up- and down-type quark fields, and for neutrinos  $\psi_R$  vanishes.

Left-handed fermions ( $\psi_L$ ) are doublets and right-handed fermions ( $\psi_R$  and  $\psi'_R$ ) are singlets, both are associated to a “weak” isospin  $Y$ , and have  $T = 1/2$  and  $T = 0$ , respectively.

Given the symmetry group  $SU(2)_L \times U(1)_Y$ , the field transformations are defined by the representation of the group. Under  $SU(2)_L$ , left-handed quarks and leptons transform like doublets in the fundamental representation. Right-handed quarks and leptons are singlets under  $SU(2)_L$  ( $T = 0$ ), so  $SU(2)$  does not act on right-handed fermions. The  $SU(2)_L$  gauge group is a non-abelian group, and the gauge invariance under  $SU(2)$  group results in the presence of three gauge fields  $W_\mu^i$  ( $i = 1, 2, 3$ ). Fermions of right chirality being  $SU(2)_L$  singlets do not interact with the  $W_\mu^i$  fields. Under  $U(1)_Y$ , associated to the weak hypercharge defined as  $Y = 2(Q - T_3)$ , all fields are singlets of hypercharge. The  $U(1)_Y$  gauge group is Abelian and a single gauge field, denoted as  $B_\mu$ , results from the  $U(1)_Y$  local gauge invariance.

As done earlier with QCD, the EW Lagrangian is built by imposing gauge conditions [7] on the Lagrangian for free fields of Tab. 1.1, using the chirality projectors  $P_L$  and  $P_R$ :

$$\mathcal{L}_{free} = \bar{\Psi}_L(i\cancel{D})\Psi_L + \bar{\psi}_R(i\cancel{D})\psi_R + \bar{\psi}'_R(i\cancel{D})\psi'_R, \quad (1.9)$$

where  $\cancel{D} = \gamma_\mu \partial^\mu$  and  $\gamma$  are the Dirac matrices, and the contribution from each generation of leptons and quark of Tab. 1.1 is implied. Imposing the gauge conditions on the left- and right-handed fields under  $SU(3)_L$  and  $U(1)_Y$  of the Lagrangian, the covariant derivatives are defined as:

$$D_L^\mu = \partial^\mu + igW_\mu^i T_i + ig' \frac{Y_L}{2} B^\mu \quad \text{and} \quad D_R^\mu = \partial^\mu + ig' \frac{Y_R}{2} B^\mu, \quad (1.10)$$

where  $Y_L$  and  $Y_R$  are the hypercharge for each quark and lepton family and are reported in Tab. 1.1,  $T_i = \sigma_i/2$ , where  $\sigma_i$  are the Pauli matrices, and  $g$  and  $g'$  are coupling constants, which define the strength of the interactions with the four gauge fields  $W_\mu^i$  and  $B_\mu$ .

Introducing the covariant derivatives, the Lagrangian acquires an interaction term:

$$\mathcal{L}_{int}^{EW} = g\bar{\psi}_L\gamma_\mu\frac{\sigma^i}{2}\psi_L W_\mu^i + ig'\bar{\psi}_L\gamma_\mu\frac{Y_L}{2}\psi_L B^\mu + g'\bar{\psi}_R\gamma_\mu\frac{Y_R}{2}\psi_R B^\mu + g'\bar{\psi}'_R\gamma_\mu\frac{Y_R}{2}\psi'_R B^\mu. \quad (1.11)$$

By comparing the experimental currents associated to the Eq. 1.11 to the ones needed to describe the electroweak interactions, it can be seen that the charged  $W^\pm$  bosons and the photon are linear combinations of the introduced gauge fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2). \quad (1.12)$$

The  $Z_\mu$  and photon  $A_\mu$  fields, corresponding to the  $Z$  and  $\gamma$  bosons can be derived from a linear superposition parameterized with the weak mixing angle, called the Weinberg angle  $\theta_W$ , given by  $\tan \theta_W = g'/g$ , and has been experimentally measured as  $\sin^2 \theta_W = 0.23122(15)$  Ref. [8]:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu \end{pmatrix}. \quad (1.13)$$

The electric charge can be expressed as  $e = g \cos \theta_W = g' \sin \theta_W$ , at leading order.

The kinetic terms of the gauge fields and the self-interaction of the weak gauge bosons are written as:

$$\mathcal{L}_{kin}^{EW} = -\frac{1}{4}W_i^{\mu\nu}W_{\mu\nu}^i - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \quad (1.14)$$

where:

$$\begin{aligned} B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu, \\ W_i^{\mu\nu} &= \partial^\mu W_i^\nu - \partial^\nu W_i^\mu + g\epsilon^{ijk}W_j^\mu W_k^\nu; \end{aligned}$$

and  $\epsilon^{ijk}$  is the Levi-Civita tensor.

From the first term of Eq. 1.14 there are obtained up to quartic interaction vertices between charged bosons or charged and neutral bosons, while the second term produces vertices with no more than two neutral bosons. Triple gauge couplings (or quartic interaction vertices) of only neutral bosons such as  $ZZZ$ ,  $ZZ\gamma$ ,  $Z\gamma\gamma$ , are then absent in the SM.

Finally, the EW Lagrangian is obtained by adding the kinematic term from Eq. 1.14:

$$\mathcal{L}_{EWK} = \mathcal{L}_{free} + \mathcal{L}_{int}^{EWK} + \mathcal{L}_{kin}^{EW}. \quad (1.15)$$

The first term is related to the kinetic energies of lepton and quarks, second term to their interactions with  $W^\pm$ ,  $Z$  bosons and the photon and the third term to the  $W^\pm$ ,  $Z$  bosons and the photon kinetic energies and self-interactions.

As it was the case with the strong interaction Lagrangian, in Eq. 1.6, any explicit mass term of the gauge fields would break the invariance, while mass terms for the fermions are also not allowed due to the left and right chiralities of the fields that would generate a mass term  $m\psi^-\psi = m(\psi_R^-\psi_L + \psi_L^-\psi_R)$  that would incorrectly mix singlets and doublets. This is in total contradiction with reality where, as shown in Fig. 1.1, weak bosons ( $W$  and  $Z$ ) [18], and almost all fermions are experimentally observed to be massive.

The solution to this problem is the Higgs-Brout-Englert mechanism [19, 20, 21] that will be discussed in the following section.

### 1.1.2 Electroweak symmetry breaking and the Higgs Mechanism

Experimental observations showed that many SM particles, in particular fermions and weak gauge bosons, have a non-zero mass. In order to reproduce the observed masses, the simple addition of a mass term in the Lagrangian of the electroweak theory is not a valid method, because this term would spoil the gauge invariance. As an example, a mass term for a  $W$  boson can be written as  $m_W^2 W_\mu^\dagger W^\mu$ , is not invariant under the gauge transformations of the  $SU(2)_L \times U(1)_Y$  group, similarly a mass term for fermions would generate a mass term  $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$  that is not invariant too. The mechanism able to account for the mass generation of the weak gauge bosons, and of the experimental massive fermions without “breaking” the gauge theory, known today as the Higgs-Brout-Englert (BEH) mechanism, or simply the Higgs mechanism. The BEH mechanism has become an essential part of the SM, and it requires a new particle of spin-0, called the Higgs Boson.

### 1.1.3 The Higgs mechanism

The BEH mechanism is based on the concept of spontaneous symmetry breaking, a phenomenon that is often observed in nature, whenever individual ground states of a system do not satisfy the symmetries of the system itself.

A classic physical example of this phenomenon is provided by ferromagnets, which can be modeled as spin on a grid. The system has a Lagrangian invariant under a certain symmetry and a non-invariant ground state. The Lagrangian describing the spin-spin interaction is invariant under tridimensional rotations ( $SO(3)$ ), but at a certain ground state this symmetry is broken. The Curie temperature ( $T_C$ ), or Curie point, is the temperature above which certain materials lose their permanent magnetic properties. For temperatures above  $T_C$  the ferromagnets are in the paramagnetic phase, which means the spin of the system is completely disordered since the spins point in arbitrary directions, as in Fig. 1.2a. In that scenario, the ground state is invariant under  $SO(3)$  rotations due to the disordered formations of the spin. However, for temperatures below  $T_C$  the ferromagnets are in the ferromagnetic phase, a spontaneous magnetisation of the system occurs aligning the spins in some specific direction, as in Fig. 1.2b, the ground state is not invariant under global rotations of the  $SO(3)$  group. The symmetry is said to be "spontaneously broken" because the ground state is not invariant under it, but the ground state chosen is only one out of the continuum of the possible ground states [22, 14].

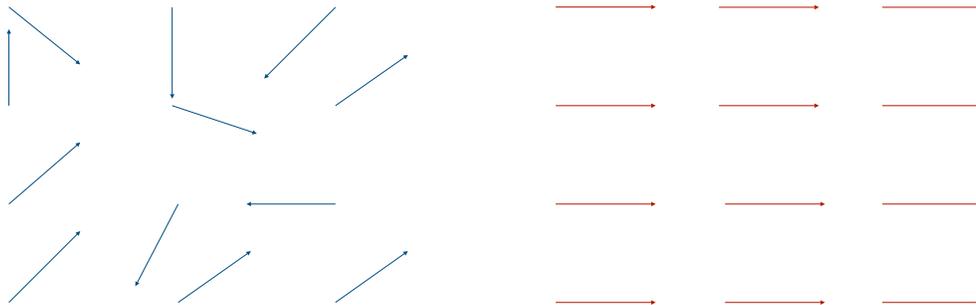


Figure 1.2: Two dimensional model of spontaneous breaking of global symmetry. (a) The ferromagnet for  $T > T_C$ . The global rotational symmetry is broken (b) The ferromagnet for  $T < T_C$ . The global rotational symmetry is spontaneously broken.

In quantum field theory it is conventional to call the ground state the vacuum. The SSB mechanism appears when the lowest-energy (vacuum) state of a theory is not invariant under the full symmetry of the corresponding Lagrangian and the symmetry is spontaneously broken (SSB) when the system goes to such a vacuum state. However, since the chosen vacuum is related to the original symmetry of the system, the initial

symmetry is said to be hidden among the multiple choices of vacuum states.

In the SM, the SSB is achieved through the BEH mechanism, proposed in 1964 independently by the Englert and Brout [19], Higgs [20] and Guralnik, Hagen and Kibble [21], and it is based on the spontaneous breaking of the electroweak symmetry.

The SSB has the important feature that the Lagrangian of the Higgs field is invariant under the  $SU(2)_L \times U(1)_Y$  group but the ground state of the theory not [6], which turns out to be degenerate, and therefore it is necessary to break the symmetry choosing one eigenstate. The SSB is achieved by injecting a new scalar field, called the Higgs field, that is symmetric under the  $SU(2)_L \times U(1)_Y$  group, that acquires a non-zero expectation value (VEV) in the vacuum state. The simplest field of this type that can be introduced is a  $SU(2)_L$  doublet of complex scalar fields with weak hypercharge  $Y = 1$  and weak isospin  $T_3 = \pm 1/2$ :

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (1.16)$$

where  $\phi^+$  and  $\phi^0$  are each complex fields:

$$\phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}}, \quad \phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}}; \quad (1.17)$$

and  $\phi_i$  ( $i = 1, 2, 3, 4$ ) represent four scalar fields that contribute.

The fields must be scalar to satisfy space isotropy, otherwise the expectation value on the vacuum would be frame-dependent. Moreover, the expectation value on the vacuum must be constant to satisfy space homogeneity.

The Lagrangian associated to this doublet must be invariant under the  $SU(2)_L \times U(1)_Y$ :

$$\mathcal{L}_{\text{BEH}} = (D_\mu \phi^\dagger)(D^\mu \phi) - V(\phi), \quad (1.18)$$

the covariant derivative, in order to obtain such invariance, since  $Y = 1$ , is constructed as:

$$D_\mu = \partial_\mu - ig \frac{1}{2} \sigma_i W_\mu^i - ig' \frac{Y}{2} B_\mu, \quad (1.19)$$

where  $g$  is the coupling constant of the  $SU(2)_L$  group,  $g'$  is the  $U(1)_Y$  coupling constant, the  $W_\mu$  and  $B_\mu$  are the gauge fields corresponding to the generators of the electroweak group, and  $\sigma = (\sigma_1, \sigma_2, \sigma_3)$  are the Pauli matrices.

The potential  $V(\phi)$  is introduced ad hoc and is expressed as :

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.20)$$

The parameter  $\lambda$  is chosen to be greater than zero to make the potential bounded from below, instead the parameter  $\mu^2$  is considered  $\mu^2 < 0$ . The Higgs potential of Eq. 1.18 has the form of the so-called ‘‘Mexican Hat’’. This form is because the potential does not have a unique minimum and the ground state with  $\phi = 0$  corresponds to a local maximum of the potential, that is to an unstable equilibrium. The shape of the Higgs potential is shown in Fig. 1.3

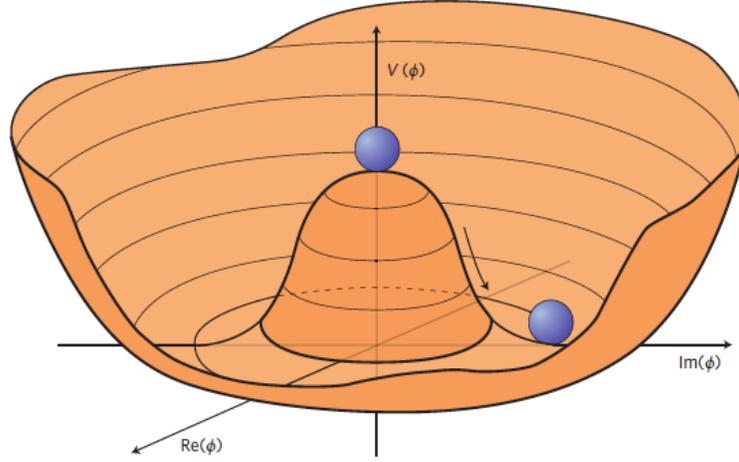


Figure 1.3: The Higgs potential is shown. The vacuum state is degenerate, with an expectation value of  $v$  and it is represented by a circle at the bottom where the minima of the potential lie. The symmetry of the electroweak interaction is spontaneously broken when the system chooses one particular vacuum state [23].

Minimizing the potential:

$$\frac{\partial}{\partial(\phi^\dagger\phi)} V(\phi^\dagger\phi) = \mu^2 + 2\lambda[(\phi_{vac}^+)^2 + (\phi_{vac}^0)^2] = 0; \quad (1.21)$$

the minimum of the potential is given by all the doublets that satisfy the condition:

$$|\phi^2| = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \Rightarrow V(\phi_0) = -\frac{\lambda}{4}v^4. \quad (1.22)$$

When a particular vacuum is chosen for  $\phi$  (Eq. 1.16), corresponding to  $\phi_3 = v$  and  $\phi_1 = \phi_2 = \phi_4 = 0$ :

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v = \sqrt{\frac{-\mu^2}{\lambda}}, \quad (1.23)$$

where  $v$  is the VEV.

All the states on which the expectation value of  $\phi_3$  is  $v$ , are related to this vacuum state by a rotation of  $SU(2)$ , and correspond to a minimum of the potential, as shown in Fig.

1.3. The ground state is clearly no longer invariant under an arbitrary  $SU(2)_L \times U(1)_Y$  transformations, leading to the spontaneous symmetry breaking of the  $SU(2)_L \times U(1)_Y$  symmetry, down to a  $U(1)$  subgroup, when a vacuum state is chosen. Since the electric charge is a conserved quantity ( $Q = 0$ ), only the neutral scalar field can acquire a vacuum expectation value, meaning that the subgroup  $U(1)_{em}$  is chosen as the unbroken subgroup, to preserve the gauge symmetry of QED. The electroweak symmetry breaking (EWSB) is thus specified as:

	$SU(2)_L$	$\times$	$U(1)_Y$	$\xrightarrow{\text{EWSB}}$	$U(1)_{em}$ ,
generators:	$T^1, T^2, T^3$		$Y_W$		$Q$
	weak isospin		weak hypercharge		electric charge
gauge coupling:	$g$		$g'$		$g_e$
gauge bosons:	$W_\mu^1, W_\mu^2, W_\mu^3$		$B_\mu$		$A_\mu$

After a particular ground state or VEV is chosen (Eq. 1.23), and the symmetry is spontaneously broken to the electromagnetic subgroup  $U(1)_{em}$ , as a consequence of the Goldstone theorem [24], since three of four generator of  $SU(2)_L \times U(1)_Y$  were broken, three massless bosons are generated.

The field expansion around the ground state of the  $SU(2)_L$  doublet (Eq. 1.16) can be parameterized as:

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{i\sigma_i}{2} \theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} = U^\dagger \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.24)$$

where  $H(x)$  and  $\theta^i(x)$  are four real scalar fields. The  $\theta^i(x)$  fields are three massless Goldstone bosons, whose existence is predicted by the Goldstone theorem. These massless fields, not observed in nature, and can they be removed with the following gauge transformation:

$$\phi \rightarrow \phi'(x) = U\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (1.25)$$

After this transformation, only the real scalar field  $H$  remains associated to the physical particle Higgs boson and ensures that the fields appearing in the Lagrangian correspond to the physical particles. Substituting  $\phi'$  from Eq. 1.25, as well as the physical  $W_\mu^\pm$ ,  $Z_\mu$  fields introduced in Eq. 1.12 and Eq. 1.13, respectively, into the Eq. 1.18

$$\mathcal{L}_{\text{BEH}} = \frac{1}{2} \partial^\mu H \partial_\mu H + \frac{1}{2} (2\mu^2) H^2 \quad (1.26)$$

$$+ \left[ \left( \frac{gv}{2} \right)^2 W^{\mu+} W_\mu^- + \frac{1}{2} \frac{(g^2 + g'^2)v^2}{4} Z^\mu Z_\mu \right] \left( 1 + \frac{H}{v} \right)^2 \quad (1.27)$$

$$- \lambda v H^3 - \frac{\lambda}{4} H^4 + \frac{\lambda}{4} v^4. \quad (1.28)$$

The first line of the Eq. 1.26 is the free Lagrangian, which describes the evolution of the Higgs field, along with its respective associated boson of mass  $m_H = \sqrt{-2\mu^2}$ , and it is a free parameter of the theory, directly related to the parameter  $\mu^2$  of the scalar potential. The second line describes the mass terms of the weak bosons ( $W$  and  $Z$ ) and the interaction of them with one or two Higgs bosons:  $\text{HWW}$ ,  $\text{HZZ}$ ,  $\text{HHWW}$  and  $\text{HHZZ}$ . The masses of the bosons  $Z$  and  $W$  are calculated as:

$$m_Z = \frac{\sqrt{g^2 + g'^2}}{2} v \quad \text{and} \quad m_W = \frac{gv}{2} = m_Z \cos \theta_w. \quad (1.29)$$

Due to the transformation performed in the Eq. 1.24, the degrees of freedom associated with the  $\theta^i(x)$  fields are turned into longitudinal degrees of freedom. Thus the Goldstone bosons are removed, since they are absorbed as additional degrees of freedom of the  $W^\pm$  and  $Z$  bosons, and this process allows the bosons to acquire mass.

The third line of Eq. 1.26 predicts  $H^3$  and  $H^4$ , which are the cubic and quartic self-interactions of the Higgs boson respectively. Finally, the absence of terms with  $A_\mu$  reflects the fact that the  $U(1)_{em}$  is not broken, thus the photon remains massless.

In order to describe the self-couplings of the Higgs boson, are used the already known parameters: the Higgs boson mass ( $m_H$ ) and its vacuum-expectation value  $v$ . The BEH potential can be rewritten in terms of a trilinear and a quadrilinear coupling as:

$$V(H) = \frac{1}{2} m_H^2 H^2 + \lambda v H^3 + \frac{1}{4} \lambda H^4 - \frac{\lambda}{4} v^4, \quad (1.30)$$

in this way the self-couplings can be written as:

$$\lambda_{HHH} = \lambda v = \frac{m_H^2}{2v} \quad \lambda_{HHHH} = \frac{\lambda}{4} = \frac{m_H^2}{8v^2} \quad (1.31)$$

The self couplings of the Higgs boson depend only on the mass of the Higgs boson and the VEV. The latter corresponds to the energy scale of the electroweak symmetry breaking and can be evaluated at the first order from the Fermi constant  $G_F$  that is precisely determined from the muon lifetime as [8]:

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g}{2\sqrt{2}}\right)^2 \frac{1}{m_W^2} \Rightarrow v = \frac{1}{\sqrt{2}G_F} \approx 246 \text{ GeV} \quad (1.32)$$

The precise measurement of the self-couplings of the Higgs boson represents a crucial test of the validity of the SM predictions, since in the SM the values of  $m_H$  and  $v$  have already been measured directly or indirectly, and any deviation from the SM predictions would imply the existence of physics beyond the Standard Model (BSM).

Until now, the fermions remain massless, however the solution to the problem of their masses lies in the fact that in the SM the Higgs field couples to fermions through the Yukawa interaction. This interaction provides a mass term to the fermions, since the coupling strength between the Higgs boson and fermions ( $y_f$ ) is proportional to the fermions masses. In terms of the right- and left-handed chiral fields ( $\Psi_L, \psi_R$  and  $\psi'_R$ ) described in Eq. 1.8, the interaction between the scalar Higgs field and the fermions interaction is described by the Yukawa Lagrangian

$$\mathcal{L}_{\text{Yukawa}} = -y_{f'}(\bar{\Psi}_L \phi \psi'_R + \bar{\psi}'_R \phi^\dagger \Psi_L) - y_f(\bar{\Psi}_L \tilde{\phi} \psi'_R + \bar{\psi}'_R \tilde{\phi}^\dagger \Psi_L), \quad (1.33)$$

where  $\tilde{\phi} = i\sigma_2 \phi^*$ , and  $\phi$  is the Higgs doublet of the spontaneously broken state from Eq. 1.25. The Yukawa Lagrangian is Lorentz and gauge invariant, which means it can be added into the SM Lagrangian.

The  $\mathcal{L}_{\text{Yukawa}}$ , after replacing explicitly the scalar doublet  $\phi$ , can be written as:

$$\mathcal{L}_{\text{Yukawa}} = - \sum_f m_f (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) \left(1 + \frac{H}{v}\right), \quad (1.34)$$

where the subscript  $f$  refers to the up and down fermions, and the mass of each fermion is calculate from the coupling  $y_f$  as

$$m_f = y_f \frac{v}{\sqrt{2}}. \quad (1.35)$$

As mentioned above, the mass of the fermions depends directly on the strengths of the interactions.

In summary, the fermions and gauge bosons initially are massless, however the BEH mechanism allows these particles to acquire mass when the Higgs boson scalar field is introduced. The equations 1.35 and 1.31 show that the Higgs boson couples to the SM particles with strength proportional to the particle mass, for fermions, and to the mass squared, for gauge bosons. As a consequence, the decays of the Higgs boson more likely to occur in the heaviest kinematically accessible particles.

## 1.2 Higgs boson phenomenology and experimental status

The observation of the Higgs boson, announced in 2012 by the ATLAS and CMS experiments at the LHC [1, 2], is the experimental confirmation of the BEH mechanism, since it allowed to measure the Higgs boson mass  $m_H$ , one of the two free parameters from the theory. The other parameter  $v$  is calculated from the Fermi constant  $G_F$  which is accurately calculated from the lifetime of the muon (Eq. 1.32). The Higgs boson mass was observed with a significance of 5.9 and 5.0 standard deviations, respectively. The discovery is based on a combination of studies in different final states performed with data from the p-p collisions with a center of mass energy  $\sqrt{s} = 7 \text{ TeV}$  and  $\sqrt{s} = 8 \text{ TeV}$ , collected in 2011 and 2012 (LHC Run I).

The Higgs boson can be produced through a number of different processes at the energies of LHC. The dominant production modes in p-p collisions in order of relevance are the gluon-gluon fusion (GGF), the vector boson fusion (VBF), the Higgs-strahlung (production in association with a W or Z boson) and in association with a pair of top quarks ( $t\bar{t}H$ ). The leading-order (LO) Feynman diagrams corresponding to these processes are shown in Fig. 1.4.

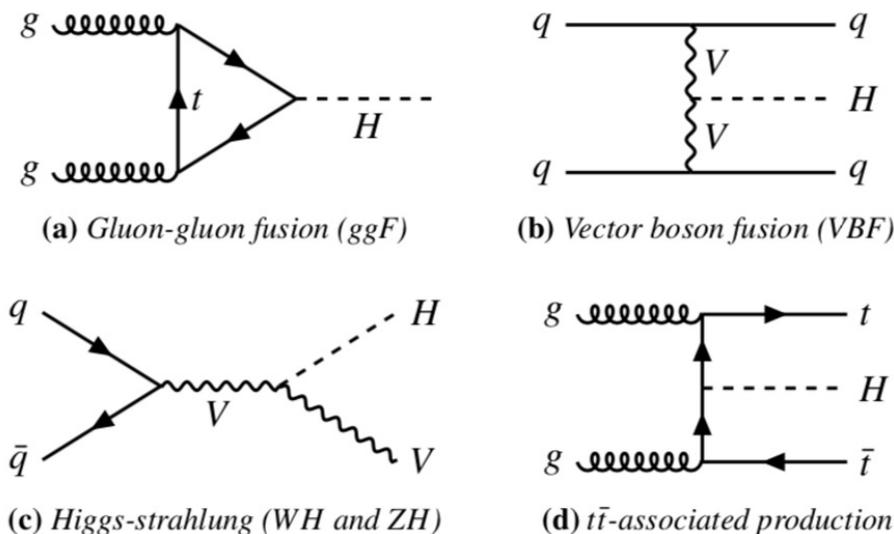


Figure 1.4: Examples of LO Feynman diagrams corresponding to the dominant mechanisms for producing the Higgs boson at the LHC. V refers to either the W or Z boson.

The cross section for producing the Higgs boson in p-p collisions at LHC as a function of the centre-of-mass energy  $\sqrt{s}$ , for the different production mechanisms, is shown Fig. 1.5 (left). The leading production mechanism is the GGF production via a loop of virtual top quarks (Fig. 1.4 (a)). The second leading production mechanism (Fig. 1.4 (b)),

the VBF production, despite the low cross section, has a peculiar topology with two forward jets well separated in pseudorapidity and with a large invariant mass, providing a useful handle to discriminate signal events from background. Both the VBF and the Higgs-strahlung production mechanisms allow to test the Higgs boson coupling with the vector bosons. The more rare mechanism, among the processes displayed in the left of Fig. 1.4, to produce a Higgs boson, is in association with a single quark top or a pair of top quarks ( $tH$  and  $t\bar{t}H$ ); however these processes allow to make a direct measurement of the magnitude of the top Yukawa coupling  $y_t$ .

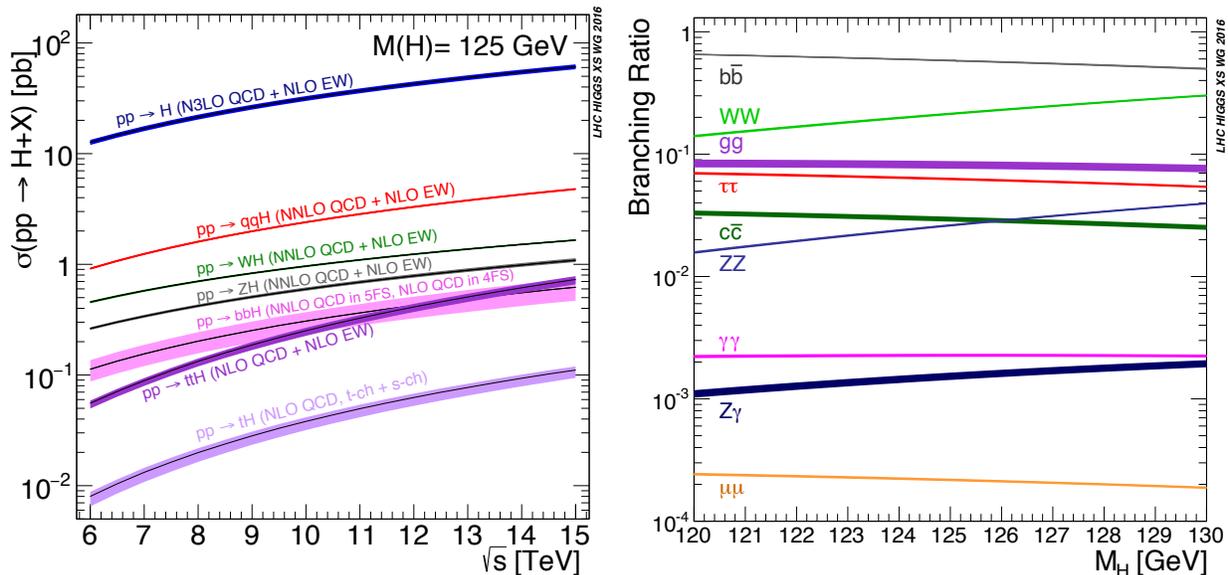


Figure 1.5: (Left) The SM Higgs boson production cross sections as a function of the center of mass energy,  $\sqrt{s}$ , for the different production mechanisms. The GGF process is indicated as H, while the VBF process as  $qqH$ . For each curve is indicated at which order of perturbation theory is obtained. (Right) Branching ratios for the main decays of the SM Higgs boson near  $m_H = 125$  GeV [25, 26].

The Higgs boson can decay into several final states, and the calculation of the branching fractions (BR) depending on its mass. The branching ratios for the main decays of the SM Higgs boson near  $m_H = 125.09$  GeV are reported in Fig. 1.5 (right) [25, 26]. Preferred decays are into heavier fermions and into W or Z bosons. The largest branching ratio is for the decay  $H \rightarrow b\bar{b}$ , for  $m_H = 125$  GeV, since the decay into a pair of top quarks is not kinematically allowed given that the top quark is heavier than the Higgs boson. Although the Higgs boson mass is less than twice the W boson mass, the second largest branching ratio is for the  $H \rightarrow W\bar{W}$  decay, where one of the W bosons is produced off-mass-shell. The Higgs boson can also indirectly decay into massless particles, such as photons and gluons, via a loop of top quarks or W bosons. A list of the experimentally

Decay mode	Branching ratio [%]
$H \rightarrow bb$	$58.09^{+0.72}_{-0.73}$
$H \rightarrow WW^*$	$21.52 \pm 0.33$
$H \rightarrow gg$	$8.18 \pm 0.42$
$H \rightarrow \tau\tau$	$6.27 \pm 0.10$
$H \rightarrow cc$	$2.88^{+0.16}_{-0.06}$
$H \rightarrow ZZ^*$	$2.641 \pm 0.04$
$H \rightarrow \gamma\gamma$	$0.2270 \pm 0.0047$
$H \rightarrow Z\gamma$	$0.1541 \pm 0.009$
$H \rightarrow \mu\mu$	$0.02171^{+0.00036}_{-0.00037}$

Table 1.2: Branching fraction of the main Higgs boson decay modes for the SM Higgs boson with  $m_H = 125.09$  GeV. The theoretical uncertainties considered combine the uncertainties on the Higgs boson partial width, the uncertainties related to the mass of the quarks, and of the value of  $\alpha_s$ . Values taken from Ref. [26].

most relevant decay modes and their corresponding BR is given in Tab. 1.2 for a Higgs boson mass  $m_H = 125.09$  GeV.

The two final states that provided most of the sensitivity to the observation of the Higgs boson announced in 2012 [1, 2], were the two photons final state ( $H \rightarrow \gamma\gamma$ ) and the four-lepton final state ( $H \rightarrow ZZ^* \rightarrow \ell^+\ell^-\ell^+\ell^-$ , with  $\ell = e, \mu$ ), despite of their small BR around 0.2% and 0.01% respectively. In fact, thanks to the good resolution of the invariant mass of their final state objects and relative low background, they are the final states with the highest sensitivity. In Fig. 1.6 the invariant mass distribution, of the four leptons final state, obtained by the (a) ATLAS and (b) CMS Collaboration, is shown; the Higgs boson is visible as an excess over the background around 125 GeV.

After a technical stop, in 2015 the LHC started a new data taking period referred to as Run II (from 2016 to 2018) delivering p-p collisions at a higher center-of-mass energy ( $\sqrt{s} = 13$  TeV). Thanks to the increase of energy and the larger amount of data collected, the ATLAS and CMS collaborations investigated less accessible properties of the Higgs boson, like more rare and challenging production processes and decay modes. For example, it was possible the identification of other final states, such as the Higgs decay into a pair of b-quarks  $H \rightarrow bb$  (BR  $\approx 57\%$ ), which is a final state more challenging

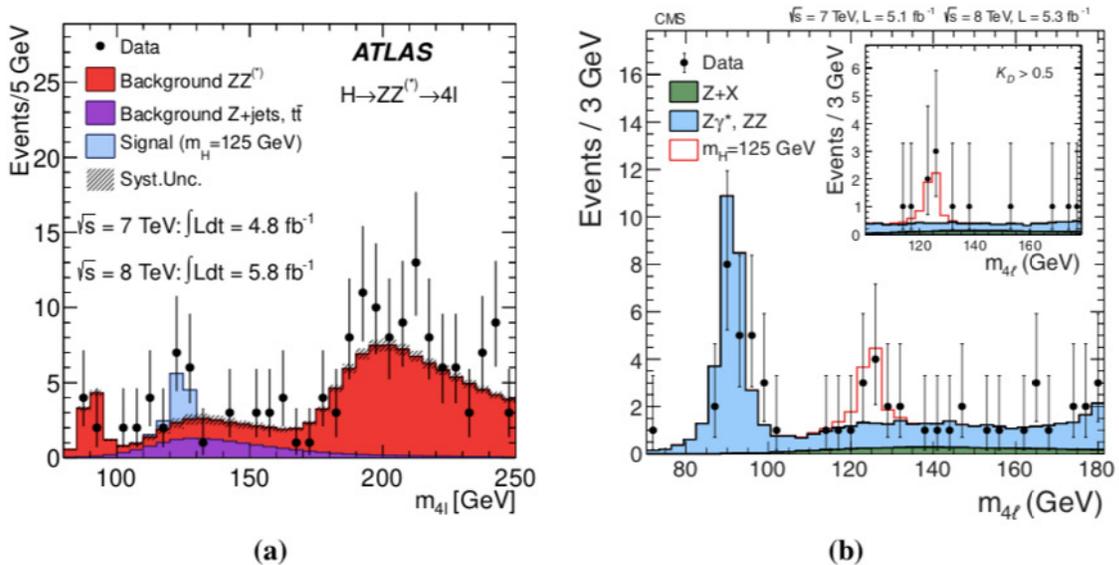


Figure 1.6: Four-lepton invariant mass of  $H \rightarrow ZZ^* \rightarrow 4\ell$  candidates for the combination of 7 TeV and 8 TeV p-p collision data recorded by (a) the ATLAS experiment [1] and (b) the CMS experiment [2]. The points represent the observed data, compared to the background expectation. The signal expectation for the SM Higgs boson with  $m_H = 125$  GeV is also shown, added to the backgrounds. The inset in the plot on the right shows the four-lepton invariant mass distribution after selection of events with  $K_D > 0.5$ , where  $K_D$  is a kinematic discriminant based on the probability ratio of the signal and background hypotheses

because the b-quark hadronize producing hadronic jets; but despite of these difficulties, the CMS collaboration announced the observation of the Higgs boson decay to bottom quarks in 2018 [27]. Also, the decay of the Higgs boson into a pair of  $\tau$  leptons was announced in 2018 by the CMS collaboration [28].

The measurements performed confirmed that the observed new particle properties are consistent with the ones predicted by the BEH mechanism: in particular the new particle is consistent with the CP even-parity and zero-spin hypotheses [29, 30] and has been shown to behave, interact and decay in many of the ways predicted by the SM [31]. As an example, the interaction of the Higgs boson with fermions and vector bosons must be, respectively, linearly and quadratically proportional to their masses and this is confirmed by ATLAS and CMS [32], as shown in Fig. 1.7; where the CMS measurements of the couplings as a function of the masses, are reported. Normalized Higgs boson coupling constants as a function of the boson or fermion masses, where  $v = 246$  GeV is the vacuum expectation value of the Higgs field. The dashed (blue) line indicates the predicted dependence on the particle mass in the case of the SM Higgs boson [33].

The discovery of the Higgs boson opened the road for precision test of the SM in the

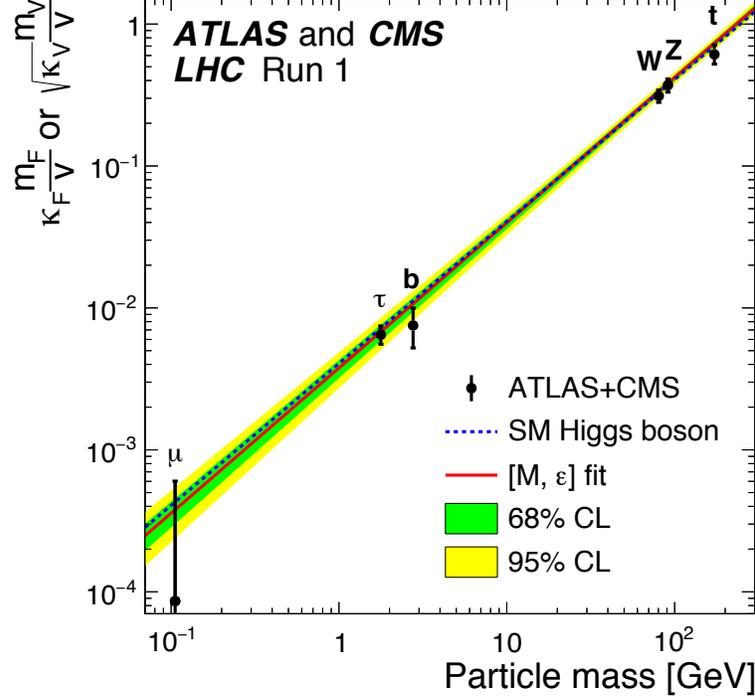


Figure 1.7: Normalized Higgs boson coupling as a function of particle mass for the combination of ATLAS and CMS data [33].

scalar sector [34]. As stated before, the properties of the new particle measured until now are in agreement with the ones predicted by the SM. The still-missing properties to be measured are the trilinear and quadrilinear self-couplings ( $\lambda_{\text{HHH}}$  and  $\lambda_{\text{HHHH}}$ ). These couplings are associated respectively to the cubic and the quartic term of the Higgs field Eq. 1.30 and their values are predicted by the SM after the measurement of the Higgs boson mass. The direct measurement of  $\lambda_{\text{HHH}}$  and  $\lambda_{\text{HHHH}}$  will provide a crucial test of the scalar sector [34] since the tiniest deviation from the SM predictions will be a hint for new physics.

The  $\lambda_{\text{HHH}}$  coupling can be probed directly with the Higgs boson pair (HH) production. Similarly, the measurement of the quadrilinear coupling  $\lambda_{\text{HHHH}}$ , a further probe of the BEH potential, requires the study of triple Higgs final state. The production of the latter is however extremely rare in the SM, with a cross section of about 80 fb at LHC for  $\sqrt{s} = 13$  TeV [35], thus it is out of the experimental reach of the LHC. In contrast, the double Higgs production cross section at the LHC amounts to about 40 fb, in p-p collision at  $\sqrt{s} = 13$  TeV, although challenging, the  $\lambda_{\text{HHH}}$  can be experimentally probed. Its measurement is one of the main objectives of the LHC physics program. This thesis is focused on the search of double Higgs production.

## 1.3 Higgs boson pair production

A very powerful way to investigate the Higgs sector is the study of the double Higgs production (HH), predicted by the SM and by many BSM scenarios. The HH production is the only process that gives direct access to the trilinear self-coupling by measuring the production of pair of Higgs bosons. However, as predicted by the SM, at LHC this is a very rare process, with a cross section more than three orders of magnitude less than the cross section for the single Higgs boson production (Tab. 1.3). The precise measurement of the Higgs boson self-coupling of the SM is out of reach on data collected in Run II by ATLAS and CMS. By estimate of Ref. [36], a precision of 50% will be achieved by the end of the high luminosity phase of LHC, unless effects from physics BSM, enhancing the cross section of double Higgs production, appear.

Even though there is a special interest in the HH production, caused by the opportunity to measure  $\lambda_{\text{HHH}}$ , its study gives also access to the Yukawa coupling  $y_t$ , and the trilinear and quartic couplings with vector bosons, as described in the Eq. 1.28. Representative Feynman diagrams of the interactions that contribute to the HH production are shown in the Figures 1.8 - 1.10. The interactions present in the processes will depend on the production mechanism from where the pair of Higgs bosons originates, and these are detailed in the following section.

### 1.3.1 HH production mechanisms at LHC

Similarly to the single Higgs boson production, the double Higgs production has the same production mechanisms, with the same order in relevance; however the HH production is around  $O(1000)$  rarer than the single Higgs production. Each of these processes involves different interactions of Higgs boson with other particles, in addition to the trilinear Higgs coupling, that will contribute to the total cross section of double Higgs production. Their effect, must be properly taken into account in order to extract a valid measurement of the  $\lambda_{\text{HHH}}$  parameter by measuring the cross section of  $pp \rightarrow \text{HH}$  at LHC. In the following, the three main modes of the HH production are described. The cross sections of the main HH production mechanisms, as a function of the of the collider center-of-mass energy  $\sqrt{s}$ , are shown in the Fig. 1.11; in addition the Tab. 1.3 summarizes the cross sections of these processes as predicted by the SM, at  $\sqrt{s} = 13$  TeV [36].

1. **Gluon fusion production (GGF)**  $gg \rightarrow \text{HH}$ , is the dominant HH production mechanism at LHC. The Higgs bosons are produced through the trilinear Higgs boson self-coupling  $\lambda_{\text{HHH}}$ , or the radiation of two on-shell Higgs bosons from heavy quark loop, so the cross section of the GGF mechanism depends on the couplings  $\lambda_{\text{HHH}}$  and  $y_t$ . These two process are denoted by their Feynman diagram as “triangle”

and “box” respectively, which are shown in Fig. 1.8. The triangle contribution (left) depends on both the  $\lambda_{HHH}$  and  $y_t$  couplings, instead the box contribution (right) depends exclusively on the  $y_t$  coupling, having a quadratically dependence on it. Even through this is the most common production mechanism, its cross section has a small value because both of the contributions from the triangle and box diagrams have similar amplitudes and they interfere destructively, turning this into a rare process.

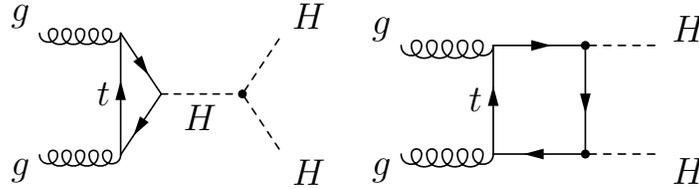


Figure 1.8: Feynman diagrams contributing to Higgs pair production at LO via gluon fusion.

2. **Vector boson fusion production (VBF)**  $qq' \rightarrow V^*V^* \rightarrow qq'HH$  (with  $V = W$  or  $Z$ ), is the second-largest production mechanism at the LHC. It involves diagrams in which a single Higgs boson (off-shell) splits into a Higgs pair (Fig. 1.9 left diagram) and diagrams in which the two Higgs bosons radiate from the virtual  $W$  or  $Z$  bosons (second and third diagram of Fig. 1.9). From Fig. 1.9, while the first diagram involves the trilinear  $H$  self-coupling, the Higgs pair production occurs through trilinear and quadrilinear interaction with vector bosons in the other diagrams. The production cross section of this process is about 20 times smaller than the gluon fusion one, but the presence of two jets in the final state provides a very clean signature, being characterized by two forward jets well-separated in pseudo-rapidity and with a large invariant mass, that can be exploited to discriminate signal from background events in the online and offline selection of this process.

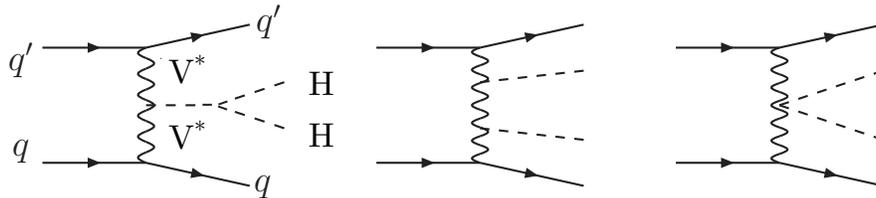


Figure 1.9: Feynman diagrams contributing to Higgs pair production at LO via VBF.

3. **Top quark pair associated production (ttHH)**  $qq'/gg \rightarrow t\bar{t} HH$ : It involves the couplings  $\lambda_{HHH}$  and  $y_t$ . In this process, two Higgs bosons can be created either

radiating from the top quarks or from the Higgs boson self-coupling. The Feynman diagrams contributing to this process are illustrated in Fig. 1.10.

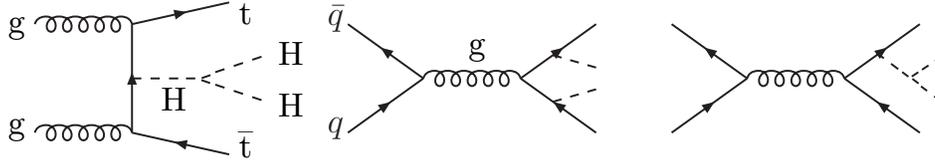


Figure 1.10: Feynman diagrams contributing to Higgs pair production at LO in association with a  $t\bar{t}$  pair.

Since the HH production is very rare, the current searches at the LHC are focused on the main production mechanisms: GGF or GGF plus VBF, the processes with dominant contribution to the total cross section for HH production. In the analysis presented in this thesis, the GGF and VBF processes are considered.

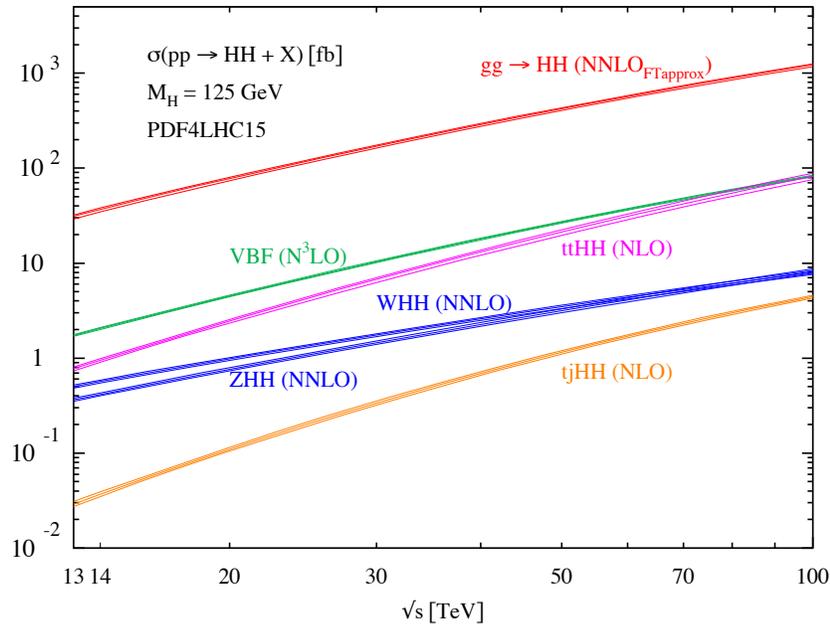


Figure 1.11: Total cross sections for production of Higgs pairs within the SM via GGF, VBF, double Higgs-strahlung and double Higgs bremsstrahlung off top quarks. PDF4LHC15 parton densities have been used. The size of the bands shows the total uncertainties originating from the scale dependence and the PDF+ $\alpha_s$  uncertainties [36].

Process	$\sigma$ [fb] ( $\sqrt{s} = 13$ TeV)
ggHH	$31.05^{+2.2\%}_{-5.00\%}$ (scale) $\pm 3.0\%$ (PDF $+\alpha_s$ ) $\pm 2.6\%$ ( $m_{\text{top}}$ )
VBF HH	$1.73^{+0.03\%}_{-0.04\%}$ (scale) $\pm 2.1\%$ (PDF $+\alpha_s$ )
HHZ	$0.363^{+3.4\%}_{-2.7\%}$ (scale) $\pm 1.9\%$ (PDF $+\alpha_s$ )
HHW <sup>+</sup>	$0.329^{+0.32\%}_{-0.41\%}$ (scale) $\pm 2.2\%$ (PDF $+\alpha_s$ )
HHW <sup>-</sup>	$0.173^{+1.2\%}_{-1.3\%}$ (scale) $\pm 2.8\%$ (PDF $+\alpha_s$ )
t $\bar{t}$ HH	$0.775^{+1.5\%}_{-4.3\%}$ (scale) $\pm 3.2\%$ (PDF $+\alpha_s$ )
tjHH	$0.0289^{+5.5\%}_{-3.6\%}$ (scale) $\pm 4.7\%$ (PDF $+\alpha_s$ )

Table 1.3: SM Cross section of the HH production modes, assuming a Higgs boson mass  $m_H = 125$  GeV [36].

### 1.3.2 Double Higgs production beyond the SM

Despite the success of the SM in describing the phenomenology observed at collider experiment so far, it is far from being a complete theory. Among its limitations are: the lack of a mechanism responsible for the matter-antimatter asymmetry in the Universe, does not provide candidate particles for dark matter and dark energy, and one of the most important missing point, does not includes among the fundamental interactions the gravitational force.

Furthermore, the SM does not explain some parts of the theoretical aspects of the electroweak sector satisfactory. For example, there is not a clear theoretical reason, that could explain the presence of three family of fermions, but only experimental evidence: the fermions are almost replicated over three families, the only distinctive property of each generation are the masses, due to modifications of the coupling with the Higgs boson, and the model does not provide an explanation of why the couplings to fermions span over many orders of magnitude. Furthermore, the value of the mass of the Higgs boson is not determined by a fundamental physical reason, like a symmetry in the theory, the value of the Higgs mass observed is the result of large divergences canceled out by a regularisation mechanism that requires an unnatural extreme fine-tuning [37, 38] if the SM is valid up to the Plank scale.

Another phenomenon that is not explained by the SM is related to the solar neutrino problem [39], which its discrepancy between the expected number of neutrinos and the number of measured neutrinos by the detectors on Earth. This problem was solved by the

inclusion, around 30 year ago, of the neutrino oscillation mechanism, which states that the neutrinos change their flavour because they are massive and their flavor states are a mixture of their mass states. In order to be able to occur the neutrino oscillation, the SM needs to incorporate three neutrinos with masses, which currently are not part of the model [12].

For all these reasons, it is believed that the current SM is not the final theory that will clear all of the open questions. Instead, there should be an extended theory, which could address the problems that the SM is incapable of. Several theories, addressed as BSM theories could provide a solution to the aforementioned problems of the SM. The BSM theories, being an extension of the SM, are formulated in order to preserve all its capabilities of describing the phenomenology observed so far, while adding the explanation to unsolved problems and predicting new phenomena at high-energy scale.

The study of the double Higgs boson production comes in handy in this context, because it offers a probe for several BSM scenarios and discriminate between possible alternatives. Two possible scenarios are presented to account for potential deviations from the SM predictions, the resonant and the non-resonant production of the Higgs pair.

In the resonant production mechanism, the production of a new particle  $X$  that will decays into a Higgs pair is expected, and the experimental signature is a resonance of  $m_X > 2m_H$ . The cross section of the HH production will be enhanced with respect to the one predicted by the SM and the new particle will be observed if the scale of the BSM physics is within the current reach of the LHC. Some of the models for resonant production of Higgs pairs will be discussed in the following.

The non-resonant production could be probed even if the scale of the BSM scenario is higher than the current centre-of-mass energy at the LHC, since a BSM signature in this case is an increase of the cross section of double Higgs production  $\sigma_{HH}$ , and a change of the kinematics of the two Higgs bosons, which could be caused by anomalous Higgs boson couplings. Both mechanisms are studied in this thesis.

### 1.3.2.1 Resonant BSM HH production

The description in the following of the resonant production models, is not meant to be an extensive summary of BSM models, it aims at showing that despite the different theoretical assumptions behind the analyzed BSM models, they can be simultaneously probed in HH production.

The resonant production is predicted by many extensions of the SM such as the Singlet model [40, 41, 42], the Two-Higgs Doublet Model (2HDM) [43], the Minimal Supersymmetric Standard Model (MSSM) [44, 45], and the Warped Extra Dimensions (WED) [46, 47] model. Although they describe different phenomenologies, all these models predict the existenc of a CP-even scalar particle of spin-0 or as a spin-2 particle ( $X$ )

decaying into a Higgs boson pair, so that the signature is that of a resonance with mass  $m_X > 2m_H$ , with an intrinsic width that is negligible with respect to the detector resolution (narrow width approximation).

For example, the WED models, inspired by the string theory, are based on the hypothesis that a finite extra spatial dimension exists; for instance, while the SM belongs to the four-dimensional space, gravity propagates also in higher dimensions in this scenario, so that its interaction in the four-dimensional space appears weaker than that of the other fundamental forces.

The relevant consequence of these models for HH searches is the presence of additional particles, belonging to a higher dimensional space, and with sizeable branching fractions to the HH final state. These new particles required are a spin-2 particle (“Bulk Graviton”, G) [48] and a spin-0 (“Radion”, R) [49], that can decay into a pair of Higgs bosons. The spin-2 Graviton is the mediator of the gravitational force and its branching fraction to an HH pair can be as large as 10% and remain constant as function of its mass,  $m_G$ . The spin-0 Radion stabilizes the size of the extra dimensions, and has a branching fraction in HH of 25% and a very little dependence on the model parameters. The experimental signature of these resonance is a peak in the invariant mass spectra of the Higgs pair candidates and an enhanced Higgs pair production cross section.

Although the other models describe different phenomenologies, they all predict the existence of CP-even scalar spin-0 or spin-2 particles, with an intrinsic width that is negligible with respect to the detector resolution and with similar signatures as those for the graviton and radion. However, the mass hypothesis of these new predicted particles span over a wide range, depending on the respective model space parameters: from 250 to 350 GeV for the 2HDM and MSSM models, and from 250 GeV to 1 TeV for the Singlet Model, and from 250 GeV to 3 TeV for the WED model. Different models can be explored at once with similar strategies, choosing a large phase space of their parameters.

### 1.3.2.2 Non-resonant BSM HH production

BSM physics can be probed in the non-resonant HH production, since its effects could be observed as contributions in the quantum loops responsible for the HH production. Evidence of this physics can be measured as a difference between the measured and predicted cross section of the HH production. If there is a difference between the cross section, will be observed through an anomalous trilinear Higgs self-coupling and in the Higgs boson kinematic properties.

In the SM, the  $\lambda_{HHH}$  coupling depends completely on the values of  $v$  and  $m_H$ , which are already measured (Eq. 1.31). However, in the BSM physics modifications to  $\lambda_{HHH}$  are expected, which could be indirectly observed as anomalous properties of the HH production. In order to parametrize the deviations of the couplings from the SM values,

coupling modifiers are introduced, in the GGF mechanism, for the trilinear Higgs self-coupling and the top Yukawa interaction, defined as:

$$\kappa_\lambda = \lambda_{\text{HHH}}/\lambda_{\text{HHH}}^{\text{SM}}, \quad \kappa_t = y_t/y_t^{\text{SM}}. \quad (1.36)$$

On the other hand, no evidence of BSM physics results in values of  $\kappa_\lambda$  and  $\kappa_t$  equal to 1, as for the SM scenario.

Any deviation of  $\kappa_\lambda$  and  $\kappa_t$  could be due to an imbalance of the amplitude of the contributions of a given production mode, for example modifications to the amplitude of the contribution in the GGF production mechanism (represented in the Fig. 1.8), in the triangle and box diagrams. In the Fig. 1.12a for the main HH production mechanisms, the impact of the deviation of  $\lambda_{\text{HHH}}$  with respect to the SM expected value  $\kappa_\lambda$  on the production the cross section is illustrated.

In the GGF HH production, the cross section normalized to the SM prediction, can be parameterized at LO, taking into consideration the triangle and box contributions, in the following way:

$$\frac{\sigma_{\text{HH}}}{\sigma_{\text{HH}}^{\text{SM}}} = 0.28\kappa_\lambda^2\kappa_t^2 - 1.37\kappa_\lambda\kappa_t^3 + 2.09\kappa_t^4, \quad (1.37)$$

where the coupling modifiers are those defined in Eq. 1.36; and the numerical coefficients are calculated in Ref. [50], in the context of the effective field theory parametrization described later in this section. Different values of  $\lambda_{\text{HHH}}$  result in substantial changes on the HH production cross section, as shown in Fig. 1.12a (gray curve, for GGF) where the cross section for the main mechanisms contributing to the double Higgs production is reported as a function of the  $\kappa_\lambda$ . The cross section of these mechanisms is sensitive to the value of the trilinear self-coupling, each with a different dependency. The curve of GGF fusion (from Eq. 1.37) has a minimum in  $\kappa_\lambda/\kappa_t = 2.45$ , corresponding to the maximum destructive interference between the box and triangle diagram and by moving from the SM value prediction to  $\kappa_\lambda = -1$ , the production cross section is enhanced by about a factor 20. Furthermore, the HH pair kinematics depends on  $\kappa_\lambda$ , as illustrated in Fig. 1.12b where the  $m_{\text{HH}}$  distribution is shown for different values of  $\kappa_\lambda$  [51].

Regarding the VBF HH production, as already mentioned, this mode gives access to  $\kappa_\lambda$ , and, in addition, to the trilinear,  $\lambda_V$  and quadrilinear,  $\lambda_{2V}$ , couplings due to the interaction of one or two Higgs bosons with two vector bosons, where V indicates the Z or W boson. Two coupling modifiers are defined as:

$$c_V = \lambda_V/\lambda_V^{\text{SM}} \quad c_{2V} = \lambda_{2V}/\lambda_{2V}^{\text{SM}}. \quad (1.38)$$

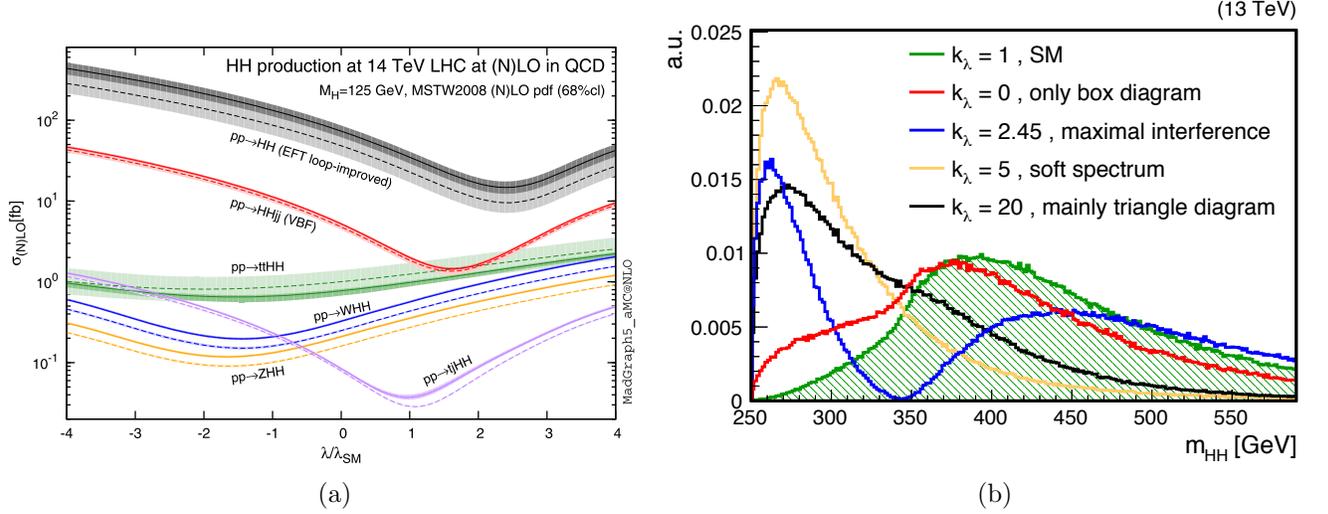


Figure 1.12: HH production cross-section as a function of the coupling modifier  $\kappa_\lambda$  for the dominant mechanisms contributing at LHC [50]. The dashed and solid lines denote respectively the LO and NLO predictions and the bands indicate the PDF and scale uncertainties added linearly. (b)  $m_{\text{HH}}$  distributions for different values of  $\kappa_\lambda$  [51]

The Fig. 1.13 shows the dependence of the di-Higgs production cross section on the coupling modifiers: the quadrilinear coupling with the vector bosons  $c_{2V}$ , and the trilinear Higgs self-coupling  $\kappa_\lambda$ . Both parameters are presented as the difference with respect to the SM value of 1, i.e. the values  $\delta\kappa_\lambda = \kappa_\lambda - 1$  and  $\delta c_{2V} = c_{2V} - 1$ . It is observed that the sensitivity of the cross section has a high dependence on  $c_{2V}$ , since a variation of an unit increases the cross section by a factor of 50. On the other hand, the  $\kappa_\lambda$  parameter has a relatively weak effect on the sensitivity of this process [52].

In practice, none of these observations invalidate the current SM as structurally complete. However, they can point to the existence of more fundamental theories at higher scales; ideally, these theories should be more general and incorporate the existing mechanisms.

### 1.3.2.3 Approaches to measure deviations from the SM predictions for HH production

Two approaches can be followed to measure the possible deviations from the SM predictions: studies can be performed either in the context of a specified model for the resonant production of di-Higgs achievable at LHC or using a model-independent effective field theory for the non resonant di-Higgs production. So far, no new particles have been observed. Assuming the mass of new particles from BSM physics out the direct reach of the LHC, BSM processes can however contribute in the quantum loops responsible for

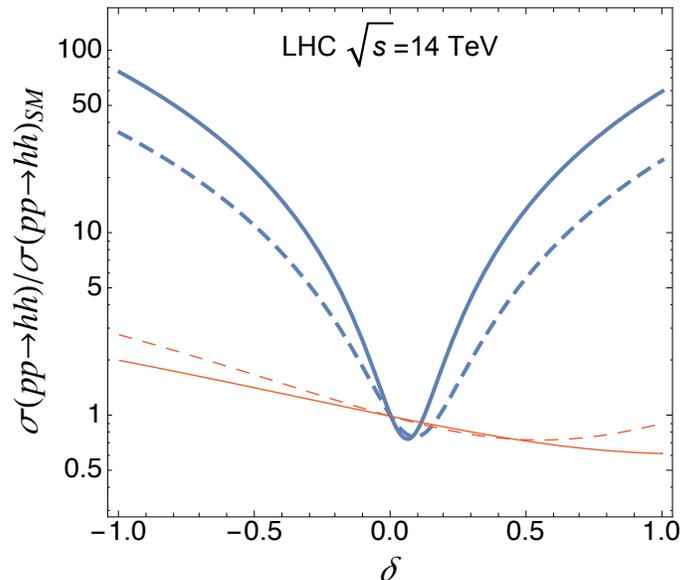


Figure 1.13: Cross section of the VBF HH production, in units of the SM value, as a function of  $\delta\kappa_\lambda$  (orange) and  $\delta c_{2V}$  (blue). The solid lines represent the cross section after applying the cuts representing the acceptance of the detector. The dotted lines represent the cross section after requiring some specific cuts related to the VBF selection [52].

HH production, modifying the kinematics of the process and enhancing the HH production rate, encoding the effects of new heavy states currently beyond experimental reach. New Physics valid at a new energy scale  $\Lambda \gg v$  can be described by complementing the Lagrangian of the SM with additional fields that act only at short distances (very heavy new particles) and at large energy scales. In this case an effective field theory (EFT) [53] approach is used. The key idea of this approach is that interactions that act at short distances can be included in the SM Lagrangian and parameterized in terms of a set of couplings modifiers with the new Lagrangian which provides an “effective” description of any underlying model. In the case of HH production, the Lagrangian will be the SM Lagrangian with corrections described by addition of local operators.

All terms of a renormalisable extended theory Lagrangian, are operators of dimension  $d \leq 4$ . Thus the effects of BSM physics, whose scale is beyond the direct reach of LHC, can be locally approximated by adding to the SM Lagrangian higher order operators ( $d > 4$ ).

The only requirement imposed on these operators is that they should be invariant under the  $SU(3)_C \times SU(2)_L \times U(1)_Y$  local transformations. Taking into account this condition the new EFT Lagrangian can be written as:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{D5} + \mathcal{L}_{D6} + \dots \quad (1.39)$$

These additional operators are suppressed by powers of a scale  $\Lambda$ , which can be seen as the scale up to which only SM fields propagate. Considering that dimension-8 operators, because of the power suppression, can be neglected, and dimension-5 and dimension-7 operators, that involve lepton number violation and neutrino mass terms can be ignored when discussing LHC processes, the EFT Lagrangian will have only dimension-4 operators, that are included into the  $\mathcal{L}_{\text{SM}}$  term, and dimension-6 operators. Thus the Eq. 1.39 can be rewritten as:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^6, \quad (1.40)$$

where the BSM physics is fully parametrized by the Wilson coefficients  $c_i$  and  $\Lambda$  is the energy scale suppressing the higher-dimensional operators. From an experimental point of view, Eq. 1.40 provides a generic parametrization to investigate several BSM signatures with a model-independent approach.

In the VBF case this translates in the modification of the SM couplings of the Feynman diagrams of Fig. 1.9, and as a consequence of the related coupling modifiers  $c_V, c_{2V}$  and  $\kappa_\lambda$ . Instead, for the GGF process additional couplings arise, as described in the following.

### GGF in EFT

In this context, there are five anomalous Higgs boson couplings relevant for HH production: the H coupling to the top quark  $y_t$ , the trilinear self-coupling  $\lambda_{\text{HHH}}$ ; and three additional couplings, which parameterize contact interactions not predicted by the SM, i.e. interactions with a substructure that is not present in the SM. These couplings are denoted as  $c_g$ ,  $c_{2g}$ , and  $c_2$ , representing respectively the interactions of a gluon pair with a single H (ggH), the interactions between two Higgs bosons with two gluons (ggHH), and with a top quark-antiquark pair (ttHH), which could arise through the mediation of very heavy new states. In this context, the diagrams of the GGF HH production, illustrated in Fig. 1.8, can differ from the SM ones, and the values of  $y_t$  and  $\lambda_{\text{HHH}}$  can be different from what is predicted by the SM. The diagrams of all the processes involved at leading order in HH production via gluon fusion are shown in Fig. 1.14, where the BSM couplings are highlighted in red.

The EFT GGH cross section of the HH production normalized to the SM prediction at LO, can be written as [50]:

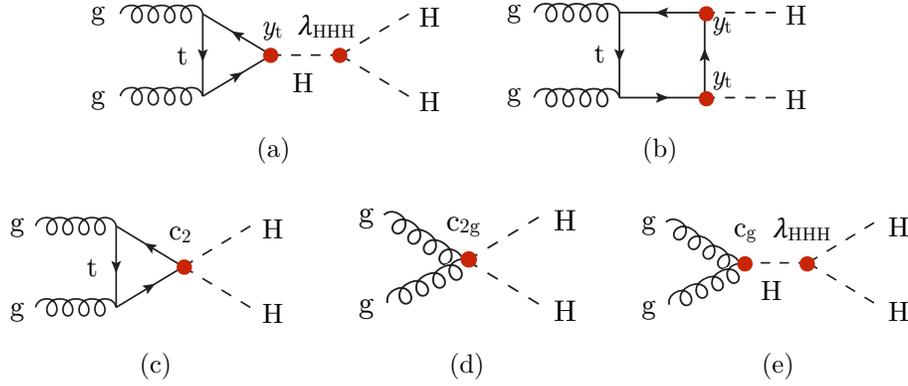


Figure 1.14: Feynman diagrams that contribute to Higgs boson pair production by gluon fusion at LO. Diagrams (a) and (b) correspond to SM-like processes, while diagrams (c), (d), and (e) correspond to pure BSM effects: (c) exploits the contact interaction of two Higgs bosons with top quarks, (d) and (e) describe contact interactions between the Higgs boson and gluons.

$$\begin{aligned}
\frac{\sigma_{HH}}{\sigma_{HH}^{\text{SM}}} &= A_1 \kappa_t^4 + A_2 c_2^2 + (A_3 \kappa_t^2 + A_4 c_g^2) \kappa_\lambda^2 + A_5 c_{2g}^2 \\
&+ (A_6 c_2 + A_7 \kappa_\lambda \kappa_t) \kappa_t^2 + (A_8 \kappa_t \kappa_\lambda + A_9 c_g \kappa_\lambda) c_2 \\
&+ A_{10} c_2 c_{2g} + (A_{11} c_g \kappa_\lambda + A_{12} c_{2g}) \kappa_t^2 \\
&+ (A_{13} \kappa_\lambda c_g + A_{14} c_{2g}) \kappa_t \kappa_\lambda + A_{15} c_g c_{2g} \kappa_\lambda.
\end{aligned} \tag{1.41}$$

In the limit where  $c_g = c_{2g} = c_2 = 0$  Eq. 1.41 is equivalent to Eq. 1.37. The values of the coefficients  $A_i$  are calculated in Ref. [50], and are determined from a simultaneous fit to the cross section obtained from MADGRAPH5\_AMC@NLO [54] simulations at LO precision.

Exploring all possible combinations of the five couplings is not feasible for an experimental search in terms of complexity of the combinations and computing time. A possible alternative approach, as discussed in Ref. [50], is introducing the definition of “shape benchmarks”, which combines the five EFT parameters whose topologies are representative for large regions of the five-dimensional parameter space. The benchmark are defined by scanning a sample of 1507 points generated in a five-dimensional grid, regrouping those with similar kinematic properties in clusters. With this procedure, 12 benchmark shapes are defined and the corresponding shapes are shown in Fig. 1.15. The combinations of the five couplings for each benchmark are reported in Tab. 1.4.

Changes in the values of the coupling do not only affect the cross section of the process, but also the kinematic properties of the HH pair. This effect is illustrated in Fig. 1.15,

---

Benchmark nr.	$\kappa_\lambda$	$\kappa_t$	$c_2$	$c_g$	$c_{2g}$
1	7.5	1.0	-1.0	0.0	0.0
2	1.0	1.0	0.5	-0.8	0.6
3	1.0	1.0	-1.5	0.0	-0.8
4	-3.5	1.5	-3.0	0.0	0.0
5	1.1	1.1	0.0	0.8	-1.0
6	2.4	1.0	0.0	0.2	-0.2
7	5.0	1.0	0.0	0.2	-0.2
8	15.0	1.1	0.0	-1.0	1.0
9	1.0	1.0	1.0	-0.6	0.6
10	10.0	1.5	-1.0	0.0	0.0
11	2.4	1.0	0.0	1.0	-1.0
12	15.0	1.0	1.0	0.0	0.0
box	0.0	1.0	0.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

---

Table 1.4: The values of the anomalous coupling parameters for the 12 benchmark models [50]. In addition the SM and the box values are shown.

which shows the distributions of the invariant mass of the Higgs pair  $m_{hh}$ , for the 12 benchmarks listed in Tab. 1.4. The shapes of the  $m_{hh}$  distributions are very different, clusters with number 1, 2, 5, 6 and 8 have two peaks in the distributions, while the others have a single peak. Furthermore, the range they cover drastically change: comparing the range of cluster 10 with the one of cluster 2, almost all the events from the the former are in the low mass region, while the latter presents a long tail, which extends to values of mass larger than 1 TeV.

## 1.4 Current searches of HH production at the LHC

Since the double Higgs production is a rare process at the LHC, it is important to study different final states of the Higgs decays. Moreover, the study of different final states of the HH production allows to probe different regions of the anomalous couplings space and of the resonant invariant mass spectrum. Therefore, a combination of the different final states is needed in order to have the best possible sensitivity for the HH production.

---

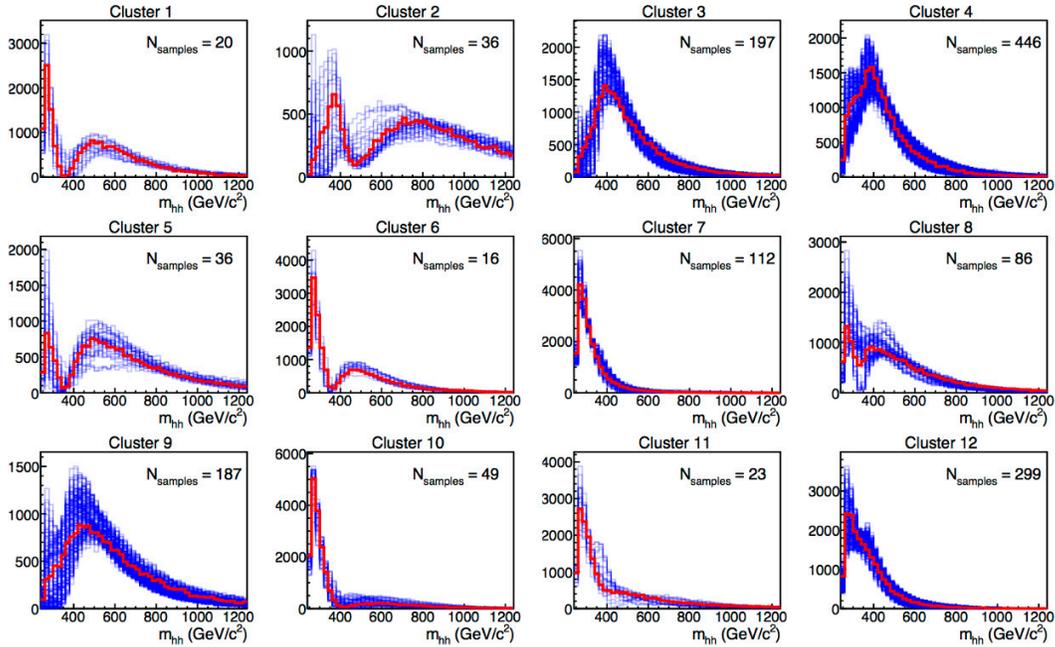


Figure 1.15: Distributions of  $m_{hh}$  of the Higgs pair system, for the clusters listed in Tab. 1.4. The red distribution represent the benchmark sample in each cluster, while the blue distribution the other members of each cluster. Cluster 3 contains the SM sample [50].

### 1.4.1 HH decay channels

The choice of the decay channels is very important, it is necessary to make compromises between the magnitude of the BR and the respective background contamination. For example, it is possible to combine final states with a b-quark pair, which has a high BR, but suffers from high backgrounds. Since the HH production cross section is rather small, at least in SM predictions, experimental HH searches focus mainly on decay channels with larger branching fractions, in order to increase the HH signal rate. Currently at the LHC several decay channels are studied, and the last results of their combination are presented in Refs. [3, 4] by the ATLAS and CMS collaborations. In Fig. 1.16 the BR of the HH final states is illustrated. Some combinations of the final states are very rare, and in those cases the BR is neglected. The BR of the Higgs bosons decays are listed in Tab. 1.2, for a Higgs boson of  $m_H = 125.09$  GeV.

In the following the four most sensitive channels in which the HH production is currently pursued at the LHC are presented:

- $HH \rightarrow bbbb$  is the final state with largest branching fraction (BR  $\simeq 34\%$ ). However, the contamination which mimic the signal signature, mainly due to the production of jets (multijet QCD) is extremely high.
- $HH \rightarrow bb\tau\tau$  has a sizeable branching fraction (BR  $\simeq 7.3\%$ ) and a quite clean

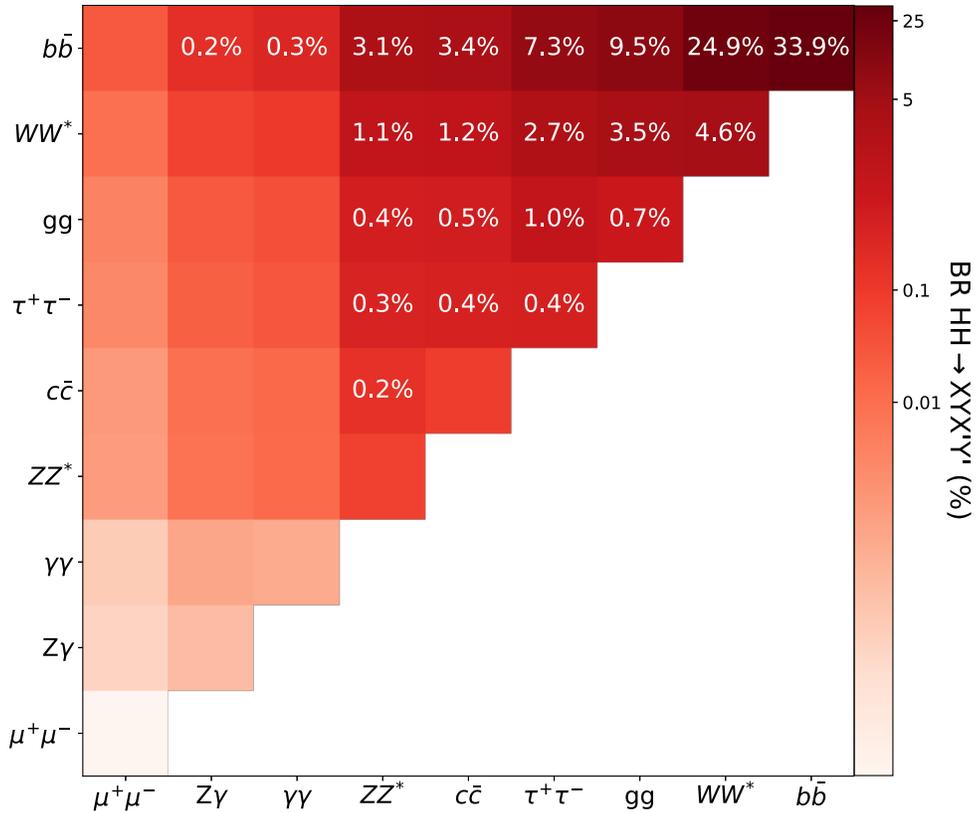


Figure 1.16: Branching fractions of the decay of an HH pair to a selected group of final states. The decay modes are shown on each axis by increasing probability. The numerical values are only shown if larger than 0.1%. The branching fractions of the Higgs boson are evaluated for  $m_H = 125.0$  GeV [36].

and efficient signature of the H decaying into a  $\tau$  pair. This decay mode is a good compromise, although it suffers from large contamination from multijet QCD events and from events where a top-antitop quark pair ( $t\bar{t}$ ), each having a large probability of decaying in a W and a b-quark, is produced.

- $HH \rightarrow bbVV$  has a relatively high total BR (BR  $\simeq 28.1\%$ ), and low background contamination compared with the  $4b$  final states, however it has a large background contribution from the top-antitop quark pair ( $t\bar{t}$ ) process. At CMS, the past and current searches focus on the  $VV$  ( $ZZ$  or  $WW$ )  $\rightarrow 2\ell 2\nu$  case, reducing by one order of magnitude the total BR ( $\simeq 2.7\%$ ).
- $HH \rightarrow bb\gamma\gamma$  although the very small BR (BR  $\simeq 0.26\%$ ) the clean signature of the photon pairs and of their kinematic results in a high signal selection efficiency and purity.

Many other final states can be studied at the LHC to improve the sensitivity to

HH production. Those listed above represent nevertheless the decay channels that are expected to be the most sensitive and whose combination can ensure the largest coverage of the possible HH topologies.

In this thesis the HH  $\rightarrow$  bb $\tau\tau$  channel is investigated in three different final states: the fully-hadronic channel HH  $\rightarrow$  bb $\tau_h\tau_h$ , where  $\tau_h$  refers to a  $\tau$  lepton decaying into hadrons plus a  $\nu_\tau$ , and the semileptonic HH  $\rightarrow$  bb $\tau_\ell\tau_h$  where one  $\tau$  lepton decays to a lighter lepton  $e$  or  $\mu$  ( $\tau_\ell$ ) plus neutrinos, while the other decays hadronically ( $\tau_h$ ). The fully-hadronic channel represents the 42% of the Higgs di-tau decay modes, while the semileptonic occurs in the 46% of the cases.

The final state bb $\tau\tau$ , is as a compromise between the decay modes HH  $\rightarrow$  bb $\gamma\gamma$  with a very clean signature but a low branching fraction, and HH  $\rightarrow$  bbbb with the highest branching fraction, but an overwhelming QCD background. Since the tau lepton is unstable and can decay to either leptons or hadrons, the searches must exploit different final states. The main challenge in the reconstruction and identification of these final states arises from the presence of neutrinos, which does not allow for a complete reconstruction of the event, and the large backgrounds requiring dedicated analysis techniques to be suppressed, as discussed in Section 4.4 of Chapter 4. However HH  $\rightarrow$  bb $\tau\tau$ , as pointed out in several studies [55, 56, 57], has the best sensitivity over many resonant and non-resonant HH signal hypotheses. The properties of the bb $\tau\tau$  decay channel and the major background contributions are discussed in Section 4.1 of Chapter 4 and throughout the rest of this thesis.

## 1.4.2 Status of experimental HH searches

Several HH searches for the non-resonant and the resonant production have been performed by both the CMS and ATLAS collaborations at the LHC. Data collected at  $\sqrt{s} = 8$  TeV (Run I) have already been used in several HH final states. Searches of HH on Run II data at  $\sqrt{s} = 13$  TeV, are ongoing, and most of the searches results of Run II cover the 2016 period of data-taking only.

The most recent results in CMS of the HH combination can be found in Ref. [4]. They are a combination of the results from the four final states mentioned in the previous section, only for the GGF production mechanism, using the data collected in 2016 by the CMS experiment at p-p collisions at  $\sqrt{s} = 13$  TeV, corresponding to an integrated luminosity of 35.9 fb $^{-1}$ . The results presented in the following are interpreted using the statistical method described in Section 4.7.

### 1.4.2.1 Non-resonant production

For the non-resonant production the ATLAS Collaboration searched, in addition of the four final states mentioned in the previous section, the di-Higgs final states  $WW^*\gamma\gamma$  and  $WW^*WW^*$ [58].

In Tab. 1.5 the the 95% CL expected and observed limits on the signal strength  $\mu = \sigma_{\text{HH}}/\sigma_{\text{HH}}^{\text{SM}}$  for all the different channels separately, and for the combination result obtained by the two Collaborations, are summarized. The most sensitive channel for ATLAS is  $bb\tau\tau$ , while for CMS it is the  $bb\gamma\gamma$  final state: the differences between the performances of the two experiments in each channel are due to different analysis strategies and optimization of the algorithms used to define the final states. The observed (expected) upper limit on the HH production has been set at 6.9 (10) times the SM by ATLAS, and at 22 (13) by CMS.

Search channel	Collaboration	95% CL observed	Upper Limit expected
bbbb	ATLAS	13	21
	CMS	75	37
$bb\gamma\gamma$	ATLAS	20	26
	CMS	24	19
$bb\tau\tau$	ATLAS	12	15
	CMS	32	25
$bbVV (\ell\nu\ell\nu)$	ATLAS	40	29
	CMS	79	89
$bbWW (\ell\nu qq)$	ATLAS	305	305
	CMS	-	-
$WW\gamma\gamma$	ATLAS	230	160
	CMS	-	-
WWWW	ATLAS	160	120
	CMS	-	-
Combined	ATLAS	6.9	10
	CMS	22	13

Table 1.5: List of HH searches at the LHC based on the p-p collisions data collected by ATLAS and CMS at 13 TeV and corresponding to about  $36 \text{ fb}^{-1}$ . Observed and expected upper limits on the SM HH production cross section are normalized to the SM prediction. The ATLAS search for  $bbVV (\ell\nu\ell\nu)$  is not included in the combination and uses  $139 \text{ fb}^{-1}$  of integrated luminosity [36].

BSM modifications to the Higgs self-coupling modifier  $\kappa_\lambda$  has been also studied by

both Collaborations, leading to an upper limit on the HH production cross section as a function of  $\kappa_\lambda$ , as shown in the Figures 1.17b and 1.17a for the various channels studied, and their combination by ATLAS and CMS respectively. The ATLAS analyses combined set an upper limit of 6.9(10) times the SM prediction for the cross section and the value of  $\kappa_\lambda$ , it is constrained between -5 and 12 by observed data and between -5.8 and 12 through the expected upper limit (Fig. 1.17a). For CMS the observed (expected) combined upper limit is about 22(13) times larger than the SM prediction and the observed (expected) constraint on the trilinear self-coupling is  $-11.8(-7.1) < \kappa_\lambda < 18.8(13.6)$  (Fig. 1.17b).

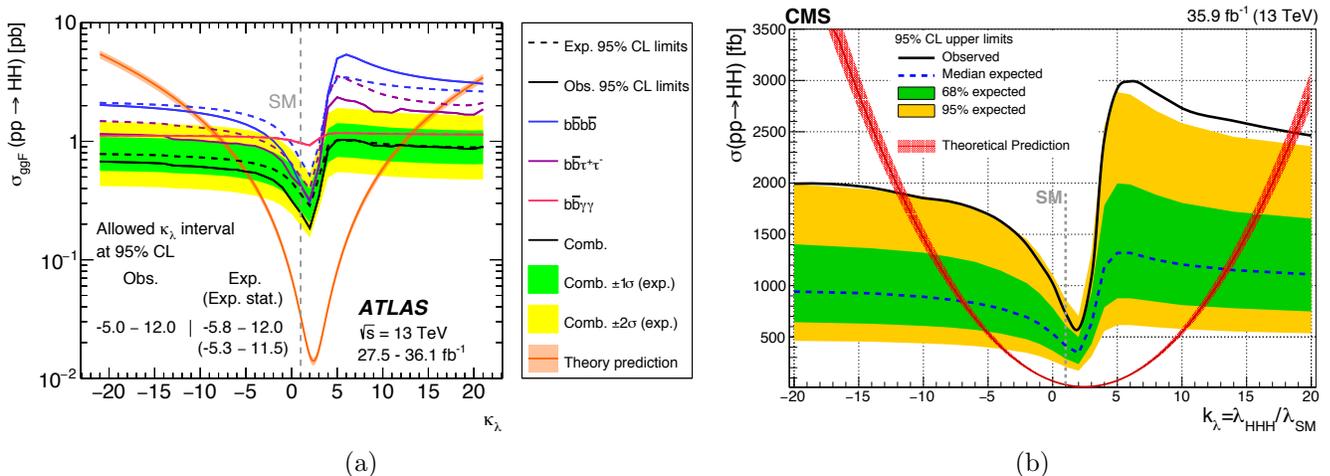


Figure 1.17: Expected and observed 95% CL upper limits on the HH production cross section as a function of  $\kappa_\lambda$  for ATLAS (a) and CMS (b) [58, 4]. The SM expectation and its uncertainty are also reported. All other Higgs boson couplings are set to their SM values [36].

As shown in Section 1.3.2.3, each benchmark scenario predicts a different  $m_{\text{hh}}$  distribution as shown in Fig. 1.15, that affects the signal acceptance and the final discriminant of the analyses determining different sensitivities for different benchmark points. The CMS experiment has adopted this approach and provided the observed and expected exclusion limits on the HH cross section for the different EFT benchmarks, which are shown in Fig. 1.18. These results shows also that sensitivities of the different channels depends on the benchmark scenario.

#### 1.4.2.2 Resonant production

For the resonant production, ATLAS and CMS explored both the spin-0 in the narrow width approximation and the spin-2 resonance hypotheses. Figs. 1.19a and 1.19c show the results for the resonant searches in CMS, for a mass range between 250 and 3000 GeV, in a spin-0 and spin-2 resonance respectively, combining the results from the four channels.

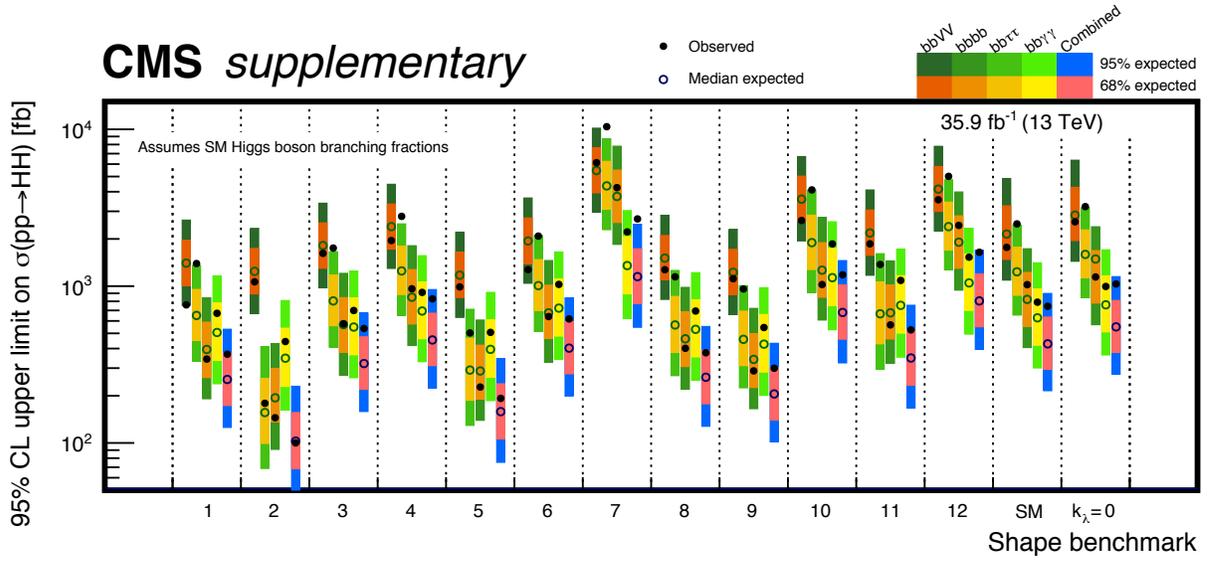


Figure 1.18: The 95% CL upper limits on  $\sigma_{HH}$  for the EFT benchmark topologies, listed in Tab. 1.4, on each bin is shown the limit for each of the 12 benchmark, and in the last 2 bins of the histogram is shown the limits for the SM and box scenarios (represented in Fig. 1.8 as the triangle and box contributions). The green and yellow bands represent the regions containing 68 and 95% deviations respectively from the expected value [4].

Under the hypothesis of a narrow-width resonance, for both spin resonances, was found no significant excess in all the mass range studied.

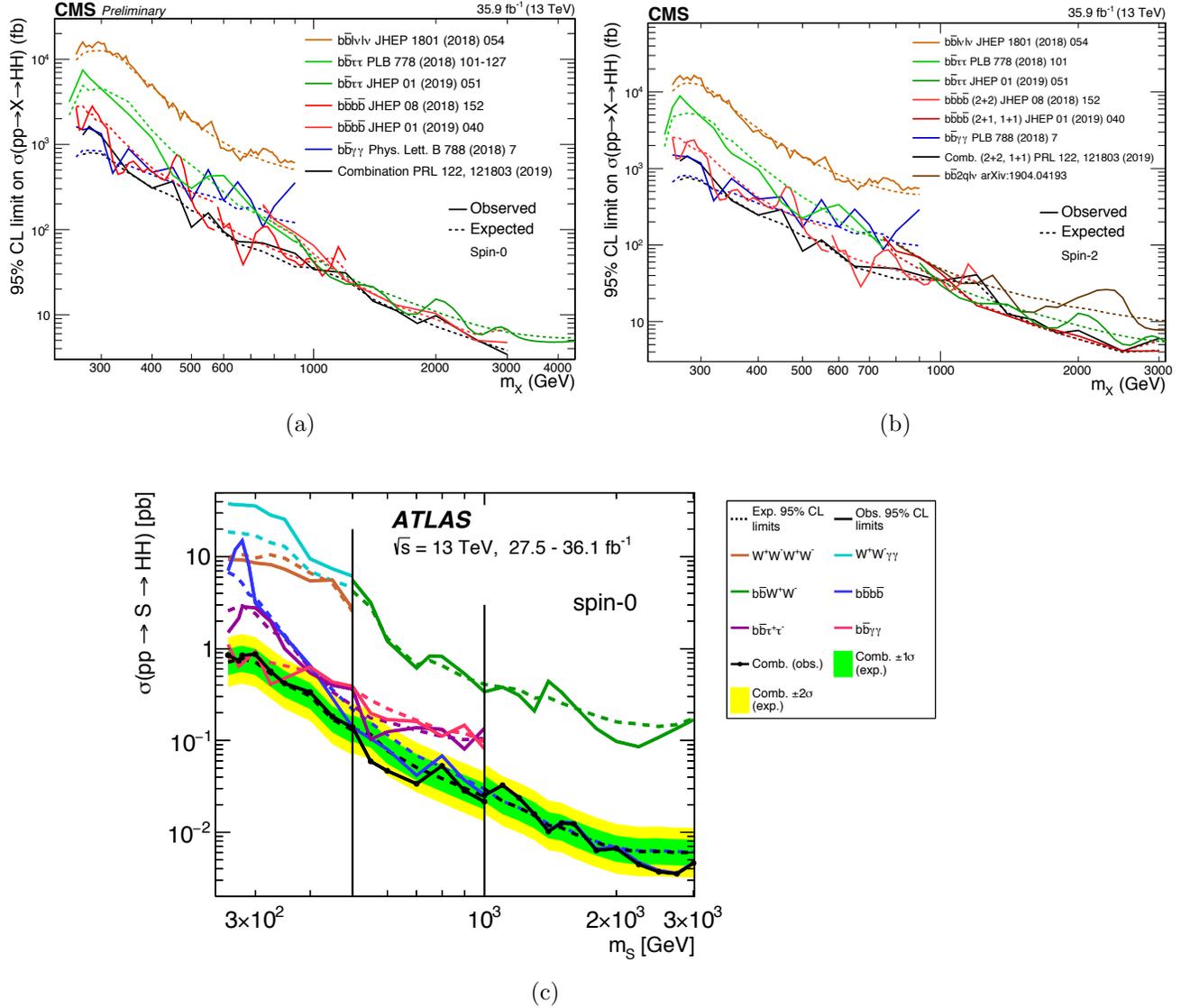


Figure 1.19: Expected and observed 95% CL exclusion limits on the production cross section of a narrow, spin-0 resonance (S or X) decaying into a pair of Higgs bosons (a) ATLAS combination and breakdown by final state for  $m_S < 3$  TeV (b) CMS combination and breakdown by final state for  $m_X < 3$  TeV and (c) spin-2 resonance decaying into a pair of Higgs bosons CMS combination and breakdown by final state [36].

# Chapter 2

## The CMS experiment at LHC

The European Organization for Nuclear Research known as CERN, is one of the largest scientific research centers and hosts the world's largest particle physics laboratory. It was created in 1952 with the idea of establishing a fundamental physics research organization in Europe. At that time, investigations were focused on understanding what was inside the atom, hence the word nuclear. Over the years this initial purpose has changed as there is much more knowledge and new research topics and therefore today the fundamental goal of CERN is to study the constituents fundamental elements of matter and the forces acting between them.

The CERN laboratories host the Large Hadron Collider (LHC), the most large complex and powerful particle accelerator in the world. Inside of the LHC, the particles collide with each others at speed close to the speed of light, and the study of these collisions can give information or clues about how the particles at this scale interact, and provide some light about the fundamental laws of the universe.

The LHC collides bunches of protons in four interaction points, where the main experiments are located: ATLAS, CMS, LHCb and ALICE. The data analyzed in this thesis were collected by the CMS detector.

This chapter describes the experimental framework used to perform this thesis. In the following, a general description of the accelerator system, the CMS detector, and the trigger system used to produce and record the data will be introduced.

### 2.1 The Large Hadron Collider

The LHC is a 27 km diameter tunnel accelerator and proton-proton (p-p) collider. It is located about 50 m to 175 m underground, below the Franco-Swiss border, and it is operated by the CERN laboratories. A beam of protons collide in the center of each detector, where the collisions will be analyzed with different physics purposes, with the main focus

---

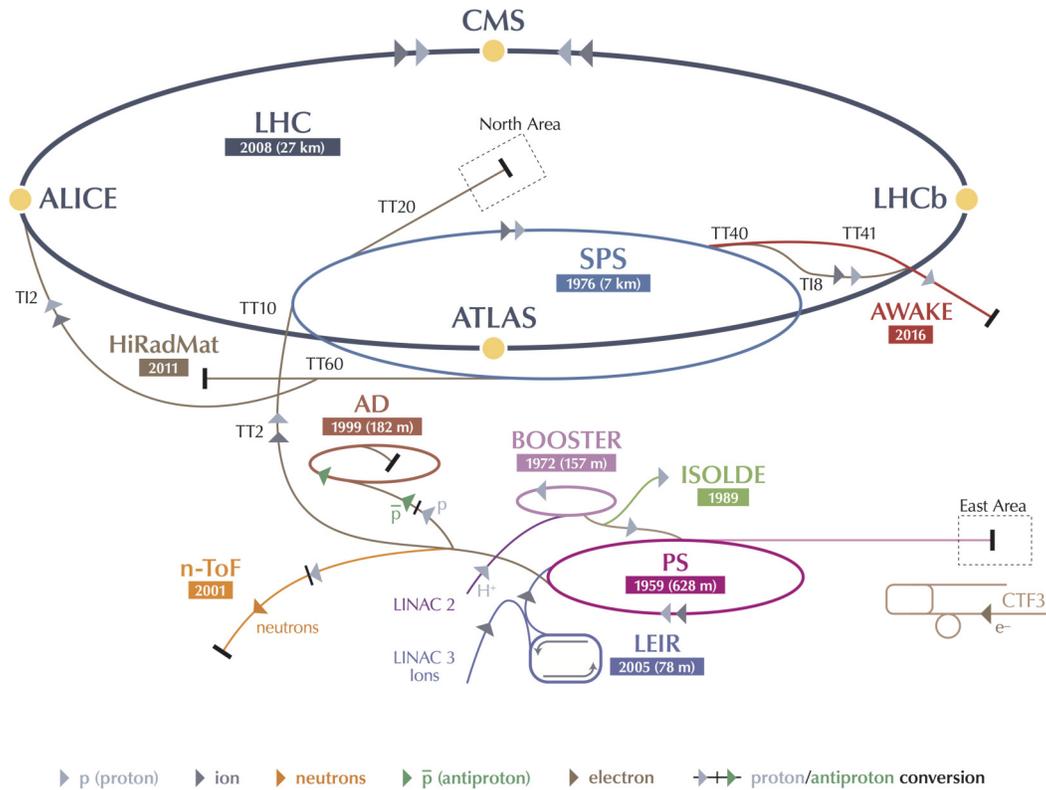


Figure 2.1: Representation of the CERN accelerator complex [60].

on probing the Standard Model theory, to exploring the TeV scale and searching for new physics.

Even though the LHC was initially designed to collide protons, there is also a physics program of heavy ions collisions, which focus on the study of the collective behavior of quarks and gluons in the form of plasma.

Before entering the LHC ring, the proton beams are accelerated, using a long chain of accelerating processes, in order to increase their energy. This process is illustrated in Fig. 2.1. The protons are produced from atoms of hydrogen, excited with the use of a strong magnetic field. After its production, the proton beam enters a linear accelerator called LINAC 2, and at this point the proton beam acquires an energy of around 50 MeV. Then, the proton beam enters the Proton Synchrotron Booster (PSB), the first circular accelerator of this process, where the proton beam accelerates up to 1.4 GeV, and right after enters the second circular accelerator, the Proton Synchrotron (PS), where the proton beam will arrive up to 25 GeV. Finally, the proton beam is directed to the Super Proton Synchrotron (SPS), where after 12 cycles, the beam is accelerated to 450 GeV, and it is ready to be injected into the LHC [59].

The LHC is the last stage of a complex system of accelerators. It is composed of two

rings in which injected protons circulate in opposite directions, with separated magnetic fields and vacuum chambers in the main arcs and with common parts at the interaction points. The original design provides protons that are collected in 2808 bunches composed of  $10^{11}$  protons each and with a time separation of 25 ns. The tunnel is not a perfect ring, but it is composed of eight arcs and eight straight sections. In the arcs are present dipole magnets and focusing quadrupoles. The dipole magnets (in total 1232) guide the protons around the accelerator ring, operating at 1.9 T reaching a maximum of 8.33 T, thanks to their superconductive characteristics. In each straight section there are superconductive radio-frequency cavities which are tuned to oscillate at 400 MHz.

Inside LHC the protons are accelerated up to 7 TeV where, thanks to the radio-frequency cavities (RF), they keep their energy. The RF cavities are also responsible for shaping the beams into proton bunches and for the distribution of the clock to all LHC experiments. The stability of the beams is obtained thanks to quadrupoles magnets (392) focusing the particles in a narrow beam. The protons are packed in bunches distanced in time by  $\Delta t$  and distributed in a structure prepared along the injection chain. A “fill” is complete when the LHC cannot accommodate any more bunches. Special quadrupoles magnets, installed on both sides of the collision points, are used to squeeze the beams and increase the proton density in the collisions. The lifetime of a fill can widely change from a few minutes to about twelve hours; however, in nominal conditions, it is limited by the luminosity lifetime, that decreases in a fill mainly due to beam losses from collisions.

In Tab. 2.1 the nominal beam parameters and their description are summarized.

The crucial parameters useful to understand the performance of a particle collider are the energy available in the center-of mass and luminosity (instantaneous ( $\mathcal{L}$ ) and integrated ( $\mathcal{L}_{int}$ )). The production rate for events of a given physical process ( $dN_{proc}/dt$ ) is proportional to the instantaneous luminosity:

$$\frac{dN_{proc}}{dt} = \mathcal{L} \cdot \sigma_{proc}, \quad (2.1)$$

where  $\sigma_{proc}$  is the cross section of the process in question, which depends on the center of mass energy of the colliding protons,  $\sqrt{s} = 2E$ . The instantaneous luminosity of a particle accelerator depends on its intrinsic features [59]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma F}{4\pi \epsilon_n \beta^*}, \quad (2.2)$$

where  $N_b$  is the number of particles per bunch,  $n_b$  the number of bunches per beam,  $f_{rev}$  the beam revolution frequency,  $\epsilon_n$  the normalized transverse beam emittance and  $\beta^*$  the beta function at the interaction point. The proton relativistic factor, defined as  $\gamma = E/m$ , and  $F$  is the geometric reduction factor of the instantaneous luminosity due to the beams crossing angle  $\theta_c$  and the transverse and longitudinal sizes of the spreads,  $\sigma_{xy}$  and  $\sigma_z$ , of the Beam Spot (BS), the luminous region produced by the collisions of the beams.

Even though the instantaneous luminosity is an important parameter, the definitive value to measure the performance of the accelerator will be given by the so-called integrated luminosity, it provides a measurement of the total amount of produced events of a given process over the total time of activity  $T$ :

$$N_{proc} = \sigma_{proc}(E) \int_0^T \mathcal{L}(t) dt. \quad (2.3)$$

High instantaneous luminosities allows processes with very low cross sections to be produced, making their searches possible.

A complication at high luminosities is the contribute of multiple interactions in the detector signals associated with a single bunch-crossing, defined as pile-up. The pile-up represents for a bunch crossing the average number of simultaneous interactions, and for p-p collisions can be calculated by:

$$\langle \text{PU} \rangle = \frac{\mathcal{L} \cdot \sigma_{pp}^{\text{inel}}}{n_b f_{\text{rev}}}, \quad (2.4)$$

where  $\sigma_{pp}^{\text{inel}}$  represents the cross section of the inelastic p-p collisions, which at  $\sqrt{s} = 13 \text{ TeV}$  has a value of 69 mb [61]. The probability that more than one p-p interaction belonging to the same bunch crossing could produce an interesting process is usually considered to be negligible. Therefore, at the analysis level, only the most energetic p-p collision per event is selected and is referred to as the primary hard interaction, while the other collisions in the event are called pile-up interactions.

The LHC is designed to have a center-of-mass energy  $\sqrt{s} = 14 \text{ TeV}$ , and an instantaneous luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  for p-p collisions.

The beams accelerated inside the LHC collide in four points along its circumference where experiments are located: ALICE (A Large Ion ColliderExperiment), ATLAS (AToroidal Lhc ApparatuS), CMS (Compact Muon Solenoid) and LHCb (Large Hadron Collider beauty), as illustrated in Fig. 2.1. The ATLAS and CMS are multi-purpose detectors, while ALICE and LHCb are focused on more specific studies: ALICE investigates the strong interactions sector of the SM and the quark-gluon plasma physics, and LHCb heavy flavour quarks physics.

The analysis presented in this thesis uses part of the data collected by the CMS experiment during the Run II (defined as the data taking period between 2015-2018) corresponding to an integrated luminosity of  $137.1 \text{ fb}^{-1}$  of p-p collisions at  $\sqrt{s} = 13 \text{ TeV}$  during the period 2016-2018.

Parameter	Description	Nominal value
$\sqrt{s}$	centre-of-mass energy	14 TeV
$\Delta t$	bunch separation	25 ns
$n_b$	bunches per beam	2808
$n_p$	particles per bunch	$1.15 \times 10^{11}$
$f_{\text{rev}}$	revolution frequency	100 kHz
$\epsilon_n$	transverse beam emittance	3.75 $\mu\text{rad}$
$\beta^*$	beta function at the interaction point	0.55 m
$\theta_c$	crossing angle at the interaction point	285 $\mu\text{rad}$
$\sigma_z$	longitudinal r.m.s. bunch size	7.55 cm
$\sigma_{xy}$	transverse r.m.s. bunch size at interaction point	16.7 $\mu\text{m}$

Table 2.1: Summary of the nominal values of the LHC parameters in p-p collisions.

### Operational history

The first collisions in the LHC were made during spring of 2010 with a center-of-mass energy of  $\sqrt{s} = 7$  TeV, and as of today the Run I and Run II phases have been completed, defined respectively as the periods from 2009 to 2013 and from 2015 to 2018. At the end of Run II, the center-of-mass energy was increased to  $\sqrt{s} = 13$  TeV.

The evolution over time of the instantaneous and integrated luminosity recorded by CMS are shown in Figs. 2.2a and 2.2b, respectively.

The distributions of the average number of interactions per crossing, defined as pile-up (Eq. 2.4), for p-p collisions, for the periods of activity 2011-2012 and 2015-2018, are shown in Fig. 2.3. In addition, in the same figure the overall mean values and the minimum bias cross sections are specified [62].

In the period of 2016-2018, part of the Run II, at the LHC the center-of-mass energy was  $\sqrt{s} = 13$  TeV, the data was delivered and collected, with a collision frequency of 40 MHz, and a bunch spacing of 50 ns. In that period, the LHC collected data from April to November of each year, and it was possible to increase the luminosity, overcoming the original LHC design value. As it can be seen in Fig. 2.2a, the measurements of luminosity made by CMS in 2016 achieved the highest values up to that moment ( $1.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ), which was then exceeded in 2017 and 2018, since a maximum value of luminosity was achieved for both years ( $\sim 2.1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ). Similarly, for the integrated luminosity, the measurements made by CMS were  $41 \text{ fb}^{-1}$  in 2016,  $49 \text{ fb}^{-1}$  in 2017 and  $68 \text{ fb}^{-1}$  in 2018, as shown in Fig. 2.2b.

Both the experiments ATLAS and CMS collected and reconstructed in a successful way the majority of the delivered luminosity,  $140 \text{ fb}^{-1}$ , approximately one order of magnitude

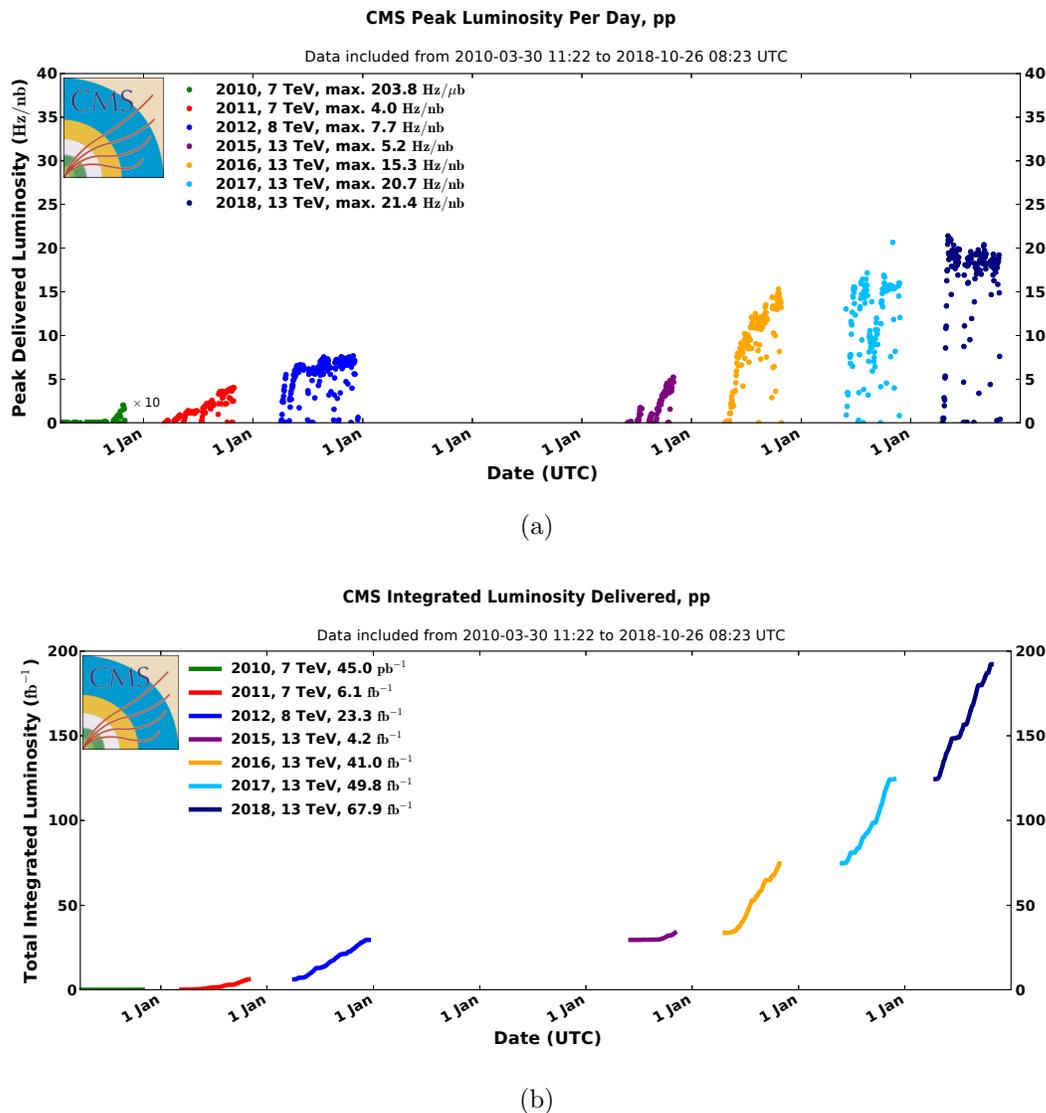


Figure 2.2: (a) Peak instantaneous luminosity (b) total integrated luminosity, recorded by the CMS experiment per year. The instantaneous luminosity is expressed in Hz/nb corresponding to  $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$  [62]

more than in Run I, that can be used for physics analyses.

In the 2018 the Run II of the LHC finished, and the LHC has entered into a phase of maintenance and upgrade, in preparation for the upcoming Run III, when the collider is expected to start collecting data again in 2021 until 2023 or 2024, and to achieve an integrated luminosity of  $300 \text{ fb}^{-1}$  <sup>1</sup>.

<sup>1</sup>Following the COVID-19, the plan for Run III is now to start in 2022 and collect  $200 \text{ fb}^{-1}$

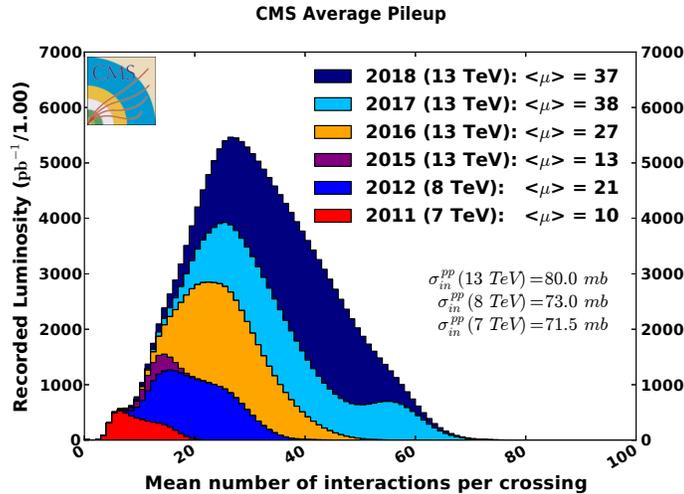


Figure 2.3: Distribution of the average pile-up profile, staked for each data taking year, measured by CMS. The average pile-up is denoted as  $\langle \mu \rangle$  (Eq. 2.4) [62].

## 2.2 The Compact Muon Solenoid experiment at LHC

In this section the main features of the Compact Muon Solenoid (CMS) [63, 64] are described. The CMS detector is designed to study the p-p and nucleon-nucleon scattering physics at the energies and luminosities of the LHC. CMS is situated at Point 5 of the LHC ring, about 100 m underground near Cessy in France, between the Jura mountains and Lake Geneva. It has a cylinder geometry, with two “endcaps” closing the “barrel”. It is 21.6 m long, it has a diameter of 14.6 m, for a total weight of 12 500 t. All components of the CMS detector will be described in the following sections.

### 2.2.1 The CMS coordinate system and useful variables

The CMS coordinate system has its origin in the nominal interaction point, located in the center of the detector. The  $z$ -axis is along the beam line and points toward the Jura mountains. The  $x$ -axis is horizontal and points toward the center of the LHC. The  $y$ -axis is vertical and points upward, perpendicular to the LHC plane. The spherical coordinates set  $(r, \theta, \phi)$  can be obtained. Using the cartesian coordinate set  $(x, y, z)$ ,  $r$  is the distance from the beam axis in the transverse plane  $r = \sqrt{x^2 + y^2}$ ;  $\phi = \tan^{-1} \frac{y}{x}$  is the azimuthal angle measured in the x-y plane in the range  $[-\pi, +\pi]$ ;  $\theta = \tan^{-1} \frac{\sqrt{x^2 + y^2}}{z}$ , is the polar angle measured from  $z$ -axis in the range  $[0, +\pi]$ . The  $x - y$  plane is referred as the transverse plane, which is used to measure important features of the particles, like energy and momentum. This coordinate system is illustrated in Figs. 2.4a and 2.4b .

With the origin of the reference frame at the collision point, in spherical coordinates,

the position of a particle can be described as  $\vec{x} = (r, \theta, \phi)$ , where  $r$  indicates the distance from the beam line,  $\theta$  is the polar angle, the angle from the positive  $z$ -axis to the beam,  $\phi$  is the azimuthal angle, the polar angle in the transverse plane, and measured from the  $x$ -axis relative to the beam axis.

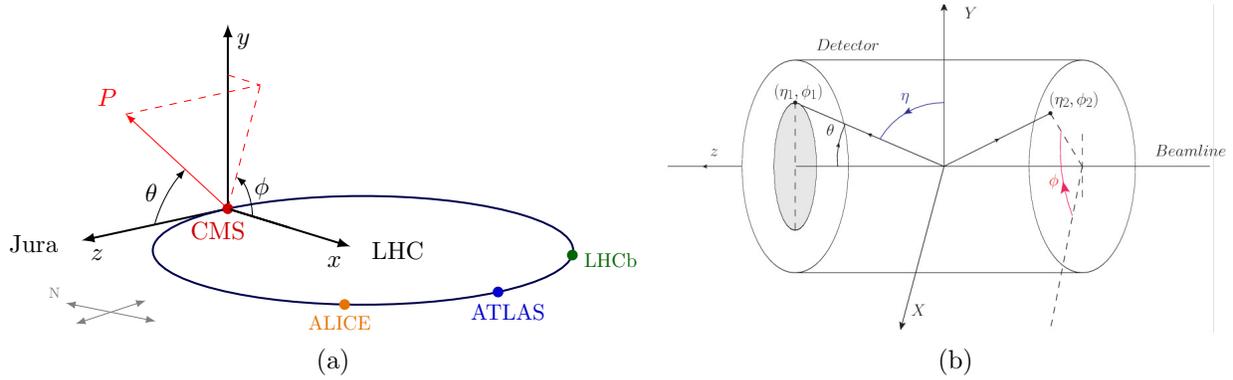


Figure 2.4: Illustration of the CMS coordinate system (a) conventional coordinate system with LHC and other detectors (b) Detector coordinate system.

In order to conveniently describe the particles it is preferable to use quantities that are not distorted by the center-of-mass boost. For example, instead of using  $\theta$  an approximation for ultra-relativistic particles of it, known as pseudorapidity  $\eta$ , is used. The pseudorapidity is the relativistic limit of the rapidity of a particle, and indicates the directions relative to the  $z$ -axis:  $y_z \equiv \tanh^{-1}(p_z/E)$ , which depends on the particle's energy  $E$  and longitudinal momentum  $p_z$ . The difference between the rapidities of two particles,  $\Delta y_z$ , is invariant under a Lorentz boost along the  $z$ -axis. Same for the difference between their pseudorapidities,  $\Delta\eta$ , in the relativistic limit, which is a good approximation to the experimental high energy physics case. Since  $\eta$  does not depend on either  $E$  or  $p_z$ , but only on  $\theta$ , the use of the  $\eta$  coordinate is more advantageous than using  $z$  or  $\theta$ . The pseudo-rapidity is used to identify different detector regions according to their position with respect to the beamline and to the nominal beam collision point.

Since both  $\Delta\phi$  and  $\Delta\eta$  are invariant under a Lorentz boost along the  $z$ -axis, it is useful to introduce a separation in the  $\eta - \phi$  space,  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ . This variable is used to define a  $z$ -boost invariant cone around a single particle or detector position, in order to study nearby detector activity.

In the same way, the momentum and energy of a particle can be described in the transverse plane, for example the transverse momentum is calculated as  $p_T = p \sin\theta$ , with  $\sin\theta = \frac{p_T}{\sqrt{p_x^2 + p_y^2 + p_z^2}}$  where  $p_x, p_y$  and  $p_z$  are the projections of the momentum of a particle  $p$  respectively along the  $x$ -,  $y$ - and  $z$ -axis.

### 2.2.2 Detector

The main features of CMS that allow to meet the requirements of the LHC physics programme are the following:

- good charged-particle  $p_T$  resolution and reconstruction efficiency in the tracker. Pixel detectors very close to the collision point (IP) are required to distinguish different Primary Vertices (PVs) in a high luminosity environment and to distinguish the Secondary Vertices (SVs) present in events with b-jets and hadronic  $\tau$  objects.
- good electromagnetic energy resolution and good electron and photon coverage. This is performed by the electromagnetic calorimeter.
- good hadrons energy resolution and hadron coverage. This is performed by the hadronic calorimeter.
- good missing transverse energy resolution. A calorimeter system with a large geometric coverage and fine segmentation is required, extending the coverage down to  $|\eta| \approx 5$ .
- good muon identification and  $p_T$  resolution.
- The CMS detector can reconstruct muons over a wide range, from  $p_T \approx 1$  GeV in the forward region, up to  $p_T \approx 1$  TeV. This is achieved by the inner tracker and the muon station systems.

Going from the beam collision point outward, the main subdetectors are the silicon tracker, the electromagnetic calorimeter, the hadronic calorimeter and the muon chambers. The detector can be subdivided into the central section, denoted as “barrel”, and two forward regions known as “endcaps”. The subdetectors that compose the CMS experiment are shown in Fig. 2.5. One of the most relevant features of CMS is the magnetic field bending power for charged particles. At the heart of CMS is the superconducting solenoid, 13 m long, 3 m radius, capable of generating a homogeneous 3.8 T magnetic field along the beam direction. The silicon inner tracker and the two calorimeters are all within the solenoid. The muon stations are placed outside the solenoid, in the return yoke, which also contains overall 1.5 m of iron along the radial axis; the bending power for a muon reaching the muon stations is 12 T m.

At the center-of-mass energy of  $\sqrt{s} = 14$  TeV, the total p-p cross-section is roughly 100 mb. With the LHC design luminosity, this implies a rate of  $10^9$  inelastic events/s. This huge amount of data is a challenge for the storage and the computing. Discerning the interesting events for physics searches from all produced events is of capital importance. CMS makes use of hardware triggers installed into the sub-detectors for reducing the

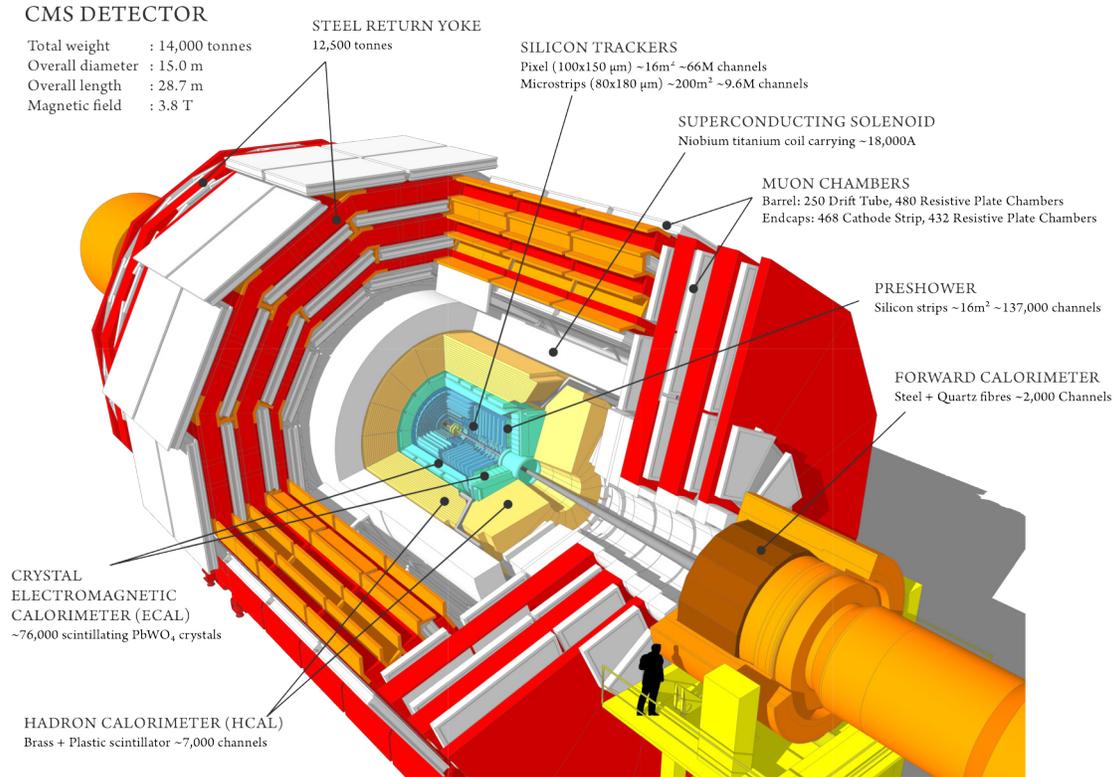


Figure 2.5: Perspective view of the CMS detector and its subdetectors [65].

event rate down to 100 kHz. After that, the events must pass software trigger selections, that reduce further the event rate down to few hundreds per second. Finally, the accepted events can be recorded, processed and stored in sites around the world, ready for physics analyses.

In the following sections a fully description of the subdetectors will be provided.

### 2.2.2.1 The Tracking System

The tracker is the closest subdetector to the interaction point, and it is designed to measure the primary and secondary vertices of the charged particles that cross it, as well as to perform measurements of the trajectories and momentum, thanks to the strong magnetic field in which the detector is embedded. The tracker system is submerged in a 3.8 T magnetic field, which is produced by the solenoid magnet. The efficient reconstruction of the vertices is very important, since allows to distinguish the origin of the hard scatter interaction producing the interesting event, from secondary vertices of in-fly decays of b-hadrons or  $\tau$  leptons, and from additional interactions created by pile-up events. The tracker system allows to detect satisfactorily the trajectories of all charged particles produced during the collision, with  $1 < p_T < 100$  GeV, and  $|\eta| = 2.5$  [66, 67].

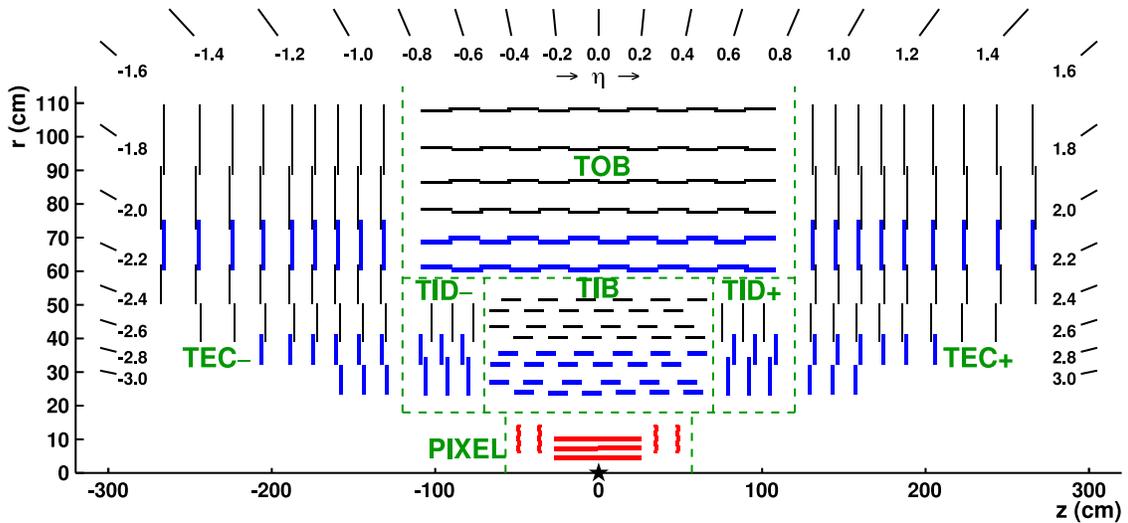


Figure 2.6: Schematic view of the original CMS inner tracking system in the  $r - z$  plane. Only the top half is shown, since the lower half is its mirror reflection about the  $r = 0$  line. The pixel modules are shown by red lines. Back-to-back strip modules, which allow a 3D hit position reconstruction, are shown by blue lines, while the rest of the strip modules are shown by black lines [67].

The main requirements for this detector are a low occupancy and a high detector granularity, in order to be fast enough to take data every 25 ns (40 MHz), the time distance between two collisions. In addition, it is important for the tracker to be light, both in the active and dead material, in order not to alter the trajectories of charged particles and to have a large redundancy to accurately reconstruct the particles trajectories. However, a fast and granular detector needs adequate numbers of readout channels and an efficient cooling, which result in dead material.

Different technologies are employed to satisfy the requirements mentioned above. The best solution was found using silicon detector technology: the tracker is composed of two concentric cylindrical parts, such that they surround the axis of the beam, the inner pixel detector and silicon strip detector. A charge particle passing through a thin layer of silicon cells produces electron-hole pairs that travel thanks to an applied electric field, giving rise to a pulse that is collected and then used to reconstruct the particle position. Silicon sensors cover the region up to  $|\eta| < 2.5$  with a radius  $r < 1.2$  m around the beam pipe and for  $|z| < 2.7$  m, for a total length of 5.8 m and a total surface of  $210 \text{ m}^2$ . The thickness of the silicon sensors crossed by a particle changes as a function of the pseudorapidity, being  $0.35 X_0$  at small  $\eta$ ,  $1.8 X_0$  in the transition region between barrel and endcap, and  $1.1 X_0$  at  $|\eta| \cong 2.5$

The pixel detector provides very low occupancy, high resolution and precise vertex reconstruction thanks to the extremely high granularity, thus they are placed in the

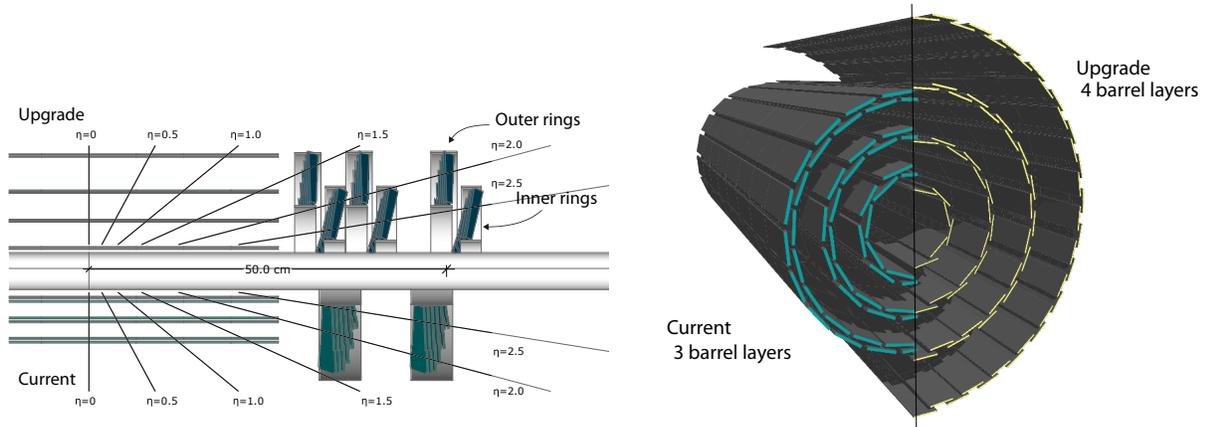


Figure 2.7: Comparison of the new pixel detector, updated during the 2016 technical stop, and the original one. Left: is shown the conceptual layout comparing the different layers and disks in the current and upgrade pixel detectors in the  $r - z$  plane, in top is shown the current updated detector and in the bottom the original. Right: transverse-oblique view comparing the pixel barrel layers in the two detectors. [68].

region closest to the beam pipe. The microstrips that make up the strip detector, cover the more extended region outside the pixels, since they allow to reduce the number of read-out channels, with an occupancy that decreases as the flux of particles with a  $1/r^2$  dependence, maintaining a good resolution. To reduce the radiation damage, the pixel and strip detectors are kept at an operation temperature of about  $-15^\circ\text{C}$  and  $-20^\circ\text{C}$  respectively, with an efficient cooling system that absorbs the heat produced by the on-board electronics.

### 2.2.2.2 The Pixel Detector

The pixel detector is located in the innermost region of the tracker, it covers a region of  $|\eta| = 2.5$ , and in the barrel  $29\text{ mm} < r < 10\text{ cm}$ , which is the area with higher flux of particles ( $\approx 10^7$  particles per second). The original system consisted in three barrel layers and two endcap disks,  $\approx 66$  million pixel cells, each of  $100 \times 150\ \mu\text{m}^2$ , grouped in 1400 sensors for a total surface of  $\approx 1.6\text{ m}^2$ . The barrel region (BPix), counts three layers, each 53 cm long and at a radius  $r = 4.4\text{ cm}$ ,  $r = 7.3\text{ cm}$ ,  $r = 10.2\text{ cm}$  respectively, while two disks constituted each endcap (FPix), made of 24 blades in a turbine-like shape each, at a radius  $r = 7.3\text{ cm}$  and  $r = 15\text{ cm}$  respectively, as schematically shown in in red in Fig. 2.6. Spatial resolutions of  $10\ \mu\text{m}$  in the transverse plane  $r - \phi$  and  $15\ \mu\text{m}$  in the  $z$ -coordinate are achieved in the barrel region, while lower resolutions ( $15\ \mu\text{m}$  and  $20\ \mu\text{m}$  respectively) are achieved in the endcaps [68].

To face the challenging conditions foreseen for 2017 data taking in terms of high rate

of collisions and elevated number of pile-up events, the pixel tracker has been completely replaced during the usual technical stop at the end of 2016, with a completely new system able to sustain an instantaneous luminosity of  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , and over 50 pile-up events [69]. The Fig. 2.7 shows a comparison between the old pixel detector and the new one, which is composed of four disks in the barrel region and three disks in each endcap. The upgraded pixel detector is not only composed of an additional layer per region, but is also closest to the beam pipe (about 3 cm for the barrel section) and designed to be lighter (together with the support and services) than the previous one, reducing the material budget as illustrated in Fig. 2.8; allowing to lower the weight of about 40% in the barrel region and 80% in the endcaps.

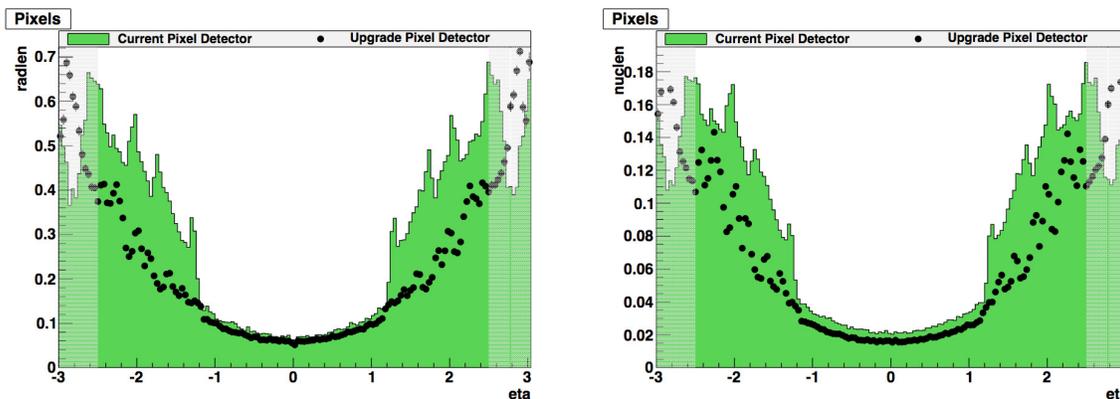


Figure 2.8: The figures shown a comparison between the Phase-0 (green histogram) and Phase-I (black points) upgrade to the tracker system. It is presented the amount of material in the Pixel detector in units of radiation length ( $X_0$ ) (left), and in units of nuclear interaction length (right) as a function of  $\eta$ . The shaded regions at high values of  $\eta$  are outside the tracker acceptance [69].

### 2.2.2.3 The Strip Detector

In the outer region  $20 < r < 110 \text{ cm}$ , where the flux of particles is reduced, a strip silicon detector is used. The strip tracker is divided in four parts (Fig.2.6): Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disk (TID) and Tracker End Cap (TEC); they are composed of single-sided and double-sided microstrip modules (stereo). In order to provide 3D measurements (made in the planes  $r - \phi$  and  $r - z$ ) the double-sided or back-to-back modules are tilted against each other by an angle of 0.1 rad. The inner parts of the strip tracker, TIB and TID, are located in the radial region 20-55 cm with  $|z| < 124 \text{ cm}$  and are composed of 4 layers and  $3 \times 2$  disks, respectively. The TIB and TID use  $320 \mu\text{m}$  thick micro-strip silicon sensors with the pitch size varying for different layers in the range  $[80,120] \mu\text{m}$  for TIB and  $[100,141] \mu\text{m}$  for TID. The resulting TIB spatial resolution is 23-35  $\mu\text{m}$  in the transverse direction. The outer strip tracker comprises 6

TOB layers, which occupy  $55 < r < 116$  cm with  $|z| < 110$  cm, and 9 x 2 TEC disks, which occupy the region  $124 < |z| < 282$  cm, extending the  $\eta$  region covered to  $\eta = 5$ . In the TOB and 3 outer TEC disks 500  $\mu\text{m}$  thick strip silicon sensors are used to improve signal-over-noise ratio, while in the 4 inner TEC disks 320  $\mu\text{m}$  sensors are used, same as TIB and TID. The pitch size varies from 122 to 183  $\mu\text{m}$  for the TOB and from 97 to 184  $\mu\text{m}$  for the TEC. The resulting TOB spatial resolution is 35–53  $\mu\text{m}$  in the transverse direction. In the first two layers of TIB and TOB, and in the transition region between different parts of the tracker, as shown in Figure 2.7, double sided modules are used. Such configuration allows to simultaneously measure transverse and longitudinal hit position. Longitudinal resolution for the stereo modules is 230  $\mu\text{m}$  for the TIB and 530  $\mu\text{m}$  for the TOB. The inner tracker covers the region  $|\eta| < 5$ , providing up to 10 high resolution measurements for charged particles with  $|\eta| < 2.4$ .

#### 2.2.2.4 The calorimeters

The calorimeter system allows to measure the energy and direction of the particles produced during the collision. In order to detect a particle that passes through a calorimeter, the particle must interact with the detector material, and transfer all or part of its energy. The energy deposit left by the particle in the detector will generate a signal that is collected, amplified and processed by means of electronic circuits, and then transferred to a data acquisition and storage system. The signal obtained through the detectors is proportional to the initial amount of energy [70].

When a high energy particle enters the detector produces a shower of secondary particles. Each secondary particle deposits its energy, and additionally produces other particles. This process occurs until a large part of its energy is absorbed, however the particle energy is not always completely absorbed. The composition and dimensions of the shower depends on the type and energy of the particle in question.

There are two types of particle showers that can occur in a calorimeter, an electromagnetic (EM) shower that can be described with the theory of QED, or a hadronic shower, which is more difficult to model, since it also involves nuclear interactions. As a consequence, it is necessary to exploit different calorimeters technologies, to be able to measure the energy of the different types of particles that can be produced in the collision and its decay products.

In the CMS detector, the calorimeters are located outside the tracker and inside the magnetic solenoid, in the inner part is located the electromagnetic calorimeter, and then the hadronic one. In the following an introduction to both calorimeters is presented.

---

### The electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [71] is a homogeneous, hermetic and highly granular calorimeter made of more than 75000 leadtungstate ( $\text{PbWO}_4$ ) crystals. This scintillating material detects the electromagnetic shower produced through Bremsstrahlung and pair production: since the intensity of the emitted light is proportional to the energy absorbed by the crystals, it is possible to measure the energy of incident photons or electrons studying the shower shape.

Since the ECAL is a homogeneous calorimeter, the  $\text{PbWO}_4$  acts as a dense interacting material and as an active scintillating medium at the same time. It was chosen for its high density ( $\rho = 8.28 \text{ g cm}^{-3}$ ), short radiation length ( $X_0 = 0.89 \text{ cm}$ ), small Molière radius <sup>2</sup> (2.2 cm), a fast response (the 80% of the light is emitted in only 25 ns), and resistance to radiations, up to 10 mrad. These characteristics allow the ECAL to be compact, fast, with a fine granularity, and to ensure an excellent containment of the electromagnetic shower within the crystals, which have a length of approximately  $25 X_0$ . ECAL is divided in two regions: barrel ECAL (EB) and endcap ECAL (EE), as shown in Figures 2.9 and 2.10.

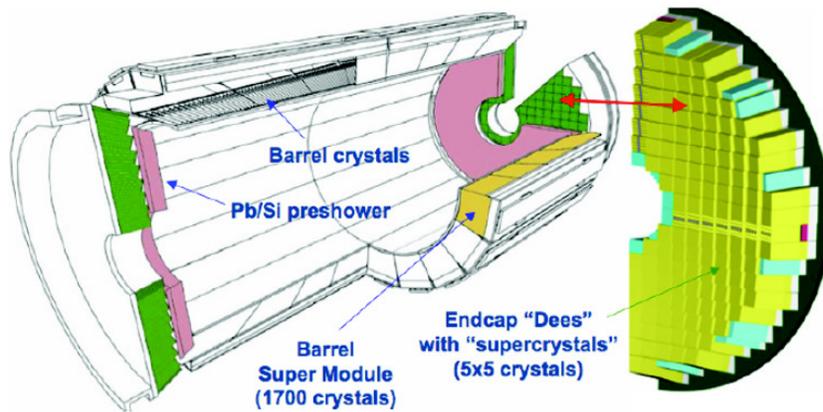


Figure 2.9: Schematic view of the electromagnetic calorimeter.

On the other hand, the  $\text{PbWO}_4$  releases only about 30 photons/MeV and this implies the need to use intrinsic gain photodiodes, able to work in high magnetic fields. Thus, in association with the crystals, silicon avalanche photodiodes in the barrel, and vacuum phototriodes in the endcaps, are used. Signals are amplified by the front-end electronics and sampled at a frequency of 40 MHz with a 12-bit analog- to-digital converter.

The barrel part (EB), with an internal radius of 129 cm, covers a region up to  $|\eta| < 1.479$ . The 61200 crystals of EB have a truncated pyramid shape which varies with  $\eta$  and

<sup>2</sup>The Molière radius  $RM$  is the radius of a cylinder containing the 90% of the electromagnetic showers in the lateral direction. It sets the transverse shower size and gives the lateral deflection of critical energy electrons after traversing one radiation length [72]

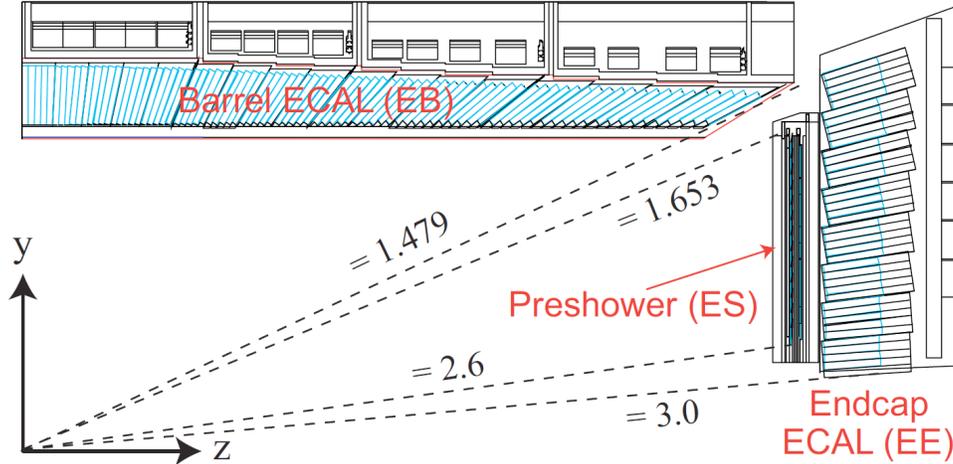


Figure 2.10: Schematic longitudinal layout of one quarter of the electromagnetic calorimeter. The barrel, endcaps and the preshower detector are shown [73].

create a  $\eta-\phi$  grid. Each crystal is long 230 mm with a section towards the interaction point wide  $22 \times 22 \text{ mm}^2$ , with a granularity equal to  $0.0174 \times 0.0174$  in  $\Delta\eta \times \Delta\phi$ , corresponding to an average of  $25.8 X_0$ .

The endcap part (EE) covers a region  $1.479 < |\eta| < 3.0$  and it is distant 314 cm from the nominal interaction point. Each endcap is divided in two parts with a D shape, named “Dees”. The crystals are all identical and have a front area wide  $28.62 \times 28.62 \text{ mm}^2$ , the back area wide  $30 \times 30 \text{ mm}^2$  and they are long 220 mm ( $24.7 X_0$ ). The crystals are grouped in  $5 \times 5$  mechanical units, called supercrystals.

A preshower sampling calorimeter, located in front of the EE calorimeter, provides a good separation between  $\pi_0$  and photons. It is made by two alternate layers of lead and silicon strip detectors. The preshower has a total thickness of 20 cm, corresponding to about  $3 X_0$ , and covers the region  $1.653 < |\eta| < 2.6$ .

Since the average  $X_0$  of the EB and EE are around  $25.8X_0$  and  $24.7X_0$ , the electromagnetic showers formed by electrons and photons of energies up to 1 TeV, are on average 98% contained.

The energy resolution can be parametrized as a function of the energy  $E$  of the incident particles:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.5)$$

where  $S$  is a stochastic term accounting for fluctuations in the number of photo-electrons produced and in the shower-containment,  $N$  is the noise term, due to the electronics and pile-up noise;  $C$  is a constant term, which is related to the calorimeter calibration, and to the energy leakage from the back side of the crystals, which has to be kept lower than

0.55% to profit of the excellent stochastic term provided by  $\text{PbWO}_4$  crystal, relevant for the Higgs searches.

Using beam tests [64], the energy resolution for ECAL up to energies of 500 GeV has been parameterized as follows:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12.0\%}{E}\right)^2 + (0.3\%)^2, \quad (2.6)$$

where E is expressed in GeV.

The radiation has also an impact on crystals, that despite their resistance to radiation, are not insensitive to it. As a consequence, their optical transparency decreases due to ionization radiation and interactions of hadrons with the crystal lattice. The first effect is recovered with time when there are no collisions, but the second one cannot be neutralized. For this reason, a laser light injection system has been used to continuously monitor transparency loss during the data-taking, and time-dependant corrections are computed and applied to maintain a good resolution on measured particle energy. Calibration of the calorimeter is also fundamental, to determine the absolute energy scale and the channel-to-channel intercalibration, that is the relative difference in scintillation light yields (up to  $\sim 15\%$  in EB crystals and  $\sim 25\%$  in EE crystals). Combined corrections computed from laboratory, cosmic-rays and in-situ measurements, lead to a precision of the percent level.

### The hadronic calorimeter

The Hadronic Calorimeter (HCAL) [74] is located beyond ECAL and lies almost entirely inside the magnet. It is fundamental to identify hadrons that pass through ECAL without being stopped.

The HCAL is a hermetic and sampling calorimeter, which means it is not made up of a single material, but rather it is composed by alternating layers of brass plates that play the role of the absorber, and plastic scintillators. It measures hadron energy deposits giving the only measurement in the detector for neutral hadrons, and improving information on charged hadron tracks, in order to reconstruct jets.

It is also important to measure the missing transverse energy (MET), being hermetic up to the maximum  $\eta$  region possible ( $|\eta| = 5$ ), in order to provide the signature for neutrinos and other otherwise undetectable particles. The HCAL surrounds the ECAL, so it receives the hadrons after they pass through the ECAL without depositing their energy completely, since in general hadrons only deposit around 30% in the ECAL. The HCAL is made of layers of absorber material with alternating tiles of plastic scintillators. In the interaction of a hadron with the absorber layer, many secondary particles are produced, which travel through the following absorber layers producing other secondary particles

themselves. The result is a shower of particles that produce light in the interaction with the scintillating material: this light is then readout giving a measurement of the initial particle energy, if the shower is fully contained in the calorimeter. The HCAL can be divided in four regions: barrel (HB), endcap (HE), forward (HF) and outer hadron calorimeters (HO), as shown in Fig. 2.11. The HB and HE of the HCAL are located between the ECAL and the internal surface of the solenoid, instead the HO and HF are outside.

### Barrel and endcap hadron calorimeters

The HB and HE are sampling calorimeters, made of C26000/cartridge brass as absorber material, a metal alloy made of copper (70%) and zinc (30%) and plastic scintillator tiles as active material. The brass was chosen because of its relatively low nuclear interaction length of 16.42 cm and because it is not ferromagnetic, and it is useful to obtain small shower dimensions. The signal from the scintillators is extracted by wavelength shifting fibres (WLS), which deliver light to the optical decoder units, where it is read out by hybrid photodiodes (HPD's). The barrel is divided in two parts made of 18 identical wedges, each of which is segmented in four azimuthal angle ( $\phi$ ) sectors. The plastic scintillator is segmented and readout in  $\eta - \phi$  towers with granularity  $0.087 \times 0.087$ , thus equivalent to the area of a  $5 \times 5$  array of the ECAL crystals. The tiles placed in the same  $\phi$  region are grouped in scintillating units, called tray. The absorber consists of a front steel plate, 40 mm thick, 8 brass plates 50.5 mm thick plus other 6 brass plates 56.5 mm thick and finally a 75 mm thick back steel plate. The effective thickness in units of nuclear absorption length ( $\lambda_l$ ), seen by a hadron entering in the barrel grows as  $1/\sin \theta$ , reaching  $10.6\lambda_l$  at  $|\eta| = 1.3$ .

The endcap calorimeters cover a region  $1.3 < |\eta| < 3$ . It is divided in two sections: eight internal towers with a  $\eta - \phi$  granularity  $0.175 \times 0.175$ , and five external towers with a granularity  $0.087 \times 0.087$ . Its hermetic structure consists of 79 mm thick brass plates with 9 mm gaps to accommodate the scintillators.

### Forward hadron calorimeters

The two forward calorimeters (HF) are placed around the beam-pipe at  $|z| = 11.2$  m and cover the region  $|\eta| < 5$ . Both HFs have a cylindrical structure with an external radius of 130 cm. Placed outside the muon chambers, it extends the coverage in the region  $3 < |\eta| < 5.2$ . It is a sampling calorimeter, whose absorber is made of steel cylindrical plates for a total of 16.5 m thickness along the  $z$ -axis and the active part uses radiation hard quartz fibers, where Cherenkov light is produced and collected by photomultiplier tubes. This configuration has been chosen in order to obtain a detector which resists to the high flux of particles in this region, resulting in a radiation dose that approaches the

100 Mrad/year for the LHC operation luminosity design. The fibers run in a longitudinal direction and are organized to form towers with  $0.175 \times 0.175 \eta - \phi$  granularity.

Because the HF provides good precision for energy measurements in the forward region, it has been used for the CMS luminosity measurements [75].

### Hadron Outer Calorimeter

The outer calorimeter (HO), outside the magnet coil, is added to improve the energy resolution of the barrel calorimeters, catching the tails of the hadron showers that are not fully contained in HB, increasing the total interaction length. It contains 10 mm thick scintillators and it is divided in five  $\eta$  regions, called ring, each covering 2.5 m along the  $z$ -axis. The central ring has two layers of scintillators (at radial distances of 3.82 m and 4.07 m, respectively) divided by a 19.5 cm thick piece of iron (the tail catcher iron). All other rings have a single scintillator layer at a radial distance of 4.07 m. The additional absorber material of the HO corresponds to an interaction length of about  $1.4 \lambda_l / \sin \theta$ . The HO allows to increase the minimal interaction length of the HCAL at  $\eta = 0$  of the HCAL to  $11.8 \lambda_l$ . The depth of the calorimeter is a function of the pseudorapidity, being  $5.25 \lambda_0$  at  $|\eta| = 0$ ,  $9.1 \lambda_0$  at  $|\eta| = 1.3$  and  $10.5 \lambda_0$  at  $|\eta| = 5$ .

The energy resolution of the HCAL in combination with the ECAL is parametrized as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2, \quad (2.7)$$

where  $E$  is the energy of the particle measured in GeV, and the values from this equation were calculated in Ref. [76], from beam test analyses, for the energy of particles in the range 2 - 350 GeV.

#### 2.2.2.5 The Magnet

One of the most distinguishing features of the CMS detector is its magnet, composed of a superconducting solenoid surrounded by an iron yoke. It provides a significant bending power, fundamental to be able to perform precise measurements of the transverse momentum of charged particles, both in the tracker and in the iron yoke. A cryogenic system maintains the magnet at an operating temperature of  $\sim 4.5$  K using flows of liquid helium. The solenoid consists of a winding with 4 layers made of the conductor NbTi. Its length, 12.5 m, provides a homogeneous field over the entire central region of the CMS, as shown in Fig. 2.10. In addition, the cold bore diameter of 6.3 m permits placing the inner tracker system, the ECAL and most of the HCAL, inside the magnet, in order to not disturb the energy measurements.

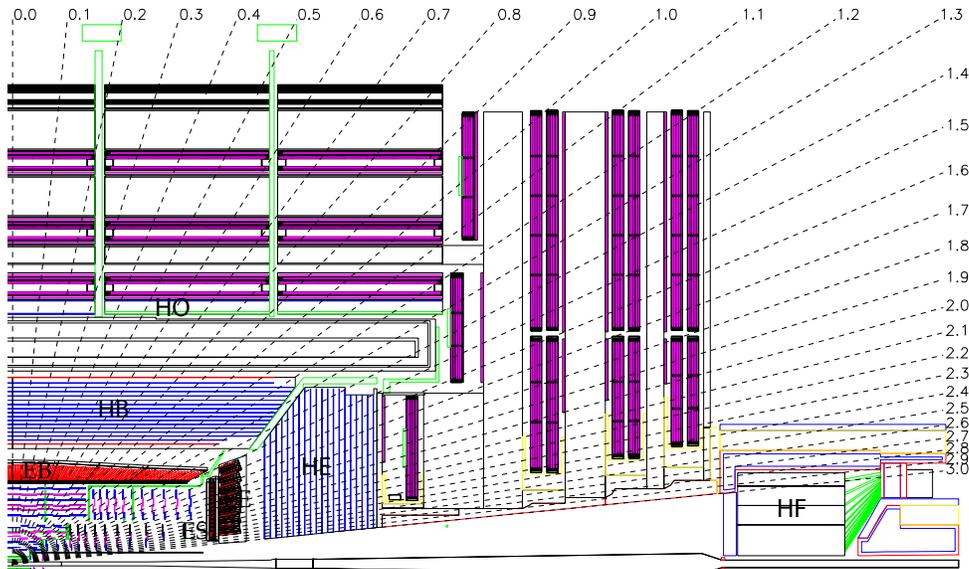


Figure 2.11: Schematic longitudinal layout of one quarter of the hadronic calorimeter. The HCAL comprises the HB, HE, HO and HF [77].

To minimize the radial extent, the magnet has a thin coil design with 0.312 m radial thickness, so that  $\Delta R/R \sim 0.1$ . As a result, during the operation of the magnet, the coil has a very high energy density per unit mass ( $11.6 \text{ kJ kg}^{-1}$ ). To maintain such an energy density without damaging the coil, and to ensure a stable magnet performance, the cryogenic system should fulfill a series of strict requirements, including a maximum temperature margin of 1.8 K with respect to the operating temperature.

The magnetic field loop is closed by a very massive iron yoke of  $10 \times 10^7 \text{ kg}$ , that extends up to 14 m in length and absorbs all particles except for muons and neutrinos. It is composed of 5 barrel wheels and 3 disks for each endcap. Each barrel wheel and endcap disk also hosts muon chambers, as described in Section 2.2.2.6. The return field is  $\approx 2 \text{ T}$ , pointing in the direction opposite to the direction of the field inside the coil. This is used to improve the momentum resolution of the muons, whose trajectories are bent in opposite directions in the tracking region and in the region of the muon system.

### 2.2.2.6 The muon detector

The muons produced from hard interactions are very energetic, thus they will be able to pass through both calorimeters without being absorbed. Therefore, it is necessary to have a special detector to be able to identify the muons and measure their properties.

The CMS muon system [78, 79], shown in Fig. 2.12, is the outer part of the detector and is designed to identify muons, the only charged particles that can fully penetrate the ECAL and HCAL detector layers. Its goals are to measure the transverse momentum and

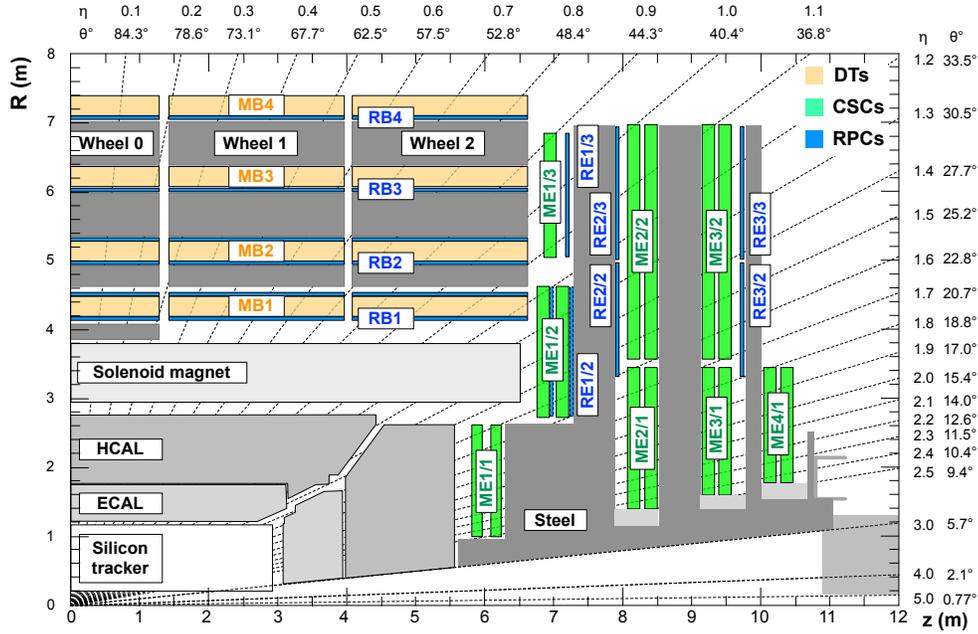


Figure 2.12: Schematic longitudinal layout of one quarter of the muon system in the  $r - z$  plane. There are shown the locations of the various muon stations and the steel disks (dark grey areas). The DT, CSC and RPC are shown in orange, green, and blue respectively[79].

trajectory of muons, thanks to the magnetic field ( $\sim 1.8$  T) created by the return yoke; to complement the inner tracker information (even if muon system standalone measurements are also possible); and thanks to its fast signal, to act as muon trigger. It covers the pseudorapidity region  $|\eta| < 2.4$  and is entirely made of gaseous detectors exploiting different technologies, based on the ionization electrons created by the passage of charged particles in a gas volume immerse in an electric field to produce the signal.

Unlike the previous subdetectors, it is a gaseous particle detector, and is based on three different detector technologies in two regions, chosen accordingly to the expected background rate and the uniformity of the magnetic field in the various regions: drift tubes in the Muon Barrel region (MB) up to  $|\eta| < 1.2$ , cathode strip chambers in the Muon Endcap region (ME) where  $0.9 < |\eta| < 2.4$ , and resistive plate chambers in both MB and ME up to  $|\eta| < 1.6$  to improve redundancy

### Drift tubes Drift tube chambers (DTs)

The DTs are rectangular ( $2 \times 2.5$  m<sup>2</sup>) detectors organized in 5 barrel sections, “wheels”, each consisting in four concentric rings of DT stations (MB1, MB2, MB3 and MB4), divided in 12 contiguous sectors, as shown in Fig. 2.12. For the first three stations, each DT chamber is composed of three Super-Layers (SL), two composed of four layers of drift

cells each, oriented to measure the muon position in the  $r - \phi$  plane, and one SL made of four layers of cells in the orthogonal direction, to measure the  $z$ -coordinate. In the MB4 stations these last layers are missing and the muon position is given in the  $r - \phi$  plane only. The basic element of this detector is thus a DT cell, which has a transverse area of  $4.2 \times 1.3 \text{ cm}^2$  with a  $50 \mu\text{m}$  stainless steel anode wire in the center. It is filled with a mixture of Ar (85%) and  $\text{CO}_2$  (15%) that provides a  $55 \mu\text{m ns}^{-1}$  drift velocity. When a muon passes through the detector, it ionizes the gas producing electrons which travel towards the anode wire: the measurement of the drift time provides the position and incident angle of the muon. The spacial resolution of each cell is  $\simeq 200 \mu\text{m}$ , for a global resolution for the chamber of about 80- 120  $\mu\text{m}$ . Consecutive layers are staggered by half a cell width to improve the coverage and efficiency (which reaches 99.8%) and provide an accurate bunch crossing identification [80].

### Cathode strip chambers (CSCs)

. The two endcap sections allow identifying muons in the region  $0.9 < |\eta| < 2.4$ . In the endcap, where the muon rates are higher than in the barrel, CSCs are used to ensure good performance in the presence of a high background level within a non-uniform magnetic field. The advantages of the CSC are its fine segmentation, fast response time, and radiation hardness. The CSCs are multi-wire proportional chambers with the cathode planes segmented in negatively- charged strips orthogonal to positively-charged anode wires. Thanks to their fan-shape, consisting in trapezoidal panels mounted on eight disks, four in each endcap, partially overlapping in the  $\phi$ -plane to improve the coverage and efficiency, they can easily be arranged in the endcap regions. The CSCs are filled with a mixture of Ar (40%),  $\text{CO}_2$  (50%) and  $\text{CF}_4$  (10%) gas that gets ionized upon the passage of a muon: the signal on wires and strips is interpolated to provide a measurement in the  $r - \phi$  plane (from the wires) and in the  $z$ -direction (thanks to the strip segmentation), with a resolution ranging from 50 to 150  $\mu\text{m}$ . The anodes also provide good temporal resolution, which is used to identify the beamcrossing time of a muon.

### Resistive plate chambers (RPCs)

The RPCs are made of four 2 mm Bakelite planes which form two gaps of 2 mm width filled with a mixture of  $\text{C}_2\text{H}_2\text{F}_4$  (95.2%),  $\text{i-C}_4\text{H}_{10}$  (4.5%) and  $\text{SF}_6$  (0.3%). These detectors are operated in avalanche regime with a high intrinsic time resolution of  $\sim 2 \text{ ns}$ . The avalanche signal produced by the muon ionization is detected by arrays of metallic strips that run parallel to the beam axis, in the outer surface of the gaps, separated from the graphite by an insulating Polyethylene Terephthalate (PET) film. The 4 RPC stations in the barrel region and the 3 RPC stations in the endcaps region provide additional

capabilities for triggering on high  $p_T$  tracks, due to the fast RPC response time, and extra information for tracks ambiguity resolution.

The spatial and time resolutions of the muon system were measured using p-p collisions data collected by the CMS [81]. The average DT chamber resolution are around 80–120  $\mu\text{m}$  and 130–390  $\mu\text{m}$ , respectively in  $r - \phi$  and  $r - z$  directions. The spatial resolution of CSC varies from 58 to 136  $\mu\text{m}$  in the strips local coordinates, corresponding to  $r - \phi$  global coordinates. For the RPC chambers the spatial resolution varies from 0.8 to 1.3 cm, similarly in the strip local coordinates. The time resolution achievable was 3 ns or better per chamber for all 3 systems. This technology allows fast access to hit information from the entire muon detector for use in a hardware muon trigger and to correctly identify the bunch crossing from which a muon originates.

### 2.2.3 The trigger system

In the LHC, at the the design instantaneous luminosity of  $\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$ , and with collisions happening every 25 ns, a rate of  $\sim 10^9$  p-p collisions are produced every second. This interaction rate is orders of magnitude higher than the maximum rate (1 kHz) that any data acquisition system can handle and store. Furthermore, the total p-p cross section (including the elastic, inelastic, and diffractive processes) is about 69 mb while interesting physics signals have a much lower cross-section (as for instance, only about 50 pb for the dominant Higgs boson production mechanism). However, even if the bandwidth to write on tape would be much larger, extracting the most interesting physics events from the large number of normal events would be required to reduce the cost and the time necessary to reconstruct data [63].

This leads to implement a trigger system that preselects events online and decides if the corresponding event information has to be kept or discarded. The trigger system has to be very efficient for the interesting events (the signal events), typically several orders of magnitudes less frequent than the background events.

The selection is made on-line by the CMS trigger and data acquisition system [82, 83] designed to collect and analyze the detector information every 25 ns, that is at every bunch crossing. A multi-level trigger is adopted: the first level, called Level-1 trigger (L1) [82], while the second (L2) and third (L3) levels, referred to as high-level trigger (HLT) [82]. The L1 trigger is hardware based and uses custom electronics to implement in hardware the events selection using information from the calorimeters and muon systems only. The decision on whether to accept or reject an event is made in 3.8  $\mu\text{s}$  from the bunch crossing and the data taking rate is reduced to about 100 kHz. Events accepted by the L1 trigger are read out to the HLT. The complete detector information, including input from the tracker, is processed by complex algorithms running on a farm of 22000 CPU cores, to further select events. The decision on whether to accept or reject an event is made in

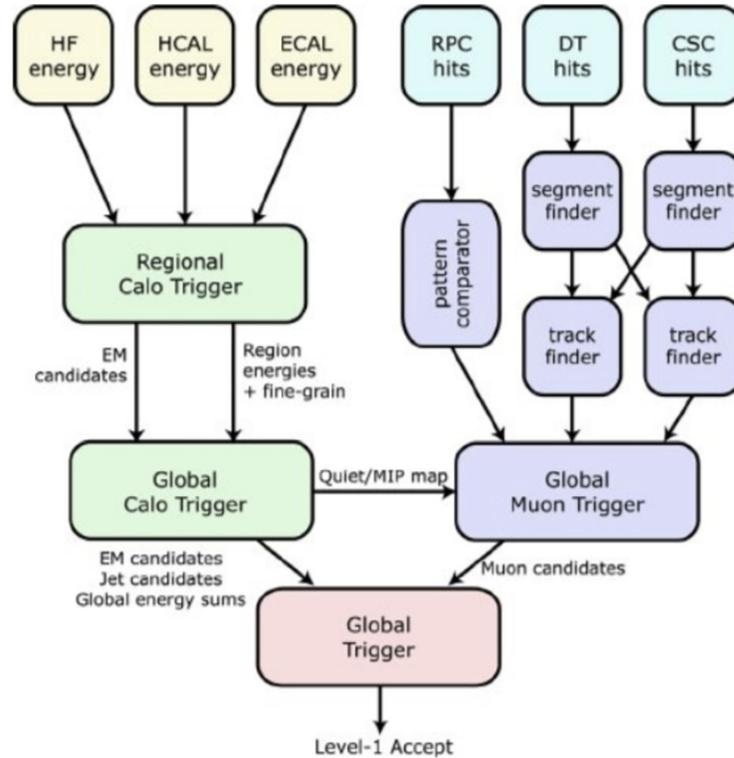


Figure 2.13: Functional block diagram of the CMS L1 trigger.

an average time of about  $220\ \mu\text{s}$  and the trigger rate is reduced below  $100\ \text{kHz}$  with a  $\sim 3\ \text{GB s}^{-1}$  Gbs throughput rate.

In the following a short description of the L1 and HLT is presented.

### 2.2.3.1 Level-1 Trigger (L1)

The first system in the data acquisition chain is the L1 trigger system [82], which is completely hardware-based, and its decisions are based only on the information collected by the calorimetric and muon systems. The L1 system has around  $3.2\ \mu\text{s}$  to decide to keep or discard data from a particular beam crossing.

The L1 trigger is composed by local, regional and global components (Fig. 2.13). During the local step, trigger primitives for each subdetector are generated. The calorimeter trigger primitive combines energies deposited in ECAL - HCAL towers (5 crystals in ECAL correspond to the granularity of HCAL) to obtain the trigger tower. The muon trigger primitives are composed by track segments, for the CSC and  $\phi$ -projection in the DT, or hit patterns, for the RPC and  $\eta$ -projection in the DT. The local information is then combined by the Regional Triggers, where calorimeter electron/photon candidates are defined, and combined DT/CSC tracks are assigned with physical parameters (like for instance  $p_T$ ).

---

In the next step, the Global Calorimeter Trigger (GCT) and Global Muon Trigger (GMT) determines L1 trigger objects based on the combined subdetectors information. The GCT trigger objects include jets and  $\tau$ -jets, the missing transverse energies, isolated and non-isolated  $e/\gamma$  candidates (since no information from the tracker systems is available at this stage, electrons and photons have almost the same experimental signature and are reconstructed as  $e/\gamma$  objects.). Then, in the last step, the L1 Global Trigger (GT) decides to accept or to reject an event based on GCT and GMT trigger objects. The event is taken if it satisfies all the requirements of at least one of the GT algorithms. The GT may execute in parallel  $\sim 130$  algorithms, including both basic algorithms, which consist on some simple  $p_T$  or  $E_T$  thresholds to a single object, and complex algorithms based on topological selections. The result of each algorithm is represented by one bit, which indicates if an event passed the algorithm requirements or not. If the L1 accept decision is made, the entire detector information is readout and is passed to the event builder network (EB), to generate a global event and then it is transferred to the HLT. The sustainable rate of the EB is 1 Tb/s. All L1 triggers that acquire data in a given LHC run are grouped together in a so-called L1 trigger menu.

As the LHC luminosity was being increased, different L1 trigger menus had to be selected, because the total rate must be maintained in a safe regime below the 1 kHz bandwidth. There are two main ways to keep the rate low when the luminosity increases: using more tight selections or using a prescale. In the first case, one can request more trigger objects: for example, instead of asking for one single L1 energy deposit in ECAL, one can request two different L1 deposits; or one can raise the energy threshold, moving from 10 GeV to 25 GeV the minimum jet energy. Another possibility is to prescale the trigger, that is a way that allows to skip triggerable events if the trigger has fired before. For example, if a 10 GeV muon trigger has a prescale of 20, it actually fires only one time every twenty events that contain a muon with  $p_T > 10$  GeV. This allows to keep triggers with low thresholds, to the detriment of the acquired integrated luminosity that, for the aforementioned trigger, would be 1/20 of the total luminosity.

### 2.2.3.2 The High Level Trigger (HLT)

The information from all the subdetectors has to be stored in buffers since the time to analyze one event can take up to few seconds (for events with high pile-up). The HLT system is formed by few hundred different software algorithms, running on every reconstructed event. Each of these algorithms, called HLT paths, are divided into different steps, that are producers or filters. Producers make objects relevant to the given HLT path (for example, one can produce electron objects from the ECAL and the tracker). To optimize the execution time, the HLT does not execute all the paths for every event. Which HLT paths should be executed is determined by the L1 results. One HLT path

---

may be seeded by several L1 algorithms. Filters take decisions on these produced objects (in the same example, a filter will accept events in which there are at least two electrons with  $p_T > 20$  GeV). An HLT path gives a final positive response if the event passes its last step. When one of the filters gives a negative response, the HLT path rejects the event and the subsequent filters and producers of that path are not executed. In this way it is possible to save time and CPU power.

Each trigger path targets a certain event topology, suitable for different studies such as the search for top-quarks, Higgs boson, supersymmetric particles, etc. and defines a sequence of modules which are run sequentially to optimize the computational execution time. The first step, L2, accesses calorimeter and muon system information only; the second step, L3, is conditioned to the L2 decision since it includes the reconstruction of the complete tracks in the tracker, a process that requires a large amount of CPU time. The trigger paths used in this analysis require a central isolated electron or muon candidate and a hadronic tau candidate, and for VBF two additional jets, as described in Section 4.3.

If an event is accepted by the HLT, it is transferred for permanent storage to the CMS Tier-0 computing centre. While the size of one event is about 0.5–1 MB, the total size of the CMS data, including measurements and simulations, is  $O(10)$  PB. Such a huge amount of information is managed by the Worldwide LHC Computing Grid (WLCG), a global computing infrastructure that provides the computing resources to store, distribute, and analyze the data generated by the LHC [84].

---

# Chapter 3

## Event reconstruction and physical objects

This chapter describes the algorithms used to reconstruct the high level detector objects, like calorimeter clusters and vertices; and then how from these the physical objects (such as muons, electrons, taus, jets and b-jets) are reconstructed, which are the input of all the data analyses, included this one.

### 3.1 High level reconstruction objects

Once the proton collision occurs, the detector collects information that will be processed by the trigger system. At the Level 1 trigger (L1), described in Section 2.2.3.1, this information consists of electric signals such as hits in the tracker and muon chambers, or energy deposits in the calorimeters. In the High Level Trigger (HLT), described in Section 2.2.3.2, this information is used to reconstruct and select interesting events at rates dictated by the LHC bunch spacing. The events accepted by the trigger system are stored, and then reprocessed offline, in order to fully exploit CPU time and computation resources, for the more refined algorithms, used to improve the reconstruction of the physical objects, and convert data into a format that can be used for physics analysis. The next two sections are dedicated to a few fundamental parts of the reconstruction algorithms used at the HLT.

#### 3.1.1 The track reconstruction

The track reconstruction is performed by a complex algorithm which estimates the trajectory and the momentum of the charged particles passing through the inner tracking system (Section 2.2.2.1). An accurate track reconstruction is a key point to obtain later

---

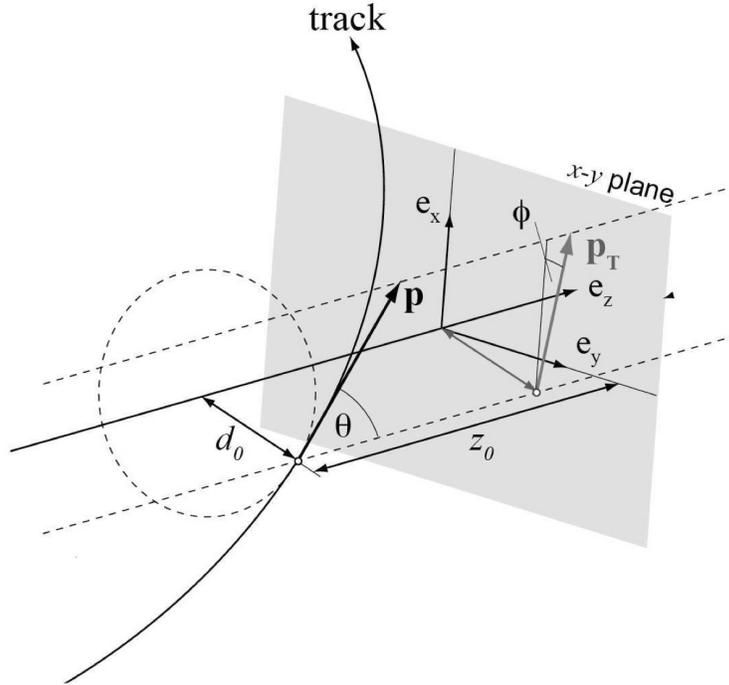


Figure 3.1: Illustration of the parameterization of the track's trajectory at a given point, by its five parameters:  $d_0$ ,  $z_0$ ,  $\phi$ ,  $\theta$  and  $p_T$ .

a high vertex reconstruction efficiency and a precise measurement of the momentum of the particles.

Inside the tracker region, thanks to the almost uniform magnetic field, the trajectory of a charged particle can be parameterized by a helix. The five parameters used by CMS to determine the track's trajectory are:  $d_0$ ,  $z_0$ ,  $\phi$ ,  $\theta$  and the  $p_T$  of the track. The transverse momentum  $p_T$  is defined at the point of coordinates  $(x_0, y_0, z_0)$ , the closest to the nominal beam axis ( $z = 0$ ) point of the track trajectory. The  $\phi$  and  $\theta$  parameters are the polar and the azimuthal angles of the track momentum vector at the impact point, respectively;  $d_0$  is the impact parameter of the track in the transverse plane, and is defined through the coordinates of the impact point  $d_0 = -y_0 \cos \theta + x_0 \sin \theta$ . An illustration of these five parameters is shown in Fig. 3.1.

To estimate these parameters, the spatial positions of the charge deposits (hits) produced by a charged particle when it passes through the layers of silicon detectors are used.

The first step of the track reconstruction is a local reconstruction, where the signals collected by the pixels and strips of the silicon detectors above a specified threshold are grouped together realizing clusters. The hits, are reconstructed from clusters using a specific algorithm for each subdetector [67], with a hit reconstruction efficiency of around 99%.

---

Using these hits, pattern recognition algorithms can reconstruct the original trajectories of the particles, associating a group of hits to a track and measuring the five parameters of the helix that best match to the observed hits. Due to the high particle multiplicity in the tracker, the track reconstruction is a computationally challenging procedure. To reduce the combinatorial complexity, an iterative tracking process is used by CMS, the Combinatorial Track Finder (CTF) [67], an extension of the Kalman Filter (KF) algorithm [85]. The CTF algorithm performs the pattern recognition and the track fitting using an iterative tracking procedure where the first iteration search for tracks with high  $p_T$  coming from the primary vertex. Once found, these tracks are removed from the following iterations to reduce the combinatorial and simplify the subsequent iterations with a complex topology (like lower  $p_T$  or with displaced vertex).

Starting from the reconstructed hits, track reconstruction can be decomposed in four steps:

1. **Seed generation:** the seeds will define the starting trajectory parameters and associated uncertainties of the potential track. The seeds are defined by two or three hits in the pixel detector. They are constructed from the inner part of the tracker and the tracks are built outwards, in order to keep high reconstruction efficiency and to facilitate the reconstruction of low-momentum tracks. All hits must belong to different detector layers.
  2. **Track finding:** this is an iterative algorithm, based on the Kalman filter method [85], that takes into account the information of the successive detection layers and updates the track parameters at each step. It starts from the trajectory seeds, and adds the hits of the outer layers updating the track parameters and their uncertainties in each layer, as well as the material crossed. This procedure is repeated until the final tracker layer is reached. Only a limited number of track candidates are accepted on the basis of number of hits and a minimum  $\chi^2$  criteria. For these tracks a search in the inward direction is then performed to find new hits, and using them to update the trajectory parameters. In order to evade duplicates of the real tracks, it is used an algorithm denoted as trajectory cleaner. This algorithm in case there are several tracks that have in common more than 19% of hits, selects the track with the highest number of hits, and in the case the tracks have the same number of hits, the track with the lowest  $\chi^2$  value is selected.
  3. **Track fitting:** once the tracks are built, to remove possible biases and increase the information of the trajectory, the trajectories are refitted with a Kalman filter and a Runge-Kutta propagator [85, 86]. First the trajectories are refitted using the Kalman filter, which in a iterative way fits the full list of hits, from the inside
-

outwards, updating the track trajectory estimate sequentially. Then, the filter procedure is repeated, initialized with the result of the first one, and performed in the opposite direction. The latter step is known as the smoothing stage, which will help to remove any bias and complete the information of the trajectory. The filtering and smoothing procedure uses a Runge–Kutta propagator, to extrapolate the trajectory from one hit to the next. Thanks to the use of the Runge–Kutta propagator, it is possible to take into account the material effects, and accommodate possible inhomogeneities in the magnetic field, which could cause the particle to not move along a perfect helix. Then, the outliers, defined as the hits incorrectly associated to the track, are withdrawn by applying requirements on the distance between the track candidate and the associated hits on the probability that a pixel hit is consistent with the fitted helix. This procedure is repeated until no outliers are found, filtering and smoothing the track in any iteration.

4. **Track selection:** in the previous stages are created reconstructed track not associated with a charged particle, defined as fake tracks. The fraction of reconstructed tracks that are fake, defined as fake rate, can be reduced by applying certain quality requirements. These requirements are based on the number of layers with hits associated to the track, the track quality ( $\chi^2$  of the final fit), and the compatibility of the tracks respect to one originated from a primary interaction vertex.

In the following a comparison of the performance of the track reconstruction, for the 2016 and 2017 detectors, is presented [67].

The track reconstruction performance can be expressed in term of tracking efficiency (the fraction of simulated charged particles that can be associated to the reconstructed tracks), fake rate (fraction of reconstructed tracks that are not associated with any simulated particle) and track parameters resolution. The performances have been evaluated using  $t\bar{t}$  simulated events corresponding to the 2016-2017 LHC data-taking period.

The efficiency of the tracking is shown in Figs. 3.2a and 3.2c as a function of simulated  $\eta$  and  $p_T$  respectively, of the track, while the rate of fake tracks is shown in Figs. 3.2b and 3.2d has a function of the reconstructed  $\eta$  and  $p_T$  of the fake track. The 2016 tracking performances are compared to the simulation of the detector with the new pixel detector installed in 2017. Only “high purity” tracks are considered for these plots. The efficiency is 95% with the new pixel detector for central  $\eta$ , and an improved efficiency with respect to 2016 is visible in the range  $1.5 < \eta < 2$ . The efficiency is flat and close to 90% for  $1 \text{ GeV} < p_T < 100 \text{ GeV}$ . The fake rate is also lower everywhere in the 2017 simulation. The tracking efficiencies and fake rate were also measured in data with similar results.

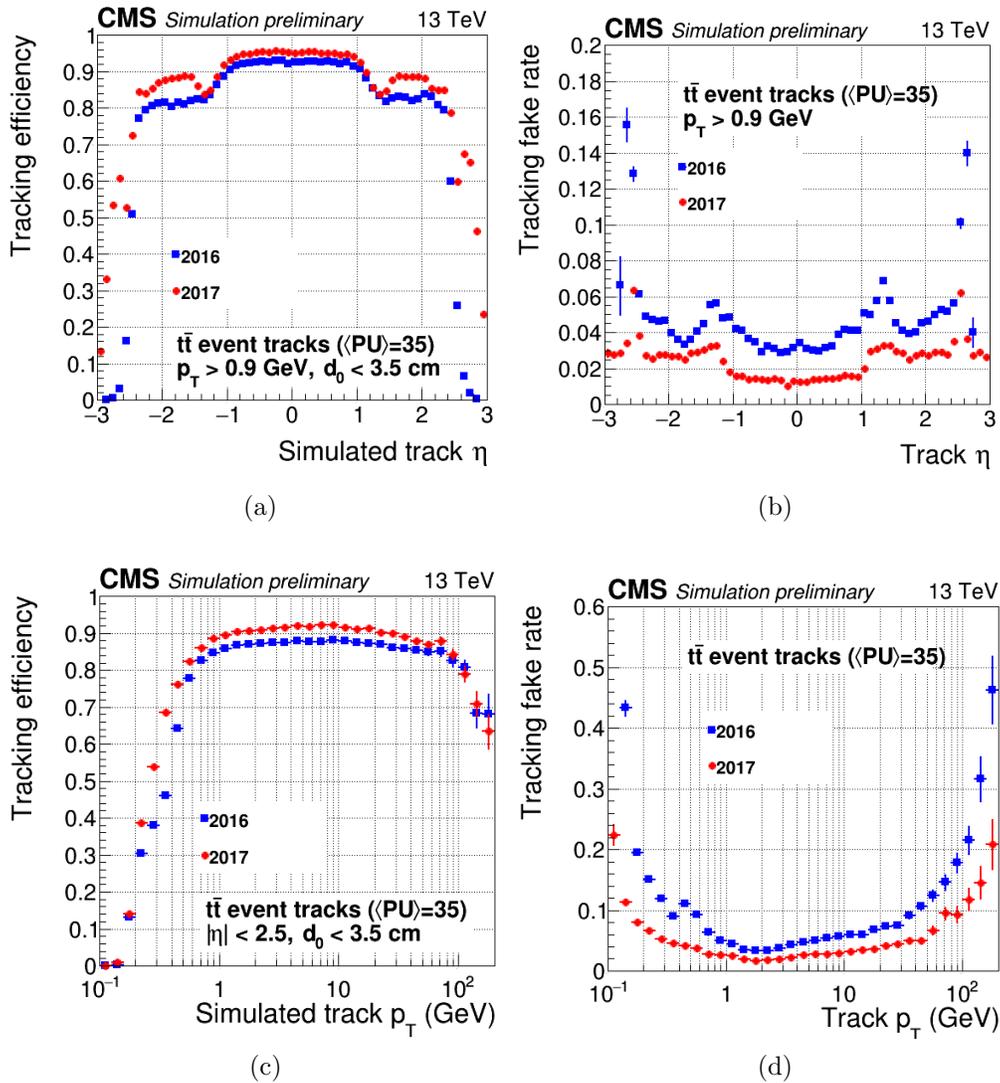


Figure 3.2: Track reconstruction efficiencies and tracking fake rate as a function of  $\eta$  are shown respectively in (a) and (b), and as a function of  $p_T$  respectively in (c) and (d), for the 2016 and 2017 tracker detectors [67].

The resolution of the track parameters is estimated calculating the track residuals defined as the difference between the reconstructed and the true (from MC simulation) track parameters. In Figs. 3.3a and 3.3b are reported the resolutions of transverse impact parameter and of the longitudinal impact parameter respectively. In every bin, the resolution is defined as the half-width of the interval in the distribution of track residuals. The resolution is of 100-200  $\mu\text{m}$  in the transverse impact parameter,  $d_0$ , and 100-500  $\mu\text{m}$  in the longitudinal,  $z_0$ . The impact parameter resolution is largely improved with the new pixel tracker, assuming ideal detector operation, this is crucial for b-tagging and in particular for this analysis where one of the two Higgs bosons produced decays in

a  $b\bar{b}$  pair. The  $p_T$  resolution is 2-4% for  $p_T$  up to one TeV and does not depend as much from the new pixel detector.

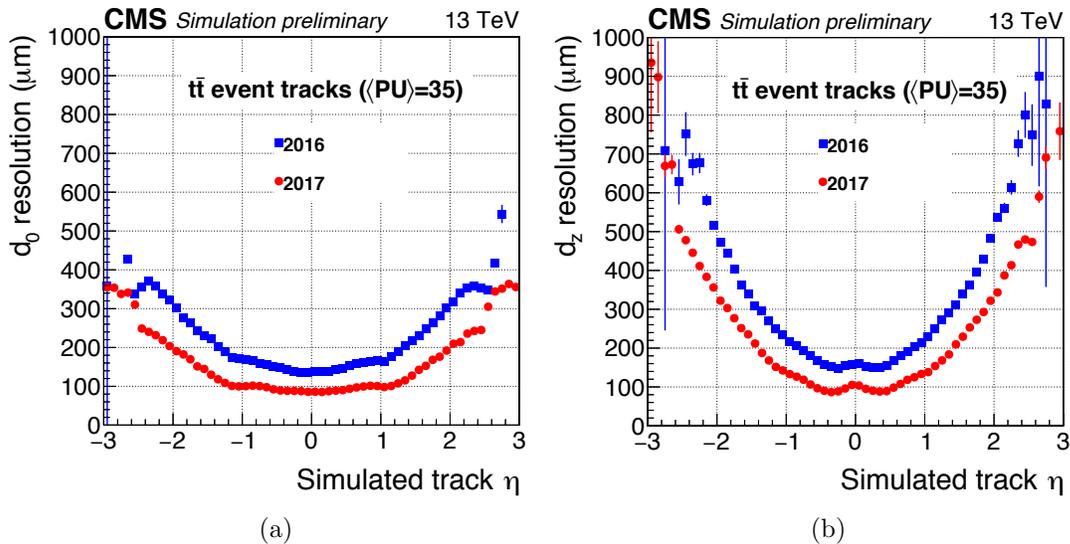


Figure 3.3: Track transverse impact parameter  $d_0$  (a) and longitudinal impact parameter  $d_z$  (b) resolution as a function of  $\eta$ , for the 2016 and 2017 tracker detector [67].

### 3.1.2 Primary vertex

For the data recorded by CMS in 2016, 2017 and 2018, the number of p-p interactions (number of reconstructed vertices) per bunch crossing were reported in Fig. 2.3 and its average value was respectively 27, 38 and 37.

The most probable scenario, in the p-p interactions produced in a bunch crossing, is where only one interesting physical process occurs. However, when the reconstruction is performed several interactions are present, hence it is necessary to separate the hard primary interactions, characterized by a high  $p_T$ , and their secondary vertices, from the interactions coming from the pile-up. In order to achieve this result, it is necessary to have a precise location of the primary vertex, as well the secondary vertex position. Once the vertices are reconstructed, one vertex is selected as the signal vertex, which will be used latter at the analysis level.

The position of each interaction vertex in the event is reconstructed in three steps [87, 67]. Firstly, the tracks satisfying the good quality criteria and no significant displacement from the beam spot selected are clustered in  $z$  using the deterministic annealing (DA) algorithm [88]. Then, the position of each vertex is obtained by fitting the associated tracks with an adaptive vertex fitter [89]. This algorithm is an iterative re-weighted Kalman filter that fits a candidate vertex starting from a collection of tracks. At the

analysis level for Run 2, the primary hard interaction vertex is selected as the vertex with the highest sum of the  $\sum p_T^2$  of all the collision products, as reconstructed using the particle-flow algorithm (Section 3.1.3), and satisfying the following quality requirements: the absolute value of its longitudinal position is  $z < 24$  cm, its transverse distance from the  $z$ -axis,  $d = \sqrt{x^2 + y^2}$ , is less than 24 cm, and on the quality of the fit.

The transverse and the longitudinal primary vertex resolution depends on the number of associated tracks, as shown in Fig. 3.4.

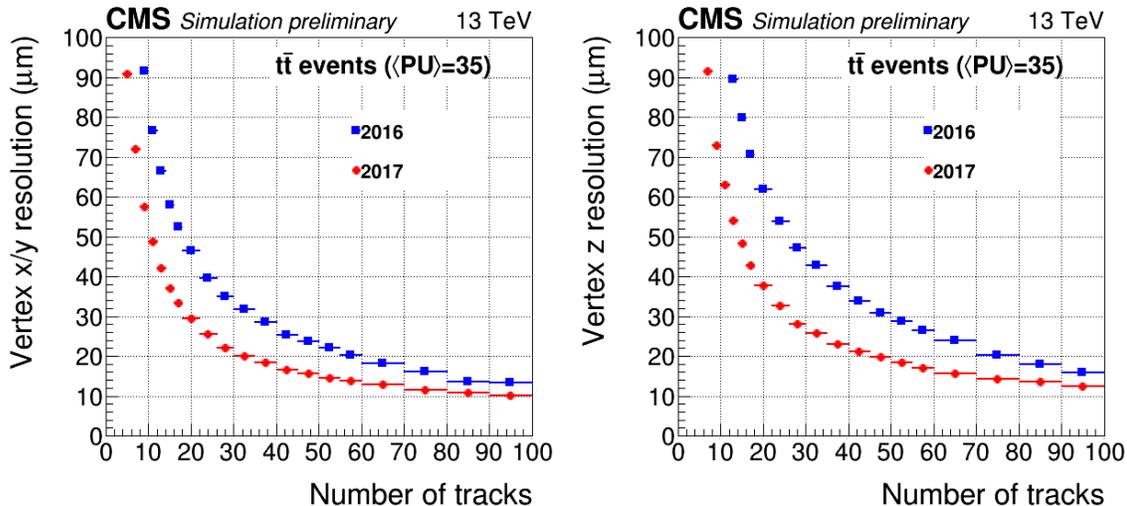


Figure 3.4: Vertex transverse (left) and longitudinal (right) resolution as a function of the number of tracks used in the vertex fit for the 2016 (blue) and 2017 simulated  $t\bar{t}$ bar samples. The 2017 detector shows better performance than 2016 detector [90].

### 3.1.3 The particle-flow algorithm

The strong magnetic field acting on charged particles, the high granularity of the tracker and ECAL detectors, the hermetic HCAL and HF, and the excellent muon system resolution make the CMS detector ideally suited to identify and measure the individual final state particles.

The particle-flow (PF) algorithm [91] is used by CMS to reconstructs all stable particles in the event by combining basic elements from all the subdetectors, performing a global event reconstruction from the raw detector data. The particles reconstructed by the PF algorithm (PF particles) are then used to reconstruct jets, hadronic  $\tau$  decays, b-jet, the missing energy, as well as to determine the isolation of the physical objects (see Section 3.2.2.2 for the definition of “isolation”): the main ingredient of the  $HH \rightarrow b\bar{b}\tau\tau$  search.

The method used by the PF algorithm is based on the typical particle signatures as shown in Fig. 3.5. From this illustration, all charged particles leave a series of hits in the

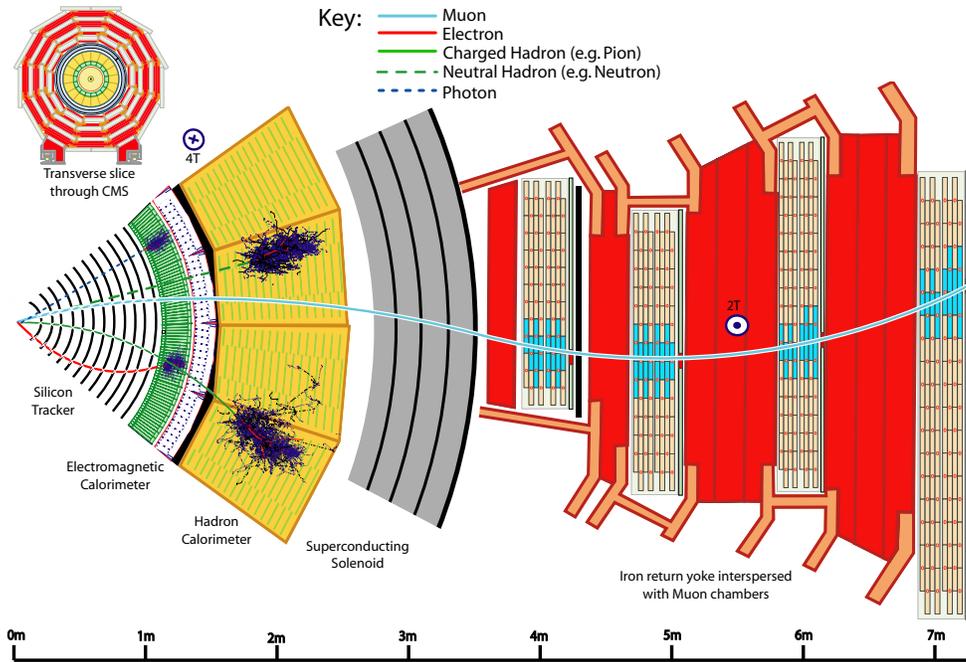


Figure 3.5: Transverse view of the CMS detector, and the experimental signatures left by the particles as they go through the detector.

tracker, which are reconstructed as part of a track; an electron leaves also an energy deposit in the ECAL, while a charged hadron in the HCAL; muons leave a track both in the tracker and in the muon system; neutral particles such as photons and neutral hadrons are reconstructed from an energy deposit in the ECAL and the HCAL respectively, and are not matched to any track in the inner tracker. The sketch of Fig. 3.5 is an ideal situation, that does not occur in real collisions due to the interactions of particles with the detector material, in particular with the tracker, producing for example bremsstrahlung emission from electrons, photon conversion to  $e^-e^+$ , hadron nuclear interaction, or multiple scattering that affects muons. The PF method, uses advanced specific algorithms to recover such effects and to perform the best object reconstruction possible.

The PF algorithm combines together the so called building elements, i.e. charged particle tracks reconstructed by the inner tracker, calorimeter clusters and muon tracks using a linking algorithm, which associates first the objects connected together by a geometrical link, to form a block denoted as PF block. Then, the elements in the PF block are combined to reconstruct and identify the particle candidates (called PR candidates) which are used to reconstruct the higher level objects employed at the analysis level. In the following a short description of the building elements and of the linking algorithm is given.

### Building elements

The Tracker Tracks are reconstructed with the iterative sequence described in Section 3.2.2. Standalone muon track reconstruction in the muon system starts by building track segments in the DTs and CSCs using pattern recognition and linear fit techniques [63]. Next, the track segments and RPC clusters are combined and refitted by a Kalman filter forming standalone muon tracks [64]

The information provided by the calorimeters allows measuring the energy and direction of the photons and stable neutral hadrons, to identify the electrons and the associated bremsstrahlung photons, and improving the energy resolution of the charged particles that have low quality or high  $p_T$  tracks. In order to have a high detection efficiency for low  $p_T$  particles and to split close energy deposits in the calorimeters, a specific clustering algorithm has been designed for the PF algorithm.

This cluster algorithm starts from the calorimetric reconstructed clusters, the cluster seeds, considering the clusters with a maximum local energy. Then topological clusters are built, grouping the neighbors cells with energy above a specific threshold. After collecting all adjacent cells, one topological cluster may contain more than one seed. Therefore, in the last algorithm step, topological clusters are split by sharing the energy deposit of each cell between the PF clusters, whose amount is equal to the number of seeds.

### Linking algorithm

A particle is expected to create several PF building elements as it passes through the detector. Therefore, the first step for the reconstruction of a particle is to connect the different PF elements collected from different subdetectors, this is done with the link algorithm. In order to avoid excessive computing time for the linking part, the elements taken into consideration to link are limited to the closest neighbors in the  $\eta - \phi$  plane. The charged particle tracks, calorimeter clusters, and muon tracks are linked together into the PF blocks. Each block links a group of elements from two sub-detectors, using a criterion to link these elements, and depending on the composition, the block is then interpreted as a particle of a particular type.

Different types of connections between the PF elements are possible, as described in Ref. [91], the most relevant are:

- **track and calorimeter cluster.** The link between a charged particle, which produces a track, and a calorimeter cluster, is allowed if the extrapolated position of the tracks trajectory, from the outermost tracker layer, is within the cluster area, determined by the union of the cells areas in the  $\eta - \phi$  plane for the HCAL and the ECAL barrel, or in the  $x - y$  plane for the ECAL endcaps and the preshower. To take into consideration several sources of errors in the measurements, caused for

example by gaps between calorimeter cells due to the effect of multiple scattering on low-momentum charged particles, the area is enlarged by up to the size of a cell in each direction. The link distance can be defined by the distance between the cluster position in the  $\eta - \phi$  plane, with respect to the position of the extrapolated track. If a track is connected to several clusters, in the ECAL or HCAL, the link with the smallest distance is chosen.

Additional links between charged particle tracks and calorimeter clusters are established to take into account bremsstrahlung by electrons. In this case, tangents to the tracks are extrapolated to the ECAL from the intersection points between the track and each of the tracker layers. If the extrapolated tangent position is within the boundaries of a cluster, as defined above, then the cluster is linked to the track as a potential bremsstrahlung photon.

- **calorimeter cluster-to-cluster.** A link between the ECAL and HCAL clusters is created if the position of the ECAL cluster, which have higher granularity, is within the HCAL cluster envelope. A similar logic is followed to link the preshower and ECAL clusters. As before, if multiple links are created, the one with the shortest distance in the  $(\eta, \phi)$  plane is selected.
- **track and muon track.** A link is established if the  $\chi^2$  of the global fit of the two tracks is less than  $\chi_{max}^2$ . If more than one combination of tracks satisfy this criterion, the combination with the smallest  $\chi^2$  is chosen. This link is named as a global muon and will be described in detail in 3.2.2 .

## Reconstruction and identification of the PF particles

In the final step, the PF algorithm reconstructs and identifies a set of particles using the block of elements. The PF particles are created in several iterations, on each iteration blocks assigned to a PF particle are excluded in the subsequent iterations.

A schematic view of the PF algorithm principle is reported in Fig. 3.5. The procedure for reconstruction and identification of a particle, or an ensemble of particles (hadrons), from each block can be summarized in the following way:

- In the first iteration, the PF muons are created from the global muons (Section 3.2.2), requiring that the combined global muon momentum is within three sigma of the momentum estimated using only the inner tracker. In the following iterations, the PF elements, tracks and clusters belonging to the muon are removed from the PF blocks.
- The electrons reconstruction can be summarized into two steps. First each track is submitted to a pre-identification step considering the tracker as a pre-shower,

since real electrons often emit bremsstrahlung photons and photons often convert to  $e^+e^-$  pairs, which in turn emit bremsstrahlung photons, etc. For this reason, the basic properties and the technical issues to be solved for the tracking and the energy deposition patterns of electrons and photons are similar. The energy of the pre-identified electrons is measured using a combination of the energy collected in the ECAL and a refit of the track, performed with a Gaussian–Sum Filter (GSF) [92]. In the second step, a multivariate discriminator assesses the compatibility with the linked ECAL clusters. This provides discrimination against charged hadrons. If the candidate passes, a PF electron is formed from the GSF track and linked to ECAL clusters, which are also removed from the block. More details about the electron reconstruction and identification is in Section 3.2.3.

- Tracks classified as neither electrons nor muons result in a charged hadron. The mass assigned is the pion mass and the momentum is from the track fit. If the energy of the linked calorimeter clusters is compatible with this momentum within uncertainties, the candidate momentum is updated to a weighted average between the track and cluster measurements, which allows improving the energy resolution for high- $p_T$  and large- $\eta$  particles. When, on the contrary, the combined calorimeter cluster energy is significantly larger than the sum of the momenta of the associated tracks, in addition to the PF charged hadrons, first a PF photon is created with an energy that correspond to the total energy excess in the ECAL and then if the remaining energy excess is larger than the calorimeter energy resolution, a PF neutral hadron is also created.
- All the not linked ECAL and HCAL clusters produce PF photons and PF neutral hadrons, respectively.

The global event description ends when all the defined blocks have been processed and the particles identified.

## 3.2 Physics object reconstruction

### 3.2.1 Muon and electron isolation

In this thesis, the final state explored includes a tau decaying leptonically in electron or muon plus neutrinos. Since it is impossible to separate the  $\tau$  signature from the electron and muon, in CMS the  $\tau$  decay in electron and neutrinos, indicated with  $\tau_e$ , is reconstructed as an electron (Section 3.2.3), while the  $\tau$  decay in muon and neutrinos, indicated with  $\tau_\mu$ , is reconstructed as a muon (Section 3.2.2)

In order to select the leptons with high efficiency and purity, and reduce the contamination of leptons originated from b- or c-quark decays within jets or decays in flight, the muons and electrons candidates are required to be isolated. This implies a small activity of particles around the lepton direction in the tracker and in the calorimeter. The variable used to define the relative lepton isolation,  $I_{\text{PR, rel}}^\ell$ , of the PF lepton candidate, is defined as the sum of the transverse momenta of the PF particles inside a  $\Delta R$  cone (0.3 for electrons and 0.4 for muons) around the lepton relative to the transverse momentum  $p_T^\ell$  of the lepton:

$$I_{\text{PR, rel}}^\ell = \sum p_T^{\text{charged}} + \max [ 0, \sum p_T^{\text{neutral had}} + \sum p_T^\gamma - p_T^{\text{pile-up}} ] / p_T^\ell \quad (3.1)$$

where  $p_T^{\text{charged}}$ ,  $p_T^{\text{neutral had}}$ , and  $p_T^\gamma$  are the scalar sum of the transverse momenta of the charged hadrons originating from the primary vertex, neutral hadrons and photons, respectively; the  $p_T^{\text{pile-up}}$  is the sum of the transverse momenta of charged hadrons not originating from the primary vertex.

### 3.2.2 Muons

The muons within the pseudorapidity range  $|\eta| < 2.4$  are detected by the CMS muon system, described in Section 2.2.2.6. All other particles except neutrinos are expected to be absorbed by the upstream calorimeters and iron yoke plates in the muon system.

As described in Section 3.1.3, the muon identification and reconstruction is the first step of the PF algorithm, since the muon identification is based on the muon chambers. Muon are reconstructed with dedicated algorithms combining information from muon chambers and tracker. The three different reconstruction strategies can be summarized as [93]:

- **Standalone Muon reconstruction** (outside-in). The standalone-muon tracks are reconstructed only with the information from the three muon subdetectors. The track segments are formed by clustering hits within the DT or CSC and they are used to seed the pattern recognition step. Matched hits and segments in DTs, CSCs, and RPCs are then fitted using a Kalman filter to build the standalone muon tracks.
- **Tracker muon reconstruction** (inside-out). All the tracker tracks are propagated to the muon system (DT or CSCs). Only tracker tracks with  $p_T > 0.5$  GeV and total momentum  $p > 2.5$  GeV are used in the extrapolation. The extrapolated positions are matched to the local track segments in the muon system based on the distances in the  $r - \phi$  plane. A tracker-muon is form if at least one of the extrapolated track matches to at least one muon segment

- **Global muon.** Tracks are obtained by matching each standalone-muon track to tracker-tracks. The track parameters of both tracks are propagated to a common surface, and a match is found if the track parameters are compatible. The hits of the matched standalone muon track and the inner track are then combined and fitted using a Kalman filter to build the global muon.

Around 99% of the muons are reconstructed as a tracker- or a global-muon, and in most of the cases as both. However, the tracker-muon has a higher reconstruction efficiency than the global-muon at low momenta ( $p_T < 5 \text{ GeV}$ ), given that the tracker muon reconstruction requires only a single muon segment in the muon system, while the global-muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station, with  $p_T > 200 \text{ GeV}$ , and typically requires segments in at least two muon stations. The muon candidates found by both approaches that share the same tracker track are merged into a single candidate.

### 3.2.2.1 Identification

A set of criteria based on PF is applied to identify muons. Three PF identifications, Loose, Medium and Tight muon ID are provided [81]. The Loose muon ID requires that muons are identified by the PF algorithm, and are either tracker or global muons. It aims to distinguish genuine muons, either prompt ones produced at the primary vertex or the non-prompt ones from hadron in-flight decay, against charged hadrons that are not stopped in the calorimeters and are misidentified as muons. The Medium muon ID adds additional requirements on top of the Loose ID in order to only identify prompt muons or muons from heavy flavor decay. The Tight muon ID aims to suppress muons from decay in flight, and from hadronic punch-through.

The Medium and Tight muon identification correspond respectively to efficiencies of  $\sim 99.7$ ,  $98.5$  and  $97\%$  in data [81]. “Loose muons” are selected by the PF algorithm and are also a tracker- or a global-muon. “Medium muons” are Loose muons with a tracker track that uses hits from more than 80% of the inner tracker layers it traverses. Different selections based on the fit  $\chi^2$  and on the compatibility between the tracker track and the muon track or segment are applied for tracker and standalone muons. “Tight muons” are loose global muons with a tracker track that uses hits from at least six layers of the inner tracker, including at least one pixel hit, and a segment matching in at least two of the muon stations. Other criteria applied is based on the global muon track fit  $\chi^2$ , and on the compatibility with the primary vertex. These criteria suppress punch-through charged hadrons and muons produced in flight. The Tight muon identification is therefore specialized in prompt muons, while the Medium identification is used both for prompt and muons in jets from heavy flavor decays.

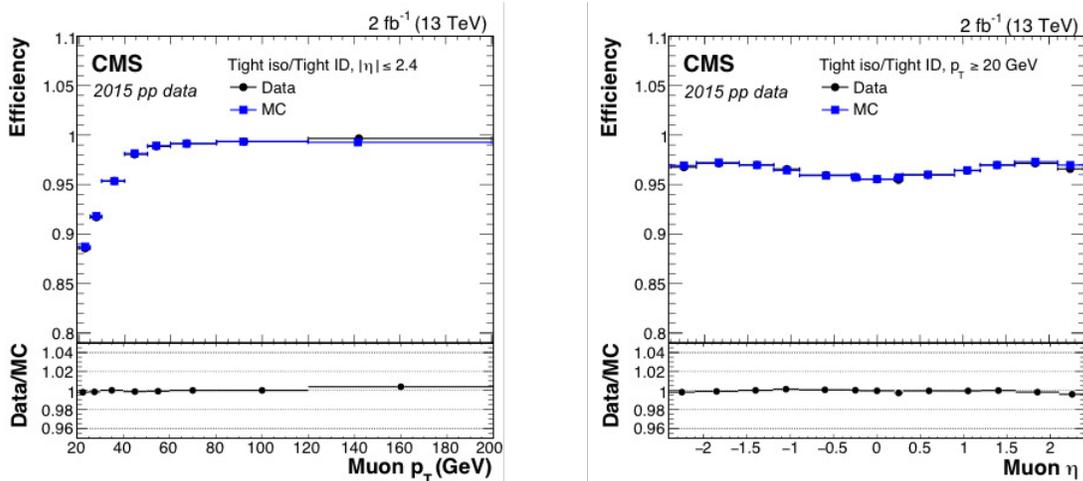


Figure 3.6: Muon reconstruction and identification efficiency measured by tag and probe method in  $Z \rightarrow \mu^+ \mu^-$  data and simulated events. Tight PF Isolation working point on top of the Tight ID (left) versus  $p_T$  for muons in the acceptance of the muon spectrometer, and (right) versus pseudorapidity for muons with  $p_T > 20$  GeV, for 2015 data (circles), simulation (squares), and the ratio (bottom inset). The statistical uncertainties are smaller than the symbols used to display the measurements [81].

In this analysis, the signal muon candidates are prompt muon and required to pass the Tight ID, while for veto muons the Medium or Tight ID is used.

### 3.2.2.2 Isolation

In this analysis, the signal muon candidates are required to have a relative isolation, defined in Eq. 3.1,  $I_{\text{PR, rel}}^\ell < 0.15$  (Tight Isolated), while for veto muons  $I_{\text{PR, rel}}^\ell < 0.3$  is required.

### 3.2.3 Electrons

The electrons deposit most of their energy in the ECAL, but they also leave hits in the inner tracker. The electron reconstruction is based on an algorithm that combines the inputs from both subsystems. The algorithm associates a reconstructed track with a cluster of energy in the ECAL and exploits both sides of the information to estimate the electron momentum. The electron reconstruction is complicated by their interaction with the inner tracker material located between the collision point and the ECAL. This causes significant bremsstrahlung along the electron trajectory, with the resulting photons possibly converting to electron pairs.

Since the electron energy usually spreads over several crystals of the ECAL, the first reconstruction step consists in the clustering of the energy deposits. Two ECAL clustering

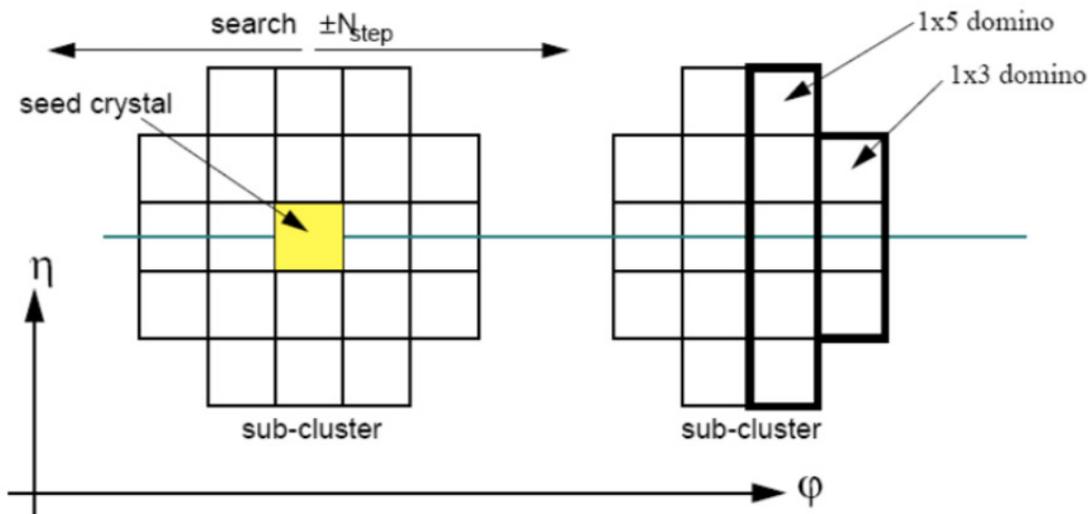


Figure 3.7: Domino construction step of Hybrid algorithm used for the electron reconstruction [94].

techniques are used, depending on the region of the ECAL: a hybrid and a multi- $5 \times 5$  algorithm. The first is used in the ECAL barrel, and the second one is used in the ECAL endcaps. Both techniques take into account the fact that radiated photons are mostly spread in the  $\phi$ -direction, while their spread in the  $\eta$ -direction is very small [94, 95].

**Barrel Hybrid algorithm.** In this case crystals are arranged in  $\eta - \phi$  directions. The algorithm starts from crystals with transverse energy deposit  $E_T$ <sup>1</sup> above a threshold and considers as a seed the crystal with the highest transverse energy. Then clusters are formed starting from crystal seeds by adding arrays of  $5 \times 1$   $\eta - \phi$  crystals in both  $\phi$ -directions. The energy threshold below which crystals are not clustered is 100 MeV. Then the reconstructed clusters belonging to the same region are grouped into a Super Cluster (SC). The idea of the hybrid algorithm is illustrated in Fig. 3.7

**End Cap Multi- $5 \times 5$  algorithm.** The crystals are not arranged in the  $\eta - \phi$  directions. The clusters are formed starting from seeds by adding to them  $5 \times 5$  neighboring crystal arrays with possible overlaps. In the case the total energy of the nearby clusters are above a certain threshold, the clusters are grouped into a SC. The SC energy is calculated as the sum of the energies of the clusters that are part of it. The position of the SC is defined as the energy-weighted mean of the cluster positions, which represents the impact point in the ECAL of a non-radiating electron, and has the same energy as the supercluster.

Dedicated track reconstruction for electrons is initialized from seeds that are likely

---

<sup>1</sup> $E_T$  is the transverse energy, defined as  $E_T = E \sin \theta$ , where  $E$  is the deposit of energy in the calorimeter cell and  $\theta$  is the polar angle position of the cell

from electron trajectories. Such seeds are found by two complementary algorithms based on either the ECAL or the tracker. The ECAL-based seeding uses the SC energy and position to estimate the electron tracks in the inner layers of the silicon tracker. Seeds are selected from the hits in the corresponding layers if they are within  $\phi-z$  (or  $\phi-r$ ) windows of the SC extrapolation. The tracker-based seeding starts with tracks reconstructed as described in Section 3.1.1. These tracks are extrapolated to the ECAL and then matched to the PF clusters based on their positions and momenta. If a match is found, the seed of the matched track is used to seed the electron track reconstruction in the next step. Otherwise, a second selection is tried among the unmatched tracks to recover potential electron tracks that have significant bremsstrahlung. Tracks with small number of hits or a large  $\chi^2$  from the Kalman filter are refitted using a Gaussian sum filter (GSF) to account for the energy loss within the tracker volume.

Compared to the ECAL-based seeding, the tracker-based seeding has higher efficiency for non-isolated electrons with low  $p_T$  or electrons in the barrel-endcap transition region. The electron seeds selected from both methods are combined to initiate track reconstruction, which is based on the combinatorial Kalman filter method. The Bethe-Heitler function [96] is used to model the electron energy loss, and the match between predicted and reconstructed hits in a layer is less restrictive during the track building process. The final track parameters are extracted by a GSF fit. The GSF models the energy loss in each layer by a mixture of Gaussian distributions.

The electron reconstruction efficiencies measured in 2017 data and in simulated  $Z/\gamma$  samples are shown in Fig. 3.8, together with the scale factors for different  $p_T$  bins as a function of  $\eta$ . They are compatible in data and simulation, giving scale factors close to one in almost the entire range.

### 3.2.3.1 Identification

Several electron identification methods are developed by CMS. In order to select the prompt isolated electrons, that constitutes the analysis signal, from background sources, the identification of the electron in this analysis uses a BDT multivariate classifier algorithm that has been updated and improved for the Run II analysis [97]. The main background sources are the electrons originated from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b- and c-quarks.

The BDT exploits observables sensitive to the amount of bremsstrahlung along the electron trajectory, the three PF isolation components (Eq. 3.1), the transverse energy density within the isolation cone, the geometrical and momentum matching between the electron trajectory and associated clusters, shower-shape observables, and electron conversion variables. The BDT is tuned for best performance and trained separately for each data taking year on a Drell-Yan plus jets MC sample under 2017 detector conditions.

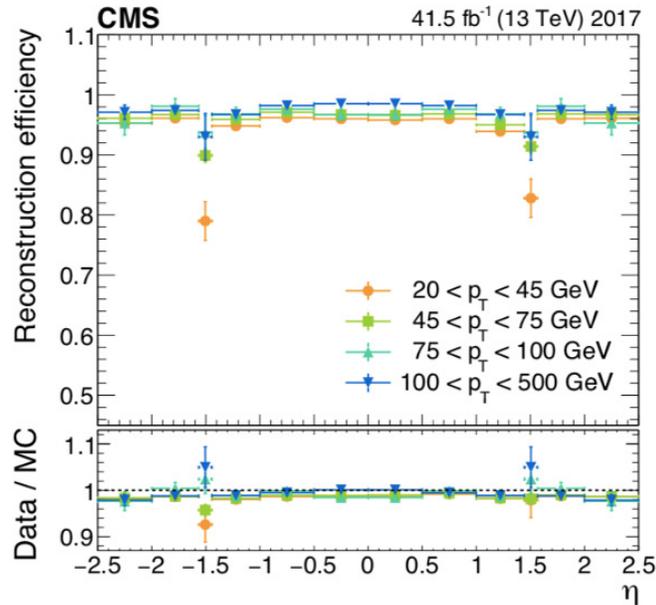


Figure 3.8: Electron reconstruction efficiency versus  $\eta$  measured in dielectron events in data (upper panel) and data-to-simulation efficiency ratios (lower panel) for the 2017 data taking period. The vertical bars on the markers represent the combined statistical and systematic uncertainties. The region  $1.44 < |\eta| < 1.57$  corresponds to the transition region between the barrel and endcap regions of ECAL and is not considered in physics analyses [97].

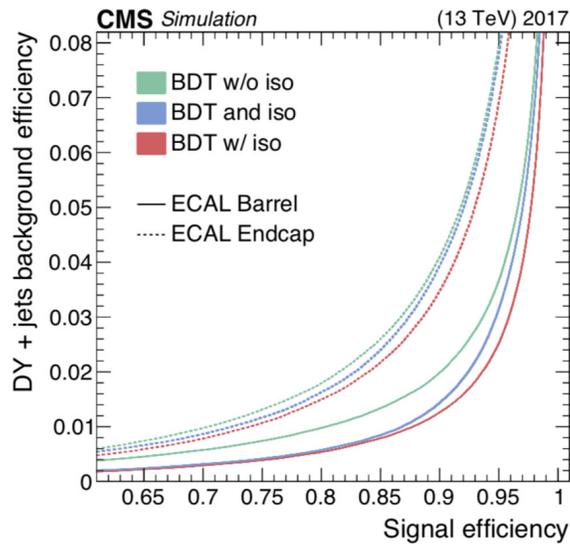


Figure 3.9: Receiver operating characteristic curves of the electron BDT-based identification [97]

Signals (backgrounds) are reconstructed electrons that match (do not match) to genuine electrons at generator level within a cone of size  $\Delta R = 0.1$ , regardless if they pass the

trigger requirements or not, with  $p_T > 10$  GeV and defines three categories, based on the electron  $\eta$ : two for the barrel ( $|\eta| < 0.8$ ) and one for the endcap.

The reconstructed electron candidate is identified as a prompt and isolated electron if the value of the discriminant (the output of the BDT) is larger than a fixed threshold, denoted as working point. The chosen threshold determines the efficiency of correctly identified and the misidentification probability for electron originating from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b- and c-quarks. In this analysis, signal electrons, are required to satisfy, for each MVA category (barrel and endcap), a tight working point with 80% of signal efficiency, provided by the E/gamma POG (Physic Object Group), as described in Section 4.4.1.1.

A comparison of the performance of the electron BDT-based discriminants, trained with and without isolation variables, is shown in Fig. 3.9.

### 3.2.4 Taus

Because of their short lifetime,  $\tau$  leptons decay before reaching the detector volume, thus the only way to reconstruct them is through their decay products. The main  $\tau$  decay modes and their branching ratios (BR) are listed in Tab. 3.1. Decays containing one and three charged hadrons are generally referred to as one- and three-prong decays, respectively.

The  $\tau$  decay modes can be divided into leptonic and hadronic decays. The leptonic  $\tau$  decays, denoted as  $\tau_\ell$  are the cases where the  $\tau$  lepton decays into an electron or muon, a tau neutrino and a neutrino associated to electron or muon (which does not enter in the reconstruction because it is undetected), and represents  $\sim 35.2\%$  of the cases. The leptonic  $\tau$  decays are reconstructed as an electron ( $\tau_e$ ), or as a muon ( $\tau_\mu$ ), described in Section 3.2.3 and Section 3.2.2 respectively; since it is hard to separate the  $\tau$  signature from the other electron and muon production processes. The remaining  $\sim 64.8\%$  of the cases is when the  $\tau$  decay contains hadrons, in this case the  $\tau$  lepton is reconstructed as a hadronic  $\tau$  jet ( $\tau_h$ ).

The constituents of the jets are analyzed in order to reconstruct and identify individual  $\tau_h$  decay modes. The reconstruction and identification are performed using respectively the hadron-plus-strips algorithm (HPS) [98, 99, 100] and the DeepTau [101] algorithms, described in the following.

#### Hadronic tau reconstruction (HPS)

The reconstruction of hadronic tau decays is seeded from PF jets. The HPS algorithm combines charged hadron constituents in the PF jet with “strips” from neutral pions. The neutral pion decays into two photons promptly. Each photon is likely to convert into

Decay mode	BR (%)	
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$	$17.41 \pm 0.04$	} 35.24 ± 0.08 %
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$	$17.83 \pm 0.04$	
$\tau^- \rightarrow h^- \nu_\tau$	$11.53 \pm 0.06$	} 64.76 %
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$25.95 \pm 0.09$	
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$9.53 \pm 0.11$	
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$9.80 \pm 0.06$	
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$	$4.80 \pm 0.06$	
Other hadronic decays	3.15	

Table 3.1: Branching ratio of the possible decay modes of the  $\tau$  lepton. The  $h$  symbol indicates a charged  $\pi^\pm$  or a  $K^\pm$ . The  $\tau^+$  lepton decays are charge conjugates of the listed  $\tau^-$  decays [8].

a pair of electron and positron in the tracker material before reaching the ECAL. As a result, the energy deposit in the ECAL due to neutral pions is likely to spread in a region narrow in  $\eta$  and wide in  $\phi$ , referred to as a “strip”. The strip is reconstructed by clustering the photon and electron PF candidates in the jet within a  $\Delta\eta \times \Delta\phi$  window. Both  $\Delta\eta$  and  $\Delta\phi$  are parameterized by the  $p_T$  of the clustered PF candidates, varying from 0.05 to 0.15 and from 0.05 to 0.3 respectively. The momentum of the strip is computed as the vectorial sum of all its constituent momenta, and its  $\eta - \phi$  coordinates are determined by a  $p_T$ -weighted average of its constituents.

The charged hadron candidates used in the HPS algorithm are required to be compatible with originating from the primary vertex. They are combined with the strips in the jet under different  $\tau$  decay mode hypotheses. Seven signatures corresponding to a single charged hadron in association to zero, one, or two strips, and three charged hadrons with no strips associated:

- $\mathbf{h}^\pm$  : single charged hadron with no strips, targeting  $\tau^\pm \rightarrow \pi^\pm \nu_\tau$ .
- $\mathbf{h}^\pm \pi^0$  : single charged hadron with one strip, targeting  $\tau^\pm \rightarrow \pi^\pm \pi^0 \nu_\tau$ .
- $\mathbf{h}^\pm \pi^0 \pi^0$  : single charged hadron with two strip, targeting  $\tau^\pm \rightarrow \pi^\pm \pi^0 \pi^0 \nu_\tau$ .
- $\mathbf{h}^\pm \mathbf{h}^\mp \mathbf{h}^\pm$  : three charged hadrons with no strips, the tracks are required to originate from the same vertex, targeting  $\tau^\pm \rightarrow h^\pm h^\mp h^\pm \nu_\tau$ .
- $\mathbf{h}^\pm \mathbf{h}^\mp \mathbf{h}^\pm \pi^0$  : targeting  $\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$  decays.

In the current documentation presented in References [98, 99, 100], the combinations of charged particles and strips, considered by the HPS algorithm, represent almost all the hadronic  $\tau$  decay modes listed in Tab. 3.1. The exceptions are the decays with low branching fraction, such as  $\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$  (BR = 4.8%), which in addition suffers from a greater contamination by jets, and decays included in the “other hadronic decays” category in the Tab. 3.1. However, in the current implementation of the HPS algorithm (not documented yet), some of these previously excluded decay modes are included. Furthermore, more inclusive decay mode definitions have been introduced, adding three charged prongs plus zero or one  $\pi^0$  with relaxed matching conditions to the previously available reconstruction modes. In addition, two recovery decay modes are available: two charged prongs plus zero or one  $\pi^0$  to recover 3-prong decays.

- $\mathbf{h^\pm h^\pm / \mp}$  : targeting  $\tau^- \rightarrow h^- h^+ h^- \nu_\tau$  decays.
- $\mathbf{h^\pm h^\pm / \mp \pi^0}$  : targeting  $\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$  decays.

The mass of the reconstructed  $\tau_h$  candidate is required to be compatible with the mass of  $\rho(770)$  meson if it is reconstructed in the  $h^\pm \pi^0$  mode, or with the mass of  $a_1(1260)$  meson if it is reconstructed in the  $h^\pm \pi^0 \pi^0$  or  $h^\pm h^\mp h^\pm$  decay mode. The total charge of the  $\tau_h$  candidate has to be  $\pm 1$ . Furthermore, a signal cone is established around the  $\tau_h$  candidate as  $\Delta R = 3 \text{ GeV}/p_T$  with  $0.05 < \Delta R < 0.01$ . The  $\tau_h$  candidate is rejected if any of its constituent charged hadrons or strips are outside the signal cone.

## Hadronic Tau identification

To discriminate reconstructed hadronic decays of tau leptons from genuine tau leptons against  $\tau_h$  candidates that originate from quark or gluon jets, electrons, or muons, the analysis presented in this thesis uses a new tau identification algorithm based on a deep neural network structure that simultaneously discriminates against jets, electrons, and muons. This algorithm, called DeepTau [101], takes as input the information from all the reconstructed particles in the vicinity of a reconstructed  $\tau_h$  candidate and employs a deep neural network with convolutional layers to efficiently process the inputs.

Before the introduction of DeepTau, the  $\tau_h$  discrimination was performed using three dedicated algorithms, one for each type of background: jets or electrons or muons from hadronic taus [99, 100].

The DeepTau discriminator was designed to be suitable for a wide range of CMS analyses with hadronic taus in the final state, and it provides a considerable improvement of the tau identification performance with respect to the before-mentioned methods.

The algorithm takes as input the information from all reconstructed particles in the vicinity of a reconstructed  $\tau_h$  candidate and employs a deep neural network with convolutional layers to efficiently process the inputs. The algorithm leads to an improved

performance across the board compared to the previously used algorithm. For example, the efficiency for genuine  $\tau_h$  to pass the discriminator against jets increases by 10–30% for a given efficiency for quark and gluon jets.

The high-level features exploited by the algorithm correspond to a large fraction to those proven to be useful in the previous classifier. These variables include the  $\tau_h$  four-momentum and charge, the numbers of charged and neutral particles used to reconstruct the  $\tau_h$  candidate, isolation variables, the compatibility of the leading  $\tau_h$  track with coming from the primary interaction vertex, the properties of a secondary vertex in case of multi-prong  $\tau_h$ , observables related to the  $\eta$  and  $\phi$  distribution of energy reconstructed in the strips, observables related to the compatibility of the  $\tau_h$  with being an electron, and the estimated pileup density in the event. In addition, information of all PF particles within a cone of size  $\Delta R < 0.5$  around the direction of the  $\tau_h$  candidate is used, including basic kinematic quantities  $(p_T, \eta, \phi)$  and observables related to calorimeter clusters and muon detectors.

The training of the DeepTau discriminator uses simulated samples of Z boson production in association with jets (Z+jets), W boson production in association with jets (W+jets), top quark pair production ( $t\bar{t}$ ), the production of heavy gauge bosons  $Z'$  with decays to tau leptons, electrons, or muons ( $Z' \rightarrow \tau\tau$ ,  $Z' \rightarrow ee$ ,  $Z' \rightarrow \mu\mu$ ), with  $m(Z')$  ranging from 1 to 5 TeV, and multijet production through the strong interaction, referred to as quantum chromodynamics (QCD) multijet production. Different samples of the same processes are used to evaluate the performance of the discriminator. The  $\tau_h$  candidates reconstructed in these samples are required to have:  $20 < p_T < 1000$  GeV,  $|\eta| < 2.3$  and the longitudinal impact parameter with respect to the primary vertex  $d_z < 0.2$  cm.

To guide usage in physics analyses and to derive suitable data-to-simulation corrections, working points are defined. The target efficiencies for the different working points are given in Tab. 3.2, and address the range of efficiencies used in a wide range of physics analyses with taus in the final state. These target efficiencies are defined as the efficiency for genuine  $\tau$  from  $H \rightarrow \tau\tau$  event sample with a transverse momentum  $30 < p_T < 70$  GeV in the  $H \rightarrow \tau\tau$  event sample and range from 40 to 98% for the DeepTau vs jets discriminator, from 99.5 to 99.95% for the DeepTau vs. muons discriminator, and from 60 to 99.5% for the DeepTau vs. electrons discriminator. The efficiencies for genuine  $\tau_h$ , electrons, and muons are generally representative also for different samples and  $p_T$  ranges, whereas the efficiencies for jets depend strongly on  $p_T$  and jet flavour. The numbers for jets in this table are hence given for illustrative purposes only since the identification efficiency is integrated over the  $p_T$ ,  $\eta$ , and flavour distributions of the jets in the given event sample.

In Figs. 3.10, 3.11a and 3.11b is shown the performance of the DeepTau, compared with previous discriminators, against jets, electrons and muons respectively, in the left for

	VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
<b>VSe</b>	60%	70%	80 %	90 %	95%	98%	99%	99.5%
<b>VSmu</b>	-	-	99.5 %	99.8 %	99.9%	99.95%	-	-
<b>VSjet</b>	40%	50%	60 %	70 %	80%	90%	95%	98%

Table 3.2: Defined working points for the DeepTau discriminator, against electrons (VSe), muons (VSmu) and jets (VSjet), and their target efficiency, evaluated with  $H \rightarrow \tau\tau$  MC events sample for  $\tau_h$  with  $p_T \in [30, 70]$  GeV [101].

$p_T < 100$  GeV and on the right for  $p_T > 100$  GeV. The working points of the discriminators are indicated by dots. The best performance is obtained with the DeepTau discriminator, against all the backgrounds.

The identification and reconstruction of the  $\tau$  lepton decays is very important for this thesis, because in the final states of the search performed two  $\tau$  are involved, which can be either two  $\tau_h$ , or  $\tau_l \tau_h$ . In both cases the two  $\tau$  leptons come from the decay of one of the Higgs bosons. Specific requirements for the analysis are detailed in Section 4.4.1.

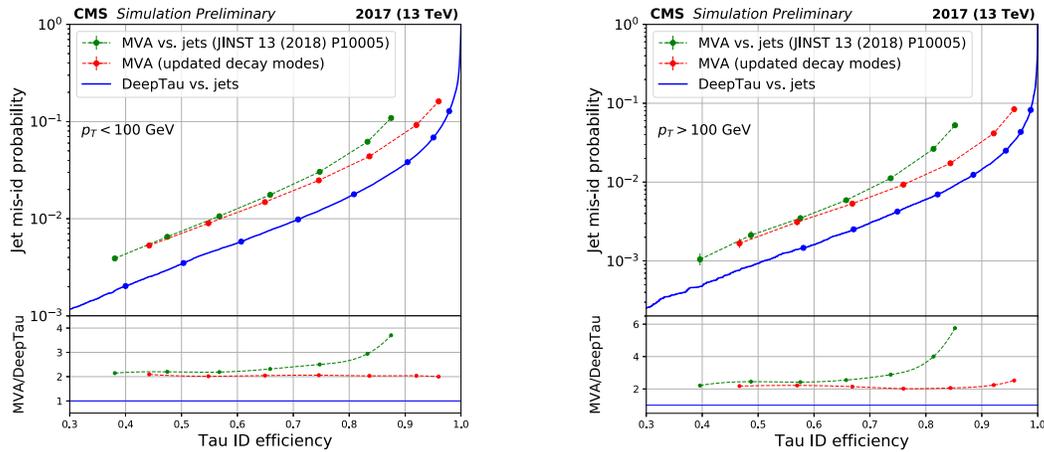


Figure 3.10: Performance of the DeepTau, compared with previous discriminators, against jets, for a reconstructed  $\tau_h$  with  $p_T < 100$  GeV (left) and  $p_T > 100$  GeV (right). The working points of the discriminators are indicated by dots, from left to right the loosest to the tighter [102].

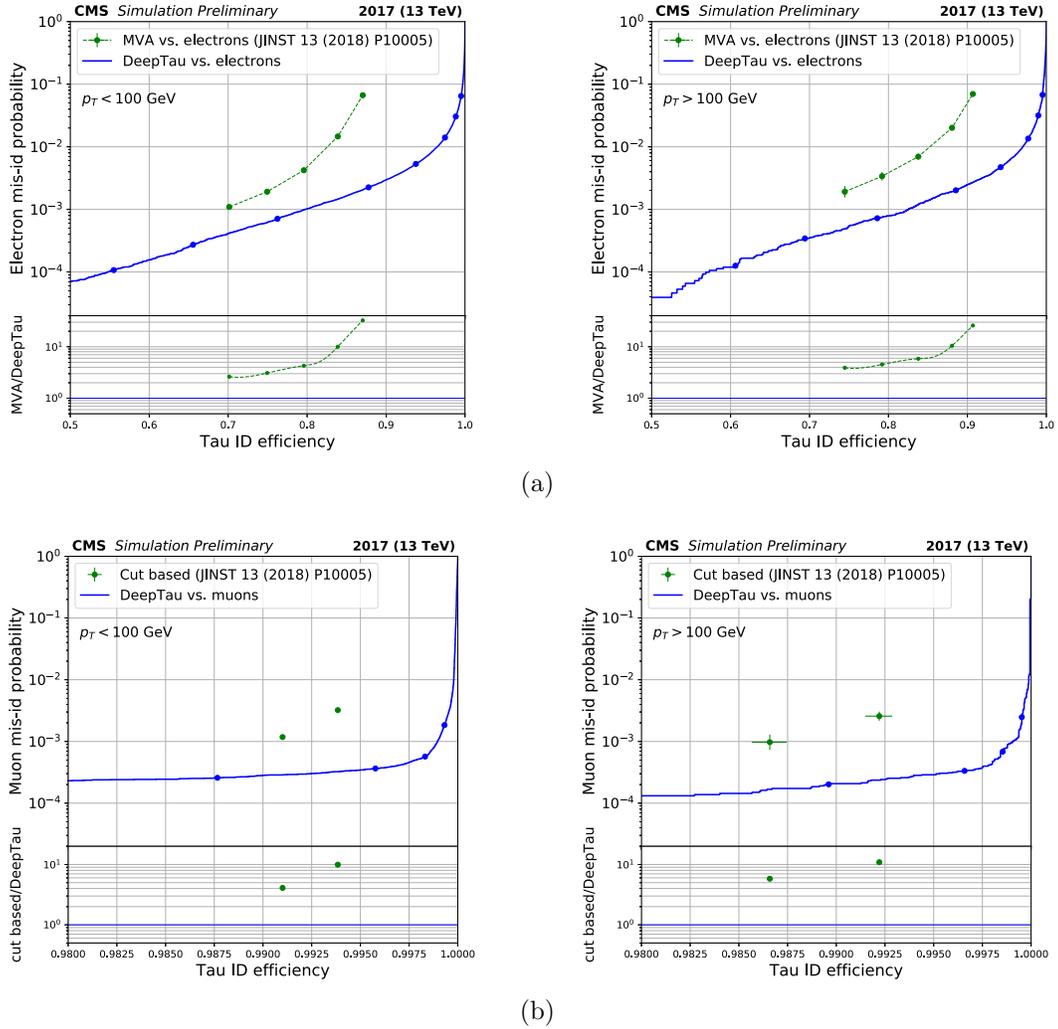


Figure 3.11: Performance of the DeepTau, compared with previous discriminators, against electrons (a) , and muons (b), for a reconstructed  $\tau_h$  with  $p_T < 100$  GeV (left) and  $p_T > 100$  GeV (right). The working points of the discriminators are indicated by dots, from left to right the loosest to the tighter [102].

### 3.2.5 Jets

Quarks and gluons, abundantly produced in p-p collisions at LHC, instantaneously fragment and through hadronization produce a collimated spray of hadrons. The non-stable hadrons created decay to stable particles which, together with the stable particles created during hadronisation, reach the detector material. The resulting shower of particles appears as a cluster of energy deposited in localized areas of the calorimeter, called jet. This is pictorially represented in Fig. 3.12.

The main challenge of an analysis using jets from the hard collision is to recover the initial energy, momentum and, possibly, the nature of the parton produced in the

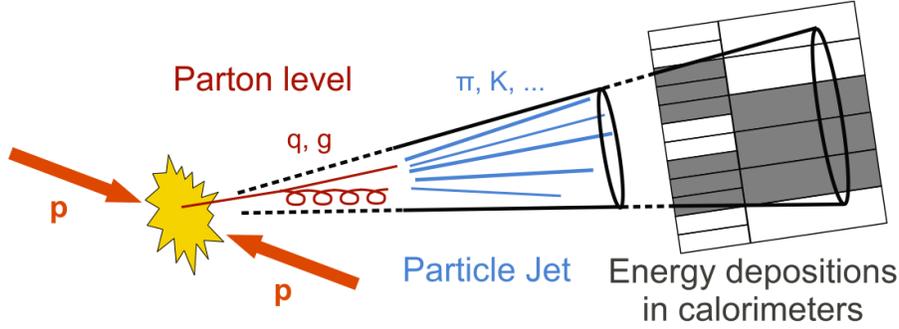


Figure 3.12: illustration of a jet produced during a p-p collision.

original hard interaction conceiving an algorithm that reconstructs jets from the hard collision, allowing a fairly accurate extrapolation of the parton properties and suppressing the contribution due to the pile-up collisions in the recorded event. In CMS, jets are reconstructed from the PF particles by the anti- $k_t$  algorithm [103], as implemented in the fastjet package [104].

The reconstruction starts from the PF candidates (Section 3.1.3), which are clustered using the anti- $k_t$  algorithm to form jets. The anti- $k_t$  is an ideal cone algorithm and uses as input the list of the PF particles. In each iteration, it recombines the current set of particles and jet-candidates, which both are referred as entries, clustering them around the hardest particle in the event in order to produce jets of an approximate conic shape. To form jets, for each pair of objects to be clustered, two quantities are evaluated in each step, the distances between pairs of entries ( $d_{ij}$ ) and a parameter used to represent the distance between an entry and the beam line ( $d_{iB}$ ):

$$d_{ij} = \min(p_{T,i}^{2n}, p_{T,j}^{2n}) \frac{\Delta R_{ij}^2}{\Delta R^2}, \quad \text{where} \quad \Delta R_{ij}^2 = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}, \quad (3.2)$$

$$d_{iB} = p_{T,i}^{-2}. \quad (3.3)$$

The parameter  $R_{ij}^2$  is a fixed distance, which defines the cone typical amplitude, and  $p_{T,i}$ ,  $\eta_i$  and  $\phi_i$  are respectively the transverse momentum, rapidity and azimuth of particle  $i$ . In the iteration process, if  $d_{iB}$  is smaller than  $\min_j d_{ij}$ , entry  $i$  is promoted to a reconstructed jet and is excluded from further iterations. Otherwise, entry  $i$  and  $j' = \min_j d_{ij}$  are merged into a new entry  $i'$  and the algorithm proceeds to the next iteration. The algorithm stops when all jet candidates are promoted to the reconstructed jets or when  $\min(d_{ij}, d_{iB})$  is below a threshold.

The standard clustering distance used at CMS is  $\Delta R = 0.4$  for Run II, while it was 0.5 at 8 TeV. However, when looking for boosted heavy particles decaying into hadrons, a  $\Delta R = 0.8$ , is used, since in this case the jets from the decay are not separated. In

the analysis presented in this thesis two collections of jets are used: the non-boosted jets reconstructed with  $\Delta R = 0.4$ , the so called AK4 jets, and the boosted jets with  $\Delta R = 0.8$ , the so called AK8 jets. The use of the collection of AK8 jets allows a more efficient reconstruction of the decays  $H \rightarrow b\bar{b}$ , when the two b jets are close to each other (boosted regime), and the hadronization cones are likely to be partially in overlap and cannot be resolved as separate AK4 jets. Unless explicitly stated otherwise, the mention of “jets” in the following denotes AK4 jets.

The raw jet 4-momentum, obtained by the anti- $k_t$  algorithm, does not represent an optimal estimation of the true jet 4-momentum at the hadron level, because of the non-linearities in the detector responses, imperfect detector modelling, noise and pile-up effects. To improve this estimation a set of corrections called “jet energy corrections” (JEC) is applied to the 4-momentum magnitude. The corrections, calculated using simulated samples, are applied to the 4-momentum of all the jets, in order to make the on average the 4-momentum equal to the one at hadron level [105].

### 3.2.5.1 Identification of b-jets

The identification of jets resulting from the hadronization of b-quarks, known as b-tagging, is the key for many physics analyses at LHC. The b-tagging, is used in the SM Higgs or in top quark analyses since their preferred decays contains b-hadrons in the final state. Searches for new physics also typically involve b-tagging, especially in theories where the third generation plays a special role. Last but not least, b-tagging is an important tool used in the search presented in this thesis, where one of the two Higgs bosons, decays into a pair of b-jets. The reconstruction of b-jets is quite challenging because they can be easily faked by jets originated by gluons and light-flavour quarks. The discrimination of b-jets against jets originated from light quarks or gluons, mainly exploits the properties of b-hadrons arising from the hadronization of the b-quark: their long lifetime ( $\sim 1.5$  ps) and their relatively large mass ( $\sim 5$  GeV). A simplified representation of a b-jet is given in Fig. 3.13.

The long lifetime means that charged tracks from the b-hadron decay have typically at the LHC a large impact parameter with respect to the primary vertex and are originated from a secondary vertex separated several millimeters from the primary vertex. A secondary vertex or multiple secondary vertices can be reconstructed inside a jet and their properties are highly discriminating variables. The relatively large mass disfavors their creation during fragmentation and hence b-hadrons are produced more likely from the higher momentum hard process than a typical light hadron. Furthermore, the number of particles produced during the fragmentation depends on the mass of the parton and hence a b-hadron will carry a larger fraction of the momentum of the parton than a light hadron. This implies that decay products within a b-jet are more collimated within the

jet cone and with a larger invariant mass.

Additionally, the larger mass of the b-hadron means it will typically decay into more particles than light hadrons. Another property of the b-hadron decay is the relatively high rate of lepton production from semileptonic decays (around 25%). These leptons can be identified thanks to their relatively high  $p_T$  relative to the b-hadron flight direction, approximated by the b-jet direction: leptons from b-hadron decays have order of GeV momenta relative to b-jet direction, because of the b-hadron mass ( $\sim 5$  GeV), while leptons in jets of other flavors tend to be closely aligned with the jet. These are the properties to be exploited by a tagging algorithm for b-jets.

The reconstruction of b-jets is the same as described for the other jets in Section 3.2.5. In CMS there are available several algorithms that, with the use of machine learning, have created powerful b-tagging tools. The currently recommended flavour-taggers are DeepCVS [106] and DeepFlavour, or also referred as DeepJet [107, 108, 109]. In the ongoing  $bb\tau\tau$  analysis [110], and in the work presented in this thesis, the b-jet identification is done with the latest CMS b-tagging algorithm DeepFlavour. The representation of the jet used by DeepFlavour aims to best exploit the description of the jet as produced by the PF algorithm.

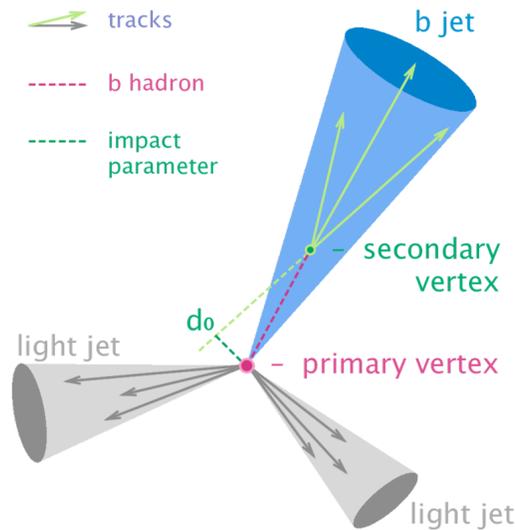


Figure 3.13: Simplified illustration of a b-jet produced in association with other light jets.

The DeepJet tagger is based on a deep neural network (DNN), which uses advanced multivariate techniques like convolutional layers to distinguish b-jets from other jets, exploiting the capacity of deep models to process a large number of particles and features. For this complex deep neural network structure, 16 properties of up to 25 charged PF jet constituents, six different kinematic variables of 25 neutral PF jet constituents, 17 properties from up to four secondary vertices matched geometrically to the jet, and global variables are used as input to the DNN, including high level features that provide a

global description of the jet suitable for b-tagging, the so-called tagging variables, and the jet kinematics  $(p_T, \eta)$ . The output of this algorithm are the probabilities for the identification of jets from different flavors, which are classified into six categories: jets containing exactly one b-hadron, at least two b-hadrons, one leptonically decaying b-hadron, at least one c-hadron but no b-hadrons, light-quark jets, and gluon-jets. The b-jet identification probability is determined, by summing the probabilities of the first three categories.

In Fig. 3.14 a comparison of the classification performance of the DeepFlavour and DeepCSV taggers is presented. Their performance is evaluated using fully hadronic  $t\bar{t}$  MC events, for jets with  $p_T > 30$  and  $p_T > 90$  GeV, on the left and right plots, respectively. The main difference between the two methods is that with respect to its predecessor DeepCSV [106], DeepJet uses also the PF neutral information. The DeepFlavour performs better than DeepCSV, especially at high jet  $p_T$ , where it improves the efficiency, defined as the true positive rate, to almost 20% at  $10^{-3}$  misidentification probability, defined as the false positive rate [107].

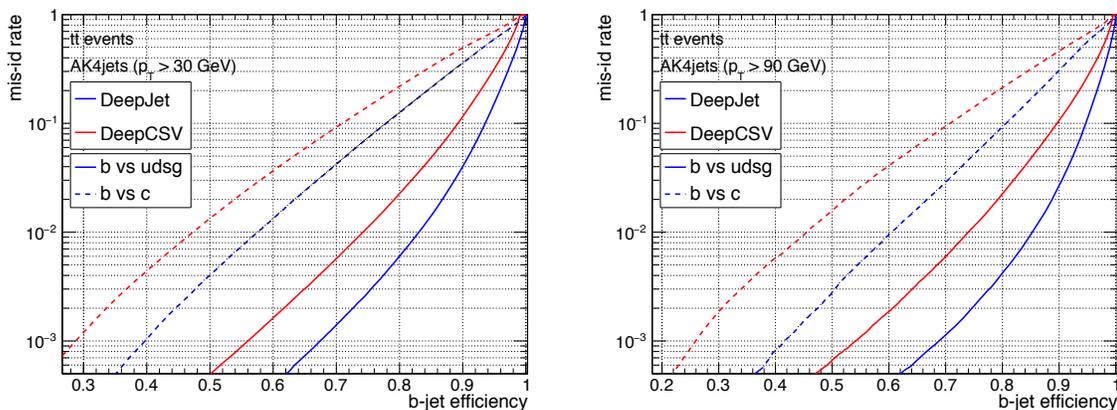


Figure 3.14: Comparison of the performance of the DeepFlavour and DeepCSV b-tagging algorithms, evaluated using  $t\bar{t}$  MC events. The dashed lines represent the performance of b-jets vs c-jets, and in solid lines b-jets vs light-jets misidentification rate. Requirements for the jet are set to  $p_T > 30$  GeV (left) and  $p_T > 90$  GeV (right) [107].

Tree working points are used: Loose, Medium and Tight, defined as the cut values in the discriminator distribution at which the rate of misidentifying a light jet as a b-jet is 10%, 1% and 0.1% respectively. The operating points are derived from the performance curves of the DeepFlavour discriminator, using QCD multijet MC events, requiring for the jets  $|\eta| < 2.5$  and  $p_T > 30$  GeV

Three different working points, delivered by the CMS b-Tag and Vertexing POG [111] for each year of the Run II data taking period, are defined for b-tagging algorithms, based on the probability of misidentifying light-flavored jets from u, s, and d quarks or

gluons as b jets: loose (10%), medium (1%), and tight (0,1%). The working points used in the analysis are described in Section 4.4.2.

### 3.2.6 Missing transverse energy

The initial protons taking part in a p-p collision only carry longitudinal momenta. Therefore, the total transverse momentum of all particles produced within the collision is zero as a result of momentum conservation. As neutrinos only interact weakly, they cannot be directly measured with the CMS detector, and their presence is inferred by the imbalance in the transverse momentum measurement, called missing transverse momentum. The missing transverse momentum vector  $\vec{p}_T^{miss}$  is defined as the negative vectorial sum of the transverse momenta of all reconstructed PF particles :

$$\vec{p}_T^{miss} = - \sum_{i=1}^{N \text{ particles}} \vec{p}_{T,i}, \quad (3.4)$$

Its magnitude is referred to as missing transverse momentum or missing transverse energy (MET).

For each event, the missing transverse momentum  $\vec{p}_T^{miss}$  is reconstructed with the PF MET algorithm [112], following the recommendations of the JetMET POG [113].

For this analysis, we applied filters, as also recommended by the JetMET POG [114], to ensure a good quality of the reconstructed MET.

As the reconstruction of PF particles is smeared by all the applied jet corrections as described in the Appendix A.3, the missing transverse momentum needs to be calibrated to account for the detector response. This is done by replacing in missing momentum the jets with the jets obtained after jet energy corrections are applied:

$$\vec{p}_T^{miss,corr} = \vec{p}_T^{miss} - \sum_{j=1}^{N_{PF} \text{ jets}} \left( \vec{p}_{T,j}^{corr} - \vec{p}_{T,j} \right) \quad (3.5)$$

# Chapter 4

## The $HH \rightarrow bb\tau\tau$ analysis

This chapter describes the analysis strategy to select and reconstruct the signal-like events: a pair of Higgs bosons, where one decays into two tau leptons, and the other into two b-jets. This specific final state presents several challenges, because it requires to reconstruct different types of physical objects. The aim is to maximize the signal efficiency and suppress the overwhelming background, in order to increase the analysis sensitivity. To achieve this goal this analysis uses the most recent techniques for the online and offline selection, reconstruction and identification of the final state physical objects, and a precise modeling of the signal and of the background. Some of these new techniques were developed specifically for this analysis. Latest results for  $HH \rightarrow bb\tau\tau$  published in the Ref. [115] were obtained using  $35.9 \text{ fb}^{-1}$  of data collected during the 2016. The new search [110], presented in the following make part of the so called Run II legacy analyses, which uses the full sample of data collected during the Run II (2016-2018) corresponding to an integrated luminosity of  $137 \text{ fb}^{-1}$ .

### 4.1 The $HH \rightarrow bb\tau\tau$ decay channel

#### 4.1.1 Signal definition

This particular search focuses on the non-resonant Higgs boson pair production investigating both the gluon fusion and the vector boson fusion production mechanism, exploiting events where one Higgs boson decays into two b-quarks and the other into two  $\tau$  lepton. The decay of the Higgs boson into a tau pair has six possible final states, and they can be categorized either as full-leptonic, semi-leptonic or hadronic; the branching ratios of these possible final states are summarized in Tab. 4.1. The three final states explored are: a hadronic channel  $HH \rightarrow bb\tau_h\tau_h$ , where  $\tau_h$  represents a  $\tau$  decaying into hadrons plus a  $\nu_\tau$ , and two semi-leptonic channels  $HH \rightarrow bb\tau_\ell\tau_h$ , where one  $\tau$  decays to a lighter lepton ( $\ell = e$  or  $\mu$ ) while the other decays into hadrons ( $\tau_h$ ). The fully-leptonic channel

---

is not considered, since it represents only the 12% of the  $H \rightarrow \tau\tau$  decays, and it suffers from considerable background contamination from the Drell-Yan electron and muon pair production, and the  $t\bar{t}$  processes.

Decay mode	BR (%)	
$\tau_e\tau_e$	3.2	Fully-leptonic decays = 12.4%
$\tau_e\tau_\mu$	6.2	
$\tau_\mu\tau_\mu$	3.0	
$\tau_e\tau_h$	23.1	Semi-leptonic decays = 45.6%
$\tau_\mu\tau_h$	22.5	
$\tau_h\tau_h$	42.0	Hadronic decay

Table 4.1:  $H \rightarrow \tau\tau$  decay modes and their branching ratio [116].

### 4.1.2 Background definition

The signal events described previously, characterized by having two tau leptons, and two b-jets in the final state, are not produced only by the double Higgs production. Unfortunately, there are several processes in the SM with the same signature final state, and those events are classified in this work as background. In particular, the hadronic channel is rather challenging at LHC due to the large background produced in association with jets by QCD processes. In order to increase the analysis sensitivity, it is important to select and identify with high efficiency and purity the signal events. This is challenging, since the background events are produced several orders of magnitude more frequently than the signal events, and the use of harsh selections or methods that remove the majority of events in the signal regions would compromise the results, spoiling the analysis sensitivity. There are several processes contributing to the background for the  $bb\tau\tau$  channel. These processes can be classified as reducible or irreducible background.

The reducible background process refers to the erroneous object identification. Among the processes with major contribution to the reducible background are: the Drell-Yan process, when a tau pair is produced in association to a light flavour quark pair, the  $t\bar{t}$  process, when the  $W$  boson produced decays into quarks, and the QCD multi-jet production, which contribute to the misidentification of quark and gluon jets as  $\tau_h$ . These background processes occur much often than the signal events, and their contribution is not negligible.

The irreducible backgrounds are defined as the background processes where the final state is the same as the signal one, in our case two b-jets and tau leptons. There is a high probability for those events to be selected as signal events. The main sources of this

type of background, for the  $bb\tau\tau$  channel are:  $t\bar{t}$  process, when the  $W$  boson decays into leptons and the two  $b$ -jets are identified and reconstructed, and the Drell-Yan process when a tau pair in association to a  $b$ -quark pair is produced.

In order to reject the several background processes, different techniques should be applied. The reducible backgrounds are handled by requiring strict quality criteria for the reconstructed physical objects, for example in order to discriminate a  $\tau_h$  against a jet requiring a tighter operational working point (Tab. 3.2), for the DeepTau discriminator, described in Section 3.2.4. In addition, the event categorization can improve the rejection of the reducible background processes, by separating the events according to the quality of the reconstructed objects, as in done in this analysis for the  $b$ -jets, described in Section 4.4.2.1. Instead, in order to reject irreducible background processes more complex techniques are needed, where the kinematic properties of the objects in the event are exploited, for example with machine learning techniques like the one presented in Ref. [117], and briefly described in Section 4.4.4.2.

The main four background processes for the  $HH \rightarrow bb\tau\tau$  final state, in order of relevance are:

- **QCD multi-jet:** it contributes to the reducible background when one or two jets, are misidentified as a  $\tau_h$  candidate. Its contribution is small in the semileptonic channels, because the selection criteria manage to reject electrons and muons coming from the jets. The cross section of this process makes this background one of the most important.
- **Drell-Yan ( $Z/\gamma^* \rightarrow \ell\ell + \text{jets}$  (with  $\ell = e, \mu, \tau$ )):** it is a large source of background given that the decay  $Z \rightarrow \tau\tau$  is very similar to the  $H \rightarrow \tau\tau$  process, so it can be misidentified if the  $\tau_h$  from the  $Z$  decay pass the selection outlined in Section 4.4.1. It is a source of reducible and irreducible background depending on the specific final state of the process and what objects are misidentified. Indeed, it represents a source of reducible background when only the pair of  $\tau_h$  candidates are misidentified as the signal one, however when the pair of leptons are produced in association with two  $b$ -jets, the later can be misidentified as the  $b$ -jet pair coming from the Higgs boson decay, and it becomes an irreducible source of background.
- **$t\bar{t} + \text{jets}$ :** each of the top-quark has a large probability [8] of decaying into a  $b$ -quark and a  $W$  boson, which in turn can decay which in turn can decay in a lepton and a neutrino, with  $\text{BR}(W \rightarrow \tau\nu_\tau) = \text{BR}(W \rightarrow \mu\nu_\mu) = \text{BR}(W \rightarrow e\nu_e) \simeq 10.8\%$ , with the same signature of the signal. The major contribution this irreducible background is when the two  $\tau_h$  come from the two  $W$  decay. Processes where the  $W$  decay into quarks contribute to reducible background.

- **W+jets** ( $W \rightarrow \ell\nu_\ell + \text{jets}$ ;) regardless its high production cross section, this represents a reducible background process, with a lower contribution with respect to the previous ones. Because, at the most, only one real  $\tau_\ell$  comes from W decay and in this case, the other three jets in the event must be misidentified as a  $\tau_h$  plus two b-jets, and this happens with a low probability.

## 4.2 Data samples

The analysis presented in this thesis uses a total of  $137.2 \text{ fb}^{-1}$  data collected at  $\sqrt{s} = 13 \text{ TeV}$  with a 25 ns bunch crossing intervals during 2016 ( $35.9 \text{ fb}^{-1}$ ), 2017 ( $41.5 \text{ fb}^{-1}$ ) and 2018 ( $59.7 \text{ fb}^{-1}$ ) at CMS [118, 119, 75]. The names of the data sets used in each of the channels are given in Tables 4.3 and 4.2 with the corresponding run range and integrated luminosity. The dataset uses only runs certified as good, where all components of the detector and accelerator are reliably working.

$\tau_h\tau_h$ channel		
Data set Name	Run-range	Luminosity [ $\text{fb}^{-1}$ ]
/Tau/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376	5.750
/Tau/Run2016C-17Jul2018-v1/MINIAOD	275656-276283	2.573
/Tau/Run2016D-17Jul2018-v1/MINIAOD	276315-276811	4.242
/Tau/Run2016E-17Jul2018-v1/MINIAOD	276831-277420	4.025
/Tau/Run2016F-17Jul2018-v1/MINIAOD	277932-278808	3.105
/Tau/Run2016G-17Jul2018-v1/MINIAOD	278820-280385	7.576
/Tau/Run2016H-17Jul2018-v1/MINIAOD	281613-284044	8.651
/Tau/Run2017B-31Mar2018-v1/MINIAOD	297047-299329	4.794
/Tau/Run2017C-31Mar2018-v1/MINIAOD	299368-302029	9.633
/Tau/Run2017D-31Mar2018-v1/MINIAOD	302031-302663	4.248
/Tau/Run2017E-31Mar2018-v1/MINIAOD	303824-304797	9.315
/Tau/Run2017F-31Mar2018-v1/MINIAOD	305040-306462	13.54
/Tau/Run2018A-17Sep2018-v1/MINIAOD	315257-316995	14.03
/Tau/Run2018B-17Sep2018-v1/MINIAOD	317080-319310	7.066
/Tau/Run2018C-17Sep2018-v1/MINIAOD	319337-320065	6.899
/Tau/Run2018D-PromptReco-v2/MINIAOD	320413-325175	31.75

Table 4.2: List of data sets analyzed for  $\tau_\mu\tau_h$  final state.

$\tau_\mu\tau_h$ channel		
Data set Name	Run-range	Luminosity [ $\text{fb}^{-1}$ ]
/SingleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376	5.750
/SingleMuon/Run2016C-17Jul2018-v1/MINIAOD	275656-276283	2.573
/SingleMuon/Run2016D-17Jul2018-v1/MINIAOD	276315-276811	4.242
/SingleMuon/Run2016E-17Jul2018-v1/MINIAOD	276831-277420	4.025
/SingleMuon/Run2016F-17Jul2018-v1/MINIAOD	277932-278808	3.105
/SingleMuon/Run2016G-17Jul2018-v1/MINIAOD	278820-280385	7.576
/SingleMuon/Run2016H-17Jul2018-v1/MINIAOD	281613-284044	8.651
/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297047-299329	4.794
/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368-302029	9.633
/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302018-302663	4.248
/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824-304797	9.315
/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040-306462	13.54
/SingleMuon/Run2018A-17Sep2018-v2/MINIAOD	315257-316995	14.03
/SingleMuon/Run2018B-17Sep2018-v1/MINIAOD	317080-319310	7.066
/SingleMuon/Run2018C-17Sep2018-v1/MINIAOD	319337-320065	6.899
/SingleMuon/Run2018D-22Jan2019-v2/MINIAOD	320413-325175	31.75
$\tau_e\tau_h$ channel		
Data set Name	Run-range	Luminosity [ $\text{fb}^{-1}$ ]
/SingleElectron/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376	5.750
/SingleElectron/Run2016C-17Jul2018-v1/MINIAOD	275656-276283	2.573
/SingleElectron/Run2016D-17Jul2018-v1/MINIAOD	276315-276811	4.242
/SingleElectron/Run2016E-17Jul2018-v1/MINIAOD	276831-277420	4.025
/SingleElectron/Run2016F-17Jul2018-v1/MINIAOD	277932-278808	3.105
/SingleElectron/Run2016G-17Jul2018-v1/MINIAOD	278820-280385	7.576
/SingleElectron/Run2016H-17Jul2018-v1/MINIAOD	281613-284044	8.651
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD	297047-299329	4.794
/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD	299368-302029	9.633
/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD	302018-302663	4.248
/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD	303818-304797	9.315
/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD	305040-306460	13.54
/EGamma/Run2018A-17Sep2018-v2/MINIAOD	315257-316995	14.03
/EGamma/Run2018B-17Sep2018-v1/MINIAOD	317080-319310	7.066
/EGamma/Run2018C-17Sep2018-v1/MINIAOD	319337-320065	6.899
/EGamma/Run2018D-22Jan2019-v2/MINIAOD	320413-325175	31.75

Table 4.3: List of data sets analyzed for the  $\tau_e\tau_h$  and  $\tau_\mu\tau_h$  final states.

### 4.3 Trigger selections

The  $\text{HH} \rightarrow \text{bb}\tau\tau$  events, from data and MC simulated samples, are selected with a set of HLT paths, described in Section 2.2.3, which target only the  $\text{H} \rightarrow \tau\tau$  decay, and require

the presence of either a muon, electron or  $\tau_h$  candidate in the event, originating from the hard interaction (Section 3.1.2), depending on the final state considered. The trigger paths are chosen in such a way to guarantee the greatest possible signal acceptance. Several trigger paths are used to increase the efficiency in the selection, since the trigger selection uses an “OR” logic, i.e. only one path is needed to select the event.

The HLT paths used are classified into four categories: double- $\tau_h$  or di-tau trigger, VBF trigger, single-lepton, and cross-lepton trigger. For the hadronic channel a double- $\tau_h$  or di-tau trigger is used, where the presence of two  $\tau_h$  is required. In addition, for the data periods of 2017 and 2018, the VBF trigger paths are included. The VBF trigger is only used in the region inaccessible to the di-tau trigger, i.e. for events where at least one of the  $\tau_h$  has  $p_T < 40$  GeV (as shown in Tab. 4.12), since the VBF trigger has a lower  $p_T$  threshold for the tau candidates (20 GeV). In addition, the VBF trigger has specific requirements on the two additional jets  $p_T$  and invariant mass: a leading jet transverse momentum greater than 115 GeV, each jet must have a  $p_T > 40$  GeV and the invariant mass of the jet pair  $m_{jj} > 620$  GeV.

For the semi-leptonic channels ( $\tau_e\tau_h, \tau_\mu\tau_h$ ) two types of HLT paths are used: the cross-lepton triggers, where only the presence of one  $\tau_h$  is required, along with one electron or muon, and the single-electron trigger, where only the presence of an electron or muon is required. The HLT trigger paths requested in data are listed in Tables 4.5, 4.6 and 4.8. The same trigger bits are requested in the MC simulated events [110]. Owing to the cross-lepton triggers, the loss of acceptance due to the tighter selections for the two semi-leptonic channels is mitigated. The trigger paths requiring one  $\tau_h$  have changed during the 2018 data-taking, since an online version of the HPS algorithm (Section 3.2.4) has been developed and deployed starting from run 317509, as marked in Tab. 4.2

In order to account for differences between data and MC efficiency of these trigger paths, a correction factor or “scale factor” (SF) for the trigger efficiency is applied to the simulated samples. The trigger efficiencies used are provided by the Tau POG in the context of the Standard Model  $H \rightarrow \tau\tau$  analysis [120]. The trigger efficiency scale factors are described in the Appendix A.

The dataset name corresponds to specific requirements for each channel, and these are reported in Tables 4.9, 4.4 and 4.7, respectively for  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  channel.

HLT path	Kinematic selection	$\tau_h$ isolation and ID
single- $e$	One isolated $e$ , $p_T > 32$ GeV, $ \eta  < 2.1$	-
	One isolated $e$ , $p_T > 35$ GeV, $ \eta  < 2.1$	-
cross $e\tau_h$	One $\tau_h$ , $p_T > 30$ GeV, $ \eta  < 2.1$ ,	Loose
	One isolated $e$ , $p_T > 24$ GeV, $ \eta  < 2.1$	

Table 4.9: Trigger selections used in the  $\tau_e\tau_h$  channel.

HLT path	Kinematic selection	$\tau_h$ isolation and ID
single- $\mu$	One isolated $\mu$ , $p_T > 24$ GeV	-
	One isolated $\mu$ , $p_T > 27$ GeV	-
cross $\mu\tau_h$	One $\tau_h$ , $p_T > 27$ GeV, $ \eta  < 2.1$ , One isolated $\mu$ , $p_T > 20$ GeV, $ \eta  < 2.1$	Loose

Table 4.4: Trigger selections used in the  $\tau_\mu\tau_h$  channel.

Channel	HLT path name	runs
$\tau_\mu\tau_h$	HLT_IsoMu22	all runs
	HLT_IsoMu22_eta2p1	all runs
	HLT_IsoTkMu22	all runs
	HLT_IsoTkMu22_eta2p1	all runs
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20	all runs
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1	all runs
$\tau_e\tau_h$	HLT_Ele25_eta2p1_WPTight_Gsf	all runs
$\tau_h\tau_h$	HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg	eras B to G
	HLT_DoubleMediumCombinedIsoPFTau35_Trk1_eta2p1_Reg	era H

Table 4.5: Trigger paths used in the  $\tau_\mu\tau_h$ ,  $\tau_e\tau_h$ , and  $\tau_h\tau_h$  channels for 2016 data and MC simulation.

Channel	HLT path name	runs
$\tau_\mu\tau_h$	HLT_IsoMu24	all runs
	HLT_IsoMu27	all runs
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1	all runs
$\tau_e\tau_h$	HLT_Ele32_WPTight_Gsf_L1DoubleEG	all runs
	HLT_Ele32_WPTight_Gsf	all runs
	HLT_Ele35_WPTight_Gsf	all runs
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1	all runs
$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg	all runs
	HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg	all runs
	HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg	all runs
	HLT_VBF_DoubleLooseChargedIsoPFTau20_Trk1_eta2p1_Reg	eras D to F

Table 4.6: Trigger paths used in the  $\tau_\mu\tau_h$ ,  $\tau_e\tau_h$ , and  $\tau_h\tau_h$  channels for 2017 data and MC simulation. Since multiple paths cover the same runs, the logical OR of all these paths within each channel is used, except for the di-tau and VBF triggers.

HLT path	Kinematic selection	$\tau_h$ isolation and ID
double- $\tau_h$	Two $\tau_h$ , $p_T > 35$ GeV, $ \eta  < 2.1$	Tight, Tight
	Two $\tau_h$ , $p_T > 40$ GeV, $ \eta  < 2.1$	Medium, Tight
	Two $\tau_h$ , $p_T > 40$ GeV, $ \eta  < 2.1$	Tight, none
VBF $H \rightarrow \tau_h\tau_h$	Two $\tau_h$ , $p_T > 20$ GeV, $ \eta  < 2.1$ , Two jets, $p_{T1} > 115$ GeV, $p_{T2} > 40$ GeV, $m_{jj} > 620$ GeV	Loose, none

Table 4.7: Trigger selections used in the  $\tau_h\tau_h$  channel. The VBF  $H \rightarrow \tau_h\tau_h$  is only used to collect events in the corresponding VBF category.

Channel	HLT path name	runs
$\tau_\mu\tau_h$	HLT_IsoMu24	all, Data and MC
	HLT_IsoMu27	all, Data and MC
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1	run < 317509 (Data)
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTauHPS27_eta2p1_CrossL1	run $\geq$ 317509 (Data) and MC
$\tau_e\tau_h$	HLT_Ele32_WPTight_Gsf	all, Data and MC
	HLT_Ele35_WPTight_Gsf	all, Data and MC
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1	run < 317509 (Data)
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTauHPS30_eta2p1_CrossL1	run $\geq$ 317509 (Data) and MC
$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg	run < 317509 (Data)
	HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg	run < 317509 (Data)
	HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg	run < 317509 (Data)
	HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg	run $\geq$ 317509 (Data) and MC
	HLT_VBF_DoubleLooseChargedIsoPFTau20_Trk1_eta2p1	run < 317509 (Data)
	HLT_VBF_DoubleLooseChargedIsoPFTauHPS20_Trk1_eta2p1	run $\geq$ 317509 (Data) and MC

Table 4.8: Trigger paths used in the  $\tau_\mu\tau_h$ ,  $\tau_e\tau_h$ , and  $\tau_h\tau_h$  channels for 2018 data and MC simulation. Since multiple trigger paths cover the same runs, the logical OR of all these paths within each channel is used, except for the di-tau and VBF triggers.

## 4.4 Analysis flow

The analysis flow is articulated in three main steps. The first step consists in the selection of the candidates used to reconstruct the  $\tau\tau$  pair from the decay of one of the Higgs bosons in the event. At this same step, the final state of the  $\tau\tau$  system is assessed according to

the decay products of the tau leptons (Section 4.4.1). The second step is the identification and selection the b-jet pair candidate from the decay of the second Higgs boson and of the VBF additional two jets. After the last step, the events are categorized in resolved, boosted and VBF categories (Section 4.4.2.1).

The selection adopted corresponds to the so-called baseline selection. Then, the signal region for each of these categories is defined, applying requirements on the invariant mass of the  $\tau$  and of the b-jet pair (Section 4.4.3) of the HH candidate.

The final categorization for the signal extraction, including a splitting of the the VBF category in five sub-categories, is done by a multi-class deep neural network, as described in Section 4.4.4.1. A different deep neural network, described in Section 4.4.4.2, is used to discriminate signal from background events in each category.

#### 4.4.1 $H \rightarrow \tau\tau$ selection

First step of the offline event selection, aim to select and identify the visible decay products of the Higgs boson decaying into a  $\tau$  pair.

The  $\tau\tau$  decay categories ( $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$ ) are determined using only offline information. In all the categories is required to have at least one  $\tau_h$ . An event is classified as  $\tau_\mu\tau_h$  if there is a muon that satisfies the requirements from the baseline selection. In the case such muon is not found, but there is an electron that fulfills the requirements of the baseline selection the event is classified as  $\tau_e\tau_h$ . In the case neither a muon nor electron that meets the requirements is found, but there is a second  $\tau_h$  found, the event is classified as  $\tau_h\tau_h$ .

In the case that more than one pair of leptons are identified as possible decay products of the  $H \rightarrow \tau\tau$  decay, it is evaluated which of the pairs has a higher purity. The purity is based on the isolation of the first leg, and in the case this does not help to discriminate, the purity is evaluated using the  $p_T$  of the first leg as the discriminating parameter. If neither of these parameters are helpful to make a discrimination, the isolation and  $p_T$  of the second leg is evaluated, in order to maximize purity, and reduce possible overlaps between final states. The pairs formed with the highest purity, which pass all the requirements of the selection are chosen. In addition, a third lepton veto is applied, to remove events with additional electrons or muons, that are not selected as the  $\tau\tau$  pair. The third lepton veto, described in Section 4.4.1.4 requirement helps reducing the Drell-Yan contribution, and also ensures that the three  $\tau\tau$  final states are mutually exclusive.

After the  $\tau\tau$  pair is selected, it is evaluated whether the event is firing the trigger path associated to the selected final states, and whether the selected offline leptons are geometrically matched to the online ones (see Section 2.2.3 for a description the trigger system in CMS, and Section 4.3 for a description of the trigger requirements in the  $bb\tau\tau$  final state). As mentioned in Section 4.3, in the semileptonic final states HLT cross

triggers are used, however the cross triggers were introduced later in the Run II period, for the 2017 and 2018 data taking periods. The introduction of the HLT cross triggers has allowed to lower the  $p_T$  threshold on the  $\tau$  leptons, and increasing the signal acceptance. Each reconstructed offline lepton is required to pass a  $p_T$  threshold, which depends on the HLT trigger path fired by the event:

$$p_T^{\text{offline}} \geq p_T^{\text{HLT}} + \text{threshold}, \quad (4.1)$$

where  $p_T^{\text{offline}}$  is the transverse momentum of the offline selected lepton,  $p_T^{\text{HLT}}$  is the  $p_T$  threshold applied at trigger level, and the threshold added is a fixed value, which takes into account the resolution of the HLT, and depends on the type of particle: 1 GeV for muons and electrons, and 5 GeV for hadronic taus. The  $p_T$  thresholds are chosen to be conservative with respect to the trigger turn-on curves and accordingly to the recommendations of the Tau POG [121].

In the Tables 4.10 to 4.12 the specific selection applied is reported, as well as the types of HLT applied for each channel and year.

#### 4.4.1.1 $\tau_e\tau_h$ final state selection

The electrons are reconstructed as described in Section 3.2.3, and the hadronic taus as described in Section 3.2.4, following the standard CMS algorithm. The legs are required to have opposite electric charge and a spatial separation  $\Delta R(\tau_e, \tau_h) > 0.5$ .

The identification of genuine prompt electrons, as described in Section 3.2.3, is based on a MVA approach [95], in combination with a Tight working point.

In addition, the reconstructed electron production vertex of the candidate is required to be compatible with the primary vertex. Thus, it is required that the transverse and longitudinal impact parameter of the electron track with respect to the primary vertex satisfies:  $|d_{xy}| < 0.045$  cm (transverse) and  $|d_z| < 0.2$  cm (longitudinal).

The  $p_T$  thresholds of both legs depends on the HLT trigger path fired, as shown in Tab. 4.10.

The  $\tau_h$  leg is required to satisfy the Tight DeepTauVSmu and the VLoose DeepTauVSe working points of the DeepTau discriminator (the DeepTau discriminator working points and corresponding efficiencies, against electrons, muons and jets, are listed in Tab. 3.2, Section 3.2.4). The isolation requirement on the  $\tau_h$  leg is the Medium working point of the DeepTauVSjet discriminator. In addition, the requirements listed before on  $|d_{xy}|$  and  $|d_z|$  are applied to the  $\tau_h$ .

The invariant mass of the  $\tau$  pair candidates  $m_{\tau\tau}$  (here and in the following) is reconstructed using the SVfit algorithm [122], that is based on a likelihood function which quantifies the level of compatibility between a Higgs mass hypothesis and the measured

momenta of the visible tau decay products plus the missing transverse energy reconstructed in the event (Section 3.2.6). The SVfit algorithm improves the resolution on the invariant mass of the di-tau pair compared with the invariant mass from the visible decay product, therefore allowing a better signal to background discrimination.

In Fig. 4.1 the distributions of the invariant mass of the  $\tau_e\tau_h$  pair, for the 2016, 2017 and 2018 data taking periods are presented. The distributions are calculated at the baseline selection level in the  $\tau_e\tau_h$  channel, with all MC corrections are applied. There is an acceptable level of agreement between data and MC.

Object	Year	HLT path	Selection
$\tau_e$	all	all	Tight MVA ID $ d_{xy}  < 0.045\text{cm}$ , $ d_z  < 0.2\text{cm}$
	2016	single-e	$p_T > 26\text{ GeV}$ , $ \eta  < 2.1$
	2017	single-e	$p_T > 33\text{ GeV}$ , $ \eta  < 2.1$
		cross $\tau_e\tau_h$	$p_T > 25\text{ GeV}$ , $ \eta  < 2.1$
2018	single-e	$p_T > 33\text{ GeV}$ , $ \eta  < 2.1$	
	cross $\tau_e\tau_h$	$p_T > 25\text{ GeV}$ , $ \eta  < 2.1$	
$\tau_h$	all	all	$ d_z  < 0.2\text{cm}$ MVAWP: Medium VS jet, Tight VS $\mu/e$
	2016	single-e	$p_T > 20\text{ GeV}$ , $ \eta  < 2.3$
	2017	single-e	$p_T > 20\text{ GeV}$ , $ \eta  < 2.3$
		cross $\tau_e\tau_h$	$p_T > 35\text{ GeV}$ , $ \eta  < 2.1$
2018	single-e	$p_T > 20\text{ GeV}$ , $ \eta  < 2.3$	
	cross $\tau_e\tau_h$	$p_T > 35\text{ GeV}$ , $ \eta  < 2.1$	
$\tau_e\tau_h$ pair	all	all	$\Delta R(e, \tau_h) > 0.5$

Table 4.10: Offline selections for the  $\tau_e\tau_h$  channel for the three years of data taking. Explanation of the selections can be found in the text.

#### 4.4.1.2 $\tau_\mu\tau_h$ final state selection

The muons are reconstructed as described in Section 3.2.2, and the hadronic taus as described in Section 3.2.4, following the standard CMS algorithm. The legs are required to have opposite electric charge and a spatial separation  $\Delta R(\tau_\mu, \tau_h) > 0.5$ .

The identification of genuine prompt muons, is based also on a MVA approach, requiring also a Tight working point, and in addition isolation requirements are applied, requiring  $I_{\text{PR, rel}}^\ell < 0.15$  (see Eq. 3.1 in Section 3.2.2.2).

It is required  $|d_{xy}| < 0.045\text{ cm}$  (transverse) and longitudinal impact parameter  $|d_z| < 0.2\text{ cm}$ . The same requirements apply for the  $\tau_h$ .

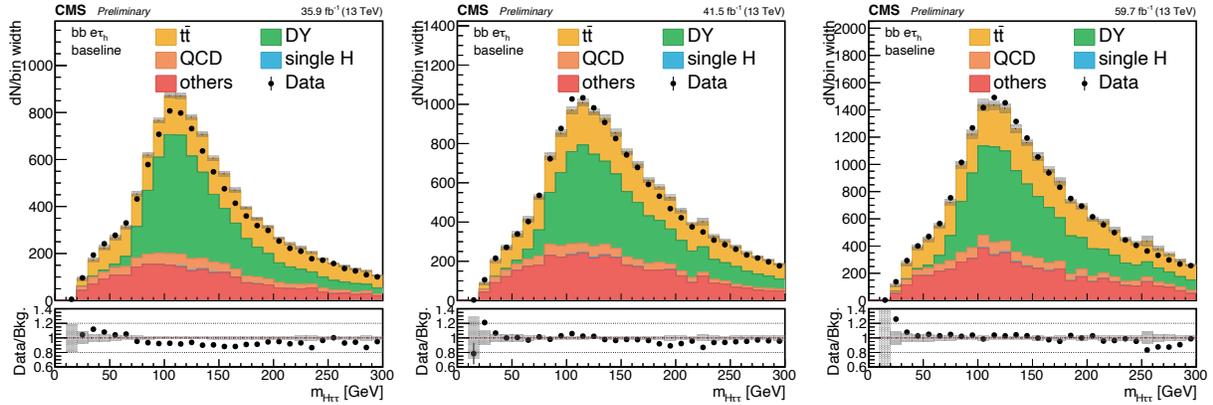


Figure 4.1: Control plots for the  $\tau_e\tau_h$  channel from left to right the 2016, 2017 and 2018 data taking periods. The shaded band in the plots represents the statistical uncertainty only.

Object	Year	HLT path	Selection
$\tau_\mu$	all	all	Tight Muon ID and $I_{\text{rel}}^\mu < 0.15$ $ d_{xy}  < 0.045\text{cm}$ , $ d_z  < 0.2\text{cm}$
	2016	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 23$ GeV, $ \eta  < 2.1$ $p_T > 20$ GeV, $ \eta  < 2.1$
	2017	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 25$ GeV, $ \eta  < 2.1$ $p_T > 21$ GeV, $ \eta  < 2.1$
	2018	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 25$ GeV, $ \eta  < 2.1$ $p_T > 21$ GeV, $ \eta  < 2.1$
$\tau_h$	all	all	$ d_z  < 0.2\text{cm}$ MVA WP: Medium VS jet, Tight VS $\mu$ , VLoose VS e
	2016	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 20$ GeV, $ \eta  < 2.3$ $p_T > 25$ GeV, $ \eta  < 2.1$
	2017	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 20$ GeV, $ \eta  < 2.3$ $p_T > 32$ GeV, $ \eta  < 2.1$
	2018	single- $\mu$ cross $\tau_\mu\tau_h$	$p_T > 20$ GeV, $ \eta  < 2.3$ $p_T > 32$ GeV, $ \eta  < 2.1$
$\tau_\mu\tau_h$ pair	all	all	$\Delta R(\mu, \tau_h) > 0.5$

Table 4.11: Offline selections for the  $\tau_\mu\tau_h$  channel for the three years of data taking.

The  $p_T$  thresholds of both legs depends on the HLT trigger path fired, as shown in Tab. 4.11.

The  $\tau_h$  leg must satisfy the Tight DeepTauVSmu and the VLoose DeepTauVSe working points of the DeepTau discriminator.

In Fig. 4.2 the distributions of the invariant mass of the  $\tau_\mu\tau_h$  pair estimated using the SVfit algorithm, for the 2016, 2017 and 2018 data taking periods are presented. The distributions are calculated at the baseline selection level in the  $\tau_\mu\tau_h$  channel, with all MC corrections are applied. There is an acceptable level of agreement between data and MC.

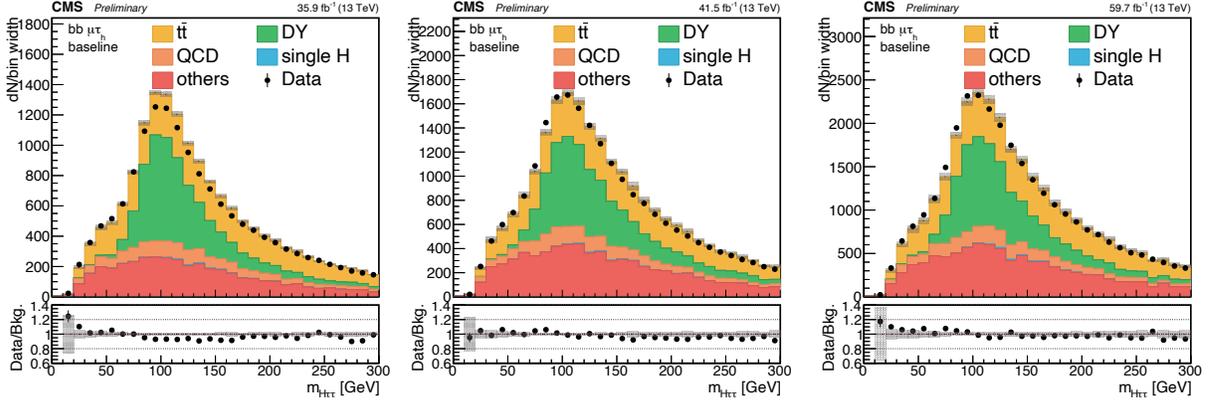


Figure 4.2: Control plots for the  $\tau_\mu\tau_h$  channel from left to right the 2016, 2017 and 2018 data taking periods. The shaded band in the plots represents the statistical uncertainty only.

#### 4.4.1.3 $\tau_h\tau_h$ final state selection

There are required two  $\tau_h$  with: opposite electric charge, an angular separation  $\Delta R(\tau_h, \tau_h) > 0.5$ , both with a longitudinal impact parameter  $|d_z| < 0.2$  cm, satisfying the VLoose DeepTauVSmu and the VVLoose DeepTauVSe working points of the discriminators, and the isolation requirement of a Medium working point of the DeepTauVSjet algorithm. In addition, similarly as for the other channels, the  $p_T$  threshold on the  $\tau_h$  depends on the HLT trigger path fired, as listed in Tab. 4.12

The two  $\tau_h$  are sorted by decreasing  $p_T$  inside the pair (i.e.  $p_T(\tau_1) > p_T(\tau_2)$ ). In the case where several pairs of  $\tau_h$  that meet the criteria are found in the event, the pair with the most isolated  $\tau_1$  is preferred, and if the pair share the same isolation value, the variable  $p_T$  of the legs is used as discriminator, choosing the pair with the highest  $p_T(\tau_1)$ , and as a last resort the isolation of  $\tau_2$  leg is used, selecting the pair with the most isolated  $\tau_2$ .

In Fig. 4.3 the distributions of the invariant mass of the  $\tau_h\tau_h$  pair estimated using the SVfit algorithm, for the 2016, 2017 and 2018 data taking periods are presented. The distributions are calculated at the baseline selection level in the  $\tau_h\tau_h$  channel, with all MC corrections are applied. There is an acceptable level of agreement between data and MC.

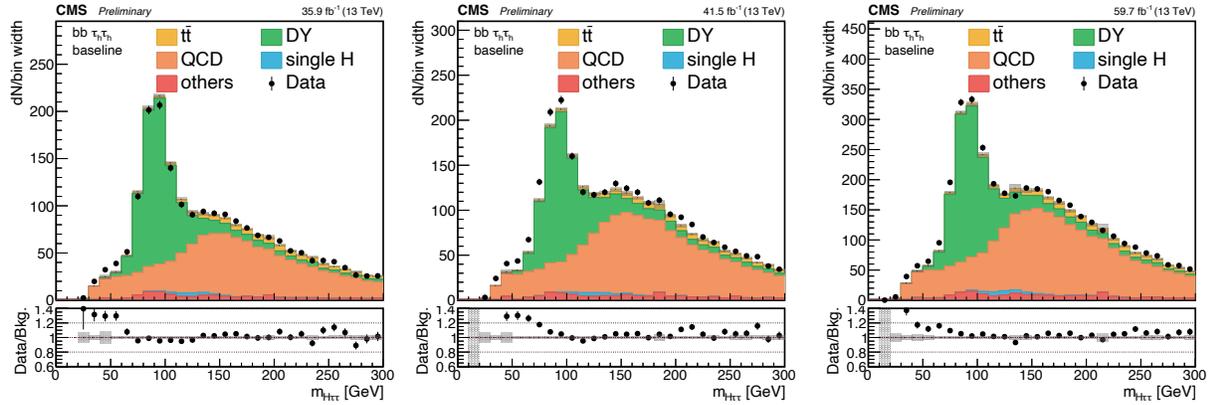


Figure 4.3: Control plots for the  $\tau_h\tau_h$  channel from left to right the 2016, 2017 and 2018 data taking periods. The shaded band in the plots represents the statistical uncertainty only.

Object	Year	HLT path	Selection
each $\tau_h$	all	all	$ d_z  < 0.2\text{cm}$ MVAWP: Medium VS jet, VLoose VS $\mu$ , VVLoose VS e
	2016	di- $\tau_h$	$p_T > 40\text{ GeV}$ , $ \eta  < 2.1$
	2017	di- $\tau_h$ VBF $H \rightarrow \tau_h\tau_h$	$p_T > 40\text{ GeV}$ , $ \eta  < 2.1$ $p_T > 25\text{ GeV}$ , $ \eta  < 2.1$
	2018	di- $\tau_h$ VBF $H \rightarrow \tau_h\tau_h$	$p_T > 40\text{ GeV}$ , $ \eta  < 2.1$ $p_T > 25\text{ GeV}$ , $ \eta  < 2.1$
$\tau_h\tau_h$ pair	all	all	$\Delta R(\tau_h, \tau_h) > 0.5$

Table 4.12: Offline selections for the  $\tau_h\tau_h$  channel for the three years of data taking. Explanation of the selections can be found in the text.

#### 4.4.1.4 Third lepton veto

In the analysis events are rejected if they contain an electron or a muon, in addition to the signal electron (muon) for the  $\tau_e\tau_h$  ( $\tau_\mu\tau_h$ ) channel, that passes the following selection:

- An electron of  $|\eta_e| < 2.5$  and  $p_T > 10\text{ GeV}$ . The electron passes the loose MVA iso-identification criteria (`mvaEleID-Fall117-iso-v2-wp90`) OR the logical AND between the loose MVA non-iso-identification criteria (`mvaEleID-Fall117-noIso-v2-wp90`) and the relative isolation requirement  $I_{\text{PR,rel}}^\ell < 0.3$ . The reconstructed electron production vertex must be close to the main primary vertex (the first of the `offlineSlimmedPrimaryVertices` collection) within a distance  $d_{xy} < 0.045\text{ cm}$  and  $d_z < 0.2\text{ cm}$ .

- A muon of  $|\eta_\mu| < 2.4$ ,  $p_T > 10$  GeV and passing the tight or medium particle-flow muon identification criteria, plus the relative isolation requirement  $I_{\text{PR, rel}}^\ell < 0.3$ . The reconstructed muon production vertex must be close to the main primary vertex (the first of the `offlineSlimmedPrimaryVertices` collection) within a distance  $d_{xy} < 0.045$  cm and  $d_z < 0.2$  cm

#### 4.4.2 H $\rightarrow$ bb and VBF-jets selection

This step of the analysis aims to select the two b-jets from the decay of the other Higgs boson. The following selection is applied in order to choose the b-jets candidates to form the H  $\rightarrow$  bb decay, and the two additional forward jets present in the case of the VBF production.

Two b-jets candidates are required with a  $p_T > 20$  GeV and  $|\eta| < 2.4$ , in order to be within the tracker acceptance. This is necessary since b-tagging algorithms use tracks and vertices information. In order to avoid overlap between the reconstructed b-jet and the  $\tau_h$  candidates, an angular separation between each jet and the  $\tau_h$  candidates of  $\Delta R(\tau_h, jet) > 0.5$  is required.

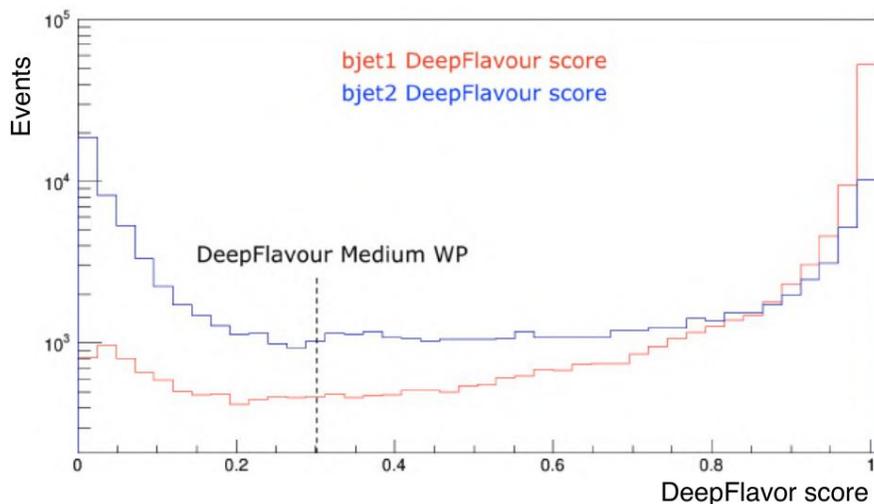


Figure 4.4: Distribution of the DeepFlavour score of the two jets candidates with the highest score value, corresponding to two real b-jet from H  $\rightarrow$  bb, using a 2017 GGF SM simulated signal sample.

At the beginning of the analysis the two jets with the highest score from DeepFlavour tagger has been used to select the pair of b-jets forming the H  $\rightarrow$  bb. As shown in Fig. 4.4, very often the second jet does not satisfy the b-tag requirement of a Medium WP of the DeepFlavour tagger. This spoils the efficiency for the signal events and increases the probability that one of the two b-jet candidate is fake.

The  $H \rightarrow bb$  selection has been improved respect to the previous published  $bb\tau\tau$  result on 2016 data [115], by developing a new algorithm to select the pair of b-jet to form the  $H \rightarrow bb$  candidate. This algorithm, called HH-Btag, is based on a recurrent neural network and is described in this thesis in Chapter 6, and in Ref. [123]. The information exploited by HH-Btag includes the scores of the DeepFlavour discriminator of jets in the event. For each event, to each b-jet candidate is assigned a score by the HH-Btag algorithm: the two jets with the highest scores are taken to be the two b-jets candidates originating from the decay of the Higgs boson. The first b-jet candidate is the one with the highest HH-BTag score, and the second b-jet candidate is the second highest HH-BTag score.

The jets that remain, once the b-jet pair candidates is selected, that have  $p_T > 30$  GeV and  $|\eta| < 4.7$ , are joined in pairs, and the pair with the highest invariant mass ( $m_{jj}$ ) is chosen as the VBF-jets candidate pair. In case the event was selected by the VBF trigger (Section 4.3), three additional requirements, are then made to the VBF-jets candidates:  $m_{jj} > 800$  GeV,  $p_T > 140$  GeV for the leading VBF-jet candidate, and  $p_T > 60$  GeV for the trailing VBF-jet.

In Fig. 4.5 the data and background event distribution as a function of the mass of the  $H \rightarrow bb$  candidate ( $m_{bb}$ ), for the three channels and data taking periods are shown. There is an acceptable level of agreement between data and MC.

#### 4.4.2.1 $H \rightarrow bb$ categories

The events are separated into three orthogonal categories, which represent different regimes for bb pair production, depending on the separations of the b-jets:

- $\Delta R(bb) > 0.8$  : jets are reconstructed as separated object with AK4, and are defined as resolved jets.
- $0.4 < \Delta R(bb) < 0.8$  : the jets partially overlap and are reconstructed both as separated (boosted jet) and as a merged jet (fat jet).
- $\Delta R(bb) < 0.4$  : these highly boosted jets are merged and reconstructed only as a fat jet.

The scenario of highly boosted jets is never reached for non-resonant signal, thus it is not relevant for most of the signals under study. Therefore, the events are divided into resolved and boosted categories in order to increase the analysis sensitivity by profiting of the different regime for the bb pair production.

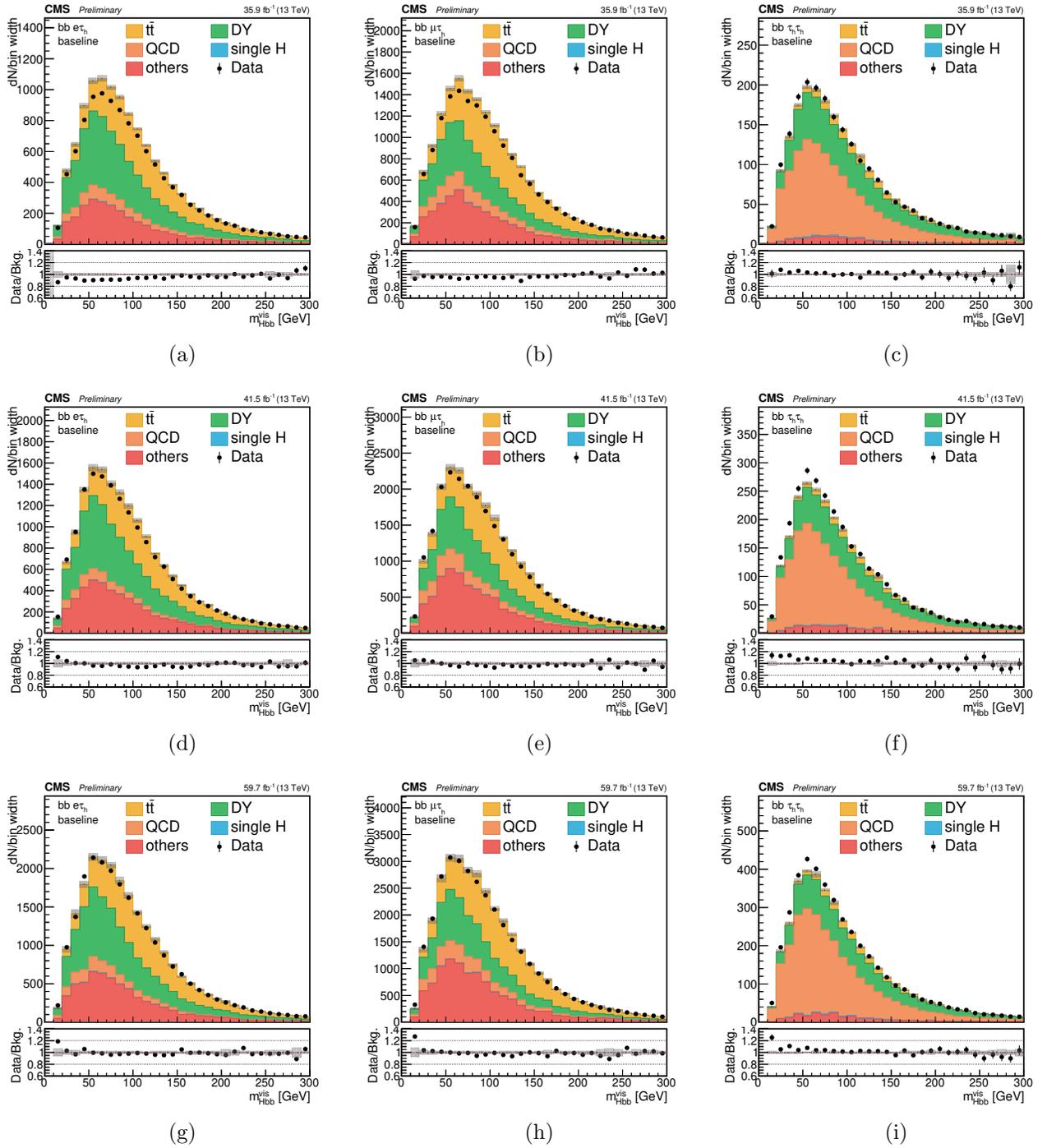


Figure 4.5: Data and background event distribution as a function of the  $H \rightarrow bb$  candidate ( $m_{bb}$ ). On each row the three channels for the same year are shown, and each column from top to bottom represent the 2016, 2017, and 2018 data taking periods, respectively.

The events from the resolved category are further categorized based on the b-tagging requirements they satisfy. The b-tagging requirement allows to reduce the contribution of the backgrounds with jets originating from light quarks. The resolved category is divided into:

- **Resolved 2jet–1tag (res1b):** contains events where only one of the two b-jet candidates passes the Medium working point of the DeepFlavour discriminator for all the final states.
- **Resolved 2jet–2tag (res2b):** contains events where both b-jet candidates passes the Medium working point of the DeepFlavour discriminator for all the final states. This category has a lower statistics with respect to the previous category, but it is the most signal-sensitive category.

An additional category, meant only for the VBF production mechanism, is introduced. Events are added to this category if they have two VBF-jet candidates, as described before, and additionally they have an invariant mass  $m_{VBF-jj} > 500$  GeV and  $\Delta\eta_{VBF-jj} > 3$ . The set of requirements for the VBF category, referred to as VBF tag, are chosen in order to find a compromise between background events suppression and signal events preservation. The impact of these selections can be seen in Fig. 4.6 where the additional two jet pair invariant mass versus angular separation is reported for two simulated samples: VBF di-Higgs SM production and  $t\bar{t}$  background. Whenever an event has two VBF-jet candidates, and it does not pass the VBF tag requirements, it is classified in the GGF categories as boosted or resolved.

### 4.4.3 HH signal region

In the previous steps, in each event the two  $\tau$  and b-jet candidates, expected in the double Higgs boson final state were selected. However, these events are still expected to be background dominated. In order to improve the separation between signal and background the invariant masses of the Higgs boson candidates  $H_{\tau\tau}$  and  $H_{bb}$  reconstructed from the  $\tau$  and b-jet pair candidates, are used.

The invariant mass of the  $\tau$  candidates ( $m_{\tau\tau}$ ) is reconstructed using the SVfit algorithm. The invariant mass of the b-jet candidates ( $m_{bb}$ ) is reconstructed using the invariant mass of the two b-jets.

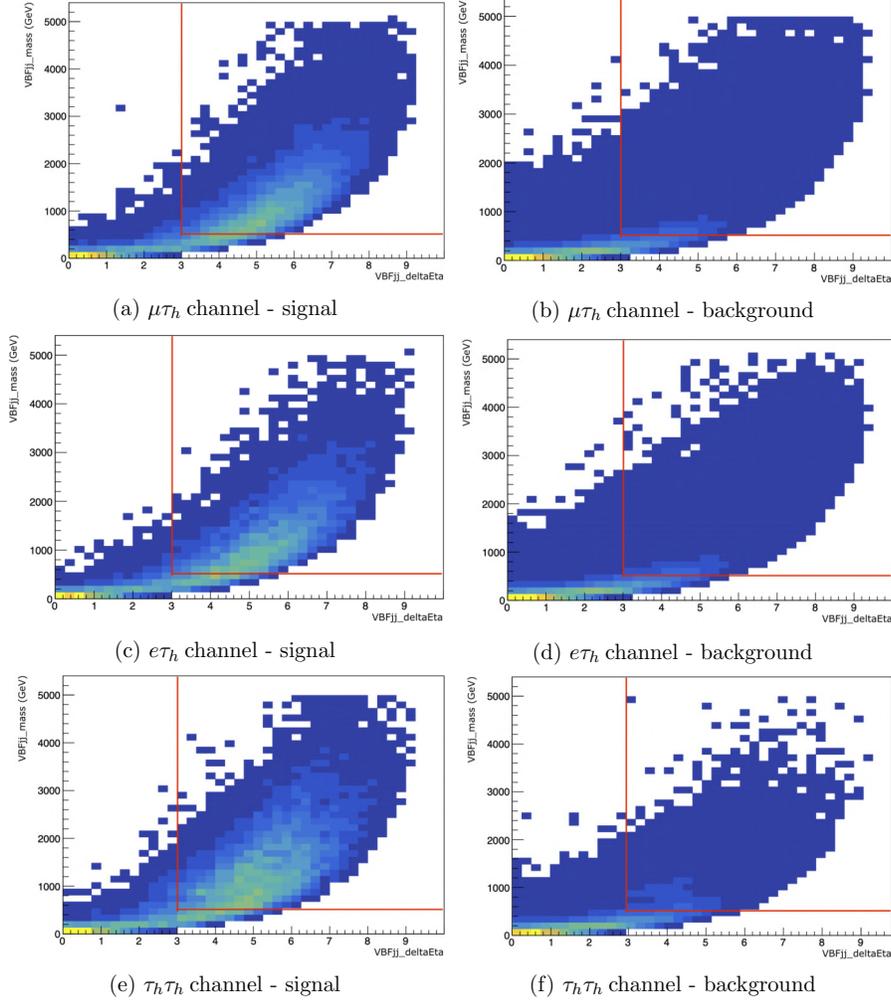


Figure 4.6: 2D distributions of  $(VBFjj_{\Delta\eta}, VBFjj_{mass})$  for the VBF SM signal (a, c, e) and for the  $t\bar{t}$  background (b, d, f), for the 2018 MC simulation. Only events belonging to the VBF category described in the text are considered.

To define the signal region in the resolved and boosted categories of the analysis an elliptical cut is imposed on the SVfit mass ( $m_{\tau\tau}$ ) and on the reconstructed invariant mass of the bb pair ( $m_{bb}$ ). The “elliptical” selection was already used in Ref. [115], leading to a region with higher signal over background ratio, compared to the previous analyses [124], where a “squared” selection was used. For the resolved category the selection is:

$$\frac{(m_{\tau\tau} - 129 \text{ GeV})^2}{(53 \text{ GeV})^2} + \frac{(m_{bb} - 169 \text{ GeV})^2}{(145 \text{ GeV})^2} < 1. \quad (4.2)$$

Instead, for the boosted category, the selection is:

$$\frac{(m_{\tau\tau} - 128 \text{ GeV})^2}{(60 \text{ GeV})^2} + \frac{(m_{bb} - 159 \text{ GeV})^2}{(94 \text{ GeV})^2} < 1. \quad (4.3)$$

The parameters from Eq. 4.2 and 4.3 are the measured resolution on the invariant mass for the  $\tau\tau$  and  $bb$  objects. The ellipse is centered in the position of the expected reconstructed Higgs boson ( $m_H = 125$  GeV) peak in the  $m_{\tau\tau}$  and  $m_{bb}$  distributions. The values of the mass cuts were defined using a random search for the offsets and resolutions, of both the  $m_{\tau\tau}$  and  $m_{bb}$  distributions, that minimized the weighted background acceptance for a weighted signal acceptance above 90%. The selection described above profits from the different kinematics of the  $HH \rightarrow bb\tau\tau$  decay in the different categories to reduce the background contributions and only, and select events that are compatible with the  $HH$  decay hypothesis.

As an example, the  $m_{bb}$  versus the  $m_{\tau\tau}$  for the signal (GGF SM di-Higgs production) and for the background is shown in Fig. 4.7 for each of the three years of data taking, before the invariant mass cuts are applied and requiring that both b-jet candidates pass the Medium WP. The region defined by the elliptical selection is also shown superimposed on the plot.

#### 4.4.4 Event categorization and signal extraction

As described in Section 4.4.2.1, three categories are defined to better identify the GGF events: res1b, res2b and boosted. A fourth category, VBF, is defined to isolate signal events produced through the Vector Boson Fusion mechanism. Since a substantial contamination from  $HH$  production via GGF might be present in the VBF category, a multi-classification approach is the most suitable option to enhance the analysis sensitivity. A multi-classifier based on a deep neural network, described in detail in Ref. [125], and summarized below 4.4.4.1, has been developed for this purpose, and separates the VBF categories in five “multi-categories”: one for genuine VBF events, one for the GGF contamination, one for  $t\bar{t}H$ , one for  $t\bar{t}$  contamination, and one for  $Z\gamma^* \rightarrow \ell\ell$  contamination. The simulated samples of the backgrounds ( $t\bar{t}H$ ,  $t\bar{t}$  and  $Z\gamma^* \rightarrow \ell\ell$  are described in Section 4.5.2)

A diagram that describes how the events are divided into the defined categories, and the discriminating variables used for signal extraction is shown in Fig. 4.8.

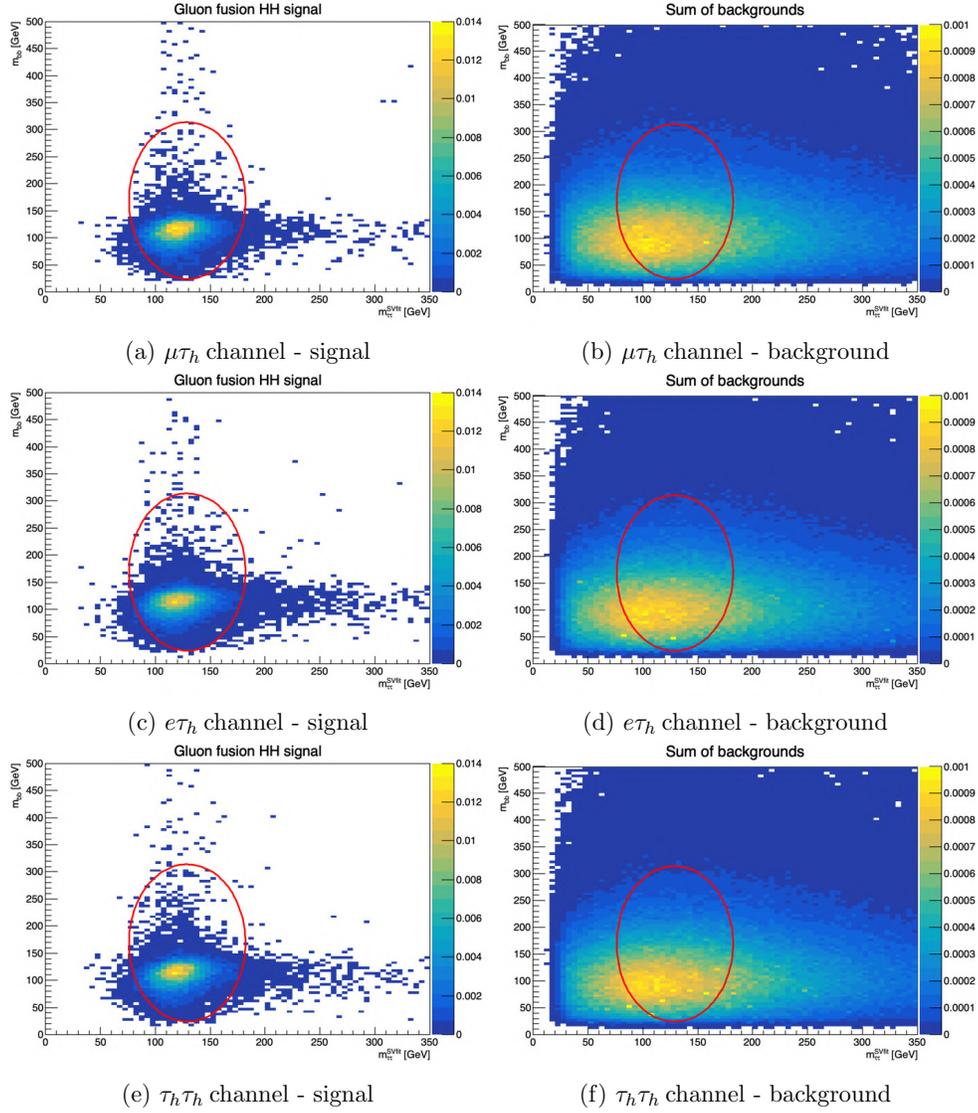


Figure 4.7: 2D distributions  $m_{bb}$  versus the  $m_{\tau\tau}$  for the Gluon Fusion SM signal (a, c, e) and for the sum of MC backgrounds (b, d, f). Plots show the normalized processes yields after the  $\tau\tau$  and  $bb$  candidates selections and before the invariant mass requirements for 2016 (a, b), 2017 (c, d), and 2018 (e, f). The presence of two  $b$ -tagged jets passing the medium WP is required; the three  $\tau\tau$  decay channels considered in the analysis are merged together. The red line shows the region selected by the resolved elliptical cut.

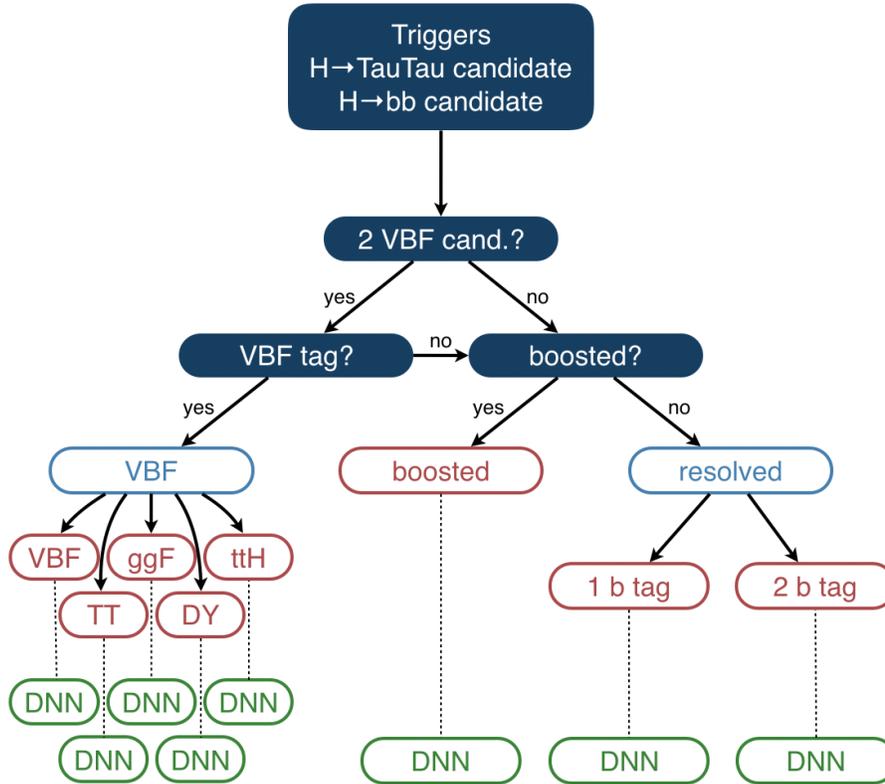


Figure 4.8: Scheme of the event categorization and discriminating variables used in each category. The events classified as VBF are split accordingly to the “most probable process” predicted by the multi-class neural network [125]. The green boxes “DNN”, represent the Gluon Fusion optimized DNN score which is used for signal extraction both in the GGF and in the VBF categories [126].

#### 4.4.4.1 Multi-class categorization for VBF categories

Despite the requirements that help reduce contamination from GGF production in the VBF categories, there is a substantial amount of GGF events that may still be included in the VBF categories. As a consequence, a multi-classification approach represents the most suitable option to further split events in this category and enhance the analysis sensitivity. This multi-classification used in the analysis is described in detail in Ref. [125].

In the multi-class categorization strategy, machine learning techniques are used to assign probability estimates for an event to belong to categories associated to any of the relevant physics processes under consideration.

### Categorization Strategy

The sensitivity of the analysis is connected to the ability to measure and constrain particular nuisance parameters, described in Section 4.7. This can be achieved by creating categories that are solely enriched with one of the relevant physics processes. By doing so, interference between involved nuisance parameters is mitigated, rendering their measurement in the fitting procedure less perturbed such that stronger model constraints can be deduced, thus leading to an enhanced analysis sensitivity.

Unlike binary classification that distinguishes between two classes of events, i.e., usually *one* signal process and *all* background processes combined, multi-class classification has the advantage that different physics processes can be treated with equal importance. This is especially desirable in scenarios where multiple signal processes with different topologies are to be examined, or where several background processes, affected by large systematic uncertainties, contribute significantly to the overall expected background.

The search for HH production in the  $bb\tau\tau$  final state is attributed both of these characteristics. On one hand, as explained above, the distinction between GGF- and VBF-induced HH production could give rise to the  $VVHH$  coupling strength that can only be probed by the latter production mode. On the other hand, the analysis is dominated by backgrounds originating from QCD multijet and Drell-Yan processes, as well as processes involving top quark pairs ( $t\bar{t}$  and  $t\bar{t}H$ ). While the two former impact the analysis greatly due to their vast rate, the two latter processes exhibit similarities in their final state with respect to the expected signal signatures and are therefore to be included in the list of background processes to distinguish. Moreover, contributions from  $t\bar{t}$  are subdivided further according to the decays of the two W bosons, namely dilepton (DL), semi-lepton (SL), and full-hadron decays (FH), and contributions from  $t\bar{t}H$  subdivided into events with either  $H \rightarrow bb$  or  $H \rightarrow \tau\tau$  decays, which both have partial similarities to the signal processes. This results in a total of six significant background processes.

The multi-class classification strategy can be described as follows with a supporting illustration shown in Fig. 4.9. Per event, a predefined set of input variables is fed into a network whose architecture and weights are subject to a training and optimization process. The network evaluates the variables and outputs per event a vector of nine floating point numbers with a sum of one, where each number describes a probability estimate for the event to originate from a particular process under investigation. These processes are defined based on the six significant background processes listed above, the HH (GGF) signal process, the HH (VBF) signal process with  $\kappa_{2V} = 1$ , and the same HH (VBF) signal process with  $\kappa_{2V} = \{0, 2\}$  (definition of the couplings modifiers  $\kappa$  is presented in Section 1.3.2.2).

The decision to split the HH (VBF) process into a SM- and a BSM-like component is based on the expected variation of the event topology for different values of  $\kappa_{2V}$ . Therefore,

only passing information about the SM configuration could lead to an artificial bias of the neural network performance towards the SM-like realization of the HH (VBF) process, impeding its ability to exclude certain BSM scenarios. Contributions from QCD are determined by means of a data-driven approach (Section 4.5.2) and are therefore not attributed an output value by the DNN.

The subsequent categorization is performed in two steps. First, the output values of the three  $t\bar{t}$ , the two  $t\bar{t}H$ , and the two HH (VBF) sub-processes, respectively, are added such that a total of five distinct outputs remains: HH (VBF), HH (GGF),  $t\bar{t}H$ ,  $t\bar{t}$ , and Drell-Yan. While this *merging* of output values is found to have only a negligible impact on the signal extraction results, it allows to pass more useful information to the neural network during the training process, eventually improving its separation performance. The eventual categorization into one of the remaining five *classes* is then derived according to the class that received the highest output value. This approach unambiguously assigns each event to a category that is intended to be enriched with events of similar topology corresponding the *most probable process*.

Following this categorization scheme, the choice of the discriminant to be used in the signal extraction procedure remains arbitrary (Fig. 4.8). It is employed the DNN output distribution, as described Sec. 4.4.4.2, due to the similarities between the GGF signal process and the VBF signal process with non-SM  $\kappa_{2V}$  values (especially  $\kappa_{2V} = \{0, 2\}$ ).

## Input Features

Table 4.13 lists 74 kinematic observables that are considered as input variables to the neural networks.

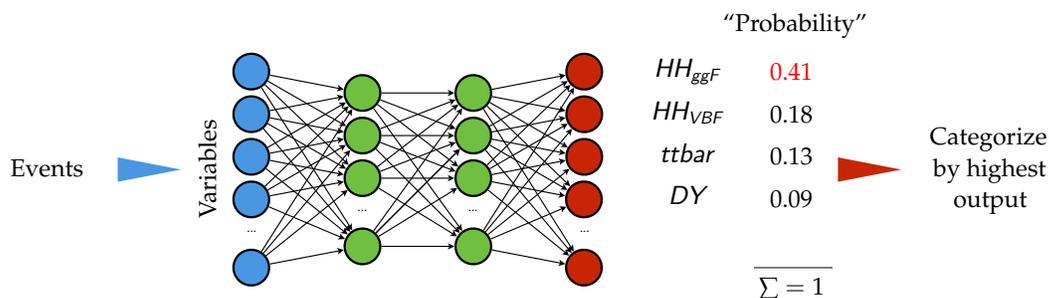


Figure 4.9: Concept of the event categorization approach using DNNs that perform a multi-class classification. Each event is attributed even values that express its probability to originate from either the  $HH_{GGF}$  or the  $HH_{VBF}$  signal, or one of the six relevant background processes to consider. An event is categorized by the process class that received the highest probability (HH (VBF) in the depicted example).

Variable name(s)	Description
$is_{201\{6,7,8\}}$	Flag denoting the campaign / year of input events.
$is_{\{e\tau, \mu\tau, \tau\tau\}}$	Flag denoting reconstructed lepton channel.
$lep\{1,2\}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the two leptons.
$bjet\{1,2\}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the two b-tagged jets.
$bjet\{1,2\}_{\{deepflavor,hbtag\}}$	b-tag and HH-b-tag of the two b-tagged jets.
$vbfjet\{1,2\}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the two VBF jets.
$vbfjet\{1,2\}_{\{deepflavor,hbtag\}}$	b-tag and HH-b-tag of the two VBF jets.
$ctjet\{1,2,3\}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the three additional central jets.
$ctjet\{1,2,3\}_{\{deepflavor,hbtag\}}$	b-tag and HH-b-tag of the three additional central jets.
$fwjet\{1,2\}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the two additional forward jets.
$met_{\{pt,\phi\}}$	Missing transverse energy and azimuthal direction.
$bh_{\{e,pt,\eta,\phi\}}$	Four-vector components of the reconstructed $H_{bb}$ candidate.
$\tau_{uh}_{\{e,pt,\eta,\phi\}}$	Four-vector components of the reconstructed $H_{\tau\tau}$ candidate.

Table 4.13: Names and descriptions of variables considered as inputs to the neural networks. In case an event exhibits less VBF jets, additional central jets or forward jets than described, the network receives 0 values instead.

#### 4.4.4.2 Deep neural network for signal extraction

The Deep neural network (DNN) is trained to classify the events in each category as either signal or background by assigning a single prediction per event. A prediction closer to zero indicates “background-like”, while a prediction closer to one indicates “signal-like” events. The DNN is composed of two discriminators, each consisting of ten neural networks trained via ten-fold stratified cross-validation of the training set. The approach used to train the DNN [127, 128, 129], follows that used in the CMS HH High Luminosity LHC projection analysis for the  $bb\tau\tau$  channel [130]. In this approach the simulated events are split evenly into two sets; each discriminator of the DNN is trained on different halves of the events. A total of 26 features (selected from a starting pull of over 100) are used as input to the DNN.

#### Input features

The final continuous input features are listed here in descending order of importance. The ranking is based on Random Forest models: features at the top are consistently important, features in the middle are usually found to be important, and features near the bottom are only important at least 30% of the time.

- DeepFlavour charm tagger  $CvsB$ , binned in working points, of the first b-jet, as defined by the BtagPOGin [131].
- DeepFlavour charm tagger  $CvsL$ , binned in working points, of the first b-jet, as

defined by the BtagPOGin [131].

- The  $\chi^2$  of the kinematic fit to the invariant mass of the  $HH \rightarrow bb\tau\tau$  system.
- The invariant mass of the  $HH \rightarrow bb\tau\tau$  system from the kinematic fit.
- The invariant mass of the  $\tau\tau$  system from the SVfit algorithm.
- The product  $\Delta R(\tau_1, \tau_2) \cdot p_T(H_{\tau\tau}^{SVfit})$ .
- The transverse mass of the first tau lepton: the  $e$  or  $\mu$  in case of semi-leptonic decays, or the most isolated  $\tau_h$  in the fully hadronic decay case.
- The transverse momentum of the second tau lepton: the  $\tau_h$  case of semi-leptonic decays, or the second most isolated  $\tau_h$  in the fully hadronic decay case.
- The  $\Delta R$  separation between the two tau lepton candidates.
- The  $\Delta\phi$  separation between the  $H \rightarrow \tau\tau$  candidate, reconstructed with the SVfit algorithm, and the MET.
- The mass of the  $H \rightarrow bb$  candidate.
- The HH-Btag score of the second b-jet.
- The invariant mass of the system  $H_{bb} + H_{\tau\tau}^{SVfit}$ .
- The  $\Delta\phi$  separation between the  $H_{bb}$  and the  $H_{\tau\tau}^{SVfit}$  systems.
- The transverse momentum of the  $H_{bb}$  system.
- The  $\Delta R$  separation between the two tau candidates in the rest frame of the  $H_{\tau\tau} + MET$  system.
- The transverse momentum of the first tau lepton: the  $e$  or  $\mu$  in case of semi-leptonic decays, or the most isolated  $\tau_h$  in the fully hadronic decay case.
- The transverse momentum of the first b-jet.
- Angle between the decay planes of the four final state particles expressed in the rest frame of the HH system.
- Cosine of theta angle between the second tau lepton and the direction of flight of  $H_{\tau\tau} + MET$  system in the  $H_{\tau\tau} + MET$  rest frame.

In addition, six categorical features are used as input to the neural network:

- If the event is boosted or not.
- The  $\tau\tau$  decay mode.
- If the event has two VBF jet candidates, as defined in Section 4.4.2.
- Highest working point of the DeepFlavour algorithm passed by the first b-jet.
- Highest working point of the DeepFlavour algorithm passed by the second b-jet.
- The year of the data taking.

## 4.5 Sample composition modeling

The sensitivity of the search of  $HH \rightarrow bb\tau\tau$ , rely on an accurate modelling of each process contributing to the background collected by the trigger (Section 4.3). At the same time, an accurate modeling of the signal and backgrounds involved is the key to optimize the analysis strategy and techniques, and to obtain a valid comparison between the observed data and the theoretical predictions. In this analysis, the signature of the signal process can be reproduced by a large variety of background processes.

The signal processes are modeled using simulated samples. The modeling of the signal for the different mechanism of production and for different scenarios is described in Section 4.5.1.

For some background processes, where the statistics necessary is prohibitive and/or the modelling uncertainties are big, different approaches are used. Three of the main sources of background (Section 4.1.2) are modeled in a such way: QCD multi-jets, Drell-Yan  $Z \rightarrow \ell\ell$  plus jets events (DY) and  $t\bar{t}$  process. The QCD modeling uses a pure Data-Driven method to determine the differential distribution (shape) and the Normalization of the QCD background. For DY ( $Z \rightarrow \tau\tau$  plus jets) and  $t\bar{t}$  the shape is obtained from simulated samples and the normalization is obtained from data. All the other background considered for this analysis are modelled using simulation. The description of the background modelling is described in Section 4.5.2.

### 4.5.1 Signal modeling

The exploration of various BSM scenarios requires the modelling of the signal for several different sets of couplings (Section 1.3.2.3, and Tab. 1.4). Since only a limited set of simulated samples can be produced, weighting techniques are implemented both for the gluon fusion and the VBF signals, in order to model additional BSM scenario starting from a small set of fully simulated BSM signals.

For gluon fusion HH samples, two different reweighting methods are used: one applies to the samples simulated at leading order (LO), it is described in paragraph 4.5.1.1 and it was also used in the 2016  $b\bar{b}\tau\tau$  analysis [132] and in the 2016 HH CMS combination [133]. On one hand, this technique provides a large number of signal events: all the LO GGF data sets are merged together and reweighted to the target BSM scenario. On the other hand, the LO modelling does not consider parton emission at the matrix element level. Thus, a new reweighting procedure to be applied to NLO samples has been adopted by all HH analyses in view of the Run II Legacy results, and it is described in paragraph 4.5.1.2.

For the GGF, the LO samples are used to produce events for the training of the machine learning techniques implemented in this analysis (Sections 4.4.4.2 and 4.4.4.1 and Chapter 6), while NLO samples are used for the final signal extraction. The VBF modelling strategy is similar to that used for GGF NLO samples, as described in paragraph 4.5.1.3. Original HH non-resonant production samples are produced centrally by CMS.

#### 4.5.1.1 LO Gluon Fusion modelling

The weighting technique for the LO gluon fusion modelling allows various scenarios to be explored in terms of the effective Lagrangian parametrization described in [50], where the Higgs pair production is regulated by the five couplings  $y_t$ ,  $\lambda_{HHH}$ ,  $c_2$ ,  $c_{2g}$  and  $c_g$ ; the variations from the SM values of the couplings are expressed as  $k_\lambda = \lambda_{HHH}/\lambda_{HHH}^{\text{SM}}$  and  $k_t = y_t/y_t^{\text{SM}}$ , as described in Section 1.3.2.3.

The two Higgs bosons are produced back-to-back in the reference frame of the center of mass; before any hadronization effect, they have identical transverse momenta and opposite azimuthal angle,  $\theta^*$ , with respect to the beam axis. At this level, the kinematics of the event is totally determined by two parameters: the invariant mass of the HH system and  $\cos\theta^*$  between one Higgs boson and the beam axis [50]. These variables are exploited by the weighting procedure.

The events of all the gluon fusion HH signal samples are combined to build a 2D distribution as a function of  $m_{HH}$  and  $|\cos\theta^*|$ , computed using simulated Higgs boson properties after the hard scatter and before hadronization effects. An identical histogram is filled using the SM signal sample only. The content of a bin  $j$  in the two bidimensional distributions, normalised to unity, is denoted as  $f_{comb}^j$  and  $f_{SM}^j$ .

The ratio of the total HH cross section over the SM prediction, can be expressed as a function of the  $j$  bin number as

$$R_{\text{HH}}^j = \frac{\sigma_{\text{HH}}^j}{\sigma_{\text{HH}}^{j,\text{SM}}} = A_1^j k_t^4 + A_2^j c_2^2 + (A_3^j k_t^2 + A_4^j c_g^2) k_\lambda + A_5^j c_{2g}^2 + \quad (4.4)$$

$$(A_6^j c_2 + A_7^j k_\lambda k_t) k_t^2 + (A_8^j k_t k_\lambda + A_9^j c_g k_\lambda) c_2 + \quad (4.5)$$

$$(A_{10}^j c_2 c_{2g} + (A_{11}^j c_g k_\lambda + A_{12}^j c_{2g}) k_t^2) + \quad (4.6)$$

$$(A_{13}^j k_\lambda c_g + A_{14}^j c_{2g}) k_t k_\lambda + A_{15}^j c_g c_{2g} k_\lambda. \quad (4.7)$$

The ratio  $R_{\text{HH}}^j$  is computed using simulated events with different sets of couplings; thus, the  $A_i^j$  coefficients are extracted from its interpolation as a function of the couplings.

Finally, event-by-event weights are computed as

$$\omega = \frac{\Omega}{\sum_n \Omega} \quad (4.8)$$

where

$$\Omega(k_\lambda, k_t, c_2, c_g, c_{2g}; j) = \frac{f_{\text{SM}}^j}{f_{\text{comb}}^j} \cdot \frac{R_{\text{HH}}^j(k_\lambda, k_t, c_2, c_g, c_{2g})}{R_{\text{HH}}(k_\lambda, k_t, c_2, c_g, c_{2g})} \quad (4.9)$$

and the sum goes over the number of simulated events; thus, only the differential event distribution is modified and not the global normalization.

For each of the identified BSM benchmarks, signal samples are produced centrally by CMS at leading order (LO) precision with MADGRAPH5\_AMC@NLO; the samples used in this search are reported for each year of data taking in Tables 6.2, 6.3 and 6.4, marked as ‘‘GGF LO’’.

#### 4.5.1.2 NLO Gluon Fusion modelling

A different procedure, based on the sum of different samples, is used to reweight NLO samples due to the presence of an additional parton at matrix level that prevents the usage of the same technique used for LO samples.

At leading order the amplitude of the gluon fusion HH production process can be written as:

$$\mathcal{A} = \kappa_\lambda \kappa_t T + \kappa_t^2 B \quad (4.10)$$

where,  $T$  and  $B$  can be directly associated to the triangle and box diagrams of Fig. 1.8.

The HH production cross section via gluon fusion can thus be written as the square of the amplitude:

$$\sigma(\kappa_\lambda, \kappa_t) \sim |\mathcal{A}|^2 = \kappa_\lambda^2 \kappa_t^2 t + \kappa_t^4 b + \kappa_\lambda \kappa_t^3 i \quad (4.11)$$

where  $t = |T|^2$ ,  $b = |B|^2$ , and  $i = |TB^* + B^*T|$ .

Eq. 4.11 can also be rewritten in a matricial form as

$$\sigma(\kappa_\lambda, \kappa_\tau) = \mathbf{c}(\kappa_\lambda, \kappa_t) \cdot \mathbf{v} \quad (4.12)$$

where  $\mathbf{c}(\kappa_\lambda, \kappa_t) = (\kappa_\lambda^2 \kappa_t^2, \kappa_t^4, \kappa_\lambda \kappa_t^3)$  is the vector of the coupling functions and  $\mathbf{v} = (t, b, i)$  is the vector of the values of the three components.

For a fixed set of  $(\kappa_\lambda, \kappa_t)$  values, the HH signal can thus be described as the linear sum of three components.

Given three samples  $(s_1, s_2, s_3)$  produced at NLO using POWHEG with a specific choice of  $\kappa_\lambda$  and, denoting with  $(c_i^1, c_i^2, c_i^3)$  the 3 elements of the coupling vector for each  $s_i$  samples, we can express Eq. 4.11 in terms of a linear combination of these three samples:

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} c_1^1 & c_1^2 & c_1^3 \\ c_2^1 & c_2^2 & c_2^3 \\ c_3^1 & c_3^2 & c_3^3 \end{pmatrix} \begin{pmatrix} t \\ b \\ i \end{pmatrix}, \quad (4.13)$$

or again in matricial form:

$$\boldsymbol{\sigma} = \mathbf{C}\mathbf{v}. \quad (4.14)$$

Given that the three values of  $\sigma_i$  are provided from the POWHEG generator, the matrix  $\mathbf{C}$  can be inverted to express:

$$\mathbf{v} = \mathbf{C}^{-1}\boldsymbol{\sigma}, \quad (4.15)$$

so that by simple substitution Eq. 4.12 becomes:

$$\sigma(\kappa_\lambda) = \mathbf{c}^T(\kappa_\lambda)\mathbf{C}^{-1}\boldsymbol{\sigma}. \quad (4.16)$$

This can be applied to build the differential  $d\sigma/dx$  distributions as well.

A generic value of  $\kappa_\lambda$  can therefore be easily modelled by manipulating a few input histograms rather than going through an event-by-event reweighting.

The samples used are reported for each year of data taking in Tab. 4.14, marked as ‘‘GGF NLO’’.

	Data set Name
GGF NLO 2016	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
GGF NLO 2017	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
GGF NLO 2018	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Table 4.14: List of non-resonant GGF NLO signal samples MC datasets for 2016, 2017 and 2018.

#### 4.5.1.3 LO Vector Boson Fusion modelling

A procedure similar to the one described above, can be applied to the VBF modelling. The HH production cross section via VBF can be written as the square of the amplitude of the LO diagrams represented in 1.9, i.e.

$$\begin{aligned}
\sigma(c_V, c_{2V}, \kappa_\lambda) &\sim |Ac_V \kappa_\lambda + Bc_V^2 + Cc_{2V}|^2 = \\
&= ac_V^2 \kappa_\lambda^2 + bc_V^4 + cc_{2V}^2 + i_{ab} c_V^3 \kappa_\lambda + i_{ac} c_V c_{2V} \kappa_\lambda + i_{bc} c_V^2 c_{2V}
\end{aligned} \tag{4.17}$$

where  $a = |A|^2$ ,  $b = |B|^2$ ,  $c = |C|^2$  and  $i_{ij}$  are the interference terms. Therefore, the cross section, as well as any differential distribution  $d\sigma/dx$ , depends on six components.

Thus, the VBF signal can be modelled through the sum of six components  $\mathbf{V} = \{a, b, c, i_{ab}, i_{ac}, i_{bc}\}$ , each scaled by a function of  $c_V$ ,  $c_{2V}$  and  $\kappa_\lambda$ ; denoting as

$$\mathbf{K} = \{c_V^2 \kappa_\lambda^2, c_{2V}^4, c_{2V}^2, c_V^3 \kappa_\lambda, c_V c_{2V} \kappa_\lambda, c_V^2 c_{2V}\}$$

the vector of the functions of the couplings, the Eq. 4.17 can be expressed as

$$\sigma = \mathbf{K}^T \mathbf{V}. \tag{4.18}$$

However, the generator used does not allow for generating the individual  $\mathbf{V}_i$  components. Instead, they can be determined by solving a system of equations using six samples corresponding to different combinations of  $(c_V, c_{2V}, \kappa_\lambda)$ . Denoting these samples as  $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$ , where  $\sigma_i = \sigma(c_{V,i}, c_{2V,i}, \kappa_{\lambda,i})$ , they can be represented as

$$\boldsymbol{\sigma} = \mathbf{M} \mathbf{V}, \quad (4.19)$$

where  $\mathbf{M}$  is the  $6 \times 6$  coefficients matrix; its solution is

$$\mathbf{V} = \mathbf{M}^{-1} \boldsymbol{\sigma}. \quad (4.20)$$

Thus, the cross section  $\sigma_{target}$  of a given  $(c_V, c_{2V}, \kappa_\lambda)$  combination can be computed as

$$\sigma_{target} = [\mathbf{K}^T \mathbf{M}^{-1}] \boldsymbol{\sigma}. \quad (4.21)$$

The Eq. 4.21 can be equally applied to build the differential distribution as a function of a given observable and for a given  $(c_V, c_{2V}, \kappa_\lambda)$  combination; in that case, the unknowns  $\mathbf{V}(x)$  are a function of the observable  $x$  and

$$h(x)_{target} = [\mathbf{K}^T \mathbf{M}^{-1}] \mathbf{h}(x), \quad (4.22)$$

where  $\mathbf{h}(x)$  contains the differential distributions. Thus, the shape of a signal can be easily obtained by manipulating a few input histograms, rather than going through an event-by-event reweighting; to do so, only six fully simulated combinations of  $(c_V, c_{2V}, k_\lambda)$  are needed. The samples used are provided centrally by CMS and they are reported, for each year of data taking respectively, in Tables 6.2, 6.3 and 6.4, marked as ‘‘VBF LO’’.

## 4.5.2 Background modeling

The main background sources, after the selection criteria applied by trigger and analysis selections are the  $t\bar{t}$ , Drell-Yan and QCD multi-jet production, which are modelled with specific approaches. Other processes contributing to the background are modeled both in shape and event yield, simply using simulated samples. Among these: single top, single Higgs boson (SM),  $t\bar{t}$ , di-boson (WW, ZZ, and WZ), tri-boson (ZZZ, ZZW, ZWW, and WWW),  $t\bar{t}$  pairs in association with a single boson or a pair of vectors bosons.

The Drell-Yan and W boson production processes in association with jets, are produced centrally by CMS at LO, using for the modelling MADGRAPH5\_aMC@NLO with MLM merging [134]. Instead, other background processes like the single top, single Higgs, and  $t\bar{t}$  backgrounds are simulated at NLO precision, modelled with POWHEG. The background processes of diboson (WW, ZZ, and WZ) and triboson (ZZZ, ZZW, ZWW, and WWW) production are simulated at NLO precision with MADGRAPH5\_aMC@NLO, POWHEG, and

MADGRAPH5\_aMC@NLO with FFX merging [135] depending on the process. The production of  $t\bar{t}$  pairs in association with a single boson or a pair of vectors bosons are simulated with MADGRAPH5\_aMC@NLO both with and without FFX merging depending on the process. In Tab. 4.15 are reported the specific simulated processes and their cross section (with the exception of the single top and single Higgs boson (SM)).

Process	Cross section [pb]	
$W \rightarrow \ell \nu_\ell + \text{jets}$	$6.15 \times 10^4$	
$Z / \gamma^* \rightarrow \ell\ell + \text{jets}$	$6.07 \times 10^3$	
$W^- \text{ jj}$	20.25	} Electroweak
$W^+ \text{ jj}$	25.62	
$Z^+ \text{ jj}$	3.987	
$t\bar{t}$	831.59	
W channel	35.9 (35.9)	} single t( $\bar{t}$ )
t-channel	80.95 (136.02)	
WW		} Di-boson
$\rightarrow 4q$	51.72	
$\rightarrow 2\ell 2q$	50.0	
$\rightarrow 2\ell 2\nu$	12.18	
ZZ		} Di-boson
$\rightarrow 2\ell 2q$	5.52	
$\rightarrow 2q 2\nu$	4.07	
$\rightarrow 4\ell$	1.26	
$\rightarrow 2\ell 2\nu$	0.564	
WZ		} Tri-boson
$\rightarrow \ell\nu qq$	10.71	
$\rightarrow 2\ell 2q$	5.60	
$\rightarrow 3\ell \nu$	4.43	
$\rightarrow \ell 3\nu$	3.06	
WWW	0.209	} Tri-boson
WWZ	0.17	
WZZ	0.057	
ZZZ	0.015	

Table 4.15: Background processes and cross sections.

### 4.5.2.1 QCD multi-jet

The use of MC simulation to evaluate the QCD contribution is disfavored due to two main factors, one is the probability for a quark or a gluon jet to be identified as a  $\tau_h$  object is very low and it has to be combined with the equally low probability to have in the event two additional jets that pass the medium working point of the b tagging discriminator. A too large number of events must be simulated to ensure a trustable simulation for the analysis. The other factor is the misidentification rate for  $\tau_h$  objects is mainly lead by detector effects that are very complex to simulate properly and can change over time in account of many external factors impossible to predict in advance.

In the analysis described in this thesis, the so-called ABCD method is adopted in order to model and estimate the QCD multi-jet background from jet enriched regions in data. The phase space of the events is divided in four regions, using the values of two uncorrelated variables, tau pair sign and tau isolation (the sub-leading tau in the  $\tau_h\tau_h$  channel). A schematic representation can be seen in Fig. 4.10.

- Region A: Represents the signal region as defined in Section 4.4.3 and contains a pair of opposite sign electric charge (OS) tau leptons (either  $\tau_{e/\mu}\tau_h$  or  $\tau_h\tau_h$ ) and where all  $\tau_h$  objects pass the medium working point of the tau isolation discriminant.
- Region B: It is defined with the same isolation selections, but the pair charge requirement is inverted (same sign or SS).
- Region C: It is composed by events with an opposite sign tau pair where  $\tau_h$  objects pass the VVLoose, VLoose or Loose working point of the tau isolation discriminant, but are required to fail the medium WP that defines the signal region. In the  $\tau_h\tau_h$  and  $\tau_e\tau_h$  channels this tau isolation selection is applied only to the  $\tau_h$  candidate present in the event, while in the  $\tau_h\tau_h$  final state it is applied only to the lowest  $p_T$   $\tau_h$  candidate selected.
- Region D: this the region with the lowest contribution from signal, has the same tau isolation criteria of region C, but it also requires that the leptons in the tau pair have the same electric charge.

The distribution of the multi-jet QCD contribution are taken from the C region, where the yields from other backgrounds, given by the MC simulations, are removed:

$$N_i = N_i^{\text{data}} - N_i^{\text{MC}}, \quad (4.23)$$

where  $N^{\text{MC}}$  is the number of background events. The number of events left are multiplied by a factor  $\kappa^{\text{iso/non-iso}}$ , which is calculated as the ratio between the event yields in the regions B/D after the subtraction of the other backgrounds.

Therefore, the ABCD method estimates the QCD contribution in the signal region, by taking its shape from the Data-MC in region C, and the yield from the Data-MC yields in regions B,C,D as:

$$N_A = N_C \times \frac{N_B}{N_D}. \quad (4.24)$$

Given the limited statistics in the boosted category (defined in Section 4.4.2.1) the correction factor ( $N_B/N_D$ ) is estimated from events that pass all the selections that define the boosted category, excluding the b-tagging requirement. In the same way, in the five VBF sub-categories, the correction factor is estimated from an inclusive category that includes all the five sub-categories themselves.

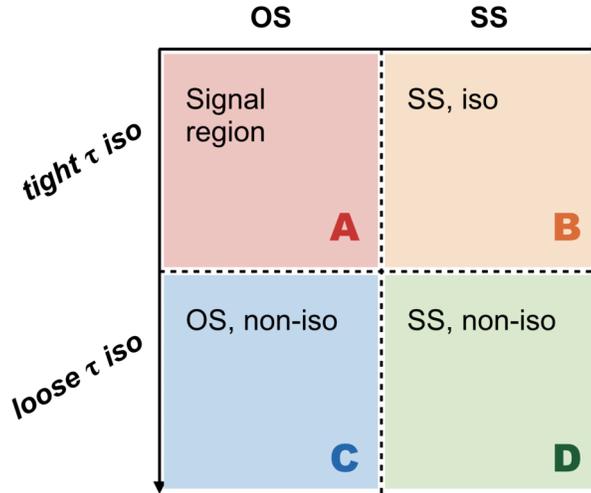


Figure 4.10: Scheme of the four regions used to estimate the multi-jet QCD background.

#### 4.5.2.2 Drell-Yan background ( $Z + \text{jets}$ )

The contribution of  $Z/\gamma^* \rightarrow \ell\ell$  ( $\ell = e, \mu, \tau$ ) plus jets represents the second background contributions in terms of production rate for our analysis. This background is modelled relying on simulation.

Given the narrow phase space studied in the analysis, the DY statistics is increased by combining inclusive sample ( $Z + \text{Jets}$ ) with complementary ones, where the emission of 1, 2, 3 or 4 additional jets, or the emission of 1 or 2 b-jets, is required. Since the full MC simulation process is quite consuming, in terms of computing time and power, in all the events of these samples the invariant mass  $m_{\ell\ell}$  is forced to be larger than 50 GeV without losing any information from low mass events that would be in any case excluded from the analysis by the selections used to define the signal region. If, on one hand, the differential distributions of the DY events show a good agreement with the

observed data, the modelling of the yield, especially when the production is in association with multiple jets, is known to be imperfect and thus requires a correction.

Scale factors for the simulation normalization are taken from 18 Control Regions (CRs) of events compatible with  $Z \rightarrow \mu\mu$  plus jets decays. In the CRs the jet selection criteria and third lepton veto are applied as done for the signal region (described in Section 4.4). There are defined 18 CRs, differently for the data and MC simulations. The data is first divided into three CRs, depending on the number of b-tagged jets (0, 1 or 2), then each of the three CRs is further divided into six subregions, based on the  $p_T$  of the reconstructed Z candidate. The  $p_T$  thresholds are year-dependent since they depend on the data-MC agreement of the  $p_T$  in the different regions. Similarly as for the data, the simulated events are divided into three templates, depending on the number of b-partons at LHE generator level (0, 1 or 2) [136, 137]. Each group is then subdivided into six templates, depending on the  $p_T$  of the generator level Z boson. One additional template, combining all other backgrounds, is also considered. Finally, the scale factors are computed by performing a simultaneous fit of the CRs and the templates, where the normalization of the simulation templates are kept floating. These scale factors are applied in the signal region.

The agreement between data and MC simulation in the  $Z \rightarrow \mu\mu$  control region, before and after applying the scale factors, is shown in Fig. 4.11 for the 2018 data taking period.

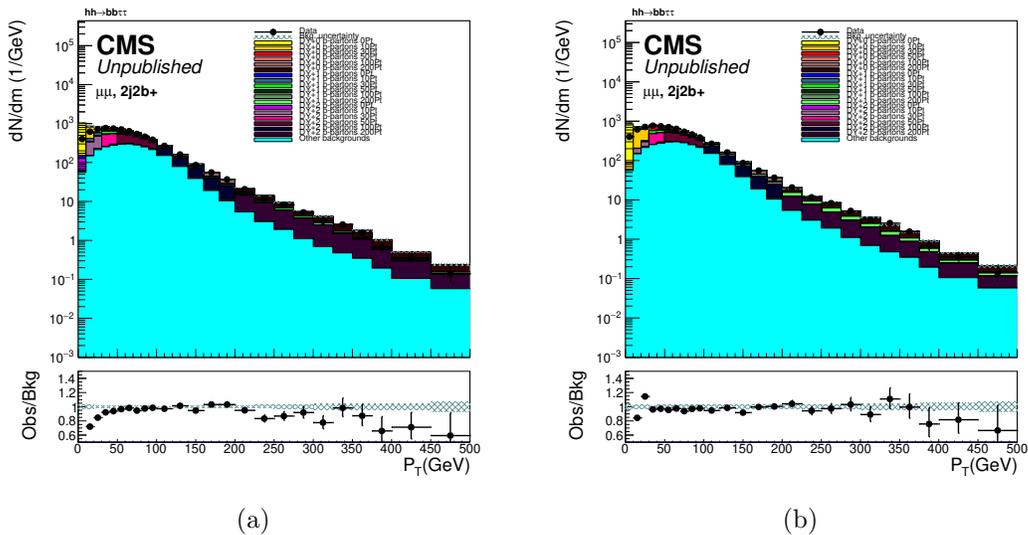


Figure 4.11: Distribution of  $p_T$  for the  $Z \rightarrow \mu\mu + 2$  b-jets control region for 2018. The black dots correspond to  $59.7 \text{ fb}^{-1}$  of data collected by the CMS experiment during 2018 at  $\sqrt{s} = 13 \text{ TeV}$  and colored filled histograms represent MC samples produced with the 2018 data-taking conditions of the CMS experiment. Before applying DY SFs is shown in (a), and after in (b). Dashed blue areas represent the uncertainty on MC.

### 4.5.2.3 $t\bar{t}$ background

The  $t\bar{t}$  background contribution is modelled relying on simulation: while the shape of the process is well described, the normalization of the background shows a disagreement with respect to the observed data. To address the discrepancy a scale factor (ttSF) is fitted from a  $t\bar{t}$  enriched control region and validated in a  $t\bar{t}$  enriched subset of the signal region. Three control regions are defined, one for each year, to contain  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$ , and  $\tau_h\tau_h$  events that satisfy the res2b selections outlined in Section 4.4.2.1 and pass a mass cut obtained inverting the one of the the resolved categories defined in equation 4.2.

The contamination from other backgrounds in the  $t\bar{t}$  enriched control region is less than 9% for all the three years. The results of the ttSF fits are reported in Table 4.16 together with their statistical error.

	2016	2017	2018
ttSF	$0.908 \pm 0.006$	$0.988 \pm 0.006$	$0.966 \pm 0.009$

Table 4.16: ttSF values and their statistical error obtained from the control regions for the three years.

The plots reported in Fig. 4.12 show the improvement obtained by the use of the ttSF for the 2016 data taking period.

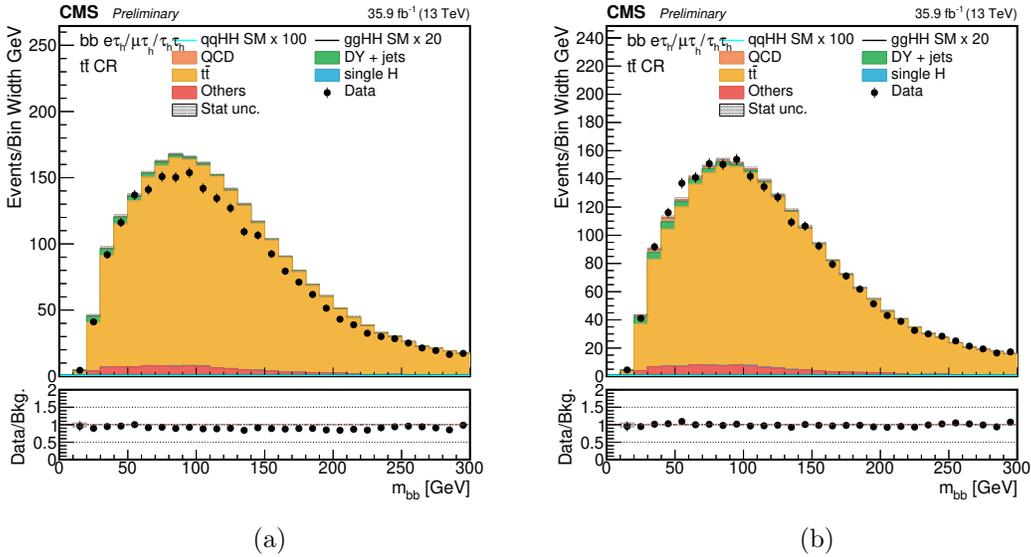


Figure 4.12: Distributions of  $m_{bb}$  in the  $t\bar{t}$  CR for the 2016 data taking period, before (a) and after (b) the application of ttSF are reported. Only the statistical uncertainty is reported.

## 4.6 Systematic uncertainties

Systematic uncertainties arise from uncertainties associated with the nature of the measurement apparatus, assumptions made by the experimenter, or the model used to make inferences based on the observed data. These uncertainties have an effect in the modeling of signal and background processes, described in Section 4.1. The systematic uncertainties can be divided into normalization and shape uncertainties: while the first affect the total event yield of the processes, i.e. the number of expected events of the modeled contribution, of a particular background or signal; the latter affect the distribution of the events. In the following a list of the uncertainties considered in the analysis is presented.

### 4.6.1 Normalization uncertainties

The following normalization uncertainties are considered:

- Different sources of uncertainties on the luminosity measurement are considered and treated as correlated. Their values are obtained from dedicated Van-der-Meer scans and stability of detector response during data taking and range between 0.1 and 2.2% for the three years. These uncertainties are applied to the signals and background processes estimated only from simulation. Since the normalizations of the  $t\bar{t}$ ,  $Z/\gamma^* \rightarrow \ell\ell$ , and multi-jet backgrounds are obtained from data, they are not subject to the luminosity uncertainties.
- Electron and muon reconstruction, isolation, and identification uncertainties are determined from the uncertainties on the Monte Carlo-to-data scale factors. ; a value of 1% for both electrons and muons is obtained. An additional uncertainty of 3% (15%) for  $\tau$  leptons with  $p_T < 100$  GeV ( $>100$  GeV) is added in the  $\tau_h\tau_h$  channel as recommended by the physics objects group. This uncertainty arises from the choice of the working point of the DeepTau algorithm used in the selection of the  $\tau\tau$  pair.
- During the 2016 and 2017 data taking years, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at  $|\eta| > 2.0$  caused a specific trigger inefficiency. For events containing an electron (a jet) with  $p_T$  larger than  $\sim 50$  GeV ( $\sim 100$  GeV), in the region  $2.5 < |\eta| < 3.0$  the efficiency loss is  $\sim 10$ – $20\%$ , depending on  $p_T$ ,  $\eta$ , and time. Correction factors were computed from data and applied to the acceptance evaluated by simulation. An uncertainty of 2% is associated to this effect.
- An uncertainty on the pile-up reweighting technique is estimated varying the applied pile-up weights value by their uncertainty. The systematic uncertainty is estimated

to be negligible in the  $\tau_h\tau_h$  final state, while it has a value of 1% in the  $\tau_\mu\tau_h$  and  $\tau_e\tau_h$  channels.

- The normalization of the  $t\bar{t}$  background is taken from a fit to a CR per year, as described in Section 4.5.2.3. The uncertainty on the scale factors obtained in these CRs is purely statistical and is always below 1%.
- The normalization of the  $Z/\gamma^* \rightarrow \ell\ell$  background is taken from a fit to 18 CRs per year, as described in Section 4.5.2.2. The uncertainties on the scale factors obtained in these CRs are propagated to the SR taking into account their correlation and range from 0.1 to 60% depending on year and CR considered.
- The multi-jet background contribution is determined from data in jet-enriched regions, as described in Section 4.5.2.1. Two normalization and one shape uncertainties are derived in order to take into account the statistical fluctuations of the jet-enriched regions.
- The uncertainties on the normalizations of the backgrounds modelled relying solely on the simulated events range from 2 to 10%.
- The theoretical uncertainty on the cross section of HH production: via gluon-gluon fusion  $+2.2/-5.5\%$  (scale), 3% (PDF $+\alpha_S$ ), 2.6% ( $m_t$ ) [138]; and via vector boson fusion:  $+0.03/-0.04\%$  (scale),  $\pm 2.1\%$  (PDF +  $\alpha_S$ ) [139]. These uncertainties are only considered when upper limits are quoted with respect to the SM and for the likelihood scans, while they are not included for the upper limits on the cross section.
- The theoretical uncertainty on the H branching ratios [140]: to  $b\bar{b}$  0.65% (theory),  $+0.72/-0.74\%$  ( $m_q$ ),  $+0.78/-0.80\%$  ( $\alpha_S$ ); and to  $\tau$  leptons  $+1.16/-1.17\%$  (theory),  $+0.98/-0.99\%$  ( $m_q$ ), 0.62% ( $\alpha_S$ ).

## 4.6.2 Shape uncertainties

In order to evaluate the effect of a systematic uncertainty on the shape of the discriminating variable it is necessary to recompute the shape after shifting each object affected by that uncertainty (like for instance, when shifting hadronic taus also the MET in the event is shifted and the SVfit algorithm has to be re-run to compute the  $H \rightarrow \tau\tau$  mass).

A shape uncertainty is assigned to each bin of the distribution to account for the finite MonteCarlo samples statistics following the Barlow-Beeston approach described in [141].

The following shape uncertainties are considered:

- The uncertainty on the measurement of the energy of  $\tau_h$  leptons. The uncertainties are derived combining low and high- $p_T$  measurements in  $Z \rightarrow \tau\tau$  and in  $W^* \rightarrow \tau\nu$  events. Four different uncertainties are included to take into account the different  $\tau_h$  decay modes considered in this analysis. When considering the uncertainty for a particular decay mode, the shift is applied only to the  $\tau_h$  candidates that are reconstructed with that particular decay mode, all other  $\tau_h$  candidates are left unchanged.
  - The jet energy scale (JES) and jet energy resolution (JER) uncertainties. For JES, 11 separate sources of uncertainties are included per year, those that appear in multiple years are treated as fully correlated, while those appearing only in one year are treated as uncorrelated. For JER, alternative templates are produced by shifting all jet related features. These shifts come from the use of scale factors to smear MC simulation jets to match the experimental energy resolution.
  - Separate uncertainties on the energy scale of electrons misidentified as  $\tau_h$  are provided to take into account two different decay modes ( $h^\pm$  and  $h^\pm\pi^0$ ). The recommended uncertainty on the energy scale of muons misidentified as  $\tau_h$  is 1%, uncorrelated in decay mode.
  - The uncertainties arising from the application of the DeepTau identification scale factors. Five uncertainties are present for the identification against jets and are calculated in tau  $p_T$  bins; all uncertainties are used in the  $\tau_\mu\tau_h$  and  $\tau_e\tau_h$  channels. Since in the  $\tau_h\tau_h$  channel both leptons are required to have  $p_T$  above a threshold, only the highest  $p_T$  bin uncertainty is applied. Two uncertainties are present in the identification against electrons, one for the barrel and one for the endcap, and are treated as uncorrelated across tau  $\eta$  bins.
  - The uncertainties on the trigger scale factors. Four different uncertainties are included to take into account the different  $\tau_h$  decay modes considered in this analysis. They are applied to the  $\tau_h$  leg of each category. Two more trigger uncertainties are used to cover the cases where the  $\tau$  decay in an electron or in a muon. Finally, one uncertainty is added for the  $\tau_h\tau_h$  final state in 2017 and 2018 to take into account the jet legs scale factors of the VBF trigger.
  - The uncertainty on the b-tagging efficiency, as function of  $p_T$  and  $\eta$ .
  - The uncertainty on the pile-up jet identification scale factors as function of  $p_T$  and  $\eta$ .
  - For the  $\tau_h\tau_h$  channel in 2017, ad hoc scale factors are applied to events with real taus. Four systematic uncertainties are defined to take into account the errors provided
-

by the fit used to derive these scale factors; since the fits are done independently for each decay mode of the tau, the four uncertainties are treated as uncorrelated.

## 4.7 Statistical interpretation

In order to evaluate the presence of HH signal in the observed data, or to set an upper limit on its production cross section if no excess over the background predictions is observed, a statistical procedure is needed. The Higgs boson searches performed by the ATLAS and CMS collaborations make use of a modified frequentist approach, referred to as  $CL_s$ , defined in Ref. [142]. In this approach, a null and an alternative hypotheses are constructed, and test against each other. The null hypothesis ( $H_b$ ) describes the background-only scenario, while the alternative hypothesis ( $H_{\mu s+b}$ ) describes a model with signal plus background.

A binned maximum likelihood fit of the DNN prediction (Section 4.4.4.2) is performed simultaneously in the signal regions of the eight categories per channel (the latter are defined in Section 4.4.4). The expected event yield of the signal and of the total background are denoted as  $s$  and  $b$  respectively, and the signal normalization is scaled by a signal strength modifier  $\mu$ . The model adopted relies on a signal strength modifier parameter that modifies the SM HH cross section by the scale  $\mu$  and leaves the HH  $\rightarrow$  bb $\tau\tau$  decay branching fraction unchanged with  $\mu = \sigma_{obs}/\sigma_{SM}$ . In the case there is not signal contribution, corresponding to the background-only hypothesis, the signal strength modifier  $\mu$  is null.

The systematic uncertainties described in Section 4.6, affect the background and signal event yields, so they are introduced as nuisance parameters  $\theta_i$  in the maximum likelihood fit, collectively denoted as  $\theta$ . Therefore the event yield for the signal and background processes can be denoted as  $s(\theta)$  and  $b(\theta)$ , respectively.

In this work, binned distributions are used for the signal extraction, thus the signal and background are defined as vectors of components  $s_k$  and  $b_k$ , corresponding to the content of each bin  $k$ . For binned distributions, the number of events predicted in the bin  $k$  by the modeling of the reference distribution can be written as the sum of signal  $\mu s$  and background  $b$  expected events:

$$n_k(\mu, \theta_i) = \mu \cdot s_k(\theta_i) + b(\theta_i). \quad (4.25)$$

The Poisson distribution representing the probability of observing  $n_k^{obs}$  data event in the same bin of the same event category, for  $n_k > 0$ , can be written as:

$$P_k(n_k^{obs}, n_k) = \frac{n_k^{n_k^{obs}}}{n_k^{obs}!} e^{-n_k}. \quad (4.26)$$

The likelihood function, of the hypothesis that the data observed is compatible with the signal plus background, is the product of the Poisson probabilities to observe  $n_k^{obs}$  events in the bins  $k$

$$\mathcal{L}(n|\mu, \theta) = \prod_k P_k(n_k^{obs}, n_k) \cdot p(\tilde{\theta}|\theta). \quad (4.27)$$

The term  $p(\tilde{\theta}|\theta)$  is the nuisance probability density function, which represents the knowledge about the values of the nuisance parameters; each term reflects the probability for the true value to be equal to  $\theta_i$ , given the best estimate  $\tilde{\theta}_i$  obtained from auxiliary measurements on control region events or directly from the MC simulation. The systematic uncertainties sources are assumed to be uncorrelated, thus the combined term is the product of the single uncertainties ( $p(\tilde{\theta}|\theta) = \prod_i p_i(\tilde{\theta}_i|\theta_i)$ ) Since the signal and background are functions of the nuisances  $\theta$ , the likelihood, from Eq. 4.27, is scaled by the nuisance probability density function.

The functional form  $p_i(\tilde{\theta}_i|\theta_i)$  depends on the type of uncertainty described. Uncertainties which arise from independent measurements, such as luminosity or trigger efficiencies, are modeled with log-normal functions:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi \ln(k)}} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln(k))^2}\right) \frac{1}{\theta} \quad (4.28)$$

where  $k$  is the parameter that defines the width of the log-normal distribution and thus represents the interval of possible variations of the observable.

Uncertainties on the template shapes are taken into account during the fit procedure, using the *Vertical Template Morphing* technique: for each quantity that affects the shape, multiple instances of the templates are produced from the simulated events by varying that quantity by  $\pm 1\sigma$  and bin-by-bin interpolation is performed between them. A nuisance that represents the variation of such quantities from the nominal value, is added to the likelihood model [143].

To set an exclusion limit on the presence of a signal, it is necessary to find the value of  $\mu$  that allows to reject the  $H_{\mu s+b}$  in favor of  $H_b$ . The statistic test used to set the exclusion limit is the likelihood ratio; corresponding to the signal plus background and to the background only hypotheses:

$$q_\mu = -2 \ln \frac{\mathcal{L}(n|\mu, \hat{\theta}_\mu)}{\mathcal{L}(n|\hat{\mu}, \hat{\theta})} \quad \text{with } 0 \leq \hat{\mu} \leq \mu, \quad (4.29)$$

where  $\hat{\theta}_\mu$  is the value of  $\theta$  that maximizes the likelihood of Eq. 4.27 for a fixed value of  $\mu$ , referred as the conditional maximum-likelihood estimator; and the denominator represents the maximized likelihood. The parameters estimators  $\hat{\theta}$  and  $\hat{\mu}$  represent the

global maximum of the likelihood; they are the values obtained when maximisation is performed on both parameter simultaneously. Higher values of  $q_\mu$  indicates an increasing incompatibility with the signal plus background hypothesis. The constraint in the upper value of the  $\hat{\mu}$  is set as a protection against the effect of downward background statistical fluctuations, so that they are not interpreted as evidence against the hypothesis of a signal with small strength; instead the lower value in the  $\hat{\mu}$  constraint is dictated by physics, since it is expected only a positive signal rate.

Given a signal strength modifier  $\mu$ , the observed value of the test statistics  $q_\mu^{obs}$  is calculated performing an statistic test using the observed data  $n$ . To measure the compatibility between the  $H_{\mu s+b}$  and  $H_b$ , the probabilities of observing  $q_\mu \geq q_\mu^{obs}$  are calculated for both hypothesis:

$$\begin{aligned} CL_{s+b}(\mu) &= P(q_\mu \geq q_\mu^{obs} | H_{\mu s+b}), \\ CL_b(\mu) &= P(q_\mu \geq q_\mu^{obs} | H_b). \end{aligned} \quad (4.30)$$

The  $CL_s$ , for a given value of  $\mu$ , is calculated as the ratio

$$CL_s = \frac{CL_{s+b}(\mu)}{CL_b(\mu)}. \quad (4.31)$$

A signal of strength  $\mu$  is said to be excluded at a CL  $(1 - \alpha)$ , if  $CL_s \leq \alpha$ . By convention as been chosen the 95% confidence level upper limit on  $\mu$ ; which is achieved by varying the value of  $\mu$  until the condition  $CL_s \leq 0.05$  is satisfied. The value of  $\mu$  obtained is converted into a limit in the signal cross section  $\sigma_{\text{HH}} \times \text{BR}(\text{HH} \rightarrow \text{bb}\tau\tau)$  by rescaling the signal normalization initially fixed.

## 4.8 Results

### 4.8.1 Inclusive GGF+VBF HH production

#### 95% CL upper limits on production cross section vs $\kappa_\lambda$

The expected sensitivity for the inclusive (GGF + VBF) HH production cross section are reported as function of  $\kappa_\lambda$  at 95% Confidence Level. The eight categories in three channels, described in Section 4.4.2.1, are fitted simultaneously in the three data taking years to obtain the final sensitivity. The expected constraint of  $\kappa_\lambda$  is  $-2.8 < \kappa_\lambda < 9.5$ , as shown in Fig. 4.15 for the full Run II combination. The expected limits for the SM point ( $\kappa_\lambda = 1$ ), including the statistical and systematic uncertainties, are listed in Tab. 4.17 and summarized in Fig. 4.13.

Expected limit	2016	2017	2018	Run 2
$\sigma_{GGF+VBF}(\text{pp} \rightarrow HH \rightarrow bb\tau\tau) / \sigma_{GGF+VBF}^{SM}$	9.03	10.38	7.16	4.55
$\sigma_{GGF+VBF}(\text{pp} \rightarrow HH \rightarrow bb\tau\tau)$ [fb]	21.61	24.82	17.12	10.88
$\sigma_{GGF+VBF}(\text{pp} \rightarrow HH)$ [fb]	296.01	340.05	234.55	149.03

Table 4.17: 95% CL expected limits for the SM point ( $\kappa_\lambda=1$ ), where  $\sigma_{GGF+VBF}^{SM} = (31.05 + 1.726) \cdot \mathcal{B}(HH \rightarrow bb\tau\tau) = 32.776 \text{ fb} \cdot 0.073 = 2.39 \text{ fb}$ .

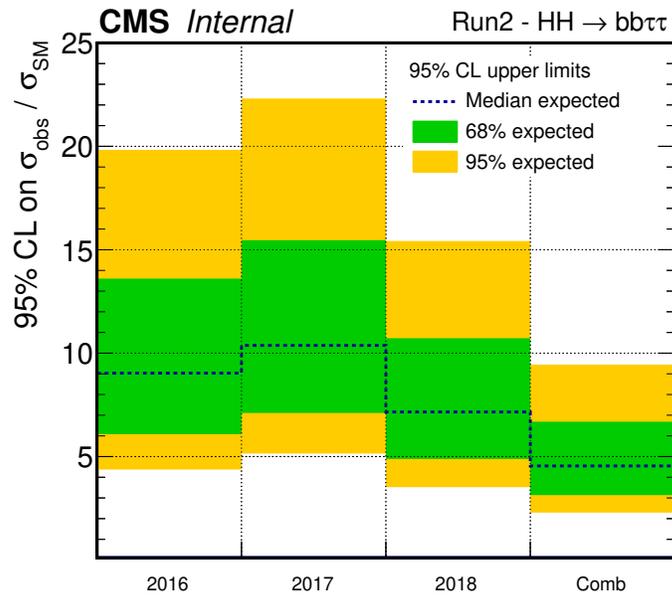


Figure 4.13: Visualization of the 95% CL expected limits, including the systematic uncertainties, for the SM point  $\kappa_\lambda=1$  [110]. as reported in Table 4.17.

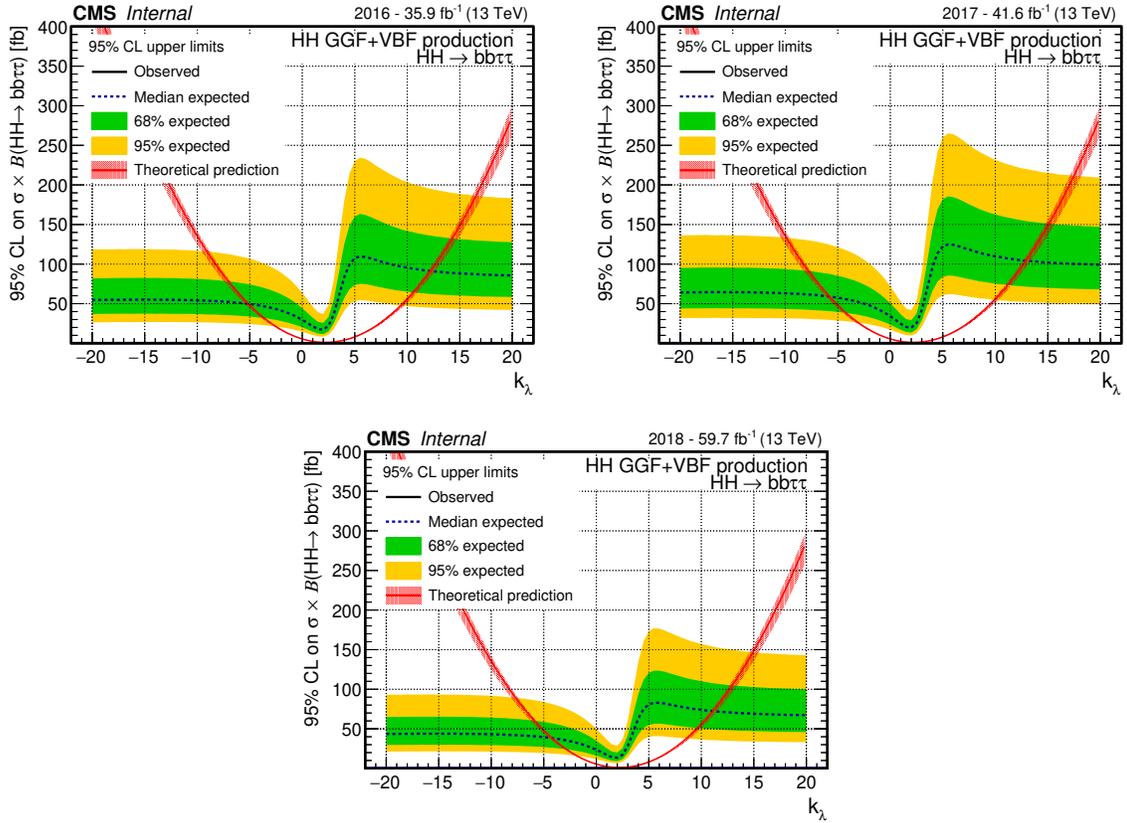


Figure 4.14: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the  $\kappa_\lambda$  coupling. The plots correspond to the combination of the channels and categories within each year: 2016 (top left), 2017 (top right) and 2018 (bottom).

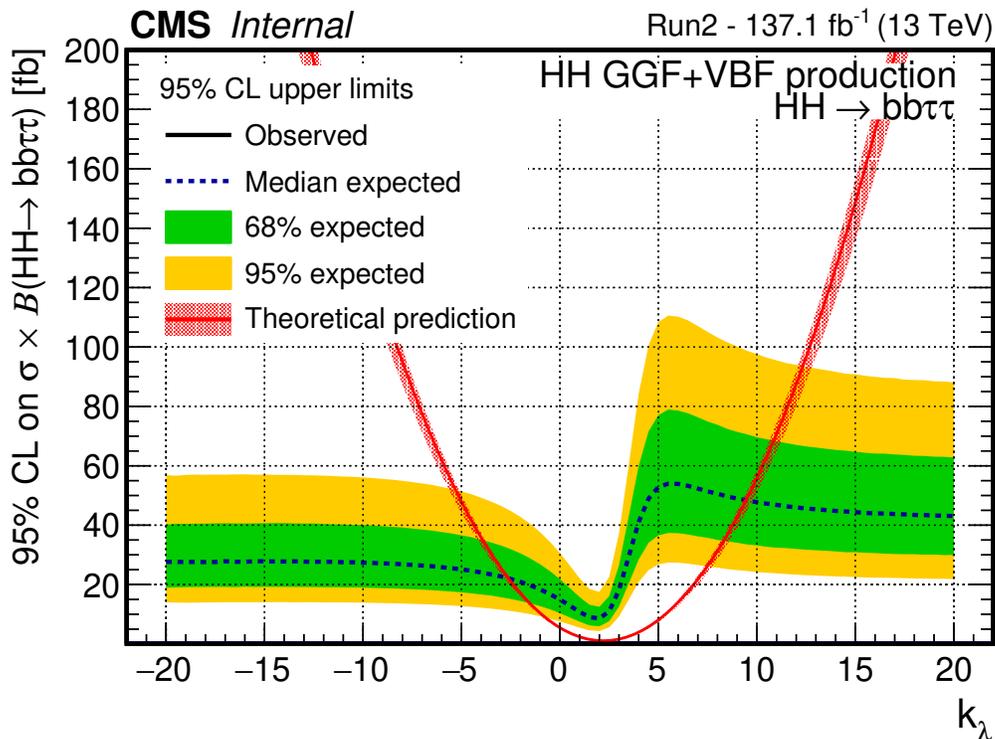


Figure 4.15: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the  $\kappa_\lambda$  coupling. The plot corresponds to the combination of all the channels and categories for the full Run II [110].

### 95 %CL upper limits for different categories and channels

The expected sensitivity for the inclusive (GGF + VBF) HH production cross section are reported here as function of  $\kappa_\lambda$  at 95% CL highlighting the contribution of different channels and categories. Results are reported for 2016, 2017 and 2018 separately in Figures 4.16, 4.17 and 4.18, respectively.

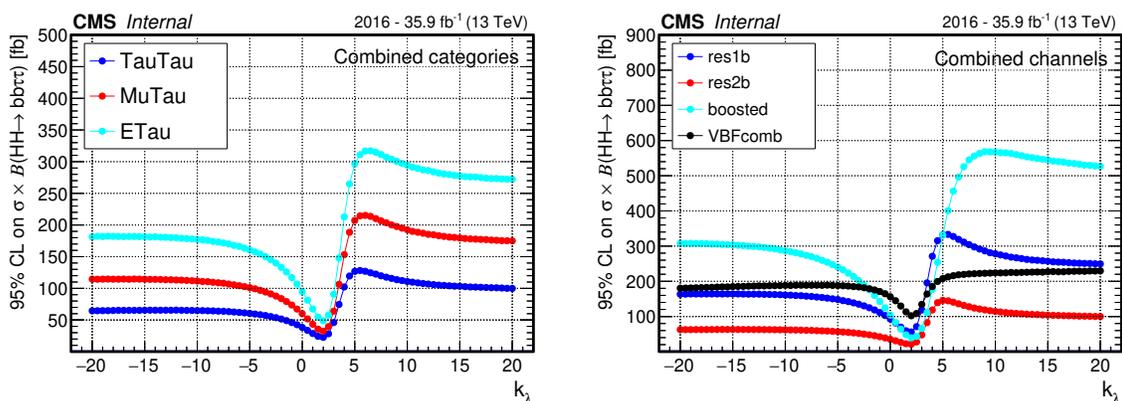


Figure 4.16: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the  $\kappa_\lambda$  coupling. The plot on the left shows the result obtained in 2016 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2016 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multi-categories in the VBF phase-space (in the plot denoted as *VBFcomb*).

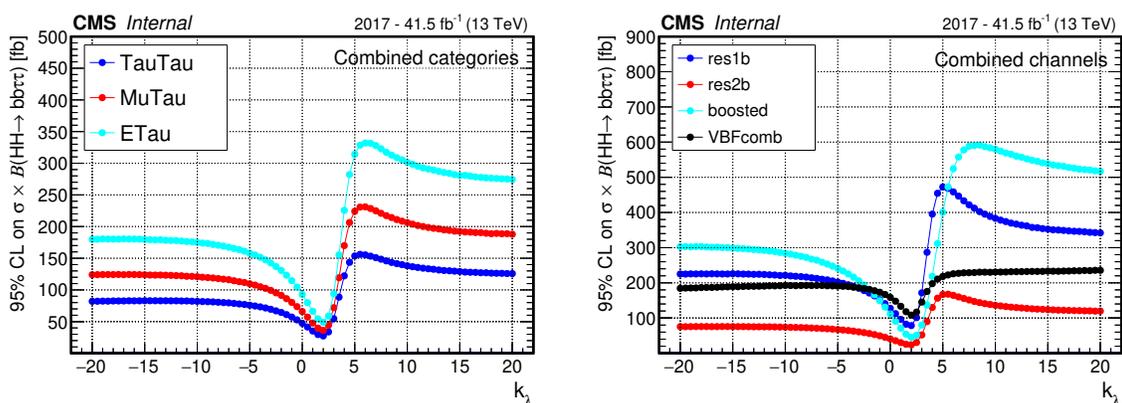


Figure 4.17: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the  $\kappa_\lambda$  coupling. The plot on the left shows the result obtained in 2017 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2017 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

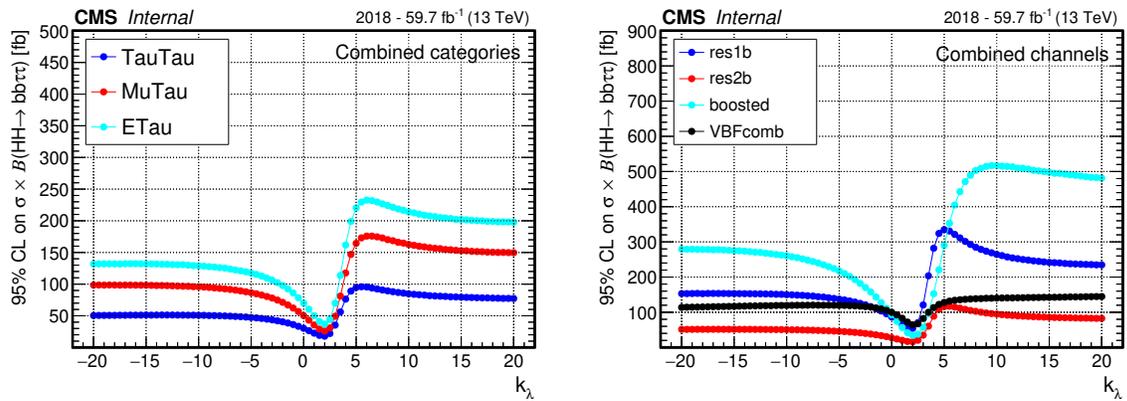


Figure 4.18: 95% CL expected exclusion of the GGF + VBF  $HH$  production cross section as a function of the  $\kappa_\lambda$  coupling. The plot on the left shows the result obtained in 2018 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2018 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as  $VBFcomb$ ).

## 4.8.2 VBF HH production

### 95 %CL upper limits on VBF production cross section vs $c_{2V}$

The expected sensitivity for the inclusive VBF HH production cross section are reported here as function of  $c_{2V}$  at 95% Confidence Level. Also for VBF studies eight categories in three channels are fitted simultaneously in the three years to obtain the final sensitivity. The expected constraint of  $c_{2V}$  is  $-0.6 < c_{2V} < 2.7$ , as shown in Fig. 4.20 for the three years, and in Fig. 4.21 for the full Run 2 combination. The expected limits for the SM point ( $c_{2V} = 1$ ), with the statistical and systematic uncertainty, are listed in Tab. 4.18. The results including systematic uncertainties are also summarized in Fig. 4.19.

Expected limit	2016	2017	2018	Run 2
$\sigma_{VBF}(\text{pp} \rightarrow \text{HH} \rightarrow \text{bb}\tau\tau)/\sigma_{VBF}^{SM}$	331.0	366.5	196.8	137.5
$\sigma_{VBF}(\text{pp} \rightarrow \text{HH} \rightarrow \text{bb}\tau\tau)$ [fb]	571.3	632.6	339.6	237.3
$\sigma_{VBF}(\text{pp} \rightarrow \text{HH})$ [fb]	41.7	46.2	24.8	17.3

Table 4.18: 95% CL expected limits for the SM point ( $c_{2V}=1$ ), where  $\sigma_{VBF}^{SM} = 1.726 \cdot \mathcal{B}(\text{HH} \rightarrow \text{bb}\tau\tau) = 1.726 \text{ fb} \cdot 0.073 = 0.126 \text{ fb}$ .

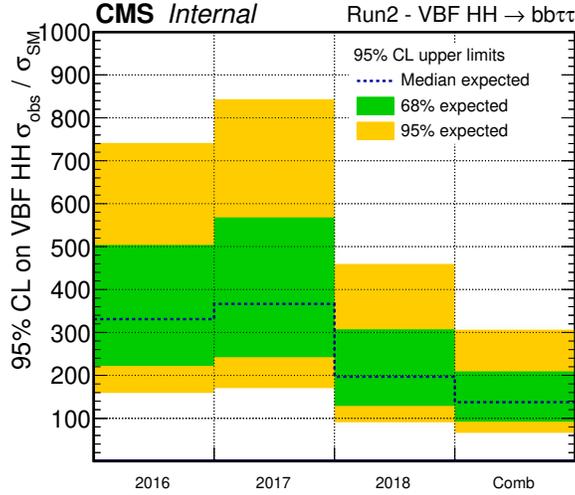


Figure 4.19: Visualization of the 95% CL expected limits, including the systematic uncertainties, for the SM point ( $c_{2V}=1$ ), as reported in Table 4.18.

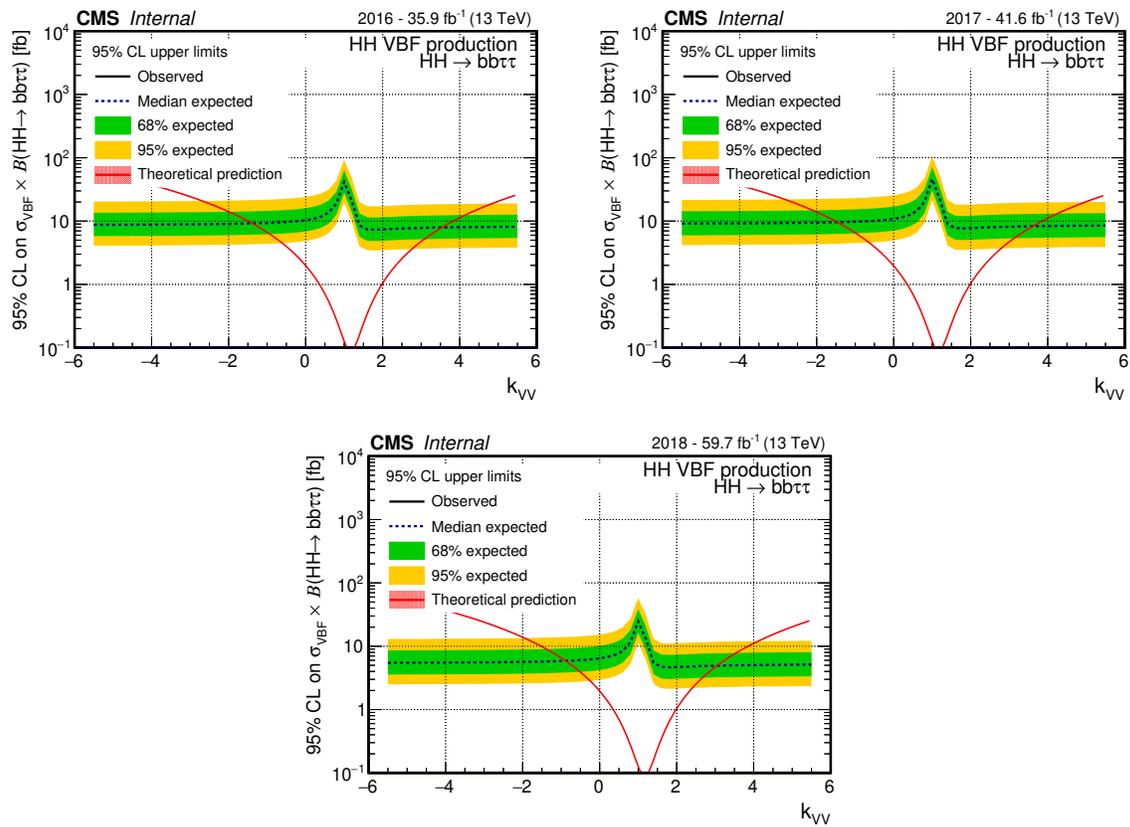


Figure 4.20: 95% CL expected exclusion of the BF HH production cross section as a function of the  $c_{2V}$  coupling (denoted as  $k_{\text{VV}}$  in the figure). The plots correspond to the combination of the channels and categories within each year: 2016 (top left), 2017 (top right) and 2018 (bottom).

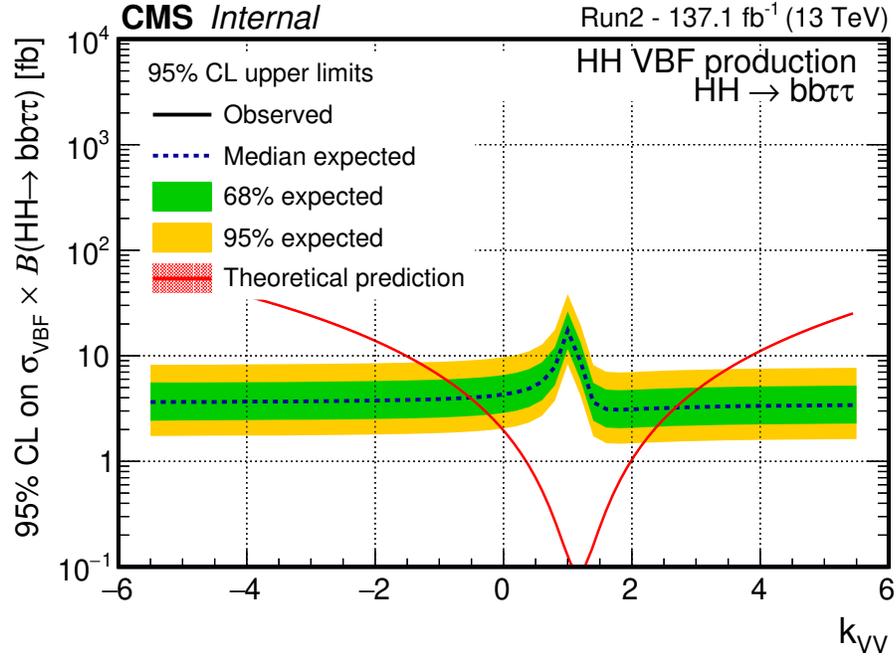


Figure 4.21: 95% CL expected exclusion of the VBF HH production cross section as a function of the  $c_{2V}$  coupling (denoted as  $k_{VV}$  in the figure). The plot corresponds to the combination of all the channels and categories for the full Run II.

#### 4.8.2.1 95 %CL upper limits on VBF production cross section for different categories and channels

the inclusive VBF HH production cross section are reported here as function of the  $c_{2V}$  coupling at 95% Confidence Level highlighting the contribution of different channels and categories. Results are reported for 2016, 2017 and 2018 separately in Figures 4.22, 4.23 and 4.24, respectively.

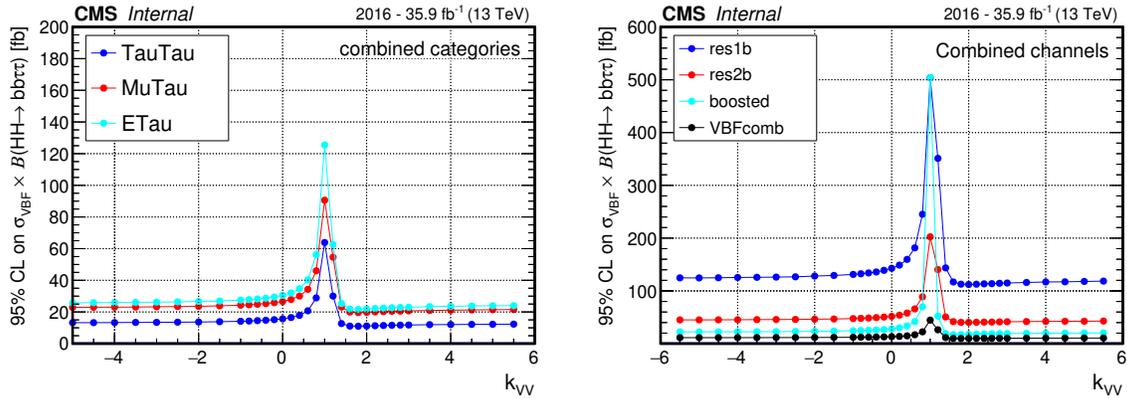


Figure 4.22: 95% CL expected exclusion of the VBF  $HH$  production cross section as a function of the  $c_{2V}$  coupling (denoted as  $k_{\text{VV}}$  in the figure). The plot on the left shows the result obtained in 2016 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2016 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as  $VBFcomb$ ).

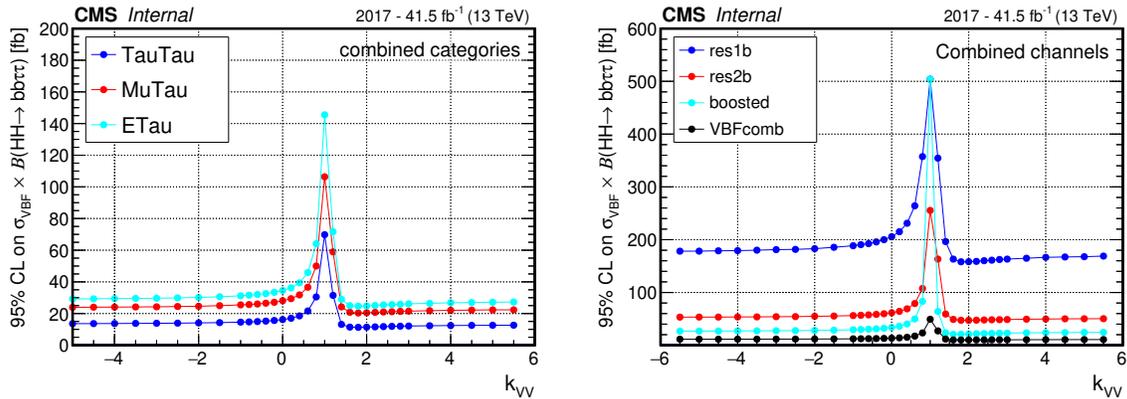


Figure 4.23: 95% CL expected exclusion of the VBF  $HH$  production cross section as a function of the  $c_{2V}$  coupling (denoted as  $k_{\text{VV}}$  in the figure). The plot on the left shows the result obtained in 2017 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2017 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as  $VBFcomb$ ).

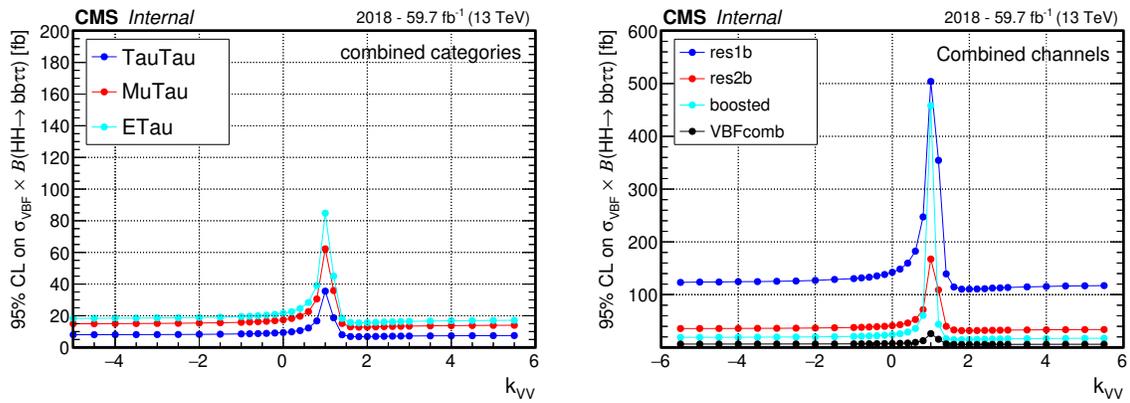


Figure 4.24: 95% CL expected exclusion of the VBF HH production cross section as a function of the  $c_{2V}$  coupling (denoted as  $k_{\text{VV}}$  in the figure). The plot on the left shows the result obtained in 2018 combining all the categories in different  $\tau\tau$  decay channels. The plot on the right shows the result obtained in 2018 combining all three  $\tau\tau$  decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

# Chapter 5

## Introduction to neural networks

In the recent years machine learning techniques have proven to be very useful in the high-energy physics field. Specifically in CMS, as mentioned in Chapter 3, one of its most relevant uses is for the identification of physical objects, where was detailed the use of these techniques for the tau (Section. 3.2.4) and jet (Section 3.2.5.1) identification. In this chapter, concepts related to the neural networks are presented. These concepts are applied in the HH-Btag algorithm, introduced in Chapter 6. In the following, a general introduction to neural networks is presented.

### 5.1 Neural networks

Artificial neural networks, simply called neural networks (NNs), are a technique for building computer algorithms that learns from data, with the purpose of being able to make classifications, predictions, and in general any type of “decision” or “choice”. It is formed by a circuit of artificial *neurons* connected to each other, where information is communicated and shared among the neurons. This idea is somehow based on how the human brain works. Very roughly, after the neurons are connected, the network is asked to solve a problem iteratively, and during each of the iterations the connections that lead to success are strengthened and diminished those that lead to failure, where the terms success and failure are referred to the accuracy of solving the problem in question. The real power of NNs can be seen when several thousand, or even million, of neurons interact with each other to solve a specific problem.

Artificial neurons, or also called *units*, are mathematical functions, which represent the elementary units of the NNs. An artificial neuron, in general way, takes an input vector  $\vec{x}$ , performs an affine transformation and produces an output, a simple representation of this is shown in Fig. 5.1a. The neural network architecture is build gathering multiple neurons in the same layer and stacking a certain number of layers. The architecture of a

---

NN, as shown in Fig. 5.1b, can be divided into: input, hidden and output layers. The input layer is the initial data fed to the NN ( $\vec{x}$ ), the hidden layers are the layers between the input and output layers, in this part will be performed all the computation, and finally the output layer will produce the result for a given input, which will be optimized in order to match with the correct value the network is trying to predict.

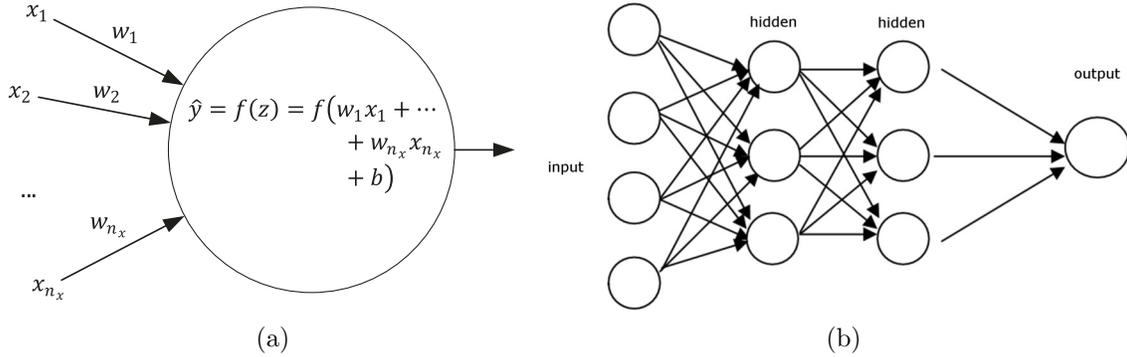


Figure 5.1: (a) Artificial neuron schematic representation. (b) Schematic representation of an neural network 2 hidden layers.

### 5.1.1 Weights

Each neuron has its own set of weights ( $w_i$ ), and each of the inputs are weighted separately ( $w_i \in R$ , with  $i = 1, 2, \dots, n_x$ , where  $n_x$  is the total number of inputs). The weights express the importance of the respective inputs to the determination of the output. After a weight is assigned to each of the inputs, a linear combination of the weights and inputs is passed to a non-linear function, known as *activation function*. The activation function has a threshold effect, so depending if the output of the function is greater or not than the threshold value, the output of the neuron will be modified accordingly. This is a crucial part, since it determines the output, its accuracy, and the computational efficiency of the training model [144]. These steps can be summarized as:

1. Combine linearly all inputs  $x_i$

$$z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b = \sum_{j=0}^{n_x} w^j x^j + b, \quad (5.1)$$

where  $b \in R$  and is denoted as the bias, which is a tunable parameter of the neuron.

2. Apply the neuron's activation function denoted as  $f$ , producing the output

$$\hat{y} = f(z) = f(w_1x_1 + w_2x_2 + \dots + w_nx_n + b). \quad (5.2)$$

One of the most commonly used activation function is the sigmoid function, since it gives as output values in the interval  $[0, 1]$ , and this is useful for models that aim to predict the probability as a result, thus only can assume values between this interval. The sigmoid function can be written as:

$$f(z) = \sigma(z) = \frac{1}{1 + e^{-z}}. \quad (5.3)$$

Others popular activation functions are shown in Fig. 5.3

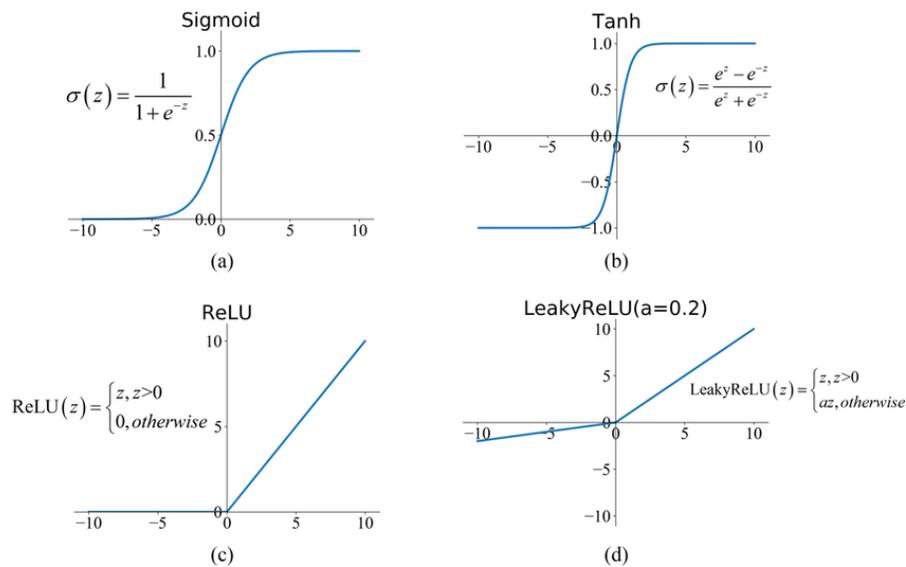


Figure 5.2: Commonly used activation functions: (a) Sigmoid (b) Tanh (c) ReLU and (d) LeakyReLU.

## 5.1.2 Loss function and optimizers

### 5.1.2.1 Loss function

The way the NNs “learn” is by finding the correct weights for each neuron, and biases of the network that minimize a chosen function, usually called the *loss function* or *cost function*, which measures the performance of a NN model. It measures the difference between two probability distributions (in our case the predictions made by the model and the ground truth), so the loss increases as the predicted probability diverges from the actual label. Different types of loss functions are available depending on the application of the model. In our case we are developing a neural network model for probabilistic classification, to select the b-jet pair coming from the  $H \rightarrow b\bar{b}$  decay, hence a loss function optimized for classification purposes is needed. One of the most commonly used

classification loss function is the binary cross-entropy. It is used specifically for binary classifications tasks, where the model is needs to evaluate a problem and answer it with only two choices.

The binary cross-entropy loss function is calculated as the average result of the categorical cross-entropy function, across all data samples:

$$L(\hat{y}_i, y_i) = -\frac{1}{N} \left[ \sum_{i=1}^N y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \right], \quad (5.4)$$

where  $N$  is equal to the output size,  $y_i$  represents the truth value, and  $\hat{y}_i$  is the  $i$ -th scalar value in the model's output. In our case  $N$  is equal to the sum of all the b-jet's candidates in all the events considered,  $y_i$  the ground truth for the  $b_i$ -jet candidate, defined in Section 6.4, which is always 0 or 1, and  $\hat{y}_i$  is the score value given to the  $b_i$ -jet and takes any value between 0 and 1 [145, 146].

### 5.1.2.2 Optimizers

In calculus, there are several ways to find the minimum of a given function analytically, however for the specific problem of the NNs where is needed to find the minimum of the weights, it is not possible due to the high number of weights in question. Instead it is necessary to exploit an optimization algorithm that uses numerical methods to solve this problem, and usually this function is refereed as the *optimization algorithm*. One of the most commonly optimization algorithm used in ML is the gradient descent (GD) optimization algorithm [147]. The GD is used to minimize the loss function by iteratively updating the values of the weights in the opposite direction of the gradient of the function with respect to the weights. Given a loss function which depends on the weights  $J(w)$ , where in this case  $w$  is the vector of weights, the value of  $w$  that minimize the loss function  $J$  can be found by the GD method following these steps:

1. Iteration 0: Choose a random initial guess for  $w_0$ .
2. Iteration n+1: The weights at  $w_{n+1}$  will be updated from the values obtained by the previous iteration n as:

$$w_{n+1} = w_n - \gamma \nabla J(w_n), \quad \text{where} \quad \nabla J(w_n) = \begin{pmatrix} \frac{\partial J}{\partial w_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial J}{\partial w_n} \end{pmatrix}. \quad (5.5)$$

The symbol  $\gamma$  represents an adequately chosen gain, usually defined as learning rate. The partial derivatives of  $J(w)$  with respect to  $w$  are calculated for each of

the weights. The GD steps enumerated above are performed iteratively until the function  $J(w)$  finds a local minimum.

The *learning rate* is a parameter that determines the step size at each iteration while moving towards the local minimum. With high learning rate it is possible to cover more ground each step, but there is a possibility of overshoot the lowest point, since the values of the weights are updated continuously. Instead, if a low learning rate is used, the calculation is performed more frequently, so the risk of missing a local minimum is lower, but computing this calculations is time-consuming, so it is necessary to choose this value wisely, so that the minimum of the function is not missed, but it does not take an infinite time to find it.

The *batch size* is an important hyperparameter that influences the dynamics of the learning algorithm, since defines in how many subgroups the training dataset will be divided. During the training, one batch is taken at a time and the predictions of the model are calculated; later these predictions are compared with the truth value and the loss function is computed. Then, as explained above, the GD is used in order to find the correct weights which minimize the loss function, this procedure is repeated until the entire training dataset has been evaluated, and this is defined as one *epoch*. The number of epochs will define the number times the learning algorithm will work through the entire training dataset. The bigger the batch size, the quicker the NN model will complete each epoch during the training, however the quality of the model may degrade if the batch size is larger, and also it is necessary to consider the computational resources available, since it may not be possible to compute in parallel all the events in the subgroup of the training dataset [148].

There are additional strategies for optimizing GD that can be used alongside any of the GD types, one of them is the *batch normalization*. The concept is similar to the normalization performed to the inputs, explained at the end of Section 6.5, since it normalizes the values of the weights for each batch, and as a consequence there is an improvement in the training speed which allows each layer of the network to learn by itself a little bit more independently of other layers [149].

There are three variants of GD which differ by the amount of data it is used to compute the gradient of the loss function, that will determine the accuracy of the updated weights and the time needed to perform an update:

- **Batch Gradient Descent:** described previously in Eq. 5.5, the weights and biases variations are calculated for each observation, but it performs the learning (i.e. updates the values of the weights and bias, only at the end of the epoch, after all the training dataset has been evaluated). This is not the most efficient algorithm to find the local minimum, since for big datasets it is computationally intensive, and time consuming.
-

- **Stochastic Gradient Descent (SGD):** calculates the gradient of the loss functions and updates the weights and biases accordingly, for each observation in the dataset. One of the positive aspects of this method is the high frequency of the updates, which allows to check how the model learning is performing, and unlike the previous method, it is not necessary to wait until all the dataset has been evaluated. However, on large datasets SGD is computationally intensive and time consuming due to the continuous updates.
- **Mini-Batch Gradient Descent:** the weights and biases are updated after each batch has been fed to the model. It is a middle point between Batch Gradient Descent and SGD. It is more computationally efficient since it performs less calculations, but still the model is updated several times during each epoch. In this method the batch size plays an important role, and must be tuned accordingly.

None of these optimization algorithms is the most efficient way to proceed, since there are many GD based algorithms available that can make the learning procedure occur faster and thus more efficient. The final optimizer of the loss function used in this work is the Adam algorithm [150]. The Adam (Adaptive Moment estimation) algorithm is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. The method computes individual adaptive learning rates for each parameters from estimates of first and second moments (mean  $\mu$  and variance  $\sigma$ ) of the gradients.

## 5.2 Recurrent neutral networks

There are several kinds of artificial neutral networks, among them one of the most often used nowadays are the Recurrent Neutral Networks (RNNs). The RNNs are a class of NNs that are powerful for modeling sequence data such as time series or natural language. A RNN works in a similar way to another type of NN called feedforward, one of the simplest forms of NNs, where the data enters at the input layer and passes through the network, layer by layer, until it arrives at the output layer, i.e. the data travels only in one direction, so there is no backpropagation.

The RNN uses the idea of the feedforward NN, but saves the output of a layer and feeds this information back to the input, in order to build a memory of the previous terms in the sequence, to predict the next term. It allows to take into account what happened previously and share the parameters and weights between the neurons. In a way RNNs can be seen as a feedforward NNs with feedback loops or backpropagation through time, which adds an additional time variable. The use of RNN allows the network to use not only the current inputs but also the inputs that it encountered earlier, and as consequence

---

in cases where the prediction was wrong, it makes small changes so that it will gradually work towards making the right prediction during the back propagation [151, 144].

Typically, in a RNN an output sequence  $(y_1, y_2, \dots, y_t)$  is computed based on its input sequence  $(x_1, x_2, \dots, x_t)$  and its previous hidden state  $(h_1, h_2, \dots, h_t)$  in this way:

$$h_t = \sigma(w_i \cdot x_t + w_h \cdot h_{t-1} + b_h), \quad (5.6)$$

$$y_t = \theta(w_0 \cdot h_t + b_h); \quad (5.7)$$

where  $\sigma$  and  $\theta$  are output and hidden activation functions, and  $b$  is the bias vector of the RNN. There are two sets of weights:  $w_i$  used for the internal memory states, and one  $w_0$  for the sequence element, at each time  $t$  the  $w_0$  does not change, instead changes the hidden state  $h_t$  which is fed to the output  $y_t$ , every time an evaluation is performed. Following Eqs. 5.6 and 5.7, the first neuron of the network does its evaluation of the input  $x_t$  at a time  $t$ , produces an output  $y_t$  and creates an internal memory state  $h_t$ , then the second neuron evaluated at a time  $t + 1$ , gets as input the next element in the sequence  $x_{t+1}$ , and the previous memory state  $h_t$ . The algorithm will continue this iteration with all the neurons in the network.

Even though RNNs have been proven to achieve remarkable results in many applications, these network architectures are typically much harder to train. Given that it is not possible to simply backpropagate the error layer by layer, it is necessary to consider the additional time component, which amplifies the vanishing and exploding gradient problem [152]. However, one of the solutions to this problem is the use of Long Short Term Memory (LSTM) [153], a novel recurrent network architecture, that in conjunction with an appropriate gradient-based learning algorithm, is capable of learning long-term dependencies.

The LSTM network is a long chain of recurrent units called memory cells, also frequently referred to as LSTM cell. In the memory cell there is a small buffer element called cell state, represented as  $C_t$ , which maintains the global information at each time-step in the chain. A LSTM cell functions as a memory used to write, read and erase information, thanks to the internal mechanisms called gates, that can regulate the flow of information, learning to keep only relevant information to make predictions and forget non relevant ones. These gates contain sigmoid activation functions, which rescale the values in an interval  $[0, 1]$ , describing how much of each component should be let through, The closer to zero a value is, the smaller its importance, so should not be let through [154]. The gates can be summarized in:

- Forget gate: controls the amount of information received from the previous memory state. It decides, when a new input  $x_t$  is encountered, if the information from the cell state  $C_{t-1}$  it is going to be “forgotten” or “kept” from the cell state  $C_t$ . The
-

information from the previous cell and current input is passed through the sigmoid function, and will act as a filter and produces a value in range  $[0, 1]$  for each element in a cell state.

- Input gate: adds new information to the cell state  $C_t$  from the input  $x_t$ . In addition to the sigmoid activation, that decides which values should be updated, a tanh activation (Fig. 5.3) will create a vector of the new candidates  $C'_t$  that will be added to the cell state.
- Output gate: controls the flow of outgoing information. The hidden state  $h_t$ , at each memory cell is decided based on the updated cell state  $C_t$ , and the output vector.

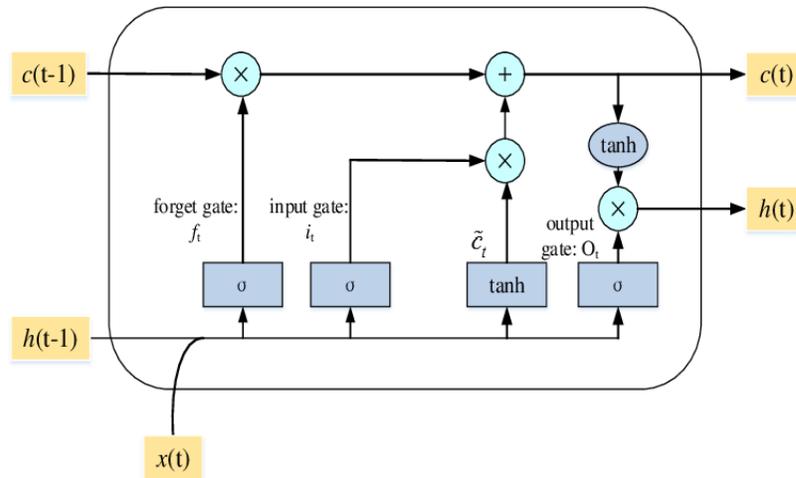


Figure 5.3: The structure of the LSTM unit [155].

# Chapter 6

## HH-BTag algorithm to select b-jets in $HH \rightarrow bb\tau\tau$ searches

This chapter describes a new machine learning algorithm, HH-BTag, meant to improve the selection of the b-jet pair in resonant and non-resonant  $HH \rightarrow bb\tau\tau$  searches. The performance of the algorithm will be evaluated and compared with respect to the standard procedure based on the use of b-tagging algorithms currently available at CMS, in simulated samples of resonant and non-resonant  $HH \rightarrow bb\tau\tau$ , and in different scenarios. The HH-Btag algorithm is used in the ongoing non-resonant search (Chapter 4) having shown better performance with respect previous method.

### 6.1 Motivation for a new algorithm

At the beginning of the analysis presented in Chapter 4, the two b-jet candidates that form the  $H \rightarrow bb$  decay (Section 4.4.2), in the  $HH \rightarrow bb\tau\tau$  final state, are selected as the two jets with the highest DeepFlavour score, that satisfy the Medium working point of the algorithm (see Fig. 4.4). The results obtained with this procedure show that in many cases the sub-leading jet is misidentified, since the true b-jet from  $H \rightarrow bb$  does not pass the Medium working point of the algorithm.

This situation motivated a study to assess the status of the reconstruction of the  $HH \rightarrow bb\tau\tau$  final state, with the aim of improving the signal extraction. The strategy for this study is to evaluate the efficiency and purity of the signal reconstruction, using as a reference the generator level objects, defined in Section 6.4, studying the match between the generator and the reconstructed physical objects (b-jet,  $\tau_\ell$  and  $\tau_h$ , described in Sections 3.2.4 and 3.2.5.1)

The available methods for the identification of the final state objects taus and b-jets are evaluated, using as a parameter the purity in the selection of the correct  $\tau$  and b-jet

---

pair candidates (Eq. 6.4). The procedure is the following:

- The reconstructed events are required to pass the kinematic selection described in Section 6.3, and the loosest working points of the DeepTau discriminator against electrons and muons.
- The reconstructed objects (taus or b-jets) are reordered according to the score given by the tagger to be evaluated.
- The two reconstructed objects with the highest scores of the tagger are selected.
- The reconstructed object is classified as correct, if a matching generator object is found within a cone of size  $\Delta R < 0.5$  around the direction of the reconstructed object.
- The event is classified as a full match, if both elements of the pair (taus or b-jets) are matched with generator objects.

The purity of the signal selection, using the best performing taggers for the b-jet and  $\tau$  selection (DeepFlavour and DeepTau, respectively), is presented in Tab. 6.1, for each channel, using a GGF 2017 SM simulated signal sample. Three scenarios are evaluated for each event: a full match in the selection of the b-jets, coming from the  $H \rightarrow b\bar{b}$  decay; a full match in the selection of the  $\tau$  pair, coming from the  $H \rightarrow \tau\tau$  decay; and a full match in the selection of both the b-jet and  $\tau$  pair.

	2 matches $H_{bb}$ [%]	2 matches $H_{\tau\tau}$ [%]	2 matches $H_{bb}$ & matches $H_{\tau\tau}$ [%]
$bb\tau_e\tau_h$	90.6	97.4	88.3
$bb\tau_\mu\tau_h$	90.5	98.7	89.5
$bb\tau_h\tau_h$	89.9	97.8	87.9

Table 6.1: Purity of the signal selection using the DeepTau and DeepFlavour taggers, for the  $\tau$  and b-jet selection.

In the  $\tau_h\tau_h$  final state, the correct selection of the  $\tau$  pair occurs  $\sim 97.8\%$  of the times, with the use of the DeepTau tagger, while the correct selection of the b-jet pair has a probability of the  $\sim 90\%$ , using the DeepFlavour tagger. The final purity for  $HH \rightarrow b\bar{b}\tau\tau$  is spoiled by the low purity in the selection of the b-jets. This issue requires the development of a new machine learning algorithm, capable of exploiting all the relevant event information in order to discriminate the two b-jets from the Higgs decay, in resonant (future) and not-resonant (ongoing) productions, in different scenarios of  $HH \rightarrow b\bar{b}\tau\tau$  searches.

## 6.2 Signal samples for HH production

For this study fully simulated samples of resonant and non-resonant HH samples are used and are described in the following.

### 6.2.1 Signal samples for non-resonant SM and BSM HH production

The signal samples include some of the BSM benchmark models of Tab. 1.4 and correspond to the original samples produced centrally by CMS at leading order (LO) precision with MADGRAPH5\_aMC@NLO. These samples are reported in tables Tables 6.2, 6.3 and 6.4, marked respectively as “GGF LO” “VBF LO”.

For the GGF process, the corresponding benchmark point is indicated in the string name. For the VBF production, the signal simulations correspond to variations of the  $\kappa_\lambda$ ,  $C_V$  and  $C_{2V}$  strength modifiers (Section 1.3.2.2) and the corresponding values of  $(\kappa_\lambda, C_V, C_{2V})$  are marked by the values of  $C3$ ,  $CV$  and  $C2V$  in the dataset name. All the samples reported, cover the phase space of non-resonant HH production used for the study described.

### 6.2.2 Signal samples for resonant HH production (BSM)

Signal samples are generated at leading order using MADGRAPH5\_aMC@NLO 2.6.0 (for 2016 simulation 2.4.2). The models describe the production through GGF or VBF of particles with narrow width (set to 1 MeV) that decay to two Higgs bosons with the mass of  $m_H = 125$  GeV.

Events are generated either for spin-0 radion production, or spin-2 graviton production. For each spin hypothesis several mass points are generated depending on the year and spin. The mass points are in a range from 250 GeV to 3000 GeV, depending on the production year and production mechanism, since both GGF or VBF production mechanisms are considered.

All the samples reported in Tabs. [6.5 - 6.12], cover the phase space of resonant HH production used for the study described.

	Dataset Name
GGF LO	/GluGluToHHTo2B2Tau_node_box.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_SM.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.2.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.9.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.10.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.11.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.12.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.13.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV.0.5.C2V.1.C3.1.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.5.C2V.1.C3.1.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.0.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.1.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.2.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.2.C3.1.13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic.v3-v2/MINIAODSIM

Table 6.2: List of non-resonant GGF and VBF signal samples MC datasets for 2016.

	Dataset Name
GGF LO	/GluGluToHHTo2B2Tau_node_SM.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.2.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.3.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.4.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.7.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.9.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node.12.13TeV-madgraph.correctedcfg/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV.1.5.C2V.1.C3.1.13TeV-madgraph/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.0.13TeV-madgraph/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.1.13TeV-madgraph/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.1.C3.2.13TeV-madgraph/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV.1.C2V.2.C3.1.13TeV-madgraph/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic.v14-v1/MINIAODSIM

Table 6.3: List of non-resonant GGF and VBF signal samples MC datasets for 2017.

	Dataset Name
GGF LO	/GluGluToHHTo2B2Tau_node_SM_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_2_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_3_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_4_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_5_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_6_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_7_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_8_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_9_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_10_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_11_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_12_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV_1.C2V_1.C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_0.5.C2V_1.C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1.5.C2V_1.C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1.C2V_1.C3_0_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1.C2V_1.C3_2_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1.C2V_2.C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Table 6.4: List of non-resonant GGF and VBF signal samples MC datasets for 2018.



	Dataset Name
VBF spin-0 LO	/VBFToRadionToHHTo2B2Tau.M-250_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-270_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v2/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-280_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-300_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v2/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-350_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v2/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-400_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v2/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-450_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-500_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-550_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-600_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-650_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-700_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-750_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v2/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-800_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-850_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-900_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1000_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1250_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1500_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1750_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM
/VBFToRadionToHHTo2B2Tau.M-2000_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM	
/VBFToRadionToHHTo2B2Tau.M-3000_narrow_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X.mc2017_realistic.v14-v1/MINIAODSIM	

Table 6.7: List of resonant VBF spin-0 signal samples MC datasets for 2017.







	Dataset Name
VBF spin-0 LO	/VBFToRadionToHHTo2B2Tau.M-250_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-260_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-270_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-280_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-300_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-320_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-350_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-400_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-450_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-500_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-550_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-600_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-650_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-700_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-750_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-800_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-900_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1000_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1250_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFToRadionToHHTo2B2Tau.M-1500_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
/VBFToRadionToHHTo2B2Tau.M-1750_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	
/VBFToRadionToHHTo2B2Tau.M-2000_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	
/VBFToRadionToHHTo2B2Tau.M-3000_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	

Table 6.11: List of resonant VBF spin-0 signal samples MC datasets for 2018.



### 6.3 Event Selection

The simulated MC samples are processed like data, to reconstruct the HH candidates. First step they have to pass is the trigger selections operated by the di-tau, VBF-, single- and cross-triggers paths (Tables 4.5, 4.6 and 4.8). An offline selection of the  $HH \rightarrow bb\tau\tau$  slightly different from the baseline selection (Section 4.4) is used. The requirements applied are more flexible than those from the baseline, where the goal was to achieve the highest signal purity; for this study, more flexible requirements allow the algorithm to study simultaneously a large number of scenarios, with different kinematics. The events are selected with the requirements reported in Tab. 6.13.

Channel	kinematic selection	$\tau_h$ and $\tau_\ell$ iso and ID
$\tau_\mu\tau_h$	$\mu : p_T > 10 \text{ GeV},  \eta  < 2.1$ $\tau_h : p_T > 20 \text{ GeV},  \eta  < 2.3$	Medium ID WP DeepTaufv2p1: Medium VSjet, Tight VSmu, VLoose VSe
$\tau_e\tau_h$	$e : p_T > 10 \text{ GeV},  \eta  < 2.1$ $\tau_h : p_T > 20 \text{ GeV},  \eta  < 2.3$	Medium ID WP DeepTaufv2p1: Medium VSjet, Tight VSmu, Tight VSe
$\tau_h\tau_h$	two $\tau_h : p_T > 20 \text{ GeV},  \eta  < 2.3$	DeepTaufv2p1: Medium VSjet, VLoose VSmu, VVLoose VSe

Table 6.13: Selection requirements.

The event is accepted if at least two  $\tau$ 's and two jet candidates are reconstructed, and in addition, the following set of requests are satisfied:

- In the semi-leptonic channels ( $\tau_e\tau_h$  and  $\tau_\mu\tau_h$ ), for the  $\tau$ 's identified as electrons or muons, the angular separation  $\Delta R$ , between the lepton (electron or muon) and the hadronic tau candidate is required to be greater than 0.1.
- The reconstructed b-jet candidates are ordered using the DeepFlavour b-tagger (Section 3.2.5.1), each jet must have  $p_T > 20$ ,  $|\eta| > 2.4$ , and  $\Delta R(\tau_h, jet) > 0.5$ . The last requirement is applied to avoid in the event the overlap between the reconstructed b-jet and  $\tau_h$  candidates (since both, are jets identified as b-jets and/or tau jets).

### 6.4 Ground Truth definition

With a supervised machine learning approach, the model is trained using the ground truth information as expected output of the network [156]. The ground truth is based

on MC truth matching, assigning a true b-jet label to the reconstructed b-jet ( $b_{reco}$ ) from the Higgs boson candidate decay if at generator level a b-jet ( $b_{gen}$ ) from the Higgs boson decay is found to satisfy some criteria.

However, the  $b_{gen}$  4-momentum must be evaluated taking into account that the direct Higgs boson daughters 4-momentum is not accurate enough for the  $H \rightarrow bb$  decay, due to the fragmentation and hadronization in the b-quark decay. In CMS, the algorithm to associate hadron flavor to reconstructed jet using MC truth is not suitable, since it does not provide information about the particle from which the jet was originated.

For this reason, a method to obtain such association was implemented in this work, using the truth information from the Higgs boson decaying into  $bb$ , and the baryons and mesons from hard processes in the b-decay.

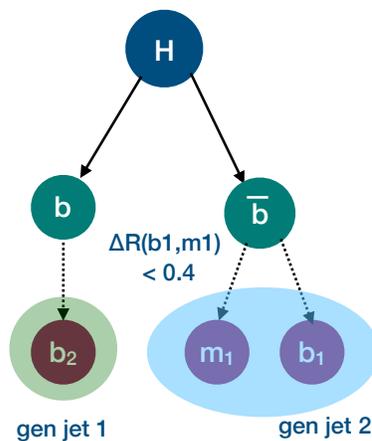


Figure 6.1: Redefinition of the generator jet used for the ground truth definition.

Most of the events have two hard hadrons, produced by the hadronization, suggesting the use of them to define b-jets at generator level; however, there is a considerable amount of events with more than two hard hadrons. In order to take into account more than two hard hadrons, the selected generator level baryons and mesons were ordered in energy and regrouped according to their angular separation in the following way:

- Generator level baryons and mesons inside a cone  $\Delta R < 0.4$  around the most energetic hadron, are grouped.
- Jets used above are excluded, and procedure is repeated until there are not hadrons ungrouped.

Each of these groups of baryons and mesons are a redefinition of generator level b-jets ( $b_{gen}$ ), a picture of this selection is represented in Fig. 6.1.

As shown in Fig. 6.2a, obtained with a SM GGF 2017 sample, using the regrouping algorithm to define  $b_{gen}$  jets, there are only three scenarios: events with two  $b_{gen}$  jets (86%

of the events), with more than two  $b_{gen}$  (10% of the events), and with only one  $b_{gen}$  jet (4% of the events). To evaluate the performance of the HH-BTag, the few events with only one (redefined)  $b_{gen}$  jet are dismissed. The Higgs boson in these type of events is very energetic, thus the jets are produced in a very narrow cone and they are not separated in to two jets from AK4 algorithm.

In Fig.6.2b the number of baryons plus mesons from hard processes versus the number of  $b_{gen}$  jets defined using the regrouping algorithm is reported (after the removing of the events with only one  $b_{gen}$  jet). It can be seen that most of the events have the two  $b_{gen}$  jets formed each one by one baryon or one meson, while the second most probable configuration is two  $b_{gen}$  jets, formed regrouping three hadrons (baryons and/or mesons).

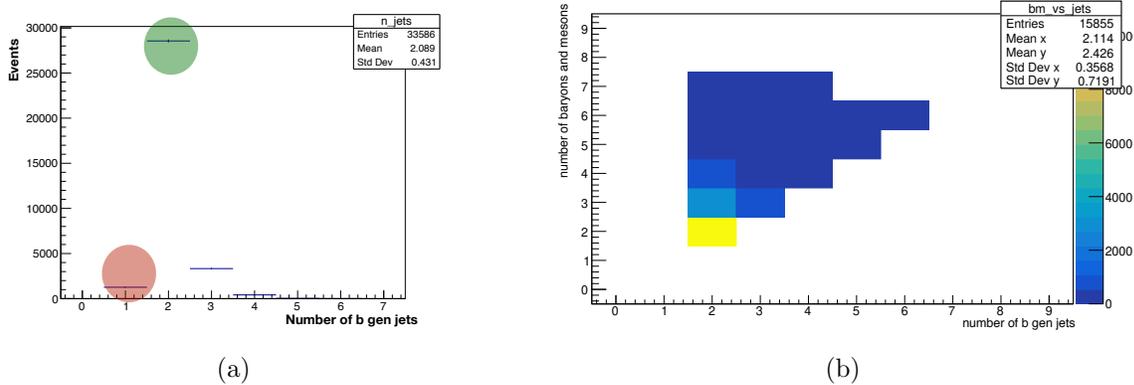


Figure 6.2: Information of the generator jets, using a SM GGF 2017 MC sample. (a) Distribution of the number of  $b_{gen}$  jets produced by the regrouping algorithm. (b) Number of hadrons from hard processes vs the number of  $b_{gen}$  jets defined using the regrouping algorithm.

In Fig.6.3, a comparison between using the two most energetic hard hadrons in the event to reconstruct the  $b_{gen}$  pair, and regrouping the hard hadrons is shown. In Fig.6.3a the distribution of the invariant mass of the  $b_{gen}$ , calculated using both reconstruction methods is reported. The reconstruction of the Higgs mass is improved using the regrouping algorithm, producing a better resolution and a peak slightly shifted to a higher mass.

In Fig.6.3b, the distribution of the relative energy loss defined as the ratio between the sum of the energies of the hard hadrons excluding the two most energetic hadrons (baryons and mesons) and the sum of the energies of all the hard hadrons is reported in red; while the same variable but defined this time as the ratio between the sum of the energies of the  $b_{gen}$  jets excluding the two most energetic, and the sum of the energies of all the  $b_{gen}$  jets, is reported in blue. As shown in this figure, the loss of energy is smaller when using the regrouping and the two most energetic  $b_{gen}$  jets are chosen: in this case,

around 80% of the energy is taken by the most two energetic  $b_{gen}$  jets.

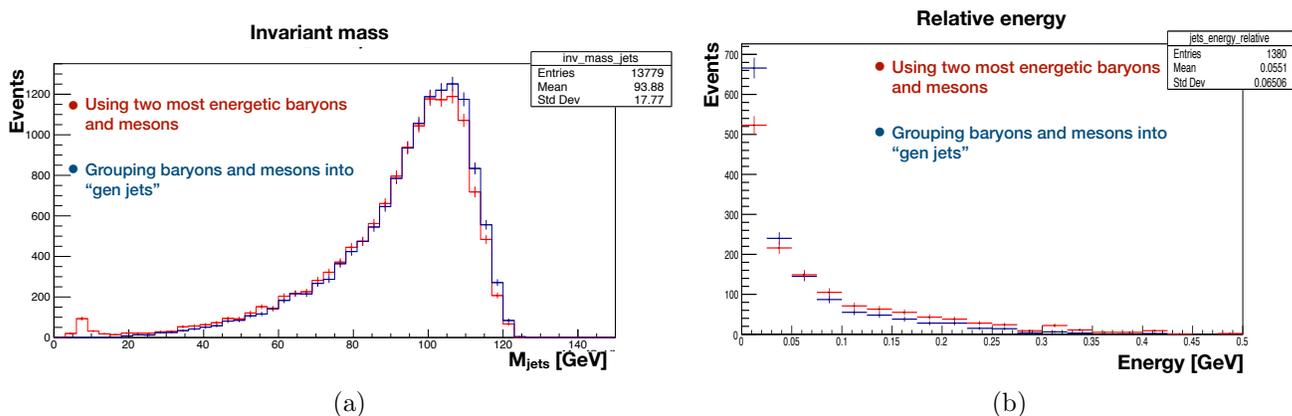


Figure 6.3: Comparison of the generator jets definitions, using a SM GGF 2017 MC sample: (a) invariant mass of the generator level jets using both methods to reconstruct the jets (b) relative energy of the gen jets.

In summary, the ground truth:

- Uses the two most energetic  $b_{gen}$  jets to perform the match.
- The true b-jet label is assigned to the reconstructed b-jet ( $b_{reco}$ ) if a matching  $b_{gen}$  is found within a cone of size  $\Delta R < 0.5$  around the direction of the  $b_{reco}$ .

## 6.5 Features and inputs

The definition and selection of the variables (features) used as inputs of a NN is a very important step in the development of trained discriminators. Several high-level variables were investigated, and the most discriminant ones were retained.

The training (Section 6.6.5) used a total of 14 variables that can be subdivided in four groups:

- Information of the b-jet candidate: score of the b-jet discriminator and kinematic variables.
- Information of the  $H_{\tau\tau}$  candidate: kinematic variables and MET.
- The angular separation between the b-jet candidate and the  $H_{\tau\tau}$  candidate.
- Global event variables.

In order to cover the entire phase space of the signal processes (resonant and not resonant, mass, spin, etc...), the kinematic variables were chosen so that they were independent of a specific scenario.

In the following, a short description of these variable.

#### 6.5.0.1 Information of the b-jet candidate

- The score of the b-jet candidates, given by the DeepFlavour b-tagger [108].
- The  $p_T$  and  $\eta$  of each of the b-jet candidate from the Higgs boson decay.
- The relative mass and energy of the b-jet candidate defined as  $\frac{M}{p_T}$  and  $\frac{E}{p_T}$ , where M is the invariant mass, E the energy, and  $p_T$  the transverse momentum of the b-jet candidate.

The use of the relative variables allows to reduce correlations with  $p_T$  compared to the use of absolute values.

#### 6.5.0.2 Information of the $H_{\tau\tau}$ candidate

- The scalar sum of the visible  $p_T$  of the  $H \rightarrow \tau\tau$  candidate, calculated as  $p_{T,H_{\tau\tau}} = \sum \tau_{1p_T} + \tau_{2p_T}$ .
- The visible  $p_T$  and pseudorapidity  $\eta$  (defined in Section 2.2.1) of the  $H \rightarrow \tau\tau$  candidate, where the visible momentum of the  $H_{\tau\tau}$  is calculated as the sum of the visible 4-momentum of the  $\tau$  pair candidates.
- The azimuthal angle difference between the MET (Section 3.2.6) and the visible  $H_{\tau\tau}$  candidate ( $\Delta\phi(MET, H_{\tau\tau,vis})$ ). For the signal events, the missing momentum usually points to the Higgs boson reconstructed from the  $\tau$  pairs. In fact, since neutrinos are principally produced in the direction of the  $\tau$  decays, signal events have a small angular separation, while for background events this variable is expected to be randomly distributed.
- The relative momentum of the missing energy, calculated as:  $\frac{p_T(MET)}{p_T(H_{\tau\tau})}$ .

#### 6.5.0.3 Angular separation between the b-jet candidate and the $H_{\tau\tau}$ candidate

The b-jet candidates are required to be out of a cone of radius  $\Delta R = 0.5$  around the  $H_{\tau\tau}$  candidate.

The variables used are related to the angular separation respectively in the pseudorapidity and azimuthal angle ( $\eta, \phi$ ):

---

- $\Delta\eta(\text{jets}, H_{\tau\tau})$ .
- $\Delta\phi(\text{jets}, H_{\tau\tau})$ .

#### 6.5.0.4 Identification of the event

These variables allow to identify the year of the sample (different behavior of data) and the decay modes of the  $\tau\tau$  pair from the Higgs decay:

- Year of the sample.
- Channel. In the final state of the  $HH \rightarrow bb \tau\tau$  analysis three decay channels for the  $H \rightarrow \tau\tau$  decay are studied:  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$ .

To help the convergence and improve the stability during the training, a scaling or normalization procedure is applied to some of the input variables (Section 5.2). A linear transformation is applied to the features that have a significantly non-Gaussian distribution on the finite validity domain, to confine their values in a range  $[-1,1]$ . For features with a relatively narrow peak in their distributions a standardization procedure is instead applied, changing the  $x$  variable to  $x \rightarrow (x - \mu)/\sigma$ , where  $\mu$  and  $\sigma$  are respectively the mean and the standard deviation of the  $x$  distribution. The standardized inputs are truncated to  $[-5, 5]$  in order to protect the training against the outliers. The purpose of the pre-processing of the inputs before creating the machine learning model is to change the numerical values in a dataset to use a common scale, without distorting differences in the range of values or losing information. Many algorithms attempt to make trends in the inputs by comparing features of the input points, however there is an issue when the features are in drastically different scales: the features with the larger scales completely dominates the others [157].

In Tab. 6.14 is reported in the first column the variable used, in the second column when the standardization procedure is applied, are reported the  $\mu$  and the  $\sigma$  of the distribution of the variable, calculated in the training using the training set (Section 6.6.5), while when linearization procedure is applied, in the third column the linear transformation applied in the range of variable is reported.

The importance of each of the variables used in the final training of the algorithm is presented in Section 6.6.5.2.

## 6.6 HH-BTag algorithm

The HH-BTag discriminant is created through a Recurrent Neural Network (RNN) (Section 5.2) implemented in Keras with TensorFlow backend [158, 159].

Variable	Std. transform		Linear interval transformation
	$\mu$	$\sigma$	
jet score DeepFlavour	-	-	-
jet $p_T$	-	-	[20, 5000] $\rightarrow$ [-1, 1]
jet $\eta$	-	-	[-2.4, 2.4] $\rightarrow$ [-1, 1]
jet $M/p_T$	0.139	0.056	-
jet $E/p_T$	1.722	0.938	-
$H_{\tau\tau} p_T$	-	-	[0, 4000] $\rightarrow$ [-1, 1]
$H_{\tau\tau,vis} p_T$	-	-	[0, 4000] $\rightarrow$ [-1, 1]
$H_{\tau\tau,vis} \eta$	-	-	[-8.5, 8.5] $\rightarrow$ [-1, 1]
$\Delta\phi(MET, H_{\tau\tau,vis})$	-	-	-
$MET(p_T)/H_{\tau\tau}(p_T)$	0.761	0.758	-
$\Delta\eta(\text{jets}, H_{\tau\tau,vis})$	-	-	[-9.1, 9.1] $\rightarrow$ [-1, 1]
$\Delta\phi(\text{jets}, H_{\tau\tau,vis})$	-	-	[-3.142, 3.142] $\rightarrow$ [-1, 1]
Year	-	-	[2016, 2018] $\rightarrow$ [-1, 1]
Channel	-	-	[0, 2] $\rightarrow$ [-1, 1]

Table 6.14: List of input variables.  $\mu$  and  $\sigma$  are the mean and the standard , used when the standardization transformation is applied. A detailed description can be found in the text.

The general idea of how the neural network architecture should be, which was be the basis for the optimization, can be described in three blocks, the range and/or type of the hyperparameters that will be listed in the following, are described in Tab. 6.15, also a scheme of this is shown in Fig. 6.4, and a description of them is presented in Chapter 5 . A more complete description of the final architecture is given in the Section 6.6.4.

- **Pre TimeDistributed Dense Block**

Hyperparameters to optimize: *number of TimeDistributed Dense layers, number of neurons, activation function, and dropout value for each of the layers.*

- **Recurrent Neutral Network Block**

Hyperparameters to optimize: *type of RNN, number of RNN layers, number of neurons and dropout value for each of the RNN layers.*

- **Post TimeDistributed Dense Block**

Same as the Pre TimeDistributed Dense Block.

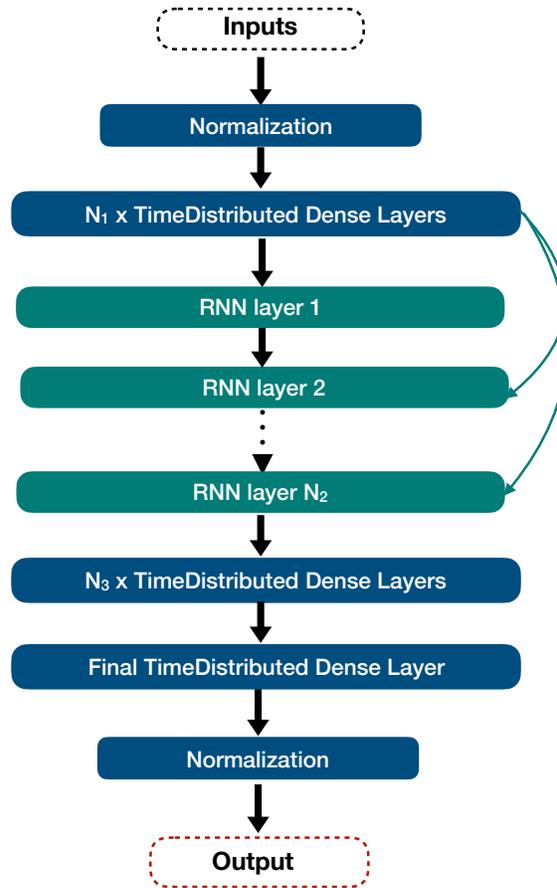


Figure 6.4: Scheme of the Neural Network.

### 6.6.1 Optimization of the NN hyperparameters

The use of machine learning algorithms involves a careful tuning of the model hyperparameters. Unfortunately, this tuning is often difficult and requires expert experience, or sometimes brute-force search and it is not time efficient. Therefore, there is a great appeal for automatic approaches that can optimize the performance of any given learning algorithm to the problem at hand.

The Bayesian optimization is a machine-learning-based optimization method, focused on objective functions that take a long time to evaluate. It is best-suited for optimization over continuous domains of less than 20 dimensions, and tolerates stochastic noise in function evaluation. This method differs from other available methods because it depends on the past evaluations to improve the search speed, thus the performance of the evaluated hyperparameters affects the future decision [160]. This optimization is well suited to our case, since we have a hyperparameter space with 14 dimensions, the target function is not convex, there is a huge number of local maxima, and the evaluation of the function takes

around one hour. The Bayesian Optimization package used for this NN can be found in Ref. [161].

A reasonable range of the hyperparameters is chosen taking into account the number of degrees of freedom and statistics of the samples evaluated. The studied hyperparameters with their description and the range of values are reported in Tab. 6.15. A diagram that represents how the hyperparameters are used inside the NN is shown in Fig. 6.4. During the training around 300 combination points inside the hyperparameter space were studied.

### 6.6.2 Theory behind Bayesian optimization

The Bayesian optimization tries to locate an input  $x$  in a certain domain  $A$ , which maximizes (or minimizes) the value of the function  $f(x)$  over the domain  $A$ :

$$\max_{x \in A} f(x), \tag{6.1}$$

where:

- The input  $x$  is one of the possible combinations of the parameters inside the domain  $A$ .
- The domain  $A$  is the set of parameters inside the range chosen, inside the  $\mathbb{R}^d$  domain, where  $d$  is the number of parameters that will be optimized.
- The objective function  $f : A \rightarrow \mathbb{R}^d$  is the function to be evaluated. In our case, the function to be maximized is the purity of the selection of the b-jet candidates pair, defined as the ratio between the number of the  $H_{bb}$  reconstructed candidates identified by the algorithm in HH  $\rightarrow$  bb $\tau\tau$  events, and the total number of  $H_{bb}$  reconstructed candidates (true plus false) tagged by the algorithm defined in Eq. 6.4.

The Bayesian optimization can be summarized in the following steps:

1. Choose a prior probability function over the space of the objective function  $f(x)$ , which represents the initial belief about the space of possible objective functions.
  2. Combine the prior and the likelihood functions to get a posterior probability function, which captures the updated beliefs about the unknown objective function.
-

Hyperparameter	Range / Functions	Optimized value
Number of TimeDistributed dense layers (Pre Block)	[0 - 5]	0
Number of neurons TimeDistributed dense layers (Pre Block)	[15 - 60]	-
Dropout value TimeDistributed dense layers (Pre Block)	[0 - 0.5]	-
Activation TimeDistributed dense layers (Pre Block)	sigmoid, relu and tanh	-
RNN type	LSTM, GRU, SimpleRNN	LSTM
Number of RNN layers	[3 -10]	5
Number of units for the RNN layers	[20 - 200]	74
Dropout RNN layers	[0 - 0.5]	0
Number of TimeDistributed dense layers (Post Block)	[0 - 5]	10
Number of neurons TimeDistributed dense layers (Post Block)	[15 - 60]	15
Dropout value TimeDistributed dense layers (Post Block)	[0 - 0.5]	0
Activation function TimeDistributed dense layers (Post Block)	sigmoid, relu and tanh	sigmoid
Exponential learning rate	[-6, -1]	-3
Optimizers	RMSprop, SGD, Adam and Nadam	Adam
Batch size	[100 - 500]	100

Table 6.15: NN Hyperparameters optimization. The NN hyperparameters to be optimized, their range or the set of functions that describe the hyperparameter and the value of the parameter or the function obtained from the optimization procedure (Section 6.6.1) are reported. The optimized values will be used by HH-BTag in the training and validation.

3. As the number of observations grows, the posterior distribution improves, and the algorithm becomes more certain of which regions in the parameter space are worth exploring.

In principle, the shape of the objective function  $f(x)$  is unknown, and compute every

single point of the function it is impossible and inefficient. However, it is possible to build a surrogate model to approximate the true objective function, using a given set of initial points. A surrogate model is a method used when an outcome of interest cannot be easily directly measured, so that a model of the outcome must be used instead. Surrogate models are popularly used to substitute black-box models which are either computationally expensive to evaluate, have noisy output behavior or do not supply gradients. When new hyperparameters are generated and the objective function score is calculated, the surrogate model is updated with this new information, bringing it closer to the true underlying model.

A popular surrogate model for Bayesian Optimization, which is used in our work, is Gaussian Process (GP), due to its simplicity and easy optimization. The GP directly models the surrogate model  $p(y|x)$  as a probability of the true objective function score ( $y$ ) and the hyperparameters ( $x$ ). It uses the available information of the pair  $(x, y)$  to construct the multivariate Gaussian distributions, and it models each point in  $x \in A$  as a Gaussian random variable. As such, it is completely characterized by the mean  $\mu(x)$  and the variance function  $\sigma^2(x)$ . As enumerated before, the GP defines a prior over functions (step 1), and after having observed some function values, posterior probability function is defined (step 2). The GP posterior is computationally cheap to evaluate and is used to propose points in the search space where sampling is likely to yield an improvement. These steps are repeated until the maximum number of iterations is reached.

Proposing sampling points in the search space is done by acquisition functions. Acquisition functions are mathematical techniques that guide how the parameter space should be explored during Bayesian optimization. They use the predicted mean and variance generated by the GP model, into a criterion that will direct the search evaluated. There is usually a trade-off between two strategies, with the goal of maximizing the acquisition function to determine the next sampling point:

- Exploitation means sampling where the surrogate model predicts a high objective, focusing on results in the vicinity of the current best results by preferring low variance values.
- Exploration means sampling at unexplored locations where the prediction uncertainty is high.

There are different scenarios, if an acquisition function does not explore enough, it is possible will only exploit sub-optimal regions, and the global optima is not going to be found. The other case is if it explores too much, since could continue evaluating points trying to improve an already adequate Gaussian process. This makes balancing exploration and exploitation vital.

---

The acquisition function used in this work is the Upper Confidence Bound (UCB), often phrased as “optimism in the face of uncertainty”, because despite the lack of knowledge respect to what could be the best hyperparameters set to construct an optimistic guess, it is chosen the action with the highest expected payoff. In the case the guess was wrong, in the next iteration, the optimistic guess will decrease and will switch to a different action continue exploring. Instead if the guess achieve a good result, will continue exploiting that region. In this way it is balanced the exploration and exploitation.

The exploration-exploitation trade-off is controlled by the variable  $\kappa$ , which captures the number of standard deviations in the upper confidence bound. The acquisition function  $\alpha(x)$  can be written as:

$$\alpha_{t+1}^{GP-UCB}(x) = \mu_t(x) + \sqrt{\kappa}\sigma_t(x), \quad (6.2)$$

for a given time  $t$ ; increasing  $\kappa$  makes the acquisition function favor points with high variance, causing more exploration. Decreasing  $\kappa$  will make the acquisition function favor points closer to the mean, causing more exploitation [162, 163].

### 6.6.3 Guided Optimization

After having evaluated a certain number of points of the parameter space, the regions where the maximum of the function might lie were reduced. In order to arrive to the best result in a shorted time, the optimization was stopped and some actions were performed before restarting, and afterwards it was stooped-restarted several times, in order to tune some minor hyperparameter. Some of these actions were:

- Specify points to be probed. The BayesianOptimization package [161] allows to specify points to be probed before the GP takes over, while initially the currents optimal points were used as a probe points.
- Reduce the number of hyperparameters to be optimized. Some of the hyperparameters were removed, and then set to their optimized values, among these: RNN type, batch size, activation function for the dense and RNN layers and optimizer.
- Reduce the exploitable range for some of the hyperparameters.
- Explicitly write the dependency between some of the hyperparameter, e.g, if the numbers of number of dense layers is set to zero, the number neurons of the dense layers is set to its minimal value.
- Each time the optimization is restarted all the previous probed pointed are read and stored as used for this round.

At the beginning of the optimization, the region where the maximum lie was unknown, thus the goal was to explore the hyperparameter-space, to achieve that, values of  $\kappa < 1$  were chosen (Eq. 6.2). In this second stage, it was clearer where optimal point could lie, so it was necessary to exploit the surroundings around the optimal points in that moment, and as a consequence values of  $\kappa$  in the interval  $[1,2]$  were used.

As an example, in Fig. 6.5, one-dimensional GPs are shown, for six of the hyperparameters optimized. First, the ten sets of hyperparameters  $s_{best}$  which yield the highest purity (Eq. 6.4) were selected. Then, for a given hyperparameter  $h_i$  at a time, its ten values  $v_{h_i}$  from the previously defined sets are taken and shown in the plot, together with the corresponding purity values. This is done in the plane in which the current maximum lies, which means that the remaining hyperparameters are set to the optimized value shown in Tab.6.15. In each plot, the dotted line represents the mean of the GP surrogate model prediction of the objective function (in our case the purity) evaluated with the ten sets of hyperparameters, the green and yellow bands indicates the regions containing the 68 and 95% confidence intervals of the purity, representing the confidence of the algorithm in its current representation of the objective function. During the optimization, the acquisition function selects the next point to be evaluated based on the uncertainty as shown in Fig. 6.2, so these plots give useful information about the current status of the optimization. Indeed, it is possible to evaluate if the optimization is in a exploration of exploitation stage, and tune some of the hyperparameters by hand, for example proposing some initial points to probe or reducing the range, these strategies were performed in the last part of the optimization.

---

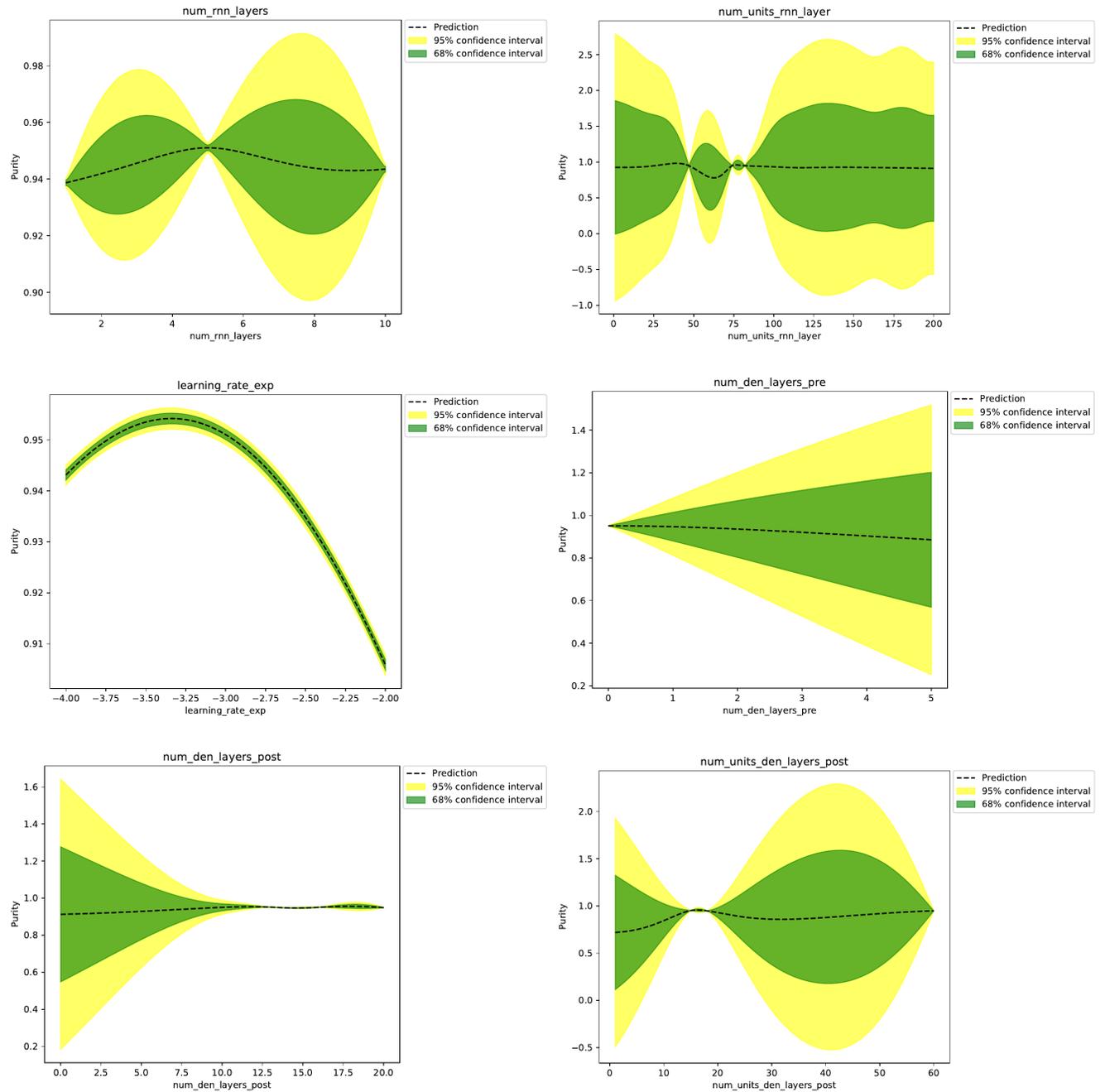


Figure 6.5: Each plot represents a one-dimensional GP for one of the hyperparameters optimized, in the plane where the maximum lies (Tab.6.15). There are shown six of the most relevant hyperparameters. In each plot, the dotted line represents the mean of the GP surrogate model prediction of the objective function (in our case the purity of the selection of the b-jet candidate pair from Eq. 6.4) evaluated with the ten sets of hyperparameters which yield the highest purity, and the shaded area shows the mean ( $\mu$ ) plus and minus the variance. The green and yellow bands indicates the regions containing the 68 and 95% confidence intervals of the purity, representing the confidence of the algorithm in its current representation of the objective function.

### 6.6.4 Final architecture

The details of the final architecture of the algorithm, obtained by the optimization of the hyperparameters, reported Tab. 6.15, are summarized in the following:

1. **Pre Block** *Number of TimeDistributed dense layers* is null.
2. **RNN Block** The *RNN type* is LSTM. In this block were used (Tab. 6.15), 5 concatenated *LSTM layers*, each of them with 74 *units*, with a sigmoid as *activation function*.

After each LSTM layer, batch normalization layer is applied. Initially, a dropout was also included, which randomly drops out nodes during training to reduce overfitting. However, the optimization concluded it was not necessary when the batch normalization was used, due to the slight regularization effect of the batch normalization. In addition, the batch normalization adds some noise to each hidden layer's activations, this is a way to avoid overfitting without losing information, which is the case for the dropout.

3. **TimeDistributed Dense layers Block** Has a number of 10 *TimeDistributed Dense layers* with 15 *units* each, with a sigmoid as *activation function*. As for the RNN block, a batch normalization layer was applied between each layer. After this, a final TimeDistributed Dense layer with only one units and with a sigmoid activation function, is set, in order to have an output between 0 and 1, corresponding to the score of each b-jet candidates.

In order to have a score of the jet pair of the event in the range  $[0,1]$ , the final output is normalized to two since for each event there should be two b-jets candidates each one with a score in the interval  $[0,1]$ .

### 6.6.5 Training

The signal samples considered in the training are both the non-resonant and the resonant HH production (Section 6.2), and the events are selected using the event selection described in Section 6.3.

The training of a ML model is the process in which the model determines the correct values of the weights that reduces the loss function, following the procedure explained in Section 5.1.2, and improves the performance measure for the task, in our case selecting correctly the b-jet pair from the  $H \rightarrow bb$  decay, a process known as the optimization of the ML.

Before performing the training it is necessary to split the dataset into three sets:

- **Training dataset:** the model is trained on this dataset, which means that this is the dataset used to minimize the loss function.
- **Validation dataset:** it serves as a guide of the performance of the model during the training. Usually in each epoch (Section 5.1.2.2, some parameters like the loss function or the measure for the task (in our case the purity of the selection of the b-jet pair candidates Eq. 6.4) are calculated using this dataset. Evaluating the training model with this set has several purposes, for example to understand how the training is performing (the evaluation for each epoch of the loss function and purity are shown in Figs. 6.6 and 6.7, respectively), choose the best model at training time, or to select the set of hyperparameters during the optimization.
- **Test dataset:** Similar to the validation dataset, provides an unbiased evaluation of the final model trained with the training dataset. However, it is used to evaluate only the final performance of the model after the training is finished.

The total number of events is divided into two samples, the even and the odd sample ( $sample_{even}$  and  $sample_{odd}$ ), depending on the number of the event. Then, two separated trainings and tests are performed. In one case, the odd sample is used in the training, while the even sample is used for the test, in the other case the even sample is used in the training, while the odd sample is used for the test. In both trainings, the validation set is taken from the training dataset, and represents the 25% of the total training dataset.

During the training, a weight is applied to each event according to its production cross-section, production mechanism, process, decay channel and integrated luminosity. This procedure allows to renormalize the events in order to have equal importance for different processes.

The training was performed using as loss function the binary cross-entropy described in Eq. 5.4 in the Section 5, minimized with the Adam algorithm [150].

The early stopping technique was used in this work. This technique, based on the evaluation of the validation set, stops the learning when the model no longer improves and it is prone to start overfitting. The metric or parameter used to monitor when to stop the training was the purity of the selection of the b-jet pair candidate (Eq. 6.4). In the early stopping technique, the number of epochs with no improvement after which the training is stopped is referred to as patience, and its value was set to 10 epochs. The model was trained with 25 epochs.

The set of weights saved during training, was the set which yields the highest value of the purity selection (Eq. 6.4). These weights will be used later to evaluate the performance of the model in Section 6.7 and the importance of the features used in the training.

### 6.6.5.1 Evaluation of the training

The evolution of the loss function and the purity (defined in Section 6.7, Eq. 6.4) as a function of the epochs completed during the training, evaluated using the training and validation sets (both sets are defined in Section 6.6.5), are shown in Fig. 6.6 and Fig. 6.7, respectively.

The best performance on the validation set, evaluated using the purity, is achieved after 15 epochs. After that, it can be observed that the purity stops improving in both trainings, especially the training performed with the odd events becomes very unstable, and the purity evaluated using the validation deteriorates (Fig. 6.7). Similar results are obtained by the loss function, in which after 15 epochs it becomes very unstable and does not have any kind of improvement (Fig. 6.6).

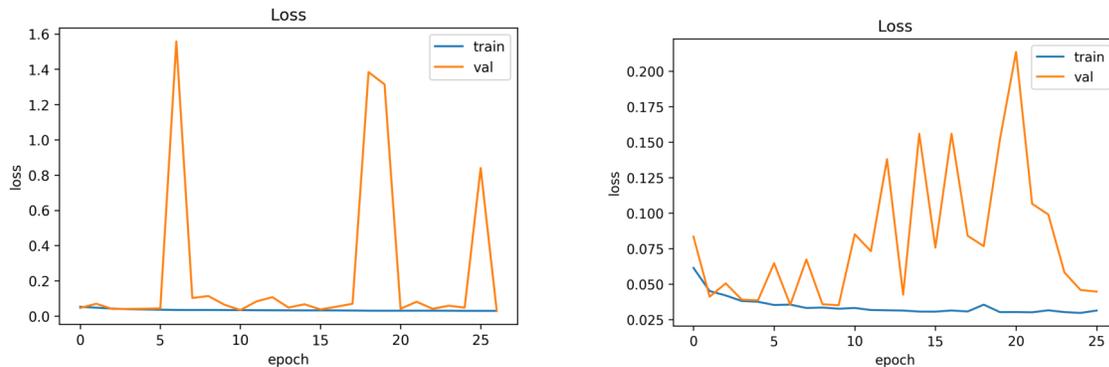


Figure 6.6: Evolution of the loss function as a function of the epoch number, for the pair events training (left) and the odd events training (right). Curves in blue represent the evaluation using the training dataset, while the evaluation using the validation dataset shown in orange.

### 6.6.5.2 Permutation feature importance

The permutation feature importance method measures the increase in the prediction error of the model when single feature values are randomly shuffled or permuted. This method was introduced by Breiman in Ref. [164]. A feature is said to be important if shuffling its values increases the model prediction's error, that means the model highly depends on this feature, and therefore the performance is negatively affected. In the opposite case, a feature is said to be unimportant if after shuffling its values, the performance of the model is not affected, which means the model is not taking advantage of this parameter, thus ignoring the feature while calculating the predictions [165].

The importance score of the  $i$ -th feature ( $f_i$ ) is defined as the reduction in performance after the values of  $f_i$  are randomly shuffled or permuted. In this work, the performance

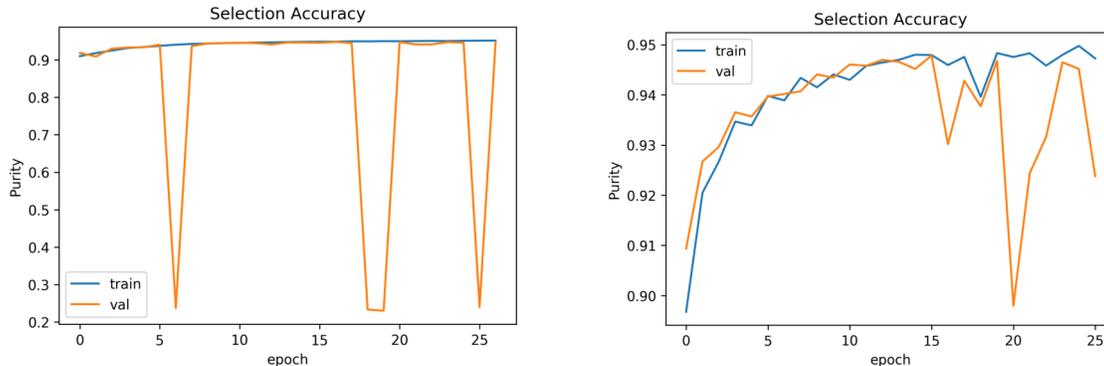


Figure 6.7: Evolution of the purity as a function of the epoch number, for the pair events training (left) and the odd events training (right). Curves in blue represent the evaluation using the training dataset, while the evaluation using the validation dataset shown in orange.

is evaluated through the purity of the selection of the correct b-jet pair candidates ( $P$ ) defined in Eq. 6.4,

$$I(f_i) = P - P_{f_i, \text{shuffled}}. \quad (6.3)$$

The model used for the purity evaluation the set of weights determined by the training. The importance score is calculated for each of the features in both trainings (made with the even and odd sample). The purity is evaluated with the test dataset of the opposite parity respect to the weights file used:  $w_{\text{even}}$  is evaluated with  $\text{sample}_{\text{odd}}$  and vice versa.

In the Fig. 6.8, is shown the importance of each of the features for both of the trainings performed. The most important features are the score of the b-jet candidate given by the DeepFlavour b-tagger,  $\eta$  of the jet candidates and the  $\Delta\phi$  between the b-jet and the  $H_{\tau\tau}$  candidates.

## 6.7 HH-BTag performance

In this section is presented a comparison of the performance of HH-BTag with respect to the standard classifiers, DeepFlavour and DeepCSV (Section 3.2.5.1).

All the processes considered for the comparison are described in Section 1, and the corresponding simulated samples are listed in Tables 6.2 - 6.12.

### Purity definition

In order to evaluate the performance of the new discriminator it is necessary to compute the predictions made by the trained model. The predictions were calculated using

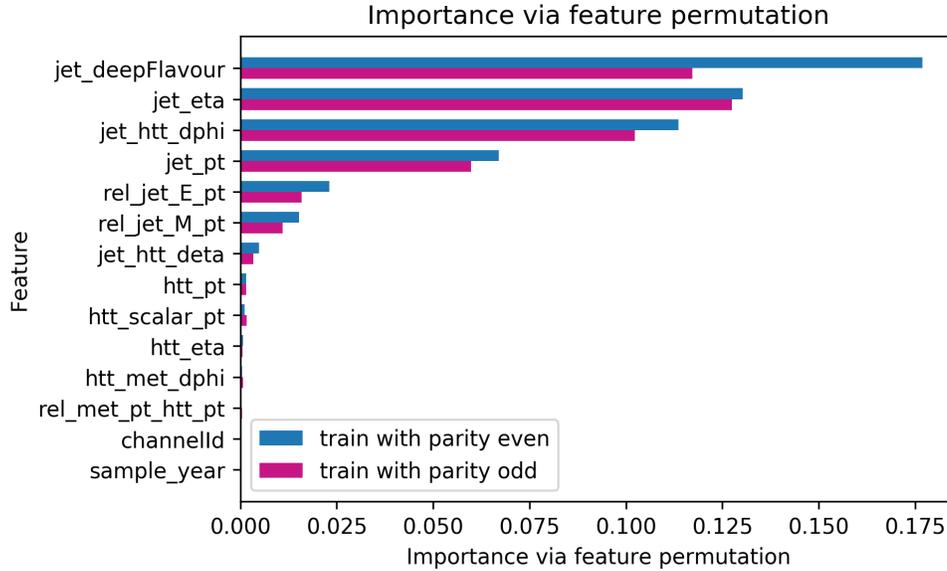


Figure 6.8: Permutation importance for the selected features used for the training of the NN. The importance score calculated using the weights from the training made with the even sample is shown in blue, and in pink using the odd sample.

separately each set of weights determined by the training ( $w_{even}, w_{odd}$ ), evaluated on the corresponding validation set (see Section 6.6.5), using as input to be labeled the sample with the opposite parity. The output of the HH-BTag assign a score in the interval  $[0, 1]$  to each of the b-jet candidates, depending on the confidence of the model that the b-jet in question is one of the b-jet’s from the  $H \rightarrow bb$  decay.

After applying the selections detailed in Section 6.3, once the candidates  $HH \rightarrow bb\tau\tau$  are reconstructed, the pair of b-jets candidates with the highest score of the classifier (which is being evaluated, i.e HH-BTag, DeepFlavour and DeepCSV) is selected. The variable used to evaluate the performance of the discriminators, is the purity of the selection of the b-jet pair candidate defined as:

$$purity^{classifier} = \frac{N_{true}^{classifier}}{N^{classifier}} \quad (6.4)$$

where  $N_{true}^{classifier}$  is the number of events in which the selection of the b-jet pair candidate matched with the ground truth (defined in Section 6.4), and  $N^{classifier}$  is the total number of events where a candidate is reconstructed and the b-jet pair forming the  $H \rightarrow bb$  candidate is selected using the chosen classifier.

The results shown in this section, are produced using the combination of the purities ( $p_{even}, p_{odd}$ ), computed with the predictions evaluated with the sets of weights ( $w_{even}, w_{odd}$ ), respectively. The combination’s values and errors were obtained in the following way:

- The value is the mean between  $p_{even}$  and  $p_{odd}$ .
- The errors are calculated using the confidence intervals for binomial proportions (using the Clopper-Pearson interval based on beta distribution) providing as input the number of successes ( $N_{true}^{classifier}$ ) and number of trials ( $N^{classifier}$ ) [166].

### 6.7.1 Gluon fusion non-resonant

The MC samples used are listed in Tables 6.2, 6.3, 6.4. These samples are produced with different combinations of the coupling modifiers  $\kappa_\lambda$  and  $\kappa_t$  and of couplings  $c_2$ ,  $c_g$  and  $c_{2g}$ . These samples includes some of the twelve benchmark modelling points, reported in Tab. 1.4. In addition to these samples, the SM and the box modelling points have in the dataset name the string “SM” or “BOX”. The purity of the three classifiers as a function of the modelling point of the GGF non-resonant process simulated is shown for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.9, 6.10 and 6.11, respectively for the data taking years 2016, 2017 and 2018.

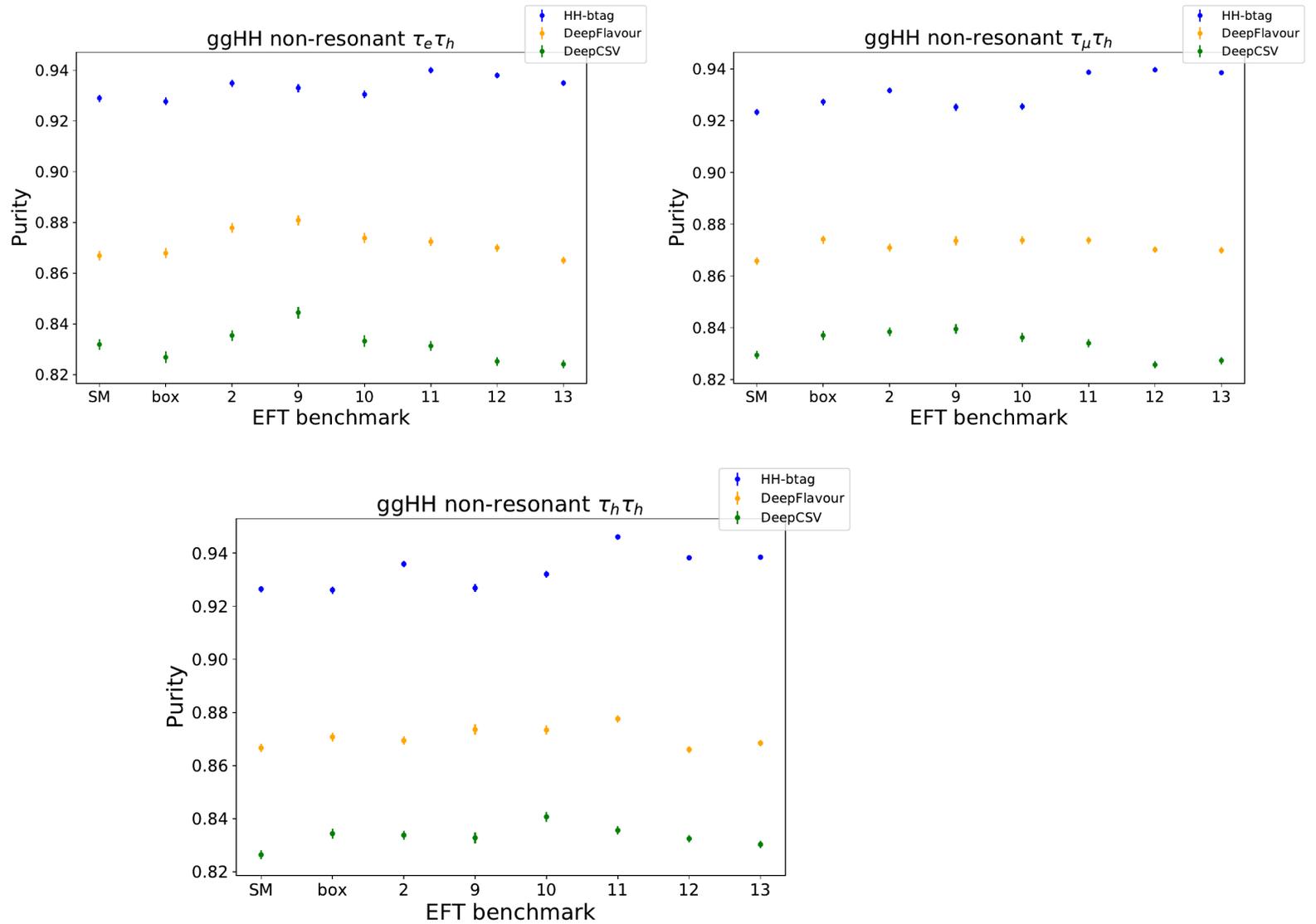


Figure 6.9: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF non-resonant processes, for 2016 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

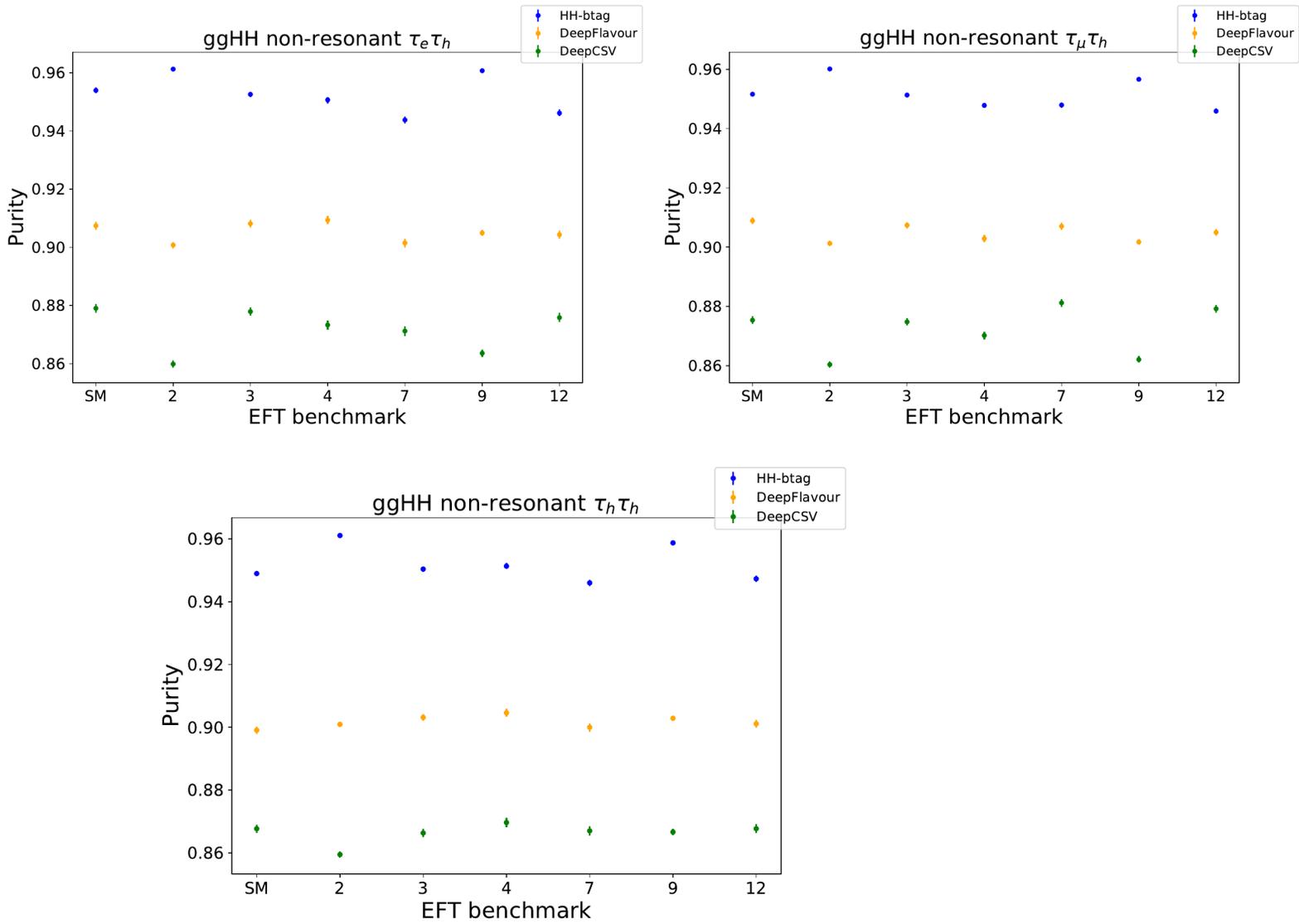


Figure 6.10: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF non-resonant processes, for 2017 for the  $\tau_e \tau_h$  (top left),  $\tau_\mu \tau_h$  (top right) and  $\tau_h \tau_h$  (bottom) channels.

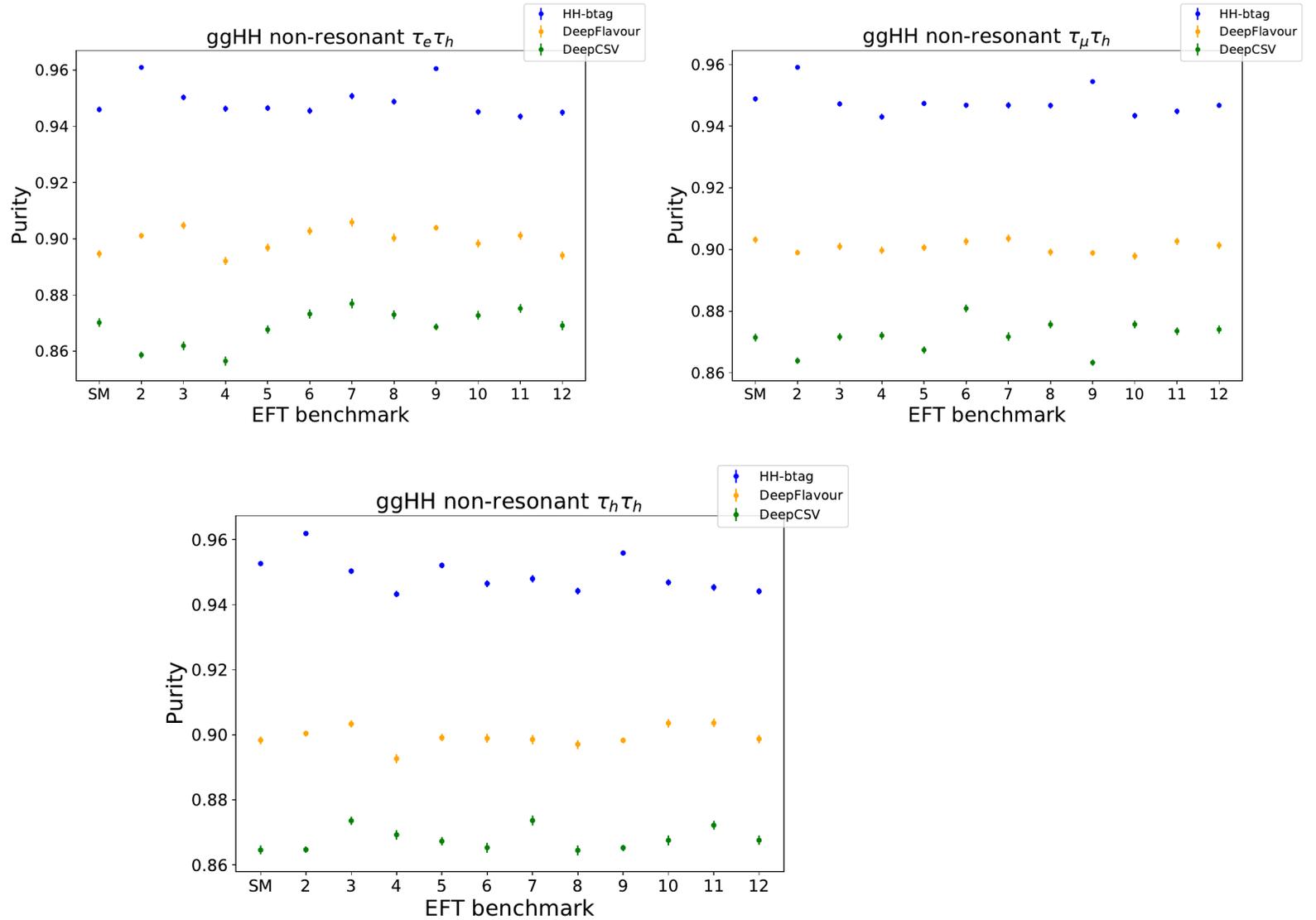


Figure 6.11: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF non-resonant processes, for 2018 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

### 6.7.2 VBF non-resonant

The MC samples used are listed in the Tables 6.2, 6.3, 6.4. These samples are produced with different combinations of the variations of the couplings ( $\lambda_V$ ,  $\lambda_{2V}$  and  $\lambda_{HHH}$ ) from the SM values, expressed for each of the coupling as: trilinear coupling with the vector bosons modifier,  $c_V = \lambda_V/\lambda_V^{SM}$ , quadrilinear coupling with the vector bosons modifier  $c_{2V} = \lambda_{2V}/\lambda_{2V}^{SM}$ , and trilinear self-coupling of the Higgs boson modifier,  $c_3 = \lambda_{HHH}/\lambda_{HHH}^{SM}$  (the latter is indicated for the GGF production as  $\kappa_\lambda$ ). The values of  $c_v$ ,  $c_{2V}$  and  $c_3$  are

written in a string in the dataset name. The string is used to represent the simulated corresponding modelling points in the abscissa of histograms. In these histograms is reported the purity of the three classifiers as a function of the VBF non-resonant process modelling point simulated, for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.12, 6.13 and 6.14, respectively for the data taking years 2016, 2017 and 2018.

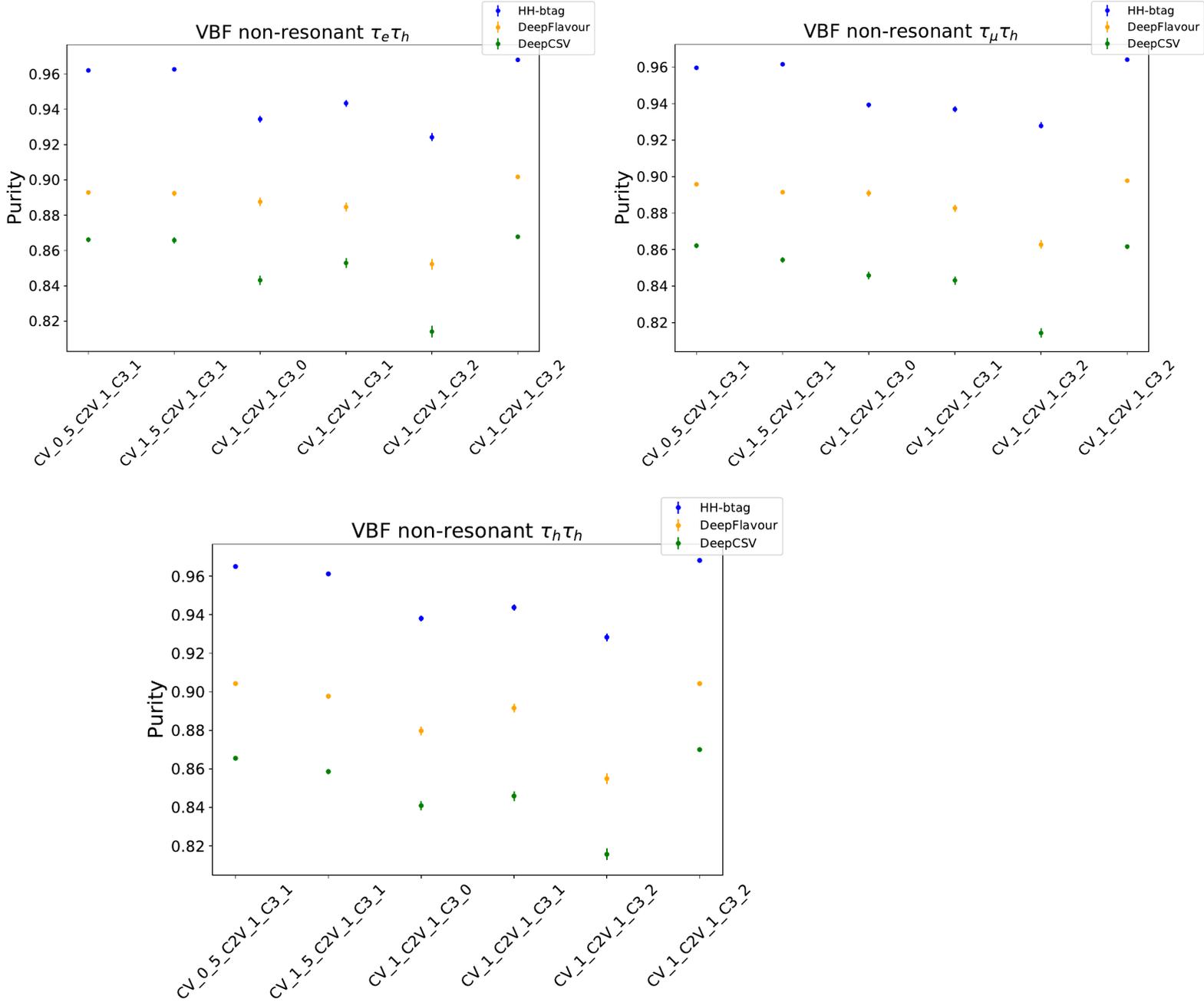


Figure 6.12: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF non-resonant processes, for 2016 for  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

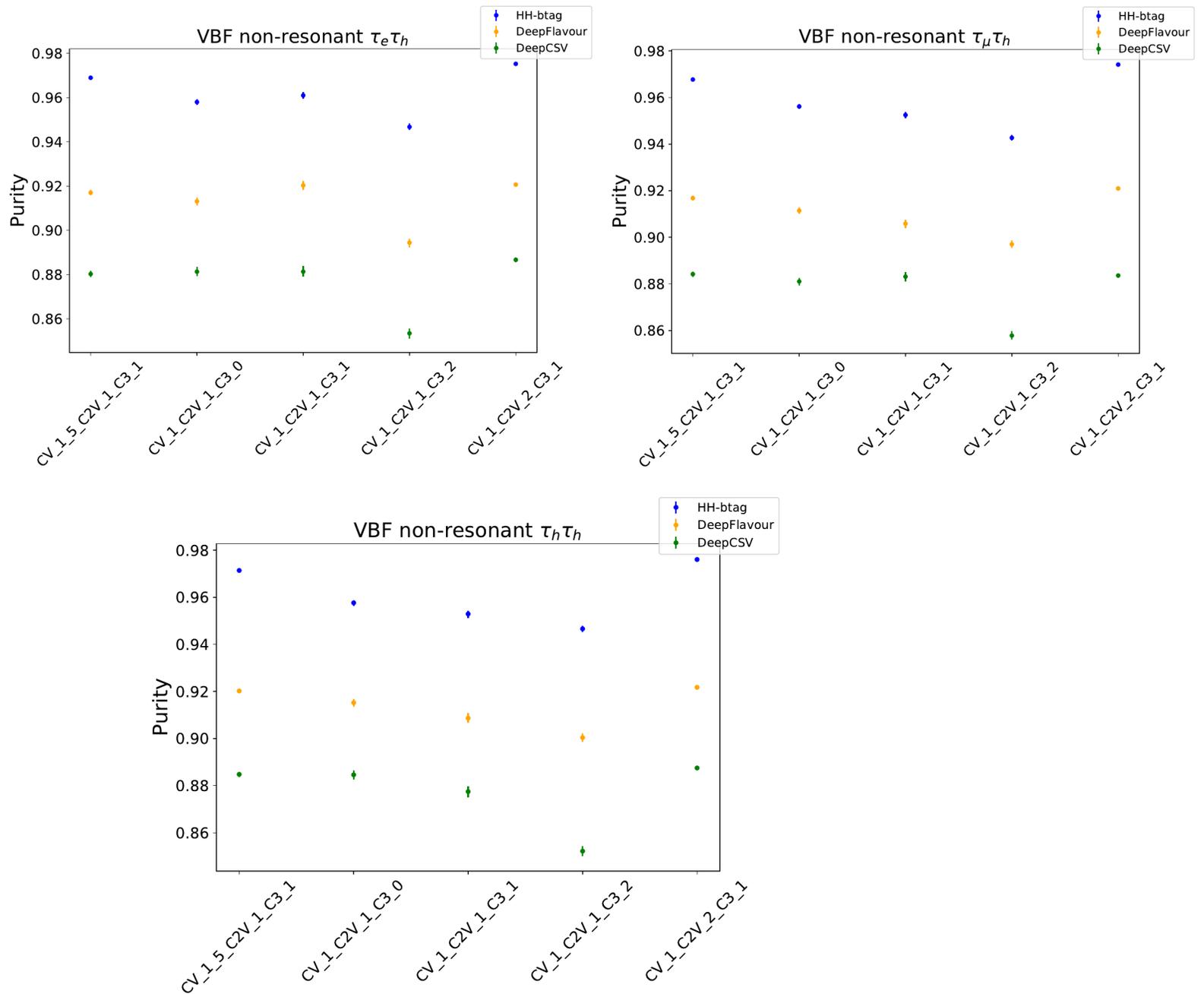


Figure 6.13: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF non-resonant processes, for 2017 for  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

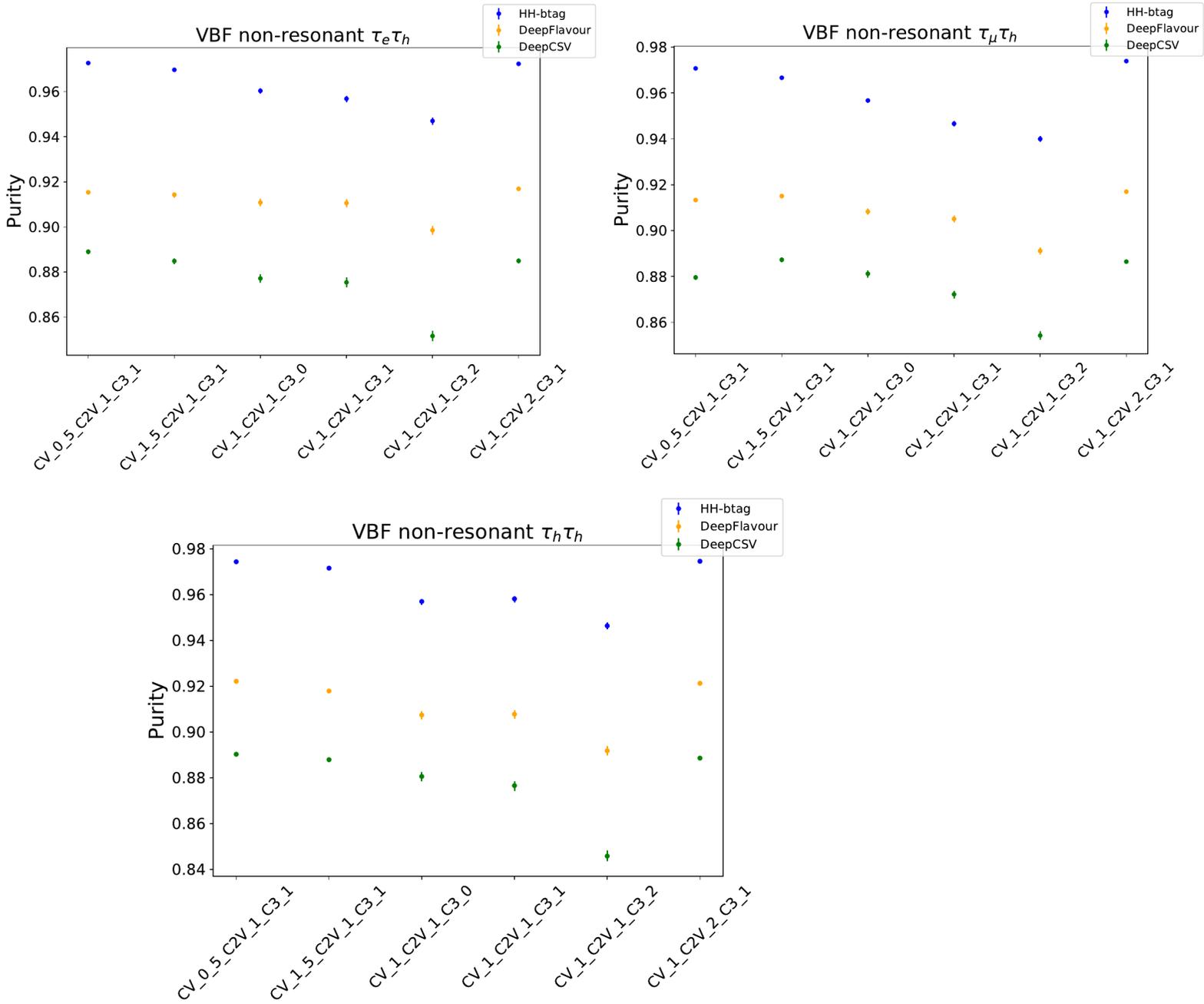


Figure 6.14: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF non-resonant processes, for 2018 for  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

### 6.7.3 Gluon fusion resonant

There are evaluated two scenarios in the resonant production, the case of a spin-0 and the case of a spin-1 resonance:

- spin-0 GGF: The purity of the three classifiers as a function of the mass of the spin-0 radion for the MC samples produced by the GGF mechanism listed in Tables 6.5, 6.8 and 6.10, is shown for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.15, 6.16 and 6.17, respectively for the data taking years 2016, 2017 and 2018.
- spin-2 GGF: The purity of the three classifiers as a function of the mass of the spin-2 graviton for the MC samples produced by the GGF mechanism listed in the Tables 6.6, 6.9 and 6.12, is shown for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.18, 6.19, and 6.20, respectively for the data taking years 2016, 2017 and 2018.

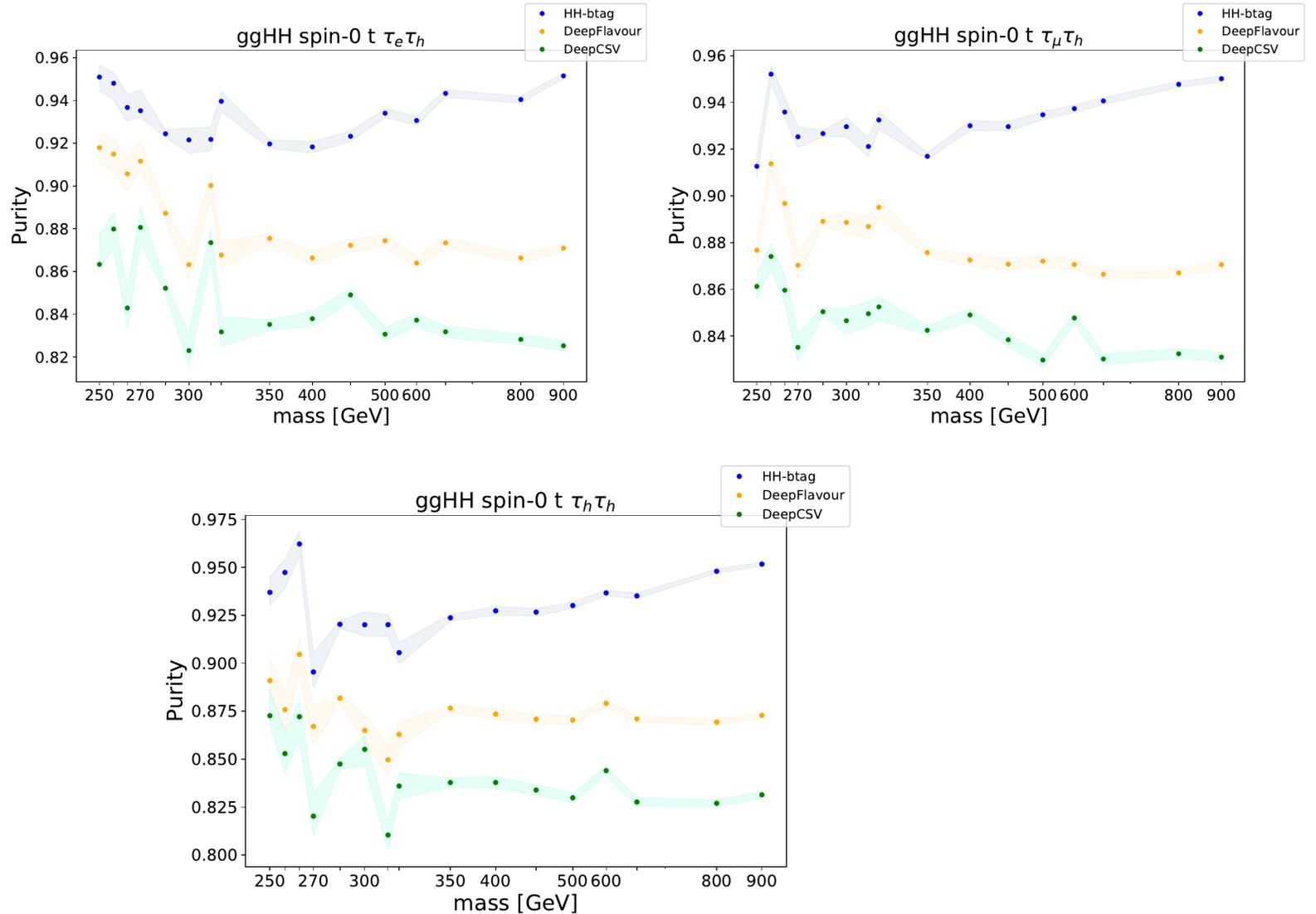


Figure 6.15: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-0 processes, for 2016 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

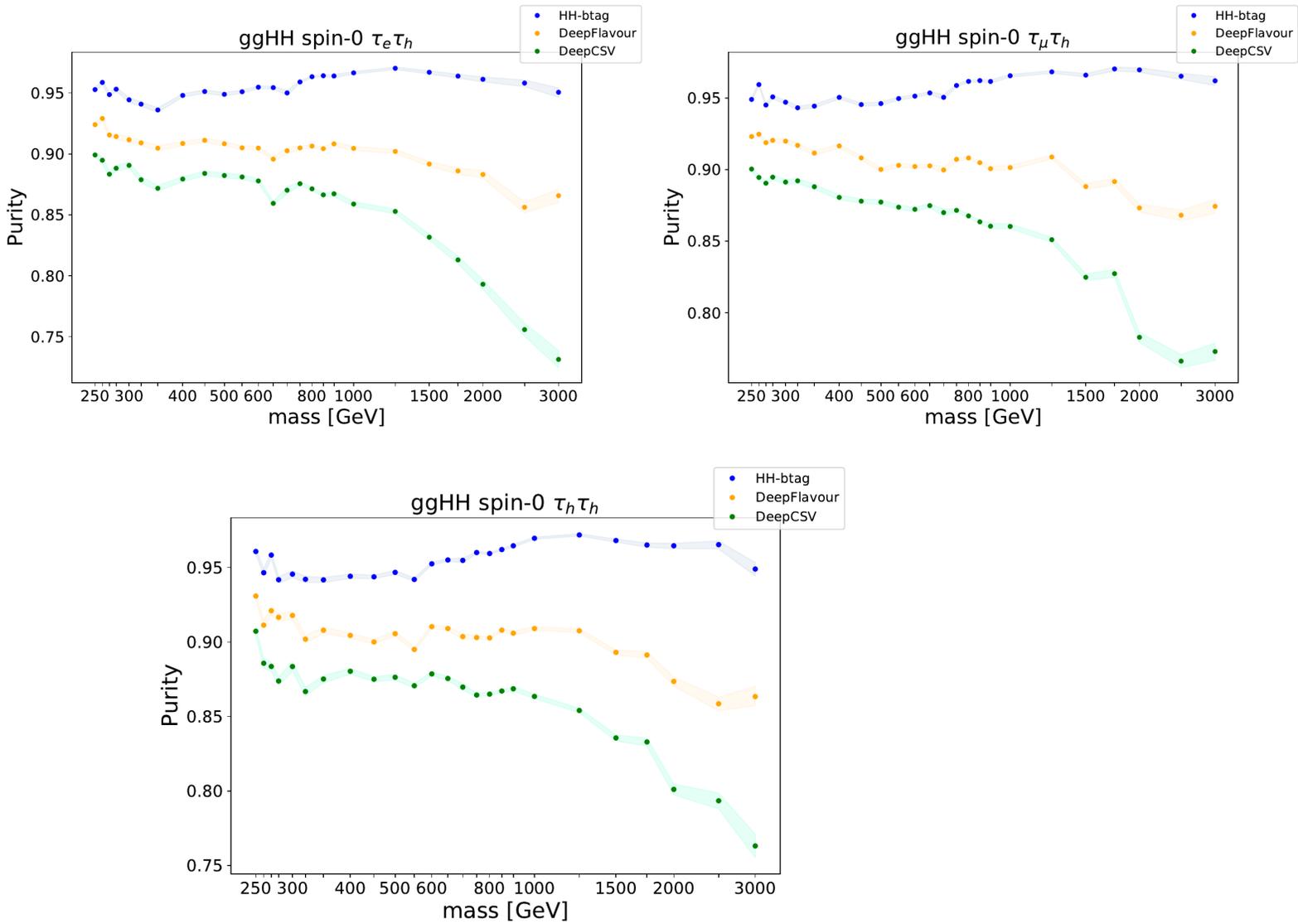


Figure 6.16: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-0 processes, for 2017 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels

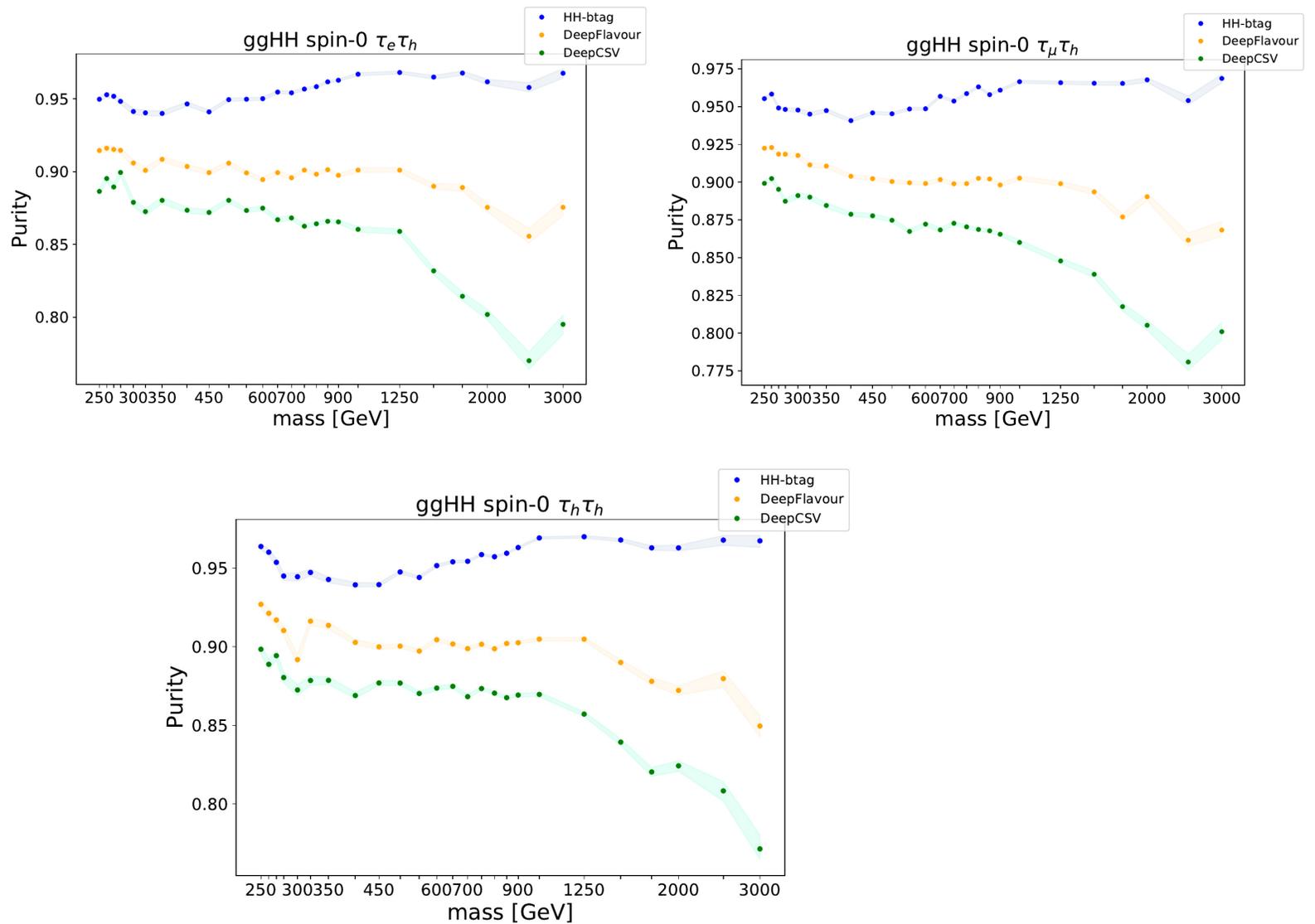


Figure 6.17: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-0 processes, for 2018 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

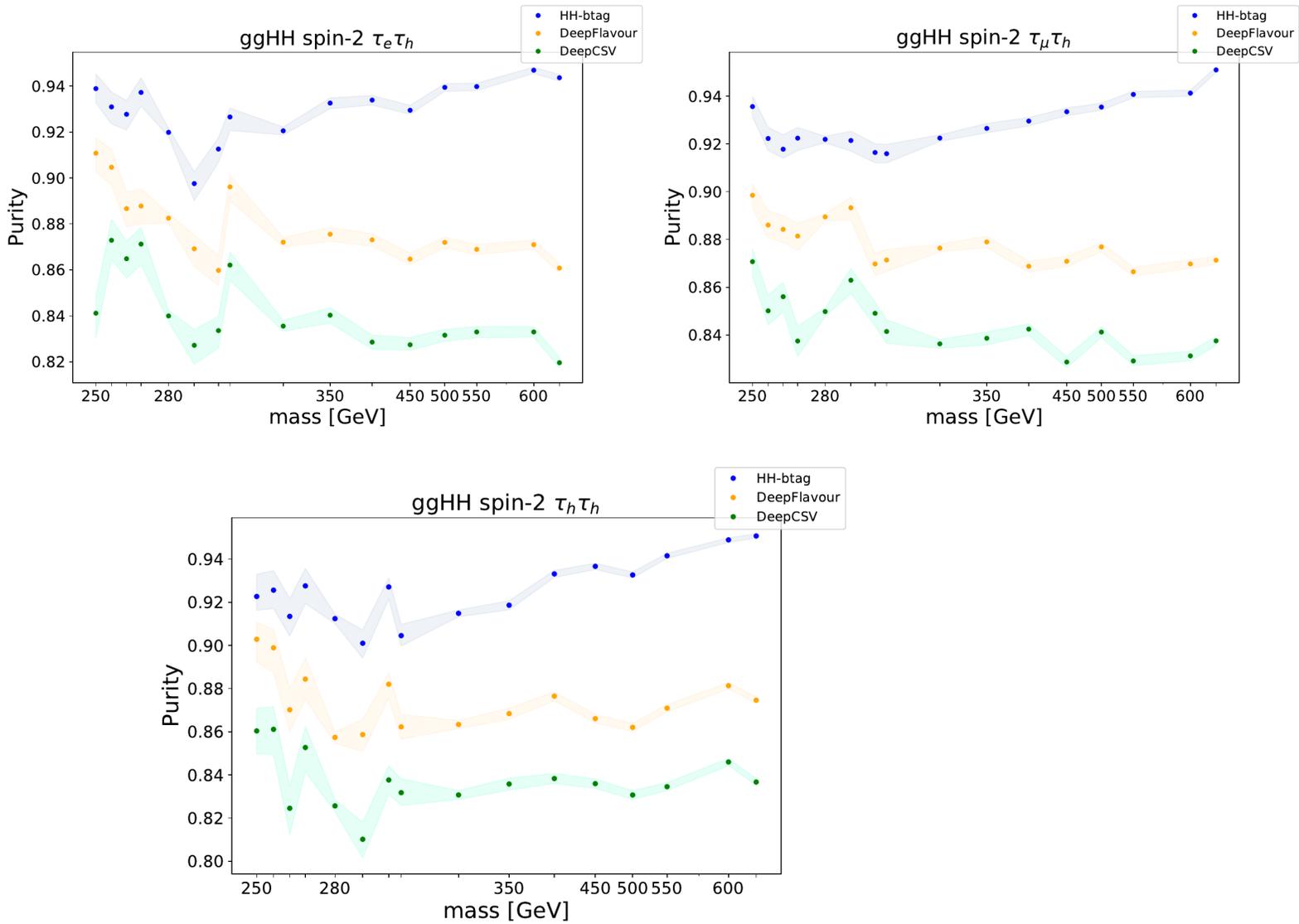


Figure 6.18: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-2 processes, for 2016 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels

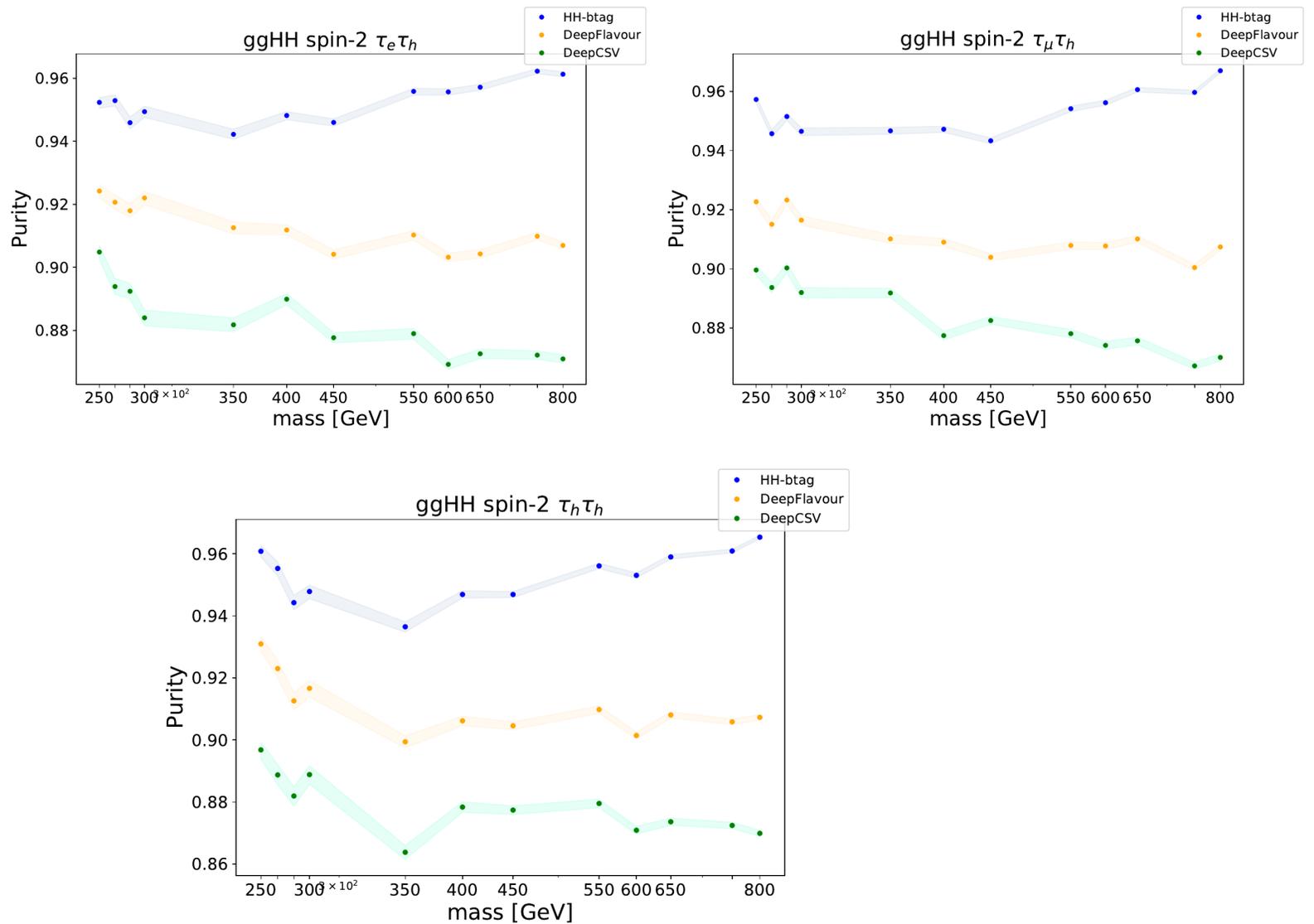


Figure 6.19: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-2 processes, for 2017 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

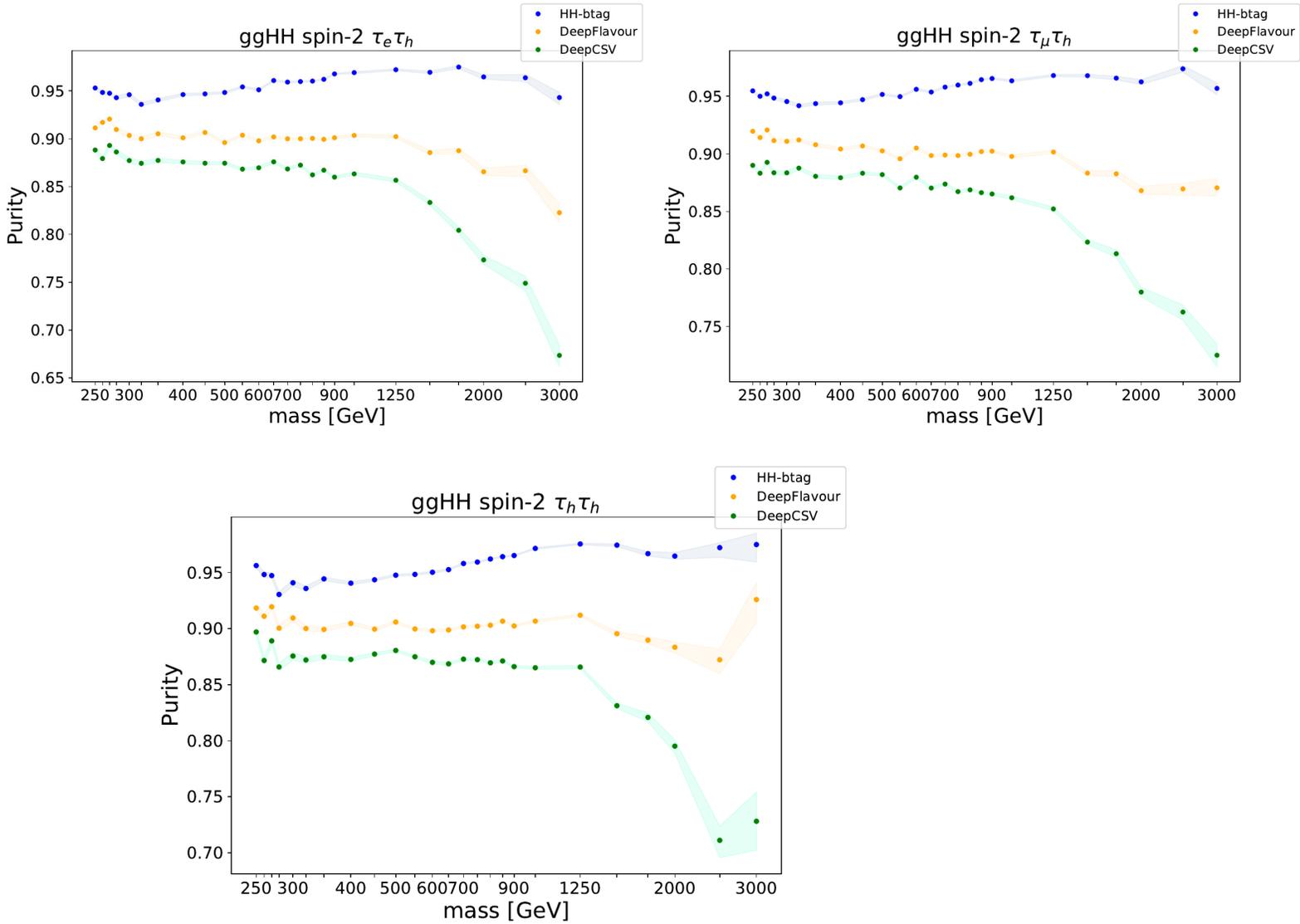


Figure 6.20: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the GGF resonant spin-2 processes, for 2018 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

#### 6.7.4 VBF resonant

There are evaluated two scenarios in the resonant production, the case of a spin-0 and the case of a spin-1 resonance:

- spin-0 VBF: The purity of the three classifiers as a function of the mass of the spin-0 radion for the MC samples produced by the VBF mechanism listed in Tables 6.7 and 6.11, is shown for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.21 and 6.22, respectively for the data taking years 2017 and 2018.

- spin-2 VBF: The purity of the three classifiers as a function of the mass of the spin-2 graviton for the MC samples produced by the VBF mechanism listed in Tables 6.9 and 6.12, is shown for the three channels  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  in Figures 6.23 and 6.24, respectively for the data taking years 2017 and 2018.

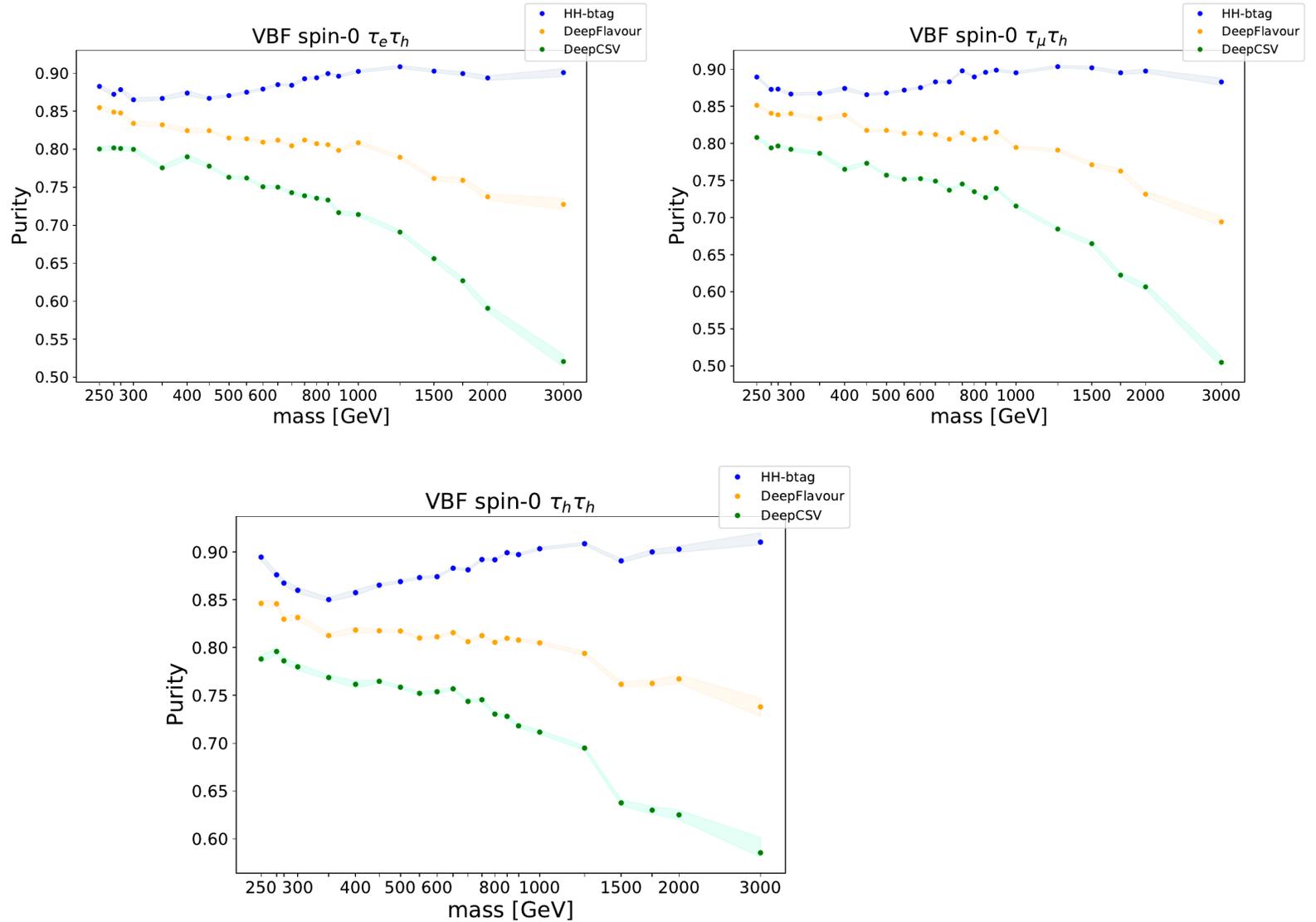


Figure 6.21: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF resonant spin-0 processes, for 2017 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

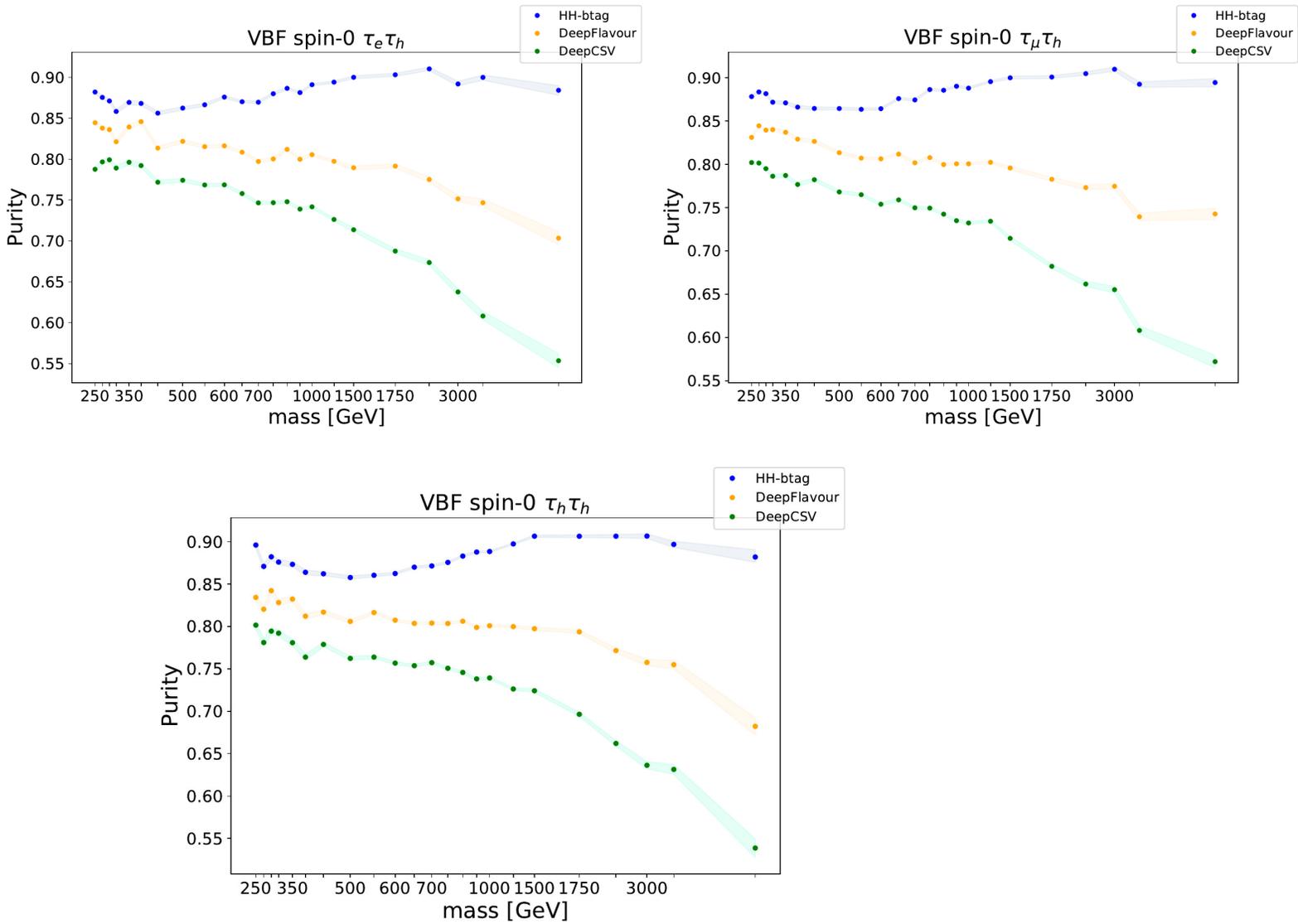


Figure 6.22: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF resonant spin-0 processes, for 2018 for the  $\tau_e \tau_h$  (top left),  $\tau_\mu \tau_h$  (top right) and  $\tau_h \tau_h$  (bottom) channels.

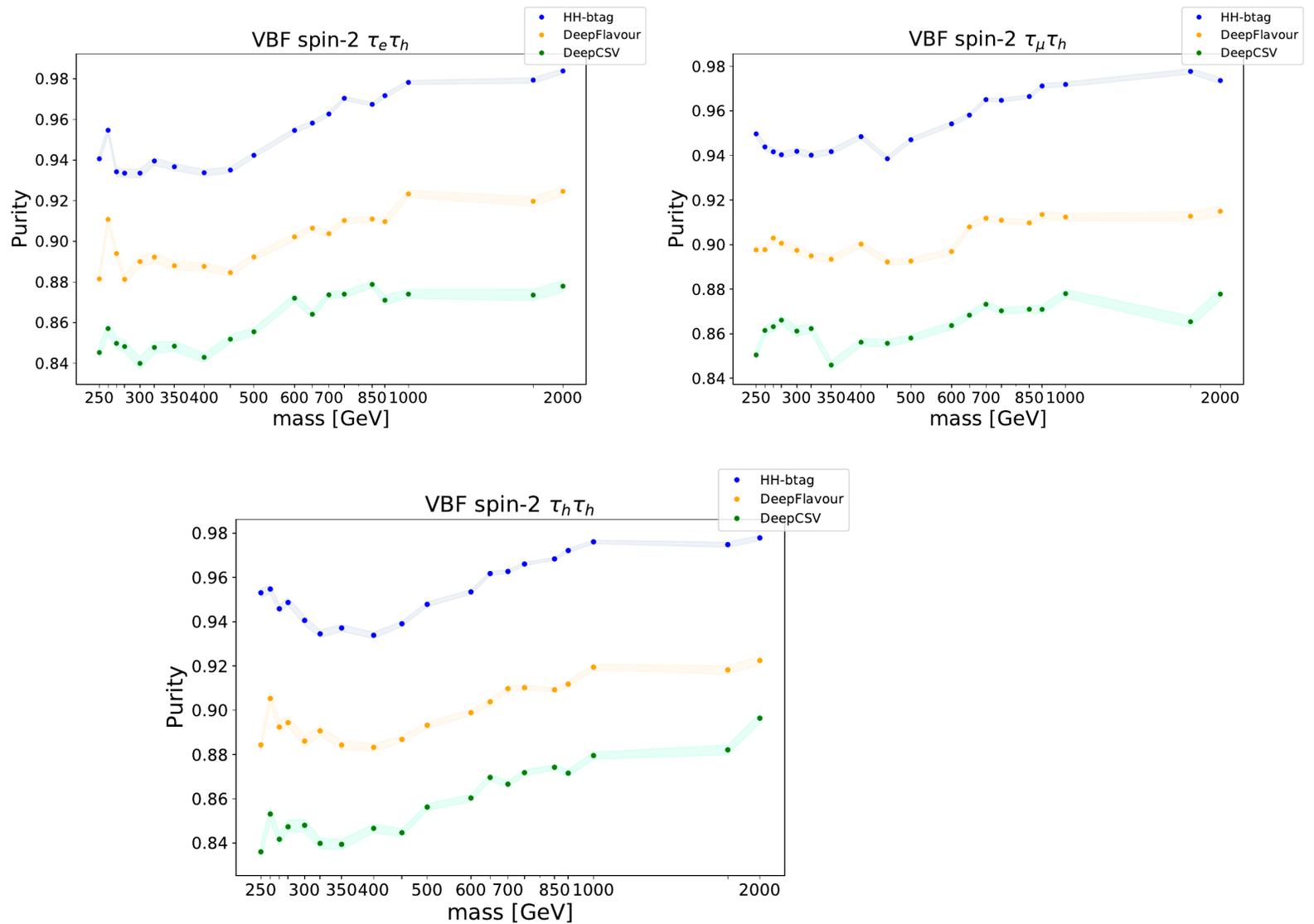


Figure 6.23: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF resonant spin-2 processes, for 2017 for the  $\tau_e\tau_h$  (top left),  $\tau_\mu\tau_h$  (top right) and  $\tau_h\tau_h$  (bottom) channels.

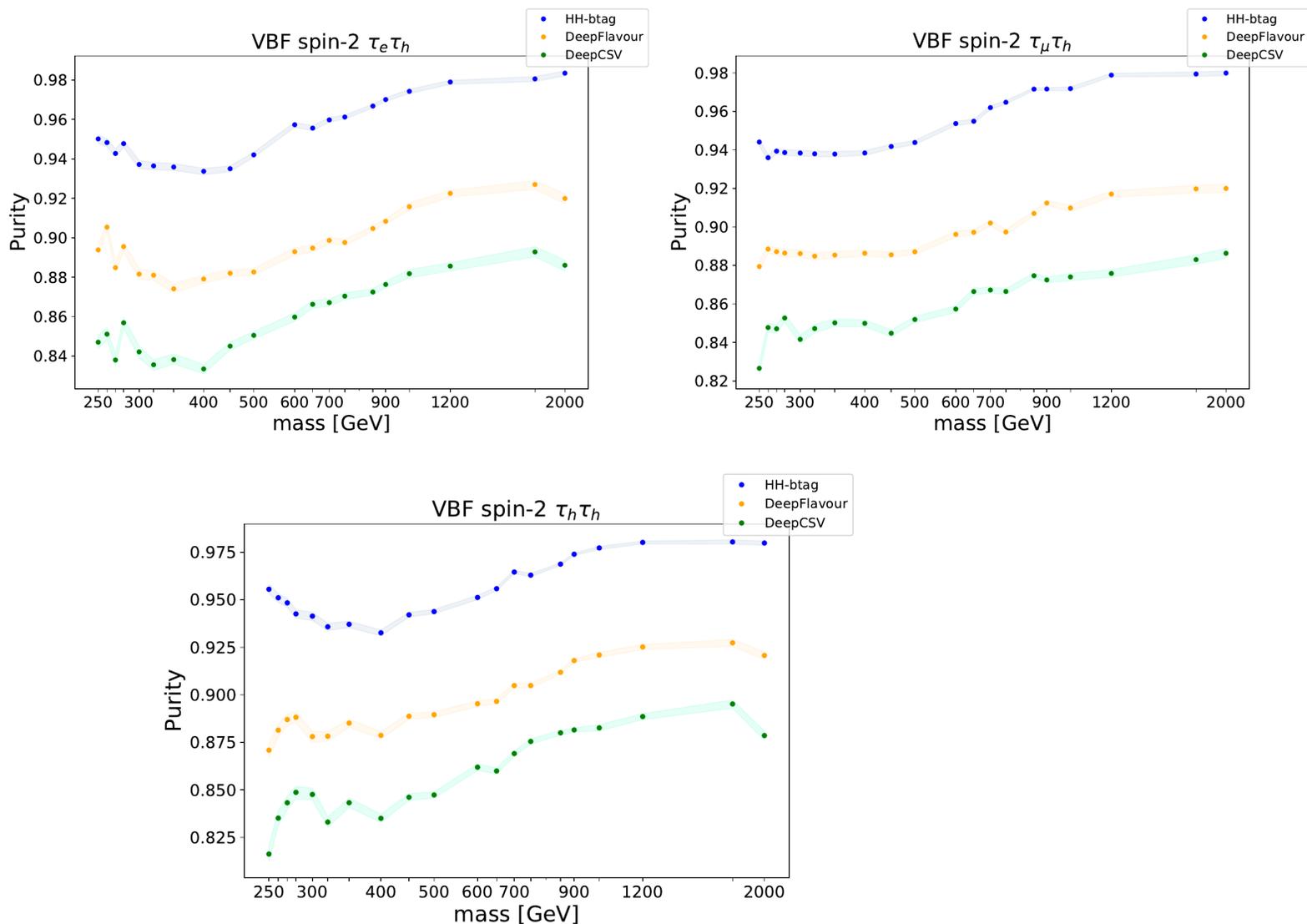


Figure 6.24: Performance of the HH-BTag, DeepFlavour and DeepCSV discriminators for the VBF resonant spin-2 processes, for 2018 for the  $\tau_e \tau_h$  (top left),  $\tau_\mu \tau_h$  (top right) and  $\tau_h \tau_h$  (bottom) channels.

### 6.7.5 Comparison with DeepFlavour algorithm

A comparison between the performances of the DeepFlavour and HH-BTag algorithms, for the selection of the b-jet pair, is performed for the  $\tau_h \tau_h$  channel in Tab. 6.16, in Fig. 6.25, for 2018 GGF and VBF SM signal samples, and for the simulated background processes DY and  $t\bar{t}$ . The events are required to satisfy the baseline analysis selection, described in Section 4.4. In addition, jet candidates are required to be classified as a hadron-based b-flavour jets, using the Jet Flavour Identification tool, provided by CMS, which allows

to identify the flavour of the reconstructed jets based on the MC truth information [167]. The events that satisfy these two requirements are denoted as pure events.

In Tab. 6.16, the percentage of events where both b-jets selected are pure, is shown for the DeepFlavour and HH-BTag taggers, together with the relative increase of the pure yield obtained with the HH-BTag tagger, with respect to the DeepFlavour tagger. When using the HH-BTag algorithm, an improvement of  $\sim 10\%$  in the relative pure yield in the SM GGF HH process, and  $\sim 6\%$  in the SM VBF HH process, is obtained. Furthermore, the relative pure yield of the  $t\bar{t}$  process decreases by 6.5%, and by 7.1% for the DY process; this arises from the capabilities of HH-BTag in discriminating the pair of b-jets from background, thanks to the learning process.

		GGF HH	VBF HH	DY	$t\bar{t}$
Pure events [%]	DeepFlavour	64.7	62.1	49.5	2.4
	HH-BTag	71.1	66.4	46.1	2.3
Relative increase pure yield [%]	HH-BTag / DeepFlavour	10.2	6.2	-7.1	-6.5

Table 6.16: Comparison between the HH-BTag and DeepFlavour tagger. Explanation of the definitions can be found in the text.

Fig. 6.25 shows the distribution of the invariant mass of the  $H \rightarrow bb$  candidate ( $m_{bb}$ ), for the GGF and VBF SM signal processes, by using the DeepFlavour (blue) or HH-BTag (red) taggers to select the b-jet pair. The HH-Btag selection improves the resolution of  $m_{bb}$ , reducing its width of (of  $\sim 25\%$  in the GGF HH SM process).

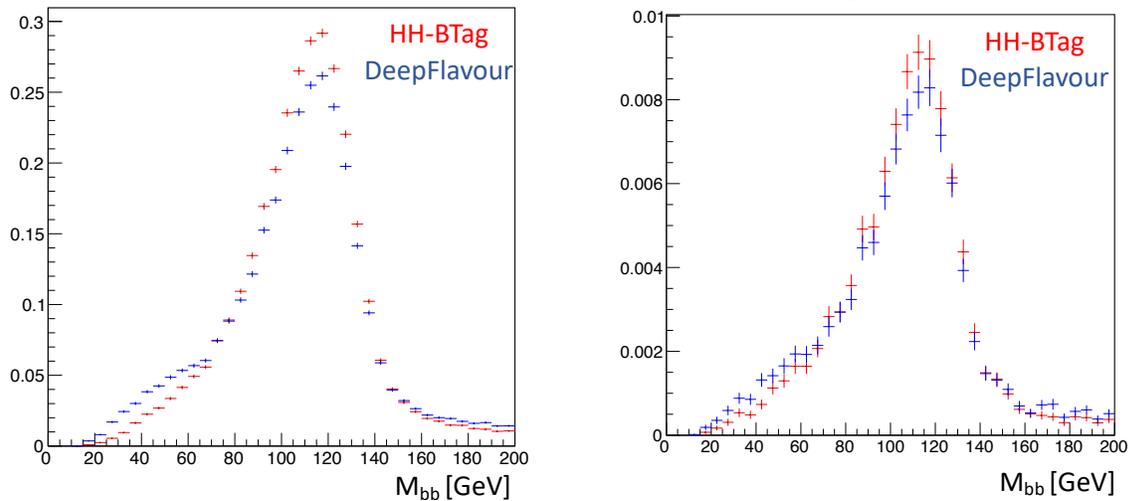


Figure 6.25: Distribution of the invariant mass of the  $H \rightarrow bb$  candidate ( $m_{bb}$ ), for the GGF (left) and VBF SM (right) signal processes.

---

## 6.8 HH-BTag Results

As shown in Figures 6.9 - 6.14, for the non-resonant production, in all the production mechanisms GGF and VBF, data taking years, and decay channels, the purity as a function of the modelling point goes from 85% to 98% for HH-BTag, while with the use of the DeepFlavour discriminator (the best performing before this study) goes from 80% to 90%.

For the resonant production, on Figures 6.15 - 6.24, the magnitude of the improvement by using HH-BTag depends strongly on the mass, spin of the resonance and mechanism of production, with the best performance achieved for the higher masses of a spin-0 resonance produced via VBF.

Furthermore, as shown in Tab. 6.16, the use of HH-BTag increases the percentage of pure events selected in signal samples, for the production mechanisms GGF and VBF, while this percentage is reduced for the background processes DY and  $t\bar{t}$ . In addition, the use of HH-BTag improves the resolution of the invariant mass of the  $H \rightarrow b\bar{b}$  candidate ( $m_{b\bar{b}}$ ).

In conclusion the overall performance has greatly improved in both, resonant and non-resonant production, when the two highest HH-BTag scored jet candidates are used in order to select the b-jet pair from the  $H_{b\bar{b}}$  decay.

---

# Conclusions

In this thesis, a search for Higgs boson pair production in the  $HH \rightarrow bb\tau\tau$  has been presented, in the non-resonant hypothesis for the  $\tau_e\tau_h$ ,  $\tau_\mu\tau_h$  and  $\tau_h\tau_h$  final states. The search uses the full Run II data from p-p collisions at a center-of-mass energy of 13 TeV recorded with the CMS detector at the LHC, corresponding to an integrated luminosity of  $137.1\text{ fb}^{-1}$ , participating to the Run II legacy analysis. The analysis presented was improved with respect the previous ones, thanks to a series of improvements, with many of them specifically developed within the analysis group. Among these, a new trigger path, operating from 2017, designed to collect VBF  $H \rightarrow \tau\tau$ , the development of several neural networks to identify the b-jets from the Higgs boson decay (HH-Btag), categorize the events and perform signal extraction. At the same time, this analysis builds up on CMS-wide innovations, such as the new DeepTau and DeepFlavour algorithms.

The internal CMS collaboration review of the analysis started in December 2020 and is still ongoing, thus the statistical analysis of the results has been carried out without the inclusion of the observed data (blind analysis), and only expected results from standard model predictions have been reported. However, waiting for the approval, the results presented for the SM predictions, which include for the first time the VBF mechanism, outperform the previously CMS published ones for  $HH \rightarrow bb\tau\tau$ , as can be seen from the results presented in Section 4.8 and Tab. 4.17. In addition, these results are the best if compared with other CMS HH searches.

The tools and techniques developed for the work presented in this thesis will provide a solid foundation for the search of resonant  $HH \rightarrow bb\tau\tau$ , where the HH-BTag shown the best performance (see Section 6.7). The resonant search in this channel is planned as one of the CMS legacy analyses, with the same team of the non-resonant analysis involved.

The high luminosity phase of the LHC (HL-LHC), collecting a dataset of about  $3000\text{ fb}^{-1}$  in a decade of p-p collisions at  $\sqrt{s} = 14\text{ TeV}$ , will results in an unprecedented opportunity to study very rare phenomena, where HH production represents a top priority of LHC research programme. Among several decay channels, the  $HH \rightarrow bb\tau\tau$  is expected to have a key role for the interesting, stimulating, and rewarding challenge of the measurement of  $\sigma_{HH}$  both in the context of the SM and of physics beyond it.

---

# Appendix A

## MC Corrections

In order to take into account wrongly simulated effects and to cope with the efficiencies of the working points of the reconstruction and identification algorithms used in the analysis, the following corrections are applied to the simulated events. Systematics uncertainties associated to these corrections are described in Section 4.6. The following corrections are based on Ref. [110].

### A.1 Pile-up reweighting

The distribution of the number of interaction vertices in the Monte Carlo events (primary vertex selection described in Section 3.1.1) do not exactly matches the one of the data. To achieve a better match between the data and simulations, the standard re-weighting procedure recommended for Run 2 [25] is applied to the MC events. The Minimum Bias cross-section equals to 69.2 mb with an uncertainty of 4.6%, and it is used as an input to produce a pile-up histogram used for the re-weighting of MC events. The re-weighting is applied event by event.

### A.2 Pile-up jet identification scale factor

Since all jets with  $p_T < 50$  GeV that do not pass the loose working point of the pile-up jet discriminator are rejected, the scale factors described in the pile-up jet identification twiki are applied as recommended [168].

### A.3 Jet energy smearing

Since measurements show that the jet energy resolution in data is worse than in the simulation, the jets in MC need to be smeared to describe the data. The smearing

---

procedure applied is the so called “hybrid” method suggested by the JetMET POG [169]: when a matching particle-level jet is found, the scaling method is used; otherwise the stochastic smearing is applied.

## A.4 L1 ECAL prefiring weight

For 2016 and 2017 MC simulation a reweighting procedure is applied as suggested on the L1ECALPrefiring twiki [170].

## A.5 Trigger scale factors

To account for differences in the data and MC efficiencies of these trigger paths, a correction factor or “scale factor” (SF) for the trigger efficiency is applied on MC events. The trigger efficiencies used are provided by the Tau POG in the context of the Standard Model  $H \rightarrow \tau\tau$  analysis [121]. The SFs depend on the  $\eta$  and  $p_T$  of the object.

For the  $\tau_h\tau_h$  final state, the trigger efficiencies and SFs are measured using  $Z \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$  events selected with a tag and probe technique and cover the logical OR of the three trigger paths used.

For the  $\tau_\mu\tau_h$  and  $\tau_e\tau_h$  final states, the SFs must take into account the efficiency of the logical OR between single and cross triggers: assuming the efficiencies of the two legs to be independent, the efficiency of the logic OR can be factorized and easily computed from the single objects efficiencies. The resulting event by event SF formula is:

$$SF = \frac{Eff_{DATA}}{Eff_{MC}}, \quad (\text{A.1})$$

where the efficiency for both MC and data is:

$$Eff = \varepsilon_L(1 - \varepsilon_\tau) + \varepsilon_l\varepsilon_\tau, \quad (\text{A.2})$$

where  $\varepsilon_L$  is the single lepton trigger efficiency;  $\varepsilon_l$  is the cross lepton trigger efficiency for the  $\tau_e$  or  $\tau_\mu$  leg; and  $\varepsilon_\tau$  is the cross lepton trigger efficiency for the  $\tau_h$  leg.

The VBF trigger paths are included for the  $\tau_h\tau_h$  final state in 2017 and 2018. Efficiencies and SFs are computed separately for the tau legs and for the jet legs. The tau legs measurements are provided by the HTT team [171], while the jet legs measurement was performed by the *bbautau* team [172] and both were approved by the Tau POG.

Efficiencies and SF are applied as a function of the  $p_T$  and decay mode of the two hadronic tau leptons; while for the jets, they are applied based on the transverse momenta of the two jets with highest invariant mass, and on their invariant mass.

## A.6 Trigger prescale weight

In order to take into account the trigger prescales used during data taking for some of the HLT paths, the events that fired only prescaled paths are corrected with weights computed as the fraction of integrated luminosity covered by the prescaled trigger over the total integrated luminosity of the year.

## A.7 b-tag scale factors

To account for discrepancies in the b tag performance in MC the whole b-tagging discriminant distribution in MC is corrected to match that in data following the procedure recommended by the b Tag and Vertexing (BTV) POG in [173].

For each MC event with a given jet configuration, the event weight  $\omega$  is computed as:

$$\omega = \prod_i^{N_{jets}} SF(D_i, p_{Ti}, \eta_i), \quad (\text{A.3})$$

where the scale factors SF are provided by the BTV POG as a function of the discriminator score, transverse momentum and pseudorapidity of the jets.

The event weights computed with this method are supposed to change only the shape of the b-tagging discriminant. Before applying any b-tag selection criteria, expected event yields should be preserved: this means that the number of events (i.e. the sum of event weights) before and after applying b-tag weights should be identical. In order to ensure this, the sum of event weights before and after applying b-tag event weights, without requiring any b-tag selection, is computed. The ratio  $r = \sum \omega_{\text{before}} / \sum \omega_{\text{after}}$  represents a phase space extrapolation and is multiplied to the b-tag event weight. The values of these  $r$  factors are reported in Table A.1.

Year	Channel	$r$ factor
2016	$\tau_h\tau_h$	1.0103
	$\tau_\mu\tau_h$	1.0081
	$\tau_e\tau_h$	1.0068
2017	$\tau_h\tau_h$	0.9547
	$\tau_\mu\tau_h$	0.9993
	$\tau_e\tau_h$	0.9949
2018	$\tau_h\tau_h$	0.9795
	$\tau_\mu\tau_h$	1.0039
	$\tau_e\tau_h$	1.0040

Table A.1: Values of the  $r$  factors used to correct the b-tag event weights and preserve the normalization of the samples.

## A.8 DeepTau scale factors for hadronic taus

Different working points of the DeepTau algorithm (3.2.4) are employed in the selection of the  $\tau\tau$  pair, as described in detail in Section 4.4.1. The scale factors related to the hadronically decaying taus for the DeepTau algorithm are applied following the recommendations of the TauPOG [120]:

- For genuine taus, the scale factors are provided binned as function of the transverse momentum of the tau in the range from 20 GeV to 1 TeV.
- For genuine electrons misidentified as taus, the scale factors are provided split into barrel and endcap.
- For genuine muons misidentified as taus, the scale factors are provided binned as function of  $\eta$ .

For the leptonic decays of the tau candidates (either  $\tau \rightarrow \mu$  or  $\tau \rightarrow e$ ) the same scale factor used by the HTT analysis are used here.

## A.9 Electron and muon scale factors

In order to correct possible disagreements between data and MC regarding the reconstruction and identification of electrons and muons in the  $\tau_e\tau_h$  and  $\tau_\mu\tau_h$  channels, specific scale factors are applied to events in these channels. These correction factors are provided, binned as function of the  $p_T$  and  $\eta$  of the leptons, by the HTT working group and are available in the LeptonEfficiencies github repository [174].

## A.10 Tau Energy Scale corrections

The tau energy scale (TES) corrections are provided by the TauPOG binned in decay modes (0, 1, 10 and 11) for genuine taus, and split in endcap and barrel for electrons misidentified as taus (decay modes 0 and 1) [120]. No energy scale correction is needed for muons misidentified as taus.

## A.11 Custom SF for the fully hadronic channel in 2017

When all the recommended corrections are applied, a satisfactory agreement between data and Monte Carlo simulation can be observed in the  $\tau_\mu\tau_h$  and  $\tau_e\tau_h$  channels for 2017. Instead, in the  $\tau_h\tau_h$  channel, a disagreement at the level of 20% still remains (see Figure A.1) especially where the Drell-Yan background contribution is higher.

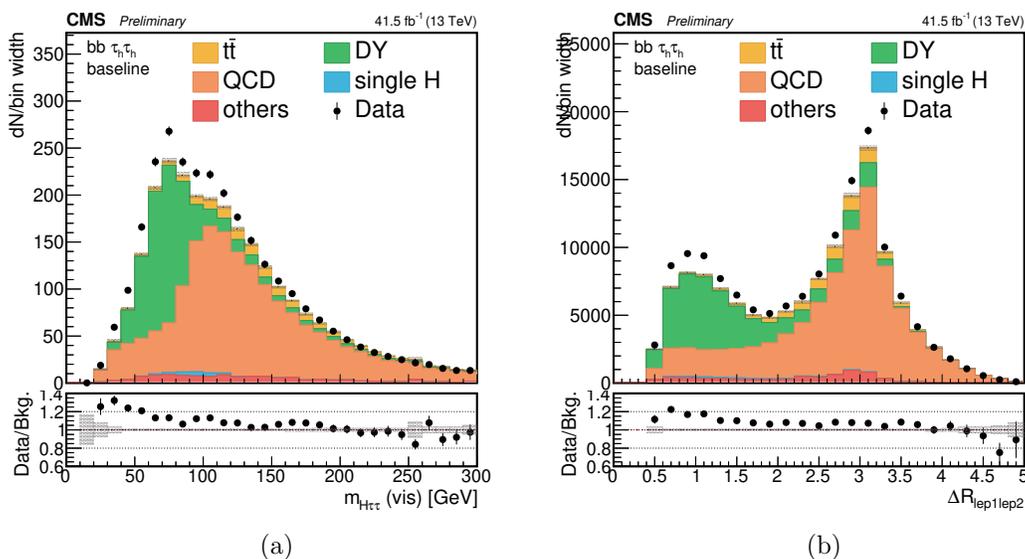


Figure A.1: Distributions of the invariant mass of the visible decay products of the  $\tau\tau$  pair (left) and of the spatial separation  $\Delta R$  between the two tau candidates (right). Events are shown for the  $\tau_h\tau_h$  channel in 2017. A clear disagreement between data and MC simulation is visible in the low mass region where the main background contribution is the Drell-Yan process. The shaded band in the plots only shows the statistical uncertainty.

A region enriched with Drell-Yan events is defined by applying the regular  $\tau_h\tau_h$  channels selections with an additional cut on the angular separation between the two taus:  $\Delta R(\tau_h, \tau_h) < 2$ . In order to gain more statistics in this sideband region the request of

having at least two good b-jet candidates ( $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.4$ ) is dropped. This region is relatively pure in genuine hadronically decaying taus and is used to derive alternative scale factors. For these events all other recommended corrections are applied.

Four different categories are defined by requiring that both the tau legs have the same decay mode (DM): a correction  $SF_{DM}$  is extrapolated within each of these categories, i.e. for each of the decay modes of the tau considered in this analysis. The simulated events are split in events with one, two and zero genuine hadronic tau leptons; the global event yield of simulated events in these categories should match the number of data events, through the variation of the multiplicative factors  $SF_{DM}^2$ , for events with two real tau leptons, and  $SF_{DM}$  for events with one real tau lepton. As the QCD estimation is affected by the changes in the event yield of the simulated backgrounds in the sidebands, the measurement is performed through a simultaneous fit in the four ABCD regions, defined in Section 4.5.2.1.

The values of the scale factors thus obtained are listed in Table A.2, and the corresponding uncertainties are obtained directly from the fit described in the previous paragraph. Their effect, once applied, is visible in Figure A.2.

Decay Mode	SF_DM	Error
0	1.078	-0.036/ + 0.034
1	1.112	-0.023/ + 0.023
10	0.984	-0.067/ + 0.063
11	0.759	-0.259/ + 0.178

Table A.2: Custom scale factors derived for the  $\tau_h\tau_h$  channel in 2017 as a function of the decay mode of the  $\tau_h$  candidates.

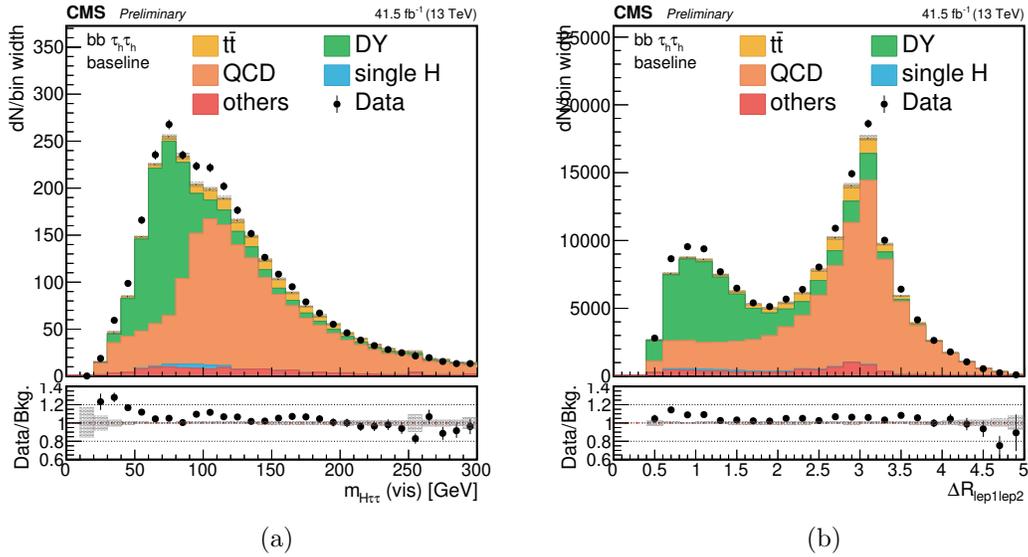


Figure A.2: Distributions of the invariant mass of the visible decay products of the  $\tau\tau$  pair (left) and of the spatial separation  $\Delta R$  between the two tau candidates (right). Events are shown for the  $\tau_h\tau_h$  channel in 2017. The shaded band in the plots only shows the statistical uncertainty. The custom scale factors are applied on top of all other recommended corrections: the disagreement between data and MC simulation is drastically reduced and only remains in a very low mass region which will be excluded from the analysis by the invariant mass selections defined in Section 4.4.3.

# Bibliography

- [1] The ATLAS Collaboration, “Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc”, *Physics Letters B* (2012) doi:10.1016/j.physletb.2012.08.020.
  - [2] The CMS Collaboration, “Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc”, *Physics Letters B* (2012) doi:10.1016/j.physletb.2012.08.021.
  - [3] The ATLAS Collaboration, “Combination of searches for Higgs boson pairs in  $pp$  collisions at 13 TeV with the ATLAS experiment.”, Technical Report ATLAS-CONF-2018-043, CERN, Geneva, Sep, 2018.
  - [4] The CMS Collaboration, “Combination of searches for higgs boson pair production in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, doi:10.1103/PhysRevLett.122.121803.
  - [5] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”. Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.
  - [6] M. D. Schwartz, “Quantum Field Theory and the Standard Model”. Cambridge University Press, 3, 2014. ISBN 978-1-107-03473-0, 978-1-107-03473-0.
  - [7] F. Halzen and A. D. Martin, “Quarks and leptons: An Introductory course in modern particle physics”. 1984. ISBN 978-0-471-88741-6.
  - [8] Particle Data Group Collaboration, “Review of particle physics”, *Phys. Rev. D* **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.
  - [9] J. W. e. Guido Altarelli (auth.), “Collider Physics within the Standard Model: A Primer”. Lecture Notes in Physics 937. Springer International Publishing, 1 edition, 2017. ISBN 978-3-319-51919-7, 978-3-319-51920-3.
  - [10] A. D. Giacomo, “Understanding color confinement”, doi:10.1051/epjconf/20147000019.
-

- 
- [11] A. Bettini, “Introduction to Elementary Particle Physics”. Cambridge University Press, 1 edition, 2008. ISBN 0521880211,9780521880213.
- [12] J. Barranco, “Some standard model problems and possible solutions”, *J. Phys. Conf. Ser.* **761 012007** (2016) doi:0.1088/1742-6596/761/1/012007.
- [13] C. Giganti, S. Lavignac, and M. Zito, “Neutrino oscillations: The rise of the pmns paradigm”, *Progress in Particle and Nuclear Physics* **98** (Jan, 2018) 1–54, doi:10.1016/j.pnpnp.2017.10.001.
- [14] A. Pich, “The Standard Model of Electroweak Interactions”, in *2010 European School of High Energy Physics*, pp. 1–50. 1, 2012. arXiv:1201.0537.
- [15] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc. C* **680519** (1968) 367–377, doi:10.1142/9789812795915\_0034.
- [16] S. L. Glashow, “Partial-symmetries of weak interactions”, *Nuclear Physics* **22** (1961), no. 4, 579–588, doi:https://doi.org/10.1016/0029-5582(61)90469-2.
- [17] M. Goldhaber, L. Grodzins, and A. W. Sunyar, “Helicity of neutrinos”, *Phys. Rev.* **109** (Feb, 1958) 1015–1017, doi:10.1103/PhysRev.109.1015.
- [18] Particle Data Group Collaboration, “Review of particle physics”, *Chinese Phys* **38** (2014) doi:10.1088/1674-1137/38/9/090001.
- [19] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [20] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- [22] S. F. Novaes, “Standard model: An introduction,”, arXiv:hep-ph/0001283.
- [23] J. Ellis, “Higgs Physics”, doi:10.5170/CERN-2015-004.117.
- [24] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961) 154–164, doi:10.1007/BF02812722.
-

- [25] LHC Higgs Cross Section Working Group Collaboration S. Heinemeyer, et al., “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group”. CERN Yellow Reports: Monographs. CERN, Geneva, Jul, 2013. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>. doi:10.5170/CERN-2013-004.
- [26] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, doi:10.23731/CYRM-2017-002, arXiv:1610.07922.
- [27] The CMS Collaboration, “Observation of higgs boson decay to bottom quarks”, *Phys. Rev. Lett.* **121** (Sep, 2018) 121801, doi:10.1103/PhysRevLett.121.121801.
- [28] The CMS Collaboration, “Observation of the higgs boson decay to a pair of tau leptons with the cms detector”, *Physics Letters B* **779** (2018) 283 – 316, doi:<https://doi.org/10.1016/j.physletb.2018.02.004>.
- [29] G. Aad et al., “Study of the spin and parity of the higgs boson in diboson decays with the atlas detector”, *The European Physical Journal C* **75** (Oct, 2015) doi:10.1140/epjc/s10052-015-3685-1.
- [30] S. Chatrchyan et al., “Study of the mass and spin-parity of the higgs boson candidate via its decays to z boson pairs”, *Physical Review Letters* **110** (Feb, 2013) doi:10.1103/physrevlett.110.081803.
- [31] A. M. Sirunyan et al., “Combined measurements of higgs boson couplings in proton–proton collisions at  $\sqrt{s} = 13$  TeV”, *The European Physical Journal C* **79** (May, 2019) doi:10.1140/epjc/s10052-019-6909-y.
- [32] G. Aad et al., “Measurements of the higgs boson production and decay rates and constraints on its couplings from a combined atlas and cms analysis of the lhc pp collision data at  $\sqrt{s} = 7$  and 8 tev”, *Journal of High Energy Physics* **2016** (Aug, 2016) doi:10.1007/jhep08(2016)045.
- [33] The CMS Collaboration, “Combined measurements of higgs boson couplings in proton-proton collisions at  $\sqrt{s} = 13$  tev”, *Eur. Phys. J. C.* **77** (2019) 421, doi:10.1140/epjc/s10052-019-6909-y.
- [34] S. Dawson, C. Englert, and T. Plehn, “Higgs physics: It ain’t over till it is over”, *Physics Reports* **816** (2019) 1–85,
-

- doi:<https://doi.org/10.1016/j.physrep.2019.05.001>. Higgs physics: It ain't over till it is over.
- [35] F. Maltoni, E. Vryonidou, and M. Zaro, “Top-quark mass effects in double and triple higgs production in gluon-gluon fusion at nlo”, *Journal of High Energy Physics* **2014** (Nov, 2014) doi:[10.1007/jhep11\(2014\)079](https://doi.org/10.1007/jhep11(2014)079).
- [36] B. D. Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, “Higgs boson potential at colliders: Status and perspectives”, *Reviews in Physics* **5** (Nov, 2020) 100045, doi:[10.1016/j.revip.2020.100045](https://doi.org/10.1016/j.revip.2020.100045).
- [37] A. R. Vieira, B. Hiller, M. C. Nemes, and M. Sampaio, “Naturalness and theoretical constraints on the higgs boson mass”, *International Journal of Theoretical Physics* **52** (Jun, 2013) 3494–3503, doi:[10.1007/s10773-013-1652-x](https://doi.org/10.1007/s10773-013-1652-x).
- [38] J. D. Wells, “Higgs naturalness and the scalar boson proliferation instability problem”, *Synthese* **194** (Dec, 2014) 477–490, doi:[10.1007/s11229-014-0618-8](https://doi.org/10.1007/s11229-014-0618-8).
- [39] D. Arnon and S. Giora, “The solar neutrino problem – an update”, *Physics Reports* **311** (Apr, 1999) 115–141, doi:[10.1016/s0370-1573\(98\)00094-5](https://doi.org/10.1016/s0370-1573(98)00094-5).
- [40] T. Binoth and J. J. van der Bij, “Influence of strongly coupled, hidden scalars on Higgs signals”, *Z. Phys.* **C75** (1997) 17–25, doi:[10.1007/s002880050442](https://doi.org/10.1007/s002880050442), arXiv:[hep-ph/9608245](https://arxiv.org/abs/hep-ph/9608245).
- [41] R. M. Schabinger and J. D. Wells, “A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider”, *Phys. Rev.* **D72** (2005) 093007, doi:[10.1103/PhysRevD.72.093007](https://doi.org/10.1103/PhysRevD.72.093007), arXiv:[hep-ph/0509209](https://arxiv.org/abs/hep-ph/0509209).
- [42] B. Patt and F. Wilczek, “Higgs-field portal into hidden sectors”, arXiv:[hep-ph/0605188](https://arxiv.org/abs/hep-ph/0605188).
- [43] G. C. Branco et al., “Theory and phenomenology of two-Higgs-doublet models”, *Phys. Rept.* **516** (2012) 1–102, doi:[10.1016/j.physrep.2012.02.002](https://doi.org/10.1016/j.physrep.2012.02.002), arXiv:[1106.0034](https://arxiv.org/abs/1106.0034).
- [44] P. Fayet, “Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino”, *Nucl. Phys.* **B90** (1975) 104–124, doi:[10.1016/0550-3213\(75\)90636-7](https://doi.org/10.1016/0550-3213(75)90636-7).
-

- [45] P. Fayet, “Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions”, *Phys. Lett.* **69B** (1977) 489, doi:10.1016/0370-2693(77)90852-8.
- [46] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped Gravitons at the LHC and Beyond”, *Phys. Rev.* **D76** (2007) 036006, doi:10.1103/PhysRevD.76.036006, arXiv:hep-ph/0701186.
- [47] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, “Searching for the Kaluza-Klein Graviton in Bulk RS Models”, *JHEP* **09** (2007) 013, doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.
- [48] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, “Phenomenology of the randall-sundrum gauge hierarchy model”, *Phys. Rev. Lett.* **84** (Mar, 2000) 2080–2083, doi:10.1103/PhysRevLett.84.2080.
- [49] C. Csáki, M. L. Graesser, and G. D. Kribs, “Radion dynamics and electroweak physics”, *Phys. Rev. D* **63** (Feb, 2001) 065002, doi:10.1103/PhysRevD.63.065002.
- [50] A. Carvalho et al., “Higgs pair production: choosing benchmarks with cluster analysis”, *Journal of High Energy Physics* **2016** (Apr, 2016) 1–28, doi:10.1007/jhep04(2016)126.
- [51] L. Cadamuro, “Search for Higgs boson pair production in the  $b\bar{b}\tau^+\tau^-$  decay channel with the CMS detector at the LHC. Recherche de la production de paires de bosons de Higgs dans le canal de désintégration  $b\bar{b}\tau^+\tau^-$  avec le détecteur CMS auprès du LHC”, Sep, 2017. Presented 05 Oct 2017. <http://cds.cern.ch/record/2292733>.
- [52] F. Bishara, R. Contino, and J. Rojo, “Higgs pair production in vector-boson fusion at the lhc and beyond”, *The European Physical Journal C* **77** (Jul, 2017) doi:10.1140/epjc/s10052-017-5037-9.
- [53] F. Goertz, A. Papaefstathiou, L. L. Yang, and J. Zurita, “Higgs boson pair production in the  $d=6$  extension of the sm”, *Journal of High Energy Physics* **2015** (4, 2015) doi:10.1007/jhep04(2015)167.
- [54] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *Journal of High Energy Physics* **2014** (Jul, 2014) doi:10.1007/jhep07(2014)079.
-

- 
- [55] M. J. Dolan, C. Englert, and M. Spannowsky, “Higgs self-coupling measurements at the lhc”, *Journal of High Energy Physics* **2012** (Oct, 2012) doi:10.1007/jhep10(2012)112.
- [56] J. Baglio et al., “The measurement of the higgs self-coupling at the lhc: theoretical status”, *Journal of High Energy Physics* **2013** (Apr, 2013) doi:10.1007/jhep04(2013)151.
- [57] A. J. Barr, M. J. Dolan, C. Englert, and M. Spannowsky, “Di-higgs final states augmt2ed – selecting hh events at the high luminosity lhc”, *Physics Letters B* **728** (Jan, 2014) 308–313, doi:10.1016/j.physletb.2013.12.011.
- [58] The ATLAS. Collaboration, “Combination of searches for higgs boson pairs in pp collisions at  $\sqrt{s}=13\text{teV}$  with the atlas detector”, *Physics Letters B* **800** (Jan, 2020) 135103, doi:10.1016/j.physletb.2019.135103.
- [59] L. Evans and P. Bryant, “LHC machine”, *Journal of Instrumentation* **3** (aug, 2008) S08001–S08001, doi:10.1088/1748-0221/3/08/s08001.
- [60] J. Haffner, “The CERN accelerator complex. Complexe des accélérateurs du CERN”, . General Photo.
- [61] A. M. Sirunyan et al., “Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13 \text{ teV}$ ”, *Journal of High Energy Physics* **2018** (Jul, 2018) doi:10.1007/jhep07(2018)161.
- [62] CMS Collaboration, “Luminosity public plots”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [63] The CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”, technical report, Geneva, 2006.
- [64] The CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [65] The CMS Collaboration, “Cutaway diagrams of CMS detector”, .
- [66] The CMS Collaboration, “Cms, tracker technical design report”, .
- [67] The CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* **9** (oct, 2014) P10009–P10009, doi:10.1088/1748-0221/9/10/p10009.
-

- [68] A. Dominguez and et al., “Cms technical design report for the pixel detector upgrade”, (9, 2012) doi:10.2172/1151650.
- [69] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”, Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep, 2012. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.
- [70] B. S. Claus Grupen, “Particle detectors”. Cambridge monographs on particle physics, nuclear physics, and cosmology 26. Cambridge University Press, 2nd ed edition, 2008. ISBN 0521840066,9780521840064,9780511388668.
- [71] The CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”, technical report, Geneva, 1997.
- [72] F. C.W., “Calorimetry in High-Energy Physics”. vol 128. Springer, Boston, MA, 1985.
- [73] A. Benaglia, “The CMS ECAL performance with examples”, *Journal of Instrumentation* **9** (feb, 2014) C02008–C02008, doi:10.1088/1748-0221/9/02/c02008.
- [74] The CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”, technical report, Geneva, 1997.
- [75] The CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019.
- [76] E. Yazgan and the CMS ECAL/HCAL Collaborations, “The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c”, *Journal of Physics: Conference Series* **160** (apr, 2009) 012056, doi:10.1088/1742-6596/160/1/012056.
- [77] C. M. S. at CERN, “Xdaq - cms online software project page”,.
- [78] The CMS Collaboration, “The CMS muon project: Technical Design Report”, technical report, Geneva, 1997.
- [79] The CMS Collaboration, “The performance of the cms muon detector in proton-proton collisions at  $s = 7$  tev at the lhc”, *Journal of Instrumentation* **8** (Nov, 2013) P11002–P11002, doi:10.1088/1748-0221/8/11/p11002.
-

- 
- [80] F. Gasparini et al., “Bunch crossing identification at lhc using a mean-timer technique”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **336** (1993), no. 1, 91–97, doi:[https://doi.org/10.1016/0168-9002\(93\)91082-X](https://doi.org/10.1016/0168-9002(93)91082-X).
- [81] A. Sirunyan et al., “Performance of the cms muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s}=13$  tev”, *Journal of Instrumentation* **13** (Jun, 2018) P06015–P06015, doi:[10.1088/1748-0221/13/06/p06015](https://doi.org/10.1088/1748-0221/13/06/p06015).
- [82] The CMS Collaboration, “CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems”, technical report.
- [83] The CMS Collaboration, “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”, technical report, Geneva, 2002.
- [84] K. Bos et al., “LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)”, technical report, Geneva, 2005.
- [85] R. Frühwirth, “Application of kalman filtering to track and vertex fitting”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987), no. 2, 444 – 450, doi:[https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [86] J. R. Cash and A. H. Karp, “A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides”, *ACM Trans. Math. Softw.* **16** (September, 1990) 201–222, doi:[10.1145/79505.79507](https://doi.org/10.1145/79505.79507).
- [87] T. Speer et al., “Vertex Fitting in the CMS Tracker”, Technical Report CMS-NOTE-2006-032, CERN, Geneva, Feb, 2006.
- [88] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (1998), no. 11, 2210–2239, doi:[10.1109/5.726788](https://doi.org/10.1109/5.726788).
- [89] W. Waltenberger, R. Frühwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (nov, 2007) N343–N356, doi:[10.1088/0954-3899/34/12/n01](https://doi.org/10.1088/0954-3899/34/12/n01).
- [90] The CMS Collaboration, “2017 tracking performance plots”,.
-

- [91] The CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* **12** (oct, 2017) P10003–P10003, doi:10.1088/1748-0221/12/10/p10003.
- [92] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc”, *Journal of Physics G: Nuclear and Particle Physics* **31** (Jul, 2005) N9–N20, doi:10.1088/0954-3899/31/9/n01.
- [93] The CMS Collaboration, “Performance of cms muon reconstruction in pp collision events at  $\sqrt{s} = 7\text{tev}$ ”, *Journal of Instrumentation* **7** (Oct, 2012) P10002–P10002, doi:10.1088/1748-0221/7/10/p10002.
- [94] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, “Electron Reconstruction in the CMS Electromagnetic Calorimeter”, Technical Report CMS-NOTE-2001-034, CERN, Geneva, Jun, 2001.
- [95] *Journal of Instrumentation* **10** (Jun, 2015) P06005–P06005, doi:10.1088/1748-0221/10/06/p06005.
- [96] H. Bethe and W. Heitler, “Adaptive vertex fitting”, *Proc. R. Soc* **A146** (1934), no. 83,.
- [97] The CMS Collaboration Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, technical report, CERN, Geneva, Dec, 2020. arXiv:2012.06888.
- [98] C. Collaboration, “Performance of  $\tau$ -lepton reconstruction and identification in CMS”, *Journal of Instrumentation* **7** (jan, 2012) P01001–P01001, doi:10.1088/1748-0221/7/01/p01001.
- [99] “Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  at CMS”, *Journal of Instrumentation* **11** (jan, 2016) P01019–P01019, doi:10.1088/1748-0221/11/01/p01019.
- [100] A. Sirunyan et al., “Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s}=13\text{ tev}$ ”, *Journal of Instrumentation* **13** (Oct, 2018) P10005–P10005, doi:10.1088/1748-0221/13/10/p10005.
- [101] K. Androsov, R. Manzoni, J. Steggemann, and R. Wolf, “DeepTau: DNN-based multi-class tau identification algorithm”, Technical Report CMS AN-19-260, CERN, Geneva, Apr, 2020.
-

- 
- [102] The CMS Collaboration, “Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons”,.
- [103] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-ktjet clustering algorithm”, *Journal of High Energy Physics* **2008** (Apr, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.
- [104] M. Cacciari, G. P. Salam, and G. Soyez, “Fastjet user manual”, *The European Physical Journal C* **72** (Mar, 2012) doi:10.1140/epjc/s10052-012-1896-2.
- [105] V. Khachatryan et al., “Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev”, *Journal of Instrumentation* **12** (Feb, 2017) P02014–P02014, doi:10.1088/1748-0221/12/02/p02014.
- [106] The CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *Journal of Instrumentation* **13** (may, 2018) P05011–P05011, doi:10.1088/1748-0221/13/05/p05011.
- [107] E. Bols et al., “Jet flavour classification using deepjet”, *Journal of Instrumentation* **15** (Dec, 2020) P12012–P12012, doi:10.1088/1748-0221/15/12/p12012.
- [108] M. Verzetti, “Machine learning techniques for jet flavour identification at cms”, *EPJ Web of Conferences* **214** (01, 2019) 06010, doi:10.1051/epjconf/201921406010.
- [109] M. Stoye, “Deep learning in jet reconstruction at CMS”, *Journal of Physics: Conference Series* **1085** (sep, 2018) 042029, doi:10.1088/1742-6596/1085/4/042029.
- [110] C. Amendola, K. Androsov, G. Bagliesi, R. Bhattacharya, F. Beaudette, F. Brivio, M. A. Ciocci, O. Davignonand, M.R. Di Domenico, M. Gallinaro, S. G. Lopez, J. L. Holgado, S. Malvezzi, J. Motta, M. Rieger, R. Salerno, Y. Sirois, G. Strong, D. Zuolo, “Search for Higgs boson pairs produced through gluon and vector boson fusion in the  $bb\tau\tau$  final state with Run-II data”, *AN-2018/121* (2020).
- [111] BTag POG, “Performance of the CMS DeepJet b-tagging algorithm (*twiki*)”.
- [112] the CMS Collaboration, “Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector”, Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.
- [113] JetMET POG, “Missing Transverse Energy (MET) (*twiki*)”.
-

- [114] JetMET POG, “MET Filter Recommendations for Run II (*twiki*)”.
  - [115] “Search for higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at  $s=13\text{TeV}$ ”, *Physics Letters B* **778** (2018) 101–127, doi:<https://doi.org/10.1016/j.physletb.2018.01.001>.
  - [116] K. Olive, “Review of particle physics”, *Chinese Physics C* **40** (oct, 2016) 100001, doi:[10.1088/1674-1137/40/10/100001](https://doi.org/10.1088/1674-1137/40/10/100001).
  - [117] The CMS Collaboration, “A DNN-based classification approach for enhancing the search for Higgs boson pair production in the  $b\bar{b}\tau\tau$  final state in pp collisions at 13 TeV”, technical report, CERN, Geneva, 2019.
  - [118] The CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
  - [119] The CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13\text{ TeV}$ ”, Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.
  - [120] Tau POG, “Tau ID recommendations for Run-2: 2016, 2017, and 2018 (*twiki*)”.
  - [121] Tau POG, “Tau Triggers (*twiki*)”.
  - [122] L. Bianchini, J. Conway, E. Friis, and C. Veelken, “Reconstruction of the higgs mass in  $h \rightarrow \tau\tau$  events by dynamical likelihood techniques”, *Journal of Physics: Conference Series* **513** (06, 2014) 022035, doi:[10.1088/1742-6596/513/2/022035](https://doi.org/10.1088/1742-6596/513/2/022035).
  - [123] K. Androsov, M. A. Ciocci, and M. R. Di Domenico, “HH-BTag a ML-based method to select b jets from  $H \rightarrow b\bar{b}$  decay, in  $HH \rightarrow b\bar{b}\tau\bar{\tau}$  searches”, Technical Report CMS AN-2019-283, CERN, Geneva.
  - [124] CMS Collaboration Collaboration, “Search for non-resonant Higgs boson pair production in the  $b\bar{b}\tau\tau$  final state using 2016 data”, technical report, CERN, Geneva, 2016.
  - [125] J. L. Holgado, S. G. Lopez, and M. Rieger, “A multiclass classification for ggF and VBF  $HH \rightarrow b\bar{b}\tau\tau$ ”, Technical Report CMS AN-2020-095, CERN, Geneva, Apr, 2020.
  - [126] M. Gallinaro and G. C. Strong, “A DNN-based classification approach for enhancing the search for Higgs boson pair production in the  $b\bar{b}\tau\tau$  final state in pp collisions at  $\sqrt{s} = 13\text{TeV}$ ”, *AN-19-188* (2019).
-

- 
- [127] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review* **65** (1957), no. 6, 386–408.
- [128] S. Linnainmaa, “The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors”, Master’s thesis, Univ. Helsinki, 1970.
- [129] P. J. Werbos, “Applications of advances in nonlinear sensitivity analysis”, in *Proceedings of the 10th IFIP Conference, 31.8 - 4.9, NYC*, pp. 762–770. 1981.
- [130] CMS Collaboration, “Prospects for HH measurements at the HL-LHC”, Technical Report CMS-PAS-FTR-18-019, CERN, Geneva, 2018.
- [131] b Tag and Vertexing POG, “Recommendation for Using b-tag Objects in Physics Analyses (*twiki*)”.
- [132] CMS Collaboration, “Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Lett.* **B778** (2018) 101–127, doi:10.1016/j.physletb.2018.01.001, arXiv:1707.02909.
- [133] The CMS Collaboration, “Combination of searches for Higgs boson pair production in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-HIG-17-030, CERN, Geneva, 2018.
- [134] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *The European Physical Journal C* **53** (Dec, 2007) 473–500, doi:10.1140/epjc/s10052-007-0490-5.
- [135] R. Frederix and S. Frixione, “Merging meets matching in mc@nlo”, *Journal of High Energy Physics* **2012** (Dec, 2012) doi:10.1007/jhep12(2012)061.
- [136] J. Alwall et al., “A standard format for les houches event files”, *Computer Physics Communications* **176** (Feb, 2007) 300–304, doi:10.1016/j.cpc.2006.11.010.
- [137] JetMET POG, “SW Guide LHE Interface (*twiki*)”.
- [138] M. Grazzini et al., “Higgs boson pair production at nnlo with top quark mass effects”, *Journal of High Energy Physics* **2018** (May, 2018) doi:10.1007/jhep05(2018)059.
-

- [139] F. A. Dreyer and A. Karlberg, “Vector-boson fusion higgs pair production at  $n^3\text{LO}$ ”, *Phys. Rev. D* **98** (Dec, 2018) 114016, doi:[10.1103/PhysRevD.98.114016](https://doi.org/10.1103/PhysRevD.98.114016).
- [140] D. de Florian et al., “Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector”, *CERN Yellow Reports: Monographs - Volume 2/2017* (2016) arXiv:1610.07922.
- [141] R. Barlow and C. Beeston, “Fitting using finite monte carlo samples”, *Computer Physics Communications* **77** (1993), no. 2, 219–228, doi:[https://doi.org/10.1016/0010-4655\(93\)90005-W](https://doi.org/10.1016/0010-4655(93)90005-W).
- [142] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, “Procedure for the LHC Higgs boson search combination in Summer 2011”, technical report, CERN, Geneva, Aug, 2011.
- [143] M. Baak, S. Gadatsch, R. Harrington, and W. Verkerke, “Interpolation between multi-dimensional histograms using a new non-linear moment morphing method”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **771** (2015) 39–48, doi:<https://doi.org/10.1016/j.nima.2014.10.033>.
- [144] S. Raschka, “Python Machine Learning”. Packt Publishing, 2015. ISBN 978-1-78355-513-0.
- [145] I. Zafar et al., “Hands-On Convolutional Neural Networks with TensorFlow: Solve computer vision problems with modeling in TensorFlow and Python”. Packt Publishing Ltd, 2018. ISBN 9781789132823,1789132827.
- [146] Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels”, in *Advances in Neural Information Processing Systems*, S. Bengio et al., eds., volume 31, pp. 8778–8788. Curran Associates, Inc., 2018.
- [147] S. Ruder, “An overview of gradient descent optimization algorithms”, 2017.
- [148] P. M. Radiuk, “Impact of training set batch size on the performance of convolutional neural networks for diverse datasets”, *Information Technology and Management Science* **20** doi:[10.1515/itms-2017-0003](https://doi.org/10.1515/itms-2017-0003).
- [149] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift”,.
-

- 
- [150] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, 2017.
- [151] U. Michelucci, “Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks”. Apress, 2018. ISBN 1484237897, 9781484237892.
- [152] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions”, doi:10.1142/S0218488598000094.
- [153] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural computation* **9** (12, 1997) 1735–80, doi:10.1162/neco.1997.9.8.1735.
- [154] S. R. R. Venna et al., “A novel data-driven model for real-time influenza forecasting”, *IEEE Access* **7** (2019) 7691–7701, doi:10.1109/ACCESS.2018.2888585.
- [155] X. Yuan, L. Li, and Y. Wang, “Nonlinear dynamic soft sensor modeling with supervised long short-term memory network”, *IEEE Transactions on Industrial Informatics* **PP** (02, 2019) 1–1, doi:10.1109/TII.2019.2902129.
- [156] I. Faran et al., “Ground truth simulation for deep learning classification of mid-resolution venus images via unmixing of high-resolution hyperspectral fenix data”, *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium* (Jul, 2019) doi:10.1109/igarss.2019.8900186.
- [157] A. B. Peter Bruce, “Practical Statistics for Data Scientist”. O’Reilly, 2017. ISBN 9781491952962.
- [158] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous systems”, 2015. Software available from tensorflow.org. <http://tensorflow.org/>.
- [159] F. Chollet et al., “Keras”, 2015. <https://github.com/fchollet/keras>.
- [160] P. I. Frazier, “A Tutorial on Bayesian Optimization”, 2018.
- [161] F. Nogueira, “Bayesian Optimization: Open source constrained global optimization tool for Python”, 2014–. <https://github.com/fmfn/BayesianOptimization>.
- [162] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, “Information-theoretic regret bounds for gaussian process optimization in the bandit setting”, *IEEE Transactions on Information Theory* **58** (May, 2012) 3250–3265, doi:10.1109/tit.2011.2182033.
-

- [163] J. Berk, S. Gupta, S. Rana, and S. Venkatesh, “Randomised Gaussian Process Upper Confidence Bound for Bayesian Optimisation”, 2020.
  - [164] L. Breiman, “Random forests”, *Machine Learning* **45** (10, 2001) doi:10.1023/a:1010933404324.
  - [165] C. Molnar, “Interpretable Machine Learning”. 2019.  
<https://christophm.github.io/interpretable-ml-book/>.
  - [166] L. D. Brown, T. T. Cai, and A. DasGupta., “Interval estimation for a binomial proportion”, *Statistical Science* **16** (05, 2001) doi:10.1214/ss/1009213286.
  - [167] BTag POG, “Jet Flavour Identification (MC Truth) (*twiki*)”.
  - [168] JetMET POG, “Jet identification in high pile-up environment (*twiki*)”.
  - [169] JetMET POG, “<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution> (*twiki*)”.
  - [170] ECAL L1 Trigger, “Reweighting recipe to emulate Level 1 ECAL prefiring (*twiki*)”.
  - [171] Tau POG, “TauTriggerSFs repository”.
  - [172] Chiara Amendola, “VBFHTauTau Trigger Scale Factors repository”.
  - [173] b Tag and Vertexing POG, “B-tagging discriminant shape calibration using event weights with a tag-and-probe method (*twiki*)”.
  - [174] CMS HTT Group, “LeptonEfficiencies repository”.
-