

1 SEARCH FOR PRODUCTION OF A HIGGS BOSON AND A SINGLE TOP  
2 QUARK IN MULTILEPTON FINAL STATES IN  $pp$  COLLISIONS AT  $\sqrt{s} = 13$   
3 TeV.

4 by

5 Jose Andres Monroy Montañez

6 A DISSERTATION

7 Presented to the Faculty of

8 The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

10 For the Degree of Doctor of Philosophy

11 Major: Physics and Astronomy

12 Under the Supervision of Kenneth Bloom and Aaron Dominguez

13 Lincoln, Nebraska

14 July, 2018

15 SEARCH FOR PRODUCTION OF A HIGGS BOSON AND A SINGLE TOP  
16 QUARK IN MULTILEPTON FINAL STATES IN pp COLLISIONS AT  $\sqrt{s} = 13$   
17 TeV.

18 Jose Andres Monroy Montañez, Ph.D.

19 University of Nebraska, 2018

20 Adviser: Kenneth Bloom and Aaron Dominguez

# **<sub>21</sub> Table of Contents**

<b><sub>22</sub> Table of Contents</b>	<b>iii</b>
<b><sub>23</sub> List of Figures</b>	<b>vii</b>
<b><sub>24</sub> List of Tables</b>	<b>x</b>
<b><sub>25</sub> 1 INTRODUCTION</b>	<b>1</b>
<b><sub>26</sub> 2 Theoretical approach</b>	<b>2</b>
27     2.1 Introduction . . . . .	2
28     2.2 Standard model of particle physics . . . . .	3
29     2.2.1 Fermions . . . . .	5
30     2.2.1.1 Leptons . . . . .	6
31     2.2.1.2 Quarks . . . . .	8
32     2.2.2 Fundamental interactions . . . . .	12
33     2.2.3 Gauge invariance. . . . .	15
34     2.2.4 Gauge bosons . . . . .	18
35     2.3 Electroweak unification and the Higgs mechanism . . . . .	19
36     2.3.1 Spontaneous symmetry breaking (SSB) . . . . .	27
37     2.3.2 Higgs mechanism . . . . .	31

38	2.3.3	Masses of the gauge bosons . . . . .	34
39	2.3.4	Masses of the fermions . . . . .	35
40	2.3.5	The Higgs field . . . . .	36
41	2.3.6	Production of Higgs bosons at LHC . . . . .	37
42	2.3.7	Higgs boson decay channels . . . . .	41
43	2.4	Experimental status of the anomalous Higgs-fermion coupling . . . .	42
44	2.5	Associated production of a Higgs boson and a single top quark . . . .	44
45	2.6	CP-mixing in $tH$ processes . . . . .	49
46	<b>3</b>	<b>The CMS experiment at the LHC</b>	<b>54</b>
47	3.1	Introduction . . . . .	54
48	3.2	The LHC . . . . .	55
49	3.3	The CMS experiment . . . . .	65
50	3.3.1	CMS coordinate system . . . . .	68
51	3.3.2	Tracking system . . . . .	70
52	3.3.3	Silicon strip tracker . . . . .	73
53	3.3.4	Electromagnetic calorimeter . . . . .	74
54	3.3.5	Hadronic calorimeter . . . . .	76
55	3.3.6	Superconducting solenoid magnet . . . . .	77
56	3.3.7	Muon system . . . . .	79
57	3.3.8	CMS trigger system . . . . .	80
58	3.3.9	CMS computing . . . . .	81
59	<b>4</b>	<b>Event generation, simulation and reconstruction</b>	<b>86</b>
60	4.1	Event generation . . . . .	87
61	4.2	Monte Carlo Event Generators . . . . .	91
62	4.3	CMS detector simulation . . . . .	93

63	4.4 Event reconstruction . . . . .	94
64	4.4.1 Particle-Flow Algorithm . . . . .	95
65	4.4.1.1 Missing transverse energy . . . . .	106
66	4.4.2 Event reconstruction examples . . . . .	107
67	<b>5 Statistical methods</b>	<b>110</b>
68	5.1 Multivariate analysis . . . . .	110
69	5.1.1 Decision trees . . . . .	113
70	5.1.2 Boosted decision trees (BDT) . . . . .	116
71	5.1.3 Overtraining. . . . .	119
72	5.1.4 Variable ranking. . . . .	119
73	5.1.5 BDT output example. . . . .	120
74	5.2 Statistical inference. . . . .	121
75	5.2.1 Nuisance parameters. . . . .	121
76	5.2.2 Maximum likelihood estimation method . . . . .	122
77	5.2.3 Hypothesis test . . . . .	123
78	5.3 exclusion limits . . . . .	124
79	5.4 asymptotic limits . . . . .	124
80	<b>6 Search for production of a Higgs boson and a single top quark in multilepton final states in pp collisions at <math>\sqrt{s} = 13</math> TeV</b>	<b>125</b>
82	6.1 Introduction . . . . .	125
83	6.2 $tHq$ signature . . . . .	127
84	6.3 Background processes . . . . .	129
85	6.4 Data and MC Samples . . . . .	131
86	6.4.1 Full 2016 dataset and MC samples . . . . .	131
87	6.4.2 Triggers . . . . .	132

88	6.5 Object Identification . . . . .	136
89	6.5.1 Jets and $b$ -jet tagging. . . . .	136
90	6.5.2 Missing Energy MET. . . . .	137
91	6.5.3 Lepton reconstruction and identification . . . . .	138
92	6.5.4 Lepton selection efficiency . . . . .	146
93	6.6 Event selection . . . . .	148
94	6.7 Background predictions . . . . .	149
95	6.8 Signal discrimination . . . . .	151
96	6.8.1 Classifiers response . . . . .	153
97	6.9 Additional discriminating variables . . . . .	157
98	<b>Bibliography</b>	<b>159</b>
99	<b>References</b>	<b>161</b>

# <sup>100</sup> List of Figures

101	2.1 Standard Model of particle physics. . . . .	4
102	2.2 Transformations between quarks . . . . .	12
103	2.3 Fundamental interactions in nature. . . . .	13
104	2.4 SM interactions diagrams . . . . .	14
105	2.5 Neutral current processes . . . . .	20
106	2.6 Spontaneous symmetry breaking mechanism . . . . .	28
107	2.7 SSB Potential form . . . . .	29
108	2.8 Potential for complex scalar field . . . . .	30
109	2.9 SSB mechanism for complex scalar field . . . . .	31
110	2.10 Proton-Proton collision . . . . .	37
111	2.11 Higgs boson production mechanism Feynman diagrams . . . . .	39
112	2.12 Higgs boson production cross section and decay branching ratios . . . . .	40
113	2.13 Two dimensional $\kappa_t$ - $\kappa_V$ plot of the coupling modifiers. ATLAS and CMS combination. . . . .	43
115	2.14 Higgs boson production in association with a top quark mechanism Feyn- man diagrams . . . . .	45
117	2.15 Cross section for $tHq$ process as a function of $\kappa_t$ . . . . .	48
118	2.16 Cross section for $tHW$ process as a function of $\kappa_{Htt}$ . . . . .	48

119	2.17 NLO cross section for $tX_0$ and $t\bar{t}X_0$ .	52
120	2.18 NLO cross section for $tWX_0$ , $t\bar{t}X_0$ .	53
121	3.1 CERN accelerator complex . . . . .	55
122	3.2 LHC protons source. First acceleration stage.	56
123	3.3 The LINAC2 accelerating system at CERN.	57
124	3.4 LHC layout and RF cavities module.	58
125	3.5 LHC dipole magnet. . . . .	60
126	3.6 Integrated luminosity delivered by LHC and recorded by CMS during 2016	62
127	3.7 LHC interaction points . . . . .	63
128	3.8 Multiple $pp$ collision bunch crossing at CMS. . . . .	65
129	3.9 Layout of the CMS detector . . . . .	66
130	3.10 CMS detector transverse slice . . . . .	67
131	3.11 CMS detector coordinate system . . . . .	69
132	3.12 CMS tracking system schematic view. . . . .	70
133	3.13 CMS pixel detector . . . . .	71
134	3.14 SST Schematic view. . . . .	73
135	3.15 CMS ECAL schematic view . . . . .	75
136	3.16 CMS HCAL schematic view . . . . .	77
137	3.17 CMS solenoid magnet . . . . .	78
138	3.18 CMS Muon system schematic view . . . . .	79
139	3.19 CMS Level-1 trigger architecture . . . . .	81
140	3.20 WLCG structure . . . . .	82
141	3.21 Data flow from CMS detector through hardware Tiers . . . . .	84
142	4.1 Event generation process. . . . .	87
143	4.2 Particle flow algorithm. . . . .	95

144	4.3	Jet reconstruction.	103
145	4.4	Jet energy corrections.	104
146	4.5	Secondary vertex in a b-hadron decay.	105
147	4.6	HIG-13-004 Event 1 reconstruction.	107
148	4.7	$e\mu$ event reconstruction.	108
149	4.8	Recorded event reconstruction.	109
150	5.1	Scatter plots-MVA event classification.	112
151	5.2	Scalar test statistical.	112
152	5.3	Decision tree.	114
153	5.4	Decision tree output example.	117
154	5.5	BDT output example.	120
155	6.1	$tHq$ event signature.	128
156	6.2	$t\bar{t}$ and $t\bar{t}W$ events signature.	130
157	6.3	Tight vs loose lepton selection efficiencies in the 2lss channel.	146
158	6.4	Tight vs loose lepton selection efficiencies in the 3l channel.	147
159	6.5	Input variables to the BDT for signal discrimination normalized.	152
160	6.6	Input variables to the BDT for signal discrimination not normalized.	154
161	6.7	BDT inputs as seen by TMVA against $t\bar{t}$ .	155
162	6.8	BDT inputs as seen by TMVA against $t\bar{t}V$ .	156
163	6.9	Correlation matrices for the input variables in the TMVA.	157
164	6.10	MVA classifiers performance.	157
165	6.11	Additional discriminating variables distributions.	159

# <sup>166</sup> List of Tables

167	2.1	Fermions of the SM. . . . .	5
168	2.2	Fermion masses. . . . .	6
169	2.3	Lepton properties. . . . .	8
170	2.4	Quark properties. . . . .	9
171	2.5	Fermion weak isospin and weak hypercharge multiplets. . . . .	11
172	2.6	Fundamental interactions features. . . . .	15
173	2.7	SM gauge bosons. . . . .	19
174	2.8	Higgs boson properties. . . . .	37
175	2.9	Predicted branching ratios for a SM Higgs boson with $m_H = 125 \text{ GeV}/c^2$ . . . . .	41
176	2.10	Predicted SM cross sections for $tH$ production at $\sqrt{s} = 13 \text{ TeV}$ . . . . .	46
177	2.11	Predicted enhancement of the $tHq$ and $tHW$ cross sections at LHC . . . . .	49
178	6.1	Signal samples and their cross section and branching fraction. . . . .	132
179	6.2	$\kappa_V$ and $\kappa_t$ combinations. . . . .	133
180	6.3	List of background samples used in this analysis (CMSSW 80X). . . . .	134
181	6.4	Leading-order $t\bar{t}W$ and $t\bar{t}Z$ samples used in the signal BDT training. . . . .	134
182	6.5	Table of high-level triggers that we consider in the analysis. . . . .	135
183	6.6	Trigger efficiency scale factors and associated uncertainties. . . . .	136
184	6.7	Effective areas, for electrons and muons. . . . .	142

185	6.8 Requirements on each of the three muon selections. . . . .	145
186	6.9 Criteria for each of the three electron selections. . . . .	145
187	6.10 MVA input discriminating variables . . . . .	153
188	6.11 TMVA input variables ranking for BDTA_GRAD method . . . . .	158
189	6.12 TMVA configuration used in the BDT training. . . . .	158
190	6.13 ROC-integral for all the testing cases. . . . .	160

<sup>191</sup> Chapter 1

<sup>192</sup> INTRODUCTION

<sup>193</sup> **Chapter 2**

<sup>194</sup> **Theoretical approach**

<sup>195</sup> **2.1 Introduction**

<sup>196</sup> The physical description of the universe is a challenge that physicists have faced by  
<sup>197</sup> making theories that refine existing principles and proposing new ones in an attempt  
<sup>198</sup> to embrace emerging facts and phenomena.

<sup>199</sup> At the end of 1940s Julian Schwinger [1] and Richard P. Feynman [2], based on  
<sup>200</sup> the work of Sin-Itiro Tomonaga [3], developed an electromagnetic theory consistent  
<sup>201</sup> with special relativity and quantum mechanics that describes how matter and light  
<sup>202</sup> interact; the so-called *quantum electrodynamics* (QED) was born.

<sup>203</sup> QED has become the guide in the development of theories that describe the uni-  
<sup>204</sup> verse. It was the first example of a quantum field theory (QFT), which is the theore-  
<sup>205</sup> tical framework for building quantum mechanical models that describes particles and  
<sup>206</sup> their interactions. QFT is composed of a set of mathematical tools that combines  
<sup>207</sup> classical fields, special relativity and quantum mechanics, while keeping the quantum  
<sup>208</sup> point particles and locality ideas.

<sup>209</sup> This chapter gives an overview of the standard model of particle physics, starting

210 with a description of the particles and interactions that compose it, followed by a  
 211 description of the electroweak interaction, the Higgs boson and the associated pro-  
 212 duction of Higgs boson and a single top quark ( $tH$ ). The description contained in  
 213 this chapter is based on References [4–6].

## 214 2.2 Standard model of particle physics

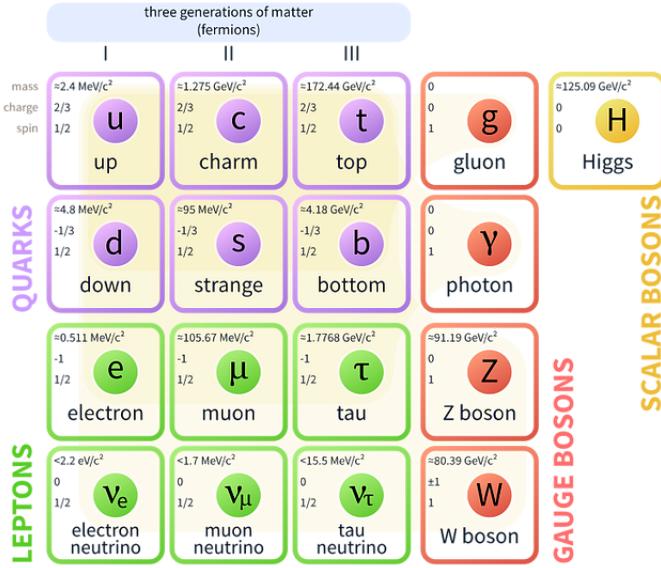
215 Particle physics at the fundamental level is modeled in terms of a collection of inter-  
 216 acting particles and fields in a theory known as the *standard model of particle physics*  
 217 (*SM*). The full picture of the SM is composed of three fields<sup>1</sup> whose excitations are  
 218 interpreted as particles called mediators or force-carriers, a set of fields whose excita-  
 219 tions are interpreted as elementary particles interacting through the exchange of those  
 220 mediators, and a field that gives the mass to elementary particles. Figure 2.1 shows  
 221 the scheme of the SM particles’ organization. In addition, for each of the particles  
 222 in the scheme there exists an antiparticle with the same mass and opposite quantum  
 223 numbers. The existence of antiparticles is a prediction of the relativistic quantum  
 224 mechanics from the solution of the Dirac equation for which a negative energy solu-  
 225 tion is also possible. In some cases a particle is its own anti-particle, like photon or  
 226 Higgs boson.

227 The mathematical formulation of the SM is based on group theory and the use of  
 228 Noether’s theorem [8] which states that for a physical system modeled by a Lagrangian  
 229 that is invariant under a group of transformations a conservation law is expected. For  
 230 instance, a system described by a time-independent Lagrangian is invariant (symmet-  
 231 ric) under time changes (transformations) with the total energy conservation law as

---

<sup>1</sup> The formal and complete treatment of the SM is out of the scope of this document, however a plenty of textbooks describing it at several levels are available in the literature. The treatment in References [5, 6] is quite comprehensive and detailed. Note that gravitational field is not included in the standard model formulation

## Standard Model of Elementary Particles



**Figure 2.1:** Schematic representation of the Standard Model of particle physics. The SM is a theoretical model intended to describe three of the four fundamental forces of the universe in terms of a set of particles and their interactions. [7].

the expected conservation law. In QED, the charge operator ( $Q$ ) is the generator of the  $U(1)$  symmetry which according to the Noether's theorem means that there is a conserved quantity; this conserved quantity is the electric charge and thus the law of conservation of electric charge is established.

In the SM, the symmetry group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  describes three of the four fundamental interactions in nature (see Section 2.2.2): strong interaction (SI), weak interaction (WI) and electromagnetic interactions (EI) in terms of symmetries associated to physical quantities:

- Strong:  $SU(3)_C$  associated to color charge
- Weak:  $SU(2)_L$  associated to weak isospin and chirality
- Electromagnetic:  $U(1)_Y$  associated to weak hypercharge and electric charge

243 It will be shown that the electromagnetic and weak interactions are combined in  
 244 the so-called electroweak interaction where chirality, hypercharge, weak isospin and  
 245 electric charge are the central concepts.

246 **2.2.1 Fermions**

247 The basic constituents of the ordinary matter at the lowest level, which form the set  
 248 of elementary particles in the SM formulation, are quarks and leptons. All of them  
 249 have spin 1/2, therefore they are classified as fermions since they obey Fermi-Dirac  
 250 statistics. There are six *flavors* of quarks and three of leptons organized in three  
 251 generations, or families, as shown in Table 2.1.

		Generation		
		1st	2nd	3rd
Leptons	Charged	Electron (e)	Moun( $\mu$ )	Tau ( $\tau$ )
	Neutral	Electron neutrino ( $\nu_e$ )	Muon neutrino ( $\nu_\mu$ )	Tau neutrino ( $\nu_\tau$ )
Quarks	Up-type	Up (u)	Charm (c)	Top (t)
	Down-type	Down (d)	Strange (s)	Bottom (b)

**Table 2.1:** Fermions of the SM. There are six flavors of quarks and three of leptons, organized in three generations, or families, composed of two pairs of closely related particles. The close relationship is motivated by the fact that each pair of particles is a member of an  $SU(2)_L$  doublet that has an associated invariance under isospin transformations. WI between leptons is limited to the members of the same generation; WI between quarks is not limited but greatly favored, to same generation members.

252

253 There is a mass hierarchy between generations (see Table 2.2), where the higher  
 254 generation particles decays to the lower one, which can explain why the ordinary  
 255 matter is made of particles from the first generation. In the SM, neutrinos are modeled  
 256 as massless particles so they are not subject to this mass hierarchy; however, today it  
 257 is known that neutrinos are massive so the hierarchy could be restated. The reason  
 258 behind this mass hierarchy is one of the most important open questions in particle

259 physics, and it becomes more puzzling when noticing that the mass difference between  
 260 first and second generation fermions is small compared to the mass difference with  
 261 respect to the third generation.

Lepton	Mass (MeV/c <sup>2</sup> )	Quark	Mass (MeV/c <sup>2</sup> )
e	0.51	u	2.2
$\mu$	105.65	c	$1.28 \times 10^3$
$\tau$	1776.86	t	$173.1 \times 10^3$
$\nu_e$	Unknown	d	4.7
$\nu_\mu$	Unknown	s	96
$\tau_\mu$	Unknown	b	$4.18 \times 10^3$

**Table 2.2:** Fermion masses [9]. Generations differ by mass in a way that has been interpreted as a mass hierarchy. Approximate values with no uncertainties are used, for comparison purpose.

262

263        Usually, the second and third generation fermions are produced in high energy  
 264 processes, like the ones recreated in particle accelerators.

265        **2.2.1.1 Leptons**

266        A lepton is an elementary particle that is not subject to the SI. As seen in Table 2.1,  
 267 there are two types of leptons, the charged ones (electron, muon and tau) and the  
 268 neutral ones (the three neutrinos). The electric charge (Q) is the property that gives  
 269 leptons the ability to participate in the EI. From the classical point of view, Q plays  
 270 a central role determining, among others, the strength of the electric field through  
 271 which the electromagnetic force is exerted. It is clear that neutrinos are not affected  
 272 by EI because they don't carry electric charge.

273        Another feature of the leptons that is fundamental in the mathematical description  
 274 of the SM is the chirality, which is closely related to spin and helicity. Helicity  
 275 defines the handedness of a particle by relating its spin and momentum such that

276 if they are parallel then the particle is right-handed; if spin and momentum are  
 277 antiparallel the particle is said to be left-handed. The study of parity conservation  
 278 (or violation) in  $\beta$ -decay has shown that only left-handed electrons/neutrinos or right-  
 279 handed positrons/anti-neutrinos are created [10]; the inclusion of that feature in the  
 280 theory was achieved by using projection operators for helicity, however, helicity is  
 281 frame dependent for massive particles which makes it not Lorentz invariant and then  
 282 another related attribute has to be used: *chirality*.

283 Chirality is a purely quantum attribute which makes it not so easy to describe in  
 284 graphical terms but it defines how the wave function of a particle transforms under  
 285 certain rotations. As with helicity, there are two chiral states, left-handed chiral (L)  
 286 and right-handed chiral (R). In the highly relativistic limit where  $E \approx p \gg m$  helicity  
 287 and chirality converge, becoming exactly the same for massless particles.

288 In the following, when referring to left-handed (right-handed) it will mean left-  
 289 handed chiral (right-handed chiral). The fundamental fact about chirality is that  
 290 while EI and SI are not sensitive to chirality, in WI left-handed and right-handed  
 291 fermions are treated asymmetrically, such that only left-handed fermions and right-  
 292 handed anti-fermions are allowed to couple to WI mediators, which is a violation of  
 293 parity. The way to translate this statement in a formal mathematical formulation is  
 294 based on the isospin symmetry group  $SU(2)_L$ .

295 Each generation of leptons is seen as a weak isospin doublet.<sup>2</sup> The left-handed  
 296 charged lepton and its associated left-handed neutrino are arranged in doublets of  
 297 weak isospin  $T=1/2$  while their right-handed partners are singlets:

$$\begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, l_R := \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, e_R, \mu_R, \tau_R, \nu_{eR}, \nu_{\mu R}, \nu_{\tau R} \quad (2.1)$$

---

<sup>2</sup> The weak isospin is an analogy of the isospin symmetry in strong interaction where neutron and proton are affected equally by strong force but differ in their charge.

298        The isospin third component refers to the eigenvalues of the weak isospin operator  
 299 which for doublets is  $T_3 = \pm 1/2$ , while for singlets it is  $T_3 = 0$ . The physical meaning  
 300 of this doublet-singlet arrangement falls in that the WI couples the two particles in  
 301 the doublet by exchanging the interaction mediator while the singlet member is not  
 302 involved in WI. The main properties of the leptons are summarized in Table 2.3.

303        Although all three flavor neutrinos have been observed, their masses remain un-  
 304 known and only some estimations have been made [11]. The main reason is that  
 305 the flavor eigenstates are not the same as the mass eigenstates which implies that  
 306 when a neutrino is created its mass state is a linear combination of the three mass  
 307 eigenstates and experiments can only probe the squared difference of the masses. The  
 308 Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix encodes the relationship  
 309 between flavor and mass eigenstates.

Lepton	$Q(e)$	$T_3$	$L_e$	$L_\mu$	$L_\tau$	Lifetime (s)
Electron (e)	-1	-1/2	1	0	0	Stable
Electron neutrino( $\nu_e$ )	0	1/2	1	0	0	Unknown
Muon ( $\mu$ )	-1	-1/2	0	1	0	$2.19 \times 10^{-6}$
Muon neutrino ( $\nu_\mu$ )	0	1/2	0	1	0	Unknown
Tau ( $\tau$ )	-1	-1/2	0	0	1	$290.3 \times 10^{-15}$
Tau neutrino ( $\tau_\mu$ )	0	1/2	0	0	1	Unknown

**Table 2.3:** Lepton properties [9]. Q: electric charge,  $T_3$ : weak isospin. Only left-handed leptons and right-handed anti-leptons participate in the WI. Anti-particles with inverted  $T_3$ , Q and lepton number complete the leptons set but are not listed. Right-handed leptons and left-handed anti-leptons, neither listed, form weak isospin singlets with  $T_3 = 0$  and do not take part in the weak interaction.

310

### 311 2.2.1.2 Quarks

312 Quarks are the basic constituents of protons and neutrons. The way quarks join to  
 313 form bound states, called *hadrons*, is through the SI. Quarks are affected by all the

314 fundamental interactions which means that they carry all the four types of charges:  
 315 color, electric charge, weak isospin and mass.

Flavor	$Q(e)$	$I_3$	$T_3$	B	C	S	T	$B'$	Y	Color
Up (u)	2/3	1/2	1/2	1/3	0	0	0	0	1/3	r,b,g
Charm (c)	2/3	0	1/2	1/3	1	0	0	0	4/3	r,b,g
Top(t)	2/3	0	1/2	1/3	0	0	1	0	4/3	r,b,g
Down(d)	-1/3	-1/2	-1/2	1/3	0	0	0	0	1/3	r,b,g
Strange(s)	-1/3	0	-1/2	1/3	0	-1	0	0	-2/3	r,b,g
Bottom(b)	-1/3	0	-1/2	1/3	0	0	0	-1	-2/3	r,b,g

**Table 2.4:** Quark properties [9]. Q: electric charge,  $I_3$ : isospin,  $T_3$ : weak isospin, B: baryon number, C: charmness, S: strangeness, T: topness,  $B'$ : bottomness, Y: hypercharge. Anti-quarks posses the same mass and spin as quarks but all charges (color, flavor numbers) have opposite sign.

316

317 Table 2.4 summarizes the features of quarks, among which the most remarkable  
 318 is their fractional electric charge. Note that fractional charge is not a problem, given  
 319 that quarks are not found isolated, but serves to explain how composed particles are  
 320 formed out of two or more valence quarks<sup>3</sup>.

321 Color charge is responsible for the SI between quarks and is the symmetry ( $SU(3)_C$ )  
 322 that defines the formalism to describe SI. There are three colors: red (r), blue (b)  
 323 and green (g) and their corresponding three anti-colors; thus each quark carries one  
 324 color unit while anti-quarks carries one anti-color unit. As explained in Section 2.2.2,  
 325 quarks are not allowed to be isolated due to the color confinement effect, hence, their  
 326 features have been studied indirectly by observing their bound states created when

- 327 • one quark with a color charge is attracted by an anti-quark with the correspond-  
 328 ing anti-color charge forming a colorless particle called a *meson*.

<sup>3</sup> Hadrons can contain an indefinite number of virtual quarks and gluons, known as the quark and gluon sea, but only the valence quarks determine hadrons' quantum numbers.

329       • three quarks (anti-quarks) with different color (anti-color) charges are attracted  
 330           among them forming a colorless particle called a *baryon* (*anti-baryon*).

331       In practice, when a quark is left alone isolated a process called *hadronization* occurs  
 332       where the quark emits gluons (see Section 2.2.4) which eventually will generate new  
 333       quark-antiquark pairs and so on; those quarks will recombine to form hadrons that  
 334       will decay into leptons. This proliferation of particles looks like a *jet* coming from  
 335       the isolated quark. More details about the hadronization process and jet structure  
 336       will be given in chapter4.

337       In the first version of the quark model (1964), M. Gell-Mann [12] and G. Zweig  
 338       [13, 14] developed a consistent way to classify hadrons according to their properties.  
 339       Only three quarks (u, d, s) were involved in a scheme in which all baryons have baryon  
 340       number  $B=1$  and therefore quarks have  $B=1/3$ ; non-baryons have  $B=0$ . Baryon  
 341       number is conserved in SI and EI which means that single quarks cannot be created  
 342       but in pairs  $q - \bar{q}$ .

343       The scheme organizes baryons in a two-dimensional space ( $I_3$  - Y); Y (hyper-  
 344       charge) and  $I_3$  (isospin) are quantum numbers related by the Gell-Mann-Nishijima  
 345       formula [15, 16]:

$$Q = I_3 + \frac{Y}{2} \quad (2.2)$$

346       where  $Y = B + S + C + T + B'$  are the quantum numbers listed in Table 2.4.

347       There are six quark flavors organized in three generations (see Table 2.1) fol-  
 348       lowing a mass hierarchy which, again, implies that higher generations decay to first  
 349       generation quarks.

	Quarks			$T_3$	$Y_W$	Leptons			$T_3$	$Y_W$
Doublets	$(\frac{u}{d'})_L$	$(\frac{c}{s'})_L$	$(\frac{t}{b'})_L$	$(\frac{1/2}{-1/2})$	$1/3$	$(\nu_e)_L$	$(\nu_\mu)_L$	$(\nu_\tau)_L$	$(\frac{1/2}{-1/2})$	-1
Singlets	$u_R$	$c_R$	$t_R$	0	$4/3$	$\nu_{eR}$	$\nu_{\mu R}$	$\nu_{\tau R}$		
	$d'_R$	$s'_R$	$b'_R$	0	$-2/3$	$e_R$	$\mu_R$	$\tau_R$	0	-2

**Table 2.5:** Fermion weak isospin and weak hypercharge multiplets. Weak hypercharge is calculated through the Gell-Mann-Nishijima formula 2.2 but using the weak isospin and charge for quarks.

351 Isospin doublets of quarks are also defined (see Table 2.5), and same as for neutrinos,  
352 the WI eigenstates are not the same as the mass eigenstates which means that  
353 members of different quark generations are connected by the WI mediator; thus, up-  
354 type quarks are coupled not to down-type quarks (the mass eigenstates) directly but  
355 to a superposition of down-type quarks ( $q'_d$ ; *the weak eigenstates*) via WI according  
356 to:

$$q'_d = V_{CKM} q_d$$

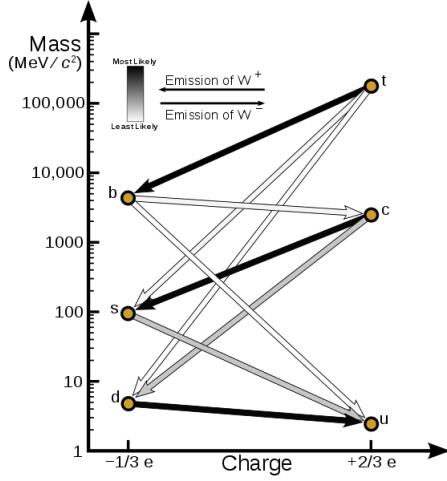
357

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.3)$$

358 where  $V_{CKM}$  is known as Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix [17,18]  
359 given by

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix}. \quad (2.4)$$

360 The weak decays of quarks are represented in the diagram of Figure 2.2; again  
361 the CKM matrix plays a central role since it contains the probabilities for the differ-



**Figure 2.2:** Transformations between quarks through the exchange of a WI. Higher generations quarks decay to first generation quarks by emitting a W boson. The arrow color indicates the likelihood of the transition according to the grey scale in the top left side which represent the CKM matrix parameters [19].

362 ent quark decay channels, in particular, note that quark decays are greatly favored  
 363 between generation members.

364 CKM matrix is a  $3 \times 3$  unitary matrix parametrized by three mixing angles and  
 365 the *CP-mixing phase*; the latter is the parameter responsible for the Charge-Parity  
 366 symmetry violation (CP-violation) in the SM. The fact that the top quark decays  
 367 almost all the time to a bottom quark is exploited in this thesis when making the  
 368 selection of the signal events by requiring the presence of a jet tagged as a jet coming  
 369 from a *b* quark in the final state.

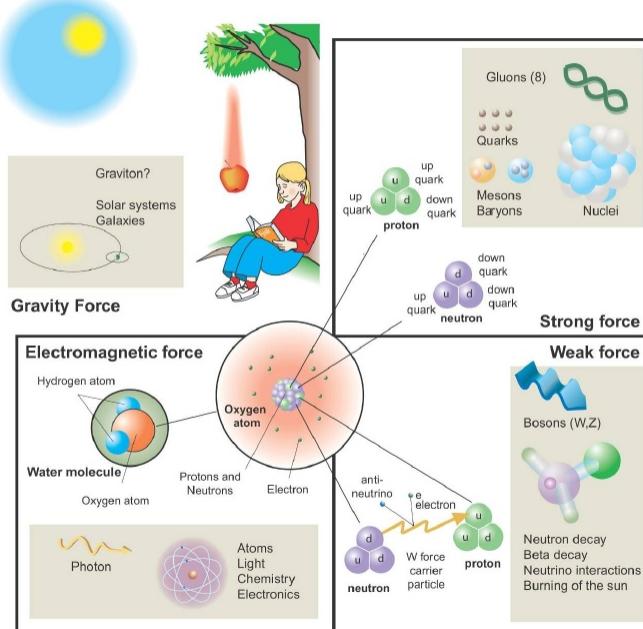
### 370 2.2.2 Fundamental interactions

371 Even though there are many manifestations of force in nature, like the ones repre-  
 372 sented in Figure 2.3, we can classify all of them in four fundamental interactions:

- 373 • *Electromagnetic interaction (EI)* affects particles that are *electrically charged*,  
 374 like electrons and protons. Figure 2.4a. shows a graphical representation, known

## Fundamental interactions.

Illustration: Typoform



Summer School KPI 15 August 2009

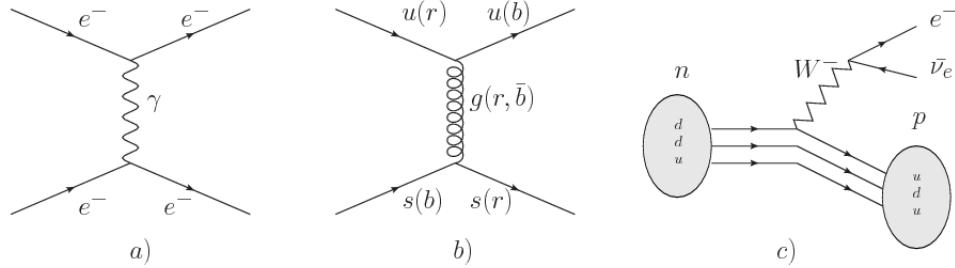
**Figure 2.3:** Fundamental interactions in nature. Despite the many manifestations of forces in nature, we can track all of them back to one of the fundamental interactions. The most common forces are gravity and electromagnetic given that all of us are subject and experience them in everyday life.

375 as *Feynman diagram*, of electron-electron scattering.

- 376     • *Strong interaction (SI)* described by Quantum Chromodynamics (QCD). Hadrons  
 377       like the proton and the neutron have internal structure given that they are com-  
 378       posed of two or more valence quarks<sup>4</sup>. Quarks have fractional electric charge  
 379       which means that they are subject to electromagnetic interaction and in the case  
 380       of the proton they should break apart due to electrostatic repulsion; however,  
 381       quarks are held together inside the hadrons against their electrostatic repulsion  
 382       by the *Strong Force* through the exchange of *gluons*. The analog to the electric  
 383       charge is the *color charge*. Electrons and photons are elementary particles as

<sup>4</sup> Particles made of four and five quarks are exotic states not so common.

384 quarks but they don't carry color charge, therefore they are not subject to SI. A  
 385 Feynman diagram for gluon exchange between quarks is shown in Figure 2.4b.



**Figure 2.4:** Feynman diagrams representing the interactions in SM; a) EI:  $e$ - $e$  scattering; b) SI: gluon exchange between quarks ; c) WI:  $\beta$ -decay

386 • *Weak interaction (WI)* described by the weak theory (WT), is responsible, for  
 387 instance, for the radioactive decay in atoms and the deuterium production  
 388 within the sun. Quarks and leptons are the particles affected by the weak  
 389 interaction; they possess a property called *flavor charge* (see 2.2.1) which can  
 390 be changed by emitting or absorbing one weak force mediator. There are three  
 391 mediators of the *weak force* known as  $Z$  boson in the case of electrically neutral  
 392 flavor changes and  $W^\pm$  bosons in the case of electrically charged flavor changes.  
 393 The *weak isospin* is the WI analog to electric charge in EI, and color charge in  
 394 SI, and defines how quarks and leptons are affected by the weak force. Figure  
 395 2.4c. shows the Feynman diagram of  $\beta$ -decay where a neutron ( $n$ ) is transformed  
 396 in a proton ( $p$ ) by emitting a  $W^-$  particle.

397 • *Gravitational interaction (GI)* described by General Theory of Relativity (GR).  
 398 It is responsible for the structure of galaxies and black holes as well as the  
 399 expansion of the universe. As a classical theory, in the sense that it can be  
 400 formulated without even appeal to the concept of quantization, it implies that  
 401 the space-time is a continuum and predictions can be made without limitation

402 to the precision of the measurement tools. The latter represents a direct con-  
 403 tradiction of the quantum mechanics principles. Gravity is deterministic while  
 404 quantum mechanics is probabilistic; despite that, efforts to develop a quantum  
 405 theory of gravity have predicted the *graviton* as mediator of the gravitational  
 406 force<sup>5</sup>.

Interaction	Acts on	Relative strength	Range (m)	Mediators
Electromagnetic (QED)	Electrically charged particles	$10^{-2}$	Infinite	Photon
Strong (QCD)	Quarks and gluons	1	$10^{-15}$	Gluon
Weak (WI)	Leptons and quarks	$10^{-6}$	$10^{-18}$	$W^\pm, Z$
Gravitational (GI)	Massive particles	$10^{-39}$	Infinite	Graviton

**Table 2.6:** Fundamental interactions features [20].

407

408 Table 2.6 summarizes the main features of the fundamental interactions. The  
 409 relative strength of the fundamental forces reveals the meaning of strong and weak;  
 410 in a context where the relative strength of the SI is 1, the EI is about hundred times  
 411 weaker and WI is about million times weaker than the SI. A good description on how  
 412 the relative strength and range of the fundamental interactions are calculated can  
 413 be found in References [20, 21]. In the everyday life, only EI and GI are explicitly  
 414 experienced due to the range of these interactions; i.e., at the human scale distances  
 415 only EI and GI have appreciable effects, in contrast to SI which at distances greater  
 416 than  $10^{-15}$ m become negligible.

### 417 2.2.3 Gauge invariance.

418 QED was built successfully on the basis of the classical electrodynamics theory (CED)  
 419 of Maxwell and Lorentz, following theoretical and experimental requirements imposed

---

<sup>5</sup> Actually a wide variety of theories have been developed in an attempt to describe gravity; some famous examples are string theory and supergravity.

420 by

- 421     • Lorentz invariance: independence on the reference frame.
- 422     • Locality: interacting fields are evaluated at the same space-time point to avoid  
423       action at a distance.
- 424     • Renormalizability: physical predictions are finite and well defined.
- 425     • Particle spectrum, symmetries and conservation laws already known must emerge  
426       from the theory.
- 427     • Local gauge invariance.

428     The gauge invariance requirement reflects the fact that the fundamental fields  
429     cannot be directly measured but associated fields which are the observables. Electric  
430     (**E**) and magnetic (**B**) fields in CED are associated with the electric scalar potential  
431      $V$  and the vector potential **A**. In particular, **E** can be obtained by measuring the  
432     change in the space of the scalar potential ( $\Delta V$ ); however, two scalar potentials  
433     differing by a constant  $f$  correspond to the same electric field. The same happens  
434     in the case of the vector potential **A**; thus, different configurations of the associated  
435     fields result in the same set of values of the observables. The freedom in choosing one  
436     particular configuration is known as *gauge freedom*; the transformation law connecting  
437     two configurations is known as *gauge transformation* and the fact that the observables  
438     are not affected by a gauge transformation is called *gauge invariance*.

439     When the gauge transformation:

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} - \Delta f \\ V &\rightarrow V - \frac{\partial f}{\partial t} \end{aligned} \tag{2.5}$$

440 is applied to Maxwell equations, they are still satisfied and the fields remain invariant.  
 441 Thus, CED is invariant under gauge transformations and is called a *gauge theory*.  
 442 The set of all gauge transformations form the *symmetry group* of the theory, which  
 443 according to the group theory, has a set of *group generators*. The number of group  
 444 generators determine the number of *gauge fields* of the theory.

445 As mentioned in the first lines of Section 2.2, QED has one symmetry group ( $U(1)$ )  
 446 with one group generator (the  $Q$  operator) and one gauge field (the electromagnetic  
 447 field  $A^\mu$ ). In CED there is not a clear definition, beyond the historical convention,  
 448 of which fields are the fundamental and which are the associated, but in QED the  
 449 fundamental field is  $A^\mu$ . When a gauge theory is quantized, the gauge fields are  
 450 quantized and their quanta are called *gauge bosons*. The word boson characterizes  
 451 particles with integer spin which obey Bose-Einstein statistics.

452 As will be detailed in Section 2.3, interactions between particles in a system can  
 453 be obtained by considering first the Lagrangian density of free particles in the sys-  
 454 tem, which of course is incomplete because the interaction terms have been left out,  
 455 and demanding global phase transformation invariance. Global phase transforma-  
 456 tion means that a gauge transformation is performed identically to every point  
 457 in the space<sup>6</sup> and the Lagrangian remains invariant. Then, the global transforma-  
 458 tion is promoted to a local phase transformation (this time the gauge transformation  
 459 depends on the position in space) and again invariance is required.

---

<sup>6</sup> Here space corresponds to the 4-dimensional space i.e. space-time.

460 Due to the space dependence of the local transformation, the Lagrangian density is  
 461 not invariant anymore. In order to reinstate the gauge invariance, the gauge covariant  
 462 derivative is introduced in the Lagrangian and with it the gauge field responsible for  
 463 the interaction between particles in the system. The new Lagrangian density is gauge  
 464 invariant, includes the interaction terms needed to account for the interactions and  
 465 provides a way to explain the interaction between particles through the exchange of  
 466 the gauge boson.

467 This recipe was used to build QED and the theories that aim to explain the  
 468 fundamental interactions.

#### 469 2.2.4 Gauge bosons

470 The importance of the gauge bosons comes from the fact that they are the force  
 471 mediators or force carriers. The features of the gauge bosons reflect those of the fields  
 472 they represent and they are extracted from the Lagrangian density used to describe  
 473 the interactions. In Section 2.3, it will be shown how the gauge bosons of the EI and  
 474 WI emerge from the electroweak Lagrangian. The SI gauge bosons features are also  
 475 extracted from the SI Lagrangian but it is not detailed in this document. The main  
 476 features of the SM gauge bosons will be briefly presented below and summarized in  
 477 Table 2.7.

- 478 • **Photon.** EI occurs when the photon couples to (is exchanged between) parti-  
 479 cles carrying electric charge; however, The photon itself does not carry electric  
 480 charge, therefore, there is no coupling between photons. Given that the photon  
 481 is massless the EI is of infinite range, i.e., electrically charged particles interact  
 482 even if they are located far away one from each other; this also implies that  
 483 photons always move with the speed of light.

- 484     • **Gluon.** SI is mediated by gluons which just as photons are massless. They  
 485       carry one unit of color charge and one unit of anticolor charge, hence, gluons  
 486       can couple to other gluons. As a result, the range of the SI is not infinite  
 487       but very short due to the attraction between gluons, giving rise to the *color*  
 488       *confinement* which explains why color charged particles cannot be isolated but  
 489       live within composite particles, like quarks inside protons.
- 490     • **W, Z.**  $W^\pm$  and Z, are massive which explains their short-range. Given that  
 491       the WI is the only interaction that can change the flavor of the interacting  
 492       particles, the W boson is the responsible for the nuclear transmutation where  
 493       a neutron is converted into a proton or vice versa with the involvement of an  
 494       electron and a neutrino (see Figure 2.4c). The Z boson is the responsible for the  
 495       neutral weak processes like neutrino elastic scattering where no electric charge  
 496       but momentum transference is involved. WI gauge bosons carry isospin charge  
 497       which makes interaction between them possible.

Interaction	Mediator	Electric charge (e)	Color charge	Weak Isospin	mass (GeV/c <sup>2</sup> )
Electromagnetic	Photon ( $\gamma$ )	0	No	0	0
Strong	Gluon (g)	0	Yes -octet	No	0
Weak	$W^\pm$	$\pm 1$	No	$\pm 1$	$80.385 \pm 0.015$
	Z	0	No	0	$91.188 \pm 0.002$

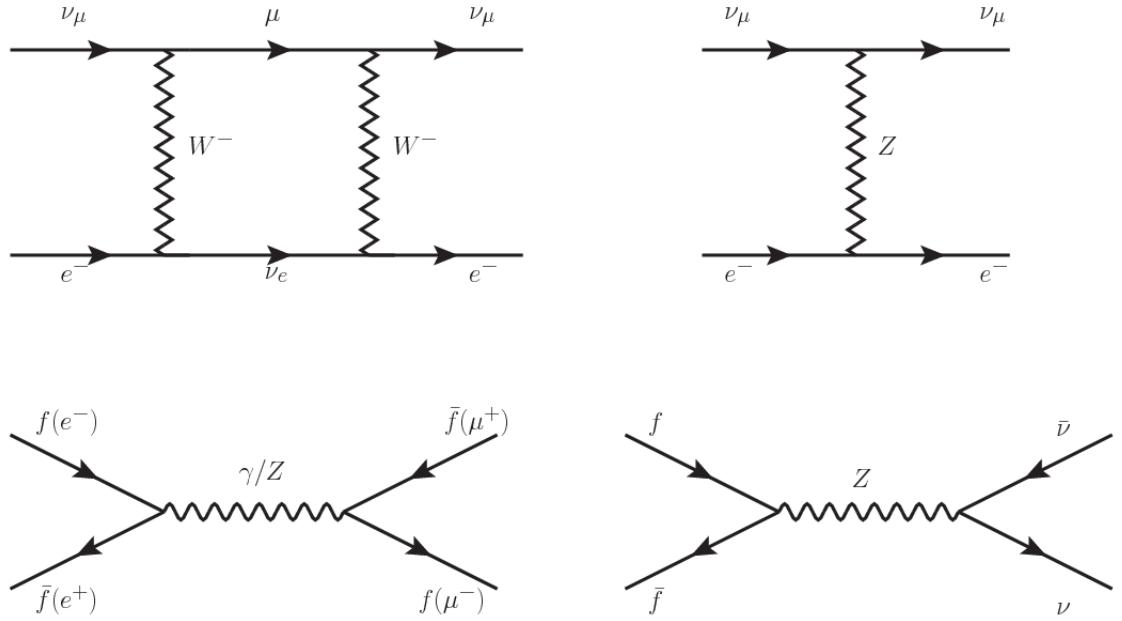
**Table 2.7:** SM gauge bosons main features [9].

498

499     **2.3 Electroweak unification and the Higgs  
 500                   mechanism**

501     Physicists dream of building a theory that contains all the interactions in one single  
 502       interaction, i.e., showing that at some scale in energy all the four fundamental inter-

actions are unified and only one interaction emerges in a *Theory of everything*. The first sign of the feasibility of such unification came from success in the construction of the CED. Einstein spent years trying to reach that full unification, which by 1920 only involved electromagnetism and gravity, with no success; however, a new partial unification was achieved in the 1960's, when S.Glashow [22], A.Salam [23] and S.Weinberg [24] independently proposed that electromagnetic and weak interactions are two manifestations of a more general interaction called *electroweak interaction* (EWI). EWI was developed by following the useful prescription provided by QED and the gauge invariance principles.



**Figure 2.5:** Top:  $\nu_\mu - e^-$  scattering going through charged currents (left) and neutral currents (right). Bottom: neutral current processes for charged fermions (left) and involving neutrinos (right). While neutral current processes involving only charged fermions can proceed through EI or WI, those involving neutrinos can only proceed via WI.

The *classic* weak theory developed by Fermi, did not have the concept of the  $W$  boson but instead it was treated as a point interaction with the dimensionful constant  $G_F$  associated with it. It works really well at low energies very far off the  $W$  mass

shell. When going up in energy, the theory of weak interactions involving the W boson is capable of explaining the  $\beta$ -decay and in general the processes mediated by  $W^\pm$  bosons. However, there were some processes like the  $\nu_\mu - e$  scattering which would require the exchange of two W bosons (see Figure 2.5 top diagrams) giving rise to divergent loop integrals and then non-finite predictions. The EWI theory, by including neutral currents involving fermions via the exchange of a neutral bosons Z, overcomes those divergences and the predictions become realistic.

Neutral weak interaction vertices conserve flavor in the same way as the electromagnetic vertices do, but additionally, the Z boson can couple to neutrinos which implies that processes involving charged fermions can proceed through EI or WI but processes involving neutrinos can proceed only through WI.

The prescription to build a gauge theory of the WI consists of proposing a free field Lagrangian density that includes the particles involved; next, by requesting invariance under global phase transformations first and generalizing to local phase transformations invariance later, the conserved currents are identified and interactions are generated by introducing gauge fields. Given that the goal is to include the EI and WI in a single theory, the group symmetry considered should be a combination of  $SU(2)_L$  and  $U(1)_{em}$ , however the latter cannot be used directly because the EI treats left and right-handed particles indistinctly in contrast to the former. Fortunately, the weak hypercharge, which is a combination of the weak isospin and the electric charge (Eqn. 2.2) is suitable to be used since it is conserved by the EI and WI. Thus, the symmetry group to be considered is

$$G \equiv SU(2)_L \otimes U(1)_Y \quad (2.6)$$

The following treatment applies to any of the fermion generations, but for sim-

538 plicity the first generation of leptons will be considered [5, 6, 25, 26].

539 Given the first generation of leptons

$$\psi_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad \psi_2 = \nu_{eR}, \quad \psi_3 = e_R^- \quad (2.7)$$

540 the charged fermionic currents are given by

$$J_\mu \equiv J_\mu^+ = \bar{\nu}_{eL} \gamma_\mu e_L, \quad J_\mu^\dagger \equiv J_\mu^- = \bar{e}_L \gamma_\mu \nu_{eL} \quad (2.8)$$

541 and the free Lagrangian is given by

$$\mathcal{L}_0 = \sum_{j=1}^3 i\bar{\psi}_j(x) \gamma^\mu \partial_\mu \psi_j(x). \quad (2.9)$$

542 Mass terms are included directly in the QED free Lagrangians since they preserve  
 543 the invariance under the symmetry transformations involved which treat left and right  
 544 handed particles similarly, however mass terms of the form

$$m_W^2 W_\mu^\dagger(x) W^\mu(x) + \frac{1}{2} m_Z^2 Z_\mu(x) Z^\mu(x) - m_e \bar{\psi}_e(x) \psi_e(x) \quad (2.10)$$

545 which represent the mass of  $W^\pm$ , Z and electrons, are not invariant under G trans-  
 546 formations, therefore the gauge fields described by the EWI are in principle massless.

547 Experiments have shown that the EWI gauge fields are not massless [27–30];  
 548 however, they have to acquire mass through a mechanism compatible with the gauge  
 549 invariance; that mechanism is known as the *Higgs mechanism* and will be considered  
 550 later in this Section. The global transformations in the combined symmetry group G  
 551 can be written as

$$\begin{aligned}
\psi_1(x) &\xrightarrow{G} \psi'_1(x) \equiv U_Y U_L \psi_1(x), \\
\psi_2(x) &\xrightarrow{G} \psi'_2(x) \equiv U_Y \psi_2(x), \\
\psi_3(x) &\xrightarrow{G} \psi'_3(x) \equiv U_Y \psi_3(x)
\end{aligned} \tag{2.11}$$

552 where  $U_L$  represent the  $SU(2)_L$  transformation acting only on the weak isospin dou-  
 553 blet and  $U_Y$  represent the  $U(1)_Y$  transformation acting on all the weak isospin mul-  
 554 tiplets. Explicitly

$$U_L \equiv \exp\left(i \frac{\sigma_i}{2} \alpha^i\right), \quad U_Y \equiv \exp(i y_i \beta) \quad (i = 1, 2, 3) \tag{2.12}$$

555 with  $\sigma_i$  the Pauli matrices and  $y_i$  the weak hypercharges. In order to promote the  
 556 transformations from global to local while keeping the invariance, it is required that  
 557  $\alpha^i = \alpha^i(x)$ ,  $\beta = \beta(x)$  and the replacement of the ordinary derivatives by the covariant  
 558 derivatives

$$\begin{aligned}
D_\mu \psi_1(x) &\equiv \left[ \partial_\mu + ig\sigma_i W_\mu^i(x)/2 + ig'y_1 B_\mu(x) \right] \psi_1(x) \\
D_\mu \psi_2(x) &\equiv \left[ \partial_\mu + ig'y_2 B_\mu(x) \right] \psi_2(x) \\
D_\mu \psi_3(x) &\equiv \left[ \partial_\mu + ig'y_3 B_\mu(x) \right] \psi_3(x)
\end{aligned} \tag{2.13}$$

559 introducing in this way four gauge fields,  $W_\mu^i(x)$  and  $B_\mu(x)$ , in the process. The  
 560 covariant derivatives (Eqn. 2.13) are required to transform in the same way as fermion  
 561 fields  $\psi_i(x)$  themselves, therefore, the gauge fields transform as:

$$\begin{aligned} B_\mu(x) &\xrightarrow{G} B'_\mu(x) \equiv B_\mu(x) - \frac{1}{g'} \partial_\mu \beta(x) \\ W_\mu^i(x) &\xrightarrow{G} W_\mu^{i\prime}(x) \equiv W_\mu^i(x) - \frac{i}{g} \partial_\mu \alpha_i(x) - \varepsilon_{ijk} \alpha_i(x) W_\mu^j(x). \end{aligned} \quad (2.14)$$

562 The G invariant version of the Lagrangian density 2.9 can be written as

$$\mathcal{L}_0 = \sum_{j=1}^3 i \bar{\psi}_j(x) \gamma^\mu D_\mu \psi_j(x) \quad (2.15)$$

563 where free massless fermion and gauge fields and fermion-gauge boson interactions  
 564 are included. The EWI Lagrangian density must additionally include kinetic terms  
 565 for the gauge fields ( $\mathcal{L}_G$ ) which are built from the field strengths, according to

$$B_{\mu\nu}(x) \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \quad (2.16)$$

$$W_{\mu\nu}^i(x) \equiv \partial_\mu W_\nu^i(x) - \partial_\nu W_\mu^i(x) - g \varepsilon^{ijk} W_\mu^j W_\nu^k \quad (2.17)$$

566 the last term in Eqn. 2.17 is added in order to hold the gauge invariance; therefore,

$$\mathcal{L}_G = -\frac{1}{4} B_{\mu\nu}(x) B^{\mu\nu}(x) - \frac{1}{4} W_{\mu\nu}^i(x) W_i^{\mu\nu}(x) \quad (2.18)$$

567 which contains not only the free gauge fields contributions, but also the gauge fields  
 568 self-interactions and interactions among them.

569 The three weak isospin conserved currents resulting from the  $SU(2)_L$  symmetry  
 570 are given by

$$J_\mu^i(x) = \frac{1}{2} \bar{\psi}_1(x) \gamma_\mu \sigma^i \psi_1(x) \quad (2.19)$$

571 while the weak hypercharge conserved current resulting from the  $U(1)_Y$  symmetry is  
 572 given by

$$J_\mu^Y = \sum_{j=1}^3 \bar{\psi}_j(x) \gamma_\mu y_j \psi_j(x) \quad (2.20)$$

573 In order to evaluate the electroweak interactions modeled by an isos triplet field  
 574  $W_\mu^i$  that couples to isospin currents  $J_\mu^i$  with strength  $g$  and additionally the singlet  
 575 field  $B_\mu$  which couples to the weak hypercharge current  $J_\mu^Y$  with strength  $g'/2$ . The  
 576 interaction Lagrangian density to be considered is

$$\mathcal{L}_I = -g J^{i\mu}(x) W_\mu^i(x) - \frac{g'}{2} J^{Y\mu}(x) B_\mu(x) \quad (2.21)$$

577 Note that the weak isospin currents are not the same as the charged fermionic cur-  
 578 rents that were used to describe the WI (Eqn. 2.8), since the weak isospin eigenstates  
 579 are not the same as the mass eigenstates, but they are closely related

$$J_\mu = \frac{1}{2}(J_\mu^1 + iJ_\mu^2), \quad J_\mu^\dagger = \frac{1}{2}(J_\mu^1 - iJ_\mu^2). \quad (2.22)$$

580 The same happens with the gauge fields  $W_\mu^i$  which are related to the mass eigen-  
 581 states  $W^\pm$  by

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2), \quad W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2). \quad (2.23)$$

582 The fact that there are three weak isospin conserved currents is an indication that  
 583 in addition to the charged fermionic currents, which couple charged to neutral leptons,  
 584 there should be a neutral fermionic current that does not involve electric charge  
 585 exchange; therefore, it couples neutral fermions or fermions of the same electric charge.  
 586 The third weak isospin current contains a term that is similar to the electromagnetic

587 current ( $j_\mu^{em}$ ), indicating that there is a relation between them and resembling the  
 588 Gell-Mann-Nishijima formula 2.2 adapted to electroweak interactions

$$Q = T_3 + \frac{Y_W}{2}. \quad (2.24)$$

589 Just as  $Q$  generates the  $U(1)_{em}$  symmetry, the weak hypercharge generates the  
 590  $U(1)_Y$  symmetry as said before. It is possible to write the relationship in terms of  
 591 the currents as

$$j_\mu^{em} = J_\mu^3 + \frac{1}{2} J_\mu^Y. \quad (2.25)$$

592 The neutral gauge fields  $W_\mu^3$  and  $B_\mu$  cannot be directly identified with the  $Z$   
 593 and the photon fields since the photon interacts similarly with left and right-handed  
 594 fermions; however, they are related through a linear combination given by

$$\begin{aligned} A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \end{aligned} \quad (2.26)$$

where  $\theta_W$  is known as the *Weinberg angle*. The interaction Lagrangian is now given by

$$\begin{aligned} \mathcal{L}_I = -\frac{g}{\sqrt{2}}(J^\mu W_\mu^+ + J^{\mu\dagger} W_\mu^-) - &\left(g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{J_\mu^Y}{2}\right) A^\mu \\ &- \left(g \cos \theta_W J_\mu^3 - g' \sin \theta_W \frac{J_\mu^Y}{2}\right) Z^\mu \end{aligned} \quad (2.27)$$

595 the first term is the weak charged current interaction, while the second term is the

596 electromagnetic interaction under the condition

$$g \sin \theta_W = g' \cos \theta_W = e, \quad \frac{g'}{g} = \tan \theta_W \quad (2.28)$$

597 contained in the Eqn.2.25; the third term is the neutral weak current.

598

599 Note that the neutral fields transformation given by the Eqn. 2.26 can be written  
600 in terms of the coupling constants  $g$  and  $g'$  as:

$$A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + g'^2}}, \quad Z_\mu = \frac{g W_\mu^3 - g' B_\mu}{\sqrt{g^2 + g'^2}}. \quad (2.29)$$

601 So far, the Lagrangian density describing the non-massive EWI is:

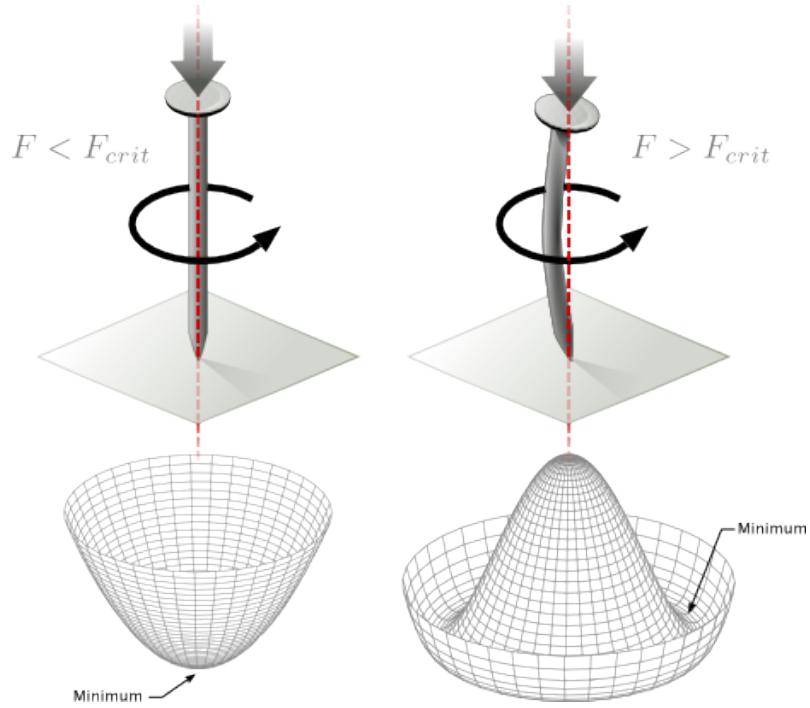
$$\mathcal{L}_{nmEWI} = \mathcal{L}_0 + \mathcal{L}_G \quad (2.30)$$

602 where fermion and gauge fields have been considered massless because their regular  
603 mass terms are manifestly non invariant under  $G$  transformations; therefore, masses  
604 have to be generated in a gauge invariant way. The mechanism by which this goal is  
605 achieved is known as the *Higgs mechanism* and is closely connected to the concept of  
606 *spontaneous symmetry breaking*.

### 607 2.3.1 Spontaneous symmetry breaking (SSB)

608 Figure 2.6 left shows a steel nail (top) which is subject to an external force; the form  
609 of the potential energy is also shown (bottom).

610 Before reaching the critical force value, the system has rotational symmetry with  
611 respect to the nail axis; however, after the critical force value is reached the nail buck-



**Figure 2.6:** Spontaneous symmetry breaking mechanism. The steel nail, subject to an external force (top left), has rotational symmetry with respect to its axis. When the external force overcomes a critical value the nail buckles (top right) choosing a minimal energy state (ground state) and thus *breaking spontaneously the rotational symmetry*. The potential energy (bottom) changes but holds the rotational symmetry; however, an infinite number of asymmetric ground states are generated and circularly distributed in the bottom of the potential [31].

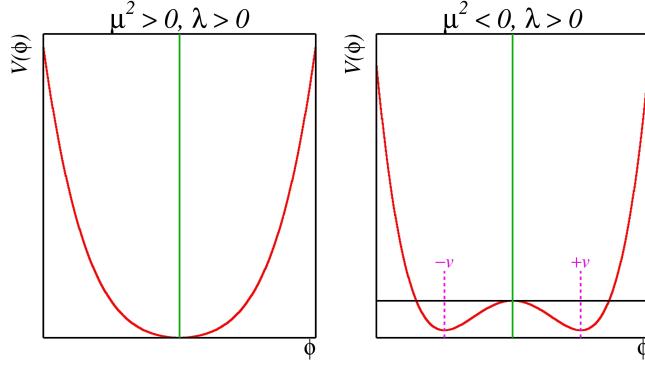
612 les (top right). The form of the potential energy (bottom right) changes appearing a  
 613 set of infinity minima but preserving its rotational symmetry. Right before the nail  
 614 buckles there is no indication of the direction the nail will bend because any of the  
 615 directions are equivalent, but once the nail bends, choosing a direction, an arbitrary  
 616 minimal energy state (ground state) is selected and it does not share the system's  
 617 rotational symmetry. This mechanism for reaching an asymmetric ground state is  
 618 known as *spontaneous symmetry breaking*.

619 The lesson from this analysis is that the way to introduce the SSB mechanism  
 620 into a system is by adding the appropriate potential to it.

621 Figure 2.7 shows a plot of the potential  $V(\phi)$  in the case of a scalar field  $\phi$

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.31)$$

622 If  $\mu^2 > 0$  the potential has only one minimum at  $\phi = 0$  and describes a scalar field  
 623 with mass  $\mu$ . If  $\mu^2 < 0$  the potential has a local maximum at  $\phi = 0$  and two minima  
 624 at  $\phi = \pm\sqrt{-\mu^2/\lambda}$  which enables the SSB mechanism to work.



**Figure 2.7:** Shape of the potential  $V(\phi)$  for  $\lambda > 0$  and:  $\mu^2 > 0$  (left) and  $\mu^2 < 0$  (right). The case  $\mu^2 < 0$  corresponds to the potential suitable for introducing the SSB mechanism by choosing one of the two ground states which are connected via reflection symmetry. [31].

625 In the case of a complex scalar field  $\phi(x)$

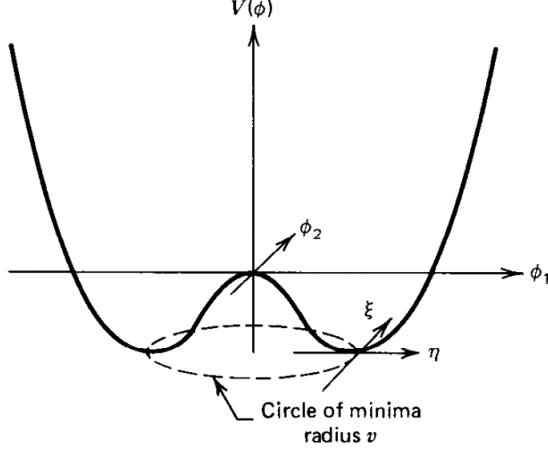
$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (2.32)$$

626 the Lagrangian (invariant under global  $U(1)$  transformations) is given by

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.33)$$

627 where an appropriate potential has been added in order to introduce the SSB.

628 As seen in Figure 2.8, the potential has now an infinite number of minima circularly  
 629 distributed along the  $\xi$ -direction which makes possible the occurrence of the SSB by  
 630 choosing an arbitrary ground state; for instance,  $\xi = 0$ , i.e.  $\phi_1 = v, \phi_2 = 0$



**Figure 2.8:** Potential for complex scalar field. There is a circle of minima of radius  $v$  along the  $\xi$ -direction [6].

$$\phi_0 = \frac{v}{\sqrt{2}} \exp(i\xi) \xrightarrow{\text{SSB}} \phi_0 = \frac{v}{\sqrt{2}} \quad (2.34)$$

631 As usual, excitations over the ground state are studied by making an expansion  
 632 about it; thus, the excitations can be parametrized as:

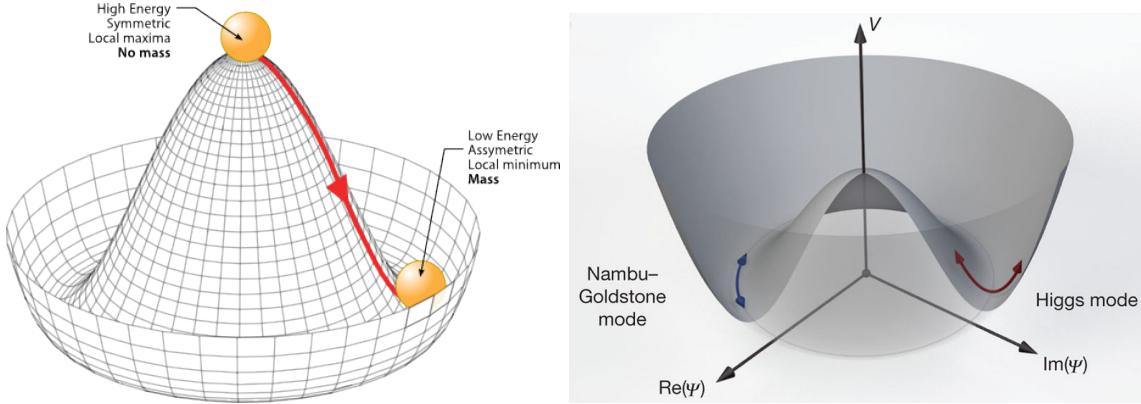
$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)) \quad (2.35)$$

633 which when substituted into Eqn. 2.33 produces a Lagrangian in terms of the new  
 634 fields  $\eta$  and  $\xi$

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu^2 \eta^2 - V(\phi_0) - \lambda v \eta (\eta^2 + \xi^2) - \frac{\lambda}{4}(\eta^2 + \xi^2)^2 \quad (2.36)$$

635 where the last two terms represent the interactions and self-interaction between the  
 636 two fields  $\eta$  and  $\xi$ . The particular feature of the SSB mechanism is revealed when  
 637 looking to the first three terms of  $\mathcal{L}'$ . Before the SSB, only the massless  $\phi$  field is

638 present in the system; after the SSB there are two fields of which the  $\eta$ -field has  
 639 acquired mass  $m_\eta = \sqrt{-2\mu^2}$  while the  $\xi$ -field is still massless (see Figure 2.9).



**Figure 2.9:** SSB mechanism for a complex scalar field [31, 32].

640 Thus, the SSB mechanism serves as a method to generate mass but as a side ef-  
 641 fect a massless field is introduced in the system. This fact is known as the Goldstone  
 642 theorem and states that a massless scalar field appears in the system for each con-  
 643 tinuous symmetry spontaneously broken. Another version of the Goldstone theorem  
 644 states that “if a Lagrangian is invariant under a continuous symmetry group  $G$ , but  
 645 the vacuum is only invariant under a subgroup  $H \subset G$ , then there must exist as many  
 646 massless spin-0 particles (Nambu-Goldstone bosons) as broken generators.” [26] The  
 647 Nambu-Goldstone boson can be understood considering that the potential in the  $\xi$ -  
 648 direction is flat so excitations in that direction are not energy consuming and thus  
 649 represent a massless state.

### 650 2.3.2 Higgs mechanism

651 When the SSB mechanism is introduced in the formulation of the EWI in an attempt  
 652 to generate the mass of the so far massless gauge bosons and fermions, an interesting  
 653 effect is revealed. In order to keep the  $G$  symmetry group invariance and generate

654 the mass of the EW gauge bosons, a G invariant Lagrangian density ( $\mathcal{L}_S$ ) has to be  
 655 added to the non massive EWI Lagrangian (Eqn. 2.30)

$$\mathcal{L}_S = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0, \mu^2 < 0 \quad (2.37)$$

$$D_\mu \phi = \left( i\partial_\mu - g \frac{\sigma_i}{2} W_\mu^i - g' \frac{Y}{2} B_\mu \right) \phi \quad (2.38)$$

656  $\phi$  has to be an isospin doublet of complex scalar fields so it preserves the G invariance;  
 657 thus  $\phi$  can be defined as:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.39)$$

658 The minima of the potential are defined by

$$\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}. \quad (2.40)$$

659 The choice of the ground state is critical. By choosing a ground state, invariant  
 660 under  $U(1)_{em}$  gauge symmetry, the photon will remain massless and the  $W^\pm$  and  $Z$   
 661 bosons masses will be generated which is exactly what is needed. In that sense, the  
 662 best choice corresponds to a weak isospin doublet with  $T_3 = -1/2$ ,  $Y_W = 1$  and  $Q = 0$   
 663 which defines a ground state with  $\phi_1 = \phi_2 = \phi_4$  and  $\phi_3 = v$ :

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 \equiv -\frac{\mu^2}{\lambda}. \quad (2.41)$$

664 where the vacuum expectation value  $v$  is fixed by the Fermi coupling  $G_F$  according  
 665 to  $v = (\sqrt{2}G_F)^{1/2} \approx 246$  GeV.

666 The G symmetry has been broken and three Nambu-Goldstone bosons will appear.

667 The next step is to expand  $\phi$  about the chosen ground state as:

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{i}{v} \sigma_i \theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \approx \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1(x) + i\theta_2(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \quad (2.42)$$

668 to describe fluctuations from the ground state  $\phi_0$ . The fields  $\theta_i(x)$  represent the  
 669 Nambu-Goldstone bosons while  $H(x)$  is known as *Higgs field*. The fundamental fea-  
 670 ture of the parametrization used is that the dependence on the  $\theta_i(x)$  fields is factored  
 671 out in a global phase that can be eliminated by taking the physical *unitary gauge*  
 672  $\theta_i(x) = 0$ . Therefore the expansion about the ground state is given by:

$$\phi(x) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (2.43)$$

673 which when substituted into  $\mathcal{L}_S$  (Eqn. 2.37) results in a Lagrangian containing the  
 674 now massive three gauge bosons  $W^\pm, Z$ , one massless gauge boson (photon) and  
 675 the new Higgs field ( $H$ ). The three degrees of freedom corresponding to the Nambu-  
 676 Goldstone bosons are now integrated into the massive gauge bosons as their lon-  
 677 gitudinal polarizations which were not available when they were massless particles.  
 678 The effect by which vector boson fields acquire mass after an spontaneous symmetry  
 679 breaking, but without an explicit gauge invariance breaking is known as the *Higgs*  
 680 *mechanism*.

681 The mechanism was proposed by three independent groups: F.Englert and R.Brout  
 682 in August 1964 [33], P.Higgs in October 1964 [34] and G.Guralnik, C.Hagen and  
 683 T.Kibble in November 1964 [35]; however, its importance was not realized until  
 684 S.Glashow [22], A.Salam [23] and S.Weinberg [24], independently, proposed that elec-  
 685 tromagnetic and weak interactions are two manifestations of a more general interac-  
 686 tion called *electroweak interaction* in 1967.

687 **2.3.3 Masses of the gauge bosons**

688 The masses of the gauge bosons are extracted by evaluating the kinetic part of La-  
 689 grangian  $\mathcal{L}_S$  in the ground state (known also as the vacuum expectation value), i.e.,

$$\left| \left( \partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - i \frac{g'}{2} B_\mu \right) \phi_0 \right|^2 = \left( \frac{1}{2} v g \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (W_\mu^3, B_\mu) \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} \quad (2.44)$$

690 comparing with the typical mass term for a charged boson  $M_W^2 W^+ W^-$

$$M_W = \frac{1}{2} v g. \quad (2.45)$$

The second term in the right side of the Eqn.2.44 comprises the masses of the neutral bosons, but it needs to be written in terms of the gauge fields  $Z_\mu$  and  $A_\mu$  in order to be compared to the typical mass terms for neutral bosons, therefore using Eqn. 2.29

$$\begin{aligned} \frac{1}{8} v^2 [g^2 (W_\mu^3)^2 - 2gg' W_\mu^3 B^\mu + g'^2 B_\mu^2] &= \frac{1}{8} v^2 [g W_\mu^3 - g' B_\mu]^2 + 0[g' W_\mu^3 + g B_\mu]^2 \quad (2.46) \\ &= \frac{1}{8} v^2 [\sqrt{g^2 + g'^2} Z_\mu]^2 + 0[\sqrt{g^2 + g'^2} A_\mu]^2 \end{aligned}$$

691 and then

$$M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}, \quad M_A = 0 \quad (2.47)$$

692 **2.3.4 Masses of the fermions**

693 The lepton mass terms can be generated by introducing a gauge invariant Lagrangian  
 694 term describing the Yukawa coupling between the lepton field and the Higgs field

$$\mathcal{L}_{Yl} = -G_l \left[ (\bar{\nu}_l, \bar{l})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} l_R + \bar{l}_R (\phi^-, \bar{\phi}^0) \begin{pmatrix} \nu_l \\ l \end{pmatrix} \right], \quad l = e, \mu, \tau. \quad (2.48)$$

695 After the SSB and replacing the usual field expansion about the ground state  
 696 (Eqn.2.41) into  $\mathcal{L}_{Yl}$ , the mass term arises

$$\mathcal{L}_{Yl} = -m_l (\bar{l}_L l_R + \bar{l}_R l_L) - \frac{m_l}{v} (\bar{l}_L l_R + \bar{l}_R l_L) H = -m_l \bar{l} \left( 1 + \frac{H}{v} \right) \quad (2.49)$$

697

$$m_l = \frac{G_l}{\sqrt{2}} v \quad (2.50)$$

698 where the additional term represents the lepton-Higgs interaction. The quark masses  
 699 are generated in a similar way as lepton masses but for the upper member of the  
 700 quark doublet a different Higgs doublet is needed:

$$\phi_c = -i\sigma_2 \phi^* = \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix}. \quad (2.51)$$

701 Additionally, given that the quark isospin doublets are not constructed in terms  
 702 of the mass eigenstates but in terms of the flavor eigenstates, as shown in Table 2.5,  
 703 the coupling parameters will be related to the CKM matrix elements; thus the quark  
 704 Lagrangian is given by:

$$\mathcal{L}_{Yq} = -G_d^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{jR} - G_u^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} u_{jR} + h.c. \quad (2.52)$$

705 with  $i, j = 1, 2, 3$ . After SSB and expansion about the ground state, the diagonal form

706 of  $\mathcal{L}_{Yq}$  is:

$$\mathcal{L}_{Yq} = -m_d^i \bar{d}_i d_i \left(1 + \frac{H}{v}\right) - m_u^i \bar{u}_i u_i \left(1 + \frac{H}{v}\right) \quad (2.53)$$

707 Fermion masses depend on arbitrary couplings  $G_l$  and  $G_{u,d}$  and are not predicted  
708 by the theory.

### 709 2.3.5 The Higgs field

710 After the characterization of the fermions and gauge bosons as well as their interac-  
711 tions, it is necessary to characterize the Higgs field itself. The Lagrangian  $\mathcal{L}_S$  in Eqn.  
712 2.37 written in terms of the gauge bosons is given by

$$\mathcal{L}_S = \frac{1}{4} \lambda v^4 + \mathcal{L}_H + \mathcal{L}_{HV} \quad (2.54)$$

713

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} m_H^2 H^2 - \frac{1}{2v} m_H^2 H^3 - \frac{1}{8v^2} m_H^2 H^4 \quad (2.55)$$

714

$$\mathcal{L}_{HV} = m_H^2 W_\mu^+ W^{\mu-} \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) \quad (2.56)$$

715 The mass of the Higgs boson is deduced as usual from the mass term in the Lagrangian  
716 resulting in:

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v \quad (2.57)$$

717 however, it is not predicted by the theory either. The experimental efforts to find the  
718 Higgs boson, carried out by the *Compact Muon Solenoid (CMS)* experiment and the *A*  
719 *Toroidal LHC Appartus (ATLAS)* experiments at the *Large Hadron Collider (LHC)*,  
720 gave great results by July of 2012 when the discovery of a new particle compatible  
721 with the Higgs boson predicted by the electroweak theory [36, 37] was announced.  
722 Although at the announcement time there were some reservations about calling the  
723 new particle the *Higgs boson*, today this name is widely accepted. The Higgs mass

724 measurement, reported by both experiments [38], is in Table 2.8.

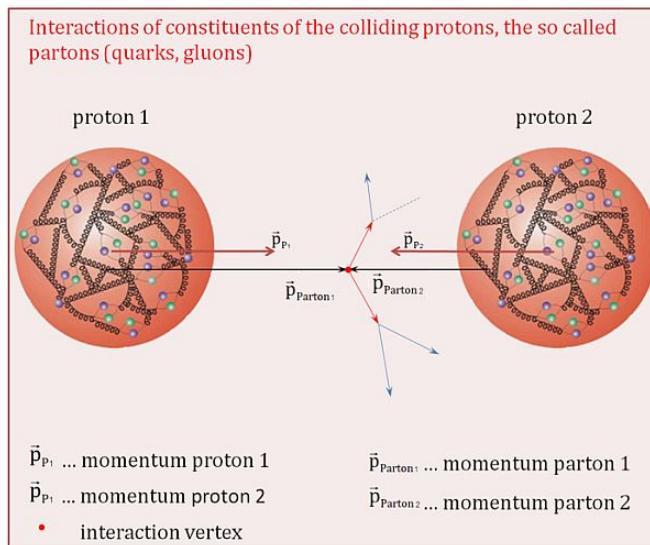
Property	Value
Electric charge	0
Color charge	0
Spin	0
Weak isospin	-1/2
Weak hypercharge	1
Parity	1
Mass (GeV/c <sup>2</sup> )	125.09±0.21 (stat.)±0.11 (syst.)

**Table 2.8:** Higgs boson properties. Higgs mass is not predicted by the theory and the value here corresponds to the experimental measurement.

725

### 726 2.3.6 Production of Higgs bosons at LHC

At the LHC, Higgs bosons are produced as a result of the collision of two counter-rotating protons beams. A detailed description of the LHC machine will be presented in chapter 3.



**Figure 2.10:** Proton-proton collision. Protons are composed of 3 valence quarks, a sea of quarks and gluons; therefore in a proton-proton collision, quarks and gluons are those who collide. [39].

729

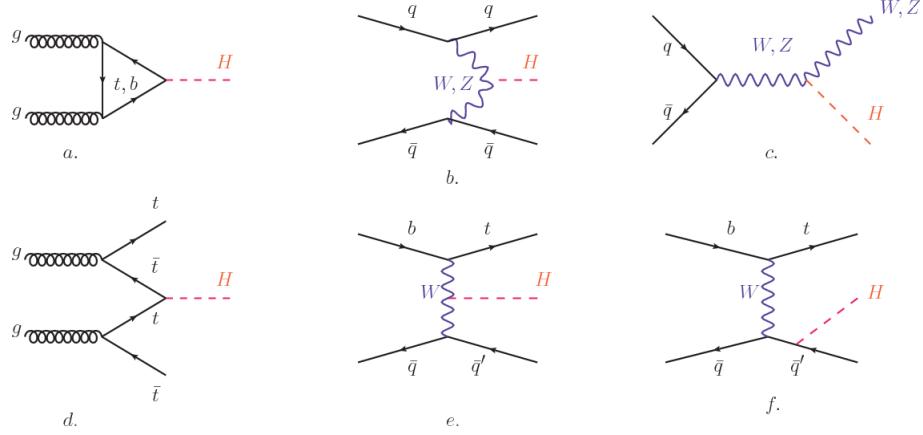
730        Protons are composed of quarks and these quarks are bound by gluons; however,  
731 what is commonly called the quark content of the proton makes reference to the  
732 valence quarks. In fact, a proton is not just a rigid entity with three balls in it all  
733 tied up with springs, but the gluons exchanged by the valence quarks tend to split  
734 spontaneously into quark-antiquark pairs or more gluons, creating *sea of quarks and*  
735 *gluons* as represented in Figure 2.10.

736        In a proton-proton ( $pp$ ) collision, the proton's constituents, quarks and gluons, are  
737 those that collide. The  $pp$  cross section depends on the momentum of the colliding  
738 particles, reason for which it is needed to know how the momentum is distributed  
739 inside the proton. Quarks and gluons are known as partons, hence, the functions that  
740 describe how the proton momentum is distributed among partons inside it are called  
741 *parton distribution functions (PDFs)*; PDFs are determined from experimental data  
742 obtained in experiments where the internal structure of hadrons is tested.

743        In addition, in physics, a common approach to study complex systems consists  
744 of starting with a simpler version of them, for which a well known description is  
745 available, and adding an additional *perturbation* which represents a small deviation  
746 from the known behavior. If the perturbation is small enough, the physical quantities  
747 associated with the perturbed system are expressed as a series of corrections to those  
748 of the simpler system. The perturbation series corresponds to an expansion in power  
749 series of a small parameter, therefore, the more terms are considered in the series (the  
750 higher order in the perturbation series), the more precise is the the description of the  
751 complex system. If the perturbation does not get progressively smaller, the strategy  
752 cannot be applied and new methods have to be employed.

753        High energy systems, like the Higgs production at LHC explored in this thesis,  
754 usually can be treated perturbatively with the expansion made in terms of the cou-  
755 pling constants. The overview presented here will be oriented specifically to the Higgs

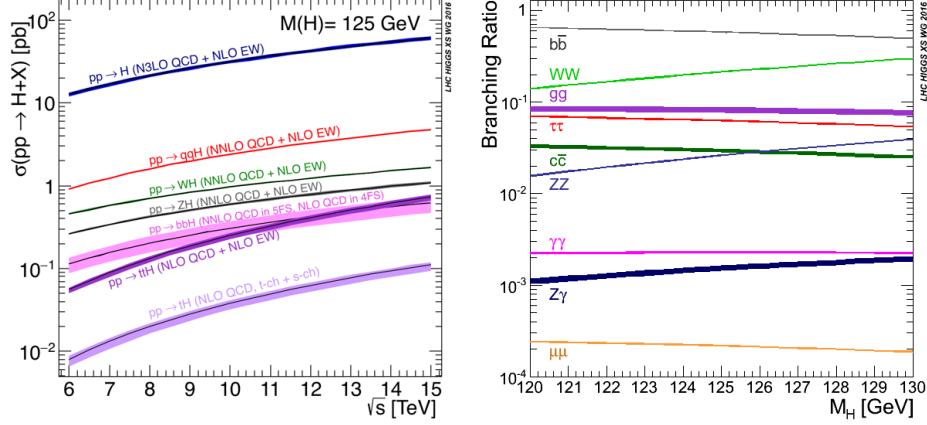
756 boson production mechanisms in  $pp$  collisions at LHC.



**Figure 2.11:** Main Higgs boson production mechanism Feynman diagrams. a. gluon-gluon fusion, b. vector boson fusion (VBF), c. Higgs-strahlung, d. Associated production with a top or bottom quark pair, e-f. associated production with a single top quark.

757 Figure 2.11 shows the Feynman diagrams for the leading order (first order) Higgs  
 758 production processes at LHC, while the cross section for Higgs production as a func-  
 759 tion of the center of mass-energy ( $\sqrt{s}$ ) for  $pp$  collisions is showed in Figure 2.12 left.  
 760 The tags NLO (next to leading order), NNLO (next to next to leading order) and  
 761 N3LO (next to next to next to leading order) make reference to the order at which  
 762 the perturbation series have been considered while the tags QCD and EW correspond  
 763 to the strong and electroweak coupling constants respectively.

764 The main production mechanism is the gluon fusion (Figure 2.11a and  $pp \rightarrow H$  in  
 765 Figure 2.12) given that gluons carry the highest fraction of momentum of the protons  
 766 in  $pp$  colliders. Since the Higgs boson does not couple to gluons, the mechanism  
 767 proceeds through the exchange of a virtual top-quark loop. Note that in this process  
 768 the Higgs boson is produced alone, turning out to be problematic for some Higgs  
 769 decays, because such absence of anything produced in association with the Higgs  
 770 represent a trouble for triggering, however, this mechanism is experimentally clean  
 771 when combined with the two-photon or the four-lepton decay channels (see Section



**Figure 2.12:** Higgs boson production cross sections (left) and decay branching ratios (right) for the main mechanisms. The VBF is indicated as  $qqH$  [40].

772 2.3.7).

773 Vector boson fusion (Figure 2.11b and  $pp \rightarrow qqH$  in Figure 2.12) has the second  
 774 largest production cross section. The scattering of two fermions is mediated by a weak  
 775 gauge boson which later emits a Higgs boson. In the final state, the two fermions tend  
 776 to be located in the central region of the detector; this kind of features are generally  
 777 used as a signature when analyzing the datasets provided by the experiments<sup>7</sup>.

778 The next production mechanism is Higgs-strahlung (Figure 2.11c and  $pp \rightarrow WH, pp \rightarrow$   
 779  $ZH$  in Figure 2.12) where two fermions annihilate to form a weak gauge boson. If the  
 780 initial fermions have enough energy, the emergent boson might emit a Higgs boson.

781 The associated production with a top or bottom quark pair and the associated  
 782 production with a single top quark (Figure 2.11d-f and  $pp \rightarrow bbH, pp \rightarrow t\bar{t}H, pp \rightarrow tH$   
 783 in Figure 2.12) have a smaller cross section than the main three mechanisms above,  
 784 but they provide a good opportunity to test the Higgs-top coupling. The analysis  
 785 reported in this thesis is developed using these production mechanisms. A detailed  
 786 description of the  $tH$  mechanism will be given in Section 2.5.

<sup>7</sup> More details about how to identify events of interest in this analysis will be given in chapter 6.

### 787 2.3.7 Higgs boson decay channels

788 When a particle can decay through several modes, also known as channels, the prob-  
 789 ability of decaying through a given channel is quantified by the *branching ratio (BR)*  
 790 of the decay channel; thus, the BR is defined as the ratio of number of decays go-  
 791 ing through that given channel to the total number of decays. In regard to the  
 792 Higgs boson decay, the BR can be predicted with accuracy once the Higgs mass is  
 793 known [41,42]. In Figure 2.12 right, a plot of the BR as a function of the Higgs mass  
 794 is presented; the largest predicted BR corresponds to the  $b\bar{b}$  pair decay channel (see  
 795 Table 2.9) given that it is the heaviest particle pair whose on-shell<sup>8</sup> production is  
 796 kinematically allowed in the decay.

Decay channel	Branching ratio	Rel. uncertainty
$H \rightarrow b\bar{b}$	$5.84 \times 10^{-1}$	$+3.2\% - 3.3\%$
$H \rightarrow W^+W^-$	$2.14 \times 10^{-1}$	$+4.3\% - 4.2\%$
$H \rightarrow \tau^+\tau^-$	$6.27 \times 10^{-2}$	$+5.7\% - 5.7\%$
$H \rightarrow ZZ$	$2.62 \times 10^{-2}$	$+4.3\% - 4.1\%$
$H \rightarrow \gamma\gamma$	$2.27 \times 10^{-3}$	$+5.0\% - 4.9\%$
$H \rightarrow Z\gamma$	$1.53 \times 10^{-3}$	$+9.0\% - 8.9\%$
$H \rightarrow \mu^+\mu^-$	$2.18 \times 10^{-4}$	$+6.0\% - 5.9\%$

**Table 2.9:** Predicted branching ratios and the relative uncertainty for some decay channels of a SM Higgs boson with  $m_H = 125\text{GeV}/c^2$  [9]; the uncertainties are driven by theoretical uncertainties for the different Higgs boson partial widths and by parametric uncertainties associated to the strong coupling and the masses of the quarks which are the input parameters. Further details on these calculations can be found in Reference [43]

797

798 Decays to other lepton and quark pairs, like electron, strange, up, and down  
 799 quark pairs not listed in the table, are also possible but their likelihood is too small

<sup>8</sup> In general, on-shell or real particles are those which satisfy the energy-momentum relation ( $E^2 - |\vec{p}|^2 c^2 = m^2 c^4$ ); off-shell or virtual particles does not satisfy it which is possible under the uncertainty principle of quantum mechanics. Usually, virtual particles correspond to internal propagators in Feynman diagrams.

800 to measure since they are very lightweight, hence, their interaction with the Higgs  
 801 boson is very weak. On other hand, the decay to top quark pairs is heavily suppressed  
 802 due to the top quark mass ( $\approx 173 \text{ GeV}/c^2$ ).

803 Decays to gluons proceed indirectly through a virtual top quark loop while the  
 804 decays to photons proceed through a virtual W boson loop, therefore, their branching  
 805 ratio is smaller compared to direct interaction decays. Same is true for the decay to  
 806 a photon and a Z boson.

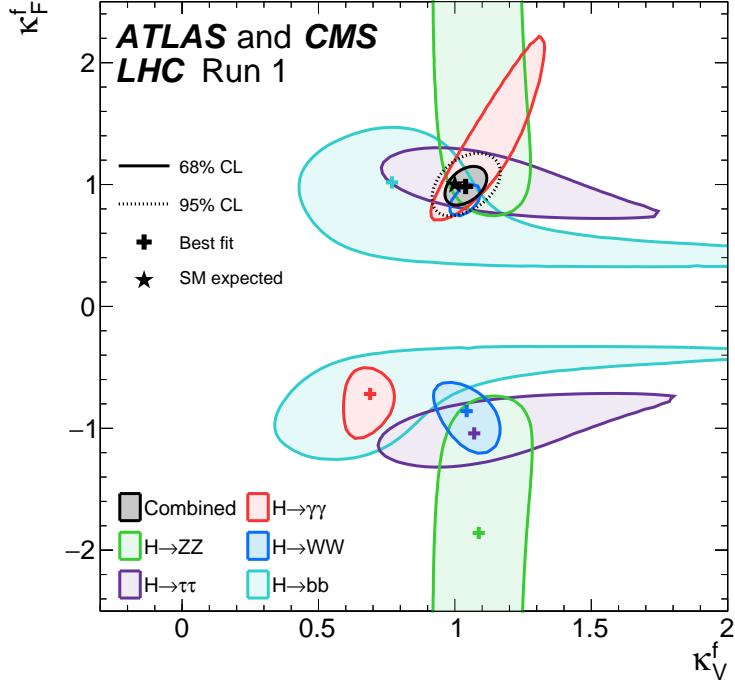
807 In the case of decays to pairs of W and Z bosons, the decay proceed with one of  
 808 the bosons being on-shell and the other being off-shell. The likelihood of the process  
 809 diminish depending on how far off-shell are the virtual particles involved, hence, the  
 810 branching ratio for W boson pairs is bigger than for Z boson pairs since Z boson mass  
 811 is bigger than W boson mass.

812 Note that the decay to a pair of virtual top quarks is possible, but the probability  
 813 is way too small.

## 814 **2.4 Experimental status of the anomalous**

### 815 **Higgs-fermion coupling**

816 ATLAS and CMS have performed analyses of the anomalous H-f coupling by making  
 817 likelihood scans for the two coupling modifiers,  $\kappa_t$  and  $\kappa_V$ , under the assumption  
 818 that  $\kappa_Z = \kappa_W \equiv \kappa_V$  and  $\kappa_t = \kappa_\tau = \kappa_b \equiv \kappa_f$ . Figure 2.13 shows the result of the  
 819 combination of ATLAS and CMS fits; also the individual decay channels combination  
 820 and the global combination results are shown. Note that from this plot there is limited  
 821 information on the sign of the coupling since the only information available about the  
 822 sign of the coupling comes from decays rather than production.



**Figure 2.13:** Combination of the ATLAS and CMS fits for coupling modifiers  $\kappa_t - \kappa_V$ ; also shown the individual decay channels combination and their global combination. No assumptions have been made on the sign of the coupling modifiers [58].

While all the channels are compatible for positive values of the modifiers, for negative values of  $\kappa_t$  there is no compatibility. The best fit for individual channels is compatible with negative values of  $\kappa_t$  except for the  $H \rightarrow bb$  channel. The best fit for the global fit yields  $\kappa_t \geq 0$  in contrast to the yields from the individual channels; the reason of this resides in the  $H \rightarrow \gamma\gamma$  coupling parameter.  $H \rightarrow \gamma\gamma$  decay proceeds through a loop of either top quarks or W bosons, hence, this channel is sensitive to  $\kappa_t$  due to the interference of these two amplitude contributions. Under the assumption that no beyond SM particles take part in the loops, a flipped sign of  $\kappa_t$  will increase the  $H \rightarrow \gamma\gamma$  branching fraction by a factor of  $\sim 2.4$  which is not supported by measurements; thus, this large asymmetry between the positive and negative coupling ratios in the  $H \rightarrow \gamma\gamma$  channel drives the yield of the global fit. Nevertheless, since the

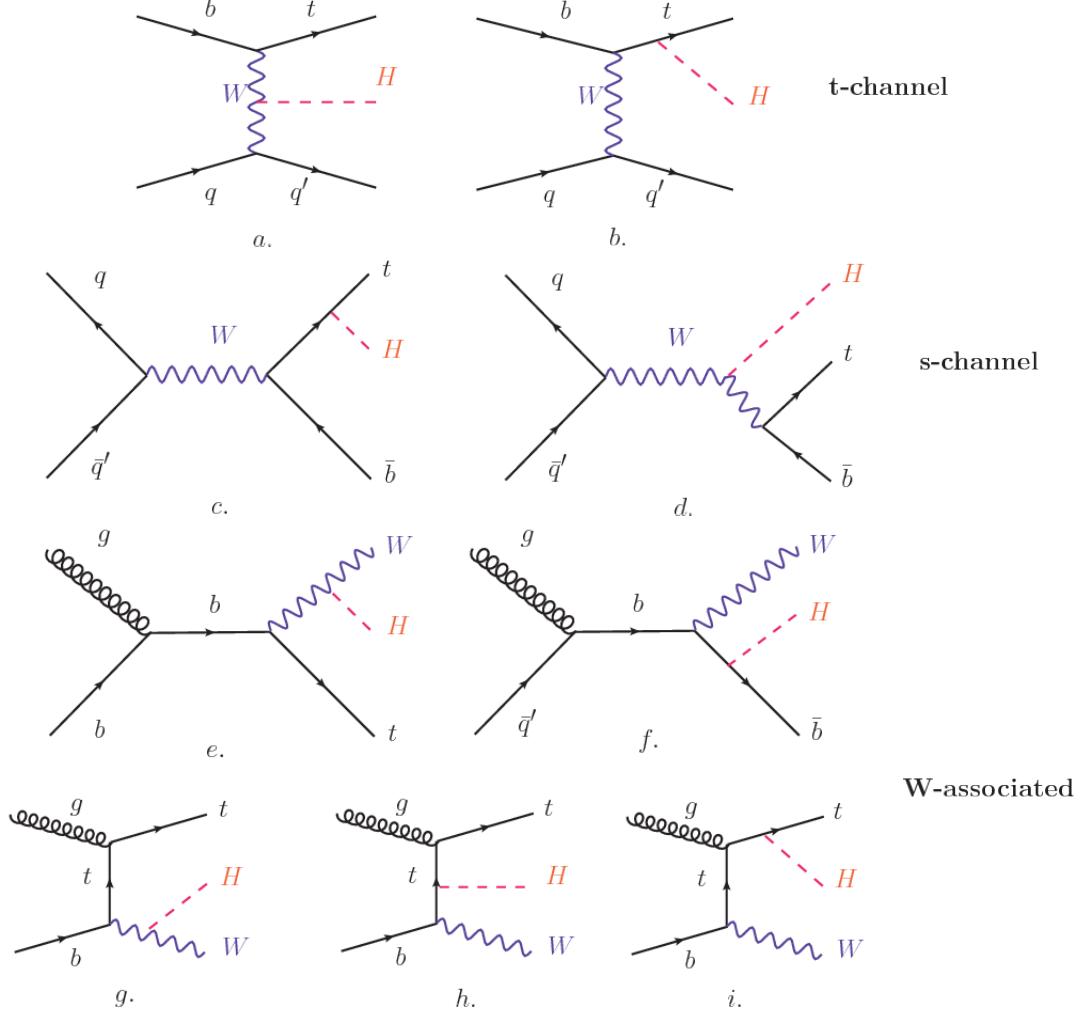
834  $H \rightarrow bb$  channel is expected to be the most sensitive channel and its best fit value of  
 835  $\kappa_t$  is positive, the global fit is well supported as shown in Reference [58].

836 Although the contributions from all the other decay channels are small compared  
 837 to the  $H \rightarrow bb$  the anomalous H-t coupling cannot be excluded completely, thus, that  
 838 motivates to look at  $tH$  processes, which can help with both, the limited information  
 839 on the sign of the H-t coupling and the access to information from the Higgs boson  
 840 production rather than from its decays.

## 841 **2.5 Associated production of a Higgs boson and a 842 single top quark**

843 The production of Higgs boson in association with a top quark has been extensively  
 844 studied [44–48]. While measurements of the main Higgs production mechanisms rates  
 845 are sensitive to the strength of the Higgs coupling to W boson or top quark, they are  
 846 not sensitive to the relative sign between the two couplings. In this thesis, the Higgs  
 847 boson production mechanism explored is the associated production with a single top  
 848 quark ( $tH$ ) which offers sensitivity to the relative sign of the Higgs couplings to W  
 849 boson and to top quark. The description given here is based on Reference [46]

A process where two incoming particles interact and produce a final state with two particles can proceed in three called channels (see, for instance, Figure 2.14 omitting the red line). The t-channel represents processes where an intermediate particle is emitted by one of the incoming particles and absorbed by the other. The s-channel represents processes where the two incoming particles merge into an intermediate particle which eventually will split into the particles in the final state. The third channel, u-channel, is similar to the t-channel but the two outgoing particles interchange their



**Figure 2.14:** Associated Higgs boson production with a top quark mechanism Feynman diagrams. a.,b. t-channel ( $tHq$ ), c.,d. s-channel ( $tHb$ ), e-i. W-associated.

roles. These three channels are connected to the so-called Mandelstam variables

$$s = (p_1 + p_2)^2 = (p'_1 + p'_2)^2 \rightarrow \text{square of the center mass-energy.} \quad (2.58)$$

$$t = (p_1 - p'_1)^2 = (p'_2 - p_2)^2 \rightarrow \text{square of the four-momentum transfer.} \quad (2.59)$$

$$u = (p_1 - p'_2)^2 = (p'_1 - p_2)^2 \rightarrow \text{square of the crossed four-momentum transfer.} \quad (2.60)$$

$$s + t + u = m_1^2 + m_2^2 + m'_1^2 + m'_2^2 \quad (2.61)$$

which relate the momentum, energy and the angles of the incoming and outgoing particles in an scattering process of two particles to two particles. The importance of the Mandelstam variables reside in that they form a minimum set of variables needed to describe the kinematics of this scattering process; they are Lorentz invariant which makes them very useful when doing calculations.

The  $tH$  production, where Higgs boson can be radiated either from the top quark or from the W boson, is represented by the leading order Feynman diagrams in Figure 2.14. The cross section for the  $tH$  process is calculated, as usual, summing over the contributions from the different Feynman diagrams; therefore it depends on the interference between the contributions. In the SM, the interference for t-channel ( $tHq$  process) and W-associated ( $tHW$  process) production is destructive [44] resulting in the small cross sections presented in Table 2.10.

tH production channel	Cross section (fb)
t-channel ( $pp \rightarrow tHq$ )	$70.79^{+2.99}_{-4.80}$
W-associated ( $pp \rightarrow tHW$ )	$15.61^{+0.83}_{-1.04}$
s-channel( $pp \rightarrow tHb$ )	$2.87^{+0.09}_{-0.08}$

**Table 2.10:** Predicted SM cross sections for  $tH$  production at  $\sqrt{s} = 13$  TeV [49, 50].

The s-channel contribution can be neglected. It will be shown that a deviation from the SM destructive interference would result in an enhancement of the  $tH$  cross section compared to that in SM, which could be used to get information about the sign of the Higgs-top coupling [46, 47]. In order to describe  $tH$  production processes, Feynman diagram 2.14b will be considered; there, the W boson is radiated by a quark in the proton and eventually it will interact with the b quark. In the high energy regime, the effective W approximation [51] is used to describe the process as the

870 emission of an approximately on-shell W and its hard scattering with the b quark;  
 871 i.e.  $Wb \rightarrow th$ . The scattering amplitude for the process is given by

$$\mathcal{A} = \frac{g}{\sqrt{2}} \left[ (\kappa_t - \kappa_V) \frac{m_t \sqrt{s}}{m_W v} A \left( \frac{t}{s}, \varphi; \xi_t, \xi_b \right) + \left( \kappa_V \frac{2m_W s}{v t} + (2\kappa_t - \kappa_V) \frac{m_t^2}{m_W v} \right) B \left( \frac{t}{s}, \varphi; \xi_t, \xi_b \right) \right], \quad (2.62)$$

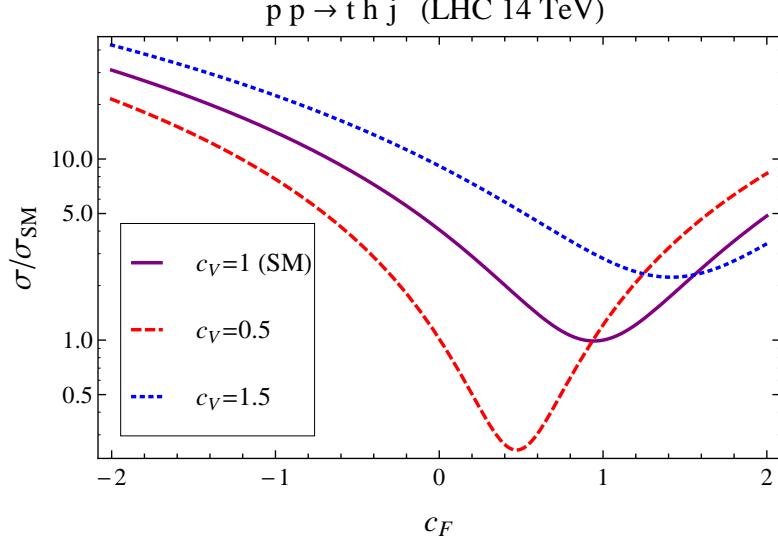
872 where  $\kappa_V \equiv g_{HVV}/g_{HVV}^{SM}$  and  $\kappa_t \equiv g_{Ht}/g_{Ht}^{SM} = y_t/y_t^{SM}$  are scaling factors that quantify  
 873 possible deviations of the couplings from the SM values, Higgs-Vector boson (H-  
 874 W) and Higgs-top (H-t) respectively, from the SM couplings;  $s = (p_W + p_b)^2$ ,  $t =$   
 875  $(p_W - p_H)^2$ ,  $\varphi$  is the Higgs azimuthal angle around the  $z$  axis taken parallel to the  
 876 direction of motion of the incoming W; A and B are functions describing the weak  
 877 interaction in terms of the chiral states  $(\xi_t, \xi_b)$  of the quarks  $b$  and  $t$ . Terms that  
 878 vanish in the high energy limit have been neglected as well as the Higgs and  $b$  quark  
 879 masses<sup>9</sup>.

880 The scattering amplitude grows with energy like  $\sqrt{s}$  for  $\kappa_V \neq \kappa_t$ , in contrast to  
 881 the SM ( $\kappa_t = \kappa_V = 1$ ), where the first term in 2.62 cancels out and the amplitude  
 882 is constant for large  $s$ ; therefore, a deviation from the SM predictions represents an  
 883 enhancement in the  $tHq$  cross section. In particular, for a SM H-W coupling and a  
 884 H-t coupling of inverted sign with respect to the SM ( $\kappa_V = -\kappa_t = 1$ ) the  $tHq$  cross  
 885 section is enhanced by a factor greater 10 as seen in the Figure 2.15 taken from  
 886 Reference [46]; Reference [52] has reported similar enhancement results.

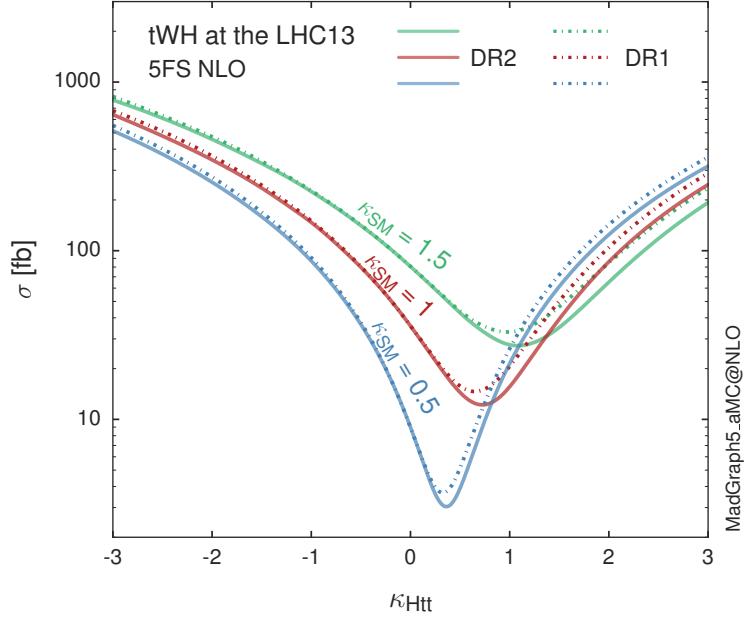
887 A similar analysis is valid for the W-associated channel but, in that case, the in-  
 888 terference is more complicated since there are more than two contributions and an ad-  
 889 ditional interference with the production of Higgs boson and a top pair process( $t\bar{t}H$ ).  
 890 The calculations are made using the so-called Diagram Removal (DR) technique where

---

<sup>9</sup> A detailed explanation of the structure and approximations used to derive  $\mathcal{A}$  can be found in Reference [46]



**Figure 2.15:** Cross section for  $tHq$  process as a function of  $\kappa_t$ , normalized to the SM, for three values of  $\kappa_V$ . In the plot  $c_f$  refers to the Higgs-fermion coupling which is dominated by the H-t coupling and represented here by  $\kappa_t$ . Solid, dashed and dotted lines correspond to  $c_V \rightarrow \kappa_V = 1, 0.5, 1.5$  respectively. Note that for the SM ( $\kappa_V = \kappa_t = 1$ ), the destructive effect of the interference is maximal.



**Figure 2.16:** Cross section for  $tHW$  process as a function of  $\kappa_{Htt}$ , for three values of  $\kappa_{SM}$  at  $\sqrt{s} = 13$  TeV.  $\kappa_{Htt}^2 = \sigma_{Htt}/\sigma_{Htt}^{SM}$  is a simple re-scaling of the SM Higgs interactions.

interfering diagrams are removed (or added) from the calculations in order to evaluate the impact of the removed contributions. DR1 was defined to neglect  $t\bar{t}H$  interference while DR2 was defined to take  $t\bar{t}H$  interference into account [53]. As shown in Figure 2.16, the  $tHW$  cross section is enhanced from about 15 fb (SM:  $\kappa_{Htt} = 1$ ) to about 150 fb ( $\kappa_{Htt} = -1$ ). Differences between curves for DR1 and DR2 help to gauge the impact of the interference with  $t\bar{t}H$ . Results of the calculations of the  $tHq$  and  $tHW$  cross sections at  $\sqrt{s} = 13$  TeV can be found in Reference [54] and a summary of the results is presented in Table 2.11.

	$\sqrt{s}$ TeV	$\kappa_t = 1$	$\kappa_t = -1$
$\sigma^{LO}(tHq)(fb)$ [46]	8	$\approx 17.4$	$\approx 252.7$
	14	$\approx 80.4$	$\approx 1042$
$\sigma^{NLO}(tHq)(fb)$ [46]	8	$18.28^{+0.42}_{-0.38}$	$233.8^{+4.6}_{-0.0}$
	14	$88.2^{+1.7}_{-0.0}$	$982.8^{+28}_{-0.0}$
$\sigma^{LO}(tHq)(fb)$ [52]	14	$\approx 71.8$	$\approx 893$
$\sigma^{LO}(tHW)(fb)$ [52]	14	$\approx 16.0$	$\approx 139$
$\sigma^{NLO}(tHq)(fb)$ [54]	8	$18.69^{+8.62\%}_{-17.13\%}$	-
	13	$74.25^{+7.48\%}_{-15.35\%}$	$848^{+7.37\%}_{-13.70\%}$
	14	$90.10^{+7.34\%}_{-15.13\%}$	$1011^{+7.24\%}_{-13.39\%}$
$\sigma^{LO}(tHW)(fb)$ [53]	13	$15.77^{+15.91\%}_{-15.76\%}$	-
$\sigma^{NLO}DR1(tHW)(fb)$ [53]	13	$21.72^{+6.52\%}_{-5.24\%}$	$\approx 150$
$\sigma^{NLO}DR2(tHW)(fb)$ [53]	13	$16.28^{+7.34\%}_{-15.13\%}$	$\approx 150$

**Table 2.11:** Predicted enhancement of the  $tHq$  and  $tHW$  cross sections at LHC for  $\kappa_V = 1$  and  $\kappa_t = \pm 1$  at LO and NLO; the cross section enhancement of more than a factor of 10 is due to the flipping in the sign of the H-t coupling with respect to the SM one.

899

## 900 2.6 CP-mixing in $tH$ processes

901 In addition to the sensitivity to sign of the H-t coupling, the  $tHq$  and  $tHW$  processes  
 902 have been proposed as a tool to investigate the possibility of a H-t coupling that does

903 not conserve CP [48, 53, 55].

904 In this thesis, the sensitivity of  $tH$ processes to CP-mixing is also studied on the  
 905 basis of References [48, 53] using the effective field theory framework where a generic  
 906 particle ( $X_0$ ) of spin-0 and a general CP violating interaction with the top quark  
 907 ( $Htt$  coupling), can couple to scalar and pseudo-scalar fermionic densities. The H-W  
 908 interaction is assumed to be SM-like. The Lagrangian modeling the H-t interaction  
 909 is given by

$$\mathcal{L}_0^t = -\bar{\psi}_t (c_\alpha \kappa_{Htt} g_{Htt} + i s_\alpha \kappa_{Att} g_{Att} \gamma_5) \psi_t X_0, \quad (2.63)$$

910 where  $\alpha$  is the CP-mixing phase,  $c_\alpha \equiv \cos \alpha$  and  $s_\alpha \equiv \sin \alpha$ ,  $\kappa_{Htt}$  and  $\kappa_{Att}$  are real  
 911 dimensionless re-scaling parameters<sup>10</sup> used to parametrize the magnitude of the CP-  
 912 violating and CP-conserving parts of the amplitude. The model defines  $g_{Htt} =$   
 913  $g_{Att} = m_t/v = y_t/\sqrt{2}$  with  $v \sim 246$  GeV the Higgs vacuum expectation value. In  
 914 this parametrization, three special cases can be recovered

915 • CP-even coupling  $\rightarrow \alpha = 0^\circ$

916 • CP-odd coupling  $\rightarrow \alpha = 90^\circ$

917 • SM coupling  $\rightarrow \alpha = 0^\circ$  and  $\kappa_{Htt} = 1$

918 The loop induced  $X_0$  coupling to gluons can also be described in terms of the  
 919 parametrization above, according to

$$\mathcal{L}_0^g = -\frac{1}{4} \left( c_\alpha \kappa_{Hgg} g_{Hgg} G_{\mu\nu}^a G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} \right) X_0. \quad (2.64)$$

920 where  $g_{Hgg} = -\alpha_s/3\pi v$  and  $g_{Agg} = \alpha_s/2\pi v$  and  $G_{\mu\nu}$  is the gluon field strength tensors.

921 Under the assumption that the top quark dominates the gluon-fusion process at LHC

---

<sup>10</sup> analog to  $\kappa_t$  and  $\kappa_V$

energies,  $\kappa_{Hgg} \rightarrow \kappa_{Htt}$  and  $\kappa_{Agg} \rightarrow \kappa_{Att}$ , so that the ratio between the gluon-gluon fusion cross section for  $X_0$  and for the SM Higgs prediction can be written as

$$\frac{\sigma_{NLO}^{gg \rightarrow X_0}}{\sigma_{NLO,SM}^{gg \rightarrow H}} = c_\alpha^2 \kappa_{Htt}^2 + s_\alpha^2 \left( \kappa_{Att} \frac{g_{Agg}}{g_{Hgg}} \right)^2. \quad (2.65)$$

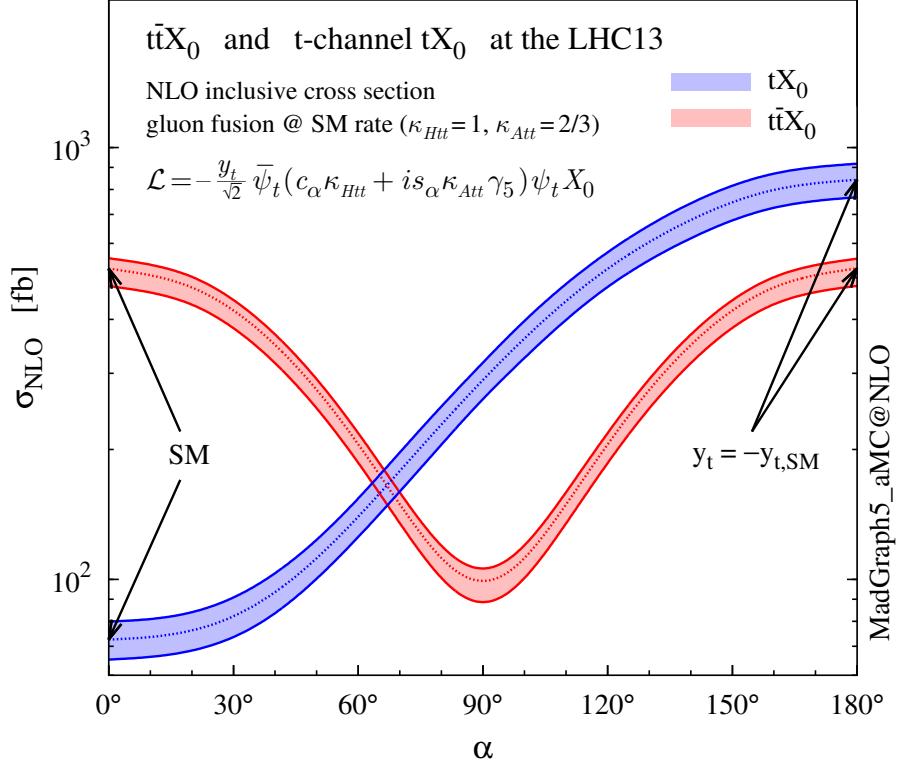
If the re-scaling parameters are set to

$$\kappa_{Htt} = 1, \quad \kappa_{Att} = \left| \frac{g_{Hgg}}{g_{Agg}} \right| = \frac{2}{3}. \quad (2.66)$$

the gluon-fusion SM cross section is reproduced for every value of the CP-mixing angle  $\alpha$ ; therefore, by imposing that condition to the Lagrangian density 2.63, the CP-mixing angle is not constrained by current data. Figure 2.17 shows the NLO cross sections for t-channel  $tX_0$ (blue) and  $t\bar{t}X_0$  (red) associated production processes as a function of the CP-mixing angle  $\alpha$ .  $X_0$  is a generic spin-0 particle with top quark CP-violating coupling. Re-scaling factors  $\kappa_{Htt}$  and  $\kappa_{Att}$  have been set to reproduce the SM gluon-fusion cross sections.

It is interesting to notice that the  $tX_0$  cross section is enhanced, by a factor of about 10, when a continuous rotation in the scalar-pseudoscalar plane is applied; this enhancement is similar to the enhancement produced when the H-t coupling is flipped in sign with respect to the SM ( $y_t = -y_{t,SM}$  in the plot), as showed in Section 2.5. In contrast, the degeneracy in the  $t\bar{t}X_0$  cross section is still present given that it depends quadratically on the H-t coupling, but more interesting is to notice that  $t\bar{t}X_0$  cross section is exceeded by  $tX_0$  cross section after  $\alpha \sim 60^\circ$ .

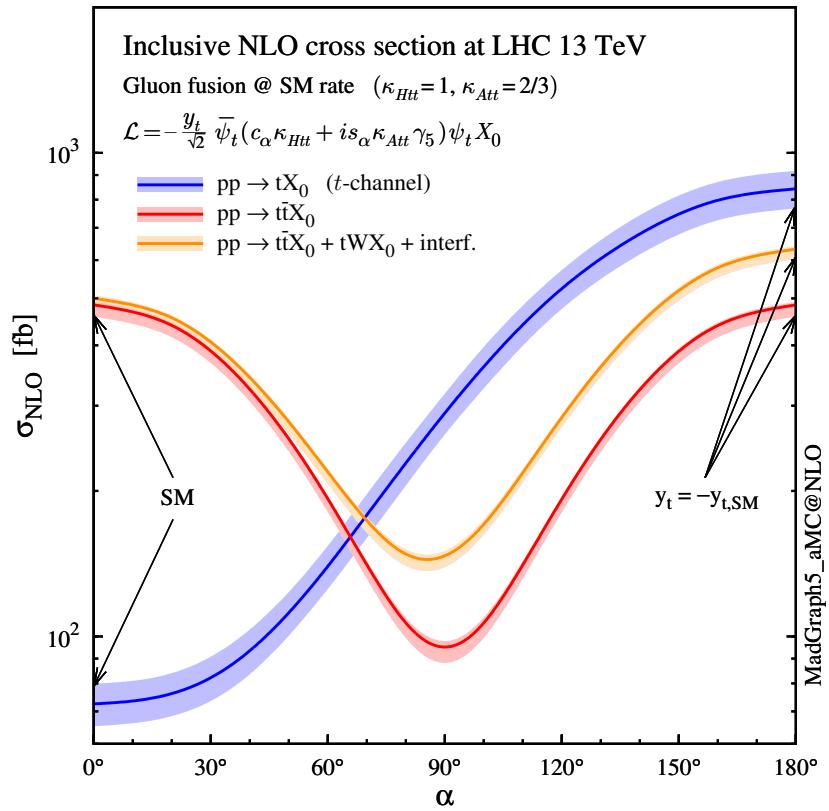
A similar parametrization can be used to investigate the  $tHW$  process sensitivity to CP-violating H-t coupling. As said in 2.5, the interference in the W-associated channel is more complicated because there are more than two contributions and also there is interference with the  $t\bar{t}H$  production process.



**Figure 2.17:** NLO cross sections for t-channel  $tX_0$ (blue) and  $t\bar{t}X_0$  (red) associated production processes as a function of the CP-mixing angle  $\alpha$ .  $X_0$  is a generic spin-0 particle with top quark CP-violating coupling [48].

943       Figure 2.18 shows the NLO cross sections for t-channel  $tX_0$ (blue),  $t\bar{t}X_0$  (red)  
 944       associated production and for the combined  $tWX_0+t\bar{t}X_0+interference$  (orange) as  
 945       a function of the CP-mixing angle. It is clear that the effect of the interference in the  
 946       combined case is the lifting of the degeneracy present in the  $t\bar{t}X_0$  production. The  
 947       constructive interference enhances the cross section from about 500 fb at SM ( $\alpha = 0$ )  
 948       to about 600 fb ( $\alpha = 180^\circ \rightarrow y_t = -y_{t,SM}$ ).

949       An analysis combining  $tHq$  and  $tHW$ processes will be made in this thesis taking  
 950       advantage of the sensitivity improvement.



**Figure 2.18:** NLO cross sections for t-channel  $tX_0$ (blue),  $t\bar{t}X_0$  (red) associated production processes and combined  $tWX_0 + t\bar{t}X_0$  (including interference) production as a function of the CP-mixing angle  $\alpha$  [48].

951 **Chapter 3**

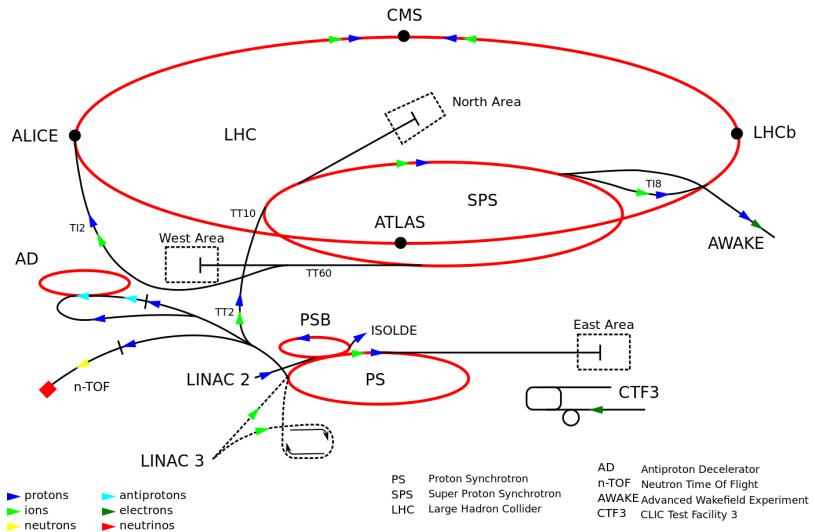
952 **The CMS experiment at the LHC**

953 **3.1 Introduction**

954 Located on the Swiss-French border, the European Council for Nuclear Research  
955 (CERN) is the largest scientific organization leading particle physics research. About  
956 13000 people in a broad range of roles including users, students, scientists, engineers,  
957 among others, contribute to the data taking and analysis, with the goal of unveiling  
958 the secrets of nature and revealing the fundamental structure of the universe. CERN  
959 is also the home of the Large Hadron Collider (LHC), the largest particle accelerator  
960 around the world, where protons (or heavy ions) traveling close to the speed of light,  
961 are made to collide. These collisions open a window to investigate how particles (and  
962 their constituents if they are composite) interact with each other, providing clues  
963 about the laws of nature. This chapter presents an overview of the LHC structure  
964 and operation. A detailed description of the CMS detector is offered, given that the  
965 data used in this thesis have been taken with this detector.

## 966 3.2 The LHC

967 With 27 km of circumference, the LHC is currently the most powerful circular accelerator  
 968 in the world. It is installed in the same tunnel where the Large Electron-Positron  
 969 (LEP) collider was located, taking advantage of the existing infrastructure. The LHC  
 970 is part of the CERN's accelerator complex composed of several successive accelerat-  
 971 ing stages before the particles are injected into the LHC ring where they reach their  
 972 maximum energy (see Figure 3.1).



**Figure 3.1:** CERN accelerator complex. Blue arrows show the path followed by protons along the acceleration process [59].

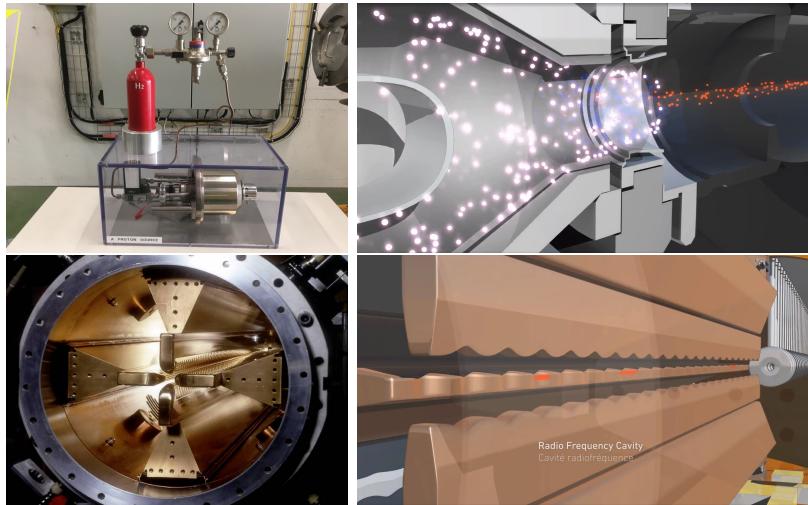
973 The LHC runs in three collision modes depending on the particles being acceler-  
 974 ated

975 • Proton-Proton collisions ( $pp$ ) for multiple physics experiments.

976 • Lead-Lead collisions ( $Pb-Pb$ ) for heavy ion experiments.

977 • Proton-Lead collisions ( $p-Pb$ ) for quark-gluon plasma experiments.

978 In this thesis only  $pp$  collisions will be considered.

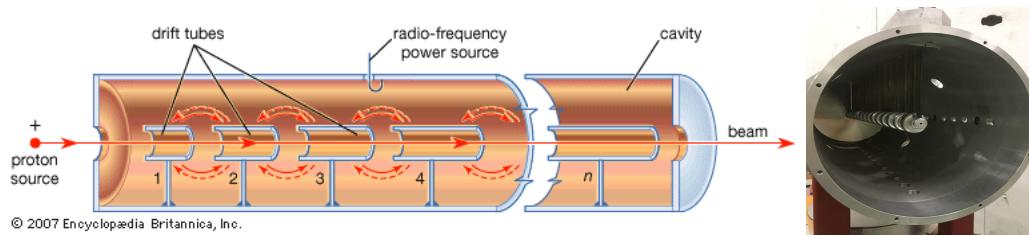


**Figure 3.2:** LHC protons source and the first acceleration stage. Top: the bottle contains hydrogen gas which is injected into the metal cylinder (white dots) to be broken down into electrons (blue dots) and protons (red dots); Bottom: the obtained protons are directed towards the radio frequency quadrupole which perform the first acceleration, focus the beam and create the bunches of protons. Left images are real pictures while right images are drawings [63, 64].

979     Collection of protons starts with hydrogen atoms taken from a bottle, containing  
 980     hydrogen gas, and injecting them in a metal cylinder; hydrogen atoms are broken  
 981     down into electrons and protons by an intense electric field (see Figure 3.2 top).  
 982     The resulting protons leave the metal cylinder towards a radio frequency quadrupole  
 983     (RFQ) that focus the beam, accelerates the protons and creates the packets of protons  
 984     called bunches. In the RFQ, an electric field is generated by a RF wave at a frequency  
 985     that matches the resonance frequency of the cavity where the electrodes are contained.  
 986     The beam of protons traveling on the RFQ axis experiences an alternating electric  
 987     field gradient that generates the focusing forces.

988     In order to accelerate the protons, a longitudinal time-varying electric field com-  
 989     ponent is added to the system; it is done by giving the electrodes a sine-like profile as  
 990     shown in Figure 3.2 bottom. By matching the speed and phase of the protons with  
 991     the longitudinal electric field the bunching is performed; protons synchronized with

992 the RFQ (synchronous protons) do not feel an accelerating force, but those protons in  
 993 the beam that have more (or less) energy than the synchronous proton (asynchronous  
 994 protons) will feel a decelerating (accelerating) force; therefore, asynchronous protons  
 995 will oscillate around the synchronous ones forming bunches of protons [61]. From the  
 996 RFQ protons emerge with energy 750 keV in bunches of about  $1.15 \times 10^{11}$  protons [62].

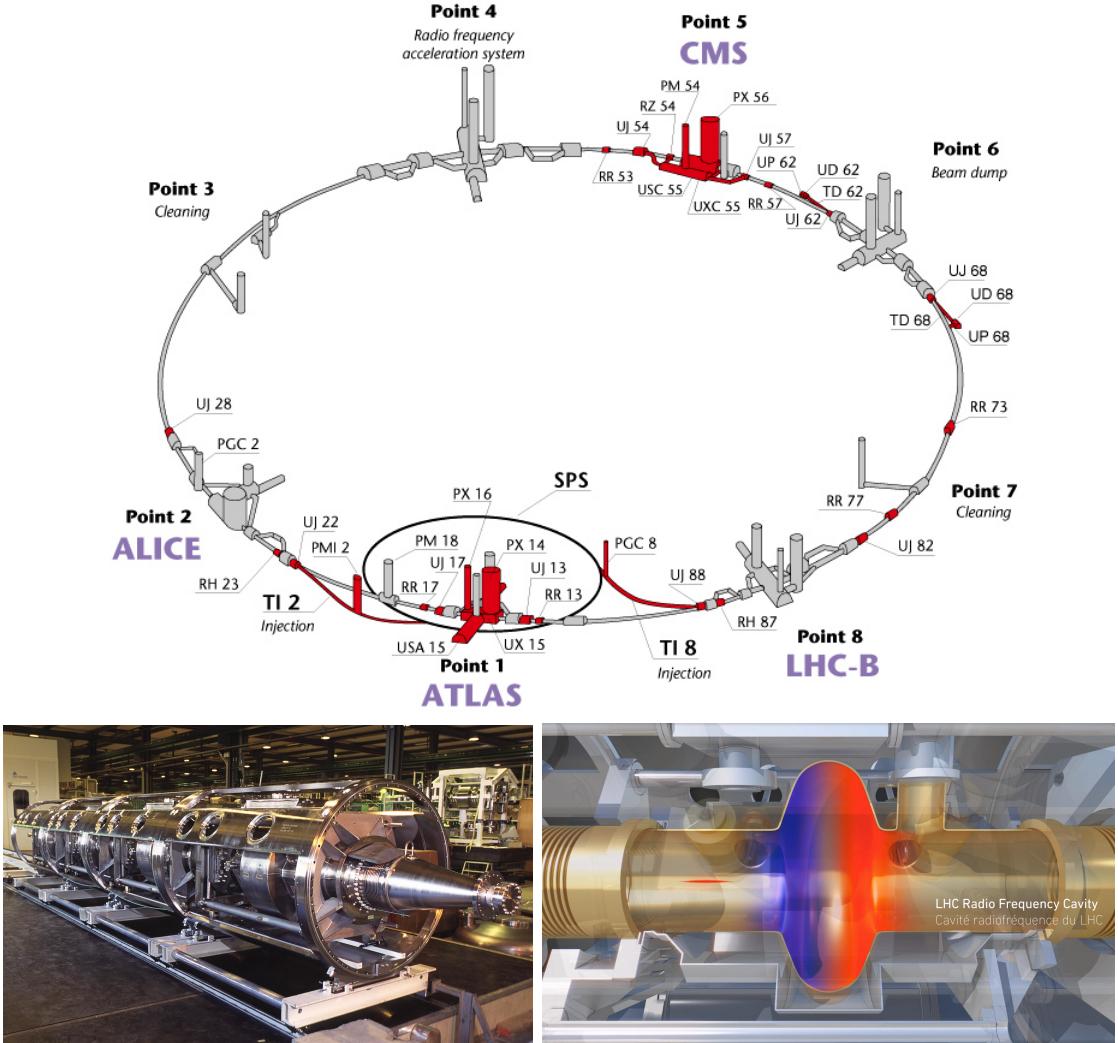


**Figure 3.3:** Left: drawing of the LINAC2 accelerating system at CERN. Electric fields generated by radio frequency (RF) create acceleration and deceleration zones inside the cavity; deceleration zones are blocked by drift tubes where quadrupole magnets focus the proton beam. Right: picture of a real RF cavity [65].

997 Proton bunches coming from the RFQ go to the linear accelerator 2 (LINAC2)  
 998 where they are accelerated to reach 50 MeV energy. In the LINAC2 stage, acceleration  
 999 is performed using electric fields generated by radio frequency which create zones  
 1000 of acceleration and deceleration as shown in Figure 3.3. In the deceleration zones,  
 1001 the electric field is blocked using drift tubes where protons are free to drift while  
 1002 quadrupole magnets focus the beam.

1003 The beam coming from LINAC2 is injected into the proton synchrotron booster  
 1004 (PSB) to reach 1.4 GeV in energy. The next acceleration is provided at the proton  
 1005 synchrotron (PS) up to 26 GeV, followed by the injection into the super proton  
 1006 synchrotron (SPS) where protons are accelerated to 450 GeV. Finally, protons are  
 1007 injected into the LHC where they are accelerated to the target energy of 6.5 TeV.

1008 PSB, PS, SPS and LHC accelerate protons using the same RF acceleration tech-  
 1009 nique described before.



**Figure 3.4:** Top: LHC layout. The red zones indicate the infrastructure additions to the LEP installations, built to accommodate the ATLAS and CMS experiments which exceed the size of the former experiments located there [60]. Bottom: LHC RF cavities. A module (left picture) accommodates 4 cavities, each like the one in the right drawing, that accelerate protons and preserve the bunch structure of the beam. The color gradient pattern represents the strength of the electric field inside the cavity; red zone corresponds to the maximum of the oscillation electric field while the blue zone corresponds to the minimum. [64, 66]

1010        The LHC has a system of 16 RF cavities located in the so-called point 4, as  
 1011        shown in Figure 3.4 top, tuned at a frequency of 400 MHz. The bottom side of  
 1012        Figure 3.4 shows a picture of a RF module composed of 4 RF cavities working in a  
 1013        superconducting state at 4.5 K; also, a representation of the accelerating electric field

1014 that accelerates the protons in the bunch is shown. The maximum of the oscillating  
 1015 electric field (red region) picks the proton bunches at the entrance of the cavity  
 1016 and keeps accelerating them through the whole cavity. The protons are carefully  
 1017 timed so that in addition to the acceleration effect the bunch structure of the beam  
 1018 is preserved.

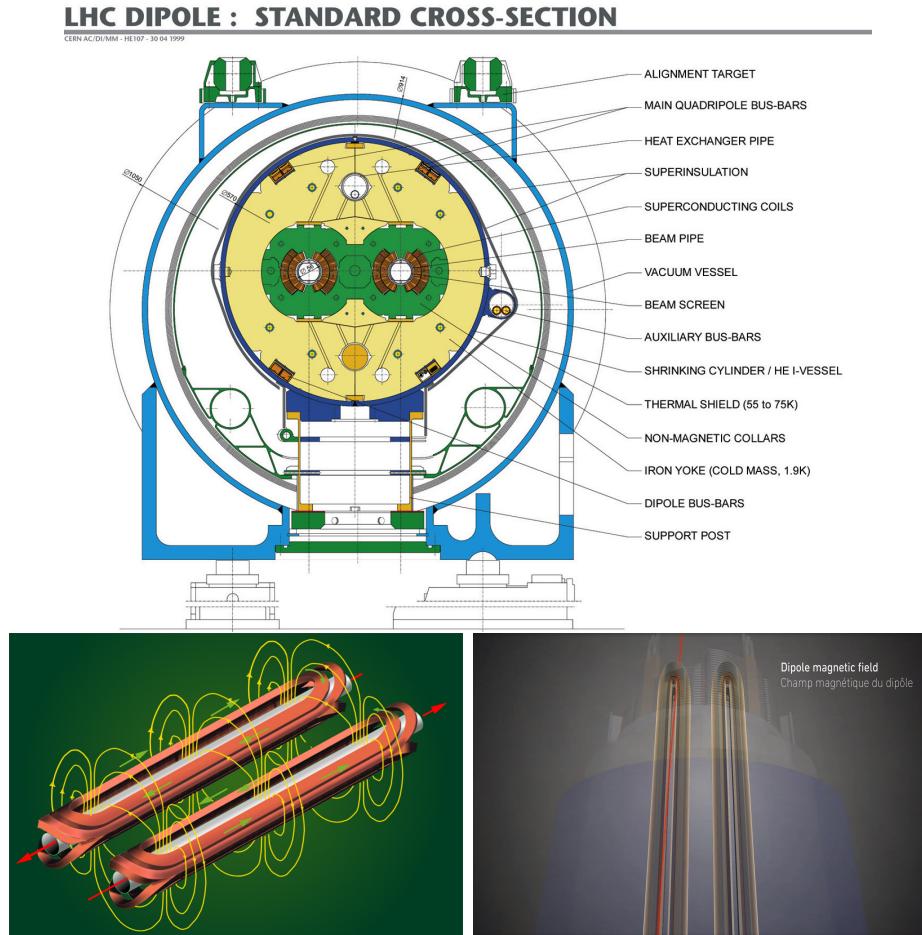
1019 While protons are accelerated in one section of the LHC ring, where the RF cavities  
 1020 are located, in the rest of their path they have to be kept in the curved trajectory  
 1021 defined by the LHC ring. Technically, LHC is not a perfect circle; RF, injection, beam  
 1022 dumping, beam cleaning and sections before and after the experimental points where  
 1023 protons collide are all straight sections. In total, there are 8 arcs 2.45 km long each  
 1024 and 8 straight sections 545 m long each. In order to curve the proton's trajectory in  
 1025 the arc sections, superconducting dipole magnets are used.

1026 Inside the LHC ring, there are two proton beams traveling in opposite directions  
 1027 in two separated beam pipes; the beam pipes are kept at ultra-high vacuum ( $\sim 10^{-9}$   
 1028 Pa) to ensure that there are no particles that interact with the proton beams. The  
 1029 superconducting dipole magnets used in LHC are made of a NbTi alloy, capable of  
 1030 transporting currents of about 12000 A when cooled at a temperature below 2K using  
 1031 liquid helium (see Figure 3.5).

1032 Protons in the arc sections of LHC feel a centripetal force exerted by the dipole  
 1033 magnets; the magnitude of magnetic field needed to keep the protons in the LHC  
 1034 curved trayectomy can be found assuming that protons travel at  $v \approx c$ , using the  
 1035 standard values for proton mass and charge and the LHC radius, as

$$F_m = \frac{mv^2}{r} = qBv \rightarrow B = 8.33T \quad (3.1)$$

1036 which is about 100000 times the Earth's magnetic field. A representation of the



**Figure 3.5:** Top: LHC dipole magnet transverse view; cooling, shielding and mechanical support are indicated. Bottom left: Magnetic field generated by the dipole magnets; note that the direction of the field inside one beam pipe is opposite with respect to the other beam pipe which guarantee that both proton beams are curved in the same direction towards the center of the ring. The effect of the dipole magnetic field on the proton beam is represented on the bottom right side [64, 67, 68].

1037 magnetic field generated by the dipole magnets is shown on the bottom left side of  
 1038 Figure 3.5. The bending effect of the magnetic field on the proton beam is shown on  
 1039 the bottom right side of Figure 3.5. Note that the dipole magnets are not curved;  
 1040 the arc section of the LHC ring is composed of straight dipole magnets of about 15  
 1041 m. In total there are 1232 dipole magnets along the LHC ring.  
 1042 In addition to the bending of the beam trajectory, the beam has to be focused. The

1043 focusing is performed by quadrupole magnets installed in a different straight section;  
 1044 in total 858 quadrupole magnets are installed along the LHC ring. Other effects like  
 1045 electromagnetic interaction among bunches, interaction with electron clouds from the  
 1046 beam pipe, the gravitational force on the protons, differences in energy among protons  
 1047 in the same bunch, among others, are corrected using sextupole and other magnetic  
 1048 multipoles.

1049 The two proton beams inside the LHC ring are made of bunches with a cylindrical  
 1050 shape of about 7.5 cm long and about 1 mm in diameter; when bunches are close to the  
 1051 interaction point (IP), the beam is focused up to a diameter of about 16  $\mu\text{m}$  in order  
 1052 to maximize the probability of collisions between protons. The number of collisions  
 1053 per second is proportional to the cross section of the bunches with the *luminosity* ( $L$ )  
 1054 as the proportionality factor, thus, the luminosity can be calculated using

$$L = fn \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (3.2)$$

1055 where  $f$  is the revolution frequency,  $n$  is the number of bunches per beam,  $N_1$  and  $N_2$   
 1056 are the numbers of protons per bunch,  $\sigma_x$  and  $\sigma_y$  are the gaussian transverse sizes of  
 1057 the bunches. With the expected parameters, the LHC expected luminosity is about

$$f = \frac{v}{2\pi r_{LHC}} \approx \frac{3 \times 10^8 \text{ m/s}}{27 \text{ km}} \approx 11.1 \text{ kHz},$$

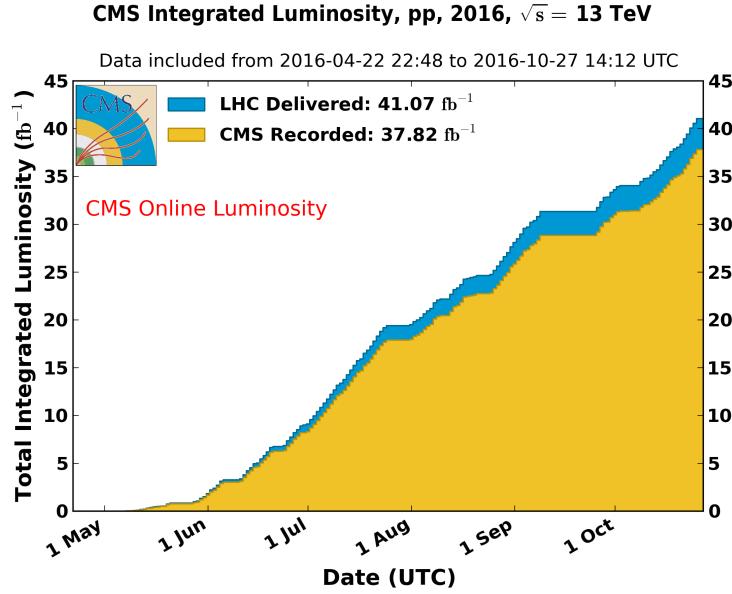
$$n = 2808$$

$$N_1 = N_2 \sim 1.5 \times 10^{11}$$

$$\sigma_x = \sigma_y = 16 \mu\text{m}$$

1058

$$L = 1.28 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1} = 1.28 \times 10^{-5} \text{ fb}^{-1} \text{ s}^{-1} \quad (3.3)$$



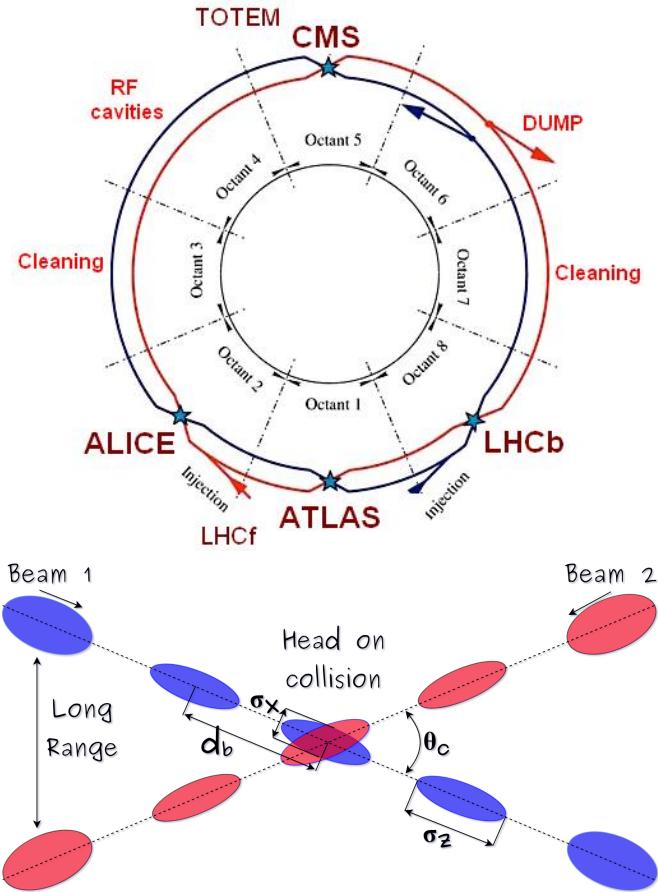
**Figure 3.6:** Integrated luminosity delivered by LHC and recorded by CMS during 2016. The difference between the delivered and the recorded luminosity is due to fails and issues occurred during the data taking in the CMS experiment [69].

1059        Luminosity is a fundamental aspect of LHC given that the bigger luminosity the  
 1060      bigger number of collisions, which means that for processes with a very small cross  
 1061      section the number of expected occurrences is increased and so the chances of being  
 1062      detected. The integrated luminosity, collected by the CMS experiment during 2016  
 1063      is shown in Figure 3.6; the data analyzed in this thesis corresponds to an integrated  
 1064      luminosity of  $35.9 \text{ fb}^{-1}$  at a center of mass-energy  $\sqrt{s} = 13 \text{ TeV}$ .

1065        One way to increase  $L$  is increasing the number of bunches in the beam. Cur-  
 1066      rently, the separation between two consecutive bunches in the beam is 7.5 m which  
 1067      corresponds to a time separation of 25 ns. In the full LHC ring the allowed number  
 1068      of bunches is  $n = 27\text{km}/7.5\text{m} = 3600$ ; however, there are some gaps in the bunch pat-  
 1069      tern intended for preparing the dumping and injection of the beam, thus, the proton  
 1070      beams are composed of 2808 bunches.

1071        Once the proton beams reach the desired energy, they are brought to cross each

1072 other producing  $pp$  collisions. The bunch crossing happens in precise places where  
 1073 the four LHC experiments are located, as seen in the top of Figure 3.7. In 2008  $pp$   
 1074 collisions of  $\sqrt{s} = 7$  TeV were performed; the energy was increased to 8 TeV in 2012  
 1075 and to 13 TeV in 2015.



**Figure 3.7:** Top: LHC interaction points. Bunch crossing occurs where the LHC experiments are located [70]. Sections indicated as cleaning are dedicated to collimate the beam in order to protect the LHC ring from collisions with protons in very spreaded bunches. Bottom: bunch crossing scheme. Since the bunch crossing is not perfectly head-on, the luminosity is reduced in a factor of 17%; adapted from Reference [84].

1076 The CMS and ATLAS experiments are multi-purpose experiments, hence, they  
 1077 are enabled to explore physics in any of the LHC collision modes. LHCb experiment  
 1078 is optimized to explore bottom quark physics, while ALICE is optimized for heavy

1079 ion collisions searches; TOTEM and LHCf are dedicated to forward physics studies;  
 1080 MoEDAL (not indicated in the Figure) is intended for monopole or massive pseudo  
 1081 stable particles searches.

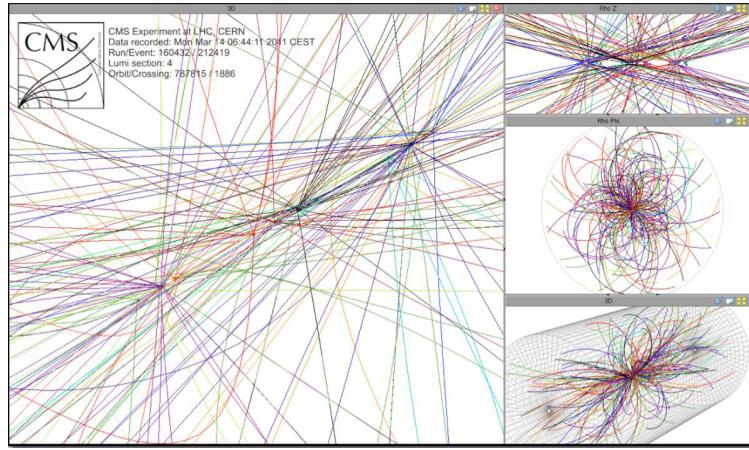
1082 At the IP there are two interesting details that need to be addressed. The first  
 1083 one is that the bunch crossing does not occur head-on but at a small crossing angle  $\theta_c$   
 1084 (280  $\mu$ rad in CMS and ATLAS) as shown in the bottom side of Figure 3.7, affecting  
 1085 the overlapping between bunches; the consequence is a reduction of about 17% in  
 1086 the luminosity (represented by a factor not included in eqn. 3.2). The second one  
 1087 is the occurrence of multiple  $pp$  collisions in the same bunch crossing; this effect is  
 1088 called *pileup* (PU). A fairly simple estimation of the PU follows from estimating the  
 1089 probability of collision between two protons, one from each of the bunches in the  
 1090 course of collision; it depends roughly on the ratio of proton size and the cross section  
 1091 of the bunch in the IP, i.e.,

$$P(pp\text{-}collision) \sim \frac{d_{proton}^2}{\sigma_x \sigma_y} = \frac{(1\text{fm})^2}{(16\mu\text{m})^2} \sim 4 \times 10^{-21} \quad (3.4)$$

1092 however, there are  $N = 1.15 \times 10^{11}$  protons in a bunch, thus the estimated number of  
 1093 collisions in a bunch crossing is

$$PU = N^2 * P(pp\text{-}collision) \sim 50pp \text{ collision per bunch crossing}, \quad (3.5)$$

1094 about 20 of which are inelastic. A multiple  $pp$  collision event in a bunch crossing at  
 1095 CMS is shown in Figure 3.8.

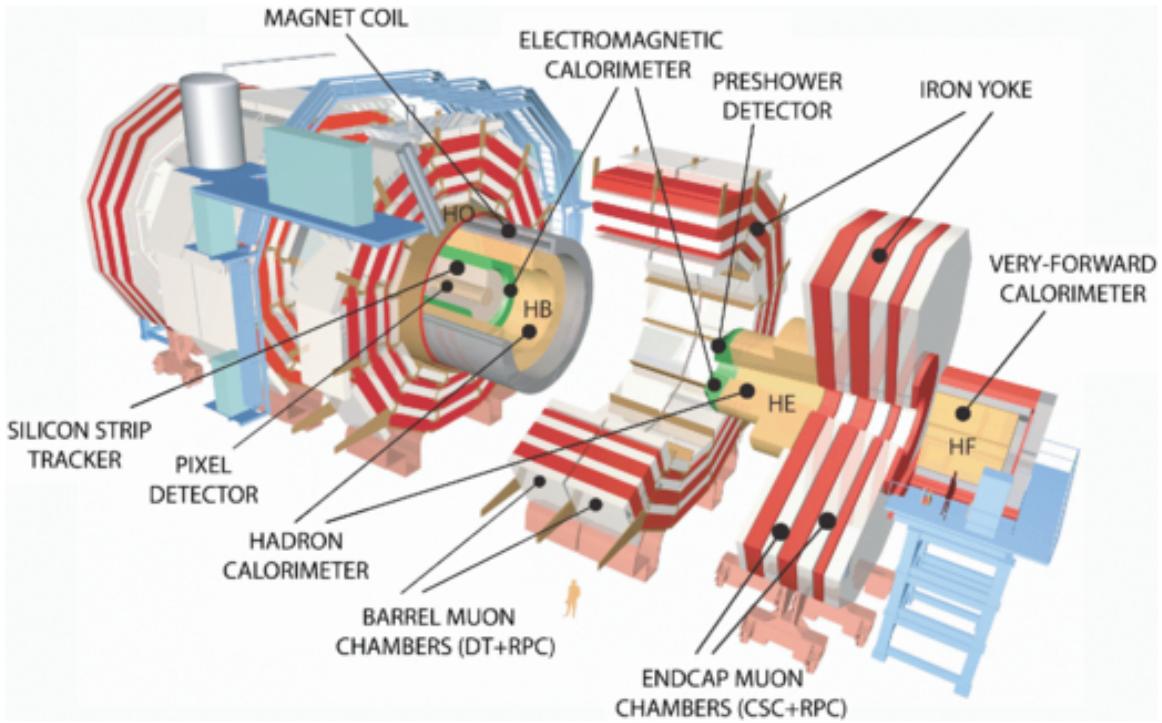


**Figure 3.8:** Multiple  $p\bar{p}$  collision bunch crossing at CMS. [71].

### 1096 3.3 The CMS experiment

1097 CMS is a general-purpose detector designed to conduct research in a wide range  
 1098 of physics from the standard model to new physics like extra dimensions and dark  
 1099 matter. Located at Point 5 in the LHC layout as shown in Figure 3.4, CMS is  
 1100 composed of several detection systems distributed in a cylindrical structure; in total,  
 1101 CMS weights about 12500 tons in a very compact 21.6 m long and 14.6 m diameter  
 1102 cylinder. It was built in 15 separate sections at the ground level and lowered to the  
 1103 cavern individually to be assembled. A complete and detailed description of the CMS  
 1104 detector and its components is given in Reference [72] on which this section is based.  
 1105 Figure 3.9 shows the layout of the CMS detector. The detection system is composed  
 1106 of (from the innermost to the outermost)

- 1107     • Pixel detector.
- 1108     • Silicon strip tracker.
- 1109     • Preshower detector.
- 1110     • Electromagnetic calorimeter.



**Figure 3.9:** Layout of the CMS detector. The several subdetectors are indicated. The central region of the detector is referred as the barrel section while the endcaps are referred as the forward sections. [73].

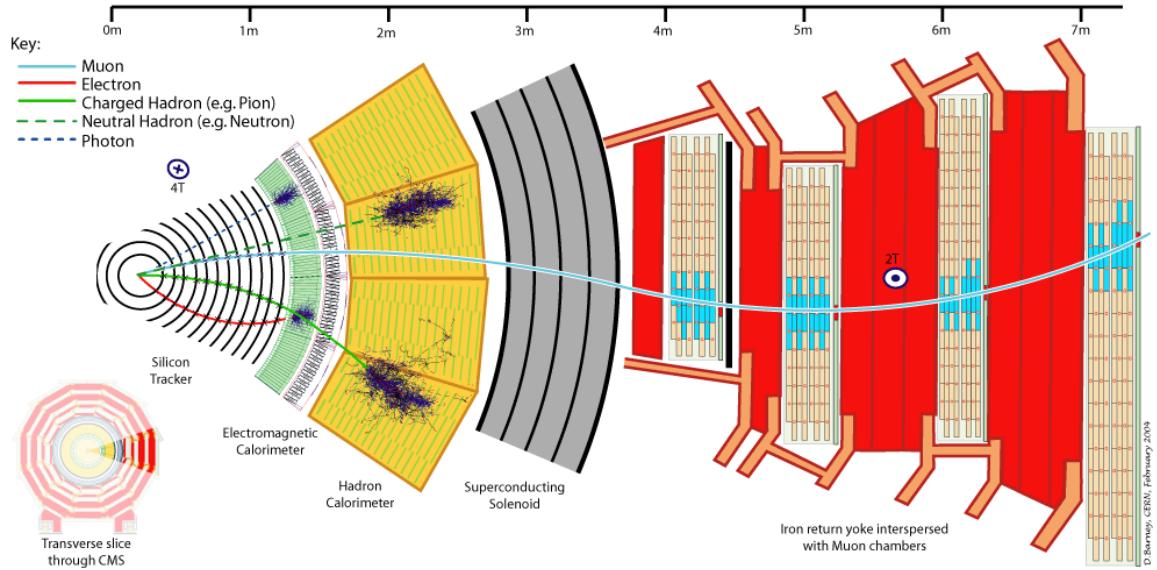
1111        • Hadronic calorimeter.

1112        • Muon chambers (barrel and endcap)

1113        The central region of the detector is commonly referred as the barrel section while  
 1114        the endcaps are referred as the forward sections of the detector; thus, each subdetector  
 1115        is composed of a barrel section and a forward section.

1116        When a  $pp$  collision happens inside the CMS detector, many different particles are  
 1117        produced, but only some of them live long enough to be detected; they are electrons,  
 1118        photons, pions, kaons, protons, neutrons and muons; neutrinos are not detected by  
 1119        the CMS detector. Thus, the CMS detector was designed to detect those particles and  
 1120        measure their properties. Figure 3.10 shows a transverse slice of the CMS detector.  
 1121        The silicon tracker (pixel detector + strip tracker) is capable to register the track of

the charged particles traversing it, while calorimeters (electromagnetic and hadronic) measure the energy of the particles that are absorbed by their materials. Considering the detectable particles, mentioned above, emerging from the IP, a basic description of the detection process is as follows.



**Figure 3.10:** CMS detector transverse slice [74].

A muon emerging from the IP, will create a track on the silicon tracker and on the muon chambers. The design of the CMS detector is driven by the requirements on the identification, momentum resolution and unambiguous charge determination of the muons; therefore, a large bending power is provided by the solenoid magnet made of superconducting cable capable of generating a 3.8 T magnetic field. The muon track is bent twice since the magnetic field inside the solenoid is directed along the  $z$ -direction but outside its direction is reversed. Muons interact very weakly with the calorimeters, therefore, it is not absorbed but escape away from the detector.

An electron emerging from the IP will create a track along the tracker which will be bent due to the presence of the magnetic field, later, it will be absorbed in the electromagnetic calorimeter where its energy is measured.

1137        A photon will not leave a track because it is neutral, but it will be absorbed in  
 1138        the electromagnetic calorimeter.

1139        A neutral hadron, like the neutron, will not leave a track either but it will lose a  
 1140        small amount of its energy during its passage through the electromagnetic calorimeter  
 1141        and then it will be absorbed in the hadron calorimeter depositing the rest of its energy.

1142        A charged hadron, like the proton or  $\pi^\pm$ , will leave a curved track on the silicon  
 1143        tracker, some of its energy in the electromagnetic calorimeter and finally will be  
 1144        absorbed in the hadronic calorimeter.

1145        A more detailed description of each detection system will be presented in the  
 1146        following sections.

### 1147        3.3.1 CMS coordinate system

1148        The coordinate system used by CMS is centered on the geometrical center of the  
 1149        detector which is the nominal IP as shown in Figure 3.11<sup>1</sup>. The  $z$ -axis is parallel  
 1150        to the beam direction, while the  $Y$ -axis pointing vertically upward, and the  $X$ -axis  
 1151        pointing radially inward toward the center of the LHC.

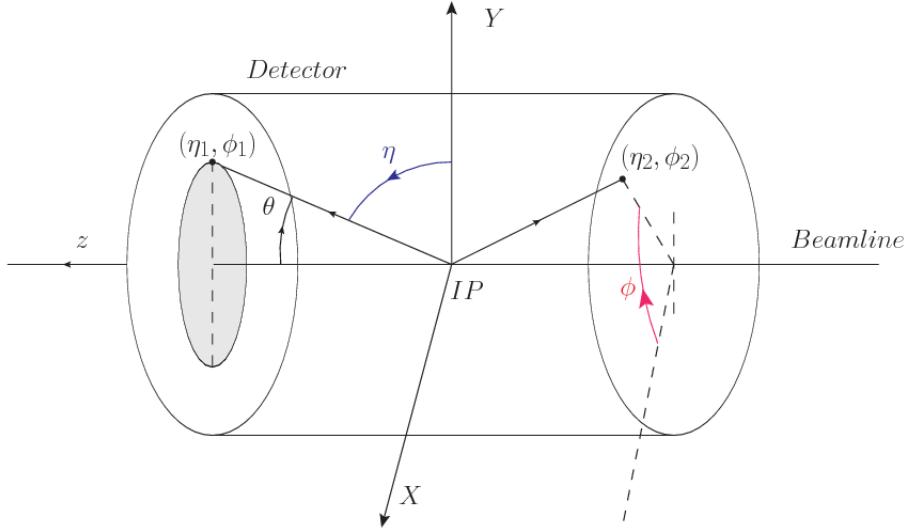
1152        In addition to the common cartesian and cylindrical coordinate systems, two co-  
 1153        ordinates are of particular utility in particle physics: rapidity ( $y$ ) and pseudorapidity  
 1154        ( $\eta$ ), defined in connection to the polar angle  $\theta$ , energy and longitudinal momentum  
 1155        component (momentum along the  $z$ -axis) according to

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad \eta = -\ln \left( \tan \frac{\theta}{2} \right) \quad (3.6)$$

1156        Rapidity is related to the angle between the  $XY$ -plane and the direction in which  
 1157        the products of a collision are emitted; it has the nice property that the difference

---

<sup>1</sup> Not all the  $pp$  interaction occur at the nominal IP because of the bunch lenght, therefore, each  $pp$  collision has its own IP location



**Figure 3.11:** CMS detector coordinate system.

1158 between the rapidities of two particles is invariant with respect to Lorentz boosts  
 1159 along the  $z$ -axis, hence, data analysis becomes more simple when based on rapid-  
 1160 ity; however, it is not simple to measure the rapidity of highly relativistic particles,  
 1161 as those produced after  $pp$  collisions. Under the highly relativistic motion approxi-  
 1162 mation,  $y$  can be rewritten in terms of the polar angle, concluding that rapidity is  
 1163 approximately equal to the pseudorapidity defined above, i.e.,  $y \approx \eta$ . Note that  $\eta$   
 1164 is easier to measure than  $y$  given the direct relationship between the former and the  
 1165 polar angle.

1166       The angular distance between two objects in the detector ( $\Delta R$ ) is commonly used  
 1167 to judge the isolation of those object; it is defined in terms of their coordinates  $(\eta_1, \phi_1)$ ,  
 1168  $(\eta_2, \phi_2)$  as

$$\Delta R = \sqrt{(\Delta\eta)^2 - (\Delta\phi)^2} \quad (3.7)$$

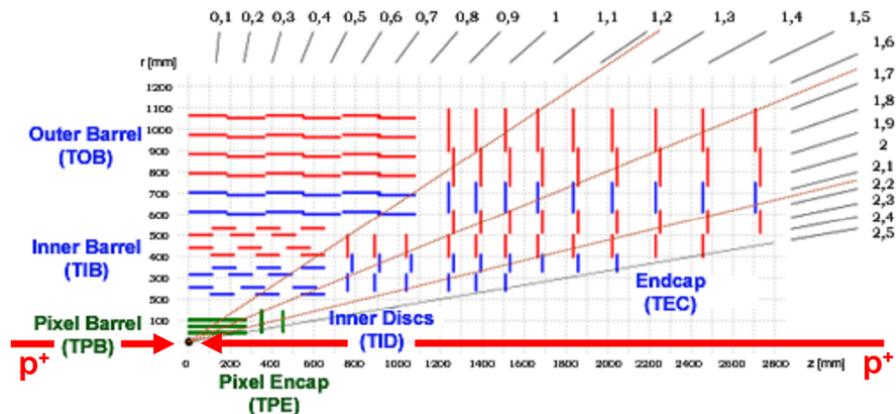
### 1169 3.3.2 Tracking system

1170 The CMS tracking system is designed to provide a precise measurement of the trajectories (*track*) followed by the charged particles created after the  $pp$  collisions; also, the  
 1171 precise reconstruction of the primary and secondary origins of the tracks (*vertices*) is  
 1172 expected in an environment where, each 25 ns, the bunch crossing produces about 20  
 1173 inelastic collisions and about 1000 particles.  
 1174

1175 Physics requirements guiding the tracking system performance include the precise characterization of events involving gauge bosons, W and Z, and their leptonic  
 1176 decays for which an efficient isolated lepton and photon reconstruction is of capital  
 1177 importance, given that isolation is required to suppress background events to a level  
 1178 that allows observations of interesting processes like Higgs boson decays or beyond  
 1179 SM events.

1181 The ability to identify and reconstruct  $b$ -jets and B-hadrons within these jets is also  
 1182 a fundamental requirement, achieved through the ability to reconstruct accurately  
 1183 displaced vertices, given that  $b$ -jets are part of the signature of top quark physics, like  
 1184 the one treated in this thesis.

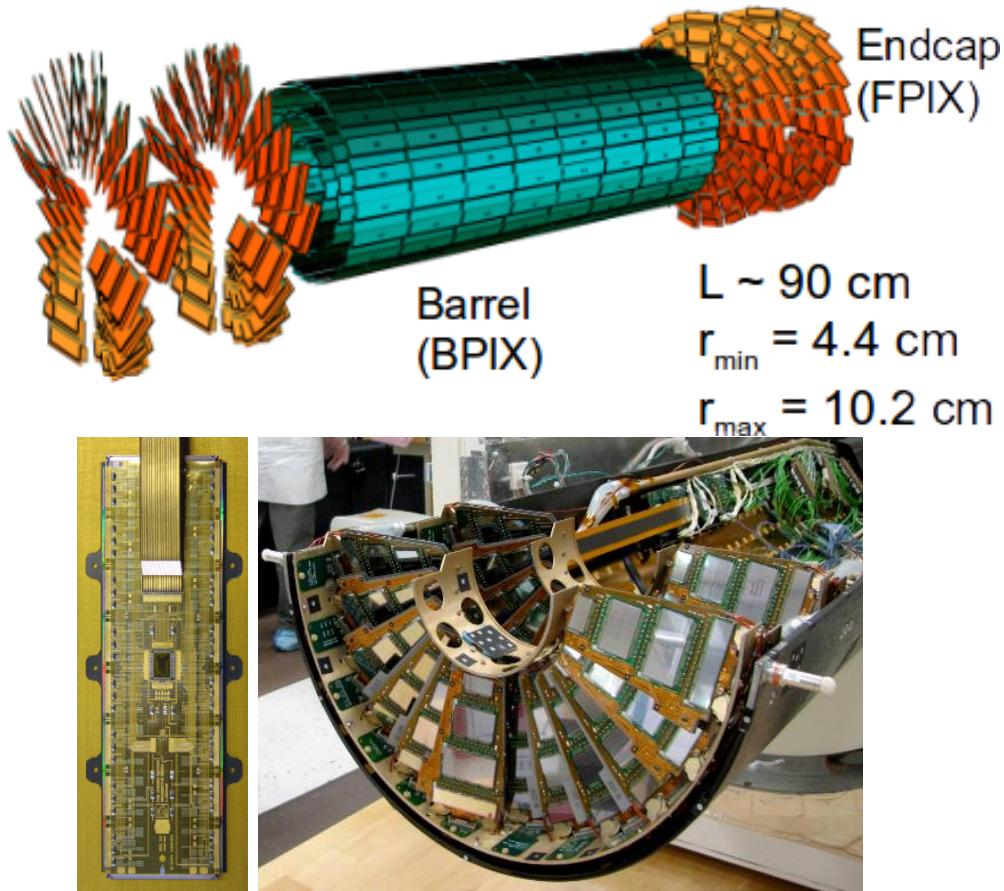
1185 An schematic view of the CMS tracking system is shown in Figure 3.12



**Figure 3.12:** CMS tracking system schematic view [76].

1186 In order to satisfy these performance requirements, the tracking system uses two  
 1187 different detector subsystems arranged in concentric cylindrical volumes, the pixel  
 1188 detector and the silicon strip tracker; the pixel detector is located in the high particle  
 1189 density region ( $r < 20\text{cm}$ ) while the silicon strip tracker is located in the medium and  
 1190 lower particle density regions  $20\text{cm} < r < 116\text{cm}$ .

1191 **Pixel detector**



**Figure 3.13:** CMS pixel detector. Top: schematic view; Bottom: pictures of a barrel(BPIX) module(left) and forward modules(right) [72].

1192 The pixel detector was replaced during the 2016-2017 extended year-end technical  
 1193 stop, due to the increasingly challenging operating conditions like the higher particle

flux and more radiation harsh environment, among others. The new one is responding as expected, reinforcing its crucial role in the successful way to fulfill the new LHC physics objectives after the discovery of the Higgs boson. Since the data sets used in this thesis were produced using the previous version of the pixel detector, it will be the subject of the description in this section. The last chapter of this thesis is dedicated to describe my contribution to the *Forward Pixel Phase 1 upgrade*.

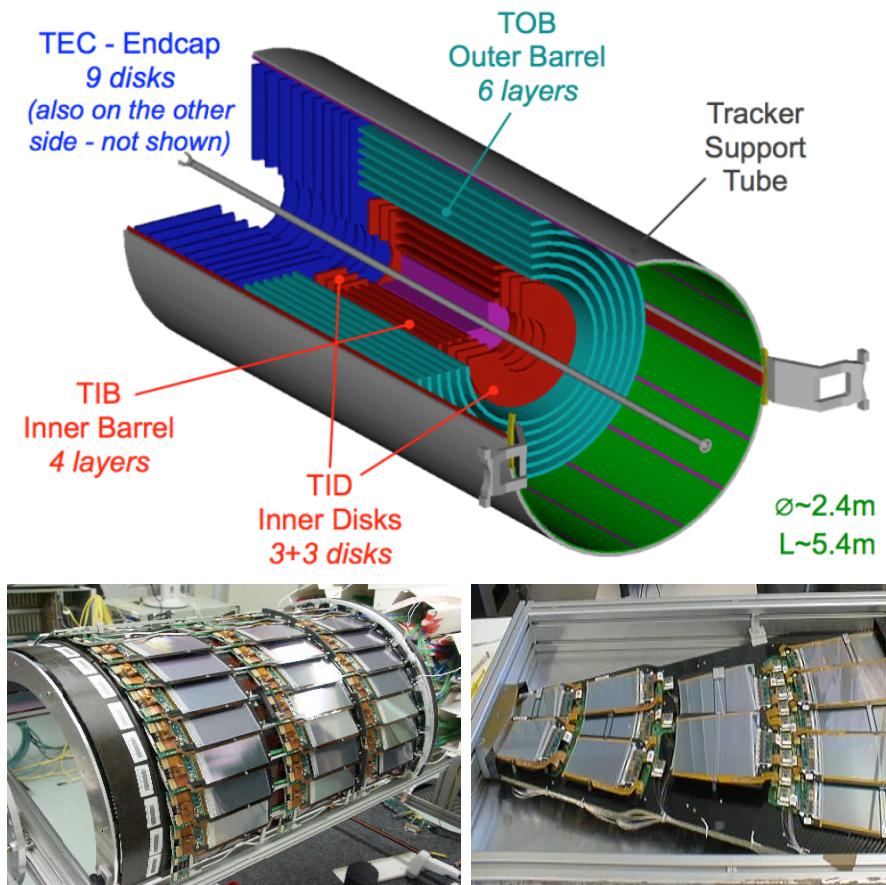
The pixel detector was composed of 1440 silicon pixel detector modules organized in three-barrel layers in the central region (BPix) and two disks in the forward region (FPix) as shown in the top side of Figure 3.13; it was designed to record efficiently and with high precision, up to  $20\mu\text{m}$  in the  $XY$ -plane and  $20\mu\text{m}$  in the  $z$ -direction, the first three space-points (*hits*) nearest to the IP region in the range  $|\eta| \leq 2.5$ . The first barrel layer was located at a radius of 44 mm from the beamline, while the third layer was located at a radius of 102 mm, closer to the strip tracker inner barrel layer (see Section 3.3.3) in order to reduce the rate of fake tracks. The high granularity of the detector is represented in its about 66 Mpixels, each of size  $100 \times 150\mu\text{m}^2$ . The transverse momentum resolution of tracks can be measured with a resolution of 1-2% for muons of  $p_T = 100$  GeV.

A charged particle passing through the pixel sensors produce ionization in them, giving energy for electrons to be removed from the silicon atoms, hence, creating electron-hole pairs. The collection of charges in the pixels generates an electrical signal that is read out by an electronic readout chip (ROC); each pixel has its own electronics which amplifies the signal. Combining the signal from the pixels activated by a traversing particle in the several layers of the detector allows one to reconstruct the particle's trajectory in 3D.

Commonly, the charge produced by traversing of a particle is collected by and shared among several pixels; by interpolating between pixels, the spatial resolution

is improved. In the barrel section the charge sharing in the  $r\phi$ -plane is due to the Lorentz effect. In the forward pixels the charge sharing is enhanced by arranging the blades in the turbine-like layout as shown in Figure 3.13 bottom left.

### 3.3.3 Silicon strip tracker



**Figure 3.14:** Top: CMS Silicon Strip Tracker (SST) schematic view. The SST is composed of the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). Each part is made of silicon strip modules; the modules in blue represent two modules mounted back-to-back and rotated in the plane of the module by a stereo angle of 120 mrad in order to provide a 3-D reconstruction of the hit positions. Bottom: pictures of the TIB (left) and TEC (right) modules [78–80].

The silicon strip tracker (SST) is the second stage in the CMS tracking system. The top side of Figure 3.14 shows a schematic of the SST. The inner tracker region is

1226 composed of the tracker inner barrel (TIB) and the tracker inner disks (TID) covering  
 1227 the region  $r < 55$  cm and  $|z| < 118$  cm. The TIB is composed of 4 layers while the TID  
 1228 is composed of 3 disks at each end. The silicon sensors in the inner tracker are 320  
 1229  $\mu\text{m}$  thick, providing a resolution of about 13-38  $\mu\text{m}$  in the  $r\phi$  position measurement.

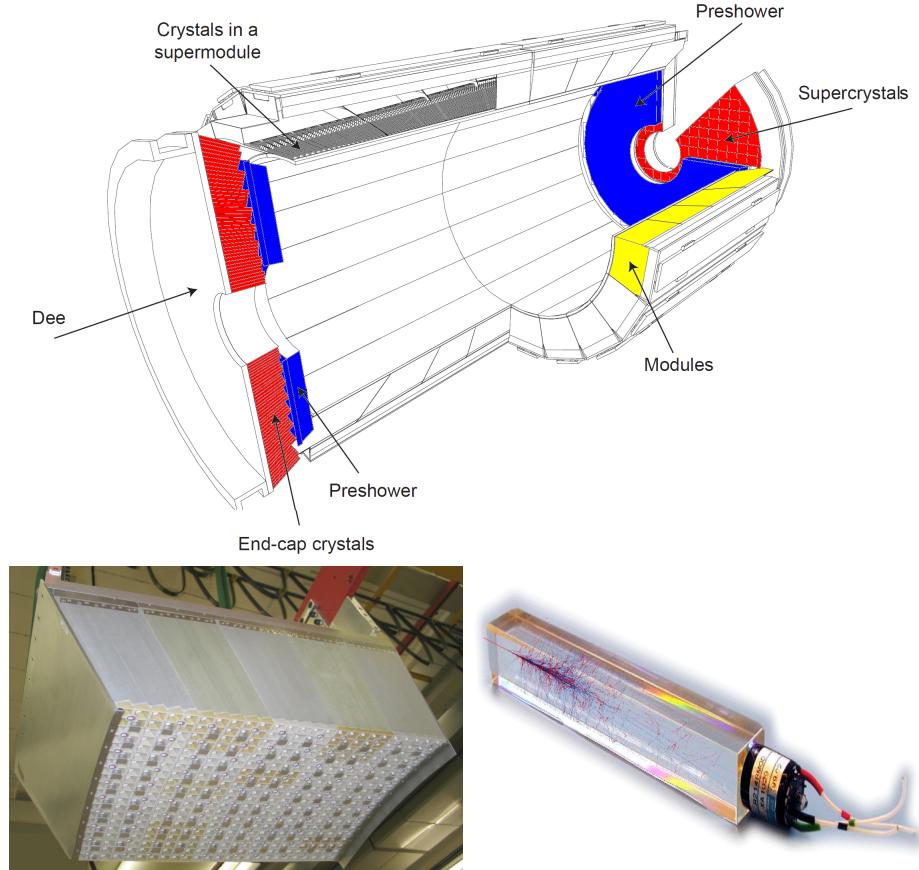
1230 The modules indicated in blue in the schematic view of Figure 3.14 are two mod-  
 1231 ules mounted back-to-back and rotated in the plane of the module by a *stereo* angle  
 1232 of 100 mrad; the hits from these two modules, known as *stereo hits*, are combined to  
 1233 provide a measurement of the second coordinate ( $z$  in the barrel and  $r$  on the disks)  
 1234 allowing the reconstruction of hit positions in 3-D.

1235 The outer tracker region is composed of the tracker outer barrel (TOB) and the  
 1236 tracker endcaps (TEC). The six layers of the TOB offer coverage in the region  $r > 55$   
 1237 cm and  $|z| < 118$  cm, while the 9 disks of the TEC cover the region  $124 < |z| < 282$   
 1238 cm. The resolution offered by the outer tracker is about 13-38  $\mu\text{m}$  in the  $r\phi$  position  
 1239 measurement. The inner four TEC disks use silicon sensors 320  $\mu\text{m}$  thick; those in  
 1240 the TOB and the outer three TEC disks use silicon sensors of 500  $\mu\text{m}$  thickness. The  
 1241 silicon strips run parallel to the  $z$ -axis and the distance between strips varies from 80  
 1242  $\mu\text{m}$  in the inner TIB layers to 183  $\mu\text{m}$  in the inner TOB layers; in the endcaps the  
 1243 wedge-shaped sensors with radial strips, whose pitch range between 81  $\mu\text{m}$  at small  
 1244 radii and 205  $\mu\text{m}$  at large radii.

1245 The whole SST has 15148 silicon modules, 9.3 million silicon strips and cover a  
 1246 total active area of about 198  $\text{m}^2$ .

### 1247 3.3.4 Electromagnetic calorimeter

1248 The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of  
 1249 electrons and photons. It is composed of 75848 lead tungstate crystals which have a



**Figure 3.15:** Top: CMS ECAL schematic view. Bottom: Module equipped with the crystals (left); ECAL crystal(right) with an artistic representation of an electromagnetic shower [72].

1250 short radiation length (0.89 cm) and fast response, since 80% of the light is emitted  
 1251 within 25 ns; however, they are combined with Avalanche photodiodes (APDs) as  
 1252 photodetectors given that crystals themselves have a low light yield ( $30\gamma/\text{MeV}$ ). A  
 1253 schematic view of the ECAL is shown in Figure 3.15.

1254 Energy is measured when electrons and photons are absorbed by the crystals  
 1255 which generates an electromagnetic *shower*, as seen in bottom right picture of the  
 1256 Figure 3.15; the shower is seen as a *cluster* of energy which depending on the amount  
 1257 of energy deposited can involve several crystals. The ECAL barrel (EB) covers the  
 1258 region  $|\eta| < 1.479$ , using crystals of depth of 23 cm and  $2.2 \times 2.2 \text{ cm}^2$  transverse

1259 section; the ECAL endcap (EE) covers the region  $1.479 < |\eta| < 3.0$  using crystals of  
 1260 depth 22 cm and transverse section of  $2.86 \times 2.86 \text{ cm}^2$ ; the photodetectors used are  
 1261 vacuum phototriodes (VPTs). Each EE is divided in two structures called *Dees*.

1262 The preshower detector (ES) is installed in front of the EE and covers the region  
 1263  $1.653 < |\eta| < 2.6$ . The ES provides a precise measurement of the position of electro-  
 1264 magnetic showers, which allows to distinguish electrons and photon signals from  $\pi^0$   
 1265 decay signals. The ES is composed of a layer of lead radiators followed by a layer of  
 1266 silicon strip sensors. The lead radiators initiate electromagnetic showers when reached  
 1267 by photons and electrons, then, the strip sensors measure the deposited energy and  
 1268 the transverse shower profiles. The full ES thickness is 20 cm.

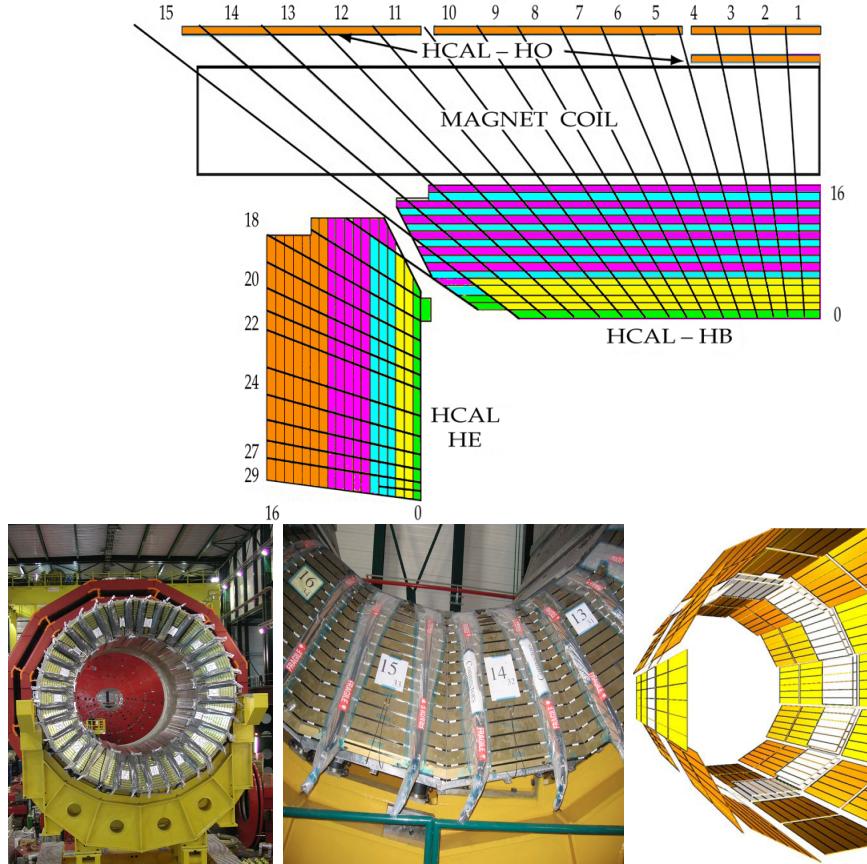
### 1269 3.3.5 Hadronic calorimeter

1270 Hadrons are not absorbed by the ECAL<sup>2</sup> but by the hadron calorimeter (HCAL),  
 1271 which is made of a combination of alternating brass absorber layers and silicon photo-  
 1272 multiplier(SiPM) layers; therefore, particles passing through the scintillator material  
 1273 produce showers, as in the ECAL, as a result of the inelastic scattering of the hadrons  
 1274 with the detector material. Since the particles are not absorbed in the scintillator,  
 1275 their energy is sampled; therefore the total energy is not measured but estimated from  
 1276 the energy clusters, which reduces the resolution of the detector. Brass was chosen  
 1277 as the absorber material due to its short interaction length ( $\lambda_I = 16.42\text{cm}$ ) and its  
 1278 non-magnetivity. Figure 3.16 shows a schematic view of the CMS HCAL.

1279 The HCAL is divided into four sections; the Hadron Barrel (HB), the Hadron  
 1280 Outer (HO), the Hadron Endcap (HE) and the Hadron Forward (HF) sections. The  
 1281 HB covers the region  $0 < |\eta| < 1.4$ , while the HE covers the region  $1.3 < |\eta| < 3.0$ .  
 1282 The HF, made of quartz fiber scintillator and steel as absorption material, covers the

---

<sup>2</sup> Most hadrons are not absorbed, but few low-energy ones might be.

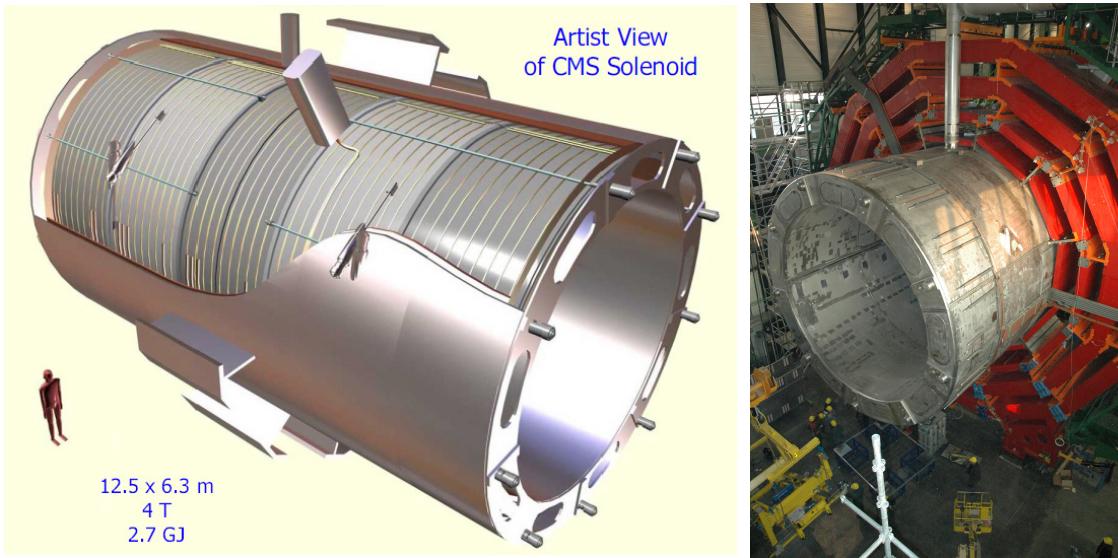


**Figure 3.16:** Top: CMS HCAL schematic view, the colors indicate the layers that are grouped into the same readout channels. Bottom: picture of a section of the HB; the absorber material is the golden region and scintillators are placed in between the absorber material (left and center). Schematic view of the HO (right). [81,82]

1283 forward region  $3.0 < |\eta| < 5.2$ . Both the HB and HF are located inside the solenoid.  
 1284 The HO is placed outside the magnet as an additional layer of scintillators with the  
 1285 purpose of measure the energy tails of particles passing through the HB and the  
 1286 magnet (see Figure 3.16 top and bottom right).

### 1287 3.3.6 Superconducting solenoid magnet

1288 The superconducting magnet installed in the CMS detector is designed to provide  
 1289 an intense and highly uniform magnetic field in the central part of the detector.

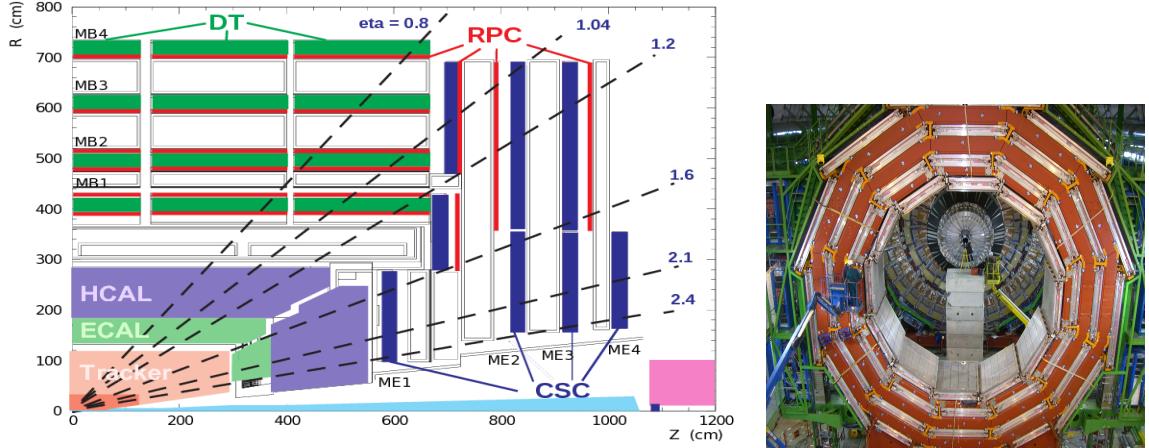


**Figure 3.17:** Artistic representation of the CMS solenoid magnet(left). The magnet is supported on an iron yoke (right) which also serves as the house of the muon detector and as mechanical support for the whole CMS detector [75].

1290 In fact, the tracking system takes advantage of the bending power of the magnetic  
 1291 field to measure with precision the momentum of the particles that traverse it; the  
 1292 unambiguous determination of the sign for high momentum muons was a driving  
 1293 principle during the design of the magnet. The magnet has a diameter of 6.3 m, a  
 1294 length of 12.5 m and a cold mass of 220 t; the generated magnetic field reaches a  
 1295 strength of 3.8T. Since it is made of Ni-Tb superconducting cable it has to operate at  
 1296 a temperature of 4.7 K by using a helium cryogenic system; the current circulating in  
 1297 the cables reaches 18800 A under normal running conditions. The left side of Figure  
 1298 3.17 shows an artistic view of the CMS magnet, while the right side shows a transverse  
 1299 view of the cold mass where the winding structure is visible.

1300 The yoke (see Figure 3.17), composed of 5 barrel wheels and 6 endcap disks made  
 1301 of iron, serves not only as the media for magnetic flux return but also provides housing  
 1302 for the muon detector system and structural stability to the full detector.

1303 **3.3.7 Muon system**



**Figure 3.18:** Left: CMS muon system schematic view; Right: one of the yoke rings with the muon DTs and RPCs installed; in the back it is possible to see the muon endcap [83].

1304 Muons are the only charged particles able to pass through all the CMS detector  
 1305 due to their low ionization energy loss; thus, muons can be separated easily from the  
 1306 high amount of particles produced in a  $pp$  collision. Also, muons are expected to be  
 1307 produced in the decay of several new particles; therefore, good detection of muons  
 1308 was one of the leading principles when designing the CMS detector.

1309 The CMS muon detection system (muon spectrometer) is embedded in the return  
 1310 yoke as seen in Figure 3.18. It is composed of three different detector types, the drift  
 1311 tube chambers (DT), cathode strip chambers (CSC), and resistive plate chambers  
 1312 (RPC); DT are located in the central region  $\eta < 1.2$  arranged in four layers of drift  
 1313 chambers filled with an Ar/CO<sub>2</sub> gas mixture.

1314 The muon endcaps are made of CSCs covering the region  $\eta < 2.4$  and filled with  
 1315 a mixture of Ar/CO<sub>2</sub>/CF<sub>4</sub>. The reason behind using a different detector type lies on  
 1316 the different conditions in the forward region like the higher muon rate and higher  
 1317 residual magnetic field compared to the central region.

1318 The third type of detector used in the muon system is a set of four disks of RPCs

1319 working in avalanche mode. The RPCs provide good spatial and time resolutions. The  
 1320 track of high- $p_T$  muon candidates is built combining information from the tracking  
 1321 system and the signal from up to six RPCs and four DT chambers.

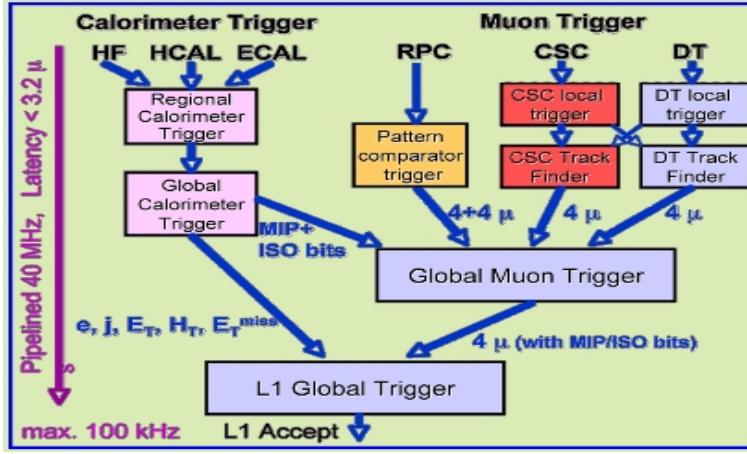
1322 The muon tracks are reconstructed from the hits in the several layers of the muon  
 1323 system.

1324 **3.3.8 CMS trigger system**

1325 CMS expects  $pp$  collisions every 25 ns, i.e., an interaction rate of 40 MHz for which  
 1326 it is not possible to store the recorded data in full. In order to handle this high event  
 1327 rate data, an online event selection, known as triggering, is performed; triggering  
 1328 reduces the event rate to 100 Hz for storage and further offline analysis.

1329 The trigger system starts with a reduction of the event rate to 100 kHz in the  
 1330 so-called *the level 1 trigger* (L1). L1 is based on dedicated programmable hardware  
 1331 like Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated  
 1332 Circuits (ASICs), partly located in the detector itself; another portion is located in  
 1333 the CMS underground cavern. Hit pattern information from the muon chambers  
 1334 and the energy deposits in the calorimeter are used to decide if an event is accepted  
 1335 or rejected, according to selection requirements previously defined, which reflect the  
 1336 interesting physics processes. Figure 3.19 shows the L1 trigger architecture.

1337 The second stage in the trigger system is called *the high-level trigger* (HLT); events  
 1338 accepted by L1 are passed to HLT in order to make an initial reconstruction of them.  
 1339 HLT is software based and runs on a dedicated server farm, using selection algorithms  
 1340 and high-level object definitions; the event rate at HLT is reduced to 100 Hz. The  
 1341 first HLT stage takes information from the muon detectors and the calorimeters to  
 1342 make the initial object reconstruction; in the next HLT stage, information from the



**Figure 3.19:** CMS Level-1 trigger architecture [84].

pixel and strip detectors is used to do first fast tracking and then full tracking online.

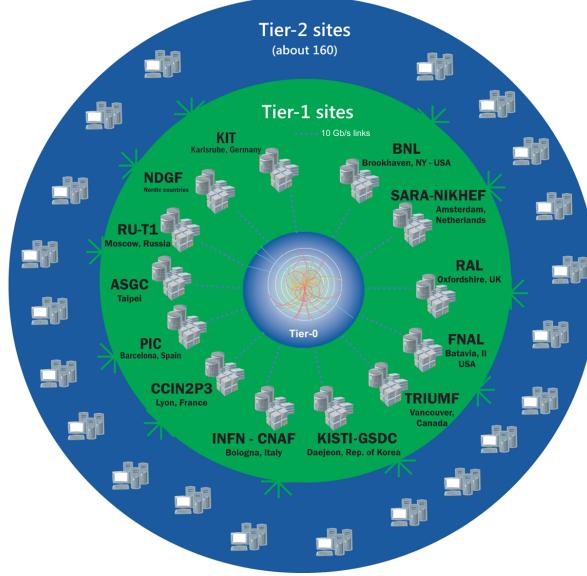
This initial object reconstruction is used in further steps of the trigger system.

Events and preliminary reconstructed physics objects from HLT are sent to be fully reconstructed at the CERN computing center known also as Tier-0 facility. Again, the pixel detector information provides high-quality seeds for the track reconstruction algorithm offline, primary vertex reconstruction, electron and photon identification, muon reconstruction,  $\tau$  identification, and b-tagging. After full reconstruction, data sets are made available for offline analyses.

### 3.3.9 CMS computing

Data coming from the experiment have to be stored and made available for further analysis; in order to cope with all the tasks implied in the offline data processing, like transfer, simulation, reconstruction and reprocessing, among others, a large computing power is required. The CMS computing system is based on the distributed architecture concept, where users of the system and physical computer centers are distributed worldwide and interconnected by high-speed networks.

The worldwide LHC computing grid (WLCG) is the mechanism used to provide



**Figure 3.20:** WLCG structure. The primary computer centers (Tier-0) are located at CERN (data center) and at the Wigner datacenter in Budapest. Tier-1 is composed of 13 centers and Tier-2 is composed of about 160 centers. [85].

1359 that distributed environment. WLCG is a tiered structure connecting computing  
 1360 centers around the world, which provides the necessary storage and computing facil-  
 1361 ities. The primary computing centers of the WLCG are located at the CERN and  
 1362 the Wigner datacenter in Budapest and are known as Tier-0 as shown in Figure 3.20.  
 1363 The main responsibilities for each tier level are [85]

- 1364     • **Tier-0:** initial reconstruction of recorded events and storage of the resulting  
   1365       datasets, the distribution of raw data to the Tier-1 centers.
- 1366     • **Tier-1:** provide storage capacity, support for the Grid, safe-keeping of a pro-  
   1367       portional share of raw and reconstructed data, large-scale reprocessing and safe-  
   1368       keeping of corresponding output, generation of simulated events, distribution  
   1369       of data to Tier 2s, safe-keeping of a share of simulated data produced at these  
   1370       Tier 2s.
- 1371     • **Tier-2:** store sufficient data and provide adequate computing power for spe-

1372       cific analysis tasks and proportional share of simulated event production and  
1373       reconstruction.

1374       Aside from the general computing strategy to manage the huge amount of data  
1375       produced by experiments, CMS uses a software framework to perform a variety of  
1376       processing, selection and analysis tasks. The central concept of the CMS data model  
1377       referred to as *event data model* (EDM) is the *Event*; therefore, an event is the unit  
1378       that contains the information from a single bunch crossing, any data derived from  
1379       that information like the reconstructed objects, and the details of the derivation.

1380       Events are passed as the input to the *physics modules* that obtain information  
1381       from them and create new information; for instance, *event data producers* add new  
1382       data into the events, *analyzers* produce an information summary from an event set,  
1383       *filters* perform selection and triggering.

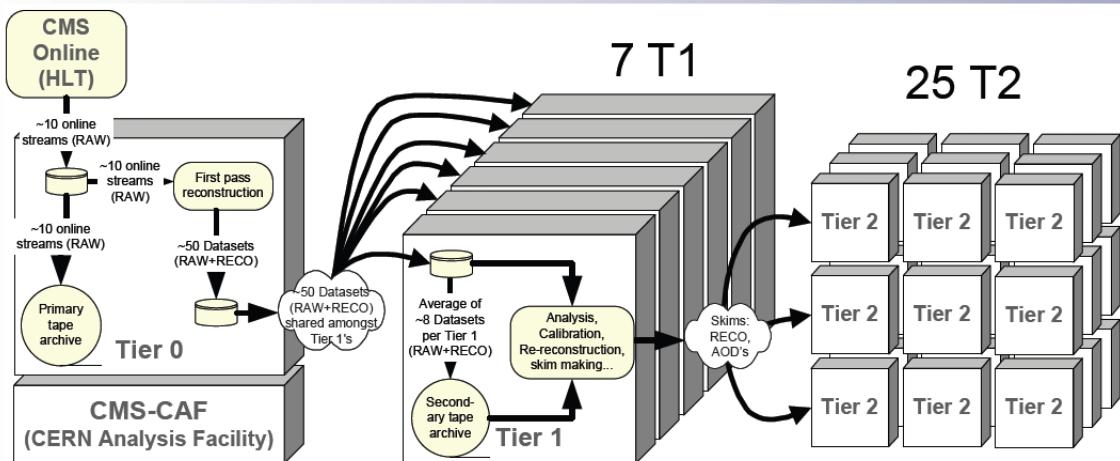
1384       CMS uses several event formats with different levels of detail and precision

1385       • **Raw format:** events in this format contain the full recorded information from  
1386       the detector as well as trigger decision and other metadata. An extended version  
1387       of raw data is used to store information from the CMS Monte Carlo simulation  
1388       tools (see Chapter 4). Raw data are stored permanently, occupying about  
1389       2MB/event

1390       • **RECO format:** events in this format correspond to raw data that have been  
1391       submitted to reconstruction algorithms like primary and secondary vertex re-  
1392       construction, particle ID, and track finding. RECO events contain physics ob-  
1393       jects and all the information used to reconstruct them; average size is about 0.5  
1394       MB/event.

- 1395     • **AOD format:** Analysis Object Data (AOD) is the data format used in the  
 1396        physics analyses given that it contains the parameters describing the high-level  
 1397        physics objects in addition to enough information to allow a kinematic refitting if  
 1398        needed. AOD events are filtered versions of the RECO events to which skimming  
 1399        or other filtering have been applied, hence AOD events are subsets of RECO  
 1400        events. Requires about 100 kB/event.
- 1401     • **Non-event data** are data needed to interpret and reconstruct events. Some  
 1402        of the non-event data used by CMS contains information about the detector  
 1403        contraction and condition data like calibrations, alignment, and detector status.

1404     Figure 3.21 shows the data flow scheme between CMS detector and tiers.



**Figure 3.21:** Data flow from CMS detector through tiers.

1405     The whole collection of software built as a framework is referred to as *CMSSW*. This  
 1406        framework provides the services needed by the simulation, calibration and alignment,  
 1407        and reconstruction modules that process event data, so that physicists can perform  
 1408        analysis. The CMSSW event processing model is composed of one executable, called  
 1409        `cmsRun`, and several plug-in modules which contains all the tools (calibration, recon-

1410 struction algorithms) needed to process an event. The same executable is used for  
1411 both detector data and Monte Carlo simulations [86].

1412 **Chapter 4**

1413 **Event generation, simulation and  
1414 reconstruction**

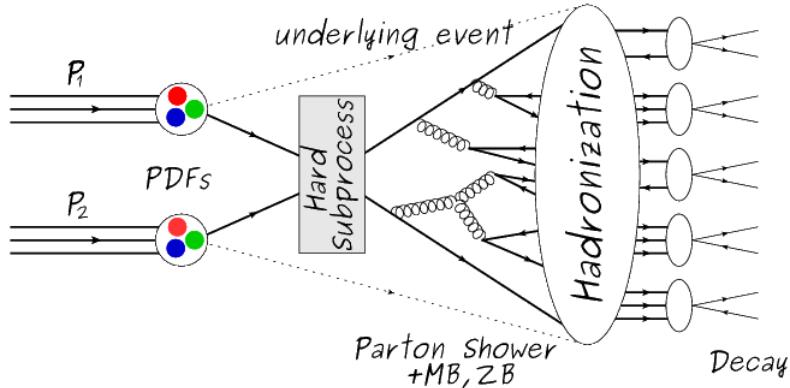
1415 The analysis of data recorded by the CMS experiment involves several stages of  
1416 data processing in order to interpret the information provided by all the detection  
1417 systems; in those stages, the particles produced after the  $pp$  collision are identified by  
1418 reconstructing their trajectories and measuring their features. In addition, the SM  
1419 provides a set of predictions that have to be compared with the experimental results;  
1420 however, in most of the cases, theoretical predictions are not directly comparable  
1421 to experimental results due to the diverse source of uncertainties introduced by the  
1422 experimental setup and theoretical approximations, among others.

1423 The strategy to face these conditions consists of the implementation of theoretical  
1424 models and statistical methods in computational algorithms to produce numerical  
1425 results that can be compared with the experimental results. These computational  
1426 algorithms are commonly known as Monte Carlo (MC) methods and, in the case of  
1427 particle physics, they are designed to apply the SM rules and produce predictions  
1428 about the physical observables measured in the experiments. Since particle physics is

1429 governed by quantum mechanics principles, predictions cannot be made from single  
 1430 events; therefore, a high number of events are *generated* and predictions are produced  
 1431 in the form of statistical distributions for the observables. Effects of the detector  
 1432 presence are included in the predictions by introducing simulations of the detector  
 1433 itself.

1434 This chapter presents a description of the event generation strategy and the tools  
 1435 used to perform the detector simulation and physics object reconstruction. A com-  
 1436 prehensive review of event generators for LHC physics can be found in Reference [87]  
 1437 on which this chapter is based.

## 1438 4.1 Event generation



**Figure 4.1:** Event generation process. The actual interaction is generated in the hard subprocess while the parton shower describes the evolution of the partons from the hard subprocess. Modified from Reference [88].

1439 The event generation is intended to simulate events that mimic the behavior of  
 1440 actual events produced in collisions; they obey a sequence of steps from the particle  
 1441 collision hard process to the decay process into the final state. Figure 4.1 shows a  
 1442 schematic view of the event generation process; the fact that the full process can be  
 1443 treated as several independent steps is motivated by the QCD factorization theorem.

1444 Generation starts by taking into account the PDFs of the incoming particles.  
 1445 Event generators offer the option to chose from several PDF sets depending on the  
 1446 particular process under simulation<sup>1</sup>; in the following,  $pp$  collisions will be consid-  
 1447 ered. The *hard subprocess* describes the actual interaction between partons from the  
 1448 incoming protons; it is represented by the matrix element connecting the initial and  
 1449 final states of the interaction. Normally, the matrix element can be written as a sum  
 1450 over Feynman diagrams with interferences between terms considered in the summa-  
 1451 tion. During the generation of the hard subprocess, the production cross section is  
 1452 calculated.

1453 The order of the Feynman diagrams involved in the calculation, i.e. the order in  
 1454 the perturbation series, determine the order to which the cross section is calculated;  
 1455 therefore, radiative corrections are included by considering higher order Feynman  
 1456 diagrams where QCD radiation dominates. Currently, cross sections calculated to LO  
 1457 do not offer a satisfactory description of the processes, i.e., the results are only reliable  
 1458 for the shape of distributions; therefore, NLO calculations have to be performed with  
 1459 the implication that the computing time needed is highly increased due to increment  
 1460 in the number of diagrams involved and their complexity.

1461 The final parton content of the hard subprocess is subjected to the *parton shower*  
 1462 which generates the gluon radiation. Parton showering evolves the partons, i.e., gluons  
 1463 split into quark-antiquark pairs and quarks with enough energy radiate gluons giv-  
 1464 ing rise to further parton multiplication, following the DGLAP (Dokshitzer-Gribov-  
 1465 Lipatov-Altarelli-Parisi) equations [90–92]. Showering continues until the energy scale  
 1466 is low enough to reach the non-perturbative limit.

1467 In the simulation of LHC processes that involve  $b$  quarks, like the single top quark  
 1468 or Higgs associated production, it is necessary to consider that the  $b$  quark is heavier

---

<sup>1</sup> Different PDF sets under customizable conditions can be plotted using the tool in Reference [89].

1469 than the proton; hence, the QCD interaction description is made in two different  
 1470 schemes [93]

- 1471     • four-flavor (4F) scheme.  $b$  quarks appear only in the final state because they  
 1472       are heavier than the proton and therefore they can be produced only from the  
 1473       splitting of a gluon into pairs or singly in association with a  $t$  quark in high  
 1474       energy-scale interactions; hence, during the simulation, the  $b$ -PDFs are set to  
 1475       zero. Calculations in this scheme are more complicated due to the presence  
 1476       of the second  $b$  quark but the full kinematics is considered already at LO and  
 1477       therefore the accuracy of the description is better.
  
- 1478     • five-flavor (5F) scheme.  $b$  quarks are considered massless, therefore they can  
 1479       appear in both initial and final states since they can now be part of the proton;  
 1480       thus, during the simulation  $b$ -PDFs are not set to zero. In this scheme, calcula-  
 1481       tions are simpler than in the 4F scheme and possible logarithmic divergences  
 1482       are absorbed by the PDFs through the DGLAP evolution.

1483     In this thesis, the  $tHq$  events are generated using the 4F scheme in order to reduce  
 1484     uncertainties, while the  $tHW$  events are generated using the 5F scheme to eliminate  
 1485     LO interference with  $t\bar{t}H$  process [53].

1486     Partons involved in the  $pp$  collision are the focus of the simulation, however, the  
 1487     rest of the partons inside the incoming protons are also affected because the remnants  
 1488     are colored objects; also, multiple parton interactions can occur. The hadronization  
 1489     of the remnants and multiple parton interactions are known as *underlying event* and  
 1490     it has to be included in the simulation. In addition, multiple  $pp$  collisions in the same  
 1491     bunch crossing (pile-up mentioned in 3.2) occurs, actually in two forms

1492     • *in-time PU* which refers to multiple  $pp$  collision in the bunch crossing but that  
 1493       are not considered as primary vertices.

1494     • *out-of-time PU* which refers to overlapping  $pp$  collisions from consecutive bunch  
 1495       crossings; this can occur due to the time delays in the detection systems where  
 1496       information from one bunch crossing is assigned to the next or previous one.

1497       While the underlying event effects are included in generation using generator-  
 1498 specific tools, PU effects are added to the generation by overlaying minimum-bias  
 1499 (MB) and zero-bias (ZB) events to the generated events. MB events are inelastic  
 1500 events from real detector data selected by using a loose trigger with as little bias as  
 1501 possible, therefore accepting a large fraction of the overall inelastic event; ZB events  
 1502 correspond to random events recorded by the detector when collisions are likely. MB  
 1503 models in-time PU and ZB models out-of-time PU.

1504       The next step in the generation process is called *hadronization*. Since particles  
 1505 with a net color charge are not allowed to exits in isolation, they have to recombine  
 1506 to form bound states. This is precisely the process by which the partons resulting  
 1507 from the parton shower arrange themselves as color singlets to form hadrons. At  
 1508 this step, the energy-scale is low and the strong coupling constant is large, therefore  
 1509 hadronization process is non-perturbative and the evolution of the partons is described  
 1510 using phenomenological models. Most of the baryons and mesons produced in the  
 1511 hadronization are unstable and hence they will decay in the detector.

1512       The last step in the generation process corresponds to the decay of the unstable  
 1513 particles generated during hadronization; it is also simulated in the hadronization  
 1514 step, based on the known branching ratios.

1515       The generated MC sample is characterized by the number of events in the sample,  
 1516 the cross section of the simulated process and the luminosity, related as  $\sigma = N_{events}L$ ;

1517 since the cross section of the generated process is fixed, the luminosity of the MC  
 1518 sample varies with the number of events in it. In order to perform a simulation at  
 1519 a given luminosity, it is necessary to *weight* the MC sample to the corresponding  
 1520 experimental luminosity of interest, such that it will be possible to make predictions  
 1521 like the number of events of signal or background that will be produced in a data  
 1522 sample corresponding to that luminosity. In this sense, the weight is the ratio between  
 1523 the desired luminosity and the original MC sample luminosity.

## 1524 4.2 Monte Carlo Event Generators

1525 The event generation described in the previous section has been implemented in  
 1526 several software packages for which a brief description is given in the following.

- 1527     • **PYTHIA 8.** This program is designed to perform the generation of high  
               1528 energy physics events that describes the collisions between particles such as  
               1529 electrons and protons. Several theories and models are implemented in it, in  
               1530 order to describe physical aspects like hard and soft interactions, parton dis-  
               1531 tributions, initial and final-state parton showers, multiple parton interactions,  
               1532 beam remnants, hadronization<sup>2</sup> and particle decay. Thanks to extensive test-  
               1533 ing, several optimized parametrizations, known as *tunings*, have been defined  
               1534 in order to improve the description of actual collisions to a high degree of pre-  
               1535 cision; for analysis at  $\sqrt{s} = 13$  TeV, the underlying event CUETP8M1 tune is  
               1536 employed [95]. The calculation of the matrix element is performed at LO which  
               1537 is not enough for the current required level of precision; therefore, PYTHIA  
               1538 is often used for parton shower, hadronization and decays, while other event  
               1539 generators are used to generate the matrix element at NLO.

---

<sup>2</sup> based in the Lund string model [94]

- 1540     • **MadGraph5\_aMC@NLO.** MadGraph is a matrix element generator which  
 1541       calculates the amplitudes for all contributing Feynman diagrams of a given  
 1542       process but does not provide a parton shower while MC@NLO incorporates  
 1543       NLO QCD matrix elements consistently into a parton shower framework; thus,  
 1544       MadGraph5\_aMC@NLO, as a merger of the two event generators MadGraph5  
 1545       and aMC@NLO, is an event generator capable to calculate tree-level and NLO  
 1546       cross sections and perform the matching of those with the parton shower. It  
 1547       is one of the most frequently used matrix element generators; however, it has  
 1548       the particular feature of the presence of negative event weights. These negative  
 1549       weights comes from interferences between Feynman diagrams and affect the  
 1550       calculation of the cross sections due to the presence of contributions of negative  
 1551       sign. Negative weights also affect the reproduction of the properties of the  
 1552       objects generated since the effective number of events used in the calculation is  
 1553       reduced, but these terms are needed in order to obtain the exact NLO results  
 1554       for total rates. [96]
- 1555     • **POWHEG.** It is an NLO matrix element generator where the hardest emis-  
 1556       sion of color charged particles is generated in such a way that the negative event  
 1557       weights issue of MadGraph5\_aMC@NLO is overcome; however, the method re-  
 1558       quires an interface with  $p_T$ -ordered parton shower<sup>3</sup> or a parton shower generator  
 1559       where this highest emission can be vetoed in order to avoid double counting of  
 1560       this highest-energetic emission. PYTHIA is a commonly matched to POWHEG  
 1561       event generator [98].
- 1562       Events resulting from the whole generation process are known as MC events.

---

<sup>3</sup> emissions in the showering are ordered in  $p_t^2$  [97]

1563 **4.3 CMS detector simulation**

1564 After generation, MC events contain the physics of the collisions but they are not  
 1565 ready to be compared to the events recorded by the experiment since these recorded  
 1566 events correspond to the response of the detection systems to the interaction with  
 1567 the particles traversing them. The simulation of the CMS detector has to be applied  
 1568 on top of the event generation; it is simulated with a MC toolkit for the simulation  
 1569 of particles passing through matter called Geant4 [99] which is also able to simulate  
 1570 the electronic signals that would be measured by all detectors inside CMS.

1571 The simulation takes the generated particles contained in the MC events as input,  
 1572 makes them pass through the simulated geometry, and models physics processes that  
 1573 particles experience during their passage through matter. The full set of results from  
 1574 particle-matter interactions corresponds to the *simulated hit*, which contains informa-  
 1575 tion about the energy loss, momentum, and position. Particles of the input event are  
 1576 called *primary*, while the particles originating from GEANT4-modeled interactions  
 1577 of primary particles with matter are called *secondary*. Simulated hits are the input  
 1578 of subsequent modules that emulate the response of the detector readout system and  
 1579 triggers. The output from the emulated detection systems and triggers is known as  
 1580 digitization [99, 100].

1581 The modeling of the CMS detector corresponds to the accurate modeling of the  
 1582 interaction among particles, the detector material, and the magnetic field. This  
 1583 simulation procedure includes the following standard steps

- 1584     • Modeling of the Interaction Region.
- 1585     • Modeling of the particle passage through the hierarchy of volumes that compose  
       1586        CMS detector and of the accompanying physics processes.

1587       • Modeling of the effect of multiple interactions per beam crossing and/or the  
 1588           effect of events overlay (pile-up simulation).

1589       • Modeling of the detector’s electronics response, signal shape, noise, calibration  
 1590           constants (digitization).

1591       In addition to the full simulation, i.e., a detailed detector simulation, a faster  
 1592       simulation (FastSim) have been developed, that may be used where much larger  
 1593       statistics are required. In FastSim, detector material effects are parametrized and  
 1594       included in the hits; those hits are used as input of the same higher-level algorithms<sup>4</sup>  
 1595       used to analyze the recorded events. In this way, comparisons between fast and full  
 1596       simulations can be performed [102].

1597       After the full detector simulation, the output events can be directly compared  
 1598       to events actually recorded in the CMS detector. The collection of MC events that  
 1599       reproduces the expected physics for a given process is known as MC sample.

## 1600 4.4 Event reconstruction

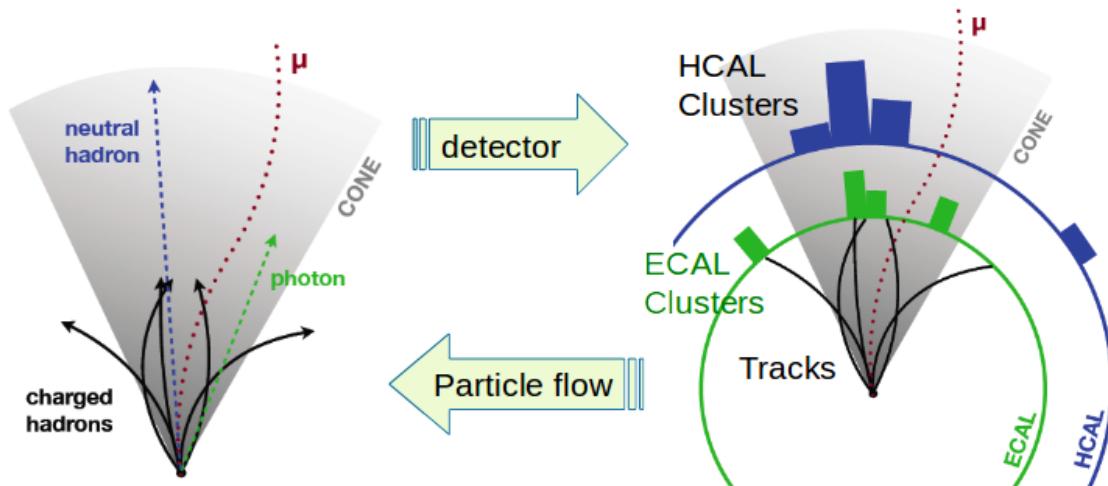
1601       The CMS experiment use the *particle-flow event reconstruction algorithm (PF)* to do  
 1602       the reconstruction of particles produced in  $pp$  collisions. Next sections will present  
 1603       a basic description of the *elements* used by PF (tracker tracks, energy clusters, and  
 1604       muon tracks), based on the References [103,104] where more detailed descriptions can  
 1605       be found.

---

<sup>4</sup> track fitting, calorimeter clustering, b tagging, electron identification, jet reconstruction and calibration, trigger algorithms which will be considered in the next sections

#### 1606 4.4.1 Particle-Flow Algorithm

1607 Each of the several subdetector systems of the CMS detector is dedicated to identify  
 1608 a specific type of particle, i.e., photons and electrons are absorbed by the ECAL  
 1609 and their reconstruction is based on ECAL information; hadrons are reconstructed  
 1610 from clusters in the HCAL while muons are reconstructed from hits in the muon  
 1611 chambers. PF is designed to correlate signals from all the detector layers (tracks and  
 1612 energy clusters) in order to reconstruct and identify each final state particle and its  
 1613 properties as sketched in Figure 4.2.



**Figure 4.2:** Particle flow algorithm. Information from the several CMS detection systems if provided as input to the algorithm which then combine it to identify and reconstruct all the particles in the final state and their properties. Reconstruction of simulated events is also performed by providing information from MC samples, detector and trigger simulation [105].

1614 For instance, a charged hadron is identified by a geometrical connection, known  
 1615 as *link*, between one or more calorimeter clusters and a track in the tracker, provided  
 1616 there are no hits in the muon system; combining several measurements allows a better  
 1617 determination of the energy and charge sign of the charged hadron.

1618 **Charged-particle track reconstruction**

1619 The strategy used by PF in order to reconstruct tracks is called *Iterative Tracking*  
 1620 which occurs in four steps:

1621 • Seed generation where initial track candidates are found by looking for a combi-  
 1622 nation of hits in the pixel detector, strip tracker, and muon chambers. In total  
 1623 ten iterations are performed, each one with a different seeding requirement.  
 1624 Seeds are used to estimate the trajectory parameters and uncertainties at the  
 1625 time of the full track reconstruction. Seeds are also considered track candidates.

1626 • Track finding using a tracking software known as Combinatorial Track Finder  
 1627 (CTF) [106]. The seed trajectories are extrapolated along the expected flight  
 1628 path of a charged particle, in agreement to the trajectory parameters obtained  
 1629 in the first step, in an attempt to find additional hits that can be assigned to  
 1630 the track candidates.

1631 • Track fitting where the found tracks are passed as input to a module which  
 1632 provides the best estimate of the parameters of each trajectory.

1633 • Track selection where track candidates are submitted to a selection which dis-  
 1634 cards those that fail a set of defined quality criteria.

1635 Iterations differ in the seeding configuration and the final track selection as elab-  
 1636 orated in References [103, 104]. In the first iteration, high- $p_T$  tracks and tracks pro-  
 1637 duced near to the interaction region are identified and those hits are masked thereby  
 1638 reducing the combinatorial complexity. Next, iterations search for more complicated  
 1639 tracks, like low- $p_T$  tracks and tracks from b hadron decays, which tend to be displaced  
 1640 from the interaction region.

1641 **Vertex reconstruction**

1642 During the track reconstruction, an extrapolation toward to the calorimeters is per-  
 1643 formed in order to match energy deposits; that extrapolation is also performed toward  
 1644 the beamline in order to find the origin of the track known as *vertex*. The vertex re-  
 1645 construction is performed by selecting from the available reconstructed tracks those  
 1646 that are consistent with being originated in the interaction region where  $pp$  collisions  
 1647 are produced. The selection involves a requirement on the number of tracker (pixel  
 1648 and strip) hits and the goodness of the track fit.

1649 Selected tracks are clustered using a *deterministic annealing algorithm (DA)*<sup>5</sup>. A  
 1650 set of candidate vertices and their associated tracks, resulting from the DA, are then  
 1651 fitted with an *adaptive vertex fitter (AVF)* to produce the best estimate of the vertices  
 1652 locations.

1653 The  $p_T$  of the tracks associated to a reconstructed vertex is added, squared and  
 1654 used to organize the vertices; the vertex with the highest squared sum is designated  
 1655 as the *primary vertex (PV)* while the rest are designated as PU vertices.

1656 **Calorimeter clustering**

1657 After traversing the CMS tracker system, electrons, photons and hadrons deposit their  
 1658 energy in the ECAL and HCAL cells. The PF clustering algorithm aims to provide  
 1659 a high detection efficiency even for low-energy particles and an efficient distinction  
 1660 between close energy deposits. The clustering runs independently in the ECAL barrel  
 1661 and endcaps, HCAL barrel and endcaps, and the two preshower layers, following two  
 1662 steps:

- 1663 • cells with an energy larger than a given seed threshold and larger than the energy

---

<sup>5</sup> DA algorithm and AVF are described in detail in References [108, 109]

1664 of the neighboring cells are identified as cluster seeds. The neighbor cells are  
 1665 those that either share a side with the cluster seed candidate, or the eight closest  
 1666 cells including cells that only share a corner with the seed candidate.

- 1667 • cells with at least a corner in common with a cell already in the cluster seed  
 1668 and with an energy above a cell threshold are grouped into topological clusters.

1669 Clusters formed in this way are known as *particle-flow clusters*. With this cluster-  
 1670 ing strategy, it is possible to detect and measure the energy and direction of photons  
 1671 and neutral hadrons as well as differentiate these neutral particles from the charged  
 1672 hadron energy deposits. In cases involving charged hadrons for which the track pa-  
 1673 rameters are not determined accurately, for instance, low-quality and high- $p_T$  tracks,  
 1674 clustering helps in the energy measurements.

## 1675 Electron track reconstruction

1676 Although the charged-particle track reconstruction described above works for elec-  
 1677 trons, they lose a significant fraction of their energy via bremsstrahlung photon radi-  
 1678 ation before reaching the ECAL; thus, the reconstruction performance depends on the  
 1679 ability to measure also the radiated energy. The reconstruction strategy, in this case,  
 1680 requires information from the tracking system and from the ECAL. Bremsstrahlung  
 1681 photons are emitted at similar  $\eta$  values to that of the electron but at different values  
 1682 of  $\phi$ ; therefore, the radiated energy can be recovered by grouping ECAL clusters in a  
 1683  $\eta$  window over a range of  $\phi$  around the electron direction. The group is called ECAL  
 1684 supercluster.

1685 Electron candidates from the track-seeding and ECAL super clustering are merged  
 1686 into a single collection which is submitted to a full electron tracking fit with a

1687 Gaussian-sum filter (GSF) [107]. The electron track and its associated ECAL su-  
 1688 percluster form a *particle-flow electron*.

1689 **Muon track reconstruction**

1690 Given that the CMS detector is equipped with a muon spectrometer capable of iden-  
 1691 tifying and measuring the momentum of the muons traversing it, the muon recon-  
 1692 struction is not specific to PF; therefore, three different muon types are defined:

- 1693     ● *Standalone muon.* A clustering of the DTs or CSCs hits is performed to form  
           1694       track segments; those segments are used as seeds for the reconstruction in the  
           1695       muon spectrometer. All DTs, CSCs, and RPCs hits along the muon trajectory  
           1696       are combined and fitted to form the full track. The fitting output is called a  
           1697       *standalone-muon track*.
- 1698     ● *Tracker muon.* Each track in the inner tracker with  $p_T$  larger than 0.5 GeV and  
           1699       a total momentum  $p$  larger than 2.5 GeV is extrapolated to the muon system.  
           1700       A *tracker muon track* corresponds to a extrapolated track that matches at least  
           1701       one muon segment.
- 1702     ● *Global muon.* When tracks in the inner tracker (inner tracks) and standalone-  
           1703       muon tracks are matched and turn out to be compatible, their hits are combined  
           1704       and fitted to form a *global-muon track*.

1705       Global muons sharing the same inner track with tracker muons are merged into  
 1706       a single candidate. PF muon identification uses the muon energy deposits in ECAL,  
 1707       HCAL, and HO associated with the muon track to improve the muon identification.

1708 **Particle identification and reconstruction**

1709 PF elements are connected by a linker algorithm that tests the connection between  
 1710 any pair of elements; if they are found to be linked, a geometrical distance that  
 1711 quantifies the quality of the link is assigned. Two elements may be linked indirectly  
 1712 through common elements. Linked elements form *PF blocks* and each PF block may  
 1713 contain elements originating from one or more particles. Links can be established  
 1714 between tracks, between calorimeter clusters, and between tracks and calorimeter  
 1715 clusters. The identification and reconstruction start with a PF block and proceed as  
 1716 follows:

1717 • Muons. An *isolated global muon* is identified by evaluating the presence of  
 1718 inner track and energy deposits close to the global muon track in the  $(\eta, \phi)$   
 1719 plane, i.e., at a particular point of the global muon track, inner tracks and  
 1720 energy deposits are sought within a radius of  $\Delta R = 0.3$  (see eqn. 3.7) from the  
 1721 muon track; if they exit and the  $p_T$  of the found track added to the  $E_T$  of the  
 1722 found energy deposit does not exceed 10% of the muon  $p_T$  then the global muon  
 1723 is an isolated global muon. This isolation condition is stringent enough to reject  
 1724 hadrons misidentified as muons.

1725 *Non-isolated global muons* are identified using additional selection requirements  
 1726 on the number of track segments in the muon system and energy deposits along  
 1727 the muon track. Muons inside jets are identified with more stringent criteria  
 1728 in isolation and momentum as described in Reference [110]. The PF elements  
 1729 associated with an identified muon are masked from the PF block.

1730 • Electrons are identified and reconstructed as described above plus some addi-  
 1731 tional requirements on fourteen variables like the amount of energy radiated,

1732 the distance between the extrapolated track position at the ECAL and the po-  
 1733 sition of the associated ECAL supercluster, among others, which are combined  
 1734 in an specialized multivariate analysis strategy that improves the electron iden-  
 1735 tification. Tracks and clusters used to identify and reconstruct electrons are  
 1736 masked in the PF block.

1737 • Isolated photons are identified from ECAL superclusters with  $E_T$  larger than 10  
 1738 GeV, for which the energy deposited at a distance of 0.15, from the supercluster  
 1739 position on the  $(\eta, \phi)$  plane, does not exceed 10% of the supercluster energy;  
 1740 note that this is an isolation requirement. In addition, there must not be links  
 1741 to tracks. Clusters involved in the identification and reconstruction are masked  
 1742 in the PF block.

1743 • Bremsstrahlung photons and prompt photons tend to convert to electron-positron  
 1744 pairs inside the tracker, therefore, a dedicated finder algorithm is used to link  
 1745 tracks that seem to originate from a photon conversion; in case those two tracks  
 1746 are compatible with the direction of a bremsstrahlung photon, they are also  
 1747 linked to the original electron track. Photon conversion tracks are also masked  
 1748 in the PF block.

1749 • The remaining elements in the PF block are used to identify hadrons. In the  
 1750 region  $|\eta| \leq 2.5$ , neutral hadrons are identified with HCAL clusters not linked  
 1751 to any track while photons from neutral pion decays are identified with ECAL  
 1752 clusters without links to tracks. In the region  $|\eta| > 2.5$  ECAL clusters linked to  
 1753 HCAL clusters are identified with a charged or neutral hadron shower; ECAL  
 1754 clusters with no links are identified with photons. HCAL clusters not used yet,  
 1755 are linked to one or more unlinked tracks and to an unlinked ECAL in order to

1756 reconstruct charged-hadrons or a combination of photons and neutral hadrons  
 1757 according to certain conditions on the calibrated calorimetric energy.

- 1758 • Charged-particle tracks may be linked together when they converge to a *sec-*  
 1759 *ondary vertex (SV)* displaced from the IP where the PV and PU vertices are  
 1760 reconstructed; at least three tracks are needed in that case, of which at most  
 1761 one has to be a track with hits in tracker region between a PV and the SV.

1762 The linker algorithm, as well as the whole PF algorithm, has been validated and  
 1763 commissioned; results from that validation are presented in Reference [103].

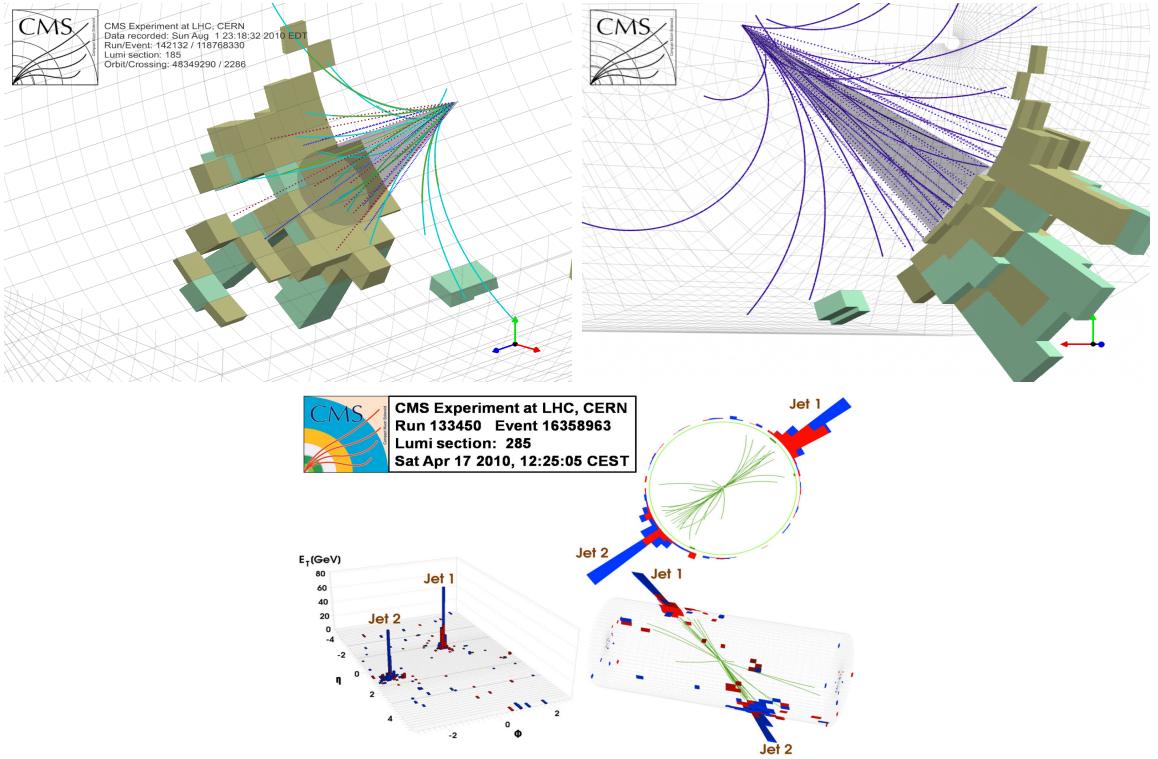
#### 1764 Jet reconstruction

1765 Quarks and gluons may be produced in the  $pp$  collisions, therefore, their hadronization  
 1766 will be seen in the detector as a shower of hadrons and their decay products in the  
 1767 form of a *jet*. The anti- $k_t$  algorithm [111] is used to perform the jet reconstruction  
 1768 by clustering those PF particles within a cone (see Figure 4.3); previously, isolated  
 1769 electrons, isolated muons, and charged particles associated with other interaction  
 1770 vertices were excluded from the clustering.

1771 The anti- $k_t$  algorithm proceeds in a sequential recombination of PF particles; the  
 1772 distance between particles  $i$  and  $j$  ( $d_{ij}$ ) and the distance between particles and the  
 1773 beam are defined as

$$d_{ij} = \min \left( \frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2} \right) \frac{\Delta_{ij}^2}{R^2}$$

$$d_{iB} = \frac{1}{k_{ti}^2} \quad (4.1)$$

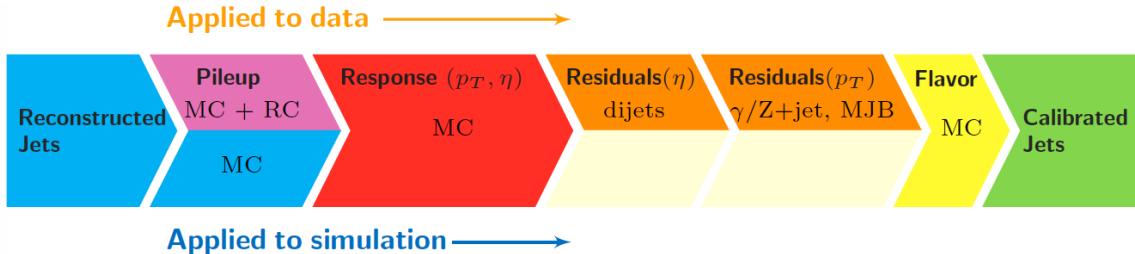


**Figure 4.3:** Jet reconstruction performed by the anti- $k_t$  algorithm. Top: Two different views of a CMS recorded event are presented. Continuous lines correspond to tracks left by charged particles in the tracker while dotted lines are the imaginary paths followed by neutral particles. The green cubes represent the ECAL cells while the blue ones represent the HCAL cells; in both cases, the height of the cube represent the amount of energy deposited in the cells [112]. Bottom: Reconstruction of a recorded event with two jets [113].

where  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ ,  $k_{ti}$ ,  $y_i$  and  $\phi_i$  are the transverse momentum, rapidity and azimuth of particle  $i$  respectively and  $R$  is the called jet radius. For all the remaining PF particles, after removing the isolated ones,  $d_{ij}$  and  $d_{iB}$  are calculated<sup>6</sup> and the smallest is identified; if it is a  $d_{ij}$ , particles  $i$  and  $j$  are replaced with a new object whose momentum is the vectorial sum of the combined particles. If the smallest distance is a  $d_{iB}$  the clustering process ends, the object  $i$  (which at this stage should be a combination of several PF particles) is declared as a *Particle-flow-jet* (PF jet) and all the associated PF particles are removed from the detector. The clustering

<sup>6</sup> Notice that this is a combinatorial calculation.

1782 process is repeated until no PF particles remain.



**Figure 4.4:** Jet energy correction diagram. Correction levels are applied sequentially in the indicated fixed order [115].

Even though jets can be reconstructed efficiently, there are some effects that are not included in the reconstruction that lead to discrepancies between the reconstructed results and the predicted results; in order to overcome these discrepancies, a factorized model has been designed in the form of jet energy corrections (JEC) [114, 115] applied sequentially as shown in the diagram of Figure 4.4.

1788 At each level, the jet four-momentum is multiplied by a scale factor based on jet  
1789 properties, i.e.,  $\eta$ , flavor, etc.

- Level 1 correction removes the energy coming from pile-up. The scale factor is determined using a MC sample of QCD dijet (2 jets) events with and without pileup overlay; it is parametrized in terms of the offset energy density  $\rho$ , jet area  $A$ , jet  $\eta$  and jet  $p_T$ . Different corrections are applied to data and MC due to the detector simulation.
  - MC-truth correction accounts for differences between the reconstructed jet energy and the MC particle-level energy. The correction is determined on a QCD dijet MC sample and is parametrized in terms of the jet  $p_T$  and  $\eta$ .
  - Residuals correct remaining small differences within jet response in data and MC. The Residuals  $\eta$ -dependent correction compares jets of similar  $p_T$  in the

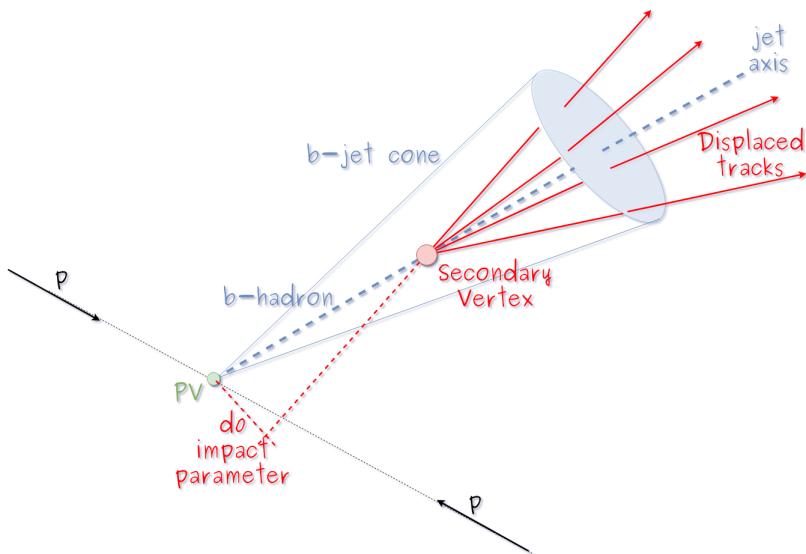
1800 barrel reference region. The Residuals  $p_T$ -dependent corrects the jet absolute  
 1801 scale (JES vs  $p_T$ ).

- 1802 • Jet-flavor corrections are derived in the same way as MC-truth corrections but  
 1803 using QCD pure flavor samples.

1804 ***b*-tagging of jets**

1805 A particular feature of the hadrons containing bottom quarks (*b*-hadrons) is that  
 1806 their lifetime is long enough to travel some distance before decaying, but it is not as  
 1807 long as those of light quark hadrons; therefore, when looking at the hadrons produced  
 1808 in  $pp$  collisions, *b*-hadrons decay typically inside the tracker rather than reaching the  
 1809 calorimeters as some light-hadrons do. As a result, a *b*-hadron decay gives rise to a  
 1810 displaced vertex (secondary vertex) with respect to the primary vertex as shown in  
 1811 Figure 4.5; the SV displacement is in the order of a few millimeters. A jet resulting  
 1812 from the decay of a *b*-hadron is called *b* jet; other jets are called light jets.

1813



**Figure 4.5:** Secondary vertex in a *b*-hadron decay.

1814 Several methods to identify  $b$ -jets ( $b$ -tagging) have been developed; the method  
 1815 used in this thesis is known as *Combined Secondary Vertex* algorithm in its second  
 1816 version (CSVv2) [116]. By using information of the impact parameter, the recon-  
 1817 structed secondary vertices, and the jet kinematics as input in a multivariate analysis  
 1818 that combines the discrimination power of each variable in one global discrimina-  
 1819 tor variable, three working points (references): loose, medium and tight, are defined  
 1820 which quantify the probabilities of mistag jets from light quarks as jets from  $b$  quarks;  
 1821 10, 1 and 0.1 % respectively. Although the mistagging probability decreases with the  
 1822 working point strength, the efficiency to correctly tag  $b$ -jets also decreases as 83, 69  
 1823 and 49 % for the respective working point; therefore, a balance needs to be achieved  
 1824 according to the specific requirements of the analysis.

1825 **4.4.1.1 Missing transverse energy**

1826 The fact that proton bunches carry momentum along the  $z$ -axis implies that for  
 1827 each event it is expected that the momentum in the transverse plane is balanced.  
 1828 Imbalances are quantified by the missing transverse energy (MET) and are attributed  
 1829 to several sources including particles escaping undetected through the beam pipe,  
 1830 neutrinos produced in weak interactions processes which do not interact with the  
 1831 detector and thus escaping without leaving a sign, or even undiscovered particles  
 1832 predicted by models beyond the SM.

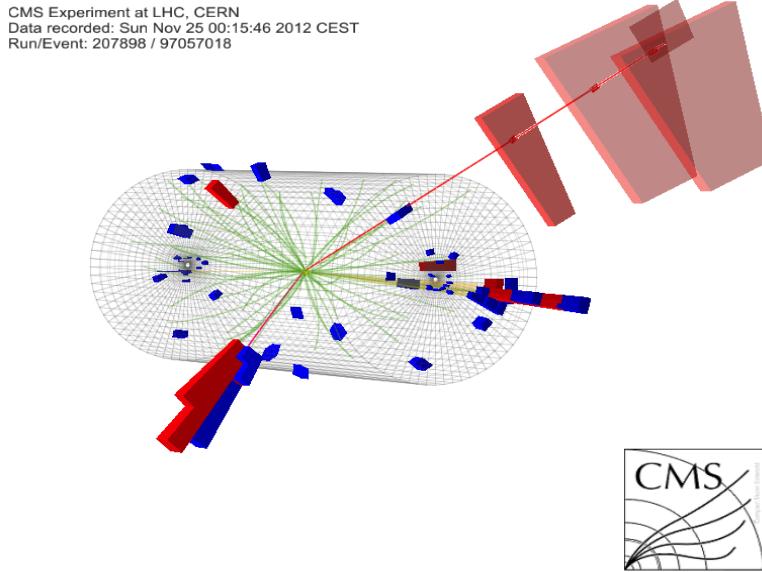
1833 The PF algorithm assigns the negative sum of the momenta of all reconstructed  
 1834 PF particles to the *particle-flow MET* according to

$$\vec{E}_T = - \sum_i \vec{p}_{T,i} \quad (4.2)$$

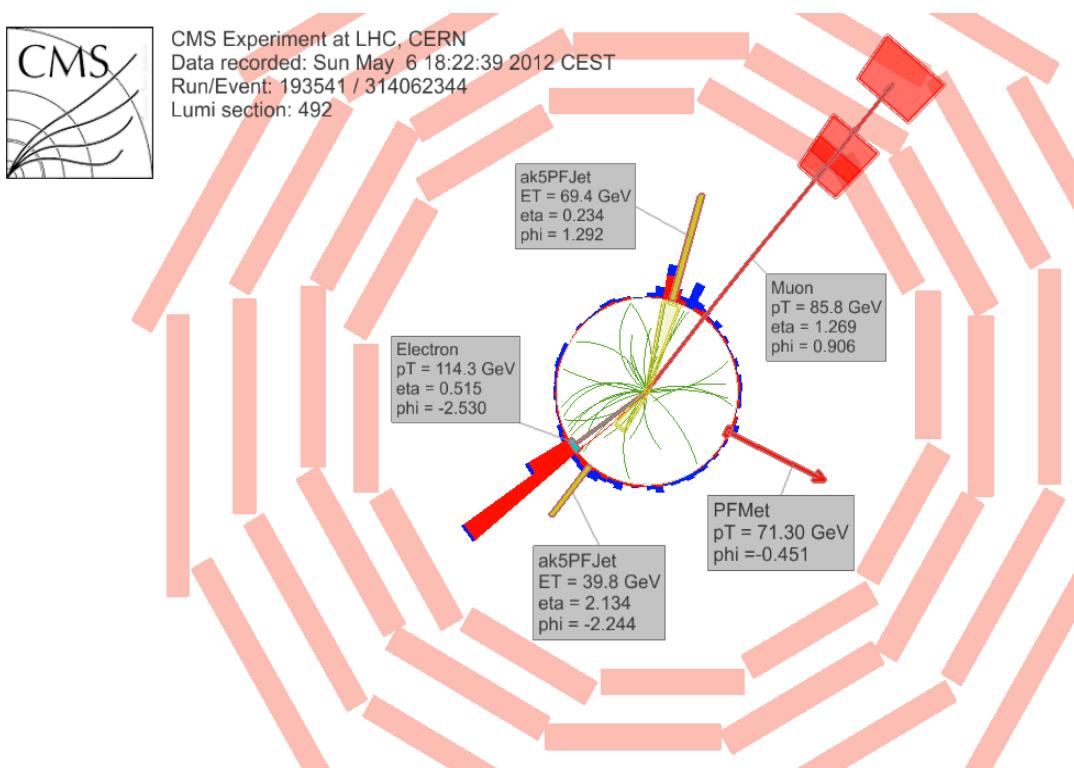
1835 JEC are propagated to the calculation of the  $\vec{E}_T$  as described in Reference [117].

1836 **4.4.2 Event reconstruction examples**

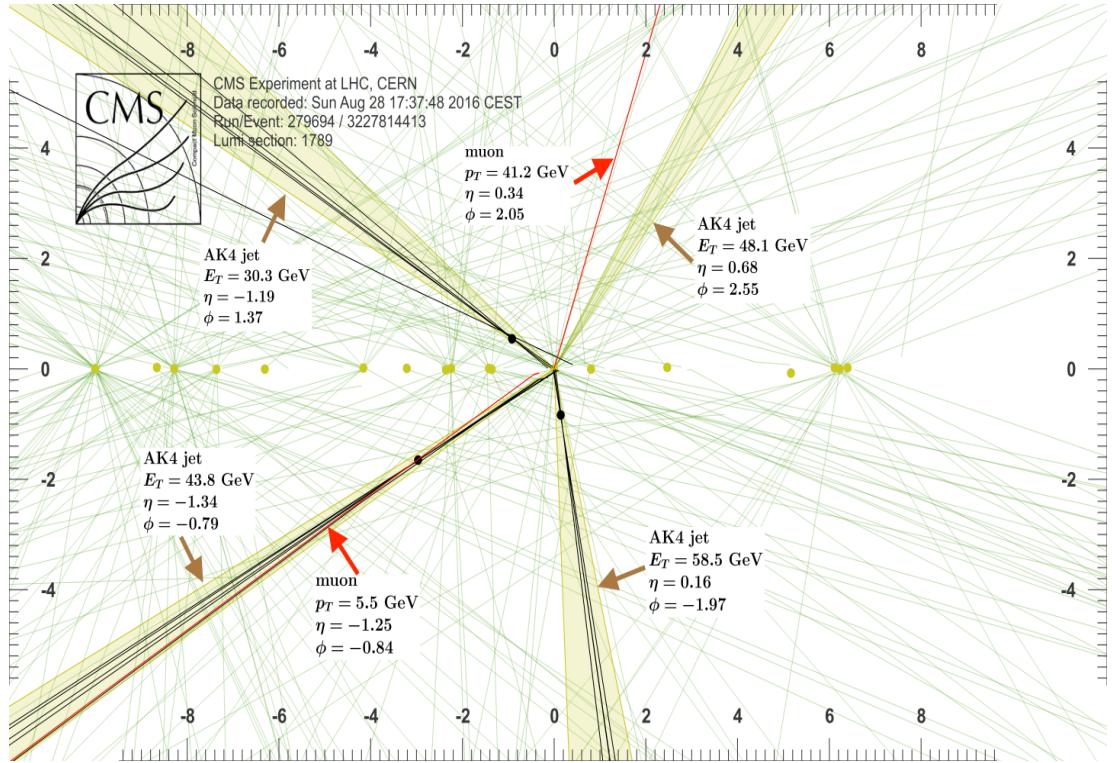
1837 Figures 4.6-4.8 show the results of the reconstruction performed on three recorded  
1838 events. Descriptions are taken directly from the source.



**Figure 4.6:** HIG-13-004 Event 1 reconstruction results; “HIG-13-004 Event 1: Event recorded with the CMS detector in 2012 at a proton-proton center-of-mass energy of 8 TeV. The event shows characteristics expected from the decay of the SM Higgs boson to a pair of  $\tau$  leptons. Such an event is characterized by the production of two forward-going jets, seen here in opposite endcaps. One of the  $\tau$  decays to a muon (red lines on the right) and neutrinos, while the other  $\tau$  decays into a charged hadron and a neutrino.” [118].



**Figure 4.7:**  $e\mu$  event reconstruction results; “An  $e\mu$  event candidate selected in 8 TeV data, as seen from the direction of the proton beams. The kinematics of the main objects used in the event selection are highlighted: two isolated leptons and two particle-flow jets. The reconstructed missing transverse energy is also displayed for reference” [119].



**Figure 4.8:** Recorded event reconstruction results; “Recorded event ( $\rho$ - $z$  projection) with three jets with  $p_T > 30$  GeV with one displaced muon track in 2016 data collected at 13 TeV. Each of the three jets has a displaced reconstructed vertex. The jet with  $p_T(j) = 43.8$  GeV,  $\eta(j) = -1.34$ ,  $\phi(j) = -0.79$  contains muon with  $p_T(\mu) = 5.5$  GeV,  $\eta(\mu) = -1.25$ ,  $\phi(\mu) = -0.84$ . Event contains reconstructed isolated muon with  $p_T(\mu) = 41.2$  GeV,  $\eta(\mu) = 0.34$ ,  $\phi(\mu) = 2.05$  and MET with  $p_T = 72.5$  GeV,  $\phi = -0.32$ . Jet candidates for a  $b$ -jet from top quark leptonic and hadronic decays are tagged by CSVv2T algorithm. One of the other two jets is tagged by CharmT algorithm. Tracks with  $p_T > 0.5$  GeV are shown. The number of reconstructed primary vertices is 18. Reconstructed  $m_T(W)$  is 101.8 GeV. Beam spot position correction is applied. Reconstructed primary vertices are shown in yellow color, while reconstructed displaced vertices and associated tracks are presented in black color. Dimensions are given in cm” [120].

1839 **Chapter 5**

1840 **Statistical methods**

1841 In the course of analyzing the data sets provided by the CMS experiment and used in  
1842 this thesis, several statistical tools have been employed; in this chapter, a description  
1843 of these tools will be presented, starting with the general statement of the multivariate  
1844 analysis method, followed by the particularities of the Boosted Decision Trees (BDT)  
1845 method and its application to the classification problem. Statistical inference methods  
1846 used will also be presented. This chapter is based mainly on the references [121–123].

1847 **5.1 Multivariate analysis**

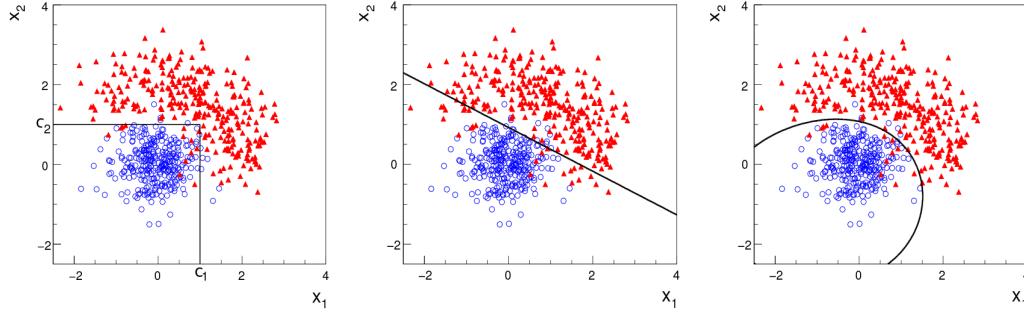
1848 Multivariate data analysis (MVA) makes reference to statistical techniques that an-  
1849 alyze data containing information of more than one variable, commonly taking into  
1850 account the effects of all variables on the response of the particular variable under  
1851 investigation, i.e., considering all the correlations between variables. MVA is em-  
1852 ployed in a variety of fields like consumer and market research, quality control and  
1853 process optimization. From a MVA it is possible to identify the dominant patterns  
1854 in the data, like groups, outliers and trends, and determine to which group a set of

1855 values belong; in the particle physics context, MVA methods are used to perform the  
 1856 selection of certain type of events, from a large data set, using a potentially large  
 1857 number of measurable properties for each event.

1858 Processes with small cross section, as the  $tHq$  process, normally are hidden behind  
 1859 more common processes; therefore, the data set results in a subset of events with  
 1860 characteristic features of interest (signal) mixed in randomly with a much larger  
 1861 number of SM events that can mimic these features of interest (background) which  
 1862 implies that it is not possible to say with certainty that a given event is signal or  
 1863 background. In that sense, the problem can be formulated as one where a set of  
 1864 events have to be classified according to some features; these features correspond to  
 1865 the measurements of several parameters like energy or momentum, organized in a  
 1866 set of *input variables*. The measurements for each event can be written in a vector  
 1867  $\mathbf{x} = (x_1, \dots, x_n)$  for which

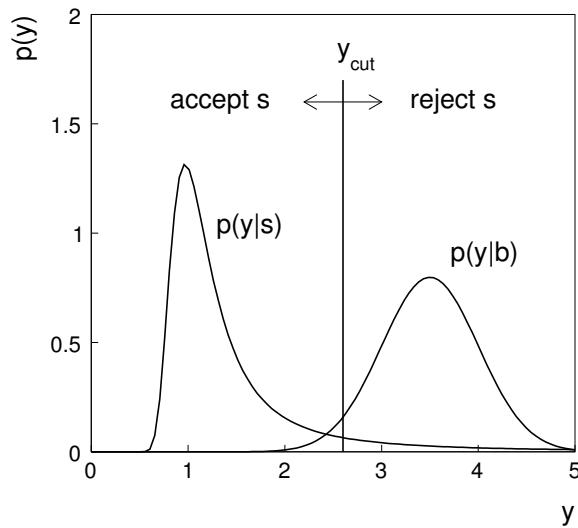
- 1868     • Signal hypotheses  $\rightarrow f(\mathbf{x}|s)$  is the probability density (*likelihood function*) that  
   1869        $\mathbf{x}$  is the set of measured values given that the events is a signal event.
- 1870     • Background hypotheses  $\rightarrow f(\mathbf{x}|b)$  is the probability density (*likelihood function*)  
   1871       that  $\mathbf{x}$  is the set of measured values given that the event is a background event.

1872     Figure 5.1 shows three ways to perform a classification of events for which mea-  
 1873       surements of two properties, two input variables, have been performed; blue circles  
 1874       represent signal events while red triangles represent background events. The classi-  
 1875       fication on (a) is *cut-based* requiring  $x_1 < c_1$  and  $x_2 < c_2$ ; usually the cut values are  
 1876       chosen according to some knowledge about the event process. In (b), the classification  
 1877       is performed by stating a cut involving a linear function of the input variables and  
 1878       so the boundary, while in (c) the the relationship between the input variables is not  
 1879       linear thus the boundary is not linear either.



**Figure 5.1:** Scatter plots-MVA event classification. Distribution of two input variables  $x_1$  and  $x_2$  measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on (a) cuts, (b) linear boundary, and (c) nonlinear boundary [121]

1880        The boundary can be parametrized in terms of the input variables such that the  
 1881        cut is set on the parametrization instead of on the variables, i.e.,  $y(\mathbf{x}) = y_{cut}$  with  
 1882         $y_{cut}$  a constant; thus, the acceptance or rejection of an event is based on what side  
 1883        of the boundary is the event located. If  $y(\mathbf{x})$  has functional form, it can be used to  
 1884        determine the probability distribution functions  $p(y|s)$  and  $p(y|b)$  and then perform  
 1885        a scalar test statistic with a single cut on the scalar variable  $y$ .



**Figure 5.2:** Distributions of the scalar test statistic  $y(\mathbf{x})$  under the signal and background hypotheses. [121]

1886       Figure 5.2 illustrates what would be the probability distribution functions under  
 1887   the signal and background hypotheses for a scalar test statistic with a cut on the  
 1888   classifier  $y$ . Notice that the tails of the distributions indicate that some signal events  
 1889   fall on the rejection region and some background events fall on the acceptance region;  
 1890   therefore, it is convenient to define the *efficiency* with which events of a given type  
 1891   are accepted, thus, the signal and background efficiencies are given by

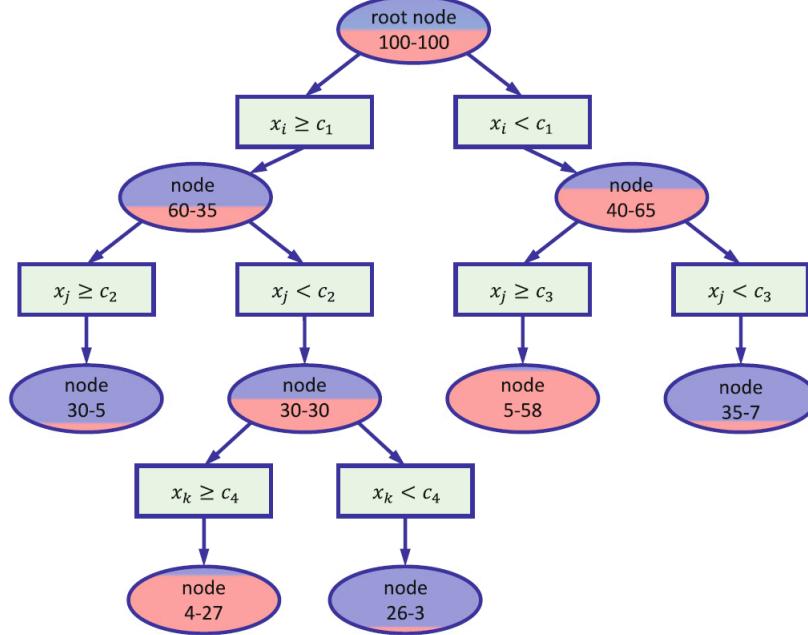
$$\varepsilon_s = P(\text{accept event}|s) = \int_A f(\mathbf{x}|s) d\mathbf{x} = \int_{-\infty}^{y_{\text{cut}}} p(y|s) dy , \quad (5.1)$$

$$\varepsilon_b = P(\text{accept event}|b) = \int_A f(\mathbf{x}|b) d\mathbf{x} = \int_{-\infty}^{y_{\text{cut}}} p(y|b) dy , \quad (5.2)$$

1892   where  $A$  is the acceptance region. Under these conditions, the background hypothesis  
 1893   corresponds to the *null hypothesis* ( $H_0$ ), the signal hypothesis corresponds to the  
 1894   *alternative hypothesis* ( $H_1$ ), the background efficiency is the significance level of the  
 1895   test, and signal efficiency is the power of the test; what is sought in an analysis is to  
 1896   maximize the power of the test relative to the significance level.

### 1897 5.1.1 Decision trees

1898   For this thesis, the implementation of the MVA strategy, described above, is per-  
 1899   formed through decision trees by using the TMVA software package [122] included in  
 1900   the the ROOT analysis framework [124]. In a simple picture, a decision tree classifies  
 1901   events according to their input variables values by setting a cut on each input variable  
 1902   and checking which events are on which side of the cut, just as proposed in the MVA  
 1903   strategy, but in addition, as a machine learning algorithm, decision trees offer the  
 1904   possibility to be trained and then perform the classification efficiently.



**Figure 5.3:** Example of a decision tree. Each node is fed with a MC sample mixing signal and background events (left-right numbers); nodes colors represent the relative amount of signal/background events [123].

1905        The training or growing of a decision tree is the process that defines the rules for  
 1906        classifying events; this process is represented in figure5.3 and consist of several steps

1907        • take MC samples of signal and background events and split them into two parts  
 1908        each; first parts form the training sample which will be used in the decision tree  
 1909        training, while the second parts form the test sample which will be used for  
 1910        testing the final classifier obtained from the training. Each event has associated  
 1911        a set of input variables  $\mathbf{x} = (x_1, \dots, x_n)$  which serve to distinguish between signal  
 1912        and background events. The training sample is taken in at the root *node*.

1913        • pick one variable, say  $x_i$   
 1914        • pick one value of  $x_i$ , each event has its own value of  $x_i$ , and split the training  
 1915        sample into two subsamples  $B_1$  and  $B_2$ ;  $B_1$  contains events for which  $x_i < c_1$

- 1916        while  $B_2$  contains the rest of the training events;
- 1917        • scan all possible values of  $x_i$  and find the splitting value that provides the *best*  
 1918        classification<sup>1</sup>, i.e.,  $B_1$  is mostly made of signal events while  $B_2$  is mostly made  
 1919        of background events.
- 1920        • It is possible that variables other than the picked one produce a better classi-  
 1921        fication, hence, all the variables have to be evaluated. Pick the next variable,  
 1922        say  $x_j$ , and repeat the scan over its possible values.
- 1923        • At the end, all the variables and their values will have been scanned, the *best*  
 1924        variable and splitting value will have been identified, say  $x_1, c_1$ , and there will  
 1925        be two nodes fed with the subsamples  $B_1$  and  $B_2$ .
- 1926        Nodes are further split by repeating the decision process until: a given number of  
 1927        final nodes is obtained, nodes are largely dominated by either signal or background  
 1928        events, or nodes has too few events to continue. Final nodes are called *leaves* and they  
 1929        are classified as signal or background leaves according to the class of the majority of  
 1930        events in them. Each *branch* in the tree corresponds to a sequence of cuts.
- 1931        The quality of the classification at each node is evaluated through a separation  
 1932        criteria; there are several of them but the *Gini Index* ( $G$ ) is the one used in the  
 1933        decision trees trained for the analysis in this thesis.  $G$  is written in terms of the  
 1934        purity ( $P$ ), i.e. the fraction of signal events, of the samples after the separation is  
 1935        made; it is given by

$$G = P(1 - P) \quad (5.3)$$

---

<sup>1</sup> Quality of the classification will be treated in the next paragraph.

1936 notice that  $P=0.5$  at the root node while  $G=0$  for pure leaves. For a node  $A$  split  
 1937 into two nodes  $B_1$  and  $B_2$  the G gain is

$$\Delta G = G(A) - G(B_1) - G(B_2) \quad (5.4)$$

1938 the *best* classification corresponds to that for which the gain of G is maximized; hence,  
 1939 the scanning over all event's variables and their values is of capital importance.

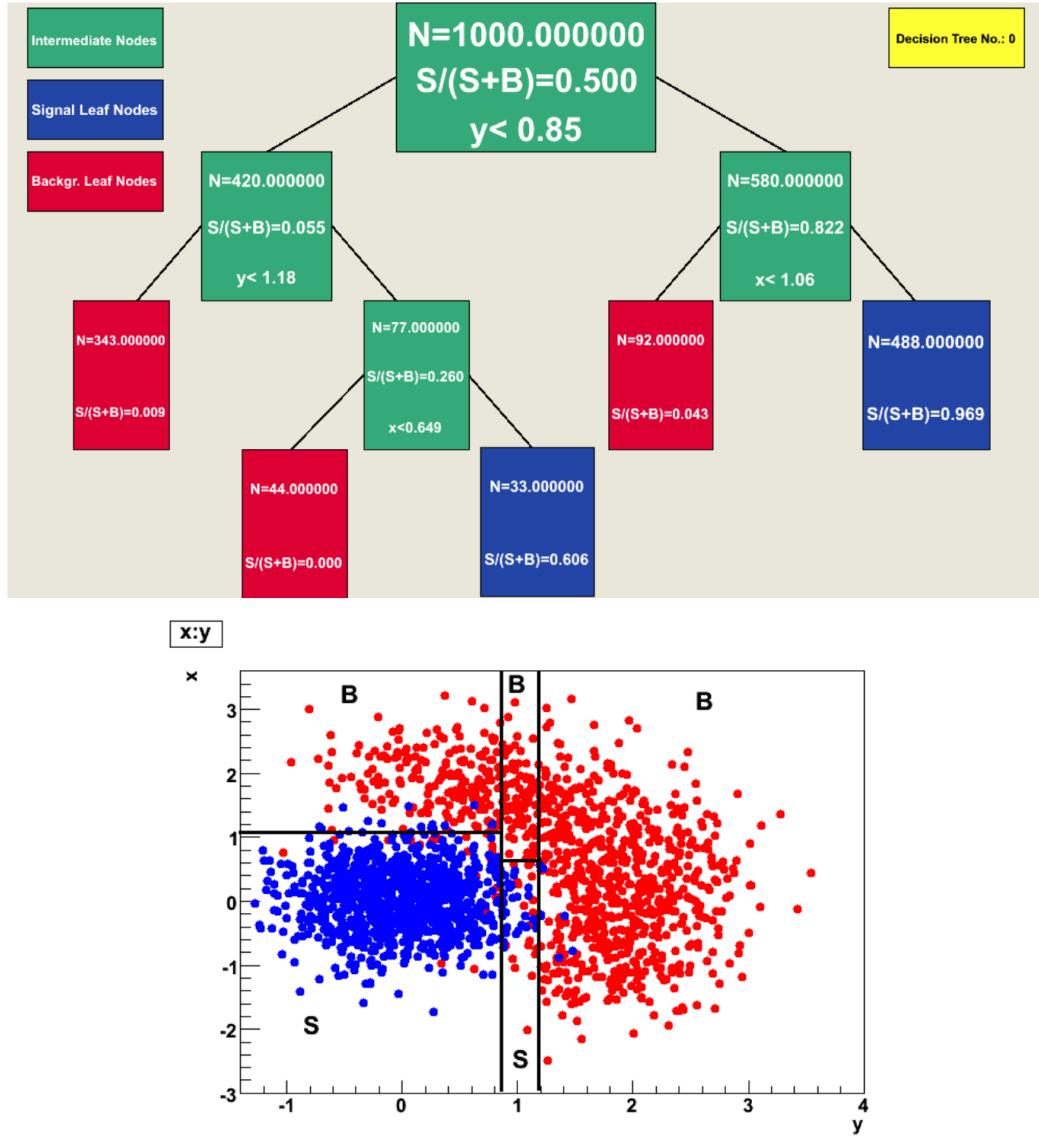
1940 In order to provide a numerical output for the classification, events in a sig-  
 1941 nal(background) leaf are assigned an score of 1(-1) each, defining in this way the  
 1942 decision tree *classifier or weak learner* as

$$f(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \text{ in signal region,} \\ -1 & \mathbf{x} \text{ in background region.} \end{cases}$$

1943 Figure 5.4 shows an example of the classification of a sample of events, containing  
 1944 two variables, performed by a decision tree.

### 1945 5.1.2 Boosted decision trees (BDT).

1946 Event misclassification occurs when a training event ends up in the wrong leaf, i.e., a  
 1947 signal event ends up in a background leaf or a background event ends up in a signal  
 1948 leaf. A way to correct it is to assign a weight to the misclassified events and train  
 1949 a second tree using the reweighted events; the event reweighting is performed by a  
 1950 boosting algorithm, events with increased weight are known as *boosted* events, in such  
 1951 a way that when used in the training of a new decision tree they get correctly classified.  
 1952 The process is repeated iteratively adding a new tree to a forest and creating a set  
 1953 of classifiers which are combined to create the next classifier; the final classifier offers



**Figure 5.4:** Example of a decision tree output. Each leaf, blue for signal events and red for background events, represent a region in the variables phase space [125].

1954 more stability<sup>2</sup> and has a smaller misclassification rate than any individual ones. The  
 1955 resulting tree collection is known as a *boosted decision tree (BDT)*.  
 1956 Thus, purity of the sample is generalized to

<sup>2</sup> Decision trees suffer from sensitivity to statistical fluctuations in the training sample which may leads to very different results with an small change in the training samples.

$$P = \frac{\sum_s w_s}{\sum_s w_s + \sum_b w_b} \quad (5.5)$$

1957 where  $w_s$  and  $w_b$  are the weights of the events; the Gini index is also generalized

$$G = \left( \sum_i^n w_i \right) P(1 - P) \quad (5.6)$$

1958 with  $n$  the number of events in the node. The final score of an event, after pass-  
 1959 ing through the forest, is calculated as the renormalized sum of all the individual  
 1960 (possibly weighted) scores; thus, high(low) score implies that the event is most likely  
 1961 signal(background).

1962 The boosting procedure, implemented in the *Gradient boosting* algorithm used  
 1963 in this thesis, produce a classifier  $F(\mathbf{x})$  which is the weighted sum of the individual  
 1964 classifiers obtained after each iteration,i.e.,

$$F(\mathbf{x}) = \sum_{m=1}^M \beta_m f(\mathbf{x}; a_m) \quad (5.7)$$

1965 where M is the number of trees in the forest. The *loss function*  $L(F, y)$  represent the  
 1966 deviation between the classifier  $F(\mathbf{x})$  response and the true value  $y$  obtained from the  
 1967 training sample (1 for signal events and -1 for background event), according to

$$L(F, y) = \ln(1 + e^{-2F(\mathbf{x})y}) \quad (5.8)$$

1968 thus, the reweighting is employed to ensure the minimization of the loss function;  
 1969 a more detailed description of the minimization procedure can be found in reference  
 1970 [126]. The final classifier output is later used as a final discrimination variable, labeled  
 1971 as *BDT output/response*.

1972 **5.1.3 Overtraining.**

1973 Decision trees offer the possibility to have as many nodes as wished in order to  
 1974 reduce the misclassification to zero (in theory); however, when a classifier is too much  
 1975 adjusted to a particular training sample, the classifier response to a slightly different  
 1976 sample may leads to a completely different classification results; this effect is known  
 1977 as *overtraining*.

1978 An alternative to reduce the overtraining in BDTs consist in pruning the tree by  
 1979 removing statistically insignificant nodes after the tree growing is completed but this  
 1980 option is not available for BDTs with gradient boosting in the TMVA-toolkit, there-  
 1981 fore, the overtraining has to be reduced by tuning the algorithm, number of nodes,  
 1982 minimum number of events in the leaves, etc. The overtraining can be evaluated  
 1983 by comparing the responses of the classifier when running over the training and test  
 1984 samples.

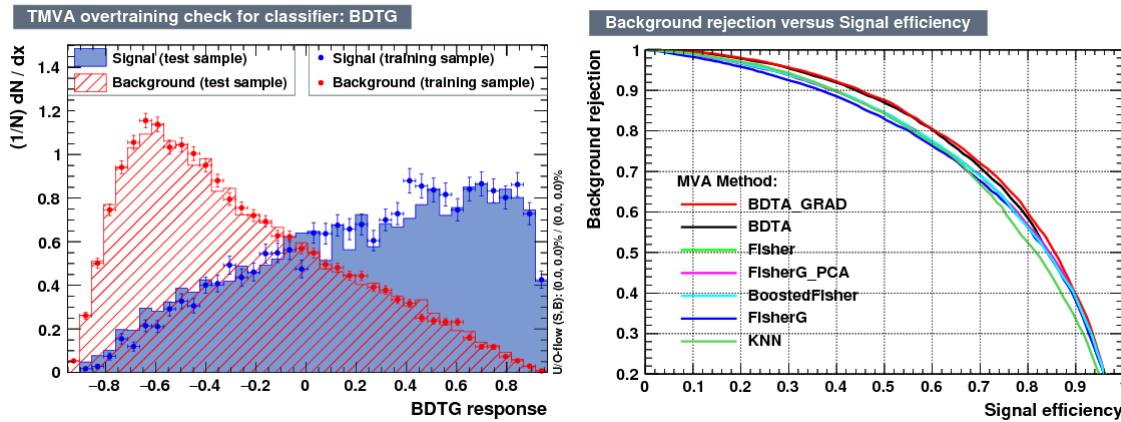
1985 **5.1.4 Variable ranking.**

1986 BDTs have the couple of particular advantages related to the input variables; on one  
 1987 side, they are relatively insensitive to the number of input variables used in the vector  
 1988  $\mathbf{x}$ . The ranking of the BDT input variables is determined by counting the number of  
 1989 times a variable is used to split decision tree nodes; in addition, the separation gain-  
 1990 squared achieved in the splitting and the number of events in the node are accounted  
 1991 by applying a weighting to that number. Thus, those variables with small or no power  
 1992 to separate signal and background events are rarely chosen to split the nodes,i.e., are  
 1993 effectively ignored.

1994 On the other side, variables correlations play an important role for some MVA  
 1995 methods like the Fisher discriminant algorithm in which the first step consist of

1996 performing a linear transformation to a phase space where the correlations between  
 1997 variables are removed; in case of BDT algorithm, correlations do not affect the per-  
 1998 formance.

1999 **5.1.5 BDT output example.**



**Figure 5.5:** Left: Output distributions for the gradient boosted decision tree (BDTG) classifier using a sample of signal( $pp \rightarrow tHq$ ) and background( $pp \rightarrow t\bar{t}$ ) events. Right: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers running over the same sample used to produce the plot on the left.

2000 Left side of figure 5.5 shows the BDT output distributions for signal( $pp \rightarrow tHq$ )  
 2001 and background( $pp \rightarrow t\bar{t}$ ) events; this plot is the equivalent to the one showed in  
 2002 figure 5.2. A forest with 800 trees, maximum depth per tree = 3, and gradient  
 2003 boosting have been used as training parameters. The BDTG classifier offers a good  
 2004 separation power; while there is a small overtraining in the signal distribution, the  
 2005 background distribution seems to be well predicted which might indicate that the  
 2006 sample is composed of more background than signal events.

2007 Right side of figure 5.5 shows the background rejection vs signal efficiency curves  
 2008 for several combinations of MVA classifiers-boosting algorithms; these curves are  
 2009 known as ROC curves and give an indication of the performance of the classifier. The

2010 best performance is achieved with the BDTG classifier (BDTA\_GRAD).

## 2011 5.2 Statistical inference.

2012 Once events are classified, the next step consists in finding the parameters that define  
 2013 the likelihood functions  $f(\mathbf{x}|s), f(\mathbf{x}|b)$  for signal and background events respectively.  
 2014 In general, likelihood functions depend not only on the measurements but also on  
 2015 parameters ( $\theta_m$ ) that define their shapes; the process of estimating these *unknown*  
 2016 *parameters* and their uncertainties from the experimental data is called *inference*.  
 2017 The likelihood function for  $N$  the events the in a sample is the combination of all the  
 2018 likelihoods functions

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N f(\mathbf{x}^i | \boldsymbol{\theta}) = \prod_{i=1}^N f(x_1^i, \dots, x_n^i; \theta_1, \dots, \theta_m) \quad (5.9)$$

2019 Thus, the estimation of the unknown parameters from experimental data samples  
 2020 is written in terms of a central value using the notation

$$\theta = \hat{\theta} \pm \delta\theta \quad (5.10)$$

2021 where the interval  $[\hat{\theta} + \delta\theta, \hat{\theta} - \delta\theta]$  is called *confidence interval*; it is usually inter-  
 2022 preted, in the limit of infinite number of experiments, as the interval where the true  
 2023 value of the unknown parameter  $\theta$  is contained with a probability of 0.6827 (if no  
 2024 other convention is stated).

### 2025 5.2.1 Nuisance parameters.

2026 The unknown parameter vector  $\boldsymbol{\theta}$  is made of two types of parameters: on one side,  
 2027 those parameters that provide information about the physical observables of interest

2028 for the experiment (*parameters of interest*); on the other side, the *nuisance parameters*  
 2029 that are not of direct interest for the experiment but that needs to be included in  
 2030 the analysis in order to achieve a satisfactory description of the data. They represent  
 2031 effects of the detector response like the finite resolutions of the detection systems,  
 2032 miscalibrations, and in general any source of uncertainty introduced in the analysis.

2033 In some cases the nuisance parameters are estimated using dedicated data samples,  
 2034 for instance data from test beams for calibration purposes, when MC samples are  
 2035 not suitable. The nuisance parameter uncertainties produce *systematic uncertainties*  
 2036 while the uncertainties associated to fluctuations in data and related to the estimation  
 2037 of the parameters of interest produce *statistical uncertainties*.

### 2038 5.2.2 Maximum likelihood estimation method

2039 The function that produce the estimate of a parameter is called *estimator*, there-  
 2040 fore, estimators are usually constructed using mathematical procedures encoded in  
 2041 algorithms. The estimation method used in this thesis is the *Maximum Likelihood*  
 2042 *Estimation* method (MLE); it is based on the combined likelihood function defined  
 2043 by eqn. 5.9 and the procedure seeks for the parameter set that corresponds to the  
 2044 maximum value of the combined likelihood function, i.e., the *maximum likelihood*  
 2045 *estimator* of the unknown parameter vector  $\boldsymbol{\theta}$  is the function that produce the vec-  
 2046 tor  $\hat{\boldsymbol{\theta}}$  for which the likelihood function  $L(\boldsymbol{\theta})$  evaluated at the measured sample  $\mathbf{x}$  is  
 2047 maximum.

2048 Usually, the logarithm of the likelihood function is used in the numerical algo-  
 2049 rithms implementations in order to avoid underflow the numerical precision of the  
 2050 computers due to the product of low likelihoods. In addition, it is usual minimize the  
 2051 negative logarithm of the likelihood function instead of maximizing the logarithm of

2052 it because in this way the procedure consist of differentiate a sum of therms and set  
 2053 the sum to zero; therefore

$$F(\boldsymbol{\theta}) = -\ln L(\boldsymbol{\theta}) = -\sum_{i=1}^N f(\mathbf{x}^i | \boldsymbol{\theta}). \quad (5.11)$$

2054 The minimization process is performed by the software MINUIT [127] imple-  
 2055 mented in the ROOT analysis framework. In case of large data samples the compu-  
 2056 tational resources needed to calculate the likelihood function are too big; therefore,  
 2057 the parameter estimation is performed using binned distributions of the variables of  
 2058 interest for which the *binned likelihood function* is given by

$$L(data|\mu, \theta) = \prod_{i=1} \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-\mu s_i(\theta) - b_i(\theta)}, \quad (5.12)$$

2059 with  $s_i$  and  $b_i$  the expected number of signal and background yields for bin  $i$  respec-  
 2060 tively,  $n_i$  is the observed number of events in the bin  $i$  and  $\mu = \sigma/\sigma_{SM}$  is the signal  
 2061 strength. Notice that the number of entries per bin follows a Poisson distribution.  
 2062 The analysis presented in this thesis is based on the binned distribution of the ratio  
 2063 signal/background obtained from the BDT outputs.

### 2064 5.2.3 Hypothesis test

2065 The test statistic mentioned in section 5.1 involving  
 2066 ; it is achieved, according to the Neyman-Pearson lemma [128],  
 2067 by defining the acceptance region such that, for  $\mathbf{x}$  inside the region, the likelihood  
 2068 ratio, i.e., the ratio of probability distribution functions for signal and background,

<sub>2069</sub> **5.3 exclusion limits**

<sub>2070</sub> **5.4 asymptotic limits**

2071 **Chapter 6**

2072 **Search for production of a Higgs  
2073 boson and a single top quark in  
2074 multilepton final states in pp  
2075 collisions at  $\sqrt{s} = 13$  TeV**

2076 **6.1 Introduction**

2077 The Higgs boson discovery, supported on experimental observations and theoretical  
2078 predictions made about the SM, gives the clue of the way in that elementary particles  
2079 acquire mass through the Higgs mechanism; therefore, knowing the Higgs mass, the  
2080 Higgs-boson and Higgs-fermion couplings can be tested. In order to test the Higgs-top  
2081 coupling, several measurements have been performed, as stated in the chapter 2, but  
2082 they are limited to measure the square of the coupling; however, the production of a  
2083 Higgs boson in association with a single top quark ( $tH$ ) not only offers access to the  
2084 sign of the coupling, but also, to the CP phase of the Higgs couplings.

2085 This chapter presents the search for the associated production of a Higgs boson  
 2086 and a single top quark events, focusing on leptonic signatures provided by the Higgs  
 2087 decay modes to  $WW$ ,  $ZZ$ , and  $\tau\tau$ ; the 13 TeV dataset produced in 2016, which  
 2088 corresponds to an integrated luminosity of  $35.9\text{fb}^{-1}$ , is used. Constraints on the sign  
 2089 of the Higgs-top coupling ( $y_t$ ) have been derived from the decay rate of Higgs boson  
 2090 to photon pairs [45] and from the cross section for associated production of Higgs and  
 2091 Z bosons via gluon fusion [129], with recent results disfavoring negative signs of the  
 2092 coupling [56, 58, 130]. It expands previous analyses performed at 8 TeV [131, 132] and  
 2093 searches for associated production of  $t\bar{t}$  pair and a Higgs boson in the multilepton final  
 2094 state channel [133]; it also complements searches in other decay channels targeting  
 2095  $H \rightarrow b\bar{b}$  [134].

2096 As shown in section 2.5, the SM cross section of the associated production of a  
 2097 Higgs boson and a single top quark ( $tHq$ ) process is driven by a destructive interfer-  
 2098 ence between two contributions (see Figure 2.14), where the Higgs couples to either  
 2099 the W boson or the top quark; however, if the sign of the Higgs-top coupling is flipped  
 2100 with respect to the SM prediction, a large enhancement of the cross section occurs,  
 2101 making this analysis sensitive to such deviation. A second process, where the Higgs  
 2102 boson and top quark are accompanied by a W boson ( $tHW$ ) has similar behavior,  
 2103 albeit with a weaker interference pattern and lower contribution to the  $tH$  cross sec-  
 2104 tion, therefore, a combination of both processes would increase the sensitivity; in  
 2105 this analysis both contributions are combined and referred as  $tH$ channel. A third  
 2106 contribution comes from  $t\bar{t}H$  process. The purpose of this analysis is to investigate  
 2107 the exclusion of the presence of the  $tH + t\bar{t}H$  processes under the assumption of the  
 2108 anomalous Higgs-top coupling modifier ( $\kappa_t = -1$ ). The analysis exploits signatures  
 2109 with two leptons of the same sign (*2lss channel*) and three leptons (*3l channel*) in  
 2110 the final state.

2111        The first sections present the characteristic  $tHq$  signature as well as the expected  
 2112   backgrounds. The MC samples, data sets, and the physics object definitions are  
 2113   then defined. Following, the background predictions, the signal extraction, and the  
 2114   statistical treatment of the selected events as well as the systematic uncertainties are  
 2115   described. The final section present the results for the exclusion limits as a function  
 2116   of the ratio of  $\kappa_t$  and the dimensionless modifier of the Higgs-vector boson  $\kappa_V$ .

## 2117   **6.2 $tHq$ signature**

2118   In order to select events of  $tHq$  process, its features are translated into a set of  
 2119   selection rules; figure 6.1 shows the Feynman diagram and an schematic view of the  
 2120    $tHq$  process from the  $pp$  collision to the final state configuration. A single top quark  
 2121   is produced accompanied by a light quark, denoted as  $q$ ; this light quark is produced  
 2122   predominantly in the forward region of the detector. The Higgs boson which can  
 2123   be either emitted by the exchanged  $W$  boson or directly by the singly produced top  
 2124   quark.

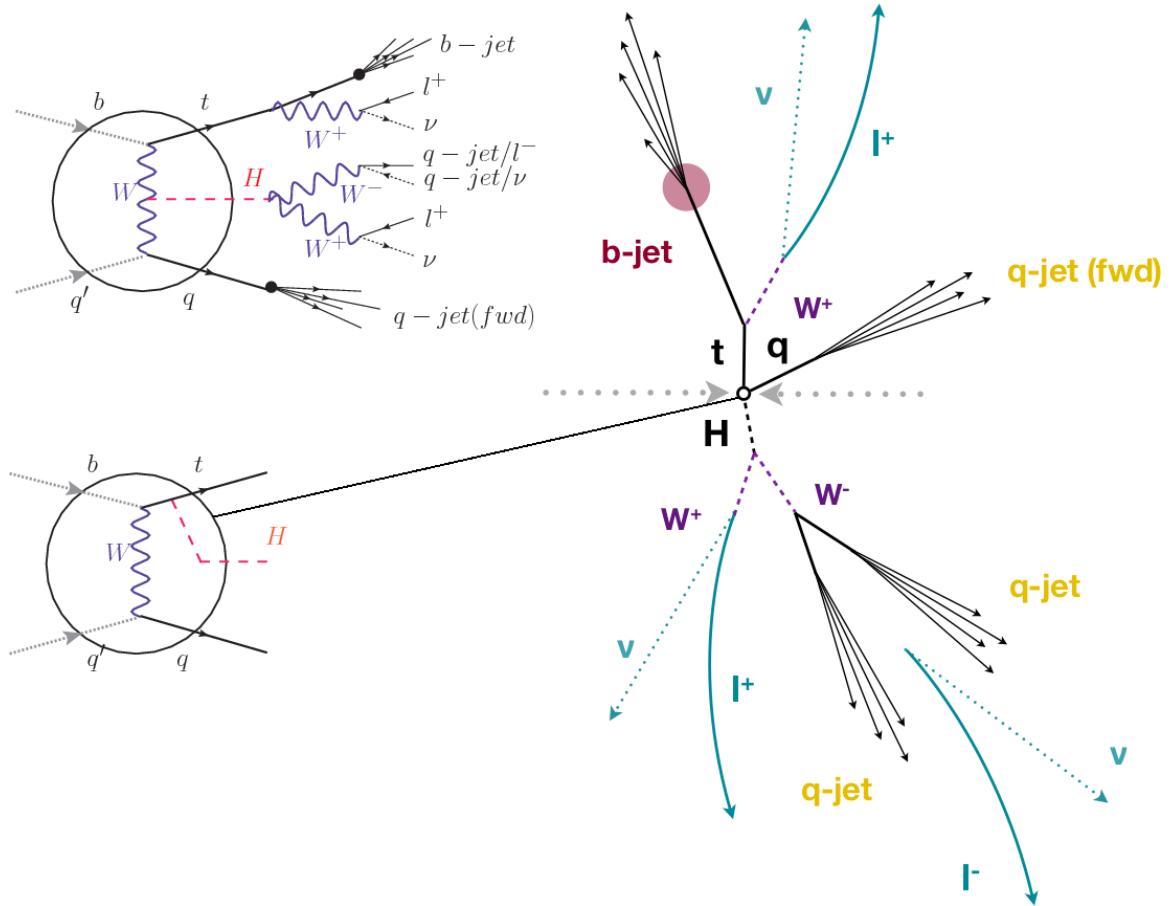
2125        The top quark and Higgs boson decay after their production in the detector due to  
 2126   their high masses/low lifetimes. The Higgs boson is required to decay into a  $W$  boson  
 2127   pair<sup>1</sup>. The top quark almost always decays into a bottom quark and a  $W$  boson, as  
 2128   encoded in the CMK matrix. The  $W$  bosons are required to decay hadronically in  
 2129   the 2lss channel case and leptonically in the 3l channel case, while  $\tau$  leptons are not  
 2130   reconstructed separately and only their leptonic decays into either electrons or muons  
 2131   are considered in this analysis.

2132        In summary, the signal process is characterized by a the final state with

- 2133        • one light-flavored forward jet,

---

<sup>1</sup> ZZ and  $\tau\tau$  decays are also include in the analysis but they are not separately reconstructed



**Figure 6.1:**  $tHq$  event signature. Left: Feynman diagram including the whole evolution up to the final state for the case of the Higgs boson emitted by the  $W$  boson (top); Feynman diagram for the case where the Higgs boson is emitted by the top quark. Right: Schematic view as it would be seen in the detector; the circle in the Feynman diagrams on the left corresponds to the circle in the center of the schematic view as indicated by the line connecting them. In the 2lss channel, one of the  $W$  bosons from the Higgs boson decays to two light-quark jets while in the 3l channel both  $W$  bosons decays to leptons.

- 2134     • one central b-jet,
  - 2135     • 2lss channel  $\rightarrow$  two leptons of the same sign, two neutrinos and two light (often  
2136       soft) jets,
  - 2137     • 3l channel  $\rightarrow$  three leptons, three neutrinos and no central light-flavored jets,
- 2138     The presence of neutrinos is inferred from the presence of MET. The analysis has

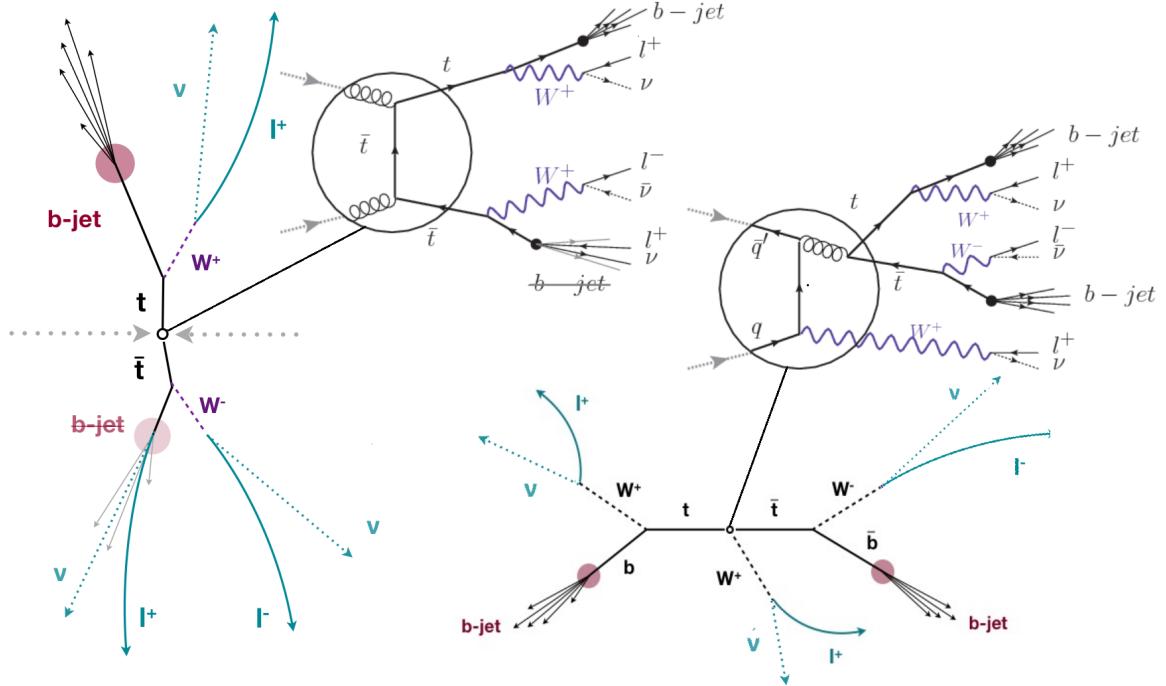
2139 been made public by CMS as a Physics Analysis Summary [135] combining the result  
 2140 for the three lepton and two lepton same-sign channels. Currently, an effort to turn  
 2141 the analysis into a paper is ongoing.

2142 **6.3 Background processes**

2143 The background processes are those that can mimic the signal signature or at least  
 2144 can be reconstructed as that as a result of certain circumstances. The backgrounds  
 2145 can be classified as

- 2146     • Irreducible backgrounds where genuine prompt leptons are produced in on-  
 2147         shell W and Z boson decays; they can be reliably estimated directly from MC  
 2148         simulated events, using higher-order cross sections or data control regions for  
 2149         the overall normalization.
- 2150     • Reducible backgrounds where at least one of the leptons is *non-prompt*, i.e., pro-  
 2151         duced within a hadronic jet, either a genuine lepton from heavy flavor decays.  
 2152         misreconstructed jets, also known as *mis-ID leptons* or *fake leptons*, are consid-  
 2153         ered non-prompt leptons as well. These non-prompt leptons leave tracks and  
 2154         hits in the detection systems as would a prompt lepton, but correlating those  
 2155         hits with nearby jets could be a way of removing them. Reducible backgrounds  
 2156         are not well predicted by simulation, and are estimated using data-driven meth-  
 2157         ods.

2158 The main sources of background events in the case of  $tHq$  process are  $t\bar{t}$  process  
 2159 and  $t\bar{t} + X(X = W, Z, \gamma)$  processes, here represented together as  $t\bar{t}V$  process. Figure  
 2160 6.2 shows the signature for  $t\bar{t}$  and  $t\bar{t}W$  processes;



**Figure 6.2:**  $t\bar{t}$  (left) and  $t\bar{t}W$  (right) events signature as they would be seen in the detector; the Feynman diagrams including the whole evolution up to the final state are also showed. The  $t\bar{t}$  process signature is very similar to that of the signal process with one fake lepton and no forward activity. The  $t\bar{t}W$  process present a higher  $b$ -jet multiplicity compared to the signal process, a prompt lepton and no forward activity.

2161        The largest contribution to irreducible backgrounds involving prompt leptons  
 2162      comes from  $t\bar{t}W$ ,  $t\bar{t}Z$ , processes for which the number of ( $b-$ )jets (( $b-$ )jet multiplicity)  
 2163      is higher than that of the signal events, while for other contributing background  
 2164      events,  $WZ$ ,  $ZZ$ , and rare SM processes like  $W^\pm W^\pm qq$ ,  $t\bar{t}t\bar{t}$ ,  $tZq$ ,  $tZW$ ,  $WWW$ ,  
 2165       $WWZ$ ,  $WZZ$ ,  $ZZZ$ , the ( $b-$ )jet multiplicity is lower compared to that of the signal  
 2166      events. None of the irreducible backgrounds present activity in the forward region of  
 2167      the detector.

2168        On the side of the reducible backgrounds, the largest contribution comes from the  
 2169       $t\bar{t}$  events which have a very similar signature to the signal events but does not present  
 2170      activity in the forward region of the detector either; A particular feature of the  $t\bar{t}$   
 2171      events is their charge-symmetry which is also a difference with the signal events.

2172        The charge misidentification plays an important role in the the 2lss channel since  
 2173        leptons in processes like  $t\bar{t}$  + jets or  $Z$  + jets can be charge misidentified, leading to  
 2174        backgrounds increments. An identification variable have been designed in order to  
 2175        reject this type of background events.

## 2176        6.4 Data and MC Samples

2177        Technical developments on the event generator side allow for an event-wise reweight-  
 2178        ing that can change the event kinematics based on specific generation parameters.  
 2179        This way not only the case of  $C_t = \sqrt{2}$ , but a whole range of  $\kappa_t$  and  $\kappa_V$  values can  
 2180        be investigated.

2181        The data considered in this analysis were collected by the CMS experiment dur-  
 2182        ing 2016 and correspond to a total integrated luminosity of  $35.9 fb^{-1}$ . Only periods  
 2183        when the CMS magnet was on were considered when selecting the data samples, that  
 2184        corresponds to the 23 Sep 2016 (Run B to G) and PromptReco (Run H) versions  
 2185        of the datasets. The MC samples used in this analysis correspond to the RunI-  
 2186        ISummer16MiniAODv2 campaign produced with CMSSW 80X. The two signal sam-  
 2187        ples (for  $tHq$  and  $tHW$ ) were produced with MG5\_aMC@NLO (version 5.222), in  
 2188        leading-order order mode, and are normalized to next-to-leading-order cross sections,  
 2189        see Tab. 6.1. Each sample is generated with a set of event weights corresponding to  
 2190        different values of  $\kappa_t$  and  $\kappa_V$  couplings as shown in Tab. 6.2.

### 2191        6.4.1 Full 2016 dataset and MC samples

2192        Different MC generators were used to generate the background processes. The dom-  
 2193        inant sources ( $t\bar{t}$ ,  $t\bar{t}W$ ,  $t\bar{t}Z$ ,  $t\bar{t}H$ ) were produced using AMC@NLO interfaced to  
 2194        PYTHIA8, and are scaled to NLO cross sections. Other processes are simulated us-

Sample	$\sigma$ [pb]	BF
/THQ_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1/	0.7927	0.324
/THW_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1/	0.1472	1.0

**Table 6.1:** Signal samples and their cross section and branching fraction used in this analysis. See Ref. [144] for more details.

ing POWHEG interfaced to PYTHIA, or bare PYTHIA (see table 6.3 and [133] for more details).

## 6.4.2 Triggers

We consider online-reconstructed events triggered by one, two, or three leptons. Single-lepton triggers are included to boost the acceptance of events where the  $p_T$  of the sub-leading lepton falls below the threshold of the double-lepton triggers. Additionally, by including double-lepton triggers in the  $\geq 3$  lepton category, as well as single-lepton triggers in all categories, we increase the efficiency, considering the logical “or” of the trigger decisions of all the individual triggers in a given category. Tab. 6.5 shows the lowest-threshold non-prescaled triggers present in the High-Level Trigger (HLT) menus for both Monte-Carlo and data in 2016.

### Trigger efficiency scale factors

The efficiency of events to pass the trigger is measured in simulation (trivially using generator information) and in the data (using event collected by an uncorrelated MET trigger). Small differences between the data and MC efficiencies are corrected by applying scale factors as shown in Tab. 6.6. The exact procedure and control plots are documented in [138] for the current analysis.

		<i>tHq</i>		<i>tHW</i>		
$\kappa_V$	$\kappa_t$	sum of weights	cross section [pb]	sum of weights	cross section [pb]	LHE weights
1.0	-3.0	35.700022	2.991	11.030445	0.6409	LHEweight_wgt[446]
1.0	-2.0	20.124298	1.706	5.967205	0.3458	LHEweight_wgt[447]
1.0	-1.5	14.043198	1.205	4.029093	0.2353	LHEweight_wgt[448]
1.0	-1.25	11.429338	0.9869	3.208415	0.1876	LHEweight_wgt[449]
1.0	-1.0		0.7927		0.1472	
1.0	-0.75	7.054998	0.6212	1.863811	0.1102	LHEweight_wgt[450]
1.0	-0.5	5.294518	0.4723	1.339886	0.07979	LHEweight_wgt[451]
1.0	-0.25	3.818499	0.3505	0.914880	0.05518	LHEweight_wgt[452]
1.0	0.0	2.627360	0.2482	0.588902	0.03881	LHEweight_wgt[453]
1.0	0.25	1.719841	0.1694	0.361621	0.02226	LHEweight_wgt[454]
1.0	0.5	1.097202	0.1133	0.233368	0.01444	LHEweight_wgt[455]
1.0	0.75	0.759024	0.08059	0.204034	0.01222	LHEweight_wgt[456]
1.0	1.0	0.705305	0.07096	0.273617	0.01561	LHEweight_wgt[457]
1.0	1.25	0.936047	0.0839	0.442119	0.02481	LHEweight_wgt[458]
1.0	1.5	1.451249	0.1199	0.709538	0.03935	LHEweight_wgt[459]
1.0	2.0	3.335034	0.2602	1.541132	0.08605	LHEweight_wgt[460]
1.0	3.0	10.516125	0.8210	4.391335	0.2465	LHEweight_wgt[461]
1.5	-3.0	45.281492	3.845	13.426212	0.7825	LHEweight_wgt[462]
1.5	-2.0	27.606715	2.371	7.809713	0.4574	LHEweight_wgt[463]
1.5	-1.5	20.476088	1.784	5.594971	0.3290	LHEweight_wgt[464]
1.5	-1.25	17.337465	1.518	4.635978	0.2749	LHEweight_wgt[465]
1.5	-1.0	14.483302	1.287	3.775902	0.2244	LHEweight_wgt[466]
1.5	-0.75	11.913599	1.067	3.014744	0.1799	LHEweight_wgt[467]
1.5	-0.5	9.628357	0.874	2.352505	0.1410	LHEweight_wgt[468]
1.5	-0.25	7.627574	0.702	1.789184	0.1081	LHEweight_wgt[469]
1.5	0.0	5.911882	0.5577	1.324946	0.08056	LHEweight_wgt[470]
1.5	0.25	4.479390	0.4365	0.959295	0.05893	LHEweight_wgt[471]
1.5	0.5	3.331988	0.3343	0.692727	0.04277	LHEweight_wgt[472]
1.5	0.75	2.469046	0.2558	0.525078	0.03263	LHEweight_wgt[473]
1.5	1.0	1.890565	0.2003	0.456347	0.02768	LHEweight_wgt[474]
1.5	1.25	1.596544	0.1689	0.486534	0.02864	LHEweight_wgt[475]
1.5	1.5	1.586983	0.1594	0.615638	0.03509	LHEweight_wgt[476]
1.5	2.0	2.421241	0.2105	1.170602	0.06515	LHEweight_wgt[477]
1.5	3.0	7.503280	0.5889	3.467546	0.1930	LHEweight_wgt[478]
0.5	-3.0	27.432685	2.260	8.929074	0.5136	LHEweight_wgt[479]
0.5	-2.0	13.956013	1.160	4.419093	0.2547	LHEweight_wgt[480]
0.5	-1.5	8.924438	0.7478	2.757611	0.1591	LHEweight_wgt[481]
0.5	-1.25	6.835341	0.5726	2.075247	0.1204	LHEweight_wgt[482]
0.5	-1.0	5.030704	0.4273	1.491801	0.08696	LHEweight_wgt[483]
0.5	-0.75	3.510528	0.2999	1.007273	0.05885	LHEweight_wgt[484]
0.5	-0.5	2.274811	0.1982	0.621663	0.03658	LHEweight_wgt[485]
0.5	-0.25	1.323555	0.1189	0.334972	0.01996	LHEweight_wgt[486]
0.5	0.0	0.656969	0.06223	0.147253	0.008986	LHEweight_wgt[487]
0.5	0.25	0.274423	0.02830	0.058342	0.003608	LHEweight_wgt[488]
0.5	0.5	0.176548	0.01778	0.068404	0.003902	LHEweight_wgt[489]
0.5	0.75	0.363132	0.03008	0.177385	0.009854	LHEweight_wgt[490]
0.5	1.0	0.834177	0.06550	0.385283	0.02145	LHEweight_wgt[491]
0.5	1.25	1.589682	0.1241	0.692099	0.03848	LHEweight_wgt[492]
0.5	1.5	2.629647	0.2047	1.097834	0.06136	LHEweight_wgt[493]
0.5	2.0	5.562958	0.4358	2.206057	0.1246	LHEweight_wgt[494]
0.5	3.0	14.843102	1.177	5.609519	0.3172	LHEweight_wgt[495]

**Table 6.2:**  $\kappa_V$  and  $\kappa_t$  combinations generated for the two signal samples and their NLO cross sections. The *tHq* cross section is multiplied by the branching fraction of the enforced leptonic decay of the top quark (0.324). See also Ref. [144].

Sample	$\sigma$ [pb]
TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
ttHJetToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8_mWCutfix /store/cmst3/group/susy/gpetrucc/13TeV/u/TTLL_m1to10_LO_NoMS_for76X/	0.2151 0.0283
WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	585.8
ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	131.3
TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	2.967
TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	2.967
TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.697
WpWpJJ_EWK-QCD_TuneCUETP8M1_13TeV-madgraph-pythia8	0.03711
ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.01398
WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.1651
WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.05565
WW_DoubleScattering_13TeV-pythia8	1.64
tZq_ll_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.0758
ST_tWll_5f_LO_13TeV-MadGraph-pythia8	0.01123
TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.009103
WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.4296
ZZTo4L_13TeV_powheg_pythia8	1.256
TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.1754
TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.1754
TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	87.3
DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	18610
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6024
WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	61526.7
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.6
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.6
ST_t-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	70.3144
ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	26.2278
ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.68064
WWTo2L2Nu_13TeV-powheg	10.481

**Table 6.3:** List of background samples used in this analysis (CMSSW 80X). In the first section of the table are listed the samples of the processes for which we use the simulation to extract the final yields and shapes, in the second section the samples of the processes we will estimate from data. The MC simulation is used to design the data driven methods and derive the associated systematics.

Sample	$\sigma$ [pb]
ttWJets_13TeV_madgraphMLM	0.6105
ttZJets_13TeV_madgraphMLM	0.5297/0.692

**Table 6.4:** Leading-order  $t\bar{t}W$  and  $t\bar{t}Z$  samples used in the signal BDT training.

Same-sign dilepton (==2 muons)
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
Same-sign dilepton (==2 electrons)
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Ele27_eta2p1_WP Loose_Gsf_v*
HLT_Ele27_WPTight_Gsf_v*
HLT_Ele25_eta2p1_WPTight_Gsf_v*
Same-sign dilepton (==1 muon, ==1 electron)
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
HLT_Ele27_WPTight_Gsf_v*
HLT_Ele25_eta2p1_WPTight_Gsf_v*
HLT_Ele27_eta2p1_WP Loose_Gsf_v*
Three lepton and Four lepton
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_v*
HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v*
HLT_TripleMu_12_10_5_v*
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
HLT_Ele27_WPTight_Gsf_v*
HLT_Ele25_eta2p1_WPTight_Gsf_v*
HLT_Ele27_eta2p1_WP Loose_Gsf_v*

**Table 6.5:** Table of high-level triggers that we consider in the analysis.

Category	Scale Factor
ee	$1.01 \pm 0.02$
e $\mu$	$1.01 \pm 0.01$
$\mu\mu$	$1.00 \pm 0.01$
3l	$1.00 \pm 0.03$

**Table 6.6:** Trigger efficiency scale factors and associated uncertainties, shown here rounded to the nearest percent.

## 2212 6.5 Object Identification

2213 In this section, the specific definitions of the physical objects in terms of the numerical  
 2214 values assigned to the reconstruction parameters are presented; thus, the provided  
 2215 details summarize and complement the descriptions presented in previous chapters.  
 2216 The object reconstruction and selection strategy used in this thesis is inherited from  
 2217 the analyses in references [133, 138], thus, the information provided in this section is  
 2218 extracted from those documents unless other references are stated.

### 2219 6.5.1 Jets and $b$ -jet tagging.

2220 In this analysis, jets are reconstructed by clustering PF candidates using the anti- $k_t$   
 2221 algorithm with parameter distance  $\Delta R = 0.4$ ; those charged hadrons that are not  
 2222 consistent with the selected primary vertex are discarded from the clustering. The  
 2223 jet energy is then corrected for the varying response of the detector as a function  
 2224 of transverse momentum  $p_T$  and pseudorapidity  $\eta$ . Jets are selected for use in the  
 2225 analysis only if they have  $p_T > 25$  GeV and are separated from any selected leptons  
 2226 by  $\Delta R > 0.4$ .

2227 Jets coming from the primary vertex and jets coming from pile-up vertices are  
 2228 distinguished using a MVA discriminator based on the differences in the jet shapes,  
 2229 in the relative multiplicity of charged and neutral components, and in the different  
 2230 fraction of transverse momentum which is carried by the hardest components. Jet

2231 tracks are also required to be compatible with the primary vertex.

2232 Jets originated from the hadronization of a  $b$  quark are selected using a MVA  
 2233 likelihood discriminant which uses track-based lifetime information and reconstructed  
 2234 secondary vertices (CSV algorithm). Only jets within the CMS tracker acceptance  
 2235 ( $\eta < 2.4$ ) are identified with this tool. Data samples are used to measure the efficiency  
 2236 of the  $b$ -jet tagging and the probability to misidentify jets from light quarks or gluons;  
 2237 in both cases the measurements are parametrized as a function of the jet  $p_T$  and  $\eta$   
 2238 and later used to correct differences between the data and MC simulation in the  $b$   
 2239 tagging performance, by applying per-jet weights to the simulation, dependent on  
 2240 the jet  $p_T$ ,  $\eta$ ,  $b$  tagging discriminator, and flavor (from simulation truth) [136]. The  
 2241 per-event weight is taken as the product of the per-jet weights, including those of the  
 2242 jets associated to the leptons. The weights are derived on  $t\bar{t}$  and Z+jets events.

2243 Two working points are defined, based on the CSV algorithm output: *loose*' work-  
 2244 ing point (CSV>0.46) with a  $b$  signal tagging efficiency of about 83% and a mistagging  
 2245 rate of about 8%; and *medium* working point (CSV>0.80) with  $b$ -tagging efficiency of  
 2246 about 69% and mistagging rate of order 1% [137]. Tagging of jets from charm quarks  
 2247 have efficiencies of about 40% and 18% for loose and medium working points re-  
 2248 spectively. Separate scale factors are applied to jets originating from bottom/charm  
 2249 quarks and from light quarks in simulated events to match the tagging efficiencies  
 2250 measured in the data.

## 2251 6.5.2 Missing Energy MET.

2252 As stated in section 4.4.1.1, the MET vector is calculated as the negative of the vector  
 2253 sum of transverse momenta of all PF candidates in the event and its magnitude is  
 2254 referred to as  $E_T^{miss}$ . Due to pile-up interactions, the performance in determining

2255 MET is degraded; in order to correct for that, the energy from the selected jets and  
 2256 leptons that compose the event is assigned to the variable  $H_T^{miss}$ . It is calculated in  
 2257 the same way as  $E_T^{miss}$  and although it has worse resolution than  $E_T^{miss}$ , it is more  
 2258 robust in the sense that it does not rely on the soft part of the event. The event  
 2259 selection uses a linear discriminator based on the two variables given by

$$E_T^{miss} LD = 0.00397 * E_T^{miss} + 0.00265 * H_T^{miss} \quad (6.1)$$

2260 taking advantage of the fact that the correlation between  $E_T^{miss}$  and  $H_T^{miss}$  is less  
 2261 for events with instrumental missing energy than for events with real missing energy.  
 2262 The working point  $E_T^{miss} LD > 0.2$  was chosen to ensure a good signal efficiency while  
 2263 keeping a good background rejection.

### 2264 6.5.3 Lepton reconstruction and identification

2265 Two types of leptons are defined in this analysis: *signal leptons* are those coming from  
 2266  $W, Z$  and  $\tau$  decays which usually are isolated from other particles; *background leptons*  
 2267 are defined as leptons produced in  $b$ -jet hadron decays, light-jets misidentification,  
 2268 and photon conversions.

2269 The process of reconstruction and identification of electron and muon candidates  
 2270 was described in chapter4, hence, the identification variables used in order to retain  
 2271 the highest possible efficiency for signal leptons while maximizing the rejection of  
 2272 background leptons are listed and described in the following sections <sup>2</sup>.

2273 The identification variables include not only observables related directly to the re-  
 2274 constructed leptons themselves, but also to the clustered energy deposits and charged  
 2275 particles in a cone around the lepton direction (jet-related variables); an initial loose

---

<sup>2</sup> the studies performed to optimize the identification are far from the scope of this thesis, therefore, only general descriptions are provided

2276 preselection of leptons candidates is performed and then an MVA discriminator, re-  
 2277 ferred to as *lepton MVA* discriminator, is used to distinguish signal leptons from  
 2278 background leptons.

2279 **Muons.**

2280 The Physics Objects Groups (POG) at CMS, are in charge of studying and defining  
 2281 the set of selection criteria applied on the course of reconstruction and identification  
 2282 of particles. These selection criteria are implemented in the CMS framework in the  
 2283 form of several object identification working points according to the strength of the  
 2284 requirements.

2285 The muon candidates are reconstructed by combining information from the tracker  
 2286 system and the muon detection system of CMS detector and the POG defined three  
 2287 working points for muon identification *MuonID* [139];

- 2288     • *POG Loose Muon ID* is a particle identified as a muon by the PF event re-  
   2289 construction and also reconstructed either as a global-muon or as an arbitrated  
   2290 tracker-muon. This identification criteria is designed to be highly efficient for  
   2291 prompt muons and for muons from heavy and light quark decays; it can be com-  
   2292 plemented by applying impact parameter cuts in analyses with prompt muon  
   2293 signals.
- 2294     • *POG Medium Muon ID* is a Loose muon with additional track-quality and  
   2295 muon-quality (spatial matching between the individual measurements in the  
   2296 tracker and the muon system) requirements. This identification criteria is de-  
   2297 signed to be highly efficient in the separation of the muons coming from decay  
   2298 in flight of heavy quarks and muons coming from B meson decays as well as  
   2299 prompt muons. An additional category *MVA Prompt ID* is defined in this iden-

2300 tification criteria directed to discriminated muons from B mesons and prompt  
 2301 muons (from W,Z and  $\tau$  decays). The Medium ID provides the same fake rate as  
 2302 the Tight Muon ID but a higher efficiency on prompt and B-decays muons. [140]

- 2303 • *POG Tight Muon ID* is a global muon with additional muon-quality require-  
 2304 ments Tight Muon ID selects a subset of the PF muons.

2305 Only muons within the muon system acceptance  $|\eta| < 2.4$  and minimum  $p_T$  of 5  
 2306 GeV are considered.

2307 **Electrons.**

2308 Electrons are reconstructed using information from the tracker and from the electro-  
 2309 magnetic calorimeter and identified by an MVA algorithm (*MVA eID* discriminant)  
 2310 using the shape of the calorimetric shower variables like the shape in  $\eta$  and  $\phi$ , the clus-  
 2311 ter circularity, widths along  $\eta$  and  $\phi$ ; track-cluster matching variables like  $E_{tot}/p_{in}$ ,  
 2312  $E_{Ele}/p_{out}$ ,  $\Delta\eta_{in}$ ,  $\Delta\eta_{out}$ ,  $\Delta\phi_{in}$ ,  $1/E - 1/p$ ; and track quality variables like  $\xi^2$  of the  
 2313 GSF tracks, the number of hits used by the GSF filter [141].

2314 A loose selection based on  $\eta$ -dependent cuts on this discriminant is used to prese-  
 2315 lect electron candidates, the full shape of the discriminant is used in the lepton MVA  
 2316 selection to separate signal leptons from background leptons (described in section  
 2317 6.5.3).

2318 In order to reject electrons from photon conversions, electron candidates with  
 2319 missing hits in the pixel tracker layers or matched to a conversion secondary vertex  
 2320 are discarded. Electrons are selected for the analysis if they have  $p_T > 7$  GeV and  
 2321 are located within the tracker system acceptance region ( $|\eta| < 2.5$ ).

2322 **Lepton vertexing and pile-up rejection.**

2323 The impact parameter in the transverse plane  $d_0$ , impact parameter along the z  
 2324 axis  $d_z$ , and the impact parameter significance in the detector space  $SIP_{3D}$ , are  
 2325 considered to perform the identification and rejection of pile-up, misreconstructed  
 2326 tracks, and background leptons from b-hadron decays; pile-up and misreconstructed  
 2327 track mitigation is achieved by imposing loose cuts on the impact parameter variables.  
 2328 The full shape of the those variables is used in a lepton MVA classifier to achieve the  
 2329 best separation between the signal and the background leptons.

2330 **Lepton isolation.**

2331 PF is able to recognize leptons from two different sources: on one side, leptons from  
 2332 the decays of heavy particles, such as W and Z bosons, which are normally isolated  
 2333 in space from the hadronic activity in the event; on the other side, leptons from the  
 2334 decays of hadrons and jets misidentified as leptons, which are not isolated as the  
 2335 former. For highly boosted systems, like the lepton and the  $b$ -jet generated in the  
 2336 semileptonic decay of a boosted top, the decay products tend to be more closer and  
 2337 sometimes they even overlap; thus, the PF standard definition of isolation in terms of  
 2338 the separation between the lepton candidates and other PF objects in the  $\eta$ - $\phi$  plane,

$$\Delta R = \sqrt{(\eta^l - \eta^i)^2 + (\phi^l - \phi^i)^2} < 0.3 \quad (6.2)$$

2339 which considers all the neutral, charged hadrons and photons in a cone around the  
 2340 leptons, is refocused to the local isolation of the leptons through the mini-isolation  
 2341  $I_{mini}$  [142] defined as the sum of particle flow candidates  $p_T$  within a cone around

2342 the lepton, corrected for the effects of pileup and divided by the lepton  $p_T$

$$I_{mini} = \frac{\sum_R p_T(h^\pm) - \max\left(0, \sum_R p_T(h^0) + p_T(\gamma) - \rho \mathcal{A}\left(\frac{R}{0.3}\right)^2\right)}{p_T(l)} \quad (6.3)$$

2343 where  $\rho$  is the pileup energy density,  $h^\pm, h^0, \gamma, l$ , represent the charged hadron, neutral  
 2344 hadrons, photons, and the lepton, respectively. The radius  $R$  of the cone depends on  
 2345 the  $p_T$  of the lepton according to

$$R = \frac{10\text{GeV}}{\min(\max(p_T(l), 50\text{GeV}), 200\text{GeV})}, \quad (6.4)$$

2346 The  $p_T$  dependence of the cone size allows for greater signal efficiency. Setting a  
 2347 cut on  $I_{mini}$  below a given threshold ensures that the lepton is locally isolated, even  
 2348 in boosted systems. The effect of pileup is mitigated using the so-called effective area  
 2349 correction  $\mathcal{A}$  listed in Table 6.7.

$ \eta $ range	$\mathcal{A}(e)$ neutral/charged	$\mathcal{A}(\mu)$ neutral/charged
0.0 - 0.8	0.1607 / 0.0188	0.1322 / 0.0191
0.8 - 1.3	0.1579 / 0.0188	0.1137 / 0.0170
1.3 - 2.0	0.1120 / 0.0135	0.0883 / 0.0146
2.0 - 2.2	0.1228 / 0.0135	0.0865 / 0.0111
2.2 - 2.5	0.2156 / 0.0105	0.1214 / 0.0091

**Table 6.7:** Effective areas, for electrons and muons used to mitigate the effect of pileup by using the so-called effective area correction.

2350 A loose cut on  $I_{mini}$  is applied to pre-select the muon and electron candidates;  
 2351 however, the full shape is used in the lepton MVA discriminator when performing the  
 2352 signal lepton selection.

2353 **Jet-related variables.**

2354 In order to reject misidentified leptons from  $b$ -jets, mostly coming from  $t\bar{t}$ +jets,  
 2355 Drell-Yan+jets, and W+jets events, the vertexing and isolation described in previous  
 2356 sections are complemented with additional variables related to the closest recon-  
 2357 structed jet to the lepton, i.e., the PF jets reconstructed<sup>3</sup> around the leptons with  
 2358  $\Delta R = \sqrt{(\eta^l - \eta^{jet})^2 + (\phi^l - \phi^{jet})^2} < 0.5$ . The identification variables used in the lep-  
 2359 ton MVA discriminator are the ratio  $p_T^l/p_T^{jet}$ , the CSV b-tagging discriminator value  
 2360 of the jet, the number of charged tracks of the jet, and the relative  $p_T$  given by

$$p_T^{rel} = \frac{(\vec{p}_{jet} - \vec{p}_l) \cdot \vec{p}_l}{\|\vec{p}_{jet} - \vec{p}_l\|}. \quad (6.5)$$

2361 **LeptonMVA discriminator.**

2362 Electrons and muons passing the basic selection process described above are referred  
 2363 to as *loose leptons*. Additional discrimination between signal leptons and background  
 2364 leptons is crucial considering that the rate of  $t\bar{t}$  production is much larger than the  
 2365 signal, hence, an overwhelming background from  $t\bar{t}$  production. To maximally ex-  
 2366 ploit the available information in each event to that end, the dedicated lepton MVA  
 2367 discriminator, based on a boosted decision tree (BDT) algorithm, has been built so  
 2368 that all the identification variables can be used together.

2369 The lepton MVA discriminator training is performed using simulated signal Loose  
 2370 leptons from the  $t\bar{t}H$  MC sample and fake leptons from the  $t\bar{t}$  +jets MC sample,  
 2371 separately for muons and electrons. The input variables used include vertexing, iso-  
 2372 lation and jet-related variables, the  $p_T$  and  $\eta$  of the lepton, the electron MVA eID  
 2373 discriminator and the muon segment-compatibility variables. An additional require-  
 2374 ment known as *tight-charge* requirement, is imposed by comparing two independent

---

<sup>3</sup> charged hadrons from PU vertices are not removed prior to the jet clustering.

measurement of the charge, one from the ECAL supercluster and the other from the tracker; thus, the consistency in the measurements of the electron charge is ensured so that events with a wrong electron charge assignment are rejected; this variable is particularly used in the 2lss channel to suppress opposite-sign events for which the charge of one of the leptons has been mismeasured. The tight-charge requirement for muons is represented by the requirement of a consistently well measured track transverse momentum given by  $\Delta p_T/p_T < 0.2$ . Leptons are selected for the final analysis if they pass a given threshold of the BDT output, and are referred to as *tight leptons* in the following.

The validation of the lepton MVA algorithm and the lepton identification variables is performed using data in various control regions; the details about that validation are not discussed here but can be found in Reference [138].

### Selection definitions

Electron and muon object identification is defined in three different sets of selections criteria; the *Loose*, *Fakeable Object*, and *Tight* selection. These three levels of selection are designed to serve for event level vetoes, the fake rate estimation application region, and the final signal selection, respectively. The  $p_T$  of fakeable objects is defined as  $0.85 \times p_T(\text{jet})$ , where the jet is the one associated to the lepton object. This mitigates the dependence of the fake rate on the momentum of the fakeable object and thereby improves the precision of the method.

Tables 6.8 and 6.9 list the full criteria for the different selections of muons and electrons.

In addition to the previously defined requirements for jets, they are required to be separated from any lepton candidates passing the fakeable object selections by  $\Delta R > 0.4$ .

Cut	Loose	Fakeable object	Tight
$ \eta  < 2.4$	✓	✓	✓
$p_T$	$> 5\text{GeV}$	$> 15\text{GeV}$	$> 15\text{GeV}$
$ d_{xy}  < 0.05 \text{ (cm)}$	✓	✓	✓
$ d_z  < 0.1 \text{ (cm)}$	✓	✓	✓
$\text{SIP}_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
is Loose Muon	✓	✓	✓
jet CSV	—	$< 0.8484$	$< 0.8484$
is Medium Muon	—	—	✓
tight-charge	—	—	✓
lepMVA $> 0.90$	—	—	✓

**Table 6.8:** Requirements on each of the three muon selections. In the cases where the cut values change between the selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated.

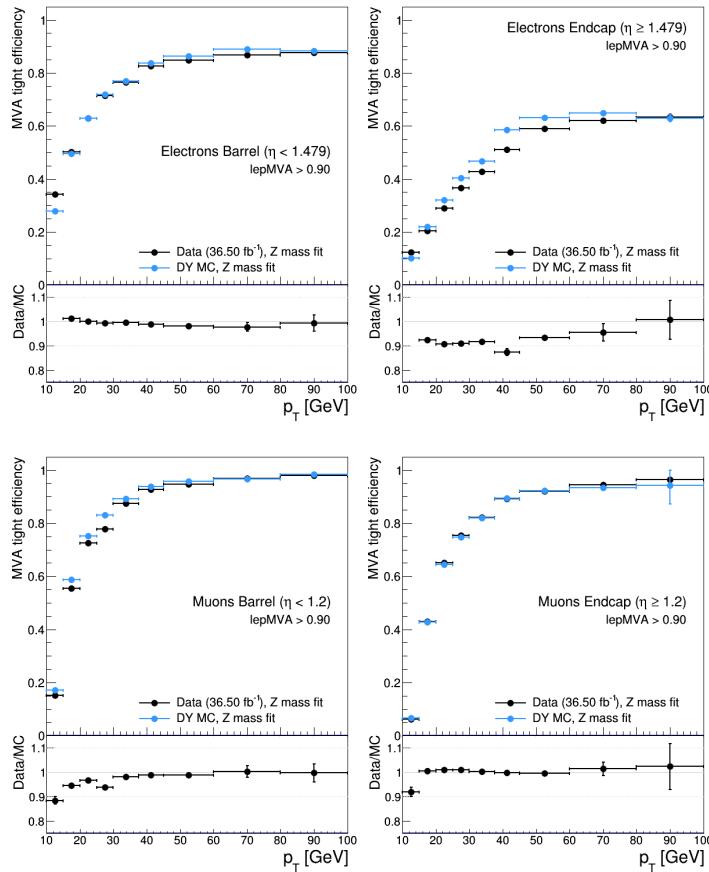
Cut	Loose	Fakeable Object	Tight
$ \eta  < 2.5$	✓	✓	✓
$p_T$	$> 7\text{GeV}$	$> 15\text{GeV}$	$> 15\text{GeV}$
$ d_{xy}  < 0.05 \text{ (cm)}$	✓	✓	✓
$ d_z  < 0.1 \text{ (cm)}$	✓	✓	✓
$\text{SIP}_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
MVA eID $> (0.0, 0.0, 0.7)$	✓	✓	✓
$\sigma_{in\eta} < (0.011, 0.011, 0.030)$	—	✓	✓
H/E $< (0.10, 0.10, 0.07)$	—	✓	✓
$\Delta\eta_{in} < (0.01, 0.01, 0.008)$	—	✓	✓
$\Delta\phi_{in} < (0.04, 0.04, 0.07)$	—	✓	✓
$-0.05 < 1/E - 1/p < (0.010, 0.010, 0.005)$	—	✓	✓
$p_T^{\text{ratio}}$	—	$> 0.5^\dagger / -$	—
jet CSV	—	$< 0.3^\dagger / < 0.8484$	$< 0.8484$
tight-charge	—	—	✓
conversion rejection	—	—	✓
Number of missing hits	$< 2$	$= 0$	$= 0$
lepton MVA $> 0.90$	—	—	✓

**Table 6.9:** Criteria for each of the three electron selections. In cases where the cut values change between selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. In some cases, the cut values change for different  $\eta$  ranges. These ranges are  $0 < |\eta| < 0.8$ ,  $0.8 < |\eta| < 1.479$ , and  $1.479 < |\eta| < 2.5$  and the respective cut values are given in the form (value<sub>1</sub>, value<sub>2</sub>, value<sub>3</sub>). For the two  $p_T^{\text{ratio}}$  and CSV rows, the cuts marked with a  $\dagger$  are applied to leptons that fail the lepton MVA cut, while the loose cut value is applied to those that pass the lepton MVA cut.

## 2400 6.5.4 Lepton selection efficiency

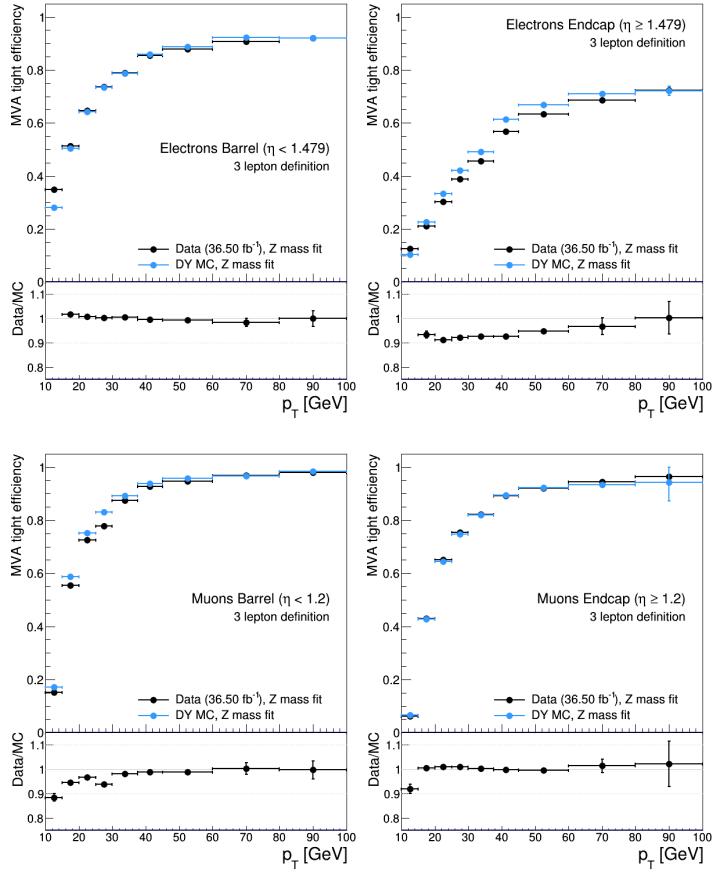
2401 Efficiencies of reconstruction and selecting loose leptons are measured both for muons  
 2402 and electrons using a tag and probe method on both data and MC, using  $Z \rightarrow \ell^+ \ell^-$   
 2403 [143]. The scale factors are derived from the ratio of efficiencies  $\varepsilon_i(p_T, \eta)$  measured  
 2404 for a given lepton in data/MC, according to

$$\rho(p_T, \eta) = \frac{\varepsilon_{\text{data}}(p_T, \eta)}{\varepsilon_{\text{MC}}(p_T, \eta)}. \quad (6.6)$$



**Figure 6.3:** Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the 2lss definition, i.e., including the tight-charge requirement.

2405 The scale factor for each event is used to correct the weight of the event in the



**Figure 6.4:** Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the 3l channel not including the tight-charge requirement.

2406 full sample; therefore, the full simulation correction is given by the product of all  
 2407 the individual scale factors. The scale factors used in this thesis are inherited from  
 2408 the reference [138] which in turns inherited them from leptonic SUSY analyses using  
 2409 equivalent lepton selections.

2410 The efficiency of applying the tight selection as defined in Tables 6.8 and 6.9, on the  
 2411 loose leptons are determined by using a tag and probe method on a sample of Drell-  
 2412 Yan enriched events. Figures 6.3 and 6.4 show the efficiencies for the 2lss channel and  
 2413 3l channel respectively. Efficiencies in the 2lss channel have been produced including  
 2414 the tight-charge requirement, while for the 3l channel it is not included. Number

2415 of passed and failed probes are determined from a fit to the invariant mass of the  
 2416 dilepton system.

2417 Simulation is corrected using these scale factors; note that they depends on  $\eta$  and  
 2418  $p_T$ .

2419 **6.6 Event selection**

2420 .

2421 .

2422 .

2423 .

2424 .

2425 .

2426 .

2427 .

2428 .

2429 . . .

2430 The analysis is designed to efficiently identify and select prompt leptons from  
 2431 on-shell W and

2432 Z boson decays and to reject non-prompt leptons from b quark decays and spurious  
 2433 lepton

2434 signatures from hadronic jets. Events are then selected in the various lepton  
 2435 channels, and are

2436 required to contain hadronic jets, some of which must be consistent with b quark  
 2437 hadronization. Finally, the signal yield is extracted by simultaneously fitting the  
 2438 output of two dedicated

2439 multivariate discriminants (trained to separate the  $tHq$  signal from the two dom-  
 2440 inant backgrounds) in all categories

2441 . Multivariate techniques are used to discriminate the signal from the dominant  
 2442 backgrounds. The analysis yields a 95% confidence level (C.L.) upper limit on the  
 2443 combined  $tH + ttH$  production cross section times branching ratio of 0.64 pb, with  
 2444 an expected limit of 0.32 pb, for a scenario with  $kt = \sqrt{1.0}$  and  $kV = 1.0$ . Values  
 2445 of  $kt$  outside the range of  $\sqrt{1.25}$  to  $\sqrt{1.60}$  are excluded at 95% C.L., assuming  $kV$   
 2446  $= 1.0$ .

2447 Dont forget to mention previous constrains to ct check reference ?? and references  
 2448 <https://link.springer.com/content/pdf/10.1007%2FJHEP01%282013%29088.pdf> (para-  
 2449 graph after eq 2)

2450 We selects events with three leptons and a  $b$ -jet tagged jet in the final state. The  $tHq$   
 2451 signal contribution is then determined in a fit of the observed data to two multivariate  
 2452 classifier outputs, each trained to discriminate against one of the two dominant back-  
 2453 grounds of events with non-prompt leptons from  $t\bar{t}$  and of associated production of  $t\bar{t}$   
 2454 and vector bosons ( $t\bar{t}W$ ,  $t\bar{t}Z$ ). The fit result is then used to set an upper limit on the  
 2455 combined  $t\bar{t}H$ ,  $tHq$  and  $tHW$  production cross section, as a function of the relative  
 2456 coupling strengths of Higgs and top quark and Higgs and W boson, respectively.

## 2457 6.7 Background predictions

2458 The modeling of reducible and irreducible backgrounds in this analysis uses the exact  
 2459 methods, analysis code, and ROOT trees used for the  $t\bar{t}H$  multilepton analysis. We  
 2460 give a brief description of the methods and refer to the documentation of that analysis  
 2461 in Refs. [133, 138] for any details.

2462 The backgrounds in three-lepton final states can be split in two broad categories:

irreducible backgrounds with genuine prompt leptons (i.e. from on-shell W and Z boson decays); and reducible backgrounds where at least one of the leptons is “non-prompt”, i.e. produced within a hadronic jet, either a genuine lepton from heavy flavor decays, or simply mis-reconstructed jets.

Irreducible backgrounds can be reliably estimated directly from Monte-Carlo simulated events, using higher-order cross sections or data control regions for the overall normalization. This is done in this analysis for all backgrounds involving prompt leptons:  $t\bar{t}W$ ,  $t\bar{t}Z$ ,  $t\bar{t}H$ ,  $WZ$ ,  $ZZ$ ,  $W^\pm W^\pm qq$ ,  $t\bar{t}t\bar{t}$ ,  $tZq$ ,  $tZW$ ,  $WWW$ ,  $WWZ$ ,  $WZZ$ ,  $ZZZ$ .

Reducible backgrounds, on the other hand, are not well predicted by simulation, and are estimated using data-driven methods. In the case of non-prompt leptons, a fake rate method is used,

Additional identification criteria are applied for electrons with  $p_T$  greater than 30 GeV to mimic the identification applied at trigger level in order to ensure consistency between the measurement region and application region of the fake-rate.

where the contribution to the final selection is estimated by extrapolating from a sideband (or “application region”) with a looser lepton definition (the fakeable object definitions in Tabs. 6.8 and 6.9) to the signal selection. The tight-to-loose ratios (or “fake rates”) are measured in several background dominated data events with dedicated triggers, subtracting the residual prompt lepton contribution using MC. Non-prompt leptons in our signal regions are predominantly produced in  $t\bar{t}$  events, with a much smaller contribution, from Drell–Yan production. The systematic uncertainty on the normalization of the non-prompt background estimation is on the order of 50%, and thereby one of the dominant limitations on the performance of multilepton analyses in general and this analysis in particular. It consists of several individual sources, such as the result of closure tests of the method using simulated

2489 events, limited statistics in the data control regions due to necessary prescaling of  
 2490 lepton triggers, and the uncertainty in the subtraction of residual prompt leptons  
 2491 from the control region.

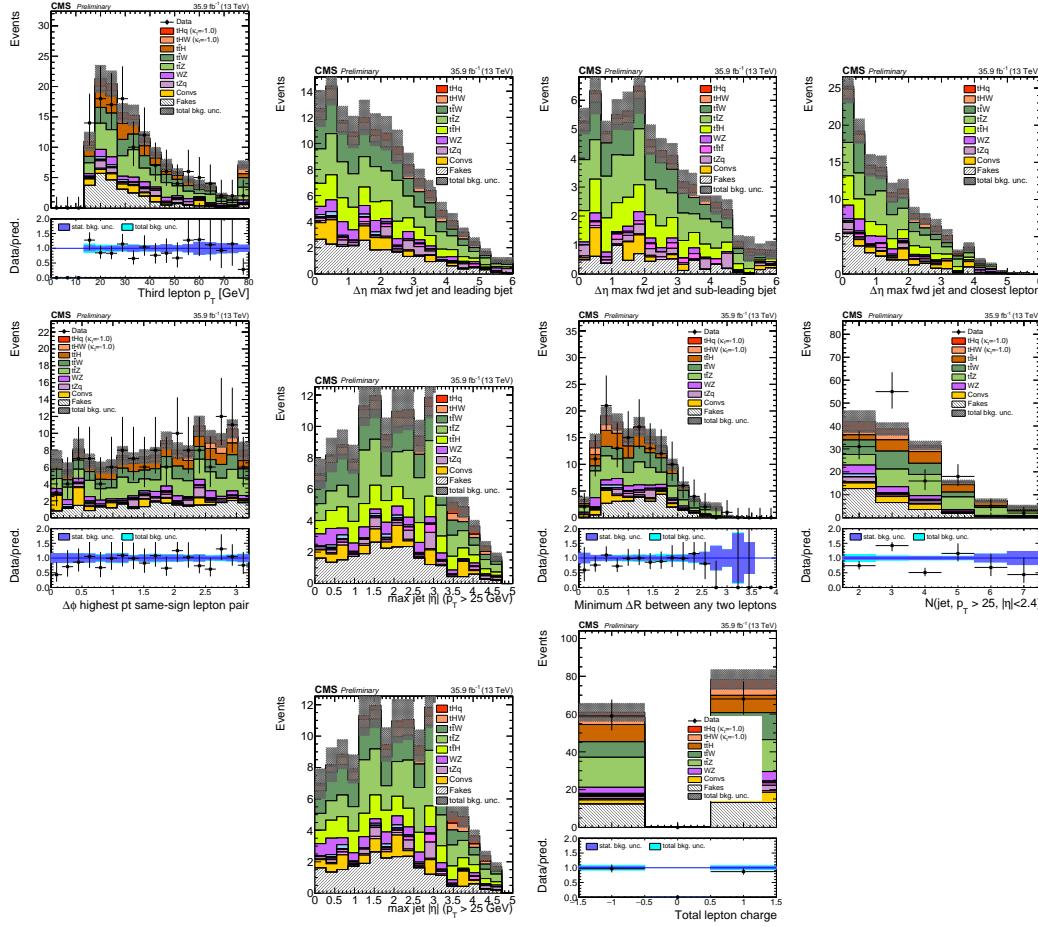
2492 The fake background where the leptons pass the looser selection are weighted  
 2493 according to how many of them fail the tight criteria. Events with a single failing  
 2494 lepton are weighted with the factor  $f/(1-f)$  for the estimate to the tight selection  
 2495 region, where  $f$  is the fake rate. Events with two failing leptons are given the negative  
 2496 weight  $-f_i f_j / (1-f_i)(1-f_j)$ , and for three leptons the weight is positive and equal  
 2497 to the product of  $f/(1-f)$  factor evaluated for each failing lepton.

2498 Figures 6.5 show the distributions of some relevant kinematic variables, normalized  
 2499 to the cross section of the respective processes and to the integrated luminosity.

## 2500 6.8 Signal discrimination

2501 The  $tHq$  signal is separated from the main backgrounds using a boosted decision  
 2502 tree (BDT) classifier, trained on simulated signal and background events. A set  
 2503 of discriminating variables are given as input to the BDT which produces a output  
 2504 distribution maximizing the discrimination power. Table 6.10 lists the input variables  
 2505 used while Figures 6.6 show their distributions for the relevant signal and background  
 2506 samples, for the three lepton channel. Two BDT classifiers are trained for the two  
 2507 main backgrounds expected in the analysis: events with prompt leptons from  $t\bar{t}W$  and  
 2508  $t\bar{t}Z$  (also referred to as  $t\bar{t}V$ ), and events with non-prompt leptons from  $t\bar{t}$ . The datasets  
 2509 used in the training are the  $tHq$  signal (see Tab. 6.1), and LO MADGRAPH samples  
 2510 of  $t\bar{t}W$  and  $t\bar{t}Z$ , in an admixture proportional to their respective cross sections (see  
 2511 Tab. 6.4).

2512 The MVA analysis consist of two stages: first a “training” where the MVA method



**Figure 6.5:** Distributions of input variables to the BDT for signal discrimination, three lepton channel, normalized to their cross section and to  $35.9\text{fb}^{-1}$ .

is trained to discriminate between simulated signal and background events, then a “test” stage where the trained algorithm is used to classify different events from the samples. The sample is obtained from a pre-selection (see Tab. ?? with pre-selection cuts). Figures 6.7 show the input variables distributions as seen by the MVA algorithm. Note that in contrast to the distributions in Fig. 6.6 only the main backgrounds ( $t\bar{t}$  from simulation,  $t\bar{t}V$ ) are included.

Note that splitting the training in two groups reveals that some variables show opposite behavior for the two background sources; potentially screening the discrimination power if they were to be used in a single discriminant. For some other variables

Variable name	Description
nJet25	Number of jets with $p_T > 25$ GeV, $ \eta  < 2.4$
MaxEtaJet25	Maximum $ \eta $ of any (non-CSV-loose) jet with $p_T > 25$ GeV
totCharge	Sum of lepton charges
nJetEta1	Number of jets with $ \eta  > 1.0$ , non-CSV-loose
detaFwdJetBJet	$\Delta\eta$ between forward light jet and hardest CSV loose jet
detaFwdJet2BJet	$\Delta\eta$ between forward light jet and second hardest CSV loose jet (defaults to -1 in events with only one CSV loose jet)
detaFwdJetClosestLep	$\Delta\eta$ between forward light jet and closest lepton
dphiHighestPtSSPair	$\Delta\phi$ of highest $p_T$ same-sign lepton pair
minDRll	minimum $\Delta R$ between any two leptons
Lep3Pt/Lep2Pt	$p_T$ of the 3 <sup>rd</sup> lepton (2 <sup>nd</sup> for ss2l)

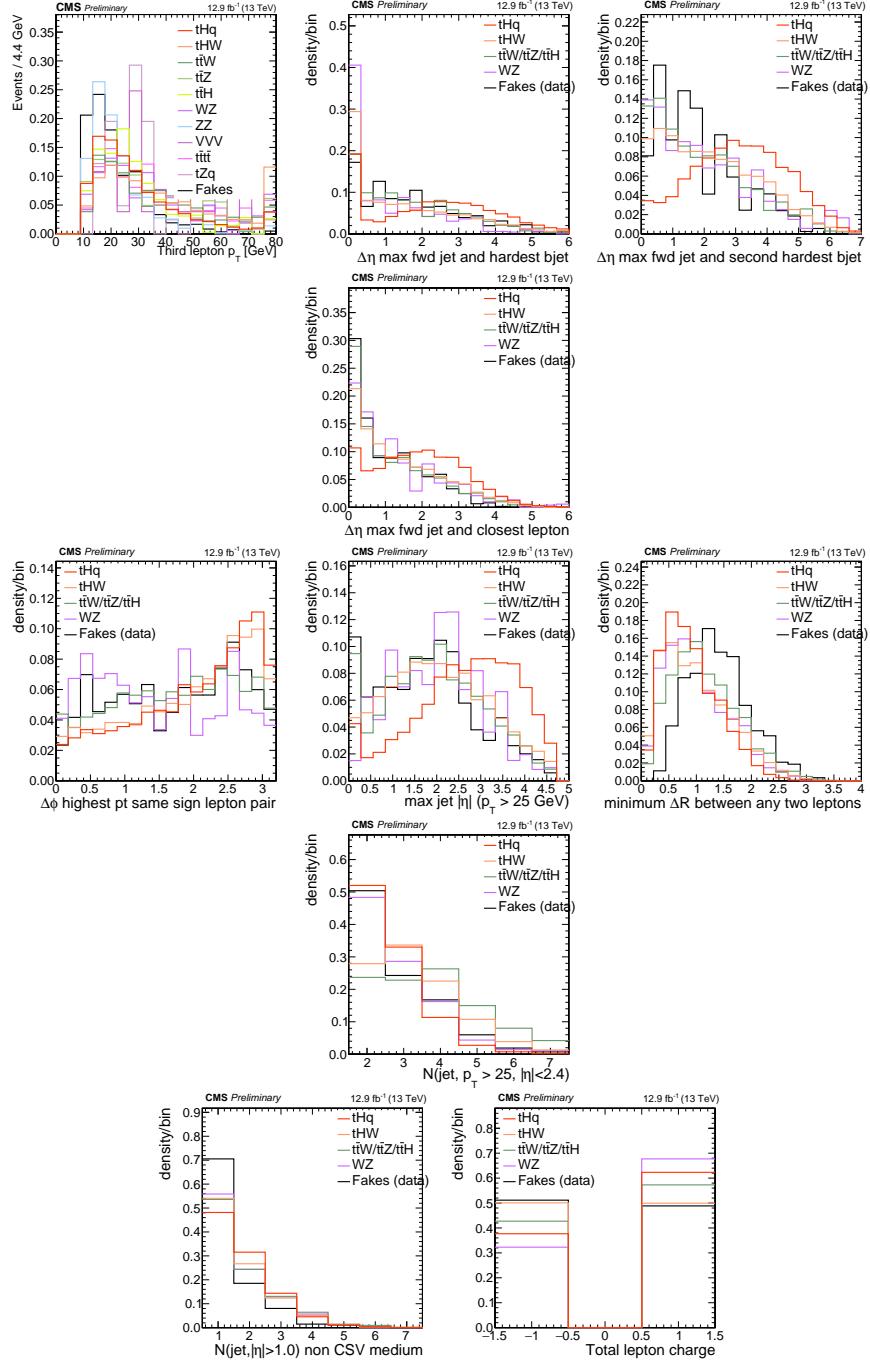
**Table 6.10:** MVA input discriminating variables

2522 the distributions are similar in both background cases.

2523 From table 6.10, it is clear that the input variables are correlated to some extend.  
 2524 These correlations play an important role for some MVA methods like the Fisher  
 2525 discriminant method in which the first step consist of performing a linear transfor-  
 2526 mation to an phase space where the correlations between variables are removed. In  
 2527 case a boosted decision tree (BDT) method however, correlations do not affect the  
 2528 performance. Figure 6.9 show the linear correlation coefficients for signal and back-  
 2529 ground for the two training cases (the signal values are identical by construction). As  
 2530 expected, strong correlations appears for variables related to the forward jet activity.  
 2531 Same trend is seen in case of the same sign dilepton channel in Figure ??.

### 2532 **6.8.1 Classifiers response**

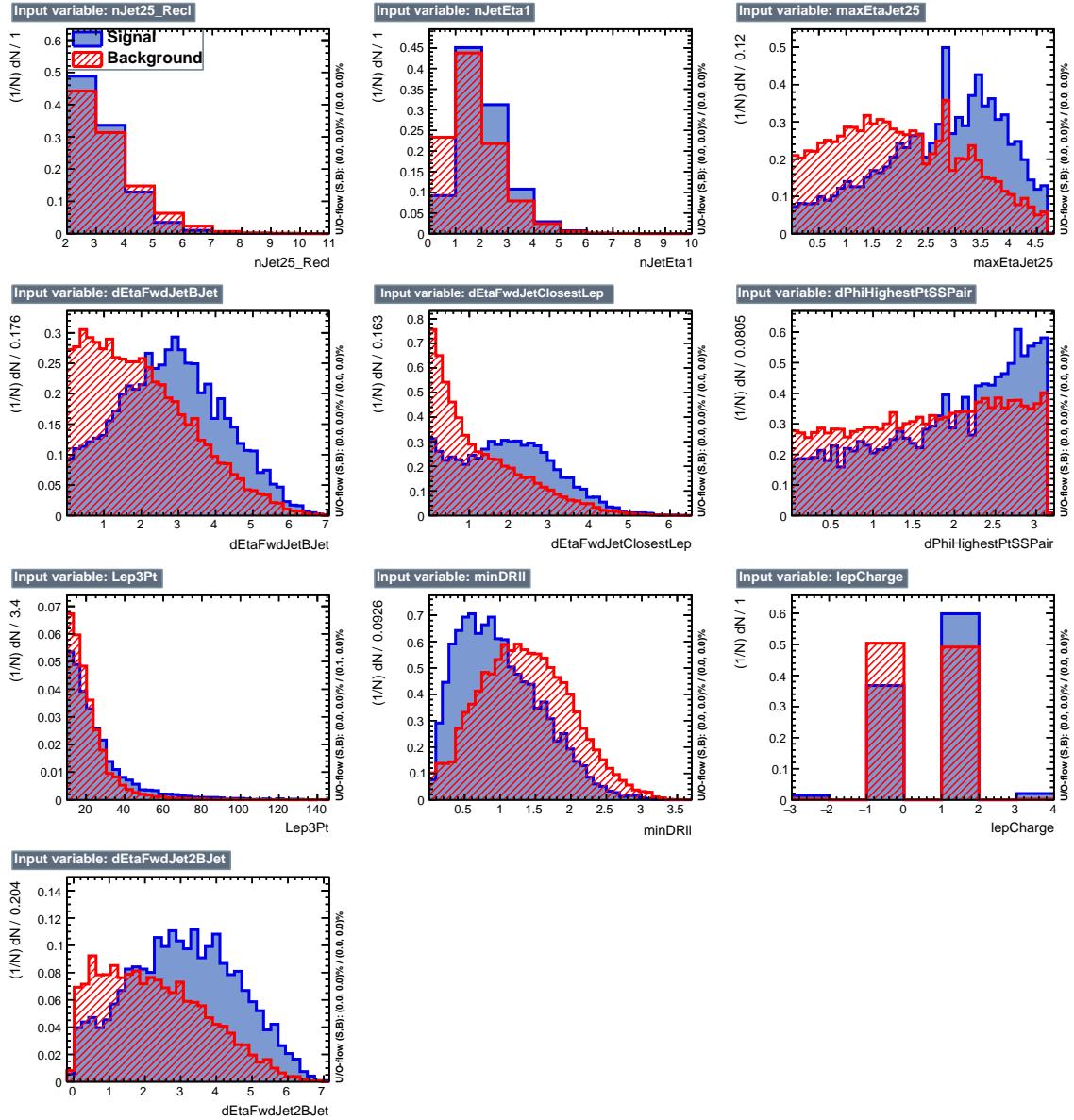
2533 Several MVA algorithms were evaluated to determine the most appropriate method  
 2534 for this analysis. The plots in Fig. 6.10 (top) show the background rejection as a  
 2535 function of the signal efficiency for  $t\bar{t}$  and  $t\bar{t}V$  trainings (ROC curves) for the different



**Figure 6.6:** Distributions of input variables to the BDT for signal discrimination, three lepton channel.

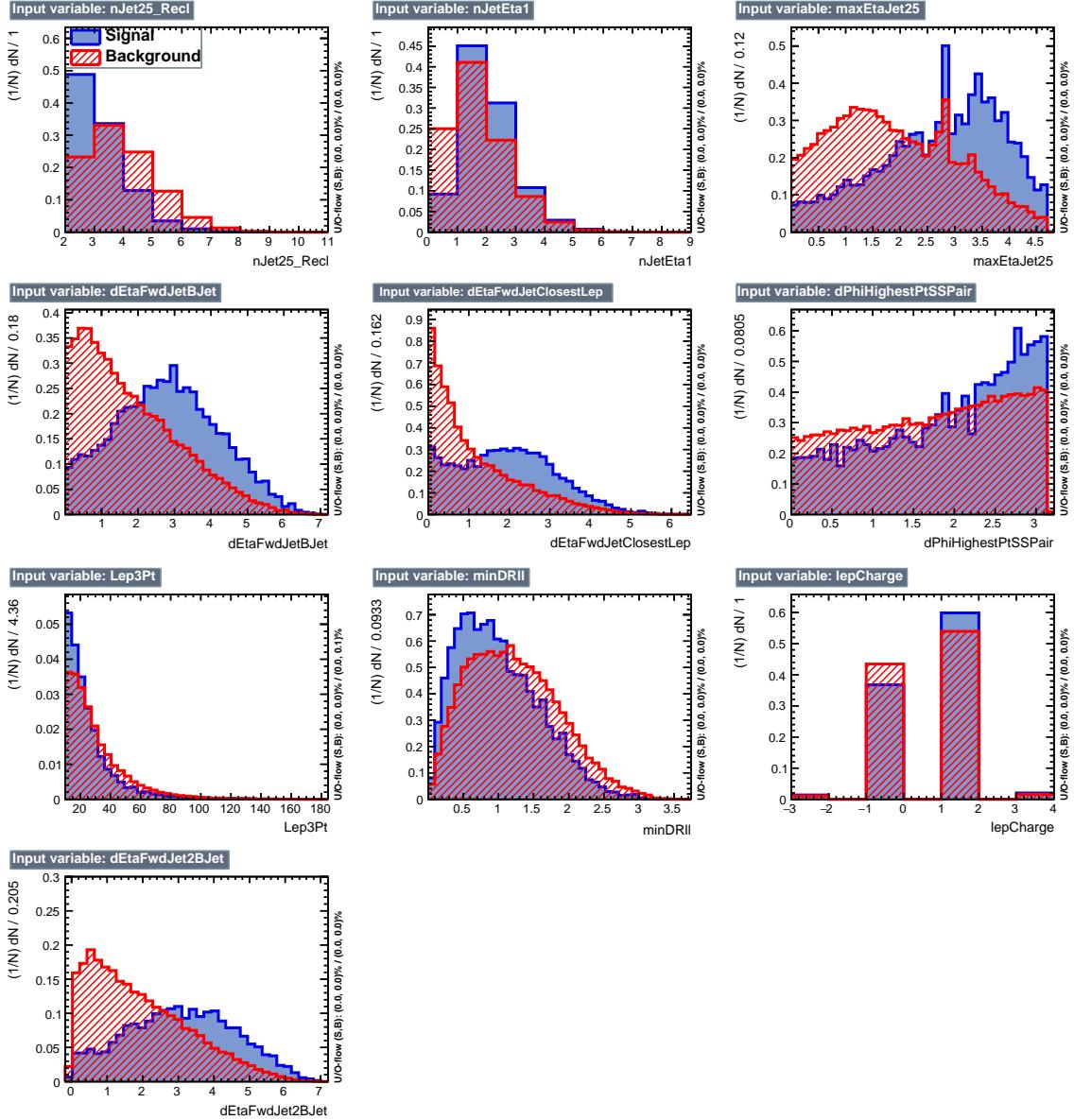
2536 algorithms that were evaluated.

2537 In both cases the gradient boosted decision tree (“BDTA\_GRAD”) classifier offers

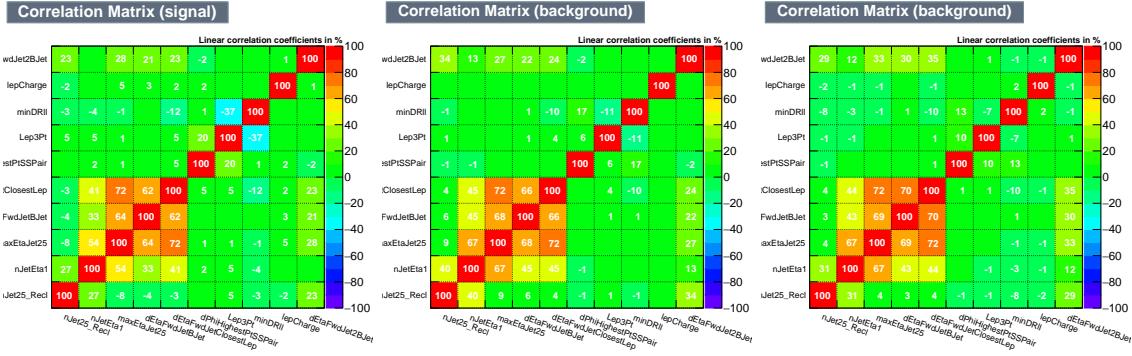


**Figure 6.7:** BDT inputs as seen by TMVA (signal, in blue, is  $tHq$ , background, in red, is  $t\bar{t}$ ) for the three lepton channel, discriminated against  $t\bar{t}$  (fakes) background.

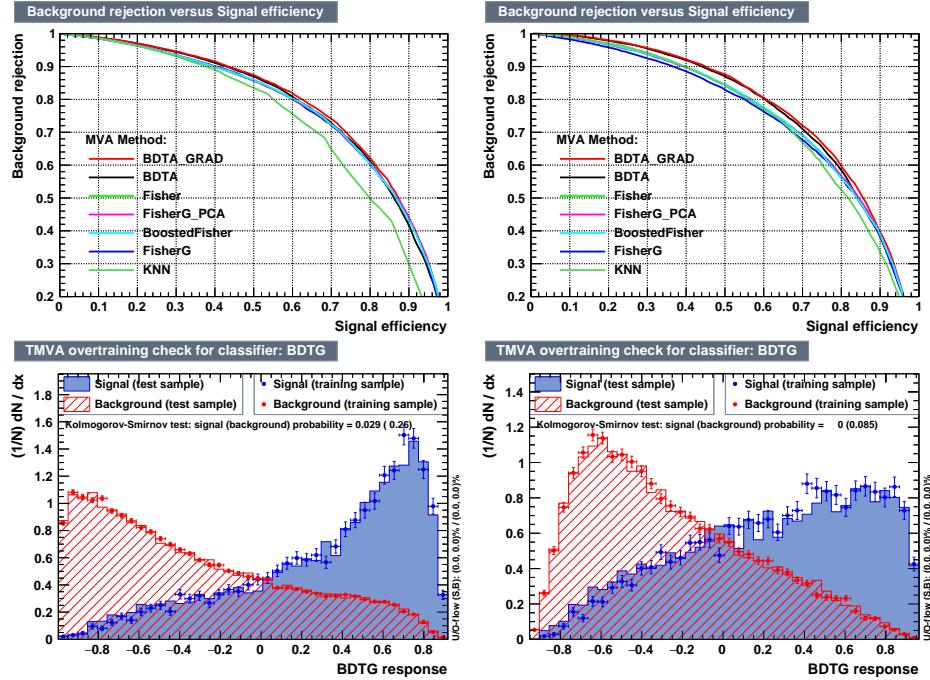
the best results, followed by an adaptive BDT classifier (“BDTA”). The BDTA\_GRAD classifier output distributions for signal and backgrounds are shown on the bottom of Fig. 6.10. As expected, a good discrimination power is obtained using default discriminator parameter values, with minimal overtraining. TMVA provides a ranking of the input variables by their importance in the classification process, shown in Tab. 6.11.



**Figure 6.8:** BDT inputs as seen by TMVA (signal, in blue, is  $tHq$ , background, in red, is  $t\bar{t}W+t\bar{t}Z$ ) for the three lepton channel, discriminated against  $t\bar{t}V$  background.



**Figure 6.9:** Signal (left),  $t\bar{t}$  background (middle), and  $t\bar{t}V$  background (right.) correlation matrices for the input variables in the TMVA analysis for the three lepton channel.



**Figure 6.10:** Top: background rejection vs signal efficiency (ROC curves) for various MVA classifiers (top) in the three lepton channel against  $t\bar{t}V$  (left) and  $t\bar{t}$  (right). Bottom: classifier output distributions for the gradient boosted decision trees, for training against  $t\bar{t}V$  (left) and against  $t\bar{t}$  (right).

## 2544 6.9 Additional discriminating variables

2545 Two additional discriminating variables were tested considering the fact that the  
 2546 forward jet in the background could come from the pileup; since we have a real forward

ttbar training			ttV training		
Rank	Variable	Importance	Variable	Importance	
1	minDRll	1.329e-01	dEtaFwdJetBJet	1.264e-01	
2	dEtaFwdJetClosestLep	1.294e-01	Lep3Pt	1.224e-01	
3	dEtaFwdJetBJet	1.209e-01	maxEtaJet25	1.221e-01	
4	dPhiHighestPtSSPair	1.192e-01	dEtaFwdJet2BJet	1.204e-01	
5	Lep3Pt	1.158e-01	dEtaFwdJetClosestLep	1.177e-01	
6	maxEtaJet25	1.121e-01	minDRll	1.143e-01	
7	dEtaFwdJet2BJet	9.363e-02	dPhiHighestPtSSPair	9.777e-02	
8	nJetEta1	6.730e-02	nJet25_Recl	9.034e-02	
9	nJet25_Recl	6.178e-02	nJetEta1	4.749e-02	
10	lepCharge	4.701e-02	lepCharge	4.116e-02	

**Table 6.11:** TMVA input variables ranking for BDTA\_GRAD method for the trainings in the three lepton channel. For both trainings the rankings show almost the same 5 variables in the first places.

---

```

TMVA.Types.kBDT
NTrees=800
BoostType=Grad
Shrinkage=0.10
!UseBaggedGrad
nCuts=50
MaxDepth=3
NegWeightTreatment=PairNegWeightsGlobal
CreateMVAPdfs

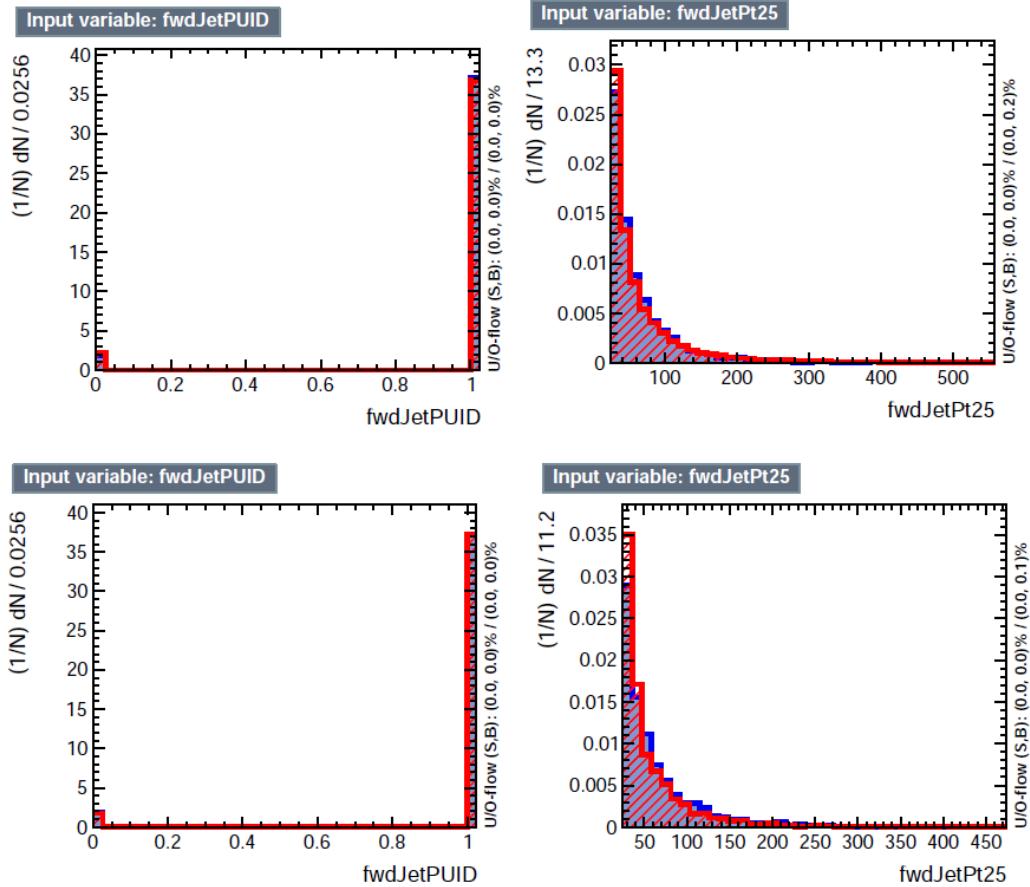
```

---

**Table 6.12:** TMVA configuration used in the BDT training.

jet in the signal, it could give some improvement in the discriminating power. The additional variables describe the forward jet momentum (fwdJetPt25) and the forward jet identification(fwdJetPUID). Distributions for these variables in the three lepton channel are shown in the figure 6.11. The forward jet identification distribution show that for both, signal and background, jets are mostly real jets.

The testing was made including in the MVA input one variable at a time, so we can evaluate the dicrimination power of each variable, and then both simultaneously. fwdJetPUID was ranked in the last place in importance (11) in both training (ttV and tt) while fwdJetPt25 was ranked 3 in the ttV training and 7 in the tt training. When training using 12 variables, fwdJetPt25 was ranked 5 and 7 in the ttV and tt



**Figure 6.11:** Additional discriminating variables distributions for ttv training (Top row) and tt training (bottom row) in the three lepton channel. The origin of the jets in the forward jet identification distribution is tagged as 0 for “pileup jets” while “real jets” are tagged as 1.

2557 trainings respectively, while `fwdJetPUID` was ranked 12 in both cases.

2558 The improvement in the discrimination performance provided by the additional  
 2559 variables is about 1%, so it was decided not to include them in the procedure. Table  
 2560 6.13 show the ROC-integral for all the testing cases we made.

ROC-integral	
base 10 var ttv	0.848
+ fwdJetPUID ttv	0.849
+ fwdJetPt25 ttv	0.856
12 var ttv	0.856
<hr/>	
base 10 var tt	0.777
+ fwdJetPUID tt	0.777
+ fwdJetPt25 tt	0.787
12 var	0.787

**Table 6.13:** ROC-integral for all the testing cases we made in the evaluation of the additional variables discriminating power. The improvement in the discrimination performance provided by the additional variables is about 1%

## 2561 References

- 2562 [1] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". Phys-  
2563       ical Review. 74 (10): 1439-61, (1948). <https://doi.org/10.1103/PhysRev.74.1439>
- 2565 [2] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". Physical  
2566       Review. 76 (6): 769-89, (1949). <https://doi.org/10.1103/PhysRev.76.769>
- 2567 [3] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum  
2568       Theory of Wave Fields". Progress of Theoretical Physics. 1 (2): 27-42, (1946).  
2569       <https://doi.org/10.1143/PTP.1.27>
- 2570 [4] D.J. Griffiths, "Introduction to electrodynamics". 4th ed. Pearson, (2013).
- 2571 [5] F. Mandl, G. Shaw. "Quantum field theory." Chichester, Wiley (2009).
- 2572 [6] F. Halzen, and A.D. Martin, "Quarks and leptons: An introductory course in  
2573       modern particle physics". New York: Wiley, (1984) .
- 2574 [7] File: Standard\_Model\_of\_Elementary\_Particle\_dark.svg. (2017, June 12)  
2575       Wikimedia Commons, the free media repository. Retrieved November 27, 2017  
2576       from <https://www.collegiate-advanced-electricity.com/single-post/2017/04/10/The-Standard-Model-of-Particle-Physics>.

- 2578 [8] E. Noether, "Invariante Variationsprobleme", Nachrichten von der Gesellschaft  
2579 der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, vol. 1918,  
2580 pp. 235-257, (1918).
- 2581 [9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)  
2582 and 2017 update.
- 2583 [10] M. Goldhaber, L. Grodzins, A.W. Sunyar "Helicity of Neutrinos", Phys. Rev.  
2584 109, 1015 (1958).
- 2585 [11] Palanque-Delabrouille N et al. "Neutrino masses and cosmology with Lyman-  
2586 alpha forest power spectrum", JCAP 11 011 (2015).
- 2587 [12] M. Gell-Mann. "A Schematic Model of Baryons and Mesons". Physics Letters.  
2588 8 (3): 214-215 (1964).
- 2589 [13] G. Zweig. "An SU(3) Model for Strong Interaction Symmetry and its Breaking"  
2590 (PDF). CERN Report No.8182/TH.401 (1964).
- 2591 [14] G. Zweig. "An SU(3) Model for Strong Interaction Symmetry and its Breaking:  
2592 II" (PDF). CERN Report No.8419/TH.412(1964).
- 2593 [15] M. Gell-Mann. "The Interpretation of the New Particles as Displaced Charged  
2594 Multiplets". Il Nuovo Cimento 4: 848. (1956).
- 2595 [16] T. Nakano, K. Nishijima. "Charge Independence for V-particles". Progress of  
2596 Theoretical Physics 10 (5): 581-582. (1953).
- 2597 [17] N. Cabibbo, "Unitary symmetry and leptonic decays" Physical Review Letters,  
2598 vol. 10, no. 12, p. 531, (1963).

- 2599 [18] M.Kobayashi, T.Maskawa, “CP-violation in the renormalizable theory of weak  
2600 interaction,” Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).
- 2601 [19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Com-  
2602 mons, the free media repository. Retrieved November 27, 2017  
2603 from [https://commons.wikimedia.org/w/index.php?title=File:  
2604 Weak\\_Decay\\_\(flipped\)\.svg&oldid=247498592](https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)\.svg&oldid=247498592).
- 2605 [20] Georgia Tech University. Coupling Constants for the Fundamental Forces(2005).  
2606 Retrieved January 10, 2018, from [http://hyperphysics.phy-astr.gsu.edu/  
2607 hbase/Forces/couple.html#c2](http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/couple.html#c2)
- 2608 [21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved Jan-  
2609 uary 10, 2018, from [https://profmattstrassler.com/articles-and-posts/  
2610 particle-physics-basics/the-known-forces-of-nature/  
2611 the-strength-of-the-known-forces/](https://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/)
- 2612 [22] S.L. Glashow. “Partial symmetries of weak interactions”, Nucl. Phys. 22 579-  
2613 588, (1961).
- 2614 [23] A. Salam, J.C. Ward. “Electromagnetic and weak interactions”, Physics Letters  
2615 13 168-171, (1964).
- 2616 [24] S. Weinberg, “A model of leptons”, Physical Review Letters, vol. 19, no. 21, p.  
2617 1264, (1967).
- 2618 [25] M. Peskin, D. Schroeder, “An introduction to quantum field theory”. Perseus  
2619 Books Publishing L.L.C., (1995).
- 2620 [26] A. Pich. “The Standard Model of Electroweak Interactions” <https://arxiv.org/abs/1201.0537>

- 2622 [27] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 122, 103 (1983).
- 2623 [28] M. Banner et al. (UA2 Collaboration), Phys. Lett. B 122, 476 (1983).
- 2624 [29] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 126, 398 (1983).
- 2625 [30] P. Bagnaia et al. (UA2 Collaboration), Phys. Lett. B 129, 130 (1983).
- 2626 [31] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved  
2627 from: <https://www.quantum-bits.org/?p=233>
- 2628 [32] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255
- 2629 [33] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge  
2630 Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964)  
2631 doi:10.1103/PhysRevLett.13.321
- 2632 [34] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Re-  
2633 view Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.
- 2634 [35] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws  
2635 and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964).  
2636 doi:10.1103/PhysRevLett.13.585.
- 2637 [36] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with  
2638 the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012).  
2639 arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021
- 2640 [37] ATLAS collaboration. “Observation of a New Particle in the Search for the Stan-  
2641 dard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters  
2642 B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

- 2643 [38] ATLAS collaboration; CMS collaboration (26 March 2015). “Combined Mea-  
 2644 surement of the Higgs Boson Mass in pp Collisions at  $\sqrt{s}=7$  and 8 TeV with  
 2645 the ATLAS and CMS Experiments”. Physical Review Letters. 114 (19): 191803.  
 2646 arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.
- 2647 [39] LHC InternationalMasterclasses“When protons collide”. Retrieved from [http://atlas.physicsmasterclasses.org/en/zpath\\_protoncollisions.htm](http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm)
- 2649 [40] CMS Collaboration, “SM Higgs Branching Ratios and Total Decay Widths (up-  
 2650 date in CERN Report4 2016)”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR> , last accessed on 17.12.2017.
- 2652 [41] R.Grant V. “Determination of Higgs branching ratios in  $H \rightarrow W^+W^- \rightarrow l\nu jj$   
 2653 and  $H \rightarrow ZZ \rightarrow l^+l^-jj$  channels”. Physics Department, University of Ten-  
 2654 nessee (Dated: October 31, 2012). Retrieved from <http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf>
- 2656 [42] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al.  
 2657 “Standard model Higgs-boson branching ratios with uncertainties”. Eur. Phys.  
 2658 J. C (2011) 71: 1753. <https://doi.org/10.1140/epjc/s10052-011-1753-8>
- 2659 [43] D. de Florian et al., LHC Higgs Cross Section Working Group,  
 2660 CERNâš2017âš002-M, arXiv:1610.07922[hep-ph] (2016).
- 2661 [44] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production  
 2662 of Higgs and single top at hadron colliders”, Phys.Rev. D64 (2001) 094023,  
 2663 [hep-ph/0106293].

- 2664 [45] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the  
 2665 top-Higgs coupling from the 8 TeV LHC data,” Journal of High Energy Physics,  
 2666 vol. 07, p. 073, (2013).
- 2667 [46] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting de-  
 2668 generacies in Higgs couplings using single top production in association with a  
 2669 Higgs boson,” Journal of High Energy Physics, vol. 05, p. 022, (2013).
- 2670 [47] T.M. Tait and C.-P. Yuan, “Single top quark production as a window to physics  
 2671 beyond the standard model”, Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].
- 2672 [48] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in  
 2673 association with a single top quark at the LHC,” European Physical Journal C,  
 2674 vol. 75, p. 267, (2015).
- 2675 [49] CMS Collaboration, “Modelling of the single top-quark production in associa-  
 2676 tion with the Higgs boson at 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV)  
 2677 [viewauth/CMS/SingleTopHiggsGeneration13TeV](#), last accessed on 16.01.2018.
- 2678 [50] CMS Collaboration, “SM Higgs production cross sections at  $\sqrt{s} =$   
 2679 13 TeV.” [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)  
 2680 [CERNYellowReportPageAt13TeV](#), last accessed on 16.01.2018.
- 2681 [51] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.
- 2682 [52] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [arXiv:1211.0499  
 2683 [hep-ph]].
- 2684 [53] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated  
 2685 production at the LHC”, European Physical Journal C, vol. 77, p. 34, (2017).  
 2686 arXiv:1607.05862

- 2687 [54] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross  
2688 Sections: 4.Deciphering the Nature of the Higgs Sector”, arXiv:1610.07922.
- 2689 [55] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top  
2690 Couplings in Associated Production”, JHEP 1404 (2014) 004, [arXiv:1312.5736].
- 2691 [56] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass  
2692 of the Higgs boson and tests of compatibility of its couplings with the standard  
2693 model predictions using proton collisions at 7 and 8 TeV,” arXiv:1412.8662.
- 2694 [57] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the  
2695 Higgs boson with the ATLAS detector using up to  $25 \text{ fb}^{-1}$  of proton-proton  
2696 collision data”, ATLAS-CONF-2014-009.
- 2697 [58] ATLAS and CMS Collaborations, “Measurements of the Higgs boson produc-  
2698 tion and decay rates and constraints on its couplings from a combined ATLAS  
2699 and CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV,” (2016).  
2700 CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.
- 2701 [59] File:Cern-accelerator-complex.svg. Wikimedia Commons, the free media repos-  
2702 itory. Retrieved January, 2018 from <https://commons.wikimedia.org/wiki/File:Cern-accelerator-complex.svg>
- 2704 [60] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”,  
2705 (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Re-  
2706 trieved from <https://cds.cern.ch/record/841542>
- 2707 [61] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-  
2708 2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762
- 2709 [62] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).

- 2710 [63] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March  
2711 1983, CERN-AC-8303511. Retrieved from <https://cds.cern.ch/record/615852>.
- 2713 [64] CERN Photographic Service “Animation of CERN’s accelerator network”, 14  
2714 October 2013. DOI: 10.17181/cds.1610170 Retrieved from <https://videos.cern.ch/record/1610170>
- 2716 [65] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17,  
2717 2013. Retrieved from <https://www.britannica.com/technology/particle-accelerator>.
- 2719 [66] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-  
2720 AC-0007016. Retrieved from <https://cds.cern.ch/record/41567>.
- 2721 [67] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”.  
2722 March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-  
2723 PHO-1998-325. Retrieved from <https://cds.cern.ch/record/841511>.
- 2724 [68] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025  
2725 retrieved from <https://cds.cern.ch/record/40524>.
- 2726 [69] CMS Collaboration “Public CMS Luminosity Information”. [https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016\\_proton\\_proton\\_13\\_TeV\\_collis](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_proton_proton_13_TeV_collis), last accessed 24.01.2018
- 2729 [70] J.L Caron. “LHC Layout” AC Collection. Legacy of AC. Pictures from 1992  
2730 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from <https://cds.cern.ch/record/841573>.

- 2732 [71] J.A. Coarasa. “The CMS Online Cluster:Setup, Operation and Maintenance  
2733 of an Evolving Cluster”. ISGC 2012, 26 February - 2 March 2012, Academia  
2734 Sinica, Taipei, Taiwan.
- 2735 [72] CMS Collaboration. “The CMS experiment at the CERN LHC” JINST 3 S08004  
2736 (2008).
- 2737 [73] CMS Collaboration. “CMS detector drawings 2012” CMS-PHO-GEN-2012-002.  
2738 Retrieved from <http://cds.cern.ch/record/1433717>.
- 2739 [74] Davis, Siona Ruth. “Interactive Slice of the CMS detector”, Aug. 2016,  
2740 CMS-OUTREACH-2016-027, retrieved from [https://cds.cern.ch/record/](https://cds.cern.ch/record/2205172)  
2741 2205172
- 2742 [75] R. Breedon. “View through the CMS detector during the cooldown of the  
2743 solenoid on February 2006. CMS Collection”, February 2006, CMS-PHO-  
2744 OREACH-2005-004, Retrieved from <https://cds.cern.ch/record/930094>.
- 2745 [76] Halyo, V. and LeGresley, P. and Lujan, P. “Massively Parallel Computing and  
2746 the Search for Jets and Black Holes at the LHC”, Nucl.Instrum.Meth. A744  
2747 (2014) 54-60, DOI: 10.1016/j.nima.2014.01.038”
- 2748 [77] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector  
2749 Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.
- 2750 [78] CMS Collaboration. “Description and performance of track and primary-vertex  
2751 reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no.  
2752 10, p. P10009,(2014).

- 2753 [79] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner Bar-  
2754 rel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from <https://cds.cern.ch/record/1431467>.  
2755
- 2756 [80] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm  
2757 and beauty hadrons Chicago, June 28-July 3 2004.  
2758
- 2759 [81] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at  
the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013.  
2760 arXiv:1307.7135
- 2761 [82] L. Veillet. “End assembly of HB with EB rails and rotation inside SX ”,Jan-  
2762 uary 2002. CMS-PHO-HCAL-2002-002. Retrieved from <https://cds.cern.ch/record/42594>.  
2763
- 2764 [83] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the UX5 cav-  
2765 ern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from <https://cds.cern.ch/record/1019185>  
2766
- 2767 [84] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web site:  
2768 Taking a closer look at LHC. Retrieved from [https://www.lhc-closer.es/taking\\_a\\_closer\\_look\\_at\\_lhc/0.lhc\\_trigger](https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger)  
2769
- 2770 [85] WLCG Project Office, “Documents & Reference - Tiers - Structure,”  
2771 (2014). <http://wlcg.web.cern.ch/documents-reference> , last accessed on  
2772 30.01.2018.  
2773
- 2774 [86] CMS Collaboration. “CMSSW Application Framework”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>,  
2775 last accesses 06.02.2018

- 2776 [87] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event gen-  
2777 erators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011
- 2778 [88] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics  
2779 of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-  
2780 3-540-71060-8 (2007)
- 2781 [89] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter”  
2782 <http://hepdata.cedar.ac.uk/pdf/pdf3.html> , last accessed on 02.02.2018.
- 2783 [90] G. Altarelli and G. Parisi. “ASYMPTOTIC FREEDOM IN PARTON LAN-  
2784 GUAGE”, Nucl.Phys. B126:298 (1977).
- 2785 [91] Yu.L. Dokshitzer. Sov.Phys. JETP 46:641 (1977)
- 2786 [92] V.N. Gribov, L.N. Lipatov. “Deep inelastic e p scattering in perturbation the-  
2787 ory”, Sov.J.Nucl.Phys. 15:438 (1972)
- 2788 [93] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a  
2789 reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).
- 2790 [94] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, “Parton fragmen-  
2791 tation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145,  
2792 1983.
- 2793 [95] CMS Collaboration, “Event generator tunes obtained from underlying event  
2794 and multiparton scattering measurements;” European Physical Journal C, vol.  
2795 76, no. 3, p. 155, (2016).
- 2796 [96] J. Alwall et. al., “The automated computation of tree-level and next-to-leading  
2797 order differential cross sections, and their matching to parton shower simula-  
2798 tions,” Journal of High Energy Physics, vol. 07, p. 079, (2014).

- 2799 [97] T. Sjöstrand and P. Z. Skands, “Transverse-momentum-ordered showers and  
 2800 interleaved multiple interactions,” European Physical Journal C, vol. 39, pp.  
 2801 129–154, (2005).
- 2802 [98] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with  
 2803 Parton Shower simulations: the POWHEG method,” Journal of High Energy  
 2804 Physics, vol. 11, p. 070, (2007).
- 2805 [99] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments  
 2806 and Methods in Physics, vol. A506, pp. 250–303, (2003).
- 2807 [100] J. Allison et.al., “Recent developments in Geant4”, Nuclear Instruments and  
 2808 Methods in Physics Research A 835 (2016) 186-225.
- 2809 [101] CMS Collaboration “Full Simulation Offline Guide”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation>, last accessed 04.02.2018
- 2811 [102] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf.  
 2812 Ser. 513 022012 (2014)
- 2813 [103] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description  
 2814 with the CMS detector”, JINST 12 P10003 (2017) <https://doi.org/10.1088/1748-0221/12/10/P10003>.
- 2816 [104] The CMS Collaboration. “ Description and performance of track and pri-  
 2817 mary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014).  
 2818 doi:10.1088/1748-0221/9/10/P10009
- 2819 [105] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides.  
 2820 Retrieved from [https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS\\_4July2012\\_Final.pdf](https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf)

- 2822 [106] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by  
2823 the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).
- 2824 [107] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of  
2825 electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf  
2826 C 0303241 (2003) TULT009 [physics/0306087].
- 2827 [108] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification,  
2828 Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.
- 2829 [109] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”,  
2830 CMS Note 2007-008 (2007).
- 2831 [110] CMS collaboration, “Performance of CMS muon reconstruction in pp collision  
2832 events at  $\sqrt{s} = 7 \text{ TeV}$  ”, JINST 7 P10002 2012, [arXiv:1206.4071].
- 2833 [111] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,”  
2834 Journal of High Energy Physics, vol. 04, p. 063, (2008).
- 2835 [112] B. Dorney. “Anatomy of a Jet in CMS”. Quantum Diaries. June  
2836 1st, 2011. Retrieved from <https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>
- 2838 [113] The CMS Collaboration.“Event Displays from the high-energy collisions at 7  
2839 TeV”, May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from <https://cds.cern.ch/record/1429614>.
- 2841 [114] The CMS collaboration. “Determination of jet energy calibration and transverse  
2842 momentum resolution in CMS”. JINST 6 P11002 (2011). <http://dx.doi.org/10.1088/1748-0221/6/11/P11002>

- 2844 [115] The CMS Collaboration, “Introduction to Jet Energy Corrections at  
2845 CMS.”. <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>, last ac-  
2846 cessed 10.02.2018.
- 2847 [116] CMS Collaboration Collaboration. “Identification of b quark jets at the CMS  
2848 Experiment in the LHC Run 2”. Tech. rep. CMS-PAS-BTV-15-001. Geneva:  
2849 CERN, (2016). <https://cds.cern.ch/record/2138504>.
- 2850 [117] CMS Collaboration Collaboration. “Performance of missing energy reconstruc-  
2851 tion in 13 TeV pp collision data using the CMS detector”. Tech. rep. CMS-PAS-  
2852 JME16-004. Geneva: CERN, 2016. <https://cds.cern.ch/record/2205284>.
- 2853 [118] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,  
2854 Retrieved from <http://cms.web.cern.ch/sites/cms.web.cern.ch/files/>  
2855 [styles/large/public/field/image/HIG13004\\_Event01\\_0.png?itok=LAwZzPHR](http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/HIG13004_Event01_0.png?itok=LAwZzPHR)
- 2856
- 2857 [119] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,  
2858 Retrieved from [http://cms.web.cern.ch/sites/cms.web.cern.ch/](http://cms.web.cern.ch/sites/cms.web.cern.ch/files/)  
2859 [files/styles/large/public/field/image/TOP12035\\_Event01.png?itok=uMdnSqzC](http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/TOP12035_Event01.png?itok=uMdnSqzC)
- 2860
- 2861 [120] K. Skovpen. “Event displays highlighting the main properties of heavy flavour  
2862 jets in the CMS Experiment”, Aug 2017, CMS-PHO-EVENTS-2017-006. Re-  
2863 trieved from <https://cds.cern.ch/record/2280025>.
- 2864 [121] G. Cowan. “Topics in statistical data analysis for high-energy physics”.  
2865 arXiv:1012.3589v1

- 2866 [122] A. Hoecker et al., “TMVA-Toolkit for multivariate data analysis”  
2867 arXiv:physics/0703039v5 (2009)
- 2868 [123] L. Lista. “Statistical Methods for Data Analysis in Particle Physics”, 2nd  
2869 ed. Springer International Publishing. (2017) <https://dx.doi.org/10.1007/978-3-319-62840-0>
- 2871 [124] I. Antcheva et al., “ROOT-A C++ framework for petabyte data storage, sta-  
2872 tistical analysis and visualization ,” Computer Physics Communications, vol.  
2873 182, no. 6, pp. 1384–1385, (2011).
- 2874 [125] Y. Coadou. “Boosted decision trees”, ESIPAP, Archamps, 9 Febru-  
2875 ary 2016. Lecture. Retrieved from [https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP\\_MVA160208-BDT.pdf](https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf)
- 2878 [126] J.H. Friedman. “Greedy function approximation: A gradient boosting ma-  
2879 chine”. Ann. Statist. Volume 29, Number 5 (2001), 1189-1232. [https://projecteuclid.org/download/pdf\\_1/euclid-aos/1013203451](https://projecteuclid.org/download/pdf_1/euclid-aos/1013203451).
- 2881 [127] F. James, M. Roos, “MINUIT: Function minimization and error analysis”. Cern  
2882 Computer Centre Program Library, Geneve Long Write-up No. D506, 1989
- 2883 [128] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests  
2884 of statistical hypotheses”. Philosophical Transactions of the Royal Society of  
2885 London. Series A, Containing Papers of a Mathematical or Physical Character.  
2886 Vol. 231 (1933), pp. 289-337

- 2887 [129] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated pro-  
 2888       duction via gluon fusion in the SM and the 2HDM”, JHEP 06 (2015) 065,  
 2889       [https://dx.doi.org/10.1007/JHEP06\(2015\)065](https://dx.doi.org/10.1007/JHEP06(2015)065), arXiv:1503.01656.
- 2890 [130] ATLAS Collaboration, “Measurements of Higgs boson produc-  
 2891       tion and couplings in diboson final states with the ATLAS  
 2892       detector at the LHC”, Phys. Lett. B726 (2013) 88–119,  
 2893       doi:10.1016/j.physletb.2014.05.011,10.1016/j.physletb.2013.08.010,  
 2894       arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].
- 2895 [131] CMS Collaboration, “Search for the associated production of a Higgs boson  
 2896       with a single top quark in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, JHEP 06  
 2897       (2016) 177,doi:10.1007/JHEP06(2016)177, arXiv:1509.08159.
- 2898 [132] B. Stieger, C. Jordà Lope et al., “Search for Associated Production of a Single  
 2899       Top Quark and a Higgs Boson in Leptonic Channels”, CMS Analysis Note CMS  
 2900       AN-14-140, 2014.
- 2901 [133] M. Peruzzi, C. Mueller, B. Stieger et al., “Search for ttH in multilepton final  
 2902       states at  $\sqrt{s} = 13$  TeV”, CMS Analysis Note CMS AN-16-211, 2016.
- 2903 [134] CMS Collaboration, “Search for H to bbar in association with a single top quark  
 2904       as a test of Higgs boson couplings at  $\sqrt{s} = 13$  TeV”, CMS Physics Analysis  
 2905       Summary CMS-PAS-HIG-16-019, 2016.
- 2906 [135] CMS Collaboration, “Search for production of a Higgs boson and a single top  
 2907       quark in multilepton final states in proton collisions at  $\sqrt{s} = 13$  TeV”, CMS  
 2908       Physics Analysis Summary CMS-PAS-HIG-17-005, 2016.

- 2909 [136] B. WG, “BtagRecommendation80XReReco”, February, 2017. <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>.
- 2910
- 2911 [137] CMS Collaboration, “Identification of b quark jets at the CMS Experiment  
2912 in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001,  
2913 2016.
- 2914 [138] M. Peruzzi, F. Romeo, B. Stieger et al., “Search for ttH in multilepton final  
2915 states at  $\sqrt{s} = 13$  TeV”, CMS Analysis Note CMS AN-17-029, 2017.
- 2916 [139] CMS Collaboration, “Baseline muon selections for Run-II.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonIdRun2>, last accessed on  
2917 24.02.2018.
- 2918
- 2919 [140] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January,  
2920 2015. <https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf>.
- 2921
- 2922 [141] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”,  
2923 October, 2014. <https://indico.cern.ch/event/298249/contribution/3/material/slides/0.pdf>.
- 2924
- 2925 [142] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic  
2926 Top Quarks at the LHC”, JHEP 03 (2011) 059, [https://dx.doi.org/10.1007/JHEP03\(2011\)059](https://dx.doi.org/10.1007/JHEP03(2011)059), arXiv:1007.2221.
- 2927
- 2928 [143] CMS Collaboration. “Tag and Probe”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe>, last accessed on 02.03.2018.
- 2929
- 2930 [144] B. Maier, “SingleTopHiggProduction13TeV”, February, 2016. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV>.
- 2931