

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

21 Table of Contents

22 Table of Contents	iii
23 List of Figures	vii
24 List of Tables	x
25 1 INTRODUCTION	1
26 2 Theoretical approach	2
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	9
32 2.2.2 Fundamental interactions	13
33 2.2.3 Gauge invariance.	16
34 2.2.4 Gauge bosons	18
35 2.3 Electroweak unification and the Higgs mechanism	20
36 2.3.1 Spontaneous symmetry breaking (SSB)	28
37 2.3.2 Higgs mechanism	32

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	Associated production of a Higgs boson and a single Top quark.	42
44	2.5	The CP-mixing in tH processes	47
45	2.6	Experimental status of the anomalous Higgs-fermion coupling.	51
46	3	The CMS experiment at the LHC	53
47	3.1	Introduction	53
48	3.2	The LHC	54
49	3.3	The CMS experiment	64
50	3.3.1	Coordinate system	66
51	3.3.2	Pixels detector	67
52	3.3.3	Silicon strip tracker	69
53	3.3.4	Electromagnetic calorimeter	71
54	3.3.5	Hadronic calorimeter	72
55	3.3.6	Superconducting solenoid magnet	74
56	3.3.7	Muon system	75
57	3.3.8	CMS trigger system	76
58	3.3.9	CMS computing	78
59	4	Event generation, simulation and reconstruction	82
60	4.1	Event generation	83
61	4.2	Monte Carlo Event Generators.	87
62	4.3	CMS detector simulation.	88

63	4.4 Event reconstruction	90
64	4.4.1 Particle-Flow Algorithm.	90
65	4.4.1.1 Missing transverse energy.	102
66	4.4.2 Event reconstruction examples	103
67	5 Statistical methods	106
68	5.1 Multivariate analysis	106
69	5.1.1 Decision trees	109
70	5.1.2 Boosted decision trees (BDT).	112
71	5.1.3 Overtraining.	115
72	5.1.4 Variable ranking.	115
73	5.1.5 BDT output example.	116
74	5.2 Statistical inference.	117
75	5.2.1 Nuisance parameters.	117
76	5.2.2 Maximum likelihood estimation method	118
77	5.2.3 Hypothesis test	119
78	5.3 exclusion limits	120
79	5.4 asymptotic limits	120
80	6 Search for production of a Higgs boson and a single top quark in multilepton final states in pp collisions at $\sqrt{s} = 13$ TeV	121
82	6.1 Introduction	121
83	6.2 tHq signature	123
84	6.3 Background processes	125
85	6.4 Data and MC Samples	127
86	6.4.1 Full 2016 dataset and MC samples	127
87	6.4.2 Triggers	128

88	6.5 Object Identification	132
89	6.5.1 Jets and b tagging.	132
90	6.5.2 Missing Energy MET.	133
91	6.5.3 Lepton reconstruction and identification	134
92	6.5.4 Lepton selection efficiency	142
93	6.6 Event selection	144
94	6.7 Background predictions	145
95	6.8 Signal discrimination	147
96	6.8.1 Classifiers response	149
97	6.9 Additional discriminating variables	153
98	Bibliography	155
99	References	157

¹⁰⁰ 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	21
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	38
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 Associated Higgs boson production mechanism Feynman diagrams	43
114	2.14 Cross section for tHq process as a function of κ_t	45
115	2.15 Cross section for tHW process as a function of κ_{Htt}	46
116	2.16 NLO cross section for tX_0 and $t\bar{t}X_0$	49
117	2.17 NLO cross section for tWX_0 , $t\bar{t}X_0$	50

118	2.18 Two dimensional κ_t - κ_V plot of the coupling modifiers. ATLAS and CMS 119 combination.	51
120	3.1 CERN accelerator complex	54
121	3.2 LHC protons source. First acceleration stage.	55
122	3.3 The LINAC2 accelerating system at CERN.	56
123	3.4 LHC layout and RF cavities module.	57
124	3.5 LHC dipole magnet.	59
125	3.6 Integrated luminosity delivered by LHC and recorded by CMS during 2016	61
126	3.7 LHC interaction points	62
127	3.8 Multiple pp collision bunch crossing at CMS.	64
128	3.9 Layout of the CMS detector	65
129	3.10 CMS detector coordinate system	66
130	3.11 CMS pixel detector schematic view.	69
131	3.12 SST Schematic view.	70
132	3.13 CMS ECAL schematic view	71
133	3.14 CMS HCAL schematic view	73
134	3.15 CMS solenoid magnet	74
135	3.16 CMS Muon system schematic view	75
136	3.17 CMS Level-1 trigger architecture	77
137	3.18 WLCG structure	79
138	3.19 Data flow from CMS detector through hardware Tiers	81
139	4.1 Event generation process.	83
140	4.2 Particle flow algorithm.	91
141	4.3 Jet reconstruction.	99
142	4.4 Jet energy corrections.	100

143	4.5	Secondary vertex in a b-hadron decay.	101
144	4.6	HIG-13-004 Event 1 reconstruction.	103
145	4.7	$e\mu$ event reconstruction.	104
146	4.8	Recorded event reconstruction.	105
147	5.1	Scatter plots-MVA event classification.	108
148	5.2	Scalar test statistical.	108
149	5.3	Decision tree.	110
150	5.4	Decision tree output example.	113
151	5.5	BDT output example.	116
152	6.1	tHq event signature.	124
153	6.2	$t\bar{t}$ and $t\bar{t}W$ events signature.	126
154	6.3	Tight vs loose lepton selection efficiencies in the 2lss channel.	142
155	6.4	Tight vs loose lepton selection efficiencies in the 3l channel.	143
156	6.5	Input variables to the BDT for signal discrimination normalized.	148
157	6.6	Input variables to the BDT for signal discrimination not normalized.	150
158	6.7	BDT inputs as seen by TMVA against $t\bar{t}$.	151
159	6.8	BDT inputs as seen by TMVA against $t\bar{t}V$.	152
160	6.9	Correlation matrices for the input variables in the TMVA.	153
161	6.10	MVA classifiers performance.	153
162	6.11	Additional discriminating variables distributions.	155

¹⁶³ List of Tables

164	2.1	Fermions of the SM.	5
165	2.2	Fermion masses.	6
166	2.3	Lepton properties.	8
167	2.4	Quark properties.	9
168	2.5	Fermion weak isospin and weak hypercharge multiplets.	11
169	2.6	Fundamental interactions features.	15
170	2.7	SM gauge bosons.	19
171	2.8	Higgs boson properties.	37
172	2.9	Predicted branching ratios for a SM Higgs boson with $m_H = 125 \text{ GeV}/c^2$	42
173	2.10	Predicted SM cross sections for tH production at $\sqrt{s} = 13 \text{ TeV}$	44
174	2.11	Predicted enhancement of the tHq and tHW cross sections at LHC	47
175	6.1	Signal samples and their cross section and branching fraction.	128
176	6.2	κ_V and κ_t combinations.	129
177	6.3	List of background samples used in this analysis (CMSSW 80X).	130
178	6.4	Leading-order $t\bar{t}W$ and $t\bar{t}Z$ samples used in the signal BDT training.	130
179	6.5	Table of high-level triggers that we consider in the analysis.	131
180	6.6	Trigger efficiency scale factors and associated uncertainties.	132
181	6.7	Effective areas, for electrons and muons.	138

182	6.8 Requirements on each of the three muon selections.	141
183	6.9 Criteria for each of the three electron selections.	141
184	6.10 MVA input discriminating variables	149
185	6.11 TMVA input variables ranking for BDTA_GRAD method	154
186	6.12 TMVA configuration used in the BDT training.	154
187	6.13 ROC-integral for all the testing cases.	156

¹⁸⁸ Chapter 1

¹⁸⁹ INTRODUCTION

¹⁹⁰ **Chapter 2**

¹⁹¹ **Theoretical approach**

¹⁹² **2.1 Introduction**

¹⁹³ The physical description of the universe is a challenge that physicists have faced by
¹⁹⁴ making theories that refine existing principles and proposing new ones in an attempt
¹⁹⁵ to embrace emerging facts and phenomena.

¹⁹⁶ At the end of 1940s Julian Schwinger [1] and Richard P. Feynman [2], based on
¹⁹⁷ the work of Sin-Itiro Tomonaga [3], developed an electromagnetic theory consistent
¹⁹⁸ with special relativity and quantum mechanics that describes how matter and light
¹⁹⁹ interact; the so-called *quantum electrodynamics* (QED) was born.

²⁰⁰ QED has become the guide in the development of theories that describe the uni-
²⁰¹ verse. It was the first example of a quantum field theory (QFT), which is the theore-
²⁰² tical framework for building quantum mechanical models that describes particles and
²⁰³ their interactions. QFT is composed of a set of mathematical tools that combines
²⁰⁴ classical fields, special relativity and quantum mechanics, while keeping the quantum
²⁰⁵ point particles and locality ideas.

²⁰⁶ This chapter gives an overview of the standard model of particle physics, starting

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

211 2.2 Standard model of particle physics

212 Particle physics at the fundamental level is modeled in terms of a collection of inter-
 213 acting particles and fields in a theory known as the *standard model of particle physics*
 214 (*SM*). The full picture of the SM is composed of three fields¹ whose excitations are
 215 interpreted as particles called mediators or force-carriers, a set of fields whose excita-
 216 tions are interpreted as elementary particles interacting through the exchange of those
 217 mediators, and a field that gives the mass to elementary particles. Figure 2.1 shows
 218 the scheme of the SM particles’ organization. In addition, for each of the particles
 219 in the scheme there exists an antiparticle with the same mass and opposite quantum
 220 numbers. The existence of antiparticles is a prediction of the relativistic quantum
 221 mechanics from the solution of the Dirac equation for which a negative energy solu-
 222 tion is also possible. In some cases a particle is its own anti-particle, like photon or
 223 Higgs boson.

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

¹ 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

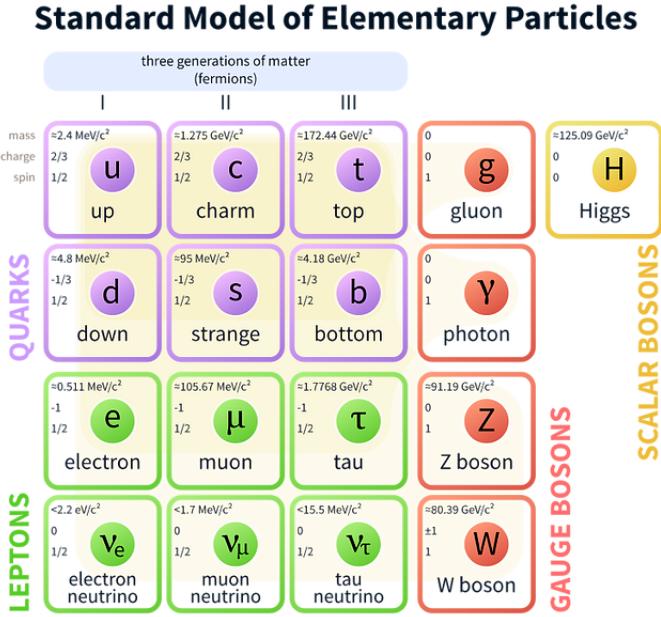


Figure 2.1: Schematic representation of the Standard Model of particle physics. The SM is a theoretical model intended to describe three of the four fundamental forces of the universe in terms of a set of particles and their interactions. [7].

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

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

243 **2.2.1 Fermions**

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

249

		Generation		
Type		1st	2nd	3rd
Leptons	Charged	Electron (e)	Moun(μ)	Tau (τ)
	Neutral	Electron neutrino (ν_e)	Muon neutrino (ν_μ)	Tau neutrino (ν_τ)
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.

250

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

256 behind this mass hierarchy is one of the most important open questions in particle
 257 physics, and it becomes more puzzling when noticing that the mass difference between
 258 first and second generation fermions is small compared to the mass difference with
 259 respect to the third generation.

Lepton	Mass (MeV/c ²)	Quark	Mass (MeV/c ²)
e	0.51	u	2.2
μ	105.65	c	1.28×10^3
τ	1776.86	t	173.1×10^3
ν_e	Unknown	d	4.7
ν_μ	Unknown	s	96
ν_τ	Unknown	b	4.18×10^3

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.

260

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

263 2.2.1.1 Leptons

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

271 Another feature of the leptons that is fundamental in the mathematical description
 272 of the SM is the chirality, which is closely related to spin and helicity. Helicity

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

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

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

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

² 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.

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

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

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

Lepton	Q(e)	T_3	L_e	L_μ	L_τ	Lifetime (s)
Electron (e)	-1	-1/2	1	0	0	Stable
Electron neutrino(ν_e)	0	1/2	1	0	0	Unknown
Muon (μ)	-1	-1/2	0	1	0	2.19×10^{-6}
Muon neutrino (ν_μ)	0	1/2	0	1	0	Unknown
Tau (τ)	-1	-1/2	0	0	1	290.3×10^{-15}
Tau neutrino (ν_τ)	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.

309 **2.2.1.2 Quarks**

310 Quarks are the basic constituents of protons and neutrons. The way quarks join to
 311 form bound states, called *hadrons*, is through the SI. Quarks are affected by all the
 312 fundamental interactions which means that they carry all the four types of charges:
 313 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.

314

315 Table 2.4 summarizes the features of quarks, among which the most remarkable
 316 is their fractional electric charge. Note that fractional charge is not a problem, given
 317 that quarks are not found isolated, but serves to explain how composed particles are
 318 formed out of two or more valence quarks³.

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

³ 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.

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

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

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

335 In the first version of the quark model (1964), M. Gell-Mann [12] and G. Zweig
 336 [13, 14] developed a consistent way to classify hadrons according to their properties.
 337 Only three quarks (u, d, s) were involved in a scheme in which all baryons have baryon
 338 number $B=1$ and therefore quarks have $B=1/3$; non-baryons have $B=0$. Baryon
 339 number is conserved in SI and EI which means that single quarks cannot be created
 340 but in pairs $q - \bar{q}$.

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

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

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

345 There are six quark flavors organized in three generations (see Table 2.1) fol-
 346 lowing a mass hierarchy which, again, implies that higher generations decay to first
 347 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	$4/3$	ν_{eR}	$\nu_{\mu R}$	$\nu_{\tau R}$		
	d'_R	s'_R	b'_R	0	$-2/3$	e_R	μ_R	τ_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.

348

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

355

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

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

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

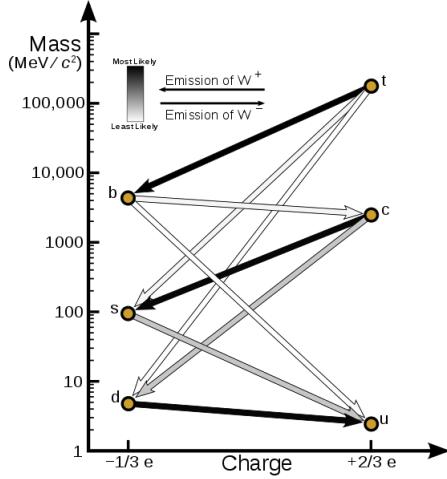


Figure 2.2: Transformations between quarks through the exchange of a WI. Higher generations quarks decay to first generation quarks by emitting a W boson. The arrow color indicates the likelihood of the transition according to the grey scale in the top left side which represent the CKM matrix parameters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

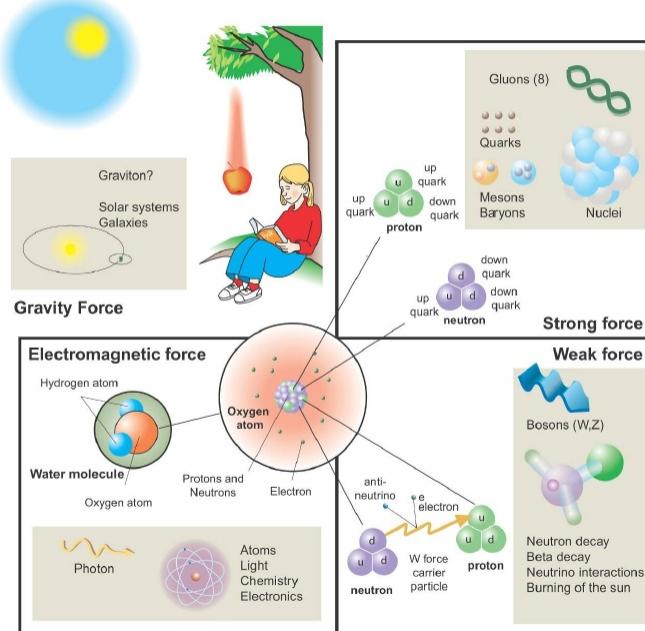
358 The weak decays of quarks are represented in the diagram of Figure 2.2; again
 359 the CKM matrix plays a central role since it contains the probabilities for the differ-
 360 ent quark decay channels, in particular, note that quark decays are greatly favored
 361 between generation members.

362

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

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.

369 2.2.2 Fundamental interactions

370 Even though there are many manifestations of force in nature, like the ones represented in Figure 2.3, we can classify all of them in four fundamental interactions:

- 372 • *Electromagnetic interaction (EI)* affects particles that are *electrically charged*,
373 like electrons and protons. Figure 2.4a. shows a graphical representation, known
374 as *Feynman diagram*, of electron-electron scattering.
- 375 • *Strong interaction (SI)* described by Quantum Chromodynamics (QCD). Hadrons
376 like the proton and the neutron have internal structure given that they are composed
377 of two or more valence quarks⁴. Quarks have fractional electric charge

⁴ Particles made of four and five quarks are exotic states not so common.

378 which means that they are subject to electromagnetic interaction and in the case
 379 of the proton they should break apart due to electrostatic repulsion; however,
 380 quarks are held together inside the hadrons against their electrostatic repulsion
 381 by the *Strong Force* through the exchange of *gluons*. The analog to the electric
 382 charge is the *color charge*. Electrons and photons are elementary particles as
 383 quarks but they don't carry color charge, therefore they are not subject to SI. A
 384 Feynman diagram for gluon exchange between quarks is shown in Figure 2.4b.

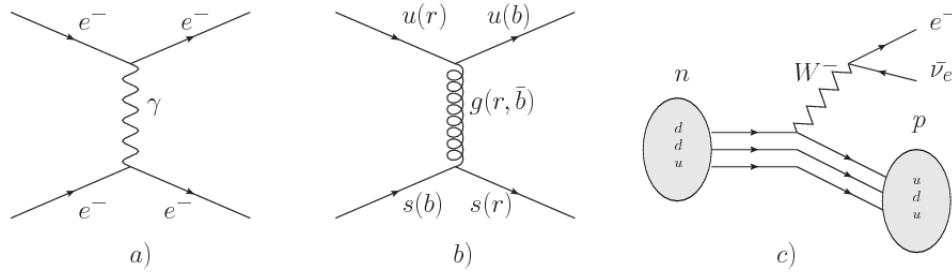


Figure 2.4: Feynman diagrams representing the interactions in SM; a) EI: e - e scattering;
 b) SI: gluon exchange between quarks ; c) WI: β -decay

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

396 • *Gravitational interaction (GI)* described by General Theory of Relativity (GR).

397 It is responsible for the structure of galaxies and black holes as well as the
 398 expansion of the universe. As a classical theory, in the sense that it can be
 399 formulated without even appeal to the concept of quantization, it implies that
 400 the space-time is a continuum and predictions can be made without limitation
 401 to the precision of the measurement tools. The latter represents a direct con-
 402 tradiction of the quantum mechanics principles. Gravity is deterministic while
 403 quantum mechanics is probabilistic; despite that, efforts to develop a quantum
 404 theory of gravity have predicted the *graviton* as mediator of the gravitational
 405 force⁵.

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].

406

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

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

415 than 10^{-15} m become negligible.

416

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
 420 by

- 421 • Lorentz invariance: independence on the reference frame.
- 422 • Locality: interacting fields are evaluated at the same space-time point to avoid
 action at a distance.
- 424 • Renormalizability: physical predictions are finite and well defined.
- 425 • Particle spectrum, symmetries and conservation laws already known must emerge
 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 (ΔV); 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^μ). 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^μ . 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 transformation
 456 invariance means that a gauge transformation is performed identically to every point

457 in the space⁶ and the Lagrangian remains invariant. Then, the global transformation
 458 is promoted to a local phase transformation (this time the gauge transformation
 459 depends on the position in space) and again invariance is required.

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

⁶ Here space corresponds to the 4-dimensional space i.e. space-time.

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 ²)
Electromagnetic	Photon (γ)	0	No	0	0
Strong	Gluon (g)	0	Yes -octet	No	0
Weak	W^\pm	± 1	No	± 1	80.385 ± 0.015
	Z	0	No	0	91.188 ± 0.002

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

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

512 The *classic* weak theory developed by Fermi, did not have the concept of the W
 513 boson but instead it was treated as a point interaction with the dimensionful constant
 514 G_F associated with it. It works really well at low energies very far off the W mass
 515 shell. When going up in energy, the theory of weak interactions involving the W
 516 boson is capable of explaining the β -decay and in general the processes mediated by
 517 W^\pm bosons. However, there were some processes like the $\nu_\mu - e$ scattering which
 518 would require the exchange of two W bosons (see Figure 2.5 top diagrams) giving
 519 rise to divergent loop integrals and then non-finite predictions. The EWI theory, by
 520 including neutral currents involving fermions via the exchange of a neutral bosons Z,
 521 overcomes those divergences and the predictions become realistic.

522 Neutral weak interaction vertices conserve flavor in the same way as the electro-
 523 magnetic vertices do, but additionally, the Z boson can couple to neutrinos which

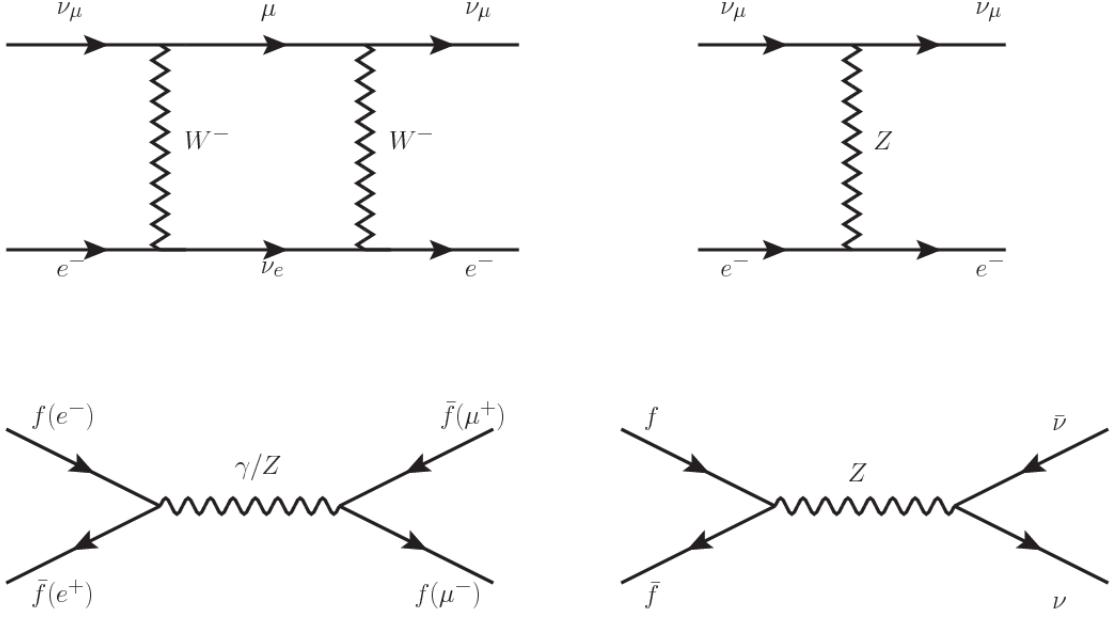


Figure 2.5: Top: $\nu_\mu - e^-$ scattering going through charged currents (left) and neutral currents (right). Bottom: neutral current processes for charged fermions (left) and involving neutrinos (right). While neutral current processes involving only charged fermions can proceed through EI or WI,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

524 implies that processes involving charged fermions can proceed through EI or WI but
 525 processes involving neutrinos can proceed only through WI.

526 The prescription to build a gauge theory of the WI consists of proposing a free
 527 field Lagrangian density that includes the particles involved; next, by requesting
 528 invariance under global phase transformations first and generalizing to local phase
 529 transformations invariance later, the conserved currents are identified and interactions
 530 are generated by introducing gauge fields. Given that the goal is to include the EI
 531 and WI in a single theory, the group symmetry considered should be a combination of
 532 $SU(2)_L$ and $U(1)_{em}$, however the latter cannot be used directly because the EI treats
 533 left and right-handed particles indistinctly in contrast to the former. Fortunately, the
 534 weak hypercharge, which is a combination of the weak isospin and the electric charge
 535 (Eqn. 2.2) is suitable to be used since it is conserved by the EI and WI. Thus, the

536 symmetry group to be considered is

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

537 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

540 Given the first generation of leptons

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

541 the charged fermionic currents are given by

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

542 and the free Lagrangian is given by

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

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

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

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

548

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

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

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

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

557 with σ_i the Pauli matrices and y_i the weak hypercharges. In order to promote the
 558 transformations from global to local while keeping the invariance, it is required that
 559 $\alpha^i = \alpha^i(x)$, $\beta = \beta(x)$ and the replacement of the ordinary derivatives by the covariant
 560 derivatives

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

576 In order to evaluate the electroweak interactions modeled by an isovector field
 577 W_μ^i which couples to isospin currents J_μ^i with strength g and additionally the singlet
 578 field B_μ which couples to the weak hypercharge current J_μ^Y with strength $g'/2$. The
 579 interaction Lagrangian density to be considered is

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

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

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

583 The same happens with the gauge fields W_μ^i which are related to the mass eigen-
 584 states W^\pm by

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

585 The fact that there are three weak isospin conserved currents is an indication that
 586 in addition to the charged fermionic currents, which couple charged to neutral leptons,
 587 there should be a neutral fermionic current that does not involve electric charge
 588 exchange; therefore, it couples neutral fermions or fermions of the same electric charge.
 589 The third weak isospin current contains a term that is similar to the electromagnetic
 590 current (j_μ^{em}), indicating that there is a relation between them and resembling the
 591 Gell-Mann-Nishijima formula 2.2 adapted to electroweak interactions

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

592 Just as Q generates the $U(1)_{em}$ symmetry, the weak hypercharge generates the
 593 $U(1)_Y$ symmetry as said before. It is possible to write the relationship in terms of
 594 the currents as

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

595 The neutral gauge fields W_μ^3 and B_μ cannot be directly identified with the Z
 596 and the photon fields since the photon interacts similarly with left and right-handed
 597 fermions; however, they are related through a linear combination given by

$$A_\mu = B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \quad (2.26)$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W$$

598 where θ_W is known as the *Weinberg angle*. The interaction Lagrangian is now given

599 by

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

600 the first term is the weak charged current interaction, while the second term is the
601 electromagnetic interaction under the condition

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

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

603

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

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

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

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

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

612 2.3.1 Spontaneous symmetry breaking (SSB)

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

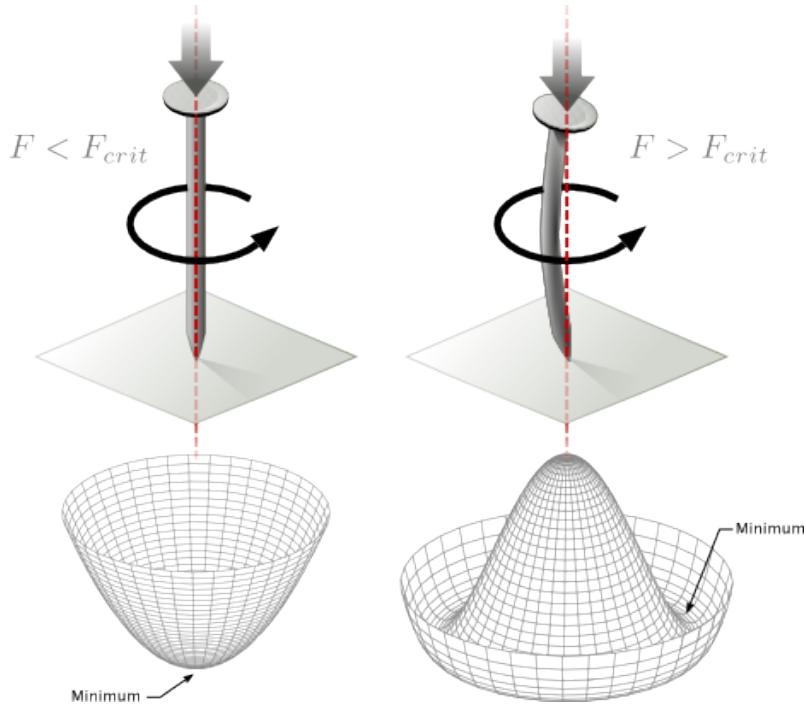


Figure 2.6: Spontaneous symmetry breaking mechanism. The steel nail, subject to an external force (top left), has rotational symmetry with respect to its axis. When the external force overcomes a critical value the nail buckles (top right) choosing a minimal energy state (ground state) and thus *breaking 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 [27].

615

616 Before reaching the critical force value, the system has rotational symmetry with
 617 respect to the nail axis; however, after the critical force value is reached the nail buck-
 618 les (top right). The form of the potential energy (bottom right) changes, preserving
 619 its rotational symmetry although its minima does not exhibit that rotational symme-
 620 try any longer. Right before the nail buckles there is no indication of the direction the

621 nail will bend because any of the directions are equivalent, but once the nail bends,
 622 choosing a direction, an arbitrary minimal energy state (ground state) is selected and
 623 it does not share the system's rotational symmetry. This mechanism for reaching an
 624 asymmetric ground state is known as *spontaneous symmetry breaking*.

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

627

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

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

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

632

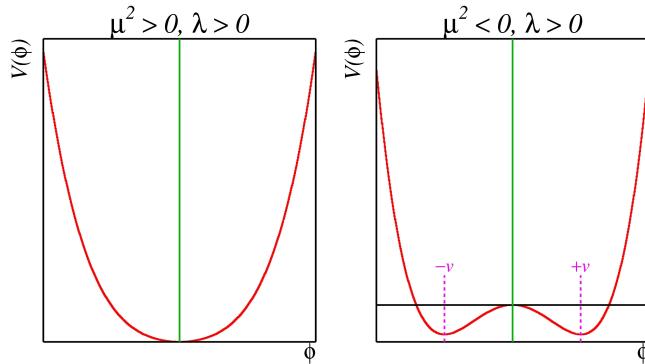


Figure 2.7: Shape of the potential $V(\phi)$ for $\lambda > 0$ and: $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The case $\mu^2 < 0$ corresponds to the potential suitable for introducing the SSB mechanism by choosing one of the two ground states which are connected via reflection symmetry. [27].

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

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

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

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

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

636

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

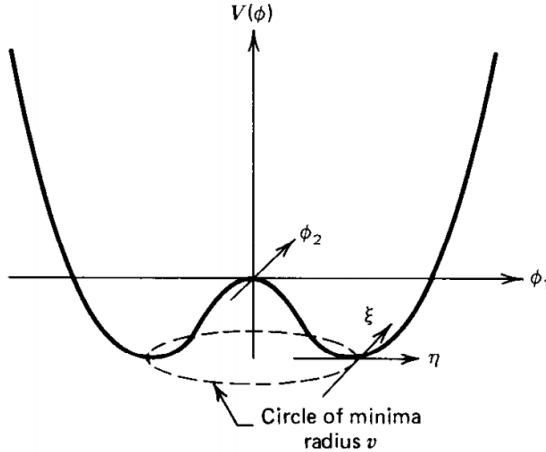


Figure 2.8: Potential for complex scalar field. There is a circle of minima of radius v along the ξ -direction [6].

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

640 As usual, excitations over the ground state are studied by making an expansion

about it; thus, the excitation can be parametrized as:

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

which when substituted into Eqn. 2.33 produces a Lagrangian in terms of the new fields η and ξ

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

where the last two terms represent the interactions and self-interaction between the two fields η and ξ . The particular feature of the SSB mechanism is revealed when looking to the first three terms of \mathcal{L}' . Before the SSB, only the massless ϕ field is present in the system; after the SSB there are two fields of which the η -field has acquired mass $m_\eta = \sqrt{-2\mu^2}$ while the ξ -field is still massless (see Figure 2.9).

649

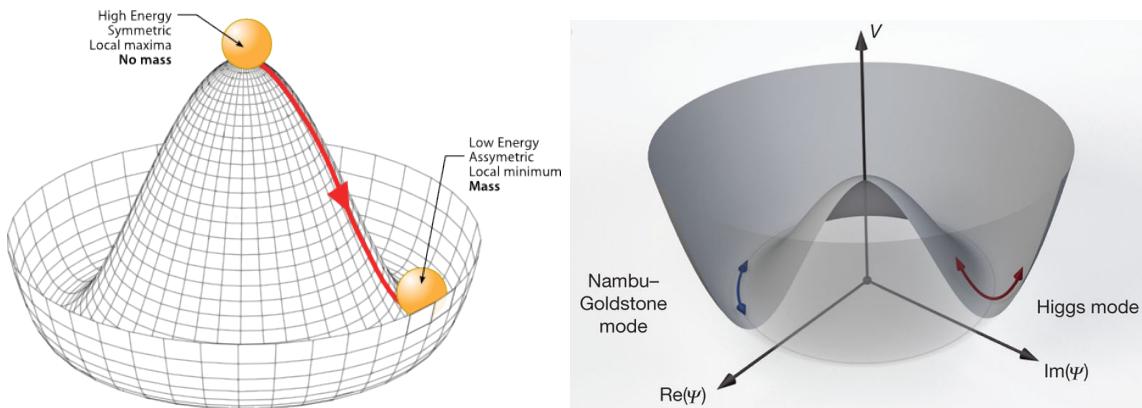


Figure 2.9: SSB mechanism for a complex scalar field [27, 28].

Thus, the SSB mechanism serves as a method to generate mass but as a side effect a massless field is introduced in the system. This fact is known as the Goldstone

theorem and states that a massless scalar field appears in the system for each continuous symmetry spontaneously broken. Another version of the Goldstone theorem states that “*if a Lagrangian is invariant under a continuous symmetry group G , but the vacuum is only invariant under a subgroup $H \subset G$, then there must exist as many massless spin-0 particles (Nambu-Goldstone bosons) as broken generators.*” [26] The Nambu-Goldstone boson can be understood considering that the potential in the ξ -direction is flat so excitations in that direction are not energy consuming and thus represent a massless state.

2.3.2 Higgs mechanism

When the SSB mechanism is introduced in the formulation of the EWI in an attempt to generate the mass of the so far massless gauge bosons and fermions, an interesting effect is revealed. In order to keep the G symmetry group invariance and generate the mass of the EW gauge bosons, a G invariant Lagrangian density (\mathcal{L}_S) has to be added to the non massive EWI Lagrangian (Eqn. 2.30)

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

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

ϕ has to be an isospin doublet of complex scalar fields so it preserves the G invariance; thus ϕ can be defined as:

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

The minima of the potential are defined by

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

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

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

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

676

677 The G symmetry has been broken and three Nambu-Goldstone bosons will appear.
 678 The next step is to expand ϕ about the chosen ground state as:

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

679 to describe fluctuations from the ground state ϕ_0 . The fields $\theta_i(x)$ represent the
 680 Nambu-Goldstone bosons while $H(x)$ is known as *Higgs field*. The fundamental fea-
 681 ture of the parametrization used is that the dependence on the $\theta_i(x)$ fields is factored
 682 out in a global phase that can be eliminated by taking the physical *unitary gauge*
 683 $\theta_i(x) = 0$. Therefore the expansion about the ground state is given by:

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

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

692

693 The mechanism was proposed by three independent groups: F Englert and R Brout
 694 in August 1964 [29], P Higgs in October 1964 [30] and G Guralnik, C Hagen and
 695 T Kibble in November 1964 [31]; however, its importance was not realized until
 696 S Glashow [22], A Salam [23] and S Weinberg [24], independently, proposed that elec-
 697 tromagnetic and weak interactions are two manifestations of a more general interac-
 698 tion called *electroweak interaction* in 1967.

699 2.3.3 Masses of the gauge bosons

700 The mass of the gauge bosons is extracted by evaluating the kinetic part of Lagrangian
 701 \mathcal{L}_S in the ground state (known also as the vacuum expectation value), i.e.,

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

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

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

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_μ and A_μ in order to be compared to the typical mass terms for neutral bosons, therefore using Eqn. 2.29

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

703 and then

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

704 2.3.4 Masses of the fermions

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

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

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

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

709

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

710 where the additional term represents the lepton-Higgs interaction. The quark masses
 711 are generated in a similar way as lepton masses but for the upper member of the

712 quark doublet a different Higgs doublet is needed:

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

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

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

717 with $i,j=1,2,3$. After SSB and expansion about the ground state, the diagonal form
 718 of \mathcal{L}_{Yq} is:

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

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

721 2.3.5 The Higgs field

722 After the characterization of the fermions and gauge bosons as well as their interac-
 723 tions, it is necessary to characterize the Higgs field itself. The Lagrangian \mathcal{L}_S in Eqn.
 724 2.37 written in terms of the gauge bosons is given by

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

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

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

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

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

729 however, it is not predicted by the theory either. The experimental efforts to find the
 730 Higgs boson, carried out by the *Compact Muon Solenoid (CMS)* experiment and the *A*
 731 *Toroidal LHC Appartus (ATLAS)* experiments at the *Large Hadron Collider(LHC)*,
 732 gave great results by July of 2012 when the discovery of a new particle compatible
 733 with the Higgs boson predicted by the electroweak theory [32, 33] was announced.
 734 Although at the announcement time there were some reservations about calling the
 735 new particle the *Higgs boson*, today this name is widely accepted. The Higgs mass
 736 measurement, reported by both experiments [34], 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 ²)	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.

737

738 2.3.6 Production of Higgs bosons at LHC

739 At LHC, Higgs boson is produced as a result of the collision of two counter-rotating
 740 protons beams. A detailed description of the LHC machine will be presented in
 741 chapter 3. *The total cross section* is a parameter that quantifies the number of pp
 742 collisions that happen when a number of protons are fired at each other. Different
 743 results can be obtained after a pp collision and for each one the *cross section* is defined

as the number of pp collisions that conclude in that particular result with respect to the number of protons fired at each other.

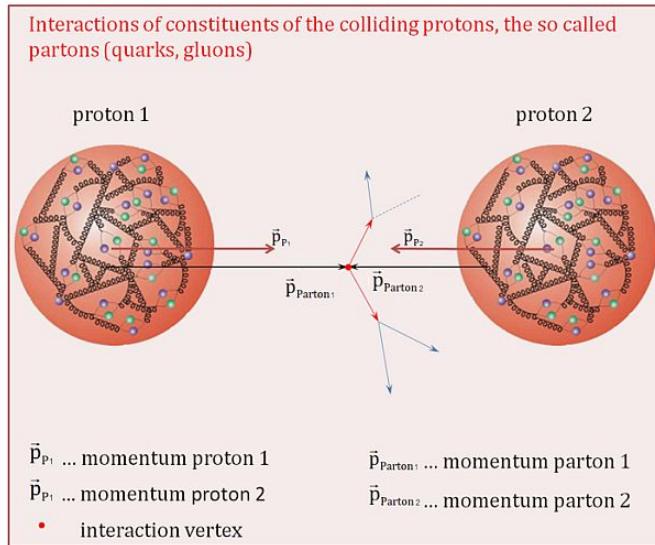


Figure 2.10: Proton-proton collision. Protons are composed of 3 valence quarks, a sea of quarks and gluons; therefore in a proton-proton collision, quarks and gluons are those who collide. [35].

745

746 Protons are composed of quarks and these quarks are bound by gluons; however,
747 what is commonly called the quark content of the proton makes reference to the
748 valence quarks. A sea of quarks and gluons is also present inside the proton as rep-
749 resented in Figure 2.10. In a proton-proton (pp) collision, the constituents (quarks
750 and gluons) are those who collide. The pp cross section depends on the momentum
751 of the colliding particles, reason for which it is needed to know how the momentum is
752 distributed inside the proton. Quarks and gluons are known as partons and the func-
753 tions that describe how the proton momentum is distributed among partons inside it
754 are called *parton distribution functions (PDFs)*; PDFs are determined from experi-
755 mental data obtained in experiments where the internal structure of hadrons is tested.

756

In addition, in physics, a common approach to study complex systems consists in

758 starting with a simpler version of them, for which a well known description is available,
 759 and add an additional *perturbation* which represents a small deviation from the known
 760 behavior. If the perturbation is small enough, the physical quantities associated with
 761 the perturbed system are expressed as a series of corrections to those of the simpler
 762 system; therefore, the more terms are considered in the series (the higher order in
 763 the perturbation series), the more precise is the the description of the complex system.

764

765 This thesis explores the Higgs production at LHC; therefore the overview presented
 766 here will be oriented specifically to the production mechanisms after pp collisions at
 767 LHC.

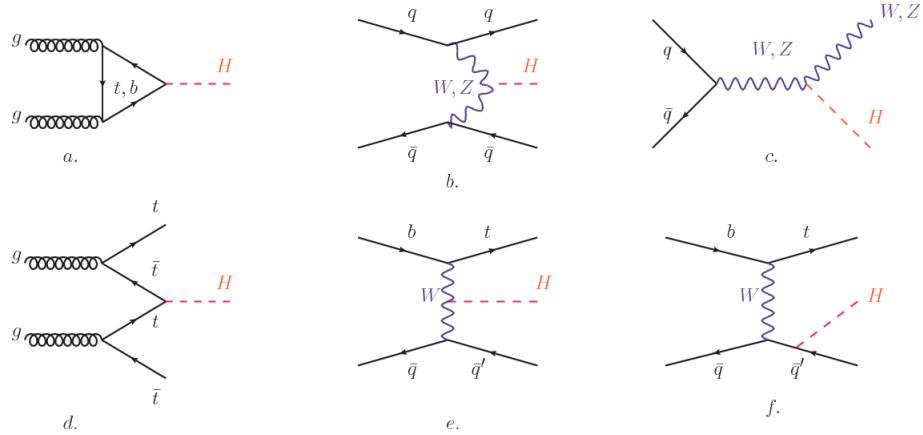


Figure 2.11: Main Higgs boson production mechanism Feynman diagrams. a. gluon-gluon fusion, b. vector boson fusion (VBF), c. Higgs-strahlung, d. Associated production with a top or bottom quark pair, e-f. associated production with a single top quark.

768 Figure 2.11 shows the Feynman diagrams for the leading order (first order) Higgs
 769 production processes at LHC, while the cross section for Higgs production as a func-
 770 tion of the center of mass-energy (\sqrt{s}) for pp collisions is showed in Figure 2.12 left.
 771 The tags NLO (next to leading order), NNLO (next to next to leading order) and
 772 N3LO (next to next to next to leading order) make reference to the order at which
 773 the perturbation series have been considered.

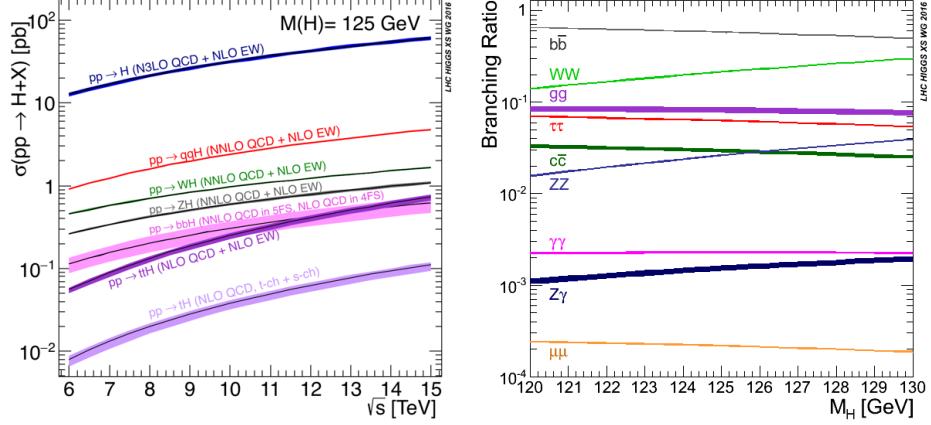


Figure 2.12: Higgs boson production cross sections (left) and decay branching ratios (right) for the main mechanisms. The VBF is indicated as qqH [36].

As shown in Eqns 2.48, 2.52 and 2.56, the strength of the Higgs-fermion interaction is proportional to the fermion mass while the strength of the Higgs-gauge boson interaction is proportional to the square of the gauge boson mass, which implies that the Higgs production and decay mechanisms are dominated by couplings $H - (W, Z, t, b, \tau)$.

The main production mechanism is the gluon fusion (Figure 2.11a and $pp \rightarrow H$ in Figure 2.12) given that gluons carry the highest fraction of momentum of the protons in pp colliders. Since the Higgs boson does not couple to gluons, the mechanism proceeds through the exchange of a virtual top-quark loop given that for it the coupling is the biggest. Note that in this process, the Higgs boson is produced alone, which makes this mechanism experimentally clean when combined with the two-photon or the four-lepton decay channels (see Section 2.3.7).

Vector boson fusion (Figure 2.11b and $pp \rightarrow qqH$ in Figure 2.12) has the second largest production cross section. The scattering of two fermions is mediated by a weak gauge boson which later emits a Higgs boson. In the final state, the two fermions tend to be located in a particular region of the detector which is used as a signature

790 when analyzing the datasets provided by the experiments. More details about how
 791 to identify events of interest in an analysis will be given in chapter 6.

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

796 The associated production with a top or bottom quark pair and the associated
 797 production with a single top quark (Figure 2.11d-f and $pp \rightarrow bbH, pp \rightarrow t\bar{t}H, pp \rightarrow tH$
 798 in Figure 2.12) have a smaller cross section than the main three mechanisms above,
 799 but they provide a good opportunity to test the Higgs-top coupling. The analysis
 800 reported in this thesis is developed using these production mechanisms. A detailed
 801 description of the tH mechanism will be given in Section 2.4.

802 **2.3.7 Higgs boson decay channels**

803 When a particle can decay through several modes, also known as channels, the prob-
 804 ability of decaying through a given channel is quantified by the *branching ratio (BR)*
 805 of the decay channel; thus, the BR is defined as the ratio of number of decays go-
 806 ing through that given channel to the total number of decays. In regard to the
 807 Higgs boson decay, the BR can be predicted with accuracy once the Higgs mass is
 808 known [37, 38]. In Figure 2.12 right, a plot of the BR as a function of the Higgs mass
 809 is presented. The largest predicted BR corresponds to the $b\bar{b}$ pair decay channel (see
 810 Table 2.9).

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

Table 2.9: Predicted branching ratios and the relative uncertainty for a SM Higgs boson with $m_H = 125\text{GeV}/c^2$. [9]

812 2.4 Associated production of a Higgs boson and a 813 single Top quark.

Associated production of Higgs boson has been extensively studied [39–43]. While measurements of the main Higgs production mechanisms rates are sensitive to the strength of the Higgs coupling to W boson or top quark, they are not sensitive to the relative sign between the two couplings. In this thesis, the Higgs boson production mechanism explored is the associated production with a single top quark (tH) which offers sensitiveness to the relative sign of the Higgs couplings to W boson and to top quark. The description given here is based on the Reference [41]

821

822 A process where two incoming particles interact and produce a final state with
823 two particles can proceed in three ways also called channels (see, for instance, Figure
824 2.13 omitting the red line). The t-channel represents processes where an intermediate
825 particle is emitted by one of the incoming particles and absorbed by the other. The
826 s-channel represents processes where the two incoming particles merge into an inter-
827 mediate particle which eventually will split into the particles in the final state. The
828 third channel, u-channel, is similar to the t-channel but the two outgoing particles

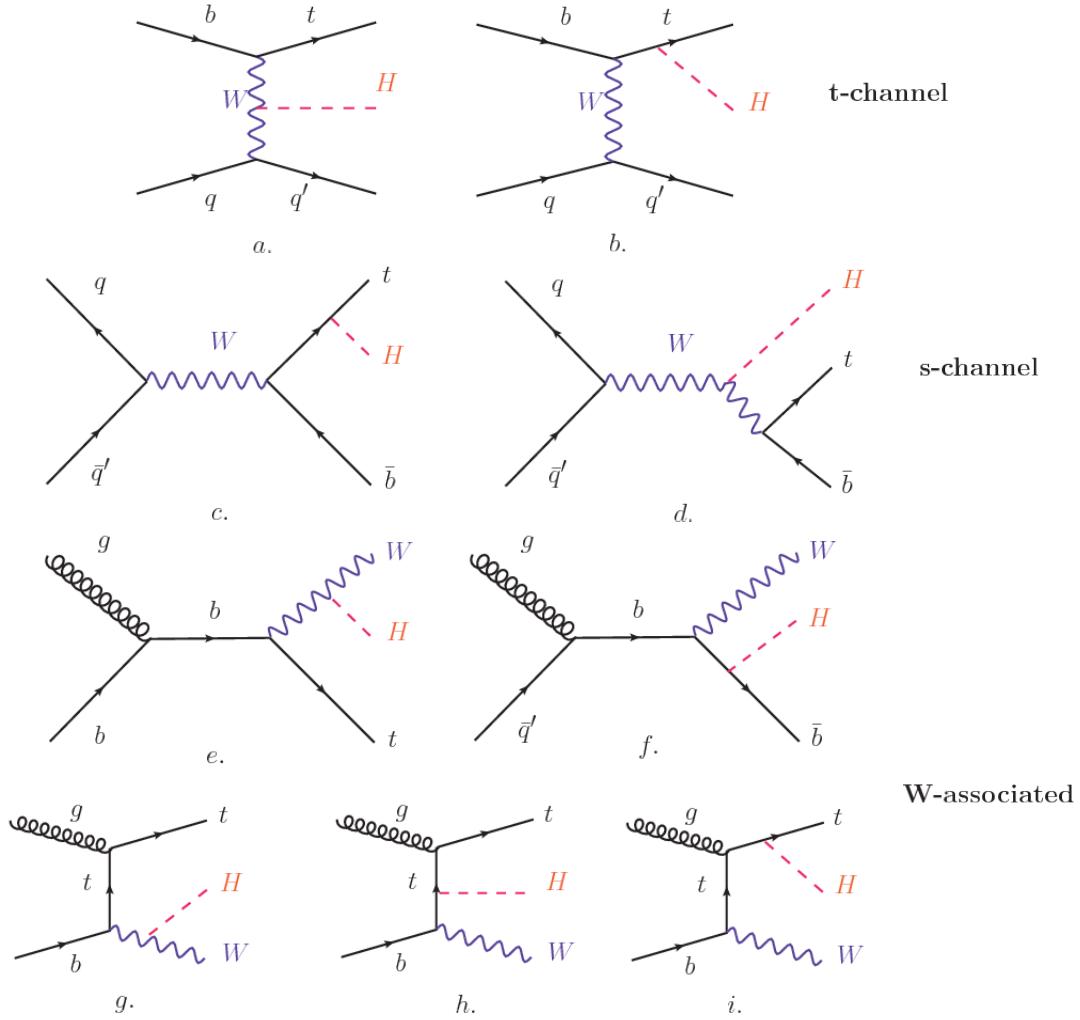


Figure 2.13: Associated Higgs boson production mechanism Feynman diagrams. a.,b. t-channel (tHq), c.,d. s-channel (tHb), e-i. W-associated.

829 interchange their roles.

830

831 The tH production, where Higgs boson can be radiated either from the top quark
 832 or from the W boson, is represented by the leading order Feynman diagrams in
 833 Figure ???. The cross section for the tH process is calculated, as usual, summing over
 834 the contributions from the different Feynman diagrams; therefore it depends on the
 835 interference between the contributions. In the SM, the interference for t-channel (tHq

process) and W-associated (tHW process) production is destructive [39] 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 [44, 45].

838

839 While the s-channel contribution can be neglected, it will be shown that a de-
 840 viation from the SM destructive interference would result in an enhancement of the
 841 tH cross section compared to that in SM, which could be used to get information
 842 about the sign of the Higgs-top coupling [41, 42]. In order to describe tH production
 843 processes, Feynman diagram 2.13b will be considered; there, the W boson is radiated
 844 by a quark in the proton and eventually it will interact with the b quark. In the
 845 high energy regime, the effective W approximation [46] allows to describe the process
 846 as the emission of an approximately on-shell W and its hard scattering with the b
 847 quark; i.e. $Wb \rightarrow th$. The scattering amplitude for the process is given by

$$\mathcal{A} = \frac{g}{\sqrt{2}} \left[(\kappa_t - \kappa_V) \frac{m_t \sqrt{s}}{m_W v} A \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) + \left(\kappa_V \frac{2m_W s}{v} \frac{1}{t} + (2\kappa_t - \kappa_V) \frac{m_t^2}{m_W v} \right) B \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) \right], \quad (2.58)$$

848 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 quan-
 849 tify possible deviations of the couplings, Higgs-Vector boson (H-W) and Higgs-top
 850 (H-t) respectively, from the SM couplings; $s = (p_W + p_b)^2$, $t = (p_W - p_H)^2$, φ is the
 851 Higgs azimuthal angle around the z axis taken parallel to the direction of motion of
 852 the incoming W; A and B are functions describing the weak interaction in terms of

853 the chiral states of the quarks b and t . Terms that vanish in the high energy limit
 854 have been neglected as well as the Higgs and b quark masses⁷.

855

856 The scattering amplitude grows with energy like \sqrt{s} for $\kappa_V \neq \kappa_t$, in contrast to
 857 the SM ($\kappa_t = \kappa_V = 1$), where the first term in 2.58 cancels out and the amplitude
 858 is constant for large s ; therefore, a deviation from the SM predictions represents an
 859 enhancement in the tHq cross section. In particular, for a SM H-W coupling and a
 860 H-t coupling of inverted sign with respect to the SM ($\kappa_V = -\kappa_t = 1$) the tHq cross
 861 section is enhanced by a factor greater 10 as seen in the Figure 2.14 taken from
 862 Reference [41]; Reference [47] has reported similar enhancement results.

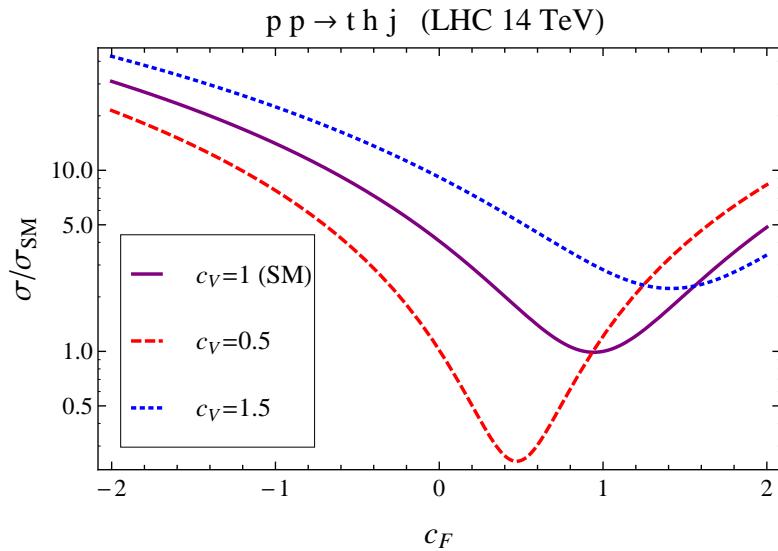


Figure 2.14: Cross section for tHq process as a function of κ_t , normalized to the SM, for three values of κ_V . In the plot c_f refers to the Higgs-fermion coupling which is dominated by the H-t coupling and represented here by κ_t . Solid, dashed and dotted lines correspond to $c_V \rightarrow \kappa_V = 1, 0.5, 1.5$ respectively. Note that for the SM ($\kappa_V = \kappa_t = 1$), the destructive effect of the interference is maximal.

863 A similar analysis is valid for the W-associated channel but, in that case, the in-
 864 terference is more complicated since there are more than two contributions and an ad-

⁷ A detailed explanation of the structure and approximations used to derive \mathcal{A} can be found in Reference [41]

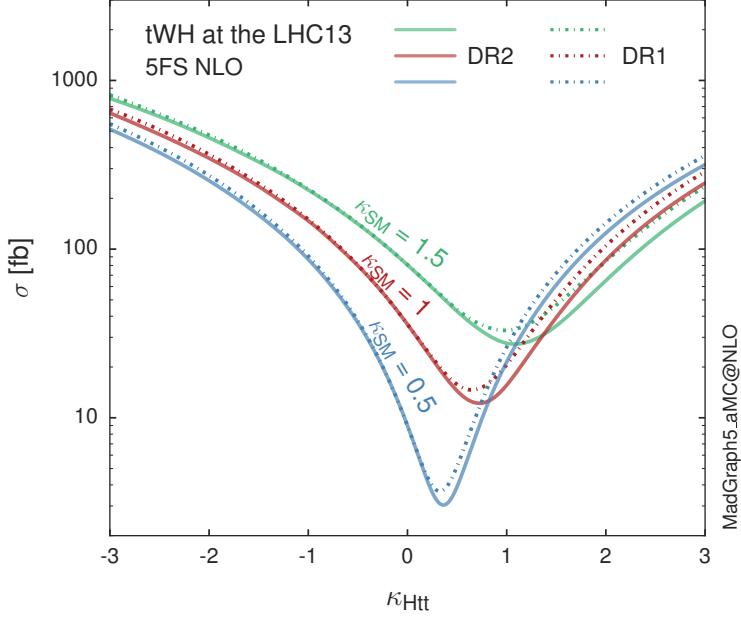


Figure 2.15: Cross section for tHW process as a function of κ_{Htt} , for three values of κ_{SM} at $\sqrt{s} = 13$ TeV. $\kappa_{Htt}^2 = \sigma_{Htt}/\sigma_{Htt}^{SM}$ is a simple re-scaling of the SM Higgs interactions.

ditional interference with the production of Higgs boson and a top pair process($t\bar{t}H$).
 The calculations are made using the so-called Diagram Removal (DR) technique where
 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 [48]. As shown in Figure
 2.15, 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 [49] and a summary of the results is presented in Table 2.11.

	\sqrt{s} TeV	$\kappa_t = 1$	$\kappa_t = -1$
$\sigma^{LO}(tHq)(\text{fb})$ [41]	8	≈ 17.4	≈ 252.7
	14	≈ 80.4	≈ 1042
$\sigma^{NLO}(tHq)(\text{fb})$ [41]	8	$18.28^{+0.42}_{-0.38}$	$233.8^{+4.6}_{-0.0}$
	14	$88.2^{+1.7}_{-0.0}$	$982.8^{+28}_{-0.0}$
$\sigma^{LO}(tHq)(\text{fb})$ [47]	14	≈ 71.8	≈ 893
$\sigma^{LO}(tHW)(\text{fb})$ [47]	14	≈ 16.0	≈ 139
$\sigma^{NLO}(tHq)(\text{fb})$ [49]	8	$18.69^{+8.62\%}_{-17.13\%}$	-
	13	$74.25^{+7.48\%}_{-15.35\%}$	$848^{+7.37\%}_{-13.70\%}$
	14	$90.10^{+7.34\%}_{-15.13\%}$	$1011^{+7.24\%}_{-13.39\%}$
$\sigma^{LO}(tHW)(\text{fb})$ [48]	13	$15.77^{+15.91\%}_{-15.76\%}$	-
$\sigma^{NLO} DR1(tHW)(\text{fb})$ [48]	13	$21.72^{+6.52\%}_{-5.24\%}$	≈ 150
$\sigma^{NLO} DR2(tHW)(\text{fb})$ [48]	13	$16.28^{+7.34\%}_{-15.13\%}$	≈ 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.

876 2.5 The CP-mixing in tH processes

877 In addition to the sensitivity to sign of the H-t coupling, tHq and tHW processes have
 878 been proposed as a tool to investigate the possibility of a H-t coupling that does not
 879 conserve CP [43, 48, 50]. Current experimental results are consistent with SM H-V
 880 and H-t couplings; however, negative H-t coupling is not excluded completely [53].

881

882 In this thesis, the sensitivity of tH processes to CP-mixing is also studied in the
 883 effective field theory framework and based in References [43, 48]; a generic particle
 884 (X_0) of spin-0 and a general CP violating interaction with the top quark, can couple
 885 to scalar and pseudo-scalar fermionic densities. The H-W interaction is assumed to
 886 be SM-like. The Lagrangian modeling the H-t interaction is given by

$$884 \quad \mathcal{L}_0^t = -\bar{\psi}_t (c_\alpha \kappa_{Htt} g_{Htt} + i s_\alpha \kappa_{Att} g_{Att} \gamma_5) \psi_t X_0, \quad (2.59)$$

887 where α is the CP-mixing phase, $c_\alpha \equiv \cos \alpha$ and $s_\alpha \equiv \sin \alpha$, κ_{Htt} and κ_{Att} are real
 888 dimensionless re-scaling parameters⁸, $g_{Htt} = g_{Att} = m_t/v = y_t/\sqrt{2}$ and $v \sim 246$ GeV
 889 is the Higgs vacuum expectation value. In this parametrization, it is easy to recover
 890 three special cases

- 891 • CP-even coupling $\rightarrow \alpha = 0^\circ$
- 892 • CP-odd coupling $\rightarrow \alpha = 90^\circ$
- 893 • SM coupling $\rightarrow \alpha = 0^\circ$ and $\kappa_{Htt} = 1$

894 The loop induced X_0 coupling to gluons can also be described in terms of the
 895 parametrization above, according to

$$\mathcal{L}_0^g = -\frac{1}{4} \left(c_\alpha \kappa_{Hgg} g_{Hgg} G_{\mu\nu}^a G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} \right) X_0. \quad (2.60)$$

896 where $g_{Hgg} = -\alpha_s/3\pi v$ and $g_{Agg} = \alpha_s/2\pi v$. Under the assumption that the top quark
 897 dominates the gluon-fusion process at LHC energies, $\kappa_{Hgg} \rightarrow \kappa_{Htt}$ and $\kappa_{Agg} \rightarrow \kappa_{Att}$,
 898 so that the ratio between the gluon-gluon fusion cross section for X_0 and for the SM
 899 Higgs prediction can be written as

$$\frac{\sigma_{NLO}^{gg \rightarrow X_0}}{\sigma_{NLO,SM}^{gg \rightarrow H}} = c_\alpha^2 \kappa_{Htt}^2 + s_\alpha^2 \left(\kappa_{Att} \frac{g_{Agg}}{g_{Hgg}} \right)^2. \quad (2.61)$$

900 If the re-scaling parameters are set to

$$\kappa_{Htt} = 1, \quad \kappa_{Att} = \left| \frac{g_{Hgg}}{g_{Agg}} \right| = \frac{2}{3}. \quad (2.62)$$

901 the gluon-fusion SM cross section is reproduced for every value of the CP-mixing
 902 angle α ; therefore, by imposing that condition to the Lagrangian density 2.59, the

⁸ analog to κ_t and κ_V

903 CP-mixing angle is not constrained by current data. Figure 2.16 shows the NLO cross
 904 sections for t-channel tX_0 (blue) and $t\bar{X}_0$ (red) associated production processes as a
 905 function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark
 906 CP-violating coupling. Re-scaling factors κ_{Htt} and κ_{Att} have been set to reproduce
 907 the SM gluon-fusion cross sections.

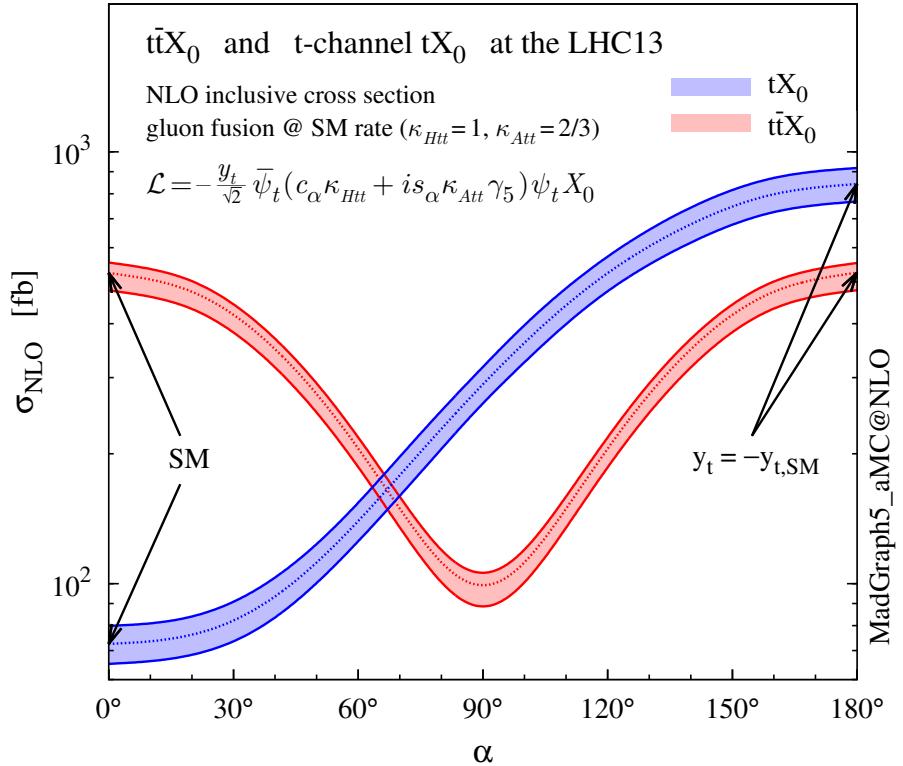


Figure 2.16: NLO cross sections for t-channel tX_0 (blue) and $t\bar{X}_0$ (red) associated production processes as a function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark CP-violating coupling [43].

908 It is interesting to notice that the tX_0 cross section is enhanced, by a factor of
 909 about 10, when a continuous rotation in the scalar-pseudoscalar plane is applied; this
 910 enhancement is similar to the enhancement produced when the H-t coupling is flipped
 911 in sign with respect to the SM ($y_t = -y_{t,SM}$ in the plot), as showed in Section 2.4. In
 912 contrast, the degeneracy in the $t\bar{X}_0$ 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 $\alpha \sim 60^\circ$.

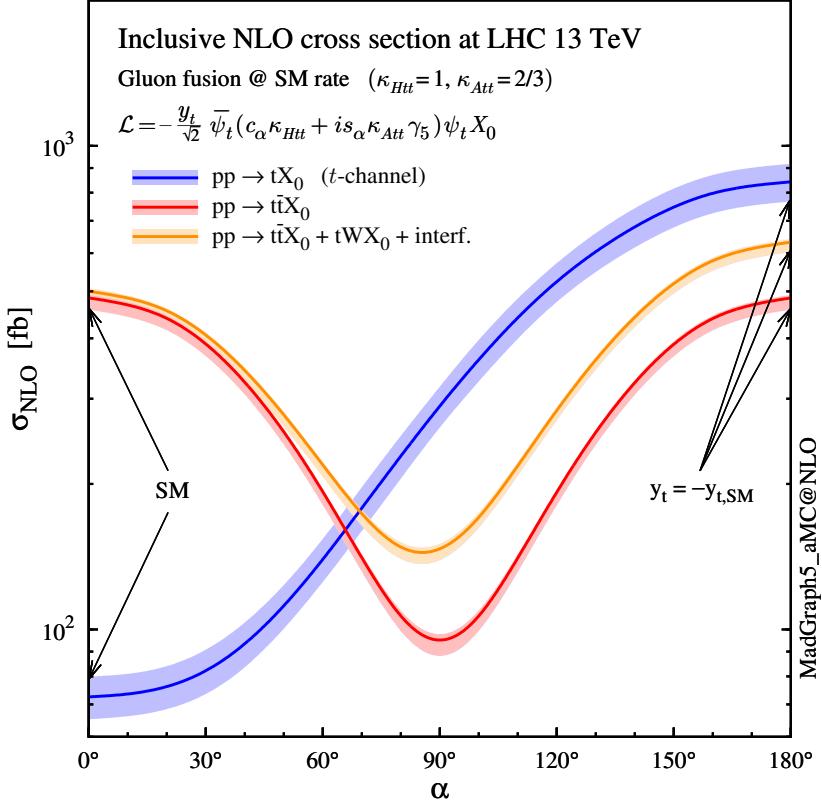


Figure 2.17: NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red) associated production processes and combined $tWX_0 + t\bar{t}X_0$ (including interference) production as a function of the CP-mixing angle α [43].

914

915 A similar parametrization can be used to investigate the tHW process sensitivity
916 to CP-violating H-t coupling. As said in 2.4, the interference in the W-associated
917 channel is more complicated because there are more than two contributions and also
918 there is interference with the $t\bar{t}H$ production process.

919

920 Figure 2.17 shows the NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red)
921 associated production and for the combined $tWX_0 + t\bar{t}X_0 +$ interference (orange) as
922 a function of the CP-mixing angle. It is clear that the effect of the interference in the

combined case is the lifting of the degeneracy present in the $t\bar{t}X_0$ production. The constructive interference enhances the cross section from about 500 fb at SM ($\alpha = 0$) to about 600 fb ($\alpha = 180^\circ \rightarrow y_t = -y_{t,SM}$).

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

2.6 Experimental status of the anomalous Higgs-fermion coupling.

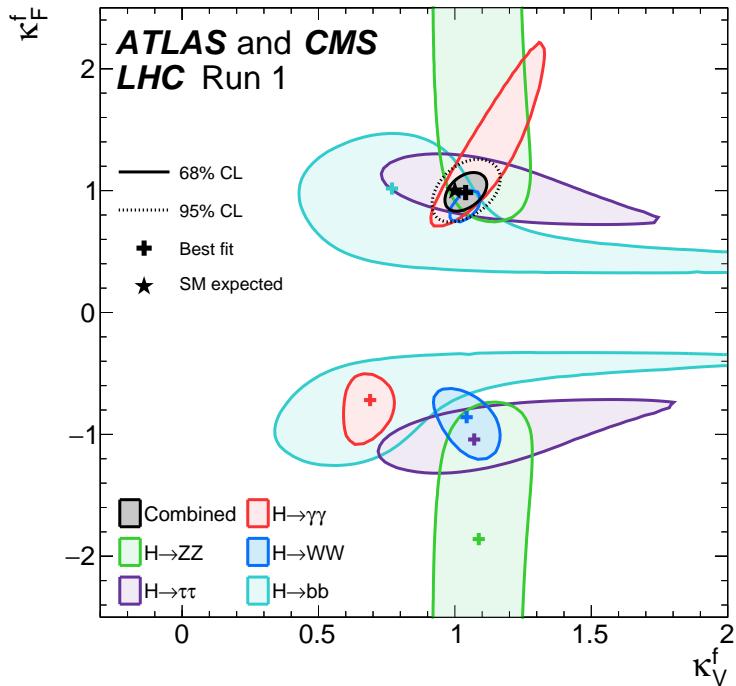


Figure 2.18: Combination of the ATLAS and CMS fits for coupling modifiers κ_t - κ_V ; also shown the individual decay channels combination and their global combination. No assumptions have been made on the sign of the coupling modifiers [53].

ATLAS and CMS have performed analysis of the anomalous H-f coupling by making likelihood scans for the two coupling modifiers, κ_t and κ_V , under the assumption

932 that $\kappa_Z = \kappa_W = \kappa_V$ and $\kappa_t = \kappa_\tau = \kappa_b = \kappa_f$. Figure 2.18 shows the result of the combi-
933 nation of ATLAS and CMS fits; also the individual decay channels combination and
934 the global combination results are shown.

935 While all the channels are compatible for positive values of the modifiers, for negative
936 values of κ_t there is no compatibility. The best fit for individual channels is compatible
937 with negative values of κ_t except for the $H \rightarrow bb$ channel which is expected to be the
938 most sensitive channel; therefore, the best fit for the global fit yields $\kappa_t \geq 0$. Thus,
939 the anomalous H-t coupling cannot be excluded completely.

940 Chapter 3

941 The CMS experiment at the LHC

942 3.1 Introduction

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

956 3.2 The LHC

957 With 27 km of circumference, the LHC is currently the largest and most powerful
 958 circular accelerator in the world. It is installed in the same tunnel where the Large
 959 Electron-Positron (LEP) collider was located, taking advantage of the existing infras-
 960 tructure. The LHC is also the larger accelerator in the CERN's accelerator complex
 961 and is assisted by several successive accelerating stages before the particles are in-
 962 jected into the LHC ring where they reach their maximum energy (see Figure 3.1).

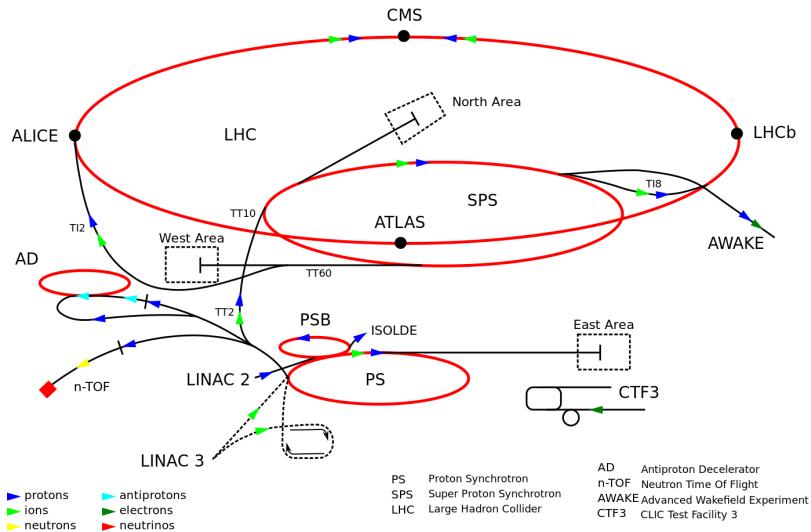


Figure 3.1: CERN accelerator complex. Blue arrows show the path followed by protons along the acceleration process [54].

963 LHC runs in three modes depending on the particles being accelerated

- 964 ● Proton-Proton collisions (pp) for multiple physics experiments.
- 965 ● Lead-Lead collisions ($Pb-Pb$) for heavy ion experiments.
- 966 ● Proton-Lead collisions ($p-Pb$) for quark-gluon plasma experiments.

967 In this thesis only pp collisions will be considered.

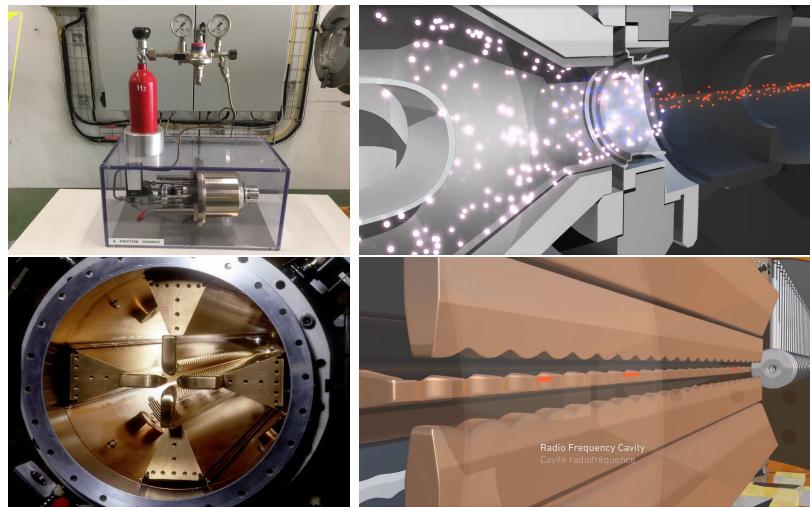


Figure 3.2: LHC protons source and the first acceleration stage. Top: the bottle contains hydrogen gas (white dots) which is injected into the metal cylinder to be broken down into electrons (blue dots) and protons (red dots); Bottom: the obtained protons are directed towards the radio frequency quadrupole which perform the first acceleration, focus the beam and create the bunches of protons. [58, 59]

969 Collection of protons starts with hydrogen atoms taken from a bottle, containing hy-
 970 drogen gas, and injecting them in a metal cylinder; hydrogen atoms are broken down
 971 into electrons and protons by an intense electric field (see Figure 3.2 top). The re-
 972 sulting protons leave the metal cylinder towards a radio frequency quadrupole (RFQ)
 973 that focus the beam, accelerates the protons and creates the packets of protons called
 974 bunches. In the RFQ, an electric field is generated by a RF wave at a frequency that
 975 matches the resonance frequency of the cavity where the electrodes are contained.
 976 The beam of protons traveling on the RFQ axis experiences an alternating electric
 977 field gradient that generates the focusing forces.

978

979 In order to accelerate the protons, a longitudinal time-varying electric field component
 980 is added to the system; it is done by giving the electrodes a sinus-like profile as shown
 981 in Figure 3.2 bottom. By matching the speed and phase of the protons with the
 982 longitudinal electric field the bunching is performed; protons synchronized with the

983 RFQ (synchronous proton) do not feel an accelerating force, but those protons in the
 984 beam that have more (or less) energy than the synchronous proton (asynchronous
 985 protons) will feel a decelerating (accelerating) force; therefore, asynchronous protons
 986 will oscillate around the synchronous ones forming bunches of protons [56]. From the
 987 RFQ protons emerge with energy 750 keV in bunches of about 1.15×10^{11} protons [57].

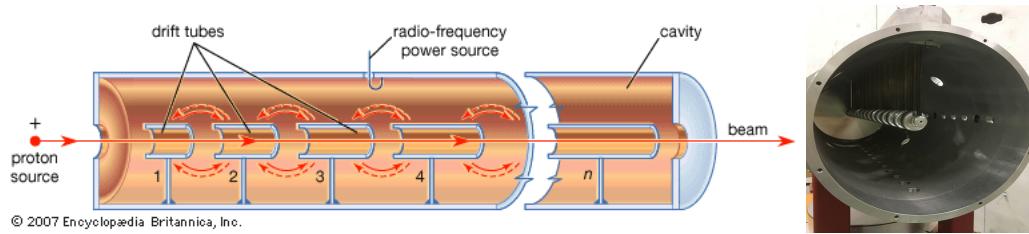


Figure 3.3: The LINAC2 accelerating system at CERN. Electric fields generated by radio frequency (RF) create acceleration and deceleration zones inside the cavity; deceleration zones are blocked by drift tubes where quadrupole magnets focus the proton beam. [60]

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

994

995 The beam coming from LINAC2 is injected into the proton synchrotron booster
 996 (PSB) to reach 1.4 GeV in energy. The next acceleration is provided at the pro-
 997 ton synchrotron (PS) up to 26 GeV, followed by the injection into the super proton
 998 synchrotron (SPS) where protons are accelerated to 450 GeV. Finally, protons are
 999 injected into the LHC where they are accelerated to the target energy of 6.5 TeV.
 1000 PSB, PS, SPS and LHC accelerate protons using the same RF acceleration technique
 1001 described before.

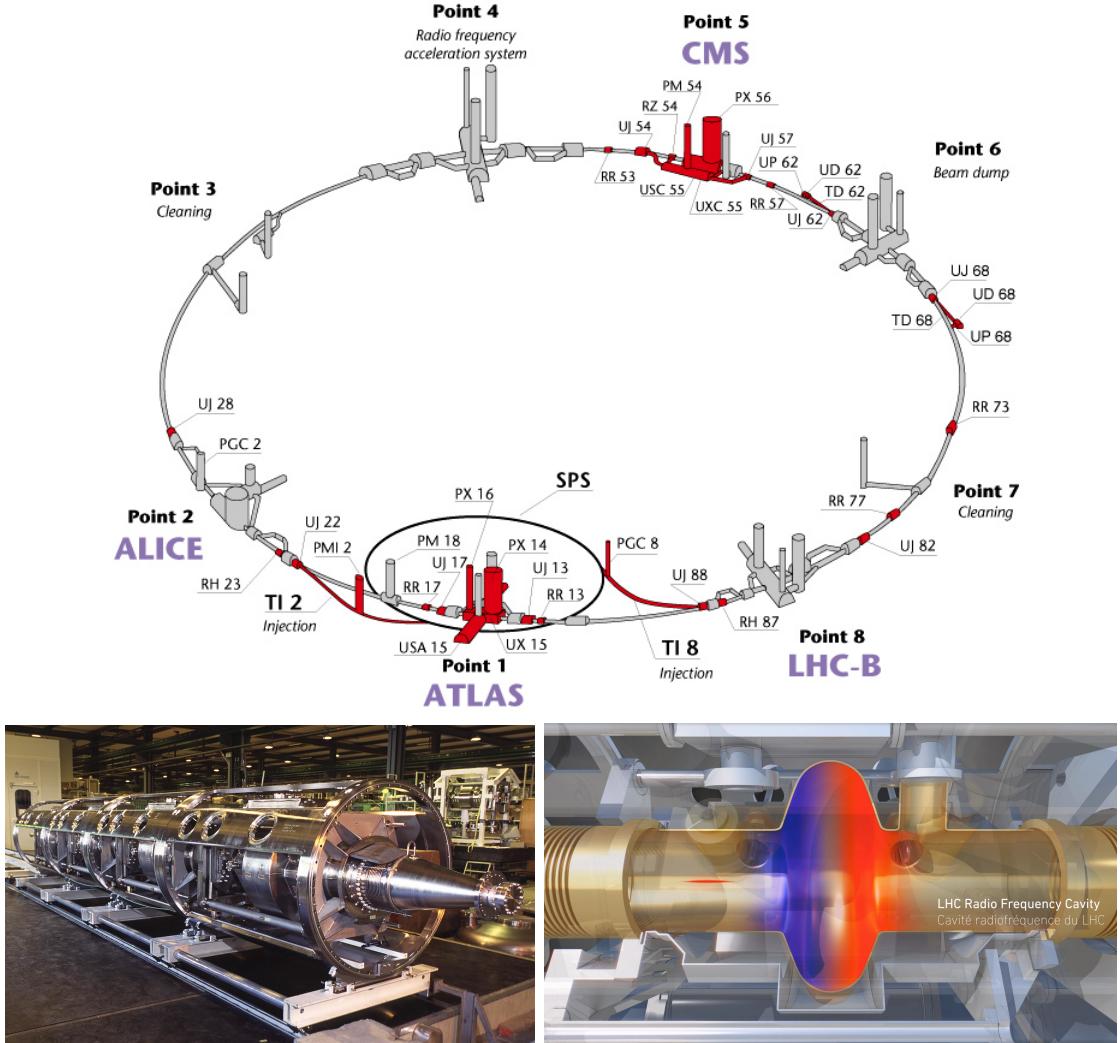


Figure 3.4: Top: LHC layout. The red zones indicate the infrastructure additions to the LEP installations, built to accommodate the ATLAS and CMS experiments which exceed the size of the former experiments located there [55]. Bottom: LHC RF cavities. A module accommodates 4 cavities that accelerate protons and preserve the bunch structure of the beam. [59, 61]

1002 LHC has a system of 16 RF cavities located in the so-called point 4, as shown in
 1003 Figure 3.4 top, tuned at a frequency of 400 MHz and the protons are carefully timed,
 1004 so in addition to the acceleration effect the bunch structure of the beam is preserved.
 1005 Bottom side of Figure 3.4 shows a picture of a RF module composed of 4 RF cavities
 1006 working in a superconducting state at 4.5 K; also is showed a representation of the

1007 accelerating electric field that accelerates the protons in the bunch.

1008

1009 While protons are accelerated in one section of the LHC ring, where the RF cavities
 1010 are located, in the rest of their path they have to be kept in the curved trajectory
 1011 defined by the LHC ring. Technically, LHC is not a perfect circle; RF, injection, beam
 1012 dumping, beam cleaning and sections before and after the experimental points where
 1013 protons collide are all straight sections. In total, there are 8 arcs 2.45 Km long each
 1014 and 8 straight sections 545 m long each. In order to curve the proton's trajectory in
 1015 the arc sections, superconducting dipole magnets are used.

1016

1017 Inside the LHC ring, there are two proton beams traveling in opposite directions in
 1018 two separated beam pipes; the beam pipes are kept at ultra-high vacuum ($\sim 10^{-9}$
 1019 Pa) to ensure that there are no particles that interact with the proton beams. The
 1020 superconducting dipole magnets used in LHC are made of a NbTi alloy, capable of
 1021 transporting currents of about 12000 A when cooled at a temperature below 2K using
 1022 liquid helium (see Figure 3.5).

1023

1024 Protons in the arc sections of LHC feel a centripetal force exerted by the dipole
 1025 magnets which is perpendicular to the beam trajectory; The magnitude of magnetic
 1026 field needed can be found assuming that protons travel at $v \approx c$, using the standard
 1027 values for proton mass and charge and the LHC radius, as

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

1028 which is about 100000 times the Earth's magnetic field. A representation of the mag-
 1029 netic field generated by the dipole magnets is shown on the bottom left side of Figure

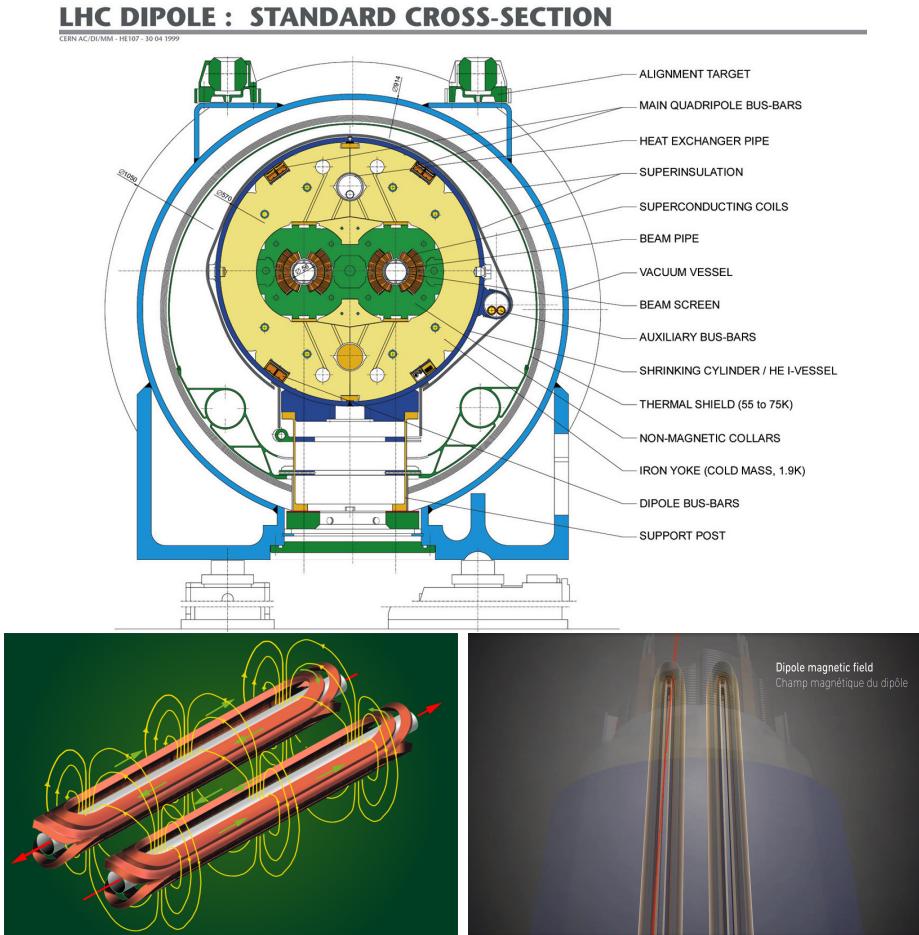


Figure 3.5: Top: LHC dipole magnet transverse view; cooling, shielding and mechanical support are indicated. Bottom left: Magnetic field generated by the dipole magnets; note that the direction of the field inside one beam pipe is opposite with respect to the other beam pipe which guarantee that both 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 [59, 62, 63].

1030 3.5. The bending effect of the magnetic field on the proton beam is shown on the
 1031 bottom right side of Figure 3.5. Note that the dipole magnets are not curved; the
 1032 arc section of the LHC ring is composed of straight dipole magnets of about 15 m.
 1033 In total there are 1232 dipole magnets along the LHC ring.
 1034
 1035 In addition to bending the beam trajectory, the beam has to be focused so it stays

1036 inside the beam pipe. The focusing is performed by quadrupole magnets installed in
 1037 a different straight section; in total 858 quadrupole magnets are installed along the
 1038 LHC ring. Other effects like electromagnetic interaction among bunches, interaction
 1039 with electron clouds from the beam pipe, the gravitational force on the protons, dif-
 1040 ferences in energy among protons in the same bunch, among others, are corrected
 1041 using sextupole and other magnetic multipoles.

1042

1043 The two proton beams inside the LHC ring are made of bunches with a cylindrical
 1044 shape of about 7.5 cm long and about 1 mm in diameter; when bunches are close
 1045 to the collision point (CP), the beam is focused up to a diameter of about 16 μm in
 1046 order to maximize the number of collisions per unit area and per second, known as
 1047 luminosity (L). Luminosity can be calculated using

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

1048 where f is the revolution frequency, n is the number of bunches per beam, N_1 and
 1049 N_2 are the numbers of protons per bunch (1.5×10^{11}), σ_x and σ_y are the gaussian
 1050 transverse sizes of the bunches. The expected luminosity is about

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

$$n = 2808$$

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

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

1051

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

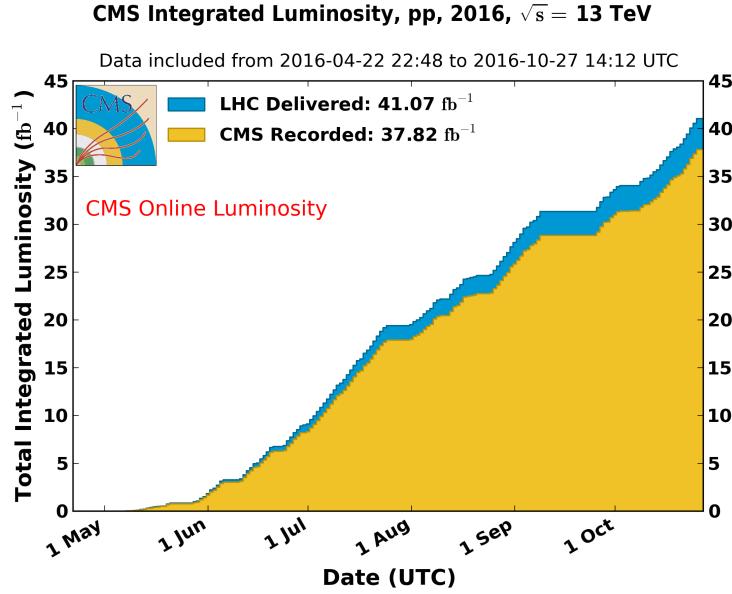


Figure 3.6: Integrated luminosity delivered by LHC and recorded by CMS during 2016. The difference between the delivered and the recorded luminosity is due to fails and issues occurred during the data taking in the CMS experiment [64].

1052 Luminosity is a fundamental aspect of LHC given that the bigger luminosity the
 1053 bigger number of collisions, which means that for processes with a very small cross
 1054 section the number of expected occurrences is increased and so the chances of being
 1055 detected. The integrated luminosity, i.e., the total luminosity, collected by the CMS
 1056 experiment during 2016 is shown in Figure 3.6; the data analyzed in this thesis cor-
 1057 responds to an integrated luminosity of 35.9 fb^{-1} at a center of mass-energy $\sqrt{s} = 13$
 1058 TeV.

1059

1060 A way to increase L is increasing the number of bunches in the beam. Currently, the
 1061 separation between two consecutive bunches in the beam is 7.5 m which corresponds
 1062 to a time separation of 25 ns. In the full LHC ring the allowed number of bunches is
 1063 $n = 27\text{km}/7.5\text{m} = 3600$; however, there are some gaps in the bunch pattern intended
 1064 for preparing the dumping and injection of the beam, thus, the proton beams are

1065 composed of 2808 bunches.

1066

1067 Once the proton beams reach the desired energy, they are brought to cross each other
 1068 producing proton-proton collisions. The bunch crossing happens in precise places
 1069 where the four LHC experiments are located, as seen in the top of Figure 3.7. In
 1070 2008, the first set of collisions involved protons with $\sqrt{s} = 7$ TeV; the energy was
 1071 increased to 8 TeV in 2012 and to 13 TeV in 2015.

1072

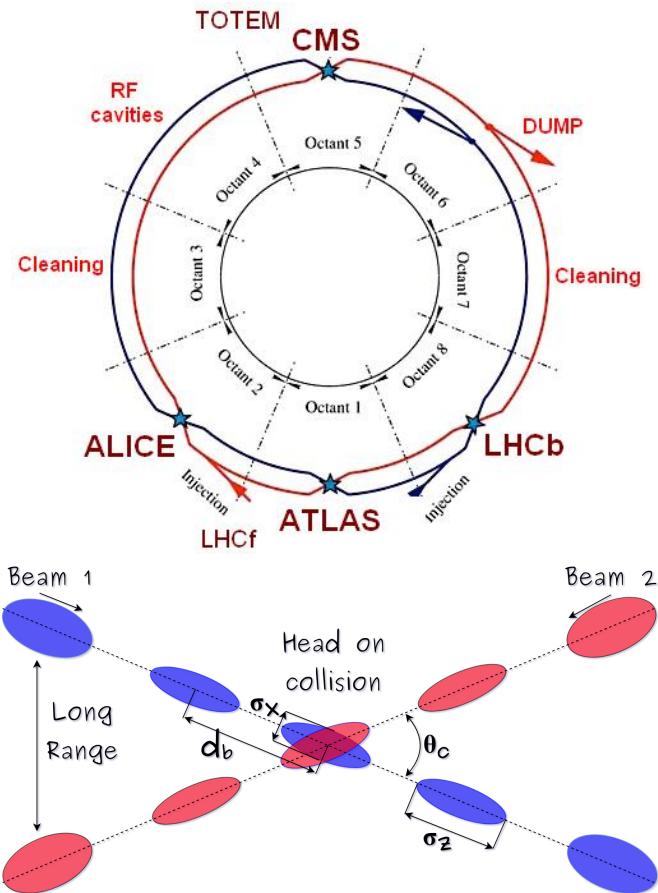


Figure 3.7: Top: LHC interaction points. Bunch crossing occurs where the LHC experiments are located [65]. Sections indicated as cleaning are dedicated to collimate the beam in order to protect the LHC ring from collisions with protons in very spreaded bunches. Bottom: bunch crossing scheme. Since the bunch crossing is not perfectly head-on, the luminosity is reduced in a factor of 17%; adapted from Reference [77].

1073 CMS and ATLAS experiments, which are multi-purpose experiments, are enabled
 1074 to explore physics in any of the collision modes. LHCb experiment is optimized
 1075 to explore bottom quark physics, while ALICE is optimized for heavy ion collisions
 1076 searches; TOTEM and LHCf are dedicated to forward physics studies; MoEDAL (not
 1077 indicated in the Figure) is intended for monopoles or massive pseudo stable particles
 1078 searches.

1079

1080 At the CP there are two interesting details that need to be addressed. The first one
 1081 is that the bunch crossing does not occur head-on but at a small crossing angle θ_c
 1082 (280 μ rad in CMS and ATLAS) as shown in the bottom side of Figure 3.7, affecting
 1083 the overlapping between bunches; the consequence is a reduction of about 17% in
 1084 the luminosity (represented by a factor not included in eqn. 3.2). The second one
 1085 is the occurrence of multiple pp collisions in the same bunch crossing; this effect is
 1086 called pile-up (PU). A fairly simple estimation of the PU follows from estimating the
 1087 probability of collision between two protons, one from each of the bunches in course
 1088 of collision; it depends roughly on the ratio of proton size and the cross section of the
 1089 bunch in the interaction point, i.e.,

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

1090 however, there are $N = 1.15 \times 10^{11}$ protons in a bunch, thus the estimated number of
 1091 collisions in a bunch crossing is

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

1092 about 20 of those pp collisions are inelastic. Each collision generates a vertex, but
 1093 only the most energetic is considered as a primary vertex; the rest are considered

1094 as PU vertices. A multiple pp collision event in a bunch crossing at CMS is showed
 1095 in Figure 3.8. Unstable particles outgoing from the primary vertex will eventually
 1096 decay; this decay vertex is known as a secondary vertex.

1097

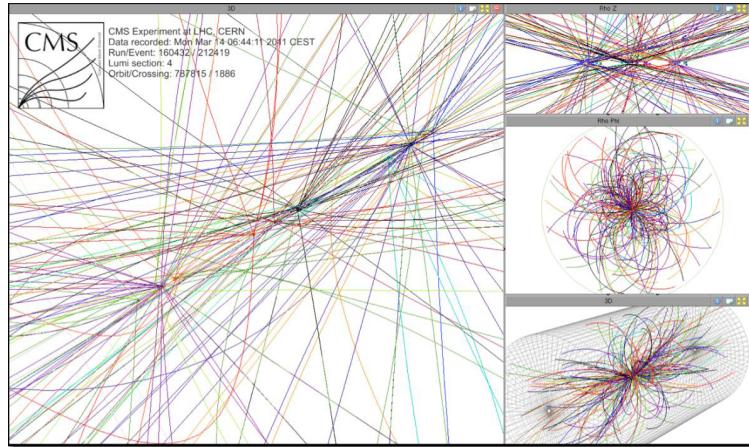


Figure 3.8: Multiple pp collision bunch crossing at CMS. Only the most energetic vertex is considered and the rest are cataloged as PU vertices [66].

1098 Next section presents a description of the CMS detector which it is the detector used
 1099 to collect the data used in this thesis.

1100 3.3 The CMS experiment

1101 CMS is a general-purpose detector designed to conduct research in a wide range
 1102 of physics from the standard model to new physics like extra dimensions and dark
 1103 matter. Located at the point 5 in the LHC layout as shown in Figure 3.4, CMS is
 1104 composed of several detection systems distributed in a cylindrical structure; in total,
 1105 CMS weights about 12500 tons in a very compact 21.6 m long and 14.6 m diameter
 1106 cylinder. It was built in 15 separate sections at the ground level and lowered to the
 1107 cavern individually to be assembled. A complete and detailed description of the CMS
 1108 detector and its components is given in Reference [67] on which this section is based

1109 on.

1110

1111 Figure 3.9 shows the layout of the CMS detector. The design is driven by the require-
 1112 ments on the identification, momentum resolution and unambiguous charge determi-
 1113 nation of the muons; therefore, a large bending power is provided by the solenoid
 1114 magnet made of superconducting cable capable to generate a 3.8 T magnetic field.
 1115 The detection system is composed of (from the innermost to the outermost)

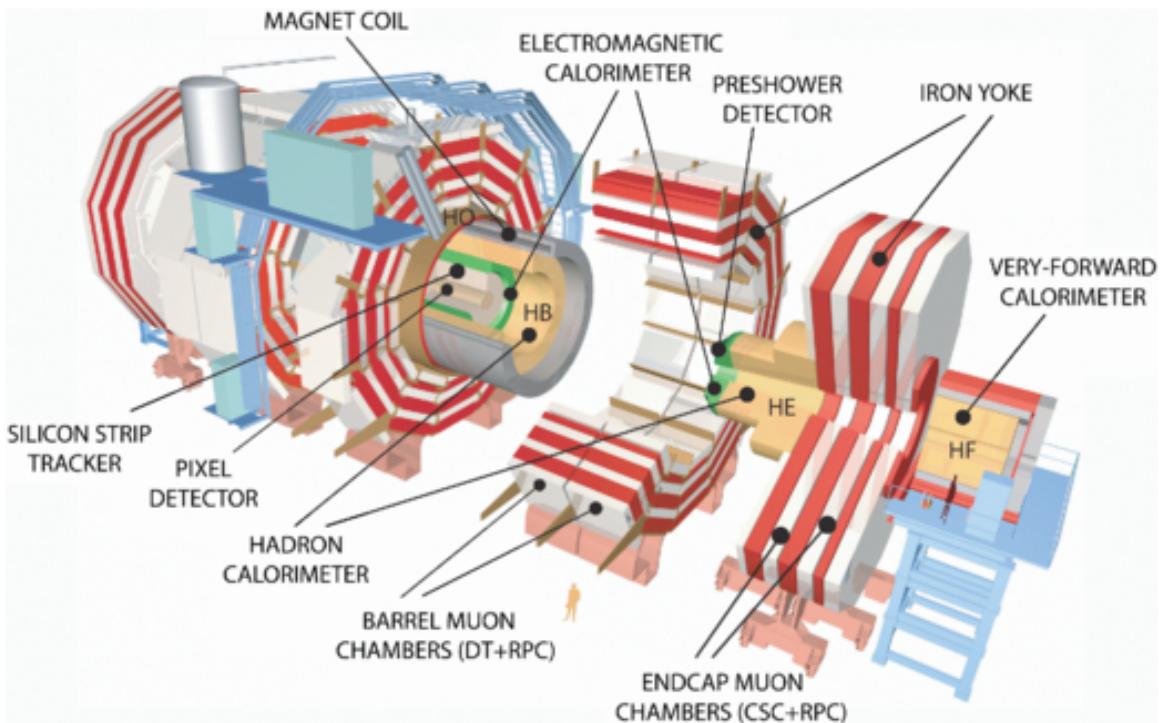


Figure 3.9: Layout of the CMS detector. The several subdetectors are indicated. The central region of the detector is referred as the Barrel section while the endcaps are referred as the forward sections. [68].

1116 • Pixel detector.

1117 • Silicon strip tracker.

1118 • Preshower detector.

1119 • Electromagnetic calorimeter.

1120 • Hadronic calorimeter.

1121 • Muon chambers (Barrel and endcap)

1122 The central region of the detector is commonly referred as the barrel section while the
 1123 endcaps are referred as the forward sections of the detector; thus, each subdetector
 1124 is composed of a barrel section and a forward section.

1125 3.3.1 Coordinate system

1126 The coordinate system used by CMS is centered in the geometrical center of the
 1127 detector which is the same as the CP as shown in Figure 3.10. The z -axis is parallel
 1128 to the beam direction, while the Y -axis pointing vertically upward, and the X -axis
 1129 pointing radially inward toward the center of the LHC.

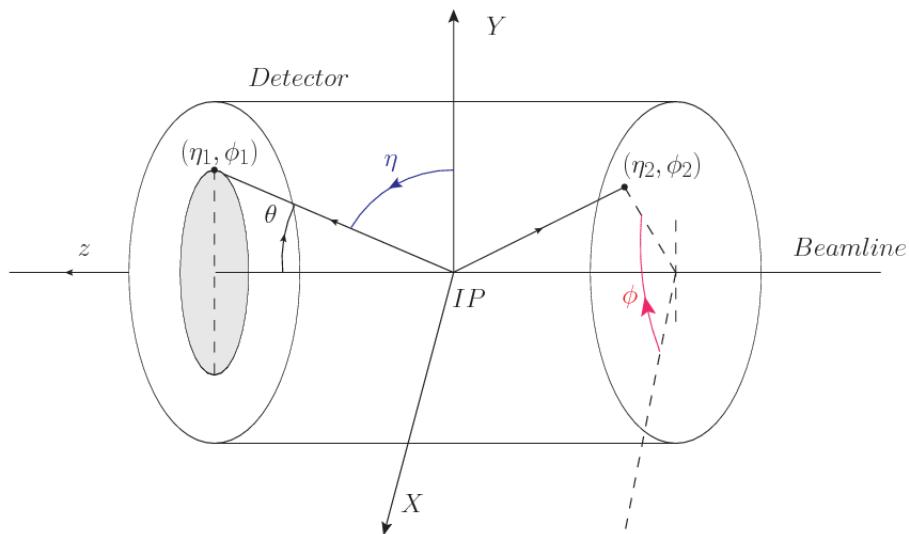


Figure 3.10: CMS detector coordinate system.

1130 In addition to the common cartesian and cylindrical coordinate systems, two coor-
 1131 dinates are of particular utility in particle physics: rapidity (y) and pseudorapidity

1132 (η) , defined in connection to the polar angle θ , energy and longitudinal momentum
 1133 component (momentum along the z -axis) according to

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

1134 Rapidity is related to the angle between the XY -plane and the direction in which the
 1135 products of a collision are emitted; it has the nice property that the difference between
 1136 the rapidities of two particles is invariant with respect to Lorentz boosts along the z -
 1137 axis. Thus, data analysis becomes more simple when based on rapidity; however, it is
 1138 not simple to measure the rapidity of highly relativistic particles, as those produced
 1139 after pp collisions. Under the highly relativistic motion approximation, y can be
 1140 rewritten in terms of the polar angle, concluding that rapidity is approximately equal
 1141 to the pseudorapidity defined above, i.e., $y \approx \eta$. Note that η is easier to measure than y
 1142 given the direct relationship between the former and the polar angle. Angular distance
 1143 between two objects in the detector (ΔR) is defined in terms of their coordinates
 1144 $(\eta_1, \phi_1), (\eta_2, \phi_2)$ as

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

1145 3.3.2 Pixels detector

1146 The CMS tracking system is designed to provide a precise measurement of the tra-
 1147 jectory (*track*) followed by the charged particles created after the pp collisions; also,
 1148 the precise reconstruction of the primary and secondary origins (*vertices*) is expected
 1149 in an environment where, each 25 ns, the bunch crossing produce about 20 inelastic
 1150 collisions and about 1000 particles. An increment in the luminosity is ongoing which
 1151 implies that the PU will increase accordingly.

1153 The pixel detector was replaced during the 2016-2017 extended year-end technical
1154 stop, due to the increasingly challenging operating conditions like the higher particle
1155 flow and more radiation harsh environment, among others. The new one is responding
1156 as expected, reinforcing its crucial role in the successful way to fulfill the new LHC
1157 physics objectives after the discovery of the Higgs boson. The last chapter of this
1158 thesis is dedicated to describe my contribution to the *Forward Pixel Phase 1 upgrade*.

1159

1160 The current pixel detector is composed of 1856 silicon pixel detector modules orga-
1161 nized in four-barrel layers in the central region and three disks in the forward region;
1162 it is designed to record efficiently and with high precision, up to $10\mu\text{m}$ in the XY -
1163 plane and $20\mu\text{m}$ in the z -direction, the first four space-points (*hits*) near to the CP
1164 region (see Figure 3.11 left side) in the range $|\eta| \leq 2.5$. The first barrel layer is located
1165 at a radius of 30 mm from the beamline, while the fourth layer is located at a radius
1166 of 160 mm closer to the strip tracker inner barrel layer (see Section 3.3.3) in order to
1167 reduce the rate of fake tracks. The high granularity of the detector is represented in
1168 its about 123 Mpixels, each of size $100 \times 150\mu\text{m}^2$, which is almost twice the channels
1169 of the old detector. The transverse momentum resolution of tracks can be measured
1170 with a resolution of 1-2% for muons of $p_T = 100$ GeV.

1171

1172 Some of the improvements with respect to the previous pixel detector include a higher
1173 average tracking efficiency and lower average fake rate as well as higher track impact
1174 parameter resolution which is fundamental in order to increase the efficiency in the
1175 identification of jets originating from b quarks (b-tagging). A significant source of
1176 improvement comes from the overall reduction in the material budget of the detector
1177 which results in fewer photon conversions and less multiple scattering from charged
1178 particles.

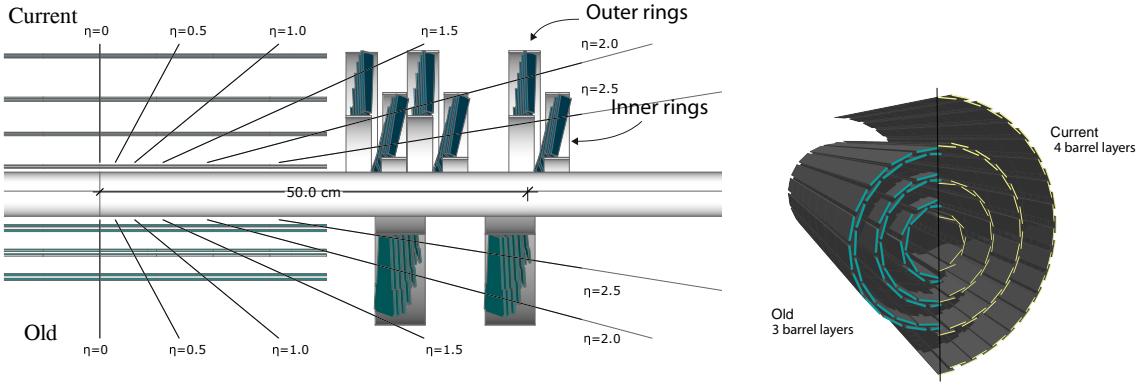


Figure 3.11: CMS pixel detector schematic view. Left: layout comparing the layers and disks in the old and current pixel detectors. Right: Transverse-oblique view comparing the pixel barrel layers in the two [70].

1179 3.3.3 Silicon strip tracker

1180 The silicon strip tracker (SST) is the second stage in the CMS tracking system. The
 1181 top side of Figure 3.12 shows a schematic of the SST. The inner tracker region is com-
 1182 posed of the tracker inner barrel (TIB) and the tracker inner disks (TID) covering the
 1183 region $r < 55$ cm and $|z| < 118$ cm. The TIB is composed of 4 layers while the TID
 1184 is composed of 3 disks at each end. The silicon sensors in the inner tracker are 320
 1185 μm thick, providing a resolution of about 13-38 μm in the $r\phi$ position measurement.

1186

1187 The modules indicated in blue in the schematic view of Figure 3.12 are two modules
 1188 mounted back-to-back and rotated in the plane of the module by a *stereo* angle of
 1189 100 mrad; the hits from these two modules, known as *stereo hits*, are combined to
 1190 provide a measurement of the second coordinate (z in the barrel and r on the disks)
 1191 allowing the reconstruction of hit positions in 3-D.

1192

1193 The outer tracker region is composed of the tracker outer barrel (TOB) and the
 1194 tracker endcaps (TEC). The six layers of the TOB offer coverage in the region $r > 55$

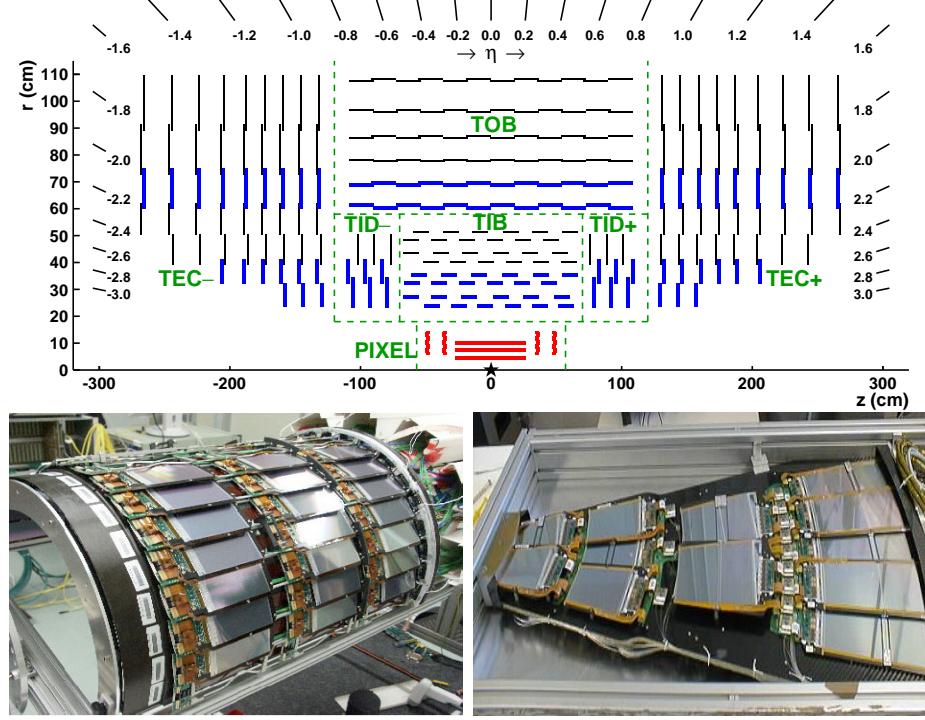


Figure 3.12: Top: CMS Silicon Strip Tracker (SST) schematic view. The SST is composed of the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). Each part is made of silicon strip modules; the modules in blue represent two modules mounted 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 [71–73].

1195 cm and $|z| < 118$ cm, while the 9 disks of the TEC cover the region $124 < |z| < 282$
 1196 cm. The resolution offered by the outer tracker is about $13\text{--}38 \mu\text{m}$ in the $r\phi$ position
 1197 measurement. The inner four TEC disks use silicon sensors $320 \mu\text{m}$ thick; those in
 1198 the TOB and the outer three TEC disks use silicon sensors of $500 \mu\text{m}$ thickness. The
 1199 silicon strips run parallel to the z -axis and the distance between strips varies from 80
 1200 μm in the inner TIB layers to $183 \mu\text{m}$ in the inner TOB layers; in the endcaps the
 1201 wedge-shaped sensors with radial strips, whose pitch range between $81 \mu\text{m}$ at small
 1202 radii and $205 \mu\text{m}$ at large radii.

1203

1204 The whole SST has 15148 silicon modules, 9.3 million silicon strips and cover a total

1205 active area of about 198 m^2 .

1206 **3.3.4 Electromagnetic calorimeter**

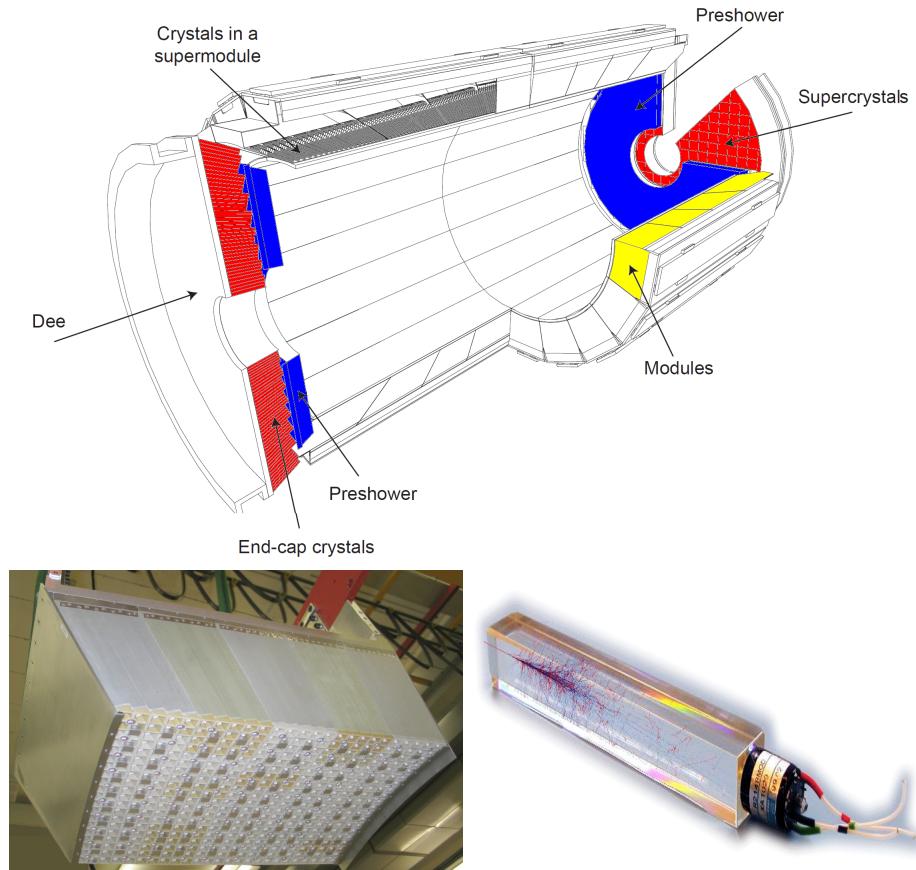


Figure 3.13: Top: CMS ECAL schematic view. Bottom: Module equipped with the crystals (left); ECAL crystal(right).

1207 The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of
 1208 electrons and photons. It is composed of 75848 lead tungstate crystals which have a
 1209 short radiation length (0.89 cm) and fast response, since 80% of the light is emitted
 1210 within 25 ns; however, they are combined with Avalanche photodiodes (APDs) as
 1211 photodetectors given that crystals themselves have a low light yield ($30\gamma/\text{MeV}$). A

1212 schematic view of the ECAL is shown in Figure 3.13.

1213

1214 Energy is measured when electrons and photons are absorbed by the crystals which
 1215 generates an electromagnetic *shower*, as seen in bottom right picture of the Figure
 1216 3.13; the shower is seen as a *cluster* of energy which depending on the amount of en-
 1217 ergy deposited can involve several crystals. The ECAL barrel (EB) covers the region
 1218 $|\eta| < 1.479$, using crystals of depth of 23 cm and $2.2 \times 2.2 \text{ cm}^2$ transverse section;
 1219 the ECAL endcap (EE) covers the region $1.479 < |\eta| < 3.0$ using crystals of depth
 1220 22 cm and transverse section of $2.86 \times 2.86 \text{ cm}^2$; the photodetectors used are vacuum
 1221 phototriodes (VPTs). Each EE is divided in two structures called *Dees*.

1222

1223 In front of the EE, it is installed the preshower detector (ES) which covers the region
 1224 $1.653 < |\eta| < 2.6$. The ES provides a precise measurement of the position of electro-
 1225 magnetic showers, which allows to distinguish electrons and photons signals from π^0
 1226 decay signals. The ES is composed of a layer of lead absorber followed by a layer of
 1227 plastic scintillators

1228 3.3.5 Hadronic calorimeter

1229 Hadrons are not absorbed by the ECAL but by the hadron calorimeter (HCAL),
 1230 which is made of a combination of alternating brass absorber layers and silicon photo-
 1231 multiplier(SiPM) layers; therefore, particles passing through the scintillator material
 1232 produce showers, as in the ECAL, as a result of the inelastic scattering of the hadrons
 1233 with the detector material. Since the particles are not absorbed in the scintillator,
 1234 their energy is sampled; therefore the total energy is not measured but estimated from
 1235 the energy clusters, which reduce the resolution of the detector. Brass was chosen

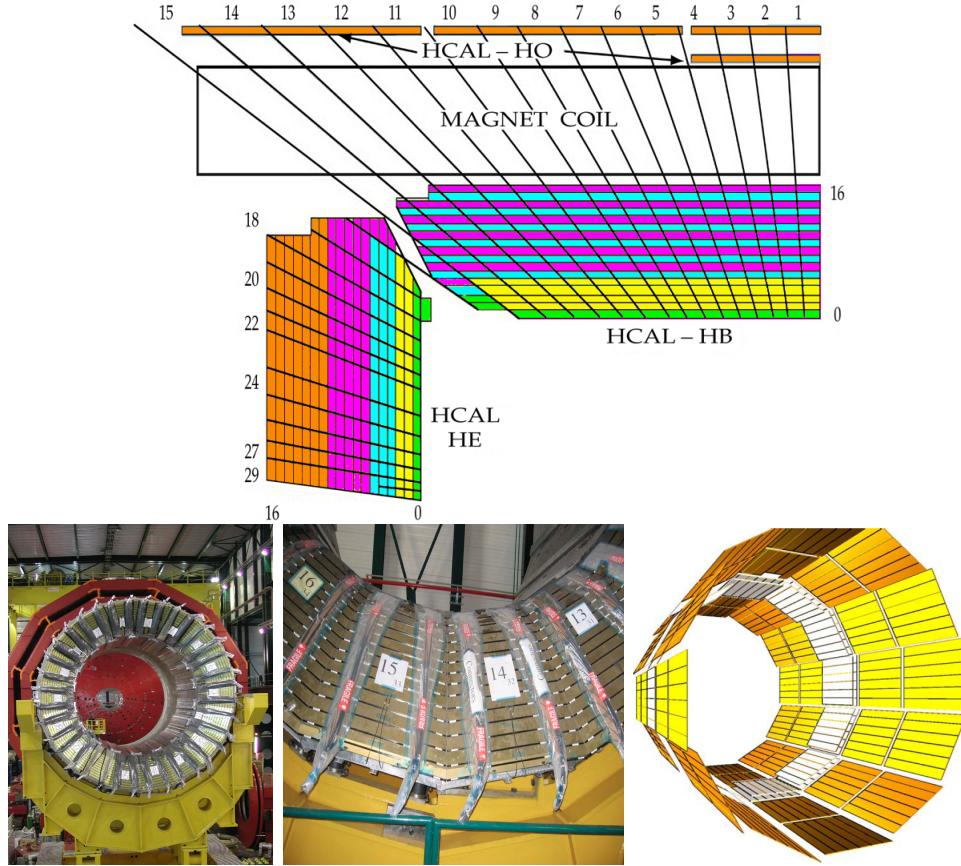


Figure 3.14: Top: CMS HCAL schematic view, the colors indicate the layers that are grouped into the same readout channels. Bottom: picture of a section of the HB; the absorber material is the golden region and scintillators are placed in between the absorber material (left and center). Schematic view of the HO (right). [74,75]

as the absorber material due to its short interaction length ($\lambda_I = 16.42\text{cm}$) and its non-magnetivity. Figure 3.14 shows a schematic view of the CMS HCAL.

1238

1239 The HCAL is divided into four sections; the Hadron Barrel (HB), the Hadron Outer
 1240 (HO), the Hadron Endcap (HE) and the Hadron Forward (HF) sections. The HB
 1241 covers the region $0 < |\eta| < 1.4$, while the HE covers the region $1.3 < |\eta| < 3.0$. The HF,
 1242 made of quartz fiber scintillator and steel as absorption material, covers the forward
 1243 region $3.0 < |\eta| < 5.2$. Both the HB and HF are located inside the solenoid. The HO

is placed outside the magnet as an additional layer of scintillators with the purpose of measure the energy tails of particles passing through the HB and the magnet (see Figure 3.14 top and bottom right). The upgrades made to the HCAL during the technical stop 2016-2017 consisted in the replacement of the photo transducer, in order to improve the efficiency.

3.3.6 Superconducting solenoid magnet

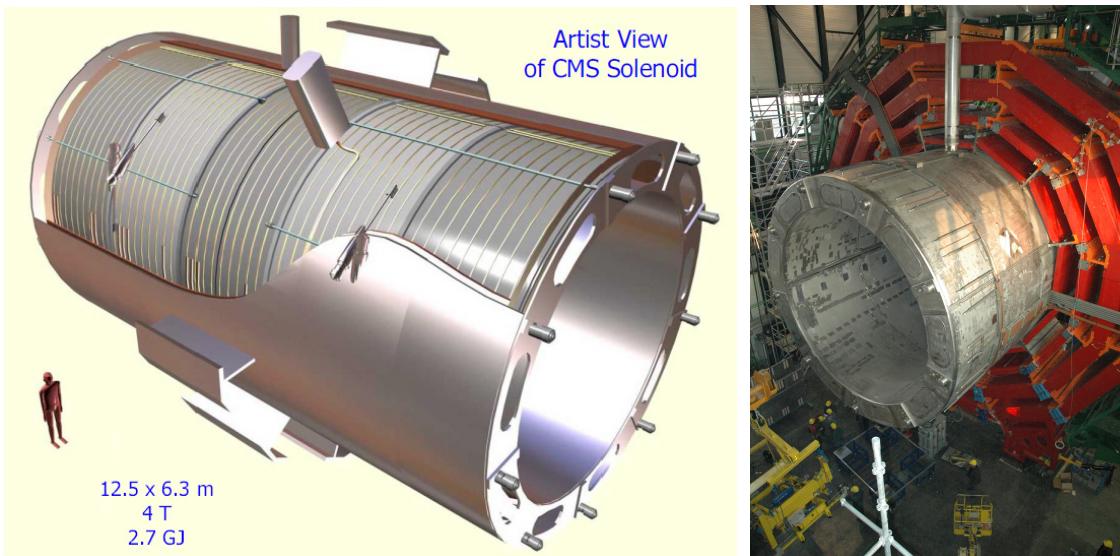


Figure 3.15: Artistic representation of the CMS solenoid magnet(left). The magnet is supported on an iron yoke (right) which also serves as the house of the muon detector and as mechanical support for the whole CMS detector [69].

The superconducting magnet installed in the CMS detector is designed to provide an intense and highly uniform magnetic field in the central part of the detector. In fact, the tracking system takes advantage of the bending power of the magnetic field to measure with precision the momentum of the particles that traverse it; the unambiguous determination of the sign for high momentum muons was a driven principle during the design of the magnet. The magnet has a diameter of 6.3 m, a length of 12.5 m and a cold mass of 220 t; the generated magnetic field reaches a strength of 3.8T.

1257 Since it is made of Ni-Tb superconducting cable it has to operate at a temperature
 1258 of 4.7 K by using a helium cryogenic system; the current circulating in the cables
 1259 reaches 18800 A under normal running conditions. The left side of Figure 3.15 shows
 1260 an artistic view of the CMS magnet, while the right side shows a transverse view of
 1261 the cold mass where the winding structure is visible.

1262

1263 The yoke (see Figure 3.15), composed of 5 barrel wheels and 6 endcap disks made
 1264 of iron, serves not only as the media for magnetic flux return but also provides the
 1265 house for the muon detector system and structural stability to the full detector.

1266 3.3.7 Muon system

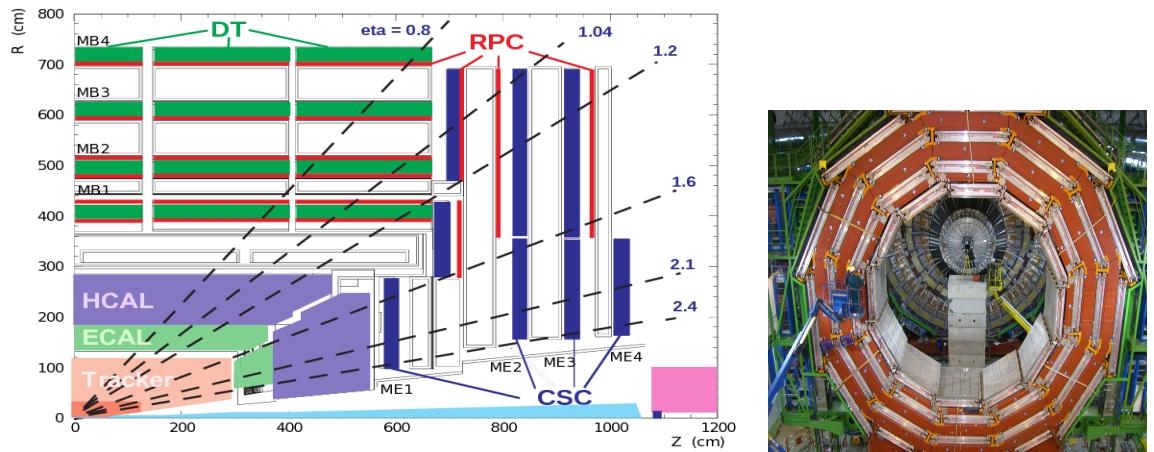


Figure 3.16: Left: CMS muon system schematic view; Right: one of the yoke rings with the muon DTs and RPCs installed; in the back it is possible to see the muon endcap [76].

1267 Muons are the only charged particles able to pass through all the CMS detector due
 1268 to their low ionization energy loss; thus, muons can be separated easily from the
 1269 high amount of particles produced in a pp collision. Also, muons are expected to be
 1270 produced in the decay of several new particles; therefore, a good detection of muons

1271 was on the leading principles when designing the CMS detector.

1272

1273 The CMS muon detection system (muon spectrometer) is embedded in the return
1274 yoke as seen in Figure 3.16. It is composed of three different detector types, the drift
1275 tube chambers (DT), Cathode strip chambers (CSC), and resistive plate chambers
1276 (RPC); DT are located in the central region $\eta < 1.2$ arranged in four layers of drift
1277 chambers filled with an Ar/CO₂ gas mixture.

1278

1279 The muon endcaps are made of CSCs covering the region $\eta < 2.4$ and filled with a
1280 mixture of Ar/CO₂/CF₄. The reason behind using a different detector type lies on
1281 the different conditions in the forward region like the higher muon rate and higher
1282 residual magnetic field compared to the central region.

1283

1284 The third type of detector used in the muon system is a set of four disks of RPCs
1285 working in avalanche mode. The RPCs provide good spatial and time resolutions.
1286 The track of $high - p_T$ muon candidates is built combining information from the
1287 tracking system and the signal from up to six RPCs and four DT chambers.
1288 The muon tracks are reconstructed from the hits in the several layers of the muon
1289 system.

1290 3.3.8 CMS trigger system

1291 Under normal conditions, CMS expects pp collisions every 25 ns, i.e., an interaction
1292 rate of 40 MHz for which it is not possible to store the recorded data in full. In order
1293 to handle this high event rate data, an online event selection, known as triggering, is
1294 performed; triggering reduce the event rate to 100 Hz for storage and further offline

1295 analysis.

1296

1297 The trigger system starts with a reduction of the event rate to 100 kHz in the so-
 1298 called *level 1 trigger (L1)*. L1 is based on dedicated programmable hardware like
 1299 Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated Cir-
 1300 cuits (ASICs), partly located in the detector itself; another portion is located in the
 1301 CMS under-ground cavern. Hit patterns information from the muon chambers and
 1302 the energy deposits in the calorimeter are used to decide if an event is accepted or
 1303 rejected, according to selection requirements previously defined, which reflect the in-
 1304 teresting physics processes. Figure 3.17 shows the L1 trigger architecture.

1305

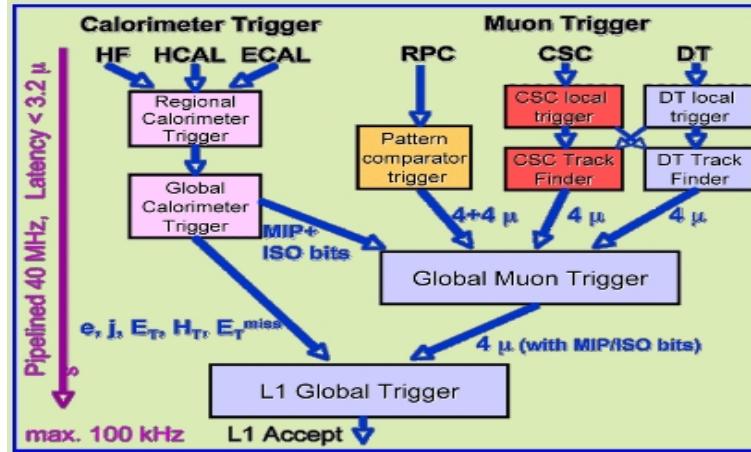


Figure 3.17: CMS Level-1 trigger architecture [77].

1306 The second stage in the trigger system is called *high-level trigger (HLT)*; events ac-
 1307 cepted by L1 are passed to HLT in order to make an initial reconstruction of them.
 1308 HLT is software based and runs on a dedicated server farm, using selection algo-
 1309 rithms and high-level object definitions; the event rate at HLT is reduced to 100 Hz.
 1310 The first HLT stage takes information from the muon detectors and the calorimeters
 1311 to make the initial object reconstruction; in the next HLT stage, information from

1312 the pixel and strip detectors is used to do first fast-tracking and then full tracking
1313 online. This initial object reconstruction is used in further steps of the trigger system.

1314

1315 Events and preliminary reconstructed physics objects from HLT are sent to be fully
1316 reconstructed at the CERN computing center. Again, the pixel detector information
1317 provides high-quality seeds for the track reconstruction algorithm offline, primary ver-
1318 tex reconstruction, electron and photon identification, muon reconstruction, τ iden-
1319 tification, and b-tagging. After full reconstruction, data sets are made available for
1320 offline analyses.

1321

1322 During the 2016-2017 technical stop, the L1 system was updated in order to improve
1323 the physics object identification by improving the algorithms and accounting for the
1324 increasing pile-up scenario.

1325 **3.3.9 CMS computing**

1326 After the data, coming from the experiment, are processed at several levels, they have
1327 to be stored and made available for further analysis; in order to cope all the tasks
1328 implied in the offline data processing, like transfer, simulation, reconstruction and
1329 reprocessing, among others, a big computing power is required. The CMS computing
1330 system is based on the distributed architecture concept, where users of the system
1331 and physical computer centers are distributed worldwide and interconnected by high-
1332 speed networks.

1333 The worldwide LHC computing grid (WLCG) is the mechanism used to provide that
1334 distributed environment. WLCG is a tiered structure connecting computing centers
1335 around the world, which provides the necessary storage and computing facilities. The

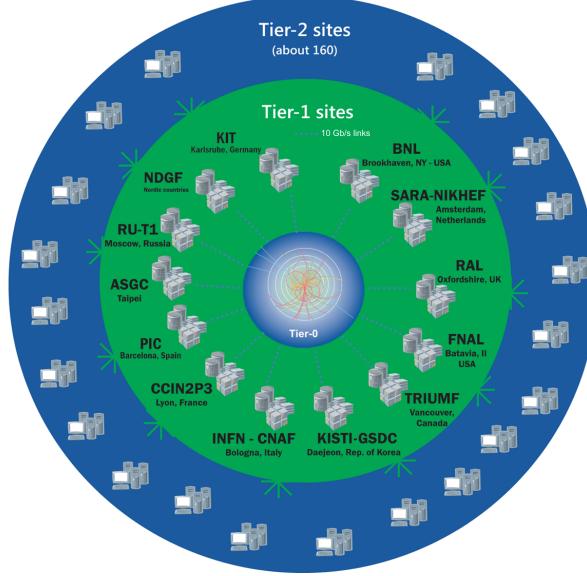


Figure 3.18: WLCG structure. The primary computer centers (Tier-0) are located at CERN (data center) and at the Wigner datacenter in Budapest. Tier-1 is composed of 13 centers and Tier-2 is composed of about 160 centers. [78].

1336 primary computing centers of the WLCG are located at the CERN and the Wigner
 1337 datacenter in Budapest and are known as Tier-0 as shown in Figure 3.18. The main
 1338 responsibilities for each tier level are [78]

- 1339 • **Tier-0:** initial reconstruction of recorded events and storage of the resulting
 1340 datasets, the distribution of raw data to the Tier-1 centers.
- 1341 • **Tier-1:** provide storage capacity, support for the Grid, safe-keeping of a pro-
 1342 portional share of raw and reconstructed data, large-scale reprocessing and safe-
 1343 keeping of corresponding output, generation of simulated events, distribution
 1344 of data to Tier 2s, safe-keeping of a share of simulated data produced at these
 1345 Tier 2s.
- 1346 • **Tier-2:** store sufficient data and provide adequate computing power for specific
 1347 analysis tasks, provide analysis requirements and proportional share of simu-
 1348 lated event production and reconstruction.

1349 Aside from the general computing strategy to manage the huge amount of data pro-
1350 duced by experiments, CMS uses a framework to perform a variety of processing,
1351 selection and analysis tasks. The central concept of the CMS data model referred to
1352 as *event data model* (EDM) is the *Event*; therefore, an event is the unit that contains
1353 the information from a single bunch crossing as well as any data derived from that
1354 information like the reconstructed objects, the details under which additional data
1355 are derived.

1356

1357 Events are passed as the input to the *physics modules* that obtain information from
1358 them and create new one; for instance, *event data producers* add new data into the
1359 events, *analyzers* produce an information summary from an event set, *filters* perform
1360 selection and triggering.

1361

1362 CMS uses several event formats with different levels of detail and precision

1363 • **Raw format:** events in this format contain the full recorded information from
1364 the detector as well as trigger decision and other metadata. An extended version
1365 of raw data is used to store information from the CMS Monte Carlo simulation
1366 tools. Raw data are stored permanently, occupying about 2MB/event

1367 • **RECO format:** events in this format correspond to raw data that have been
1368 submitted to reconstruction algorithms like primary and secondary vertex re-
1369 construction, particle ID, track-finding. RECO events contain physical objects
1370 and all the information used to reconstruct them; average size is about 0.5
1371 MB/event.

1372 • **AOD format:** Analysis Object Data (AOD) is the data format used in the
1373 physics analyses given that it contains the parameters describing the high-level

1374 physics objects in addition to enough information to allow a kinematic refitting if
 1375 needed. AOD events are filtered versions of the RECO events to which skimming
 1376 or other kind processes have been applied. Requires about 100 kB/event.

1377 • **Non-event data** are data needed to interpret and reconstruct events. Some
 1378 of the non-event data used by CMS contains information about the detector
 1379 contraction and condition data like calibrations, alignment, and detector status.

1380 Figure 3.19 shows the data flow scheme between CMS detector and hardware tiers.

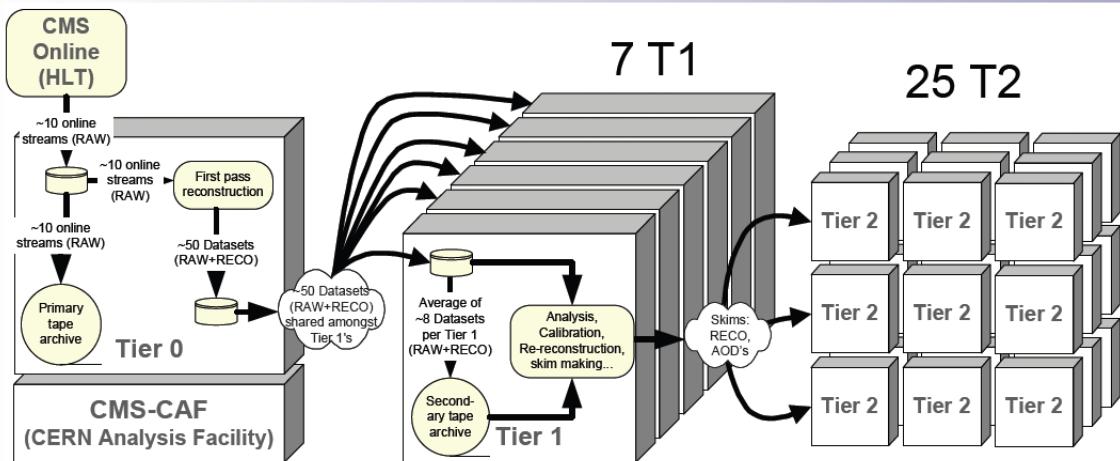


Figure 3.19: Data flow from CMS detector through hardware Tiers.

1381 The whole collection of software built as a framework is referred to as *CMSSW*. This
 1382 framework provides the services needed by the simulation, calibration and alignment,
 1383 and reconstruction modules that process event data, so that physicists can perform
 1384 analysis. The CMSSW event processing model is composed of one executable, called
 1385 cmsRun, and several plug-in modules which contains all the tools (calibration, recon-
 1386 struction algorithms) needed to process an event. The same executable is used for
 1387 both detector and Monte Carlo data [79].

1388 **Chapter 4**

1389 **Event generation, simulation and
1390 reconstruction**

1391 The process of analyzing data recorded by the CMS experiment involves several stages
1392 where the data are processed in order to interpret the information provided by all
1393 the detection systems; in those stages, the particles produced after the pp collision
1394 are identified by reconstructing their trajectories and measuring their features. In
1395 addition, the SM provides a set of predictions that have to be compared with the
1396 experimental results; however, in most of the cases, theoretical predictions are not
1397 directly comparable to experimental results due to the diverse source of uncertainties
1398 introduced by the experimental setup and theoretical approximations, among others.

1399

1400 The strategy to face these conditions consists in using statistical methods imple-
1401 mented in computational algorithms to produce numerical results that can be con-
1402 trasted with the experimental results. These computational algorithms are commonly
1403 known as Monte Carlo (MC) methods and, in the case of particle physics, they are
1404 designed to apply the SM rules and produce predictions about the physical observ-

ables measured in the experiments. Since particle physics is governed by quantum mechanics principles, predictions are not allowed from single events; therefore, a high number of events are *generated* and predictions are produced in the form of statistical distributions for the observables. Effects of the detector presence are included in the predictions by introducing simulations of the detector itself.

1410

1411 This chapter presents a description of the event generation strategy and the tools
 1412 used to perform the detector simulation and physics objects reconstruction. A com-
 1413 prehensive review of event generators for LHC physics can be found in Reference [80]
 1414 on which this chapter is based.

1415 4.1 Event generation

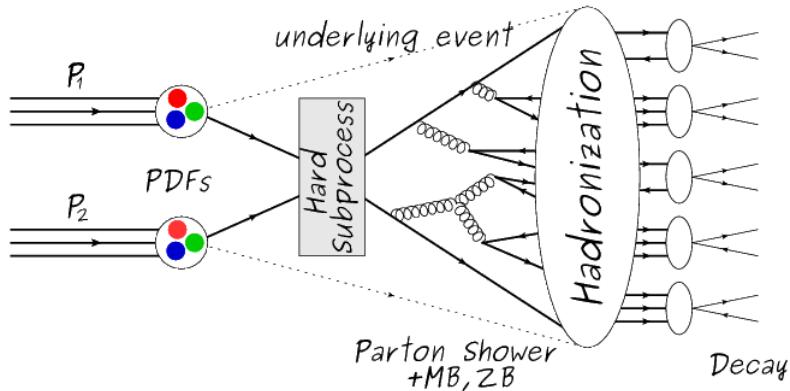


Figure 4.1: Event generation process. The actual interaction is generated in the hard subprocess. The parton shower describes the evolution of the partons from the hard subprocess. Modified from Reference [81].

1416 The event generation is intended to create events that mimic the behavior of actual
 1417 events produced in collisions; they obey a sequence of steps from the particles collision
 1418 hard process to the decay process into the final state. Figure 4.1 shows a schematic
 1419 view of the event generation process; the fact that the full process can be treated as

1420 several independent steps is motivated by the QCD factorization theorem.

1421

1422 Generation starts by taking into account the PDFs of the incoming particles. Event
 1423 generators offer the option to chose from several PDF sets depending on the particu-
 1424 lar process under simulation¹; in the following, pp collisions will be considered. The
 1425 *hard subprocess* describes the actual interaction between partons from the incoming
 1426 protons; it is represented by the matrix element connecting the initial and final states
 1427 of the interaction. Normally, the matrix element can be written as a sum over Feyn-
 1428 man diagrams and consider interferences between terms in the summation. During
 1429 the generation of the hard subprocess, the production cross section is calculated.

1430

1431 The order to which the cross section is calculated depends on the order of the Feyn-
 1432 man diagrams involved in the calculation; therefore, radiative corrections are included
 1433 by considering a higher order Feynman diagrams where QCD radiation dominates.
 1434 Currently, cross sections calculated to LO do not offer a satisfactory description of the
 1435 processes, i.e., the results are only reliable for the shape of distributions; therefore,
 1436 NLO calculations have to be performed with the implication that the computing time
 1437 needed is highly increased.

1438

1439 The final parton content of the hard subprocess is subjected to the *parton shower*
 1440 which generates the gluon radiation. Parton shower evolves the partons, i.e., glouns
 1441 split into quark-antiquark pairs and quarks with enough energy radiate gluons giv-
 1442 ing rise to further parton multiplication, following the DGLAP (Dokshitzer-Gribov-
 1443 Lipatov-Altarelli-Parisi) equations. Showering continues until the energy scale is low
 1444 enough to reach the non-perturbative limit.

¹ Tool in Reference [82] allows to plot different PDF sets under customizable conditions.

1445

1446 In the simulation of LHC processes that involve b quarks, like the single top quark
 1447 or Higgs associated production, it is needed to consider that the b quark is heavier
 1448 than the proton; hence, the QCD interaction description is made in two different
 1449 schemes [83]

1450 • four-flavor (4F) scheme. b quarks appear only in the final state because they
 1451 are heavier than the proton and therefore they can be produced only from the
 1452 splitting of a gluon into pairs or singly in association with a t quark in high
 1453 energy-scale interactions; furthermore, during the simulation, the b -PDFs are set
 1454 to zero. Calculations in this scheme are more complicated due to the presence
 1455 of the second b quark but the full kinematics is considered already at LO and
 1456 therefore the accuracy of the description is better.

1457 • five-flavor (5F) scheme. b quarks are considered massless, therefore they can
 1458 appear in both initial and final states since they can now be part of the proton;
 1459 thus, during the simulation b -PDFs are not set to zero. In this scheme, calcu-
 1460 lations are simpler than in the 4F scheme and possible logarithmic divergences
 1461 are absorbed by the PDFs through the DGLAP evolution.

1462 In this thesis, the tHq events are generated using the 4F scheme in order to reduce
 1463 uncertainties, while the tHW events are generated using the 5F scheme to eliminate
 1464 LO interference with $t\bar{t}H$ process [48].

1465

1466 Partons involved in the pp collision are the focus of the simulation, however, the rest
 1467 of the partons inside the incoming protons are also affected because the remnants are
 1468 colored objects; also, multiple parton interactions can occur. The hadronization of
 1469 the remnants and multiple parton interactions are known as *underlying event* and it

1470 has to be included in the simulation. In addition, multiple pp collisions in the same
 1471 bunch crossing (pile-up mentioned in 3.2) occurs, actually in two forms

1472 • *in-time PU* which refers to multiple pp collision in the bunch crossing but that
 1473 are not considered as primary vertices.

1474 • *Out-of-time PU* which refers to overlapping pp collisions from consecutive bunch
 1475 crossings; this can occur due to the time-delays in the detection systems where
 1476 information from one bunch crossing is assigned to the next or previous one.

1477 While the underlying event effects are included in generation using generator-specific
 1478 tools, PU effects are added to the generation by overlaying Minimum-bias (MB) and
 1479 Zero-bias (ZB) events to the generated events. MB events are inelastic events se-
 1480 lected by using a loose trigger with as little bias as possible, therefore accepting a
 1481 large fraction of the overall inelastic event; ZB events correspond to random events
 1482 recorded by the detector when collisions are likely. MB models in-time PU and ZB
 1483 models out-of-time PU.

1484

1485 The next step in the generation process is called *hadronization*. Since particles with
 1486 a net color charge are not allowed to exits isolated, they have to recombine to form
 1487 bound states. This is precisely the process by which the partons resulting from the
 1488 parton shower arrange themselves as color singlets to form hadrons. At this step, the
 1489 energy-scale is low and the strong coupling constant is large, therefore hadronization
 1490 process is non-perturbative and the evolution of the partons is described using phe-
 1491 nomenological models. Most of the baryons and mesons produced in the hadronization
 1492 are unstable and hence they will decay in the detector.

1493

1494 The last step in the generation process corresponds to the decay of the unstable
 1495 particles generated during hadronization; it is also simulated in the hadronization
 1496 step, based on the known branching ratios.

1497 **4.2 Monte Carlo Event Generators.**

1498 The event generation described in the previous section has been implemented in
 1499 several software packages for which a brief description is given.

1500 • **PYTHIA 8.** It is a program designed to perform the generation of high energy
 1501 physics events which describes the collisions between particles such as electrons
 1502 and protons. Several theories and models are implemented in it, in order to
 1503 describe physical aspects like hard and soft interaction, parton distributions,
 1504 initial and final-state parton showers, multiple parton interactions, beam rem-
 1505 nants, hadronization² and particle decay. Thanks to extensive testing, several
 1506 optimized parametrizations, known as *tunings*, have been defined in order to
 1507 improve the description of actual collisions to a high degree of precision; for
 1508 analysis at $\sqrt{s} = 13$ TeV, the underline event CUETP8M1 tune is employed [85].
 1509 The calculation of the matrix element is performed at LO which is not enough
 1510 for the current required level of precision; therefore, pythia is often used for
 1511 parton shower, hadronization and decays, while other event generators are used
 1512 to generate the matrix element at NLO.

1513 • **MadGraph5_aMC@NLO.** MadGraph is a matrix element generator which
 1514 calculates the amplitudes for all contributing Feynman diagrams of a given pro-
 1515 cess but does not provide a parton shower while MC@NLO incorporates NLO

² based in the Lund string model [84]

1516 QCD matrix elements consistently into a parton shower framework; thus, Mad-
 1517 Graph5_aMC@NLO, as a merger of the two event generators MadGraph5 and
 1518 aMC@NLO, is an event generator capable to calculate tree-level and NLO cross
 1519 sections and perform the matching of those with the parton shower. It is one of
 1520 the most frequently used matrix element generators; however, it has the partic-
 1521 ular feature of the presence of negative event weights which reduce the number
 1522 of events used to reproduce the properties of the objects generated [86].

1523

1524 • **POWHEG.** It is an NLO matrix element generator where the hardest emis-
 1525 sion of color charged particles is generated in such a way that the negative event
 1526 weights issue of MadGraph5_aMC@NLO is overcome; however, the method re-
 1527 quires an interface with p_T -ordered parton shower or a parton shower generator
 1528 where this highest emission can be vetoed in order to avoid double counting of
 1529 this highest-energetic emission. PYTHIA is a commonly matched to POWHEG
 1530 event generator [87].

1531 Events resulting from the whole generation process are known as MC events.

1532 4.3 CMS detector simulation.

1533 After generation, MC events contain the physics of the collisions but they are not
 1534 ready to be compared to the events recorded by the experiment since these recorded
 1535 events correspond to the response of the detection systems to the interaction with
 1536 the particles traversing them. The simulation of the CMS detector has to be applied
 1537 on top of the event generation; it is simulated with a MC toolkit for the simulation
 1538 of particles passing through matter called Geant4 which is also able to simulate the

1539 electronic signals that would be measured by all detectors inside CMS.

1540

1541 The simulation takes the generated particles contained in the MC events as input,
1542 makes them pass through the simulated geometry, and models physics processes that
1543 particles experience during their passage through matter. The full set of results from
1544 particle-matter interactions corresponds to the simulated hit which contains informa-
1545 tion about the energy loss, momentum and position. Particles of the input event are
1546 called *primary*, while the particles originating from GEANT4-modeled interactions of
1547 a primary particle with matter are called a *secondary*. Simulated hits are the input
1548 of subsequent modules that emulate the response of the detector readout system and
1549 triggers. The output from the emulated detection systems and triggers is known as
1550 digitization [88,89].

1551

1552 The modeling of the CMS detector corresponds to the accurate modeling of the
1553 interaction among particles, the detector material, and the magnetic field. This
1554 simulation procedure includes the following standard steps

- 1555 • Modeling of the Interaction Region.
- 1556 • Modeling of the particle passage through the hierarchy of volumes that compose
1557 CMS detector and of the accompanying physics processes.
- 1558 • Modeling of the effect of multiple interactions per beam crossing and/or the
1559 effect of events overlay (Pile-Up simulation).
- 1560 • Modeling of the detector's electronics response, signal shape, noise, calibration
1561 constants (digitization).

1562 In addition to the full simulation, i.e., a detailed detector simulation, a faster simu-
 1563 lation (FastSim) have been developed, that may be used where much larger statistics
 1564 are required. In FastSim, detector material effects are parametrized and included in
 1565 the hits; those hits are used as input of the same higher-level algorithms³ used to an-
 1566 alyze the recorded events. In this way, comparisons between fast and full simulations
 1567 can be performed [91].

1568

1569 After the full detector simulation, the output events can be directly compared to
 1570 events actually recorded in the CMS detector. The collection of MC events that
 1571 reproduces the expected physics for a given process is known as MC sample.

1572 **4.4 Event reconstruction.**

1573 The CMS experiment use the *particle-flow event reconstruction algorithm (PF)* to do
 1574 the reconstruction of particles produced in pp collisions. Next sections will present
 1575 a basic description of the *Elements* used by PF (tracker tracks, energy clusters, and
 1576 muon tracks), based in the References [92, 93] where more detailed descriptions can
 1577 be found.

1578 **4.4.1 Particle-Flow Algorithm.**

1579 Each of the several sub detection systems of the CMS detector is dedicated to identify
 1580 a specific type of particles, i.e., photons and electrons are absorbed by the ECAL
 1581 and their reconstruction is based on ECAL information; hadrons are reconstructed
 1582 from clusters in the HCAL while muons are reconstructed from hits in the muon

³ track fitting, calorimeter clustering, b tagging, electron identification, jet reconstruction and calibration, trigger algorithms which will be considered in the next sections

1583 chambers. PF is designed to correlate signals from all the detector layers (tracks and
 1584 energy clusters) in order to reconstruct and identify each final state particle and its
 1585 properties as sketched in Figure 4.2.

1586

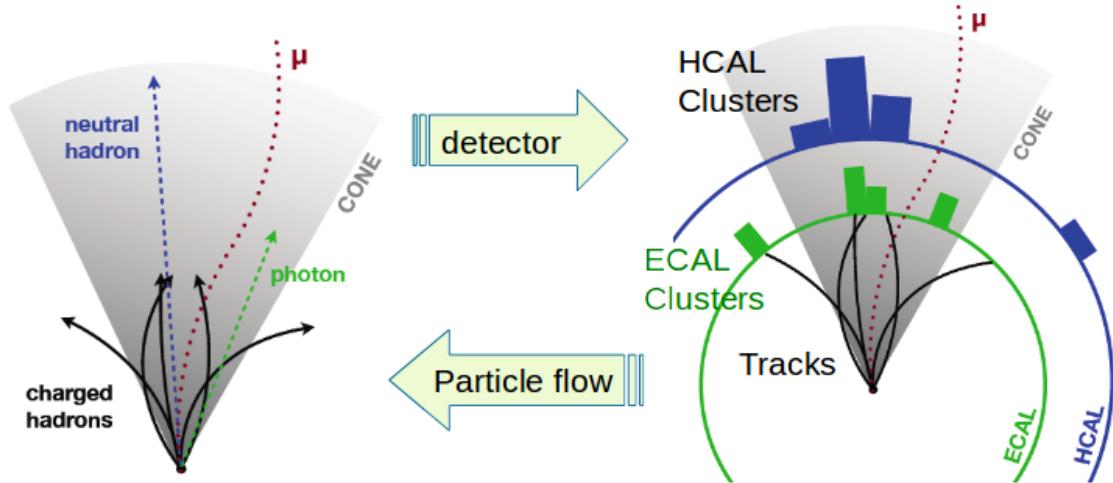


Figure 4.2: Particle flow algorithm. Information from the several CMS detection systems is provided as input to the algorithm which then combine it to identify and reconstruct all the particles in the final state and their properties. Reconstruction of simulated events is also performed by providing information from MC samples, detector and trigger simulation [94].

1587 For instance, a charged hadron is identified by a geometrical connection, known as
 1588 *link*, between one or more calorimeter clusters and a track in the tracker, provided
 1589 there are no hits in the muon system; combining several measurements allows a better
 1590 determination of the energy and charge sign of the charged hadron.

1591 Charged-particle track reconstruction.

1592 The strategy used by PF in order to reconstruct tracks is called *Iterative Tracking*
 1593 which occurs in four steps

- 1594 • Seed generation where initial track candidates are found by looking for a combi-
 1595 nation of hits in the pixel detector, strip tracker, and muon chambers. In total

1596 ten iterations are performed, each one with a different seeding requirement.
 1597 Seeds are used to estimate the trajectory parameters and uncertainties at the
 1598 time of the full track reconstruction. Seeds are also considered track candidates.

- 1599 • Track finding using a tracking software known as Combinatorial Track Finder
 1600 (CTF) [95]. The seed trajectories are extrapolated along the expected flight
 1601 path of a charged particle, in agreement to the trajectory parameters obtained
 1602 in the first step, in an attempt to find additional hits that can be assigned to
 1603 the track candidates.
- 1604 • Track-fitting where the found tracks are passed as input to a module which
 1605 provides the best estimate of the parameters of each trajectory.
- 1606 • Track selection where track candidates are submitted to a selection which dis-
 1607 cards those that fail a set of defined quality criteria.

1608 Iterations differ in the seeding configuration and the final track selection as elaborated
 1609 in References [92, 93]. In the first iteration, high p_T tracks and tracks produced near
 1610 to the interaction region are identified and those hits are masked thereby reducing
 1611 the combinatorial complexity. Next, iterations search for more complicated tracks,
 1612 like low p_T tracks and tracks from b hadron decays, which tend to be displaced from
 1613 the interaction region.

1614 **Vertex reconstruction.**

1615 During the track reconstruction, an extrapolation toward to the calorimeters is per-
 1616 formed in order to match energy deposits; that extrapolation is performed also toward
 1617 the beamline in order to find the origin of the track known as *vertex*. The vertex re-
 1618 construction is performed by selecting from the available reconstructed tracks, those

1619 that are consistent with being originated in the interaction region where pp collisions
 1620 are produced. The selection involves a requirement on the number of tracker (pixel
 1621 and strip) hits and the goodness of the track fit.

1622

1623 Selected tracks are clustered using a *deterministic annealing algorithm* (DA)⁴. A set
 1624 of candidate vertices and their associated tracks, resulting from the DA, are then fit-
 1625 ted with an *adaptive vertex fitter* (AVF) to produce the best estimate of the vertices
 1626 locations.

1627

1628 The p_T of the tracks associated to a reconstructed vertex is added, squared and used
 1629 to organize the vertices; the vertex with the highest squared sum is designated as the
 1630 *primary vertex* (PV) while the rest are designated as PU vertices.

1631 **Calorimeter clustering.**

1632 After traversing the CMS tracker system, electrons, photons and hadrons deposit their
 1633 energy in the ECAL and HCAL cells. The PF clustering algorithm aims to provide
 1634 a high detection efficiency even for low-energy particles and an efficient distinction
 1635 between close energy deposits. The clustering runs independently in the ECAL barrel
 1636 and endcaps, HCAL barrel and endcaps, and the two preshower layers, following two
 1637 steps

- 1638 • cells with an energy larger than a given seed threshold and larger than the energy
 1639 of the neighboring cells are identified as cluster seeds. The neighbor cells are
 1640 those that either share a side with the cluster seed candidate, or the eight closest
 1641 cells including cells that only share a corner with the seed candidate.

⁴ DA algorithm and AVF are described in detail in References [97, 98]

1642 • cells with at least a corner in common with a cell already in the cluster seed
 1643 and with an energy above a cell threshold are grouped into topological clusters.

1644 Clusters formed in this way are known as *particle-flow clusters*. With this clustering
 1645 strategy, it is possible to detect and measure the energy and direction of photons and
 1646 neutral hadrons as well as differentiate these neutral particles from the charged hadron
 1647 energy deposits. In cases involving charged hadrons for which the track parameters
 1648 are not determined accurately, for instance, low-quality and high- p_T tracks, clustering
 1649 helps in the energy measurements.

1650 **Electron track reconstruction.**

1651 Although the charged-particle track reconstruction described above works for elec-
 1652 trons, they lose a significant fraction of their energy via bremsstrahlung photon radi-
 1653 ation before reaching the ECAL; thus, the reconstruction performance depends on the
 1654 ability to measure also the radiated energy. The reconstruction strategy, in this case,
 1655 requires information from the tracking system and from the ECAL. Bremsstrahlung
 1656 photons are emitted at similar η values to that of the electron but at different values
 1657 of ϕ ; therefore, the radiated energy can be recovered by grouping ECAL clusters in a
 1658 η window over a range of ϕ around the electron direction. The group is called ECAL
 1659 supercluster.

1660

1661 Electron candidates from the track-seeding and ECAL super clustering are merged
 1662 into a single collection which is submitted to a full electron tracking fit with a
 1663 Gaussian-sum filter (GSF) [96]. The electron track and its associated ECAL su-
 1664 percluster form a *particle-flow electron*.

1665 **Muon track reconstruction.**

1666 Given that the CMS detector is equipped with a muon spectrometer capable to iden-
 1667 tify and measure the momentum of the muons traversing it, the muon reconstruction
 1668 is not specific to PF; therefore, three different muon types are defined

- 1669 • *Standalone muon.* A clustering on the DTs or CSCs hits is performed to form
 1670 track segments; those segments are used as seeds for the reconstruction in the
 1671 muon spectrometer. All DTs, CSCs, and RPCs hits along the muon trajectory
 1672 are combined and fitted to form the full track. The fitting output is called a
 1673 *standalone-muon track*.
- 1674 • *Tracker muon.* Each track in the inner tracker with p_T larger than 0.5 GeV and
 1675 a total momentum p larger than 2.5 GeV is extrapolated to the muon system.
 1676 A *tracker muon track* corresponds to a extrapolated track that matches at least
 1677 one muon segment.
- 1678 • *Global muon.* When tracks in the inner tracker (inner tracks) and standalone-
 1679 muon tracks are matched and turn out being compatibles, their hits are com-
 1680 bined and fitted to form a *global-muon track*.

1681 Global muons sharing the same inner track with tracker muons are merged into a
 1682 single candidate. PF muon identification uses the muon energy deposits in ECAL,
 1683 HCAL, and HO associated with the muon track to improve the muon identification.

1684 **Particle identification and reconstruction.**

1685 PF elements are connected by a linker algorithm that tests the connection between any
 1686 pair of elements; if they are found to be linked, a geometrical distance that quantifies
 1687 the quality of the link is assigned. Two elements may be linked indirectly through

1688 common elements. Linked elements form *PF blocks* and each PF block may contain
 1689 elements originating in one or more particles. Links can be established between
 1690 tracks, between calorimeter clusters, and between tracks and calorimeter clusters.
 1691 The identification and reconstruction start with a PF block and proceed as follows

1692 • Muons. An *isolated global muon* is identified by evaluating the presence of
 1693 inner track and energy deposits close to the global muon track in the (η, ϕ)
 1694 plane, i.e., in a particular point of the global muon track, inner tracks and
 1695 energy deposits are sought within a radius of $\Delta R = 0.3$ (see eqn. 3.7) from the
 1696 muon track; if they exit and the p_T of the found track added to the E_T of the
 1697 found energy deposit does not exceed 10% of the muon p_T then the global muon
 1698 is an isolated global muon. This isolation condition is stringent enough to reject
 1699 hadrons misidentified as muons.

1700 *Non-isolated global muons* are identified using additional selection requirements
 1701 on the number of track segments in the muon system and energy deposits along
 1702 the muon track. Muons inside jets are identified with more stringent criteria
 1703 in isolation and momentum as described in Reference [99]. The PF elements
 1704 associated with an identified muon are masked from the PF block.

1705 • Electrons are identified and reconstructed as described above plus some addi-
 1706 tional requirements on fourteen variables like the amount of energy radiated,
 1707 the distance between the extrapolated track position at the ECAL and the po-
 1708 sition of the associated ECAL supercluster, among others, which are combined
 1709 in an specialized multivariate analysis strategy that improves the electron iden-
 1710 tification. Tracks and clusters used to identify and reconstruct electrons are
 1711 masked in the PF block.

- 1712 • Isolated photons are identified from ECAL superclusters with E_T larger than 10
 1713 GeV, for which the energy deposited at a distance of 0.15, from the supercluster
 1714 position on the (η, ϕ) plane, does not exceed 10% of the supercluster energy;
 1715 note that this is an isolation requirement. In addition, there must not be links
 1716 to tracks. Clusters involved in the identification and reconstruction are masked
 1717 in the PF block.

- 1718 • Bremsstrahlung photons and prompt photons tend to convert to electron-positron
 1719 pairs inside the tracker, therefore, a dedicated finder algorithm is used to link
 1720 tracks that seem to originate from a photon conversion; in case those two tracks
 1721 are compatible with the direction of a bremsstrahlung photon, they are also
 1722 linked to the original electron track. Photon conversion tracks are also masked
 1723 in the PF block.

- 1724 • The remaining elements in the PF block are used to identify hadrons. In the
 1725 region $|\eta| \leq 2.5$, neutral hadrons are identified with HCAL clusters not linked
 1726 to any track while photons from neutral pion decays are identified with ECAL
 1727 clusters without links to tracks. In the region $|\eta| > 2.5$ ECAL clusters linked to
 1728 HCAL clusters are identified with a charged or neutral hadron shower; ECAL
 1729 clusters with no links are identified with photons. HCAL clusters not used yet,
 1730 are linked to one or more unlinked tracks and to an unlinked ECAL in order to
 1731 reconstruct charged-hadrons or a combination of photons and neutral hadrons
 1732 according to certain conditions on the calibrated calorimetric energy.

- 1733 • Charged-particle tracks may be liked together when they converge to a *sec-
 1734 ondary vertex (SV)* displaced from the interaction point where the PV and PU
 1735 vertices are reconstructed; at least three tracks are needed in that case, of which
 1736 at most one has to be an incoming track with hits in tracker region between a

1737 PV and the SV.

1738

1739 The linker algorithm, as well as the whole PF algorithm, has been validated and
 1740 commissioned; results from that validation are presented in the Reference [92].

1741 **Jet reconstruction.**

1742 Quarks and gluons may be produced in the pp collisions, therefore, their hadronization
 1743 will be seen in the detector as a shower of hadrons and their decay products in the
 1744 form of a *jet*. The anti- k_t algorithm [100] is used to perform the jet reconstruction
 1745 by clustering those PF particles within a cone (see Figure 4.3); previously, isolated
 1746 electrons, isolated muons, and charged particles associated with other interaction
 1747 vertices are excluded from the clustering.

1748 The anti- k_t algorithm proceeds in a sequential recombination of PF particles; the
 1749 distance between particles i and j (d_{ij}) and the distance between particles and the
 1750 beam are defined as

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

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

1751 where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti}, y_i and ϕ_i are the transverse momentum, ra-
 1752 pidity and azimuth of particle i respectively and R is the called jet radius. For all
 1753 the remaining PF particles, after removing the isolated ones, d_{ij} and d_{iB} are calcu-
 1754 lated⁵ and the smallest is identified; if it is a d_{ij} , particles i and j are replaced with

⁵ Notice that this is a combinatorial calculation.

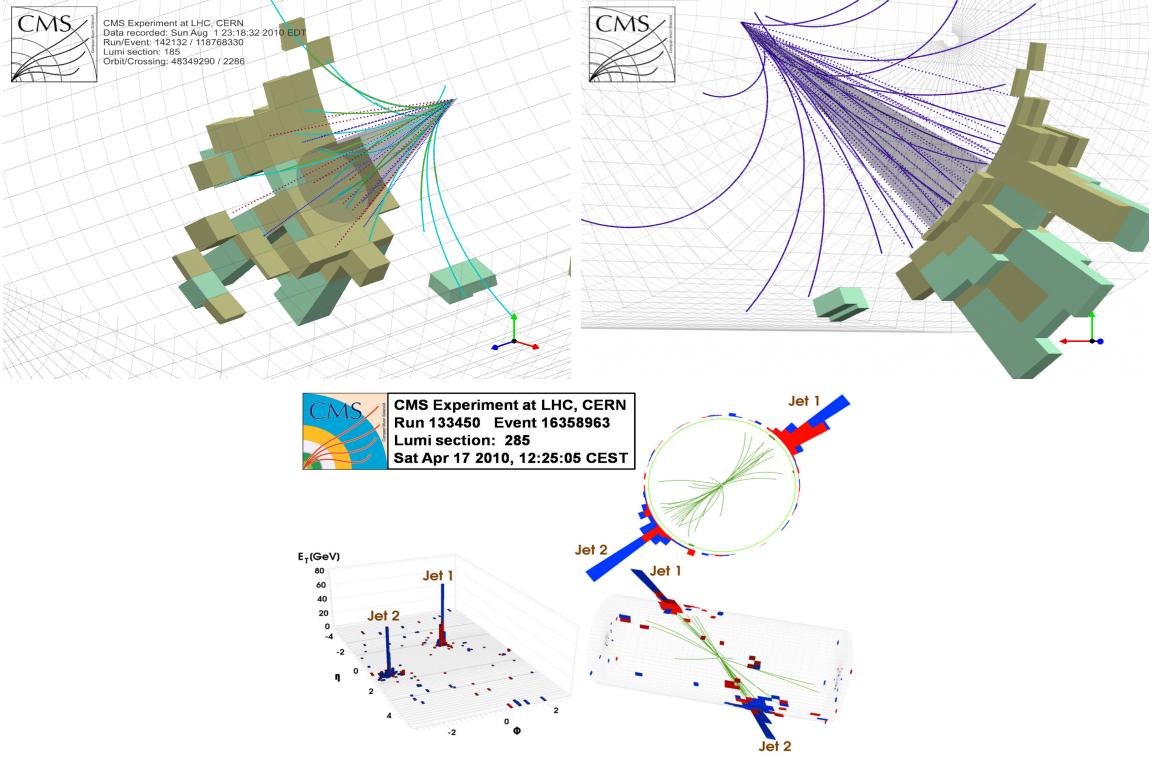


Figure 4.3: Jet reconstruction performed by the anti- k_t algorithm. Top: Two different views of a CMS recorded event are presented. Continuous lines correspond to tracks left by charged particles in the tracker while dotted lines are the imaginary paths followed by neutral particles. The green cubes 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 [101]. Bottom: Reconstruction of a recorded event with two jets [102].

1755 a new object whose momentum is the vectorial sum of the combined particles. If the
 1756 smallest distance is a d_{iB} the clustering process ends, the object i (which at this stage
 1757 should be a combination of several PF particles) is declared as a *Particle-flow-jet* (PF
 1758 jet) and all the associated PF particles are removed from the detector. The clustering
 1759 process is repeated until no PF particles remain.

1760 Even though jets can be reconstructed efficiently, there are some effects that are not in-
 1761 cluded in the reconstruction and that lead to discrepancies between the reconstructed
 1762 results and the predicted results; in order to overcome these discrepancies, a factor-
 1763 ized model has been designed in the form of jet energy corrections (JEC) [103, 104]

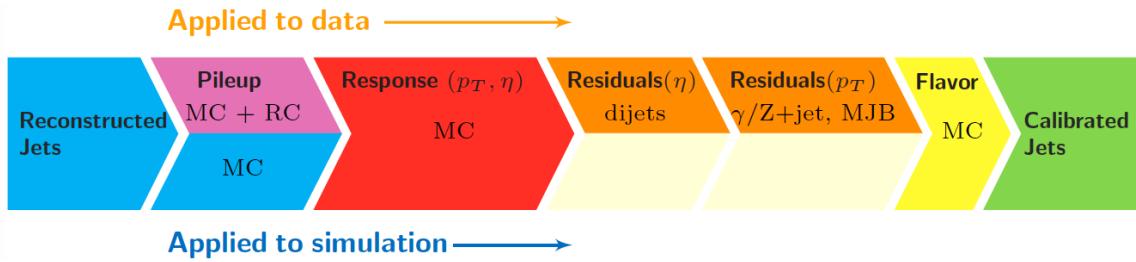


Figure 4.4: Jet energy correction diagram. Correction levels are applied sequentially in the indicated fixed order [104].

1764 applied sequentially as shown in the diagram of Figure 4.4.

¹⁷⁶⁵ At each level, the jet four-momentum is multiplied by a scaling factor based on jet

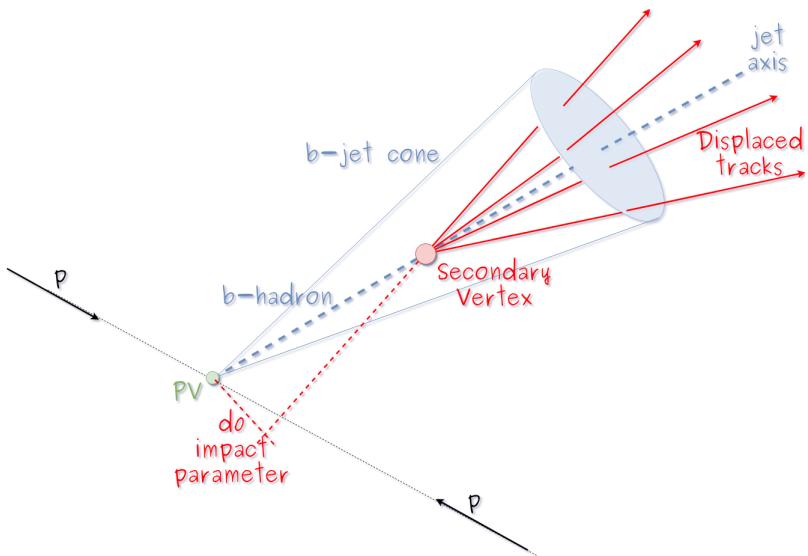
1766 properties, i.e., η , flavor, etc.

- 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 ρ , jet area A , jet η 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 η .
 - Residuals correct remaining small differences within jet response in data and MC. The Residuals η -dependent correction compares jets of similar p_T in the barrel reference region. The Residuals p_T -dependent correct the jet absolute scale (JES vs p_T).
 - Jet-flavor corrections are derived in the same way as MC-truth corrections but using QCD pure flavor samples.

1781 ***b*-tagging of jets.**

1782 A particular feature of the hadrons containing bottom quarks (*b*-hadrons) is that
 1783 their lifetime is long enough to travel some distance before decaying, but it is not as
 1784 long as those of light quark hadrons; therefore, when looking at the hadrons produced
 1785 in pp collisions, *b*-hadrons decay typically inside the tracker rather than reaching the
 1786 calorimeters as some light-hadrons do. As a result, a *b*-hadron decay gives rise to a
 1787 displaced vertex (secondary vertex) with respect to the primary vertex as shown in
 1788 Figure 4.5; the SV displacement is in the order of a few millimeters. A jet resulting
 1789 from the decay of a *b*-hadron is called *b* jet; other jets are called light jets.

1790

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

1791 Several methods to identify *b*-jets (*b*-tagging) have been developed; the method used
 1792 in this thesis is known as *Combined Secondary Vertex* algorithm in its second version
 1793 (CSVv2) [105]. By using information of the impact parameter, the reconstructed
 1794 secondary vertices, and the jet kinematics as input in a multivariate analysis that
 1795 combines the discrimination power of each variable in one global discriminator vari-

able, three working points (references): loose, medium and tight, are defined which quantify the probabilities of mistag jets from light quarks as jets from b quarks; 10, 1 and 0.1 % respectively. Although the mistagging probability decreases with the working point strength, the efficiency to correctly tag b -jets also decreases as 83, 69 and 49 % for the respective working point; therefore, a balance needs to be achieved according to the specific requirements of the analysis.

4.4.1.1 Missing transverse energy.

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

1810

1811 The PF algorithm assigns the negative sum of the momenta of all reconstructed PF
1812 particles to the *particle-flow MET* according to

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

1813 JEC are propagated to the calculation of the \vec{E}_T as described in the Reference [106].

1814

1815 **4.4.2 Event reconstruction examples**

1816 Figures 4.6-4.8 show the results of the reconstruction performed on 3 recorded events.

1817 Descriptions are taken directly from the source.

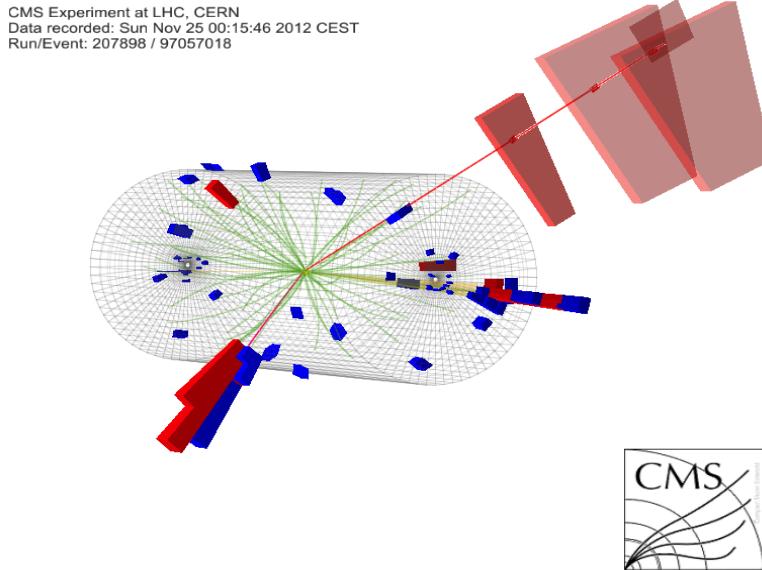


Figure 4.6: HIG-13-004 Event 1 reconstruction results; “HIG-13-004 Event 1: Event recorded with the CMS detector in 2012 at a proton-proton center-of-mass energy of 8 TeV. The event shows characteristics expected from the decay of the SM Higgs boson to a pair of τ leptons. Such an event is characterized by the production of two forward-going jets, seen here in opposite endcaps. One of the τ decays to a muon (red lines on the right) and neutrinos, while the other τ decays into a charged hadron and a neutrino.” [107].

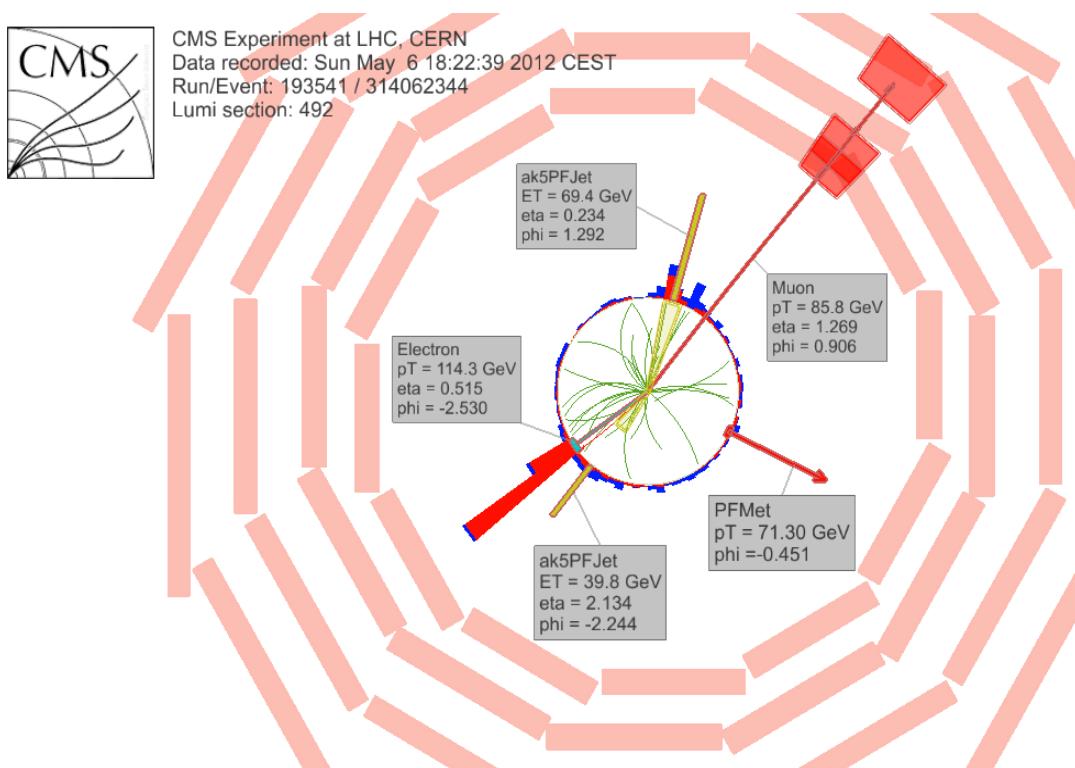


Figure 4.7: $e\mu$ event reconstruction results; “An $e\mu$ event candidate selected in 8 TeV data, as seen from the direction of the proton beams. The kinematics of the main objects used in the event selection are highlighted: two isolated leptons and two particle-flow jets. The reconstructed missing transverse energy is also displayed for reference” [108].

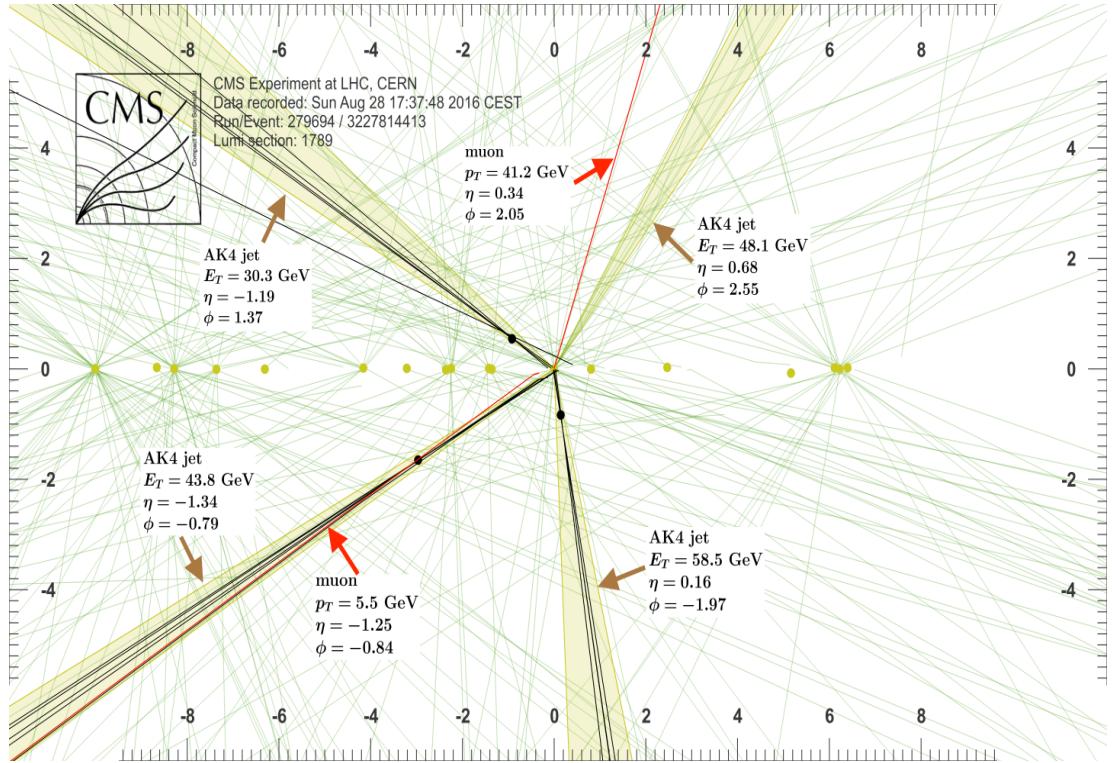


Figure 4.8: Recorded event reconstruction results;“Recorded event (ρ - z projection) with three jets with $p_T > 30$ GeV with one displaced muon track in 2016 data collected at 13 TeV. Each of the three jets has a displaced reconstructed vertex. The jet with $p_T(j) = 43.8$ GeV, $\eta(j) = -1.34$, $\phi(j) = -0.79$ 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” [109].

1818 **Chapter 5**

1819 **Statistical methods**

1820 In the course of analyzing the data sets provided by the CMS experiment and used in
1821 this thesis, several statistical tools have been employed; in this chapter, a description
1822 of these tools will be presented, starting with the general statement of the multivariate
1823 analysis method, followed by the particularities of the Boosted Decision Trees (BDT)
1824 method and its application to the classification problem. Statistical inference methods
1825 used will also be presented. This chapter is based mainly on the references [110–112].

1826 **5.1 Multivariate analysis**

1827 Multivariate data analysis (MVA) makes reference to statistical techniques that an-
1828 alyze data containing information of more than one variable, commonly taking into
1829 account the effects of all variables on the response of the particular variable under
1830 investigation, i.e., considering all the correlations between variables. MVA is em-
1831 ployed in a variety of fields like consumer and market research, quality control and
1832 process optimization. From a MVA it is possible to identify the dominant patterns
1833 in the data, like groups, outliers and trends, and determine to which group a set of

1834 values belong; in the particle physics context, MVA methods are used to perform the
 1835 selection of certain type of events, from a large data set, using a potentially large
 1836 number of measurable properties for each event.

1837 Processes with small cross section, as the tHq process, normally are hidden behind
 1838 more common processes; therefore, the data set results in a subset of events with
 1839 characteristic features of interest (signal) mixed in randomly with a much larger
 1840 number of SM events that can mimic these features of interest (background) which
 1841 implies that it is not possible to say with certainty that a given event is signal or
 1842 background. In that sense, the problem can be formulated as one where a set of
 1843 events have to be classified according to some features; these features correspond to
 1844 the measurements of several parameters like energy or momentum, organized in a
 1845 set of *input variables*. The measurements for each event can be written in a vector
 1846 $\mathbf{x} = (x_1, \dots, x_n)$ for which

- 1847 • Signal hypotheses $\rightarrow f(\mathbf{x}|s)$ is the probability density (*likelihood function*) that
 1848 \mathbf{x} is the set of measured values given that the event is a signal event.
- 1849 • Background hypotheses $\rightarrow f(\mathbf{x}|b)$ is the probability density (*likelihood function*)
 1850 that \mathbf{x} is the set of measured values given that the event is a background event.

1851 Figure 5.1 shows three ways to perform a classification of events for which mea-
 1852 surements of two properties, two input variables, have been performed; blue circles
 1853 represent signal events while red triangles represent background events. The classi-
 1854 fication in (a) is *cut-based* requiring $x_1 < c_1$ and $x_2 < c_2$; usually the cut values are
 1855 chosen according to some knowledge about the event process. In (b), the classification
 1856 is performed by stating a cut involving a linear function of the input variables and
 1857 so the boundary, while in (c) the relationship between the input variables is not
 1858 linear thus the boundary is not linear either.

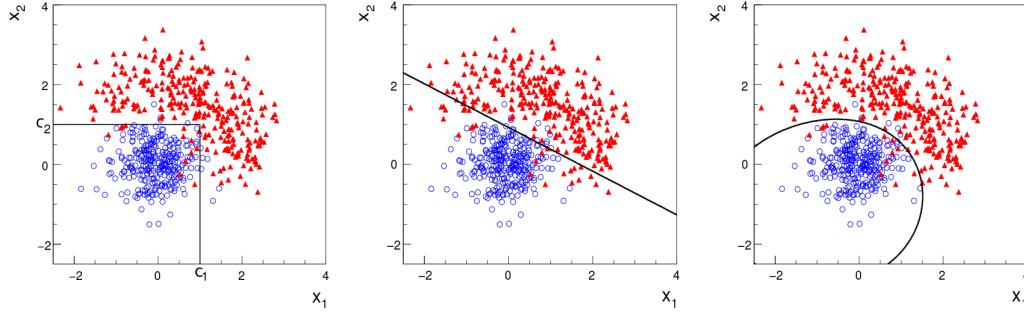


Figure 5.1: Scatter plots-MVA event classification. Distribution of two input variables x_1 and x_2 measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on (a) cuts, (b) linear boundary, and (c) nonlinear boundary [110]

1859 The boundary can be parametrized in terms of the input variables such that the
 1860 cut is set on the parametrization instead of on the variables, i.e., $y(\mathbf{x}) = y_{cut}$ with
 1861 y_{cut} a constant; thus, the acceptance or rejection of an event is based on what side
 1862 of the boundary is the event located. If $y(\mathbf{x})$ has functional form, it can be used to
 1863 determine the probability distribution functions $p(y|s)$ and $p(y|b)$ and then perform
 1864 a scalar test statistic with a single cut on the scalar variable y .

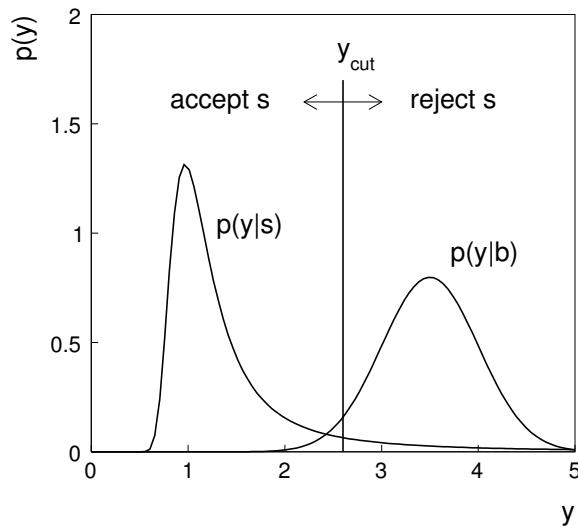


Figure 5.2: Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses. [110]

1865 Figure 5.2 illustrates what would be the probability distribution functions under
 1866 the signal and background hypotheses for a scalar test statistic with a cut on the
 1867 classifier y . Notice that the tails of the distributions indicate that some signal events
 1868 fall on the rejection region and some background events fall on the acceptance region;
 1869 therefore, it is convenient to define the *efficiency* with which events of a given type
 1870 are accepted, thus, the signal and background efficiencies are given by

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

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

1871 where A is the acceptance region. Under these conditions, the background hypothesis
 1872 corresponds to the *null hypothesis* (H_0), the signal hypothesis corresponds to the
 1873 *alternative hypothesis* (H_1), the background efficiency is the significance level of the
 1874 test, and signal efficiency is the power of the test; what is sought in an analysis is to
 1875 maximize the power of the test relative to the significance level.

1876 5.1.1 Decision trees

1877 For this thesis, the implementation of the MVA strategy, described above, is per-
 1878 formed through decision trees by using the TMVA software package [111] included in
 1879 the the ROOT analysis framework [113]. In a simple picture, a decision tree classifies
 1880 events according to their input variables values by setting a cut on each input variable
 1881 and checking which events are on which side of the cut, just as proposed in the MVA
 1882 strategy, but in addition, as a machine learning algorithm, decision trees offer the
 1883 possibility to be trained and then perform the classification efficiently.

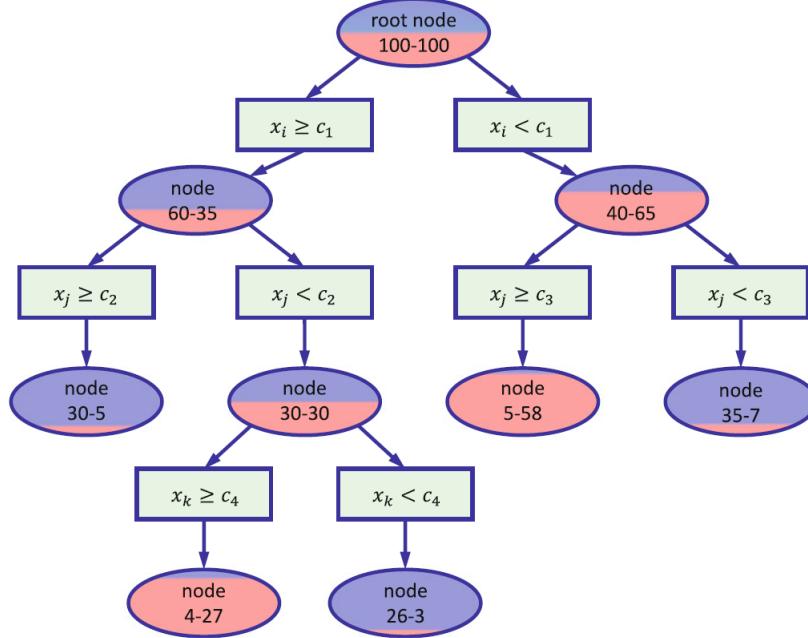


Figure 5.3: Example of a decision tree. Each node is fed with a MC sample mixing signal and background events (left-right numbers); nodes colors represent the relative amount of signal/background events [112].

1884 The training or growing of a decision tree is the process that defines the rules for
 1885 classifying events; this process is represented in figure5.3 and consist of several steps

1886 • take MC samples of signal and background events and split them into two parts
 1887 each; first parts form the training sample which will be used in the decision tree
 1888 training, while the second parts form the test sample which will be used for
 1889 testing the final classifier obtained from the training. Each event has associated
 1890 a set of input variables $\mathbf{x} = (x_1, \dots, x_n)$ which serve to distinguish between signal
 1891 and background events. The training sample is taken in at the root *node*.

1892 • pick one variable, say x_i
 1893 • pick one value of x_i , each event has its own value of x_i , and split the training
 1894 sample into two subsamples B_1 and B_2 ; B_1 contains events for which $x_i < c_1$

1895 while B_2 contains the rest of the training events;

1896 • scan all possible values of x_i and find the splitting value that provides the *best*
 1897 classification¹, i.e., B_1 is mostly made of signal events while B_2 is mostly made
 1898 of background events.

1899 • It is possible that variables other than the picked one produce a better classi-
 1900 fication, hence, all the variables have to be evaluated. Pick the next variable,
 1901 say x_j , and repeat the scan over its possible values.

1902 • At the end, all the variables and their values will have been scanned, the *best*
 1903 variable and splitting value will have been identified, say x_1, c_1 , and there will
 1904 be two nodes fed with the subsamples B_1 and B_2 .

1905 Nodes are further split by repeating the decision process until: a given number of
 1906 final nodes is obtained, nodes are largely dominated by either signal or background
 1907 events, or nodes has too few events to continue. Final nodes are called *leaves* and they
 1908 are classified as signal or background leaves according to the class of the majority of
 1909 events in them. Each *branch* in the tree corresponds to a sequence of cuts.

1910 The quality of the classification at each node is evaluated through a separation
 1911 criteria; there are several of them but the *Gini Index* (G) is the one used in the
 1912 decision trees trained for the analysis in this thesis. G is written in terms of the
 1913 purity (P), i.e. the fraction of signal events, of the samples after the separation is
 1914 made; it is given by

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

¹ Quality of the classification will be treated in the next paragraph.

1915 notice that $P=0.5$ at the root node while $G=0$ for pure leaves. For a node A split
 1916 into two nodes B_1 and B_2 the G gain is

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

1917 the *best* classification corresponds to that for which the gain of G is maximized; hence,
 1918 the scanning over all event's variables and their values is of capital importance.

1919 In order to provide a numerical output for the classification, events in a sig-
 1920 nal(background) leaf are assigned an score of 1(-1) each, defining in this way the
 1921 decision tree *classifier or weak learner* as

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

1922 Figure 5.4 shows an example of the classification of a sample of events, containing
 1923 two variables, performed by a decision tree.

1924 5.1.2 Boosted decision trees (BDT).

1925 Event misclassification occurs when a training event ends up in the wrong leaf, i.e., a
 1926 signal event ends up in a background leaf or a background event ends up in a signal
 1927 leaf. A way to correct it is to assign a weight to the misclassified events and train
 1928 a second tree using the reweighted events; the event reweighting is performed by a
 1929 boosting algorithm, events with increased weight are known as *boosted* events, in such
 1930 a way that when used in the training of a new decision tree they get correctly classified.
 1931 The process is repeated iteratively adding a new tree to a forest and creating a set
 1932 of classifiers which are combined to create the next classifier; the final classifier offers

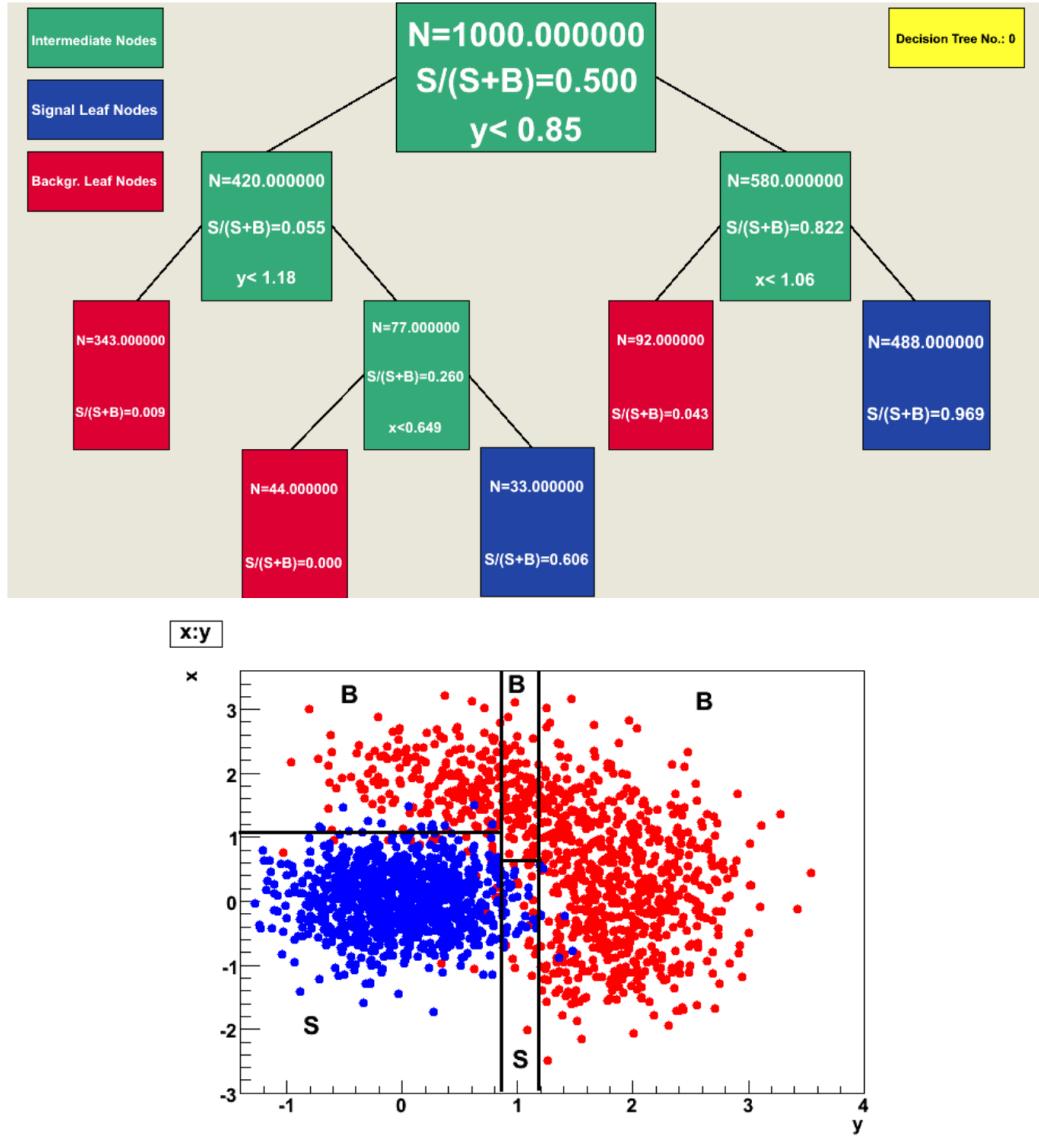


Figure 5.4: Example of a decision tree output. Each leaf, blue for signal events and red for background events, represent a region in the variables phase space [114].

1933 more stability² and has a smaller misclassification rate than any individual ones. The
 1934 resulting tree collection is known as a *boosted decision tree (BDT)*.
 1935 Thus, purity of the sample is generalized to

² 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.

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

1936 where w_s and w_b are the weights of the events; the Gini index is also generalized

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

1937 with n the number of events in the node. The final score of an event, after pass-
 1938 ing through the forest, is calculated as the renormalized sum of all the individual
 1939 (possibly weighted) scores; thus, high(low) score implies that the event is most likely
 1940 signal(background).

1941 The boosting procedure, implemented in the *Gradient boosting* algorithm used
 1942 in this thesis, produce a classifier $F(\mathbf{x})$ which is the weighted sum of the individual
 1943 classifiers obtained after each iteration,i.e.,

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

1944 where M is the number of trees in the forest. The *loss function* $L(F, y)$ represent the
 1945 deviation between the classifier $F(\mathbf{x})$ response and the true value y obtained from the
 1946 training sample (1 for signal events and -1 for background event), according to

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

1947 thus, the reweighting is employed to ensure the minimization of the loss function;
 1948 a more detailed description of the minimization procedure can be found in reference
 1949 [115]. The final classifier output is later used as a final discrimination variable, labeled
 1950 as *BDT output/response*.

1951 **5.1.3 Overtraining.**

1952 Decision trees offer the possibility to have as many nodes as wished in order to
 1953 reduce the misclassification to zero (in theory); however, when a classifier is too much
 1954 adjusted to a particular training sample, the classifier response to a slightly different
 1955 sample may leads to a completely different classification results; this effect is known
 1956 as *overtraining*.

1957 An alternative to reduce the overtraining in BDTs consist in pruning the tree by
 1958 removing statistically insignificant nodes after the tree growing is completed but this
 1959 option is not available for BDTs with gradient boosting in the TMVA-toolkit, there-
 1960 fore, the overtraining has to be reduced by tuning the algorithm, number of nodes,
 1961 minimum number of events in the leaves, etc. The overtraining can be evaluated
 1962 by comparing the responses of the classifier when running over the training and test
 1963 samples.

1964 **5.1.4 Variable ranking.**

1965 BDTs have the couple of particular advantages related to the input variables; on one
 1966 side, they are relatively insensitive to the number of input variables used in the vector
 1967 \mathbf{x} . The ranking of the BDT input variables is determined by counting the number of
 1968 times a variable is used to split decision tree nodes; in addition, the separation gain-
 1969 squared achieved in the splitting and the number of events in the node are accounted
 1970 by applying a weighting to that number. Thus, those variables with small or no power
 1971 to separate signal and background events are rarely chosen to split the nodes,i.e., are
 1972 effectively ignored.

1973 On the other side, variables correlations play an important role for some MVA
 1974 methods like the Fisher discriminant algorithm in which the first step consist of

1975 performing a linear transformation to a phase space where the correlations between
 1976 variables are removed; in case of BDT algorithm, correlations do not affect the per-
 1977 formance.

1978 **5.1.5 BDT output example.**

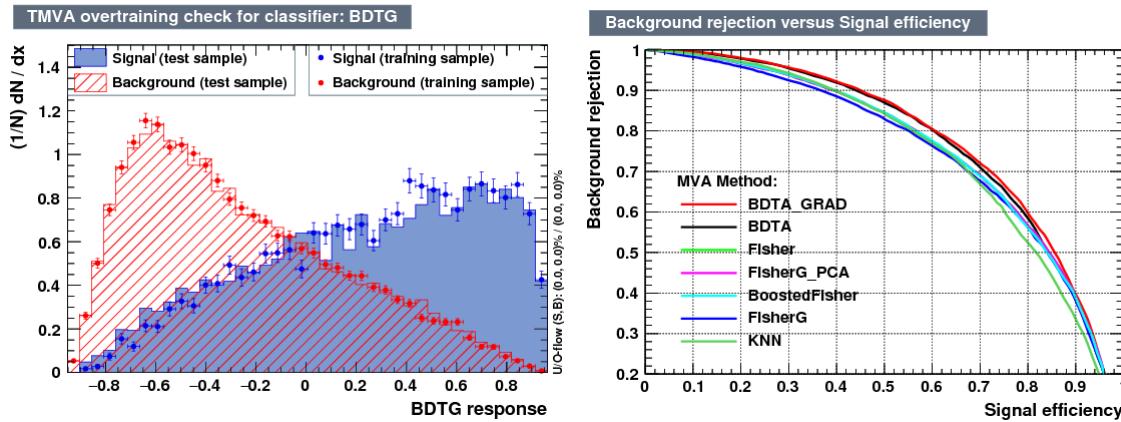


Figure 5.5: Left: Output distributions for the gradient boosted decision tree (BDTG) classifier using a sample of signal($pp \rightarrow tHq$) and background($pp \rightarrow t\bar{t}$) events. Right: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers running over the same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

1979 Left side of figure 5.5 shows the BDT output distributions for signal($pp \rightarrow tHq$)
 1980 and background($pp \rightarrow t\bar{t}$) events; this plot is the equivalent to the one showed in
 1981 figure 5.2. A forest with 800 trees, maximum depth per tree = 3, and gradient
 1982 boosting have been used as training parameters. The BDTG classifier offers a good
 1983 separation power; while there is a small overtraining in the signal distribution, the
 1984 background distribution seems to be well predicted which might indicate that the
 1985 sample is composed of more background than signal events.

1986 Right side of figure 5.5 shows the background rejection vs signal efficiency curves
 1987 for several combinations of MVA classifiers-boosting algorithms; these curves are
 1988 known as ROC curves and give an indication of the performance of the classifier. The

1989 best performance is achieved with the BDTG classifier (BDTA_GRAD).

1990 5.2 Statistical inference.

1991 Once events are classified, the next step consists in finding the parameters that define
 1992 the likelihood functions $f(\mathbf{x}|s), f(\mathbf{x}|b)$ for signal and background events respectively.
 1993 In general, likelihood functions depend not only on the measurements but also on
 1994 parameters (θ_m) that define their shapes; the process of estimating these *unknown*
 1995 *parameters* and their uncertainties from the experimental data is called *inference*.
 1996 The likelihood function for N the events the in a sample is the combination of all the
 1997 likelihoods functions

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

1998 Thus, the estimation of the unknown parameters from experimental data samples
 1999 is written in terms of a central value using the notation

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

2000 where the interval $[\hat{\theta} + \delta\theta, \hat{\theta} - \delta\theta]$ is called *confidence interval*; it is usually inter-
 2001 preted, in the limit of infinite number of experiments, as the interval where the true
 2002 value of the unknown parameter θ is contained with a probability of 0.6827 (if no
 2003 other convention is stated).

2004 5.2.1 Nuisance parameters.

2005 The unknown parameter vector $\boldsymbol{\theta}$ is made of two types of parameters: on one side,
 2006 those parameters that provide information about the physical observables of interest

2007 for the experiment (*parameters of interest*); on the other side, the *nuisance parameters*
 2008 that are not of direct interest for the experiment but that needs to be included in
 2009 the analysis in order to achieve a satisfactory description of the data. They represent
 2010 effects of the detector response like the finite resolutions of the detection systems,
 2011 miscalibrations, and in general any source of uncertainty introduced in the analysis.

2012 In some cases the nuisance parameters are estimated using dedicated data samples,
 2013 for instance data from test beams for calibration purposes, when MC samples are
 2014 not suitable. The nuisance parameter uncertainties produce *systematic uncertainties*
 2015 while the uncertainties associated to fluctuations in data and related to the estimation
 2016 of the parameters of interest produce *statistical uncertainties*.

2017 5.2.2 Maximum likelihood estimation method

2018 The function that produce the estimate of a parameter is called *estimator*, there-
 2019 fore, estimators are usually constructed using mathematical procedures encoded in
 2020 algorithms. The estimation method used in this thesis is the *Maximum Likelihood*
 2021 *Estimation* method (MLE); it is based on the combined likelihood function defined
 2022 by eqn. 5.9 and the procedure seeks for the parameter set that corresponds to the
 2023 maximum value of the combined likelihood function, i.e., the *maximum likelihood*
 2024 *estimator* of the unknown parameter vector $\boldsymbol{\theta}$ is the function that produce the vec-
 2025 tor $\hat{\boldsymbol{\theta}}$ for which the likelihood function $L(\boldsymbol{\theta})$ evaluated at the measured sample \mathbf{x} is
 2026 maximum.

2027 Usually, the logarithm of the likelihood function is used in the numerical algo-
 2028 rithms implementations in order to avoid underflow the numerical precision of the
 2029 computers due to the product of low likelihoods. In addition, it is usual minimize the
 2030 negative logarithm of the likelihood function instead of maximizing the logarithm of

2031 it because in this way the procedure consist of differentiate a sum of therms and set
 2032 the sum to zero; therefore

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

2033 The minimization process is performed by the software MINUIT [116] imple-
 2034 mented in the ROOT analysis framework. In case of large data samples the compu-
 2035 tational resources needed to calculate the likelihood function are too big; therefore,
 2036 the parameter estimation is performed using binned distributions of the variables of
 2037 interest for which the *binned likelihood function* is given by

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

2038 with s_i and b_i the expected number of signal and background yields for bin i respec-
 2039 tively, n_i is the observed number of events in the bin i and $\mu = \sigma/\sigma_{SM}$ is the signal
 2040 strength. Notice that the number of entries per bin follows a Poisson distribution.
 2041 The analysis presented in this thesis is based on the binned distribution of the ratio
 2042 signal/background obtained from the BDT outputs.

2043 5.2.3 Hypothesis test

2044 The test statistic mentioned in section 5.1 involving
 2045 ; it is achieved, according to the Neyman-Pearson lemma [117],
 2046 by defining the acceptance region such that, for \mathbf{x} inside the region, the likelihood
 2047 ratio, i.e., the ratio of probability distribution functions for signal and background,

₂₀₄₈ **5.3 exclusion limits**

₂₀₄₉ **5.4 asymptotic limits**

2050 **Chapter 6**

2051 **Search for production of a Higgs**

2052 **boson and a single top quark in**

2053 **multilepton final states in pp**

2054 **collisions at $\sqrt{s} = 13 \text{ TeV}$**

2055 **6.1 Introduction**

2056 The Higgs boson discovery, supported on experimental observations and theoretical
2057 predictions made about the SM, gives the clue of the way in that elementary particles
2058 acquire mass through the Higgs mechanism; therefore, knowing the Higgs mass, the
2059 Higgs-boson and Higgs-fermion couplings can be tested. In order to test the Higgs-top
2060 coupling, several measurements have been performed, as stated in the chapter 2, but
2061 they are limited to measure the square of the coupling; however, the production of a
2062 Higgs boson in association with a single top quark (tH) not only offers access to the
2063 sign of the coupling, but also, to the CP phase of the Higgs couplings.

2064 This chapter presents the search for the associated production of a Higgs boson
 2065 and a single top quark events, focusing on leptonic signatures provided by the Higgs
 2066 decay modes to WW , ZZ , and $\tau\tau$; the 13 TeV dataset produced in 2016, which
 2067 corresponds to an integrated luminosity of 35.9fb^{-1} , is used. Constraints on the sign
 2068 of the Higgs-top coupling (y_t) have been derived from the decay rate of Higgs boson
 2069 to photon pairs [40] and from the cross section for associated production of Higgs and
 2070 Z bosons via gluon fusion [118], with recent results disfavoring negative signs of the
 2071 coupling [51, 53, 119]. It expands previous analyses performed at 8 TeV [120, 121] and
 2072 searches for associated production of $t\bar{t}$ pair and a Higgs boson in the multilepton final
 2073 state channel [122]; it also complements searches in other decay channels targeting
 2074 $H \rightarrow b\bar{b}$ [123].

2075 As shown in section 2.4, the SM cross section of the associated production of a
 2076 Higgs boson and a single top quark (tHq) process is driven by a destructive interfer-
 2077 ence between two contributions (see Figure 2.13), where the Higgs couples to either
 2078 the W boson or the top quark; however, if the sign of the Higgs-top coupling is flipped
 2079 with respect to the SM prediction, a large enhancement of the cross section occurs,
 2080 making this analysis sensitive to such deviation. A second process, where the Higgs
 2081 boson and top quark are accompanied by a W boson (tHW) has similar behavior,
 2082 albeit with a weaker interference pattern and lower contribution to the tH cross sec-
 2083 tion, therefore, a combination of both processes would increase the sensitivity; in
 2084 this analysis both contributions are combined and referred as tH channel. A third
 2085 contribution comes from $t\bar{t}H$ process. The purpose of this analysis is to investigate
 2086 the exclusion of the presence of the $tH + t\bar{t}H$ processes under the assumption of the
 2087 anomalous Higgs-top coupling modifier ($\kappa_t = -1$). The analysis exploits signatures
 2088 with two leptons of the same sign (*2lss channel*) and three leptons (*3l channel*) in
 2089 the final state.

2090 The first sections present the characteristic tHq signature as well as the expected
 2091 backgrounds. The MC samples, data sets, and the physics object definitions are
 2092 then defined. Following, the background predictions, the signal extraction, and the
 2093 statistical treatment of the selected events as well as the systematic uncertainties are
 2094 described. The final section present the results for the exclusion limits as a function
 2095 of the ratio of κ_t and the dimensionless modifier of the Higgs-vector boson κ_V .

2096 6.2 tHq signature

2097 In order to select events of tHq process, its features are translated into a set of
 2098 selection rules; figure 6.1 shows the Feynman diagram and an schematic view of the
 2099 tHq process from the pp collision to the final state configuration. A single top quark
 2100 is produced accompanied by a light quark, denoted as q ; this light quark is produced
 2101 predominantly in the forward region of the detector. The Higgs boson which can
 2102 be either emitted by the exchanged W boson or directly by the singly produced top
 2103 quark.

2104 The top quark and Higgs boson decay after their production in the detector due to
 2105 their high masses/low lifetimes. The Higgs boson is required to decay into a W boson
 2106 pair¹. The top quark almost always decays into a bottom quark and a W boson, as
 2107 encoded in the CMK matrix. The W bosons are required to decay hadronically in
 2108 the 2lss channel case and leptonically in the 3l channel case, while τ leptons are not
 2109 reconstructed separately and only their leptonic decays into either electrons or muons
 2110 are considered in this analysis.

2111 In summary, the signal process is characterized by a the final state with

- 2112 • one light-flavored forward jet,

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

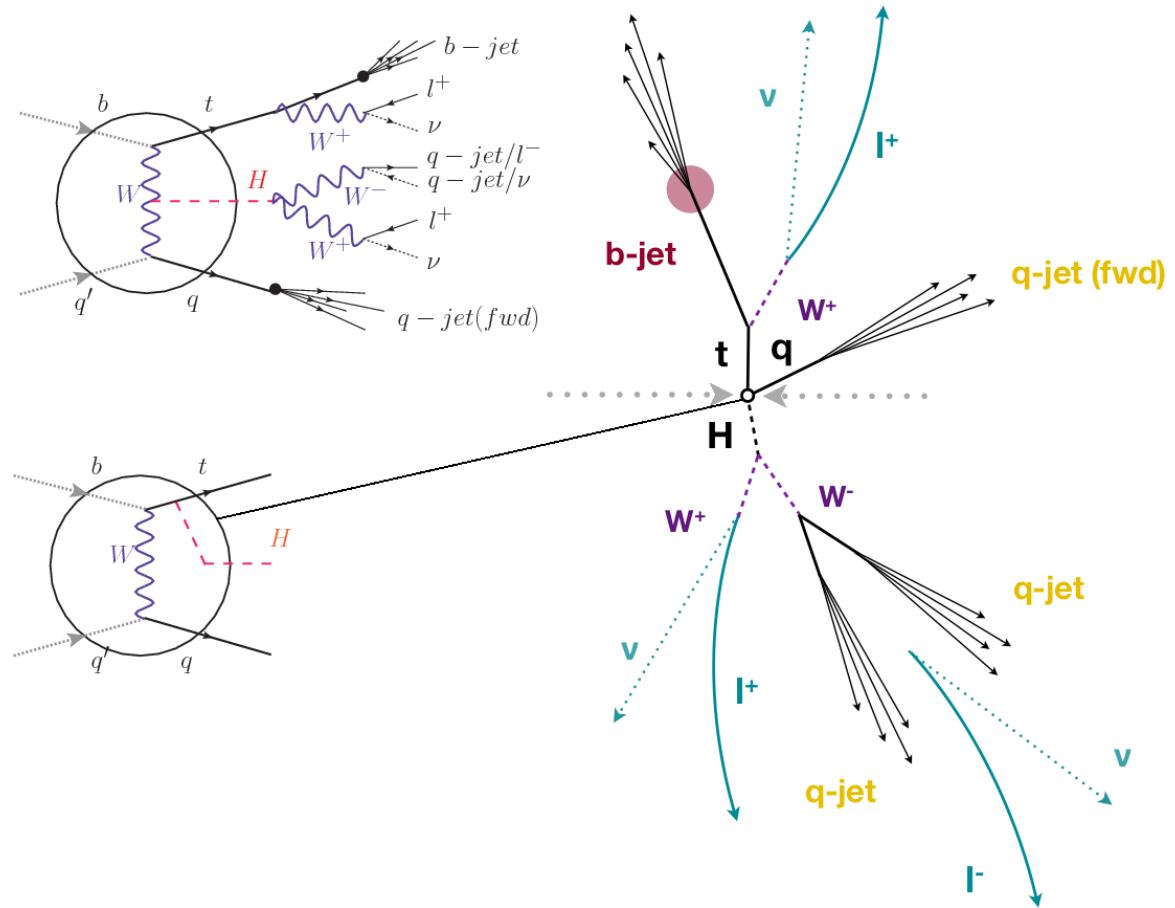


Figure 6.1: tHq event signature. Left: Feynman diagram including the whole evolution up to the final state for the case of the Higgs boson emitted by the W boson (top); Feynman diagram for the case where the Higgs boson is emitted by the top quark. Right: Schematic view as it would be seen in the detector; 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.

- 2113 • one central b-jet,
- 2114 • 2lss channel \rightarrow two leptons of the same sign, two neutrinos and two light (often
2115 soft) jets,
- 2116 • 3l channel \rightarrow three leptons, three neutrinos and no central light-flavored jets,
- 2117 The presence of neutrinos is inferred from the presence of MET. The analysis has

2118 been made public by CMS as a Physics Analysis Summary [124] combining the result
 2119 for the three lepton and two lepton same-sign channels. Currently, an effort to turn
 2120 the analysis into a paper is ongoing.

2121 **6.3 Background processes**

2122 The background processes are those that can mimic the signal signature or at least
 2123 can be reconstructed as that as a result of certain circumstances. The backgrounds
 2124 can be classified as

- 2125 • Irreducible backgrounds where genuine prompt leptons are produced in on-
 2126 shell W and Z boson decays; they can be reliably estimated directly from MC
 2127 simulated events, using higher-order cross sections or data control regions for
 2128 the overall normalization.
- 2129 • Reducible backgrounds where at least one of the leptons is *non-prompt*, i.e., pro-
 2130 duced within a hadronic jet, either a genuine lepton from heavy flavor decays.
 2131 misreconstructed jets, also known as *mis-ID leptons* or *fake leptons*, are consid-
 2132 ered non-prompt leptons as well. These non-prompt leptons leave tracks and
 2133 hits in the detection systems as would a prompt lepton, but correlating those
 2134 hits with nearby jets could be a way of removing them. Reducible backgrounds
 2135 are not well predicted by simulation, and are estimated using data-driven meth-
 2136 ods.

2137 The main sources of background events in the case of tHq process are $t\bar{t}$ process
 2138 and $t\bar{t} + X(X = W, Z, \gamma)$ processes, here represented together as $t\bar{t}V$ process. Figure
 2139 6.2 shows the signature for $t\bar{t}$ and $t\bar{t}W$ processes;

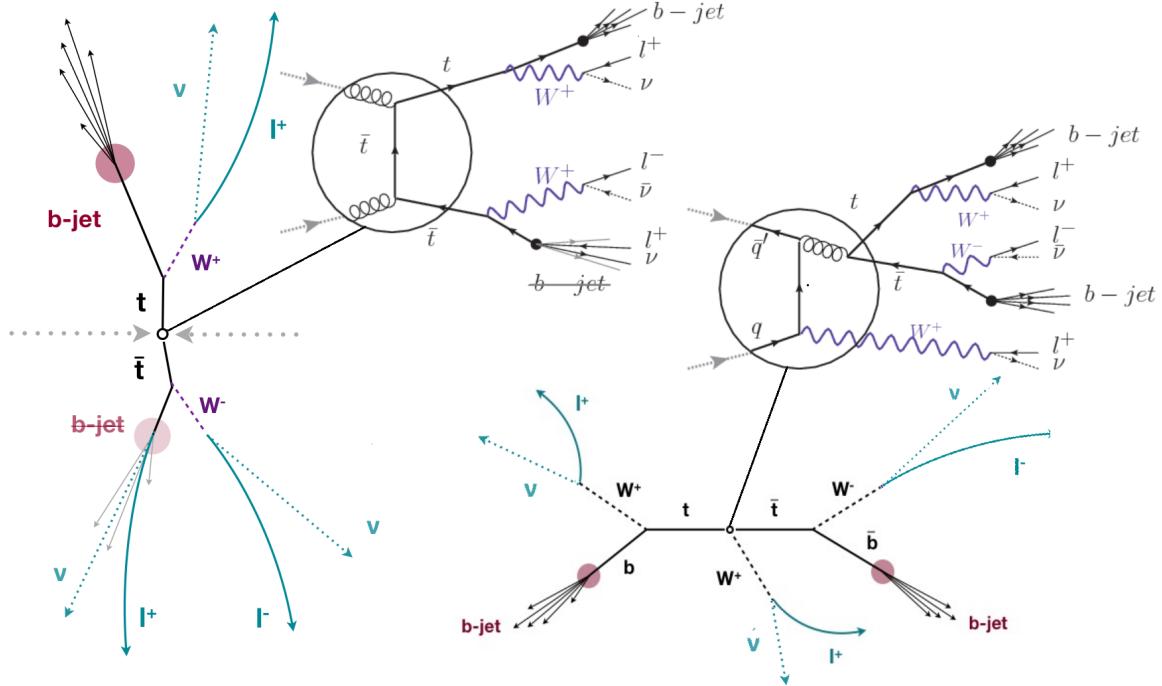


Figure 6.2: $t\bar{t}$ (left) and $t\bar{t}W$ (right) events signature as they would be seen in the detector; the Feynman diagrams including the whole evolution up to the final state are also showed. The $t\bar{t}$ process signature is very similar to that of the signal process with one fake lepton and non 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.

2140 The largest contribution to irreducible backgrounds involving prompt leptons
 2141 comes from $t\bar{t}W$, $t\bar{t}Z$, processes for which the number of ($b-$)jets (($b-$)jet multiplicity)
 2142 is higher than that of the signal events, while for other contributing background
 2143 events, WZ , ZZ , and rare SM processes like $W^\pm W^\pm qq$, $t\bar{t}t\bar{t}$, tZq , tZW , WWW ,
 2144 WWZ , WZZ , ZZZ , the ($b-$)jet multiplicity is lower compared to that of the signal
 2145 events. None of the irreducible backgrounds present activity in the forward region of
 2146 the detector.

2147 On the side of the reducible backgrounds, the largest contribution comes from the
 2148 $t\bar{t}$ events which have a very similar signature to the signal events but does no present
 2149 activity in the forward region of the detector either; A particular feature of the $t\bar{t}$
 2150 events is their charge-symmetry which is also a difference with the signal events.

2151 The charge misidentification plays an important role in the the 2lss channel since
 2152 leptons in processes like $t\bar{t}$ + jets or Z + jets can be charge misidentified, leading to
 2153 backgrounds increments. An identification variable have been designed in order to
 2154 reject this type of background events.

2155 6.4 Data and MC Samples

2156 Technical developments on the event generator side allow for an event-wise reweight-
 2157 ing that can change the event kinematics based on specific generation parameters.
 2158 This way not only the case of $C_t = \sqrt{2}$, but a whole range of κ_t and κ_V values can
 2159 be investigated.

2160 The data considered in this analysis were collected by the CMS experiment dur-
 2161 ing 2016 and correspond to a total integrated luminosity of $35.9 fb^{-1}$. Only periods
 2162 when the CMS magnet was on were considered when selecting the data samples, that
 2163 corresponds to the 23 Sep 2016 (Run B to G) and PromptReco (Run H) versions
 2164 of the datasets. The MC samples used in this analysis correspond to the RunI-
 2165 ISummer16MiniAODv2 campaign produced with CMSSW 80X. The two signal sam-
 2166 ples (for tHq and tHW) were produced with MG5_aMC@NLO (version 5.222), in
 2167 leading-order order mode, and are normalized to next-to-leading-order cross sections,
 2168 see Tab. 6.1. Each sample is generated with a set of event weights corresponding to
 2169 different values of κ_t and κ_V couplings as shown in Tab. 6.2.

2170 6.4.1 Full 2016 dataset and MC samples

2171 Different MC generators were used to generate the background processes. The dom-
 2172 inant sources ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$) were produced using AMC@NLO interfaced to
 2173 PYTHIA8, and are scaled to NLO cross sections. Other processes are simulated us-

Sample	σ [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. [133] for more details.

ing POWHEG interfaced to PYTHIA, or bare PYTHIA (see table 6.3 and [122] 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 ≥ 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 [127] for the current analysis.

		<i>tHq</i>		<i>tHW</i>		
κ_V	κ_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: κ_V and κ_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. [133].

Sample	σ [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	σ [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 ± 0.02
e μ	1.01 ± 0.01
$\mu\mu$	1.00 ± 0.01
3l	1.00 ± 0.03

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

2191 6.5 Object Identification

2192 In this section, the specific definitions of the physical objects in terms of the numerical
 2193 values assigned to the reconstruction parameters are presented; thus, the provided
 2194 details summarize and complement the descriptions presented in previous chapters.
 2195 The object reconstruction and selection strategy used in this thesis is inherited from
 2196 the analyses in references [122, 127], thus, the information provided in this section is
 2197 extracted from those documents unless other references are stated.

2198 6.5.1 Jets and b tagging.

2199 In this analysis, jets are reconstructed by clustering PF candidates using the anti- k_t
 2200 algorithm with parameter distance $\Delta R = 0.4$; those charged hadrons that are not
 2201 consistent with the selected primary vertex are discarded from the clustering. The
 2202 jet energy is then corrected for the varying response of the detector as a function
 2203 of transverse momentum p_T and pseudorapidity η . Jets are selected for use in the
 2204 analysis only if they have $p_T > 25$ GeV and are separated from any selected leptons
 2205 by $\Delta R > 0.4$.

2206 Jets coming from the primary vertex and jets coming from pile-up vertices are
 2207 distinguished using a MVA discriminator based on the differences in the jet shapes,
 2208 in the relative multiplicity of charged and neutral components, and in the different
 2209 fraction of transverse momentum which is carried by the hardest components. Jet

2210 tracks are also required to be compatible with the primary vertex.

2211 Jets originated from the hadronization of a b quark are selected using a MVA
 2212 likelihood discriminant which uses track-based lifetime information and reconstructed
 2213 secondary vertices (CSV algorithm). Only jets within the CMS tracker acceptance
 2214 ($\eta < 2.4$) are identified with this tool. Data samples are used to measure the efficiency
 2215 of the b tagging and the probability to misidentify jets from light quarks or gluons;
 2216 in both cases the measurements are parametrized as a function of the jet p_T and η
 2217 and later used to correct differences between the data and MC simulation in the b
 2218 tagging performance, by applying per-jet weights to the simulation, dependent on
 2219 the jet p_T , η , b tagging discriminator, and flavor (from simulation truth) [125]. The
 2220 per-event weight is taken as the product of the per-jet weights, including those of the
 2221 jets associated to the leptons. The weights are derived on $t\bar{t}$ and Z+jets events.

2222 Two working points are defined, based on the CSV algorithm output: *loose*' work-
 2223 ing point (CSV>0.46) with a b signal tagging efficiency of about 83% and a mistagging
 2224 rate of about 8%; and *medium* working point (CSV>0.80) with b -tagging efficiency of
 2225 about 69% and mistagging rate of order 1% [126]. Tagging of jets from charm quarks
 2226 have efficiencies of about 40% and 18% for loose and medium working points re-
 2227 spectively. Separate scale factors are applied to jets originating from bottom/charm
 2228 quarks and from light quarks in simulated events to match the tagging efficiencies
 2229 measured in the data.

2230 6.5.2 Missing Energy MET.

2231 As stated in section 4.4.1.1, the MET vector is calculated as the negative of the vector
 2232 sum of transverse momenta of all PF candidates in the event and its magnitude is
 2233 referred to as E_T^{miss} . Due to pile-up interactions, the performance in determining

2234 MET is degraded; in order to correct for that, the energy from the selected jets and
 2235 leptons that compose the event is assigned to the variable H_T^{miss} . It is calculated in
 2236 the same way as E_T^{miss} and although it has worse resolution than E_T^{miss} , it is more
 2237 robust in the sense that it does not rely on the soft part of the event. The event
 2238 selection uses a linear discriminator based on the two variables given by

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

2239 taking advantage of the fact that the correlation between E_T^{miss} and H_T^{miss} is less
 2240 for events with instrumental missing energy than for events with real missing energy.
 2241 The working point $E_T^{miss} LD > 0.2$ was chosen to ensure a good signal efficiency while
 2242 keeping a good background rejection.

2243 6.5.3 Lepton reconstruction and identification

2244 Two types of leptons are defined in this analysis: *signal leptons* are those coming from
 2245 W, Z and τ decays which usually are isolated from other particles; *background leptons*
 2246 are defined as leptons produced in b hadron decays, light-jets misidentification, and
 2247 photon conversions.

2248 The process of reconstruction and identification of electron and muon candidates
 2249 was described in chapter4, hence, the identification variables used in order to retain
 2250 the highest possible efficiency for signal leptons while maximizing the rejection of
 2251 background leptons are listed and described in the following sections ².

2252 The identification variables include not only observables related directly to the re-
 2253 constructed leptons themselves, but also to the clustered energy deposits and charged
 2254 particles in a cone around the lepton direction (jet-related variables); an initial loose

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

2255 preselection of leptons candidates is performed and then an MVA discriminator, re-
 2256 ferred to as *lepton MVA* discriminator, is used to distinguish signal leptons from
 2257 background leptons.

2258 **Muons.**

2259 The Physics Objects Groups (POG) at CMS, are in charge of studying and defining
 2260 the set of selection criteria applied on the course of reconstruction and identification
 2261 of particles. These selection criteria are implemented in the CMS framework in the
 2262 form of several object identification working points according to the strength of the
 2263 requirements.

2264 The muon candidates are reconstructed by combining information from the tracker
 2265 system and the muon detection system of CMS detector and the POG defined three
 2266 working points for muon identification *MuonID* [128];

- 2267 • *POG Loose Muon ID* is a particle identified as a muon by the PF event re-
 2268 construction and also reconstructed either as a global-muon or as an arbitrated
 2269 tracker-muon. This identification criteria is designed to be highly efficient for
 2270 prompt muons and for muons from heavy and light quark decays; it can be com-
 2271 plemented by applying impact parameter cuts in analyses with prompt muon
 2272 signals.
- 2273 • *POG Medium Muon ID* is a Loose muon with additional track-quality and
 2274 muon-quality (spatial matching between the individual measurements in the
 2275 tracker and the muon system) requirements. This identification criteria is de-
 2276 signed to be highly efficient in the separation of the muons coming from decay
 2277 in flight of heavy quarks and muons coming from B meson decays as well as
 2278 prompt muons. An additional category *MVA Prompt ID* is defined in this iden-

2279 tification criteria directed to discriminated muons from B mesons and prompt
 2280 muons (from W,Z and τ decays). The Medium ID provides the same fake rate as
 2281 the Tight Muon ID but a higher efficiency on prompt and B-decays muons. [129]

- 2282 • *POG Tight Muon ID* is a global muon with additional muon-quality require-
 2283 ments Tight Muon ID selects a subset of the PF muons.

2284 Only muons within the muon system acceptance $|\eta| < 2.4$ and minimum p_T of 5
 2285 GeV are considered.

2286 **Electrons.**

2287 Electrons are reconstructed using information from the tracker and from the electro-
 2288 magnetic calorimeter and identified by an MVA algorithm (*MVA eID* discriminant)
 2289 using the shape of the calorimetric shower variables like the shape in η and ϕ , the clus-
 2290 ter circularity, widths along η and ϕ ; track-cluster matching variables like E_{tot}/p_{in} ,
 2291 E_{Ele}/p_{out} , $\Delta\eta_{in}$, $\Delta\eta_{out}$, $\Delta\phi_{in}$, $1/E - 1/p$; and track quality variables like ξ^2 of the
 2292 GSF tracks, the number of hits used by the GSF filter [130].

2293 A loose selection based on η -dependent cuts on this discriminant is used to prese-
 2294 lect electron candidates, the full shape of the discriminant is used in the lepton MVA
 2295 selection to separate signal leptons from background leptons (described in section
 2296 6.5.3).

2297 In order to reject electrons from photon conversions, electron candidates with
 2298 missing hits in the pixel tracker layers or matched to a conversion secondary vertex
 2299 are discarded. Electrons are selected for the analysis if they have $p_T > 7$ GeV and
 2300 are located within the tracker system acceptance region ($|\eta| < 2.5$).

2301 **Lepton vertexing and pile-up rejection.**

2302 The impact parameter in the transverse plane d_0 , impact parameter along the z
 2303 axis d_z , and the impact parameter significance in the detector space SIP_{3D} , are
 2304 considered to perform the identification and rejection of pile-up, misreconstructed
 2305 tracks, and background leptons from b-hadron decays; pile-up and misreconstructed
 2306 track mitigation is achieved by imposing loose cuts on the impact parameter variables.
 2307 The full shape of the those variables is used in a lepton MVA classifier to achieve the
 2308 best separation between the signal and the background leptons.

2309 **Lepton isolation.**

2310 PF is able to recognize leptons from two different sources: on one side, leptons from
 2311 the decays of heavy particles, such as W and Z bosons, which are normally isolated
 2312 in space from the hadronic activity in the event; on the other side, leptons from the
 2313 decays of hadrons and jets misidentified as leptons, which are not isolated as the
 2314 former. For highly boosted systems, like the lepton and the b -jet generated in the
 2315 semileptonic decay of a boosted top, the decay products tend to be more closer and
 2316 sometimes they even overlap; thus, the PF standard definition of isolation in terms of
 2317 the separation between the lepton candidates and other PF objects in the η - ϕ plane,

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

2318 which considers all the neutral, charged hadrons and photons in a cone around the
 2319 leptons, is refocused to the local isolation of the leptons through the mini-isolation
 2320 I_{mini} [131] defined as the sum of particle flow candidates p_T within a cone around

2321 the lepton, corrected for the effects of pileup and divided by the lepton p_T

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

2322 where ρ is the pileup energy density, h^\pm, h^0, γ, l , represent the charged hadron, neutral
 2323 hadrons, photons, and the lepton, respectively. The radius R of the cone depends on
 2324 the p_T of the lepton according to

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

2325 The p_T dependence of the cone size allows for greater signal efficiency. Setting a
 2326 cut on I_{mini} below a given threshold ensures that the lepton is locally isolated, even
 2327 in boosted systems. The effect of pileup is mitigated using the so-called effective area
 2328 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.

2329 A loose cut on I_{mini} is applied to pre-select the muon and electron candidates;
 2330 however, the full shape is used in the lepton MVA discriminator when performing the
 2331 signal lepton selection.

2332 **Jet-related variables.**

2333 In order to reject misidentified leptons from b -jets, mostly coming from $t\bar{t}$ +jets,
 2334 Drell-Yan+jets, and W+jets events, the vertexing and isolation described in previous
 2335 sections are complemented with additional variables related to the closest recon-
 2336 structed jet to the lepton, i.e., the PF jets reconstructed³ around the leptons with
 2337 $\Delta R = \sqrt{(\eta^l - \eta^{jet})^2 + (\phi^l - \phi^{jet})^2} < 0.5$. The identification variables used in the lep-
 2338 ton MVA discriminator are the ratio p_T^l/p_T^{jet} , the CSV b-tagging discriminator value
 2339 of the jet, the number of charged tracks of the jet, and the relative p_T given by

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

2340 **LeptonMVA discriminator.**

2341 Electrons and muons passing the basic selection process described above are referred
 2342 to as *loose leptons*. Additional discrimination between signal leptons and background
 2343 leptons is crucial considering that the rate of $t\bar{t}$ production is much larger than the
 2344 signal, hence, an overwhelming background from $t\bar{t}$ production. To maximally ex-
 2345 ploit the available information in each event to that end, the dedicated lepton MVA
 2346 discriminator, based on a boosted decision tree (BDT) algorithm, has been built so
 2347 that all the identification variables can be used together.

2348 The lepton MVA discriminator training is performed using simulated signal Loose
 2349 leptons from the $t\bar{t}H$ MC sample and fake leptons from the $t\bar{t}$ +jets MC sample,
 2350 separately for muons and electrons. The input variables used include vertexing, iso-
 2351 lation and jet-related variables, the p_T and η of the lepton, the electron MVA eID
 2352 discriminator and the muon segment-compatibility variables. An additional require-
 2353 ment known as *tight-charge* requirement, is imposed by comparing two independent

³ charged hadrons from PU vertices are not removed prior to the jet clustering.

2354 measurement of the charge, one from the ECAL supercluster and the other from the
 2355 tracker; thus, the consistency in the measurements of the electron charge is ensured
 2356 so that events with a wrong electron charge assignment are rejected; this variable is
 2357 particularly used in the 2lss channel to suppress opposite-sign events for which the
 2358 charge of one of the leptons has been mismeasured. The tight-charge requirement for
 2359 muons is represented by the requirement of a consistently well measured track trans-
 2360 verse momentum given by $\Delta p_T/p_T < 0.2$. Leptons are selected for the final analysis
 2361 if they pass a given threshold of the BDT output, and are referred to as *tight leptons*
 2362 in the following.

2363 The validation of the lepton MVA algorithm and the lepton identification variables
 2364 is performed using data in various control regions; the details about that validation
 2365 are not discussed here but can be found in Reference [127].

2366 Selection definitions

2367 Electron and muon object identification is defined in three different sets of selections
 2368 criteria; the *Loose*, *Fakeable Object*, and *Tight* selection. These three levels of selection
 2369 are designed to serve for event level vetoes, the fake rate estimation application region,
 2370 and the final signal selection, respectively. The p_T of fakeable objects is defined as
 2371 $0.85 \times p_T(\text{jet})$, where the jet is the one associated to the lepton object. This mitigates
 2372 the dependence of the fake rate on the momentum of the fakeable object and thereby
 2373 improves the precision of the method.

2374 Tables 6.8 and 6.9 list the full criteria for the different selections of muons and
 2375 electrons.

2376 In addition to the previously defined requirements for jets, they are required to
 2377 be separated from any lepton candidates passing the fakeable object selections by
 2378 $\Delta R > 0.4$.

Cut	Loose	Fakeable object	Tight
$ \eta < 2.4$	✓	✓	✓
p_T	$> 5\text{GeV}$	$> 15\text{GeV}$	$> 15\text{GeV}$
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (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$ (cm)	✓	✓	✓
$ d_z < 0.1$ (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^{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 η 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 (value₁, value₂, value₃). For the two p_T^{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.

2379 6.5.4 Lepton selection efficiency

2380 Efficiencies of reconstruction and selecting loose leptons are measured both for muons
 2381 and electrons using a tag and probe method on both data and MC, using $Z \rightarrow \ell^+ \ell^-$
 2382 [132]. The scale factors are derived from the ratio of efficiencies $\varepsilon_i(p_T, \eta)$ measured
 2383 for a given lepton in data/MC, according to

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

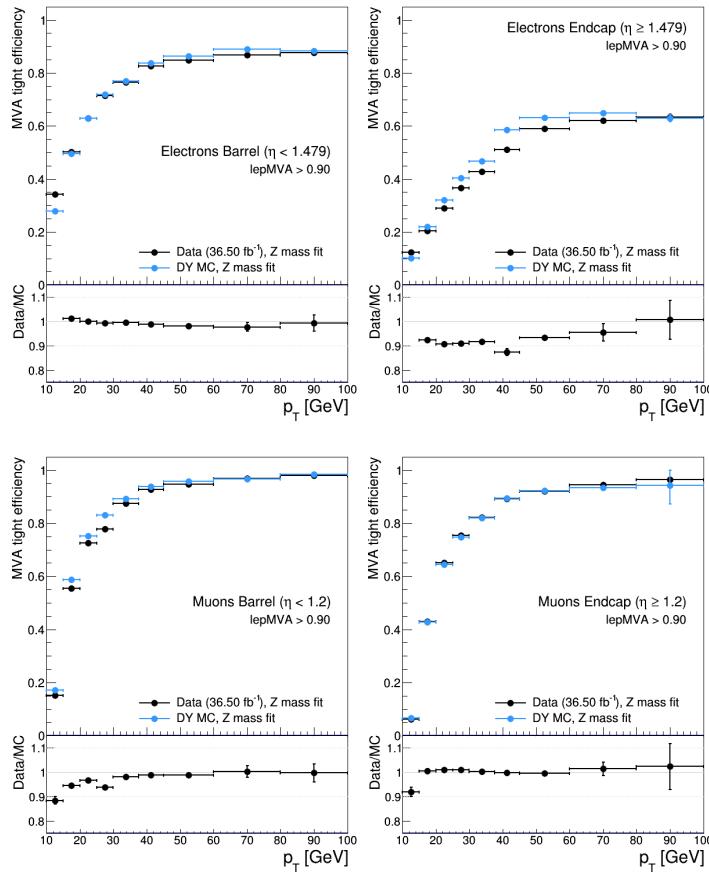


Figure 6.3: Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the 2lss definition, i.e., including the tight-charge requirement.

2384 The scale factor for each event is used to correct the weight of the event in the

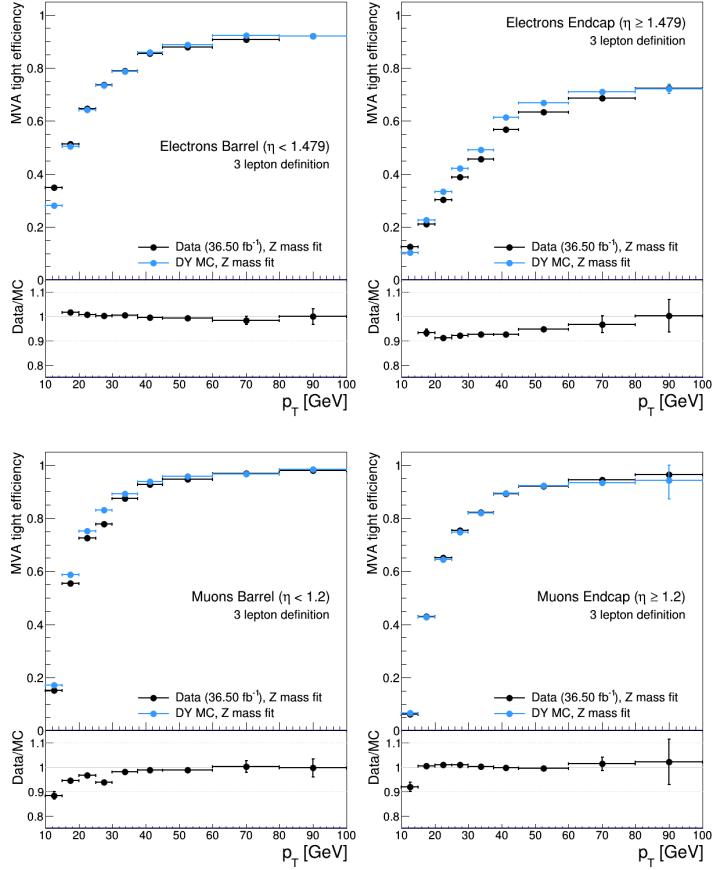


Figure 6.4: Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the 3l channel not including the tight-charge requirement.

2385 full sample; therefore, the full simulation correction is given by the product of all
 2386 the individual scale factors. The scale factors used in this thesis are inherited from
 2387 the reference [127] which in turns inherited them from leptonic SUSY analyses using
 2388 equivalent lepton selections.

2389 The efficiency of applying the tight selection as defined in Tables 6.8 and 6.9, on the
 2390 loose leptons are determined by using a tag and probe method on a sample of Drell-
 2391 Yan enriched events. Figures 6.3 and 6.4 show the efficiencies for the 2lss channel and
 2392 3l channel respectively. Efficiencies in the 2lss channel have been produced including
 2393 the tight-charge requirement, while for the 3l channel it is not included. Number

2394 of passed and failed probes are determined from a fit to the invariant mass of the
 2395 dilepton system.

2396 Simulation is corrected using these scale factors; note that they depends on η and
 2397 p_T .

2398 **6.6 Event selection**

2399 .

2400 .

2401 .

2402 .

2403 .

2404 .

2405 .

2406 .

2407 .

2408 . . .

2409 The analysis is designed to efficiently identify and select prompt leptons from
 2410 on-shell W and

2411 Z boson decays and to reject non-prompt leptons from b quark decays and spurious
 2412 lepton

2413 signatures from hadronic jets. Events are then selected in the various lepton
 2414 channels, and are

2415 required to contain hadronic jets, some of which must be consistent with b quark
 2416 hadronization. Finally, the signal yield is extracted by simultaneously fitting the
 2417 output of two dedicated

2418 multivariate discriminants (trained to separate the tHq signal from the two dom-
 2419 inant backgrounds) in all categories

2420 . Multivariate techniques are used to discriminate the signal from the dominant
 2421 backgrounds. The analysis yields a 95% confidence level (C.L.) upper limit on the
 2422 combined $tH + ttH$ production cross section times branching ratio of 0.64 pb, with
 2423 an expected limit of 0.32 pb, for a scenario with $kt = \sqrt{1.0}$ and $kV = 1.0$. Values
 2424 of kt outside the range of $\sqrt{1.25}$ to $\sqrt{1.60}$ are excluded at 95% C.L., assuming kV
 2425 $= 1.0$.

2426 Dont forget to mention previous constrains to ct check reference ?? and references
 2427 <https://link.springer.com/content/pdf/10.1007%2FJHEP01%282013%29088.pdf> (para-
 2428 graph after eq 2)

2429 We selects events with three leptons and a b tagged jet in the final state. The tHq
 2430 signal contribution is then determined in a fit of the observed data to two multivariate
 2431 classifier outputs, each trained to discriminate against one of the two dominant back-
 2432 grounds of events with non-prompt leptons from $t\bar{t}$ and of associated production of $t\bar{t}$
 2433 and vector bosons ($t\bar{t}W$, $t\bar{t}Z$). The fit result is then used to set an upper limit on the
 2434 combined $t\bar{t}H$, tHq and tHW production cross section, as a function of the relative
 2435 coupling strengths of Higgs and top quark and Higgs and W boson, respectively.

2436 6.7 Background predictions

2437 The modeling of reducible and irreducible backgrounds in this analysis uses the exact
 2438 methods, analysis code, and ROOT trees used for the $t\bar{t}H$ multilepton analysis. We
 2439 give a brief description of the methods and refer to the documentation of that analysis
 2440 in Refs. [122, 127] for any details.

2441 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

2468 events, limited statistics in the data control regions due to necessary prescaling of
 2469 lepton triggers, and the uncertainty in the subtraction of residual prompt leptons
 2470 from the control region.

2471 The fake background where the leptons pass the looser selection are weighted
 2472 according to how many of them fail the tight criteria. Events with a single failing
 2473 lepton are weighted with the factor $f/(1-f)$ for the estimate to the tight selection
 2474 region, where f is the fake rate. Events with two failing leptons are given the negative
 2475 weight $-f_i f_j / (1-f_i)(1-f_j)$, and for three leptons the weight is positive and equal
 2476 to the product of $f/(1-f)$ factor evaluated for each failing lepton.

2477 Figures 6.5 show the distributions of some relevant kinematic variables, normalized
 2478 to the cross section of the respective processes and to the integrated luminosity.

2479 6.8 Signal discrimination

2480 The tHq signal is separated from the main backgrounds using a boosted decision
 2481 tree (BDT) classifier, trained on simulated signal and background events. A set
 2482 of discriminating variables are given as input to the BDT which produces a output
 2483 distribution maximizing the discrimination power. Table 6.10 lists the input variables
 2484 used while Figures 6.6 show their distributions for the relevant signal and background
 2485 samples, for the three lepton channel. Two BDT classifiers are trained for the two
 2486 main backgrounds expected in the analysis: events with prompt leptons from $t\bar{t}W$ and
 2487 $t\bar{t}Z$ (also referred to as $t\bar{t}V$), and events with non-prompt leptons from $t\bar{t}$. The datasets
 2488 used in the training are the tHq signal (see Tab. 6.1), and LO MADGRAPH samples
 2489 of $t\bar{t}W$ and $t\bar{t}Z$, in an admixture proportional to their respective cross sections (see
 2490 Tab. 6.4).

2491 The MVA analysis consist of two stages: first a “training” where the MVA method

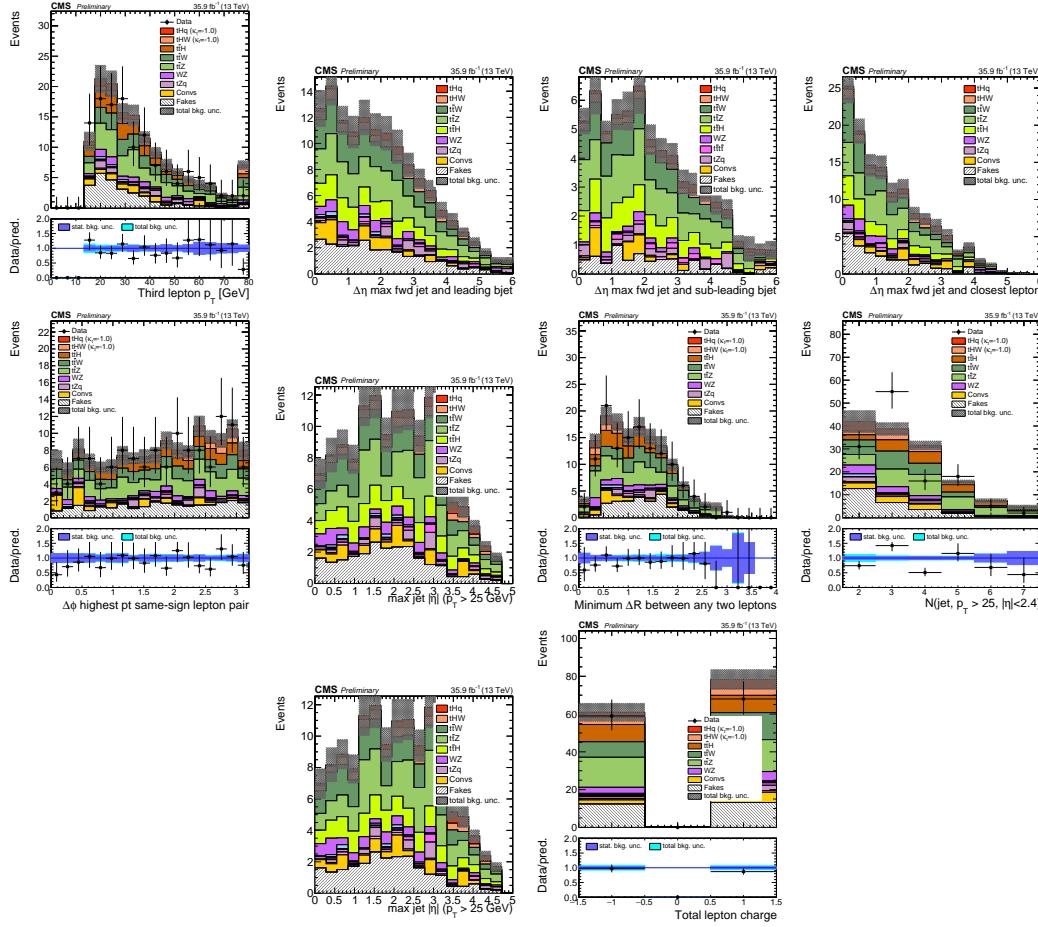


Figure 6.5: Distributions of input variables to the BDT for signal discrimination, three lepton channel, normalized to their cross section and to 35.9fb^{-1} .

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 ΔR between any two leptons
Lep3Pt/Lep2Pt	p_T of the 3 rd lepton (2 nd for ss2l)

Table 6.10: MVA input discriminating variables

2501 the distributions are similar in both background cases.

2502 From table 6.10, it is clear that the input variables are correlated to some extend.
 2503 These correlations play an important role for some MVA methods like the Fisher
 2504 discriminant method in which the first step consist of performing a linear transfor-
 2505 mation to an phase space where the correlations between variables are removed. In
 2506 case a boosted decision tree (BDT) method however, correlations do not affect the
 2507 performance. Figure 6.9 show the linear correlation coefficients for signal and back-
 2508 ground for the two training cases (the signal values are identical by construction). As
 2509 expected, strong correlations appears for variables related to the forward jet activity.
 2510 Same trend is seen in case of the same sign dilepton channel in Figure ??.

2511 6.8.1 Classifiers response

2512 Several MVA algorithms were evaluated to determine the most appropriate method
 2513 for this analysis. The plots in Fig. 6.10 (top) show the background rejection as a
 2514 function of the signal efficiency for $t\bar{t}$ and $t\bar{t}V$ trainings (ROC curves) for the different

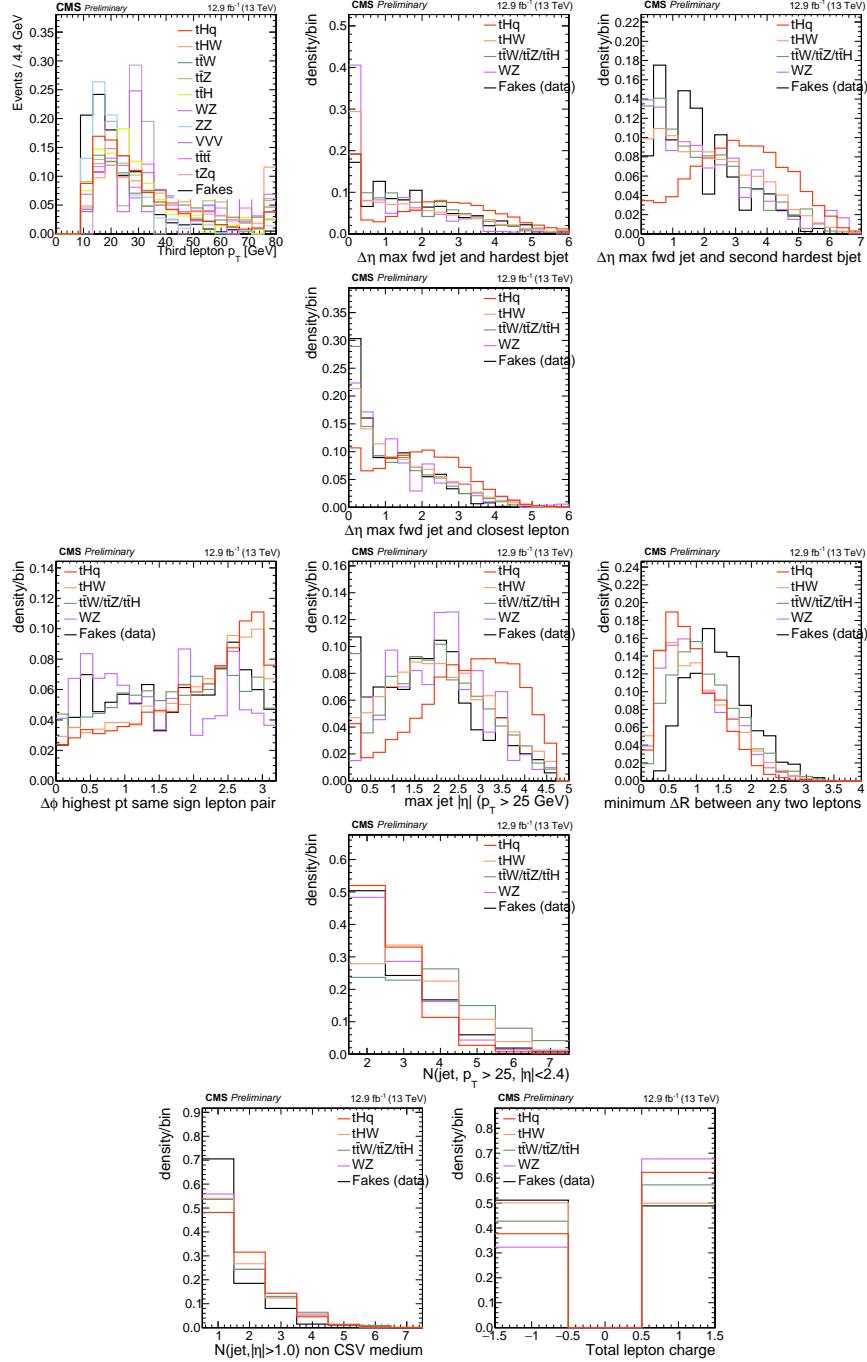


Figure 6.6: Distributions of input variables to the BDT for signal discrimination, three lepton channel.

algorithms that were evaluated.

In both cases the gradient boosted decision tree (“BDTA_GRAD”) classifier offers

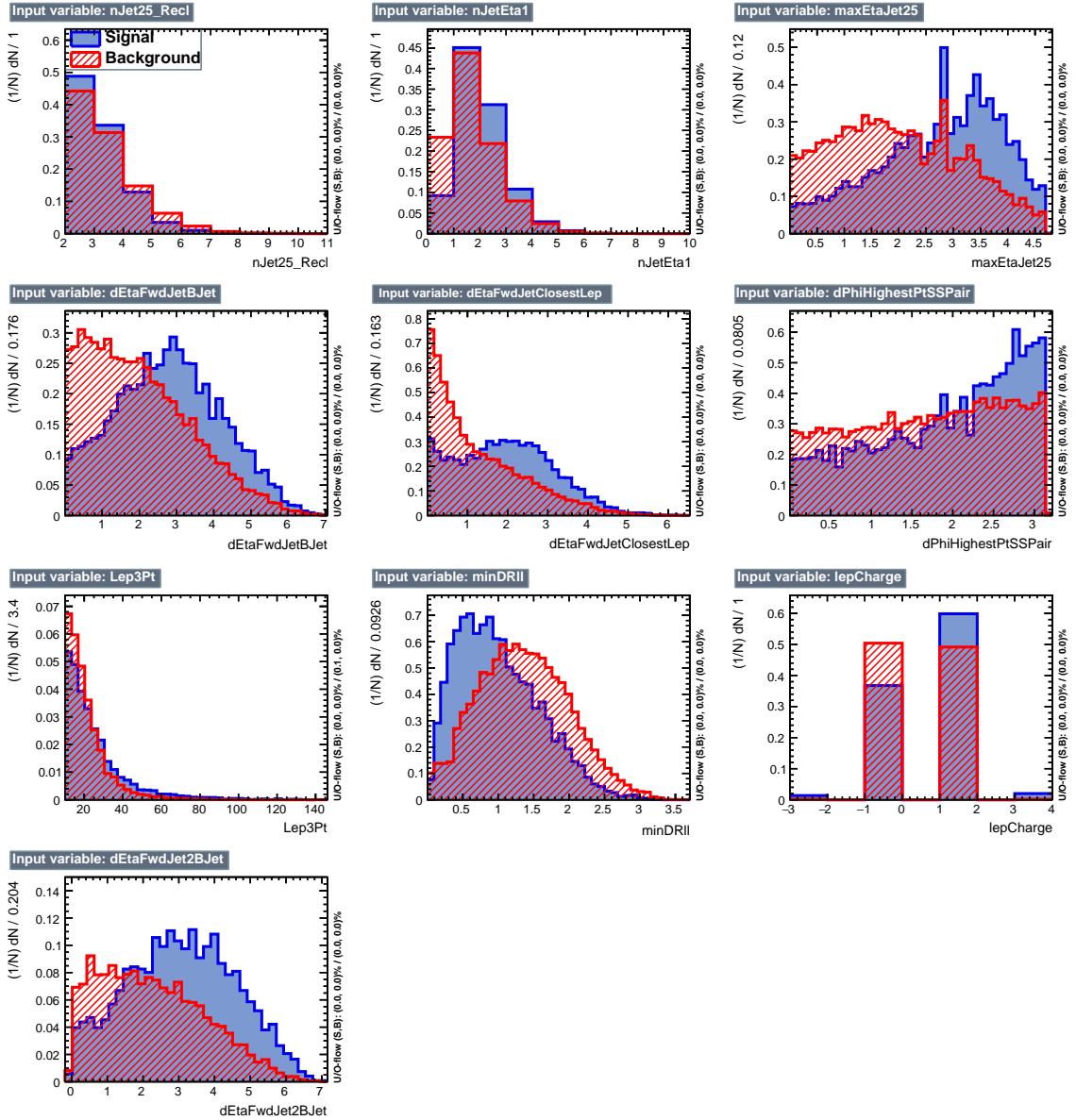


Figure 6.7: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}$) for the three lepton channel, discriminated against $t\bar{t}$ (fakes) background.

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.

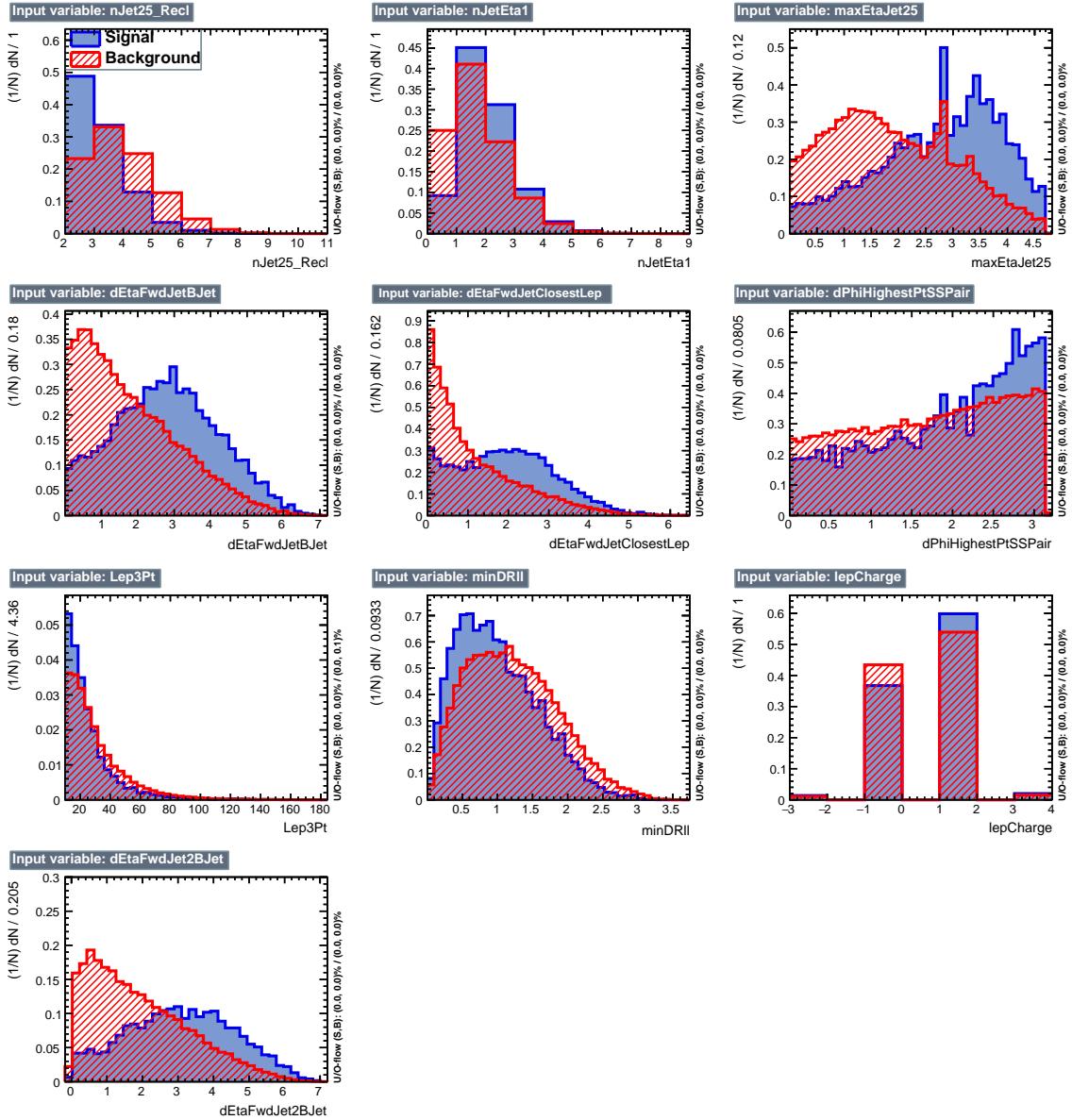


Figure 6.8: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}W+t\bar{t}Z$) for the three lepton channel, discriminated against $t\bar{t}V$ background.

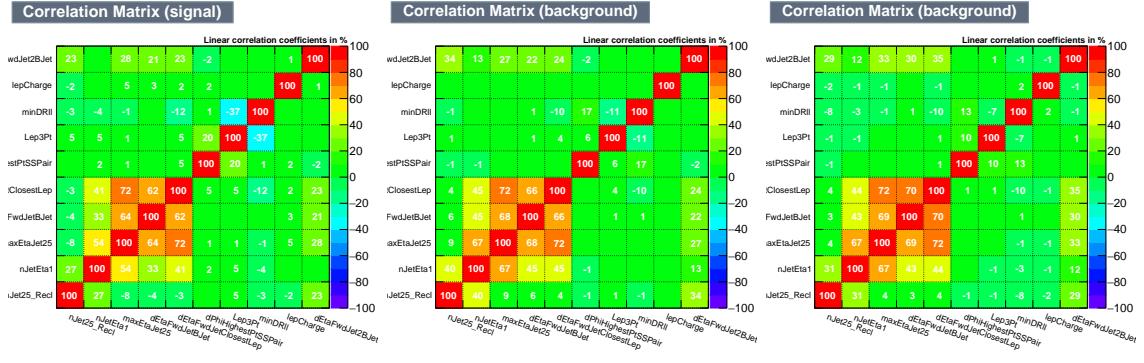


Figure 6.9: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the three lepton channel.

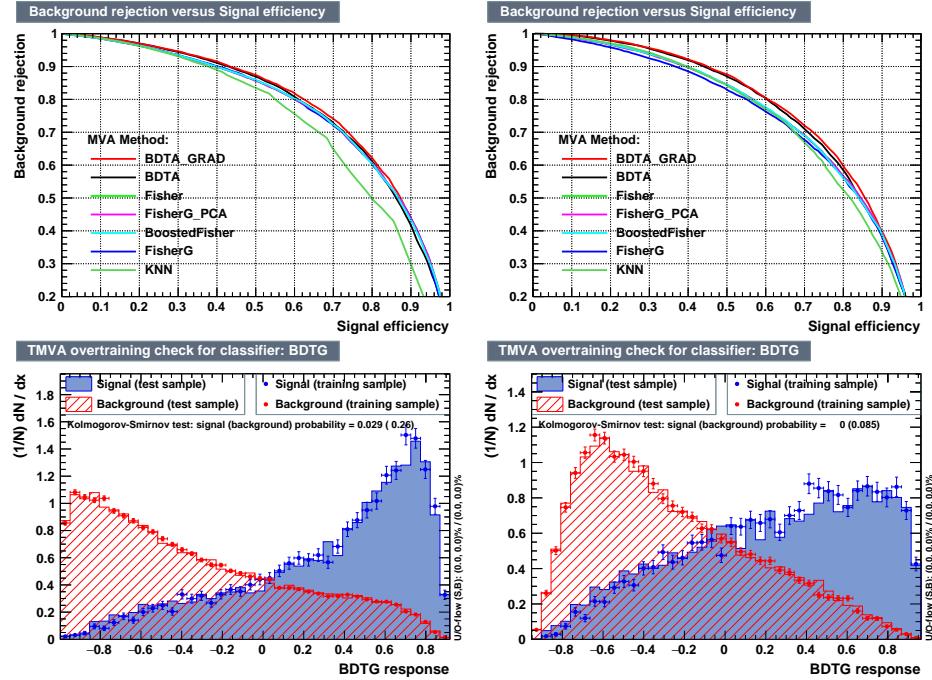


Figure 6.10: Top: background rejection vs signal efficiency (ROC curves) for various MVA classifiers (top) in the three lepton channel against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distributions for the gradient boosted decision trees, for training against $t\bar{t}V$ (left)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

2523 6.9 Additional discriminating variables

2524 Two additional discriminating variables were tested considering the fact that the
 2525 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

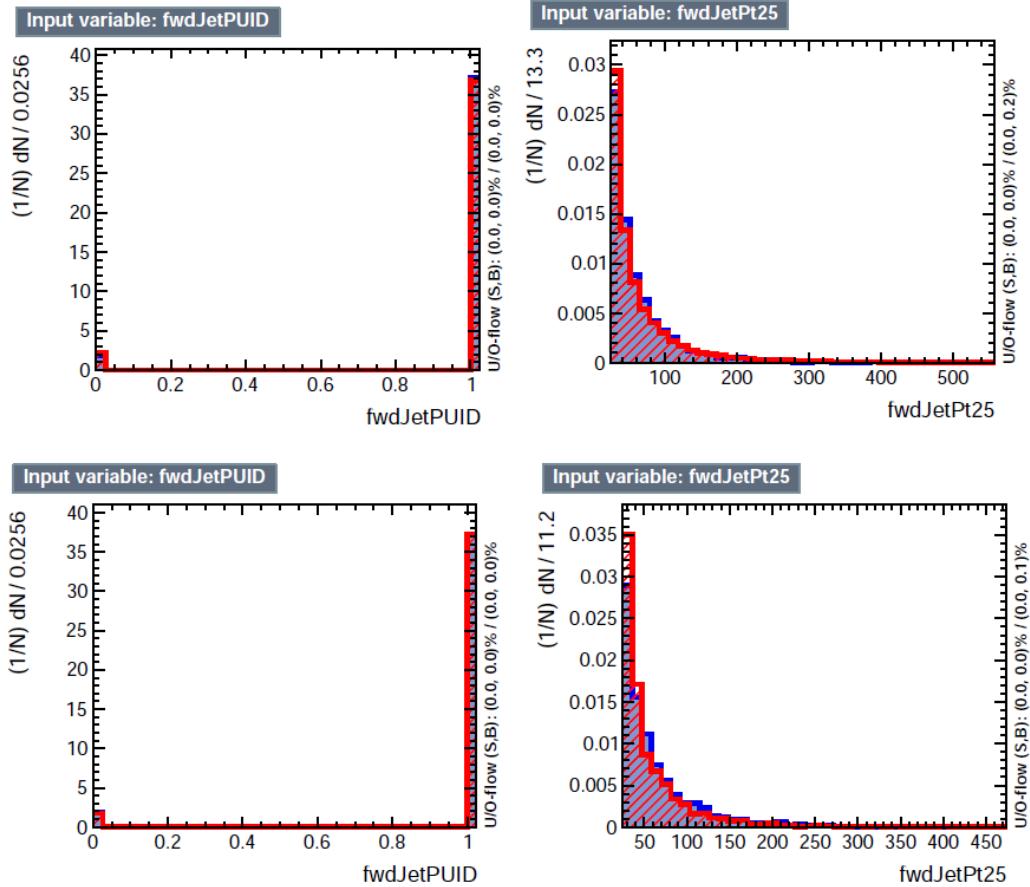


Figure 6.11: Additional discriminating variables distributions for ttv training (Top row) and tt training (bottom row) in the three lepton channel. The origin of the jets in the forward jet identification distribution is tagged as 0 for “pileup jets” while “real jets” are tagged as 1.

2536 trainings respectively, while `fwdJetPUID` was ranked 12 in both cases.

2537 The improvement in the discrimination performance provided by the additional
 2538 variables is about 1%, so it was decided not to include them in the procedure. Table
 2539 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%

2540 **References**

- 2541 [1] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". Phys-
 2542 ical Review. 74 (10): 1439-61, (1948).
- 2543 [2] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". Physical
 2544 Review. 76 (6): 769-89, (1949).
- 2545 [3] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum
 2546 Theory of Wave Fields". Progress of Theoretical Physics. 1 (2): 27-42, (1946).
- 2547 [4] D.J. Griffiths, "Introduction to electrodynamics". 4th ed. Pearson, (2013).
- 2548 [5] F. Mandl, G. Shaw. "Quantum field theory." Chichester, Wiley (2009).
- 2549 [6] F. Halzen, and A.D. Martin, "Quarks and leptons: An introductory course in
 2550 modern particle physics". New York: Wiley, (1984) .
- 2551 [7] File: Standard_Model_of_Elementary_Particle_dark.svg. (2017, June 12)
 2552 Wikimedia Commons, the free media repository. Retrieved November
 2553 27, 2017 from <https://www.collegiate-advanced-electricity.com/single-post/2017/04/10/The-Standard-Model-of-Particle-Physics>.
- 2555 [8] E. Noether, "Invariante Variationsprobleme", Nachrichten von der Gesellschaft
 2556 der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, vol. 1918,
 2557 pp. 235-257, (1918).

- 2558 [9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)
2559 and 2017 update.
- 2560 [10] M. Goldhaber, L. Grodzins, A.W. Sunyar “Helicity of Neutrinos”, Phys. Rev.
2561 109, 1015 (1958).
- 2562 [11] Palanque-Delabrouille N et al. “Neutrino masses and cosmology with Lyman-
2563 alpha forest power spectrum”, JCAP 11 011 (2015).
- 2564 [12] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. Physics Letters.
2565 8 (3): 214-215 (1964).
- 2566 [13] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking”
2567 (PDF). CERN Report No.8182/TH.401 (1964).
- 2568 [14] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking:
2569 II” (PDF). CERN Report No.8419/TH.412(1964).
- 2570 [15] M. Gell-Mann. “The Interpretation of the New Particles as Displaced Charged
2571 Multiplets”. Il Nuovo Cimento 4: 848. (1956).
- 2572 [16] T. Nakano, K, Nishijima. “Charge Independence for V-particles”. Progress of
2573 Theoretical Physics 10 (5): 581-582. (1953).
- 2574 [17] N. Cabibbo, “Unitary symmetry and leptonic decays” Physical Review Letters,
2575 vol. 10, no. 12, p. 531, (1963).
- 2576 [18] M.Kobayashi, T.Maskawa, “CP-violation in the renormalizable theory of weak
2577 interaction,” Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).
- 2578 [19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Com-
2579 mons, the free media repository. Retrieved November 27, 2017 from

- 2580 https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)
2581 .svg&oldid=247498592.
- 2582 [20] Georgia Tech University. Coupling Constants for the Fundamental
2583 Forces(2005). Retrieved January 10, 2018, from http://hyperphysics.phy-
2584 astr.gsu.edu/hbase/Forces/couple.html#c2
- 2585 [21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved Jan-
2586 uary 10, 2018, from https://profmattstrassler.com/articles-and-posts/particle-
2587 physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/
- 2588 [22] S.L. Glashow. “Partial symmetries of weak interactions”, Nucl. Phys. 22 579-
2589 588, (1961).
- 2590 [23] A. Salam, J.C. Ward. “Electromagnetic and weak interactions”, Physics Letters
2591 13 168-171, (1964).
- 2592 [24] S. Weinberg, “A model of leptons”, Physical Review Letters, vol. 19, no. 21, p.
2593 1264, (1967).
- 2594 [25] M. Peskin, D. Schroeder, “An introduction to quantum field theory”. Perseus
2595 Books Publishing L.L.C., (1995).
- 2596 [26] A. Pich. “The Standard Model of Electroweak Interactions”
2597 https://arxiv.org/abs/1201.0537
- 2598 [27] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved
2599 from: https://www.quantum-bits.org/?p=233
- 2600 [28] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255

- 2601 [29] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge
 2602 Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964)
 2603 doi:10.1103/PhysRevLett.13.321
- 2604 [30] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Re-
 2605 view Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.
- 2606 [31] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws
 2607 and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964).
 2608 doi:10.1103/PhysRevLett.13.585.
- 2609 [32] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with
 2610 the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012).
 2611 arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021
- 2612 [33] ATLAS collaboration. “Observation of a New Particle in the Search for the Stan-
 2613 dard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters
 2614 B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.
- 2615 [34] ATLAS collaboration; CMS collaboration (26 March 2015). “Combined Mea-
 2616 surement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with
 2617 the ATLAS and CMS Experiments”. Physical Review Letters. 114 (19): 191803.
 2618 arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.
- 2619 [35] LHC InternationalMasterclasses“When protons collide”. Retrieved from http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm
- 2621 [36] CMS Collaboration, “SM Higgs Branching Ratios and Total Decay Widths (up-
 2622 date in CERN Report4 2016)”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR> , last accessed on 17.12.2017.

- 2624 [37] R.Grant V. “Determination of Higgs branching ratios in $H \rightarrow W^+W^- \rightarrow l\nu jj$
2625 and $H \rightarrow ZZ \rightarrow l^+l^-jj$ channels”. Physics Department, University of Tennessee
2626 (Dated: October 31, 2012). Retrieved from <http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf>
- 2628 [38] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al.
2629 “Standard model Higgs-boson branching ratios with uncertainties”. Eur. Phys.
2630 J. C (2011) 71: 1753. <https://doi.org/10.1140/epjc/s10052-011-1753-8>
- 2631 [39] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production
2632 of Higgs and single top at hadron colliders”, Phys.Rev. D64 (2001) 094023,
2633 [hep-ph/0106293].
- 2634 [40] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the
2635 top-Higgs coupling from the 8 TeV LHC data,” Journal of High Energy Physics,
2636 vol. 07, p. 073, (2013).
- 2637 [41] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting de-
2638 generacies in Higgs couplings using single top production in association with a
2639 Higgs boson,” Journal of High Energy Physics, vol. 05, p. 022, (2013).
- 2640 [42] T.M. Tait and C.-P. Yuan, “Single top quark production as a window to physics
2641 beyond the standard model”, Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].
- 2642 [43] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in
2643 association with a single top quark at the LHC,” European Physical Journal C,
2644 vol. 75, p. 267, (2015).

- 2645 [44] CMS Collaboration, “Modelling of the single top-quark production in associa-
 2646 tion with the Higgs boson at 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV)
 2647 [viewauth/CMS/SingleTopHiggsGeneration13TeV](#), last accessed on 16.01.2018.
- 2648 [45] CMS Collaboration, “SM Higgs production cross sections at $\sqrt{s} =$
 2649 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)
 2650 [view/LHCPhysics/CERNYellowReportPageAt13TeV](#), last accessed on 16.01.2018.
- 2651 [46] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.
- 2652 [47] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [[arXiv:1211.0499](https://arxiv.org/abs/1211.0499)
 2653 [hep-ph]].
- 2654 [48] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated
 2655 production at the LHC”, European Physical Journal C, vol. 77, p. 34, (2017).
 2656 [arXiv:1607.05862](https://arxiv.org/abs/1607.05862)
- 2657 [49] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross
 2658 Sections: 4.Deciphering the Nature of the Higgs Sector”, [arXiv:1610.07922](https://arxiv.org/abs/1610.07922).
- 2659 [50] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top
 2660 Couplings in Associated Production”, JHEP 1404 (2014) 004, [[arXiv:1312.5736](https://arxiv.org/abs/1312.5736)].
- 2661 [51] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass
 2662 of the Higgs boson and tests of compatibility of its couplings with the standard
 2663 model predictions using proton collisions at 7 and 8 TeV,” [arXiv:1412.8662](https://arxiv.org/abs/1412.8662).
- 2664 [52] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the
 2665 Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton
 2666 collision data”, ATLAS-CONF-2014-009.

- 2667 [53] ATLAS and CMS Collaborations, “Measurements of the Higgs boson produc-
 2668 tion and decay rates and constraints on its couplings from a combined ATLAS
 2669 and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,” (2016).
 2670 CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.
- 2671 [54] File:Cern-accelerator-complex.svg. Wikimedia Commons,
 2672 the free media repository. Retrieved January, 2018 from
 2673 <https://commons.wikimedia.org/wiki/File:Cern-accelerator-complex.svg>
- 2674 [55] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”,
 2675 (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Re-
 2676 trieval from <https://cds.cern.ch/record/841542>
- 2677 [56] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-
 2678 2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762
- 2679 [57] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).
- 2680 [58] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March
 2681 1983, CERN-AC-8303511. Retrieved from <https://cds.cern.ch/record/615852>.
- 2682 [59] CERN Photographic Service “Animation of CERN’s accelerator net-
 2683 work”, 14 October 2013. DOI: 10.17181/cds.1610170 Retrieved from
 2684 <https://videos.cern.ch/record/1610170>
- 2685 [60] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17, 2013. Re-
 2686 trieval from <https://www.britannica.com/technology/particle-accelerator>.
- 2687 [61] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-
 2688 AC-0007016. Retrieved from <https://cds.cern.ch/record/41567>.

- 2689 [62] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”.
2690 March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-
2691 PHO-1998-325. Retrieved from <https://cds.cern.ch/record/841511>
- 2692 [63] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025
2693 retrieved from <https://cds.cern.ch/record/40524>.
- 2694 [64] CMS Collaboration “Public CMS Luminosity Information”.
2695 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016>
2696 _proton_proton_13_TeV_collis, last accessed 24.01.2018
- 2697 [65] J.L. Caron. “LHC Layout” AC Collection. Legacy of AC. Pictures
2698 from 1992 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from
2699 <https://cds.cern.ch/record/841573>.
- 2700 [66] J.A. Coarasa. “The CMS Online Cluster: Setup, Operation and Maintenance
2701 of an Evolving Cluster”. ISGC 2012, 26 February - 2 March 2012, Academia
2702 Sinica, Taipei, Taiwan.
- 2703 [67] CMS Collaboration. “The CMS experiment at the CERN LHC” JINST 3 S08004
2704 (2008).
- 2705 [68] CMS Collaboration. “CMS detector drawings 2012” CMS-PHO-GEN-2012-002.
2706 Retrieved from <http://cds.cern.ch/record/1433717>.
- 2707 [69] R. Breedon. “View through the CMS detector during the cooldown of the
2708 solenoid on February 2006. CMS Collection”, February 2006, CMS-PHO-
2709 OREACH-2005-004, Retrieved from <https://cds.cern.ch/record/930094>.
- 2710 [70] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector
2711 Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.

- 2712 [71] CMS Collaboration. “Description and performance of track and primary-vertex
2713 reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no.
2714 10, p. P10009,(2014).
- 2715 [72] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner
2716 Barrel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from
2717 <https://cds.cern.ch/record/1431467>.
- 2718 [73] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm
2719 and beauty hadrons Chicago, June 28-July 3 2004.
- 2720 [74] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at
2721 the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013.
2722 arXiv:1307.7135
- 2723 [75] L. Veillet. “End assembly of HB with EB rails and rotation in-
2724 side SX ”,January 2002. CMS-PHO-HCAL-2002-002. Retrieved from
2725 <https://cds.cern.ch/record/42594>.
- 2726 [76] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the
2727 UX5 cavern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from
2728 <https://cds.cern.ch/record/1019185>
- 2729 [77] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web
2730 site: Taking a closer look at LHC. Retrieved from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger
- 2732 [78] WLCG Project Office, “Documents & Reference - Tiers - Structure,” (2014).
2733 <http://wlcg.web.cern.ch/documents-reference> , last accessed on 30.01.2018.

- 2734 [79] CMS Collaboration. “CMSSW Application Framework”,
 2735 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>,
 2736 last accesses 06.02.2018
- 2737 [80] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event gen-
 2738 erators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011
- 2739 [81] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics
 2740 of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-
 2741 3-540-71060-8 (2007)
- 2742 [82] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter.”
 2743 <http://hepdata.cedar.ac.uk/pdf/pdf3.html> , last accessed on 02.02.2018.
- 2744 [83] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a
 2745 reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).
- 2746 [84] B. Andersson, G. Gustafson, G.Ingelman and T. Sjostrand, “Parton fragmen-
 2747 tation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145,
 2748 1983.
- 2749 [85] CMS Collaboration, “Event generator tunes obtained from underlying event
 2750 and multiparton scattering measurements;” European Physical Journal C, vol.
 2751 76, no. 3, p. 155, (2016).
- 2752 [86] J. Alwall et. al., “The automated computation of tree-level and next-to-leading
 2753 order differential cross sections, and their matching to parton shower simula-
 2754 tions,” Journal of High Energy Physics, vol. 07, p. 079, (2014).

- 2755 [87] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with
2756 Parton Shower simulations: the POWHEG method,” Journal of High Energy
2757 Physics, vol. 11, p. 070, (2007).
- 2758 [88] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments
2759 and Methods in Physics, vol. A506, pp. 250–303, (2003).
- 2760 [89] J.Allison et.al.,“Recent developments in Geant4”, Nuclear Instruments and
2761 Methods in Physics Research A 835 (2016) 186-225.
- 2762 [90] CMS Collaboration “Full Simulation Offline Guide”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation>, last accessed 04.02.2018
- 2764 [91] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf.
2765 Ser. 513 022012 (2014)
- 2766 [92] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description
2767 with the CMS detector”, JINST 12 P10003 (2017) <https://doi.org/10.1088/1748-0221/12/10/P10003>.
- 2769 [93] The CMS Collaboration. “ Description and performance of track and pri-
2770 mary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014).
2771 doi:10.1088/1748-0221/9/10/P10009
- 2772 [94] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides.
2773 Retrieved from https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf
- 2775 [95] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by
2776 the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).

- 2777 [96] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of
 2778 electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf
 2779 C 0303241 (2003) TULT009 [physics/0306087].
- 2780 [97] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification,
 2781 Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.
- 2782 [98] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”,
 2783 CMS Note 2007-008 (2007).
- 2784 [99] CMS collaboration, “Performance of CMS muon reconstruction in pp collision
 2785 events at $\sqrt{s} = 7$ TeV ”, JINST 7 P10002 2012, [arXiv:1206.4071].
- 2786 [100] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,”
 2787 Journal of High Energy Physics, vol. 04, p. 063, (2008).
- 2788 [101] B. Dorney. “Anatomy of a Jet in CMS”. Quantum Diaries. June
 2789 1st, 2011. Retrieved from <https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>
- 2791 [102] The CMS Collaboration.“Event Displays from the high-energy collisions at 7
 2792 TeV”, May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from <https://cds.cern.ch/record/1429614>.
- 2794 [103] The CMS collaboration. “Determination of jet energy calibration and transverse
 2795 momentum resolution in CMS”. JINST 6 P11002 (2011). <http://dx.doi.org/10.1088/1748-0221/6/11/P11002>
- 2797 [104] The CMS Collaboration, “Introduction to Jet Energy Corrections at
 2798 CMS.”. <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>, last ac-
 2799 cessed 10.02.2018.

- 2800 [105] CMS Collaboration Collaboration. “Identification of b quark jets at the CMS
2801 Experiment in the LHC Run 2”. Tech. rep. CMS-PAS-BTV-15-001. Geneva:
2802 CERN, (2016). <https://cds.cern.ch/record/2138504>.
- 2803 [106] CMS Collaboration Collaboration. “Performance of missing energy reconstruc-
2804 tion in 13 TeV pp collision data using the CMS detector”. Tech. rep. CMS-PAS-
2805 JME16-004. Geneva: CERN, 2016. <https://cds.cern.ch/record/2205284>.
- 2806 [107] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,
2807 Retrieved from <http://cms.web.cern.ch/sites/cms.web.cern.ch/files/>
2808 styles/large/public/field/image/HIG13004_Event01_0.png?itok=
2809 LAwZzPHR
- 2810 [108] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,
2811 Retrieved from <http://cms.web.cern.ch/sites/cms.web.cern.ch/>
2812 files/styles/large/public/field/image/TOP12035_Event01.png?itok=
2813 uMdnSqzC
- 2814 [109] K. Skovpen. “Event displays highlighting the main properties of heavy flavour
2815 jets in the CMS Experiment”, Aug 2017, CMS-PHO-EVENTS-2017-006. Re-
2816 trieval from <https://cds.cern.ch/record/2280025>.
- 2817 [110] G. Cowan. “Topics in statistical data analysis for high-energy physics”.
2818 arXiv:1012.3589v1
- 2819 [111] A. Hoecker et al., “TMVA-Toolkit for multivariate data analysis”
2820 arXiv:physics/0703039v5 (2009)

- 2821 [112] L. Lista. “Statistical Methods for Data Analysis in Particle Physics”, 2nd
2822 ed. Springer International Publishing. (2017) <https://dx.doi.org/10.1007/978-3-319-62840-0>
- 2824 [113] I. Antcheva et al., “ROOT-A C++ framework for petabyte data storage, sta-
2825 tistical analysis and visualization ,” Computer Physics Communications, vol.
2826 182, no. 6, pp. 1384â€¢1385, (2011).
- 2827 [114] Y. Coadou. “Boosted decision trees”, ESIPAP, Archamps, 9 Febru-
2828 ary 2016. Lecture. Retrieved from https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf
- 2831 [115] J.H. Friedman. “Greedy function approximation: A gradient boosting ma-
2832 chine”. Ann. Statist. Volume 29, Number 5 (2001), 1189-1232. https://projecteuclid.org/download/pdf_1/euclid-aos/1013203451.
- 2834 [116] F. James, M. Roos, “MINUIT: Function minimization and error analysis”. Cern
2835 Computer Centre Program Library, Geneve Long Write-up No. D506, 1989
- 2836 [117] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests
2837 of statistical hypotheses”. Philosophical Transactions of the Royal Society of
2838 London. Series A, Containing Papers of a Mathematical or Physical Character.
2839 Vol. 231 (1933), pp. 289-337
- 2840 [118] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated pro-
2841 duction via gluon fusion in the SM and the 2HDM”, JHEP 06 (2015) 065,
2842 [https://dx.doi.org/10.1007/JHEP06\(2015\)065](https://dx.doi.org/10.1007/JHEP06(2015)065), arXiv:1503.01656.

- 2843 [119] ATLAS Collaboration, “Measurements of Higgs boson produc-
 2844 tion and couplings in diboson final states with the ATLAS
 2845 detector at the LHC”, Phys. Lett. B726 (2013) 88âš119,
 2846 doi:10.1016/j.physletb.2014.05.011,10.1016/j.physletb.2013.08.010,
 2847 arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].
- 2848 [120] CMS Collaboration, “Search for the associated production of a Higgs boson
 2849 with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV”, JHEP 06
 2850 (2016) 177,doi:10.1007/JHEP06(2016)177, arXiv:1509.08159.
- 2851 [121] B. Stieger, C. Jordà Lope et al., “Search for Associated Production of a Single
 2852 Top Quark and a Higgs Boson in Leptonic Channels”, CMS Analysis Note CMS
 2853 AN-14-140, 2014.
- 2854 [122] M. Peruzzi, C. Mueller, B. Stieger et al., “Search for ttH in multilepton final
 2855 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-16-211, 2016.
- 2856 [123] CMS Collaboration, “Search for H to bbar in association with a single top quark
 2857 as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis
 2858 Summary CMS-PAS-HIG-16-019, 2016.
- 2859 [124] CMS Collaboration, “Search for production of a Higgs boson and a single top
 2860 quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV”, CMS
 2861 Physics Analysis Summary CMS-PAS-HIG-17-005, 2016.
- 2862 [125] B. WG, “BtagRecommendation80XReReco”, February, 2017. <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>.

- 2864 [126] CMS Collaboration, “Identification of b quark jets at the CMS Experiment
2865 in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001,
2866 2016.
- 2867 [127] M. Peruzzi, F. Romeo, B. Stieger et al., “Search for ttH in multilepton final1
2868 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-17-029, 2017.
- 2869 [128] CMS Collaboration, “Baseline muon selections for Run-II.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonIdRun2>, last accessed on
2870 24.02.2018.
- 2872 [129] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January,
2873 2015. <https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf>.
- 2875 [130] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”,
2876 October, 2014. <https://indico.cern.ch/event/298249/contribution/3/material/slides/0.pdf>.
- 2878 [131] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic
2879 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.
- 2881 [132] CMS Collaboration. “Tag and Probe”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe>, last accessed on 02.03.2018.
- 2883 [133] B. Maier, “SingleTopHiggProduction13TeV”, February, 2016.
2884 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV>.