

# Search for multiple Higgs boson production in hadronic final states at the Large Hadron Collider

---

**Chitroda, Bhakti Kanulal**

**Doctoral thesis / Doktorski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:911256>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-03-17**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)





University of Zagreb

FACULTY OF SCIENCE

Bhakti Kanulal Chitroda

**Search for multiple Higgs boson production in  
hadronic final states at the Large Hadron Collider**

DOCTORAL DISSERTATION

Zagreb, 2024.





University of Zagreb

FACULTY OF SCIENCE

Bhakti Kanulal Chitroda

**Search for multiple Higgs boson production in  
hadronic final states at the Large Hadron Collider**

DOCTORAL DISSERTATION

Supervisor:  
Dinko Ferenc̆ek, Ph.D.

Zagreb, 2024.





Sveučilište u Zagrebu

PRIRODOSLOVNO MATEMATIČKI FAKULTET

Bhakti Kanulal Chitroda

**Potruga za višestrukom produkcijom Higgsovih  
bozona u hadronskim konačnim stanjima na  
Velikom hadronskom sudarivaču**

DOKTORSKI RAD

Mentor:  
dr. sc. Dinko Ferenček

Zagreb, 2024.



## Supervisor Information

Dinko Ferenček

### Education

2006–2011	<b>Ph.D., Physics</b> , University of Maryland, USA
2001–2006	<b>M.Sc., Physics</b> , University of Zagreb, Croatia

### Work Experience

2021– Present	<b>Senior Research Associate</b> , Ruđer Bošković Institute, Zagreb, Croatia
2015–2021	<b>Research Associate</b> , Ruđer Bošković Institute, Zagreb, Croatia
2011–2015	<b>Postdoctoral Research Associate</b> , Rutgers, The State University of New Jersey, USA
2007–2011	<b>Research Assistant</b> , University of Maryland, USA
2006–2007	<b>Teaching Assistant</b> , University of Maryland, USA

### Selected Publications

1. CMS Collaboration, “Measurement of simplified template cross sections of the Higgs boson produced in association with W or Z bosons in the  $H \rightarrow b\bar{b}$  decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Rev. D* **109** (2024) 092011, [arXiv:2312.07562](https://arxiv.org/abs/2312.07562)
2. CMS Collaboration, “Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quarks final state with boosted topology”, *Phys. Lett. B* **842** (2023) 137392, [arXiv:2204.12413](https://arxiv.org/abs/2204.12413)



- 17 3. Ferenčec D., Roguljić M., Starodumov A., “Production, calibration, and perfor-  
18 mance of the layer 1 replacement modules for the CMS pixel detector”, *Proceedings*  
19 *of The 29th International Workshop on Vertex Detectors (VERTEX2020)*, JPS Conf.  
20 Proc. **34** (2021) 010023
- 21 4. Roguljić M., Starodumov A., Karadzhinova-Ferrer A., Ferenčec D., Ahmed A. A.,  
22 Jara-Casas L. M., “Low dose rate  $^{60}\text{Co}$  facility in Zagreb”, *Proceedings of The*  
23 *28th International Workshop on Vertex Detectors (VERTEX2019)*, PoS Vertex2019  
24 (2020) 066
- 25 5. Majer M., Roguljić M., Knežević, Ž., Starodumov A., Ferenčec D., Brigljević V.,  
26 Mihaljević B., “Dose mapping of the panoramic  $^{60}\text{Co}$  gamma irradiation facility  
27 at the Ruđer Bošković Institute – Geant4 simulation and measurements”, *Appl.*  
28 *Radiat. Isot.* **154** (2019) 108824
- 29 6. CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in  
30 pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, arXiv:1712.07158
- 31 7. CMS Collaboration, “Search for dijet resonances in proton-proton collisions at  $\sqrt{s} =$   
32 13 TeV and constraints on dark matter and other models”, *Phys. Lett. B* **769** (2017)  
33 520, arXiv:1611.03568
- 34 8. CMS Collaboration, “Search for narrow resonances in dijet final states at  $\sqrt{s} = 8$  TeV  
35 with the novel CMS technique of data scouting”, *Phys. Rev. Lett.* **117** (2016)  
36 031802, arXiv:1604.08907
- 37 9. CMS Collaboration, “Search for heavy resonances decaying to two Higgs bosons in fi-  
38 nal states containing four b quarks”, *Eur. Phys. J. C* **76** (2016) 371, arXiv:1602.08762
- 39 10. CMS Collaboration, “Search for narrow resonances decaying to dijets in proton-  
40 proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Rev. Lett.* **116** (2016) 071801, arXiv:1512.01224
- 41 11. CMS Collaboration, “Search for pair-produced vector-like B quarks in pp collisions  
42 at  $\sqrt{s} = 8$  TeV”, *Phys. Rev. D* **93** (2016) 112009, arXiv:1507.07129
- 43 12. CMS Collaboration, “Search for vector-like T quarks decaying to top quarks and  
44 Higgs bosons in the all-hadronic channel using jet substructure”, *JHEP* **06** (2015)  
45 080, arXiv:1503.01952

- 46 13. CMS Collaboration, “Search for resonances and quantum black holes using dijet  
47 mass spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Rev. D* **91** (2015)  
48 [052009](#), [arXiv:1501.04198](#)
- 49 14. CMS Collaboration, “Search for pair-produced resonances decaying to jet pairs in  
50 proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Lett. B* **747** (2015) 98, [arXiv:1412.7706](#)
- 51 15. CMS Collaboration, “Search for narrow resonances and quantum black holes in  
52 inclusive and b-tagged dijet mass spectra from pp collisions at  $\sqrt{s} = 7$  TeV”, *JHEP*  
53 **01** (2013) 013, [arXiv:1210.2387](#)
- 54 16. CMS Collaboration, “Search for First Generation Scalar Leptoquarks in the  $evjj$   
55 Channel in pp Collisions at  $\sqrt{s} = 7$  TeV”, *Phys. Lett. B* **703** (2011) 246, [arXiv:1105.5237](#)
- 56 17. CMS Collaboration, “Missing transverse energy performance of the CMS detector”,  
57 *JINST* **6** (2011) 09001, [arXiv:1106.5048](#)
- 58 18. CMS Collaboration, “Search for Pair Production of First-Generation Scalar Lepto-  
59 quarks in pp Collisions at  $\sqrt{s} = 7$  TeV”, *Phys. Rev. Lett.* **106** (2011) 201802,  
60 [arXiv:1012.4031](#)
- 61 19. V. Brigljević et. al., “Study of di-boson production with the CMS detector at LHC”,  
62 *J. Phys. G* **34** (2007) N269-N295



## Acknowledgements

64 This thesis work is the result of numerous people's encouragement and advice. I am  
65 extremely grateful of each and every one of them.

66 First and foremost, I would like to sincerely thank my supervisor, Dr. Dinko Ferenček,  
67 for his unwavering support and motivation, without which this research project would  
68 not have been feasible. Thank you for helping me become a more confident person and  
69 for enhancing my communication skills. His thoughtfully organised digital and physical  
70 life, coupled with his curiosity to understand the core reason behind any odd outcomes,  
71 will always serve as an inspiration to me. You are excellent mentor and I am grateful for  
72 the opportunity to learn as much from you. There is much more to say, but to sum up,  
73 I would say that because it was such an exciting period overall, five years did not seem  
74 like a long journey.

75 Along with my supervisor, I am grateful to Dr. Vuko Brigljević, Dr. Andrey Starodumov,  
76 and Dr. Devdatta Majumdar for their advice and assistance during the early stages of  
77 my research. I appreciate their tolerance for my shoddy communication skills. All of  
78 them have helped me understand the physics concepts I needed. Thank you for always  
79 believing in me and correcting me on my mistakes.

80 I want to thank my senior PhD fellows, Dr. Matej Roguljić and Dr. Saswat Mishra for  
81 their discussions and help. I'll always be envious of Matej's discipline, organized days,  
82 and presentation skills. I still have a lot to learn from him. I thank Tanja and Antonija  
83 for great coffees with interesting conversations from which I have learnt different things.  
84 I will always be grateful for the group at IRB for helping me during my sick days and  
85 giving me time to recover. They have been tremendously helpful in making my first time  
86 living abroad feel like home.

87 This research is a result of collaboration with the CMS in association with CERN. I  
88 want to thank all of the engineers, scientists, and students for their hard work and pro-  
89 activeness in solving any technical difficulties we face during this research. I would also  
90 like to express my gratitude to the CMS collaboration for providing opportunities to  
91 participate in various physics collaborative research activities to advance my career as an  
92 experimental high-energy physicist. And personally, I want to thank Chris, Alessandro,

93 Nick, Aliya, Krunal, Adinda, Pirmin, Sam, and CMS VHbb team, Atanu, Giulia, Danek,  
94 Lea, Sahiti, and Pixel online team, Maxime, and  $HHbb\gamma\gamma$  team.

95 I am thankful to my family for giving me the opportunity and support. Especially my  
96 parents, for their continuous efforts and sacrifices for my upbringing and education. Nidhi  
97 and Tilak for taking care of parents while I am away from the home. I would want to  
98 express deep appreciation to Dr. Vaishali Joshi, my school physics teacher, for igniting  
99 my interest in Physics.

100 I have always been very shy to make new friends, moving to new country was hard enough  
101 that I had to make new friends as well. From bottom of my heart, I am grateful to Iris,  
102 Lara, Navin, Meghana, Akshay, Priyanka, Hrishi, Nayeem, Siddharth, Diwakar, Gaurav  
103 and everyone else for the great time. I can't even imagine going through pandemic without  
104 the zoom calls with my friends in India, Chandra Shekhar, Vismay, Sharan, Vishnu, Vidhi,  
105 Namrata, Binal, Rushabh, Avik. I thank you all for the casual but insightful discussions,  
106 be it on Physics or on life.

107 Nidhi and Zankhana, you are the biggest support I have. Words won't be enough to  
108 express my gratitude towards them. They have heard all my excitements, blabbers, silly  
109 complaints, just sharing with them would be enough to deal with anything and the support  
110 they gave me in return is unthinkable.

111 Thank you everyone for beautiful memories.

## Abstract

112

113 After the discovery of the Higgs boson, one of the main purposes of the physics program  
114 at the LHC has been a thorough characterization of its properties. Among important  
115 properties of Higgs lies trilinear and quartic Higgs self-coupling, accessible through di-  
116 Higgs and triple Higgs production respectively. This thesis primarily reports on two  
117 highly significant and intriguing studies that were performed utilizing data from the Run-  
118 2(2016–2018) proton–proton collision at a center-of-mass energy of 13 TeV collected by  
119 the CMS experiment at the LHC. The first study concerns the search for the non-resonant  
120 H pair produced through gluon-gluon fusion (ggF) and Vector Boson Fusion (VBF) modes  
121 in two photons and two b-jets final state. The  $b\bar{b}\gamma\gamma$  final state is one of the most sensitive  
122 channels to the HH signal thanks to a relatively clean signature of two high-energy photons  
123 and large branching ratio of two heavy-flavor high-energy jets (b-jets). A 95% confidence  
124 level (CL) upper limit on the product of Higgs boson pair production cross section and  
125 branching fraction are derived as a function of  $\kappa_\lambda$  parameter. The constraints on the  
126 anomalous coupling of a pair of H with a pair of gauge bosons was determined for the  
127 first time by the CMS collaboration.

128 The second study focuses on the search for triple Higgs boson production in three b-jet  
129 pairs final state. Given the extremely low standard model (SM) cross-section for triple-  
130 Higgs production, the Two-Real Singlet Model (TRSM) is considered as an extension of  
131 SM. TRSM introduces two real scalars X and Y, which decay to three SM Higgs bosons  
132 which decays further to three pair of b-quarks through the process  $X \rightarrow YH \rightarrow HHH \rightarrow$   
133  $b\bar{b}b\bar{b}b\bar{b}$ . The search is performed in mass ranges of X (1–4 TeV) and Y (300–2800 GeV)  
134 where the H is highly Lorentz-boosted. In this kinematic regime, decayed b-quark pairs  
135 are collimated enough to allow the reconstruction of H using single large-area jets. We  
136 have considered two topologies: one topology where all three Higgs bosons are boosted and  
137 another where two of the three Higgs bosons are boosted. This analysis is still ongoing.  
138 A scan will be performed in a two dimensional plane spanned by the invariant mass of the  
139 two large-area jets associated to Y, and the invariant mass of three large-area jets used  
140 to reconstruct X.

141 Keywords: LHC, CMS, standard model, Higgs boson, Higgs self-coupling, TRSM, boosted  
142 objects, b-quarks, photons, Higgs-pair, triple Higgs



144 **Uvod**

145 Postoje četiri temeljne sile koje upravljaju svim interakcijama u svemiru, a rezultirale  
146 su stvaranjem svijeta u kojem živimo. To su – jaka nuklearna, elektromagnetska, slaba  
147 nuklearna, i gravitacijska sila, u padajućem redoslijedu relativnih jakosti. Zajedno, te  
148 sile čine manje od 5% ukupnog maseno-energetskog sadržaja svemira, u obliku obične  
149 materije i energije. To znači da je više od 95% svemira neistraženo područje koje ljudi  
150 vrlo malo ili nimalo ne razumiju.

151 Standardni model (SM) elementarnih čestica vrhunac je znanstvenih napora da se tri od  
152 četiri temeljne sile prirode (sve osim gravitacije) usklade u zajednički teorijski okvir. To  
153 je pomoglo u razumijevanju zakona koji upravljaju različitim temeljnim interakcijama  
154 i evolucije svemira od vremena Velikog praska do njegovog današnjeg oblika. U SM-u  
155 postoje tri generacije fermiona koji sačinjavaju materiju, a svaka se sastoji od nabijenog  
156 leptona (elektron, mion i tau), leptonskog neutrina i para kvarkova (u i d; s i c; t i b)  
157 . Dakle, postoji šest leptona i šest kvarkova; svaka od ovih dvanaest čestica ima svoju  
158 antičesticu. Osim toga, postoji ukupno pet bozona nositelja sile ( $W^\pm$ ,  $Z$ ,  $\gamma$ ,  $g$ ) sa spinom  
159 1, koji odgovaraju trima silama prirode. Naposljetku, skalarni Higgsov bozon stupa u  
160 interakciju kroz svoje polje s ostalim elementarnim česticama i dodjeljuje im njihove  
161 mase. Međutim, Higgsov bozon ne stupa u interakciju s neutrinima i stoga su neutriini  
162 bez mase u SM.

163 SM je vrlo uspješna prekretnica u našem trenutnom razumijevanju različitih pojava koje  
164 se događaju u našem svemiru. Kompletan okvir SM-a, koji danas poznajemo, razvio se  
165 tijekom proteklih 50 godina eksperimenata, uspješno prolazeći kroz nebrojena rigorozna  
166 testiranja različitih pretpostavki modela. Od uspješnog opisa elektronskog magnetskog  
167 dipolnog momenta do preciznosti od 11 signifikantnih znamenki, do predviđanja posto-  
168 janja novih čestica, SM je nastavio držati svoje mjesto u polju fizike čestica kada je riječ  
169 o opisivanju zakona svemira. Međutim, to još uvijek nije potpuna opis prirode.

170 Dokazi o oscilacijama neutrina iz eksperimenta Super Kamiokande utvrdili su da najmanje  
171 dva od tri neutrina moraju imati vrlo malu masu različitu od nule. Mehanizam kojim  
172 SM neutriini postižu svoju masu, i koji ju čini iznimno malom, trenutno je misterija. 27%



173 maseno-energijskog sadržaja svemira je u obliku tamne tvari, za što postoje neizravni  
174 kozmološki dokazi iz rotacijskih krivulja galaksija ili gravitacijskih leća, za koju nema  
175 izravnog objašnjenja ili čestice kandidata u SM-u. Preostalih 68% svemira zajednički se  
176 naziva tamnom energijom, za koju ne postoje ni izravni ni neizravni zaključci o njezinu  
177 podrijetlu i mikroskopska svojstvima. To bi mogla biti potpuno nova, peta temeljna sila  
178 prirode, s vrlo potisnutom interakcijom s česticama iz SM-a. Zanimljivo je da SM također  
179 ne uključuje četvrtu temeljnu interakciju, gravitacijsku silu, u svoj okvir kvantne teorije  
180 polja i stoga ne može objasniti zašto je gravitacija  $\mathcal{O}(10^{-38})$  puta slabija od jake sile.  
181 Osim ovih, postoje i nepotpuna i nezadovoljavajuća objašnjenja koja stoje iza asimetrije  
182 materije i antimaterije u svemiru, činjenice da postoje tri generacije fermiona materije i  
183 hijerarhija mase među njima i još mnogo toga. Sve to ukazuju na činjenicu da bi SM  
184 mogao biti efektivna teorija ili niskoenergetska manifestacija još temeljnije teorije prirode  
185 koja djeluje na mnogo višim energetske skalama.

186 Gornji nedostaci dokazuju da smo, čak i nakon što smo na pravom putu da razumijemo  
187 svemir, još uvijek daleko od temeljne znanosti koja stoji iza svega. Zbog toga fizičari  
188 razvijaju nove teorije koje bi mogle reproducirati opažanja SM-a uz pružanje objašnjenja  
189 gore navedenih tajanstvenih pojava. One su poznate kao teorije izvan Standardnog modela  
190 (BSM). U tu skupinu spadaju supersimetrija (SUSY), iskrivljene ekstra dimenzije (WED)  
191 i mnogi drugi modeli. Međutim, postavlja se pitanje gdje tražiti novu fiziku? Jedna od  
192 mogućnosti je da koristimo Higgsov bozon kao stepenicu za dolazak do vrata nove fizike.

193 Nakon otkrića Higgsovog bozona, jedna od glavnih svrha istraživačkog programa na LHC-  
194 u bila je temeljita karakterizacija njegovih svojstava. Razlika između mjerenja i odgo-  
195 varajućih očekivanja standardnog modela (SM) bila bi jasan znak nove fizike. Među  
196 najosjetljivijim parametrima za novu fiziku, dostupnima na LHC-u, su trilinearno samo-  
197 sprezanje Higgsovog bozona  $\lambda_{HHH}$ , kvartično samosprezanje Higgsovog bozona  $\lambda_{HHHH}$ ,  
198 vezanje Higgsovog bozona na top kvark  $y_t$ , i sprezanje između dva vektorska bozona i  
199 dva Higgsova bozona  $c_{2\nu}$ . Parametri  $\lambda_{HHH}$  i  $\lambda_{HHHH}$  nastaju u Standardnom modelu iz  
200 razvoja potencijala polja Higgsovog bozona oko njegove vakuumske očekivane vrijednosti.  
201 Stoga njihovo izravno mjerenje pruža temeljni test predviđanja SM-a za oblik potencijala  
202 polja Higgsovog bozona. Najprikladniji procesi za izravno mjerenje  $\lambda_{HHH}$  i  $\lambda_{HHHH}$  su pro-  
203 dukcija para i tripleta Higgsovih bozona. Parametar  $y_t$  definira najjaču spregu Higgsovog

204 bozona s fermionima i temeljan je, primjerice, za određivanje (meta)stabilnosti vakuuma  
205 SM-a. Dva procesa osjetljiva na  $y_t$  su produkcija  $t\bar{t}H$  i para Higgsovih bozona. Sprezanje  
206  $c_{2V}$  očekuje se u SM-u kao posljedica mehanizma spontanog narušavanja simetrije. Pro-  
207 dukcija para Higgsovih bozona putem fuzije vektorskih bozona najosjetljiviji je proces na  
208 parametar  $c_{2V}$  dostupan na LHC-u.

209 Višestruke produkcije Higgsovog bozona ključne su za mjerenja važnih parametara. Ovaj  
210 rad sastoji se od dva glavna dijela: 1) potrage za nerezonantnom produkcijom para Hig-  
211 gsovih bozona i 2) rezonantne produkcije tripleta Higgsovih bozona. U SM-u moguća  
212 je samo nerezonantna produkcija tripleta Higgsovih bozona s vrlo malim udarnim pres-  
213 jekom. Stoga se razmatraju modifikacije SM-a, kao što su EFT vezanja i modifikatori za  
214 produkciju para Higgsovih bozona te Two-Real Singlet Model (TRSM) za rezonantnu pro-  
215 dukciju tripleta Higgsovih bozona što predstavlja proširenje SM-a s dva dodatna realna  
216 skalara. U SM-u na izmjerenoj masi Higgsovog bozona,  $H \rightarrow b\bar{b}$  kanal raspada ima najveći  
217 omjer grananja. Dakle, da bi se poboljšala statistika za produkciju tripleta Higgsovih  
218 bozona, razmatra se konačno stanje gdje se sva tri Higgsova bozona raspadaju u par  $b$   
219 kvarkova parova ( $X \rightarrow YH \rightarrow HHH \rightarrow b\bar{b}b\bar{b}$ ). S druge strane, konačno stanje  $b\bar{b}\gamma\gamma$  jedan  
220 je od najosjetljivijih kanala za HH signal zahvaljujući relativno čistom potpisu dva fotona  
221 visoke energije i dva visokoenergetska mlaza teškog okusa ( $b$ -mlazovi) ( $pp \rightarrow HH \rightarrow b\bar{b}\gamma\gamma$ ).

## 222 **Ekperimentalni postav**

223 LHC je najveći i najsnažniji akcelerator čestica na svijetu te može postići sudare protona  
224 s energijom centra mase ( $\sqrt{s}$ ) do 14 TeV i sudare teških iona s energijom centra mase  
225 do 2,76 TeV po nukleonu. Među glavnim ciljevima istraživačkog programa LHC-a bila  
226 je potraga za Higgsovim bozonom koji je prvi put uočen 2012. godine. LHC je kružni  
227 akcelerator s opsegom od oko 27 km. Nalazi se u blizini Ženeve na granici Francuske i  
228 Švicarske u tunelu na dubini između 50 i 175 m ispod zemlje.

229 Jedan od detektora dizajniranih za rekonstrukciju sudara s LHC-a je Compact Muon  
230 Solenoid (CMS). CMS detektor nalazi se u jednoj od četiri točke interakcije duž prstena  
231 LHC-a. Trenutačni istraživački program CMS-a usmjeren je na karakterizaciju Higgsovog  
232 bozona kao i na precizna mjerenja parametara SM-a, posebno elektroslabog sektora. U  
233 isto vrijeme, potrage za BSM fenomenima do energetske skale TeV također su ključne u  
234 istraživačkom programu CMS-a. CMS detektor ima 13 m dug supravodljivi solenoid s

235 unutarjim polumjerom od 5,9 m. Solenoid može osigurati jednolično magnetsko polje  
236 od 3,8 T unutar svog cilindra. Ovo snažno magnetsko polje omogućuje visokoprecizna  
237 mjerenja količine gibanja nabijenih čestica.

238 Komponente CMS detektora počevši od točke interakcije su:

- 239 • Unutarnji detektor tragova: visoko segmentirani detektor napravljen od silicijskih  
240 piksela koristi se za rekonstrukciju primarnih točaka sudara i moguće sekundarnih  
241 vrhova nastalih raspadima kratkoživućih čestica. Izvan piksel detektora nalazi se  
242 detektor sa silicijskim trakama za praćenje putanja nabijenih čestica unutar mag-  
243 netskog polja i mjerenje njihove transverzalne količine gibanja.
- 244 • Elektromagnetski kalorimetar (ECAL): homogeni kalorimetar napravljen od scintili-  
245 rajućeg kristala za mjerenje energije elektromagnetskog pljuska potaknutog fotonima  
246 i elektronima.
- 247 • Hadronski kalorimetar (HCAL): uzorkujući kalorimetar sa slojevima mjedenog radi-  
248 jatora koji se izmjenjuju sa slojevima plastičnog scintilatora. Svrha HCAL-a je  
249 mjerenje energije hadrona.
- 250 • Mionski sustav: sustav od tri različite detektorske tehnologije za učinkovitu rekon-  
251 strukciju miona kao i precizno mjerenje njihove količine gibanja, posebice za mione  
252 s visokim  $p_T$ . Mionske komore nalazi se izvan solenoida, uglavljene u željezni jaram  
253 za povrat magnetskog toka.

## 254 Analiza podataka

255 Ovaj doktorski rad usredotočen je na potrage za produkcijom parova i tripleta Higgsovih  
256 bozona. Istraživački rad možemo podijeliti u dva dijela:

- 257 1. nerezonantna produkcija para Higgsovih bozona u konačnom stanju  $b\bar{b}\gamma\gamma$ ,
- 258 2. rezonantna produkcija tripleta Higgsovih bozona u konačnom stanju  $b\bar{b}b\bar{b}b\bar{b}$ .

259 **Nerezonantna produkcija para Higgsovih bozona i kasniji raspad na par fotona**  
260 **i  $b$  kvarkova**

261 Potraga za nerezonantnom produkcijom parova Higgsovih bozona u konačnom stanju  
262  $b\bar{b}\gamma\gamma$  iskorištava podatke koje je prikupio CMS detektor u sudarima protona s energijom  
263 u centru mase od 13 TeV, za ukupni integrirani luminozitet od  $137 \text{ fb}^{-1}$ . Utvrđeno je  
264 da je produkcija HH u skladu sa SM-om. Postavljene su granice na anomalne vrijednosti  
265 parametara Higgsovih vezanja. Parametri koji se razmatraju su trilinearno samosprezanje  
266 Higgsovog bozona  $\lambda_{HHH}$ , konstanta vezanja dva Higgsova bozona s dva vektorska bozona  
267  $c_{2V}$  i Yukawino vezanje Higgsovog bozona s top kvarkom  $y_t$ .

268 Postoje dva glavna načina produkcije, fuzija gluona (ggF HH) i fuzija vektorskih bozona  
269 (VBF HH) koji se razmatraju s nekim modificiranim EFT parametrima.  $HH \rightarrow b\bar{b}\gamma\gamma$   
270 konačno stanje sastoji se od dva fotona visoke energije i dva b-mlaza visoke energije.  
271 Za VBF HH topologiju, dva dodatna mlaza s velikom razlikom u pseudorapiditetu su  
272 producirana u konačnom stanju. Glavni izvori pozadine su procesi  $\gamma\gamma + \text{mlaz}$  proces i  
273  $\gamma + \text{mlaz}$  s jednim mlazom pogrešno identificiranim kao foton. Drugi važni izvori pozadine  
274 su procesi  $t\bar{t}\gamma\gamma$  i  $t\bar{t}\gamma$ . Događaji pojedinačne produkcije Higgsovog bozona koji se raspada  
275 u par fotona predstavljaju dodatni važan izvor pozadine za HH pretragu. Posebno je  
276 izražena kontaminacija događaja  $t\bar{t}H(\gamma\gamma)$ , koji u konačnom stanju imaju dva fotona i dva  
277 b-mlaza, u područjima HH signala.

278 Kandidati su prvenstveno odabrani zahtijevajući dva rekonstruirana fotona visoke en-  
279 ergije. Kako bi se povećala osjetljivost na HH signal, događaji su klasificirani u ekskluzivne  
280 kategorije od kojih svaka cilja na specifičan mehanizam produkcije para Higgsovih bozona,  
281 tj. ggF HH i VBF HH. Osim toga, razvijeni su klasifikatori multivarijatne analize (MVA)  
282 za izolaciju svakog signala od njegove pozadine u ciljanoj kategoriji. Radi jasnoće, tijek  
283 analize sažet je na Slici 1. Potklasifikacija događaja na temelju rezultata MVA klasifika-  
284 tora provodi se u svakoj kategoriji. U kategorijama ggF HH i VBF HH, događaji se dalje  
285 klasificiraju korištenjem invarijantne mase sustava četiri tijela koji se sastoji od para fo-  
286 tona i para mlazova koji tvore  $H \rightarrow \gamma\gamma$  i  $H \rightarrow b\bar{b}$  kandidate za poboljšanje osjetljivosti na  
287 nekoliko BSM scenarija.

288 Signal se identificira kao vrh u distribuciji  $m_{\gamma\gamma}$  i vrh u distribuciji  $m_{jj}$ , oba na vrijednosti  
289  $m_H$ . Za oba procesa, pozadina se modelira iz podataka korištenjem parametarskog modela  
290 pozadine. Kako bi se izmjerili parametri od interesa, provodi se istovremena prilagodba  
291 raspodjelama  $m_{\gamma\gamma}$  i  $m_{jj}$ , za koje se pretpostavlja da nisu u korelaciji, u svakoj od kategorija

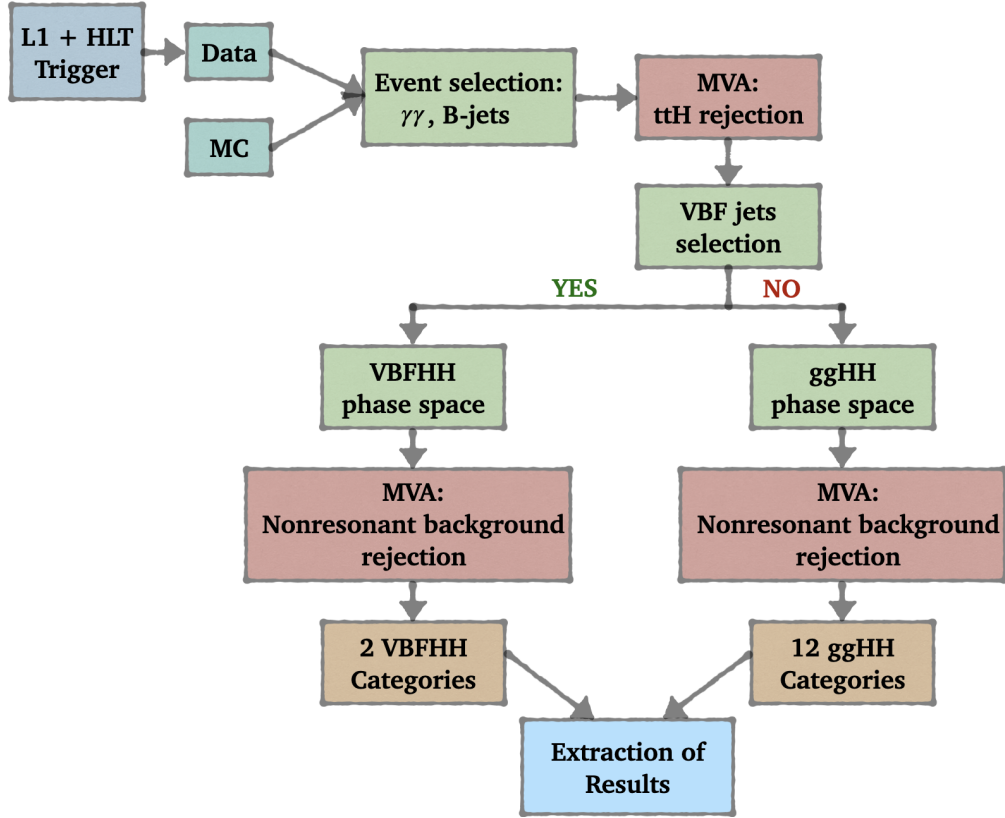


Figure 1: Shematski prikaz tijeka analize.

292 obogaćenih HH signalom.

## 293 Rezultati

294 HH kategorije (dvanaest ggF HH i dvije VBF HH kategorije) uključene su u ekstrakciju  
 295 jakosti HH signala. Očekivane i opažene jakosti signala navedene su u Tab. 1.

Parametar	Očekivano	Opaženo
$\mu_{HH}$	$1.0^{+2.7}_{-1.9}$	$2.7^{+2.6}_{-2.0}$
$\mu_{ggF HH}$ fixing $\mu_{VBF HH} = 1$	$1.0^{+2.7}_{-1.9}$	$2.8^{+2.7}_{-2.0}$
$\mu_{VBF HH}$ fixing $\mu_{ggF HH} = 1$	$1.0^{+91.3}_{-65.1}$	$10.2^{+97.21}_{-61.6}$

Table 1: Očekivana i opažena jakost signala za ukupni HH signal (ggF HH + VBF HH), ggF HH i VBF HH.

296 Budući da nema naznaka HH signala, 95% C.L. gornje granice na  $\sigma_{ggF HH} \times \text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$   
 297 izvedene su kao funkcija parametra  $\kappa_\lambda$ , kao što je prikazano na lijevom panelu Slike 2.  
 298 Ovisnost gornje granice o  $\kappa_\lambda$  određena je varijacijom distribucije  $\widetilde{M}_x$  signala ggF HH  
 299 koja modificira populaciju kategorija, a time i osjetljivost na taj signal. Konkretno, kate-

300 gorije s visokim  $\widetilde{M}_x$  pružaju veću osjetljivost od kategorija s niskim  $\widetilde{M}_x$  jer imaju manju  
 301 kontaminaciju pozadinom. Za  $\kappa_\lambda$  vrijednosti u intervalu  $[0,6]$ , destruktivna interferen-  
 302 cija između dijagrama proizvodnje ggF HH s "kutijastom" petljom top kvarkova i onog s  
 303 tro-Higgsovim vrhom je maksimalna. To uzrokuje snažnu varijaciju distribucije  $\widetilde{M}_x$  koja  
 304 migrira od najvišeg energetskeg spektra na oko  $\kappa_\lambda = 2$  do najmekšeg spektra na oko  $\kappa_\lambda$   
 305  $= 5$ . Uspoređujući s teorijskim predviđanjem, rezultirajuće granice na parametar  $\kappa_\lambda$  su:

$$\begin{aligned} \text{Opaženo: } & -3,26 < \kappa_\lambda < 8,48 \text{ na } 95\% \text{ C.L.} \\ \text{Očekivano: } & -2,61 < \kappa_\lambda < 8,28 \text{ na } 95\% \text{ C.L.} \end{aligned} \tag{1}$$

306 Isti postupak se provodi za dobivanje granica na parametar  $c_{2V}$ , kao što je vidljivo  
 307 na desnom panelu Slike 2. U ovom slučaju gornja granica je izvedena na  $\sigma_{VBF\ HH} \times$   
 308  $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$  jer je osjetljivost na  $c_{2V}$  u potpunosti ograničena na VBF HH proces.  
 309 Kao i za parametar  $\kappa_\lambda$ , varijacija gornje granice kao funkcija vrijednosti  $c_{2V}$  određena  
 310 je odgovarajućom varijacijom distribucije  $\widetilde{M}_x$ . U ovom slučaju, interferencija između tri  
 311 dijagrama proizvodnje VBF HH čini da spektar distribucije  $\widetilde{M}_x$  migrira na visoke energije  
 312 čim  $c_{2V}$  odstupa od svog SM predviđanja, povećavajući osjetljivost na VBF HH signal.  
 313 Rezultirajuća granice na parametar  $c_{2V}$  su:

$$\begin{aligned} \text{Opaženo: } & -1,31 < c_{2V} < 3,45 \text{ na } 95\% \text{ C.L.} \\ \text{Očekivano: } & -0,96 < c_{2V} < 3,07 \text{ na } 95\% \text{ C.L.} \end{aligned} \tag{2}$$

314 Ovo je prva granica na parametar  $c_{2V}$  postavljena od strane eksperimenta CMS. Kanal  
 315  $HH \rightarrow \gamma\gamma b\bar{b}$  pruža dobru osjetljivost i na parametar  $c_V$ .

316 Ukratko, utvrđeno je da su svi rezultati kompatibilni s predviđanjima SM-a. HH procesu  
 317 nije opažen i postavljen je gornja granica za njegov udarni presjek. Dodatni podaci bit će  
 318 prikupljeni detektorom CMS tijekom Run 3 faze LHC-a, što je ekvivalentno integriranom  
 319 luminozitetu od oko  $300 \text{ fb}^{-1}$ . Ovaj skup podataka će poboljšati osjetljivost na HH signal  
 320 i na parametre vezanja. Preliminarne studije daju projiciranu granicu (očekivani SM) na  
 321 inkluzivni udarni presjek za ggF HH od oko  $3,6 \times \text{SM}$  za kraj Runa 3. Run 3 neće dati  
 322 potrebnu količinu podataka za naznaku HH procesa. Naznaka (SM) HH procesa očekuju  
 323 se tijekom faze visokog luminoziteta LHC-a.

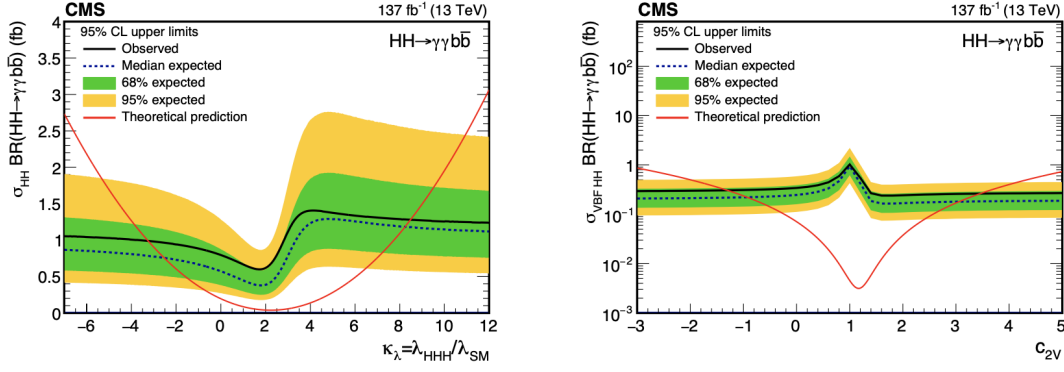


Figure 2: Očekivane i opažene gornje granice na 95% C.L. na udarni presjek za HH produkciju pomnožen s  $BR(HH \rightarrow \gamma\gamma b\bar{b})$  dobivene za različite vrijednosti  $\kappa_\chi$  i  $c_{2V}$  na lijevoj odnosno desnoj strani. Zelena i žuta vrpca predstavljaju odstupanje očekivane granice za jednu odnosno dvije standardne devijacije. Crvene linije prikazuju teorijska predviđanja.

### 324 Rezonantna produkcija tripleta Higgsovih bozona i raspad na šest $b$ kvarkova

325 Kvartično samosprezanje Higgsovog bozona može se izmjeriti iz produkcije tripleta Hig-  
 326 gsovih bozona. Međutim, zbog vrlo malog udarnog presjeka ( $\sim 80$  ab pri  $\sqrt{s} = 13$  TeV),  
 327 potraga za produkcijom tripleta Higgsovih bozona u kontekstu SM-a izvan je dosega LHC-  
 328 a.

329 TRSM ekstenzija SM-a proširuje skalarni sektor SM-a dodatnim skalarnim poljima koja  
 330 se transformiraju kao singleti u odnosu na baždarne grupe SM-a. U TRSM-u udarni  
 331 presjek za gluonsku fuziju  $pp \rightarrow HHH$  pojačan je rezonantnom produkcijom skalara  $X$ .  
 332 Fokusiramo se na scenarij u kojem se stanje  $H$  identificira s Higgsovim bozonom iz SM-a,  
 333 a  $Y$  i  $X$  su dva nova teža skalara koji zadovoljavaju sljedeću hijerarhiju masa

$$2m_H < m_Y < (m_X - m_H). \quad (3)$$

334 To rezultira s dva skalarna singleta  $X$  i  $Y$  koji se raspadaju na Higgsov bozon  $H$  iz  
 335 SM-a. Budući da  $H \rightarrow b\bar{b}$  ima najveći omjer grananja, da bismo povećali statistiku, raz-  
 336 matramo slučaj kad se sva tri Higgsova bozona raspadaju na parove  $b$  kvarkova. Ovo  
 337 daje  $X \rightarrow Y(HH)H \rightarrow b\bar{b}b\bar{b}b\bar{b}$ . Ovi  $b$  kvarkovi se eksperimentalno opažaju kao hadronski  
 338 mlazovi. Ako Higgsov bozon miruje ili ima malu količinu gibanja, dva su  $b$  kvarka iz nje-  
 339 gova raspada dobro odvojena i bit će detektirana kao dva različita mlaza. Međutim, ako  
 340 Higgsov bozon ima veliku količinu gibanja, dva mlaza će biti kolimirana u smjeru gibanja

341 Higgsovog bozona. Ako je količina gibanja Higgsovog bozona dovoljno velika, dva mlaza  
342 će se stopiti u jedan veliki mlaz ("debeli" mlaz). Potraga se provodi u rasponima masa  
343  $X$  (1-4 TeV) i  $Y$  (300-2800 GeV) gdje je  $H$  u ultrarelativističkom režimu. U ovom kine-  
344 matičkom režimu parovi  $b$  kvarkova su dovoljno kolimirani da omoguće rekonstrukciju  $H$   
345 pomoću jednog "debelog" mlaza. Ovdje će kombinacija tri Higgsova bozona dati različite  
346 topologije, od kojih nas zanimaju topologije u kojima su sva tri ili dva od tri Higgsova  
347 bozona u ultrarelativističkom režimu.

348 Glavna pozadina ovog procesa su višestruka produkcija hadronskih mlazova unutar SM-a  
349 te produkcija para top kvarkova. Koristili smo *2D-alphabet* metodu za procjenu oblika i  
350 normalizacije pozadine u signalnom području. Ova analiza je još u tijeku. Skeniranje će  
351 se izvršiti u dvodimenzionalnoj ravnini razapetoj masama skalara  $X$  i  $Y$ .





352 **Contents**

353 **Supervisor Information**

354 **Acknowledgements**

355 **Abstract**

356 **Prošireni sažetak rada**

357	<b>1 Introduction</b>	<b>1</b>
358	1.1 The Standard Model of Elementary Particle Physics . . . . .	5
359	1.1.1 Electromagnetic Interactions . . . . .	8
360	1.1.2 Strong Interactions . . . . .	8
361	1.1.3 Weak Interactions . . . . .	10
362	1.2 The Brout-Englert-Higgs Mechanism . . . . .	11
363	1.3 Higgs Self Coupling . . . . .	14
364	<b>2 The Higgs Boson Phenomenology</b>	<b>16</b>
365	2.1 Higgs Boson Searches in the pre-LHC Era . . . . .	16
366	2.2 Search for Higgs Boson at LHC . . . . .	17
367	2.2.1 Higgs Boson at LHC . . . . .	17

---

368	2.2.2 Higgs Boson Discovery . . . . .	21
369	2.3 Interpretation of LHC Data . . . . .	24
370	2.3.1 Signal Strength $\mu$ . . . . .	24
371	2.3.2 kappa-Framework ( $\kappa$ ) . . . . .	25
372	2.4 HH Production . . . . .	26
373	2.5 Higgs Boson Couplings . . . . .	30
374	2.5.1 Higgs Boson Trilinear Self-Coupling . . . . .	30
375	2.5.2 Yukawa Coupling of the Higgs Boson to the top Quark . . . . .	31
376	2.5.3 Higgs Boson Couplings to the Vector Bosons . . . . .	33
377	2.5.4 BSM Higgs Boson Couplings . . . . .	33
378	2.6 TRSM and HHH Production . . . . .	36
379	<b>3 Experimental Setup</b>	<b>38</b>
380	3.1 The Large Hadron Collider . . . . .	39
381	3.1.1 Accelerator complex . . . . .	40
382	3.1.2 Nominal Design Parameters . . . . .	41
383	3.2 Compact Muon Solenoid . . . . .	45
384	3.2.1 CMS coordinate system . . . . .	47
385	3.2.2 Tracker . . . . .	48
386	3.2.3 Calorimeters . . . . .	50
387	3.2.4 Superconducting Solenoid Magnet . . . . .	53
388	3.2.5 Muon Chambers . . . . .	54
389	3.2.6 Trigger System . . . . .	55

390	3.3 CMS simulation tools . . . . .	59
391	<b>4 Physics Object Reconstruction and Identification</b>	<b>62</b>
392	4.1 Track and Primary Vertex Reconstruction . . . . .	63
393	4.2 Particle Flow . . . . .	65
394	4.3 Photon Reconstruction and Identification . . . . .	69
395	4.3.1 Photon Identification . . . . .	69
396	4.3.2 Photon Energy Regression . . . . .	70
397	4.4 Jet Reconstruction . . . . .	72
398	4.4.1 Jet Energy Corrections (JEC) . . . . .	75
399	4.4.2 Jet Energy Resolution (JER) . . . . .	76
400	4.4.3 Jet $b$ tagging . . . . .	77
401	4.4.4 Boosted Higgs boson jet tagging . . . . .	81
402	<b>5 Search for Non-resonant Higgs Pair Production in <math>b\bar{b}\gamma\gamma</math> Final State</b>	<b>85</b>
403	5.1 Introduction . . . . .	85
404	5.2 Search Strategy . . . . .	86
405	5.3 Data Samples and Simulated Events . . . . .	88
406	5.3.1 Trigger Requirements . . . . .	88
407	5.3.2 Simulated Samples . . . . .	91
408	5.4 Event Selection . . . . .	94
409	5.4.1 Photon Reconstruction . . . . .	94
410	5.4.2 Jet Reconstruction . . . . .	94
411	5.4.3 Energy Regression for b-Jets . . . . .	95

---

412	5.4.4	Selection of the $H \rightarrow \gamma\gamma$ Candidate . . . . .	100
413	5.4.5	Selection of the $H \rightarrow b\bar{b}$ Candidate . . . . .	100
414	5.4.6	Requirements for the VBF HH Topology . . . . .	100
415	5.5	Background Rejection . . . . .	101
416	5.5.1	$t\bar{t}H(\gamma\gamma)$ Background Rejection . . . . .	102
417	5.5.2	Non-resonant Background Rejection . . . . .	106
418	5.6	Event Categorization . . . . .	110
419	5.6.1	VBF HH-enriched Categories . . . . .	111
420	5.6.2	ggF HH-enriched Categories . . . . .	112
421	5.7	Statistical Analysis . . . . .	113
422	5.7.1	Likelihood Definition . . . . .	116
423	5.7.2	Modeling of HH Processes . . . . .	119
424	5.7.3	Background Modeling . . . . .	121
425	5.8	Systematic Uncertainties . . . . .	125
426	5.9	Results . . . . .	128
427	5.9.1	Search for HH Process . . . . .	129
428	5.9.2	Constraints on the BSM Benchmark Hypotheses . . . . .	130
429	5.9.3	Constraints on the Higgs Boson Couplings . . . . .	130
430	5.10	Conclusion . . . . .	140
431	<b>6</b>	<b>Resonant Triple Higgs Production and Decay to Six b-Quarks</b>	<b>142</b>
432	6.1	Introduction . . . . .	142
433	6.2	Search Strategy . . . . .	143

434	6.3 Data Samples and Simulated Events . . . . .	144
435	6.3.1 Data Samples and Trigger Requirements . . . . .	144
436	6.3.2 Simulated Signal Samples . . . . .	145
437	6.3.3 Simulated Background Samples . . . . .	145
438	6.4 Event Selection . . . . .	149
439	6.4.1 Dijet mass vs Trijet mass . . . . .	151
440	6.5 Background Estimation . . . . .	151
441	6.5.1 2D Alphabet . . . . .	151
442	6.5.2 2DAlphabet fits . . . . .	154
443	6.6 Future Goals . . . . .	156
444	<b>7 Conclusion</b>	<b>157</b>
445	<b>Bibliography</b>	<b>158</b>
446	<b>A Search for VHbb</b>	<b>173</b>
447	A.1 Signal and backgroud processes . . . . .	173
448	A.1.1 Signal . . . . .	173
449	A.1.2 Background processes . . . . .	174
450	A.2 Analysis strategy . . . . .	175
451	A.3 Data Samples and simulated events . . . . .	177
452	A.3.1 Data Samples and trigger requirements . . . . .	177
453	A.3.2 Simulated events . . . . .	178
454	A.4 Event selection . . . . .	178

455	A.4.1 Simplified template cross-section (STXS) bins . . . . .	179
456	A.4.2 Resolved analysis selection . . . . .	179
457	A.4.3 Boosted analysis selection . . . . .	183
458	A.5 Multivariate discriminants . . . . .	183
459	A.5.1 DNN . . . . .	184
460	A.6 Results . . . . .	185
461	A.7 Summary . . . . .	189
462	<b>Curriculum vitae</b>	<b>193</b>

# List of Tables

463

464	1	Očekivana i opažena jakost signala za ukupni HH signal (ggF HH + VBF	
465		HH), ggF HH i VBF HH. . . . .	
466	2.1	Coupling values for the twelve defined BSM benchmarks. . . . .	35
467	3.1	Design parameters of the LHC accelerator in proton-proton collisions [54]. .	43
468	4.1	Difference between DEEPCSV and DEEPJET. . . . .	79
469	5.1	List of the simulated ggF HH and VBF HH BSM samples for anomalous	
470		coupling values. The same setup of the corresponding SM sample is used	
471		for the simulation. . . . .	92
472	5.2	List of the simulated SM Higgs boson production samples. . . . .	93
473	5.3	List of the simulated background samples. . . . .	93
474	5.4	Caption . . . . .	94
475	5.5	Summary of the baseline selection criteria. . . . .	100
476	5.6	Optimized BDT-score and $\widetilde{M}_x$ selections for the HH categories. In all the	
477		categories the selection $t\bar{t}H$ -score $> 0.26$ is also applied. . . . .	114
478	5.7	Impact of the systematic uncertainties on the signal strengths in percentage.	126
479	5.8	Expected and observed signal strength for inclusive HH(ggF HH + VBF	
480		HH), ggF HH and VBF HH . . . . .	129



481	5.9	The observed and expected upper limits. . . . .	130
482	5.10	Best fit values for the 1D and 2D likelihood scans of the Higgs coupling	
483		parameters. For the 1D scans the $1\sigma$ uncertainties and the 95% confidence	
484		intervals are also quoted. . . . .	139
485	6.1	HLT Trigger Paths applied . . . . .	145
486	6.2	List of $m_Y$ points for each $m_X$ hypothesis considered in the analysis. . . . .	146
487	6.3	Summary of the Monte Carlo background samples used for 2016. . . . .	147
488	6.4	Summary of the Monte Carlo background samples used for 2017 and 2018. . . . .	148
489	A.1	Definition of the SR and CRs for the resolved selection in the 0-lepton	
490		channel. If the same selection is applied in all SRs and CRs, this is indicated	
491		by the -//- symbol in the latter. The $M_{jj}$ and momenta variables have units	
492		of GeV. . . . .	181
493	A.2	Definition of the SR and CRs for the resolved selection of the 1-lepton	
494		channel. If the same selection is applied in all SRs and CRs, this is indicated	
495		by the -//- symbol in the latter. The $M_{jj}$ and momenta variables have units	
496		of GeV. . . . .	181
497	A.3	Definition of the SR and CRs for the resolved selection in the 2-lepton	
498		channel. If the same selection is applied in all SRs and CRs, this is indicated	
499		by the -//- symbol in the latter. The $M_{jj}$ , $M(V)$ , and momenta variables	
500		have units of GeV. . . . .	182
501	A.4	Selection criteria for the SR and CRs in the boosted topology for 0-, 1-	
502		, and 2-lepton channels. The DeepAK8bbVsLight designation represents	
503		the DEEPAK8 discriminant for the light-quark flavor discrimination node.	
504		The $m_{SD}$ and $M(V)$ variables have units of GeV. . . . .	183
505	A.5	Classes used for the 0/1-lepton multi-DNN classifier. . . . .	184

506	A.6	Predicted and measured values of the product of the cross section and	
507		branching fractions in the V(leptonic)H STXS process scheme. The SM	
508		predictions for each bin are calculated using the inclusive values reported	
509		in Ref. [132]. The uncertainties shown are the combined statistical and	
510		systematic components. . . . .	186
511	A.7	The sources of systematic uncertainty in the inclusive signal strength mea-	
512		surement and their positive and negative values. . . . .	190



# 513 List of Figures

514	1	Shematski prikaz tijekom analize. . . . .	
515	2	Očekivane i opažene gornje granice na 95% C.L. na udarni presjek za HH	
516		produkciju pomnožen s $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$ dobivene za različite vrijednosti	
517		$\kappa_\lambda$ i $c_{2V}$ na lijevoj odnosno desnoj strani. Zelena i žuta vrpca predstavljaju	
518		odstupanje očekivane granice za jednu odnosno dvije standardne devijacije.	
519		Crvene linije prikazuju teorijska predviđanja. . . . .	
520	1.1	Diagram representing the elementary particles of the Standard Model.	
521		Matter is constituted by three generations of quarks (in purple) and leptons	
522		(in green), while the interactions amongst them are governed by the gauge	
523		bosons (in red). The Higgs boson (in yellow) is responsible for the masses	
524		of the particles. . . . .	6
525	1.2	Higgs interactions. . . . .	15
526	1.3	Possible interpretation of the Higgs field in SM and BSM. . . . .	15
527	2.1	Left: Leading order Feynman diagram of ggH process, middle: VBF Higgs	
528		production and right: Higgsstrahlung (VH) process of Higgs boson in as-	
529		sociation with a vector boson. . . . .	18
530	2.2	Representative tree-level Feynman diagrams of the $t\bar{t}H$ production. . . . .	19
531	2.3	Representative tree-level Feynman diagrams of the tH process. Diagram	
532		on the left contains $c_V$ and the right one has $y_t$ . . . . .	20

533	2.4	Left - Production cross section of a SM Higgs boson of mass $m_H = 125$ GeV as a function of center-of-mass energy $\sqrt{s}$ . Right - Production cross section of a SM Higgs boson at $\sqrt{s} = 14$ TeV as a function of Higgs Mass. The lines with different colors correspond to the different production modes with a certain order of accuracy, while the band across each lines give the uncertainty of the calculation. Figure taken from [29]. . . . .	20
534			
535			
536			
537			
538			
539	2.5	Branching ratios of the Higgs boson in the SM as a function of the mass. Figure taken from [29]. . . . .	21
540			
541	2.6	Left: The diphoton invariant mass distribution weighted by the $S/(S + B)$ value of its category, the peak around 125 GeV from the SM the Higgs boson contribution on top of the continuum diphoton background is shown with red solid line. Right: Distribution of the four-lepton invariant mass for the $ZZ \rightarrow 4l$ analysis, the Higgs boson with a mass of $m_H$ 125 GeV has been shown in red solid line sitting on the background. The figure is taken from [2]. . . . .	23
542			
543			
544			
545			
546			
547			
548	2.7	Feynman diagrams for the SM gluon-gluon fusion di-Higgs production (left) triangle diagram (right) box diagram. . . . .	27
549			
550	2.8	Higgs pair invariant mass distribution at LO for the different contributions (box and triangle) to the ggF HH production mechanism and their interference. Figure taken from [34] . . . . .	27
551			
552			
553	2.9	Feynman diagrams of the processes contributing to the production of Higgs boson pairs via BSM ggF at LO. The diagram on the left involves the contact interaction of two Higgs bosons with two top quarks ( $c_2$ ), the middle diagram shows the contact interaction between the Higgs bosons and two gluons ( $c_{2g}$ ), and the diagram on the right describes the contact interactions between the Higgs boson and gluons ( $c_g$ ). . . . .	28
554			
555			
556			
557			
558			
559	2.10	Feynman diagrams that contribute to the production of Higgs boson pairs via VBF at LO. On the left the diagram involving the HHH vertex ( $\lambda_{HHH}$ ), in the middle the diagram with two HVV vertices ( $c_V$ ), and on the right the diagram with the HHVV vertex ( $c_{2V}$ ). . . . .	29
560			
561			
562			

563	2.11 SM di-Higgs decay branching ratio (BR) for $m_H = 125$ GeV . . . . .	30
564	2.12 <i>Left</i> : cross sections of the main HH production modes as a function of	
565	$\kappa_\lambda$ . <i>Right</i> : mHH distributions for the ggF HH process for different $\kappa_\lambda$	
566	hypotheses. The distributions are all normalized to unity. . . . .	31
567	2.13 Variations of the ggF HH cross sections as a function of $\kappa_t$ . . . . .	32
568	2.14 Variations of the VBF HH cross section as a function of $c_{2V}$ . . . . .	34
569	2.15 Generator-level distributions of di-Higgs boson mass for the clustered bench-	
570	marks from [46] are shown. The red distributions correspond to the chosen	
571	benchmark sample in each cluster, while the blue ones describe the other	
572	members of each cluster. . . . .	36
573	2.16 Feynman diagram showing the gluon-gluon production mode of a heavy	
574	scalar $X$ followed by its decay process $X \rightarrow Y(HH)H$ . . . . .	37
575	3.1 Illustration of the accelerator complex at CERN. Protons are accelerated to	
576	increasing energies at the LINAC 2 (in light pink), Booster (in light pink),	
577	PS (in dark pink), SPS (in light blue) and LHC (in grey) accelerators. The	
578	counter-circulating proton beams at the LHC collide in the centre of the	
579	CMS, ATLAS, LHCb and ALICE detectors [48]. . . . .	41
580	3.2 Cross section of an LHC dipole magnet [49]. . . . .	42
581	3.3 Average number of simultaneous pp interactions per bunch crossing (left)	
582	and integrated luminosity collected by the CMS experiment (right) by year	
583	of data taking [56]. . . . .	44
584	3.4 Schematic view of the CMS detector and its subcomponents [57] . . . . .	46
585	3.5 An illustration of the CMS detector with spherical co-ordinate system [58].	47
586	3.6 Sketch of one quarter of the Phase-1 CMS tracking system in r-z view. The	
587	pixel detector is shown in green, while single-sided and double-sided strip	
588	modules are depicted as red and blue segments, respectively. [60]. . . . .	49

---

589	3.7	Schematic view of the ECAL showing the cylindrical barrel closed by the	
590		two endcap regions with one half endcap displayed [64]. . . . .	51
591	3.8	A schematic view of one quarter of the CMS HCAL, showing the positions	
592		of its four major components: the hadron barrel (HB), the hadron endcap	
593		(HE), the hadron outer (HO), and the hadron forward (HF) calorimeters	
594		[66]. . . . .	52
595	3.9	Schematic view, in the r-z plane, of one quadrant of the CMS detector,	
596		with the axis parallel to the beam (z) running horizontally and the radius	
597		(r) increasing upward. The interaction region is at the lower left corner.	
598		The position of the present RPC chambers is shown in blue. The RPCs	
599		are both in the barrel and in the endcaps of CMS. The DT chambers are	
600		labeled MB and the CSC chambers are labeled ME. The steel disks are	
601		displayed as dark gray areas. [67]. . . . .	55
602	3.10	Flowchart of CMS trigger system. Image reproduced from Ref. [68]. . . . .	56
603	3.11	Schematic layout of L1 trigger system of the CMS experiment [69]. . . . .	58
604	4.1	Transverse slice of the CMS detector, showing the experimental signatures	
605		of the different final-state particles [79]. . . . .	63
606	4.2	A collision recorded by the CMS detector during the 2016 data taking [81].	
607		The lines correspond to the reconstructed tracks, while the dots represent	
608		the reconstructed interaction vertices. . . . .	64
609	4.3	The particle-flow algorithm combines information from various subdetec-	
610		tors to provide a global description of the event in terms of electrons, pho-	
611		tons, muons, charged hadrons, and neutral hadrons. . . . .	66
612	4.4	Performance of the photon identification based on cut-based and MVA	
613		based approach. The three points for the cut-based method refer to the	
614		three different working points: loose, medium, and tight [85]. . . . .	71
615	4.5	Reconstructed $Z \rightarrow e^+e^-$ invariant mass distribution before and after the	
616		energy regression correction for the barrel (left) and endcap (right). [85]. .	72

617	4.6	JER scale factor as a function of absolute value of pseudorapidity ( $ \eta $ ) [90].	76
618	4.7	An illustration of b-hadron decay and corresponding impact parameter (IP)	
619		[91]. . . . .	77
620	4.8	DEEPCSV and DEEPJET performance curves for AK4 jets [95]. . . . .	80
621	4.9	Performance comparison in terms of ROC curves on the Higgs boson taggers	
622		[98]. . . . .	83
623	5.1	The distribution of $M(\gamma\gamma)$ (left) and $M(jj)$ (right) for signal and back-	
624		ground events. All contributions are normalized to a unitary area. . . . .	87
625	5.2	(left) Distribution of $\widetilde{M}_x$ for some benchmark signal datasets, and for the	
626		resonant and non-resonant backgrounds. (right) Distribution of $\widetilde{M}_x$ for	
627		SM-like HH production and for the different background components. All	
628		contributions are normalized to a unit area. . . . .	88
629	5.3	Scheme of the analysis workflow. . . . .	89
630	5.4	Re-weighting factors in $(R_9, \eta)$ for $Z \rightarrow e^+e^-$ selected events with respect	
631		to $H \rightarrow \gamma\gamma$ events. . . . .	90
632	5.5	Diphoton trigger efficiency measured on 2018 data for $Z \rightarrow e^+e^-$ events	
633		using tag-and-probe method. . . . .	91
634	5.6	Mass resolution in this analysis for SM ggF HH samples merging all years	
635		weighted by the luminosity: without regression (grey), with L1 regression	
636		(red), and with L1+L2 regression (blue). . . . .	96
637	5.7	The $m_{jj}$ distributions in the data, the ZZ signal, and the DY and $t\bar{t}$ +jets	
638		backgrounds before BDT selection for the $e^+e^-$ (left) and the $\mu^+\mu^-$ (right)	
639		channels. . . . .	97
640	5.8	The signal and background efficiencies as a function of the BDT score	
641		(upper left), ROC curve (upper right) and the overtraining check (lower). . .	98



642	5.9	The $m_{jj}$ distributions in the data, the ZZ signal, and the DY and $t\bar{t}$ +jets	
643		backgrounds with BDT $> 0.2$ selection for the $e^+e^-$ (left) and the $\mu^+\mu^-$	
644		(right) channels. . . . .	98
645	5.10	Data and MC events with a Crystal-Ball. TOP - $E_T^{miss} > 40$ (effect observed	
646		and expected), BOTTOM - $E_T^{miss} < 40$ (no effect observed and expected).	
647		Left - with L1 regression, right - with L1+L2 regression. Results for 2017	
648		data sample. . . . .	99
649	5.11	Distributions of $m_{\gamma\gamma}$ (left) and $m_{jj}$ (right) for the selected HH candidates	
650		for data and for the simulated single and double Higgs processes. . . . .	101
651	5.12	Angular variables used in the training, from left to right: $\Delta R_{min}(\gamma, b-jet)$ ,	
652		$ \cos\theta_{HH}^{CS} $ and $ \cos\theta_{jj} $ . These variables are used for the training of the $t\bar{t}H$	
653		discriminant. . . . .	103
654	5.13	Major variables used in the training to reject events with a leptonic-decay	
655		W boson, from left to right top to bottem: $E_T^{miss}$ , $\Delta\phi(p_T^{miss}, b-jet1)$ ,	
656		$\Delta\phi(p_T^{miss}, b-jet2)$ and the transverse momentum of the leading and sub-	
657		leading electrons and muons. . . . .	104
658	5.14	$\chi_{top}^2$ variables in training to reject events with a hadronic decay W boson,	
659		for events with at least 2 additional jets (left plot) and 4 additional jets	
660		(right plot) besides the two b-tagged jets. . . . .	104
661	5.15	<i>Left</i> : $t\bar{t}H$ -score distribution for all the main background components from	
662		the MC simulations of the 2016 and 2017 events. <i>Right</i> : Comparison be-	
663		tween data and MC simulation of the $t\bar{t}H$ -score distribution for the HH	
664		candidates with $t\bar{t}H$ -score $> 0.26$ . . . . .	105
665	5.16	Distribution of $ \cos\theta_{HH}^{CS} $ (left), $p_T^{\gamma\gamma}/m_{\gamma\gamma jj}$ (center), and the b-tag score for	
666		the leading jet (right) of the selected HH candidates for the data as well	
667		as the simulated $t\bar{t}+X$ , $\gamma\gamma$ +jets, $\gamma$ +jets, and multijet events excluding the	
668		signal region $120 < m_{\gamma\gamma} < 130$ GeV. The distribution of the simulated	
669		$HH \rightarrow b\bar{b}\gamma\gamma$ events is also shown with a red line. . . . .	107

670	5.17	The distribution of the MVA output to discriminate the ggF HH signal	
671		from the continuum background for the selected data and simulated events	
672		in the ggF HH signal region (including the selection MVA score $> 0.37$ ).	. 108
673	5.18	Distribution of $m_{jj}^{VBF}$ (left), the quark-gluon likelihood (center), and the	
674		$C_{bb}$ centrality variable for the selected VBF HH candidates for the data as	
675		well as the simulated $t\bar{t}+X$ , $\gamma\gamma$ +jets, $\gamma$ +jets, and multijet events, excluding	
676		the signal region $120 < m_{\gamma\gamma} < 130$ GeV. The distributions of the simulated	
677		ggF HH and VBF HH events are also shown with red and purple lines,	
678		respectively. . . . .	. 110
679	5.19	MVA multiclassifier output relative to the VBF HH class for the data events	
680		as well as for the VBF HH and single Higgs simulated events, for the $\widetilde{M}_x$	
681		$< 500$ GeV and $\widetilde{M}_x > 500$ GeV on the left and right, respectively. . . . .	. 110
682	5.20	$\widetilde{M}_x$ distribution for the VBF HH simulated events for $c_{2V} = 1$ and $c_{2V} =$	
683		0 in red and blue, respectively. . . . .	. 111
684	5.21	$\widetilde{M}_x$ distribution for the SM ggF HH and the main background MC events.	
685		All the distributions are normalized to one. . . . .	. 113
686	5.22	Visualized categorization scheme for the VBF HH (left) and ggF HH (right)	
687		analysis. . . . .	. 114
688	5.23	<i>Left:</i> Expected categories composition in terms Higgs boson processes.	
689		<i>Right:</i> Expected $S/(S + B)$ in $\pm 1\sigma_{eff}$ for each category. $S$ is referred to	
690		the Higgs boson process target of each category, and $B$ is the sum of the	
691		expected Higgs boson background processes and of the expected continuum	
692		background. . . . .	. 115

693	5.24 $m_{\gamma\gamma}$ modeling for the ggF HH and VBF HH process on the left and right,	
694	respectively, for the best resolution (high $\widetilde{M}_x$ and high BDT score) category	
695	in the 2018 dataset. The open squares represent simulated events and the	
696	blue lines are the corresponding models. Also shown are the $\sigma_{eff}$ value	
697	(half the width of the narrowest interval containing 68.3% of the invariant	
698	mass distribution) and the corresponding interval as a gray band, and the	
699	full width at half the maximum (FWHM) and the corresponding interval	
700	as a double arrow. . . . .	120
701	5.25 $m_{jj}$ modeling for the ggF HH and VBF HH process on the left and right,	
702	respectively, for the best resolution (high $\widetilde{M}_x$ and high BDT score) category	
703	in the 2018 dataset. . . . .	121
704	5.26 Two dimensional ( $m_{jj}, m_{\gamma\gamma}$ ) distribution for the selected HH candidates of	
705	the simulated ggF HH events using the 2018 dataset. . . . .	121
706	5.27 $m_{jj}$ modeling for the ggH (top-left), qqH (top-right), VH (bottom-left),	
707	$t\bar{t}H$ (bottom-center), and tHq (bottom-right) processes, in the ggF HH-	
708	enriched category with the highest BDT score value. . . . .	123
709	5.28 The $m_{\gamma\gamma}$ (upper row) and $m_{jj}$ (bottom row) distribution for the selected	
710	events in data (black points) is shown for the two VBF HH categories with	
711	the curves corresponding to the signal + background fit (solid red), the	
712	single Higgs boson and the non-resonant processes H+B (solid blue) and	
713	the background only (dashed black), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$	
714	uncertainties in the fitted background. . . . .	131
715	5.29 The $m_{\gamma\gamma}$ distribution for the selected events in data (black points) is shown	
716	for the twelve ggF HH categories with the curves corresponding to the signal	
717	+ background fit (solid red) and the background only (dashed red), with	
718	bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background. .	132
719	5.30 The $m_{jj}$ distribution for the selected events in data (black points) is shown	
720	for the twelve ggF HH categories with the curves corresponding to the	
721	signal + background fit (solid red) and the background only (dashed red),	
722	with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.	133

723	5.31	The $m_{\gamma\gamma}$ and $m_{jj}$ distribution for the selected events in data (black points)	
724		weighted by $S/(S+B)$ with the curves corresponding to the signal + back-	
725		ground fit (solid red), the single Higgs boson and the non-resonant processes	
726		H+B (solid blue) and the background only (dashed black), with bands cov-	
727		ering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background. . . . .	133
728	5.32	Expected and observed two dimensional likelihood scan of the $\mu_{ggF HH}$ (y-	
729		axis) and $\mu_{VBF HH}$ (x-axis) parameters, on the left and on the right side,	
730		respectively. The exclusion contours are also shown for the 68% (solid black	
731		line) and 95% (dashed black line) confidence levels. . . . .	134
732	5.33	Expected and observed 95% CL upper limits on the $\sigma_{ggF HH} \times \text{BR}(HH \rightarrow$	
733		$\gamma\gamma b\bar{b})$ for the BSM benchmark models, shown with transparent and solid	
734		circles, respectively. The green and yellow bands represent the one and two	
735		standard deviation extensions beyond the expected limit, respectively. . . .	134
736	5.34	Expected and observed 95% CL upper limits on the SM-like VBF HH	
737		production cross section times $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$ obtained for different	
738		values of $\kappa_\lambda$ and $c_{2V}$ on the left and right side, respectively. The green	
739		and yellow bands represent the one and two standard deviation extensions	
740		beyond the expected limit, respectively. The red lines show the theoretical	
741		predictions. . . . .	136
742	5.35	Expected and observed likelihood-scan of the $\kappa_\lambda$ parameter, on the left and	
743		right sides, respectively. The likelihood scan is shown including the $t\bar{t}H$	
744		categories (orange line) or using only the HH categories (blue line). The	
745		likelihood values corresponding to the 68% and 95% confidence levels are	
746		represented by the lower and upper horizontal dashed grey lines, respectively.	137
747	5.36	Expected and observed likelihood-scan of the $\kappa_t$ parameter, on the left and	
748		right sides, respectively. The likelihood scan is shown including the $t\bar{t}H$	
749		categories (orange line) or using only the HH categories (blue line). The	
750		likelihood values corresponding to the 68% and 95% confidence levels are	
751		represented by the lower and upper horizontal dashed grey lines, respectively.	137

752	5.37	Expected and observed likelihood-scan of the $c_{2V}$ parameter, on the left	
753		and right sides, respectively. The likelihood values corresponding to the	
754		68% and 95% confidence levels are represented by the lower and upper	
755		horizontal dashed grey lines, respectively. . . . .	138
756	5.38	Expected (left) and observed (right) two dimensional likelihood-scans of the	
757		$(\kappa_\lambda, \kappa_t)$ parameters including the $t\bar{t}H$ categories (orange) and using only	
758		the HH categories (blue). The exclusion regions at 68% and 95% confidence	
759		level are represented with the solid and the dashed lines, respectively. The	
760		region where the $\sigma_{t\bar{t}H}(\kappa_\lambda, \kappa_t)$ parametrization is not reliable is highlighted	
761		in gray. . . . .	139
762	5.39	Expected (left) and observed (right) two dimensional likelihood-scans of the	
763		$(\kappa_\lambda, c_{2V})$ parameters. The exclusion regions at 68% and 95% confidence	
764		level are represented with the solid and the dashed lines, respectively. . . .	140
765	6.1	Fraction of generated boosted Higgs boson candidates in $(m_X, m_Y)$ plane.	144
766	6.2	Simulated Signal samples for 2017. Blue circles are centrally produced	
767		samples at CMS from which green circled ones are used for this analysis. .	146
768	6.3	Event selection criteria for boosted and semi-boosted topologies. . . . .	150
769	6.4	Cut-flow for boosted and semi-boosted signal region with pass and fail	
770		categories. . . . .	150
771	6.5	Efficiency map for boosted and semi-boosted signal region pass category. .	150
772	6.6	Trijet mass (left) and dijet mass (right) distribution for boosted (upper)	
773		and semi-boosted (lower) topologies. . . . .	152
774	6.7	Illustration of 2D alphabet method. . . . .	153
775	6.8	2D Alphabet background only fit for fail (upper) and pass (lower) categories	
776		in validation region for boosted topology. . . . .	154
777	6.9	2DAlphabet background only fit for fail (upper) and pass (lower) categories	
778		in validation region for semi-boosted topology. . . . .	155

779	6.10 2D Alphabet background only fit for fail (upper) and pass (lower, using toy	
780	data) categories in validation region for boosted topology. . . . .	155
781	6.11 2DAlphabet background only fit for fail (upper) and pass (lower, using toy	
782	data) categories in validation region for semi-boosted topology. . . . .	156
783	A.1 VH (top left) and ggZH (top right and bottom) process Feynman diagrams.	174
784	A.2 Simplified schematic of the analysis strategy using a simultaneous fit of	
785	templates derived from MC simulations. . . . .	176
786	A.3 Overview of the STXS bins for the three VH production modes [130]. The	
787	vertical axis reflects the $p_T(V)$ bin ranges and the horizontal axis the num-	
788	ber of additional jets. . . . .	180
789	A.4 Left: Resolved 0-lepton and 1-lepton channels selection scheme. Right:	
790	Resolved 2-lepton channel selection scheme. . . . .	182
791	A.5 Left: Boosted 0-lepton and 1-lepton channels selection scheme. Right:	
792	Boosted 2-lepton channel selection scheme. . . . .	184
793	A.6 Distribution of the HFDNN scores in the 0-lepton (left) and 1-lepton,	
794	(right) $Z + b$ and $W + b$ heavy-flavor CRs for the 2016 dataset, after the fit	
795	to data. The output nodes target enrichment in the V+light-quark (first	
796	bin), V+c (second bin), V+b (third bin), $V + b\bar{b}$ (fourth bin), single top	
797	quark (fifth bin), and $t\bar{t}$ (sixth bin) backgrounds. The lower plots display	
798	the ratio of the data to the MC expectations. The vertical bars on the	
799	points represent the statistical uncertainty in the ratio, and the hatched	
800	area shows the MC uncertainty. . . . .	185

- 801 A.7 Signal strengths (points) for the 0-, 1-, and 2-lepton channels (left) and the  
802  $ZH$  and  $WH$  production modes (right). The horizontal red and blue bars  
803 on the points represent the systematic and total uncertainties, respectively.  
804 The combined inclusive signal strength is shown by the vertical line, with  
805 the green band giving the 68% confidence interval. The results combine the  
806 2016–2018 data-taking years. The first and the second uncertainty values  
807 correspond to the statistical and systematic uncertainties, respectively. . . . 187
- 808 A.8 STXS signal strengths from the analysis of the 2016–2018 data. The ver-  
809 tical dashed line corresponds to the SM value of the signal strength. The  
810 first and the second uncertainty values correspond to the statistical and  
811 systematic uncertainties, respectively. . . . . 187
- 812 A.9 Measured values of  $\sigma\mathcal{B}$ , defined as the product of the  $VH$  production cross  
813 sections multiplied by the branching fractions of  $V \rightarrow$  leptons and  $H \rightarrow b\bar{b}$ ,  
814 evaluated in the same STXS bins as for the signal strengths, combining all  
815 years. In the lower panel, the ratio of the observed results, with associated  
816 uncertainties, to the SM expectations is shown. If the observed signal  
817 strength for a given STXS bin is negative, no value is plotted for  $\sigma\mathcal{B}$  in the  
818 upper panel. . . . . 188
- 819 A.10 Post-fit distributions of the DNN discriminant in the  $250 < p_{\text{T}}(V) <$   
820  $400\text{GeV}$  category of the 0-lepton (top left), 1-lepton (top right) and 2-  
821 lepton (bottom) channels for the electron final state using the 2018 data  
822 set. The background contributions after the maximum likelihood fit are  
823 shown as filled histograms. The Higgs boson signal is also shown as a filled  
824 histogram, and is normalized to the signal strength shown in Fig. A.8.  
825 The hatched band indicates the combined statistical and systematic un-  
826 certainty in the sum of the signal and background templates. The ratio of  
827 the data to the sum of the fitted signal and background is shown in the  
828 lower panel. The distributions that enter the maximum likelihood fit use  
829 the same binning as shown here. . . . . 189

830 A.11 Distributions of signal, background, and observed data event yields sorted  
831 into bins of similar signal-to-background ratio, as given by the result of the  
832 fit to the multivariate discriminants in the resolved and boosted categories.  
833 All events in the signal regions of the 2016–2018 data set are included. The  
834 red histogram indicates the Higgs boson signal assuming SM yields ( $\mu = 1$ )  
835 and the sum of all backgrounds is given by the gray histogram. The lower  
836 panel shows the ratio of the observed data to the background expectation,  
837 with the total uncertainty in the background prediction indicated by the  
838 gray hatching. The red line indicates the sum of signal assuming the SM  
839 prediction plus background contribution, divided by the background. . . . 190





# Chapter 1

## Introduction

The universe has always been like a mystery box for us. The deeper we dig into it, our curiosity gets increased. Humans have been attempting to unravel this mystery for many ages. The foremost curiosity is the science behind the origin of the universe that could explain all the observed physics phenomena.

After studying the results from many experiments, scientists have constructed a particle theory named the Standard Model (SM) that describes elementary constituents of the universe and their interactions. There is no doubt that SM is the most successful theory so far after the discovery of Higgs particle by the two largest collaborations ATLAS [1] and CMS [2, 3], at CERN. However, SM only explains about 5% of the universe; the remaining 95% indicates that SM might be an effective theory that shows a low energy signature of physics existing at a high energy scale [4]. It does not explain many physics observations, e.g., the gravitational force does not fit within it. However, this is not the only drawback. The SM also does not provide any explanation for dark matter and dark energy [5, 6, 7], hierarchy problem of the electroweak SM [8], neutrino masses [9], amongst others.

The above drawbacks prove that even after being on the right track to unravel the universe, we are still far from the fundamental science behind it. This is why physicists are developing new theories that could reproduce the SM observations with providing an explanation of the above mysterious phenomena. These are known as beyond the Standard Model (BSM) theories. It comprises super-symmetry (SUSY) [10], warped extra dimen-

862 sion (WED) [11] and many other models. However, the question is where to look for new  
863 physics? One possibility is that we use the Higgs boson as a stairway to reach the door  
864 of new physics.

865 After discovering Higgs boson at the Large Hadron Collider (LHC) [12], the SM has  
866 become a complete theory. The couplings of Higgs boson with the SM gauge bosons  
867 and fermions (known as Yukawa couplings) have also been measured within a certain  
868 uncertainty by studying various production and decay modes. Considering these uncer-  
869 tainties, we still have experimentally allowed phase space supporting the existence of the  
870 BSM physics as extension of SM. Hence, SM precision measurements are essential in this  
871 prospect.

872 Following direct and indirect search techniques, colliders can play a crucial role in the  
873 search for the new physics (might appear at a very high energy scale) and SM precision  
874 measurements (any observed deviation from the SM predictions will be an indication for  
875 the new physics). Thus, the goals of LHC include new physics searches at high energy  
876 scales and SM precision measurements, including studies related to the self-interactions  
877 of the Higgs boson as the SM predicts that Higgs boson interacts with itself via trilinear  
878 and quartic self-coupling. However, these self couplings are yet to be determined. Ex-  
879 perimentally, the trilinear coupling can be directly measured using the Higgs boson pair  
880 production mode  $pp \rightarrow HH$ , also known as **Di-Higgs production** and quartic coupling  
881 can be measured using **Triple Higgs production**.

882 Within SM, non-resonant production is the only process for multiple-Higgs production,  
883 while the resonant multiple-Higgs production has its own importance to look for the new  
884 BSM particles. Briefly, we can understand the importance of multiple-Higgs production  
885 in two ways:

- 886 1. Non-resonant multiple-Higgs production: It is a direct probe for the SM Higgs  
887 trilinear self-coupling. This approach is also suitable for BSM effective field theory  
888 (EFT) searches where resonance might appear at the large TeV scale, but we look  
889 for its low energy signatures.
- 890 2. Resonant multiple-Higgs production: There are BSM theories that provide the so-  
891 lutions to the SM inconsistencies like hierarchy problems, dark matter, etc. The

---

892 predicted resonances by these BSM models with an enhanced cross section directly  
893 couple to Higgs boson, which might be easier to probe using the direct search meth-  
894 ods.

895 We have explored the non-resonant di-Higgs production mode with the decay channel, one  
896 Higgs boson decays into a pair of the bottom quark, and another one decays into a pair of  
897 photon resulting in  $b\bar{b}\gamma\gamma$  final state. The  $H \rightarrow b\bar{b}$  has the large branching fraction among  
898 all Higgs boson decay modes, but the SM multi-jet backgrounds make it a challenging  
899 final state to perform any study. On the other side, despite of low  $H \rightarrow \gamma\gamma$  branching  
900 ratio, it has very low background contamination. Thus,  $HH \rightarrow b\bar{b}\gamma\gamma$  keeps benefit of high  
901 purity and selection efficiency.

902 And for the resonant triple-Higgs production mode, the cross section is very small that  
903 it is better to have large branching ratio and thus, the decay channel,  $HHH \rightarrow b\bar{b}b\bar{b}b\bar{b}$  is  
904 preferred. This search is motivated by the BSM theories such as Two-real-scalar-singlet  
905 extension of the SM (TRSM) [13] which provide explanation for some of the shortcomings  
906 of the SM and, among others, postulate additional scalar particles.

907 This thesis work focuses on the searches using both the di-Higgs production and triple-  
908 Higgs production. We can divide this research work into two parts,

- 909 1. non-resonant di-Higgs study in  $b\bar{b}\gamma\gamma$  final state with proton-proton collision data.
- 910 2. resonant triple-Higgs study in boosted  $b\bar{b}b\bar{b}b\bar{b}$  final state.

911 Two primary production modes, vector boson fusion (VBF) and gluon-gluon fusion (ggF),  
912 are investigated in non-resonant di-Higgs research. SM Higgs couplings such as the cou-  
913 pling between the Higgs boson and the top quark ( $y_t$ ) and the trilinear Higgs self-coupling  
914  $\lambda_{HHH}$  are accessible through both production modes. Simultaneously, considering pro-  
915 duction modes modified by BSM allows access to EFT couplings. Study of the production  
916 of Higgs boson pairs via BSM ggF gives access to two Higgs bosons and two gluons ( $c_{2g}$ ),  
917 between one Higgs boson and two gluons ( $c_g$ ), and between two Higgs bosons and two  
918 top quarks ( $c_2$ ). Additionally, via BSM VBF gives access to the coupling of Higgs bosons  
919 pair with vector bosons pair,  $C_{2V} \sim HHVV$  and the coupling of a single H with a pair of  
920 vector bosons  $C_V \sim HVV$ . The couplings are further explained with Feynman diagrams in

921 section 2.4. When required,  $t\bar{t}H$  production results were incorporated with ggF HH and  
922 VBF HH data in order to increase the sensitivity of  $y_t$  coupling at the end.

923 The search for di-Higgs production has two major obstacles: a small cross-section and  
924 irreducible background. To separate signal events from background events, MVAs were  
925 trained and the signal was divided into different categories to increase sensitivity. Despite  
926 this, no significant excesses over the background of double Higgs production events were  
927 found, thus upper limits on the HH cross sections were extracted. The observed upper  
928 limit on the inclusive HH production cross section is 7.7 times SM and corresponds to  
929 the most stringent result achieved by the CMS experiment to date [14]. In addition,  
930 constraints on Higgs coupling parameters were put with 95% confidence level, which  
931 is the most stringent constraint among the published results. Current results show no  
932 significant deviation from SM. An evidence of a SM HH process is expected during the  
933 high-luminosity phase of the LHC. The non-resonant HH analysis was published in 2021:

- 934 • CMS Collaboration, **Search for non-resonant Higgs boson pair production**  
935 **in final states with two bottom quarks and two photons in proton-proton**  
936 **collisions at  $\sqrt{s} = 13$  TeV**, JHEP 03 (2021) 257, DOI [10.1007 / JHEP03\(2021\)257](https://doi.org/10.1007/JHEP03(2021)257)

937 The triple-Higgs production analysis is still under process. The event selection and  
938 boosted categories are defined for the selected channel. The 2D-Alphabet method is  
939 used to model background [15, 16]. The next steps are to define systematic uncertainties  
940 for theory model and experimental setup and extract the signal or extract upper limit for  
941 the cross-section.

942 This thesis is structured as follows. Chapter 1 aims to introduce the theory of elemen-  
943 tary particle physics, focusing on its aspects relevant to the Higgs boson sector as it is  
944 directly related to the search for a signature of new physics presented in this thesis. The  
945 Higgs boson at LHC and the phenomenology of multiple-Higgs production is described in  
946 Chapter 2. Chapter 3 convey the details of the collider and experiment that provide us  
947 data used for the research work. Chapter 4 explains how to identify and measure particles  
948 present in the event using the signals left in the detector. Chapter 5 contains the first  
949 part of research work where non-resonant HH production in  $b\bar{b}\gamma\gamma$  final state is described  
950 with the results. Chapter 6 details the second part of the research work, the search for

951 resonant triple-Higgs production in boosted  $6b$  final state. In Chapter 7, we conclude with  
952 a detailed summary of the main results from both the analyses performed in Chapters 5  
953 and 6. The Appendix A describes additional work I have done apart from multiple-Higgs  
954 analysis. It includes my work on Higgs bosons decaying into a  $b\bar{b}$  quark pair, produced  
955 in association with a vector boson (VHbb analysis) which is not part of the main thesis.  
956 This analysis was published in 2024:

- 957 • CMS Collaboration, **Measurement of simplified template cross sections of**  
958 **the Higgs boson produced in association with W or Z bosons in the**  
959  **$H \rightarrow b\bar{b}$  decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV**, Phys.  
960 Rev. D 109 (2024) 092011, DOI [10.1103/PhysRevD.109.092011](https://doi.org/10.1103/PhysRevD.109.092011)

## 961 1.1 The Standard Model of Elementary Particle 962 Physics

963 The SM of particle physics was developed throughout the second half of the 20th century  
964 and provides a description of the elementary particles and their fundamental interactions  
965 [17]. Its theoretical framework is built upon the mathematical foundations of quantum  
966 field theory (QFT) and gauge symmetries, refined by the constant back and forth between  
967 theory and experiment. It is well corroborated by the experimental observations, and its  
968 predictive power was further consolidated with the discovery of the Higgs boson by the  
969 ATLAS and CMS experiments at the LHC on the 4th July 2012 [1, 2], almost half a  
970 century after it was postulated.

971 The SM is the name given to a theory of fundamental particles back in the 1970s. It  
972 incorporated all that was known about subatomic particles at the time and predicted the  
973 existence of additional particles as well. Two types of particles are included in the SM: the  
974 building blocks of matter, also known as *matter* particles, and the intermediate interaction  
975 particles, or *force carriers*. The first group is composed of *fermions*, whereas the second  
976 group is composed by *bosons*, which are the particles exchanged by the fermions during  
977 interactions.

978 Every elementary particle in the SM is characterized by a few quantum numbers which are  
 979 conserved in the fundamental interactions. These are unique invariant masses, an electric  
 980 charge (in the units of  $e$ ), and a spin quantum number, which is equal to half integral  
 981 ( $\frac{1}{2}$ ,  $\frac{3}{2}$ ,  $\frac{5}{2}$ , etc.) for fermions, whereas a whole integer (0, 1, 2, etc.) for bosons. The  
 982 modern-day visualization of the SM, where all the fundamental particles are strategically  
 983 placed, according to their designated roles in the nature, is shown in Fig. 1.1.

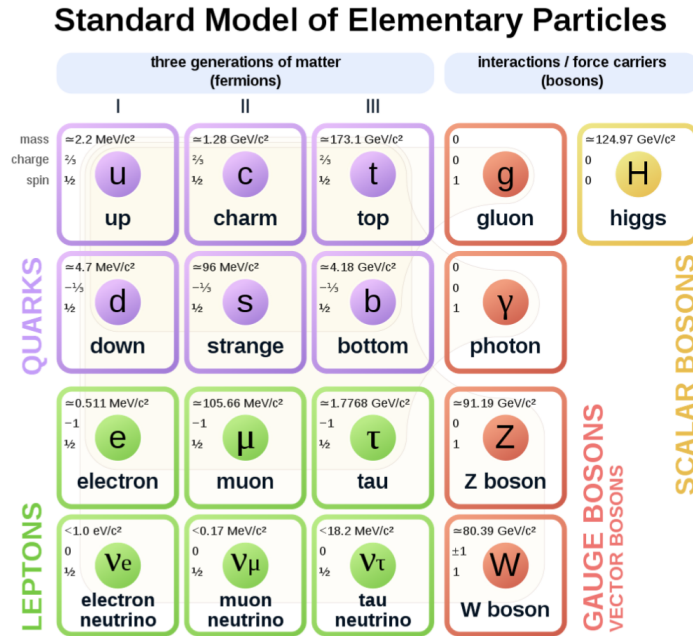


Figure 1.1: Diagram representing the elementary particles of the Standard Model. Matter is constituted by three generations of quarks (in purple) and leptons (in green), while the interactions amongst them are governed by the gauge bosons (in red). The Higgs boson (in yellow) is responsible for the masses of the particles.

984 Quarks are the only SM particles that are subject to the three forces: the electromag-  
 985 netic, the weak and the strong. Each quark carries a flavour, which is subject to the  
 986 electroweak interaction, and a colour, which is subject to the strong interaction. The  
 987 latter is described by the Quantum Chromodynamics (QCD) theory. A property of this  
 988 theory is the colour confinement, through which quarks do not exist as free states and can  
 989 only be experimentally observed as bound states. Hence, they form mesons, which are  
 990 quark-antiquark states, and baryons, which are composed by three quarks. Both bound  
 991 states are denoted as hadrons. Although quarks are confined in hadrons, they are asymp-  
 992 totically free particles, meaning the strong coupling becomes weaker when the momentum  
 993 transfer is large. This property allows the fundamental interactions between them to be  
 994 studied in proton colliders such as the LHC.

995 The leptons, as the quarks, are divided in three families, but they are only subject to the  
996 electromagnetic and the weak interactions. The charged leptons of the three families are  
997 the electron ( $e$ ), muon ( $\mu$ ) and tau lepton ( $\tau$ ), respectively. The electron is the lightest  
998 one and is stable. Each lepton is paired to a neutrino of the same flavour ( $\nu_e, \nu_\mu, \nu_\tau$ ),  
999 which is electrically neutral and is massless in the classical SM formulation. However, the  
1000 observation of neutrino flavour oscillations implies that neutrinos have non-zero masses.  
1001 Being electrically neutral, neutrinos interact with the matter only via the weak force, and  
1002 consequently they are not directly detectable at collider experiments. Their presence can  
1003 nonetheless be inferred via the energy imbalance of the event.

1004 There are two types of bosons in the SM: vector bosons ( $W^\pm, Z, \gamma$ ) with spin = 1 and a  
1005 scalar Higgs boson (H) with spin = 0. The vector bosons are the force-carrier particles of  
1006 the fundamental interactions, viz.  $\gamma$  for electromagnetic interaction, eight types of gluons  
1007 ( $g$ ) for the strong interaction, and  $W^\pm$  or  $Z$  boson for the weak interaction. Last but  
1008 not the least, the Higgs boson is responsible for generating masses to all the SM particles  
1009 (including itself). The generation of the mass of the bosons and fermions is explained  
1010 by the phenomenon of spontaneous symmetry breaking of the electroweak theory, which  
1011 results from the postulation of the existence of the Higgs boson. The phenomenon of  
1012 spontaneous symmetry breaking is presented in Section 1.2.

1013 Any system is described by a Lagrangian density, or simply Lagrangian, which encodes the  
1014 propagation of these fields and the interactions between them, based on a basic underlying  
1015 symmetry, the gauge invariance. This invariance means the Lagrangian is invariant under  
1016 local transformations which form certain Lie groups. The Lie groups which give rise to  
1017 the interactions described by the SM are the  $SU(2) \otimes U(1)$  and  $SU(3)$  corresponding to  
1018 electroweak and strong interactions, respectively. The SM does not include the description  
1019 of the gravitational interaction, but this force can be neglected at the considered energies:  
1020 its intensity is 25 orders of magnitude lower than the weak force, the weakest within  
1021 the SM. The interactions between SM particles are described by a Lagrangian involving  
1022 the corresponding quantum fields. The SM Lagrangian can be described as a sum of  
1023 Lagrangian for the three interactions: Electromagnetic, strong and weak interactions.  
1024 Throughout this thesis, natural units ( $\hbar = c = 1$ ) are used.



### 1025 1.1.1 Electromagnetic Interactions

1026 The electromagnetic interaction is described by Quantum Electrodynamics (QED). It ex-  
1027 plains phenomena involving electrically charged particles interacting by means of exchange  
1028 of a photon. It is an abelian gauge theory with the symmetry group  $U(1)$ .

1029 The Lagrangian of the QED is given as,

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - eQ\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.1)$$

1030 where  $\psi$  is a fermionic field, involving both quarks and leptons. The first term corresponds  
1031 to the free Lagrangian of a massive spin -  $\frac{1}{2}$  field, whereas the second term is the mass  
1032 term. The third term arises from the introduction of the  $U(1)$  covariant derivative, namely

$$D_\mu = \partial_\mu + ieQA_\mu, \quad (1.2)$$

1033 and corresponds to the interaction between the photon, represented by the gauge potential  
1034  $A_\mu$ , and the fermion. The strength of the interaction is proportional to the charge  $eQ$  of  
1035 the fermion,  $Q$  being the quantum number associated to this interaction and  $e$  the electron  
1036 charge. The last term in Eq. 1.1 corresponds to the free propagation of the photon, where  
1037  $F_{\mu\nu}$  is the electromagnetic or Maxwell tensor defined as  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ . The photon  
1038 is massless, but it does not interact with itself, since QED is an abelian theory.

### 1039 1.1.2 Strong Interactions

1040 The strong interaction is governed by QCD, the theory that describes the interactions  
1041 between quarks and gluons. QCD is a non-abelian gauge theory based on a local gauge  
1042 symmetry group called  $SU(3)$ . It describes the Colour charge (C) associated to this  
1043 group, and it can take three values: red, green and blue. Quarks being spin -  $\frac{1}{2}$  fermions,  
1044 they satisfy the Dirac equation and hence the free-field Lagrangian is given by the Dirac  
1045 Lagrangian

$$\mathcal{L} = \bar{q}(i\gamma^\mu\partial_\mu - m)q, \quad (1.3)$$

1046

1047 where  $q$  corresponds to the quark field,  $m$  to its mass and  $\gamma^\mu$  to the Dirac matrices. The  
 1048 symbol  $\partial_\mu$  denotes the partial derivative with respect to the spacetime coordinates.

1049 In order for the Lagrangian to be invariant under the transformation, the derivative  $\partial_\mu$   
 1050 has to be re-defined to the so-called covariant derivative  $D_\mu$  as

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a, \quad (1.4)$$

1051 where  $g_s$  is a real constant in the parameters of the transformation. The term  $\frac{\lambda_a}{2}$  cor-  
 1052 responds to  $3 \times 3$  traceless hermitian matrices, the so-called Gell-Mann matrices, which  
 1053 generate the group, the gauge vector fields  $G_\mu^a$  correspond to the eight gluons that mediate  
 1054 the strong force. The overall QCD Lagrangian is

$$\mathcal{L}_{\text{QCD}} = i\bar{q}\gamma^\mu\partial_\mu q - m\bar{q}q - g_s\bar{q}\gamma^\mu\frac{\lambda_a}{2}qG_\mu^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a \quad (1.5)$$

1055 with the summation over all quark fields involved. The first two terms in Eq. 1.5 are  
 1056 as described in Eq. 1.1.2. The third term arises from the introduction of the covariant  
 1057 derivative and describes interaction of the gluon with a quark and an antiquark. The  
 1058 strength of the interaction is parametrized by the constant  $g_s$ , usually redefined as the  
 1059 strong coupling constant  $\alpha_s = g_s^2/4\pi$ . This constant has the property of asymptotic  
 1060 freedom: it becomes very small when the energy transfer is large enough, leading to  
 1061 a quasi-free behaviour of the quarks and gluons. Finally, the fourth term represents  
 1062 the propagation of the gluons; upon expansion, it leads to 3-gluons and 4-gluons self-  
 1063 interactions. Gluons must be massless, otherwise adding a mass term  $m^2 G_\mu^a G_\mu^a$  would  
 1064 lead to a gauge non-invariant Lagrangian.

### 1.1.3 Weak Interactions

The weak interaction is described with the non-abelian gauge group SU(2) group. It is a unique theory: unlike other interactions, it has the peculiarity that it violates parity. This is accounted for in the theoretical description by the property of the chirality of a fermion field, which introduces a vector-axial structure in the Lagrangian of the weak force. The chirality is a Lorentz-invariant quantity corresponding to the eigenvalues of the operator  $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ , which can be -1 or +1, giving rise to the so-called left ( $\psi_L$ ) and right ( $\psi_R$ ) chirality fields, represented as SU(2) doublets and SU(2) singlets, respectively. The left and right components of a fermion field  $\psi$  are obtained by applying the  $P_L$  and  $P_R$  projectors

$$\begin{aligned}\psi &= \psi_L + \psi_R \\ \psi_L &= P_L\psi = \frac{1}{2}(1 - \gamma^5)\psi \\ \psi_R &= P_R\psi = \frac{1}{2}(1 + \gamma^5)\psi.\end{aligned}\tag{1.6}$$

where  $\frac{1}{2}(1 - \gamma^5)$  and  $\frac{1}{2}(1 + \gamma^5)$  are the left- and right-handed projection operators respectively. Each fermionic field of the SM is represented as one left chirality doublet ( $\Psi_L$ ) and two right chirality singlets ( $\psi_R, \psi'_R$ ). For the first family of fermions, the fields of the electron-neutrino pair are expressed as

$$\Psi_L(x) = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}; \quad \psi_R(x) = \nu_{eR}, \quad \psi'_R(x) = e_R;\tag{1.7}$$

the up-down quark pair is expressed as

$$\Psi_L(x) = \begin{pmatrix} u_L \\ d_L \end{pmatrix}; \quad \psi_R(x) = u_R, \quad \psi'_R(x) = d_R.\tag{1.8}$$

The same holds for the other two families. Under this notation, the weak Lagrangian for a spin -  $\frac{1}{2}$  field can be written as

$$\mathcal{L}_{\text{weak}} = i\bar{\Psi}_L\gamma^\mu D_\mu\Psi_L + i\bar{\psi}_R\gamma^\mu D_\mu\psi_R + i\bar{\psi}'_R\gamma^\mu D_\mu\psi'_R - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu},\tag{1.9}$$

1082 The quantum number associated to the SU(2) group is the weak isospin, which has three  
 1083 components  $I_{1,2,3}$ . The right chirality fields  $\psi_R$  and  $\psi'_R$  have a third isospin component  
 1084 of  $I_3 = 0$  as they are singlets under SU(2); the left-handed field  $\Psi_L$  has  $I_3 = +1/2$  and  
 1085  $I_3 = -1/2$  for the upper and lower components, since they form a doublet under SU(2).

## 1086 1.2 The Brout-Englert-Higgs Mechanism

1087 Higgs field and electroweak symmetry breaking (EWSB) was first proposed in the mid-  
 1088 sixties by three independent groups: Robert Brout and François Englert;[\[18\]](#) by Peter  
 1089 Higgs;[\[19\]](#) and by Gerald Guralnik, C. R. Hagen, and Tom Kibble.[\[20, 21, 22\]](#). The Brout-  
 1090 Englert-Higgs (BEH) mechanism was introduced as a solution to generate the gauge boson  
 1091 masses and explain the fermion masses, which were not accounted for in the electroweak  
 1092 gauge formalism. It is based on the concept of spontaneous symmetry breaking, a phe-  
 1093 nomenon that is often observed in nature in which a physical system whose equations of  
 1094 motion are symmetric, ends up in an asymmetric state. In particular, it describes systems  
 1095 where the Lagrangian obeys certain symmetries, but an individual ground state of the  
 1096 system does not exhibit the symmetries of the system itself.

1097 The Lagrangian mass terms for the fermions, for the  $W^\pm$ , and for the Z bosons should be:

$$\begin{aligned} \mathcal{L}_{mf} &= m_f (\bar{\Psi}_L \Psi_R + \bar{\Psi}_R \Psi_L) \text{ for the fermions} \\ \mathcal{L}_{mV} &= m_W W^{\mu+} W_\mu^- + \frac{m_Z}{2} Z^\mu Z_\mu \text{ for the gauge bosons} \end{aligned} \tag{1.10}$$

1098 However, the  $\mathcal{L}_{mf}$  term violates the SU(2) gauge symmetry while the  $\mathcal{L}_{mV}$  term violates  
 1099 both the SU(2) and the U(1) gauge symmetries that were assumed to build the electroweak  
 1100 model. The Brout-Englert-Higgs mechanism allows to naturally introduce the mass terms  
 1101 in the SM Lagrangian within the initial assumption of the gauge symmetries. In particular,  
 1102 a new SU(2) doublet  $\Phi$  is defined as:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \tag{1.11}$$

1103 The  $\Phi^+$  and  $\Phi^0$  are complex scalar fields. Their superscripts correspond to their electric

1104 charge as it will be proven later on. The  $\Phi$  field is introduced in the Lagrangian through  
 1105 a kinetic term:

$$\mathcal{L}_{\Phi kin} = D_\mu \Phi D^\mu \Phi \quad (1.12)$$

1106 and a "potential" term:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \quad \text{with } \mu^2 > 0 \text{ and } \lambda > 0 \quad (1.13)$$

1107 The  $V(\Phi)$  potential has a set of degenerate minima defined by the condition:

$$|\Phi|^2 = \frac{\mu^2}{2\lambda} = \frac{v^2}{2} \quad \text{with } v = \mu/\sqrt{\lambda} \quad (1.14)$$

1108 Therefore,  $v/\sqrt{2}$  is the vacuum expectation value of the  $\Phi$  field. The generic  $\Phi$  vacuum  
 1109 state can be written as:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.15)$$

1110 The kinetic term of Eq. 1.12 can be expressed with respect to the  $\Phi$  vacuum state of Eq.  
 1111 1.15 obtaining:

$$\mathcal{L}_{\Phi kin} = \frac{1}{2} \partial^\mu H \partial_\mu H + \left(1 + \frac{H}{v}\right)^2 \left[ \frac{g^2 v^2}{4} W^{+\mu} W_\mu^- + \frac{1}{2} \frac{(g^2 + g'^2) 2v^2}{4} Z^\mu Z_\mu \right] \quad (\text{A.25}) \quad (1.16)$$

1112 This result has two important consequences:

- 1113 • The non-zero vacuum expectation value of the  $\Phi$  field introduces in the SM La-  
 1114 grangian the mass terms for the  $W^\pm$  and  $Z$  bosons. The  $W^\pm$  and  $Z$  bosons masses  
 1115 can be related to the  $v$  constant as:

$$m_W = \frac{gv}{2} \quad \text{and} \quad m_Z = \frac{\sqrt{g^2 + g'^2} v}{2} = \frac{m_W}{\cos(\theta_w)} \quad (1.17)$$

1116 In addition, the  $v$  value can be related to the Fermi constant  $G_F$ , experimentally  
 1117 measured with a very good precision:

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}} = 246 \text{ GeV} \quad (1.18)$$

1118 • In the Lagrangian, new interaction terms between one or two  $H$  fields and a vector  
 1119 boson pair arise. The corresponding vertices are proportional to the squared mass  
 1120 of the vector bosons.

1121 In the same way, the potential term of Eq. 1.13 can be expressed with respect to the  $\Phi$   
 1122 vacuum state of Eq. 1.15 obtaining:

$$V(\langle\Phi\rangle) = \frac{1}{2}(2\lambda v^2)H^2 + \lambda v H^3 + \frac{\lambda}{4}H^4 - \frac{\lambda}{4}v^4 \quad (1.19)$$

1123 This result has important consequences:

1124 • A mass term for the  $H$  field arises in the Lagrangian. Therefore, the  $H$  scalar field  
 1125 describes a boson with spin 0, without electric charge, and with a mass  $m_H = \sqrt{2\lambda}v$ .  
 1126 This particle takes the name of Higgs boson.

1127 • Tri-linear and quadri-linear  $H$  self-coupling terms arise in the Lagrangian with a  
 1128 coupling constant proportional to  $\lambda v$  and  $\lambda/4$ , respectively.

1129 It is worth to notice that the functional form for the Higgs boson potential of Eq. 1.13  
 1130 was chosen arbitrarily as the lowest order polynomial ensuring the  $SU(2) \otimes U(1)$  gauge  
 1131 symmetries and providing a vacuum expectation value different from zero. The actual  
 1132 functional form can differ. Therefore, in order to test the accuracy of the model predic-  
 1133 tions, it is fundamental to perform precision measurements of the  $m_H$  value and of the  
 1134 trilinear and the quadrilinear Higgs boson self-coupling constants.

### 1.3 Higgs Self Coupling

The Higgs boson discovery at the LHC in 2012 [1][2] and the subsequent campaign of measurements of its properties [23, 24], have provided a wonderful confirmation of our understanding of elementary particles and their interactions. So far, the predictions of the Standard Model (SM) for the Higgs boson couplings to the vector bosons and to third generation fermions are in spectacular agreement with observations. On the other hand, its interactions with lighter sectors, such as the first and second generation quarks and leptons, are still to be confirmed.

More precisely, the SM Higgs boson couplings to fundamental fermions are linearly proportional to the fermion masses, whereas the couplings to bosons are proportional to the square of the boson masses. The SM Higgs boson couplings to gauge bosons and fermions, as well as the Higgs boson self-coupling, are shown in the Fig.1.2. Because of charge-neutral and color-singlet property, the Higgs boson does not couple at tree level to the mass-less photons and gluons. Its coupling to gluons is induced at leading order by a one-loop process in which it couples to a virtual  $t\bar{t}$  pair (with minor contributions from the other lighter quarks). Likewise, the Higgs boson coupling to photons is also generated via loops.

One key sector, which is currently very weakly constrained and could very easily hide or be connected to new physics, is the scalar potential. In the SM, the Higgs scalar potential is fixed by just two low energy parameters, the Higgs mass ( $m_H \simeq 125$  GeV) and the Fermi constant  $G_F$  (or equivalently the vacuum expectation value  $v \simeq 246$  GeV). The scalar potential can be written in terms of the Higgs trilinear ( $\lambda_3$ ) and quartic ( $\lambda_4$ ) self-couplings :

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_3 v H^3 + \frac{1}{4}\lambda_4 H^4 \quad (1.20)$$

where in the SM,  $\lambda_3 = \lambda_4 = m_H^2/2v^2 \equiv \lambda_{SM}$ . In particular, higher-point Higgs boson self interactions are forbidden in SM. The measurement of the parameters that describe the shape of the Higgs potential are therefore a milestone in the quest of understanding the mechanism of the electroweak symmetry breaking and of exploration of new physics.

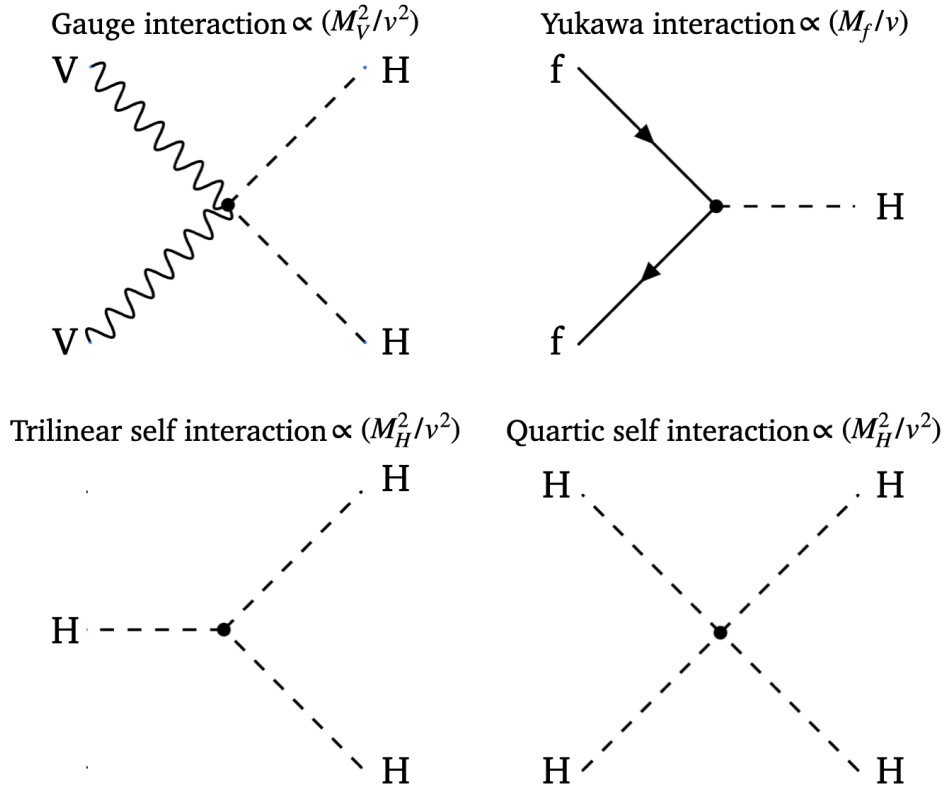


Figure 1.2: Higgs interactions.

1162 The interpretation of possible scenario with new physics is shown in Fig.1.3. As from SM,  
 1163  $\lambda_3 = \lambda_4$  is theoretical prediction and not yet experimentally observed. Thus, there are  
 1164 chances that  $\lambda_4$  can be different from  $\lambda_3$  which can create second stable or meta-stable  
 1165 minima for Higgs field and can direct towards new physics, new solution for the mystery  
 1166 of 95% unknown universe.

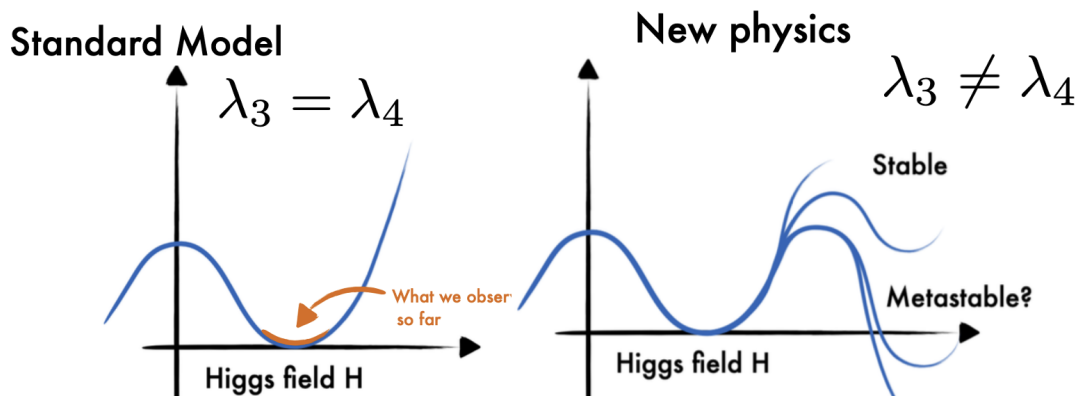


Figure 1.3: Possible interpretation of the Higgs field in SM and BSM.



# Chapter 2

## The Higgs Boson Phenomenology

### 2.1 Higgs Boson Searches in the pre-LHC Era

Various experiments in the 1970s and 1980s confirmed the general structure of the Standard Model and, broadly, the predictions concerning the gauge sector. However, the scalar sector, signifying the generation of mass via the Higgs mechanism, remained to be established experimentally. In other words, the existence of a new type of fundamental scalar particle, the Higgs boson remained questionable. This affirmation requiring discovery of the particle continued to be elusive in the next few decades making the SM to be an incomplete description. The mass of the Higgs boson,  $M_H$ , is not predicted from the theory, though the nature of the interaction and other relevant aspects are. Various considerations allowed a wide possible range of  $M_H$ , up to about 750 GeV. [25]

The Higgs boson was searched for extensively in the experiments at the large-electron-positron (LEP) collider at CERN, near Geneva, Switzerland using different production processes depending on the centre-of-mass energy ( $\sqrt{s}$ ). In the LEP1 (1984-1994) era production of Z boson with subsequent decay through  $e^-e^+ \rightarrow Z \rightarrow qq$  channel was the main target. In the LEP2 (1994-2004) era with higher  $\sqrt{s}$  the Higgsstrahlung process -  $e^-e^+ \rightarrow Z^* \rightarrow ZH$  opened up and the Higgs boson was searched for in the  $b\bar{b}$  final state due to the largest decay branching ratio. This was combined with leptonic and hadronic decay modes of Z, providing good event statistics.

1187 For the highest value of  $\sqrt{s} = 206$  GeV at LEP2, the kinematic considerations allowed  
1188 a maximum value for  $M_H$  to be about 115 GeV. In spite of some experimental hints of  
1189 possible production of the Higgs boson, the final conclusion from LEP by the turn of the  
1190 century (2000) was a lower limit on the mass:  $M_H > 114.4$  GeV at 95% confidence level  
1191 (CL). [26]

1192 The Higgs boson search continued extensively at the Tevatron proton-antiproton collider  
1193 at Fermilab. Despite being a hadron machine, it was more challenging to exploit the  
1194 related Higgsstrahlung production. The experimentally sensitive mass region was limited  
1195 essentially between 140 to 180 GeV for production modes initiated via quark-antiquark  
1196 pair. Just before the physics analysis started at the LHC, the CDF and D0 experiments  
1197 at the Tevatron excluded the mass range of the Higgs boson 162-166 GeV at the 95% CL.  
1198 [27] By the time the Higgs boson was discovered at the LHC in 2012, Tevatron data also  
1199 hinted at an excess of 3 standard deviations in the mass range of 115-140 GeV.

## 1200 2.2 Search for Higgs Boson at LHC

1201 By early 1990s search capability for the Higgs boson became a major benchmark for the  
1202 planned LHC experiments. A search had to be made across the entire allowed range of  
1203 masses; from around a mass of approximately 50 GeV, the lower limit at the time, up  
1204 to its largest possible value of approximately 1000 GeV. Since the LHC was capable of  
1205 producing the Higgs boson of wide range of masses, the mandate was to hunt out the  
1206 particle and resolve the issue of electroweak symmetry breaking. Accordingly, the search  
1207 strategy and hence the detector design were focused on specific final states for different  
1208 mass regions.

### 1209 2.2.1 Higgs Boson at LHC

1210 At the LHC, the Higgs boson can be produced by several mechanisms over the whole pos-  
1211 sible mass range, due to the availability of highly energetic partons. The major production  
1212 modes are briefly discussed below:

1213 1. **Gluon-Gluon Fusion (ggH)** - The Higgs boson production mechanism with the  
 1214 largest cross section is the gluon fusion. It contributes almost 88 (85)% of the total  
 1215 Higgs cross section at the LHC center-of-mass energy of  $\sqrt{s}=7(13)$  TeV. In the SM,  
 1216 the direct coupling between the Higgs boson and gluons are not allowed, so the  
 1217 production of the Higgs boson proceeds via virtual quark loops. The dominant  
 1218 contribution comes from the exchange of a virtual top quark while contributions  
 1219 from lighter quarks propagating in the loop are suppressed due to their lower masses.  
 1220 This process thus indirectly gives access to the top quark Yukawa coupling ( $y_t$ ) from  
 1221 the virtual loop. A representative Feynman diagram of ggH process at the leading  
 1222 order (LO) is shown in Fig.2.1 (left).

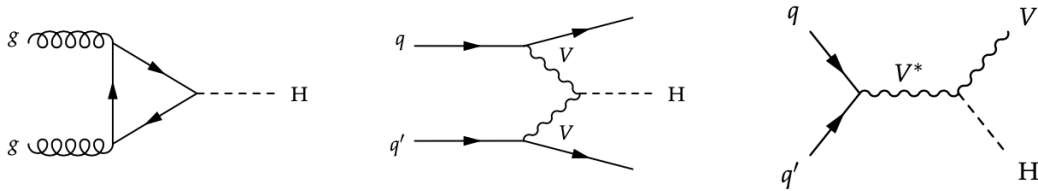


Figure 2.1: Left: Leading order Feynman diagram of ggH process, middle: VBF Higgs production and right: Higgsstrahlung (VH) process of Higgs boson in association with a vector boson.

1223 2. **Vector Boson Fusion (VBF)** - The mechanism with the second-largest cross  
 1224 section is VBF with reduced cross section by about a factor of ten than the ggH  
 1225 production process. The colliding parton pair simultaneously radiates two vector  
 1226 bosons which are fused to produce the the Higgs boson in the central region as  
 1227 shown in Fig.2.1 (middle). The outgoing quarks continue almost along the original  
 1228 direction. The distinctive topology of the event and the kinematics of the final  
 1229 state makes this process very unique and can be exploited to distinguish them from  
 1230 overwhelming backgrounds and used as a clean environment not only for Higgs  
 1231 searches but also for the determination of the Higgs boson couplings. This process  
 1232 gives direct access to the  $HVV = c_V$  coupling.

1233 3. **Higgsstrahlung (VH)** - In this case the the Higgs boson is produced in association  
 1234 with a weak interaction gauge boson (V), ie., either  $W^\pm$  or Z as shown in Fig.2.1  
 1235 (right). As in case of VBF, the VH production mode also provides the access to  $c_V$ .  
 1236 As neither the Higgs boson nor the vector bosons are stable particles, their decay

1237 channels have to be considered. Tagging the leptonic decay of V, the search for the  
 1238 hadronic decay of H, i.e.  $H \rightarrow b\bar{b}$  is possible due to reduction of the QCD induced  
 1239 multijet backgrounds.

1240 **4. In association with top pair ( $t\bar{t}H$ )** - The Higgs boson production in association  
 1241 with  $t\bar{t}$  provides a direct probe of the Higgs-top Yukawa coupling,  $y_t$ , in contrast  
 1242 with the  $y_t$  measurement from the virtual top quark loop. Since top quark mass is  
 1243 much more than  $m_H/2$ , kinematically the Higgs boson cannot decay to top quark  
 1244 pair, and hence this coupling cannot be measured from the decay of the Higgs boson  
 1245 to top pair. Representative Feynman diagrams of  $t\bar{t}H$  process are shown in Fig.2.2

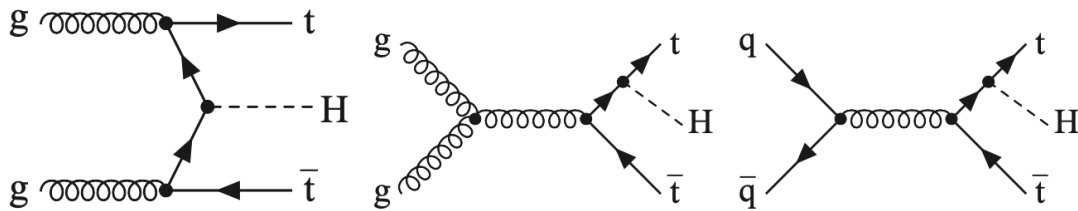


Figure 2.2: Representative tree-level Feynman diagrams of the  $t\bar{t}H$  production.

1246 **5. In association with single top (tH)** - Production in association with a single top  
 1247 quark, the Higgs boson can be radiated either from the exchanged W boson or from  
 1248 the top quark in the two dominant leading order processes as shown in Fig. 2.3.  
 1249 The relative sign between the two Higgs boson couplings,  $y_t$  and  $c_V$ , decides the sign  
 1250 of the interference terms of the two diagrams; thus can bring valuable information,  
 1251 in particular regarding the sign of the top Yukawa coupling. It is to be noted that  
 1252 this process has not yet been observed experimentally at the LHC. The current  
 1253 signal strength of the tH has been observed from the multilepton final state study  
 1254 by the CMS experiment to be  $5.7 \pm 2.7(\text{stat}) \pm 3.0(\text{syst})$ [28] times the SM predicted  
 1255 value of 0.0724 pb with observed(expected) significance amounts to 1.4(0.3) for tH  
 1256 production.

1257 The cross sections for the production of the Higgs boson with their theoretical uncertain-  
 1258 ties are shown in Fig.2.4.

1259 After the production of H, it subsequently decays within a short lifetime of  $\mathcal{O}(10^{-22})$ s[30].  
 1260 Driven mainly by the value of  $M_H$  and the coupling constants of the Higgs boson to vector

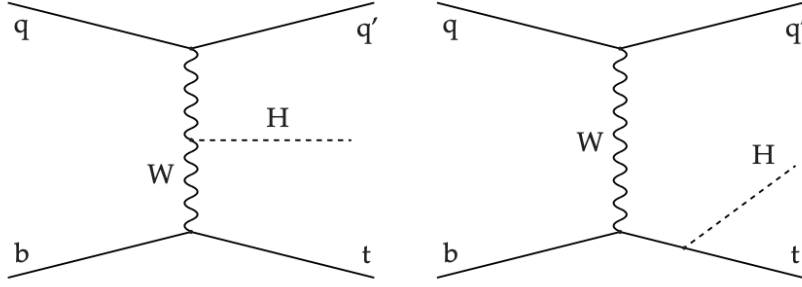


Figure 2.3: Representative tree-level Feynman diagrams of the tH process. Diagram on the left contains  $c_V$  and the right one has  $y_t$ .

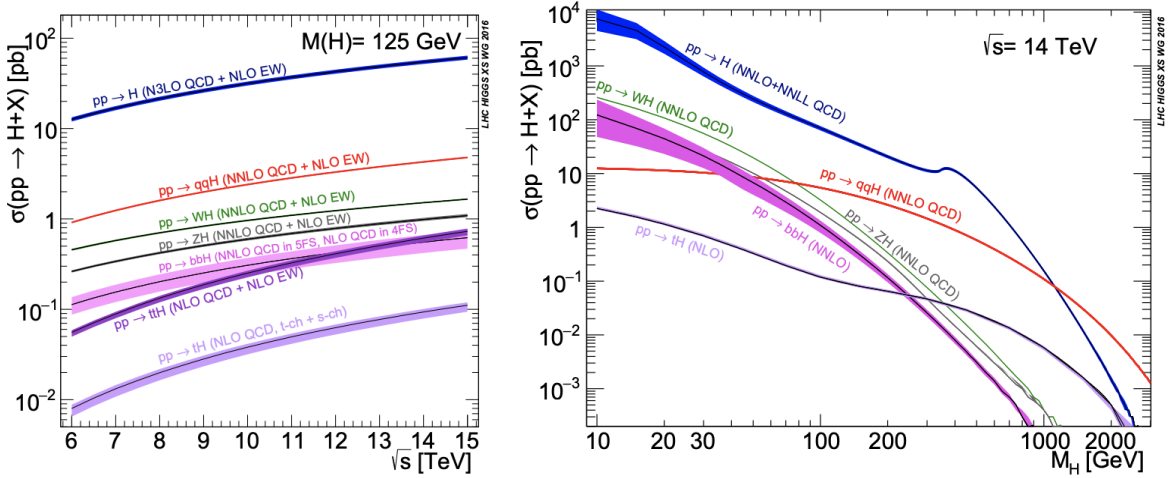


Figure 2.4: Left - Production cross section of a SM Higgs boson of mass  $m_H = 125$  GeV as a function of center-of-mass energy  $\sqrt{s}$ . Right - Production cross section of a SM Higgs boson at  $\sqrt{s} = 14$  TeV as a function of Higgs Mass. The lines with different colors correspond to the different production modes with a certain order of accuracy, while the band across each lines give the uncertainty of the calculation. Figure taken from [29].

1261 bosons and fermions, various decay final states are possible. Interestingly, the measured  
 1262 mass of  $M_H$  allows, fortunately, a large variety of possible decay channels, most of which  
 1263 can be detected experimentally.

1264 Of course, at a hadron collider like the LHC, the experimental challenges for measurements  
 1265 are greater for the hadronic final states. The general purpose detectors at the LHC  
 1266 are designed to be maximally sensitive to non-hadronic final states involving photons,  
 1267 electrons and muons, such that the discovery milestone could be reached even with limited  
 1268 data.

1269 The branching ratio ( $Br$ ) to a particular decay final state ( $H \rightarrow xx$ ) is defined as the  
 1270 ratio of decay width of the Higgs boson to that particular decay mode to the total decay

1271 width.

$$Br(xx) = \frac{\Gamma_{xx}}{\Gamma_{total}} \quad (2.1)$$

1272 Fig. 2.5 shows the  $Br$  of the Higgs boson in different final states for  $M_H = 125$  GeV.  
 1273 The dominant decay modes of the Higgs boson with mass of 125 GeV are  $H \rightarrow b\bar{b}$  and  
 1274  $H \rightarrow WW^+$ , followed by  $H \rightarrow gg$ ,  $H \rightarrow \tau^+\tau^-$ ,  $H \rightarrow cc$  and  $H \rightarrow ZZ^*$ . With much  
 1275 smaller rates follow the Higgs boson decays into  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow \gamma Z$  and  $H \rightarrow \mu^+\mu^-$ . The  
 1276 importance of studying the the Higgs boson in different final states has been described in  
 1277 the later sections of this thesis.

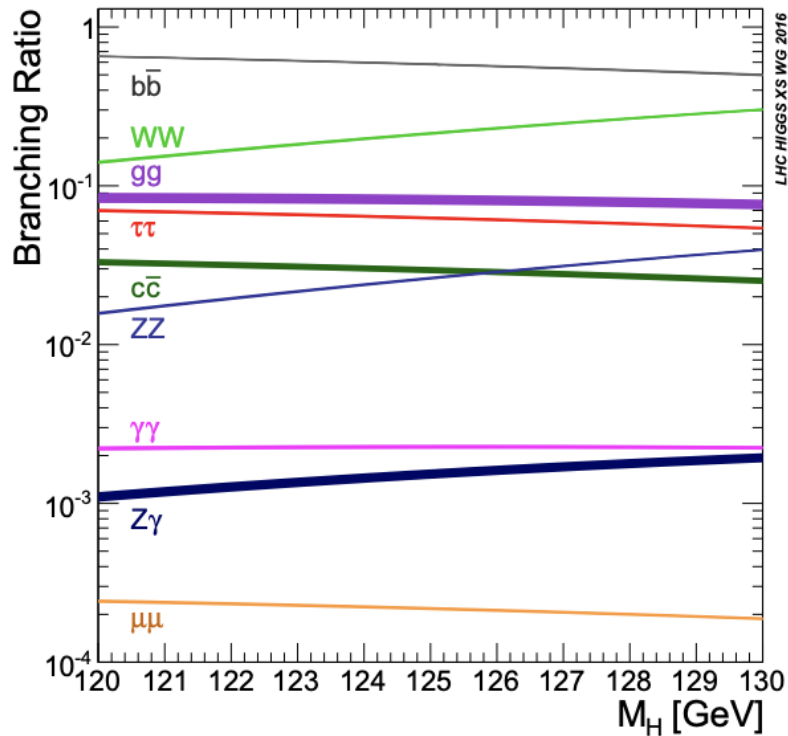


Figure 2.5: Branching ratios of the Higgs boson in the SM as a function of the mass. Figure taken from [29].

## 1278 2.2.2 Higgs Boson Discovery

1279 The LHC physics journey started in earnest in April 2010, when the first proton-proton  
 1280 collisions at an unprecedented centre-of-mass energy of  $\sqrt{s}=7$  TeV, 3.5 times larger than  
 1281 at the previous most powerful hadron collider - the Tevatron. The collision energy was

1282 raised to  $\sqrt{s}=8$  TeV in 2012. The first LHC data-taking period ('Run 1') covered about  
1283 3 years, from April 2010 to February 2013.

1284 A large amount of data, about 5 billion events, from the examination of some 2000 trillion  
1285 proton–proton interactions, was recorded in Run 1 by each of the two experiments, ATLAS  
1286 and CMS. Owing to the excellent performance, ATLAS and CMS were able to 'reproduce'  
1287 50 years of particle physics in less than 1 year of operation.

1288 Undoubtedly, the most striking result to emerge from the ATLAS[1] and CMS[2] experi-  
1289 ments is the discovery of a new heavy boson with a mass of approximately 125 GeV. The  
1290 Higgs boson discovery was announced on July 4, 2012. Humongous efforts from a large  
1291 community consisting of accelerator engineers, theoretical and experimental physicists  
1292 matched by computing experts made it possible; this discovery has been truly termed  
1293 as *a big leap for human kind*. This hallmark result established the last part of the SM  
1294 particle spectrum which was missing for several decades and resolved the mystery about  
1295 the mass generation of the weak gauge bosons and the fermions.

1296 The ATLAS Collaboration reported the existence of a neutral scalar boson with a mass  
1297 measured at  $126.0 \pm 0.4(\text{stat}) \pm 0.4(\text{syst})$  GeV with a signal significance of 5.9 standard  
1298 deviations corresponding to a background fluctuation probability of  $1.7 \times 10^{-9}$  [1]. The  
1299 analysis was based on the accumulated data of  $4.8 \text{ fb}^{-1}$  collected at  $\sqrt{s} = 7$  TeV in 2011  
1300 and  $5.8 \text{ fb}^{-1}$  at  $\sqrt{s} = 8$  TeV in 2012. In parallel, the CMS Collaboration also established  
1301 an excess of events corresponding to a neutral resonance production at the mass of  $125.3$   
1302  $\pm 0.4(\text{stat}) \pm 0.5(\text{syts})$  GeV with a signal significance of 5.9 standard deviations based on  
1303 analysis of  $5.1 \text{ fb}^{-1}$  and  $5.7 \text{ fb}^{-1}$  of data collected at  $\sqrt{s} = 7$  and 8 TeV, respectively [2].  
1304 For both the experiments the analysis sensitivities were mostly driven by the di-photon  
1305 ( $H \rightarrow \gamma\gamma$ ) and four-leptons ( $H \rightarrow ZZ^* \rightarrow 4l$ ) final states with excellent, high resolution  
1306 ( $\Delta m/m \sim 1 - 2\%$ ) measurements. The discovery plots from CMS Collaboration are  
1307 presented in Fig. 2.6. As the newly discovered particle decays to a pair of photons, it  
1308 ensures that its intrinsic spin cannot be one unit and it belongs to the bosonic family.

1309 As mentioned already, the intrinsic mass of the Higgs boson is a free parameter and for  
1310 the measured value of the  $m_H$ , SM can predict almost all the important properties of  
1311 H. At the same time, for scenarios beyond the SM, the Higgs field structure is extended  
1312 and thereby existence of multiple physical Higgs bosons are predicted. In some models,

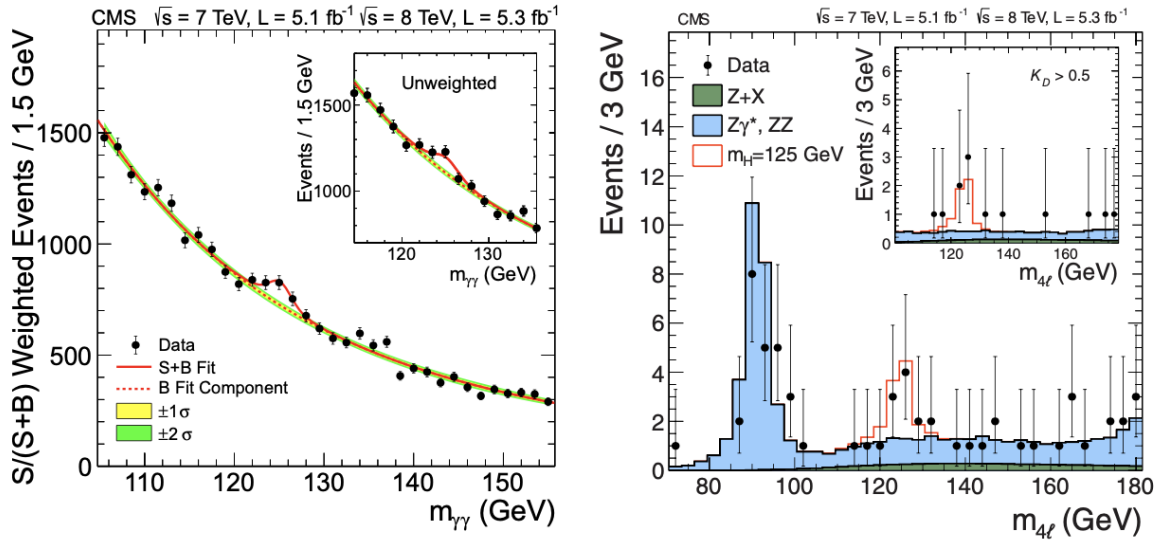


Figure 2.6: Left: The diphoton invariant mass distribution weighted by the  $S/(S + B)$  value of its category, the peak around 125 GeV from the SM the Higgs boson contribution on top of the continuum diphoton background is shown with red solid line. Right: Distribution of the four-lepton invariant mass for the  $ZZ \rightarrow 4l$  analysis, the Higgs boson with a mass of  $m_H$  125 GeV has been shown in red solid line sitting on the background. The figure is taken from [2].

1313 eg. minimal supersymmetric extension of SM (MSSM), the lightest member of the Higgs  
 1314 boson family resemble the SM particle, although with slight differences in some of the  
 1315 properties, like couplings to various particles. Hence, even after observing the existence  
 1316 of the Higgs boson at  $m_H \sim 125$  GeV, the dilemma was whether the discovered resonance  
 1317 belongs to the SM or not. To resolve this, all the properties of the Higgs boson must be  
 1318 measured thoroughly and compared with the attributes in the SM. Both the ATLAS and  
 1319 the CMS experiments have been studying painstakingly various properties of the Higgs  
 1320 boson utilizing both the Run 1 and Run 2 data.

1321 The wisdom gained during last one decade is extremely rich and unexpected to a good  
 1322 extent. The exemplary works of the collider physics community has made many interesting  
 1323 measurements possible. All the measurements are compatible with the predictions of SM  
 1324 so far. However the current level of uncertainties still allows the particle to belong to  
 1325 certain physics scenarios beyond SM, although the specific nature of them cannot be  
 1326 judged.



## 1327 2.3 Interpretation of LHC Data

1328 Before further moving on, it is important to know some basic notations or framework for  
1329 interpretation of LHC data. It will help understand next chapters.

### 1330 2.3.1 Signal Strength $\mu$

1331 This is the first and the most simplified theoretical framework developed to interpret the  
1332 LHC data. To understand it with example, for a particular mass hypothesis of H, the  
1333 expected number of signal events ( $s(M_H)$ ) in a particular decay mode can be written as:

$$s(M_H) = \sigma_{SM}(M_H) \cdot Br \cdot L \cdot \epsilon \cdot A \quad (2.2)$$

1334 Here  $\sigma_{SM}(M_H)$  is the production cross section,  $Br$  is the branching ratio of the particular  
1335 the Higgs boson decay mode,  $L$  is the integrated luminosity of the data being used and  
1336  $\epsilon$  and  $A$  are the efficiency and the geometrical acceptance of the experiment. In a total  
1337 of  $n$  number of observed events with  $b$  number of background events predicted from SM,  
1338  $\sigma_{SM}(M_H)$  can be fitted like:

$$n = \mu \cdot s(M_H) + b \quad (2.3)$$

1339 where  $\mu$  is called signal strength which is defined as the ratio of the observed value of  
1340  $\sigma \cdot Br$  to its expectation predicted from the SM.

$$\mu = \frac{(\sigma \cdot Br)_{obs}}{(\sigma \cdot Br)_{SM}} \quad (2.4)$$

1341 Normally, a measured value of  $\mu = 1$  corresponds to the SM prediction, while a devia-  
1342 tion indicates the effects of the beyond the SM physics (BSM). Nevertheless, there are  
1343 uncertainties in the measurements which have both the statistical and the systematic  
1344 components. A lot of effort goes into improving the precision.

### 1345 2.3.2 kappa-Framework ( $\kappa$ )

1346 In simple words,  $\kappa$  is the ratio of quantity to its SM prediction and mainly used to check  
 1347 agreement with SM ( $\kappa = 1$ ); if not, there is scope for new physics.

1348 The  $\kappa$ -framework introduced in Ref. [31] can be considered as a special case of the  
 1349 SMEFT [29] to consistently parametrize the Higgs boson production cross section and  
 1350 decay width in presence of anomalous Higgs coupling values. In the  $\kappa$ -framework, only  
 1351 the EFT operators whose effect is the modification of the SM couplings are considered,  
 1352 while the other EFT operators are assumed to be negligible. In addition, the new operators  
 1353 are assumed to impact only on the coupling strength, and not on the tensor structure of  
 1354 the coupling.

1355 The couplings of the Higgs boson to the massive SM particles (neutrinos are ignored)  
 1356 are taken into account both at the production as well as at the decay vertices. The  
 1357 production cross section, the total width and the decay branching ratio to a particular  
 1358 mode are scaled separately by the relevant scale factors called  $\kappa$ . Considering a process  
 1359  $ii \rightarrow H \rightarrow ff$ , where the Higgs boson is produced with a cross section of  $\sigma_{ii}$  followed by  
 1360 decay  $H \rightarrow ff$  with total decay width of  $\Gamma_H$  and partial width of  $\Gamma_{ff}$ , under narrow-width  
 1361 approximation one can write :

$$\begin{aligned}
 (\sigma \cdot BR)(ii \rightarrow H \rightarrow ff) &= \sigma(ii \rightarrow H) \cdot BR(H \rightarrow ff) \\
 &= \sigma(ii \rightarrow H) \cdot \frac{\Gamma_{ff}}{\Gamma_H} \\
 &= \sigma^{SM}(ii \rightarrow H) \cdot \frac{\Gamma_{ff}^{SM}}{\Gamma_H^{SM}} \cdot \frac{\kappa_i^2 \cdot \kappa_f^2}{\kappa_H^2}
 \end{aligned} \tag{2.5}$$

1362  $\kappa_i$  appears due to the couplings at the production side, while  $\kappa_f$  is the coupling modifier  
 1363 for the coupling between the the Higgs boson and its decay products and, finally,  $\kappa_H$  is  
 1364 the coupling modifier for the total decay width, since some of the couplings are yet to be  
 1365 established.

1366 It is to be noted that experimentally we are only estimating the deviations of the couplings  
 1367 wrt SM via  $\kappa$  measurements; we are not directly measuring the individual couplings. In  
 1368 various measurements there is also an implicit assumptions that the couplings do not

1369 “run” or vary across different datasets collected at different energies.

1370 Under zero-width assumption,  $\kappa$  parameters are defined to parametrize the modification  
 1371 of the Higgs couplings strengths in such a way that  $\kappa_i^2 = \sigma_i/\sigma_i^{SM}$ , or  $\kappa_i^2 = \Gamma_i/\Gamma_i^{SM}$ .  
 1372 The  $\kappa_b$ ,  $\kappa_t$ ,  $\kappa_\tau$ ,  $\kappa_\mu$ ,  $\kappa_W$  and  $\kappa_Z$  define the coupling modifiers of the Higgs boson to the  
 1373 bottom quark, top quark,  $\tau$  lepton,  $\mu$  lepton, W boson, and Z boson, respectively. In  
 1374 addition the  $H \rightarrow \gamma\gamma$  and the ggH vertices can be considered as effective vertices with  
 1375 coupling modifiers  $\kappa_\gamma$  and  $\kappa_g$ , respectively, or they can be expressed in term of the particles  
 1376 contributing inside the loops. Such loops are dominated by the top quark contribution,  
 1377 and for the  $H \rightarrow \gamma\gamma$  also by the W boson contribution. The  $\kappa_\gamma$  and  $\kappa_g$  parameters are  
 1378 typically used to probe whether BSM particles contribute to the effective  $H \rightarrow \gamma\gamma$  and  
 1379 ggH vertices.

## 1380 2.4 HH Production

1381 As mentioned above, experimentally, the trilinear coupling can be directly measured using  
 1382 the Higgs boson pair production mode  $pp \rightarrow HH$ , also known as di-Higgs production.  
 1383 Within the SM, non-resonant production is the only process for di-Higgs production,  
 1384 while the resonant di-Higgs production has its own importance to look for the new BSM  
 1385 particles.

1386 At the LHC, the main production mode of the di-Higgs process is through gluon gluon  
 1387 fusion which produces almost 95% of the HH events. Similar to the single-H production,  
 1388 there are other subdominant modes of HH production which can also be probed at the  
 1389 LHC by utilizing special properties of the concerned processes.

- 1390 • **Gluon Gluon Fusion (ggF HH)** - The dominant production mode of the HH is  
 1391 via gluon pair fusion with a cross section of about 31.05 fb at N2LO accuracy in  
 1392 QCD [32, 33] at a centre-of-mass energy of 13 TeV; it is about 1000 times smaller  
 1393 than the single-H production. Fig. 2.7 shows the leading order diagrams of the ggF  
 1394 HH process, where Higgs bosons are produced via a heavy quark loop and contain  
 1395 the t-quark Yukawa coupling ( $y_t$ ). The first diagram called triangle diagram contains  
 1396  $\lambda_{HHH}$  and  $y_t$ , while the second box diagram only contains  $y_t$ .

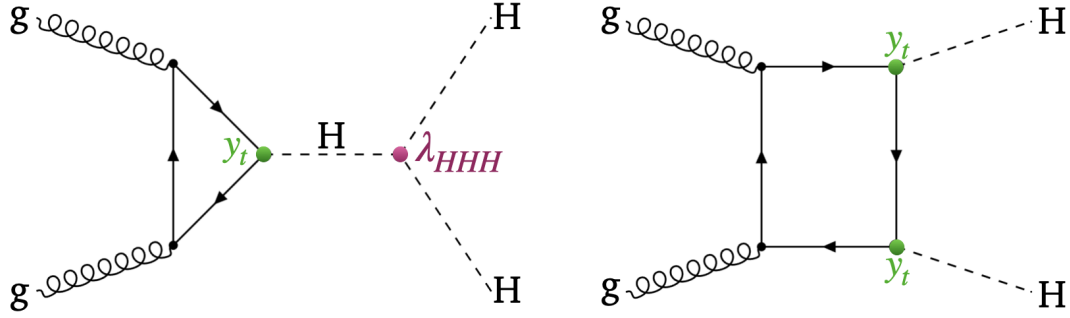


Figure 2.7: Feynman diagrams for the SM gluon-gluon fusion di-Higgs production (left) triangle diagram (right) box diagram.

1397  
1398  
1399  
1400  
1401

The cross section and kinematics of the ggF HH process depends on the  $\lambda_{HHH}$  and  $y_t$ . The reason behind a smaller cross section value is not only the small  $\lambda_{HHH}^{SM}$  value ( $=0.13$ ), but also box and triangle diagrams have opposite signs leading to destructive interference. The contribution of the individual diagrams and the interference term is shown in Fig. 2.8 as a function of HH invariant mass.

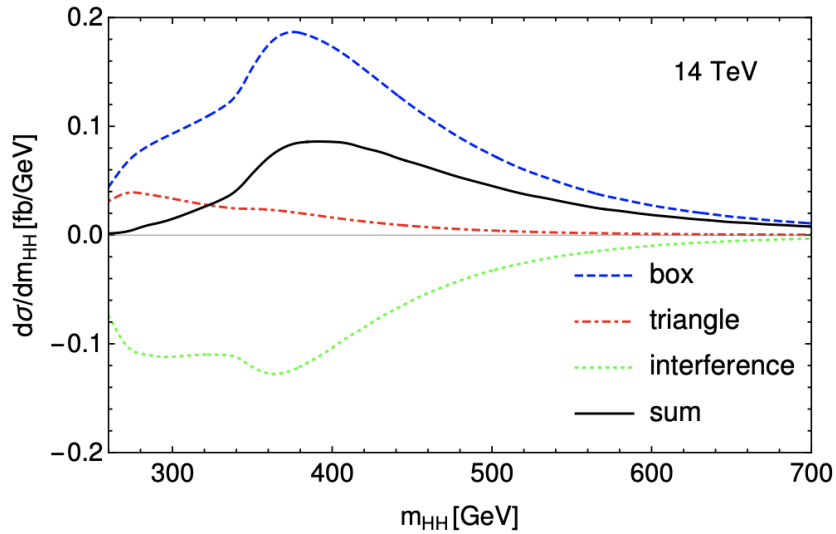


Figure 2.8: Higgs pair invariant mass distribution at LO for the different contributions (box and triangle) to the ggF HH production mechanism and their interference. Figure taken from [34]

1402  
1403  
1404  
1405  
1406

Contributions from physics beyond the SM (BSM) can significantly enhance the HH production cross section, as well as change the kinematical properties of the produced Higgs boson pair, and consequently those of the decay products. The modification of the properties of non-resonant HH production via ggF from BSM effects can be parameterized through an effective Lagrangian that extends the SM

1407 one with dimension-6 operators[29, 35]. This parameterization results in five cou-  
 1408 plings:  $\lambda_{HHH}$ , the coupling between the Higgs boson and the top quark ( $y_t$ ), and  
 1409 three additional couplings not present in the SM. Those three couplings represent  
 1410 contact interactions between two Higgs bosons and two gluons ( $c_{2g}$ ), between one  
 1411 Higgs boson and two gluons ( $c_g$ ), and between two Higgs bosons and two top quarks  
 1412 ( $c_2$ ). The Feynman diagrams contributing to BSM ggF HH production at leading  
 1413 order (LO) are shown in Fig. 2.9.

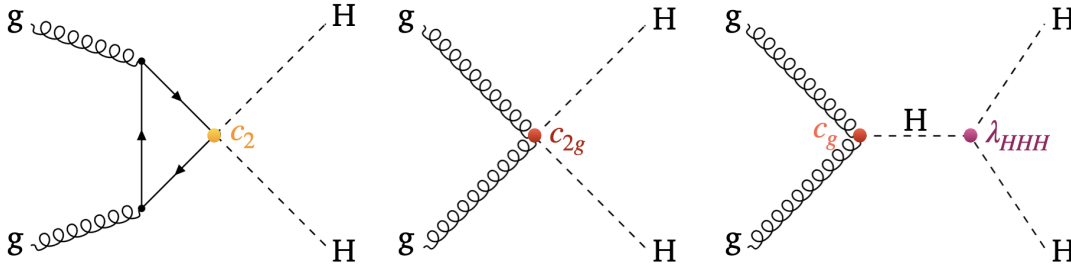


Figure 2.9: Feynman diagrams of the processes contributing to the production of Higgs boson pairs via BSM ggF at LO. The diagram on the left involves the contact interaction of two Higgs bosons with two top quarks ( $c_2$ ), the middle diagram shows the contact interaction between the Higgs bosons and two gluons ( $c_{2g}$ ), and the diagram on the right describes the contact interactions between the Higgs boson and gluons ( $c_g$ ).

1414 As mentioned before,  $\lambda_{HHH}^{SM}$  has small value but BSM model allows larger values of  
 1415  $\lambda_{HHH}$ . We can define  $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$  and for SM production modes,  $\kappa_\lambda = 1$ .

- 1416 • **Vector Boson Fusion (VBF HH)** - HH production via VBF process is the sub-  
 1417 leading mode, where a soft emission of two massive vector bosons ( $V = W, Z$ ) from  
 1418 the colliding partons (quarks) is followed by their fusion leading to the hard scat-  
 1419 tering  $VV \rightarrow HH$ . At  $\sqrt{s} = 13$  TeV, the cross section of VBF HH process is 1.73  
 1420 fb at N3LO QCD accuracy [36, 37]. The representative Feynman diagrams for the  
 1421 VBF HH process at LO are shown in Fig. 2.10.

1422 The most interesting fact about VBF HH process is the unique and direct access to  
 1423 the coupling of Higgs bosons pair with vector bosons pair,  $c_{2V}$ , whereas the other  
 1424 two diagrams contains the self-coupling  $\lambda_{HHH}$  which is mainly constrained from  
 1425 measurements of HH production via ggF and the coupling of a single Higgs boson  
 1426 with a pair of vector bosons  $c_V$  is constrained by measurements of vector boson  
 1427 associated production of a single Higgs boson and the decay of the Higgs boson to  
 1428 a pair of bosons. Despite having a very small cross section, a very small change

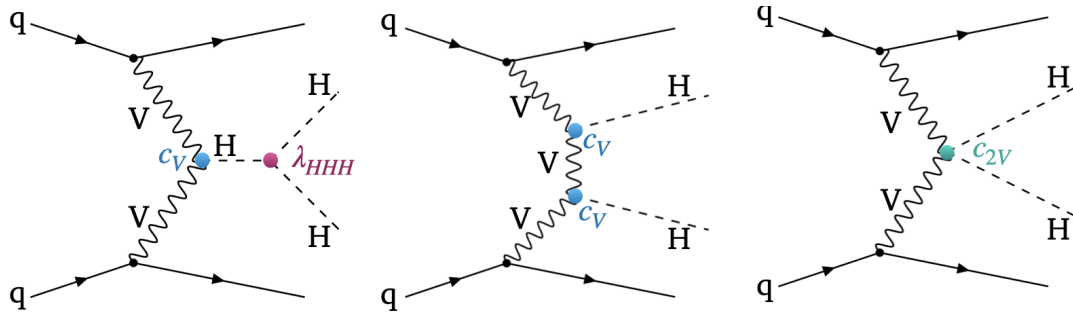


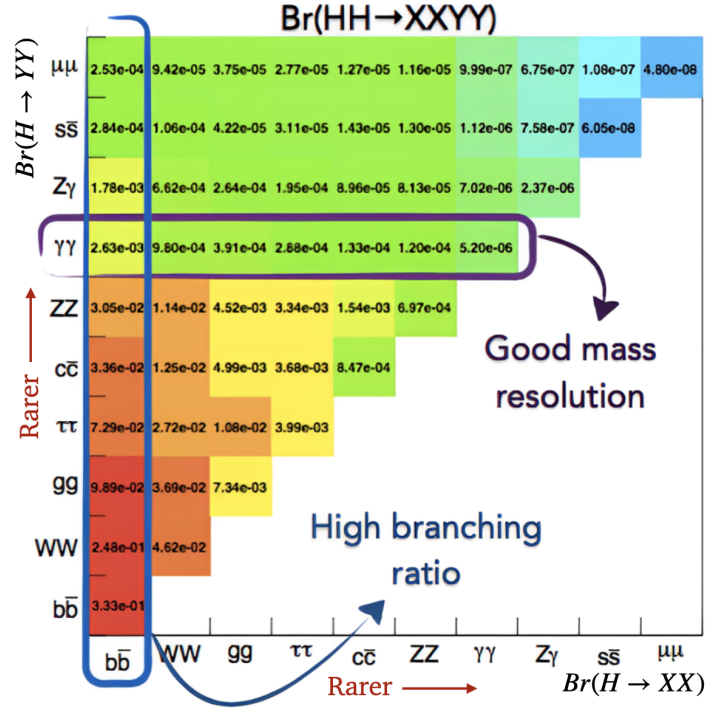
Figure 2.10: Feynman diagrams that contribute to the production of Higgs boson pairs via VBF at LO. On the left the diagram involving the HHH vertex ( $\lambda_{HHH}$ ), in the middle the diagram with two HVV vertices ( $c_V$ ), and on the right the diagram with the HHV vertex ( $c_{2V}$ ).

1429 in the couplings can induce a striking increase of the cross section; this enhanced  
 1430 sensitivity potentially adds extra impact on the measurement and the constraints  
 1431 on  $\lambda_{HHH}$  and  $c_{2V}$ . Anomalous values of  $c_{2V}$  can be investigated to establish the  
 1432 presence of the HHVV-mediated process as a probe of BSM physics.

1433 There are other possible production modes like VHH and  $t\bar{t}HH$  that have very small  
 1434 cross-section at the LHC  $\sqrt{s} = 13$  TeV. They can be included in future projects with  
 1435 higher centre-of-mass energy and luminosity.

1436 In the search for di-Higgs boson production, the crucial step is to choose a decay channel.  
 1437 This choice depends on the purity, selection efficiency and branching ratio of the channel.  
 1438 Fig. 2.11 shows the possible decay channels, which are explored and analyzed by various  
 1439 experiments to understand di-Higgs physics.

1440 In this thesis, we will focus on the decay channel  $HH \rightarrow bb\gamma\gamma$  as  $H \rightarrow bb$  has the largest  
 1441 branching ratio but high background rates and  $H \rightarrow \gamma\gamma$  has excellent mass resolution  
 1442 with lower background but low branching ratio. In a way, these two channels complement  
 1443 each other and we are trying to get the best out of each.


 Figure 2.11: SM di-Higgs decay branching ratio (BR) for  $m_H = 125$  GeV

## 1444 2.5 Higgs Boson Couplings

### 1445 2.5.1 Higgs Boson Trilinear Self-Coupling

1446 The parameter  $\lambda_{HHH}$  is the Higgs boson trilinear self-coupling. It arises in the SM from  
 1447 the expansion of the Higgs field potential around its vacuum expectation value  $v/\sqrt{2}$ . In  
 1448 the SM theory, the predicted  $\lambda_{HHH}$  value at the leading order is

$$\lambda_{HHH}^{SM} = \frac{m_H^2}{2v^2} = (1.291 \pm 0.003) \cdot 10^{-1} \quad (2.6)$$

1449 where  $m_H$  is the Higgs boson mass with a measured value  $m_H = 125.10 \pm 0.14$  GeV  
 1450 [38]. The value of  $v$  predicted by the SM is 246.22 GeV derived with an extremely good  
 1451 precision from the Fermi coupling constant. The direct measurement of the  $\lambda_{HHH}$  value  
 1452 provides a consistency test of the spontaneous symmetry breaking mechanism. On the  
 1453 other hand, several BSM theories predicts anomalous values of the  $\lambda_{HHH}$  value, such as  
 1454 the composite Higgs models [29, 39] and in general the Higgs-portal models [40]. It is  
 1455 practical to define  $\kappa_\lambda$  as the ratio of a BSM and the SM value.

1456 As discussed before, SM cross section of HH production is very small. However, anomalous  
 1457  $\lambda_{HHH}$  values could significantly increase the cross section, as visible in Fig. 2.12, and make  
 1458 the ggF HH process observable. As visible in the same figure, the cross sections of other  
 1459 HH production modes, i.e. the HH associated production with a  $t\bar{t}$  quark pair  $t\bar{t}HH$ ,  
 1460 and the HH associated production with a vector boson VHH, are also sensitive to the  
 1461  $\lambda_{HHH}$  parameter. However, the cross section of these processes is much smaller than  
 1462  $\sigma_{ggHH}$ , hence also the corresponding sensitivity to  $\lambda_{HHH}$  is reduced. In this work, only  
 1463 the qqHH (VBF HH) mechanism is considered along with the ggF HH mechanism, while  
 1464 the other HH production modes are neglected. In case of anomalous  $\lambda_{HHH}$  values, the  
 1465 HH invariant mass ( $m_{HH}$ ) distribution could also be strongly modified, as shown in Fig.  
 1466 2.12. This feature is properly exploited in the analysis to increase the sensitivity to the  
 1467  $\lambda_{HHH}$  parameter.

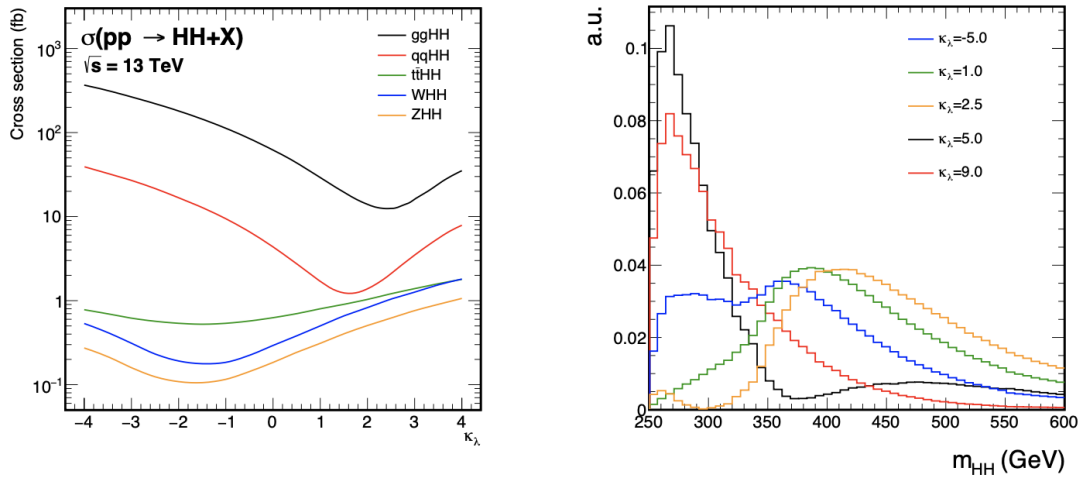


Figure 2.12: *Left*: cross sections of the main HH production modes as a function of  $\kappa_\lambda$ . *Right*:  $m_{HH}$  distributions for the ggF HH process for different  $\kappa_\lambda$  hypotheses. The distributions are all normalized to unity.

## 2.5.2 Yukawa Coupling of the Higgs Boson to the top Quark

1469 The Yukawa coupling of the Higgs boson to the top quark is predicted by the Standard  
 1470 Model with a strength:

$$y_t^{SM} = \frac{\sqrt{2}m_t}{v} = 0.992 \pm 0.002 \quad (2.7)$$

1471 where  $m_t$  is the top quark mass with a value  $m_t = 172.76 \pm 0.30$  [38]. As the Yukawa  
 1472 coupling of the Higgs boson with a fermion is proportional to the fermions mass,  $y_t$  is



1473 the strongest coupling of the Higgs boson with a fermion. The fact that its predicted  
 1474 value is close to the unity suggests that the interaction of the Higgs boson with the top  
 1475 quark might have some special role not disclosed in the SM. The Higgs boson - top quark  
 1476 interaction induces also very large corrections to the SM Higgs boson potential which  
 1477 could produce additional minima in the Higgs field potential. The minima values are  $y_t$ -  
 1478 dependent and determine the SM vacuum (meta)stability [41]. On the other hand, some  
 1479 BSM theories predict deviations of  $y_t$  from the SM prediction up to 20 - 30% [42]. The  
 1480 measurement of the ggF HH and of the  $H \rightarrow \gamma\gamma$  decay width provide access to the  $y_t$   
 1481 parameter through loop-induced processes. In such processes other BSM phenomena, e.g.  
 1482 new particles in the loop, could modify the final observable, enhancing or compensating  
 1483 the effect of an anomalous  $y_t$  value. The variation of the HH cross sections for anomalous  
 1484 values of  $y_t$  is shown in Fig. 2.13, where for simplicity the  $\kappa_t$  parameter is defined as  
 1485  $\kappa_t \equiv y_t/y_t^{SM}$ .

1486 The ggF HH cross section  $\sigma_{ggHH}$  is very sensitive to  $|\kappa_t| \gg 1$ , because in that case the  
 1487 Feynman diagram with a box loop of top quarks is the dominating matrix element, thus  
 1488  $\sigma_{ggHH}$  scales as  $\kappa_t^4$ .

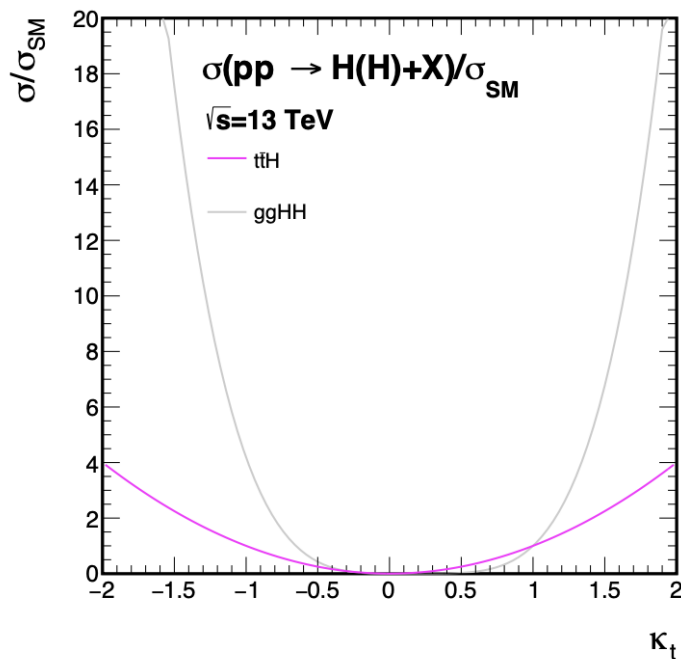


Figure 2.13: Variations of the ggF HH cross sections as a function of  $\kappa_t$ .

### 1489 2.5.3 Higgs Boson Couplings to the Vector Bosons

1490 The spontaneous symmetry breaking mechanism induces the couplings  $c_Z$  ( $c_W$ ) between  
 1491 one Higgs boson and two Z (W) bosons, as well as the couplings  $c_{2Z}$  ( $c_{2W}$ ) between two  
 1492 Higgs bosons and two Z (W) bosons. In the assumption that the BSM phenomena affect  
 1493 in the same way the  $c_Z$  and  $c_W$  couplings, it is practical to define  $c_V$  and  $c_{2V}$  as:

$$\begin{aligned} c_V &= c_Z/c_Z^{SM} = c_W/c_W^{SM} \\ c_{2V} &= c_{2Z}/c_{2Z}^{SM} = c_{2W}/c_{2W}^{SM} \end{aligned} \quad (2.8)$$

1494 The measured value of  $c_V$  is consistent with the SM prediction with an uncertainty of  
 1495 about 10% [43]. In the SM, considerations of perturbative unitarity of the VBF HH cross  
 1496 section [44] require that the  $c_{2V}$  and the  $c_V$  values are related through the relation:

$$c_{2V} = c_V^2 \quad (2.9)$$

1497 It is fundamental to test the relation in Eq. 2.9 because its violation, hence a violation  
 1498 of the perturbative unitarity, would be a clear signature of a BSM dynamics. In fact, a  
 1499 BSM dynamics would be required at a higher energy scale to re-establish the unitarity.  
 1500 Therefore, a direct observation of the  $c_{2V}$  coupling, and a measurement of its strength, is  
 1501 a very important step toward the full characterization of the Higgs boson properties. The  
 1502 observation of the  $c_{2V}$  coupling is challenging because the most sensitive physics process,  
 1503 accessible at the LHC, is the VBF HH whose SM cross section is expected to be only 1.73  
 1504 fb. Thus, the collected data are expected to provide only an upper limit to it. However,  
 1505 anomalous values of  $c_{2V}$  could significantly increase the VBF HH cross section, as shown  
 1506 in Fig. 2.14, making the process observable.

### 1507 2.5.4 BSM Higgs Boson Couplings

1508 Theoretical considerations [45] indicate that the scale of the new physics  $\Lambda$ , e.g. the  
 1509 mass of new particles not predicted by the SM, could be at the TeV scale. From the  
 1510 experimental point of view, the direct searches performed so far using the LHC data

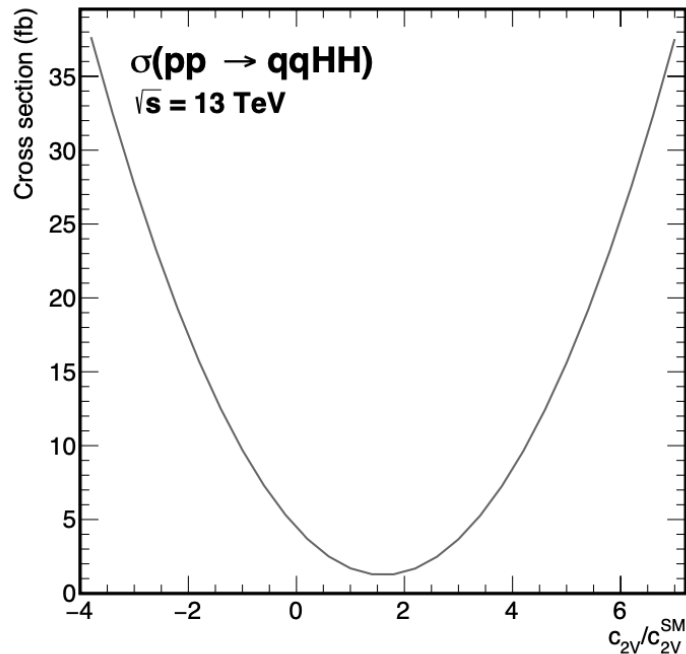


Figure 2.14: Variations of the VBF HH cross section as a function of  $c_{2V}$ .

1511 exclude the presence of BSM resonances typically up to around 1 TeV. The possibility  
 1512 to probe masses beyond 1 TeV with direct searches is limited at LHC by the available  
 1513 center-of-mass energy and collected data. However, the indirect probe of high energy  
 1514 BSM phenomena at a smaller and accessible energy scale is possible thanks to radiative  
 1515 or perturbative effects. The SM effective field theory (SMEFT [29]) approach allows a  
 1516 quasi model-independent description of a phenomenon at an energy scale  $E \ll \Lambda$ . The  
 1517 only remnants of the high-energy dynamics are in the low-energy couplings and in the  
 1518 symmetries of the EFT. With the SMEFT approach new operators are added to the SM  
 1519 Lagrangian. Such operators are built using the SM particle fields and ensuring the SM  
 1520 Gauge and Lorentz invariance. As a consequence, new effective couplings between the  
 1521 SM particles and modifications of the SM coupling constants could arise.

1522 As visible in Fig. 2.9, the HH production via gluon fusion, is sensitive to five Higgs EFT  
 1523 coupling constants  $\kappa_\lambda$ ,  $\kappa_t$ ,  $c_g$ ,  $c_{2g}$ , and  $c_2$  controlling the strength of the corresponding  
 1524 EFT operators. The  $c_{2g}$  and  $c_2$  couplings are effective couplings that can be induced by  
 1525 loops dominated by new heavy BSM particles. The impact of the EFT couplings on the  
 1526 HH observables is double:

- 1527 • They induce a variation, typically an increase, of the inclusive HH production cross

1528 section.

- 1529 • They significantly modify the differential HH cross section. In particular, the dis-  
1530 tribution of the di-Higgs invariant mass  $m_{HH}$  can dramatically change in case of  
1531 anomalous couplings.

1532 Therefore, the measured HH cross section together with the  $m_{HH}$  differential information  
1533 can be used to constrain the EFT parameters.

### 1534 BSM Benchmarks to Probe the Sensitivity

1535 Because of the small HH cross sections values, the HH processes are not sufficient to  
1536 simultaneously constrain the  $\kappa_\lambda$ ,  $\kappa_t$ ,  $c_2$ ,  $c_g$  and  $c_{2g}$  couplings. For this reason, twelve  
1537 points in the five parameters space are selected to be representative of the HH kinematics  
1538 for all the possible anomalous couplings scenarios [20,21]. Such points are called BSM  
1539 benchmarks. Typically in the HH searches, the data compatibility with each specific BSM  
1540 benchmark is tested. If no BSM evidences are found, upper limits on the benchmark cross  
1541 sections are extracted. The coupling values for each benchmark is reported in Tab. 2.1,  
1542 while the corresponding  $m_{HH}$  distributions are visible in Fig. 2.15.

Benchmark	$\kappa_\lambda$	$\kappa_t$	$c_2$	$c_g$	$c_{2g}$
0	7.5	1.0	-1.0	0.0	0.0
1	1.0	1.0	0.5	-0.8	0.6
2	1.0	1.0	-1.5	0.0	-0.8
3	-3.5	1.5	-3.0	0.0	0.0
4	1.0	1.0	0.0	0.8	-1.0
5	2.4	1.0	0.0	0.2	-0.2
6	5.0	1.0	0.0	0.2	-0.2
7	15.0	1.0	0.0	-1.0	1.0
8	1.0	1.0	1.0	-0.6	0.6
9	10.0	1.5	-1.0	0.0	0.0
10	2.4	1.0	0.0	1.0	-1.0
11	15.0	1.0	1.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

Table 2.1: Coupling values for the twelve defined BSM benchmarks.

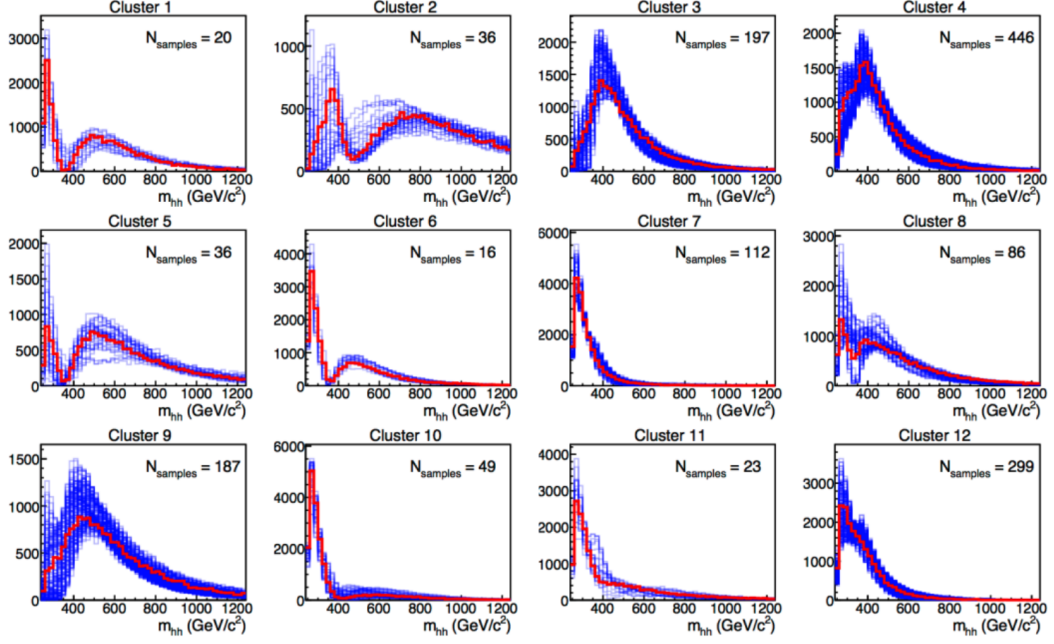


Figure 2.15: Generator-level distributions of di-Higgs boson mass for the clustered benchmarks from [46] are shown. The red distributions correspond to the chosen benchmark sample in each cluster, while the blue ones describe the other members of each cluster.

## 2.6 TRSM and HHH Production

1543

1544 The quartic self-coupling of the Higgs boson can be measured from the production of  
 1545 triplet Higgs bosons. However, due to the very small impact cross section ( $\sim 80$  ab at  
 1546  $\sqrt{s} = 13$  TeV) the search for triplet Higgs boson production in the context of the SM is  
 1547 beyond the reach of the LHC.

1548 The two-real-scalar-singlet extension of the SM (TRSM) [13] extends the scalar sector of  
 1549 the SM by additional scalar fields that transform as singlets under the SM gauge group.  
 1550 In the TRSM, the gluon-fusion  $pp \rightarrow HHH$  cross section is enhanced via the resonant  
 1551 production of  $X$ .

1552 One of the simplest ways to realise this is through models that extend the SM scalar  
 1553 sector by two additional singlet fields. The most general extension of the SM by  $n$  real  
 1554 scalar singlet fields  $\phi_i (i \in [1, \dots, n])$  has a scalar potential of the form

$$V(\Phi, \phi_i) = V_{\text{singlets}}(\Phi, \phi_i) + V_{SM}(\Phi) \quad (2.10)$$

1555 Here,  $\Phi$  describes the scalar  $SU(2)_L$  doublet field of the SM and  $V_{SM}$  denotes the scalar  
 1556 potential of the SM. It is not possible to write down gauge invariant and renormalizable in-  
 1557 teractions between a scalar singlet and any of the SM fermions. The singlets will therefore  
 1558 only interact with the SM Higgs boson through the couplings of the scalar potential.

1559 We consider here a specific version, TRSM [47], where in addition two  $\mathbb{Z}_2$  symmetries are  
 1560 imposed, leading to a reduction of the available number of degrees of freedom. Depending  
 1561 on the masses, the heaviest scalar can decay to the two lighter scalars, which in turn decay  
 1562 to SM particles with the branching fractions depending on their masses. Here, however,  
 1563 we will focus on the scenario where the state  $H$  is identified with the SM-like Higgs boson,  
 1564 and  $Y$  and  $X$  are two new heavier scalars obeying the mass hierarchy,

$$2m_H < m_Y < (m_X - m_H) \tag{2.11}$$

1565 This results in two real scalar singlets  $X$  and  $Y$  decaying to the SM Higgs boson  $H$ . The  
 1566 Feynman diagram for gluon-gluon production of said process is shown in Fig. 2.16

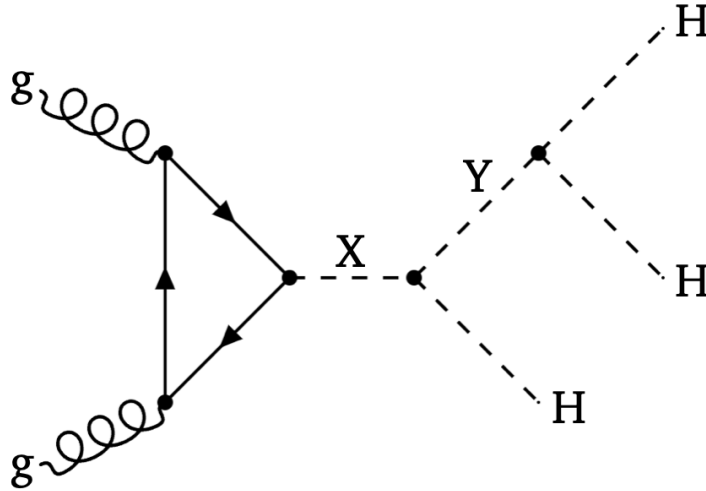


Figure 2.16: Feynman diagram showing the gluon-gluon production mode of a heavy scalar  $X$  followed by its decay process  $X \rightarrow Y(HH)H$

1567 As mentioned in earlier chapters,  $H \rightarrow b\bar{b}$  has highest branching ratio and as triple-  
 1568 Higgs production has lower cross-section, we will focus on the case where all three Higgs  
 1569 bosons are decaying to  $b\bar{b}$  resulting in  $X \rightarrow Y(HH)H \rightarrow b\bar{b}b\bar{b}b\bar{b}$ .

## Chapter 3

# Experimental Setup

CERN, or the European Laboratory for Particle Physics, is an international research centre that operates the largest particle physics laboratory in the world. It sits astride the Franco-Swiss border west of Geneva and was founded in 1954 by twelve European countries. It was initially dedicated to the fields of nuclear and particle physics: its original name stands for Conseil Européen pour la Recherche Nucléaire, or European Council for Nuclear Research. Today, as our understanding of matter goes much deeper than the nucleus, the laboratory is oriented towards particle physics research. It has become an example of international scientific collaboration, with more than 13000 collaborators of over 100 nationalities representing more than 500 universities and institutes. As particle physics demands the ultimate in performance, CERN is at the forefront of technology development and knowledge transfer, and most notably served as the birthplace of the World Wide Web (WWW) in 1989.

CERN's current major facility is the Large Hadron Collider (LHC), the largest and most powerful particle accelerator ever built. It is a circular proton accelerator designed to reach a centre-of-mass energy of 14 TeV. Built between 1998 and 2008, the design of the LHC was largely driven to profit from the pre-existing CERN infrastructures: the LHC is installed in a 26.7 km long tunnel that was built to host its predecessor, the Large Electron Positron (LEP) collider, located between 45 m and 170 m below ground level. The LHC hosts two beam-pipes where protons circulate in opposite directions, and which are brought to collision at four interaction points, where four particle detectors

1592 are installed. At one of this points sits the Compact Muon Solenoid (CMS) experiment,  
1593 a general-purpose detector designed to explore a broad range of physics processes, from  
1594 precision electroweak measurements to searches of supersymmetric particles. It is with  
1595 the proton collision data collected by this detector that the analysis described in thesis  
1596 was conducted

1597 This chapter gives an overview of the LHC accelerator and the CMS detector. Section  
1598 3.1 reviews the design and parameters of the LHC. The CMS sub-detector structure with  
1599 trigger system is presented in Section 3.2.

## 1600 3.1 The Large Hadron Collider

1601 The LHC was designed to deliver proton-proton ( $pp$ ) collisions at an unprecedented max-  
1602 imum centre-of-mass energy of  $\sqrt{s} = 14$  TeV with a very high instantaneous luminosity of  
1603  $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . It was conceived to investigate the nature of the spontaneous symmetry  
1604 breaking through the search of the Higgs boson, which was observed by the ATLAS and  
1605 CMS collaborations in 2012 [1, 2]. Additionally, it was intended to scan the accessible  
1606 phase space in the search of new phenomena beyond the SM, aiming at favouring or ruling  
1607 out the postulated scenarios. Complementary to the proton runs, a physics program of  
1608 heavy ion collisions (Pb-Pb) is also carried out with the goal of studying the collective  
1609 behaviour of quarks and gluons in plasma.

1610 The realization of the LHC constituted a two decade-long international journey. Its first  
1611 proposal dates back to 1984 with the official recognition of the project, subsequently  
1612 approved in 1994 and inaugurated in 2008. Two eras of physics operations have already  
1613 been conducted: Run 1, which lasted from 2009 to 2013, and Run 2, from 2015 to 2018.  
1614 The LHC is currently in the third data-taking era (Run 3) that will take place from  
1615 2022 to 2025. After that, the LHC and the accelerator complex will undergo a profound  
1616 upgrade towards the High Luminosity LHC (HL-LHC), scheduled to start in 2029. A  
1617 description of the accelerator complex and operations is given in the following.



### 3.1.1 Accelerator complex

The complete accelerator complex is illustrated in Fig. 3.1. The LHC is the last ring in a chain of particle accelerators, built well before the LHC and upgraded to meet its stringent requirements. The first step of the chain consists in the extraction of protons from a bottle of hydrogen gas making use of a strong electric field. The protons are then sent to a Radio Frequency Quadrupole (RFQ), where they are grouped into bunches and accelerated until they reach an energy of 750 keV. After that, the protons are supplied to the Linear Accelerator (LINAC 2), which brings the proton beam to an energy of about 50 MeV. The particles then arrive to the first circular collider, the Proton Synchrotron Booster (PSB), a 150 m ring that accelerates the beam up to an energy of 1.4 GeV and increases the intensity of the proton bunches. Next, the beam enters the Proton Synchrotron (PS) and then the Super Proton Synchrotron, two circular accelerators of 620 m and 6912 m in length which raise the energy of the beam to 26 GeV and 450 GeV, respectively.

The proton bunches are fed into the LHC with fast kicker magnets, which split the beam into two parallel beamlines that travel in opposite directions in the LHC tunnel. Once in the LHC, the beams are further accelerated to their maximal energy. The acceleration is performed in the high frequency accelerating cavities, placed in eight 545 m long straight sections along the ring. The trajectory of the beam is bent with 1232 superconducting dipole magnets placed throughout eight 2.45 km long arcs. These magnets generate a field of 8.3 T and need to be cooled down to a temperature of 1.9 K (-271.25°C) with superfluid helium-4. This structure is shown in Fig. 3.2, and the LHC contains 1232 such magnets for bending the beam. Aside from the dipole magnets, the LHC contains many other magnets, mainly quadrupole for correcting and stabilizing the beams to keep the particles focused in narrow beams.

Once the proton beam reaches the nominal energy and the beam is stabilized, protons are brought to collide at four different points along the LHC instrumented with particle detectors. ATLAS (A Toroidal LHC Apparatus) [50] and CMS (Compact Muon Solenoid) [51] are multipurpose detectors which can measure the products of both proton and heavy-ion collisions. They are installed in the diametrically opposite points of the LHC, where the highest instantaneous luminosity is achieved. The LHCb (LHC beauty) [52] experiment

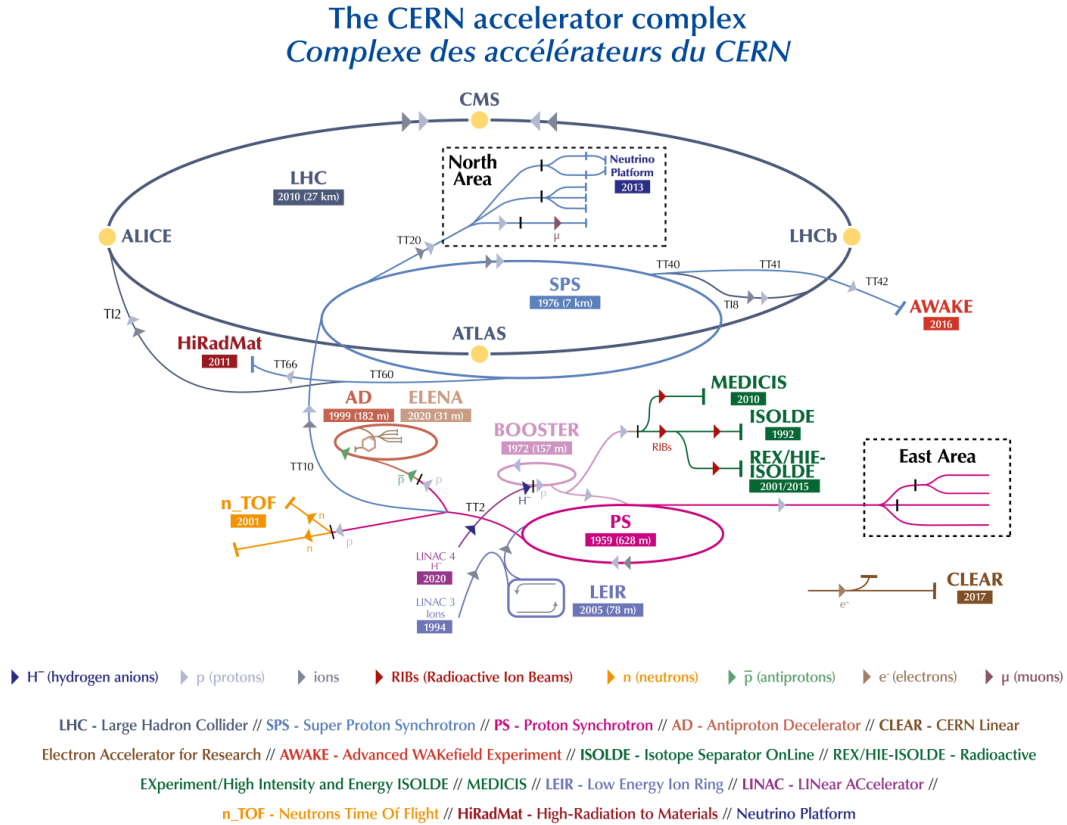


Figure 3.1: Illustration of the accelerator complex at CERN. Protons are accelerated to increasing energies at the LINAC 2 (in light pink), Booster (in light pink), PS (in dark pink), SPS (in light blue) and LHC (in grey) accelerators. The counter-circulating proton beams at the LHC collide in the centre of the CMS, ATLAS, LHCb and ALICE detectors [48].

1649 consists of an asymmetric single-arm detector devoted to heavy flavour quarks physics;  
 1650 its primary goal is to search for evidence of new physics in charge-parity (CP) violation  
 1651 and rare decays. The last experiment, ALICE (A Large Ion Collider Experiment) [53],  
 1652 was designed to cope with very high particle multiplicities and is mainly devoted to the  
 1653 study of quark-gluon plasma in heavy-ion collisions.

### 1654 3.1.2 Nominal Design Parameters

1655 The LHC accelerates protons, which are charged, composite and stable particles. Being  
 1656 fundamental particles, electrons would be more appropriate for precision measurements,  
 1657 but protons have the advantage that they suffer much smaller synchrotron radiation losses  
 1658 due to their higher mass. This type of radiation is emitted when a charged particle is

## LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

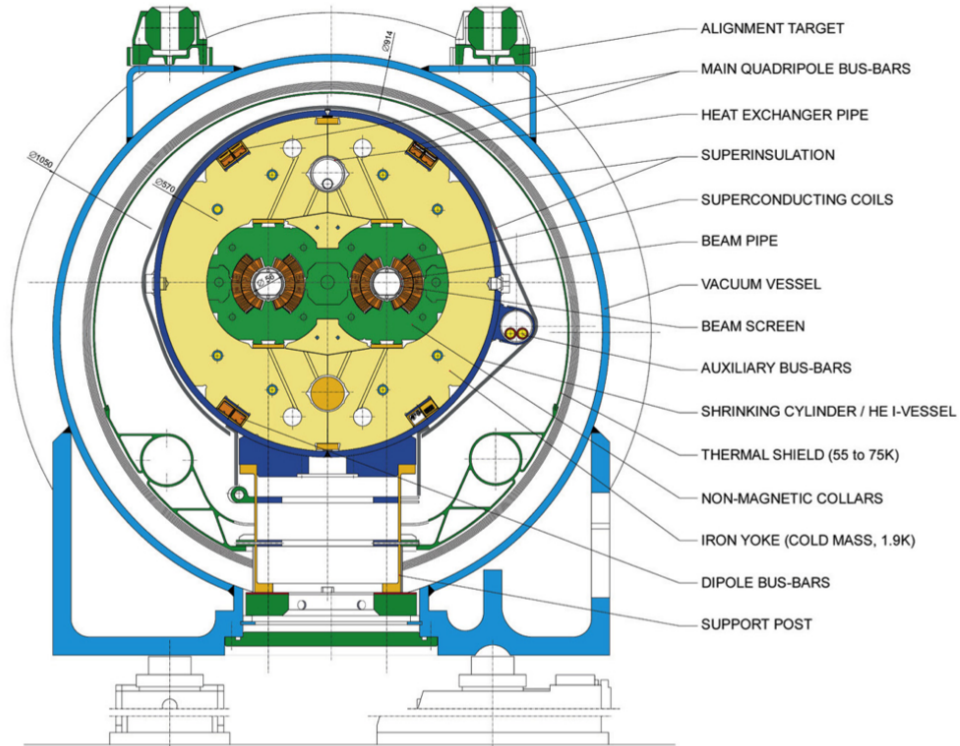


Figure 3.2: Cross section of an LHC dipole magnet [49].

1659 accelerated radially; it produces energy losses which limit the maximum reachable energy  
 1660 in a circular collider. Thus a proton collider can achieve much higher energy than an  
 1661 electron collider, the maximum value being limited by the capacity of the magnets to  
 1662 maintain the protons in the circular trajectory. The design *centre-of-mass* energy of the  
 1663 proton-proton collisions at the LHC is  $\sqrt{s} = 14$  TeV, meaning each beam has an energy of  
 1664 7 TeV. An important fraction of the momentum of the proton is carried by the sea quarks  
 1665 and gluons that compose it; it is therefore possible to generate interesting physics without  
 1666 colliding protons with their antiparticles, which are much more difficult to produce.

1667 Alongside the beam energy, a key parameter of the LHC machine is the *instantaneous*  
 1668 *luminosity*  $\mathcal{L}$ , which characterizes the collision rate and serves as an indicator of its  
 1669 performance. It relates the number of events per unit time  $\partial N/\partial t$  produced for a given  
 1670 process with its cross section  $\sigma$  via

$$\frac{\partial N}{\partial t} = \mathcal{L} \times \sigma. \quad (3.1)$$

1671 A large instantaneous luminosity is essential to produce low probability processes such  
1672 as  $t\bar{t}H$  and  $tH$ , but it also represents a challenge for the data acquisition system. Upon  
1673 integration of  $\mathcal{L}$  over time, one obtains the *integrated luminosity*,  $L = \int \mathcal{L} dt$ , which  
1674 characterizes the amount of data produced. The instantaneous luminosity is usually  
1675 expressed in units of  $\text{cm}^{-2}\text{s}^{-1}$ , while the integrated luminosity is expressed in units of  
1676 inverse picobarns ( $\text{pb}^{-1}$ ) or femtobarns ( $\text{fb}^{-1}$ ).

1677 Under the assumption that the two counter-rotating beams are identical, the instanta-  
1678 neous luminosity relates to the beam properties as

$$\mathcal{L} = \frac{N_p^2 n_b f \gamma_r}{4\pi \epsilon_n \beta^*} F, \quad (3.2)$$

1679 where  $N_p$  is the number of protons per bunch and  $n_b$  is the number of bunches. The  
1680 symbol  $f$  represents the revolution frequency of the bunches and  $\gamma_r$  is the relativistic  
1681 factor. The transverse emittance  $\epsilon_n$  characterizes the confinement of the beam in space  
1682 and momentum, whereas the beta function  $\beta^*$  represents its focus at the interaction point.  
1683 Finally,  $F$  is a geometric factor which accounts for the luminosity reduction due to the  
1684 crossing-angle of the beams at the interaction point ( $F \leq 1$ ). The values of the LHC  
1685 design parameters are given in Tab. 3.1.

Symbol	Parameter	Nominal value
$\sqrt{s}$	Centre-of-mass energy	14 TeV
$\Delta t$	Bunch spacing	25 ns
$\mathcal{L}$	Instantaneous luminosity	$1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$
$n_b$	Number of bunches per beam	2808
$N_p$	Number of protons per bunch	$1.15 \times 10^{11}$
$f$	Revolution frequency	11245 Hz
$\epsilon_n$	Transverse emittance	$3.75 \mu\text{m rad}$
$\beta^*$	Beta function	0.55 m

Table 3.1: Design parameters of the LHC accelerator in proton-proton collisions [54].

1686 In its nominal design, the LHC accelerates and collides as many as 2808 proton bunches  
1687 per beam, each bunch containing about 115 billion protons. The bunches are grouped  
1688 in trains of 48 bunches ("48b" scheme) spaced in intervals of 25 ns each, and circulate  
1689 around the ring about 11000 times per second, only 3.1 m/s slower than the speed of light.  
1690 Of these, 2544 bunches collide at the CMS interaction point at a bunch collision rate of  
1691 40 MHz. This configuration yields a luminosity of  $\sim 1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  at the beginning of

1692 the fill, defined as the point when the proton injection is complete and the LHC cannot  
 1693 accommodate any more bunches.

1694 Any collider with high instantaneous luminosity faces an important drawback: the *pileup*  
 1695 (PU), defined as the number of simultaneous interactions taking place in each bunch  
 1696 crossing. The average PU is directly proportional to the instantaneous luminosity and  
 1697 relates to the beam properties as

$$\langle PU \rangle = \frac{\mathcal{L} \sigma_{pp}^{inel}}{n_b f} \quad (3.3)$$

1698 where  $\sigma_{pp}^{inel}$  is the inelastic *pp* cross section, which amounts to 69 mb at  $\sqrt{s} = 13$  TeV  
 1699 [55], leading to a nominal average pileup of  $\sim 22$  interactions per bunch crossing at  
 1700 the LHC, frequently exceeded during Run 2 operations. High pileup values result in a  
 1701 very high detector occupancy that degrades the efficiency and resolution of the particle  
 1702 reconstruction. The average number of simultaneous proton-proton (*pp*) interactions per  
 1703 bunch crossing (pileup) is shown on the left in Fig. 3.3.

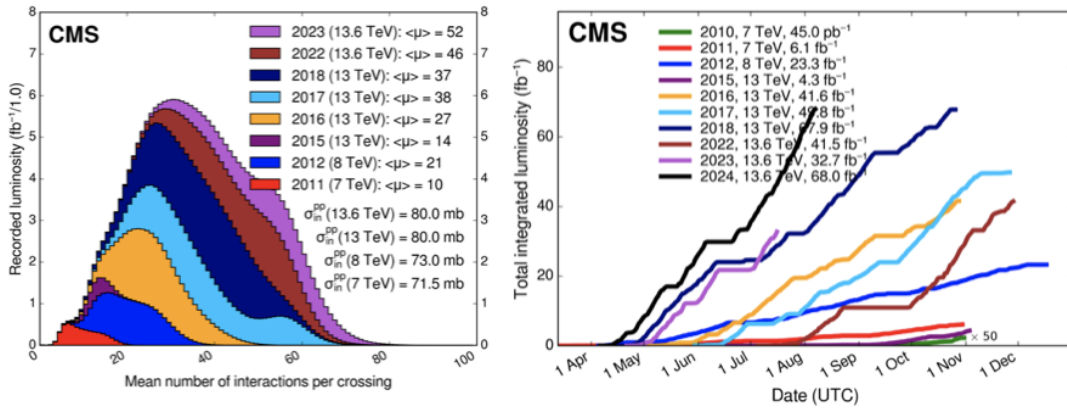


Figure 3.3: Average number of simultaneous *pp* interactions per bunch crossing (left) and integrated luminosity collected by the CMS experiment (right) by year of data taking [56].

1704 So far, the LHC has completed two successful runs of data taking: Run 1 (2010-2012), at  
 1705 the centre-of-mass energies of 7 and 8 TeV , and Run 2 (2015-2018), at the centre-of-mass  
 1706 energy of 13 TeV. Run 3 (2022-2025) is ongoing with 13.6 TeV centre-of-mass energy. A  
 1707 total integrated luminosity of  $340.6 \text{ fb}^{-1}$  was delivered to CMS by the LHC by August  
 1708 2024,  $163.6 \text{ fb}^{-1}$  of which at  $\sqrt{s}=13$  TeV during Run 2. Furthermore, a peak luminosity  
 1709 of above  $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  was achieved during Run 2, which amounts to twice the design  
 1710 value. The integrated luminosity delivered to the CMS experiment by year of data taking

1711 is shown in Fig. 3.3.

1712 The Long Shutdown 3 (LS3) is expected to start in 2026, concluding the *Phase 1* of the  
1713 LHC. During this period, the LHC and the CMS detector will undergo a profound upgrade  
1714 towards the High Luminosity LHC (HL-LHC), which will run in *Phase 2*. The goal of the  
1715 upgraded machine is to reach a peak instantaneous luminosity of  $5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , which  
1716 yields to a total integrated luminosity of about  $3000 \text{ fb}^{-1}$  after a decade of operations,  
1717 enhancing significantly the sensitivity to rare phenomena. The unprecedented collision  
1718 rate will produce an average pileup of  $\sim 140$ . In its ultimate configuration, the machine  
1719 could be pushed to an instantaneous luminosity of  $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , corresponding to  
1720 an average pileup of  $\sim 200$ .

## 1721 3.2 Compact Muon Solenoid

1722 The CMS (*Compact Muon Solenoid*) [51] detector is a cylindrical detector with a length  
1723 of 21.6 m and a diameter of 14.6 m. The term "compact" in its name refers to its weight of  
1724 14,500 tons, which is more than twice the weight of the ATLAS detector, which is roughly  
1725 twice as large. It is situated at the LHC point 5, close to the French village of Cessy, 100  
1726 meters below the surface in a cavern. CMS is a general-purpose detector that was initially  
1727 designed to precisely reconstruct the Higgs boson decay products. It is constructed with  
1728 many concentric subdetectors that complement each other in the characterization of the  
1729 different particles generated from the pp interactions. One of the key features of CMS is  
1730 the intense magnetic field induced by the solenoid magnet (see Fig. 3.4). It is possible to  
1731 precisely measure the momenta, trajectories, and interaction vertices of traversing charged  
1732 particles using pixel and strip trackers in close proximity to the interaction point since  
1733 these particles are bent under the effect of the field. The tracking system is surrounded  
1734 by electromagnetic and hadronic calorimeters, which are meant to measure and absorb  
1735 the energy of electrons, photons, and hadrons. Muons move through the calorimeters and  
1736 are picked up by the muon tracking systems at the edge of CMS.

1737 The LHC instantaneous luminosity sets a very difficult goal for the detector during the  
1738 data collecting. With an instantaneous luminosity of about  $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , a peak pileup  
1739 (PU) of almost 50 extra pp interactions is expected during collisions. This additional

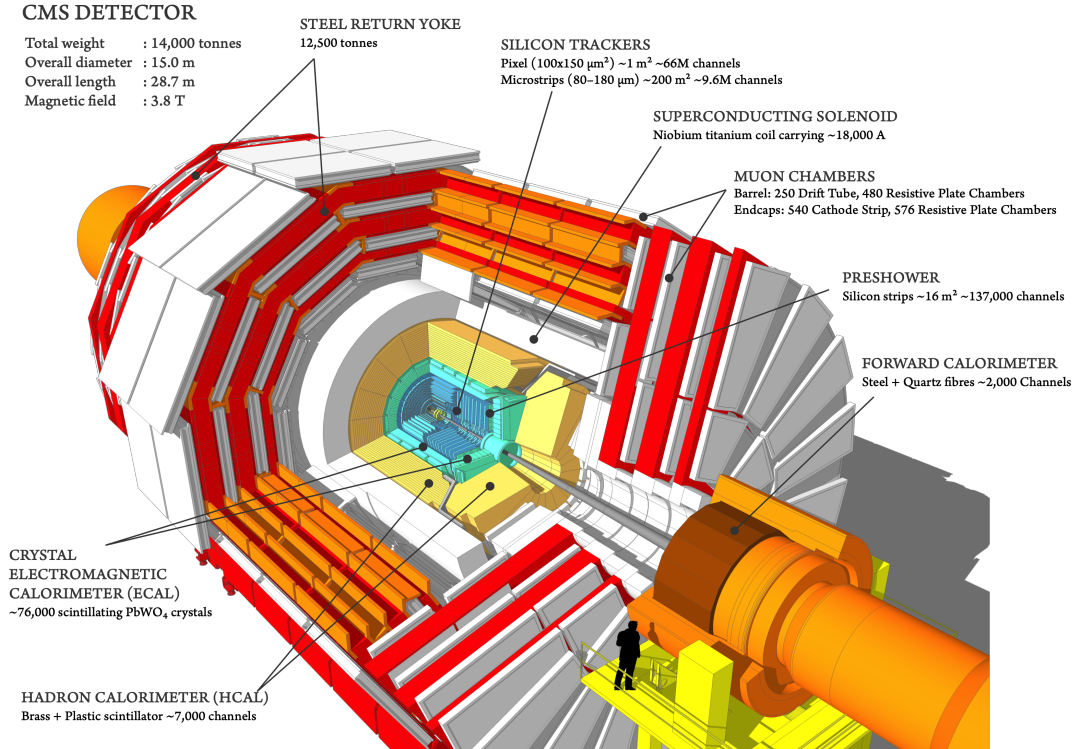


Figure 3.4: Schematic view of the CMS detector and its subcomponents [57]

1740 activity raises detector occupancy while lowering detector performance. Every 25 ns,  
 1741 collisions occur at the CMS's center at the interaction point (IP). A bunch crossing (in-  
 1742 time pileup) can have multiple interactions occurring within it. The hadronic activity  
 1743 that is not derived from the hard scattering process, which are referred to as underlying  
 1744 events. The underlying events are usually softer and can be separated from the signals  
 1745 of interest provided the high granularity, fast reaction, and wide solid angle of CMS,  
 1746 equipped with radiation-hard detectors and electronics. Additionally, the experimental  
 1747 signatures overlap (out-of-time pileup) happens. This indicates that new proton bunches  
 1748 are colliding at the CMS core before the decay products from the previous collision have  
 1749 reached the active chambers of the detectors.

1750 In achieving the wide range of physics goals, the main challenges for CMS are good  
 1751 electron and muon detection, high trigger efficiency and offline tagging of leptons and  
 1752 jets associated with b quarks, good electromagnetic energy resolution and good dijet,  
 1753 diphoton and dielectron mass resolution, and identification of missing transverse energy.  
 1754 Apart from efficient readout electronics performance, by construction the detector is prone  
 1755 to radiation damage, particularly in the forward regions.

1756 An overview of the CMS design and performance is given in the following subsections.

### 1757 3.2.1 CMS coordinate system

1758 The CMS coordinate system is right-handed as given in Fig. 3.5. The origin of the  
 1759 CMS coordinate system is located within the detector, with the y-axis pointing vertically  
 1760 upward, the x-axis pointing radially inward to the middle of the LHC ring, and the z-axis  
 1761 pointing along the beam direction. The detector design can be understood using the  
 1762 spherical coordinate system where  $r$  is the distance from the z-axis,  $\phi$  is the azimuthal  
 1763 angle (measured from the x-axis in the x-y plane), and  $\theta$  is the polar angle (measured  
 1764 from the z-axis).

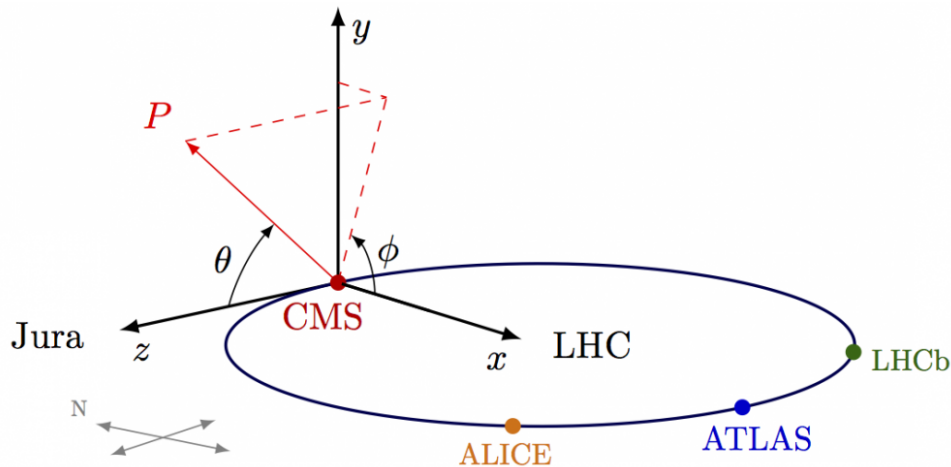


Figure 3.5: An illustration of the CMS detector with spherical co-ordinate system [58].

1765 A useful kinematic variable, which is often used in the LHC physics is rapidity ( $y$ ), where  
 1766 the rapidity difference of two particles ( $\Delta y = y_1 - y_2$ ) is invariant under Lorentz boosts  
 1767 along the z axis,  $y$  is defined for a particle having momentum 4-vector ( $E, p_x, p_y, p_z$ ) as:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (3.4)$$

1768 For a massless particle or for a particle with relatively small mass compared to its momen-  
 1769 tum ( $E \approx |\vec{p}| \gg m$ ), the definition of the rapidity can be approximated by a quantity,  
 1770 called pseudorapidity ( $\eta$ ), the pseudorapidity difference of two particles ( $\Delta \eta = \eta_1 - \eta_2$ )  
 1771 is also invariant under Lorentz boosts along the z axis. Pseudorapidity can be defined in  
 1772 terms of the polar angle  $\theta$  as,



$$\eta = -\ln \tan \frac{\theta}{2} \quad (3.5)$$

1773 The value of  $\eta$  is zero at the central part of the detector for polar angle =  $90^\circ$ , and  
 1774  $\eta = \pm\infty$  along the beam direction,  $\theta = 0, \pi$ . CMS detector covers a region up to  $\eta \sim \pm 5$ ,  
 1775 which corresponds to an angle of  $0.8^\circ$  wrt beam line.

1776 The transverse momentum is defined as the momentum projection on to the x-y plane and  
 1777 given as  $p_T = \sqrt{p_x^2 + p_y^2}$ . The usage of  $\eta$  and  $p_T$  at hadron colliders is motivated by the  
 1778 facts that the pseudorapidity difference and  $p_T$  are invariant under Lorentz boosts along  
 1779 the z axis. The beams enter along  $\pm z$ -axis within the detector; therefore, the transverse  
 1780 component for colliding particles is zero. Thus, after the collision, all outgoing particles  
 1781 should have the sum of transverse momenta equal to zero by following the momentum  
 1782 conservation law.

### 1783 3.2.2 Tracker

1784 The CMS tracker [59] sub-detector consists of a cylinder of 5.8 m in length and 2.6 m in  
 1785 diameter, placed at the inner most part at the detector. The main goal of the tracking  
 1786 system is to reconstruct the tracks of the charged particles as precise as possible and to  
 1787 reconstruct the position of the secondary vertices which is crucial for the study of the long  
 1788 lived particles (life time  $\tau > 1$  ps ) and to tag the quark-flavor of the jets. The tracking  
 1789 system was built under several basic requirements.

- 1790 • In the p-p collision mode of the LHC, the number of pileup events are very high, so  
 1791 the tracking system needs to reconstruct the tracks from the high particle flux and  
 1792 then needs to associate them to the correct vertices at the origin.
- 1793 • As the tracker system is closest to the detector interaction point, it experiences an  
 1794 enormously high radiation dose. So it is desirable that the tracking detectors are  
 1795 radiation hard.
- 1796 • For tracking, the measurement should be non destructive. Hence, the material  
 1797 budget is expected to be minimal to avoid the energy losses and multiple scattering

1798

of particles inside the tracker material before reaching to the calorimeter sectors.

1799

CMS exploits two different types of tracking sub-system: pixel detector and silicon strip

1800

detectors. Fig. 3.6 shows the longitudinal view of one quarter of the CMS tracker system

1801

with both types of CMS tracker subsystems.

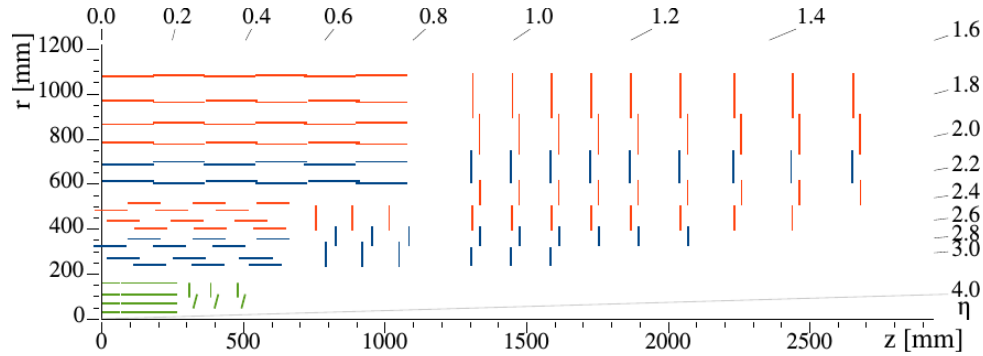


Figure 3.6: Sketch of one quarter of the Phase-1 CMS tracking system in  $r$ - $z$  view. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. [60].

1802

Given the LHC collision conditions, the main challenges for the tracker system are gran-

1803

ularity, response time, and radiation hardness. It helps to detect the charged particles by

1804

constructing their trajectory and measuring their momentum. The design of the tracker is

1805

a result of a compromise between providing the best detector performance and keeping the

1806

amount of inactive material as low as possible. The latter is a critical feature of the tracker,

1807

as a higher amount of passive material generates multiple scattering, bremsstrahlung, pho-

1808

ton conversion and nuclear interactions. These distort the measurement of the trajectory

1809

in the tracker and the measurement of the energy in the calorimeters just after.

1810

### Silicon Pixel Detector

1811

The pixel detector is the closest part of the tracker to the collision point and divided into

1812

124 million pixels with size of  $100\mu\text{m}$  by  $150\mu\text{m}$ . The barrel region of the detector has

1813

four layers with radii of 3 cm, 7 cm, 11 cm, and 16 cm and three discs on either side

1814

of the barrel (endcap regions) with distance of 29 cm, 39.6 cm, and 51.6 cm from the

1815

interaction point.

1816

Pixel detector provides three dimensional position measurement of the charged particles

1817

very close to the beam pipe. Due to radiation damage, the modules of the innermost layer

1818 of BPIX were substituted with new ones during the long shutdown 2 in 2021 [61]. I was  
1819 part of the installation and recommissioning of upgraded pixel detector.

## 1820 **Silicon Strip Detector**

1821 The silicon strip detector is the outer part of the tracking system with a coarse resolution in  
1822 position measurements than the pixel detector. The particle flux decreases with increasing  
1823 radial distance from the interaction point. The innermost part of the strip detector  
1824 consists of 4 concentric cylindrical layers of tracker inner barrel (TIB) and 3 tracker inner  
1825 disks (TID) on each side. Next layer is the outer silicon strip detector, divided into two  
1826 parts (i) tracker outer barrel (TOB) consisting of six silicon strip layers and (ii) tracker  
1827 end caps (TEC) consisting of 9 disks, each containing up to seven concentric rings of  
1828 silicon strips. Additionally, in the modules of the first two layers (rings) of TIB (TID)  
1829 and TOB and first, second and fifth rings of the TEC, a second micro-strip detector  
1830 module is mounted back-to-back to the first with a stereo angle of 100 mrad, called the  
1831 double sided or stereo module.

## 1832 **3.2.3 Calorimeters**

1833 Understanding what happened at the collision point depends on having information about  
1834 the energy of the different particles created in each collision. Two different types of  
1835 "calorimeters" in CMS are used to collect this data [62].

### 1836 **Electromagnetic Calorimeter (ECAL)**

1837 The CMS has a homogeneous electromagnetic calorimeter (ECAL) [63] made of fine-  
1838 grained 75,848 lead tungstate ( $\text{PbWO}_4$ ) crystals, just outside of the tracker system. The  
1839 homogeneous medium minimizes the sampling fluctuations and provides a better energy  
1840 resolution for photons and electrons. It is highly transparent and scintillates as electrons  
1841 and photons pass through it. In other words, it generates light in proportion to the  
1842 particle's energy. The crystals emit 80% of their light in less than 25 nanoseconds which  
1843 is the nominal time between successive bunch crossings at the LHC; this satisfies the

1844 requirement for quick detector response under LHC conditions. Since the light yield of  
 1845  $\text{PbWO}_4$  is temperature-dependent, a cooling system is needed to keep the crystals at  $\sim 18$   
 1846 degree Celsius. The photo-detectors are used to detect this scintillated light and convert  
 1847 it into an amplified electric signal.

1848 The crystals are arranged in barrel region (EB), covering pseudorapidity up to  $|\eta| = 1.48$ ,  
 1849 and in two endcap regions (EE), covering up to  $|\eta| = 3.0$ . EB has a crystal length of 230  
 1850 mm (220 mm in EE) corresponding to 26 (25) radiation lengths. The crystals on the front  
 1851 face have a transverse dimension of  $2.2 \times 2.2 \text{ cm}^2$  in EB ( $2.86 \times 2.86 \text{ cm}^2$  in EE). The total  
 1852 volume of the crystal is  $11 \text{ m}^3$ , and its weight is 92 t. The barrel calorimeter is divided  
 1853 into 36 supermodules, each with 1,700 crystals. The endcaps are divided into two dees,  
 1854 each with 3,662 crystals as shown in Fig. 3.7.

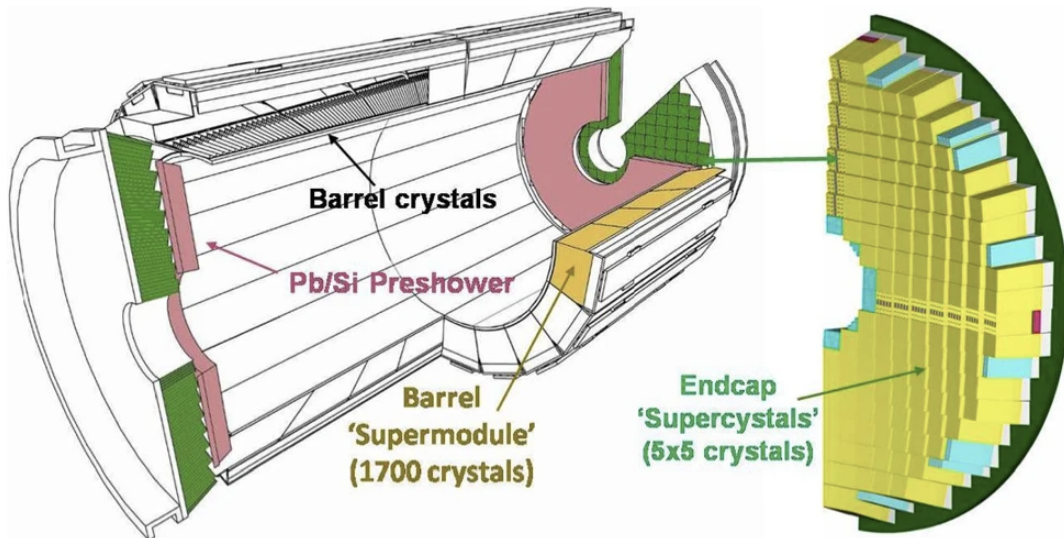


Figure 3.7: Schematic view of the ECAL showing the cylindrical barrel closed by the two endcap regions with one half endcap displayed [64].

1855 The photon separation is improved by a preshower detector (ES) based on lead ab-  
 1856 sorber and silicon strips sensors (4,288 sensors, 137,216 strips,  $1.9061 \text{ mm}^2$  with x-y view)  
 1857 mounted in front of the endcaps at  $1.65 < |\eta| < 2.6$ . The ES has a cumulative thickness  
 1858 of around three radiation lengths. It resolves the highly collimated photon pairs coming  
 1859 from the light and short-lived  $\pi^0$ - meson decay, which are not possible to resolve using  
 1860 only ECAL.

1861 The ECAL energy reconstruction is crucial for the rare physics searches with final states  
 1862 having charged leptons and photons, such as,  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ \rightarrow 4l$  and many others.

1863 **Hadron Calorimeter (HCAL)**

1864 Hadron calorimeter [65] completes the CMS calorimetric systems. The sampling calorimeter  
 1865 is made up of the active material (4 mm thick plastic scintillator tiles) placed be-  
 1866 tween copper absorber plates. HCAL determines a particle's location, energy, and arrival  
 1867 time as the particle passes through calorimeter. The active elements are read out using  
 1868 wavelength-shifting (WLS) plastic fibres. The scintillating light is collected by fibres and  
 1869 fed into readout boxes, where photo-detectors amplify the signal. The total amount of  
 1870 light in a given area, known as tower, is a measure of a particle's energy which is summed  
 1871 up over several layers of scintillator tiles in depth.

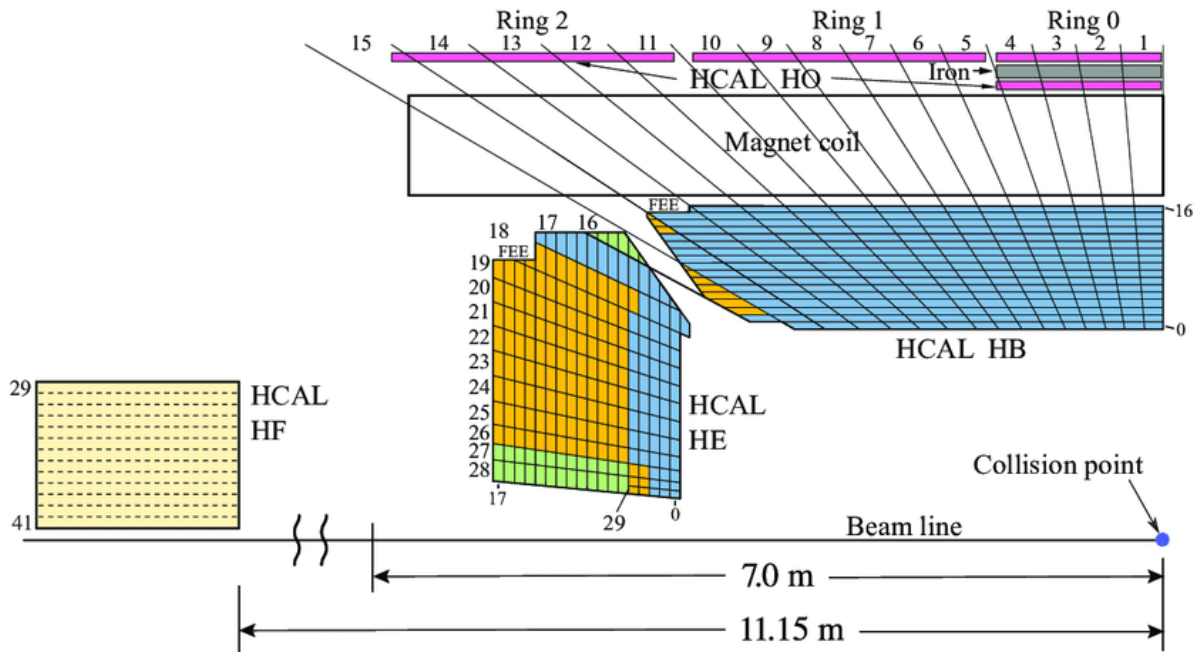


Figure 3.8: A schematic view of one quarter of the CMS HCAL, showing the positions of its four major components: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters [66].

1872 HCAL also consists of two parts: the barrel region (HB) and the endcap region (HE) as  
 1873 shown in Fig. 3.8. The absorber plates are 5 cm thick in the HB region and 8 cm thick  
 1874 in the HE region. The depth of the barrel HB is around 79 cm or 5.15 nuclear interaction  
 1875 length. Nuclear interaction length is the mean distance travelled by a hadronic particle  
 1876 before undergoing an inelastic nuclear interaction.

1877 As the material of the electromagnetic and hadronic calorimeters in the barrel may not  
 1878 provide enough stopping power for highly energetic particles in the central region ( $|\eta| <$

1879 1.4), the detector is complemented by an outer hadronic calorimeter (HO) located outside  
1880 the solenoid, composed solely of scintillating material. The CMS also uses a separate  
1881 forward calorimeter (HF) 6 m downstream of the HE endcaps. It extends the hermeticity  
1882 of the central HCAL system to a pseudorapidity of 5.0 (as needed for an excellent missing  
1883 transverse energy measurement). Quartz fibres are used as the active medium, and they  
1884 are contained in a copper absorber matrix. It is specifically sensitive to Cherenkov light  
1885 from neutral pions due to the quartz fibre active element. As a result, it has the unique  
1886 and attractive property of providing a highly localized response to hadronic showers.

1887 Along with measuring the energy of hadrons, HCAL also allows the detection of non-  
1888 interacting and uncharged particles as missing transverse energy (MET). Measuring these  
1889 particles is crucial because the measurement can reveal whether new particles have formed,  
1890 such as the supersymmetric particles (much heavier versions of the standard particles).  
1891 Some decay products of these new particles leave no trace of their existence in any part  
1892 of the CMS detector. To detect them, the HCAL must be hermetic, which means it must  
1893 catch any particle that emerges from the collisions to the greatest extent possible. We  
1894 can deduce the existence of the invisible particles if we see particles fly out on one side of  
1895 the detector but not on the other side, with an imbalance in momentum and energy.

#### 1896 3.2.4 Superconducting Solenoid Magnet

1897 A complex arrangement of niobium-titanium (Nb-Ti) coils, capable of carrying a current  
1898 of 19.5 kA and cooled by liquid helium, works as a superconducting solenoid magnet and  
1899 generates a 3.8 T magnetic field. It consists of four layers NbTi coils with 542 turns in  
1900 each of the layers. It is kept inside a liquid Helium cryostat at an operational temperature  
1901 of  $-268.65^\circ$  to reach the state of superconductivity. This magnet has an inner diameter of  
1902 6 m and a length of 12.5 m. The tracker and calorimeters are entirely contained inside  
1903 it. It is the main feature of the CMS detector, which bends the path of charged particles  
1904 while passing through the magnetic field.

1905 The magnetic field  $\vec{B}$  provokes the bending of the paths of the particles of non-zero charge  
1906  $q$  and speed  $\vec{v}$  in the transverse plane via the Lorentz force  $\vec{F}_L = q(\vec{v} \times \vec{B})$ ; the charge and  
1907 momentum of a particle can be inferred from this bending, alongside the measurement

performed by the tracker. To achieve the highest precision, the magnetic field must be accurately characterized over the entire volume of the experiment. The curvature of the path within the tracker depends on the energy and mass of the charged particle. It helps in particle identification and provides good momentum resolution. A 14 m iron return yoke surrounds the magnet coils and returns the magnetic flux through the muon chambers.

### 3.2.5 Muon Chambers

As the name of the detector “Compact Muon Solenoid” indicates, muon physics is a vital task for CMS. Muons are the charged leptons similar to electrons but 200 times heavier. Despite being a charged particle, it can penetrate the detector for several meters as it interacts weakly and deposits little energy within calorimetric systems. Therefore, the outer part of the detector is entirely covered by muon chambers [67] to detect muons, almost the only surviving particles reaching the muon chambers. There are four muon stations outside the solenoid and interleaved with iron return yoke plates. They are used to reconstruct the hits made by muons while passing through them. Muons also leave hits within the tracker. The strong solenoidal magnetic field bends the muon track which helps in measuring muon’s momentum. The hits within the tracker are combined with hits within the muon chambers for energetic muons.

The muon system contains gas ionization chambers. There are 1400 chambers in total. The 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) monitor the particles’ positions and provide a trigger, and 610 resistive plate chambers (RPCs) form a redundant trigger network that quickly determines whether or not to hold the acquired muon event. All these components are robust and capable of suppressing background noise. The muon barrel (MB) region contains RPCs and DTs, while the endcap contains RPCs and CSCs. The arrangements depend on the muon rate in MB (muon barrel) and ME (muon endcap) region. A cross-sectional view of the muon system is shown in Fig. 3.9.

The DTs cover the pseudorapidity region  $|\eta| < 1.2$ . They can reconstruct the muon track from its hits within the stations with excellent time resolution and efficiency. A gas mixture of 85% Ar+15% CO<sub>2</sub> is surrounded by a gold-plated stainless-steel anode wire in each cell, resulting in a drift time of 380 ns.

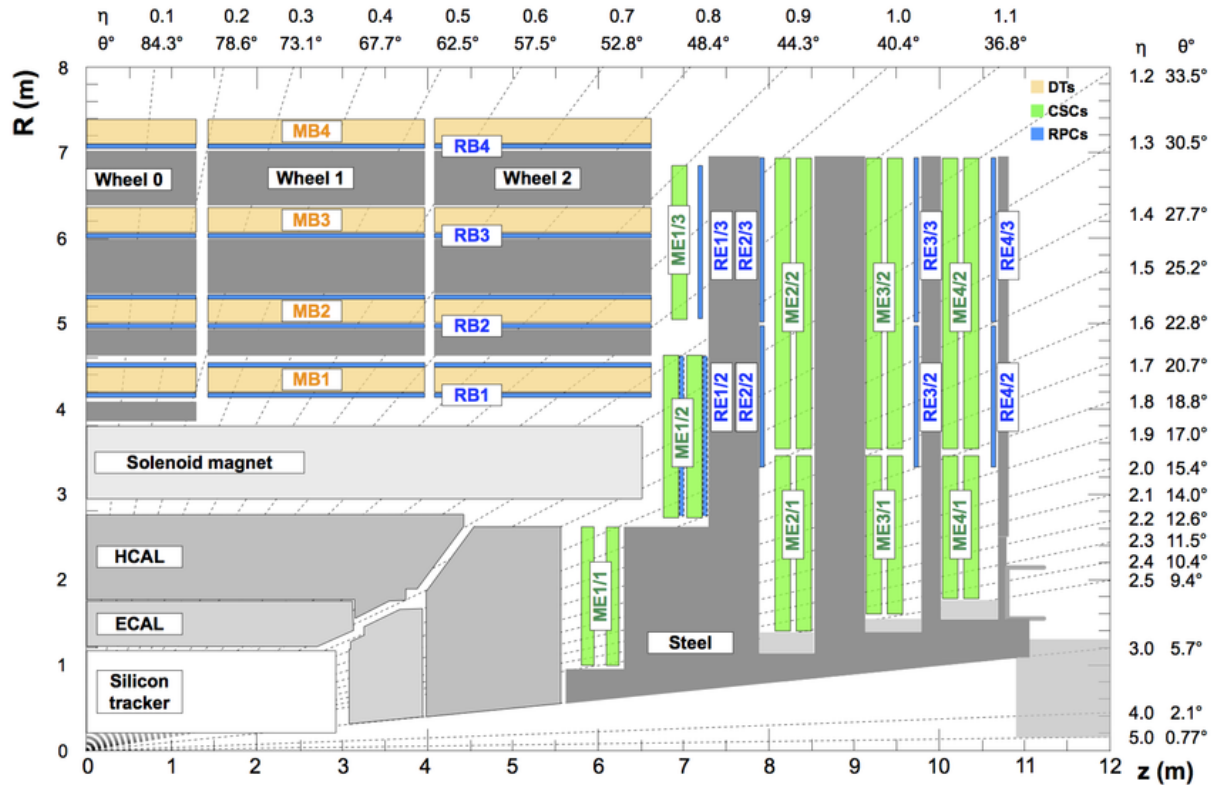


Figure 3.9: Schematic view, in the  $r$ - $z$  plane, of one quadrant of the CMS detector, with the axis parallel to the beam ( $z$ ) running horizontally and the radius ( $r$ ) increasing upward. The interaction region is at the lower left corner. The position of the present RPC chambers is shown in blue. The RPCs are both in the barrel and in the endcaps of CMS. The DT chambers are labeled MB and the CSC chambers are labeled ME. The steel disks are displayed as dark gray areas. [67].

1937 The CSCs cover a pseudorapidity region  $0.9 < |\eta| < 2.4$ . In each endcap, the 468 trape-  
 1938 zoidal CSCs are arranged into four stations. Six anode planes are interleaved among seven  
 1939 cathode panels in each chamber, with wires running azimuthally. The ME chambers use  
 1940 a gas admixture of 50%  $\text{CO}_2$  + 40% Ar and 10%  $\text{CF}_4$ .

1941 The RPCs are interspersed in both the MB and ME covering  $|\eta| < 1.9$  region. The RPCs  
 1942 are made of two resistive Bakelite plates separated by a gas volume. They provide an  
 1943 independent triggering system and a fast response with good time resolution (less than  
 1944 25 ns) for muons.

### 1945 3.2.6 Trigger System

1946 At the LHC, the proton-proton collision occurs at very high luminosity, which leads to the  
 1947 production of rare physics signals at an appreciable rate. However, most of the collisions



1948 are soft (low energy), so they do not produce any interesting physics events. Also, the  
 1949 size of each event is around 1 MB, and the frequency of collisions is 40 MHz, i.e., 40 TB of  
 1950 data per second get generated during the collisions. Considering the fact that in this huge  
 1951 data collection only a few events are of physics interest, a trigger system is used to select  
 1952 potentially interesting events. Only this fraction of data is stored on a computer disk  
 1953 for subsequent analysis. The full trigger system decreases the rate of interesting events  
 1954 to about 1 thousand per second. A series of trigger levels are used to achieve this. The  
 1955 detector stores all of the data from each crossing in buffers. A small amount of key data  
 1956 is used to perform a fast, approximate calculation to identify features of interest such as  
 1957 high-energy jets, muons, or missing energy. The levels are known as “L1-trigger” or level-1  
 1958 trigger and “HLT” or high level trigger as given in Fig. 3.10.

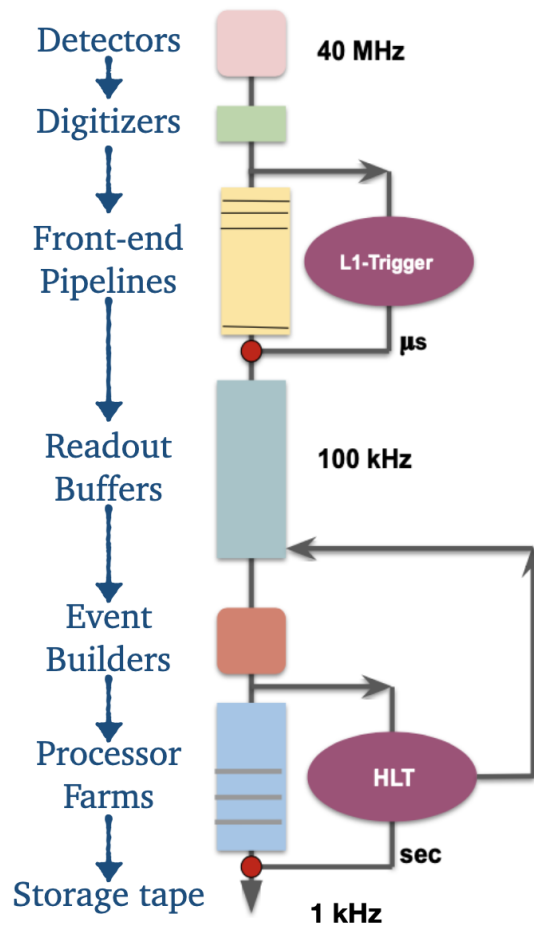


Figure 3.10: Flowchart of CMS trigger system. Image reproduced from Ref. [68].

#### 1959 **L1-trigger**

1960 The L1-trigger is based on hardware. It uses a rapid and completely automated method  
1961 that scans the basic signs of interesting physics, such as particles with high energy or  
1962 rare combinations. From the 40 million events, 100k events are selected at this level with  
1963 a latency of few microseconds using a simplified readout of the calorimeters and muon  
1964 subdetectors.

1965 A simple schematic of the CMS L1-trigger is given in Fig. 3.11. The trigger primitives  
1966 (TP) from ECAL and HCAL as well as muon detectors (drift tubes (DT), cathode strip  
1967 chambers (CSC), and resistive-plate chambers (RPC)) are processed in several steps until  
1968 the combined event information is evaluated. After this, a decision is made to accept the  
1969 event. The information from regional calorimeter triggers (RCTs) is combined to make  
1970 up the L1 global calorimeter trigger (GCT). The RCT receives the transverse energies of  
1971  $e/\gamma$  objects from ECAL and of jets from HCAL. The L1 jet reconstruction algorithm is  
1972 based on a square approach: it considers the energy deposit in a  $9\times 9$  trigger tower area  
1973 centered on a local maximum. The RCT processes this data in parallel and sends objects  
1974 and their energy information as outputs. The GCT sorts the objects using their energy  
1975 information and classifies them as isolated, non-isolated, central, forward jets, and several  
1976 global quantities.

1977 To ensure good coverage and redundancy, each of the three muon detector systems partic-  
1978 ipates in the L1 muon trigger. The front-end trigger electronics of DTs and CSCs identify  
1979 tracks (hits) and transmit them to regional track finders. They further identify muons  
1980 based on pattern recognition algorithms and measure their energy. In the overlap region  
1981 of the DT track finder and CSC track finder, the information is shared for efficient cov-  
1982 erage. For RPC hits, the information is sent to pattern comparator trigger logic boards  
1983 via front-end electronics that identify muon candidates. The three regional track finders  
1984 sort the muon candidates that have been detected and send them to the global muon  
1985 trigger (GMT) with their pT and position information. The GMT then combines muon  
1986 candidates identified by multiple systems to exclude candidates that pass multiple muon  
1987 triggers. The GMT also conducts a consistency assignment so that candidates can be re-  
1988 jected at the final trigger stage if their quality is poor and they can only be reconstructed  
1989 by one muon track finder.

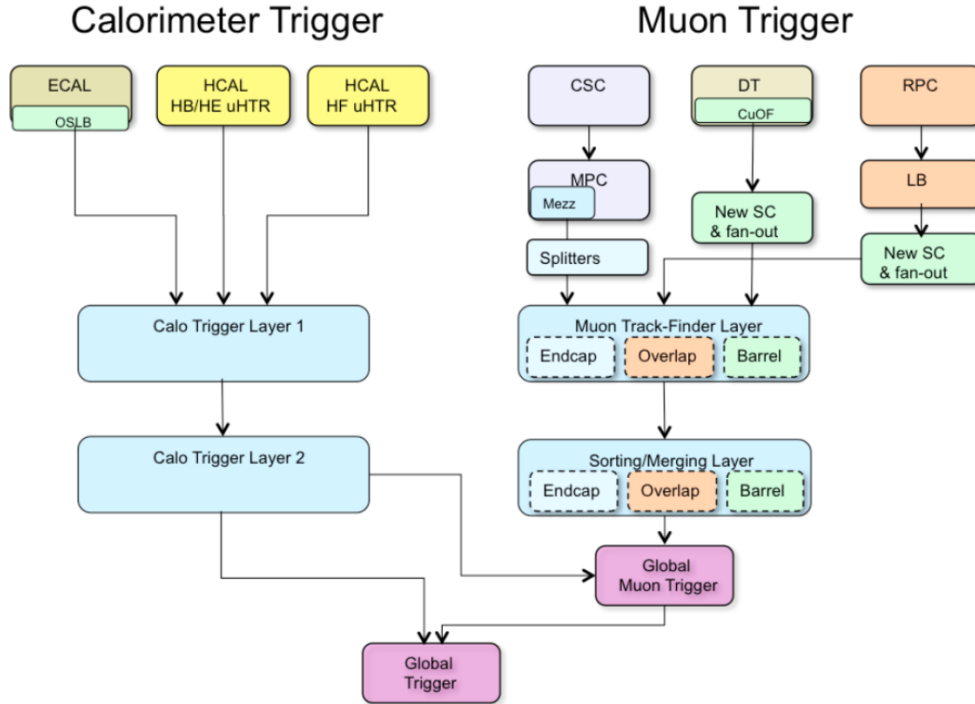


Figure 3.11: Schematic layout of L1 trigger system of the CMS experiment [69].

1990 The global trigger (GT) completes the CMS L1 trigger scheme by implementing a menu  
 1991 of triggers. The decision from GT is sent to the tracker, ECAL, HCAL or muon system  
 1992 via the trigger, timing and control (TTC) system. Finally, a set of selections are imposed  
 1993 on these reconstructed objects at the L1 level to pick good events.

### 1994 High-level Trigger (HLT)

1995 The HLT consists of a farm of processors running a version of the full event reconstruction  
 1996 software optimized for fast processing. It reduces the event rate down to 1 kHz from  
 1997 100kHz. Objects such as electrons, muons, and jets are reconstructed for each event, and  
 1998 identification criteria are used to select only interesting events for data analysis. After  
 1999 the HLT level selection, the data is stored on tape for further analyses.

2000 In the HLT operation, the data from the readout buffers are sent to a processor farm. This  
 2001 trigger level is made up of a series of increasingly complex filters. The filtering process uses  
 2002 the complete detector information from all the subdetectors, starting from reconstruction  
 2003 to selection. In simple words, the HLT considers full data events to decide if the event  
 2004 should be kept or not. In order to create datasets with different physics signatures, the

2005 final stage of HLT processing involves the reconstruction and event filtering. The time  
2006 duration it takes to process an event varies depending on the algorithms used. The average  
2007 time between events is about 60 ms, but some events can take up to a second.

2008 The data acquisition system (DAQ) reads data from various subsystems for offline storage  
2009 after the HLT decision. A complete sequence of L1 and HLT selection criteria, including  
2010 any prescale, is referred to as a trigger path. Unlike in the case of offline analysis, the  
2011 trigger selection is a non-reversible process, and discarded events cannot be recovered.

## 2012 3.3 CMS simulation tools

2013 The simulation plays a fundamental role in data analysis to perform any measurement  
2014 or extract any relevant physics parameter. It consists of complete information on the  
2015 physics process used for event generation and corresponding particle content. For event  
2016 generation, Monte Carlo (MC) based event generators are used. They use numerical  
2017 MC based techniques to produce collisions at the high energy as they occur in the LHC.  
2018 The MC event generators provide a complete picture of the collision process from ini-  
2019 tial to final stages, including the strongly inelastic interaction, the radiation process,  
2020 parton-hadronization, and the underlying event description. After the physical events are  
2021 produced using information from theoretical models incorporated into the generator, the  
2022 effects that a detector introduces into the basic theory must be considered. The choice of  
2023 event generator depends on the physics process.

2024 In simulation, each event has different probability that corresponds to its differential  
2025 cross section. While generating events at next-to-leading order (NLO) in perturbation  
2026 calculation, the infra-red (IR) divergences in the real-emission corrections and virtual  
2027 corrections are taken into account using IR subtraction method. The contribution of the  
2028 soft singularities makes the matrix element of real-emission finite. Therefore, the method  
2029 adds a subtraction term to make virtual corrections finite. The events are generated  
2030 separately for Born and real-emission phase space. For the events simulating real-emission,  
2031 event weights become negative if simulation over-estimates real-emission matrix element.  
2032 The negative weights contributes as the negative differential cross section term for real-  
2033 emission. This results in reduction of effective simulation statistics. Generally the fraction

2034 of events with negative weights is small, but still problematic while doing higher order  
2035 calculation that requires more computational resources per event. Therefore, the event  
2036 generators that take care of the negative weights issue are used for NLO event generation  
2037 [70, 71]. CMS physics simulations are mostly based on the following event generators:

2038 **MADGRAPH:** MADGRAPH [72] generates the matrix element for the multi-particle fi-  
2039 nal state process. The updated version allows for matrix element calculations at NLO  
2040 accuracy (controls theoretical uncertainty) and provides a technique for parton shower  
2041 matching. Depending upon the Lagrangian of any renormalizable or effective theory, it  
2042 can generate events for any physics process predicted by that theory.

2043 **POWHEG:** POWHEG [73, 74] also produces NLO-accurate calculations of the hard scat-  
2044 tering sub-processes. PowHeg has the characteristic of being free from negative event  
2045 weights. The Refs. [71, 73, 75] have detailed description of PowHeg NLO calculations.

2046 **PYTHIA:** The PYTHIA [76] is a standard tool for the generation of events in high-energy  
2047 collisions, which works for various SM and BSM processes. It contains the models of hard  
2048 processes and initial- and final-state parton showers, matching and merging methods  
2049 between hard processes and parton showers, multiparton interactions, beam remnants,  
2050 string fragmentation, and particle decays.

2051 The last step of the simulation process is to pass the generated events through the de-  
2052 tector response simulation. In CMS, it is done using GEANT4 [77] toolkit. The detector  
2053 simulation includes the detector geometry, particle interactions with detector's materials,  
2054 magnetic field effects, and real conditions during detector operation. The additional pp  
2055 interactions (pileup) during collisions are also superimposed to the event at hit level using  
2056 pileup mixing module. The module deals with a sequence of bunch crossings to properly  
2057 simulate the contributions affecting the in-time bunch crossing. The next step in the  
2058 event is modeling the response of the detector readout electronics, which is also known as  
2059 digitization. The digitized signals are further processed in reconstruction of higher level  
2060 physics objects, such as charged particle tracks, photons, etc. used in physics analyses.

2061 Most of the aspects of simulations are integrated with the standard analysis software  
2062 known as CMSSW [78]. The MC events are reconstructed using the same methods used  
2063 for real data, allowing for a consistent comparison between the data and the simulation.

2064 We will briefly study these reconstruction algorithms in the next chapter.

## Chapter 4

# Physics Object Reconstruction and Identification

The reconstruction of an event implies the identification of all stable and visible particles produced in a proton-proton interaction. Particles are identified based on their specific signatures in the CMS detector, as illustrated in Fig. 4.1 and their kinematic properties are measured by combining the information from the various subdetectors. The event reconstruction within CMS is performed by reconstructing each final state particles with a particle flow algorithm, whether they are leptons, hadrons or photons.

After the collision, particles enter the tracker after leaving the beam interaction point, where signals (hits) in the sensitive tracker layers are used to reconstruct charged-particle trajectories (tracks) and their origins (vertices). The tracker is embedded in a magnetic field, which bends the trajectories and allows for measuring charged particle electric charge and momentum. ECAL absorbs electrons and photons as they pass through it. The corresponding electromagnetic showers are observed as energy clusters in neighbouring cells, which are used to calculate the particle's energy and direction. Hadrons, both charged and neutral, may induce a hadronic shower in the ECAL, which is then fully absorbed in the hadron calorimeter (HCAL). The energies and directions are calculated using corresponding clusters. Muons pass through the calorimeters with very few interactions. Muons generate hits in the muon detectors, located outside the calorimeters. Neutrinos do not interact with the CMS detector and escape undetected.

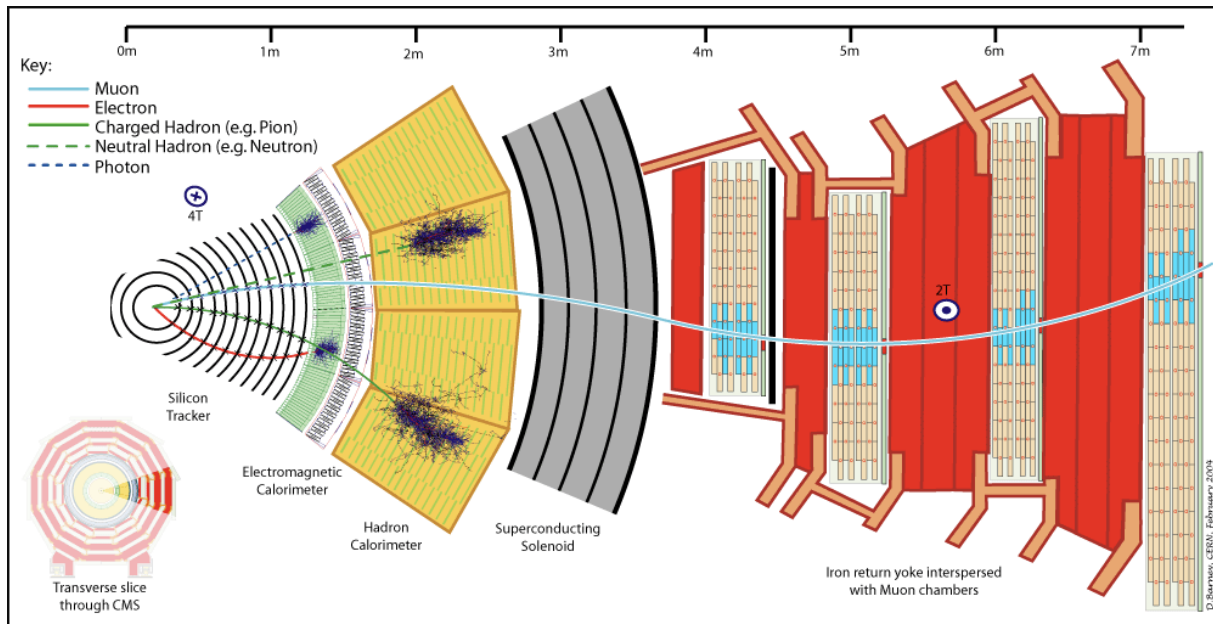


Figure 4.1: Transverse slice of the CMS detector, showing the experimental signatures of the different final-state particles [79].

2086 All the above information from subdetectors are the basis of CMS reconstruction algo-  
 2087 rithms.

## 2088 4.1 Track and Primary Vertex Reconstruction

2089 For any data analysis, it is essential to understand how the tracks and origin of the  
 2090 charged particle tracks (also known as primary vertex) can be identified when dealing  
 2091 with the large number of PU interactions. At every bunch crossing, the collisions between  
 2092 protons give rise to interaction vertices which spread along the beam axis around the  
 2093 nominal interaction point, which can be reconstructed by combining information from  
 2094 reconstructed tracks, as illustrated in Fig. 4.2. Tracks are then associated to vertices, and  
 2095 the primary vertex (PV) is defined as the one with the largest value of summed physics-  
 2096 object  $p_T^2$ . The other vertices are referred to as pileup vertices. A detailed procedure of  
 2097 CMS track and vertex reconstruction is given in Ref. [80].

2098 Tracks are essential for determining the production vertex of charged particles and mea-  
 2099 suring their momenta. For their reconstruction, firstly, the hits within the pixel and  
 2100 strip detectors are determined. They estimate the momentum and position parameters  
 2101 (longitudinal and transverse parameters) of the particles. For this estimation, tracking



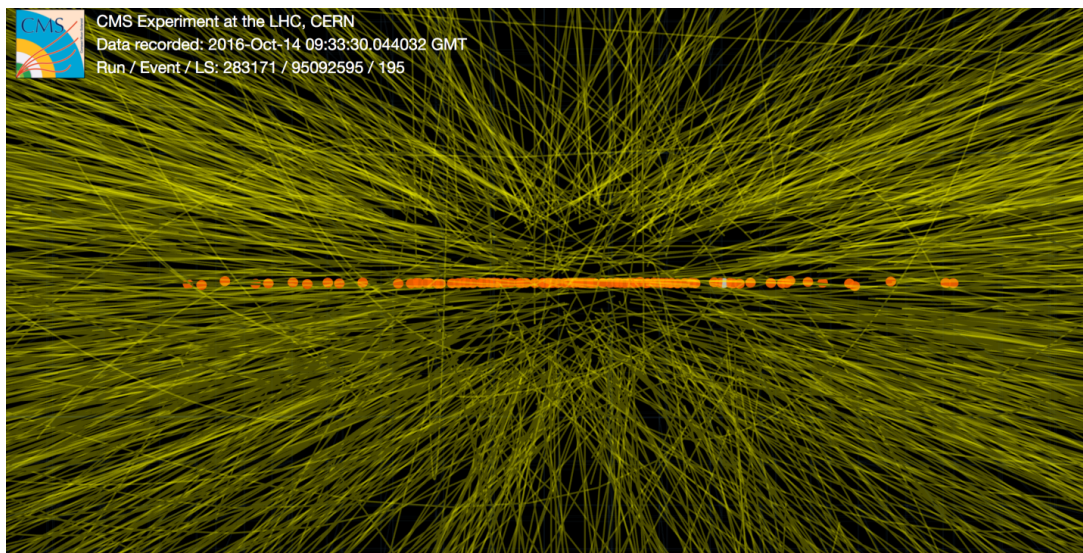


Figure 4.2: A collision recorded by the CMS detector during the 2016 data taking [81]. The lines correspond to the reconstructed tracks, while the dots represent the reconstructed interaction vertices.

2102 algorithm called combinatorial track finder (CTF) is used.

2103 The CTF is based on the concept of iterative tracking. At each iteration, track seeds are  
2104 formed using a limited number of pixel hits, and an initial estimate of the track parameters  
2105 is derived. A track-finding algorithm follows, based on the method of Kalman filter  
2106 [82]. The filter uses track seeds to extrapolate the trajectory of the track to subsequent  
2107 detector layers, taking into account the effect of multiple Coulomb scattering. At each  
2108 layer, compatible hits are added to the trajectory, the track parameters are recalculated,  
2109 and the resulting trajectory is extrapolated to the next layer. Finally, tracks that do not  
2110 satisfy goodness-of-fit criteria are discarded. Tracks that are easier to reconstruct, e.g.  
2111 those with large  $p_T$  (and therefore less pronounced curvature) and produced close to the  
2112 interaction point, are reconstructed first. Hits associated to these tracks are then removed,  
2113 reducing the complexity of subsequent iterations. Multiple iterations are performed, each  
2114 time with decreasingly stringent requirements on the track seeds. In particular, the last  
2115 iterations are optimized to reconstruct tracks with lower  $p_T$  or with larger displacement  
2116 from the interaction vertex. The efficiency of reconstructing tracks with  $p_T > 1$  GeV is  
2117 found to be larger than 99% for isolated muons over the entire coverage of the tracker,  
2118 while efficiencies for electrons and pions range between 80 and 99%, depending on the  
2119 track pseudorapidity. Fake rates are at the level of 5% in the barrel and of up to 15% in  
2120 the endcap and transition regions. The resolution on the track  $p_T$  significantly depends

2121 on the  $p_T$  and  $\eta$  of the tracks, and is below 1% for central muons with  $p_T$  between 1 and  
2122 10 GeV [80].

2123 Primary vertex reconstruction depends on the selected tracks where it finds a common  
2124 meeting point (vertex) among a set of tracks. It aims to determine the position and  
2125 associated uncertainty of all proton-proton interaction vertices, including the signal vertex  
2126 and any vertices from pileup collisions. It consists of three steps:

- 2127 1. Tracks selection
- 2128 2. Clustering of tracks originated from the same interaction vertex
- 2129 3. Track fitting for each vertex's location using its corresponding tracks

2130 The inclusive vertex finder (IVF) algorithm is used for reconstructing secondary vertices  
2131 (SV). Based on their separation in three dimensions, it clusters tracks around seeds with  
2132 high impact parameter significance. Then, an outlier-resistant fit of a common vertex of  
2133 all the tracks in a cluster yields the SV position. Tracks are then re-associated to either  
2134 the primary or the secondary vertex based on their compatibility and the SV position is  
2135 fitted again using only the remaining tracks if there are at least two tracks remaining. SVs  
2136 aren't employed directly in this study, but in the event that they can be reconstructed,  
2137 they offer significant jet flavor differentiation and are thus used in b-tagging discriminant  
2138 like DeepCSV.

## 2139 4.2 Particle Flow

2140 The Particle Flow (PF) algorithm aims to reconstruct and classify all of the particles  
2141 from a collision by integrating the information from the various subdetectors in the most  
2142 optimal way. For each collision, the set of the reconstructed and identified particles by the  
2143 algorithm (PF candidates) provides a global event description that leads to phenomenal  
2144 CMS performance for jet and  $\tau$  hadronic decay reconstruction, MET determination, and e  
2145 and  $\mu$  identification. This method also identifies particles from PU interactions; therefore,  
2146 it is used to build efficient PU mitigation techniques. The output of the PF consists of  
2147 a list of candidates classified as electrons, photons, muons, charged hadrons, or neutral

2148 hadrons, as illustrated in Fig. 4.3. The algorithm consists of two separate steps: the  
 2149 reconstruction of the PF elements, and the link between the reconstructed elements.

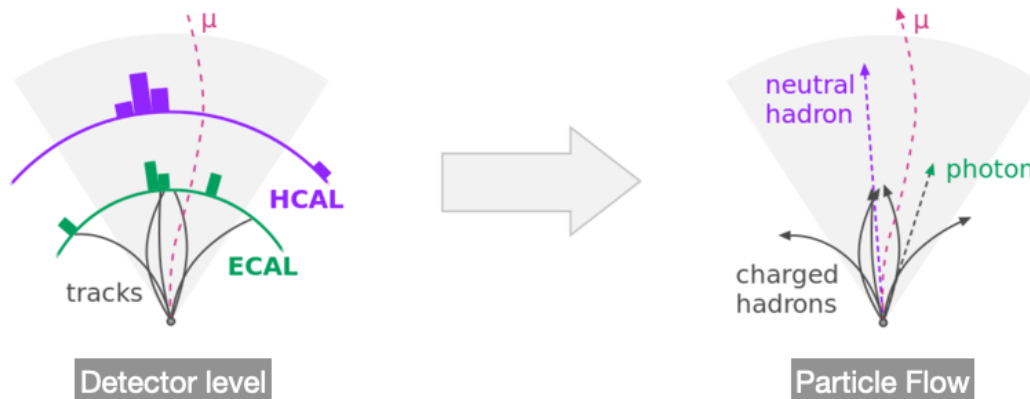


Figure 4.3: The particle-flow algorithm combines information from various subdetectors to provide a global description of the event in terms of electrons, photons, muons, charged hadrons, and neutral hadrons.

2150 Any physics study is conducted purely based on final decay products, including the fea-  
 2151 tures of the physics process. In this part, we study how the PF algorithm works for the  
 2152 final observable states that leave signals in the detector.

- 2153 • **Electrons (e) and Photons ( $\gamma$ ):** Nearly all of the energy of electrons and photons  
 2154 is deposited in the ECAL, where electrons also create hits in the tracker layers.  
 2155 Extrapolation from the last measured hit in the tracker to any cluster in the ECAL is  
 2156 used to link them. The signals in the ECAL crystals are reconstructed by subtracting  
 2157 the PU contributions. This technique has been used for both the HLT and offline  
 2158 event reconstruction during the entire LHC Run-2 data-taking period.

2159 While travelling through the tracker material in front of the ECAL, almost 60%  
 2160 of the photons start to convert into a pair of electron-positron through the pair-  
 2161 production mechanism. Furthermore, the produced electron or positron experiences  
 2162 bremsstrahlung photon radiation. So the energy measurement of a photon from  
 2163 a particular single crystal deposition is not possible. Hence, the photon-electron  
 2164 reconstruction algorithm starts by grouping the crystals into  $3 \times 3$  cluster in  $\eta - \phi$   
 2165 space around the most energetic one (called seed crystal). In the presence of the  
 2166 CMS magnetic field the trajectories of the electrons and positrons are bent leading  
 2167 to photon radiations spread radially over  $\phi$  direction. To catch the corresponding

2168 energy deposits, multiple ECAL clusters are needed to be combined to produce a  
2169 super cluster (SC). In CMS this step is done by two different algorithms (i) mus-  
2170 tache algorithm, which is particularly useful to properly measure the low energy  
2171 deposits, and (ii) refined algorithm, it uses the tracking information of the extrap-  
2172 olated bremsstrahlung tangents and the conversion tracks of the electron-positron  
2173 pair to match with the SC position to combine into a single candidate.

2174 All these ECAL clusters, superclusters, electrons tracks, and tracks from photon  
2175 conversion are used as input to link this element into a block of particles. Start-  
2176 ing from electron tracks or superclusters, respectively, the blocks are divided into  
2177 electron and photon. At this point the supercluster is called a refined superclus-  
2178 ter. A further track selection criteria are applied to these object to reconstruct “PF  
2179 electron”. Without passing track selection, the particle is labelled as “PF photons”.

- 2180 • **Muons ( $\mu$ ):** Muons are reconstructed using information from the tracker and the  
2181 muon systems in CMS [67]. The reconstruction is performed with three following  
2182 methods:

- 2183 – The local reconstruction is the first step in the muon reconstruction chain.  
2184 First, digitized electronic signals are used to recreate hits in DTs, CSCs, and  
2185 RPCs. Hits are then matched within each DT and CSC chamber to form  
2186 segments (track stubs) using the Kalman filter method. The reconstructed  
2187 muons are labelled as “standalone muons”.

- 2188 – A search is performed for tracks that fit each standalone muon track among  
2189 those reconstructed in the inner tracking system, with the best-matching  
2190 tracker track being chosen. The track fitting, using all hits in both tracks,  
2191 is performed for each tracker track-standalone muon pair, again using the  
2192 Kalman filter technique. The result is a collection of objects referred to as  
2193 “global muons”.

- 2194 – Tracker muon tracks are built from the inner tracker trajectory reconstruc-  
2195 tion. The tracker-muon algorithm is beneficial for identifying low-pT muons  
2196 that may not leave enough hits in the muon stations to be reconstructed as a  
2197 standalone muon. Tracker muons should not be used without additional spec-  
2198 ifications because the default conditions for tagging a tracker track as “tracker

2199 muon” are pretty loose.

2200 The resulting muon candidate collections are used as input for PF muon identifica-  
2201 tion. The PF muon reconstruction has been fine-tuned to distinguish muons within  
2202 jets with high accuracy, resulting in a low rate of false positives due to misidentified  
2203 charged hadrons.

2204 • **Charged and neutral hadrons:** As the collisions occur at a high energy, the  
2205 process may end up with having partons in the final state carrying colour charge.  
2206 Since they can not exist in a free state, they hadronize to produce stable colourless  
2207 hadrons as a result of QCD confinement. The produced hadrons appear to move in  
2208 the same direction as the parton they originate from, creating collimated bunches  
2209 of particles known as jets. After the identification of muons, electrons, and isolated  
2210 photons and their extraction from the PF blocks, the remaining particles to be  
2211 detected are hadrons originating from jet fragmentation and hadronization.

2212 • **Missing Transverse Energy (MET):** The MET is identified as momentum im-  
2213 balance in the transverse direction and defined as the negative vector sum of the  
2214 transverse momentum of the reconstructed PF candidates in the event [83, 84]. It  
2215 originates from weakly interacting neutrinos or any BSM particles that hardly leave  
2216 any signals within the detector. Since W bosons, top quarks, and tau leptons may  
2217 decay into neutrinos, CMS uses MET to reconstruct them. Furthermore, several  
2218 BSM physics models, such as dark matter models, supersymmetric models, and  
2219 models with warped extra dimensions, predict the presence of particles that might  
2220 be invisible and can carry momentum. Accurate MET reconstruction is complex be-  
2221 cause it requires the precise reconstruction of all visible particles in an event. The  
2222 CMS detector meets these requirements with its highly granular electromagnetic  
2223 calorimeters, hermetic hadronic calorimeters, redundant muon systems, and silicon  
2224 trackers in a strong magnetic field.

2225 The next Sections provide a detailed description of the object reconstruction of photons  
2226 and jet formed by  $b$  quark, which are majorly used for the physics analysis explored in  
2227 this thesis.

## 4.3 Photon Reconstruction and Identification

The photon candidates are reconstructed from the energy deposition in the ECAL crystals, that are not linked with the charged tracks coming from the tracker as discussed in Sec. 4.2.

### 4.3.1 Photon Identification

After the reconstruction of the photons, there is always a finite probability that a photon can be faked by a jet. To distinguish a prompt photon from a fake photon CMS has developed two different methods. One is the cut-based method, where some threshold values on different shower shape and isolation variables are applied. The second one is the multivariate technique, where using the similar identification variables, BDT is trained to discriminate the prompt photons against the fake ones. For precision measurements, the latter one is more optimal. The identification variables used as input for the multivariate photon identification technique [85] can be described as follows:

- **Isolation variables** ( $Iso_{ph}$ ,  $Iso_{ch}$ ): These two are the isolation variables obtained by sum of the transverse momenta of the electromagnetic candidates and charged hadrons within a  $\Delta R = 0.3$  isolation cone in  $\eta - \phi$  plane around photon object. The isolation thresholds depend on the energy of the photon objects.
- **Shower shape variables:** Another strategy for rejecting high-electromagnetic-content jets is to take advantage of the ECAL electromagnetic shower shape. The energetic jets with photons from hadronic decay make a wider shower within ECAL in comparison to an isolated single photon. The following are two of the most relevant variables used for photon identification depending upon the geometric shower shape from prompt and background photons:

**H/E Ratio:** The ratio of energy stored in the HCAL in a cone of radius  $\Delta R = 0.15$  around the supercluster direction compared to the energy of the photon candidates is known as the H/E ratio. For low energy photons, HCAL contribution comes due to HCAL noise and pileup, while for high energy photons, it is due to leakage of photons through the inter-module gaps.

2256  $\sigma_{i\eta i\eta}$  : This variable gives crystals with energy deposits of at least 0.9% of  $E_{5\times 5}$   
 2257 (the energy deposited in a  $5 \times 5$  crystal matrix around the most energetic crystal)  
 2258 contribute to  $\sigma_{i\eta i\eta}$  defined as:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5\times 5} \omega_i (\eta_i - \bar{\eta}_{5\times 5})^2}{\sum_i^{5\times 5} \omega_i}} \quad (4.1)$$

2259 where  $\eta_i$  is the pseudorapidity of the  $i$ th crystal,  $\bar{\eta}_{5\times 5}$  is mean pseudorapidity of  
 2260 crystal matrix, and  $\omega_i$  is weight factor defined as  $\max(0, 4.7 + \ln(E_i/E_{5\times 5}))$ , and  
 2261 is nonzero if  $\ln(E_i/E_{5\times 5}) > -4.7$ , which corresponds to the  $E_i > 0.9\%$  of the total  
 2262 energy of  $5 \times 5$  cluster. This ensures that only the crystals above a noise thresholds  
 2263 are included in this variables. The distributions of  $\sigma_{i\eta i\eta}$  of a prompt photon or  
 2264 electrons are narrow compared to the fake photons coming from jets.

- 2265 •  **$R_9$** : The sum of the energy deposition of the  $3 \times 3$  crystals centered on the most  
 2266 energetic seed crystal in the supercluster divided by the total energy deposition on  
 2267 that supercluster. The shape of  $R_9$  distribution for the unconverted photons has  
 2268 high value close to unity whereas photons that convert before reaching ECAL have  
 2269 a lower  $R_9$  value.

2270 A BDT is trained using very loose selections on photon identification variables with the  
 2271 median energy per unit area ( $\rho$ ),  $\eta$  and uncorrected energy of photon supercluster as input.  
 2272 A comparison of the performance between cut-based identification and BDT identification  
 2273 for photons is shown in Fig. 4.4 which clearly shows for a fix background misidentification  
 2274 rate, MVA-based photon identification performs better. The clustering algorithms allow  
 2275 to achieve reconstruction of about 95% of photon energy deposits.

### 2276 4.3.2 Photon Energy Regression

2277 Due to shower leakage, detector gaps, and dead crystals, the ECAL is not suitable for  
 2278 collecting all of the energy deposited by photons. The noise causes systematic varia-  
 2279 tions in measured ECAL energy and reduce photon energy resolution. A multivariate  
 2280 technique-based correction is used to minimize the impact of these losses. The target of  
 2281 this regression is the ratio between particle-level and reconstructed level photon energy,

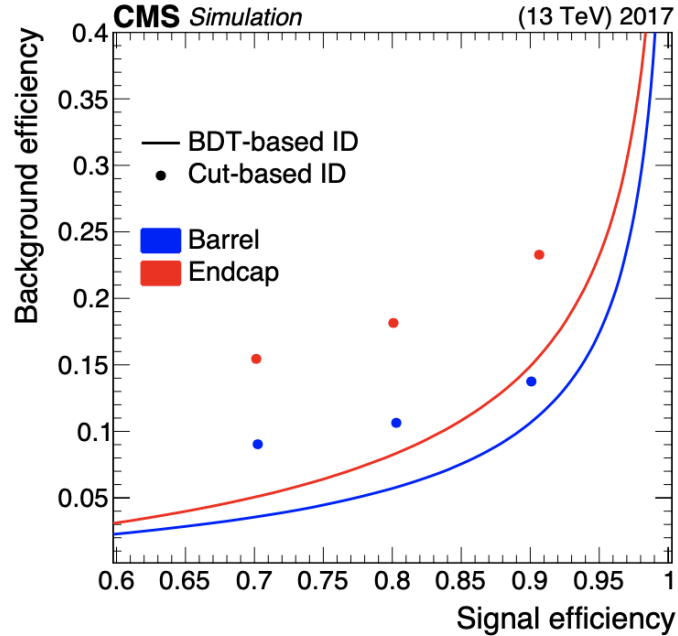


Figure 4.4: Performance of the photon identification based on cut-based and MVA based approach. The three points for the cut-based method refer to the three different working points: loose, medium, and tight [85].

2282 and its output correction factors are applied to reconstructed energy of data and simulated  
2283 events to obtain the best estimate of the true energy.

2284 There remains a slight variation in energy scales and resolutions for data and simulation  
2285 after applying the energy regression, which is fixed by scaling and smearing correction  
2286 factors. The studies are performed using simulated  $Z \rightarrow e^+e^-$  events. For photons, only  
2287 ECAL reconstructed information of the events is used.

2288 The photon energy scales are adjusted by changing the scale observed in simulated events  
2289 to match the data scale. The results of fitting the invariant mass  $m_{ee}$  distributions in  
2290 different eta regions, obtained from data and simulated events separately, are compared  
2291 to derive a scale offset. This method extracts corrections to both the energy resolution in  
2292 the simulation and the scale for the data in bins of  $|\eta|$  and  $R_9$  in the second stage. It fixes  
2293 the residual discrepancy between data and simulation in  $m_{ee}$  distributions by applying an  
2294 energy Gaussian spreading function to simulated events.

2295 Depending on the pseudorapidity region and energy loss in the detector material, the  
2296 ultimate energy resolution after all corrections (regression and scale corrections) ranges



2297 from 2 to 5%. The performance of the regression is shown in Fig. 4.5 for barrel and  
 2298 endcaps.

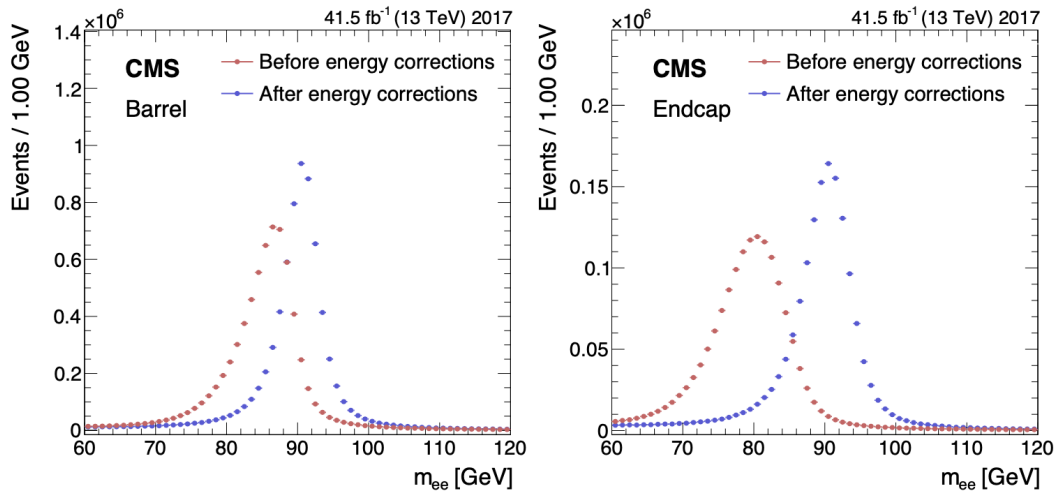


Figure 4.5: Reconstructed  $Z \rightarrow e^+e^-$  invariant mass distribution before and after the energy regression correction for the barrel (left) and endcap (right). [85].

## 2299 4.4 Jet Reconstruction

2300 In CMS, jets are reconstructed using a sequential recombination algorithm known as anti-  
 2301  $k_T$  [86]. The algorithm is designed to be safe against infrared and collinear singularities,  
 2302 i.e. insensitive to soft and collinear gluon emission, and collinear gluon splitting. The  
 2303 jets are reconstructed by clustering their constituents together using various algorithms  
 2304 that follows collinear and infrared safety principles [87]. A final state parton from hard-  
 2305 scattering and hadrons can have multiple collinear splitting or soft emissions which create  
 2306 infrared collinear (IRC) divergences. In theory, these divergences get cancel out with  
 2307 one order loop correction. But this cancellation does not happen within jet algorithm. If  
 2308 algorithm is not IRC safe, it will result in unrealistic infinite cross-section. Also, jet defined  
 2309 by any algorithm should be invariant of choice of algorithm. Particle-flow candidates are  
 2310 sequentially recombined following the procedure described below. A distance  $d_{ij}$  between  
 2311 entities (i.e. particles or combinations of particles)  $i$  and  $j$  is introduced together with a  
 2312 distance  $d_{iB}$  between entity  $i$  and the beam. These two quantities are defined, respectively,  
 2313 as:

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.2)$$

$$d_{iB} = p_{Ti}^{-2},$$

2314 where  $\Delta R_{ij}^2 = (\Delta y_{ij})^2 + (\Delta \phi_{ij})^2$ , and  $p_{Ti}$  is the transverse momentum of entity  $i$ . The  
 2315 distance parameter  $R$  controls the size of the jet and can be chosen arbitrarily. The  
 2316 quantities  $d_{ij}$  and  $d_{iB}$  are calculated for all possible combinations of entities in the event.  
 2317 If the smallest  $d_{ij}$  is smaller than the smallest  $d_{iB}$ , entities  $i$  and  $j$  are combined to form  
 2318 a new entity. Otherwise, entity  $i$  is removed and called a jet. The procedure is iterated  
 2319 until no particles are left. The jet momentum is then defined as the vectorial sum of the  
 2320 momenta of all particles in the jet.

2321 The algorithm allows clustering of the hard particles. The algorithm involves a combi-  
 2322 nation of energy and angle in its distance measure. The achievement of anti- $k_T$  is that  
 2323 it gives circular-shaped jets without using a cone-based jet algorithm. The anti- $k_T$  algo-  
 2324 rithm is widely used to define a set of collimated particles as a PF jet. The spread area  
 2325 of the jet is given in terms of jet distance parameter  $R$ . The value of  $R$  can be different  
 2326 for different collision energies. For the boosted scenario where hadrons get produced at  
 2327 very high energy, it is hard to separate the particles from two different hadronic jets.  
 2328 Therefore, the bunches of particles are merged and reconstructed using a large  $R$  in the  
 2329 algorithm. They are called fat or large-area jets.

2330 CMS generally uses AK4 PF jets, i.e., anti- $k_T$  jets clustered with distance parameter  $R$   
 2331  $= 0.4$ . For large-area jets, the value of  $R$  is taken 0.8, and they are called AK8 PF jets.  
 2332 Despite the excellent functionality of the jet clustering algorithm, we get discrepancies  
 2333 between the energy of partons and the reconstructed energy of jets due to the efficiency  
 2334 and acceptance of the detector. That is why jet energy corrections (JEC) and jet energy  
 2335 resolution correction (JER) are applied on the reconstructed jets including uncertainty  
 2336 associated with them. In order to mitigate the effect of pileup interactions, charged par-  
 2337 ticles associated to pileup vertices are removed prior to the jet clustering. This procedure  
 2338 is known as charged hadron subtraction [88]. The energy of the jets is calibrated using  
 2339 a factorized approach where subsequent corrections that account for different effects are  
 2340 applied.

2341 There are two types of PF Jets used in CMS analyses. For PF CHS (Charged Hadron  
2342 Subtraction) jets, charged particles from non-primary vertices (pileup) are removed before  
2343 clustering. Another one is PF PUPPI (Pile-Up Per Particle Identification) jets which use  
2344 the PUPPI [89] algorithm. Apart from PF jets and PF MET, Calo jets/MET and TRK  
2345 jets/met is also used for various studies. Calo objects are reconstructed using the energies  
2346 in calorimeter towers, and their direction and TRK objects are reconstructed using hits  
2347 information from the inner tracker.

2348 For LHC studies, pileup is a big challenge. PUPPI [89] is one of the new ideas proposed  
2349 for pileup mitigation. The algorithm uses global information (like PV) of an event and  
2350 local information (like tracks) at particle level to identify pileup. Within this algorithm, a  
2351 weight is calculated, using this global and local information, for each particle in the event.  
2352 It is calculated by a shape parameter  $\alpha$ , which distinguishes parton shower-like radiation  
2353 from pileup-like radiation for every particle in the event. Apart from the shape parameter,  
2354  $p_T$  also helps for pileup mitigation as  $p_T$  distribution for pileup falls much faster. The  
2355 tracking information also helps to differentiate between charged tracks that originate  
2356 from the primary vertex (PV) and the charged tracks that originate from the pileup. The  
2357 PF algorithm can be used to relate these tracks to particles. Using PF, particles can  
2358 be sorted into three class: charged hadrons from PV, charged hadrons from pileup and  
2359 neutral hadrons from both pileup and PV. For all neutral particles, the algorithm assumes  
2360 them to be originated from PV and assign weights depending upon their  $p_T$ . As a result,  
2361 it gives low weights to soft pileup contributions, which lowers the event's neutral pileup  
2362 contribution. Hence, PUPPI even works for the region where tracking is not available.

2363 The PUPPI weights are further used to rescale the four momenta of the particles. Ide-  
2364 ally, this weight is one for particles from hard scattering, and zero for pileup originated  
2365 particles. However, in the real picture, the weights can be in fractional form, depending  
2366 upon the particles' properties. Particles with a very small weight are discarded. Using  
2367 PUPPI weighted particles, one can perform jet clustering without any other treatment for  
2368 pileup. This algorithm has proven very efficient to correct jet  $p_T$ , jet mass and missing  
2369 transverse energy in a high pileup collision conditions.

### 2370 4.4.1 Jet Energy Corrections (JEC)

2371 The jet energy corrections (JEC) are used to calibrate the energy of the jet and correct  
2372 the value as much as possible to match the corresponding particle-level jets. In-time  
2373 pileup from charged particles can be subtracted in regions where tracking is available, by  
2374 removing tracks compatible with a pileup vertex. This procedure keeps only the tracks  
2375 associated to the primary vertex and tracks which can't be associated to any vertex. The  
2376 corrected jets are now called charged hadron subtracted (CHS) PF jets and this step is  
2377 limited only upto  $|\eta| < 2.5$  due to the tracker acceptance.

2378 Even after the CHS correction, a significant contribution from the pileup remains in the  
2379 jets. The first step of it is called L1 Pileup correction to remove the energy coming from  
2380 the pileup particles clustered inside the jet. The pileup offset correction is determined from  
2381 the simulation of a QCD dijet event sample with and without the pileup contribution,  
2382 parameterized as a function of jet  $p_T$  and  $\eta$  of the jet, the energy density ( $\rho$ ) of the  
2383 event and jet area. The residual difference between the data and the MC is corrected  
2384 using a random cone method applied in zero bias data, which is collected by the CMS  
2385 experiment without using any external trigger. In the random cone method, many jets  
2386 are reconstructed by clustering particles in randomly placed cone, the average energy of  
2387 these jets are mostly due to the detector response and the pileup effects. An uncertainty  
2388 of 5% on the jet response is applied based on data-simulation comparison in the random  
2389 cone method.

2390 The second step is called L2L3 MC-truth corrections, which are obtained from the sim-  
2391 ulation as a function of the  $p_T$  and  $\eta$  of the jet, by comparing the average energy of the  
2392 reconstructed jet with a geometrically matched particle-level jet energy. The correction  
2393 factors are then applied on the jets both in the simulation and the data. After application  
2394 of the simulation-based L2L3 correction factors, the residual differences between the data  
2395 and the simulation is accounted for in terms of jet energy scale (JES). JES is determined  
2396 in two steps. If the jets are in the central region of  $|\eta| < 1.3$ , the JES is determined  
2397 from the photon+jet and  $Z(\rightarrow e^+e^-/\mu^+\mu^-)$ +jet events where  $p_T$  of the photons and the  
2398  $Z$  boson can be well measured. The JES in the forward regions are obtained from a QCD  
2399 dijet event where two jets are expected to be balanced by each other and the leading jet  
2400 is required to be in the central region. The JES is estimated for the forward jet with

2401 respect to the well calibrated central jet. The residual discrepancy of JES between data  
 2402 and simulation is appeared as data-to-simulation correction factor.

#### 2403 4.4.2 Jet Energy Resolution (JER)

2404 The jet energy resolution (JER) is derived using the principle of  $p_T$  balance in dijet,  
 2405 photon+jet and Z+jet events, where the jet energy response has been derived with respect  
 2406 to a reference object i.e. central jet, photon and Z boson, respectively. For each jet, the  
 2407 response in a given  $p_T$  and  $\eta$  range is modelled with a double-sided Crystal Ball function  
 2408 (DSCB); the width of the DSCB gives the measurement of JER. The JER in data is found  
 2409 to be worse than in simulation. To match the JER in simulation with data, a correction  
 2410 factor is added to the resolution. The data-to-simulation correction factors are extracted  
 2411 using data-based methods and are used to smear the simulated jet resolution. These  
 2412 correction factors are derived for a jet in different  $\eta$  regions and shown in Fig. 4.6.

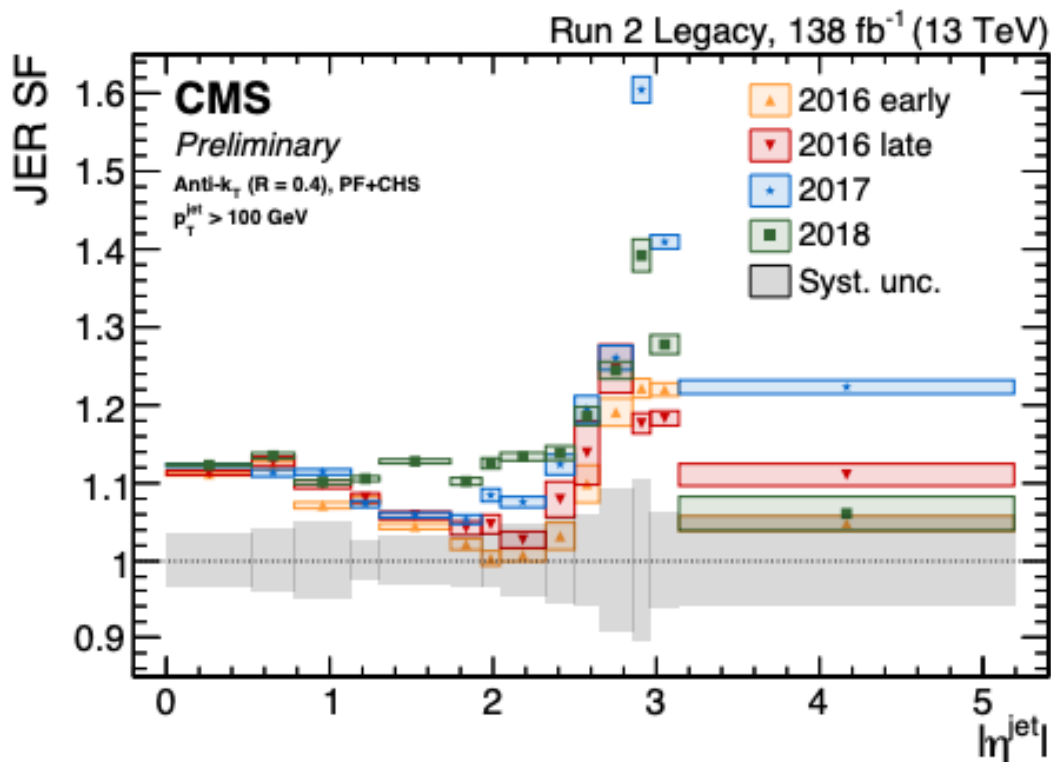


Figure 4.6: JER scale factor as a function of absolute value of pseudorapidity ( $|\eta|$ ) [90].

2413 **4.4.3 Jet  $b$  tagging**

2414 As the name reflects,  $b$  tagging aims to identify a heavy quark jet from a light flavour  
 2415 jet. Identification of the origin of jets is critical for studying and characterizing different  
 2416 channels, such as top quark/Higgs boson events and a variety of new physics scenarios.  
 2417 The long lifetime, high mass, and hard fragmentation function of  $b$  quarks and the exist-  
 2418 tence of soft leptons from semileptonic  $b$  hadron decays are used to develop CMS  $b$  tagging  
 2419 algorithms.

2420 Semileptonic decays of  $b$  hadrons give rise to  $b$  jets that contain a muon. A cascade decay  
 2421 of  $b \rightarrow c \rightarrow l$  also gives muon in the final state. Since the CMS muon systems identify  
 2422 the origin of a muons with high efficiency and resolution, this information helps to tag  
 2423 the  $b$  jets. Due to long lifetime,  $b$  hadron travels some distance until it decays at a point  
 2424 called the secondary vertex (SV) shown in Fig. 4.7. With the high resolution of the CMS  
 2425 tracking system, it is possible to reconstruct the SV. The distance between the PV and  
 2426 the SV is called the flight distance. In the SV vertex finding process, the tracks associated  
 2427 with PV are not considered.

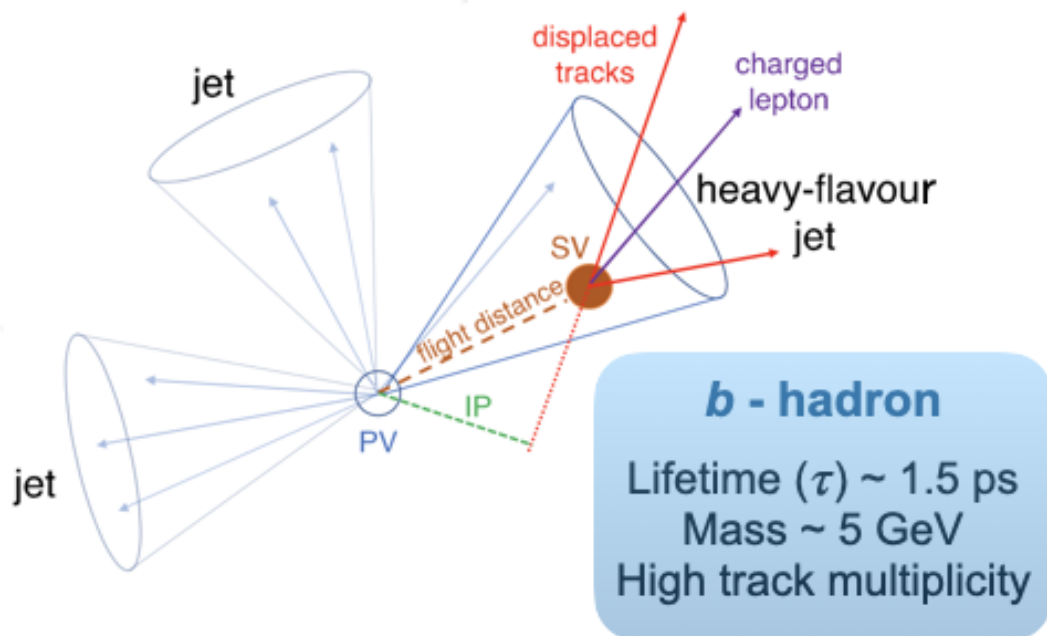


Figure 4.7: An illustration of  $b$ -hadron decay and corresponding impact parameter (IP) [91].

2428 The distance of closest approach of a track to the primary vertex is known as the impact  
 2429 parameter (IP) given in Fig. 4.7. IP's sign is positive/negative if the track passes closest to

its associated jet direction down/upstream of the PV. It is calculated in three-dimensions, which benefits from a good x-y-z resolution from the pixel detector. Due to the long lifetime, the IP from  $b$  jets is mostly positive, while for light jets, the impact parameter remains symmetric around zero. A tight selection on impact parameter helps to reject the tracks associated with the background process.

The important variables for  $b$  tagging algorithms are the IP significance of the tracks, the position of the secondary vertex and transverse momentum of muon with respect to jet direction. CMS used jet-probability (JP) (uses impact parameter significance of the tracks) and combined secondary vertex (CSVv2) (combines the information of displaced tracks with the information of secondary vertices associated with the jet using a multivariate technique) taggers during Run-1 [92]. For Run-2, the new versions of the  $b$  tagging algorithms known as DEEPCSV and DEEPJET have been developed using deep neural network (DNN) [93] training.

**DEEPCSV and DEEPJET:** DEEPCSV uses input tracks and secondary vertex information similar to JP and CSVv2 taggers including track-variables for up to six tracks,  $p_T$ ,  $\eta$  information of jets to learn the correlation between jet kinematics and other input variables. The DNN training is performed using jets with  $p_T$  up to 1 TeV and within the tracker acceptance. A mixture of  $t\bar{t}$  and multijet events is used so that the training could learn about heavy flavour and light flavour jets and discriminate them. The neural network uses four hidden layers that are fully connected. A normalized exponential function is used to activate the nodes in the last layer so that the output value can be interpreted as a likelihood/probability  $P(f)$  for a specific jet flavour category. There are five such categories known as jet having one  $b$  hadron  $P(b)$ , at least two  $b$  hadrons  $P(bb)$ , one  $c$  hadron without any  $b$  hadron  $P(c)$ , at least two  $c$  hadrons without any b-hadron  $P(cc)$  and without any b/c hadron  $P(udsg)$ . A  $b$  discriminator  $P(b)+P(bb)$  is combined to tag  $b$  jets in physics analyses. The performance of DEEPCSV improves with the 2017 and 2018 data-taking years.

Apart from DEEPCSV, CMS has developed one more DNN-based  $b$  tagger DEEPJET [94], which uses a large number (approximately 650) of jet variables as input. The inputs used here contain the information about all PF candidates clustered inside the jet and they can be classified into four different types:

- 2461 1. The global variables related to the jet level information: jet 4-momentum, number  
 2462 of constituent particles, primary and secondary vertices,
- 2463 2. Charge particle information: charged track  $p_T$ ,  $\eta$ , 2D and 3D impact parameters  
 2464 etc.,
- 2465 3. Neutral hadron information,
- 2466 4. The variables related to the secondary vertex: tracks associated with SV, etc.

2467 It is a multiclass discriminator which can differentiate b, c, uds and gluon jets by utilizing  
 2468 the same set of input variables used in DEEPCSV which results in better performance  
 2469 for 2016, 2017 and 2018 data-taking years. Unlike DEEPCSV, it has seven output cate-  
 2470 gories. Four are similar to DEEPCSV and the rest of the three P(lepb), P(g) and P(uds)  
 2471 discriminate against  $b$  jet having lepton or jet originating from gluon or light quark. The  
 2472 difference between both DNN-based  $b$  taggers is explained in Tab. 4.1 and performance  
 2473 curves are shown in Fig. 4.8.

	<b>DEEPCSV</b>	<b>DEEPJET</b>
<b>Input</b>	Information about displaced tracks and SV	Particle flow candidates
<b>Additional feature</b>	Use more charge particle tracks	Use soft lepton information
	—	Discriminate uds jets from gluon jets
	—	Recovers degradation at high momentum
	Performance degrades for phase-0 pixel configuration (2016)	Similar performance for phase-0 and phase-1 pixel configurations
<b>Category</b>	5 P(b/bb/c/cc/uds)	7 P(b/bb/c/cc/uds/g/lepb)
<b><math>b</math>-discriminator</b>	P(b + bb)	P(b + bb + lepb)

Table 4.1: Difference between DEEPCSV and DEEPJET.

2474 Within an analysis, we can use either a selection on  $b$  discriminant or full shape of  $b$   
 2475 discriminant. The selection is applied depending upon the Loose, Medium or Tight work-  
 2476 ing point corresponding to 10%, 1% or 0.1% misidentification probability. While using



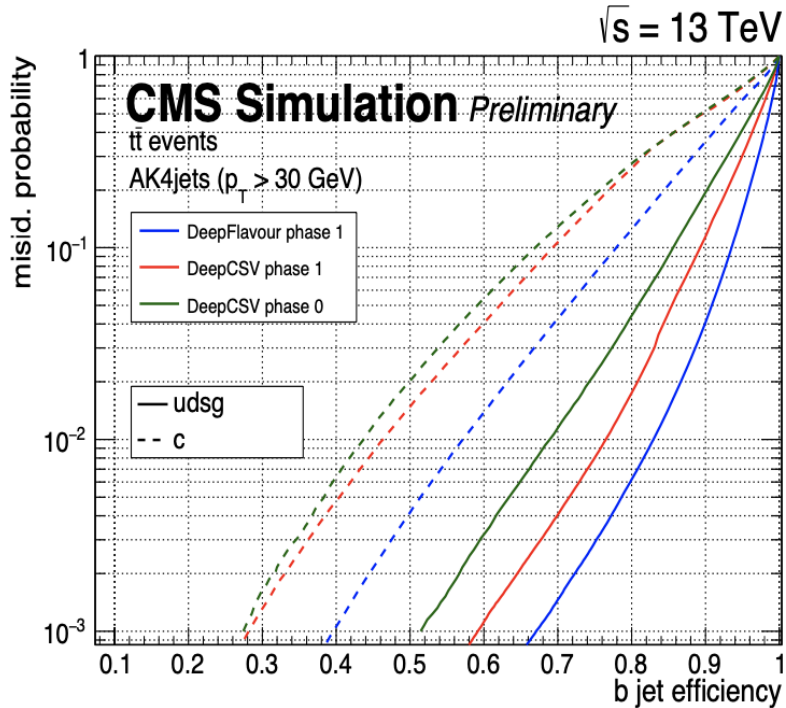


Figure 4.8: DEEPCSV and DEEPJET performance curves for AK4 jets [95].

2477  $b$  tagging algorithms, we need to consider how differently the algorithms behave for jets  
 2478 reconstructed in simulations and data. For this purpose, CMS provides the scale factors  
 2479 (SF) to

- 2480 (a) correct separately for the final yield of jets tagged as heavy or light flavour (working  
 2481 point based calibration), or
- 2482 (b) correct for the whole shape of the discriminator (shape calibration or reshaping), if  
 2483 the analysis needs it (for example, as an input to MVA training).

2484 The SFs of a jet depends on its flavour,  $p_T$  and  $\eta$  (and jet discriminator value for re-  
 2485 shaping). Depending upon the properties of the jet, SFs are calculated and applied to  
 2486 simulations to minimize discrepancy with data. In my work, DEEPJET was used in the  
 2487  $HH \rightarrow b\bar{b}\gamma\gamma$  analysis and for resolved jets in the  $HHH \rightarrow 6b$  analysis, whereas DEEPCSV  
 2488 was used in the VHbb analysis.

#### 4.4.4 Boosted Higgs boson jet tagging

In Chapter 6, BSM triple-Higgs boson searches are solely explored in the boosted regime. Therefore, understanding of the boosted Higgs jet identification is essential. As the SM predicts the highest branching ratio for  $b\bar{b}$  as the final state for Higgs boson decay, boosted  $H \rightarrow b\bar{b}$  tagging might be crucial for new physics searches. The approach of tagging a  $b$  jet in a boosted regime is different from general AK4  $b$  jet tagging. In boosted regime, jet mass is the first jets' property that distinguishes signal jet from background one. The jet mass of the signal should peak around resonance mass (Higgs boson in our case). However, this happens only in an ideal picture. In reality, the jet mass distribution gets worse, and its peak gets shifted due to soft emissions, pileup and underlying events. For  $H \rightarrow b\bar{b}$  decay, QCD multi-jet processes (mainly gluon splitting in  $b\bar{b}$ ) are dominant backgrounds. The background jets acquire mass through showering, which grows as a function of transverse momentum. Boosted hadronic objects keep a different energy pattern than QCD jets of comparable invariant mass. This is a motivation to go beyond jet mass and exploit jet substructure [96]. It removes soft radiation contamination and identifies and selects features related to hard decay.

The main idea of  $H$  boson tagging is to reconstruct a large-area jet with jet substructure (grooming and tagging) algorithms that remove soft contribution and try to understand the subject structure and shape of the jet, which discriminate signal jets from background jets. At this stage, selected jet is called the Higgs jet. In the end, using boosted  $b$  tagging algorithms, subjects are identified as  $b$  jets, and the Higgs jet is identified as a jet with two  $b$ -hadrons. Let's understand all these steps one by one.

- **Jet grooming:** When a boosted large-area jet originates from a Higgs boson, its mass should peak near Higgs boson mass. Jet grooming removes background contamination and related component from the jet mass calculation. Soft drop declustering is one of the grooming technique. It suppresses wide-angle soft radiation from a jet in order to mitigate the effects of contamination from initial state radiation, underlying event, and multiple hadron scattering (pileup).

Starting from AK8 reclustered jet with radius  $R_0$ , it does declustering of the last

2518

step and drop the soft contributions with following condition:

$$\text{softdrop condition} = \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left( \frac{\Delta R_{12}}{R_0} \right)^\beta \quad (4.3)$$

2519

where,  $p_{T_i}$  is transverse momenta of constituents,  $\Delta R_{12}$  is distance in  $\eta - \phi$  plane,  $z_{cut}$

2520

is the soft drop threshold, and  $\beta$  is angular exponent. The efficiency and performance

2521

of grooming depends on  $z_{cut}$  and  $\beta$  ( $\beta \rightarrow \infty$  for ungroomed jet). For  $\beta < 0$ , it works

2522

as tagger vetoing jets that do not have two well-separated hard prongs. While for

2523

$\beta > 0$ , it works as groomer changing the constituents of a jet. This technique is

2524

IRC safe and removes all soft contributions. CMS uses this technique with  $\beta$  set to

2525

0 as the standard choice for jet grooming.

2526

- **Jet shapes:** QCD jets have a fundamentally different pattern of energy deposits

2527

in the detector when compared to a boosted Higgs jet. To identify an N sub-jet

2528

structure within a jet, we define a variable called the N-subjettiness  $\tau_N$  [97] as:

$$\tau_N = \frac{1}{d_0} \sum_k p_T^k \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\} \quad (4.4)$$

2529

$$d_0 = \sum_k p_T^k R_0. \quad (4.5)$$

2530

Here,  $k$  runs over the constituent particles in a given jet,  $p_T^k$  are their transverse

2531

momenta,  $\Delta R_{J,k}$  is the distance in the  $\eta - \phi$  plane between a candidate subjet- $J$

2532

and a constituent particle  $k$  and  $R_0$  is the jet radius. For  $\tau_N \approx 0$ , all the radiation

2533

is along the candidate subjet, and the jet has N (or fewer) subjets. For  $\tau_N \gg 0$ , a

2534

large fraction of the jet energy is scattered away from the candidate subjet, which

2535

implies the jet has at least N+1 subjets.  $\tau_2$  is expected to be small for Higgs jets.

2536

However, QCD jets can randomly have small values of  $\tau_2$  too, and to increase the

2537

discrimination power,  $\tau_{21} = \tau_2/\tau_1$  is used as an identifier for Higgs jets. Smaller the

2538

value of  $\tau_{21}$ , more likely it is that the jet is a Higgs jet.

2539

- **boosted  $b$  tagging:** For boosted jets,  $b$  tagging can be applied either on the AK8

2540

jet or its subjets. Double- $b$  tagger algorithm includes the probability of each subjet

2541

being a  $b$  jet. This is achieved by applying the DeepJet  $b$ -tagging algorithm, which

2542

returns the probability of the subjet having come from a  $b$  quark,  $b\bar{b}$  pair, leptonic

2543  $b$  decay,  $c$  quark, light-flavour quark, or gluon. These probabilities along with the  
 2544 soft-drop mass of the subjet are also included in classification of the large-area jet.  
 2545 In order to classify a hadronically decaying particle through a single large-area jet,  
 2546 the DeepAK8 [98] algorithm defines five main categories: W, Z, H, t, and other. The  
 2547 algorithm’s goal is multi-classification of jets by exploiting particle-level information  
 2548 directly. The DeepAK8 algorithm uses a large number of variables, both low- and  
 2549 high-level, but not all variables are treated in the same way. The architecture of  
 2550 the algorithm consists of two steps: In the first step, the input variables are split in  
 2551 two lists and processed separately with two classifiers. In the second step, the two  
 2552 previous outputs are combined through a third classifier.

2553 An alternative DeepAK8 algorithm, DeepAK8-MD, has been developed to be largely  
 2554 decorrelated from the mass of jets while providing an efficiency similar to that of  
 2555 the mass-correlated version. The ROC curves in Fig. 4.9 show the performances of  
 2556 double-b, DeepAK8 and DeepAK8-MD on the same simulated dataset.

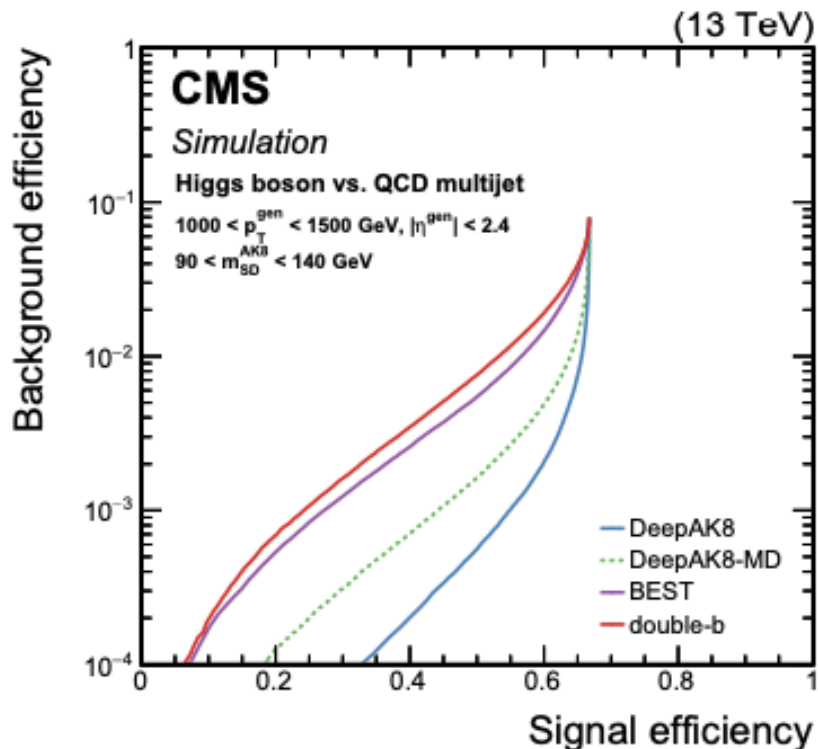


Figure 4.9: Performance comparison in terms of ROC curves on the Higgs boson taggers [98].

2557 The ParticleNet algorithm [99] is a Dynamic Graph Convolutional Neural Network  
 2558 (DGCNN) trained to classify jets according to their flavour type. Compared to the

2559 previous state-of-the-art jet tagging algorithms, like DeepAK8 tagger, ParticleNet  
 2560 relies on a highly sophisticated jet representation and network architecture. The  
 2561 ParticleNet algorithm treats a jet as an unordered set (permutation-invariant) of  
 2562 its constituents ("particle cloud"). This technique provides a natural representation  
 2563 of the jet, unlike other algorithms that treat jets as images or particles ordered by  
 2564 their  $p_T$ . The graph network architecture allows the algorithm to efficiently explore  
 2565 the correlations between the various jet constituents. Particle-flow candidates and  
 2566 secondary vertices within the jet cone are used as inputs to the algorithm.

2567 The ParticleNet algorithm has two versions:

- 2568 – the "nominal" version is designed for maximum performance but may introduce  
 2569 sculpting in the mass spectrum of the background jets,
- 2570 – the "mass decorrelated" (MD) version which is designed to be largely decorre-  
 2571 lated with respect to the mass of a jet, at the cost of slight degradation in the  
 2572 discrimination power.

2573 The tagger assigns a set of output classifier scores for each jet, corresponding to  
 2574 the probability of the jet originating from a resonance that decays into a pair of  
 2575 quarks ( $X \rightarrow b\bar{b}$ ,  $X \rightarrow c\bar{c}$ ,  $X \rightarrow q\bar{q}$ , i.e., X decaying to light quarks), or non-  
 2576 resonant quark-and-gluon jet ( $QCD_{bb}$ ,  $QCD_b$ ,  $QCD_{cc}$ ,  $QCD_c$  and  $QCD_{others}$ ). To  
 2577 focus on the discrimination power between  $X \rightarrow b\bar{b}$  and QCD jets, the  $b\bar{b}$ -tagging  
 2578 discriminant  $T_{Xbb}$  that we use for this analysis is defined to be:

$$T_{Xbb} = \frac{S}{S+B} = \frac{P_{Xbb}}{P_{Xbb} + P_{QCD}} \quad (4.6)$$

2579 In this work, ParticleNet tagger is used to tag boosted Higgs jets for triple-Higgs  
 2580 production searches and DeepAK8 tagger was used in the VHbb boosted region.

## Chapter 5

# Search for Non-resonant Higgs Pair

## Production in $b\bar{b}\gamma\gamma$ Final State

### 5.1 Introduction

With the increasingly large dataset of proton-proton collisions events delivered at the LHC, rare physics processes become experimentally accessible.

This chapter presents the search for the non-resonant production of Higgs boson pairs in the  $b\bar{b}\gamma\gamma$  final state. The work exploits the data collected by the CMS detector in proton-proton collisions with a center-of-mass energy of 13 TeV, for data collected during Run2 (2016-2018) with a total integrated luminosity of  $137\text{ fb}^{-1}$ . The analysis targets the main HH production modes: the gluon-gluon fusion (ggF HH) production mode and the vector boson fusion mode (VBF HH) as described in Section 2.4. Both modes are analyzed following similar strategies. After reducing the nonresonant  $\gamma\gamma b\bar{b}$  background and the background coming from single Higgs boson production in association with a top quark-antiquark pair ( $t\bar{t}H$ ), the events are categorized into ggF HH and VBF HH enriched signal regions using a multivariate technique. The signal is extracted from a fit to the invariant masses of the Higgs boson candidates in the  $b\bar{b}$  and  $\gamma\gamma$  final states. An orthogonality in the event selection criteria for the two processes has been maintained properly between the ggF HH and VBF HH analysis and described in the following sections whenever needed.

2600 The  $b\bar{b}\gamma\gamma$  final state has a combined branching ratio of  $2.63 \pm 0.06 \times 10^{-3}$  [29] for  $m_H = 125$   
 2601 GeV, which can be derived from the individual branching ratio of the  $H \rightarrow b\bar{b}$  and  $H \rightarrow \gamma\gamma$   
 2602 decay,  $2 \times Br(H \rightarrow b\bar{b}) \times Br(H \rightarrow \gamma\gamma) = 2 \times 0.58 \times 0.00223$ . This channel is one of the  
 2603 most sensitive to HH production because of the large SM branching fraction of  $H \rightarrow b\bar{b}$ ,  
 2604 the good mass resolution of the  $H \rightarrow \gamma\gamma$  candidate, and relatively low background rates.

## 2605 5.2 Search Strategy

2606 For signal events, the distributions of the invariant mass of diphoton  $M(\gamma\gamma)$  and dijet  
 2607  $M(jj)$  peak around the mass of the Higgs boson (125 GeV). Photon energy, and con-  
 2608 sequently  $M(\gamma\gamma)$ , can be reconstructed with high precision thanks to the CMS electro-  
 2609 magnetic calorimeter, while worse resolution (roughly one order of magnitude worse) is  
 2610 expected for  $M(jj)$ .

2611 The analysis strategy is therefore based on extracting the number of signal and background  
 2612 events using a parametric fit to the invariant mass of the diphoton system  $M(\gamma\gamma)$ , fitting  
 2613 simultaneously  $M(jj)$ . In addition, 1D-analysis strategy was implemented : using a  
 2614 parametric fit to the invariant mass of the diphoton system  $M(\gamma\gamma)$  and using the  $M(jj)$   
 2615 information in the multivariate discriminant used to separate signal or background. Both  
 2616 strategies were studied, finally the 2D simultaneous fit to  $M(\gamma\gamma)$  and  $M(jj)$  was chosen  
 2617 as the analysis strategy due to its better sensitivity.

2618 For  $M(\gamma\gamma)$  and  $M(jj)$  different distributions are expected for two types of backgrounds:

- 2619 • For the non-resonant background, mainly  $\gamma\gamma$ +jets and  $\gamma$ +jets and top events,  
 2620  $M(\gamma\gamma)$ , and  $M(jj)$  have a falling spectrum. This is a dominant background af-  
 2621 ter the preselection is applied.
- 2622 • For resonant background (SM Single Higgs production),  $M(\gamma\gamma)$  is peaking at the  
 2623 Higgs mass, since the two photons are the decay products of the Higgs boson.  
 2624 Among the different production modes for SM single Higgs production, the two most  
 2625 relevant for this analysis are gluon-gluon fusion (ggH) and associated production  
 2626 with top quarks ( $t\bar{t}H$ ), where the Higgs bosons decaying into a pair of photons. For

2627 ggH production, additional jets originate from gluon splitting into a pair of b-quarks  
 2628 and additional light jets from radiation mistagged as b-jets. For  $t\bar{t}H$  production, the  
 2629 two b-jets originate directly from the decay of the top pair. The  $M(jj)$  distribution  
 2630 for resonant backgrounds is a falling spectrum. Even if the background rejection,  
 2631 in this case, is high, the cross-section of the process is several orders of magnitude  
 2632 larger than the signal one, causing a non-negligible contribution of this background.

2633 The distribution of  $M(\gamma\gamma)$  and  $M(jj)$  for simulated signal and background events can be  
 2634 seen in Fig. 5.1, where all contributions are normalized to a unit area.

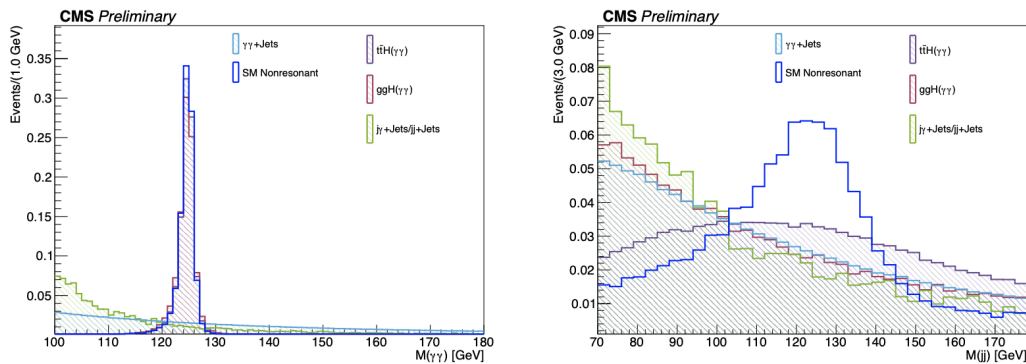


Figure 5.1: The distribution of  $M(\gamma\gamma)$  (left) and  $M(jj)$  (right) for signal and background events. All contributions are normalized to a unitary area.

2635 To improve the sensitivity of the search, MVA techniques are used to distinguish the ggF  
 2636 and VBF HH signal from the dominant nonresonant background. The output of the MVA  
 2637 classifiers is then used to define mutually exclusive analysis categories targeting VBF and  
 2638 ggF HH production. The HH signal is extracted from a fit to the invariant masses of the  
 2639 two Higgs boson candidates in the  $(M(\gamma\gamma), M(jj))$  plane simultaneously in all categories.

2640 The  $H \rightarrow b\bar{b}$  and  $H \rightarrow \gamma\gamma$  candidates can be used to reconstruct the double Higgs system  
 2641 (HH). In this analysis, the following variable is used as a proxy to a 4-body mass:

$$\widetilde{M}_x = M(jj\gamma\gamma) - M(jj) - M(\gamma\gamma) + 2m_H \quad (5.1)$$

2642 where  $m_H = 125$  GeV is the mass of the Higgs boson and  $M(jj\gamma\gamma)$  is the invariant  
 2643 mass of the two Higgs boson candidates, is particularly sensitive to different values of  
 2644 the Higgs couplings. This definition of  $\widetilde{M}_x$  is less dependent on the dijet and diphoton



2645 energy resolution than  $M(jj\gamma\gamma)$  under the assumption that the dijet and diphoton pairs  
 2646 originates from a Higgs boson [100].

2647 The distribution of  $\widetilde{M}_x$  for SM and some representative BSM benchmark points is shown  
 2648 in Fig.5.2 (left). It can be seen that different values of the couplings can introduce large  
 2649 variations in the shape of the four-body mass. Fig. 5.2 (right) shows the  $\widetilde{M}_x$  distribution  
 2650 for the individual background components and the SM signal.

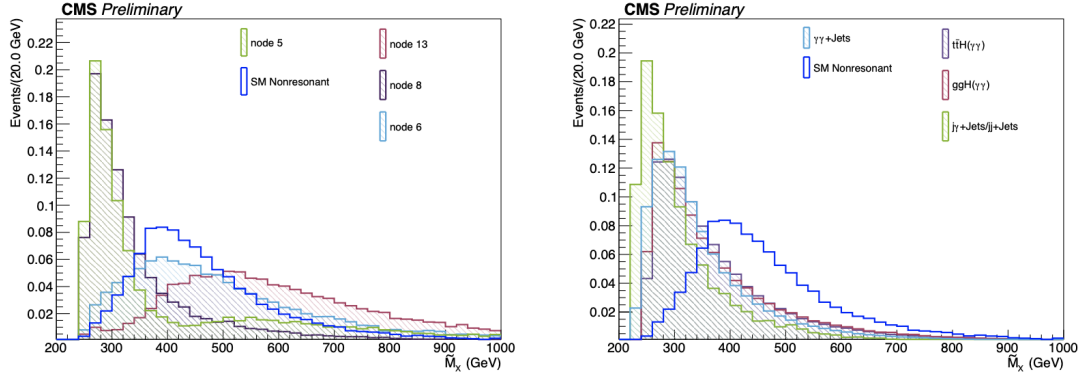


Figure 5.2: (left) Distribution of  $\widetilde{M}_x$  for some benchmark signal datasets, and for the resonant and non-resonant backgrounds. (right) Distribution of  $\widetilde{M}_x$  for SM-like HH production and for the different background components. All contributions are normalized to a unit area.

2651 As can be seen from the distributions, a lower cut on this variable would noticeably  
 2652 reduce the background contribution without losing efficiency on the SM sample; on the  
 2653 other hand, for some benchmarks points, this would result in a considerable loss in the  
 2654 efficiency. Instead of having a single cut, this variable is therefore used to categorize  
 2655 events, with the goal of maximizing the sensitivity to all considered signal hypotheses:  
 2656 SM and BSM.

2657 The analysis will follow the flowchart shown in Fig. 5.3

## 2658 5.3 Data Samples and Simulated Events

### 2659 5.3.1 Trigger Requirements

2660 At the L1 trigger level, the events are selected requiring one or two electromagnetic  
 2661 ( $e/\gamma$ ) candidates. In case of a single ( $e/\gamma$ ) candidate, the minimum  $p_T$  is required to be

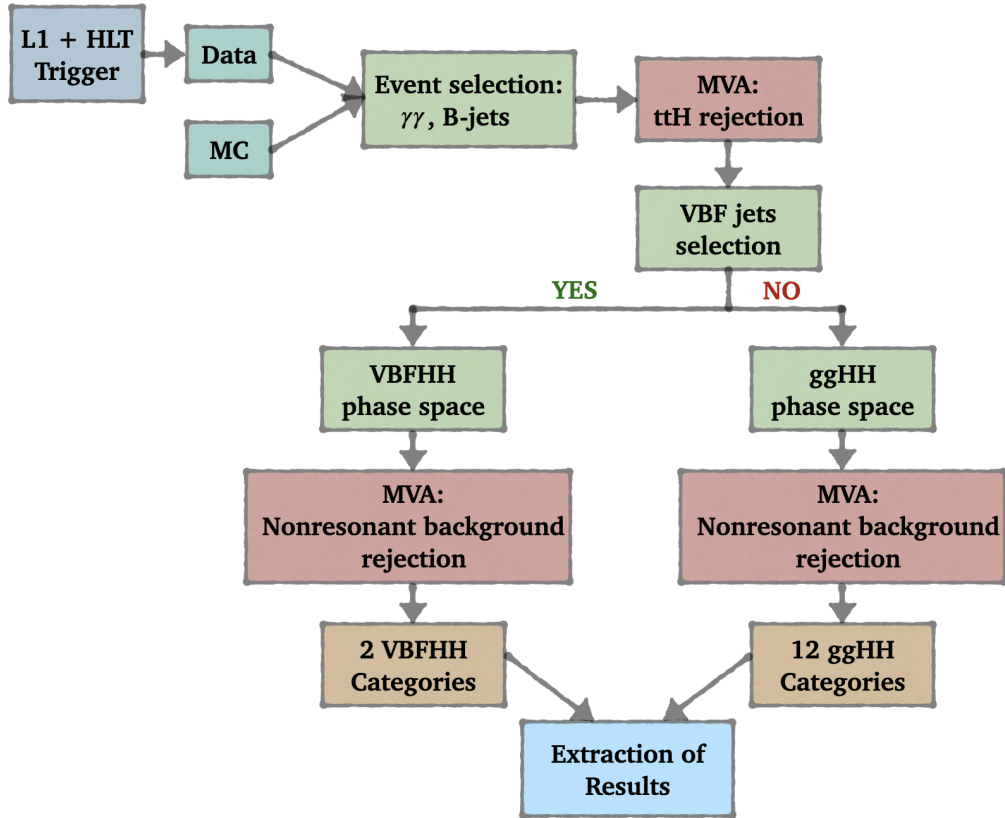


Figure 5.3: Scheme of the analysis workflow.

2662 40 GeV (30-32 GeV in case of an isolated candidate) to maintain the trigger rate at a  
 2663 sustainable level. For the events with a double ( $e/\gamma$ ) candidate, the  $p_T$  thresholds for the  
 2664 leading(subleading) candidates are set to 23(10) GeV in the 2016 data-taking period. In  
 2665 the following years(2017-2018), featuring a higher luminosity, the  $p_T$  thresholds are raised  
 2666 to 25(14) GeV to limit the trigger rate.

2667 Relying on a refined reconstruction of the events, the HLT trigger requirements are tighter  
 2668 than the L1 trigger. Furthermore, the inclusion of the tracker information allows the  
 2669 separation of diphoton from dielectron candidates in dedicated trigger paths. In the HLT  
 2670 diphoton trigger, used for this analysis, the  $p_T$  variable is required to be higher than  
 2671 30(18) GeV for the leading(subleading) photon in 2016, and than 30(22) GeV in 2017 and  
 2672 2018.

2673 In order to reduce the contamination from misidentified jet, additional selections are  
 2674 applied on the Hadronic over electromagnetic energy ratio (H/E ratio), on the photons  
 2675 isolation, and on the photons shower shapes. For good quality of photons the value of  
 2676 H/E is expected to be small.  $R_9$  is the sum of the energy deposition of the  $3 \times 3$  crystals

2677 centered on the most energetic seed crystal in the supercluster divided by the total energy  
 2678 deposition on that supercluster.

2679 The efficiency of the trigger selections is estimated from the data and then used to scale the  
 2680 MC simulations which do not include the trigger effects. In particular, a Tag and Probe  
 2681 (T&P) method with  $Z \rightarrow e^+e^-$  events is used [101]. The "probe" electron is treated as  
 2682 a photon candidate, i.e., ignoring the track information. The efficiency is estimated in  
 2683 intervals of  $p_T$ ,  $\eta$  and  $R_9$ . The different kinematics of the  $Z \rightarrow e^+e^-$  and  $H \rightarrow \gamma\gamma$  events,  
 2684 and the different interaction of electrons and photons with the material upstream the  
 2685 ECAL, resulting in a different shower shape, are properly taken into account. A rescaling  
 2686 of the "probe" electrons weights in intervals of  $\eta$  and  $R_9$ , denser than the ones considered  
 2687 to compute the efficiencies, is performed to match the corresponding distributions of the  
 2688  $H \rightarrow \gamma\gamma$  photons. The scale factors are derived from MC simulations and are shown in  
 2689 Fig. 5.4. The main effect of the scale factors consists in a shift of the  $R_9$  distribution to  
 2690 values close to unity. The shifting is more pronounced in the region  $1 < |\eta| < 1.5$  where  
 2691 the upstream material (beampipe, tracker and support structures) is maximum.

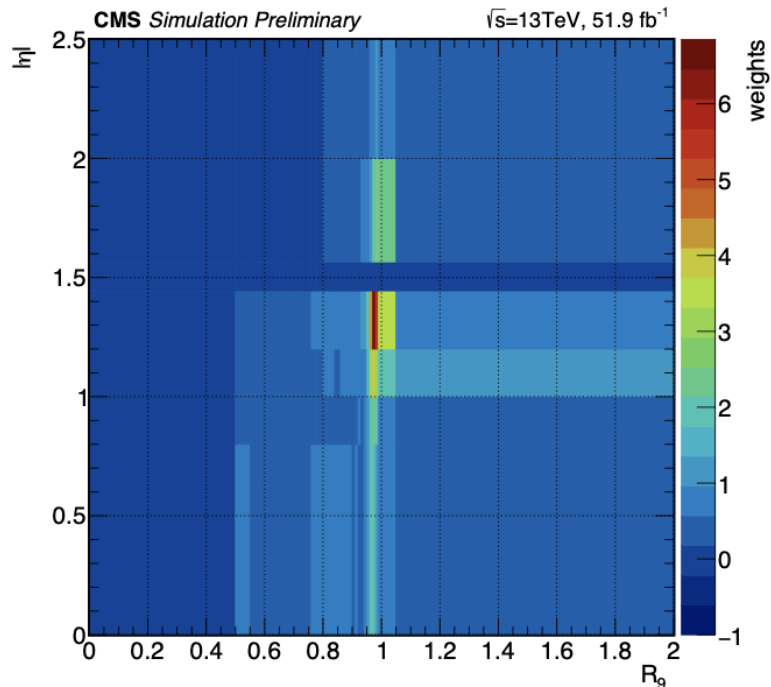


Figure 5.4: Re-weighting factors in  $(R_9, \eta)$  for  $Z \rightarrow e^+e^-$  selected events with respect to  $H \rightarrow \gamma\gamma$  events.

2692 The HLT efficiencies have been derived for both leading and subleading photons with

2693 respect to different  $p_T$ ,  $R_9$  regions separately in two  $\eta$  regions ECAL barrel (EB) and  
 2694 ECAL endcap (EE) directly from data. The efficiencies for 2018 leading and subleading  
 2695 photons are shown in Fig. 5.5.

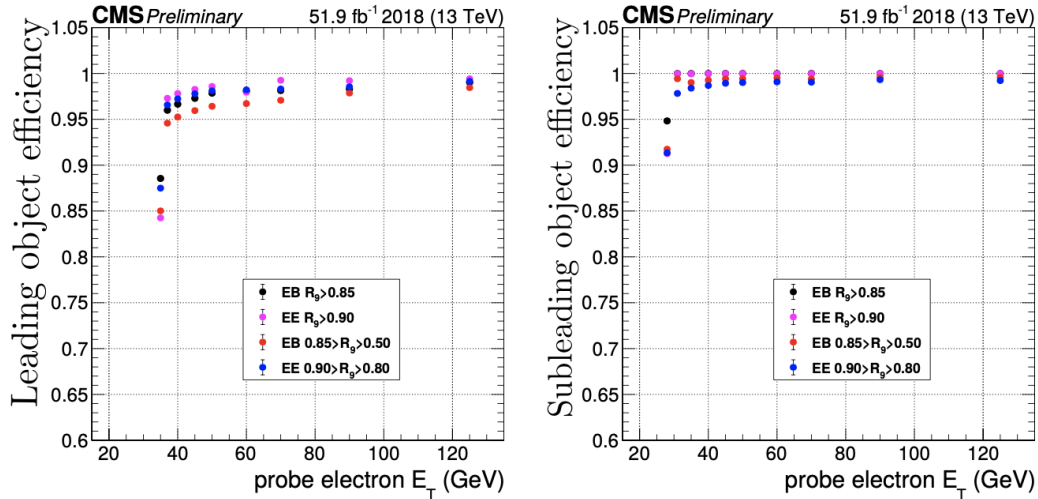


Figure 5.5: Diphoton trigger efficiency measured on 2018 data for  $Z \rightarrow e^+e^-$  events using tag-and-probe method.

### 2696 5.3.2 Simulated Samples

2697 The sample are simulated using the MADGRAPH5 generator for LO accuracy in QCD,  
 2698 and MADGRAPH5 AMCNLO [102] or POWHEG BOX2 [73] generators for up to NLO  
 2699 accuracy in QCD. The generators are interfaced with PYTHIA8 [76] that performs the  
 2700 parton showering and the hadronization. The PYTHIA tuning CUETP8M1 and CP5  
 2701 [103, 104] is used for the underlying events modeling of 2016 and 2017-2018 samples,  
 2702 respectively. The PDFs are taken from the NNPDF3.0 [105] (2016) and NNPDF3.1 [106]  
 2703 (2017 and 2018) sets. The particles interaction with the detector and the subsequent  
 2704 readout is simulated using the Geant4 package [77].

### 2705 Higgs Boson Production Samples

2706 The ggF HH samples are generated using POWHEG BOX2 at the NLO accuracy in  
 2707 QCD and including the full top quark mass dependence [107]. The VBF HH samples are  
 2708 generated using MADGRAPH5 AMCNLO at the LO accuracy. While the two extra light

2709 quark jets at LO cause the VBF HH event topology to diverge from ggF HH, at NNLO  
 2710 the extra gluon radiations cause ggF HH to imitate the VBF HH signals, and nearly 30%  
 2711 of ggF HH events migrate to VBF HH selection.

2712 The SM ggF HH and VBF HH samples are generated along with BSM samples with  
 2713 anomalous ( $\kappa_\lambda, \kappa_t, c_V, c_{2V}$ ) coupling values, as shown in Tab. 5.1.

Mechanism	Coupling values	Cross section (fb) (including k-factor)
ggF HH(SM)	$\kappa_\lambda = 1, \kappa_t = 1$	$3.105 \cdot 10^1$
ggF HH	$\kappa_\lambda = 0, \kappa_t = 1$	$6.973 \cdot 10^1$
ggF HH	$\kappa_\lambda = 2.45, \kappa_t = 1$	$1.312 \cdot 10^1$
VBF HH(SM)	$\kappa_\lambda = 1, c_V = 0, c_{2V} = 0$	$1.73 \cdot 10^0$
VBF HH	$\kappa_\lambda = 1, c_V = 1, c_{2V} = 2$	$1.42 \cdot 10^1$
VBF HH	$\kappa_\lambda = 2, c_V = 1, c_{2V} = 2$	$1.42 \cdot 10^0$
VBF HH	$\kappa_\lambda = 0, c_V = 1, c_{2V} = 1$	$4.61 \cdot 10^0$
VBF HH	$\kappa_\lambda = 1, c_V = 1.5, c_{2V} = 1$	$6.60 \cdot 10^1$
VBF HH	$\kappa_\lambda = 1, c_V = 1, c_{2V} = 0$	$2.71 \cdot 10^1$

Table 5.1: List of the simulated ggF HH and VBF HH BSM samples for anomalous coupling values. The same setup of the corresponding SM sample is used for the simulation.

2714 In addition, ggF HH samples are generated at LO for the BSM benchmarks described in  
 2715 Section 2.5.4 using MADGRAPH5 AMC@NLO. The twelve BSM benchmarks are added  
 2716 together to increase the statistical precision, and then reweighed to any coupling config-  
 2717 uration ( $\kappa_\lambda, \kappa_t, c_V, c_{2V}$ ) using the generator-level information.

2718 We apply a global k-factor to the generated ggF HH and VBF HH signal samples to  
 2719 scale the cross section to NNLO and next-to-NNLO accuracy respectively. The k-factor is  
 2720 obtained for the cross section prediction in the SM and applied to all considered scenarios.  
 2721 The k-factor for the ggF HH cross section depends on the invariant mass of the two Higgs  
 2722 bosons, however, within the region of sensitivity of this analysis, this effect is covered by  
 2723 the total scale uncertainty.

2724 **Background Samples**

2725 The dominant backgrounds in this search are irreducible prompt diphoton production  
 2726 ( $\gamma\gamma$ +jets) and the reducible background from  $\gamma$ +jets events, where the jets are misiden-  
 2727 tified as isolated photons and b jets. Although these backgrounds are estimated using  
 2728 data-driven methods, simulated samples are used for the training of multivariate discrim-  
 2729 inants and the optimization of the analysis categories. Single Higgs boson production,  
 2730 where the Higgs boson decays to a pair of photons, is considered as a resonant background.  
 2731 These production processes are simulated at NLO in QCD precision. The cross sections  
 2732 and decay branching fractions are taken from ref. [29]. The contribution from the other  
 2733 single H decay modes is negligible.

2734 The list of the samples used is shown in Tab. 5.2 and 5.3. Further selections are applied  
 2735 at the generator level to enrich the sample with signal-like events. In particular, for the  
 2736  $\gamma$ +jets at least one jet is required to have a high fraction of electromagnetic energy to  
 2737 mimic a  $e/\gamma$  object.

Mechanism	Generator	Cross section (fb) (including k-factor)
ggH	POWHEG BOX2	$4.41 \cdot 10^4$
qqH	POWHEG BOX2	$3.78 \cdot 10^3$
VH	MADGRAPH5 AMCNLO	$2.25 \cdot 10^3$
tHq	POWHEG BOX2	$7.4 \cdot 10^2$
$b\bar{b}H$	MADGRAPH5 AMCNLO	$5.3 \cdot 10^2$
$t\bar{t}H$	MADGRAPH5 AMCNLO	$5.1 \cdot 10^2$

Table 5.2: List of the simulated SM Higgs boson production samples.

Mechanism	Generator	Cross section(pb)
$\gamma\gamma$ +jets ( $m_{\gamma\gamma} > 80$ GeV)	Sherpa	88.36
$\gamma\gamma$ +jets ( $m_{\gamma\gamma} > 80$ GeV, 1 b-jet)	Sherpa	0.8185
$\gamma\gamma$ +jets ( $m_{\gamma\gamma} > 80$ GeV, 2 b-jets)	Sherpa	0.4874
$\gamma$ +jets ( $m_{\gamma\gamma} > 80$ GeV, $p_T^\gamma > 40$ GeV, EM-enriched)	Sherpa	874.2

Table 5.3: List of the simulated background samples.

## 2738 5.4 Event Selection

### 2739 5.4.1 Photon Reconstruction

2740 Photon candidates are reconstructed as part of the global event reconstruction with par-  
 2741 ticle flow as described in Sec. 4.3, using the algorithms provided centrally from the  
 2742 CMS E/gamma group. A preselection is applied. This loose selection requires cuts on  
 2743 shower shapes, kinematics and isolation variables slightly tighter than the trigger ones to  
 2744 improve Data/MC comparison. A multivariate identification method, based on photon  
 2745 shower-shape and kinematic variables, is used to separate the signal from background pho-  
 2746 tons. A very loose cut on photon ID is also applied. Tab. 5.4 summarizes the preselection  
 2747 cuts. Scale factors are used to cover Data/MC discrepancies related to the identification  
 2748 criteria.

Requirements	Leading Photon	Subleading Photon
$P_T$	30 GeV	20 GeV
$ \eta $	$ \eta  < 1.44$ or $1.55 <  \eta  < 2.5$	
Shower shape and Isolation	$R_9 > 0.8$ or $Iso_{ch} < 20$ GeV or $Iso_{ch}/p_T < 0.3$	
Identification	H/E < 0.08 and MVA ID > -0.9	

Table 5.4: Caption

### 2749 5.4.2 Jet Reconstruction

2750 Jets are reconstructed using the anti-kt algorithm with a distance parameter  $R = 0.4$ .  
 2751 The jet candidates in the event, after passing the ID requirements (Loose ID : 2016, Tight  
 2752 ID : 2017 and 2018 - both ID correspond to efficiency > 99%), must have  $p_T > 25$  GeV,  
 2753  $|\eta| < 2.4$  (2.5) for 2016 (2017-2018), so that they are within the tracker of CMS and can  
 2754 be tagged as coming from b quarks. The jets must also be outside the photon cone with  
 2755 a  $\Delta R(j, \gamma) > 0.4$ .

## 2756 DeepJet b-Tagger

2757 Within CMS, a deep neural network (DNN) classifier (DeepJet [91]) is used. The al-  
2758 gorithm exploits the distance between the production and the decay position of the  
2759 b-hadrons, of the order of 1 mm, resulting in a secondary vertex resolvable from the  
2760 primary interaction vertex. The algorithm exploits also the fact that the b hadron decay  
2761 can produce leptons (and undetected neutrinos) with high  $p_T$ . Finally, because of the  
2762 high bottom quark mass, the b-jet have a wider  $\eta - \phi$  extension than a light flavor jet.  
2763 The DeepJet score is used for the b-jets candidates selection and as input for the MVA  
2764 classifiers.

### 2765 5.4.3 Energy Regression for b-Jets

#### 2766 Level 1: DNN b-jet Energy Regression

2767 A b-jet energy regression is developed to improve b-jets resolution and, therefore, the  
2768 invariant mass of the two jets coming from the  $H \rightarrow b\bar{b}$  decay. The energy correction,  
2769 for each jet, is computed through a regression-based on a deep neural network, trained  
2770 on jet properties and jet composition information. The b-jet energy correction and b-jet  
2771 resolution estimator are output simultaneously by the neural network (NN). This allows  
2772 to correct the measured jet  $p_T$  and to use the resolution estimator to improve the analysis  
2773 sensitivity, as described later. This regression was developed for the CMS Collaboration  
2774 [108] and is analysis independent. The b-jet energy regression improves single jet energy  
2775 resolution by about 15%, and dijet invariant mass resolution by about 20-25% depending  
2776 on the phase-space. The regression technique was validated on data for the CMS  $H \rightarrow b\bar{b}$   
2777 discovery.

#### 2778 Level 2: $m_{jj}$ Oriented Energy Regression for b-j=Jets

2779 A regression to improve the mass resolution of the reconstructed  $M_{H \rightarrow b\bar{b}}$  was developed  
2780 (Level 2 or L2 regression). The approach is to improve the resolution beyond what was  
2781 achieved using the b jet  $p_T$  regression (Level 1 or L1 regression), which used as inputs jet-  
2782 related variables. However, the  $M_{H \rightarrow b\bar{b}}$  resolution can be further improved by using event



2783 variables, particularly the missing transverse energy  $E_T^{miss}$  and the kinematic variables of  
 2784 the reconstructed physics objects.

2785 The comparison of the mass resolution in this analysis, obtained including L2 and L1  
 2786 regression, to mass resolutions without these corrections are shown in Fig. 5.6. The L1  
 2787 regression affects both the resolution ( $\sigma$ ) and the position of the peak ( $\mu$ ), while the L2  
 2788 leaves the latter unchanged. The L1 regression moves the  $m_{jj}$  peak position by 5.5 GeV  
 2789 (5%) closer to the expected Higgs mass. Overall the b-jet energy corrections improve the  
 2790  $m_{jj}$  resolution of about 25%.

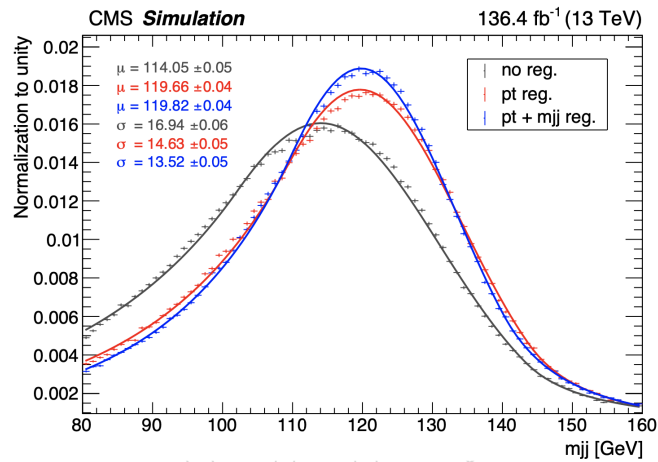


Figure 5.6: Mass resolution in this analysis for SM ggF HH samples merging all years weighted by the luminosity: without regression (grey), with L1 regression (red), and with L1+L2 regression (blue).

## 2791 Validation of $M_{H \rightarrow b\bar{b}}$ Regression

2792 To validate the L2 regression on  $HH \rightarrow \gamma\gamma b\bar{b}$ , similar  $ZZ \rightarrow l^+l^- b\bar{b}$  was considered. where  
 2793  $l$  is  $e$  or  $\mu$  leptons.  $ZZ$  events were obtained by applying the following selection criteria.  
 2794 The samples were obtained using the framework of the  $VH \rightarrow b\bar{b}$  analysis.

- 2795 • Trigger in the  $e^+e^-$  channel: HLT\_Ele23\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\_DZ
- 2796 • Triggers in the  $\mu^+\mu^-$  channel: HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL
- 2797 HLT\_Mu17\_TrkIsoVVL\_TkMu8\_TrkIsoVVL
- 2798 HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL\_DZ
- 2799 HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL\_DZ

- 2800 • Lepton  $p_T > 25$  GeV and  $|\eta| < 2.5$
- 2801 •  $85 < M(e^+e^-) \leq 97$  GeV
- 2802 •  $p_T(ll) > 50$  GeV
- 2803 • Jet  $p_T > 25$  GeV and  $|\eta| < 2.5$
- 2804 • Primary jet with DeepCSV score  $> 0.4941$  (medium) and secondary jet with
- 2805 DeepCSV score  $> 0.1522$  (loose) working points.

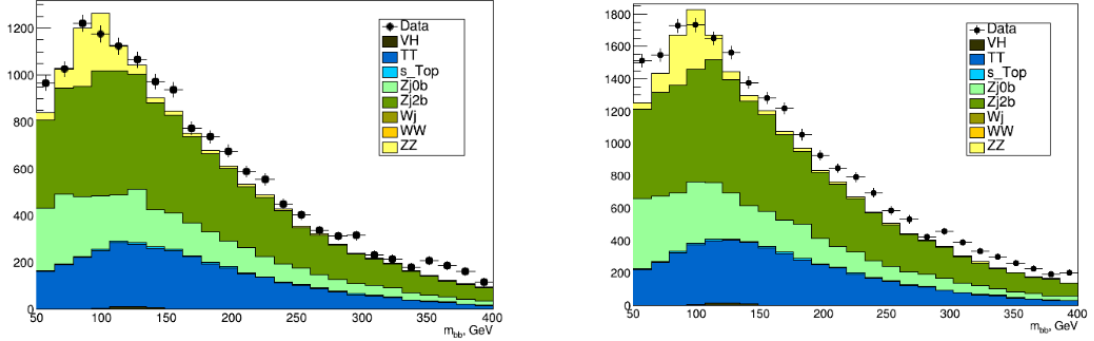


Figure 5.7: The  $m_{jj}$  distributions in the data, the ZZ signal, and the DY and  $t\bar{t}$ +jets backgrounds before BDT selection for the  $e^+e^-$  (left) and the  $\mu^+\mu^-$  (right) channels.

2806 The  $m_{jj}$  distributions after these selections are shown in Fig. . DY with b jets and  $t\bar{t}$ +jets  
 2807 make up the majority of events. The signal ZZ also has a sizeable component. The S/B  
 2808 ratio (the S here being ZZ) is 0.21. To improve this, a BDT was trained to reject the DY  
 2809 and tt+jets background. The following variables were used:

$$\begin{aligned}
 &-\Delta\eta(j, j), \Delta\eta(l, l), \Delta\eta(jj, ll), \\
 &-m_{jj}/p_T(jj), m_u/p_T(ll), \\
 &-p_T(l_1)/m_u, p_T(l_2)/m_u, p_T(j_1)/m_{jj}, p_T(j_2)/m_{jj}, \\
 &-b\text{-jet}_1 \text{ DeepCSV}, b\text{-jet}_2 \text{ DeepCSV}, \\
 &-\Delta R(j, j), \\
 &-p_T\text{-sum of all jets} \\
 &-m_u
 \end{aligned}$$

2810 The signal and background efficiencies as a function of the BDT score, the ROC curve,  
 2811 and the overtraining checks are shown in Fig. 5.8. Based on the ROC curve, one working  
 2812 points is explored:  $\text{BDT} > 0.2$  corresponding to a signal efficiency of about 60% and  
 2813 background efficiency of about 20%. The  $m_{jj}$  distribution after this cut is shown in Fig.  
 2814 5.9. The S/B ratio after BDT cut is 0.35 for  $e^+e^-$  channel and 0.30 for the  $\mu^+\mu^-$  channel.

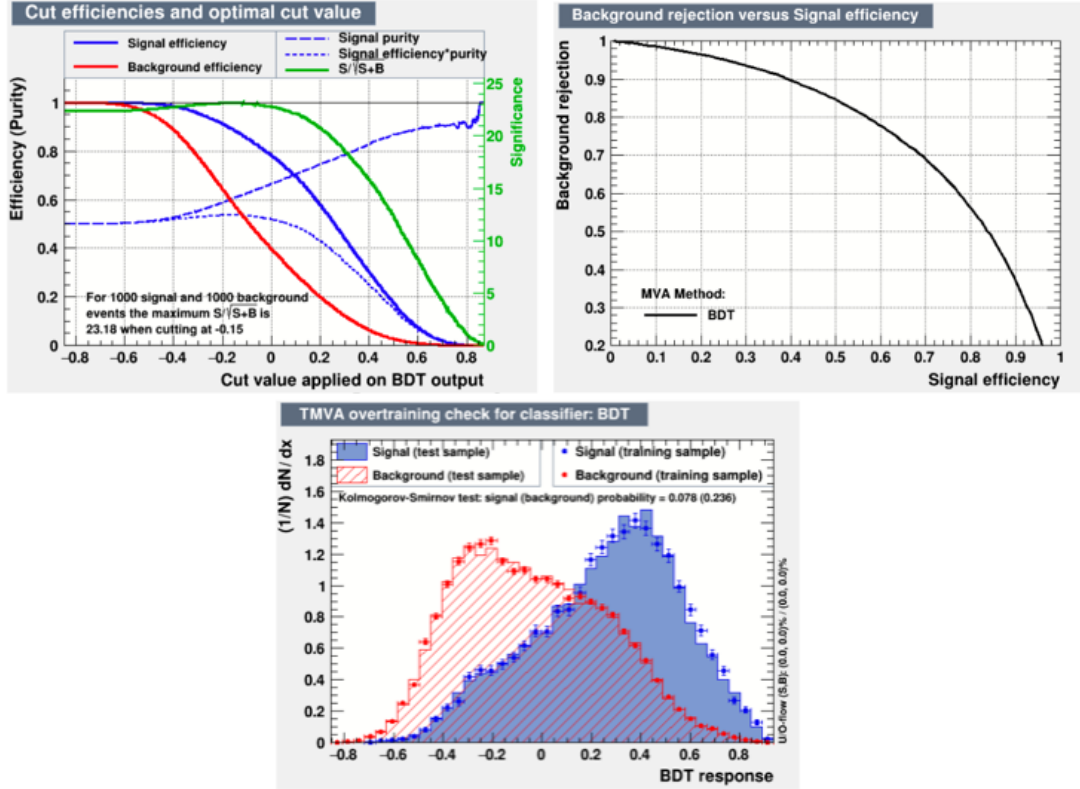


Figure 5.8: The signal and background efficiencies as a function of the BDT score (upper left), ROC curve (upper right) and the overtraining check (lower).

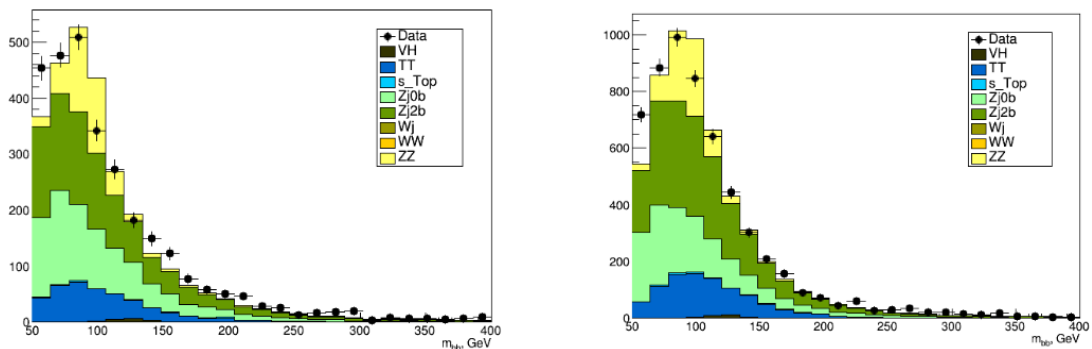


Figure 5.9: The  $m_{jj}$  distributions in the data, the ZZ signal, and the DY and  $t\bar{t}$ +jets backgrounds with  $\text{BDT} > 0.2$  selection for the  $e^+e^-$  (left) and the  $\mu^+\mu^-$  (right) channels.

2815 The validation is performed, looking at the  $p_T$  balance properties. For jets, the invariant  
 2816 mass of the dijet system can be connected to the transverse momentum of the dijet system.

## 5.4. Event Selection

2817 Therefore a regression that improves  $m_{jj}$  resolution also improves to some extent the  
 2818 momentum balance in the event, that may be distorted due to the presence of neutrinos.  
 2819 Traditionally the RMS value of two distributions:  $p_{T_{bb}} \cdot C_{p_T}^{reg}/p_{T_{\gamma\gamma}}$  and  $p_{T_{bb}}/p_{T_{\gamma\gamma}}$ , provides  
 2820 the confirmation of  $E_T^{miss}$  effect on the  $p_T$  balance. Here, we propose a slightly modified  
 2821 method of " $p_T$ -balance" using  $\frac{p_T(l)-p_T(b\bar{b})}{p_T(l)+p_T(b\bar{b})}$ . This distribution is by construction more  
 2822 symmetric, and easier to used to measure the effect of the L2 regression.

2823 In Fig. 5.10 the distribution of  $\frac{p_T(l)-p_T(b\bar{b})}{p_T(l)+p_T(b\bar{b})}$  is shown for  $E_T^{miss} > 40$  GeV and  $E_T^{miss} < 40$   
 2824 GeV. The improvement is studied separately for data and for MC. A visible effect in the  
 2825 reduction of the resolution of  $\frac{p_T(l)-p_T(b\bar{b})}{p_T(l)+p_T(b\bar{b})}$  is observed in events with large  $E_T^{miss}$ . This  
 2826 improvement in data is compatible with the one observed in MC. We can conclude out of  
 2827 it that the L2 regression is working as expected.

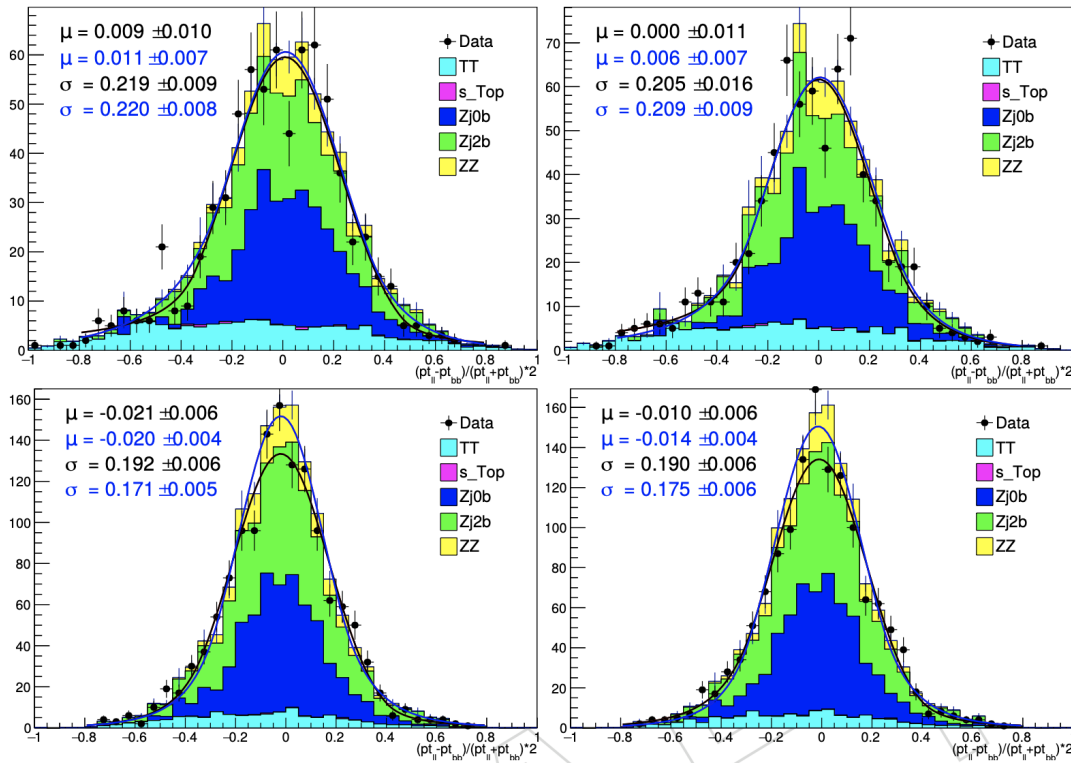


Figure 5.10: Data and MC events with a Crystal-Ball. TOP -  $E_T^{miss} > 40$  (effect observed and expected), BOTTOM -  $E_T^{miss} < 40$  (no effect observed and expected). Left - with L1 regression, right - with L1+L2 regression. Results for 2017 data sample.

#### 2828 5.4.4 Selection of the $H \rightarrow \gamma\gamma$ Candidate

2829 The  $H \rightarrow \gamma\gamma$  candidate is built using the photons passing the identification criteria  
 2830 described in Section 5.4.1. If more than two photons are present, the two photons with  
 2831 the highest  $p_T$  are selected. The leading and subleading photons are required to have  
 2832  $p_T^\gamma/m_{\gamma\gamma} > 1/3$  and  $p_T^\gamma/m_{\gamma\gamma} > 1/4$ , respectively. In addition, the diphoton invariant mass  
 2833 is required to be in the window  $100 \text{ GeV} < m_{\gamma\gamma} < 180 \text{ GeV}$ . The selections on the photon  
 2834  $p_T^\gamma/m_{\gamma\gamma}$  ratio was proven to prevent distortions of the  $m_{\gamma\gamma}$  spectrum on the low mass side  
 2835 with a negligible loss of efficiency on the  $H \rightarrow \gamma\gamma$  signal.

#### 2836 5.4.5 Selection of the $H \rightarrow b\bar{b}$ Candidate

2837 The jets passing the identification selections are also required to have  $|\eta| < 2.4$  and  $|\eta|$   
 2838  $< 2.5$  for 2016 and 2017-2018 datasets, respectively, and  $p_T > 25 \text{ GeV}$  for both jets. An  
 2839 angular distance from the two selected photons to the jet must be at least 0.4. The two  
 2840 b-jets with the highest b-tag score are selected to build the  $H \rightarrow b\bar{b}$  candidate in the HH  
 2841 categories while minimum cut on Deep Jet b-tag discriminator is  $> 0$ . Finally, the  $m_{jj}$   
 2842 value is required to be in the window  $70 \text{ GeV} < m_{jj} < 190 \text{ GeV}$ .

2843 A summary of the preselection criteria for jets and photons can be found in Tab. 5.5.

Photons		b-jets	
Variable	Selection	Variable	Selection
$p_T^{\gamma 1}$ [GeV]	$> m_{\gamma\gamma}/3$	$p_T$ [GeV]	$> 25$
$p_T^{\gamma 2}$ [GeV]	$> m_{\gamma\gamma}/4$	$\Delta R_{\gamma j}$	$> 0.4$
$ \eta $	$< 2.5$	$ \eta $	$< 2.4$
$m_{\gamma\gamma}$ [GeV]	[100,180]	$m_{b\bar{b}}$ [GeV]	[70,190]
		DeepJet Score	$> 0$

Table 5.5: Summary of the baseline selection criteria.

#### 2844 5.4.6 Requirements for the VBF HH Topology

2845 The events with a  $H \rightarrow \gamma\gamma$  and a  $H \rightarrow b\bar{b}$  candidate (HH candidates) are required to have  
 2846 at least two additional jets passing the tight PU ID selection. The jets are required to  
 2847 have  $p_T > 40$  (30) GeV for leading (subleading),  $|\eta| < 4.7$ , and an angular distance  $\Delta R >$

2848 0.4 from each of the two selected photons and b-jets. The two jets with the highest dijet  
 2849 invariant mass are selected as VBF jets candidates.

## 2850 5.5 Background Rejection

2851 The distribution of the  $m_{\gamma\gamma}$  and  $m_{jj}$  distributions after requiring the presence of a  $H \rightarrow \gamma\gamma$   
 2852 and a  $H \rightarrow b\bar{b}$  candidate in the events are visible in Fig. 5.11. The signal appears as  
 2853 a peak at 125 GeV in the two distributions smeared by the experimental resolution. A  
 2854 continuum background mainly from  $\gamma\gamma$ +jets and  $\gamma$ +jets events dominates the HH signal  
 2855 region. The single Higgs production processes represent another important source of  
 2856 background because it is resonant in the  $m_{\gamma\gamma}$  distribution as the HH signal.

2857 In order to maximize the separation of the signals from the background contaminations,  
 2858 specific MVA-based strategies were developed. MVA classifiers based on different archi-  
 2859 tectures are trained using the simulated events. The classifier outputs are used to define  
 2860 the signal regions and also to classify the events in exclusive categories. The MC simula-  
 2861 tion events are divided in two subsets. One subset of events is used for the MVA trainings,  
 2862 the other subset is used for the MVA outputs validation, the category optimization, and  
 2863 the signal modeling.

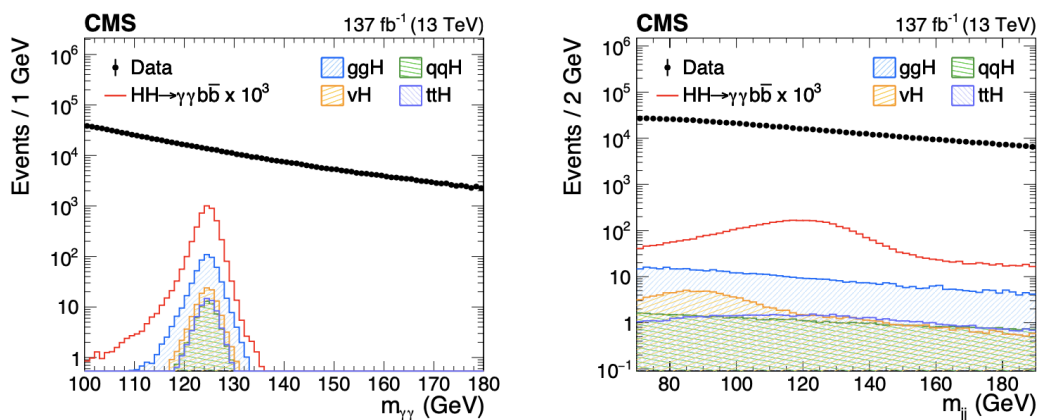


Figure 5.11: Distributions of  $m_{\gamma\gamma}$  (left) and  $m_{jj}$  (right) for the selected HH candidates for data and for the simulated single and double Higgs processes.

### 2864 5.5.1 $t\bar{t}H(\gamma\gamma)$ Background Rejection

2865 With two photons resonant on  $m_H$  and two b-jets in the final state, the  $t\bar{t}H(\gamma\gamma)$  process  
 2866 was found to be one of the main backgrounds in the most sensitive HH categories. There-  
 2867 fore, a specific strategy was developed to discriminate the  $t\bar{t}H(\gamma\gamma)$  from the  $HH \rightarrow b\bar{b}\gamma\gamma$   
 2868 mechanisms. In particular, a DNN classifier is trained using the simulated  $t\bar{t}H$  events as  
 2869 background and the combination of all the twelve simulated ggF HH benchmarks samples  
 2870 (including the SM) as signal. The twelve benchmarks are combined with the same weight.  
 2871 It was verified that this strategy improves the BSM ggF HH selection efficiency with a  
 2872 negligible impact on the SM ggF HH efficiency. This strategy makes also the classifier  
 2873 suitable to separate the  $t\bar{t}H$  from the VBF HH events, not included in the training be-  
 2874 cause the corresponding MC simulation became available only at an advanced stage of  
 2875 the analysis. The classifier exploits the angular variables related to the different topo-  
 2876 logies of the two processes as well as the presence of a W boson, decaying hadronically or  
 2877 leptonically, originated by the top quark decay.

2878 The discriminant uses a combination of low-level information from the individual PF  
 2879 candidates and high-level features describing kinematic properties of the event. The  
 2880 kinematic variables used in the DNN training can be classified in three groups: angular  
 2881 variables, variables to distinguish semileptonic decays of W bosons produced in the top  
 2882 quark decay, and variables to distinguish hadronic decays of W bosons.

#### 2883 Angular Variables

- 2884 • The minimum angular distance between one of the two selected photons and one of  
 2885 the two selected jets  $\Delta R_{min}(\gamma, b - jet)$ .
- 2886 • Helicity angle:  $|\cos\theta_{HH}^{CS}|$ . It is the Collins-Soper angle [109] between the direction  
 2887 of the  $H \rightarrow \gamma\gamma$  candidate and the average beam direction in the HH center-of-mass  
 2888 frame.
- 2889 • The cosine of the angle in the dijet rest-frame between the leading jet and the beam  
 2890 axis  $|\cos\theta_{jj}|$ .

2891 The distributions of these angular variables are shown in Fig. 5.12.

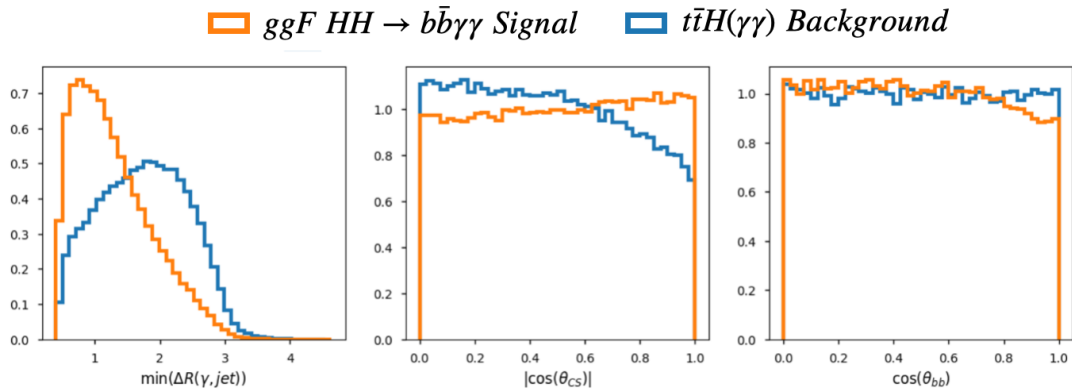


Figure 5.12: Angular variables used in the training, from left to right:  $\Delta R_{min}(\gamma, b - jet)$ ,  $|\cos\theta_{HH}^{CS}|$  and  $|\cos\theta_{jj}|$ . These variables are used for the training of the  $t\bar{t}H$  discriminant.

### 2892 Variables to Reject Events with a Semileptonic Decay of the W Boson

- 2893 • The  $E_T^{miss}$  absolute value and its azimuthal angles with the selected b-jets
- 2894  $\Delta\phi(p_T^{miss}, b - jet)$ .
- 2895 • The pT of the leading and subleading electrons and muons of the event, if any,

2896 Events with a leptonic decay W boson are expected to have significant MET due to the  
 2897 presence of neutrinos. Leptonic decay W boson event also could have leptons recon-  
 2898 structed in the final state. Thus the four vectors of reconstructed leptons with  $p_T > 10$   
 2899 GeV are also included in the training. The distributions of most of these variables are  
 2900 shown in Fig. 5.13.

### 2901 Variables to Reject Events with a Hadronic Decay of the W Boson

- 2902 • A top quark sensitive variable  $\chi_{top}^2$  defined for the events with at least two additional
- 2903 jets as:

$$\chi_{top}^2 = \min_{j1j2jb} \left[ \left( \frac{m_W - m_{j1j2}}{0.1 \cdot m_W} \right)^2 + \left( \frac{m_t - m_{jbj1j2}}{0.1 \cdot m_t} \right)^2 \right] \quad (5.2)$$

2904 Where j1 and j2 are two among the additional jets and jb is one of the two selected b-  
 2905 jets.  $m_W$  and  $m_t$  are the true mass values of the W boson and the top quark, taken to  
 2906 be 80.3 GeV and 173.5 GeV, respectively. Among all the possible jet combinations,  
 2907 the one minimizing the quantity is chosen. In case of four or more additional jets,  
 2908 an additional  $\chi_{top}^2$  variable is built in the same way with the remaining jets.



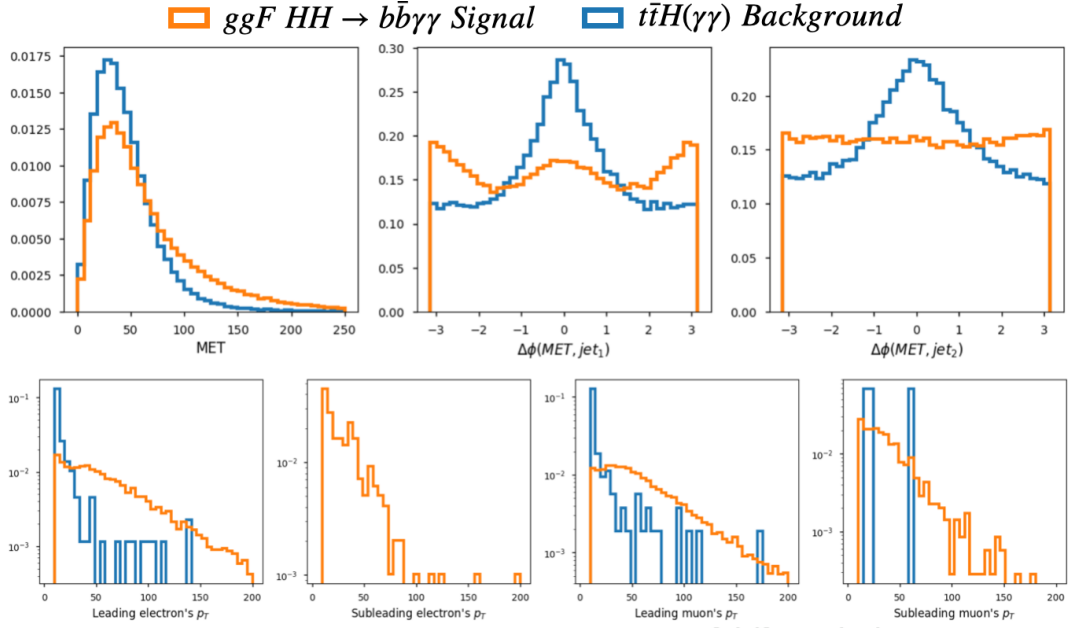


Figure 5.13: Major variables used in the training to reject events with a leptonic-decay W boson, from left to right top to bottom:  $E_T^{miss}$ ,  $\Delta\phi(p_T^{miss}, b - jet1)$ ,  $\Delta\phi(p_T^{miss}, b - jet2)$  and the transverse momentum of the leading and subleading electrons and muons.

2909 This variable is calculated for events with at least 2 additional jets and 4 additional jets  
 2910 besides the two b-tagged jets, and the distributions in these two cases are shown in Fig.  
 2911 5.14.

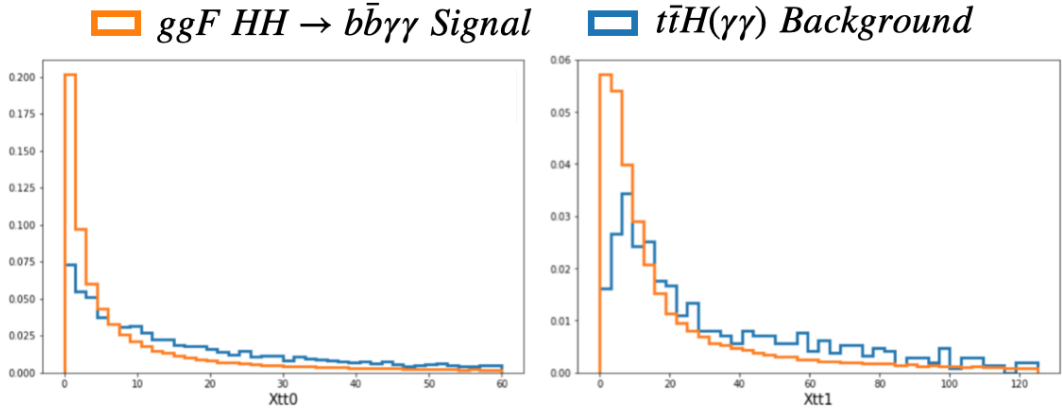


Figure 5.14:  $\chi_{top}^2$  variables in training to reject events with a hadronic decay W boson, for events with at least 2 additional jets (left plot) and 4 additional jets (right plot) besides the two b-tagged jets.

2912 A selection on the output score of this DNN ( $t\bar{t}H$ -score) is applied to reject the  $t\bar{t}H$  events  
 2913 from the HH enriched categories. The selection on the  $t\bar{t}H$ -score is optimized together  
 2914 with the ggF HH category boundaries definition, and separately with the VBF HH one,  
 2915 to provide the best sensitivity. The chosen working point ensures a  $t\bar{t}H$  rejection of about

2916 80% with an efficiency of about 95% on the ggF HH signal. A good performance is also  
 2917 achieved with respect to the VBF HH signal. In particular, after the VBF HH additional  
 2918 requirements presented in Section 5.4.6, the chosen working point provides a  $t\bar{t}H$  rejection  
 2919 of about 85% with 90% efficiency on the VBF HH signal.

## 2920 Validation of the Method

2921 In order to validate the training variables for the  $t\bar{t}H$  discriminant, their distributions in  
 2922 data and MC simulation were compared. The comparison included the preselected events  
 2923 with a  $H \rightarrow \gamma\gamma$  and a  $H \rightarrow b\bar{b}$  candidate (HH candidates) outside the diphoton invariant  
 2924 mass region  $115 \text{ GeV} < m_{\gamma\gamma} < 135 \text{ GeV}$  containing the expected signal. The distributions  
 2925 from simulation were found compatible with the distributions observed in data. The  
 2926 distribution of the  $t\bar{t}H$ -score for the different background sources was also studied by  
 2927 means of MC simulations considering for simplicity the 2016 and 2017 datasets, as shown  
 2928 in Fig. 5.15 (left). The peak at  $t\bar{t}H$ -score = 0, corresponding to  $t\bar{t}H$ -like events, is  
 2929 populated by all the  $t\bar{t} + (X)$  processes, while the peak at  $t\bar{t}H$ -score = +1 is populated  
 2930 by the HH-signal and all the other backgrounds, hence dominated by the  $\gamma\gamma$ +jets events.  
 2931 Considering the full Run 2 dataset, the  $t\bar{t}H$ -score distributions for the selected data events,  
 2932 and for the ggF HH and single Higgs simulated events, are found compatible as shown in  
 2933 Fig. 5.15.

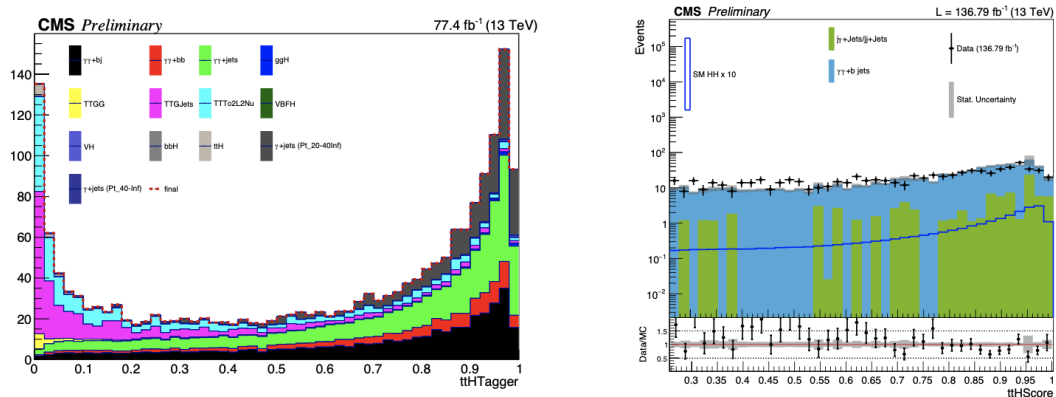


Figure 5.15: *Left*:  $t\bar{t}H$ -score distribution for all the main background components from the MC simulations of the 2016 and 2017 events. *Right*: Comparison between data and MC simulation of the  $t\bar{t}H$ -score distribution for the HH candidates with  $t\bar{t}H$ -score  $> 0.26$ .

## 5.5.2 Non-resonant Background Rejection

### Background Reduction in the ggF HH Signal Region

The dominating background for the ggF HH signal region consists in the  $\gamma(\gamma)$ +jets processes. The optimal background rejection is achieved through an MVA approach using a BDT classifier trained using MC simulated events. The variables optimized for this training exploit the HH system kinematic, the identification variables to reject background processes with jets misidentified as photons or b-jets, and the resolution variables to account for the resonant nature of the signal. In particular, the variables chosen are:

- The  $H \rightarrow \gamma\gamma$  and  $H \rightarrow b\bar{b}$  candidates kinematics described by  $p_T^\gamma/m_{\gamma\gamma}$  for each of the two selected photons, and  $p_T^j/m_{jj}$  for each of the two selected jets;
- The transverse balance of the HH system consisting in  $p_T^{\gamma\gamma}/m_{\gamma\gamma jj}$  and  $p_T^{jj}/m_{\gamma\gamma jj}$  where  $p_T^{\gamma\gamma}$  and  $p_T^{jj}$  are the diphoton and dijet transverse momentum respectively, and  $m_{\gamma\gamma jj}$  is the four-objects invariant mass;
- The  $|\cos\theta_{HH}^{CS}|$  and  $|\cos\theta_{jj}|$  helicity angles as defined for the  $t\bar{t}H$  discriminant, and in addition, the cosine of the angle in the diphoton rest-frame between the leading photon and the beam axis  $|\cos\theta_{\gamma\gamma}|$ ;
- The angular separation of the objects, i.e.,  $\Delta R_{min}(\gamma, b - jet)$  as defined for  $t\bar{t}H$  discriminant, and additionally, the  $\Delta R$  between the other selected photon and jet;
- The b-tag score provided by the DeepJet algorithm for the two selected jets;
- The photon ID output for the two selected photons;
- The energy resolution for the two selected photons, and the diphoton energy resolution estimated by the photon energy regression algorithm;
- The energy resolution for the two selected jets, and the dijet energy resolution estimated by the b-jet energy regression algorithms;
- The global event energy density  $\rho$  to account for the different pileup conditions.

## 5.5. Background Rejection

2959 The distributions of some of the most discriminating variables for data and simulated  
 2960 events are shown in Fig. 5.16. As for the  $t\bar{t}H$  discriminant, this BDT is trained using a  
 2961 combination of the twelve ggF HH BSM benchmarks samples as signal, and the  $\gamma(\gamma)$ +jets  
 2962 as background. Each event used for the training is weighed by the inverse of the esti-  
 2963 mated diphoton and dijet energy resolutions because the events with good resolutions are  
 2964 expected to provide the highest sensitivity to the HH signal.

2965 In order to prevent for any sculpting of the  $m_{\gamma\gamma}$  and  $m_{jj}$  distributions, the  $p_T^\gamma$  and  $p_T^j$   
 2966 variables are provided as input for the BDT training scaled by  $m_{\gamma\gamma}$  and  $m_{jj}$ , respectively.  
 2967 This is fundamental for a correct estimation of the signal and background. The training  
 2968 is performed separately for the three years of data taking because the different detector  
 2969 conditions have modified the variables distributions and correlations. This strategy offsets  
 2970 the differences across the years, providing very similar BDTs output distributions for the  
 2971 three years, both for signal and backgrounds. Therefore, it is possible to merge the three  
 2972 distributions and then uniformly optimize the BDT selections. The distributions of the  
 2973 BDT output for signal and background are very well separated. In order to avoid problems  
 2974 of numerical precision when defining optimal signal-enriched regions, the BDT output is  
 2975 transformed such that the signal distribution is uniform. This transformation is applied  
 2976 to all events, both in simulation and data. The distribution of the MVA output for data  
 2977 and simulated events is shown in Fig. 5.17.

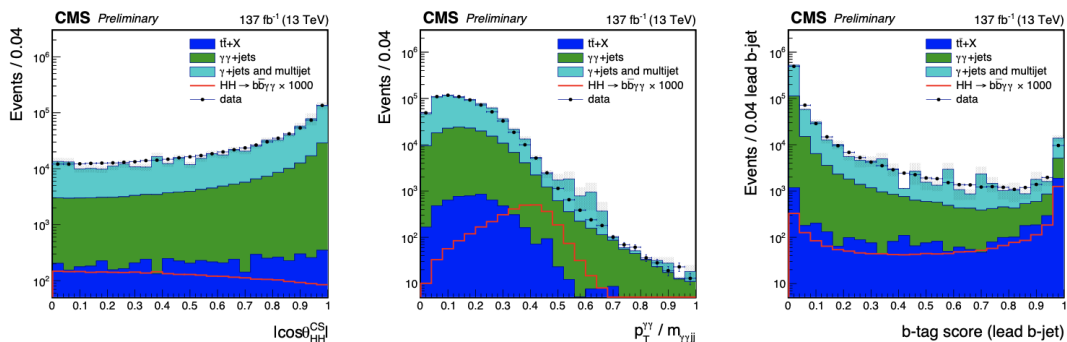


Figure 5.16: Distribution of  $|\cos\theta_{HH}^{CS}|$  (left),  $p_T^{\gamma\gamma}/m_{\gamma\gamma jj}$  (center), and the b-tag score for the leading jet (right) of the selected HH candidates for the data as well as the simulated  $t\bar{t}+X$ ,  $\gamma\gamma$ +jets,  $\gamma$ +jets, and multijet events excluding the signal region  $120 < m_{\gamma\gamma} < 130$  GeV. The distribution of the simulated  $HH \rightarrow b\bar{b}\gamma\gamma$  events is also shown with a red line.

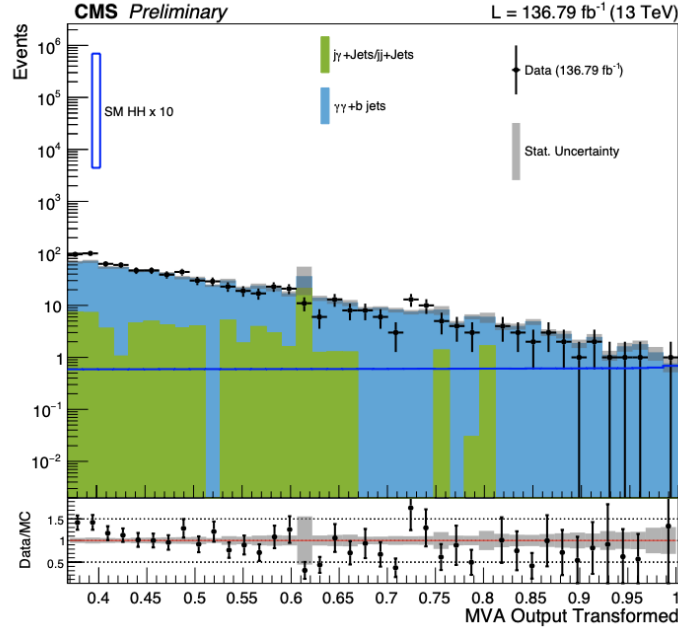


Figure 5.17: The distribution of the MVA output to discriminate the ggF HH signal from the continuum background for the selected data and simulated events in the ggF HH signal region (including the selection MVA score  $> 0.37$ ).

## 2978 Background Reduction from the VBF HH Signal Region

2979 Analogously to the ggF HH process, the background for the VBF HH is dominated by  
 2980  $\gamma(\gamma)$ +jets processes. In addition, the ggF HH (+jets) process represents a contamination  
 2981 in the VBF HH signal region limiting the sensitivity to the vector boson fusion production  
 2982 mode. Therefore, an MVA approach is used to optimize the separation of the VBF HH  
 2983 events from the continuum background and from the ggF HH events. In particular, a  
 2984 BDT multi-classifier is trained using the simulated events to discriminate between three  
 2985 classes of processes: VBF HH, ggF HH, and  $\gamma(\gamma)$ +jets. Beside the variables already  
 2986 optimized for the background rejection from the ggF HH signal region, additional variables  
 2987 exploiting the VBF HH kinematic are included in the training. Such variables improve the  
 2988 separation of the two HH production modes and also the continuum background rejection.  
 2989 In particular, the VBF jets are produced in opposite directions at large pseudorapidities  
 2990 recoiling against the diHiggs object. The VBF jets feature also a  $p_T$  typically larger than  
 2991 the background jets as well as a large dijet invariant mass. The additional variables are:

- 2992 • The dijet invariant mass  $m_{jj}^{VBF}$ ;
- 2993 • The VBF jets kinematic described by their  $p_T/m_{jj}^{VBF}$  and pseudorapidities  $\eta_1^{VBF}$

2994 and  $\eta_2^{VBF}$ ;

- 2995 • Product and difference of pseudorapidity of the two VBF jets;
- 2996 • Quark-gluon likelihood [110] of the two VBF jets to discriminate between jets orig-  
2997 inating from quarks and from gluons;
- 2998 • Minimum angular distance between the VBF jets and the selected photons  
2999  $\Delta R_{min}(j^{VBF}, \gamma)$ , or the selected b-jets  $\Delta R_{min}(j^{VBF}, b - jet)$ ;
- 3000 • The diHiggs kinematics described by the HH transverse momentum  $p_T^{HH}$  and the  
3001  $\widetilde{M}_x$  variable.
- 3002 • Centrality variable for the  $H \rightarrow \gamma\gamma$  and a  $H \rightarrow b\bar{b}$  candidates defined as:

$$C_{xx} = \exp \left[ -\frac{4}{\left(\eta_1^{VBF} - \eta_2^{VBF}\right)^2} \left( \eta_{xx} - \frac{\eta_1^{VBF} + \eta_2^{VBF}}{2} \right)^2 \right] \quad \text{with } xx = \gamma\gamma, bb \quad (5.3)$$

3003 because the two Higgs boson candidates are typically produced centrally with respect  
3004 to the VBF jets. where H is the Higgs boson candidate reconstructed either from  
3005 diphoton or dijet pairs, and  $\eta_1^{VBF}$  and  $\eta_2^{VBF}$  are the pseudorapidities of the two  
3006 VBF-tagged jets.

3007 The distributions of some of the most discriminating additional variables for data and  
3008 simulated events are shown in Fig. 5.18. The training is performed separately in two  
3009 four-body mass categories because of the different VBF HH kinematics (see Section 5.6.1).  
3010 The four-body mass variable  $\widetilde{M}_x$  is defined as Eq. 5.1. This definition reduces the impact  
3011 of the jet and photon energy resolutions on the reconstructed four-body mass.

3012 Two  $\widetilde{M}_x$  categories are defined by the selections  $\widetilde{M}_x < 500$  GeV and  $\widetilde{M}_x > 500$  GeV. In  
3013 both categories the training is performed using as signal a combination of the SM VBF  
3014 HH sample and the BSM VBF HH sample with  $c_{2V} = 0$ . This strategy ensures a good  
3015 sensitivity both to the SM and to the BSM hypotheses, especially to the  $c_{2V} = 0$  case.  
3016 The distribution of the BDT multiclassifier output relative to the VBF HH class for the  
3017 data events (VBF HH-BDT score) as well as for the VBF HH and single Higgs simulated  
3018 events is shown in Fig. 3.12 for the two  $\widetilde{M}_x$  categories.

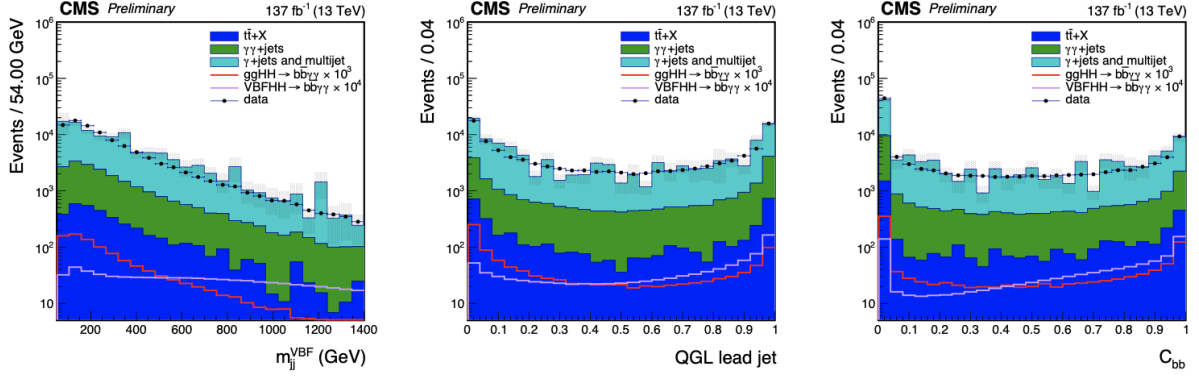


Figure 5.18: Distribution of  $m_{jj}^{VBF}$  (left), the quark-gluon likelihood (center), and the  $C_{bb}$  centrality variable for the selected VBF HH candidates for the data as well as the simulated  $t\bar{t}+X$ ,  $\gamma\gamma$ +jets,  $\gamma$ +jets, and multijet events, excluding the signal region  $120 < m_{\gamma\gamma} < 130$  GeV. The distributions of the simulated ggF HH and VBF HH events are also shown with red and purple lines, respectively.

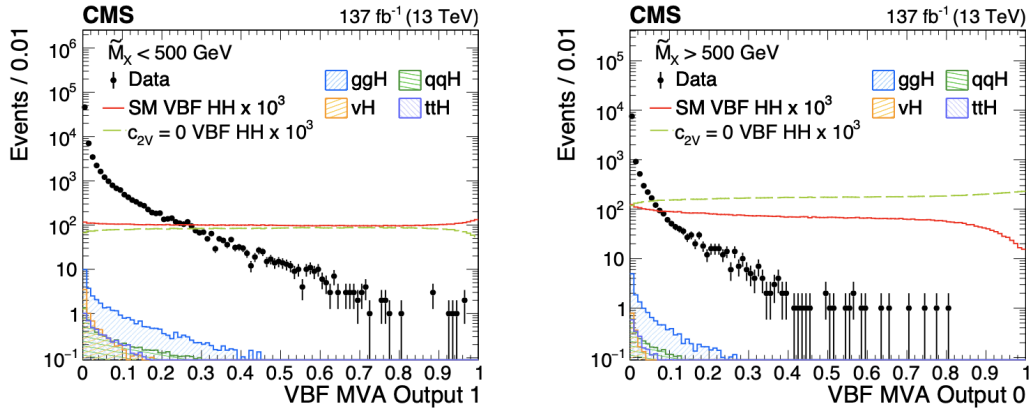


Figure 5.19: MVA multiclassifier output relative to the VBF HH class for the data events as well as for the VBF HH and single Higgs simulated events, for the  $\widetilde{M}_x < 500$  GeV and  $\widetilde{M}_x > 500$  GeV on the left and right, respectively.

## 3019 5.6 Event Categorization

3020 In order to maximize the sensitivity of the search, events are split into different categories  
 3021 according to the output of the MVA classifier and the mass of the Higgs boson pair system  
 3022  $\widetilde{M}_x$ . The  $\widetilde{M}_x$  distribution changes significantly for different BSM hypotheses, as shown in  
 3023 Fig. 5.2. Therefore, a categorization of HH events in  $\widetilde{M}_x$  creates signal regions sensitive  
 3024 to multiple theoretical scenarios. In the search for VBF HH production, the categories in  
 3025  $\widetilde{M}_x$  are defined before the MVA is trained, as described in section 5.5.2. For the categories  
 3026 that target ggF HH production, categories in  $\widetilde{M}_x$  are defined after the MVA is trained.

### 3027 5.6.1 VBF HH-enriched Categories

3028 In the first place, the HH candidates are tested for the VBF HH-enriched categories. The  
 3029 VBF HH categories are given highest priority because the VBF HH production has the  
 3030 smallest cross section among the signals considered, about 15 times smaller than the ggF  
 3031 HH cross section. Therefore, a high selection efficiency on this signal is fundamental.  
 3032 At the same time, a selection on the VBF HH-BDT score mitigates the ggF HH (+jets)  
 3033 events migration to the VBF HH-enriched categories.

3034 The HH candidates passing the additional VBF HH requirements are classified in two  
 3035  $\widetilde{M}_x$  categories to improve the sensitivity both to the SM signal and to the anomalous  
 3036  $c_{2V}$  hypothesis. The VBF HH analysis is optimized to maximize the sensitivity to the  
 3037 SM VBF HH signal and, at the same time, to the VBF HH signal for anomalous  $c_{2V}$   
 3038 values. In particular, the result of the ATLAS experiment for the  $HH \rightarrow b\bar{b}b\bar{b}$  channel  
 3039 [111] indicates that the experimental sensitivity is close to the exclusion of  $c_{2V} = 0$  at  
 3040 95% confidence level. Therefore, the analysis is optimized to achieve the best sensitivity  
 3041 both for  $c_{2V} = 0$  and  $c_{2V} = 1$  (SM). For this reason, the value chosen as boundary of the  
 3042 two categories is  $\widetilde{M}_x = 500$  GeV. As visible in Fig. 5.20, the  $\widetilde{M}_x < 500$  GeV category is  
 3043 especially sensitive to the SM signal, while the  $\widetilde{M}_x > 500$  GeV category is more populated  
 3044 by events produced with  $c_{2V} = 0$ .

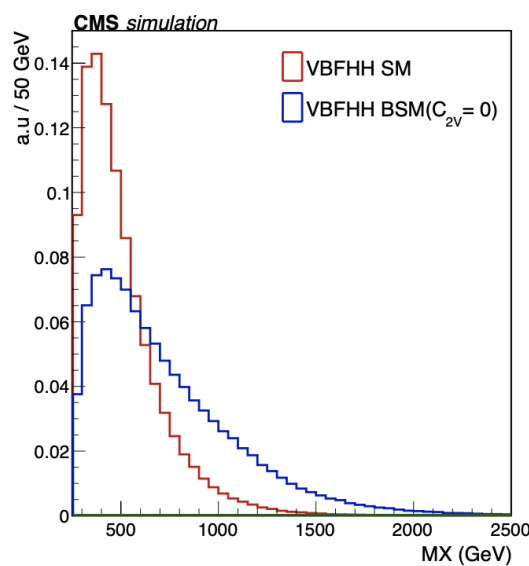


Figure 5.20:  $\widetilde{M}_x$  distribution for the VBF HH simulated events for  $c_{2V} = 1$  and  $c_{2V} = 0$  in red and blue, respectively.



3045 The selections on the  $t\bar{t}H$ -score and on VBF HH-BDT score are simultaneously optimized  
 3046 to maximize the expected significance on the VBF HH signal. The expected significance  
 3047 is estimated as the sum in quadrature over the two categories of  $S/\sqrt{B}$ , where S and B  
 3048 are the expected VBF HH signal and background yields in each category, respectively.  
 3049 Both S and B are estimated using MC events in the  $122 < m_{\gamma\gamma} < 128$  GeV region. The  
 3050 signal considered for the optimization consist in the same mixture of the SM VBF HH  
 3051 sample and the  $c_{2V} = 0$  VBF HH sample used for the BDT training. The signal times  
 3052 the branching ratio is normalized to the expected excludable cross section estimated of  
 3053 0.5 fb. The number of expected events in the sidebands (outside  $115 < m_{\gamma\gamma} < 135$   
 3054 GeV) of each category is required to be higher than 6. This number is found to be the  
 3055 minimum for a data-driven background modeling with a sufficient accuracy (see Section  
 3056 5.7.3). This constraint is found to be the factor mostly controlling the optimization of the  
 3057 VBF HH-BDT score boundaries. The significance is around its maximum for  $t\bar{t}H$ -score  
 3058 values in the range [0.2-0.3]. Therefore, for simplicity the  $t\bar{t}H$ -score  $> 0.26$  selection is  
 3059 chosen, identically to the selection applied for the ggF HH categories. The VBF HH-BDT  
 3060 score boundaries for the categories definition are summarized in Tab. 5.6. The visualized  
 3061 version of categories and constraints is shown in Fig. 5.22 (Left).

## 3062 5.6.2 ggF HH-enriched Categories

3063 The HH candidate events that do not pass the VBF HH category selections are tested for  
 3064 the ggF HH-enriched categories. The ggF HH-BDT score is used to reject the background-  
 3065 like events and to classify the remaining events in three exclusive categories. The bound-  
 3066 aries of the categories, along with the  $t\bar{t}H$ -score selection, are simultaneously optimized  
 3067 with the same procedure used for the VBF HH categories (Section 5.6.1). It was verified  
 3068 that the same  $t\bar{t}H$ -score selection for all the three categories makes the optimization more  
 3069 robust without a significant worsening of the expected significance.

3070 Within each of the three defined BDT score categories, four  $\widetilde{M}_x$  exclusive categories are  
 3071 defined to improve the sensitivity to several BSM scenarios. As discussed in Section  
 3072 2.5.1, the four-body mass is highly sensitive to the BSM benchmarks as well as to the  
 3073 anomalous  $\kappa_\lambda$  hypothesis. The distribution for the SM ggF HH and the main backgrounds  
 3074 MC events is shown in Fig. 5.21. The optimization of the  $3 \times 4 \widetilde{M}_x$  boundaries is performed

3075 simultaneously with the same procedure adopted for the MVA boundaries optimization.

3076 Unlike the VBF HH categories optimization, this optimization is performed with respect to  
 3077 the SM (ggF HH) signal. It was verified that this does not penalize the BSM sensitivity  
 3078 thanks to the dense  $\widetilde{M}_x$  categorization. The number of categories was also optimized  
 3079 repeating the procedure with a different number of BDT-score and  $\widetilde{M}_x$  categories between  
 3080 one and four. The BDT-score and  $\widetilde{M}_x$  selections for the categories is summarized in Tab.  
 3081 5.6. As for the VBF HH categories, the optimization of the boundaries for the high-purity  
 3082 categories is controlled by the constraint on the minimum number of expected background  
 3083 events in the sidebands. The visualized version of categories and constraints is shown in  
 3084 Fig. 5.22 (Right).

3085 The expected composition of the categories, estimated through simulation is shown in  
 3086 Fig. 5.23

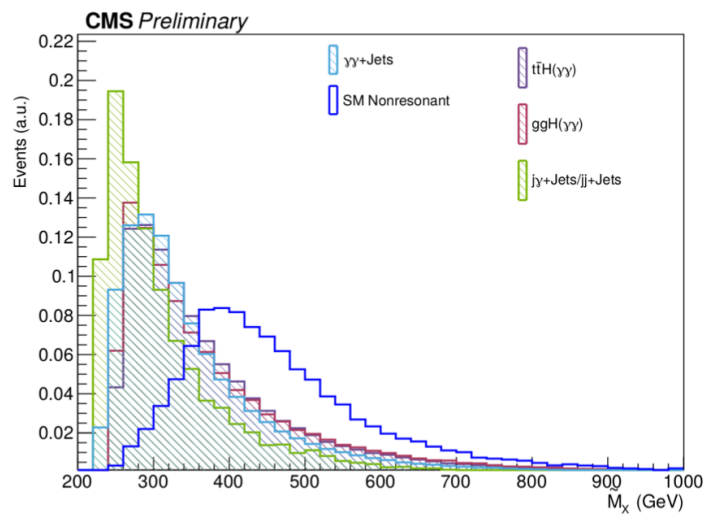


Figure 5.21:  $\widetilde{M}_x$  distribution for the SM ggF HH and the main background MC events. All the distributions are normalized to one.

## 3087 5.7 Statistical Analysis

3088 The data analysis aims to determine the compatibility of the experimental observation  
 3089 with the "signal + background" hypothesis against the "background only" hypothesis, or  
 3090 viceversa. In case a signal is observed, the data are used to measure the corresponding

Category	MVA	$\widetilde{M}_x$ (GeV)
VBF HH CAT 0	0.52-1.00	>500
VBF HH CAT 1	0.86-1.00	250-500
ggF HH CAT 0	0.78-1.0	>600
ggF HH CAT 1		510-600
ggF HH CAT 2		385-510
ggF HH CAT 3		250-385
ggF HH CAT 4	0.62-0.78	>540
ggF HH CAT 5		360-540
ggF HH CAT 6		330-360
ggF HH CAT 7		250-315
ggF HH CAT 8	0.37-0.62	>585
ggF HH CAT 9		375-585
ggF HH CAT 10		330-375
ggF HH CAT 11		250-330

Table 5.6: Optimized BDT-score and  $\widetilde{M}_x$  selections for the HH categories. In all the categories the selection  $t\bar{t}H$ -score  $> 0.26$  is also applied.

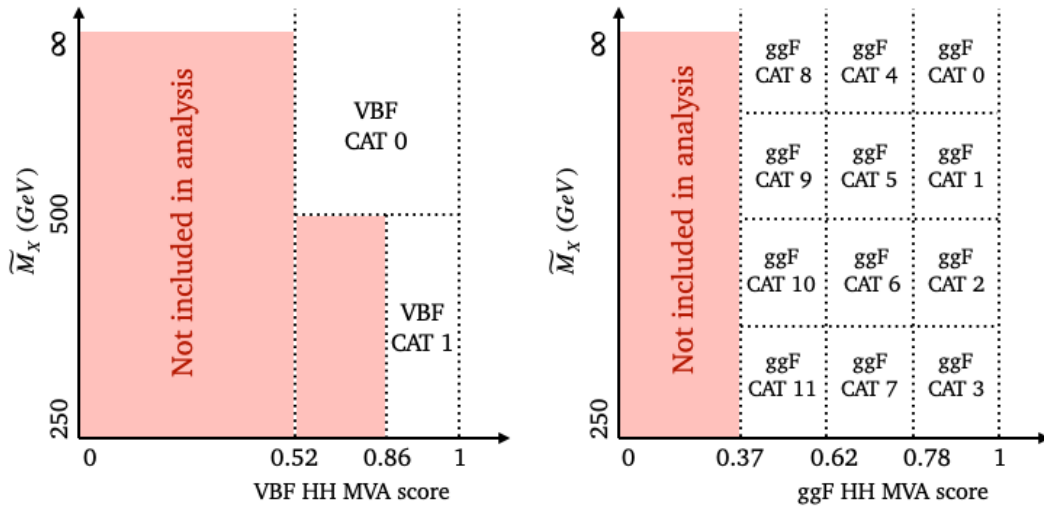


Figure 5.22: Visualized categorization scheme for the VBF HH (left) and ggF HH (right) analysis.

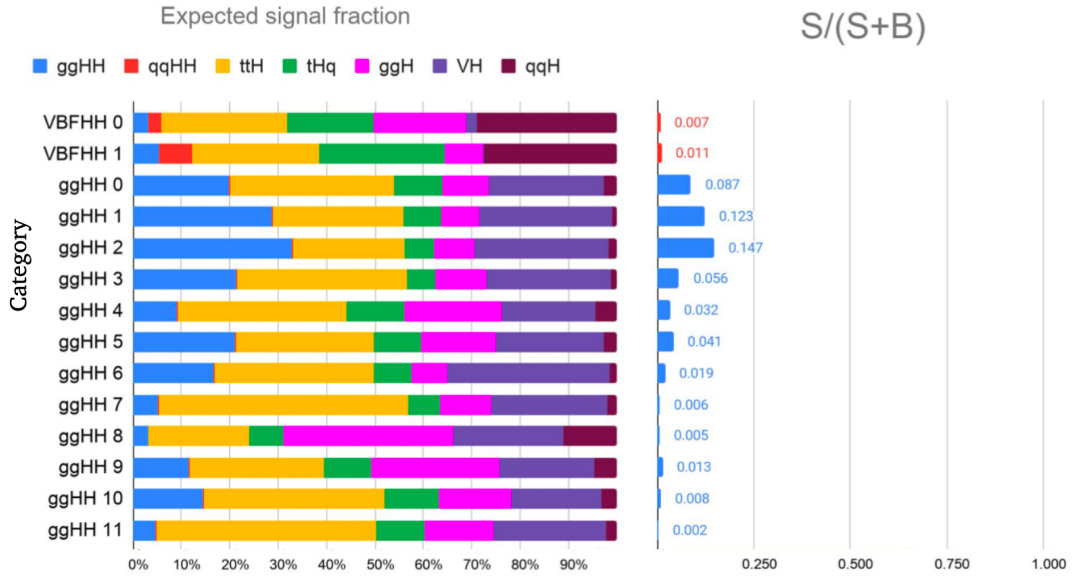


Figure 5.23: *Left*: Expected categories composition in terms Higgs boson processes. *Right*: Expected  $S/(S+B)$  in  $\pm 1\sigma_{eff}$  for each category.  $S$  is referred to the Higgs boson process target of each category, and  $B$  is the sum of the expected Higgs boson background processes and of the expected continuum background.

3091 signal strength  $\mu$  defined as:

$$\mu = \frac{\sigma_{obs}}{\sigma_{SM}} \quad (5.4)$$

3092 Setting  $\mu = 0$  corresponds to the background-only model where as  $\mu = 1$  is the Standard  
 3093 Model expectation. The measured cross section's compatibility with the SM prediction is  
 3094 demonstrated by a  $\mu$  value being consistent, within the uncertainties, with one. In case  
 3095 the signal is not observed, the data are used to set an upper limit on its cross section.  
 3096 In this analysis, multiple signals are tested: the ggF HH signal ( $\mu_{ggF\ HH}$ ), the VBF HH  
 3097 signal ( $\mu_{VBF\ HH}$ ), and the inclusive HH production (ggF HH + VBF HH) signal ( $\mu_{HH}$ ).

3098 Alternatively, the data can be interpreted in terms of Higgs boson coupling modifiers  
 3099 ( $\kappa_\lambda, \kappa_t, c_V, c_{2V}$ ). Due to the limited number of considered Higgs production and decay  
 3100 channels, at most two coupling constants are measured simultaneously, fixing the other  
 3101 couplings to the SM prediction. The measurement can be performed under the assumption  
 3102 of a SM HH signal, or no HH signals. The statistical analysis adopts a maximum likelihood  
 3103 method described in Section 5.7.1 while the signal and background modeling are described  
 3104 in Section 5.7.2 and Section 5.7.3, respectively.

### 5.7.1 Likelihood Definition

A likelihood function is used as test statistics. The likelihood is split into two terms:

$$\mathcal{L} = \mathcal{L}_{HH} \cdot p(\theta|\tilde{\theta}) \quad (5.5)$$

where  $\mathcal{L}_{HH}$  is the likelihood functions corresponding to the HH enriched categories. Defined  $\theta$  as the vector of all the nuisance parameters,  $p(\theta|\tilde{\theta})$  is the distribution of  $\theta$  given the true values (Bayesian interpretation)  $\tilde{\theta}$ . For the nuisance description, the frequentist approach common to the CMS and ATLAS experiment is used [112].

The  $\mathcal{L}_{HH}$  factor is built exploiting the resonant nature of the HH signal in the  $m_{jj}$  and  $m_{\gamma\gamma}$  distributions. Given the low statistic regime, the  $\mathcal{L}_{HH}$  function is built as an unbinned likelihood. Therefore,  $\mathcal{L}_{HH}$  is defined for each analysis category as:

$$\begin{aligned} \mathcal{L} = & k^{-1} \prod_{i \in \text{events}} \left[ \sum_{\substack{j=ggF, HH, \\ VBF, HH}} \mu_j S_j f_j(m_{\gamma\gamma}^i, m_{jj}^i) + \sum_{\substack{j=t\bar{t}H, tHq \\ ggH, VH, \\ qqH}} S_j f_j(m_{\gamma\gamma}^i, m_{jj}^i) \right. \\ & \left. + B f_B(m_{\gamma\gamma}^i, m_{jj}^i|\theta) \right] \cdot \exp\left( \sum_{\substack{j=ggF, HH, \\ VBF, HH}} \mu_j S_j + \sum_{\substack{j=t\bar{t}H, tHq \\ ggH, VH, \\ qqH}} S_j + B \right) \end{aligned} \quad (5.6)$$

where  $k$  is the total number of observed events,  $S_j$  is the number of events predicted by the SM for the  $j$ -th Higgs process, and  $f_j$  is the corresponding two dimensional ( $m_{\gamma\gamma}$ ,  $m_{jj}$ ) parametric pdf.  $B$  is the expected number of continuum background events and  $f_B$  is the corresponding parametric pdf. For simplicity, the dependence of the  $S_j$ ,  $f_j$ , and  $f_B$  quantities from the nuisance parameters  $\theta$  is omitted from Eq. 5.6. The correlation between the  $m_{\gamma\gamma}$  and  $m_{jj}$  variables is found to be negligible both for the signals and the backgrounds. Therefore, the two dimensional models can be factorized as:

$$f_j(m_{\gamma\gamma}, m_{jj}) = f_j^{\gamma\gamma}(m_{\gamma\gamma}) f_j^{jj}(m_{jj}) \quad (5.7)$$

where  $f_j^{\gamma\gamma}$  and  $f_j^{jj}$  are the one dimensional  $m_{\gamma\gamma}$  and  $m_{jj}$  models for the  $j$ -th Higgs process.

3122 For the measurement of the coupling modifiers, the signal strengths are fixed to one while  
 3123 the number of expected events for the single and double Higgs processes  $S_j$  is expressed  
 3124 as a function of the  $(\kappa_\lambda, \kappa_t, c_V, c_{2V})$  parameters. In particular, the SM and BSM ggF HH  
 3125 and VBF HH samples are properly combined to provide a per-category description of the  
 3126 signals rate variations. For the single Higgs processes a parametric description of the total  
 3127 cross section variation as a function of the coupling modifiers is used for all the categories.  
 3128 Although modifications of the  $p_T$  spectrum of the single Higgs processes, especially for  
 3129  $t\bar{t}H$ , are expected in case of anomalous  $\kappa_\lambda$  values, an explicit  $p_T$  categorization is not  
 3130 performed. Since the  $p_T$  distribution is found to be similar in all the categories, the  
 3131 inclusive cross section variation provides sufficient accuracy for the description of the  
 3132 anomalous coupling effects. A future extension of this work can be a  $p_T$  classification of  
 3133 the single Higgs events to improve the sensitivity to anomalous couplings. The  $H \rightarrow \gamma\gamma$   
 3134 and  $H \rightarrow b\bar{b}$  branching ratios variations for anomalous couplings are also considered.

### 3135 Estimation of the Parameters of Interest

3136 Let a generic  $\mu$  be the parameter of interest. The likelihood estimators of  $\mu$  and  $\theta$ , namely  
 3137  $\hat{\mu}$  and  $\hat{\theta}$ , are the values simultaneously maximizing the likelihood function, or equivalently  
 3138 the values minimizing the negative logarithm of the likelihood function (log-likelihood  $L$ ).  
 3139 The latter is much easier to compute from the algorithmic point of view. In order to  
 3140 estimate the uncertainty on  $\hat{\mu}$ , a "profile likelihood" is defined as:

$$\mathcal{L}_{prof}(\mu) = -2 \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (5.8)$$

3141 Where  $\mathcal{L}$  is the likelihood defined in Eq. 5.5,  $\hat{\hat{\theta}}$  represents the set of nuisance values  
 3142 maximizing  $\mathcal{L}$  for a given  $\mu$  and the denominator is maximized over the full parameter  
 3143 space. The 68% and 95% confidence intervals on  $\mu$  correspond to the  $\mu$  values satisfying  
 3144 the condition  $\mathcal{L}_{prof}(\mu) < 1$  and  $\mathcal{L}_{prof}(\mu) < 3.84$ , respectively.

### 3145 Hypothesis Testing to Set an Upper Limit

3146 In case no evidence of the double Higgs production is found with data, an upper limit can  
 3147 be set on the ggF HH, VBF HH, and HH cross sections. To do that, the following test  
 3148 statistic is defined:

$$q_\mu = -\log \frac{\mathcal{L}(\mu, \hat{\theta})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \text{ with } 0 \leq \hat{\mu} \leq \mu \quad (5.9)$$

3149 where  $\mathcal{L}$  is the likelihood,  $\hat{\mu}$  and  $\hat{\theta}$  are the signal strength and the nuisances values maxi-  
 3150 mizing  $\mathcal{L}$ , while  $\hat{\theta}$  is the set of nuisance values maximizing  $\mathcal{L}$  for a given value of  $\mu$ .

3151 Depending on the cross section of interest, the  $\mu$  variable is referred either to the ggF HH,  
 3152 or VBF HH, or the inclusive HH signal. The constraint  $\hat{\mu} \leq 0$  is to avoid the unphysical  
 3153 situation of negative signals, while the constraint  $\mu \leq \hat{\mu}$  is required to avoid to use upward  
 3154 fluctuations of the data against the signal hypothesis. The modified frequentist criterion  
 3155 [113] is adopted for the limit extraction. The level of disagreement of the observed data  
 3156 with a given hypothesis is quantified through a "p-value" which is the probability to  
 3157 obtain results worse than or equal to the one observed under the given hypothesis.

3158 In particular, given the observed value of the test statistic  $q_\mu^{obs}$ , two p-values  $p_\mu$  and  $p_b$  can  
 3159 be derived for the signal plus background and background only hypotheses respectively:

$$p_\mu = Prob(q_\mu > q_\mu^{obs} | signal + background) \quad 1 - p_b = Prob(q_\mu > q_\mu^{obs} | background only) \quad (5.10)$$

3160 Such p-values are computed using the asymptotic properties of the test statistics [114].  
 3161 This avoids the computationally expensive procedure of the MC toy generation to explic-  
 3162 itly derive the  $q_\mu$  distributions. The  $p_\mu$  value is not used directly for the limit extraction  
 3163 because it is not sufficiently robust against background under-fluctuations. Such under-  
 3164 fluctuations could lead to exclude small values of  $\mu$  even if the sensitivity to the signal  
 3165 would not be sufficient. In order to prevent for that effect, the  $CL_s(\mu)$  quantity is defined  
 3166 as:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \quad (5.11)$$

3167 A signal strength  $\mu$  is said to be excluded at a confidence level (CL)  $\alpha$  if  $CL_s(\mu) < 1 - \alpha$ .  
 3168 The value commonly chosen for  $\alpha$  is 95%. The limit on  $\mu$  is in fact a limit on the cross

3169 section normalized to the corresponding SM prediction.

### 3170 Hypothesis Testing to Quantify an Excess

3171 The following test statistics is defined to quantify the excess of events:

$$q_0 = -\log \frac{\mathcal{L}(\mu, \hat{\theta})|_{\mu=0}}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \text{ with } \hat{\mu} > 0 \quad (5.12)$$

3172 This is in fact the likelihood computed for the background-only hypothesis normalized by  
 3173 the likelihood value for the best fit point. Given the observed value of the test statistic  
 3174  $q_0^{obs}$ , the p-value is then defined as:

$$p_0 = Prob(q_0 \geq q_0^{obs} | backgroundonly) \quad (5.13)$$

3175 This is the probability that a background fluctuation gives an excess larger than or equal  
 3176 to the observed one. For a more direct interpretation, the  $p_0$  probability is converted to  
 3177 a  $Z$  significance by expressing it as a one-sided Gaussian integral:

$$p_0 = \int_Z^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (5.14)$$

3178 It is customary to consider a significance larger than  $3\sigma$  as an evidence, and a significance  
 3179 larger than  $5\sigma$  as an observation of a signal above the background.

### 3180 5.7.2 Modeling of HH Processes

3181 For each analysis category and each Higgs production mechanism, the MC simulations  
 3182 are used to derive the expected number of events ( $S_j$ ) and to model the  $m_{\gamma\gamma}$  distribution  
 3183 ( $f_j^{\gamma\gamma}$ ), and, the  $m_{jj}$  distribution ( $f_j^{jj}$ ).

3184 The  $m_{\gamma\gamma}$  peak is modeled as a sum of up to five gaussian functions. Examples of the  $m_{\gamma\gamma}$   
 3185 models for the HH processes are visible in Fig. 5.24.



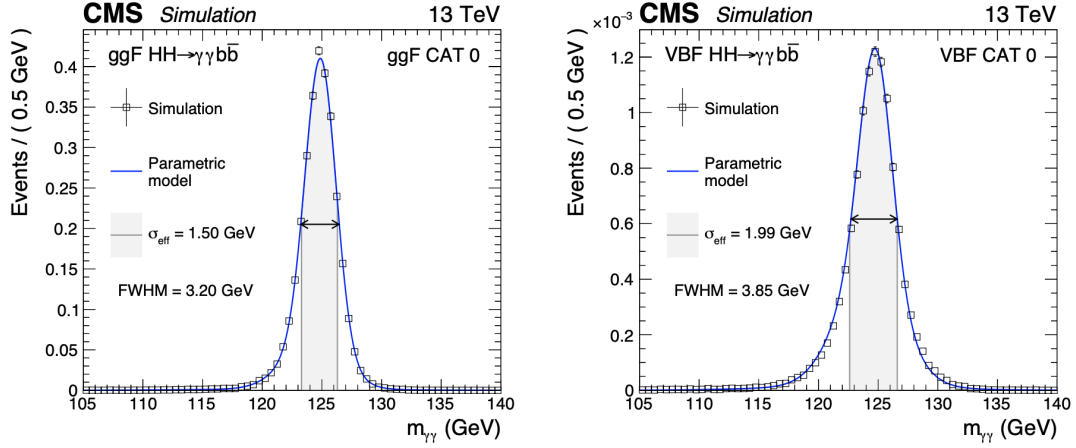


Figure 5.24:  $m_{\gamma\gamma}$  modeling for the ggF HH and VBF HH process on the left and right, respectively, for the best resolution (high  $\widetilde{M}_x$  and high BDT score) category in the 2018 dataset. The open squares represent simulated events and the blue lines are the corresponding models. Also shown are the  $\sigma_{eff}$  value (half the width of the narrowest interval containing 68.3% of the invariant mass distribution) and the corresponding interval as a gray band, and the full width at half the maximum (FWHM) and the corresponding interval as a double arrow.

3186 The  $m_{jj}$  distribution for the VBF HH and ggF HH processes is modeled with a double-  
 3187 sided Crystal Ball (CB) function which is a CB function with two independent exponential  
 3188 tails instead of one, as shown in Fig. 5.25. This function is found to provide an adequate  
 3189 description of the  $m_{jj}$  peak with its left and right tails related to the jet energy resolution.

3190 The functions parameters are determined for each category through a fit to the selected  
 3191 simulated events. The final signal model is built as the product of the obtained  $m_{\gamma\gamma}$  and  
 3192  $m_{jj}$  distributions.

3193 The assumption of no-correlation between the  $m_{\gamma\gamma}$  and  $m_{jj}$  variables hypothesis is verified  
 3194 comparing the two dimensional  $(m_{\gamma\gamma}, m_{jj})$  distribution of the simulated events, visible in  
 3195 Fig. 5.26 for the simulated ggF HH events from the 2018 dataset, to the product of  
 3196 the derived  $m_{jj}$  and  $m_{\gamma\gamma}$  models. The difference of the two distributions is found to be  
 3197 negligible within the statistical uncertainties related to the expected number of signal  
 3198 events.

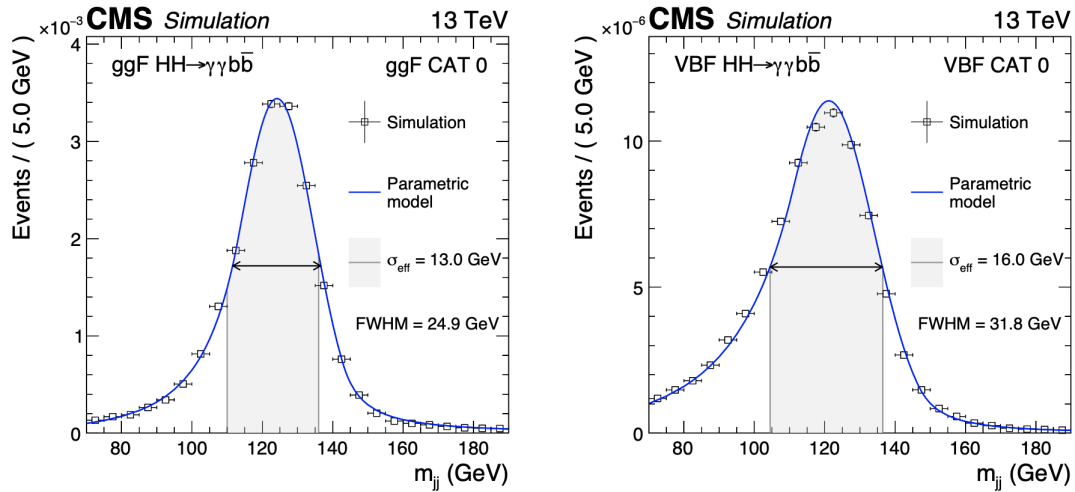


Figure 5.25:  $m_{jj}$  modeling for the ggF HH and VBF HH process on the left and right, respectively, for the best resolution (high  $\widetilde{M}_x$  and high BDT score) category in the 2018 dataset.

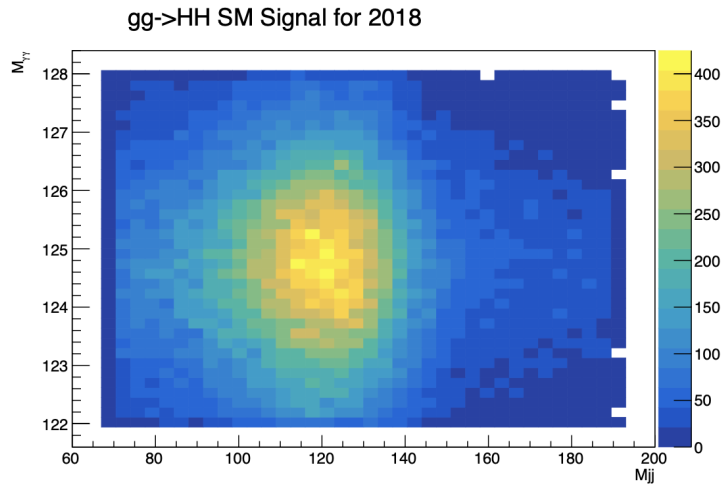


Figure 5.26: Two dimensional  $(m_{jj}, m_{\gamma\gamma})$  distribution for the selected HH candidates of the simulated ggF HH events using the 2018 dataset.

### 3199 5.7.3 Background Modeling

#### 3200 Single Higgs Background

3201 For the Single Higgs categories, the  $m_{jj}$  pdf parametrization depends on the specific  
3202 production mechanism:

- 3203 • The  $m_{jj}$  distribution for the VH process, consisting in a peak in correspondence  
3204 of the vector bosons masses, is modeled through a standard CB function. This  
3205 function, with a lower number of free parameter than a double-sided CB, provides

the required robustness and accuracy for the VH modeling. The simulated VH events are in fact affected by higher statistical uncertainties because of the limited number of selected events.

- The  $m_{jj}$  distributions for the ggH and qqH processes have a smooth falling shape, thus they are parametrized by Bernstein polynomials.
- The  $t\bar{t}H$  and tHq events kinematics feature a  $m_{jj}$  distribution peaking at about 110-120 GeV. Thus, a gaussian function is used to model their  $m_{jj}$  distributions.

The number of simulated events for the single-Higgs production mechanisms in the HH signal regions is limited, especially for the ggH, VH, and qqH processes. Therefore, the simulated events of the three data-taking years within the same BDT-score category are merged together, and a common  $m_{jj}$  model is extracted to improve the model accuracy. It was verified that the  $m_{jj}$  models across different  $\widetilde{M}_x$  categories and different years are compatible within the uncertainties. Examples of the  $m_{jj}$  models for the high BDT score category are visible in Fig. 5.27. In order to improve the accuracy of the  $m_{\gamma\gamma}$  modeling of the ggH, VH, and qqH processes in the HH signal regions, in case the number of MC entries is less than 500, the used  $m_{\gamma\gamma}$  model is the same one used for the  $t\bar{t}H$  process in that category. Then the normalization is set according to the expected yield for that specific process.

## Continuum Background

The continuum background is modeled in each category through a fit of the  $m_{\gamma\gamma}$  (and  $m_{jj}$ ) distribution of the selected data events. Since the parametrization of the underlying model is not known, a specific method called "envelope method" [115] is used to choose the background parametrization estimating also the related uncertainty.

The envelope method considers the choice of the background functional form as an additional (discrete) nuisance parameter to be included in the likelihood definition. In particular, an integer number called "envelope index" is used to select a specific functional form among the set of given parametrizations. Therefore, the negative log-likelihood is minimized with respect to all the nuisance parameters including the envelope index. The

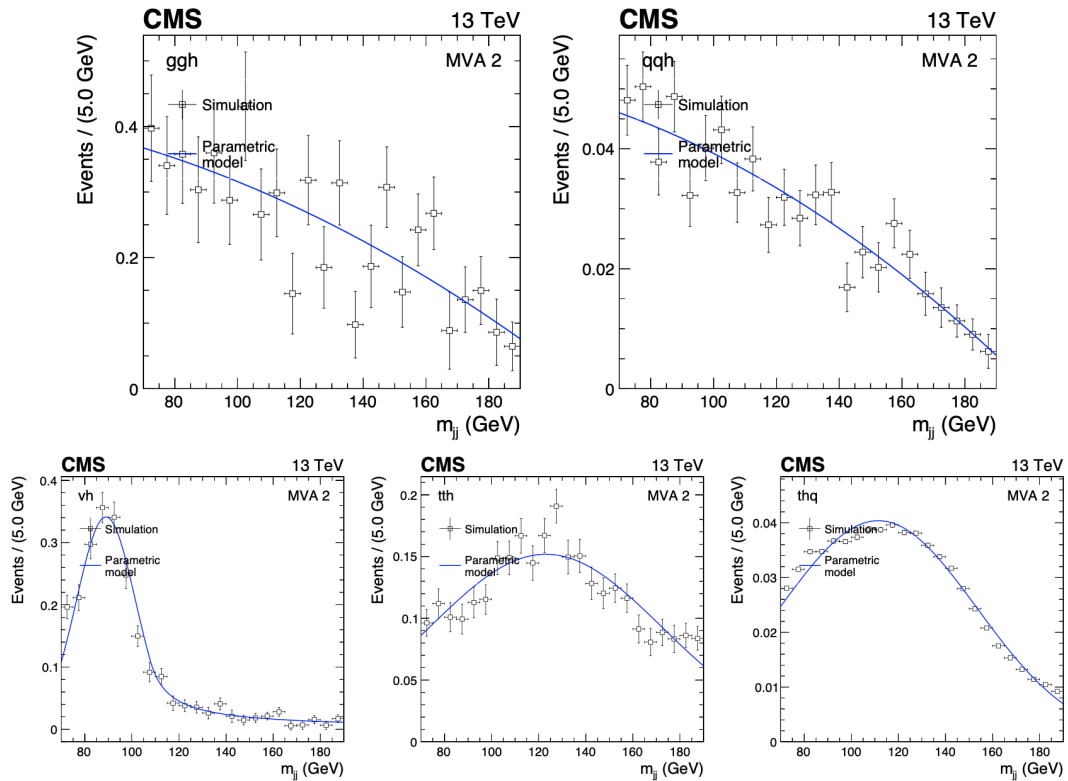


Figure 5.27:  $m_{jj}$  modeling for the ggH (top-left), qqH (top-right), VH (bottom-left),  $t\bar{t}H$  (bottom-center), and tHq (bottom-right) processes, in the ggF HH-enriched category with the highest BDT score value.

3234 impact of this additional nuisance parameter is a broadening of the profile likelihood of  
 3235 Eq. 5.8 corresponding to an increase of the uncertainty in the parameter estimation.  
 3236 This is expected because a nuisance parameter corresponds to a loss of information in the  
 3237 measurement.

3238 For a correct usage of the envelope method, a proper choice of the set of functions is crucial.  
 3239 The set of functions has to provide a negligible bias in the estimation of the parameter  
 3240 of interest and a consistent estimation of the corresponding uncertainty (coverage). The  
 3241 studies presented in Ref. [115] show that for an exponentially falling background such as  
 3242 the  $m_{\gamma\gamma}$  background distribution in the  $H \rightarrow \gamma\gamma$  analysis, a good coverage is provided by  
 3243 the following set of function families:

Power law sum:  $f(x) = p_0x^{p_1} + p_2x^{p_3} + p_4x^{p_5} + \dots$

Exponential sum:  $f(x) = p_0e^{p_1x} + p_2e^{p_3x} + p_4e^{p_5x} + \dots$

Laurent series:  $f(x) = \sum_{i=0}^N p_i/x^i$

Polynomial:  $f(x) = \sum_{i=0}^N p_ix^i$

3244 A second important aspect is how to compare functions with different degrees of freedom  
 3245 in terms of data agreement. By construction, a function with a higher number of free  
 3246 parameters within the same family is able to better describe the data, thus the likelihood  
 3247 minimization will select it. However, it is more sensitive to the data fluctuations. There-  
 3248 fore, a correction to the likelihood penalizing the higher order functions is defined to make  
 3249 the method robust against the background fluctuations. In Ref. [115], a good correction  
 3250 for the  $H \rightarrow \gamma\gamma$  case is found to be:

$$L_{corr} = L + N_{par} \tag{5.15}$$

3251 where  $L$  is the negative log-likelihood and  $N_{par}$  is the number of free parameters of the  
 3252 function considered.

3253 For the HH categories, the envelope method is used for the description both of the  $m_{\gamma\gamma}$   
 3254 and the  $m_{jj}$  backgrounds. In the assumption of no-correlations between the  $m_{\gamma\gamma}$  and  
 3255  $m_{jj}$  variables, the generalization of the envelope method for the 2D fit is straightforward.  
 3256 The same function families defined for the  $m_{\gamma\gamma}$  description can be used also for the  $m_{jj}$   
 3257 variable because the background shape is analogous. The correction defined in Eq. 5.15  
 3258 accounts for the sum of free parameters of the  $m_{jj}$  and  $m_{\gamma\gamma}$  distributions.

3259 From the practical point of view, the numerical minimization with discrete parameters  
 3260 is not reliable. Therefore, for each 2D combination of  $(m_{\gamma\gamma}, m_{jj})$  envelope indexes, a  
 3261 minimization is performed and the envelope is built afterwards. In order to reduce the  
 3262 computing time required for the minimization, a preliminary procedure determines the  
 3263 optimal order of each function family in each category. All the functions with up to six

3264 free parameters are considered. The functions with order higher than the optimal one are  
 3265 then not considered for the envelope construction.

3266 The range for the  $m_{\gamma\gamma}$  and  $m_{jj}$  fits are  $100 \text{ GeV} < m_{\gamma\gamma} < 180 \text{ GeV}$ , and  $70 \text{ GeV} < m_{\gamma\gamma} <$   
 3267  $190 \text{ GeV}$ . However, for the two ggF HH enriched categories with the lowest  $\widetilde{M}_x$  at low BDT  
 3268 score, the fit region is reduced to  $90 \text{ GeV} < m_{\gamma\gamma} < 190 \text{ GeV}$  because the background shape  
 3269 is not well described. It was verified that with the reduced  $m_{jj}$  range the background  
 3270 modeling remains robust and that the bias induced on the expected signal strength is  
 3271 negligible. The changes of the expected upper limits on the HH cross section and of the  
 3272 constraints on anomalous couplings are found to be below 1%.

## 3273 5.8 Systematic Uncertainties

3274 The analysis is statistically limited, therefore the impact of the systematics uncertainties  
 3275 on the result is small. In particular, as we will see, the upper limit on  $\mu_{HH}$  including  
 3276 the systematics uncertainties is only 2% worse than the limit computed considering only  
 3277 the statistical uncertainties. The impact of the systematics uncertainties consists in a  
 3278 modification of the single and double Higgs yields in the categories, due to the event  
 3279 loss or migration across categories. The only systematic uncertainty associated with the  
 3280 continuum background, estimated from data, is the choice of its parametric modeling  
 3281 within the envelope method. The impact of the systematics uncertainties, quoted for  
 3282 simplicity on  $\mu_{HH}$ , is shown in Tab. 5.7, while their description is listed below:

- 3283 • QCD scale: it accounts for the uncertainty on the renormalization and factorization  
 3284 scale. They are set accordingly to Ref. [29], [32], and [116] for the single Higgs, ggF  
 3285 HH, and VBF HH processes, respectively. In particular, the uncertainties associated  
 3286 with the ggF HH and  $t\bar{t}H$  QCD scales, whose  $1\sigma$  variations change the total cross  
 3287 sections of up to 5 and 9%, respectively, represent the dominant uncertainties for  
 3288 this analysis.
- 3289 • Parton distribution functions (PDF) modeling: it is computed according to the  
 3290 PDF4LHC15 prescriptions [117]. The PDF modeling affects the total number of  
 3291 events and also the event categorization since it modifies the number of jets produced

Systematic uncertainty	$\Delta\mu_{HH}$ (%)
QCD scale	+7/-2
PDF modeling + $\alpha_s$	3
Branching ratio	3
Parton shower modeling	<1
Luminosity	3
Photon preselection	2
Per-photon $\sigma_E/E$	1
Photon ID	<1
Trigger	<1
Photon energy & res.	<1
Jet energy & res.	<1
b-tag efficiency	<1
Pileup jet ID	<1

Table 5.7: Impact of the systematic uncertainties on the signal strengths in percentage.

in association with the Higgs boson signals. The event migration is computed using the NNPDF30 set with the MC2hessian method [118].

- $\alpha_s$  value: the uncertainty on the QCD coupling constant is computed along with the PDF modeling uncertainty using the PDF4LHC15 prescriptions. The  $\alpha_s$  value affects in fact also the PDF modeling, thus the impact of the two uncertainties is estimated together.
- Parton shower modeling: this uncertainty is considered for the VBF HH process because different showering schemes can significantly change some of the VBF observables [119], hence the total number of events and their classification. This uncertainty is conservatively estimated as the full symmetrized difference in yields in each category obtained from VBF HH MC samples generated with different parton shower ISR and FSR configurations.
- Uncertainty in the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow b\bar{b}$  branching fractions which amounts to about 3 and 0.5%, respectively, according to Ref. [29].

The dominant experimental systematic uncertainties are:

- Photon identification BDT score: it accounts for the residual data-simulation discrepancy of the photon ID BDT score distribution. Such a discrepancy is ascribed to the limited accuracy of the regression used to correct the BDT input variables

3310 to cover the residual discrepancies between data and simulation. It is estimated  
3311 through the procedure described in Section 5.4.1.

3312 • Photon energy scale and resolution: it accounts for the residual discrepancy between  
3313 the data and MC simulation after the corrections. It accounts for effects such as  
3314 non-linearities of the light collection, the different shower-shape of electrons, used  
3315 to derive the correction, and photons, as well as different trainings in the energy  
3316 regression and variation of the binning used to derive the correction.

3317 • Per-photon energy resolution estimate: this variable is computed by the photon  
3318 energy regression. Its impact on the event selection and classification is estimated  
3319 by varying the resolution of  $\pm 5\%$  around its nominal value.

3320 • Jet energy scale and resolution corrections: the energy scale and resolution of jets  
3321 is measured using the  $p_T$  balance in  $Z(ee)+\text{jets}$ ,  $Z(\mu\mu)+\text{jets}$ ,  $\gamma+\text{jets}$ , and multijet  
3322 events [110]. The uncertainty on the calibration is a few percent and depends on  
3323  $p_T$  and  $\eta$ . The impact of jet energy scale uncertainties in event yields is estimated  
3324 by varying the jet energy corrections within their uncertainties, ranging between  
3325 1 and 3% in central barrel, and propagating the effect to the final result. Some  
3326 sources of the jet energy scale uncertainty are fully (anti-)correlated, while others  
3327 are considered uncorrelated.

3328 • Jet b-tagging: such uncertainties are computed comparing the distribution of the b  
3329 tagging efficiency between data and simulation. The efficiency on light flavour jets  
3330 is measured using an inclusive multijet sample, while the efficiency on heavy-flavour  
3331 jets is measured using muon-enriched jet samples, and  $t\bar{t}$  plus one or two leptons  
3332 samples [91]. The uncertainties include the statistical component on the estimate  
3333 of the fraction of heavy and light flavour jets in data and simulation.

3334 • Trigger efficiency: as discussed in Section 5.3.1, the efficiency of the trigger selection  
3335 and the corresponding uncertainty is measured with  $Z \rightarrow e^+e^-$  events using a T&P  
3336 technique. An additional uncertainty is introduced to account for a gradual shift  
3337 in the timing of the inputs of the ECAL L1 trigger in the region  $|\eta| > 2$ , causing a  
3338 specific trigger inefficiency during the 2016 and 2017 data taking periods. Photons  
3339 and also jets are affected by this inefficiency, which has a small impact.



- 3340 • Photon preselection: as discussed in Section 5.4.1, the photon preselection efficiency  
3341 (including the electron veto efficiency) is estimated using  $Z \rightarrow e^+e^-$  and  $Z \rightarrow$   
3342  $\mu^+\mu^-\gamma$  events with a T&P method. The uncertainty on the scale factors derived to  
3343 match the efficiency of the simulation to the one measured with data is propagated  
3344 throughout the analysis.
- 3345 • Integrated luminosity: the related uncertainties are determined through auxiliary  
3346 measurements by the CMS luminosity monitoring for the 2016–2018 data-taking  
3347 years [120, 121, 122]. The uncertainties across the different years of data-taking  
3348 are partially correlated to account for common sources of uncertainty in the lumi-  
3349 nosity measurement schemes. The total 2016–2018 integrated luminosity has an  
3350 uncertainty of 1.8%.
- 3351 • Pileup jet ID output score: it accounts for the differences between data and simula-  
3352 tion in the distribution of the pileup jet ID variable. Only the VBF jets are affected.  
3353 This uncertainty is estimated by comparing the score of jets in Z+jets events in data  
3354 and simulation in intervals of  $p_T$  and  $\eta$ .

3355 Other systematics uncertainties impact the signal strength by less than 1% and are thus  
3356 negligible with regard to the ones described above. They include uncertainties on lepton  
3357 identification and isolation efficiencies, on the correct vertex assignment efficiency, and on  
3358 the missing transverse momentum.

## 3359 5.9 Results

3360 This work led to two results: the search for the double Higgs production presented in  
3361 Section 5.9.1, and the measurement of the Higgs boson couplings presented in Section  
3362 5.9.3. All the results were found compatible with the SM predictions. In particular, no  
3363 significant excesses over the background of double Higgs production events were found,  
3364 thus upper limits on the HH cross sections were extracted.

3365 A simultaneous fit of the HH and the  $t\bar{t}H$  cross sections is also performed to improve  
3366 the sensitivity on the  $\lambda)HHH$  and the  $y_t$  parameters, and to simultaneously measure the

3367 two parameters. The  $t\bar{t}H$  and HH processes are intrinsically correlated because they both  
 3368 depend on the  $\lambda)HHH$  and  $y_t$  constants.

### 3369 5.9.1 Search for HH Process

3370 The HH categories (twelve ggF HH and two VBF HH categories) are included in the  
 3371 extraction of the HH production yield. A likelihood fit is performed using the likelihood  
 3372 defined in Eq. 3.5 to measure the  $\mu_{HH}$  signal modifier. The fit to the  $m_{\gamma\gamma}$  and  $m_{jj}$  distri-  
 3373 butions for two categories of VBF HH is shown in Fig. 5.28 and for twelve categories of  
 3374 ggF HH is shown in Fig. 5.29 and 5.30 respectively. The weighted distribution of events  
 3375 while merging all categories according to the factor  $S/(S+B)$  is shown in Fig. 5.31. Anal-  
 3376 ogous fits are performed to measure either the  $\mu_{ggF\ HH}$  or the  $\mu_{VBF\ HH}$  parameter fixing  
 3377 the other parameter to one. The observed and expected signal strengths are reported in  
 3378 Tab. 5.8.

Parameter	Expected	Observed
$\mu_{HH}$	$1.0^{+2.7}_{-1.9}$	$2.7^{+2.6}_{-2.0}$
$\mu_{ggF\ HH}$ fixing $\mu_{VBF\ HH} = 1$	$1.0^{+2.7}_{-1.9}$	$2.8^{+2.7}_{-2.0}$
$\mu_{VBF\ HH}$ fixing $\mu_{ggF\ HH} = 1$	$1.0^{+91.3}_{-65.1}$	$10.2^{+97.21}_{-61.6}$

Table 5.8: Expected and observed signal strength for inclusive HH(ggF HH + VBF HH), ggF HH and VBF HH

3379 Due to a small excess observed in data, the observed signal strengths are larger than  
 3380 one, but still compatible with the SM within the uncertainties. The inclusive  $\mu_{HH}$  signal  
 3381 strength is dominated by the ggF HH process because of the larger cross section, as  
 3382 visible by the large uncertainty on  $\mu_{VBF\ HH}$  compared to the uncertainty on  $\mu_{ggF\ HH}$ .  
 3383 Alternatively,  $\mu_{ggF\ HH}$  and  $\mu_{VBF\ HH}$  are measured simultaneously, as shown in Fig. 5.32.

3384 Since no evidences of double Higgs production were found, upper limits on the corre-  
 3385 sponding cross sections are extracted using the procedure described in Section 5.7.1. The  
 3386 observed and expected upper limits are presented in Tab. 5.9.

3387 The upper limit on  $\sigma_{ggF\ HH}$  of  $7.8\times\text{SM}$  is the best result from the CMS experiment to  
 3388 date, and it is comparable to the constraint set by the ATLAS experiment combining the

	Upper limit at 95% C.L.		
	Expected	Observed	Best published result (Observed)
$\sigma_{HH}$	$5.2\times\text{SM}$	$7.7\times\text{SM}$	-
$\sigma_{ggF HH}$ fixing $\mu_{VBF HH} = 1$	$5.3\times\text{SM}$	$7.8\times\text{SM}$	$6.9\times\text{SM}$ (HH comb. with $36 \text{ fb}^{-1}$ [123])
$\sigma_{VBF HH}$ fixing $\mu_{ggF HH} = 1$	$208\times\text{SM}$	$225\times\text{SM}$	$840\times\text{SM}$ ( $HH \rightarrow b\bar{b}b\bar{b}$ with $126 \text{ fb}^{-1}$ [111])

Table 5.9: The observed and expected upper limits.

3389 most sensitive HH decay channels with the 2016 dataset. This is also the first upper limit  
3390 on  $\sigma_{VBF HH}$  set by the CMS experiment, which improves the best constraint set by the  
3391 ATLAS experiment.

## 3392 5.9.2 Constraints on the BSM Benchmark Hypotheses

3393 Upper limits are also set on the twelve BSM benchmark hypotheses. The ggF HH events  
3394 simulated at the LO for each benchmark hypothesis are used to derive the expected  
3395 number of events as well as the signal model. Only the ggF HH categories are considered,  
3396 and the VBF HH process is neglected in the statistical interpretation of the data. The  
3397 observed and expected upper limits at 95% C.L. on the BSM benchmark hypotheses are  
3398 shown in Fig. 5.33. The observed limits are slightly larger than the expected limits  
3399 because of the small excess of events found in the ggF HH-enriched categories. The  
3400 expected limits span from about 0.1 to 1 fb. The different sensitivity is due to the  
3401 kinematics of each specific BSM benchmark hypothesis, which change the population of  
3402 the expected signal events in the four-body mass categories.

## 3403 5.9.3 Constraints on the Higgs Boson Couplings

3404 Since no evidence of the HH signal are found, the 95% C.L. upper limits on  $\sigma_{ggF HH} \times$   
3405  $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$  are derived as a function of the  $\kappa_\lambda$  parameter, as shown in the left  
3406 panel of Fig. 5.34. The upper limit dependence on  $\kappa_\lambda$  is determined by the variation  
3407 of the  $\widetilde{M}_x$  distribution of the ggF HH signal that modifies the categories population,

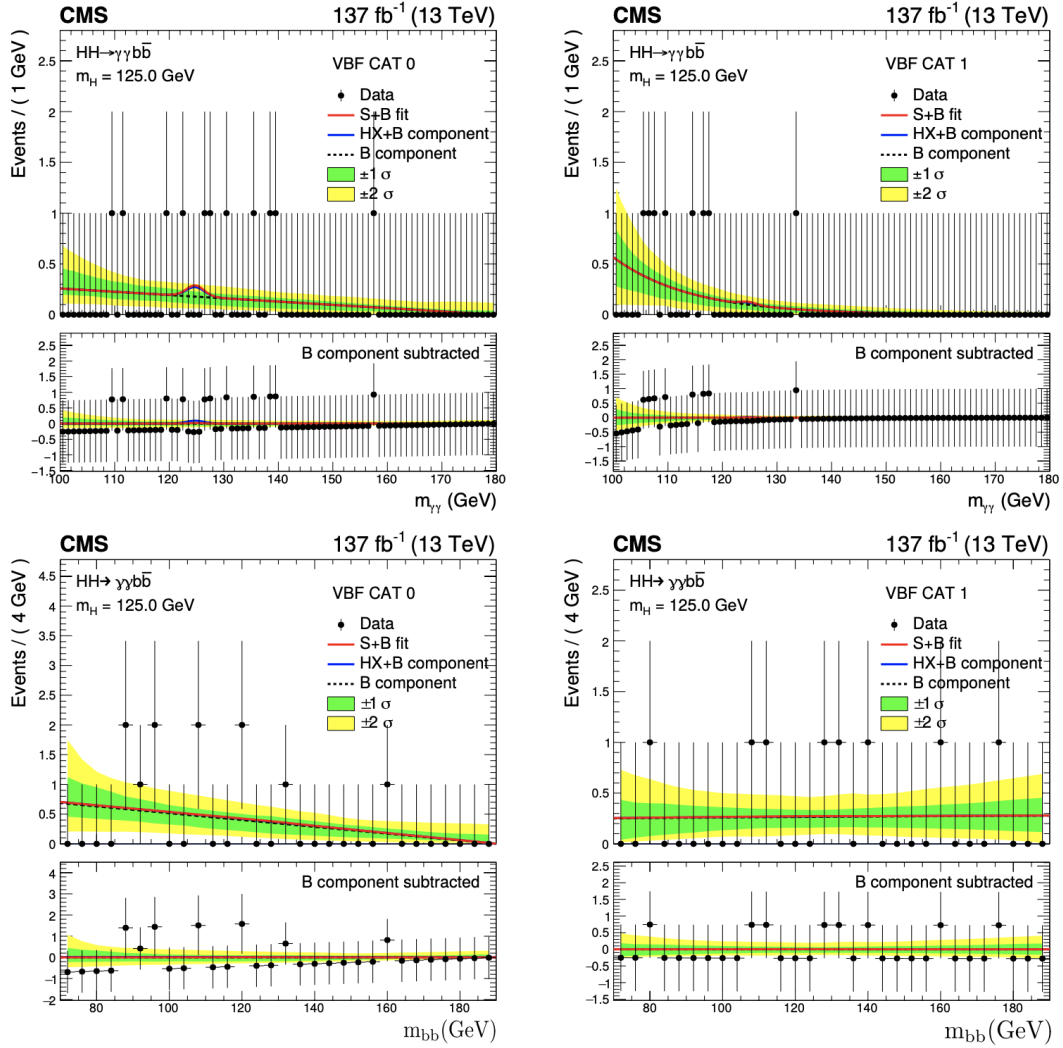


Figure 5.28: The  $m_{\gamma\gamma}$  (upper row) and  $m_{jj}$  (bottom row) distribution for the selected events in data (black points) is shown for the two VBF HH categories with the curves corresponding to the signal + background fit (solid red), the single Higgs boson and the non-resonant processes H+B (solid blue) and the background only (dashed black), with bands covering the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainties in the fitted background.

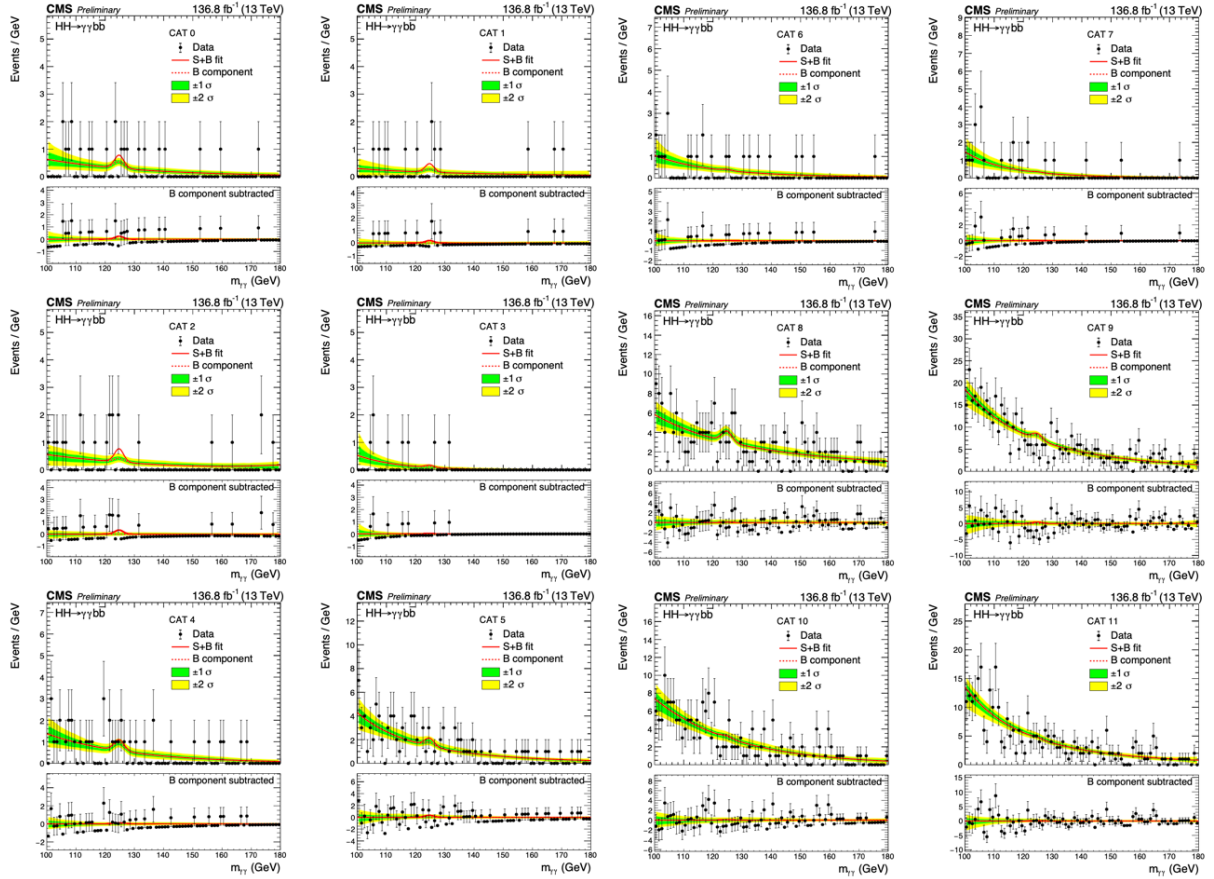


Figure 5.29: The  $m_{\gamma\gamma}$  distribution for the selected events in data (black points) is shown for the twelve ggF HH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainties in the fitted background.

## 5.9. Results

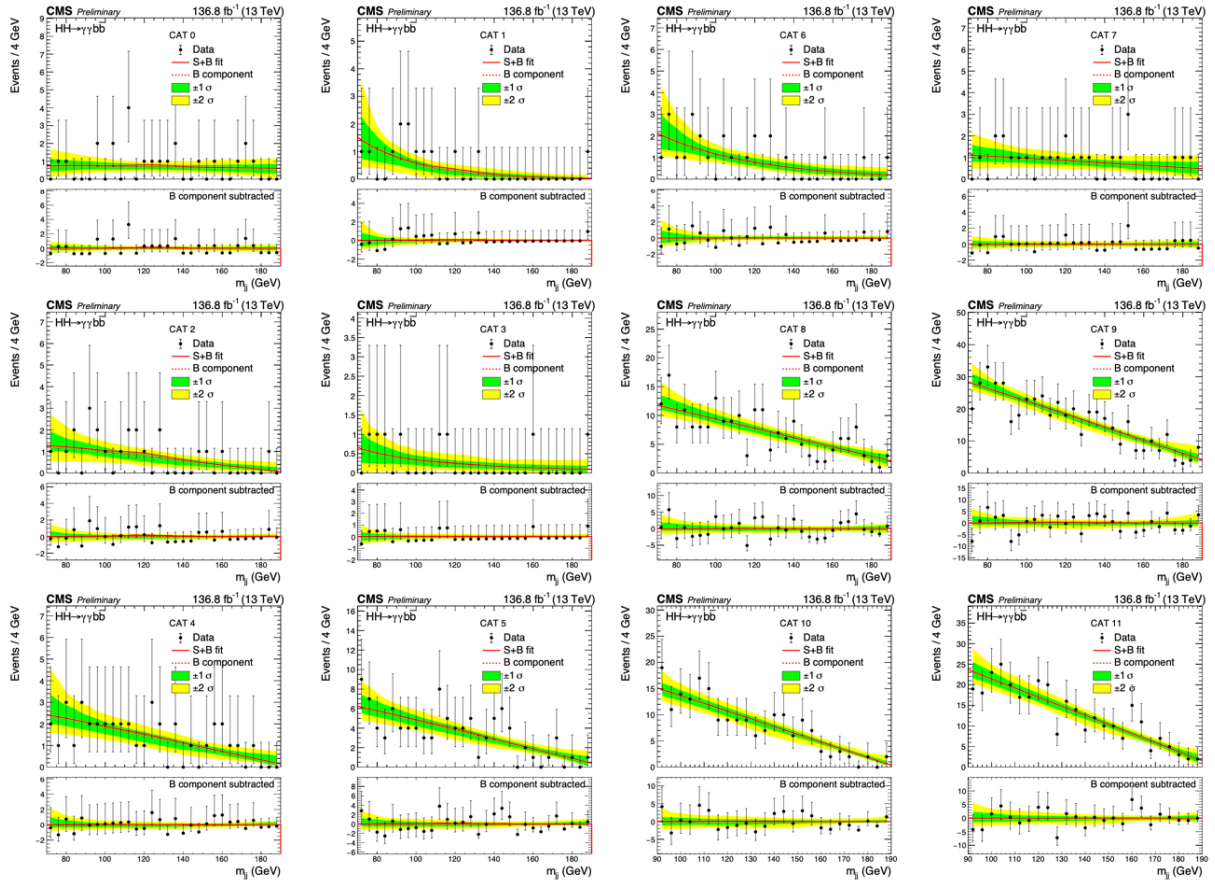


Figure 5.30: The  $m_{jj}$  distribution for the selected events in data (black points) is shown for the twelve ggF HH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainties in the fitted background.

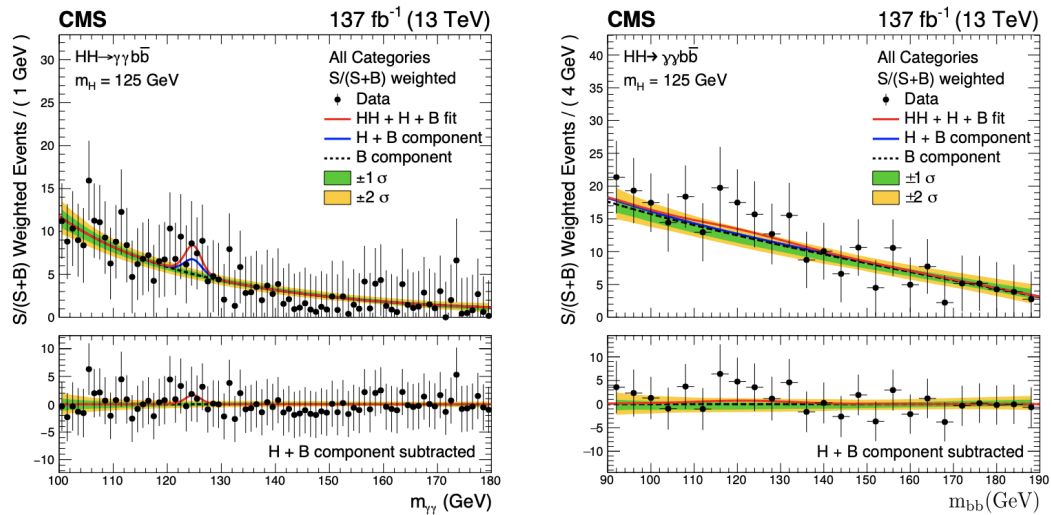


Figure 5.31: The  $m_{\gamma\gamma}$  and  $m_{jj}$  distribution for the selected events in data (black points) weighted by  $S/(S+B)$  with the curves corresponding to the signal + background fit (solid red), the single Higgs boson and the non-resonant processes H+B (solid blue) and the background only (dashed black), with bands covering the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainties in the fitted background.

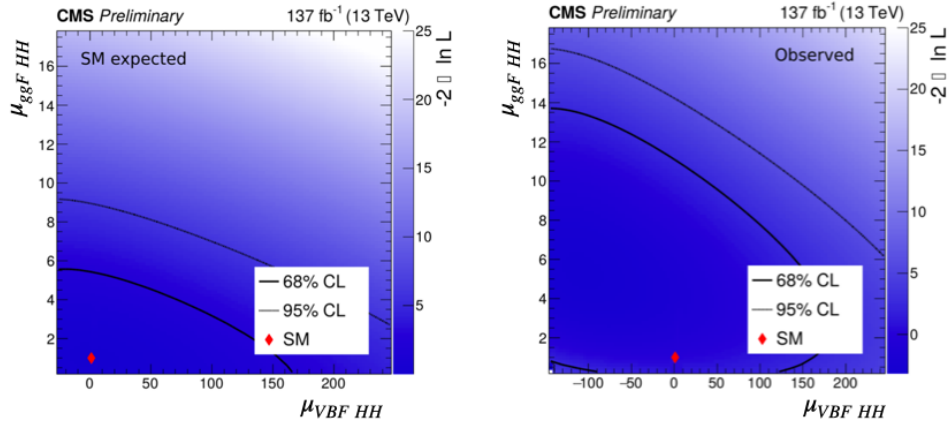


Figure 5.32: Expected and observed two dimensional likelihood scan of the  $\mu_{ggF_{HH}}$  (y-axis) and  $\mu_{VBF_{HH}}$  (x-axis) parameters, on the left and on the right side, respectively. The exclusion contours are also shown for the 68% (solid black line) and 95% (dashed black line) confidence levels.

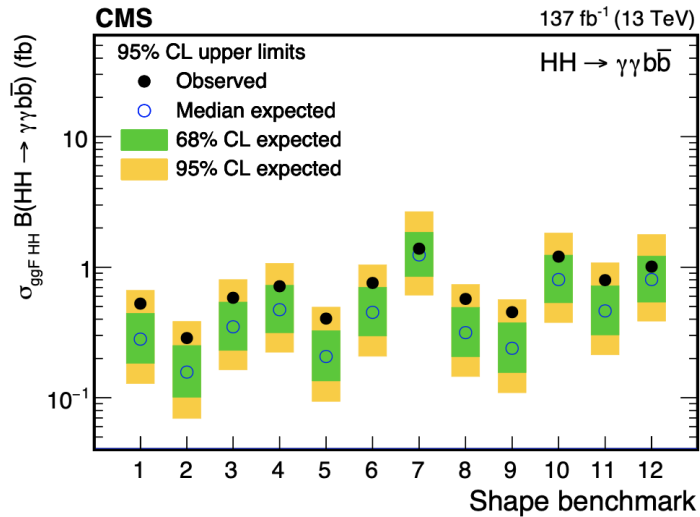


Figure 5.33: Expected and observed 95% CL upper limits on the  $\sigma_{ggF_{HH}} \times \text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$  for the BSM benchmark models, shown with transparent and solid circles, respectively. The green and yellow bands represent the one and two standard deviation extensions beyond the expected limit, respectively.

3408 thus the sensitivity to that signal. In particular, the high- $\widetilde{M}_x$  categories provides a higher  
 3409 sensitivity than the low- $\widetilde{M}_x$  categories because they have a smaller continuum background  
 3410 contamination. For  $\kappa_\lambda$  values in the  $[0,6]$  interval, the destructive interference between  
 3411 the ggF HH production diagram with a box loop of top quarks and the one with the  
 3412 triHiggs vertex is maximum. This causes a strong variation of the  $\widetilde{M}_x$  distribution which  
 3413 migrates from the highest energy spectrum at about  $\kappa_\lambda = 2$  to the softest spectrum at  
 3414 about  $\kappa_\lambda = 5$ . Comparing with the theoretical prediction, the resulting constraint on the  
 3415  $\kappa_\lambda$  parameter is found to be:

$$\begin{aligned} \text{Observed: } & -3.26 < \kappa_\lambda < 8.48 \text{ at } 95\% \text{ C.L.} \\ \text{Expected: } & -2.61 < \kappa_\lambda < 8.28 \text{ at } 95\% \text{ C.L.} \end{aligned} \tag{5.16}$$

3416 This result improves the existing most stringent constraint using the HH signal ( $-5 <$   
 3417  $\kappa_\lambda < 12$  at 95% CL), from the ATLAS HH combination with the 2016 dataset [123] ( $36$   
 3418  $fb^{-1}$ ).

3419 The same procedure is performed to extract a constraint on the  $c_{2V}$  parameter, as visible  
 3420 in the right panel of Fig. 5.34. In this case the upper limit is derived on  $\sigma_{VBF\ HH} \times$   
 3421  $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$  because the  $c_{2V}$  sensitivity is totally retained by the VBF HH process.  
 3422 As for the  $\kappa_\lambda$  parameter, the variation of the upper limit as a function of the  $c_{2V}$  value is  
 3423 determined by the corresponding  $\widetilde{M}_x$  distribution variation. In this case, the interference  
 3424 between the three VBF HH production diagrams makes the  $\widetilde{M}_x$  distribution spectrum  
 3425 migrate to high energies as soon as  $c_{2V}$  deviates from its SM prediction, enhancing the  
 3426 sensitivity to the VBF HH signal. The resulting constraint on the  $c_{2V}$  parameter is found  
 3427 to be :

$$\begin{aligned} \text{Observed: } & -1.31 < c_{2V} < 3.45 \text{ at } 95\% \text{ C.L.} \\ \text{Expected: } & -0.96 < c_{2V} < 3.07 \text{ at } 95\% \text{ C.L.} \end{aligned} \tag{5.17}$$

3428 This is the first upper limit on the  $c_{2V}$  parameter set by the CMS experiment. The  
 3429  $HH \rightarrow \gamma\gamma b\bar{b}$  channel provides a good sensitivity also to the  $c_V$  parameter.



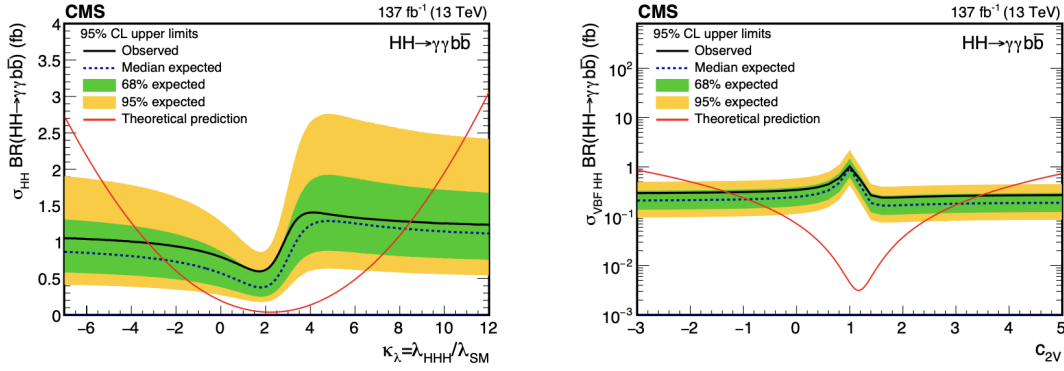


Figure 5.34: Expected and observed 95% CL upper limits on the SM-like VBF HH production cross section times  $\text{BR}(HH \rightarrow \gamma\gamma b\bar{b})$  obtained for different values of  $\kappa_\lambda$  and  $c_{2V}$  on the left and right side, respectively. The green and yellow bands represent the one and two standard deviation extensions beyond the expected limit, respectively. The red lines show the theoretical predictions.

### 3430 Likelihood Scan for the Coupling Measurements

3431 In order to measure the  $(\kappa_\lambda, \kappa_t, c_V, c_{2V})$  parameters a profile likelihood is defined as:

$$3432 \quad \mathcal{L}_{prof}(\kappa_\lambda, \kappa_t, c_V, c_{2V}) = -2 \log \frac{\mathcal{L}(\kappa_\lambda, \kappa_t, c_V, c_{2V}, \hat{\theta})}{\mathcal{L}(\hat{\kappa}_\lambda, \hat{\kappa}_t, \hat{c}_V, \hat{c}_{2V}, \hat{\theta})} \quad (5.18)$$

3432 Where  $\mathcal{L}$  is the likelihood defined in Eq. 5.5 expressed as a function of the Higgs couplings.  
 3433 The  $\hat{\kappa}_\lambda, \hat{\kappa}_t, \hat{c}_V, \hat{c}_{2V}$ , and  $\hat{\theta}$  are the parameters values maximizing  $\mathcal{L}$ , i.e. their best estimate.  
 3434 Instead,  $\hat{\theta}$  is the set of nuisance values maximizing  $\mathcal{L}$  for a given set of  $(\kappa_\lambda, \kappa_t, c_V, c_{2V})$   
 3435 values.

3436 Since the Higgs boson production and decay channels considered are insufficient to con-  
 3437 strain the full set of couplings, only one or two parameters are measured at a time while  
 3438 the other ones are fixed to the SM prediction. The profile likelihood is used to extract  
 3439 also the exclusion regions at 68% and 95% confidence level.

3440 The  $\mathcal{L}_{prof}$  value as a function of the parameter of interest (likelihood scan) is reported in  
 3441 Fig. 5.35 and 5.36 for the  $\kappa_\lambda$  and  $\kappa_t$  parameters, respectively. Each figure compares the  
 3442 likelihood scan with and without the inclusion of the  $t\bar{t}H$  categories.

3443 For the  $\kappa_\lambda$ -scan, when considering only the HH categories a difference is observed between  
 3444 the expected and the observed result because the small excess of events observed especially

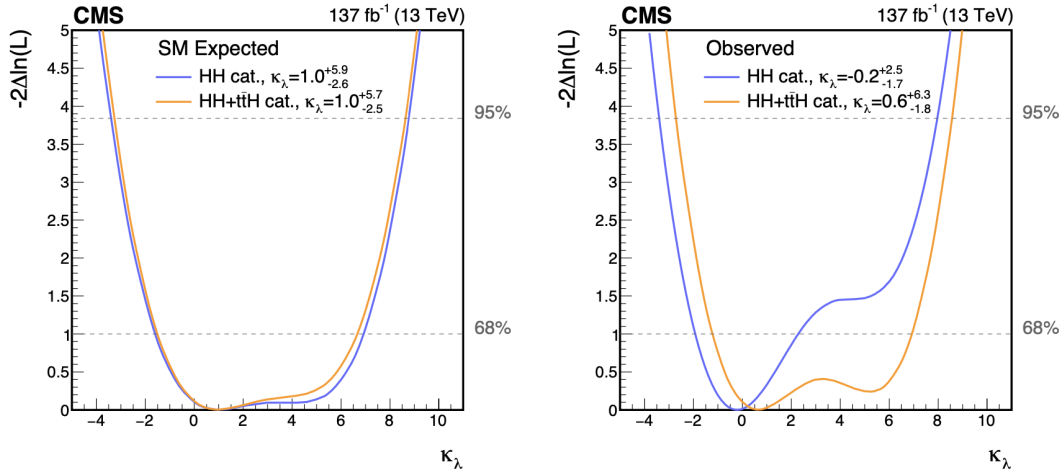


Figure 5.35: Expected and observed likelihood-scan of the  $\kappa_\lambda$  parameter, on the left and right sides, respectively. The likelihood scan is shown including the  $t\bar{t}H$  categories (orange line) or using only the HH categories (blue line). The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

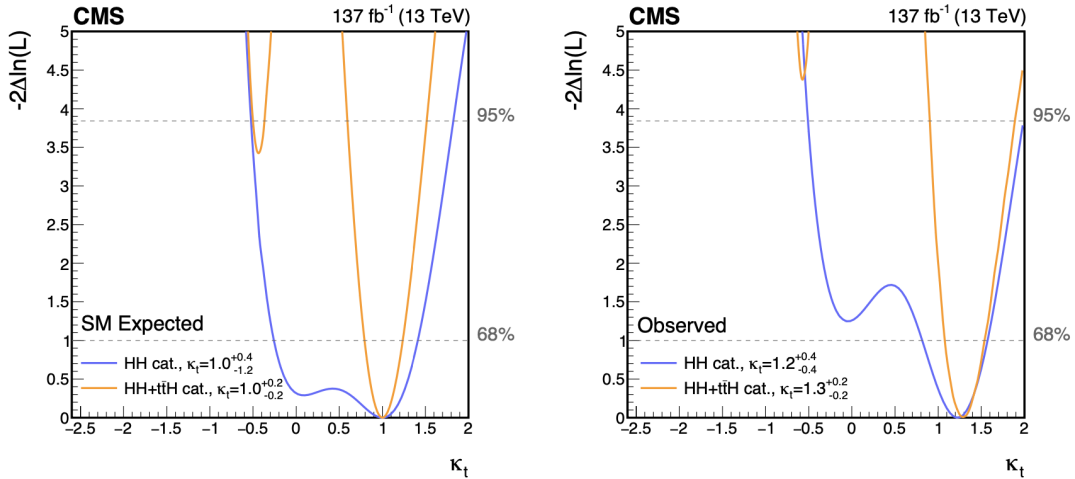


Figure 5.36: Expected and observed likelihood-scan of the  $\kappa_t$  parameter, on the left and right sides, respectively. The likelihood scan is shown including the  $t\bar{t}H$  categories (orange line) or using only the HH categories (blue line). The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

3445 in the ggF HH categories at medium  $\widetilde{M}_x$ , tends to exclude  $\kappa_\lambda$  values moderately larger  
 3446 than one. The inclusion of the  $t\bar{t}H$  categories brings the observed likelihood close to the  
 3447 expectation of the SM. In fact, the  $t\bar{t}H$  categories help constrain the  $t\bar{t}H$  contamination  
 3448 in the HH categories. In particular, an excess of events is observed also in the  $t\bar{t}H$   
 3449 categories according to Ref. [124] that measured a signal strength of  $\mu_{t\bar{t}H} = 1.4$  with an  
 3450 uncertainty of about  $\pm 30\%$ . As a consequence, the small excess of events observed in  
 3451 the HH categories is partially attributed to the excess of  $t\bar{t}H$  signal that tends instead to  
 3452 favour a  $\kappa_\lambda$  value larger than one.

3453 Due to the sensitivity of both the  $t\bar{t}H$  cross section and the  $H \rightarrow \gamma\gamma$  branching ratio to  $\kappa_t$ ,  
 3454 the addition of the  $t\bar{t}H$  categories greatly improves the sensitivity to the coupling for the  
 3455  $\kappa_t$ -scan. Specifically, for  $\kappa_t = \pm 1$ , the  $t\bar{t}H$  cross section and roughly the HH production  
 3456 cross section are symmetric. On the other hand, the sensitivity to a negative  $\kappa_t$  value  
 3457 is enhanced by the highly asymmetric  $H \rightarrow \gamma\gamma$  branching ratio dependency on  $\kappa_t$ . This  
 3458 allows the exclusion of a negative value of  $\kappa_t$  at 95% confidence level.

3459 The  $c_{2V}$  scan is shown in Fig. 5.37. For this parameter the sensitivity comes entirely from  
 3460 the VBF HH cross section, thus the contribution from the  $t\bar{t}H$  categories is completely  
 3461 negligible. A  $c_V$  likelihood scan is out of the scope for this analysis because there are  
 3462 not experimental categories targeting the VH and qqH events that dominates the  $c_V$   
 3463 sensitivity.

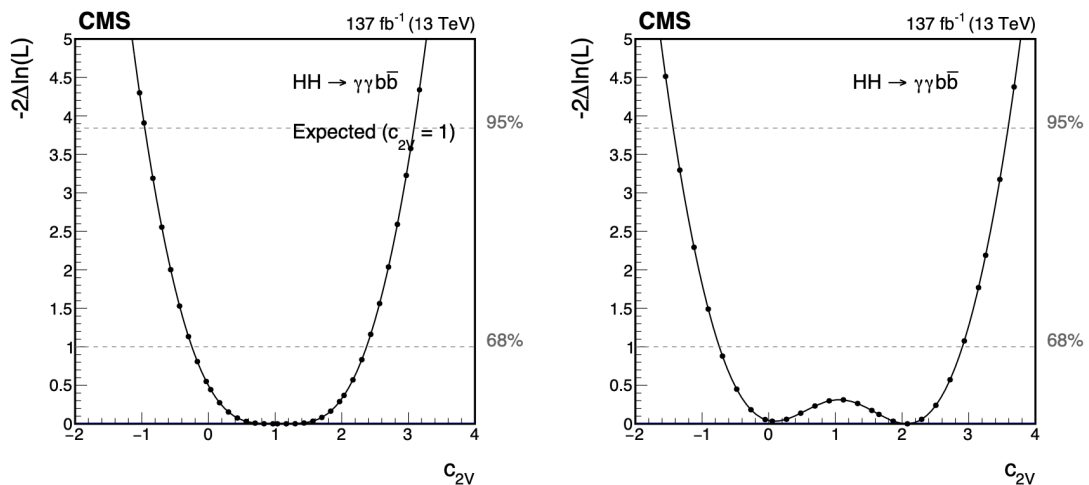


Figure 5.37: Expected and observed likelihood-scan of the  $c_{2V}$  parameter, on the left and right sides, respectively. The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

3464 The two dimensional likelihood scan of the  $(\kappa_\lambda, \kappa_t)$  parameters is reported in Fig. 5.38.  
 3465 In the figure, the regions where the parametrization of  $\sigma_{t\bar{t}H}$  ( $\kappa_\lambda, \kappa_t$ ) is not reliable are  
 3466 shown with a gray band. As expected, the improvement from the  $t\bar{t}H$  categories is large.  
 3467 The result is found compatible with the SM.

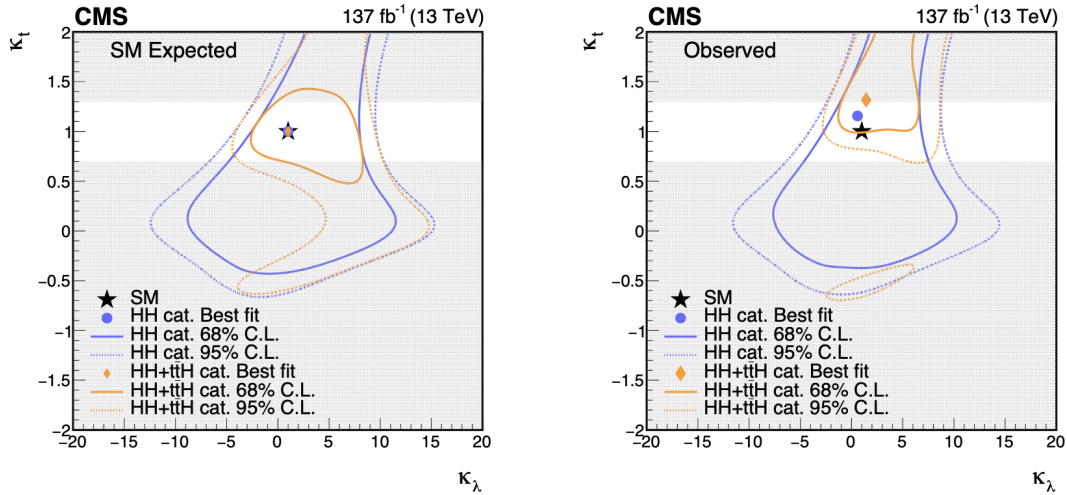


Figure 5.38: Expected (left) and observed (right) two dimensional likelihood-scans of the  $(\kappa_\lambda, \kappa_t)$  parameters including the  $t\bar{t}H$  categories (orange) and using only the HH categories (blue). The exclusion regions at 68% and 95% confidence level are represented with the solid and the dashed lines, respectively. The region where the  $\sigma_{t\bar{t}H}$  ( $\kappa_\lambda, \kappa_t$ ) parametrization is not reliable is highlighted in gray.

3468 The  $(\kappa_\lambda, c_{2V})$  likelihood scan is shown in Fig. 5.39. The sensitivity on those two pa-  
 3469 rameters is dominated by the two considered HH production mechanisms. The coupling  
 3470 measurements are summarized in Tab. 5.10.

Parameters		Best fit $\pm 1\sigma$	Interval at 95% C.L.
$\kappa_\lambda$	obs	$0.6^{+6.3}_{-1.8}$	[-2.7, 8.6]
	exp	$1.0^{+5.7}_{-2.5}$	[-3.3, 8.6]
$\kappa_t$	obs	$1.3^{+0.2}_{-0.2}$	[0.90, 1.90]
	exp	$1.0^{+0.2}_{-0.2}$	[-0.50, -0.37] $\cup$ [0.59, 1.52]
$c_{2V}$	obs	$2.1^{+0.8}_{-2.8}$	[-1.4, 3.6]
	exp	$1.0^{+1.2}_{-1.2}$	[-2.0, 3.1]
$(\kappa_\lambda, \kappa_t)$	obs	(1.4, 1.3)	-
	exp	(1, 1)	-
$(\kappa_\lambda, c_{2V})$	obs	(0.0, 0.3)	-
	exp	(1, 1)	-

Table 5.10: Best fit values for the 1D and 2D likelihood scans of the Higgs coupling parameters. For the 1D scans the  $1\sigma$  uncertainties and the 95% confidence intervals are also quoted.

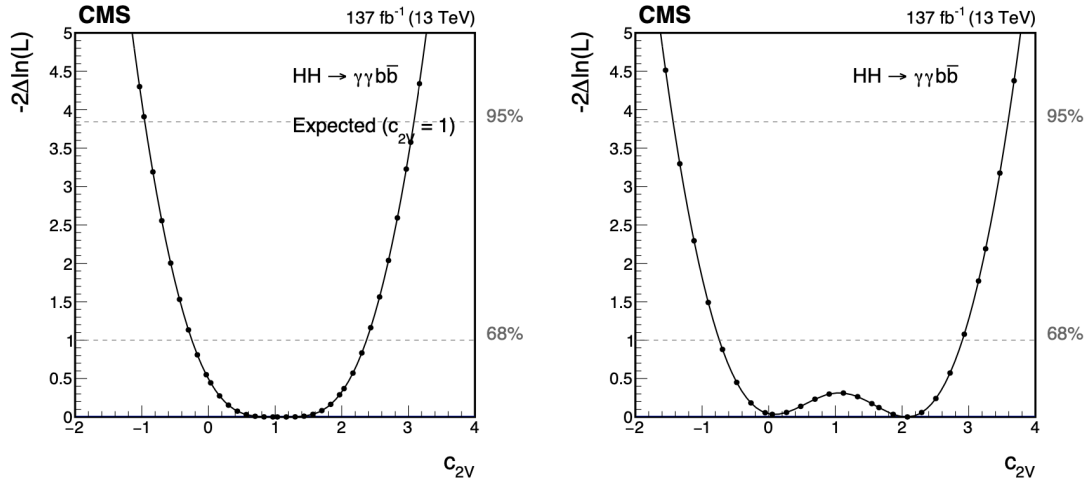


Figure 5.39: Expected (left) and observed (right) two dimensional likelihood-scans of the  $(\kappa_\lambda, c_{2V})$  parameters. The exclusion regions at 68% and 95% confidence level are represented with the solid and the dashed lines, respectively.

## 5.10 Conclusion

3471

3472 In summary, all the results were found compatible with the SM predictions. No evidence  
 3473 of the HH process is found, thus, an upper limit was set to its cross section. The observed  
 3474 upper limit on the inclusive HH production cross section is  $7.7 \times \text{SM}$  and corresponds to  
 3475 the most stringent result achieved by the CMS experiment to date. In the assumption of  
 3476 no HH signals, constraints on anomalous  $\kappa_\lambda$  values, and, for the first time with the data  
 3477 collected by the CMS experiment, on  $c_{2V}$  values are set. The observed constraints at 95%  
 3478 C.L. are  $-3.26 < \kappa_\lambda < 8.48$ , which is the most stringent constraint among the published  
 3479 results [123, 125], and  $-1.31 < c_{2V} < 3.45$ .

3480 In the assumption of a HH signal, measurements of the  $\kappa_\lambda$  and  $\kappa_t$  parameters are performed  
 3481 through a combination of HH and  $t\bar{t}H$  enriched categories. The measured values with  
 3482 one standard deviation uncertainty are  $\kappa_\lambda = 0.6^{+6.3}_{-1.8}$  and  $\kappa_t = 1.3 \pm 0.2$ . A simultaneous  
 3483 measurement of the  $\kappa_\lambda$  and  $\kappa_t$  parameters is performed and the result is shown in Fig.  
 3484 [5.38](#)

3485 Further data will be collected by the CMS experiment during the Run 3, equivalent to  
 3486 an integrated luminosity of about  $300 \text{ fb}^{-1}$ . This dataset will improve the sensitivity to  
 3487 the HH signal and to the coupling parameters. Preliminary studies provide a projected  
 3488 limit (expected SM) on the inclusive ggF HH cross section of about  $3.6 \times \text{SM}$  for the end

3489 of Run 3. Run 3 will not provide the required amount of data for an evidence of the HH  
3490 process. An evidence of a (SM) HH process is expected during the high-luminosity phase  
3491 of the LHC.

## Chapter 6

# Resonant Triple Higgs Production and Decay to Six b-Quarks

### 6.1 Introduction

The Higgs boson is the simplest manifestation of the Brout-Englert-Higgs mechanism featuring a complex scalar field that generates mass through its interaction with other particles. Extensions to the scalar sector of the standard model (SM) provide prospective methods for observing physics beyond the standard model (BSM). Many theoretical models postulate extensions featuring additional scalar fields, yielding a rich phenomenology comprising additional scalar bosons, such as in the next-to-minimal supersymmetric standard model (NMSSM) [126], the two real scalar singlet model (TRSM) [47], and models with warped extra dimensions [127].

This analysis explores one of many possible extensions and investigates the production of a BSM resonance that decays into a Higgs boson and a BSM scalar, motivated by the TRSM, which proposes three scalar bosons, the lightest of which is the familiar SM Higgs boson. In this framework, the heaviest BSM resonance, denoted by  $X$ , decays into the BSM scalar, denoted  $Y$ , and the SM Higgs boson, following the mass hierarchy  $m_X > m_Y > m_H$  where  $m_Y$  is large enough that the decay of  $Y \rightarrow HH$  is kinematically available.

The search will consider solely the scenario in which the Higgs bosons decay into pairs of

3511  $b$  quarks.

## 3512 6.2 Search Strategy

3513 The objective of this analysis is to search for the asymmetric multi-scalar production pro-  
 3514 cess in which a new resonance (X) decays into a Higgs boson (H) and an extra scalar (Y),  
 3515 where Y decays into two Higgs bosons and the three Higgs bosons decay into pairs of b  
 3516 quarks, resulting in the six b final state. The search is performed in mass ranges of X (1–4  
 3517 TeV) and Y (300–2800 GeV) where the H is highly Lorentz-boosted. In this kinematic  
 3518 regime, decayed b-quark pairs are collimated enough to allow the reconstruction of H using  
 3519 single large-area jets (fatjets). Combination of three Higgs bosons with different energies  
 3520 gives different topologies for different mixes as shown in Fig. 6.1 where fraction of gener-  
 3521 ated boosted Higgs bosons are displayed. The fraction is ratio of  $f_{gen} = N^{boostedH} / N^{Higgs}$   
 3522 where,  $N^{boostedH}$  is number of generated boosted Higgs bosons and  $N^{Higgs}$  is total number  
 3523 of Higgs bosons. The generated boosted Higgs bosons are selected using  $|\eta| < 2$  and dis-  
 3524 tance between two generated b-quark  $\Delta R(b, b) < 0.8$  cuts. The Fig. 6.1 shows different  
 3525 topologies are concentrated in different phase space depending on the  $m_X$  and  $m_Y$ . Out  
 3526 of which, two topologies are considered: one topology where all three Higgs bosons are  
 3527 boosted and another where two of the three Higgs bosons are boosted.

3528 The analyzed mass range is for  $m_X$  from 1 TeV to 4 TeV and  $m_Y$  from 250 GeV 2.8  
 3529 TeV. Events are selected by implementing a cut-based approach and the identified jets  
 3530 are paired to reconstruct the H candidates or in boosted case, fatjets are identified to  
 3531 reconstruct the H candidate.

3532 One of the most significant challenges presented in this analysis arises from the prediction  
 3533 of background events, of which the predominant expected processes are quantum chro-  
 3534 modynamic (QCD) multijet production and top quark pair ( $t\bar{t}$ ) production. Monte Carlo  
 3535 (MC) simulations of QCD multijet events are precise only to leading order (LO) and are  
 3536 therefore typically inaccurate, motivating the development of a data-driven method de-  
 3537 signed to accurately model the background shape and event yields. 2D-alphabet method  
 3538 is used to model background explained in Sec. 6.5.1



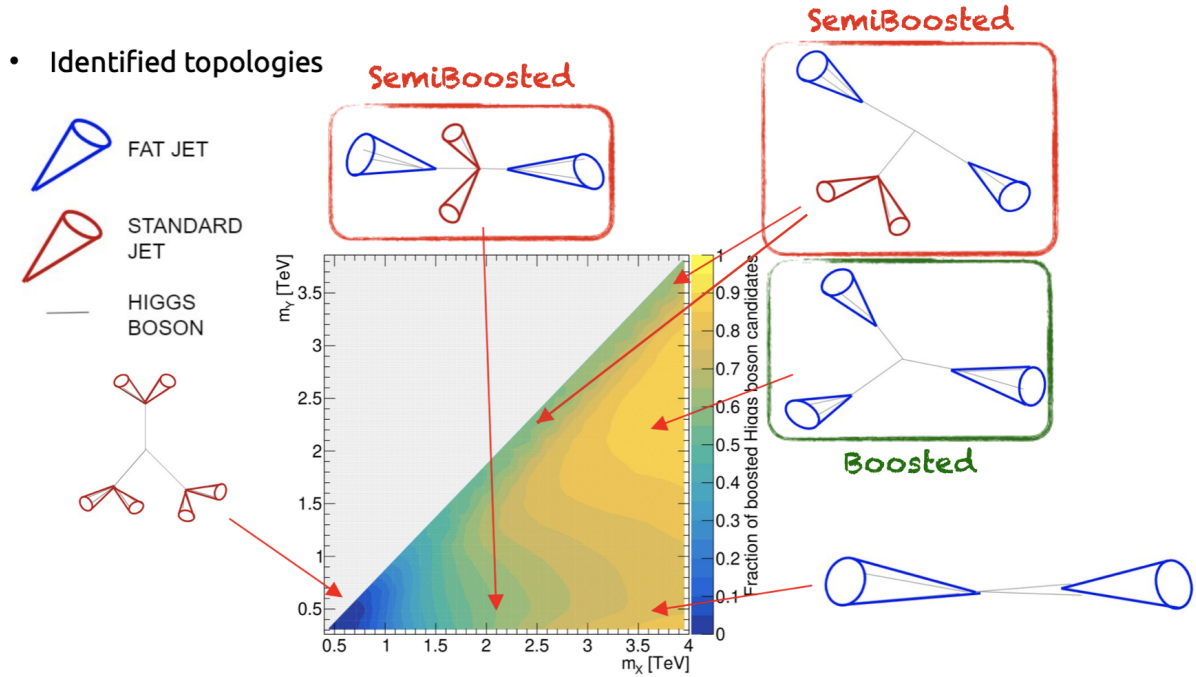


Figure 6.1: Fraction of generated boosted Higgs boson candidates in  $(m_X, m_Y)$  plane.

3539 Signal and background event distributions are plotted as a function of  $M_{jjj}$  correspond-  
 3540 ing to reconstructed mass of X and  $M_{jj}$  - reconstructed mass of Y with consideration  
 3541 of applicable systematic variations in order to extract the expected upper limit on the  
 3542 production cross section for the array  $(m_X, m_Y)$ . The data analysis will be carried out  
 3543 independently but in a similar fashion for the 2016, 2017, and 2018 datasets and the full  
 3544 Run 2 result will be a statistical combination of the independent results. The final results  
 3545 are still to be determined and published.

## 3546 6.3 Data Samples and Simulated Events

### 3547 6.3.1 Data Samples and Trigger Requirements

3548 The analysis is performed on the full Run 2 dataset, corresponding to an integrated  
 3549 luminosity of  $137.6 \text{ fb}^{-1}$  collected at  $\sqrt{s} = 13 \text{ TeV}$  in 2016, 2017, and 2018. Datasets  
 3550 produced in Ultra Legacy Campaign NanoAODv9 are used as listed below:

3551 `/JetHT/Run2017B-UL2017_MiniAODv2_NanoAODv9-v1/NANOAO`

3552 `/JetHT/Run2017C-UL2017_MiniAODv2_NanoAODv9-v1/NANOAO`

3553 /JetHT/Run2017D-UL2017\_MiniAODv2\_NanoAODv9-v1/NANOAOD

3554 /JetHT/Run2017E-UL2017\_MiniAODv2\_NanoAODv9-v1/NANOAOD

3555 /JetHT/Run2017F-UL2017\_MiniAODv2\_NanoAODv9-v1/NANOAOD

3556 The events are characterized by highly energetic fatjets arising from collimated b-quark  
 3557 pairs of boosted Higgs bosons. Multijet event selection at the trigger level is achieved by  
 3558 choosing High Level Trigger (HLT) paths featuring large transverse momentum jets. Tab.  
 3559 6.1 lists the trigger paths used in this analysis.

Year	HLT Trigger Paths
2016	HLT_PFHT900_v* OR HLT_PFJet450_v*
2017	HLT_PFHT1050_v* OR HLT_AK8PFJet500_v* OR HLT_PFJet500_v*
2018	HLT_PFHT1050_v* OR HLT_AK8PFJet500_v* OR HLT_PFJet500_v*

Table 6.1: HLT Trigger Paths applied

### 3560 6.3.2 Simulated Signal Samples

3561  $X \rightarrow Y(HH)H$  Monte Carlo samples have been produced following the recommendation  
 3562 from the LHC Higgs Cross Section working group using MADGRAPH5\_AMC@NLO 2.6.5  
 3563 to simulate event generation and PYTHIA8 to simulate the hadronization process. The  
 3564 X and Y resonances are assumed to be narrow and the generated width is set to 1 MeV.  
 3565 The generated samples are produced in context of an NMSSM model implemented.

3566 The individual dataset samples were centrally produced with the standard CMS full sim-  
 3567 ulation procedure. List of  $m_Y$  points considered for each  $m_X$  hypothesis considered for the  
 3568 analysis is listed in Tab. 6.2. Fig. 6.2 gives idea about covered phase-space of used and  
 3569 generated samples for 2017 where empty region denotes kinematically forbidden region  
 3570 ( $2m_H < m_Y < (m_X - m_H)$ )

### 3571 6.3.3 Simulated Background Samples

3572 Monte Carlo background samples are generated to determine the background composition  
 3573 98 but are not used to estimate the expected background shapes or yield. A full list of  
 3574 the MC background samples is provided in Tab. 6.3 for year 2016 and Tab. 6.4 for years  
 3575 2017 and 2018.

$m_X$	$m_Y$
1000	300, 600, 800
1200	300, 600, 800, 1000
1600	300, 600, 800, 1000, 1200, 1400
2000	300, 600, 800, 1000, 1200, 1600, 1800
2500	300, 600, 800, 1000, 1200, 1600, 2000, 2200
3000	300, 600, 800, 1000, 1200, 1600, 2000, 2500, 2800
3500	300, 600, 800, 1000, 1200, 1600, 2000, 2500, 2800
4000	300, 600, 800, 1000, 1200, 1600, 2000, 2500, 2800

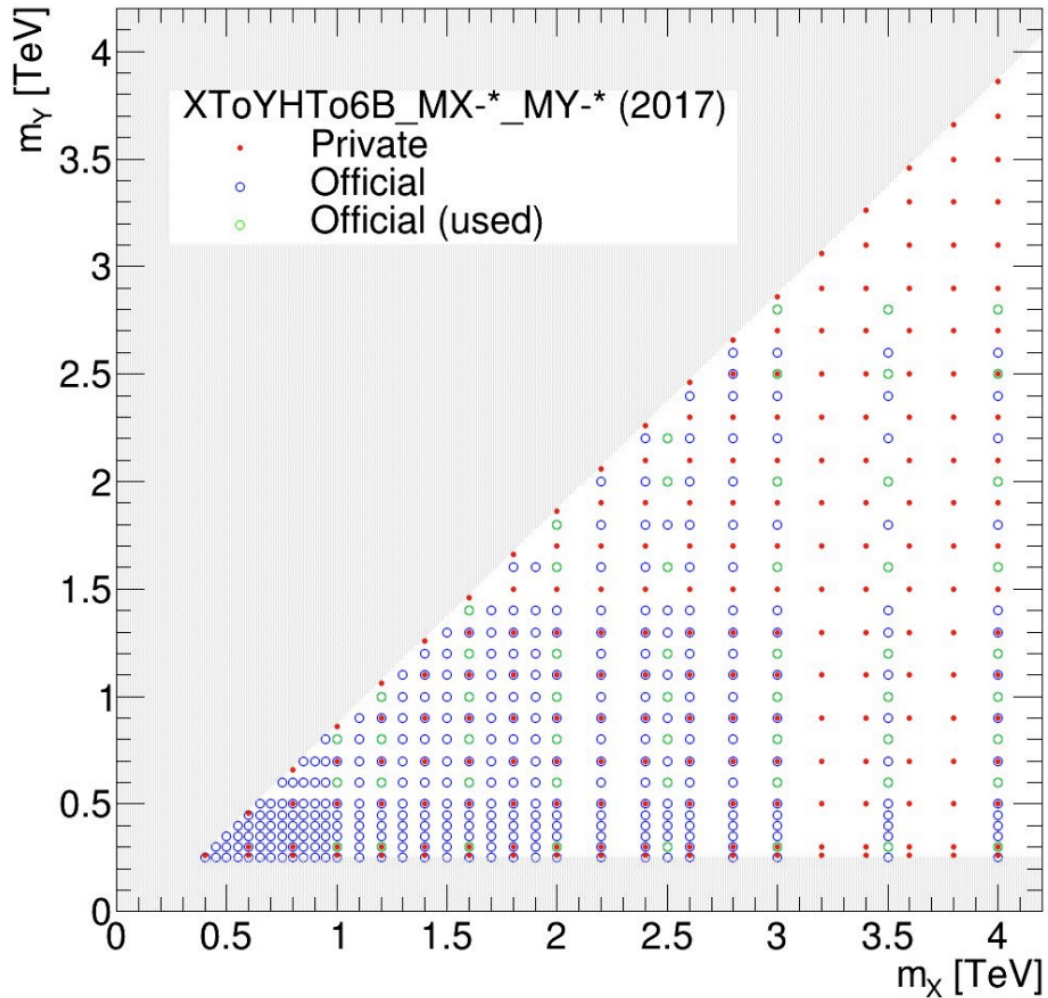
 Table 6.2: List of  $m_Y$  points for each  $m_X$  hypothesis considered in the analysis.


Figure 6.2: Simulated Signal samples for 2017. Blue circles are centrally produced samples at CMS from which green circled ones are used for this analysis.

Year	Process	Dataset Name
2016	TTbarHadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X _mcRun2_asymptotic_v17-v1/NANOADSIM
	TTbarSemileptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9 -106X_mcRun2_asymptotic_v17-v1/NANOADSIM
	QCD, $700 < H_T < 1000$ GeV	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOADSIM
	QCD, $1000 < H_T < 1500$ GeV	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOADSIM
	QCD, $1500 < H_T < 2000$ GeV	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOADSIM
	QCD, $H_T > 2000$ GeV	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOADSIM
	TTbarHadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X _mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM
	TTbarSemileptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9 -106X_mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM
	QCD, $700 < H_T < 1000$ GeV	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM
	QCD, $1000 < H_T < 1500$ GeV	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM
2016-APV	QCD, $1500 < H_T < 2000$ GeV	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM
	QCD, $H_T > 2000$ GeV	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL16 NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOADSIM

Table 6.3: Summary of the Monte Carlo background samples used for 2016.

Year	Process	Dataset Name
2017	TTbarHadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	TTbarSemileptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	QCD, $700 < H_T < 1000$ GeV	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	QCD, $1000 < H_T < 1500$ GeV	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	QCD, $1500 < H_T < 2000$ GeV	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	QCD, $H_T > 2000$ GeV	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOADSIM
	TTbarHadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOADSIM
	TTbarSemileptonic	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOADSIM
	QCD, $700 < H_T < 1000$ GeV	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOADSIM
	QCD, $1000 < H_T < 1500$ GeV	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOADSIM
2018	QCD, $1500 < H_T < 2000$ GeV	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOADSIM
	QCD, $H_T > 2000$ GeV	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOADSIM

Table 6.4: Summary of the Monte Carlo background samples used for 2017 and 2018.

## 6.4 Event Selection

This analysis focuses on two main topology as mentioned before. The first, boosted topology is where all three Higgs bosons are lorentz boosted and each Higgs boson decays into AK8jet consisting of two b-quarks which gives three AK8jets. Thus, for signal region, at least three AK8jets are required with  $p_T > 250$  GeV,  $|\eta| < 2.5$ , and  $m_{SD} \in [100, 150]$ . To select validation region close to signal region phase space, it has same selection criteria as signal region except two AK8jets with highest  $p_T$  are required to be outside the Higgs mass window with  $m_{SD} > 50$ .

The second topology considered is semi-boosted topology. It has two out of three Higgs bosons lorentz boosted, each decaying to one AK8jet consisting two b-quarks and one Higgs boson decays into two separate b-quark jets gives two AK8jets and pair of AK4 b-jets. To avoid double counts, semi-boosted topology requires exactly two AK8jets present in the event. For signal region, events with exactly two AK8jets with  $p_T > 250$  GeV,  $|\eta| < 2.5$ , and  $m_{SD} \in [100, 150]$  are required. Two AK4jets are required with  $p_T^{jet} > 30$  GeV,  $|\eta^{jet}| < 2.5$ , DeepJet score  $> 0.0532$  (loose WP), to separate them from already selected two AK8jets,  $\Delta R(AK4, AK8) > 0.8$  and invariant mass of jet pair,  $m_{dijet} \in [90, 150]$ . For the validation region, exactly same cuts except two AK8jets are required to be outside the Higgs mass window with  $m_{SD} > 50$ .

For background estimation, explained in Section 6.5.1, events are divided into pass and fail category based on ParticleNet XbbvsQCD score. If one of more AK8jet passes the loose working point (WP) cut (ParticleNet XbbvsQCD score  $> 0.9105$ ), the event falls into the pass category. For the cases where no jet passes the cut, the event is in fail category. For both boosted and semi-boosted topology, it is applied in the same manner. Fig. 6.3 gives overview of selection cuts for boosted and semi-boosted topologies respectively.

Fig. 6.4 shows cut-flow for boosted and semi-boosted signal region with pass and fail categories is given. Fig. 6.5 gives efficiency map of boosted topology (left) with maximum efficiency around 20% and semi-boosted topology (right) with maximum efficiency around 15%. In large parts of the mass plane boosted and semi-boosted selections complement one another.

Boosted signal region (SR_b)	
$\geq 3$ AK8 jets with $p_T > 250$ GeV $ \eta  < 2.5$ $100 < m_{\text{softdrop}} < 150$ GeV	
<b>Categories:</b> based on ParticleNet score $> 0.9105$ (loose WP)	
Fail	Pass
0 jets passing the cut	$\geq 1$ jets passing the cut

Semiboosted signal region (SR_sb)	
$= 2$ AK8 jets as in the boosted selection $\geq 1$ AK4 jet pair with $p_T^{\text{jet}} > 30$ GeV $ \eta^{\text{jet}}  < 2.5$ DeepB $^{\text{jet}} > 0.0532$ (loose WP) $\Delta R(\text{AK4}, \text{AK8}) > 0.8$ $90 < m_{\text{dijet}} < 150$ GeV	
<b>Categories:</b> as in the boosted selection	
Fail	Pass
0 AK8 jets passing the cut	$\geq 1$ AK8 jets passing the cut

Boosted validation region (VR_b)	
Same as the signal region but with the 2 leading jets (in $p_T$ ) outside the Higgs mass window and having $m_{\text{softdrop}} > 50$ GeV	

Semiboosted validation region (VR_sb)	
Same as the signal region but with the 2 AK8 jets outside the Higgs mass window and having $m_{\text{softdrop}} > 50$ GeV	

Figure 6.3: Event selection criteria for boosted and semi-boosted topologies.

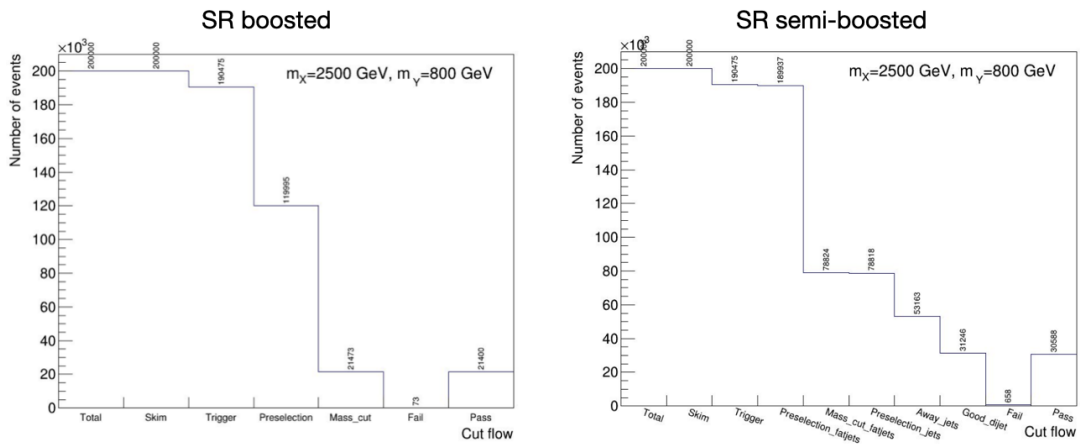


Figure 6.4: Cut-flow for boosted and semi-boosted signal region with pass and fail categories.

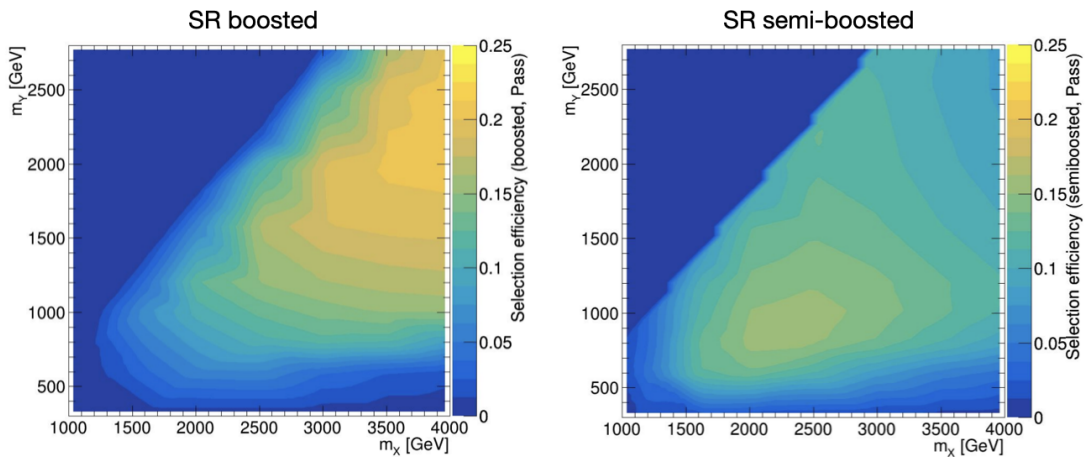


Figure 6.5: Efficiency map for boosted and semi-boosted signal region pass category.

### 3605 **6.4.1 Dijet mass vs Trijet mass**

3606 As mentioned in Section 6.2, presence of two heavy resonances X and Y in  $m_{jj}$  vs  $m_{jjj}$   
3607 distribution indicates the signal. The analysis is to search for 2D bump corresponding to  
3608 X and Y resonances. This 2D plane consists of dijet mass ( $m_{jj}$ ) and trijet mass ( $m_{jjj}$ ).  
3609 After the event selection, we get three Higgs boson candidates and invariant mass of all  
3610 three Higgs boson gives reconstructed mass of X resonance - trijet mass.

3611 The Y resonance decays into two Higgs bosons. From three reconstructed Higgs boson  
3612 candidates, there can be three possible pairs. There is no general way to find the right  
3613 combination of Higgs boson candidate pair from the Y resonance. To overcome this issue,  
3614 all three pair combinations are considered and for dijet mass distribution, there are three  
3615 entries per event. It is still expected to see mass peak at  $m_Y$ .

3616 Fig. 6.6 shows the distribution of trijet mass and dijet mass for mass points  $m_X = 2500$ ,  
3617  $m_Y = 600, 800, 1200, 2000$  for boosted and semi-boosted topologies.

## 3618 **6.5 Background Estimation**

3619 Since background events are not accurately modeled by Monte Carlo simulations, a data  
3620 driven background estimation procedure is developed and used to model the background  
3621 in the signal region. Background modeling is performed using 2D Alphabet method.

### 3622 **6.5.1 2D Alphabet**

3623 2D Alphabet is a framework to construct the workspace for a specific type of background  
3624 estimate, provide input to the Combine statistical tool, plot the 2D distributions from  
3625 the fit result, and provide the infrastructure to test this result.

3626 The name of the framework is derived from its data-driven background estimate of com-  
3627 binatorial backgrounds that are otherwise poorly modeled by Monte Carlo simulation. In  
3628 many cases, the background being modeled is QCD multijet production and as a default.



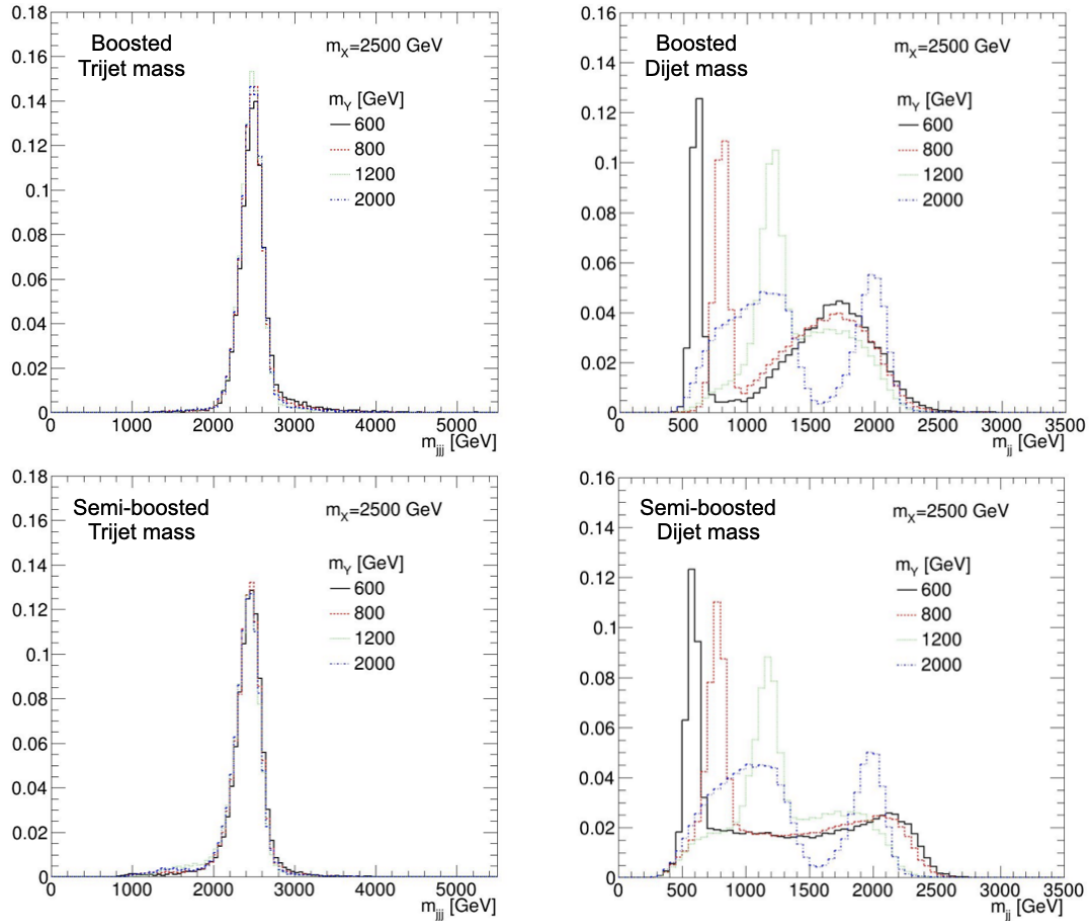


Figure 6.6: Trijet mass (left) and dijet mass (right) distribution for boosted (upper) and semi-boosted (lower) topologies.

3629 However, depending on the selection, there may be other backgrounds accounted for as  
 3630 well such as  $t\bar{t}$  in our case.

3631 The data driven background estimation method is a two dimensional version. It uses an  
 3632 analytic *transfer function* to *transfer* the background contribution in a control region to  
 3633 the contribution in the signal region. If the shapes of the background distributions in the  
 3634 control region and signal region are identical, then the transfer function is just a constant  
 3635 factor which only changes the normalization from one region to the other.

3636 The events are divided in two categories: pass - signal rich and fail - signal depleted  
 3637 category. These categories are further divided in signal region (SR) and validation region  
 3638 (VR). Validation region is used to validate the background estimation method. This gives  
 3639 four separate phase-space which consists of three signal depleted and one signal rich region  
 3640 shown in Fig. 6.7.

3641 The ABCD method measures data distributions in selection regions A, B, and D (Fig.  
 3642 6.7) which are enriched in background and depleted of signal. In the figure, var1 and var2  
 3643 are the selection variables and the C region is the signal region of the analysis. Binned  
 3644 distributions are created in some third variable (var3) for each of the four regions. The  
 3645 ratio of  $A/B$  and  $C/D$  are assumed to be equal and therefore,  $A/B * D = C$ . Here,  $A/B$   
 3646 is the transfer function.

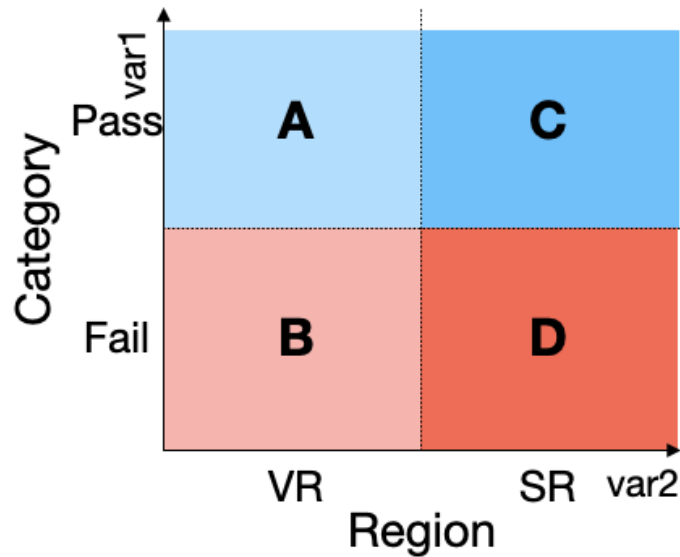


Figure 6.7: Illustration of 2D alphabet method.

3647 To create QCD estimate, measure the binned transfer function from A and B and then  
 3648 weight events in region D to get the estimate along your var3 axis of the QCD. a smooth  
 3649  $R_{P/F}$  transfer function relates event yields of the data-driven background components in  
 3650 the pass (P) and fail (F) categories. For QCD background, event yields in  $i^{th}$  bin for fail  
 3651 category can be given as:

$$n_F^{QCD}(i) = n_F^{data}(i) - n_F^{bkg,MC}(i) \quad (6.1)$$

3652 Here, last term is well modeled background component other than QCD taken from MC  
 3653 simulation. For fully data-driven background estimation for all backgrounds combined,  
 3654 one can set  $n_F^{bkg,MC}(i) = 0$ . Transfer function is derived from the ratio of background  
 3655 components in pass and fail category.

$$n_P^{QCD}(i) = n_F^{QCD}(i) \cdot R_{P/F}(i) \quad (6.2)$$

3656 where  $R_{P/F}$  transfer function is modeled as a simple low-order 2D polynomial.

## 3657 6.5.2 2DAlphabet fits

3658 From event selection, we have signal and validation regions with pass and fail categories.  
 3659 The first step for background estimation is to derive transfer function from validation  
 3660 regions. Fig. 6.8 and 6.9 show 2D Alphabet background only fit for pass and fail cate-  
 3661 gories in validation region for boosted and semi-boosted topology respectively using 2017  
 3662 samples.

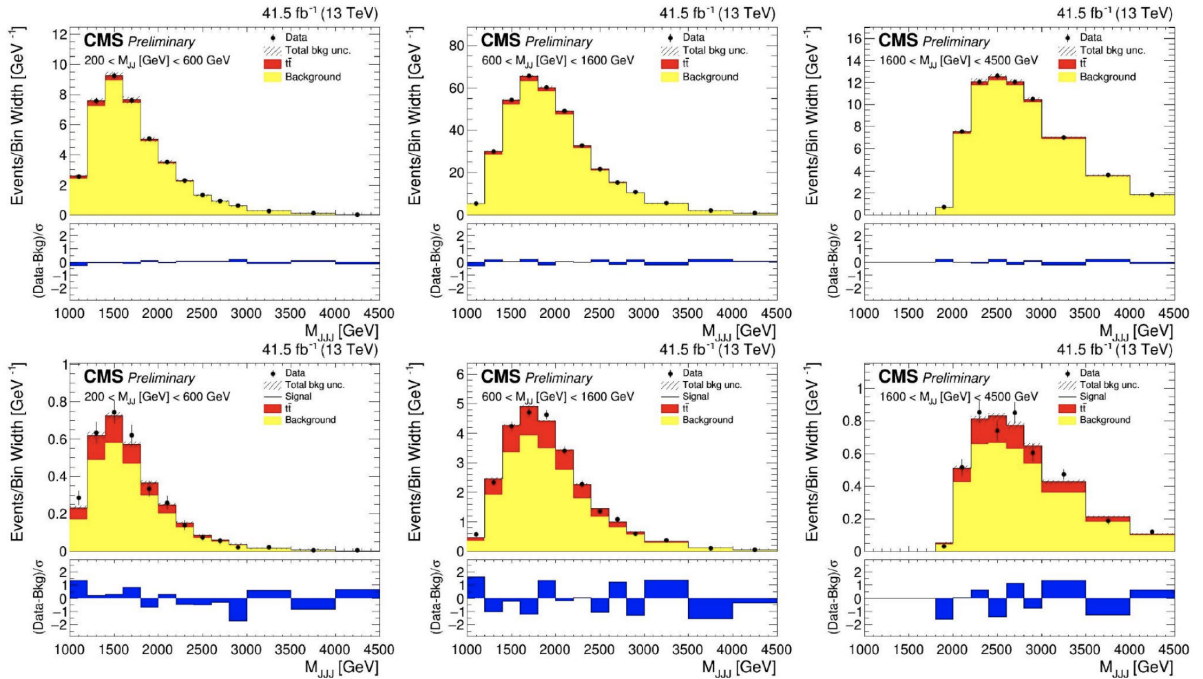


Figure 6.8: 2D Alphabet background only fit for fail (upper) and pass (lower) categories in validation region for boosted topology.

3663 After deriving transfer function from validation region, the 2DAlphabet method would  
 3664 formally require unblinding the signal region (SR) for in-situ  $R_{P/F}$  transfer function mea-  
 3665 surement. It is only possible after unblinding the analysis. Before that, signal-depleted  
 3666 SR fail category can be unblinded and use the data together with the validation region  
 3667 (VR)  $R_{P/F}$  transfer function to obtain toy data in the SR pass category assuming that  
 3668  $R_{P/F}$  transfer function is same for validation region and signal region.

## 6.5. Background Estimation

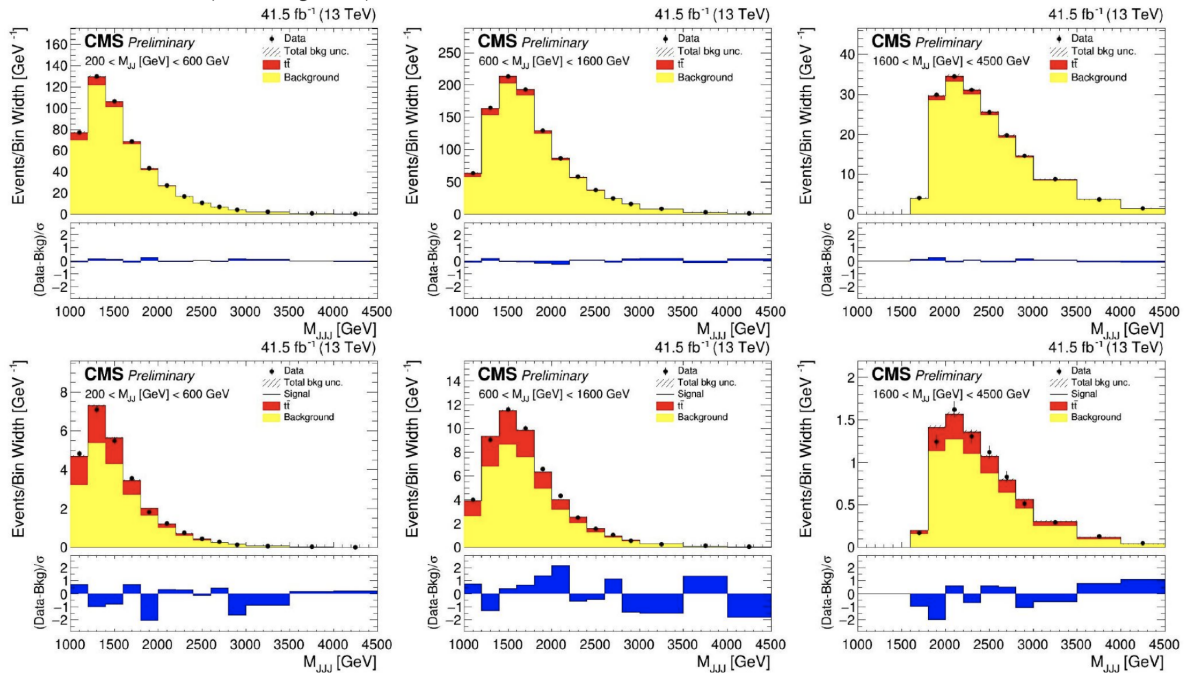


Figure 6.9: 2DAlphabet background only fit for fail (upper) and pass (lower) categories in validation region for semi-boosted topology.

3669 Fig. 6.10 and 6.11 show 2DAlphabet background only fit for pass and fail categories in  
 3670 signal region (using generated toy data in pass category) for boosted and semi-boosted  
 3671 topology respectively using 2017 samples.

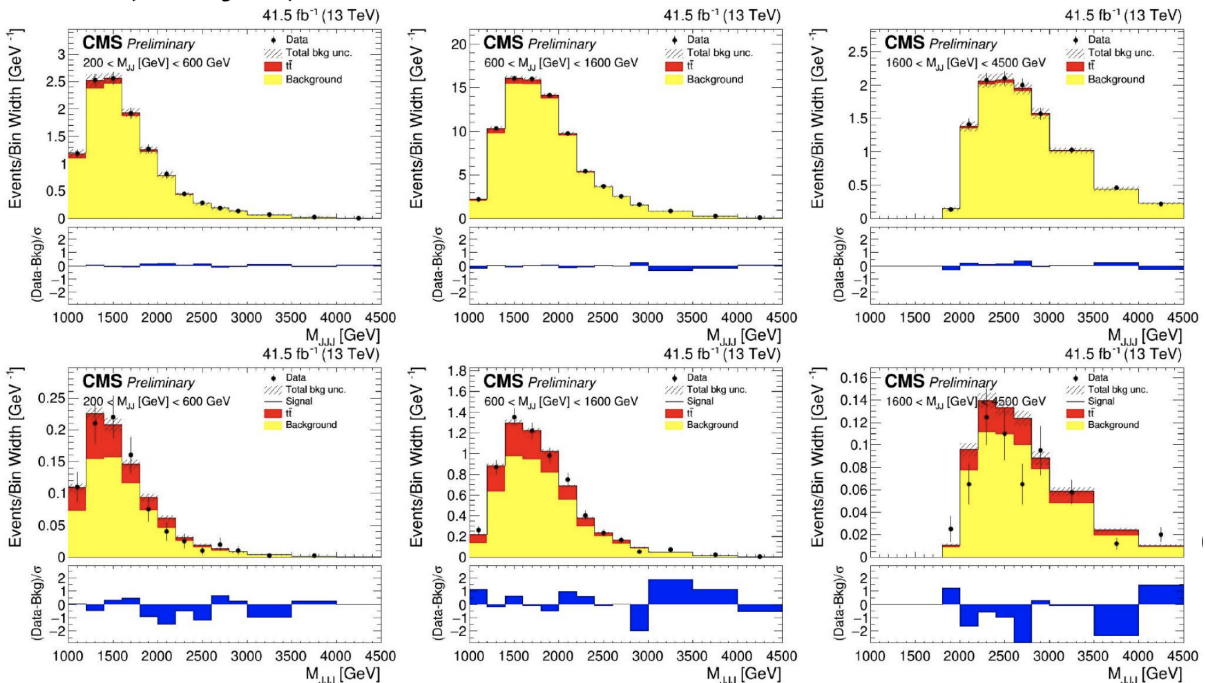


Figure 6.10: 2D Alphabet background only fit for fail (upper) and pass (lower, using toy data) categories in validation region for boosted topology.

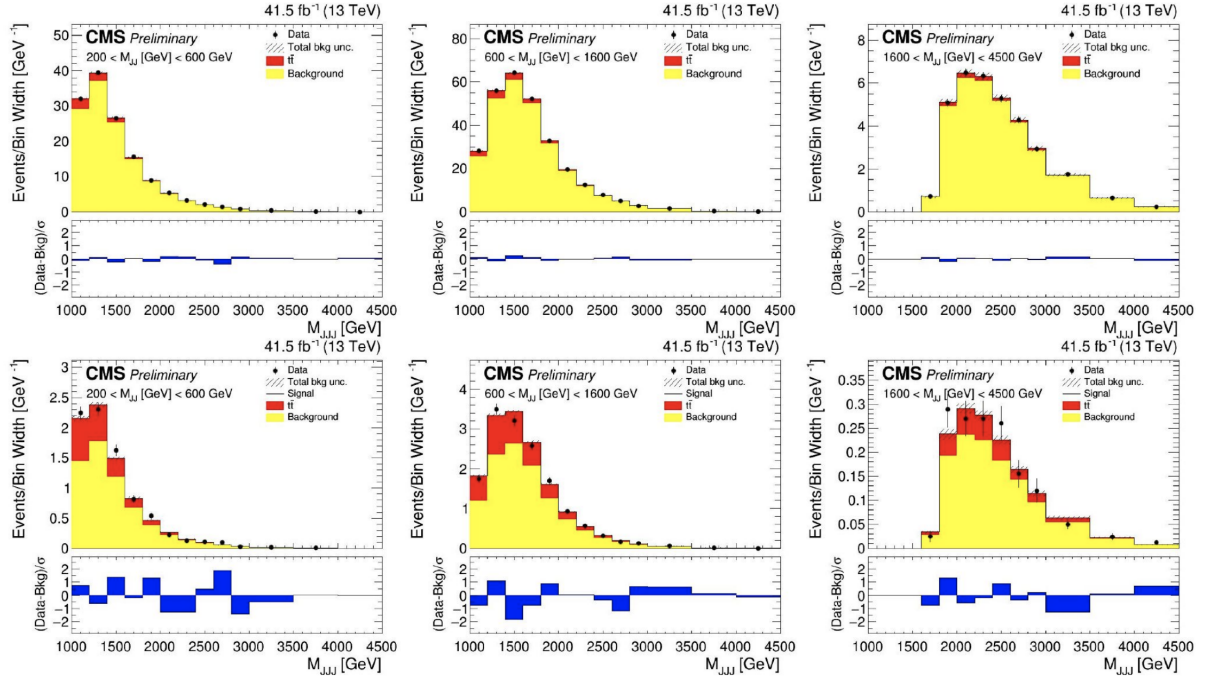


Figure 6.11: 2DAlphabet background only fit for fail (upper) and pass (lower, using toy data) categories in validation region for semi-boosted topology.

## 6.6 Future Goals

3672

3673 Several sources of systematic uncertainties are to be considered in the analysis. Next  
 3674 steps are to implement systematic uncertainties and check for different ways to increase  
 3675 signal sensitivity. The end goal is to extract the expected upper limit on the production  
 3676 cross section for the array  $(m_X, m_Y)$ .

## Chapter 7

### Conclusion

The search for the non-resonant production of Higgs boson pairs in the  $HH \rightarrow b\bar{b}\gamma\gamma$  final state is presented. The data analyzed were collected by the CMS detector in proton-proton collisions with a center-of-mass energy of 13 TeV, for a total integrated luminosity of  $137 \text{ fb}^{-1}$ . The considered di-Higgs production processes are the ggF HH, which is the main production process, and the VBF HH.

A b-jet energy regression is developed to improve b-jets resolution and, therefore, the invariant mass of the two jets coming from the  $H \rightarrow b\bar{b}$  decay. It is a key element to achieve the best sensitivity because it determines the width of the  $H \rightarrow b\bar{b}$  peak in the dijet mass distribution, which is used for the signal measurement. Overall the b-jet energy corrections improve the  $m_{jj}$  resolution of about 25%. This regression technique was validated using CMS  $H \rightarrow b\bar{b}$  discovery data.

No significant deviations from the SM predictions are found for the  $HH \rightarrow b\bar{b}\gamma\gamma$  process in the two production mechanisms considered. In particular, no evidence of the HH signal is found, thus an upper limit on its cross section is set. The observed (expected) upper limit on the inclusive  $HH \rightarrow b\bar{b}\gamma\gamma$  cross section is 7.7 (5.2) times the Standard Model. This result is the most stringent upper limit on the HH cross section set by the CMS experiment to date. Constraints on anomalous values of the Higgs coupling parameters  $\kappa_\lambda$ ,  $\kappa_t$ , and  $c_{2V}$  are also extracted. In the hypothesis of no HH signal and all the other Higgs couplings equal to their SM value, the  $\kappa_\lambda$  parameter is constrained in the range [-3.3, 8.5] at 95% confidence level. This is the most stringent constraint to the  $\kappa_\lambda$  parameter

3699 from a HH search, to date. Under the same assumptions, the  $c_{2V}$  parameter is constrained  
3700 in the range  $[-1.3, 3.5]$  at 95% confidence level.

3701 The search for resonant production of triple Higgs boson considering TRSM model is  
3702 presented. As an extension of SM, TRSM introduces two new real scalars  $X$  and  $Y$  to  
3703 enhance the SM cross-section of triple Higgs boson production. The process of two real  
3704 scalar singlets  $X$  and  $Y$  decaying to the SM Higgs boson  $H$  is considered. Since  $H \rightarrow b\bar{b}$   
3705 has the highest branching ratio, to maximize the statistics, we are considering all three  
3706 Higgs bosons decay to b-quark pairs. This gives,  $X \rightarrow Y(HH)H \rightarrow b\bar{b}b\bar{b}b\bar{b}$ . The search is  
3707 performed in mass ranges of  $X$  (1–4 TeV) and  $Y$  (300–2800 GeV) where the  $H$  is highly  
3708 Lorentz-boosted. In this kinematic regime, decayed b-quark pairs are collimated enough  
3709 to allow the reconstruction of  $H$  using single fat jet. Here, combination of three Higgs  
3710 bosons will give different topology, out of which we have considered topology where all  
3711 three Higgs bosons are boosted, and two of the three Higgs bosons are boosted.

3712 The major background for this process is QCD multijets and top quark production. We  
3713 have used 2D-alphabet method to estimate the background shape in the Signal region.  
3714 This analysis is still ongoing. A scan will be performed in a two dimensional plane spanned  
3715 by the mass of the two large-area jets associated to  $Y$ , and the invariant mass of three  
3716 large-area jets used to reconstruct  $X$ .

3717

# Bibliography

3718

- [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”. In: (2012). DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). eprint: [arXiv:1207.7214](https://arxiv.org/abs/1207.7214) (cit. on pp. 1, 5, 14, 22, 39).

3719

3720

3721

3722

- [2] The CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: (2012). DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). eprint: [arXiv:1207.7235](https://arxiv.org/abs/1207.7235) (cit. on pp. 1, 5, 14, 22, 23, 39).

3723

3724

3725

- [3] CMS Collaboration. “Observation of a new boson with mass near 125 GeV in pp collisions at  $\sqrt{s} = 7$  and 8 TeV”. In: (2013). DOI: [10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081). eprint: [arXiv:1303.4571](https://arxiv.org/abs/1303.4571) (cit. on p. 1).

3726

3727

3728

- [4] Ilaria Brivio and Michael Trott. “The Standard Model as an Effective Field Theory”. In: (2017). DOI: [10.1016/j.physrep.2018.11.002](https://doi.org/10.1016/j.physrep.2018.11.002). eprint: [arXiv:1706.08945](https://arxiv.org/abs/1706.08945) (cit. on p. 1).

3729

3730

3731

- [5] Steven W. Allen, August E. Evrard, and Adam B. Mantz. “Cosmological Parameters from Observations of Galaxy Clusters”. In: (2011). DOI: [10.1146/annurev-astro-081710-102514](https://doi.org/10.1146/annurev-astro-081710-102514). eprint: [arXiv:1103.4829](https://arxiv.org/abs/1103.4829) (cit. on p. 1).

3732

3733

3734

- [6] Edvige Corbelli and Paolo Salucci. “The Extended Rotation Curve and the Dark Matter Halo of M33”. In: (1999). DOI: [10.1046/j.1365-8711.2000.03075.x](https://doi.org/10.1046/j.1365-8711.2000.03075.x). eprint: [arXiv:astro-ph/9909252](https://arxiv.org/abs/astro-ph/9909252) (cit. on p. 1).

3735

3736

3737

- [7] Ze-Long Yi and Tong-Jie Zhang. “Constraints on holographic dark energy models using the differential ages of passively evolving galaxies”. In: (2006). DOI: [10.1142/S0217732307020889](https://doi.org/10.1142/S0217732307020889). eprint: [arXiv:astro-ph/0605596](https://arxiv.org/abs/astro-ph/0605596) (cit. on p. 1).

3738

3739



- 3740 [8] Fred Jegerlehner. *The hierarchy problem of the electroweak Standard Model revis-*  
3741 *ited*. 2013. eprint: [arXiv:1305.6652](https://arxiv.org/abs/1305.6652) (cit. on p. 1).
- 3742 [9] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. “Search for Neutrinos  
3743 from the Sun”. In: *Physical Review Letters* 20.21 (May 1968), pp. 1205–1209. ISSN:  
3744 0031-9007. DOI: [10.1103/physrevlett.20.1205](https://doi.org/10.1103/physrevlett.20.1205). URL: [http://dx.doi.org/10.](http://dx.doi.org/10.1103/PhysRevLett.20.1205)  
3745 [1103/PhysRevLett.20.1205](http://dx.doi.org/10.1103/PhysRevLett.20.1205) (cit. on p. 1).
- 3746 [10] J. Wess and B. Zumino. “Supergauge transformations in four dimensions”. In: *Nu-*  
3747 *clear Physics B* 70.1 (Feb. 1974), pp. 39–50. ISSN: 0550-3213. DOI: [10.1016/0550-](https://doi.org/10.1016/0550-3213(74)90355-1)  
3748 [3213\(74\)90355-1](https://doi.org/10.1016/0550-3213(74)90355-1). URL: [http://dx.doi.org/10.1016/0550-](http://dx.doi.org/10.1016/0550-3213(74)90355-1)  
3749 [3213\(74\)90355-1](http://dx.doi.org/10.1016/0550-3213(74)90355-1) (cit. on p. 1).
- 3750 [11] Lisa Randall and Raman Sundrum. “A Large Mass Hierarchy from a Small Extra  
3751 Dimension”. In: (1999). DOI: [10.1103/PhysRevLett.83.3370](https://doi.org/10.1103/PhysRevLett.83.3370). eprint: [arXiv:hep-](https://arxiv.org/abs/hep-ph/9905221)  
3752 [ph/9905221](https://arxiv.org/abs/hep-ph/9905221) (cit. on p. 2).
- 3753 [12] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation*  
3754 3.08 (Aug. 2008), S08001–S08001. ISSN: 1748-0221. DOI: [10.1088/1748-0221/](https://doi.org/10.1088/1748-0221/3/08/S08001)  
3755 [3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001). URL: <http://dx.doi.org/10.1088/1748-0221/3/08/S08001>  
3756 (cit. on p. 2).
- 3757 [13] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. “Two-real-scalar-singlet ex-  
3758 tension of the SM: LHC phenomenology and benchmark scenarios”. In: (2019). DOI:  
3759 [10.1140/epjc/s10052-020-7655-x](https://doi.org/10.1140/epjc/s10052-020-7655-x). eprint: [arXiv:1908.08554](https://arxiv.org/abs/1908.08554) (cit. on pp. 3,  
3760 36).
- 3761 [14] CMS Collaboration. “Search for nonresonant Higgs boson pair production in final  
3762 states with two bottom quarks and two photons in proton-proton collisions at  $\sqrt{s} =$   
3763 13 TeV”. In: (2021). DOI: [10.1007/JHEP03\(2021\)257](https://doi.org/10.1007/JHEP03(2021)257). eprint: [arXiv:2011.12373](https://arxiv.org/abs/2011.12373)  
3764 (cit. on p. 4).
- 3765 [15] L. Corcodilos. *2DAlphabet - Background modeling framework*. URL:  
3766 [https://github.com/CMS-H3PO/2DAlphabet/blob/](https://github.com/CMS-H3PO/2DAlphabet/blob/285680706c2f0bc92c9199803655ed0274831a1e/2DAlphabet.md)  
3767 [285680706c2f0bc92c9199803655ed0274831a1e/2DAlphabet.md](https://github.com/CMS-H3PO/2DAlphabet/blob/285680706c2f0bc92c9199803655ed0274831a1e/2DAlphabet.md) (cit. on  
3768 p. 4).

- 3769 [16] L. Corcodilos. *The 2D Alphabet background modeling method and its use in the*  
3770 *search for an excited bottom quark*. 2021. URL: [https://j10p-stage.library.](https://j10p-stage.library.jhu.edu/server/api/core/bitstreams/74b17791-7ef7-4451-a077-dff7818f82bb/content)  
3771 [jhu.edu/server/api/core/bitstreams/74b17791-7ef7-4451-a077-](https://j10p-stage.library.jhu.edu/server/api/core/bitstreams/74b17791-7ef7-4451-a077-dff7818f82bb/content)  
3772 [dff7818f82bb/content](https://j10p-stage.library.jhu.edu/server/api/core/bitstreams/74b17791-7ef7-4451-a077-dff7818f82bb/content) (cit. on p. 4).
- 3773 [17] M. J. Herrero. *The Standard Model*. 1998. eprint: [arXiv:hep-ph/9812242](https://arxiv.org/abs/hep-ph/9812242) (cit. on  
3774 p. 5).
- 3775 [18] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector  
3776 Mesons”. In: *Physical Review Letters* 13.9 (Aug. 1964), pp. 321–323. ISSN: 0031-  
3777 9007. DOI: [10.1103/physrevlett.13.321](https://doi.org/10.1103/physrevlett.13.321). URL: [http://dx.doi.org/10.1103/](http://dx.doi.org/10.1103/PhysRevLett.13.321)  
3778 [PhysRevLett.13.321](http://dx.doi.org/10.1103/PhysRevLett.13.321) (cit. on p. 11).
- 3779 [19] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Physical*  
3780 *Review Letters* 13.16 (Oct. 1964), pp. 508–509. ISSN: 0031-9007. DOI: [10.1103/](https://doi.org/10.1103/physrevlett.13.508)  
3781 [physrevlett.13.508](https://doi.org/10.1103/physrevlett.13.508). URL: <http://dx.doi.org/10.1103/PhysRevLett.13.508>  
3782 (cit. on p. 11).
- 3783 [20] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws  
3784 and Massless Particles”. In: *Physical Review Letters* 13.20 (Nov. 1964), pp. 585–  
3785 587. ISSN: 0031-9007. DOI: [10.1103/physrevlett.13.585](https://doi.org/10.1103/physrevlett.13.585). URL: [http://dx.doi.](http://dx.doi.org/10.1103/PhysRevLett.13.585)  
3786 [org/10.1103/PhysRevLett.13.585](http://dx.doi.org/10.1103/PhysRevLett.13.585) (cit. on p. 11).
- 3787 [21] Gerald S. Guralnik. “The History of the Guralnik, Hagen and Kibble development  
3788 of the Theory of Spontaneous Symmetry Breaking and Gauge Particles”. In: (2009).  
3789 DOI: [10.1142/S0217751X09045431](https://doi.org/10.1142/S0217751X09045431). eprint: [arXiv:0907.3466](https://arxiv.org/abs/0907.3466) (cit. on p. 11).
- 3790 [22] Tom Kibble. “Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism”. In: *Schol-*  
3791 *arpedia* 4.1 (2009), p. 6441. ISSN: 1941-6016. DOI: [10.4249/scholarpedia.6441](https://doi.org/10.4249/scholarpedia.6441).  
3792 URL: <http://dx.doi.org/10.4249/scholarpedia.6441> (cit. on p. 11).
- 3793 [23] CMS Collaboration. “Combined measurements of Higgs boson couplings in proton-  
3794 proton collisions at  $\sqrt{s} = 13$  TeV”. In: (2018). DOI: [10.1140/epjc/s10052-019-](https://doi.org/10.1140/epjc/s10052-019-6909-y)  
3795 [6909-y](https://doi.org/10.1140/epjc/s10052-019-6909-y). eprint: [arXiv:1809.10733](https://arxiv.org/abs/1809.10733) (cit. on p. 14).
- 3796 [24] ATLAS Collaboration. “Combined measurements of Higgs boson production and  
3797 decay using up to  $80 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV collected  
3798 with the ATLAS experiment”. In: (2019). DOI: [10.1103/PhysRevD.101.012002](https://doi.org/10.1103/PhysRevD.101.012002).  
3799 eprint: [arXiv:1909.02845](https://arxiv.org/abs/1909.02845) (cit. on p. 14).

- 3800 [25] G. Ridolfi. *Search for the Higgs boson: theoretical perspectives*. <https://cds.cern.ch/record/506136>. [Frascati Phys. Ser. 22 (2001) 291–304.] (cit. on p. 16).
- 3801
- 3802 [26] G. Abbiendi. “Search for the Standard Model Higgs Boson at LEP”. In: (2003). DOI: [10.1016/S0370-2693\(03\)00614-2](https://doi.org/10.1016/S0370-2693(03)00614-2). eprint: [arXiv:hep-ex/0306033](https://arxiv.org/abs/hep-ex/0306033) (cit. on
- 3803
- 3804 p. 17).
- 3805 [27] The CDF Collaboration et al. *Combined CDF and D0 Upper Limits on Standard*
- 3806 *Model Higgs-Boson Production with 2.1-5.4 fb<sup>-1</sup> of Data*. 2009. eprint: [arXiv:0911.3930](https://arxiv.org/abs/0911.3930) (cit. on p. 17).
- 3807
- 3808 [28] CMS Collaboration. “Measurement of the Higgs boson production rate in asso-
- 3809 ciation with top quarks in final states with electrons, muons, and hadronically
- 3810 decaying tau leptons at  $\sqrt{s} = 13$  TeV”. In: (2020). DOI: [10.1140/epjc/s10052-](https://doi.org/10.1140/epjc/s10052-021-09014-x)
- 3811 [021-09014-x](https://doi.org/10.1140/epjc/s10052-021-09014-x). eprint: [arXiv:2011.03652](https://arxiv.org/abs/2011.03652) (cit. on p. 19).
- 3812 [29] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the
- 3813 Nature of the Higgs Sector”. In: (2016). DOI: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). eprint:
- 3814 [arXiv:1610.07922](https://arxiv.org/abs/1610.07922) (cit. on pp. 20, 21, 25, 28, 30, 34, 86, 93, 125, 126, 178).
- 3815 [30] CMS Collaboration. “Measurement of the Higgs boson width and evidence of its
- 3816 off-shell contributions to ZZ production”. In: (2022). DOI: [10.1038/s41567-022-](https://doi.org/10.1038/s41567-022-01682-0)
- 3817 [01682-0](https://doi.org/10.1038/s41567-022-01682-0). eprint: [arXiv:2202.06923](https://arxiv.org/abs/2202.06923) (cit. on p. 19).
- 3818 [31] The LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross
- 3819 Sections: 3. Higgs Properties”. In: (2013). DOI: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004). eprint:
- 3820 [arXiv:1307.1347](https://arxiv.org/abs/1307.1347) (cit. on p. 25).
- 3821 [32] Massimiliano Grazzini et al. “Higgs boson pair production at NNLO with top
- 3822 quark mass effects”. In: (2018). DOI: [10.1007/JHEP05\(2018\)059](https://doi.org/10.1007/JHEP05(2018)059). eprint: [arXiv:](https://arxiv.org/abs/1803.02463)
- 3823 [1803.02463](https://arxiv.org/abs/1803.02463) (cit. on pp. 26, 125).
- 3824 [33] J. Baglio et al. “ $gg \rightarrow HH$ : Combined Uncertainties”. In: (2020). DOI: [10.1103/](https://doi.org/10.1103/PhysRevD.103.056002)
- 3825 [PhysRevD.103.056002](https://doi.org/10.1103/PhysRevD.103.056002). eprint: [arXiv:2008.11626](https://arxiv.org/abs/2008.11626) (cit. on p. 26).
- 3826 [34] B. Di Micco et al. “Higgs boson potential at colliders: status and perspectives”. In:
- 3827 (2019). DOI: [10.1016/j.revip.2020.100045](https://doi.org/10.1016/j.revip.2020.100045). eprint: [arXiv:1910.00012](https://arxiv.org/abs/1910.00012) (cit. on
- 3828 p. 27).

- 3829 [35] Florian Goertz et al. “Higgs boson pair production in the D=6 extension of the  
3830 SM”. In: (2014). DOI: [10.1007/JHEP04\(2015\)167](https://doi.org/10.1007/JHEP04(2015)167). eprint: [arXiv:1410.3471](https://arxiv.org/abs/1410.3471) (cit.  
3831 on p. 28).
- 3832 [36] Ling Liu-Sheng et al. “NNLO QCD corrections to Higgs pair production via vector  
3833 boson fusion at hadron colliders”. In: (2014). DOI: [10.1103/PhysRevD.89.073001](https://doi.org/10.1103/PhysRevD.89.073001).  
3834 eprint: [arXiv:1401.7754](https://arxiv.org/abs/1401.7754) (cit. on p. 28).
- 3835 [37] Frédéric A. Dreyer and Alexander Karlberg. “Fully differential Vector-Boson Fusion  
3836 Higgs Pair Production at Next-to-Next-to-Leading Order”. In: (2018). DOI: [10.](https://doi.org/10.1103/PhysRevD.99.074028)  
3837 [1103/PhysRevD.99.074028](https://doi.org/10.1103/PhysRevD.99.074028). eprint: [arXiv:1811.07918](https://arxiv.org/abs/1811.07918) (cit. on p. 28).
- 3838 [38] P A Zyla et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experi-*  
3839 *mental Physics* 2020.8 (Aug. 2020). ISSN: 2050-3911. DOI: [10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104).  
3840 URL: <http://dx.doi.org/10.1093/ptep/ptaa104> (cit. on pp. 30, 31).
- 3841 [39] Johannes Braathen and Shinya Kanemura. “Leading two-loop corrections to the  
3842 Higgs boson self-couplings in models with extended scalar sectors”. In: (2019). DOI:  
3843 [10.1140/epjc/s10052-020-7723-2](https://doi.org/10.1140/epjc/s10052-020-7723-2). eprint: [arXiv:1911.11507](https://arxiv.org/abs/1911.11507) (cit. on p. 30).
- 3844 [40] Stefano Di Vita et al. “A global view on the Higgs self-coupling”. In: (2017). DOI:  
3845 [10.1007/JHEP09\(2017\)069](https://doi.org/10.1007/JHEP09(2017)069). eprint: [arXiv:1704.01953](https://arxiv.org/abs/1704.01953) (cit. on p. 30).
- 3846 [41] Fedor Bezrukov and Mikhail Shaposhnikov. “Why should we care about the top  
3847 quark Yukawa coupling?” In: (2014). DOI: [10.1134/S1063776115030152](https://doi.org/10.1134/S1063776115030152). eprint:  
3848 [arXiv:1411.1923](https://arxiv.org/abs/1411.1923) (cit. on p. 32).
- 3849 [42] Da Liu, Ian Low, and Carlos E. M. Wagner. “Modification of Higgs Couplings in  
3850 Minimal Composite Models”. In: (2017). DOI: [10.1103/PhysRevD.96.035013](https://doi.org/10.1103/PhysRevD.96.035013).  
3851 eprint: [arXiv:1703.07791](https://arxiv.org/abs/1703.07791) (cit. on p. 32).
- 3852 [43] A. M. Sirunyan et al. “Combined measurements of Higgs boson couplings in pro-  
3853 ton–proton collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 79.5  
3854 (May 2019). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-6909-y](https://doi.org/10.1140/epjc/s10052-019-6909-y). URL:  
3855 <http://dx.doi.org/10.1140/epjc/s10052-019-6909-y> (cit. on p. 33).
- 3856 [44] Fady Bishara, Roberto Contino, and Juan Rojo. “Higgs pair production in vector-  
3857 boson fusion at the LHC and beyond”. In: (2016). DOI: [10.1140/epjc/s10052-](https://doi.org/10.1140/epjc/s10052-017-5037-9)  
3858 [017-5037-9](https://doi.org/10.1140/epjc/s10052-017-5037-9). eprint: [arXiv:1611.03860](https://arxiv.org/abs/1611.03860) (cit. on p. 33).

- 3859 [45] S. Dawson and H. Haber. *Electroweak Symmetry Breaking and Physics Beyond the*  
3860 *Standard Model*. 1996. eprint: [arXiv:hep-ph/9604354](https://arxiv.org/abs/hep-ph/9604354) (cit. on p. 33).
- 3861 [46] Alexandra Carvalho et al. “Higgs Pair Production: Choosing Benchmarks With  
3862 Cluster Analysis”. In: (2015). DOI: [10.1007/JHEP04\(2016\)126](https://doi.org/10.1007/JHEP04(2016)126). eprint: [arXiv:](https://arxiv.org/abs/1507.02245)  
3863 [1507.02245](https://arxiv.org/abs/1507.02245) (cit. on p. 36).
- 3864 [47] Andreas Papaefstathiou, Tania Robens, and Gilberto Tetlalmatzi-Xolocotzi.  
3865 “Triple Higgs Boson Production at the Large Hadron Collider with Two Real  
3866 Singlet Scalars”. In: (2020). DOI: [10.1007/JHEP05\(2021\)193](https://doi.org/10.1007/JHEP05(2021)193). eprint: [arXiv:](https://arxiv.org/abs/2101.00037)  
3867 [2101.00037](https://arxiv.org/abs/2101.00037) (cit. on pp. 37, 142).
- 3868 [48] E. Mobs. *The CERN accelerator complex*. 2018. eprint: [https://cds.cern.ch/](https://cds.cern.ch/record/2636343)  
3869 [record/2636343](https://cds.cern.ch/record/2636343) (cit. on p. 41).
- 3870 [49] AC Team. *Diagram of an LHC dipole magnet*. 1999. URL: [https://cds.cern.ch/](https://cds.cern.ch/record/40524)  
3871 [record/40524](https://cds.cern.ch/record/40524) (cit. on p. 42).
- 3872 [50] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron  
3873 Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. ISSN:  
3874 1748-0221. DOI: [10.1088/1748-0221/3/08/s08003](https://doi.org/10.1088/1748-0221/3/08/s08003). URL: [http://dx.doi.org/](http://dx.doi.org/10.1088/1748-0221/3/08/S08003)  
3875 [10.1088/1748-0221/3/08/S08003](http://dx.doi.org/10.1088/1748-0221/3/08/S08003) (cit. on p. 40).
- 3876 [51] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal*  
3877 *of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. ISSN: 1748-0221. DOI: [10.](https://doi.org/10.1088/1748-0221/3/08/s08004)  
3878 [1088/1748-0221/3/08/s08004](https://doi.org/10.1088/1748-0221/3/08/s08004). URL: [http://dx.doi.org/10.1088/1748-](http://dx.doi.org/10.1088/1748-0221/3/08/S08004)  
3879 [0221/3/08/S08004](http://dx.doi.org/10.1088/1748-0221/3/08/S08004) (cit. on pp. 40, 45).
- 3880 [52] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instru-*  
3881 *mentation* 3.08 (Aug. 2008), S08005–S08005. ISSN: 1748-0221. DOI: [10.1088/1748-](https://doi.org/10.1088/1748-0221/3/08/s08005)  
3882 [0221/3/08/s08005](https://doi.org/10.1088/1748-0221/3/08/s08005). URL: [http://dx.doi.org/10.1088/1748-](http://dx.doi.org/10.1088/1748-0221/3/08/S08005)  
3883 [0221/3/08/S08005](http://dx.doi.org/10.1088/1748-0221/3/08/S08005) (cit. on p. 40).
- 3884 [53] The ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Jour-*  
3885 *nal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. ISSN: 1748-0221. DOI:  
3886 [10.1088/1748-0221/3/08/s08002](https://doi.org/10.1088/1748-0221/3/08/s08002). URL: [http://dx.doi.org/10.1088/1748-](http://dx.doi.org/10.1088/1748-0221/3/08/S08002)  
3887 [0221/3/08/S08002](http://dx.doi.org/10.1088/1748-0221/3/08/S08002) (cit. on p. 41).
- 3888 [54] Oliver Sim Brüning et al. *LHC Design Report*. en. 2004. DOI: [10.5170/CERN-](https://doi.org/10.5170/CERN-2004-003-V-1)  
3889 [2004-003-V-1](https://doi.org/10.5170/CERN-2004-003-V-1). URL: <http://cds.cern.ch/record/782076> (cit. on p. 43).

- 3890 [55] CMS Collaboration. “Measurement of the inelastic proton-proton cross section at  
3891  $\sqrt{s} = 13$  TeV”. In: (2018). DOI: [10.1007/JHEP07\(2018\)161](https://doi.org/10.1007/JHEP07(2018)161). eprint: [arXiv:  
3892 1802.02613](https://arxiv.org/abs/1802.02613) (cit. on p. 44).
- 3893 [56] CMS Collaboration. *Public luminosity results*. 2024 (last visited August 2024). URL:  
3894 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (cit.  
3895 on p. 44).
- 3896 [57] T. Sakuma. *Cutaway diagrams of CMS detector*. 2019. URL: [https://cds.cern.  
3897 ch/record/2665537](https://cds.cern.ch/record/2665537) (cit. on p. 46).
- 3898 [58] I. Neutelings. *CMS Wiki Pages, How to draw diagrams in LaTeX with TikZ*. 2022.  
3899 URL: <https://wiki.physik.uzh.ch/cms/latex:tikz> (cit. on p. 47).
- 3900 [59] CMS Collaboration et al. *The CMS tracker system project: Technical Design Re-*  
3901 *port*. 1997. URL: <https://cds.cern.ch/record/368412> (cit. on p. 48).
- 3902 [60] CMS Collaboration. *CMS Tracker Detector Performance Results*. 2024 (last visited  
3903 August 2024). URL: [https://twiki.cern.ch/twiki/bin/view/CMSPublic/  
3904 DPGResultsTRK](https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK) (cit. on p. 49).
- 3905 [61] Dinko Ferenčec, Matej Roguljić, and Andrey Starodumov. “Production, Calibra-  
3906 tion, and Performance of the Layer 1 Replacement Modules for the CMS Pixel  
3907 Detector”. In: *Proceedings of the 29th International Workshop on Vertex Detectors*  
3908 *(VERTEX2020)*. Journal of the Physical Society of Japan, 2021. DOI: [10.7566/  
3909 jpscp.34.010023](https://doi.org/10.7566/JPSCP.34.010023). URL: <http://dx.doi.org/10.7566/JPSCP.34.010023> (cit. on  
3910 p. 50).
- 3911 [62] CMS Collaboration. *CMS workflow*. URL: <https://cms.cern/detector> (cit. on  
3912 p. 50).
- 3913 [63] D. J. A. Cockerill. *The CMS Electromagnetic Calorimeter at the LHC*. 2008. eprint:  
3914 [arXiv:0810.0381](https://arxiv.org/abs/0810.0381) (cit. on p. 50).
- 3915 [64] D. Abadjiev et al. “Autoencoder-Based Anomaly Detection System for Online Data  
3916 Quality Monitoring of the CMS Electromagnetic Calorimeter”. In: *Computing and*  
3917 *Software for Big Science* 8.1 (2024). ISSN: 2510-2044. DOI: [10.1007/s41781-024-  
3918 00118-z](https://doi.org/10.1007/s41781-024-00118-z). URL: <http://dx.doi.org/10.1007/s41781-024-00118-z> (cit. on  
3919 p. 51).

- 3920 [65] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Re-*  
3921 *port*. 1997. URL: <https://cds.cern.ch/record/357153> (cit. on p. 52).
- 3922 [66] CMS Collaboration. “Calibration of the CMS hadron calorimeters using proton-
- 3923 proton collision data at  $\sqrt{s} = 13$  TeV”. In: (2019). DOI: [10.1088/1748-0221/15/](https://doi.org/10.1088/1748-0221/15/05/P05002)  
3924 [05/P05002](https://doi.org/10.1088/1748-0221/15/05/P05002). eprint: [arXiv:1910.00079](https://arxiv.org/abs/1910.00079) (cit. on p. 52).
- 3925 [67] CMS Collaboration. “Performance of the CMS muon detector and muon recon-
- 3926 struction with proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: (2018). DOI: [10.](https://doi.org/10.1088/1748-0221/13/06/P06015)  
3927 [1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015). eprint: [arXiv:1804.04528](https://arxiv.org/abs/1804.04528) (cit. on pp. 54, 55,  
3928 67).
- 3929 [68] CMS Collaboration and C. Seez. *The CMS trigger system*. 2003. URL: [https :](https://cds.cern.ch/record/687840/files/cr03_008.pdf)  
3930 [//cds.cern.ch/record/687840/files/cr03\\_008.pdf](https://cds.cern.ch/record/687840/files/cr03_008.pdf) (cit. on p. 56).
- 3931 [69] CMS Collaboration, Darin Acosta, and Austin Ball. *CMS Technical Design Report*  
3932 *for the Level-1 Trigger Upgrade*. 2013. URL: [https://cds.cern.ch/record/](https://cds.cern.ch/record/1556311)  
3933 [1556311](https://cds.cern.ch/record/1556311) (cit. on p. 58).
- 3934 [70] Katharina Danziger, Stefan Höche, and Frank Siegert. *Reducing negative weights*  
3935 *in Monte Carlo event generation with Sherpa*. 2021. eprint: [arXiv:2110.15211](https://arxiv.org/abs/2110.15211)  
3936 (cit. on p. 60).
- 3937 [71] Paolo Nason. “A New Method for Combining NLO QCD with Shower Monte Carlo
- 3938 Algorithms”. In: (2004). DOI: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040). eprint: [arXiv:](https://arxiv.org/abs/hep-ph/0409146)  
3939 [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146) (cit. on p. 60).
- 3940 [72] Johan Alwall et al. “MadGraph 5 : Going Beyond”. In: (2011). DOI: [10.1007/](https://doi.org/10.1007/JHEP06(2011)128)  
3941 [JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128). eprint: [arXiv:1106.0522](https://arxiv.org/abs/1106.0522) (cit. on p. 60).
- 3942 [73] Simone Alioli et al. “A general framework for implementing NLO calculations in
- 3943 shower Monte Carlo programs: the POWHEG BOX”. In: (2010). DOI: [10.1007/](https://doi.org/10.1007/JHEP06(2010)043)  
3944 [JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043). eprint: [arXiv:1002.2581](https://arxiv.org/abs/1002.2581) (cit. on pp. 60, 91, 178).
- 3945 [74] Carlo Oleari. “The POWHEG-BOX”. In: (2010). DOI: [10.1016/j.nuclphysbps.](https://doi.org/10.1016/j.nuclphysbps.2010.08.016)  
3946 [2010.08.016](https://doi.org/10.1016/j.nuclphysbps.2010.08.016). eprint: [arXiv:1007.3893](https://arxiv.org/abs/1007.3893) (cit. on pp. 60, 178).
- 3947 [75] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD compu-
- 3948 tations with Parton Shower simulations: the POWHEG method”. In: (2007). DOI:  
3949 [10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070). eprint: [arXiv:0709.2092](https://arxiv.org/abs/0709.2092) (cit. on p. 60).

- 3950 [76] Torbjörn Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: (2014). DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). eprint: [arXiv:1410.3012](https://arxiv.org/abs/1410.3012) (cit. on pp. 60, 91).
- 3951
- 3952 [77] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and*  
3953 *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors*  
3954 *and Associated Equipment* 506.3 (July 2003), pp. 250–303. ISSN: 0168-9002. DOI:  
3955 [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: [http://dx.doi.org/10.1016/S0168-](http://dx.doi.org/10.1016/S0168-9002(03)01368-8)  
3956 [9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8) (cit. on pp. 60, 91).
- 3957 [78] CMS Collaboration. *CMSSW Application Framework*. URL: [https://twiki.cern.](https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework)  
3958 [ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework](https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework) (cit. on p. 60).
- 3959 [79] Siona Ruth Davis. *Interactive Slice of the CMS detector*. 2016. URL: [https://cds.](https://cds.cern.ch/record/2205172/)  
3960 [cern.ch/record/2205172/](https://cds.cern.ch/record/2205172/) (cit. on p. 63).
- 3961 [80] CMS Collaboration. “Description and performance of track and primary-vertex  
3962 reconstruction with the CMS tracker”. In: (2014). DOI: [10.1088/1748-0221/9/](https://doi.org/10.1088/1748-0221/9/10/P10009)  
3963 [10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009). eprint: [arXiv:1405.6569](https://arxiv.org/abs/1405.6569) (cit. on pp. 63, 65).
- 3964 [81] T. Mc Cauley. *Collisions recorded by the CMS detector on 14 Oct 2016 during*  
3965 *the high pile-up fill*. 2016. URL: <https://cds.cern.ch/record/2231915> (cit. on  
3966 p. 64).
- 3967 [82] P. Billoir and S. Qian. “Simultaneous pattern recognition and track fitting by  
3968 the Kalman filtering method”. In: *Nuclear Instruments and Methods in Physics*  
3969 *Research Section A: Accelerators, Spectrometers, Detectors and Associated Equip-*  
3970 *ment* 294.1–2 (Sept. 1990), pp. 219–228. ISSN: 0168-9002. DOI: [10.1016/0168-](https://doi.org/10.1016/0168-9002(90)91835-y)  
3971 [9002\(90\)91835-y](https://doi.org/10.1016/0168-9002(90)91835-y). URL: [http://dx.doi.org/10.1016/0168-9002\(90\)91835-Y](http://dx.doi.org/10.1016/0168-9002(90)91835-Y)  
3972 (cit. on p. 64).
- 3973 [83] CMS Collaboration. “Performance of the CMS missing transverse energy recon-  
3974 struction in pp data at  $\sqrt{s} = 8$  TeV”. In: (2014). DOI: [10.1088/1748-0221/](https://doi.org/10.1088/1748-0221/10/02/P02006)  
3975 [10/02/P02006](https://doi.org/10.1088/1748-0221/10/02/P02006). eprint: [arXiv:1411.0511](https://arxiv.org/abs/1411.0511) (cit. on p. 68).
- 3976 [84] CMS Collaboration. “Missing transverse energy performance of the CMS detector”.  
3977 In: (2011). DOI: [10.1088/1748-0221/6/09/P09001](https://doi.org/10.1088/1748-0221/6/09/P09001). eprint: [arXiv:1106.5048](https://arxiv.org/abs/1106.5048)  
3978 (cit. on p. 68).



- 3979 [85] CMS Collaboration. “Electron and photon reconstruction and identification with  
3980 the CMS experiment at the CERN LHC”. In: (2020). DOI: [10.1088/1748-0221/  
3981 16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014). eprint: [arXiv:2012.06888](https://arxiv.org/abs/2012.06888) (cit. on pp. 69, 71, 72).
- 3982 [86] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- $k_t$  jet clustering  
3983 algorithm”. In: (2008). DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). eprint: [arXiv:  
3984 0802.1189](https://arxiv.org/abs/0802.1189) (cit. on p. 72).
- 3985 [87] Gavin P. Salam. “Towards Jetography”. In: (2009). DOI: [10.1140/epjc/s10052-  
3986 010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). eprint: [arXiv:0906.1833](https://arxiv.org/abs/0906.1833) (cit. on p. 72).
- 3987 [88] CMS Collaboration. *Pileup Removal Algorithms*. 2014. URL: [https://cds.cern.  
3988 ch/record/1751454/](https://cds.cern.ch/record/1751454/) (cit. on p. 73).
- 3989 [89] Daniele Bertolini et al. “Pileup Per Particle Identification”. In: (2014). DOI: [10.  
3990 1007/JHEP10\(2014\)059](https://doi.org/10.1007/JHEP10(2014)059). eprint: [arXiv:1407.6013](https://arxiv.org/abs/1407.6013) (cit. on p. 74).
- 3991 [90] Garvita Agarwal. “Jet Energy Scale and Resolution Measurements in CMS”. In:  
3992 (2023). DOI: [10.22323/1.414.0652](https://doi.org/10.22323/1.414.0652). eprint: [arXiv:2301.02175](https://arxiv.org/abs/2301.02175) (cit. on p. 76).
- 3993 [91] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in  
3994 pp collisions at 13 TeV”. In: (2017). DOI: [10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011).  
3995 eprint: [arXiv:1712.07158](https://arxiv.org/abs/1712.07158) (cit. on pp. 77, 95, 127).
- 3996 [92] CMS Collaboration. “Identification of b-quark jets with the CMS experiment”. In:  
3997 (2012). DOI: [10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013). eprint: [arXiv:1211.4462](https://arxiv.org/abs/1211.4462) (cit. on  
3998 p. 78).
- 3999 [93] Daniel Guest et al. “Jet Flavor Classification in High-Energy Physics with Deep  
4000 Neural Networks”. In: (2016). DOI: [10.1103/PhysRevD.94.112002](https://doi.org/10.1103/PhysRevD.94.112002). eprint: [arXiv:  
4001 1607.08633](https://arxiv.org/abs/1607.08633) (cit. on p. 78).
- 4002 [94] Emil Bols et al. “Jet Flavour Classification Using DeepJet”. In: (2020). DOI: [10.  
4003 1088/1748-0221/15/12/P12012](https://doi.org/10.1088/1748-0221/15/12/P12012). eprint: [arXiv:2008.10519](https://arxiv.org/abs/2008.10519) (cit. on p. 78).
- 4004 [95] CMS Collaboration. *Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector*. 2018. URL: [https://cds.cern.ch/  
4005 record/2627468](https://cds.cern.ch/record/2627468) (cit. on p. 80).
- 4007 [96] Mrinal Dasgupta et al. “Towards an understanding of jet substructure”. In: (2013).  
4008 DOI: [10.1007/JHEP09\(2013\)029](https://doi.org/10.1007/JHEP09(2013)029). eprint: [arXiv:1307.0007](https://arxiv.org/abs/1307.0007) (cit. on p. 81).

- 4009 [97] Jesse Thaler and Ken Van Tilburg. “Identifying Boosted Objects with N-  
4010 subjettiness”. In: (2010). DOI: [10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015). eprint: [arXiv:1011.2268](https://arxiv.org/abs/1011.2268) (cit. on p. 82).
- 4012 [98] CMS Collaboration. “Identification of heavy, energetic, hadronically decaying par-  
4013 ticles using machine-learning techniques”. In: (2020). DOI: [10.1088/1748-0221/15/06/P06005](https://doi.org/10.1088/1748-0221/15/06/P06005). eprint: [arXiv:2004.08262](https://arxiv.org/abs/2004.08262) (cit. on p. 83).
- 4015 [99] Huilin Qu and Loukas Gouskos. “Jet tagging via particle clouds”. In: *Physical*  
4016 *Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: [10.1103/PhysRevD.101.056019](https://doi.org/10.1103/PhysRevD.101.056019). URL: <http://dx.doi.org/10.1103/PhysRevD.101.056019> (cit. on  
4017 p. 83).
- 4019 [100] Nilanjana Kumar and Stephen P. Martin. “LHC search for di-Higgs decays of  
4020 stoponium and other scalars in events with two photons and two bottom jets”. In:  
4021 (2014). DOI: [10.1103/PhysRevD.90.055007](https://doi.org/10.1103/PhysRevD.90.055007). eprint: [arXiv:1404.0996](https://arxiv.org/abs/1404.0996) (cit. on  
4022 p. 88).
- 4023 [101] The CMS Collaboration. “Measurement of the Inclusive W and Z Production Cross  
4024 Sections in pp Collisions at  $\sqrt{s} = 7$  TeV”. In: (2011). DOI: [10.1007/JHEP10\(2011\)](https://doi.org/10.1007/JHEP10(2011)132)  
4025 [132](https://doi.org/10.1007/JHEP10(2011)132). eprint: [arXiv:1107.4789](https://arxiv.org/abs/1107.4789) (cit. on p. 90).
- 4026 [102] J. Alwall et al. “The automated computation of tree-level and next-to-leading order  
4027 differential cross sections, and their matching to parton shower simulations”. In:  
4028 (2014). DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). eprint: [arXiv:1405.0301](https://arxiv.org/abs/1405.0301) (cit. on pp. 91,  
4029 178).
- 4030 [103] CMS Collaboration. “Extraction and validation of a new set of CMS PYTHIA8  
4031 tunes from underlying-event measurements”. In: (2019). DOI: [10.1140/epjc/](https://doi.org/10.1140/epjc/s10052-019-7499-4)  
4032 [s10052-019-7499-4](https://doi.org/10.1140/epjc/s10052-019-7499-4). eprint: [arXiv:1903.12179](https://arxiv.org/abs/1903.12179) (cit. on p. 91).
- 4033 [104] CMS Collaboration. “Event generator tunes obtained from underlying event and  
4034 multiparton scattering measurements”. In: (2015). DOI: [10.1140/epjc/s10052-](https://doi.org/10.1140/epjc/s10052-016-3988-x)  
4035 [016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x). eprint: [arXiv:1512.00815](https://arxiv.org/abs/1512.00815) (cit. on p. 91).
- 4036 [105] The NNPDF Collaboration et al. “Parton distributions for the LHC Run II”. In:  
4037 (2014). DOI: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040). eprint: [arXiv:1410.8849](https://arxiv.org/abs/1410.8849) (cit. on p. 91).

- 4038 [106] The NNPDF Collaboration et al. “Parton distributions from high-precision collider  
4039 data”. In: (2017). DOI: [10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5). eprint: [arXiv:  
4040 1706.00428](https://arxiv.org/abs/1706.00428) (cit. on p. 91).
- 4041 [107] G. Buchalla et al. “Higgs boson pair production in non-linear Effective Field Theory  
4042 with full  $m_t$ -dependence at NLO QCD”. In: (2018). DOI: [10.1007/JHEP09\(2018\)  
4043 057](https://doi.org/10.1007/JHEP09(2018)057). eprint: [arXiv:1806.05162](https://arxiv.org/abs/1806.05162) (cit. on p. 91).
- 4044 [108] CMS Collaboration. “A deep neural network for simultaneous estimation of b jet  
4045 energy and resolution”. In: (2019). DOI: [10.1007/s41781-020-00041-z](https://doi.org/10.1007/s41781-020-00041-z). eprint:  
4046 [arXiv:1912.06046](https://arxiv.org/abs/1912.06046) (cit. on p. 95).
- 4047 [109] John C. Collins and Davison E. Soper. “Angular distribution of dileptons in high-  
4048 energy hadron collisions”. In: *Physical Review D* 16.7 (Oct. 1977), pp. 2219–2225.  
4049 ISSN: 0556-2821. DOI: [10.1103/physrevd.16.2219](https://doi.org/10.1103/physrevd.16.2219). URL: [http://dx.doi.org/  
4050 10.1103/PhysRevD.16.2219](http://dx.doi.org/10.1103/PhysRevD.16.2219) (cit. on p. 102).
- 4051 [110] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. 2017. eprint:  
4052 <https://cds.cern.ch/record/2256875> (cit. on pp. 109, 127).
- 4053 [111] ATLAS Collaboration. “Search for the  $HH \rightarrow b\bar{b}b\bar{b}$  process via vector-boson fu-  
4054 sion production using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS  
4055 detector”. In: (2020). DOI: [10.1007/JHEP01\(2021\)145](https://doi.org/10.1007/JHEP01(2021)145). eprint: [arXiv:2001.05178  
4056](https://arxiv.org/abs/2001.05178) (cit. on pp. 111, 130).
- 4057 [112] The ATLAS, CMS Collaborations, and The LHC Higgs Combination Group. *Pro-  
4058 cedure for the LHC Higgs boson search combination in Summer 2011*. 2011. eprint:  
4059 [TechnicalReportCMS-NOTE-2011-005,ATL-PHYS-PUB-2011-11](https://arxiv.org/abs/TechnicalReportCMS-NOTE-2011-005,ATL-PHYS-PUB-2011-11) (cit. on p. 116).
- 4060 [113] Thomas Junk. “Confidence Level Computation for Combining Searches with Small  
4061 Statistics”. In: (1999). DOI: [10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2). eprint: [arXiv:  
4062 hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006) (cit. on p. 118).
- 4063 [114] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”.  
4064 In: (2010). DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). eprint: [arXiv:1007.1727  
4065](https://arxiv.org/abs/1007.1727) (cit. on p. 118).
- 4066 [115] P. D. Dauncey et al. “Handling uncertainties in background shapes: the discrete  
4067 profiling method”. In: (2014). DOI: [10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015). eprint:  
4068 [arXiv:1408.6865](https://arxiv.org/abs/1408.6865) (cit. on pp. 122–124).

- 4069 [116] Frédéric A. Dreyer and Alexander Karlberg. “Vector-Boson Fusion Higgs Pair Pro-  
4070 duction at N<sup>3</sup>LO”. In: (2018). DOI: [10.1103/PhysRevD.98.114016](https://doi.org/10.1103/PhysRevD.98.114016). eprint: [arXiv:  
4071 1811.07906](https://arxiv.org/abs/1811.07906) (cit. on p. 125).
- 4072 [117] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: (2015).  
4073 DOI: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001). eprint: [arXiv:1510.03865](https://arxiv.org/abs/1510.03865) (cit. on  
4074 p. 125).
- 4075 [118] Stefano Carrazza et al. “An unbiased Hessian representation for Monte Carlo  
4076 PDFs”. In: *The European Physical Journal C* 75.8 (Aug. 2015). ISSN: 1434-6052.  
4077 DOI: [10.1140/epjc/s10052-015-3590-7](https://doi.org/10.1140/epjc/s10052-015-3590-7). URL: [http://dx.doi.org/10.1140/  
4078 epjc/s10052-015-3590-7](http://dx.doi.org/10.1140/epjc/s10052-015-3590-7) (cit. on p. 126).
- 4079 [119] Barbara Jäger et al. “Parton-shower effects in Higgs production via Vector-Boson  
4080 Fusion”. In: (2020). DOI: [10.1140/epjc/s10052-020-8326-7](https://doi.org/10.1140/epjc/s10052-020-8326-7). eprint: [arXiv:  
4081 2003.12435](https://arxiv.org/abs/2003.12435) (cit. on p. 126).
- 4082 [120] CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking  
4083 Period*. 2017. eprint: [TechnicalReportCMS-PAS-LUM-17-001](https://arxiv.org/abs/1706.02668) (cit. on p. 128).
- 4084 [121] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period  
4085 at  $\sqrt{s} = 13$  TeV*. 2018. eprint: [TechnicalReportCMS-PAS-LUM-17-004](https://arxiv.org/abs/1806.02668) (cit. on  
4086 p. 128).
- 4087 [122] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period  
4088 at  $\sqrt{s} = 13$  TeV*. 2019. eprint: [TechnicalReportCMS-PAS-LUM-18-002](https://arxiv.org/abs/1906.02668) (cit. on  
4089 p. 128).
- 4090 [123] ATLAS Collaboration. “Combination of searches for Higgs boson pairs in  $pp$  col-  
4091 lisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: (2019). DOI: [10.1016/j.  
4092 physletb.2019.135103](https://doi.org/10.1016/j.physletb.2019.135103). eprint: [arXiv:1906.02025](https://arxiv.org/abs/1906.02025) (cit. on pp. 130, 135, 140).
- 4093 [124] A. M. Sirunyan et al. “Measurements of  $t\bar{t}H$  Production and the CP Structure of  
4094 the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton  
4095 Decay Channel”. In: *Physical Review Letters* 125.6 (Aug. 2020). ISSN: 1079-7114.  
4096 DOI: [10.1103/physrevlett.125.061801](https://doi.org/10.1103/physrevlett.125.061801). URL: [http://dx.doi.org/10.1103/  
4097 PhysRevLett.125.061801](http://dx.doi.org/10.1103/PhysRevLett.125.061801) (cit. on p. 138).

- 4098 [125] CMS Collaboration. “Combination of searches for Higgs boson pair production in  
4099 proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: (2018). DOI: [10.1103/PhysRevLett.](https://doi.org/10.1103/PhysRevLett.122.121803)  
4100 [122.121803](https://doi.org/10.1103/PhysRevLett.122.121803). eprint: [arXiv:1811.09689](https://arxiv.org/abs/1811.09689) (cit. on p. 140).
- 4101 [126] Ulrich Ellwanger, Cyril Hugonie, and Ana M. Teixeira. “The Next-to-Minimal Su-  
4102 persymmetric Standard Model”. In: *Physics Reports* 496.1–2 (Nov. 2010), pp. 1–77.  
4103 ISSN: 0370-1573. DOI: [10.1016/j.physrep.2010.07.001](https://doi.org/10.1016/j.physrep.2010.07.001). URL: [http://dx.doi.](http://dx.doi.org/10.1016/j.physrep.2010.07.001)  
4104 [org/10.1016/j.physrep.2010.07.001](http://dx.doi.org/10.1016/j.physrep.2010.07.001) (cit. on p. 142).
- 4105 [127] Alexandra Carvalho. *Gravity particles from Warped Extra Dimensions, predictions*  
4106 *for LHC*. 2014. eprint: [arXiv:1404.0102](https://arxiv.org/abs/1404.0102) (cit. on p. 142).
- 4107 [128] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”.  
4108 In: (2012). DOI: [10.1007/JHEP12\(2012\)061](https://doi.org/10.1007/JHEP12(2012)061). eprint: [arXiv:1209.6215](https://arxiv.org/abs/1209.6215) (cit. on  
4109 p. 178).
- 4110 [129] CERN. *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs*  
4111 *cross sections: 4. Deciphering the nature of the Higgs sector*. en. 2017. DOI: [10.](https://doi.org/10.23731/CYRM-2017-002)  
4112 [23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). URL: [https://e-publishing.cern.ch/index.php/](https://e-publishing.cern.ch/index.php/CYRM/issue/view/32)  
4113 [CYRM/issue/view/32](https://e-publishing.cern.ch/index.php/CYRM/issue/view/32) (cit. on p. 179).
- 4114 [130] A. Tumasyan et al. “Measurement of simplified template cross sections of the Higgs  
4115 boson produced in association with W or Z bosons in the  $H \rightarrow b\bar{b}$  decay channel  
4116 in proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *Physical Review D* 109.9 (2024).  
4117 ISSN: 2470-0029. DOI: [10.1103/physrevd.109.092011](https://doi.org/10.1103/physrevd.109.092011). URL: [http://dx.doi.](http://dx.doi.org/10.1103/PhysRevD.109.092011)  
4118 [org/10.1103/PhysRevD.109.092011](http://dx.doi.org/10.1103/PhysRevD.109.092011) (cit. on p. 180).
- 4119 [131] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous*  
4120 *Distributed Systems*. 2016. eprint: [arXiv:1603.04467](https://arxiv.org/abs/1603.04467) (cit. on p. 184).
- 4121 [132] CERN. *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs*  
4122 *cross sections: 4. Deciphering the nature of the Higgs sector*. en. 2017. DOI: [10.](https://doi.org/10.23731/CYRM-2017-002)  
4123 [23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). URL: [https://e-publishing.cern.ch/index.php/](https://e-publishing.cern.ch/index.php/CYRM/issue/view/32)  
4124 [CYRM/issue/view/32](https://e-publishing.cern.ch/index.php/CYRM/issue/view/32) (cit. on p. 186).

# 4125 Appendix A

## 4126 Search for VHbb

4127 Higgs boson decaying into a pair of b quarks ( $H \rightarrow b\bar{b}$ ) has a branching ratio of 58% among  
4128 all Higgs decays. The Higgs boson has 4 major production modes, namely gluon-gluon  
4129 fusion (ggF), vector-boson fusion (VBF), associated production with a vector boson (VH),  
4130 and associated production with a pair of top quarks (ttH) as explained in 2.2. The VH  
4131 production processes represent the most sensitive production mode for the reconstruction  
4132 of  $H \rightarrow b\bar{b}$  decay. The cross-section of these production modes are not the largest among  
4133 the Higgs production processes, but due to the leptonic decay modes of vector bosons  
4134  $Z \rightarrow \nu\nu$ ,  $W \rightarrow l\nu$  and  $Z \rightarrow ll$  the triggering and background rejection is more efficient.  
4135 The Higgs boson can be reconstructed from a pair of b-jets which are identified using b-  
4136 tagging algorithms. The main challenges in measuring  $VH \rightarrow b\bar{b}$  come from background  
4137 modeling, efficiency in tagging b-jets and measuring its momentum and energy resolution.

### 4138 A.1 Signal and background processes

#### 4139 A.1.1 Signal

4140 The Feynman diagrams for the signal process are shown in Fig. A.1.

4141 The vector boson in the quark-induced VH process can both be a Z and a W boson.

4142 The gluon-induced process is at this order only possible for Z bosons via fermion loops.

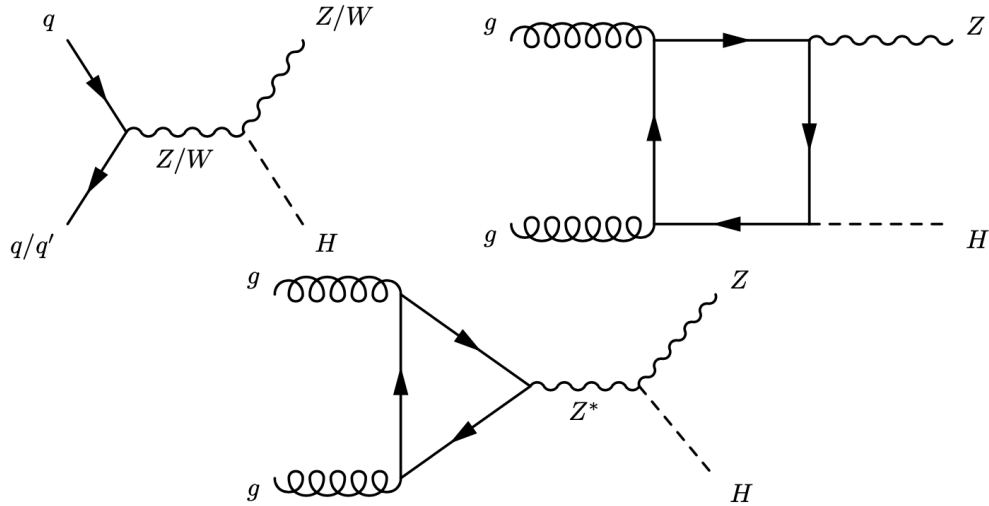


Figure A.1: VH (top left) and  $ggZH$  (top right and bottom) process Feynman diagrams.

4143 Through the box-type diagram in figure 4.2 top right, the  $ggZH$  process is sensitive also  
 4144 to the coupling to top quarks, which are the dominating fermion contribution in the loop.  
 4145 Following the different decay modes of the vector boson, the analysis is divided into three  
 4146 channels

- 4147 • 0-lepton ( $Z\nu\nu$ ): Z boson decays to two neutrinos
- 4148 • 1-lepton ( $Wl\nu$ ): W boson decays to lepton and neutrino
- 4149 • 2-lepton ( $Zll$ ): Z boson decays to two leptons (opposite charge, same flavor)

4150 The 1- and 2-lepton channel are further divided into electron and muon channels. Taus  
 4151 are not explicitly reconstructed, but due to the 35% leptonic decay branching ratio, a  
 4152 fraction of events will end up in the electron or muon channels.

## 4153 A.1.2 Background processes

4154 Final states of  $VHbb$  process involves 2 b-jets and 2 leptons (excluding taus). Based on  
 4155 these final states, following processes have similar final state signature and are accounted  
 4156 to be major backgrounds in this analysis:

- 4157 • **V+jets:** Two quarks or gluons produce a vector boson and radiate off a gluon which  
 4158 can produce a  $b\bar{b}$  quark pair, which resembles the  $b\bar{b}$  pair from the Higgs decay. It has

4159 a falling distribution in  $m_{jj}$  since the jets originate from a mass-less gluon instead  
4160 of the Higgs boson. In general the V+jets background is divided into three flavors:  
4161 V+light jets, V+c jets and V+b jets based on counting B- and D-hadrons above  
4162 25GeV within detector acceptance ( $|\eta| < 2.6$ ). If multiple hadrons of different flavor  
4163 are present, the flavor is defined by the heaviest quark.

4164 • **Top quark production:** Most of the times a top quark decays to a W-boson and  
4165 a b-quark. The consequent decay mode of the W boson defines how this background  
4166 contributes. The  $t\bar{t}$  production with hadronically decaying W bosons contributes to  
4167 the 0-lepton channel, but the additional jets activity is higher than in VH produc-  
4168 tion. If one of the W-bosons decays leptonically the final state is similar to 1-lepton  
4169 channel. The final states with both of the W-bosons decaying to leptons contribute  
4170 to the 2-lepton channel.

4171 The single-top processes are manifested similarly to the  $t\bar{t}$  production, but the kine-  
4172 matics is closer to the signal process, which makes it harder to suppress despite the  
4173 relatively low production cross-section.

4174 • **Diboson:** The diboson processes WZ and ZZ can produce the same final state as  
4175 the VH process, when a Z-boson decays to  $b\bar{b}$  and the other vector boson follows the  
4176 leptonic decay mode. The main observable that helps reducing this background is  
4177 the invariant mass of b-quark pairs, which is peaked around the Z-boson mass.

4178 • **QCD:** The QCD events are abundant at the LHC and the b-quark pairs can be  
4179 easily produced from the QCD interaction. If other particles in the event are mis-  
4180 reconstructed, the QCD processes can contribute to all channels in the analysis.

## 4181 A.2 Analysis strategy

4182 The general strategy of the analysis is to determine a signal strength modifier  $\mu$  from  
4183 the observed data by a simultaneous fit of signal and background templates in signal and  
4184 control regions. The signal region is selected to have a high signal efficiency and the control  
4185 regions to have a good purity in the individual backgrounds. The templates are derived  
4186 from a Monte Carlo (MC) simulation. All variations due to detector and calibration effects



4187 are included, either with detailed shape information or as multiplicative normalization  
 4188 factors. The strengths of each source of variation is steered by nuisance parameters in  
 4189 the likelihood function. The signal strength modifier  $\mu$  is implemented in the fit as a free-  
 4190 floating (unconstrained) parameter. Furthermore, for important background processes,  
 4191 normalization factors are considered as nuisance parameters in the fit and are set as  
 4192 free parameters so that the data points from collision reduce the impacts of modeling  
 4193 uncertainties in simulations.

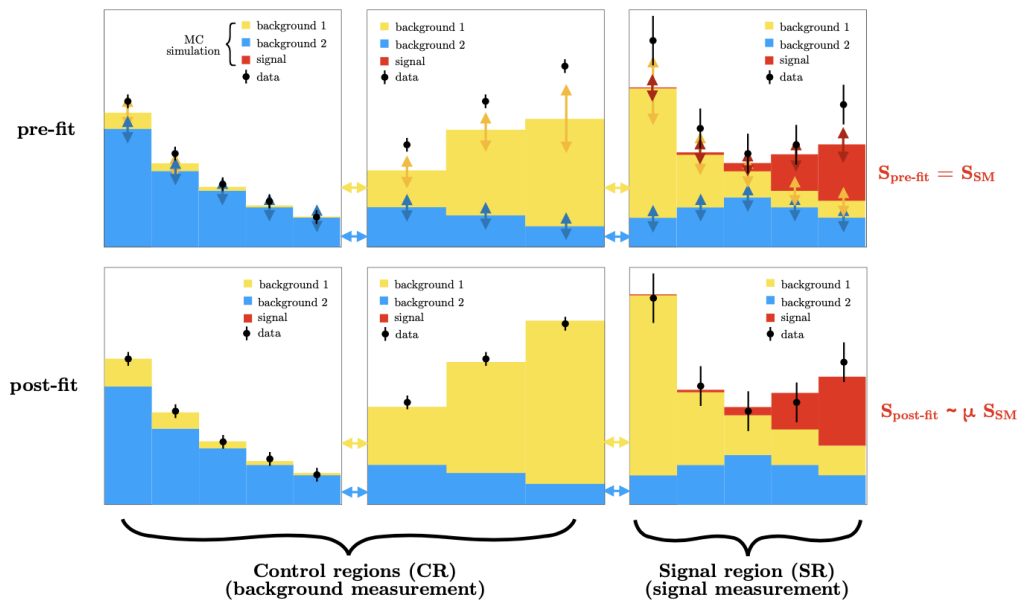


Figure A.2: Simplified schematic of the analysis strategy using a simultaneous fit of templates derived from MC simulations.

4194 To demonstrate this strategy, an example with two background processes and three regions  
 4195 of phase space is shown in Fig. A.2, the collision data and simulations pass through a  
 4196 selection criterion, to create separate templates for the signal and control regions. Then  
 4197 these templates are simultaneously fitted to the data to extract the signal strength  $\mu$ ,  
 4198 and the normalization factors for each background contributions. The top row in Fig.  
 4199 A.2 shows the signal and background predictions before the fit has been run (pre-fit). It  
 4200 shows some discrepancies between prediction and observation in this example, mostly in  
 4201 the middle column, where "background 1" dominates. Such discrepancies can originate  
 4202 from limitations in the simulation. After the fit (post-fit), the results in the bottom row  
 4203 show the discrepancy has been removed due to the maximum likelihood parameters found  
 4204 by the fit, in this example scaling up "background 1".

4205 The post-fit signal strength modifier  $\mu$  is called the parameter of interest (POI) and the  
4206 most-important result of the fit. It denotes the ratio of the observed and the predicted  
4207 (Standard Model) cross-sections of the signal. The normalization of the signal can not  
4208 only be affected by  $\mu$ , but also by other nuisance parameters of the fit, e.g., systematic  
4209 uncertainties.

4210 To obtain a low uncertainty on the post-fit  $\mu$ , a good separation of the signal and back-  
4211 ground templates in the signal region is needed. This is achieved by using the output of  
4212 a multi-variate classifier as observable for the templates.

## 4213 A.3 Data Samples and simulated events

### 4214 A.3.1 Data Samples and trigger requirements

4215 This analysis is performed using the full Run 2 CMS data with the combined luminosity  
4216 of  $137.6 \text{ fb}^{-1}$ . In the CMS experiment data is collected using a two-level trigger system,  
4217 described in Section 3.2.6. For each data taking year a set of High Level Trigger (HLT)  
4218 paths with the lowest threshold is selected.

4219 In the 0-lepton channel, events are selected with the trigger that requires the presence of  
4220 MET and MHT ( $= |\sum_{jets} \vec{p}_T|$ , where jets are required to satisfy  $p_T > 30 \text{ GeV}$  and  $|\eta| <$   
4221  $5$ .) with thresholds 110 GeV in 2016 and 120 GeV in 2017 and 2018. The MET and MHT  
4222 are constructed with the jets passing tight identification criteria. In the 1-lepton channel,  
4223 the presence of an isolated lepton is required. The  $p_T$  threshold for the HLT paths used  
4224 to trigger the isolated electron are 27 GeV in 2016 and 32 GeV in 2017 and 2018. For the  
4225 muon paths the thresholds are 27 GeV in 2017 data-taking period and 24 GeV in 2016  
4226 and 2018.

4227 For the 2-lepton channel, the double-muon and double-electron triggers are used. The  
4228  $p_T$  thresholds for the leading muon is 17 GeV and 8 GeV for the sub-leading muon. An  
4229 online requirement on the dimuon invariant mass is applied to remove the contribution  
4230 from low-mass resonances. For the electron 2-lepton channel, the  $p_T$  thresholds are 23  
4231 GeV and 12 GeV for the leading and sub-leading electrons, respectively.

### 4232 A.3.2 Simulated events

4233 The signal samples for the quark induced production of ZH and WH are generated by the  
 4234 POWHEG v2 [73, 74] event generator extended with the MiNLO procedure [64, 80] at  
 4235 NLO accuracy. The gluon induced signal samples on the other hand have leading order  
 4236 accuracy and are produced with POWHEG v2. Signal yields are scaled to an inclusive  
 4237 cross-section calculated up to NNLO [29].

4238 The di-boson samples WZ, ZZ and WW are produced at NLO with the MADGRAPH5\_aMC@NLO  
 4239 [102] using FxFx merging scheme [128]. The  $t\bar{t}$  and single-top in t-channel are simulated  
 4240 with POWHEG v2, and for single-top in s-channel and tW production the POWHEG v1  
 4241 is used.

4242 To reduce the statistical uncertainties from the generation process, events are produced  
 4243 with higher numbers than is expected to appear for that processes in the real collision  
 4244 event. To retrieve the correct number of events, a weight is assigned to each event,

$$w_{\text{event}} = \sigma \times \mathcal{L} \times \frac{w_{\text{generator}}}{\sum w_{\text{generator}}}, \quad (\text{A.1})$$

4245 where the  $w_{\text{generator}}$  is assigned by the Monte Carlo generator for each generated event and  
 4246 are not always constant, in some Monte Carlo generators these weights include negative  
 4247 values when next-to-leading order (NLO) accuracies are included.

## 4248 A.4 Event selection

4249 The signal and control regions are defined as follows, the signal region (SR) enriched  
 4250 in  $VH(b\bar{b})$ , the  $t\bar{t}$  control region where  $t\bar{t}$  has the highest contribution, the V+HF for  
 4251 the vector boson associated with heavy-flavour jets and the V+LF for the vector boson  
 4252 associated with light-flavour jets. These regions are partitioned according to the STXS  
 4253 framework.

4254 The definition of the boosted region, which has a single fat b-jet instead of the expected  
 4255 two resolved b-jets. Some events can be reconstructed both in the resolved and in the

4256 boosted analysis (called overlap event), therefore those events have to be placed in either  
 4257 the resolved or the boosted regions to avoid double counting. The most efficient scheme is  
 4258 considered where overlap events are assigned to the resolved categories, unless the event  
 4259 would move from the boosted signal region to a resolved control region, then it is assigned  
 4260 to the boosted signal region.

#### 4261 A.4.1 Simplified template cross-section (STXS) bins

4262 The simplified template cross sections (STXS) scheme [129] is designed to lessen the effect  
 4263 of theory dependence in the measurements and to make it simple to compare theoretical  
 4264 models with the observations, provide a consistent scheme for these measurements. The  
 4265 STXS framework offers the best features of signal strength measurements and also provides  
 4266 new features such as identification of a BSM-specific phase space. The STXS kinematic  
 4267 regions, also called bins, are defined at the generator level for each production process. The  
 4268 ultimate goals are to maximise the experimental sensitivity, to minimize the dependence  
 4269 on theoretical uncertainties, and to isolate BSM effects.

4270 In Fig. A.3, the latest recommended categorisation for VH mode is shown. The STXS  
 4271 bins are defined using the transverse momenta of vector boson  $p_T(V)$  and the number of  
 4272 additional jets. The VH process is split into three channels  $qq \rightarrow ZH$ ,  $qq \rightarrow WH$  and  
 4273  $gg \rightarrow ZH$ . Each of them is consequently separated into four  $p_T(V)$  regions: 0–75, 75–150,  
 4274 150–250, >250 GeV. The 150–250 GeV bin is additionally split by the number of additional  
 4275 jets ( $n_{jet}$ ) with  $p_T > 30$  GeV: 0 jets, and at least one additional jet. The  $p_T(V) > 250$  GeV  
 4276 bin represents the region sensitive to the BSM effects. The dashed boundaries are defined  
 4277 to consider further splitting if possible experimentally. The STXS bins are supposed to  
 4278 be merged by the experiments if a lack of sensitivity for the proposed binning is observed.

#### 4279 A.4.2 Resolved analysis selection

4280 Signal region selection is used to separate the signal from the background while conserving  
 4281 improved signal purity. The reconstructed di-jet invariant mass ( $m_{jj}$ ), the number of  
 4282 additional jets, and the b-tag discriminator score for the jets in the event are crucial

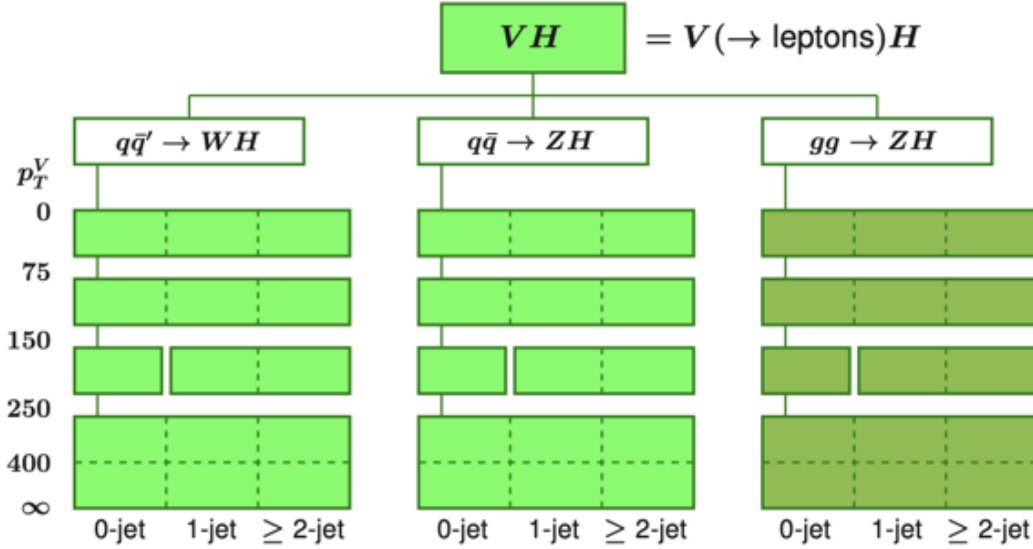


Figure A.3: Overview of the STXS bins for the three  $VH$  production modes [130]. The vertical axis reflects the  $p_T(V)$  bin ranges and the horizontal axis the number of additional jets.

4283 event variables for identifying signals from the background. The signal region is further  
 4284 divided into several bins with different S/B ratios using an MVA-classifier (DNN) used  
 4285 in the signal region. Therefore, the multivariate classifier is assigned for generating the  
 4286 higher purity signal phase space, while relatively loose cuts are applied for the signal  
 4287 region itself.

4288 A loose channel specific pre-selection is applied in the very beginning of the analysis chain.  
 4289 It is not very specific to the  $VH$  signal process but rather to make sure all the needed  
 4290 objects like jets and leptons are present. At the next step the events passing the channel  
 4291 selections are sub-categorised into events from signal regions (SR), defined as the region  
 4292 with enhanced signal efficiency, and the control regions, defined to constrain the leading  
 4293 background and enriched in the corresponding background process. The summary of the  
 4294 selection for signal and control regions for three channels: 0-lepton, 1-lepton, and 2-lepton  
 4295 are given in Tab. A.1, Tab. A.2, and Tab. A.3, respectively. The selection procedure is  
 4296 illustrated in Fig. A.4.

Variable	SR	Z+HF	Z+LF	$t\bar{t}$
Common selection:				
$\min(p_T^{miss}, H_T^{miss})$	>100	-/-	-/-	-/-
$p_T^{miss}$	>170	-/-	-/-	-/-
$p_T(j_1)$	>60	-/-	-/-	-/-
$p_T(j_2)$	>35	-/-	-/-	-/-
$p_T(j, j)$	>120	-/-	-/-	-/-
$\Delta\phi(Z, jj)$	>2.0	-/-	-/-	-/-
$\Delta\phi(p_T^{\vec{miss}}, j)$	>0.5	-/-	-/-	-/-
SR/CR difference:				
$N_{aj}$	$\leq 1$	$\leq 1$	$\leq 1$	$\geq 2$
$M_{jj}$	$\in[90-150]$	$\notin[90-150]$	-	-
$\text{btag}_{\max}$	>medium	>medium	<medium	>medium
$\text{btag}_{\min}$	>loose	>loose	<loose	>loose
$\Delta\phi(p_T^{\vec{miss}}, p_T^{\vec{miss}}_{trk})$	<0.5	<0.5	<0.5	-
$\min \Delta\phi(p_T^{\vec{miss}}, j)$	-	-	-	$< \pi/2$

Table A.1: Definition of the SR and CRs for the resolved selection in the 0-lepton channel. If the same selection is applied in all SRs and CRs, this is indicated by the -/- symbol in the latter. The  $M_{jj}$  and momenta variables have units of GeV.

Variable	SR	W+HF	W+LF	$t\bar{t}$
Common selection:				
$p_T(j, j)$	>100	-/-	-/-	-/-
$p_T(V)$	>150	-/-	-/-	-/-
$N_{al}$	<1	-/-	-/-	-/-
$p_T(j_1)$	>25	-/-	-/-	-/-
$p_T(j_2)$	>25	-/-	-/-	-/-
$\Delta\phi(\text{lep}, p_T^{\vec{miss}})$	<2	-/-	-/-	-/-
SR/CR difference:				
$\text{btag}_{\max}$	>medium	>medium	[loose-medium]	>tight
$\text{btag}_{\min}$	>loose	-	-	-
$M_{jj}$	$\in[90-150]$	$\in[150-250]$ and $<90$	<250	<250
$N_{aj}$	<2	<2	-	>1
$\frac{p_T^{miss}}{\sigma(p_T^{miss})}$	-	>2	>2	-
$\Delta\phi(H, V)$	>2.5	-	-	-

Table A.2: Definition of the SR and CRs for the resolved selection of the 1-lepton channel. If the same selection is applied in all SRs and CRs, this is indicated by the -/- symbol in the latter. The  $M_{jj}$  and momenta variables have units of GeV.

Variable	SR	Z+HF	Z+LF	$t\bar{t}$
$p_T(V)$	$>75$	-//-	-//-	-//-
$btag_{\max}$	$>\text{medium}$	$>\text{medium}$	$<\text{loose}$	$>\text{tight}$
$btag_{\min}$	$>\text{loose}$	$>\text{loose}$	$<\text{loose}$	$>\text{loose}$
$M(V)$	$\in[75-105]$	$\in[85-97]$	$\in[75-105]$	$\in[10-75]$ and $>120$
$M_{jj}$	$\in[90-150]$	$\notin[90-150]$	$\in[90-150]$	-
$p_T^{\text{miss}}$	-	$<60$	-	-
$\Delta\phi(H, V)$	-	$>2.5$	$>2.5$	-

Table A.3: Definition of the SR and CRs for the resolved selection in the 2-lepton channel. If the same selection is applied in all SRs and CRs, this is indicated by the -//- symbol in the latter. The  $M_{jj}$ ,  $M(V)$ , and momenta variables have units of GeV.

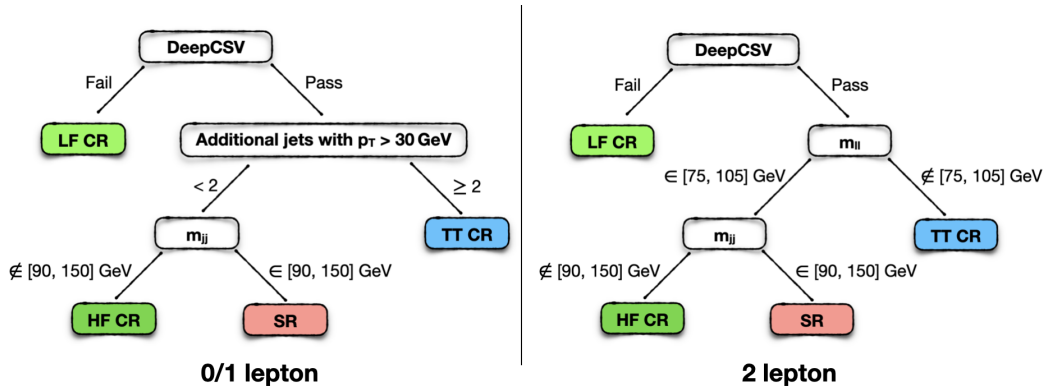


Figure A.4: Left: Resolved 0-lepton and 1-lepton channels selection scheme. Right: Resolved 2-lepton channel selection scheme.

### 4297 A.4.3 Boosted analysis selection

4298 The vector boson selection in the boosted analysis follows exactly the same procedure as  
4299 for the resolved analysis described in Section A.4.2. The Higgs boosted decay topology  
4300 is considered for the vector boson momentum range of  $p_T(V) > 250$  GeV in all analysis  
4301 channels. Boosted analysis is most relevant for high  $p_T$  bins of the STXS scheme while  
4302 reducing multi-jet background events. Selection criteria for the SR and CRs in the boosted  
4303 topology for 0-, 1-, and 2-lepton channels is given in Tab. A.4. The selection procedure  
4304 for boosted analysis is illustrated in Fig. A.5.

Variable	SR	V+HF	V+LF	$t\bar{t}$
0-lepton				
DeepAK8bbVsLight	>0.8	>0.8	<0.8	>0.8
$m_{SD}$	$\in[90-150]$	$\in[50-90]$ or $\in[150-250]$	>50	>50
$N_{al}$	=0	=0	=0	>0
$N_{aj}$	=0	=0	=0	>0
1-lepton				
DeepAK8bbVsLight	>0.8	>0.8	<0.8	>0.8
$m_{SD}$	$\in[90-150]$	$\in[50-90]$ or $\in[150-250]$	>50	>50
$N_{al}$	=0	=0	=0	>0
$N_{aj}$	=0	=0	=0	>0
2-lepton				
DeepAK8bbVsLight	>0.8	>0.8	<0.8	>0.8
$m_{SD}$	$\in[90-150]$	$\in[50-90]$ or $\in[150-250]$	$\in[90-150]$	>50
$m(V)$	$\in[75-105]$	$\in[75-105]$	$\in[75-105]$	$\notin[75-105]$

Table A.4: Selection criteria for the SR and CRs in the boosted topology for 0-, 1-, and 2-lepton channels. The DeepAK8bbVsLight designation represents the DEEPAK8 discriminant for the light-quark flavor discrimination node. The  $m_{SD}$  and  $M(V)$  variables have units of GeV.

## 4305 A.5 Multivariate discriminants

4306 The signal region selection enriches in signal the phase space. The multivariate analysis  
4307 techniques allow to further improve the signal versus background discrimination power.  
4308 Three multivariate methods are used in this analysis: a deep neural networks (DNN)  
4309 binary classifier for the resolved signal region, a multi-class DNN in the V+HF control



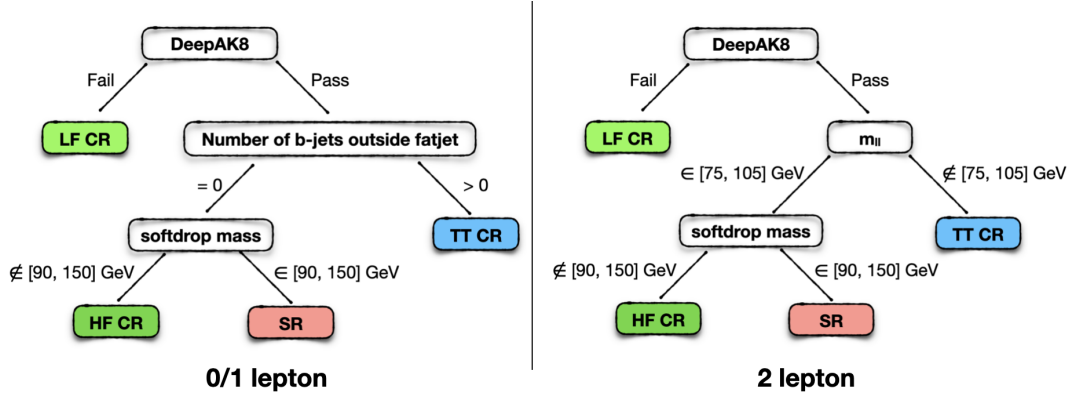


Figure A.5: Left: Boosted 0-lepton and 1-lepton channels selection scheme. Right: Boosted 2-lepton channel selection scheme.

4310 regions to improve the separation of different backgrounds, and a boosted decision tree  
 4311 (BDT) technique for the binary classification in the boosted signal region.

### 4312 A.5.1 DNN

4313 For the resolved Higgs decay topology, a signal vs. background DNN classifier is trained  
 4314 for each channel separately. The output in the signal region is used in the fit for all the  
 4315 channels. For the 0-lepton and 1-leptons channel  $V + HF$  control region, a multi-class  
 4316 DNN classifier is used.

4317 The tensorflow framework [131] was used to train a 6 hidden layer DNN classifier, with  
 4318 each layer having 512, 256, 128, 64, 64 and 64 nodes. For 2-classes DNN all signal processes  
 4319 were grouped into signal class, and all of the background processes into a background class.  
 4320 In the multi-class DNN instead of signal and background output nodes, the classification  
 4321 is performed according to the 5 leading background processes listed in Tab. A.5. A  
 4322 background class is assigned to each event, if the corresponding class probability is the  
 largest.

0	$V+udsg$
1	$V+c$
2	$V+b$
3	Single top
4	$t\bar{t}$

Table A.5: Classes used for the 0/1-lepton multi-DNN classifier.

4323

4324 In both 2-class DNN and multi-class DNN the same architecture and the same set of  
4325 input features are used. The agreement of data and simulation for all of the MVA input  
4326 variables is studied and found to be sufficient.

4327 Fig. A.6 shows the HFDNN discriminant in the 0- and 1-lepton heavy-flavor CRs, after a  
4328 maximum likelihood fit to the data. This is a simultaneous fit of all SRs and CRs in the  
4329 analysis. The DNN score is used as a discriminating variable in each resolved SR, while  
4330 different strategies are used in the resolved CRs.

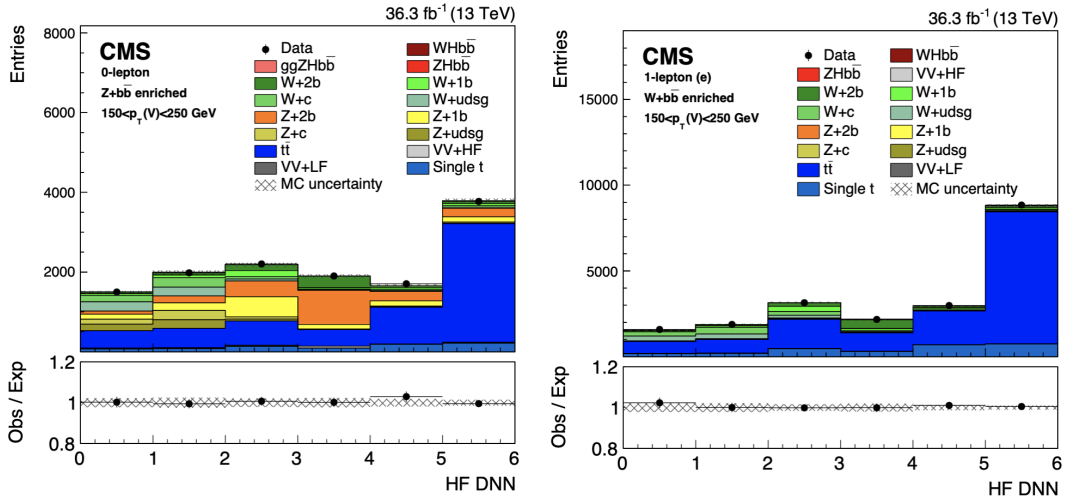


Figure A.6: Distribution of the HFDNN scores in the 0-lepton (left) and 1-lepton, (right)  $Z + b$  and  $W + b$  heavy-flavor CRs for the 2016 dataset, after the fit to data. The output nodes target enrichment in the  $V + \text{light-quark}$  (first bin),  $V + c$  (second bin),  $V + b$  (third bin),  $V + b\bar{b}$  (fourth bin), single top quark (fifth bin), and  $t\bar{t}$  (sixth bin) backgrounds. The lower plots display the ratio of the data to the MC expectations. The vertical bars on the points represent the statistical uncertainty in the ratio, and the hatched area shows the MC uncertainty.

## 4331 A.6 Results

4332 The inclusive signal strength extracted from a simultaneous maximum likelihood fit of  
4333 the SRs and CRs, combining all three data-taking years, is  $\mu = 1.15^{+0.22}_{-0.20}$ , where the  
4334 uncertainties include both the statistical and systematic components. The individual  
4335 signal strengths are  $\mu = 1.43 \pm 0.37$ ,  $\mu = 0.68 \pm 0.36$ , and  $\mu = 1.23 \pm 0.30$  for the 2016,  
4336 2017, and 2018 data-taking years, respectively. Figure A.7 shows the signal strengths per  
4337 analysis channel, as well as the signal strengths split by production mode ( $ZH$  or  $WH$ ).

4338 The  $p$ -value compatibility of the individual deviations of the three analysis channels from  
 4339 the SM expectation ( $\mu = 1$ ) is 64%, while the  $p$ -value compatibility of the three analysis  
 4340 channels with the inclusive  $VH$ ,  $H \rightarrow b\bar{b}$  signal strength is 84%.

4341 The measured signal strengths in the different STXS bins, fitting all data-taking years  
 4342 (2016–2018) are shown in Fig. A.8. These results are interpreted in Fig. A.9 as  $\sigma\mathcal{B}$ , the  
 4343 product of the production cross sections and the branching fractions for  $V \rightarrow$  leptons and  
 4344  $H \rightarrow b\bar{b}$ . To convert the results to measurements of the production cross section alone,  
 4345 theoretical uncertainties that modify the overall cross section of the individual STXS bins,  
 4346 or the inclusive cross section, are removed from the fit. These measured cross sections,  
 4347 along with the SM predictions, are given in Table A.6. The local inclusive observed  
 4348 (expected) significance of the measured  $ZH$  and  $WH$  signals, over the background-only  
 4349 expectation, is found to be 6.3 (5.6) standard deviations when taking into account all  
 4350 three data-taking years. Examples of post-fit distributions of the DNN output scores in  
 4351 the SRs of the 2018 data set are shown in Fig. A.10 for the 0-, 1-, and 2-lepton channels  
 4352 in the category targeting the  $250 < p_T(V) < 400\text{GeV}$  STXS bin. Figure A.11 shows  
 4353 the distribution of events in all channels, sorted according to the observed value of  $\log_{10}$   
 4354 (S/B), for the three data-taking years combined; here, the signal (S) and background  
 4355 (B) yields are determined from the discriminant scores used in the resolved and boosted  
 4356 analyses.

STXS bin	Expected $\sigma\mathcal{B}$ [fb]	Observed $\sigma\mathcal{B}$ [fb]	Best-fit $\mu$
ZH $75 < p_T(Z) < 150$ GeV	$50.0 \pm 5.3$	$71 \pm 38$	$1.4 \pm 0.8$
ZH $150 < p_T(Z) < 250$ GeV 0 jets	$9.0 \pm 1.4$	$3.8 \pm 4.1$	$0.4 \pm 0.5$
ZH $150 < p_T(Z) < 250$ GeV $\geq 1$ jets	$10.1 \pm 2.2$	$< 0$	$-0.6 \pm 1.0$
ZH $250 < p_T(Z) < 400$ GeV	$4.5 \pm 0.9$	$6.9 \pm 2.2$	$1.5 \pm 0.5$
ZH $p_T(Z) > 400$ GeV	$0.9 \pm 0.1$	$1.6 \pm 0.6$	$1.8 \pm 0.8$
WH $150 < p_T(W) < 250$ GeV	$24.9 \pm 1.8$	$6 \pm 16$	$0.2 \pm 0.7$
WH $250 < p_T(W) < 400$ GeV	$6.3 \pm 0.5$	$11.9 \pm 3.8$	$1.9 \pm 0.6$
WH $p_T(W) > 400$ GeV	$1.4 \pm 0.1$	$2.7 \pm 1.1$	$1.9 \pm 0.8$

Table A.6: Predicted and measured values of the product of the cross section and branching fractions in the V(leptonic)H STXS process scheme. The SM predictions for each bin are calculated using the inclusive values reported in Ref. [132]. The uncertainties shown are the combined statistical and systematic components.

4357 Tab. A.7 shows the contribution, in terms of absolute uncertainties, to the uncertainty in  
 4358 the measured inclusive signal strength originating from the various sources of systematic  
 4359 uncertainty. This contribution for a given group of uncertainties is defined as the difference

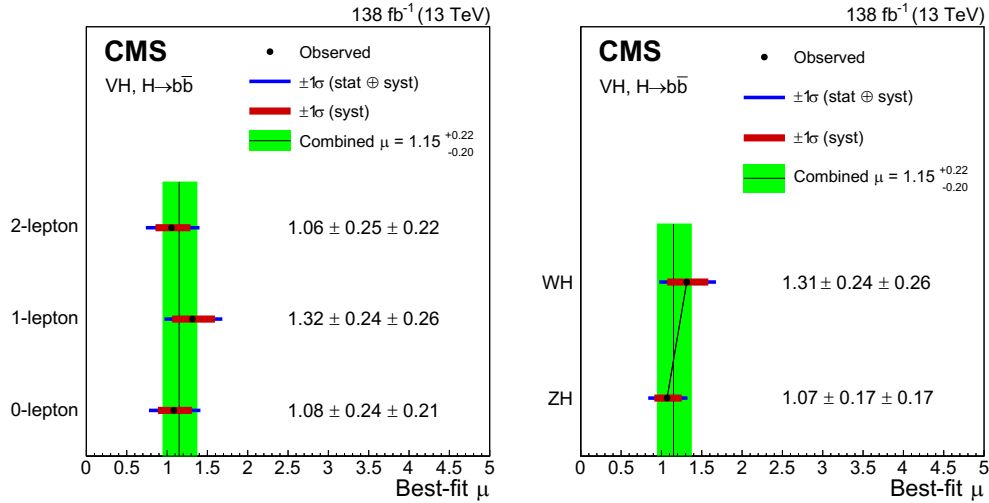


Figure A.7: Signal strengths (points) for the 0-, 1-, and 2-lepton channels (left) and the  $ZH$  and  $WH$  production modes (right). The horizontal red and blue bars on the points represent the systematic and total uncertainties, respectively. The combined inclusive signal strength is shown by the vertical line, with the green band giving the 68% confidence interval. The results combine the 2016–2018 data-taking years. The first and the second uncertainty values correspond to the statistical and systematic uncertainties, respectively.

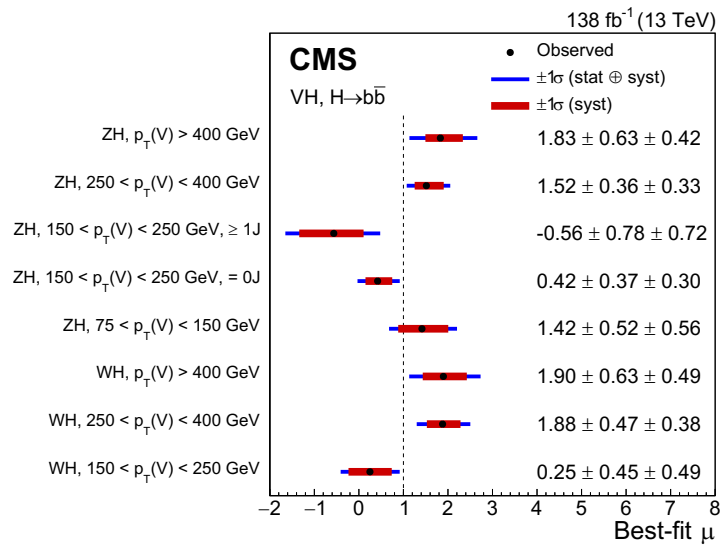


Figure A.8: STXS signal strengths from the analysis of the 2016–2018 data. The vertical dashed line corresponds to the SM value of the signal strength. The first and the second uncertainty values correspond to the statistical and systematic uncertainties, respectively.

4360 in quadrature between the total uncertainty in the signal strength and the uncertainty  
4361 in the signal strength with the nuisance parameters of the corresponding group fixed to  
4362 their best-fit values. The total statistical uncertainty is defined as the uncertainty in the  
4363 signal strength when all the constrained nuisance parameters are fixed to their best-fit  
4364 values, while the total systematic uncertainty is defined as the difference in quadrature  
4365 between the total uncertainty in the signal strength and the total statistical uncertainty.

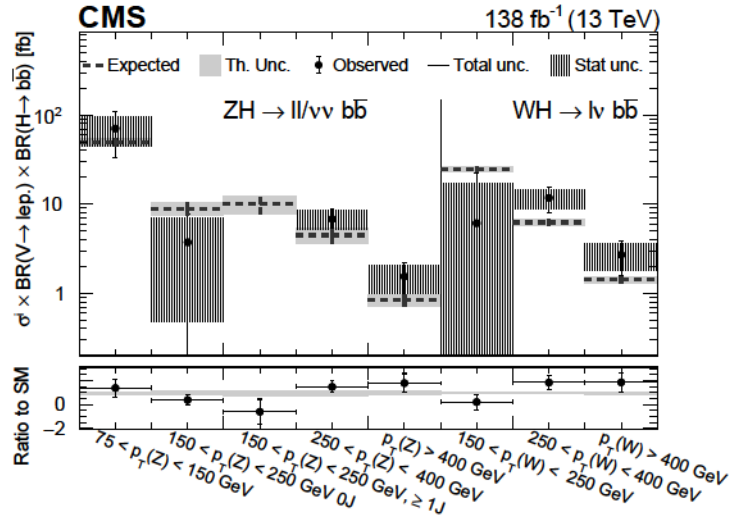


Figure A.9: Measured values of  $\sigma\mathcal{B}$ , defined as the product of the  $VH$  production cross sections multiplied by the branching fractions of  $V \rightarrow \text{leptons}$  and  $H \rightarrow b\bar{b}$ , evaluated in the same STXS bins as for the signal strengths, combining all years. In the lower panel, the ratio of the observed results, with associated uncertainties, to the SM expectations is shown. If the observed signal strength for a given STXS bin is negative, no value is plotted for  $\sigma\mathcal{B}$  in the upper panel.

4366 Tab. A.7 breaks the total uncertainty down into the following sources.

- 4367 • Theoretical uncertainties in the signal and background components.
- 4368 • Limited size of simulated samples.
- 4369 • Simulation modeling, including uncertainty sources associated with the modeling of
- 4370 the  $V$ +jets background components. Additionally, the  $p_T(V)$  migration uncertain-
- 4371 ties are included in this category.
- 4372 • Experimental uncertainties (b tagging, integrated luminosity, JES and JER, lepton
- 4373 identification, and trigger). The JES and JER components include the dedicated
- 4374 uncertainty in mass scale and smearing that is applied for jets subject to the b jet
- 4375 energy regression.

4376 The limited size of the NLO  $V$ +jets samples is the largest contribution to the overall  $VH$ ,

4377  $H \rightarrow b\bar{b}$  signal strength uncertainty.

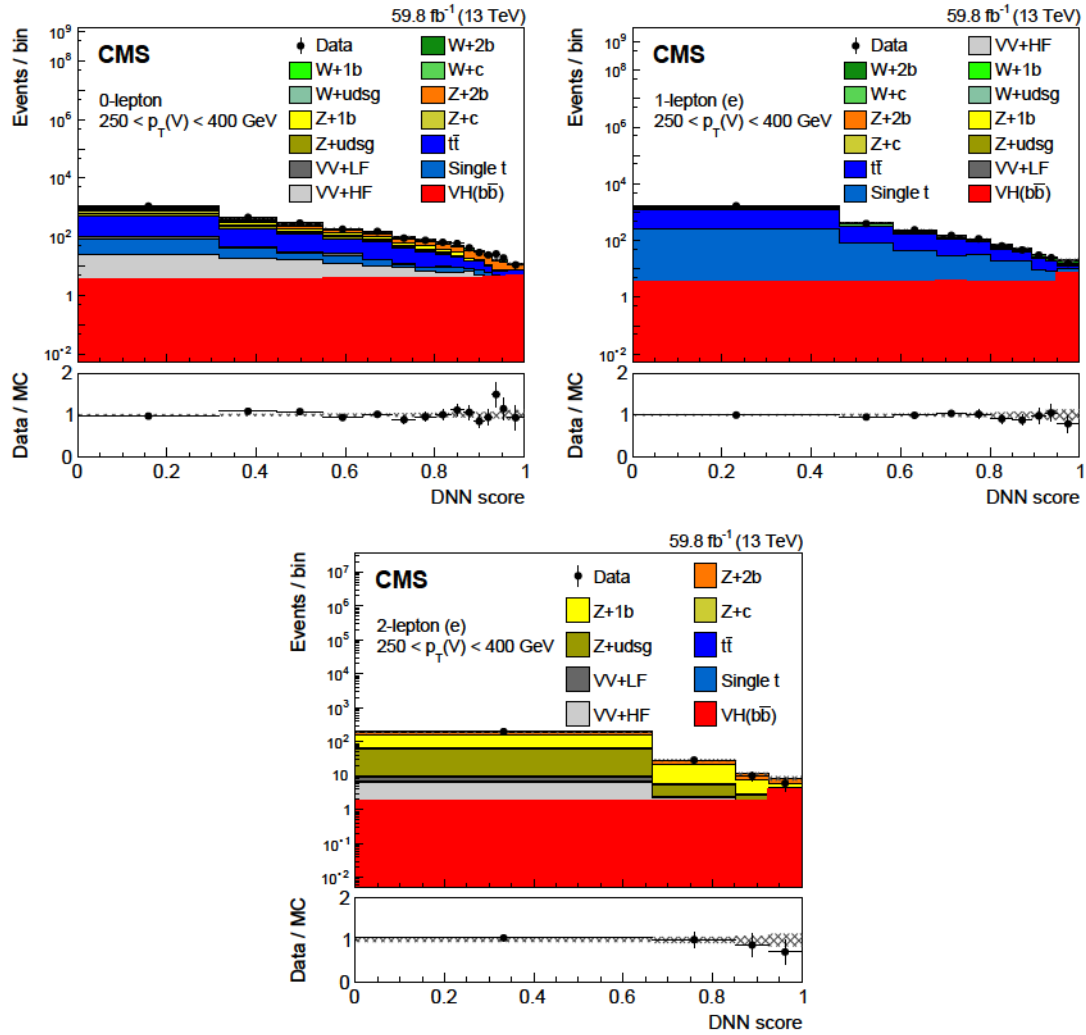


Figure A.10: Post-fit distributions of the DNN discriminant in the  $250 < p_T(V) < 400\text{GeV}$  category of the 0-lepton (top left), 1-lepton (top right) and 2-lepton (bottom) channels for the electron final state using the 2018 data set. The background contributions after the maximum likelihood fit are shown as filled histograms. The Higgs boson signal is also shown as a filled histogram, and is normalized to the signal strength shown in Fig. A.8. The hatched band indicates the combined statistical and systematic uncertainty in the sum of the signal and background templates. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel. The distributions that enter the maximum likelihood fit use the same binning as shown here.

## 4378 A.7 Summary

4379 Measurements are presented of the cross section for the associated production of the 125  
4380 GeV Higgs boson and a W or Z boson, where the Higgs boson decays to  $b\bar{b}$  and the vector  
4381 bosons decay to leptons. Proton-proton collision data collected by the CMS experiment  
4382 during 2016–2018 at  $\sqrt{s} = 13\text{TeV}$  are used, corresponding to an integrated luminosity  
4383 of  $138\text{fb}^{-1}$ . Five decay channels are analyzed, and both resolved as well as merged-jet

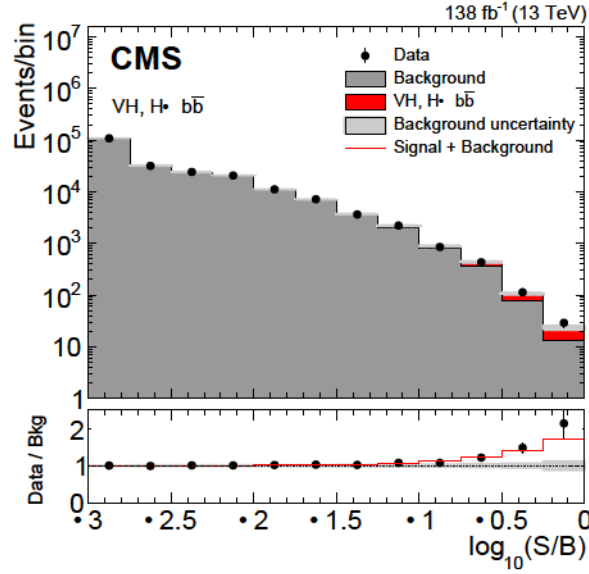


Figure A.11: Distributions of signal, background, and observed data event yields sorted into bins of similar signal-to-background ratio, as given by the result of the fit to the multivariate discriminants in the resolved and boosted categories. All events in the signal regions of the 2016–2018 data set are included. The red histogram indicates the Higgs boson signal assuming SM yields ( $\mu = 1$ ) and the sum of all backgrounds is given by the gray histogram. The lower panel shows the ratio of the observed data to the background expectation, with the total uncertainty in the background prediction indicated by the gray hatching. The red line indicates the sum of signal assuming the SM prediction plus background contribution, divided by the background.

	$\Delta\mu$	
Background (theory)	+0.043	-0.043
Signal (theory)	+0.088	-0.059
MC sample size	+0.078	-0.078
Simulation modeling	+0.059	-0.059
b tagging	+0.050	-0.046
Jet energy resolution	+0.036	-0.028
Int. luminosity	+0.032	-0.027
Jet energy scale	+0.025	-0.025
Lepton ident.	+0.008	-0.007
Trigger ( $p_T^{miss}$ )	+0.002	-0.001

Table A.7: The sources of systematic uncertainty in the inclusive signal strength measurement and their positive and negative values.

4384 topology are employed in each vector boson decay mode. An additional subcategorization  
 4385 in the transverse momentum of the vector boson and the number of additional jets in  
 4386 the event is applied to maximize the sensitivity of different simplified template cross  
 4387 section bins. The overall signal strength, combining all analysis categories, is found to be  
 4388  $\mu = 1.15_{-0.20}^{+0.22}$ . The production of the Higgs boson in association with a vector boson and

4389 decays to bottom quark pairs is established with an observed (expected) significance of  
4390 6.3 (5.6) standard deviations.





4392 **Education**

2019-Current **PhD** in EXPERIMENTAL HIGH ENERGY PHYSICS,  
University of Zagreb, Croatia

2015-2017 **Master of Science** in PHYSICS,  
4393 Major: Nuclear and Particle Physics,  
The Maharaja Sayajirao University of Baroda, India

2012-2015 **Bachelor of Science** in PHYSICS,  
The Maharaja Sayajirao University of Baroda, India

4394 **Employment History**

2019-Current **Research Assistant**  
Ruder Bošković Institute, Zagreb, Croatia

4395

2017-2019 **Scientific Project Assistant**  
Institute of Physics, Bhubaneswar, India

4396 **Research Experience**

4397 2019-  
*Current* **PhD Thesis**  
**Search for multiple Higgs boson production in hadronic final states at the Large Hadron Collider**  
Supervisor: Dr. Dinko Ferenček  
*Ruder Bošković Institute, Zagreb, Croatia*

OCT 2017 | **High level trigger development of CMS Experiment at LHC,**  
*Jan 2019* | **CERN**  
Supervisor: Dr. Aruna Kumar Nayak  
*Institute of Physics, Bhubaneswar, India*

4398 | [MSc Thesis](#)  
2016-2017 | **Exotic Hadrons**  
Supervisor: Prof. J. P. Singh  
*The Maharaja Sayajirao University, Vadodara, India*

SUMMER | **Investigation of CP Violating observables in charged B decay**  
2016 | Supervisor: Dr. Jim Libby  
*Indian Institute of Technology Madras (IIT-M), Chennai, India*

2013-2015 | **Procedural Understanding in Physics: Behaviour of soft springs**  
Supervisor: Dr. Rajesh B. Khaparde  
*Homi Bhabha Centre for Science Education (HBCSE), TIFR, Mumbai, India*

## 4399 Publications

- 4400
- ▷ CMS Collaboration, **Search for non-resonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at  $\sqrt{s} = 13$  TeV**, JHEP 03 (2021) 257, DOI [10.1007/JHEP03\(2021\)257](https://doi.org/10.1007/JHEP03(2021)257)
  - 4401 ▷ CMS Collaboration, **Measurement of simplified template cross sections of the Higgs boson produced in association with W or Z bosons in the  $H \rightarrow b\bar{b}$  decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV**, Phys. Rev. D 109 (2024) 092011, DOI [10.1103/PhysRevD.109.092011](https://doi.org/10.1103/PhysRevD.109.092011)

## Workshops and Schools Attended

- JULY 2023 **HHH Workshop** Dubrovnik, Croatia.
- AUG. 2021 **Baltic School of High-Energy Physics and Accelerator Technologies (HEP&AT)** Riga Technical University (RTU), Klapkalnciems, Latvia.
- MAR. 2021 **Tracker Upgrade DAQ School** online + hardware exercises at Institut Ruder Bošković, Zagreb, Croatia.
- JULY 2020 **Summer School on Machine Learning in High Energy Physics (MLHEP - 2020)** online by EPFL, Lausanne, Switzerland.
- SEP. 2019 **CMS Physics Object School (CMSPOS)** at RWTH, Aachen, Germany.
- 4403 NOV. 2017 **XI SERC School on Experimental High Energy Physics** at NISER, Khurda, India.
- DEC. 2016 **Winter School on Astro-particle Physics (WAPP – 2016)** at Cosmic Ray Laboratory, Ooty, India.
- DEC. 2015 **Winter School on Astro-particle Physics (WAPP – 2015)** at Bose Institute, Darjeeling, India.
- NOV. 2013 **Refresher Experimental Physics Course** at Indian Academy of Sciences, Bangalore, India.
- JUNE 2013 **NIUS Physics(10)** at Homi Bhabha Centre for Science Education, TIFR, Mumbai, India.

## Academic Achievements

- ◇ INnovation in Science Pursuit for Inspired REsearch (INSPIRE) Scholarship  
Department of Science and Technology, Govt. of India, 2012-2017.
- 4405 ◇ Qualified CSIR-UGC NET Lecturership (Rank - 65), June-2017, Government of India.
- ◇ Qualified All India Engineering Entrance Examination (AIEEE),
- ◇ National Means Cum-Merit Scholarship (NMMS), Government of India, 2008-2010.

## Skills

### Technical Skills

Python,  $\text{\LaTeX}$ , C++, Pandas(Software), Numpy, Awkward Array, Git, Matplotlib, Machine Learning, Data Science, Data Analysis, Statistical Data Analysis

### Soft Skills

Independent and critical thinking, Fast learning , Logical Problem-solving, Time Management, Communication, Flexible