

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	8
³²	2.2.2 Fundamental interactions	13
³³	2.2.3 Gauge bosons	18
³⁴	2.3 Electroweak unification and the Higgs mechanism	20
³⁵	2.3.1 Spontaneous symmetry breaking (SSB)	28
³⁶	2.3.2 Higgs mechanism	32
³⁷	2.3.3 Masses of the gauge bosons	35

38	2.3.4	Masses of the fermions	36
39	2.3.5	The Higgs field	37
40	2.3.6	Production of Higgs bosons at LHC	38
41	2.3.7	Higgs boson decay channels	42
42	2.4	Associated production of a Higgs boson and a single Top quark.	43
43	2.5	The CP-mixing in tH processes	47
44	2.6	Experimantal status of the anomalous Higg-fermion coupling.	52
45	3	The CMS experiment at the LHC	54
46	3.1	Introduction	54
47	3.2	The LHC	55
48	3.3	The CMS experiment	65
49	3.3.1	Coordinate system	67
50	3.3.2	Pixels detector	68
51	3.3.3	Silicon strip tracker	70
52	3.3.4	Electromagnetic calorimeter	72
53	3.3.5	Hadronic calorimeter	73
54	3.3.6	Superconducting solenoid magnet	75
55	3.3.7	Muon system	76
56	3.3.8	CMS trigger system	77
57	3.3.9	CMS computing	79
58	4	Event generation, simulation and reconstruction	83
59	4.1	Event generation	84
60	4.2	Monte Carlo Event Generators.	88
61	4.3	CMS detector simulation.	89
62	4.4	Event reconstruction.	91

63	4.4.1	Particle-Flow Algorithm.	92
64	4.4.2	Event reconstruction examples	104
65	5	Statistical methods	107
66	5.1	Multivariate analysis	107
67	5.1.1	Boosted decision trees (BDT)	110
68	5.2	...	112
69	5.3	MVA methods, NN, BDT, boosting, overtraining, variable ranking	114
70	5.4	statistical inference, likelihood parametrization	114
71	5.5	nuisance parameters	114
72	5.6	exclusion limits	114
73	5.7	asymptotic limits	114
74	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	115
76	6.1	Introduction	115
77	6.2	Data and MC Samples	117
78	6.2.1	Full 2016 dataset and MC samples	117
79	6.2.2	Triggers	120
80	6.2.2.1	Trigger efficiency scale factors	120
81	6.3	Object Identification and event selection	121
82	6.3.1	Jets and b tagging	121
83	6.3.2	Lepton selection	122
84	6.3.3	Lepton selection efficiency	123
85	6.4	Background predictions	124
86	6.5	Signal discrimination	125
87	6.5.1	Classifiers response	129

88	6.6 Additional discriminating variables	132
89	Bibliography	133
90	References	135

⁹¹ List of Figures

92	2.1	Standard model of particle physics.	4
93	2.2	Transformations between quarks	12
94	2.3	Fundamental interactions in nature.	13
95	2.4	SM interactions diagrams	14
96	2.5	Neutral current processes	21
97	2.6	Spontaneous symmetry breaking mechanism	29
98	2.7	SSB Potential form	30
99	2.8	Potential for complex scalar field	31
100	2.9	SSB mechanism for complex scalar field	32
101	2.10	Proton-Proton collision	39
102	2.11	Higgs boson production mechanism Feynman diagrams	40
103	2.12	Higgs boson production cross section and decay branching ratios	41
104	2.13	Associated Higgs boson production mechanism Feynman diagrams	43
105	2.14	Cross section for tHq process as a function of κ_t	46
106	2.15	Cross section for tHW process as a function of κ_{Htt}	47
107	2.16	NLO cross section for tX_0 and $t\bar{t}X_0$.	50
108	2.17	NLO cross section for tWX_0 , $t\bar{t}X_0$.	51

109	2.18 Two dimentional κ_t - κ_V plot of the coupling modifiers. ATLAS and CMS 110 combination.	52
111	3.1 CERN accelerator complex	55
112	3.2 LHC protons source. First acceleration stage.	56
113	3.3 The LINAC2 accelerating system at CERN.	57
114	3.4 LHC layout and RF cavities module.	58
115	3.5 LHC dipole magnet.	60
116	3.6 Integrated luminosity delivered by LHC and recorded by CMS during 2016	62
117	3.7 LHC interaction points	63
118	3.8 Multiple pp collision bunch crossing at CMS.	65
119	3.9 Layout of the CMS detector	66
120	3.10 CMS detector coordinate system	67
121	3.11 CMS pixel detector schematic view.	70
122	3.12 SST Schematic view.	71
123	3.13 CMS ECAL schematic view	72
124	3.14 CMS HCAL schematic view	74
125	3.15 CMS solenoid magnet	75
126	3.16 CMS Muon system schematic view	76
127	3.17 CMS Level-1 trigger architecture	78
128	3.18 WLCG structure	80
129	3.19 Data flow from CMS detector through hardware Tiers	82
130	4.1 Event generation process.	84
131	4.2 Particle flow algorithm.	92
132	4.3 Jet reconstruction.	100
133	4.4 Jet energy corrections.	101

134	4.5	Secondary vertex in a b-hadron decay.	103
135	4.6	HIG-13-004 Event 1 reconstruction.	104
136	4.7	$e\mu$ event reconstruction.	105
137	4.8	Recorded event reconstruction.	106
138	5.1	Scatter plots-MVA event classification.	109
139	5.2	Scalar test statistical.	109
140	6.1	The two leading-order diagrams of tHq production.	116
141	6.2	Input variables to the BDT for signal discrimination normalized.	126
142	6.3	Input variables to the BDT for signal discrimination not normalized.	128
143	6.4	BDT inputs as seen by TMVA against $t\bar{t}$.	129
144	6.5	BDT inputs as seen by TMVA against $t\bar{t}V$.	130
145	6.6	Correlation matrices for the input variables in the TMVA.	131
146	6.7	MVA classifiers performance.	131
147	6.8	Additional discriminating variables distributions.	133

¹⁴⁸ List of Tables

149	2.1	Fermions of the SM.	5
150	2.2	Fermion masses.	6
151	2.3	Leptons properties.	9
152	2.4	Quarks properties.	9
153	2.5	Fermion weak isospin and weak hypercharge multiplets.	11
154	2.6	Fundamental interactions features.	15
155	2.7	SM gauge bosons.	20
156	2.8	Higgs boson properties.	38
157	2.9	Predicted branching ratios for a SM Higgs boson with $m_H = 125 \text{ GeV}/c^2$	42
158	2.10	Predicted SM cross sections for tH production at $\sqrt{s} = 13 \text{ TeV}$	44
159	2.11	Predicted enhancement of the tHq and tHW cross sections at LHC	48
160	6.1	Signal samples and their cross section and branching fraction.	117
161	6.2	κ_V and κ_t combinations.	118
162	6.3	List of background samples used in this analysis (CMSSW 80X).	119
163	6.4	Leading-order $t\bar{t}W$ and $t\bar{t}Z$ samples used in the signal BDT training.	119
164	6.5	Table of high-level triggers that we consider in the analysis.	120
165	6.6	Trigger efficiency scale factors and associated uncertainties.	121
166	6.7	Requirements on each of the three muon selections.	122

167	6.8 Criteria for each of the three electron selections.	123
168	6.9 MVA input discriminating variables	127
169	6.10 TMVA input variables ranking for BDTA_GRAD method	132
170	6.11 TMVA configuration used in the BDT training.	132
171	6.12 ROC-integral for all the testing cases.	134

¹⁷² Chapter 1

¹⁷³ INTRODUCTION

¹⁷⁴ **Chapter 2**

¹⁷⁵ **Theoretical approach**

¹⁷⁶ **2.1 Introduction**

¹⁷⁷ The physical description of the universe is a challenge that physicists have faced by
¹⁷⁸ making theories that refine existing principles and proposing new ones in an attempt
¹⁷⁹ to embrace emerging facts and phenomena.

¹⁸⁰

¹⁸¹ At the end of 1940s Julian Schwinger [1] and Richard P. Feynman [2], based in the
¹⁸² work of Sin-Itiro Tomonaga [3], developed an electromagnetic theory consistent with
¹⁸³ special relativity and quantum mechanics that describes how matter and light inter-
¹⁸⁴ act; the so-called “quantum eletrodynamics” (QED) had born.

¹⁸⁵

¹⁸⁶ QED has become the guide in the development of theories that describe the universe.
¹⁸⁷ It was the first example of a quantum field theory (QFT), which is the theoretical
¹⁸⁸ framework for building quantum mechanical models that describes particles and their
¹⁸⁹ interactions. QFT is composed of a set of mathematical tools that combines classical
¹⁹⁰ fields, special relativity and quantum mechanics, while keeping the quantum point

191 particles and locality ideas.

192 This chapter gives an overview of the standard model of particle physics, starting
 193 with a description of the particles and interactions that compose it, followed by a
 194 description of the electroweak interaction, the Higgs boson and the associated pro-
 195 duction of Higgs boson and a single top quark (tH). The description contained in
 196 this chapter is based on references [4–6].

197 2.2 Standard model of particle physics

198 Particle physics at the fundamental level is modeled in terms of a collection of in-
 199 teracting particles and fields in a theory known as the “standard model of particle
 200 physics (SM)”¹.

201

202 The full picture of the SM is composed of three fields², whose excitations are inter-
 203 preted as particles called mediators or force-carriers; a set of fields, whose excitations
 204 are interpreted as elementary particles, interacting through the exchange of those
 205 mediators and a field that gives the mass to elementary particles. Figure 2.1 shows
 206 an scheme of the SM particles organization. In addition to the particles in the scheme
 207 (but not listed in it), their corresponding anti-particles, with opposite quantum num-
 208 bers, are also part of the picture; some particles are their own anti-particles, like
 209 photon or Higgs, or anti-particle is already listed like in the W^+ and W^- case.

210

211 The mathematical formulation of the SM is based on group theory and the use of
 212 Noether’s theorem [8] which states that for a physical system modeled by a Lagrangian

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 [?] is quite comprehensive and detailed.

² Note that gravitational field is not included in the standard model formulation

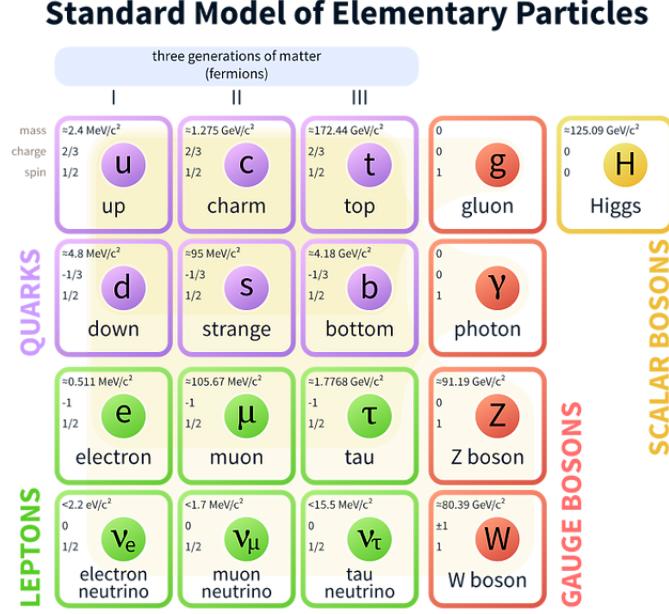


Figure 2.1: Schematic representation of the Standard model of particle physics. SM is a theoretical model intended to describe three of the four fundamental forces of the universe in terms of a set of particles and their interactions. [7].

that is invariant under a group of transformations a conservation law is expected. For instance, a system described by a time-independent Lagrangian is invariant (symmetric) under time changes (transformations) with the total energy conservation law as 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.

220

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:

- 225 • Strong: $SU(3)_C$ associated to color charge
- 226 • Weak: $SU(2)_L$ associated to weak isospin and chirality
- 227 • Electromagnetic: $U(1)_Y$ associated to weak hypercharge and electric charge
- 228 It will be shown that the electromagnetic and weak interactions are combined in
 229 the so-called electroweak interaction where chirality, hypercharge, weak isospin and
 230 electric charge are the central concepts.

231 **2.2.1 Fermions**

232 The basic constituents of the ordinary matter at the lowest level, which form the set
 233 of elementary particles in the SM formulation, are quarks and leptons. All of them
 234 have spin 1/2, therefore they are classified as fermions since they obey Fermi-Dirac
 235 statistics. There are six “flavors” of quarks and three of leptons organized in three
 236 generations, or families, as shown in table 2.1.

237

		Generation		
		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 WI between leptons is limited to the members of the same generation; WI between quarks is not limited but greatly favoured, to same generation members.

238

239 There is a mass hierarchy between generations (see table 2.2), where the higher gener-
 240 ation particles decays to the lower one, which can explain why the ordinary matter is

made of particles in the first generation. In the SM, neutrinos are modeled as massless particles so they are not subject to this mass hierarchy; however, today it is known that neutrinos are massive so the hierarchy could be restated. The reason behind this mass hierarchy is one of the most important open questions in particle physics, and it becomes more puzzling when noticing that the mass difference between first and second generation fermions is small compared to the mass difference with 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
τ_μ	Unknown	b	4.18×10^3

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

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

2.2.1.1 Leptons

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

258 by EI because they don't carry electric charge.

259

260 Another feature of the leptons that is fundamental in the mathematical description
261 of the SM is the chirality, which is closely related to spin and helicity. Helicity defines
262 the handedness of a particle by relating its spin and momentum such that if they
263 are parallel then the particle is right-handed; if spin and momentum are antiparallel
264 the particle is said to be left-handed. The study of parity conservation (or viola-
265 tion) in β -decay has shown that only left-handed electrons/neutrinos or right-handed
266 positrons/anti-neutrinos are created [10]; the inclusion of that feature in the theory
267 was achieved by using projection operators for helicity, however, helicity is frame de-
268 pendent for massive particles which makes it not Lorentz invariant and then another
269 related attribute has to be used: *chirality*.

270

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

276

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

284

285 Each generation of leptons is seen as a weak isospin doublet.³ The left-handed charged
 286 lepton and its associated left-handed neutrino are arranged in doublets of weak isospin
 287 T=1/2 while their right-handed partners are singlets:

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

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

293

294 Altough all three flavor neutrinos have been observed, their masses remain unknown
 295 and only some estimations have been made [11]. The main reason is that the fla-
 296 vor eigenstates are not the same as the mass eigenstates which implies that when
 297 a neutrino is created its mass state is a linear combination of the three mass eigen-
 298 states and experiments can only probe the squared difference of the masses. The
 299 Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix encode the relationship
 300 between flavor and mass eigenstates.

301

302 2.2.1.2 Quarks

303 Quarks are the basic constituents of protons and neutrons. The way quarks join to
 304 form bound states, called “hadrons”, is through the SI. Quarks are affected by all the

³ 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.

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 (τ_μ)	0	1/2	0	0	1	Unknown

Table 2.3: Leptons 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.

305 fundamental interactions which means that they carry all the four types of charges:
 306 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: Quarks 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.

307
 308 Table 2.4 summarizes the features of quarks, among which the most particular is
 309 their fractional electric charge. Note that fractional charge is not a problem, given
 310 that quarks are not found isolated, but serves to explain how composed particles are
 311 formed out of two or more valence quarks⁴.
 312

⁴ 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.

313 Color charge is the responsible for the SI between quarks and is the symmetry
 314 ($SU(3)_C$) that defines the formalism to describe SI. There are three colors: red (r),
 315 blue(b) and green(g) and their corresponding three anti-colors; thus each quark carries
 316 one color unit while anti-quarks carries one anti-color unit. As said above, quarks are
 317 not allowed to be isolated due to the color confinement effect, therefore their features
 318 have been studied indirectly by observing their bound states created when:

- 319 • one quark with a color charge is attracted by an anti-quark with the correspond-
 320 ing anti-color charge forming a colorless particle called a “meson.”
- 321 • three quarks (anti-quarks) with different color (anti-color) charges are attracted
 322 among them forming a colorless particle called a “baryon(anti-baryon).”

323 In practice, when a quark is left alone isolated a process called “hadronization” occurs
 324 where the quark emits gluons (see section 2.2.3) which eventually will generate new
 325 quark-antiquark pairs and so on; those quarks will recombine to form hadrons that
 326 will decay into leptons. This proliferation of particles looks like a “jet” coming from
 327 the isolated quark. More details about the hadronization process and jet structure
 328 will be given in chapter4.

329 In the first version of the quark model (1964), M. Gell-Mann [12] and G. Zweig
 330 [13, 14] developed a consistent way to classify hadrons according to their properties.
 331 Only three quarks (u, d, s) were involved in a scheme in which all baryons have
 332 baryon number $B=1$ and therefore quarks have $B=1/3$; non-baryons have $B=0$. The
 333 scheme organizes baryons in a two-dimensional space ($I_3 - Y$); Y (hypercharge) and I_3
 334 (isospin) are quantum numbers related by the Gell-Mann-Nishijima formula [15, 16]:

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

335 where $Y = B + S + C + T + B'$ are the quantum numbers listed in table 2.4. Baryon
 336 number is conserved in SI and EI which means that single quarks cannot be created
 337 but in pairs $q - \bar{q}$.

338

339 There are six quark flavors organized in three generations (see table 2.1) following a
 340 mass hierarchy which, again, implies that higher generations decay to first generation
 341 quarks.

	Quarks			T_3	Y_W	Leptons			T_3	Y_W
Doublets	$(\begin{smallmatrix} u \\ d' \end{smallmatrix})_L$	$(\begin{smallmatrix} c \\ s' \end{smallmatrix})_L$	$(\begin{smallmatrix} t \\ b' \end{smallmatrix})_L$	$(\begin{smallmatrix} 1/2 \\ -1/2 \end{smallmatrix})$	1/3	$(\begin{smallmatrix} \nu_e \\ e \end{smallmatrix})_L$	$(\begin{smallmatrix} \nu_\mu \\ \mu \end{smallmatrix})_L$	$(\begin{smallmatrix} \nu_\tau \\ \tau \end{smallmatrix})_L$	$(\begin{smallmatrix} 1/2 \\ -1/2 \end{smallmatrix})$	-1
Singlets	u_R	c_R	t_R	0	4/3	ν_{eR}	$\nu_{\mu R}$	$\nu_{\tau R}$	0	-2

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

342

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

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

348

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

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

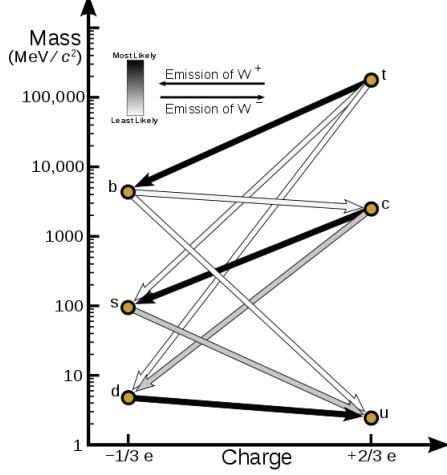


Figure 2.2: Transformations between quarks through the exchange of a WI. Higher generations quarks decay to first generation quarks by emitting a W boson. The arrow color indicates the likelihood of the transition according to the grey scale in the top left side which represent the CKM matrix parameters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

350 The weak decays of quarks are represented in the diagram of figure 2.2; again the
 351 CKM matrix plays a central role since it contains the probabilities for the different
 352 quark decay channels, in particular, note that quark decays are greatly favored be-
 353 tween generation members.

354

355 CKM matrix is a 3×3 unitary matrix parametrized by three mixing angles and
 356 the *CP-mixing phase*; the latter is the parameter responsible for the Charge-Parity
 357 symmetry violation (CP-violation) in the SM. The fact that the b quark decays almost
 358 all the times to a top quark is exploited in this thesis when making the selection of
 359 the signal events by requiring the presence of a jet tagged as a jet coming from a
 360 b quark in the final state. The effect of the *CP-mixing phase* on the cross section of
 361 associated production of Higgs boson and a single top process is also explored in this
 362 thesis.

363 2.2.2 Fundamental interactions

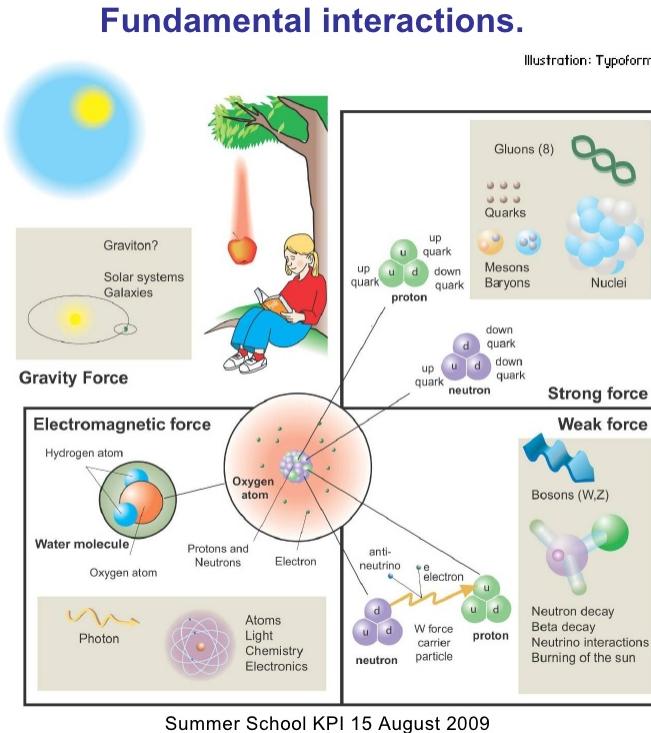


Figure 2.3: Fundamental interactions in nature. Despite the many manifestations of forces in nature, we can track all of them back to one of the fundamental interactions. The most common forces are gravity and electromagnetic given that all of us are subject and experience them in everyday life.

364 Even though there are many manifestations of force in nature, like the ones repre-

365 sented in figure 2.3, we can classify all of them into one of four fundamental interac-

366 tions:

- 367 • *Electromagnetic interaction (EI)* affects particles that are “electrically charged,”
- 368 like electrons and protons. It is described by QED combining quantum mechan-
- 369 ics, special relativity and electromagnetism in order to explain how particles
- 370 with electric charge interact through the exchange of photons, therefore, one
- 371 says that “Electromagnetic Force” is mediated by “photons”. Figure 2.4a. shows

372 a graphical representation, known as “feynman diagram”, of electron-electron
 373 scattering.

- 374 • *Strong interaction (SI)* described by Quantum Chromodynamics (QCD). Hadrons
 375 like proton and neutron have internal structure given that they are composed
 376 of two or more valence quarks⁵. Quarks have fractional electric charge which
 377 means that they are subject to electromagnetic interaction and in the case of the
 378 proton they should break apart due to electrostatic repulsion; however, quarks
 379 are held together inside the hadrons against their electrostatic repulsion by the
 380 “Strong Force” through the exchange of “gluons.” The analog to the electric
 381 charge is the “color charge”. Electrons and photons are elementary particles
 382 as quarks but they don’t carry color charge, therefore they are not subject to
 383 SI. The feynman diagram for gluon exchange between quarks is shown in figure
 384 2.4b.

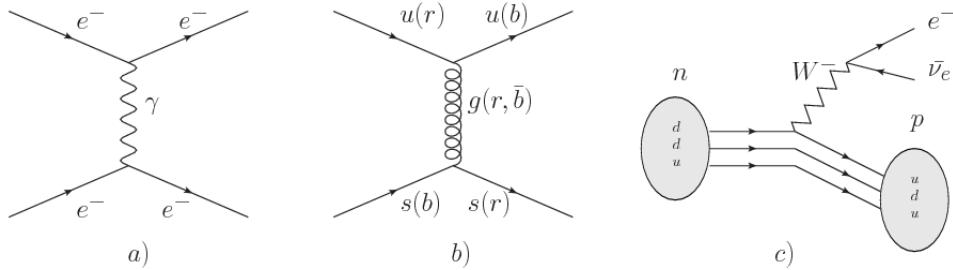


Figure 2.4: Feynman diagrams representing the interactions in SM; a) EI: e-e scattering; b) SI: gluon exchange between quarks ; c) WI: β -decay

- 385 • *Weak interaction (WI)* described by the weak theory (WT), is responsible, for
 386 instance, for the radioactive decay in atoms and proton-proton (pp) fusion
 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

⁵ particles made of four and five quarks are exotic states not so common.

390 mediators of the “weak force” known as “Z” boson in the case of electrically
 391 neutral changes and “ W^\pm ” bosons in the case of electrically charged changes.
 392 The “weak isospin” is the WI analog to electric charge in EI, and color charge
 393 in 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 newtron (n) is transformed
 395 in a proton (p) by emmiting a W^- particle. Since this thesis is in the frame
 396 of the electroweak interaction, a more detailed description of it will be given in
 397 section 2.3

398 • *Gravitational interaction (GI)* described by General Theory of Relativity (GR).
 399 It is responsible for the structure of galaxies and black holes as well as the
 400 expansion of the universe. As a classical theory, in the sense that it can be for-
 401 mulated without even appeal to the concept of quantization, it implies that the
 402 spacetime is a continuum and predictions can be made without limitation to the
 403 precision of the measurement tools. The latter represent a direct contradiction
 404 of the quantum mechanics principles. Gravity is deterministic while quantum
 405 mechanics is probabilistic; despite that, efforts to develop a quantum theory of
 406 gravity have predicted the “graviton” as mediator of the Gravitational force⁶.

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

Table 2.6: Fundamental interactions features [20].

407

⁶ Actually a wide variety of theories have been developed in an attempt to describe gravity; some famous examples are string theory and supergravity.

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

417

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
 423 action at a distance.
- 424 • renormalizability: physical predictions are finite and well defined
- 425 • particle spectrum, symmetries and conservation laws already known must emerge
 426 from the theory.
- 427 • gauge invariance.

428 The gauge invariance requirement reflects the fact that the fundamental fields cannot
 429 be directly measured but associated fields which are the observables. Electric (“E”)
 430 and magnetic (“B”) fields in CED are associated with the electric scalar potential

431 “V” and the vector potential “A”. In particular, \mathbf{E} can be obtained by measuring
 432 the 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 in
 434 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
 436 one particular configuration is known as “gauge freedom”; the transformation law con-
 437 necting two configurations is known as “gauge transformation” and the fact that the
 438 observables are not affected by a gauge transformation is called “gauge invariance”.

439

440 When the gauge transformation:

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} - \Delta f \\ V &\rightarrow V - \frac{\partial f}{\partial t} \end{aligned} \tag{2.4}$$

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

446

447 As mentioned in the first lines of section 2.2, QED has one symmetry group ($U(1)$)
 448 with one group generator (the Q operator) and one gauge field (the electromagnetic
 449 field A^μ). In CED there is not a clear definition, beyond the historical convention, of
 450 which fields are the fundamental and which are the associated, but in QED it is clear
 451 that the fundamental field is A^μ . When a gauge theory is quantized, the gauge field

452 is quantized and its quanta is called “gauge boson”. The word boson characterizes
 453 particles with integer spin which obvey Bose-einstein statistics.

454

455 As will be detailed in section 2.3, interactions between partcles in a system can be
 456 obtained by considering first the Lagrangian density of free particles in the system,
 457 which of course is incomplete because the interaction terms have been left out, and
 458 demanding global phase transformation invariance. Global phase transformation in-
 459 variance means that a gauge transformation is performed identically to every point
 460 in the space⁷ and the Lagrangian remains invariant. Then, the global transformation
 461 is promoted to a local phase transformation (this time the gauge transformation de-
 462 pends on the position in space) and again invariance is required.

463

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

471 This recipe was used to build QED and the theories that aim to explain the funda-
 472 mental interactions.

473 **2.2.3 Gauge bosons**

474 The importance of the gauge bosons comes from the fact that they are the force
 475 mediators or force carriers. The features of the gauge bosons reflect those of the

⁷ Here space corresponds to the 4-dimensional space i.e. space-time.

476 fields they represent and they are extracted from the Lagrangian density used to
477 describe the interactions. In section 2.3, it will be shown how the gauge bosons of the
478 EI and WI emerge from the electroweak Lagrangian. The SI gauge bosons features
479 are also extracted from the SI Lagrangian but it is not detailed in this document. The
480 main features of the SM gauge bosons will be briefly presented below and summarized
481 in table 2.7.

- 482 • **Photon.** EI occurs when the photon couples to (is exchanged between) particles
483 carrying electric charge; however, the photon itself does not carry electric charge,
484 therefore, there is no coupling between photons. Given that the photon is
485 massless the EI is of infinite range, i.e., electrically charged particles interact
486 even if they are located far away one from each other; this also implies that
487 photons always move with the speed of light.
- 488 • **Gluon.** SI is mediated by gluons which, same as photons, are massless. They
489 carry one unit of color charge and one unit of anticolor charge which means that
490 gluons couple to other gluons. As a result, the range of the SI is not infinite
491 but very short due to the attraction between gluons, giving rise to the “color
492 confinement” which explains why color charged particles cannot be isolated but
493 live within composited particles, like quarks inside protons.
- 494 • **W, Z.** The WI mediators, W^\pm and Z, are massive which explains their short-
495 range. Given that the WI is the only interaction that can change the flavor
496 of the interacting particles, the W boson is the responsible for the nuclear
497 transmutation where a neutron is converted in a proton or vice versa with the
498 involvement of an electron and a neutrino (see figure 2.4c). The Z boson is the
499 responsible of the neutral weak processes like neutrino elastic scattering where

500 no electric charge but momentum transference is involved. WI gauge bosons
 501 carry isospin charge which makes possible the interaction between them.

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 Z	± 1 0	No No	± 1 0	80.385 ± 0.015 91.188 ± 0.002

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

502

503 **2.3 Electroweak unification and the Higgs
 504 mechanism**

505 Physicists dream of building a theory that contains all the interactions in one single
 506 interaction, i.e., showing that at some scale in energy all the four fundamental in-
 507 teractions are unified and only one interaction emerges in a “Theory of everything”.
 508 The first sign of the feasibility of such unification comes from success in the con-
 509 struction of the CED. Einstein spent years trying to reach that dream, which by
 510 1920 only involved electromagnetism and gravity, with no success; however, a new
 511 partial unification was achieved in the 1960’s, when S.Glashow [22], A.Salam [23] and
 512 S.Weinberg [24] independently proposed that electromagnetic and weak interactions
 513 are two manifestations of a more general interaction called “electroweak interaction
 514 (EWT)”. Both, QCD and EWT, were developed in parallel and following the useful
 515 prescription provided by QED and the gauge invariance principles.

516

517 The theory of weak interactions was capable of explaining the β -decay and in general
 518 the processes mediated by W^\pm bosons. However, there were some processes like the

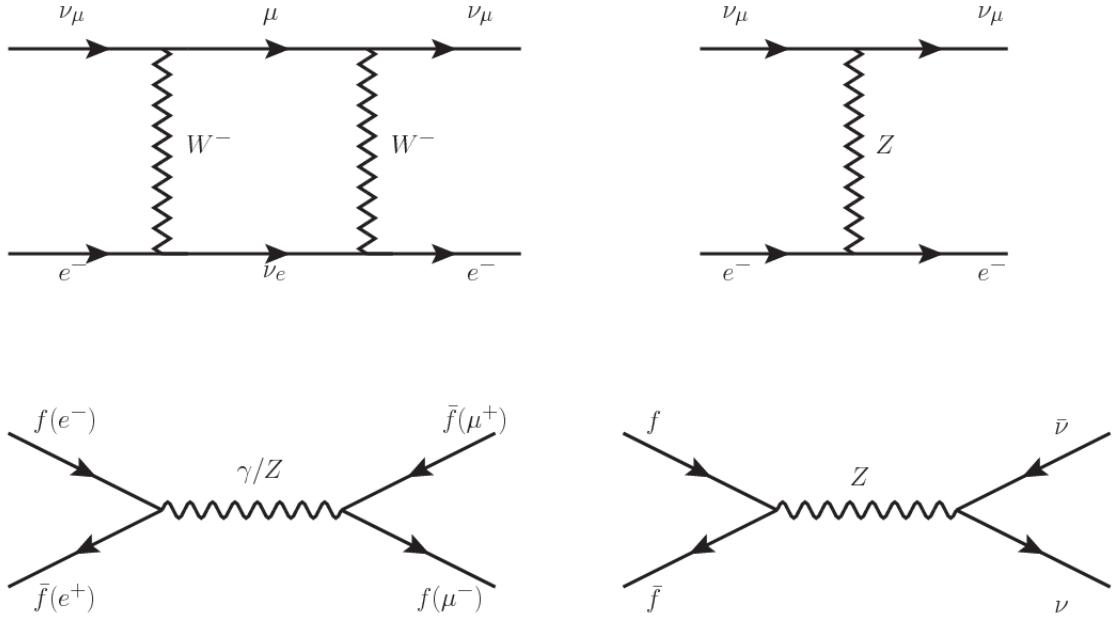


Figure 2.5: Top: $\nu_\mu - e^-$ scattering going through charged currents (left) and neutral currents (right). Bottom: neutral current processes for charged fermions (left) and involving neutrinos (right). While neutral current processes involving only charged fermions can proceed through EI or WI,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

519 “ $\nu_\mu - e$ scattering” which would require the exchange of two W bosons (see figure 2.5
520 top diagrams) giving rise to divergent loop integrals and then non finite predictions.
521 By including neutral currents involving fermions via the exchange of neutral bosons
522 Z , those divergences are compensated and the predictions become realistic.

523
524 Neutral weak interaction vertices conserve flavor in the same way as the electromag-
525 netic vertices do, but additionally, the Z boson can couple to neutrinos which implies
526 that processes involving charged fermions can proceed through EI or WI but processes
527 involving neutrinos can proceed only through WI.

528
529 The prescription to build a gauge theory of the WI consists of proposing a free field
530 Lagrangian density that includes the particles involved; next, by requesting invari-

531 ance under global phase transformations first and generalizing to local phase trans-
 532 formations invariance later, the conserved currents are identified and interactions are
 533 generated by introducing gauge fields. Given that the goal is to include the EI and
 534 WI in a single theory, the group symmetry considered should be a combination of
 535 $SU(2)_L$ and $U(1)_{em}$, however the latter cannot be used directly because the EI treats
 536 left and right-handed particles indistinctly in contrast to the former. Fortunately, the
 537 weak hypercharge, which is a combination of the weak isospin and the electric charge
 538 (eqn 2.2) is suitable to be used since it is conserved by the EI and WI. Thus, the
 539 symmetry group to be considered is

$$G \equiv SU(2)_L \otimes U(1)_Y \quad (2.5)$$

540 The following treatment applies to any of the fermion generations, but for simplicity
 541 the first generation of leptons will be considered [5, 6, 25, 26].

542

543 Given the first generation of leptons

$$\psi_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad \psi_2 = \nu_{eR}, \quad \psi_3 = e_R^- \quad (2.6)$$

544 the charged fermionic currents are given by

$$J_\mu \equiv J_\mu^+ = \bar{\nu}_{eL} \gamma_\mu e_L, \quad J_\mu^\dagger \equiv J_\mu^- = \bar{e}_L \gamma_\mu \nu_{eL} \quad (2.7)$$

545 and the free Lagrangian is given by

$$\mathcal{L}_0 = \sum_{j=1}^3 i\bar{\psi}_j(x) \gamma^\mu \partial_\mu \psi_j(x). \quad (2.8)$$

546 Mass terms are included directly in the QED and QCD free Lagrangians since they

547 preserve the invariance under the symmetry transformations involved which treat
 548 left-handed and right-handed similarly, however mass terms of the form

$$m_W^2 W_\mu^\dagger(x) W^\mu(x) + \frac{1}{2} m_Z^2 Z_\mu(x) Z^\mu(x) - m_e \bar{\psi}_e(x) \psi_e(x) \quad (2.9)$$

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

551

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

$$\begin{aligned} \psi_1(x) &\xrightarrow{G} \psi'_1(x) \equiv U_Y U_L \psi_1(x), \\ \psi_2(x) &\xrightarrow{G} \psi'_2(x) \equiv U_Y \psi_2(x), \\ \psi_3(x) &\xrightarrow{G} \psi'_3(x) \equiv U_Y \psi_3(x) \end{aligned} \quad (2.10)$$

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

$$U_L \equiv \exp\left(i \frac{\sigma_i}{2} \alpha^i\right), \quad U_Y \equiv \exp(i y_i \beta) \quad (i = 1, 2, 3) \quad (2.11)$$

560 with σ_i the Pauli matrices and y_i the weak hypercharges. In order to promote the
 561 transformations from global to local while keeping the invariance, it is required that

562 $\alpha^i = \alpha^i(x)$, $\beta = \beta(x)$ and the replacement of the ordinary derivatives by the covariant
 563 derivatives

$$\begin{aligned} D_\mu \psi_1(x) &\equiv \left[\partial_\mu + ig\sigma_i W_\mu^i(x)/2 + ig'y_1 B_\mu(x) \right] \psi_1(x) \\ D_\mu \psi_2(x) &\equiv \left[\partial_\mu + ig'y_2 B_\mu(x) \right] \psi_2(x) \\ D_\mu \psi_3(x) &\equiv \left[\partial_\mu + ig'y_3 B_\mu(x) \right] \psi_3(x) \end{aligned} \quad (2.12)$$

564 introducing in this way four gauge fields, $W_\mu^i(x)$ and $B_\mu(x)$, in the process. The
 565 covariant derivatives (eqn 2.12) are required to transform in the same way as fermion
 566 fields $\psi_i(x)$ themselves, therefore, the gauge fields transform as:

$$\begin{aligned} B_\mu(x) &\xrightarrow{G} B'_\mu(x) \equiv B_\mu(x) - \frac{1}{g'} \partial_\mu \beta(x) \\ W_\mu^i(x) &\xrightarrow{G} W_\mu^{i\prime}(x) \equiv W_\mu^i(x) - \frac{i}{g} \partial_\mu \alpha_i(x) - \varepsilon_{ijk} \alpha_i(x) W_\mu^j(x). \end{aligned} \quad (2.13)$$

567 The G invariant version of the Lagrangian density 2.8 can be written as

$$\mathcal{L}_0 = \sum_{j=1}^3 i\bar{\psi}_j(x) \gamma^\mu D_\mu \psi_j(x) \quad (2.14)$$

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

$$B_{\mu\nu}(x) \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \quad (2.15)$$

$$W_{\mu\nu}^i(x) \equiv \partial_\mu W_\nu^i(x) - \partial_\nu W_\mu^i(x) - g\varepsilon^{ijk}W_\mu^j W_\nu^k \quad (2.16)$$

571 the last term in eqn. 2.16 is added in order to hold the gauge invariance; therefore,

$$\mathcal{L}_G = -\frac{1}{4}B_{\mu\nu}(x)B^{\mu\nu}(x) - \frac{1}{4}W_{\mu\nu}^i(x)W_i^{\mu\nu}(x) \quad (2.17)$$

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

574

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

$$J_\mu^i(x) = \frac{1}{2}\bar{\psi}_1(x)\gamma_\mu\sigma^i\psi_1(x) \quad (2.18)$$

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

$$J_\mu^Y = \sum_{j=1}^3 \bar{\psi}_j(x)\gamma_\mu y_j\psi_j(x) \quad (2.19)$$

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

$$\mathcal{L}_I = -g J^{i\mu}(x) W_\mu^i(x) - \frac{g'}{2} J^{Y\mu}(x) B_\mu(x) \quad (2.20)$$

583 Note that the weak isospin currents are not the same as the charged fermionic currents
 584 that were used to describe the WI (eqn 2.7), since the weak isospin eigenstates are
 585 not the same as the mass eigenstates, but they are closely related

$$J_\mu = \frac{1}{2}(J_\mu^1 + iJ_\mu^2), \quad J_\mu^\dagger = \frac{1}{2}(J_\mu^1 - iJ_\mu^2). \quad (2.21)$$

586 The same happens with the gauge fields W_μ^i which are related to the mass eigenstates
 587 W^\pm by

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2), \quad W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2). \quad (2.22)$$

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

$$Q = T_3 + \frac{Y_W}{2}. \quad (2.23)$$

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

$$j_\mu^{em} = J_\mu^3 + \frac{1}{2} J_\mu^Y. \quad (2.24)$$

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

$$\begin{aligned} A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \end{aligned} \quad (2.25)$$

601 where θ_W is known as the “Weinberg angle.” The interaction Lagrangian is now given
 602 by

$$\mathcal{L}_I = -\frac{g}{\sqrt{2}}(J^\mu W_\mu^+ + J^{\mu\dagger} W_\mu^-) - \left(g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{J_\mu^Y}{2}\right) A^\mu - \left(g \cos \theta_W J_\mu^3 - g' \sin \theta_W \frac{J_\mu^Y}{2}\right) Z^\mu \quad (2.26)$$

603 the first term is the weak charged current interaction, while the second term is the
 604 electromagnetic interaction under the condition

$$g \sin \theta_W = g' \cos \theta_W = e, \quad \frac{g'}{g} = \tan \theta_W \quad (2.27)$$

605 contained in the eqn.2.24; the third term is the neutral weak current.

606

607 Note that the neutral fields transformation given by the eqn. 2.25 can be written in
 608 terms of the coupling constants g and g' as:

$$A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + g'^2}}, \quad Z_\mu = \frac{g W_\mu^3 - g' B_\mu}{\sqrt{g^2 + g'^2}} \quad (2.28)$$

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

$$\mathcal{L}_{nmEWI} = \mathcal{L}_0 + \mathcal{L}_G \quad (2.29)$$

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

615 2.3.1 Spontaneous symmetry breaking (SSB)

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

618

619 Before reaching the critical force value, the system has rotational symmetry with re-
 620 spect to the nail axis; however, after the critical force value is reached the nail buckles
 621 (top right). The form of the potential energy (bottom right) changes, preserving its
 622 rotational symmetry although its minima does not exhibit that rotational symmetry
 623 any longer. Right before the nail buckles there is no indication of the direction the
 624 nail will bend because any of the directions are equivalent, but once the nail bends,
 625 choosing a direction, an arbitrary minimal energy state (ground state) is selected and
 626 it does not share the system’s rotational symmetry. This mechanism for reaching an
 627 asymmetric ground state is known as “*spontaneous symmetry breaking*”.

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

630

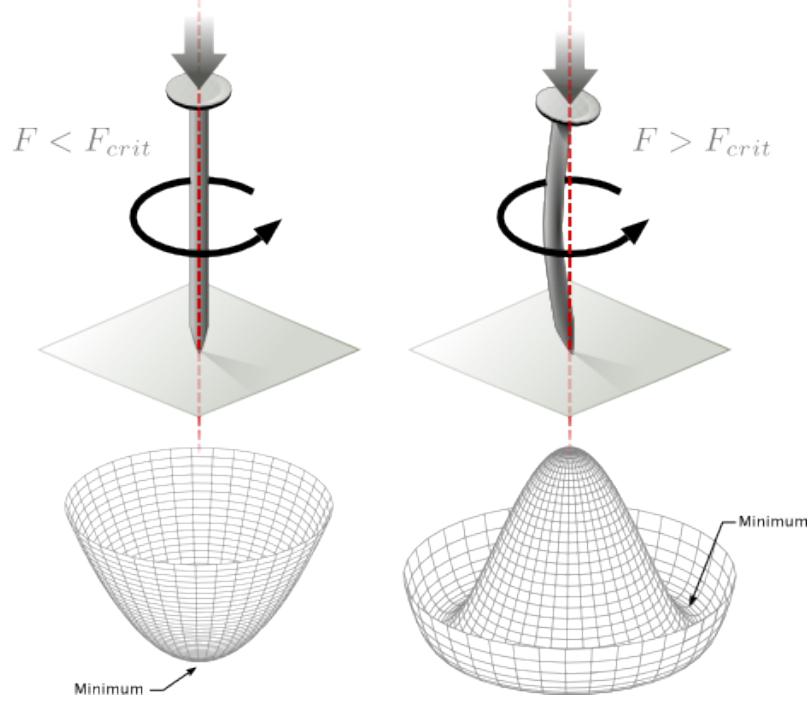


Figure 2.6: Spontaneous symmetry breaking mechanism. The steel nail, subject to an external force (top left), has rotational symmetry with respect to its axis. When the external force overcomes a critical value the nail buckles (top right) choosing a minimal energy state (ground state) and thus “*breaking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

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

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.30)$$

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

635

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

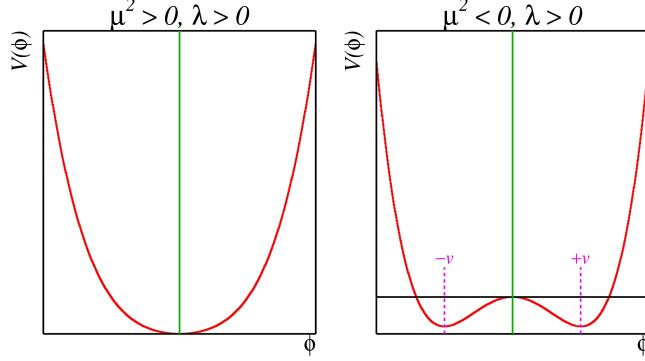


Figure 2.7: Shape of the potential $V(\phi)$ for $\lambda > 0$ and: $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The case $\mu^2 < 0$ corresponds to the potential suitable for introducing the SSB mechanism by choosing one of the two ground states which are connected via reflection symmetry. [27].

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (2.31)$$

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

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.32)$$

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

639

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

$$\phi_0 = \frac{v}{\sqrt{2}} \exp(i\xi) \xrightarrow{\text{SSB}} \phi_0 = \frac{v}{\sqrt{2}} \quad (2.33)$$

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

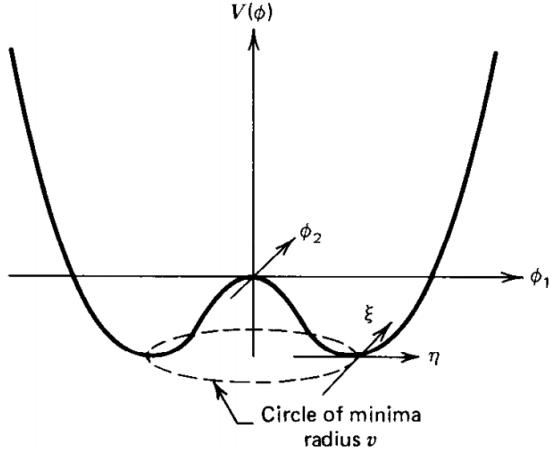


Figure 2.8: Potential for complex scalar field. There is a circle of minima of radius v along the ξ -direction [6].

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

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)) \quad (2.34)$$

645 which when substituted into eqn. 2.32 produces a Lagrangian in terms of the new
646 fields η and ξ

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu^2 \eta^2 - V(\phi_0) - \lambda v \eta (\eta^2 + \xi^2) - \frac{\lambda}{4}(\eta^2 + \xi^2)^2 \quad (2.35)$$

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

652

653 Thus, the SSB mechanism serves as a method to generate mass but as a side effect a

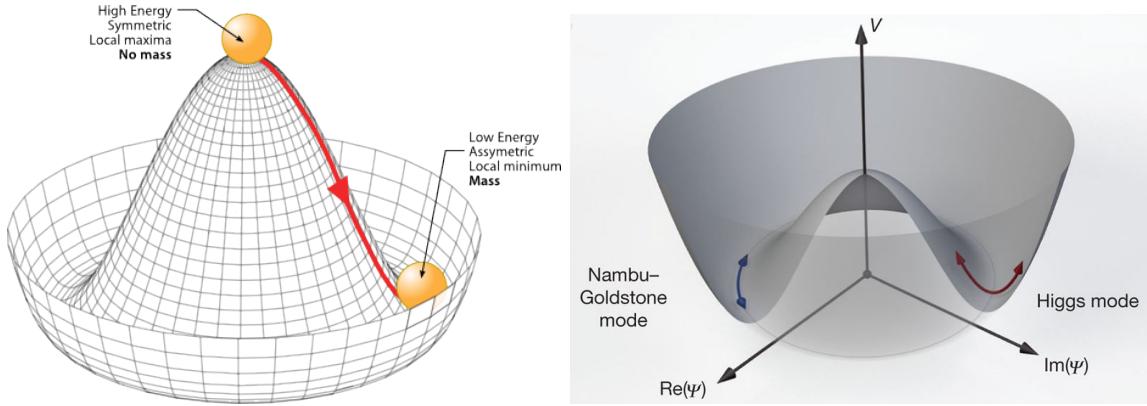


Figure 2.9: SSB mechanism for a complex scalar field [27, 28].

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

663 2.3.2 Higgs mechanism

664 When the SSB mechanism is introduced in the formulation of the EWI in an attempt
 665 to generate the mass of the so far massless gauge bosons and fermions, an interesting
 666 effect is revealed. In order to keep the G symmetry group invariance and generate
 667 the mass of the EW gauge bosons, a G invariant Lagrangian density (\mathcal{L}_S) has to be
 668 added to the non massive EWI Lagrangian (eqn. 2.29)

$$\mathcal{L}_S = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0, \mu^2 < 0 \quad (2.36)$$

$$D_\mu \phi = \left(i\partial_\mu - g \frac{\sigma_i}{2} W_\mu^i - g' \frac{Y}{2} B_\mu \right) \phi \quad (2.37)$$

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

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.38)$$

671 The minima of the potential are defined by

$$\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}. \quad (2.39)$$

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

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 \equiv -\frac{\mu^2}{\lambda}. \quad (2.40)$$

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

679

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

681 The next step is to expand ϕ about the chosen ground state as:

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{i}{v} \sigma_i \theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \approx \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1(x) + i\theta_2(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \quad (2.41)$$

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

$$\phi(x) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (2.42)$$

687 which when substituted into \mathcal{L}_S (eqn. 2.36) results in a Lagrangian containing the now
 688 massive three gauge bosons W^\pm, Z , one massless gauge boson (photon) and the new
 689 Higgs field (H). The three degrees of freedom corresponding to the Nambu-Goldstone
 690 bosons are now integrated into the massive gauge bosons as their longitudinal po-
 691 larizations which were not available when they were massless particles. The effect
 692 by which vector boson fields acquire mass after an spontaneous symmetry breaking,
 693 but without an explicit gauge invariance breaking is known as the “*Higgs mechanism*”.

694

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

701 **2.3.3 Masses of the gauge bosons**

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

$$\left| \left(\partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - i \frac{g'}{2} B_\mu \right) \phi_0 \right|^2 = \left(\frac{1}{2} v g \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (W_\mu^3, B_\mu) \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} \quad (2.43)$$

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

$$M_W = \frac{1}{2} v g. \quad (2.44)$$

The second term in the right side of the eqn.2.43 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. 2.28

$$\begin{aligned} \frac{1}{8} v^2 [g^2 (W_\mu^3)^2 - 2gg' W_\mu^3 B^\mu + g'^2 B_\mu^2] &= \frac{1}{8} v^2 [g W_\mu^3 - g' B_\mu]^2 + 0[g' W_\mu^3 + g B_\mu]^2 \quad (2.45) \\ &= \frac{1}{8} v^2 [\sqrt{g^2 + g'^2} Z_\mu]^2 + 0[\sqrt{g^2 + g'^2} A_\mu]^2 \end{aligned}$$

705 and then

$$M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}, \quad M_A = 0 \quad (2.46)$$

706 **2.3.4 Masses of the fermions**

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

$$\mathcal{L}_{Yl} = -G_l \left[(\bar{\nu}_l, \bar{l})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} l_R + \bar{l}_R (\phi^-, \bar{\phi}^0) \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L \right], \quad l = e, \mu, \tau. \quad (2.47)$$

709 After the SSB and replacing the usual field expansion about the ground state (eqn.2.40)
 710 into \mathcal{L}_{Yl} , the mass term arises

$$\mathcal{L}_{Yl} = -m_l (\bar{l}_L l_R + \bar{l}_R l_L) - \frac{m_l}{v} (\bar{l}_L l_R + \bar{l}_R l_L) H = -m_l \bar{l} l \left(1 + \frac{H}{v} \right) \quad (2.48)$$

711

$$m_l = \frac{G_l}{\sqrt{2}} v \quad (2.49)$$

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

$$\phi_c = -i\sigma_2 \phi^* = \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix}. \quad (2.50)$$

715 Additionally, given that the quark isospin doublets are not constructed in terms of
 716 the mass eigenstates but in terms of the flavor eigenstates, as shown in table2.5, the
 717 coupling parameters will be related to the CKM matrix elements; thus the quark
 718 Lagrangian is given by:

$$\mathcal{L}_{Yq} = -G_d^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{jR} - G_u^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} u_{jR} + h.c. \quad (2.51)$$

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

720 of \mathcal{L}_{Yq} is:

$$\mathcal{L}_{Yq} = -m_d^i \bar{d}_i d_i \left(1 + \frac{H}{v}\right) - m_u^i \bar{u}_i u_i \left(1 + \frac{H}{v}\right) \quad (2.52)$$

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

723 2.3.5 The Higgs field

724 After the characterization of the fermions and gauge bosons as well as their interac-
725 tions, it is necessary to characterize the Higgs field itself. The Lagrangian \mathcal{L}_S in eqn.
726 2.36 written in terms of the gauge bosons is given by

$$\mathcal{L}_S = \frac{1}{4} \lambda v^4 + \mathcal{L}_H + \mathcal{L}_{HV} \quad (2.53)$$

727

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} m_H^2 H^2 - \frac{1}{2v} m_H^2 H^3 - \frac{1}{8v^2} m_H^2 H^4 \quad (2.54)$$

728

$$\mathcal{L}_{HV} = m_H^2 W_\mu^+ W^{\mu-} \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) \quad (2.55)$$

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

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v \quad (2.56)$$

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

738 The Higgs mass measurement, reported by both experiments [34], is in table 2.8.

Property	Value
Electric charge	0
Colour charge	0
Spin	0
Weak isospin	-1/2
Weak hypercharge	1
Parity	1
Mass (GeV/c^2)	$125.09 \pm 0.21 \text{ (stat.)} \pm 0.11 \text{ (syst.)}$

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

739

740 2.3.6 Production of Higgs bosons at LHC

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

748 Protons are composed of quarks and these quarks are bound by gluons; however,
 749 what is commonly called the quark content of the proton makes reference to the
 750 valence quarks. A sea of quarks and gluons is also present inside the proton as repre-
 751 sented in figure 2.10. In a proton-proton (pp) collision, the constituents (quarks and
 752 gluons) are those who collide. The pp cross section depends on the momentum of
 753 the colliding particles, reason for which it is needed to know how the momentum is
 754 distributed inside the proton. Quarks and gluons are known as partons and the func-
 755 tions that describe how the proton momentum is distributed among partons inside it

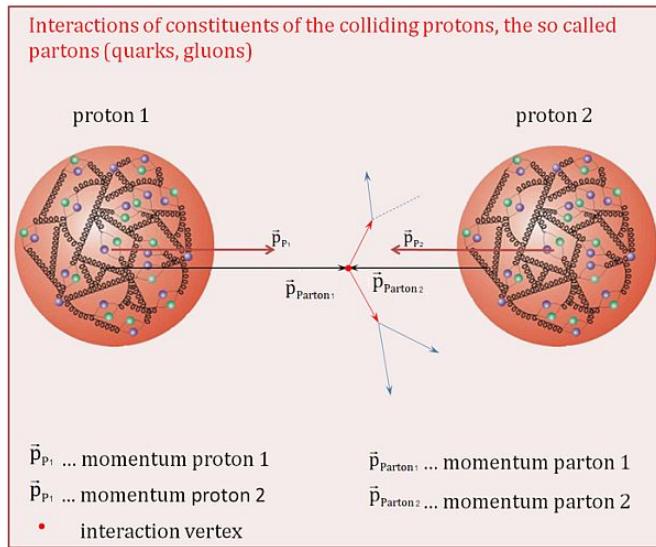


Figure 2.10: Proton-proton collision. Protons are composed of 3 valence quarks, a sea of quarks and gluons; therefore in a proton-proton collision, quarks and gluons are those who collide. [35].

756 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.

758

In addition, in physics, a common approach to study complex systems consists in starting with a simpler version of them, for which a well known description is available, and add an additional “perturbation” which represents a small deviation from the known behavior. If the perturbation is small enough, the physical quantities associated with the perturbed system are expressed as a series of corrections to those of the simpler system; therefore, the more terms are considered in the series (the higher order in the perturbation series), the more precise is the description of the complex system.

767

768 This thesis explores the Higgs production at LHC; therefore the overview presented
769 here will be oriented specifically to the production mechanisms after pp collisions at

770 LHC.

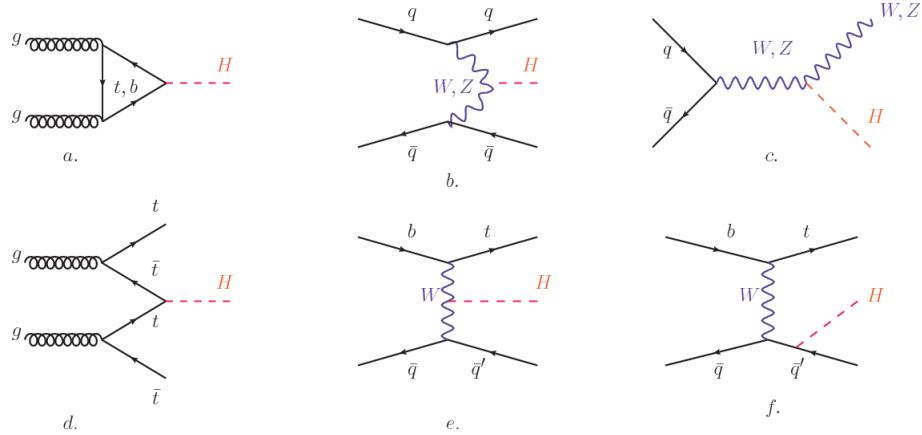


Figure 2.11: Main Higgs boson production mechanism Feynman diagrams. a. gluon-gluon fusion, b. vector boson fusion (VBF), c. Higgs-strahlung, d. Associated production with a top or bottom quark pair, e-f. associated production with a single top quark.

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

777 As shown in eqns 2.47, 2.51 and 2.55, the strength of the Higgs-fermion interaction
 778 is proportional to the fermion mass while the strength of the Higgs-gauge boson
 779 interaction is proportional to the square of the gauge boson mass, which implies
 780 that the Higgs production and decay mechanisms are dominated by couplings $H -$
 781 (W, Z, t, b, τ) .

782 The main production mechanism is the gluon fusion (figure 2.11a and $pp \rightarrow H$ in figure
 783 2.12) given that gluons carry the highest fraction of momentum of the protons in pp
 784 colliders. Since the Higgs boson does not couple to gluons, the mechanism proceeds
 785 through the exchange of a virtual top-quark loop given that for it the coupling is

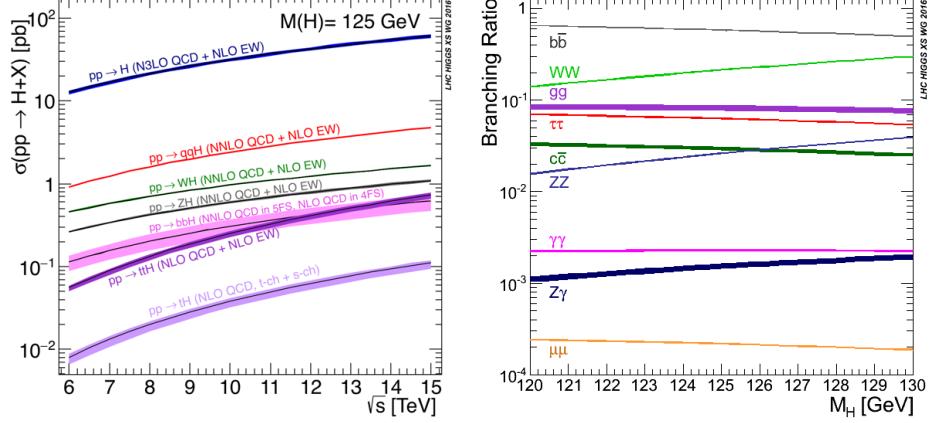


Figure 2.12: Higgs boson production cross sections (left) and decay branching ratios (right) for the main mechanisms. The VBF is indicated as qqH [36].

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 when analyzing the datasets provided by the experiments. More details about how to identify events of interest in an analysis will be given in chapter 6.
 The next production mechanism is Higgs-strahlung (figure 2.11c and $pp \rightarrow WH, pp \rightarrow ZH$ in figure 2.12) where two fermions annihilate to form a weak gauge boson. If the initial fermions have enough energy, the emergent boson eventually will emit a Higgs boson.
 The associated production with a top or bottom quark pair and the associated production with a single top quark (figure 2.11d-f and $pp \rightarrow bbH, pp \rightarrow t\bar{t}H, pp \rightarrow tH$ in figure 2.12) have a smaller cross section than the main three mechanisms above,

802 but they provide a good opportunity to test the Higgs-top coupling. The analysis
 803 reported in this thesis is developed using these production mechanisms. A detailed
 804 description of the tH mechanism will be given in section 2.4.

805 **2.3.7 Higgs boson decay channels**

806 When a particle can decay through several modes, also known as channels, the
 807 probability of decaying through a given channel is quantified by the “branching ratio
 808 (BR)” of the decay channel; thus, the BR is defined as the ratio of number of decays
 809 going through that given channel to the total number of decays. In regard to the
 810 Higgs boson decay, the BR can be predicted with accuracy once the Higgs mass is
 811 known [37, 38]. In figure 2.12 right, a plot of the BR as a function of the Higgs mass
 812 is presented. The largest predicted BR corresponds to the $b\bar{b}$ pair decay channel (see
 813 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 $m_H = 125 GeV/c^2$. [9]

815 **2.4 Associated production of a Higgs boson and a
816 single Top quark.**

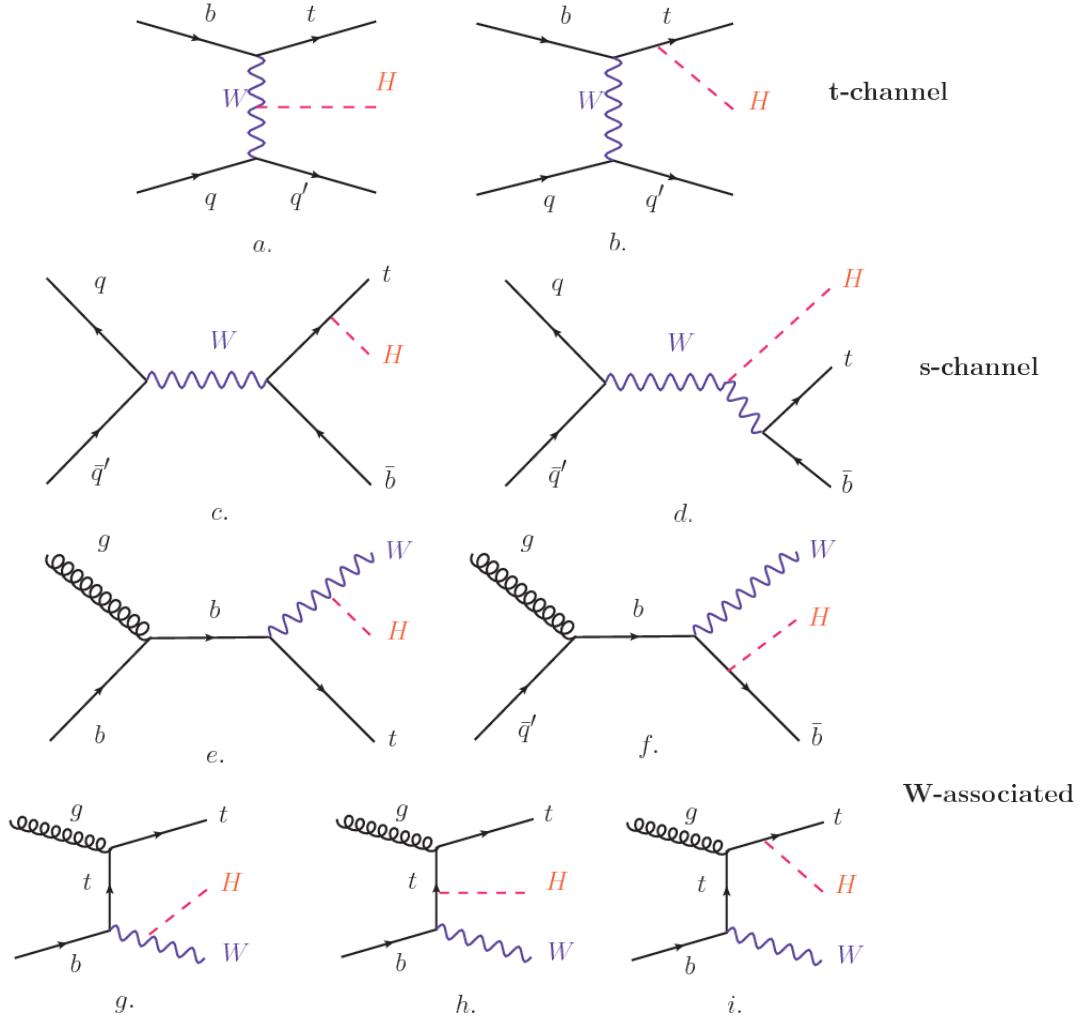


Figure 2.13: Associated Higgs boson production mechanism Feynman diagrams. a.,b. t-channel (tHq), c.,d. s-channel (tHb), e-i. W-associated.

817 Associated production of Higgs boson has been extensively studied [39–43]. While
818 measurements of the main Higgs production mechanisms rates are sensitive to the
819 strength of the Higgs coupling to W boson or top quark, they are not sensitive to the
820 relative sign between the two couplings. In this thesis, the Higgs boson production

821 mechanism explored is the associated production with a single top quark (*th*) which
 822 offers sensitiveness to the relative sign of the Higgs couplings to W boson and to top
 823 quark. The description given here is based on the reference [41]

824

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

833

834 The *th* production, where Higgs boson can be radiated either from the top quark or
 835 from the W boson, is represented by the leading order Feynman diagrams in figure
 836 2.13. The cross section for the *th* process is calculated, as usual, summing over
 837 the contributions from the different feynman diagrams; therefore it depends on the
 838 interference between the contributions. In the SM, the interference for t-channel (tHq
 839 process) and W-associated (tHW process) production is destructive [39] resulting in
 840 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].

841

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

$$\mathcal{A} = \frac{g}{\sqrt{2}} \left[(\kappa_t - \kappa_V) \frac{m_t \sqrt{s}}{m_W v} A \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) + \left(\kappa_V \frac{2m_W s}{v} \frac{1}{t} + (2\kappa_t - \kappa_V) \frac{m_t^2}{m_W v} \right) B \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) \right], \quad (2.57)$$

851 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-
 852 tify possible deviations of the couplings, Higgs-Vector boson (H-W) and Higgs-top
 853 (H-t) respectively, from the SM couplings; $s = (p_W + p_b)^2$, $t = (p_W - p_H)^2$, φ is the
 854 Higgs azimuthal angle around the z axis taken parallel to the direction of motion of
 855 the incoming W; A and B are funtions describing the weak interaction in terms of
 856 the chiral states of the quarks b and t . Terms that vanish in the high energy limit
 857 have been neglected as well as the Higgs and b quark masses⁸.

858

859 The scattering amplitude grows with energy like \sqrt{s} for $\kappa_V \neq \kappa_t$, in contrast to
 860 the SM ($\kappa_t = \kappa_V = 1$), where the first term in 2.57 cancels out and the amplitude
 861 is constant for large s ; therefore, a deviation from the SM predictions represents an
 862 enhancement in the tHq cross section. In particular, for a SM H-W coupling and a H-t

⁸ A detailed explanation of the structure and approximations used to derive \mathcal{A} can be found in reference [41]

863 coupling of inverted sign with respect to the SM ($\kappa_V = -\kappa_t = 1$) the tHq cross section
 864 is enhanced by a factor greater 10 as seen in the figure 2.14 taken from reference [41];
 865 reference [47] has reported similar enhancement results.

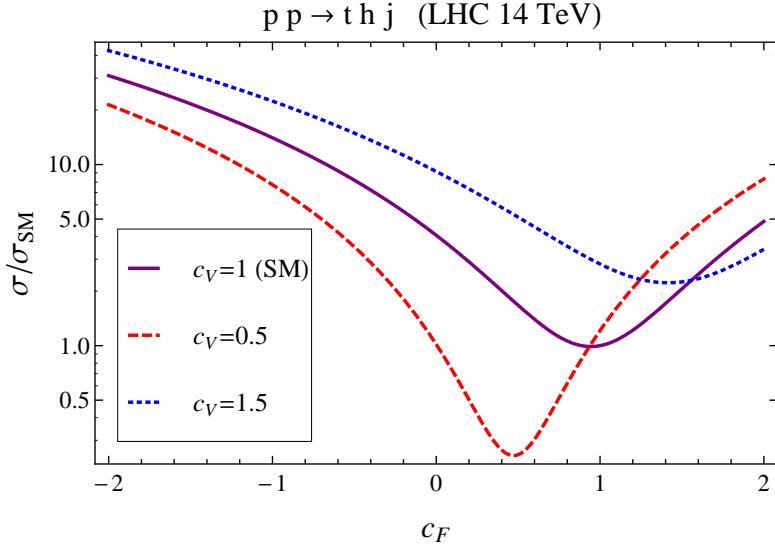


Figure 2.14: Cross section for tHq process as a function of κ_t , normalized to the SM, for three values of κ_V . In the plot c_f refers to the Higgs-fermion coupling which is dominated by the H-t coupling and represented here by κ_t . Solid, dashed and dotted lines correspond to $c_V \rightarrow \kappa_V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

866 A similar analysis is valid for the W-associated channel but, in that case, the inter-
 867 ference is more complicated since there are more than two contributions and an ad-
 868 ditional interference with the production of Higgs boson and a top pair process($t\bar{t}H$).
 869 The calculations are made using the so-called Diagram Removal (DR) technique where
 870 interfering diagrams are removed (or added) from the calculations in order to evaluate
 871 the impact of the removed contributions. DR1 was defined to neglect $t\bar{t}H$ interference
 872 while DR2 was defined to take $t\bar{t}H$ interference into account [48]. As shown in figure
 873 2.15, the tHW cross section is enhanced from about 15 fb (SM: $\kappa_{Htt} = 1$) to about
 874 150 fb ($\kappa_{Htt} = -1$). Differences between curves for DR1 and DR2 help to gauge the
 875 impact of the interference with $t\bar{t}H$.

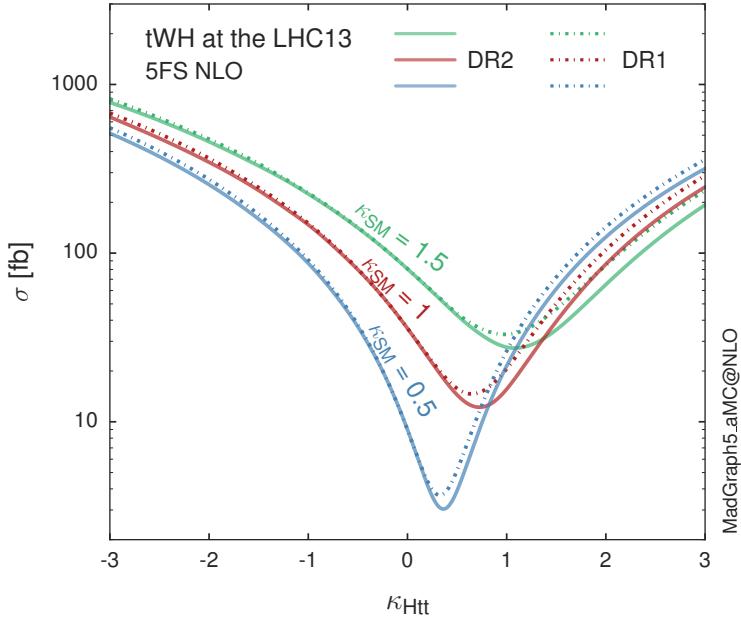


Figure 2.15: Cross section for tHW process as a function of κ_{Htt} , for three values of κ_{SM} at $\sqrt{s} = 13$ TeV. $\kappa_{Htt}^2 = \sigma_{Htt}/\sigma_{Htt}^{SM}$ is a simple rescaling of the SM Higgs interactions.

876 Results of the calculations of the tHq and tHW cross sections at $\sqrt{s} = 13$ TeV can be
 877 found in reference [49] and a summary of the results is presented in table 2.11.

878

879 2.5 The CP-mixing in tH processes

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

884

885 In this thesis, the sensitivity of th processes to CP-mixing is also studied in the
 886 effective field theory framework and based in references [43, 48]; a generic particle

	\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 flippling in the sign of the H-t coupling with respect to the SM one.

887 (X_0) of spin-0 and a general CP violating interaction with the top quark, can couple
 888 to scalar and pseudoscalar fermionic densities. The H-W interaction is assumed to
 889 be SM-like. The Lagrangian modeling the H-t interaction is given by

$$\mathcal{L}_0^t = -\bar{\psi}_t (c_\alpha \kappa_{Htt} g_{Htt} + i s_\alpha \kappa_{Att} g_{Att} \gamma_5) \psi_t X_0, \quad (2.58)$$

890 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
 891 dimensionless rescaling parameters⁹, $g_{Htt} = g_{Att} = m_t/v = y_t/\sqrt{2}$ and $v \sim 246$ GeV is
 892 the Higgs vacuum expectation value. In this parametrization, it is easy to recover
 893 three special cases

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

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

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

⁹ analog to κ_t and κ_V

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

$$\mathcal{L}_0^g = -\frac{1}{4} \left(c_\alpha \kappa_{Hgg} g_{Hgg} G_{\mu\nu}^a G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} \right) X_0. \quad (2.59)$$

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

$$\frac{\sigma_{NLO}^{gg \rightarrow X_0}}{\sigma_{NLO,SM}^{gg \rightarrow H}} = c_\alpha^2 \kappa_{Htt}^2 + s_\alpha^2 \left(\kappa_{Att} \frac{g_{Agg}}{g_{Hgg}} \right)^2. \quad (2.60)$$

903 If the rescaling parameters are set to

$$\kappa_{Htt} = 1, \quad \kappa_{Att} = \left| \frac{g_{Hgg}}{g_{Agg}} \right| = \frac{2}{3}. \quad (2.61)$$

904 the gluon-fusion SM cross section is reproduced for every value of the CP-mixing
 905 angle α ; therefore, by imposing that condition to the Lagrangian density 2.58, the
 906 CP-mixing angle is not constrained by current data. Figure 2.16 shows the NLO cross
 907 sections for t-channel tX_0 (blue) and $t\bar{t}X_0$ (red) associated production processes as
 908 a function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark
 909 CP-violating coupling. Rescaling factors κ_{Htt} and κ_{Att} have been set to reproduce
 910 the SM gluon-fusion cross sections.

911 It is interesting to notice that the tX_0 cross section is enhanced, by a factor of
 912 about 10, when a continuous rotation in the scalar-pseudoscalar plane is applied; this
 913 enhancement is similar to the enhancement produced when the H-t coupling is flipped
 914 in sign with respect to the SM ($y_t = -y_{t,SM}$ in the plot), as showed in section 2.4. In

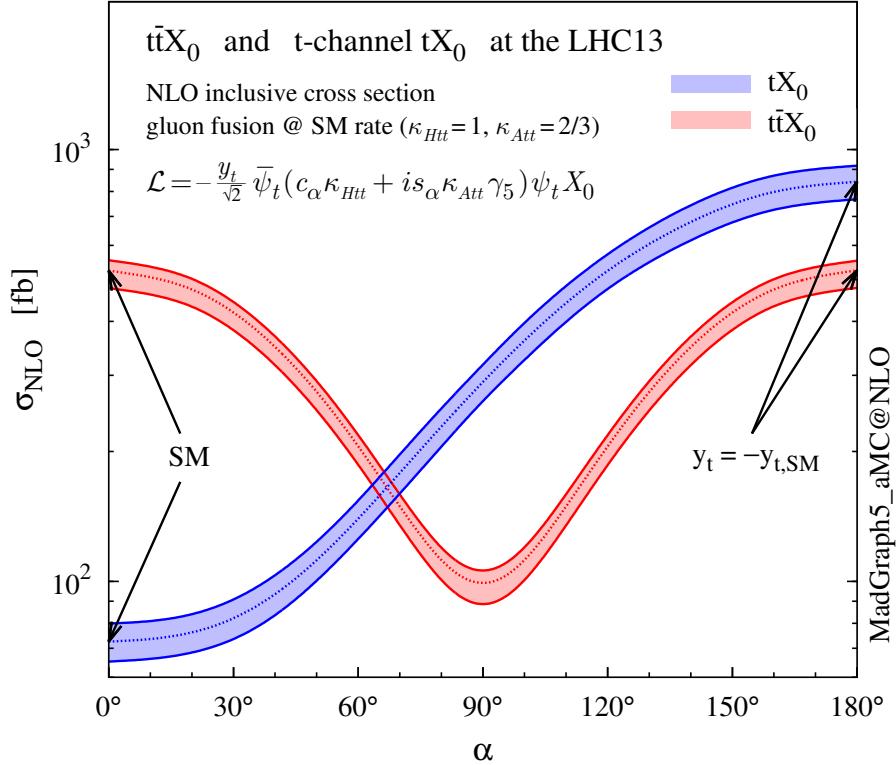


Figure 2.16: NLO cross sections for t-channel tX_0 (blue) and $t\bar{t}X_0$ (red) associated production processes as a function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark CP-violating coupling [43].

915 contrast, the degeneracy in the $t\bar{t}X_0$ cross section is still present given that it depends
 916 quadratically on the H-t coupling, but more interesting is to notice that $t\bar{t}X_0$ cross
 917 section is exceeded by tX_0 cross section after $\alpha \sim 60^\circ$.

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

922

923 Figure 2.17 shows the NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red) asso-
 924 ciated production and for the combined $tWX_0 + t\bar{t}X_0 +$ interference (orange) as a

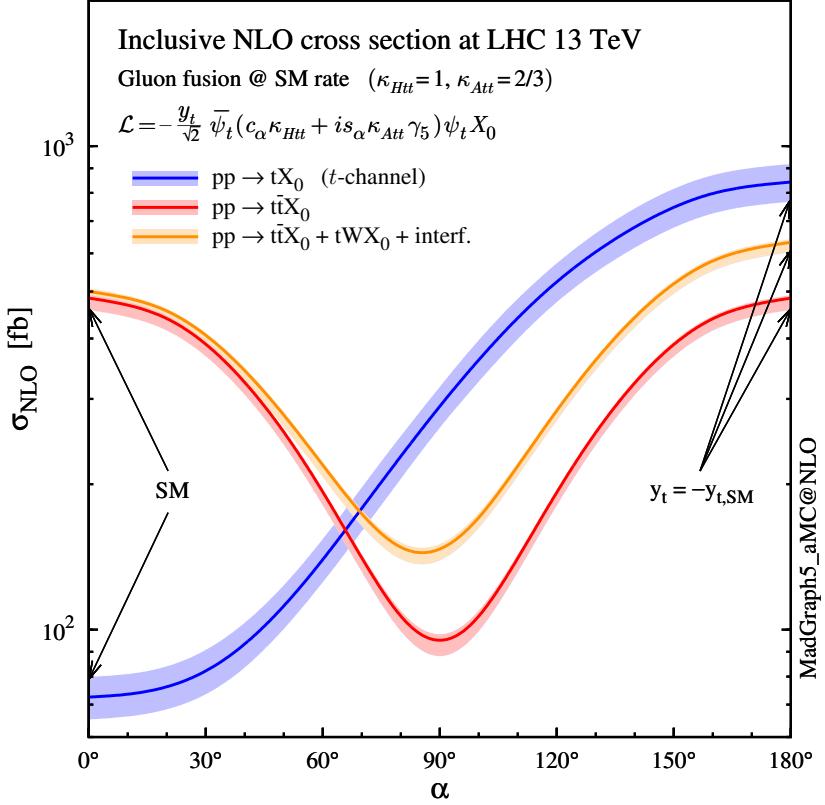


Figure 2.17: NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red) associated production processes and combined $tWX_0 + t\bar{t}X_0$ (including interference) production as a function of the CP-mixing angle α [43].

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.

931 **2.6 Experimantal status of the anomalous**
 932 **Higg-fermion coupling.**

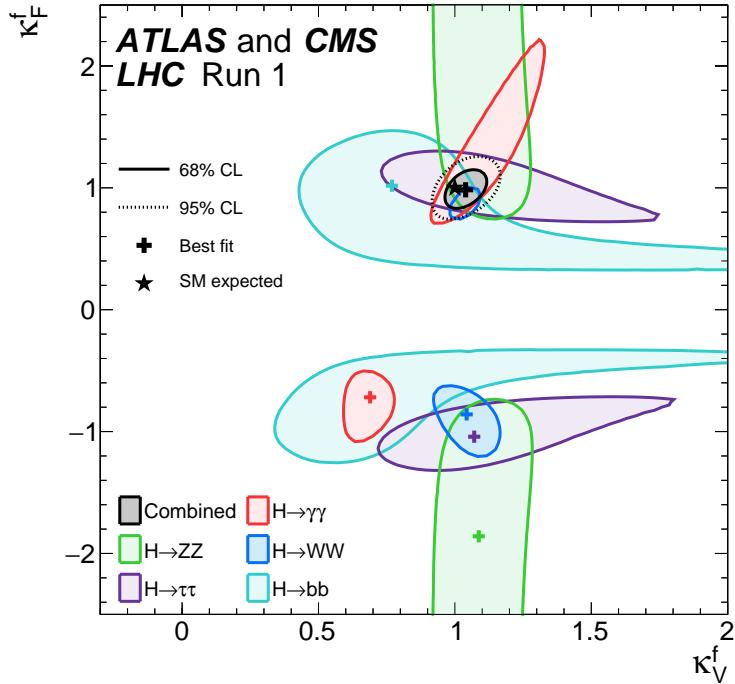


Figure 2.18: Combination of the ATLAS and CMS fits for coupling modifiers κ_t - κ_V ; also shown the individual decay channels combination and their global combination. No assumptions have been made on the sign of the coupling modifiers [53].

933 ATLAS and CMS have performed analysis of the anomalous H-f coupling by making
 934 likelihood scans for the two coupling modifiers, κ_t and κ_V , under the assumption that
 935 $\kappa_Z = \kappa_W = \kappa_V$ and $\kappa_t = \kappa_\tau = \kappa_b = \kappa_f$. Figure 2.18 shows the result of the combination
 936 of ATLAS and CMS fits; also the individual decay channels combination and the
 937 global combination results are shown.
 938 While all the channels are compatible for positive values of the modifiers, for negative
 939 values of κ_t there is no compatibility. The best fit for individual channels is compatible
 940 with negative values of κ_t except for the $H \rightarrow bb$ channel which is expected to be the

941 most sensitive channel; therefore, the best fit for the global fit yields $\kappa_t \geq 0$. Thus,
942 the anomalous H-t coupling cannot be excluded completely.

943 Chapter 3

944 The CMS experiment at the LHC

945 3.1 Introduction

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

959 3.2 The LHC

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

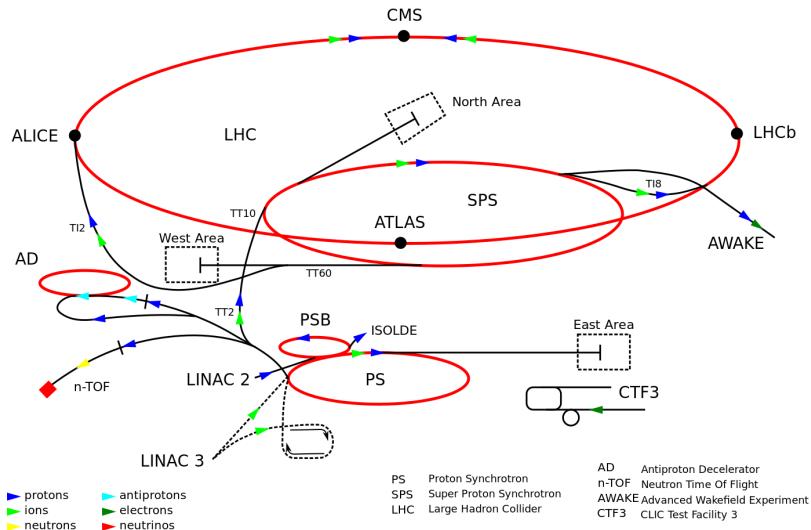


Figure 3.1: CERN accelerator complex. Blue arrows show the path followed by protons along the acceleration process [54].

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

- 967 • Proton-Proton collisions (pp) for multiple physics experiments.
- 968 • Lead-Lead collisions ($Pb-Pb$) for heavy ion experiments.
- 969 • Proton-Lead collisions ($p-Pb$) for quark-gluon plasma experiments.

970 In this thesis only pp collisions will be considered.

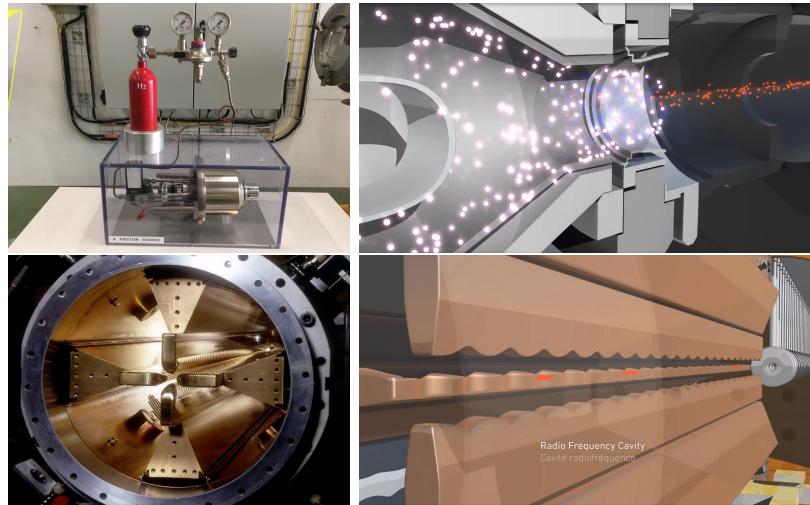


Figure 3.2: LHC protons source and the first acceleration stage. Top: the bottle contains hydrogen gas (white dots) which is injected into the metal cylinder to be broken down into electrons(blue dots) and protons(red dots); Bottom: the obtained protons are directed towards the radio frequency quadrupole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

972 Collection of protons starts with hydrogen atoms taken from a bottle, containing hy-
 973 drogen gas, and injecting them in a metal cylinder; hydrogen atoms are broken down
 974 into electrons and protons by an intense electric field (see figure3.2 top). The result-
 975 ing protons leave the metal cylinder towards a radio frequency quadrupole (RFQ)
 976 that focus the beam, accelerates the protons and creates the packets of protons called
 977 bunches. In the RFQ, an electric field is generated by a RF wave at a frequency that
 978 matches the resonance frequency of the cavity where the electrodes are contained.
 979 The beam of protons traveling on the RFQ axis experiences an alternating electric
 980 field gradient that generates the focusing forces.

981

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

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

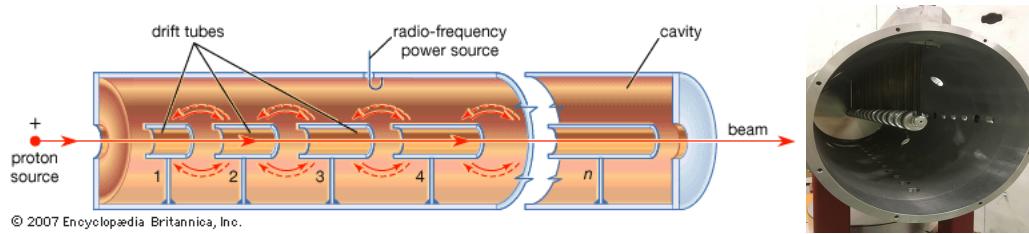


Figure 3.3: The LINAC2 accelerating system at CERN. Electric fields generated by radio frequency (RF) create acceleration and deceleration zones inside the cavity; deceleration zones are blocked by drift tubes where quadrupole magnets focus the proton beam. [60]

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

997

998 The beam coming from LINAC2 is injected into the proton synchrotron booster
 999 (PSB) to reach 1.4 GeV in energy. The next acceleration is provided at the pro-
 1000 ton synchrotron (PS) up to 26 GeV, followed by the injection into the super proton
 1001 synchrotron (SPS) where protons are accelerated to 450 GeV. Finally, protons are
 1002 injected into the LHC where they are accelerated to the target energy of 6.5 TeV.
 1003 PSB, PS, SPS and LHC accelerate protons using the same RF acceleration technique
 1004 described before.

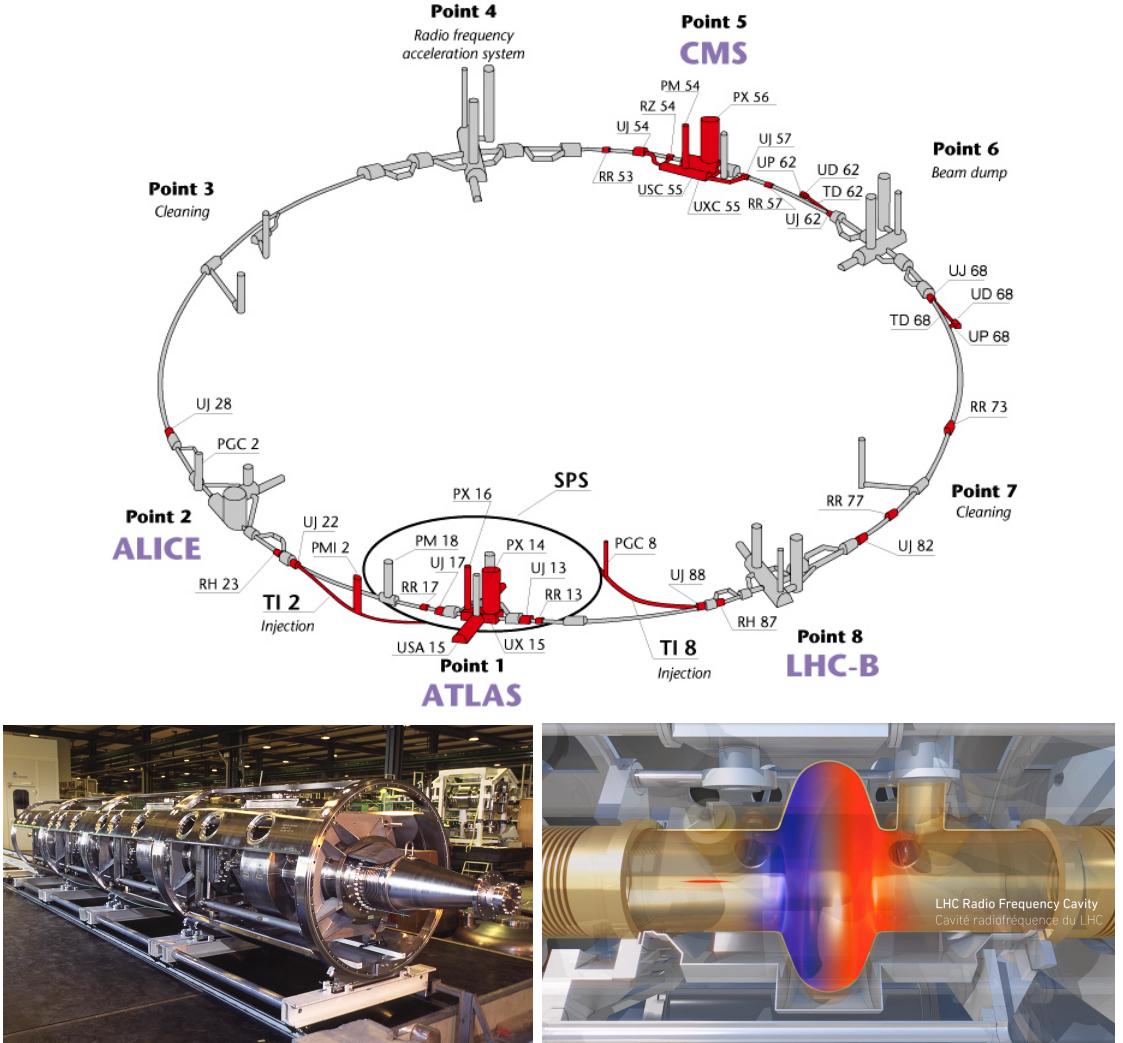


Figure 3.4: Top: LHC layout. The red zones indicate the infrastructure additions to the LEP installations, built to accommodate the ATLAS and CMS experiments which exceed the size of the former experiments located there [55]. Bottom: LHC RF cavities. A module accommodates 4 cavities that accelerate protons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

1005 LHC has a system of 16 RF cavities located in the so-called point 4, as shown in
 1006 figure 3.4 top, tuned at a frequency of 400 MHz and the protons are carefully timed,
 1007 so in addition to the acceleration effect the bunch structure of the beam is preserved.
 1008 Bottom side of figure 3.4 shows a picture of a RF module composed of 4 RF cavities
 1009 working in a superconducting state at 4.5 K; also is showed a representation of the

1010 accelerating electric field that accelerates the protons in the bunch.

1011

1012 While protons are accelerated in one section of the LHC ring, where the RF cavities
 1013 are located, in the rest of their path they have to be kept in the curved trajectory
 1014 defined by the LHC ring. Technically, LHC is not a perfect circle; RF, injection, beam
 1015 dumping, beam cleaning and sections before and after the experimental points where
 1016 protons collide are all straight sections. In total, there are 8 arcs 2.45 Km long each
 1017 and 8 straight sections 545 m long each. In order to curve the proton's trajectory in
 1018 the arc sections, superconducting dipole magnets are used.

1019

1020 Inside the LHC ring, there are two proton beams traveling in opposite directions in
 1021 two separated beam pipes; the beam pipes are kept at ultra-high vacuum ($\sim 10^{-9}$
 1022 Pa) to ensure that there are no particles that interact with the proton beams. The
 1023 superconducting dipole magnets used in LHC are made of a NbTi alloy, capable of
 1024 transporting currents of about 12000 A when cooled at a temperature below 2K using
 1025 liquid helium (see figure 3.5).

1026

1027 Protons in the arc sections of LHC feel a centripetal force exerted by the dipole
 1028 magnets which is perpendicular to the beam trajectory; The magnitude of magnetic
 1029 field needed can be found assuming that protons travel at $v \approx c$, using the standard
 1030 values for proton mass and charge and the LHC radius, as

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

1031 which is about 100000 times the Earth's magnetic field. A representation of the mag-
 1032 netic field generated by the dipole magnets is shown on the bottom left side of figure

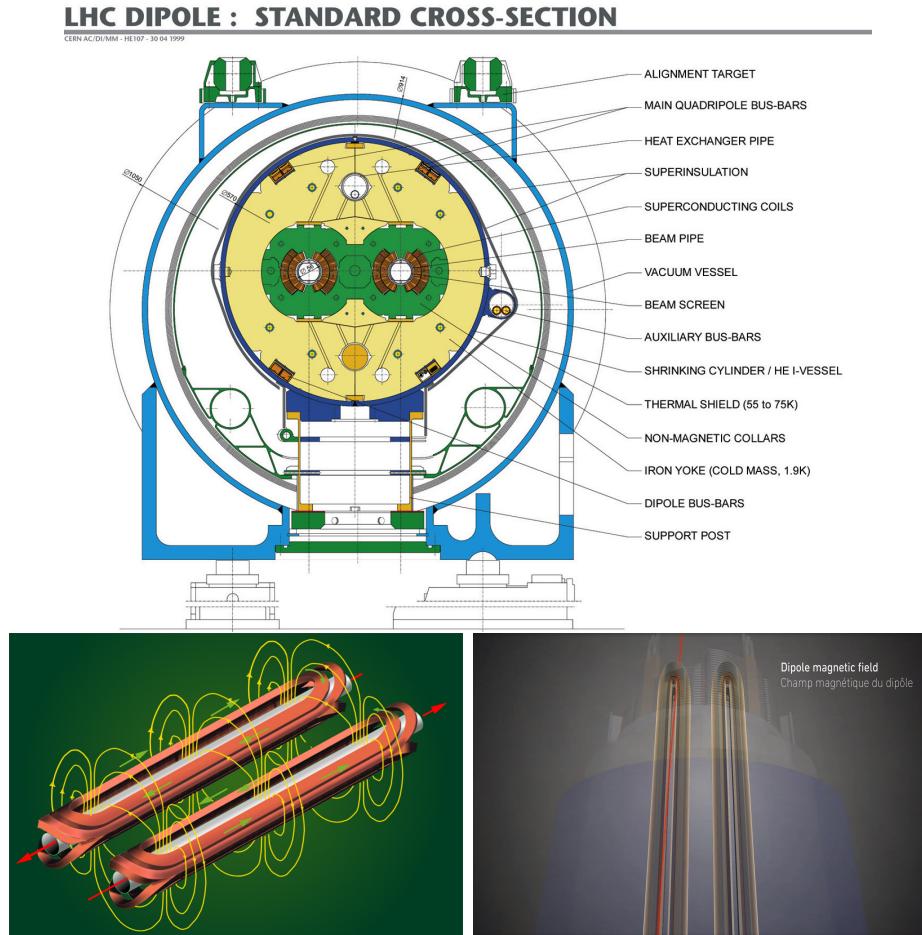


Figure 3.5: Top: LHC dipole magnet transverse view; cooling, shielding and mechanical support are indicated. Bottom left: Magnetic field generated by the dipole magnets; note that the direction of the field inside one beam pipe is opposite with respect to the other beam pipe which guarantee that both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

1033 3.5. The bending effect of the magnetic field on the proton beam is shown on the
 1034 bottom right side of figure 3.5. Note that the dipole magnets are not curved; the arc
 1035 section of the LHC ring is composed of straight dipole magnets of about 15 m. In
 1036 total there are 1232 dipole magnets along the LHC ring.

1037

1038 In addition to bending the beam trajectory, the beam has to be focused so it stays

1039 inside the beam pipe. The focusing is performed by quadrupole magnets installed in
 1040 a different straight section; in total 858 quadrupole magnets are installed along the
 1041 LHC ring. Other effects like electromagnetic interaction among bunches, interaction
 1042 with electron clouds from the beam pipe, the gravitational force on the protons, dif-
 1043 ferences in energy among protons in the same bunch, among others, are corrected
 1044 using sextupole and other magnetic multipoles.

1045

1046 The two proton beams inside the LHC ring are made of bunches with a cylindrical
 1047 shape of about 7.5 cm long and about 1 mm in diameter; when bunches are close
 1048 to the collision point (CP), the beam is focused up to a diameter of about 16 μm in
 1049 order to maximize the number of collisions per unit area and per second, known as
 1050 luminosity (L). Luminosity can be calculated using

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

1051 where f is the revolution frequency, n is the number of bunches per beam, N_1 and
 1052 N_2 are the numbers of protons per bunch (1.5×10^{11}), σ_x and σ_y are the gaussian
 1053 transverse sizes of the bunches. The expected luminosity is about

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

$$n = 2808$$

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

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

1054

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

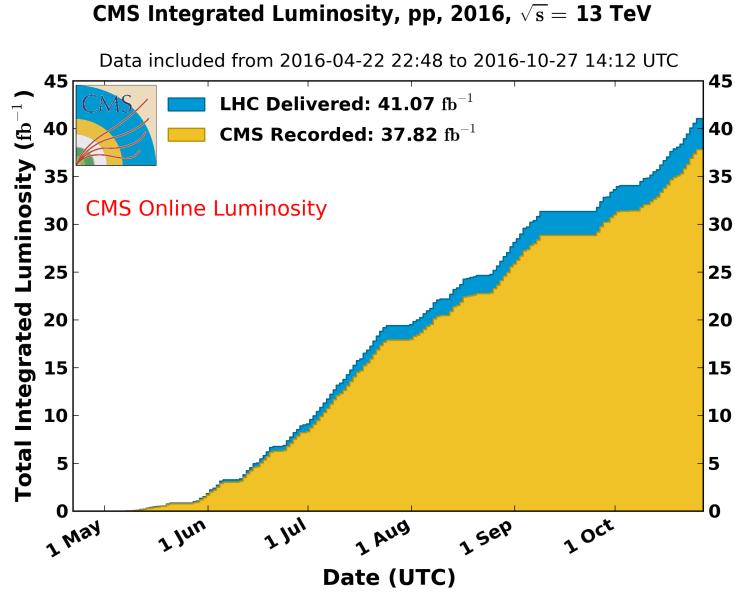


Figure 3.6: Integrated luminosity delivered by LHC and recorded by CMS during 2016. The difference between the delivered and the recorded luminosity is due to fails and issues occurred during the data taking in the CMS experiment [64].

1055 Luminosity is a fundamental aspect of LHC given that the bigger luminosity the
 1056 bigger number of collisions, which means that for processes with a very small cross
 1057 section the number of expected occurrences is increased and so the chances of being
 1058 detected. The integrated luminosity, i.e., the total luminosity, collected by the CMS
 1059 experiment during 2016 is shown in figure 3.6; the data analyzed in this thesis corre-
 1060 sponds to an integrated luminosity of 35.9 fb^{-1} at a center of mass-energy $\sqrt{s} = 13$
 1061 TeV.

1062

1063 A way to increase L is increasing the number of bunches in the beam. Currently, the
 1064 separation between two consecutive bunches in the beam is 7.5 m which corresponds
 1065 to a time separation of 25 ns. In the full LHC ring the allowed number of bunches is
 1066 $n = 27\text{km}/7.5\text{m} = 3600$; however, there are some gaps in the bunch pattern intended
 1067 for preparing the dumping and injection of the beam, thus, the proton beams are

1068 composed of 2808 bunches.

1069

1070 Once the proton beams reach the desired energy, they are brought to cross each other
 1071 producing proton-proton collisions. The bunch crossing happens in precise places
 1072 where the four LHC experiments are located, as seen in the top of figure 3.7. In 2008,
 1073 the first set of collisions involved protons with $\sqrt{s} = 7$ TeV; the energy was increased
 1074 to 8 TeV in 2012 and to 13 TeV in 2015.

1075

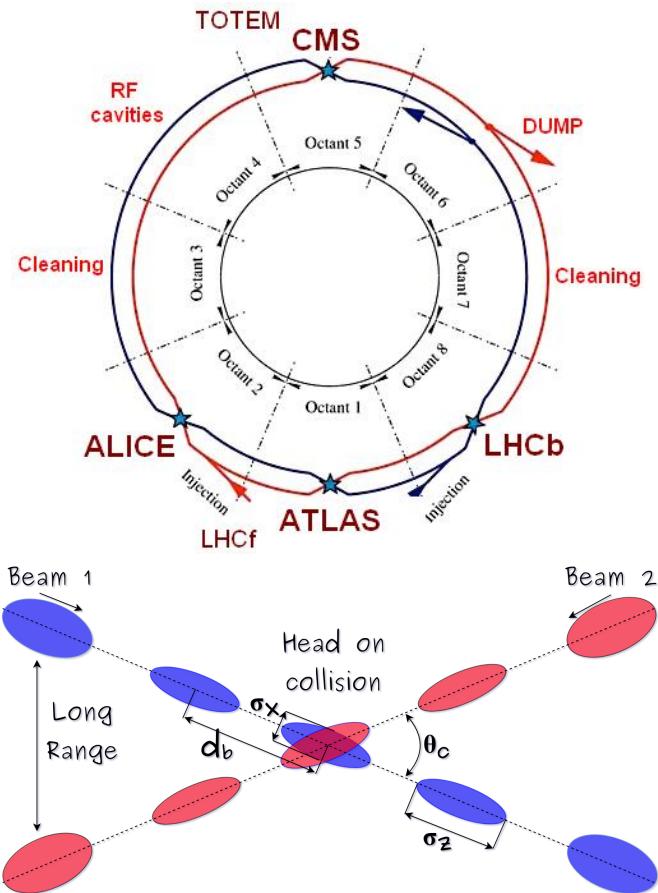


Figure 3.7: Top: LHC interaction points. Bunch crossing occurs where the LHC experiments are located [65]. Sections indicated as cleaning are dedicated to collimate the beam in order to protect the LHC ring from collisions with protons in very spreaded bunches. Bottom: bunch crossing scheme. Since the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

1076 CMS and ATLAS experiments, which are multi-purpose experiments, are enabled
 1077 to explore physics in any of the collision modes. LHCb experiment is optimized
 1078 to explore bottom quark physics, while ALICE is optimized for heavy ion collisions
 1079 searches; TOTEM and LHCf are dedicated to forward physics studies; MoEDAL (not
 1080 indicated in the figure) is intended for monopoles or massive pseudo stable particles
 1081 searches.

1082

1083 At the CP there are two interesting details that need to be addressed. The first one
 1084 is that the bunch crossing does not occur head-on but at a small crossing angle “ θ_c ”
 1085 (280 μ rad in CMS and ATLAS) as shown in the bottom side of figure 3.7, affecting
 1086 the overlapping between bunches; the consequence is a reduction of about 17% in
 1087 the luminosity (represented by a factor not included in eqn: 3.2). The second one
 1088 is the occurrence of multiple pp collisions in the same bunch crossing; this effect is
 1089 called pile-up (PU). A fairly simple estimation of the PU follows from estimating the
 1090 probability of collision between two protons, one from each of the bunches in course
 1091 of collision; it depends roughly on the ratio of proton size and the cross section of the
 1092 bunch in the interaction point, i.e.,

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

1093 however, there are $N = 1.15 \times 10^{11}$ protons in a bunch, thus the estimated number of
 1094 collisions in a bunch crossing is

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

1095 about 20 of those pp collisions are inelastic. Each collision generates a vertex, but
 1096 only the most energetic is considered as a primary vertex; the rest are considered as

1097 PU vertices. A multiple pp collision event in a bunch crossing at CMS is showed in
 1098 figure3.8. Unstable particles outgoing from the primary vertex will eventually decay;
 1099 this decay vertex is known as a secondary vertex.

1100

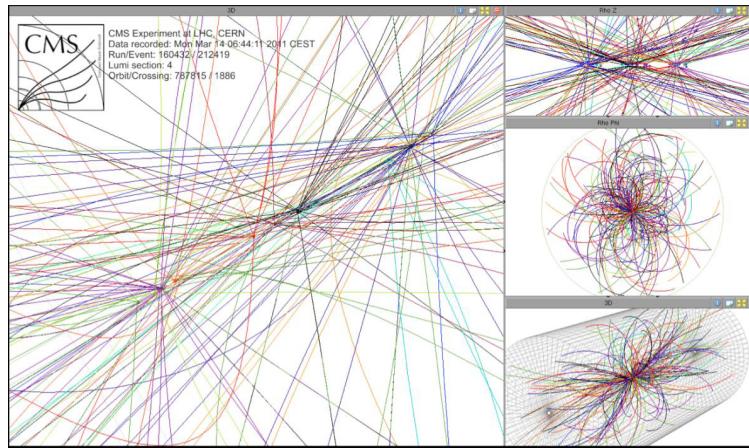


Figure 3.8: Multiple pp collision bunch crossing at CMS. Only the most energetic vertex is considered and the rest are catalogued as PU vertices [66].

1101 Next section presents a description of the CMS detector which it is the detector used
 1102 to collect the data used in this thesis.

1103 3.3 The CMS experiment

1104 CMS is a general-purpose detector designed to conduct research in a wide range
 1105 of physics from the standard model to new physics like extra dimensions and dark
 1106 matter. Located at the point 5 in the LHC layout as shown in figure 3.4, CMS is
 1107 composed of several detection systems distributed in a cylindrical structure; in total,
 1108 CMS weights about 12500 tons in a very compact 21.6 m long and 14.6 m diameter
 1109 cylinder. It was built in 15 separate sections at the ground level and lowered to the
 1110 cavern individually to be assembled. A complete and detailed description of the CMS

1111 detector and its components is given in reference [67] on which this section is based on.

1112

1113 Figure 3.9 shows the layout of the CMS detector. The design is driven by the require-
 1114 ments on the identification, momentum resolution and unambiguous charge determi-
 1115 nation of the muons; therefore, a large bending power is provided by the solenoid
 1116 magnet made of superconducting cable capable to generate a 3.8 T magnetic field.
 1117 The detection system is composed of (from the innermost to the outermost)

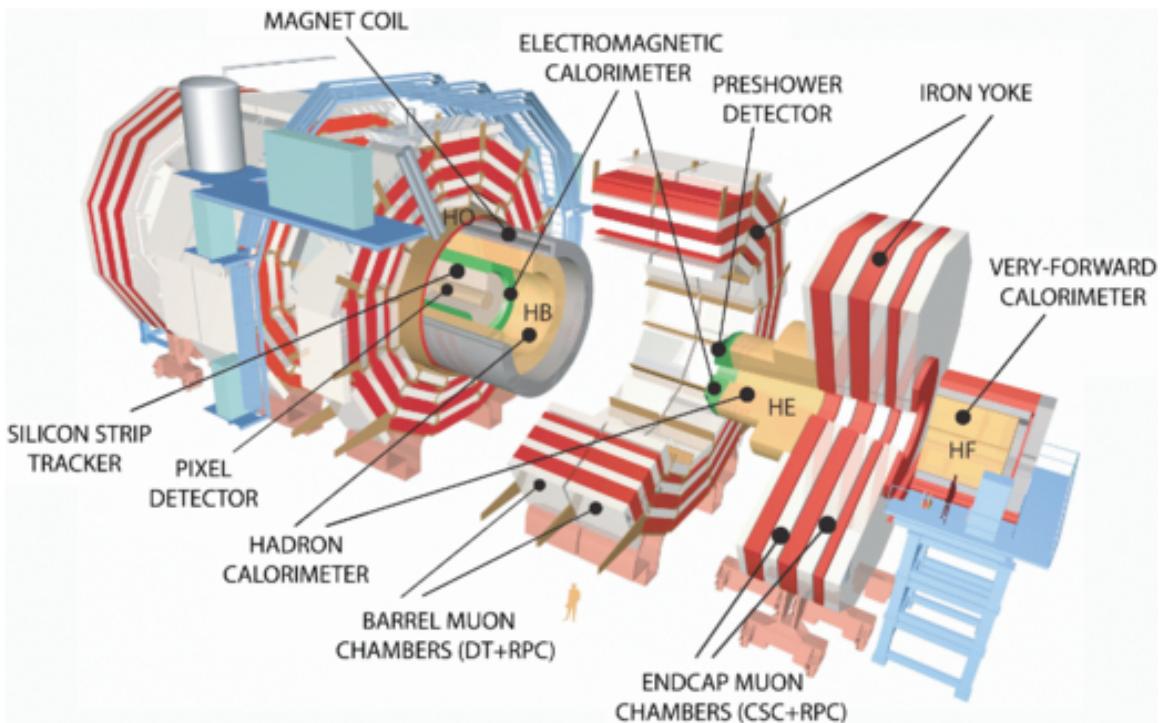


Figure 3.9: Layout of the CMS detector. The several subdetectors are indicated. The central region of the detector is referred as the Barrel section while the endcaps are referred as the forward sections. [68].

1118 • Pixel detector.

1119 • Silicon strip tracker.

1120 • Preshower detector.

1121 • Electromagnetic calorimeter.

1122 • Hadronic calorimeter.

1123 • Muon chambers (Barrel and endcap)

1124 The central region of the detector is commonly referred as the barrel section while the
 1125 endcaps are referred as the forward sections of the detector; thus, each subdetector
 1126 is composed of a barrel section and a forward section.

1127 3.3.1 Coordinate system

1128 The coordinate system used by CMS is centered in the geometrical center of the
 1129 detector which is the same as the CP as shown in figure3.10. The z -axis is parallel
 1130 to the beam direction, while the Y -axis pointing vertically upward, and the X -axis
 1131 pointing radially inward toward the center of the LHC.

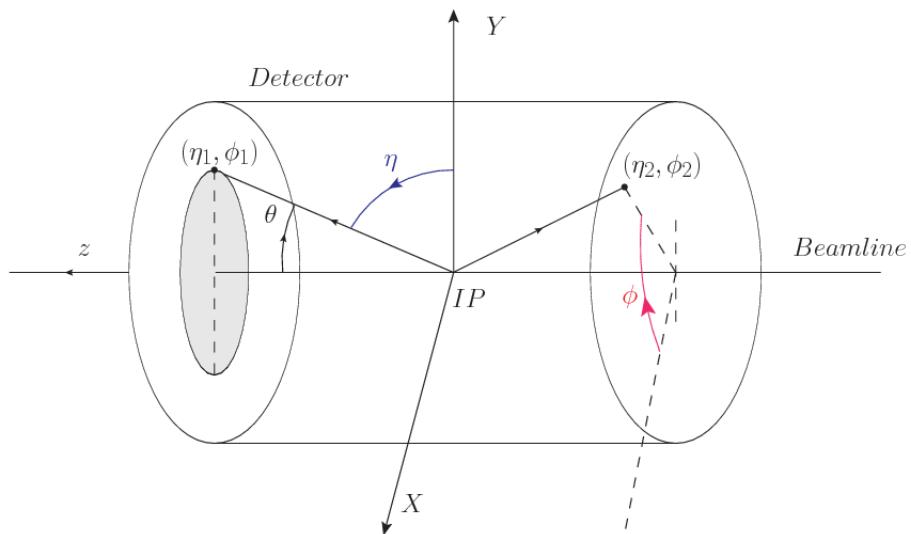


Figure 3.10: CMS detector coordinate system.

1132 In addition to the common cartesian and cylindrical coordinate systems, two coordi-
 1133 nates are of particular utility in particle physics: rapidity(y) and pseudorapidity(η),

1134 defined in connection to the polar angle θ , energy and longitudinal momentum com-
 1135 ponent (momentum along the z -axis) according to

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

1136 Rapidity is related to the angle between the XY -plane and the direction in which the
 1137 products of a collision are emitted; it has the nice property that the difference between
 1138 the rapidities of two particles is invariant with respect to Lorentz boosts along the z -
 1139 axis. Thus, data analysis becomes more simple when based on rapidity; however, it is
 1140 not simple to measure the rapidity of highly relativistic particles, as those produced
 1141 after pp collisions. Under the highly relativistic motion approximation, y can be
 1142 rewritten in terms of the polar angle, concluding that rapidity is approximately equal
 1143 to the pseudorapidity defined above, i.e., $y \approx \eta$. Note that η is easier to measure than y
 1144 given the direct relationship between the former and the polar angle. Angular distance
 1145 between two objects in the detector (ΔR) is defined in terms of their coordinates
 1146 $(\eta_1, \phi_1), (\eta_2, \phi_2)$ as

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

1147 3.3.2 Pixels detector

1148 The CMS tracking system is designed to provide a precise measurement of the tra-
 1149 jectory (*track*) followed by the charged particles created after the pp collisions; also,
 1150 the precise reconstruction of the primary and secondary origins (*vertices*) is expected
 1151 in an environment where, each 25 ns, the bunch crossing produce about 20 inelastic
 1152 collisions and about 1000 particles. An increment in the luminosity is ongoing which
 1153 implies that the PU will increase accordingly.

1155 The pixel detector was replaced during the 2016-2017 extended year-end technical
 1156 stop, due to the increasingly challenging operating conditions like the higher particle
 1157 flow and more radiation harsh environment, among others. The new one is respond-
 1158 ing as expected, reinforcing its crucial role in the successful way to fulfill the new
 1159 LHC physics objectives after the discovery of the Higgs boson. The last chapter of
 1160 this thesis is dedicated to describe my contribution to the “Forward Pixel Phase 1
 1161 upgrade”.

1162

1163 The current pixel detector is composed of 1856 silicon pixel detector modules orga-
 1164 nized in four-barrel layers in the central region and three disks in the forward region;
 1165 it is designed to record efficiently and with high precision, up to $10\mu\text{m}$ in the XY -
 1166 plane and $20\mu\text{m}$ in the z -direction, the first four space-points (*hits*) near to the CP
 1167 region (see figure 3.11 left side) in the range $|\eta| \leq 2.5$. The first barrel layer is located
 1168 at a radius of 30 mm from the beamline, while the fourth layer is located at a radius
 1169 of 160 mm closer to the strip tracker inner barrel layer (see section 3.3.3) in order to
 1170 reduce the rate of fake tracks. The high granularity of the detector is represented in
 1171 its about 123 Mpixels, each of size $100 \times 150\mu\text{m}^2$, which is almost twice the channels
 1172 of the old detector. The transverse momentum resolution of tracks can be measured
 1173 with a resolution of 1-2% for muons of $p_T = 100$ GeV.

1174

1175 Some of the improvements with respect to the previous pixel detector include a higher
 1176 average tracking efficiency and lower average fake rate as well as higher track impact
 1177 parameter resolution which is fundamental in order to increase the efficiency in the
 1178 identification of jets originating from b quarks (b-tagging). A significant source of
 1179 improvement comes from the overall reduction in the material budget of the detector
 1180 which results in fewer photon conversions and less multiple scattering from charged

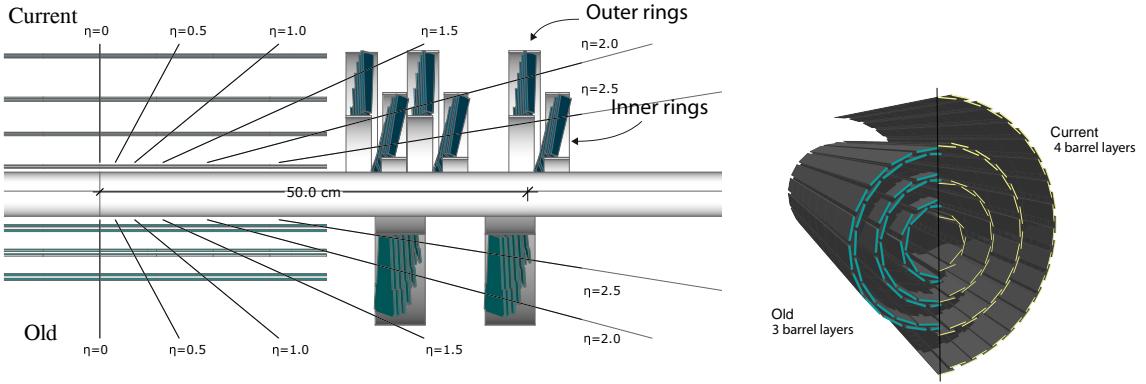


Figure 3.11: CMS pixel detector schematic view. Left: layout comparing the layers and disks in the old and current pixel detectors. Right: Transverse-oblique view comparing the pixel barrel layers in the two [70].

1181 particles.

1182 3.3.3 Silicon strip tracker

1183 The silicon strip tracker (SST) is the second stage in the CMS tracking system. The
 1184 top side of figure 3.12 shows a schematic of the SST. The inner tracker region is com-
 1185 posed of the tracker inner barrel (TIB) and the tracker inner disks (TID) covering the
 1186 region $r < 55$ cm and $|z| < 118$ cm. The TIB is composed of 4 layers while the TID
 1187 is composed of 3 disks at each end. The silicon sensors in the inner tracker are 320
 1188 μm thick, providing a resolution of about 13-38 μm in the $r\phi$ position measurement.
 1189

1190 The modules indicated in blue in the schematic view of figure 3.12 are two modules
 1191 mounted back-to-back and rotated in the plane of the module by a “stereo” angle of
 1192 100 mrad; the hits from these two modules, known as “stereo hits”, are combined to
 1193 provide a measurement of the second coordinate (z in the barrel and r on the disks)
 1194 allowing the reconstruction of hit positions in 3-D.

1195

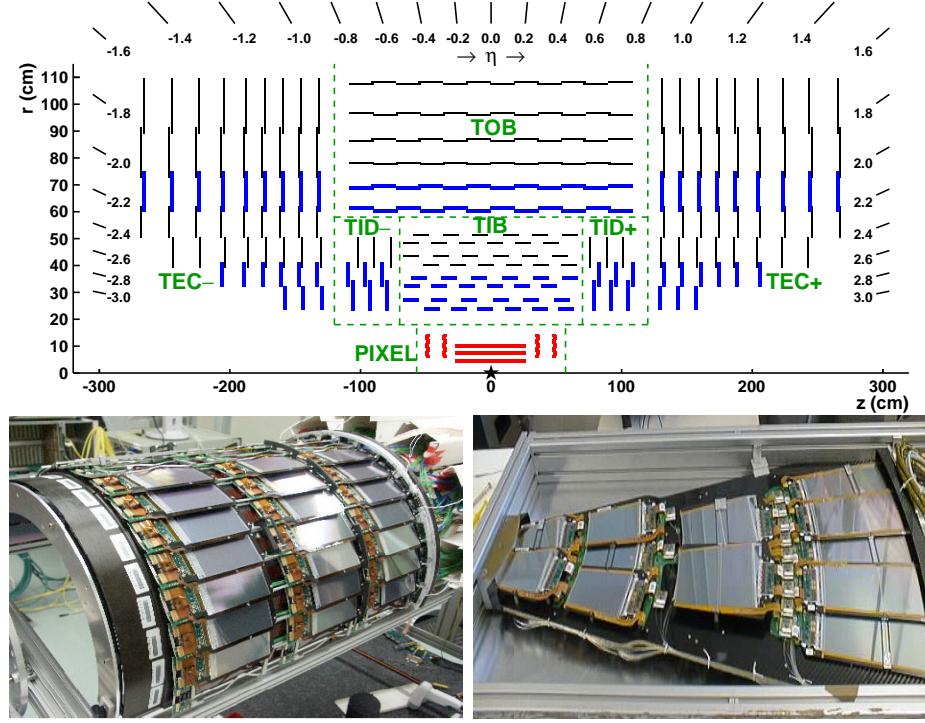


Figure 3.12: Top: CMS Silicon Strip Tracker (SST) schematic view. The SST is composed of the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). Each part is made of silicon strip modules; the modules in blue represent two modules mounted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

1196 The outer tracker region is composed of the tracker outer barrel (TOB) and the
 1197 tracker endcaps (TEC). The six layers of the TOB offer coverage in the region $r > 55$
 1198 cm and $|z| < 118$ cm, while the 9 disks of the TEC cover the region $124 < |z| < 282$
 1199 cm. The resolution offered by the outer tracker is about $13\text{--}38 \mu\text{m}$ in the $r\phi$ position
 1200 measurement. The inner four TEC disks use silicon sensors $320 \mu\text{m}$ thick; those in
 1201 the TOB and the outer three TEC disks use silicon sensors of $500 \mu\text{m}$ thickness. The
 1202 silicon strips run parallel to the z -axis and the distance between strips varies from 80
 1203 μm in the inner TIB layers to $183 \mu\text{m}$ in the inner TOB layers; in the endcaps the
 1204 wedge-shaped sensors with radial strips, whose pitch range between $81 \mu\text{m}$ at small
 1205 radii and $205 \mu\text{m}$ at large radii.

1206

1207 The whole SST has 15148 silicon modules, 9.3 million silicon strips and cover a total
 1208 active area of about 198 m^2 .

1209 **3.3.4 Electromagnetic calorimeter**

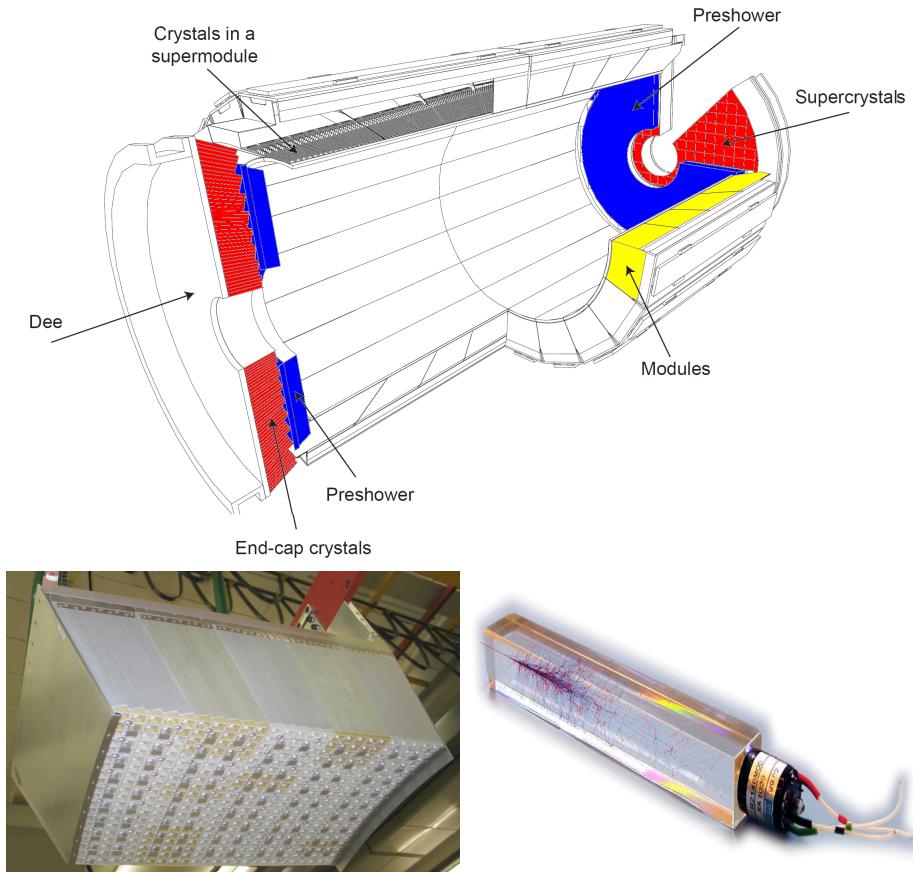


Figure 3.13: Top: CMS ECAL schematic view. Bottom: Module equipped with the crystals (left); ECAL crystal(right).

1210 The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of
 1211 electrons and photons. It is composed of 75848 lead tungstate crystals which have a
 1212 short radiation length (0.89 cm) and fast response, since 80% of the light is emitted
 1213 within 25 ns; however, they are combined with Avalanche photodiodes (APDs) as

1214 photodetectors given that crystals themselves have a low light yield ($30\gamma/\text{MeV}$). A
 1215 schematic view of the ECAL is shown in figure 3.13.

1216

1217 Energy is measured when electrons and photons are absorbed by the crystals which
 1218 generates an electromagnetic “shower”, as seen in bottom right picture of the fig-
 1219 ure3.13; the shower is seen as a *cluster* of energy which depending on the amount
 1220 of energy deposited can involve several crystals. The ECAL barrel (EB) covers the
 1221 region $|\eta| < 1.479$, using crystals of depth of 23 cm and $2.2 \times 2.2 \text{ cm}^2$ transverse
 1222 section; the ECAL endcap (EE) covers the region $1.479 < |\eta| < 3.0$ using crystals
 1223 of depth 22 cm and transverse section of $2.86 \times 2.86 \text{ cm}^2$; the photodetectors used
 1224 are vacuum phototriodes (VPTs). Each EE is divided in two structures called “Dees”.

1225

1226 In front of the EE, it is installed the preshower detector (ES) which covers the region
 1227 $1.653 < |\eta| < 2.6$. The ES provides a precise measurement of the position of electro-
 1228 magnetic showers, which allows to distinguish electrons and photons signals from π^0
 1229 decay signals. The ES is composed of a layer of lead absorber followed by a layer of
 1230 plastic scintillators

1231 3.3.5 Hadronic calorimeter

1232 Hadrons are not absorbed by the ECAL but by the hadron calorimeter (HCAL),
 1233 which is made of a combination of alternating brass absorber layers and silicon photo-
 1234 multiplier(SiPM) layers; therefore, particles passing through the scintillator material
 1235 produce showers, as in the ECAL, as a result of the inelastic scattering of the hadrons
 1236 with the detector material. Since the particles are not absorbed in the scintillator,
 1237 their energy is sampled; therefore the total energy is not measured but estimated from

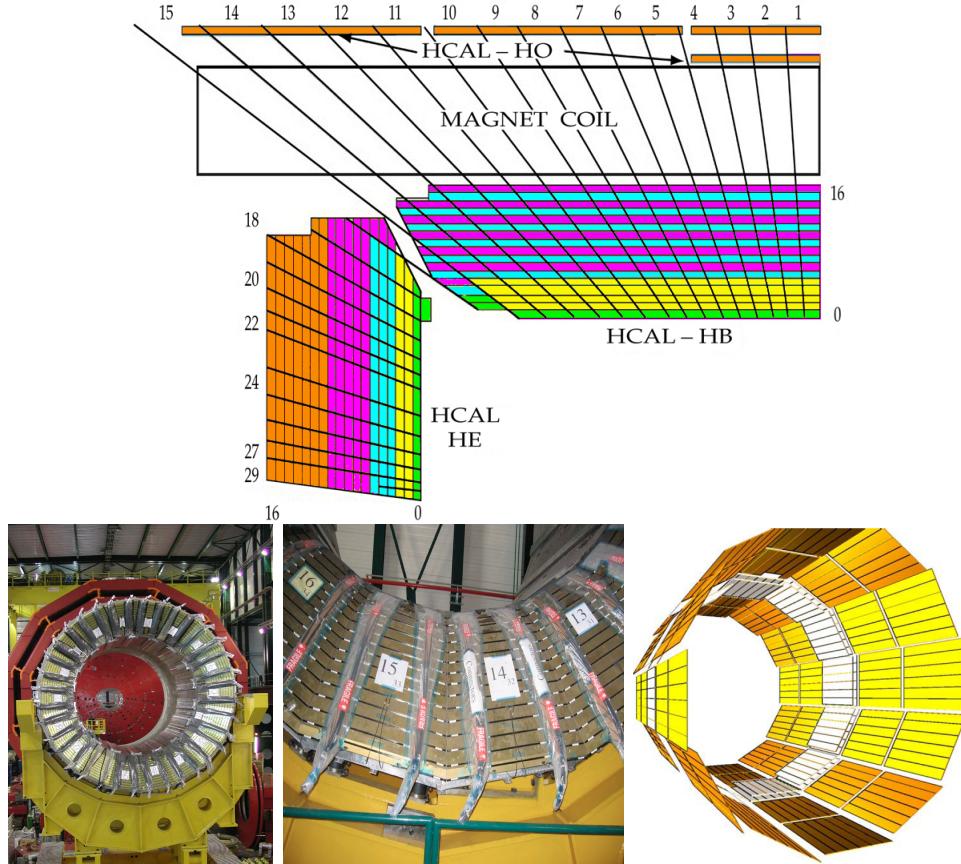


Figure 3.14: Top: CMS HCAL schematic view, the colors indicate the layers that are grouped into the same readout channels. Bottom: picture of a section of the HB; the absorber material is the golden region and scintillators are placed in between the absorber material (left and center). Schematic view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

1238 the energy clusters, which reduce the resolution of the detector. Brass was chosen
 1239 as the absorber material due to its short interaction length ($\lambda_I = 16.42\text{cm}$) and its
 1240 non-magnetivity. Figure 3.14 shows a schematic view of the CMS HCAL.

1241

1242 The HCAL is divided into four sections; the Hadron Barrel (HB), the Hadron Outer
 1243 (HO), the Hadron Endcap (HE) and the Hadron Forward (HF) sections. The HB
 1244 covers the region $0 < |\eta| < 1.4$, while the HE covers the region $1.3 < |\eta| < 3.0$. The HF,
 1245 made of quartz fiber scintillator and steel as absorption material, covers the forward

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

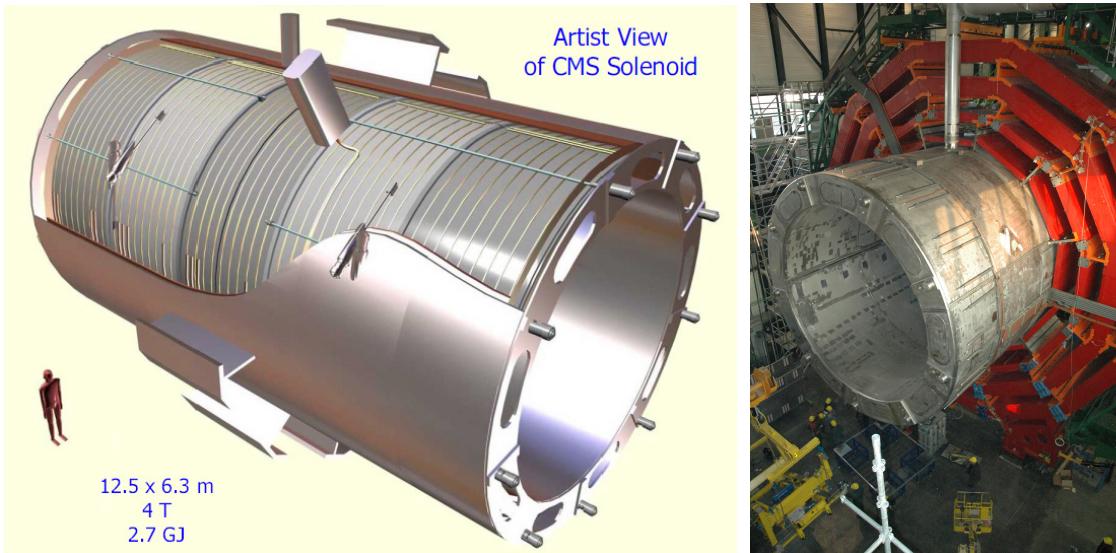


Figure 3.15: Artistic representation of the CMS solenoid magnet(left). The magnet is supported on an iron yoke (right) which also serves as the house of the muon detector and as mechanical support for the whole CMS detector [69].

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

1259 m and a cold mass of 220 t; the generated magnetic field reaches a strength of 3.8T.
 1260 Since it is made of Ni-Tb superconducting cable it has to operate at a temperature
 1261 of 4.7 K by using a helium cryogenic system; the current circulating in the cables
 1262 reaches 18800 A under normal running conditions. The left side of figure 3.15 shows
 1263 an artistic view of the CMS magnet, while the right side shows a transverse view of
 1264 the cold mass where the winding structure is visible.

1265

1266 The yoke (see figure 3.15), composed of 5 barrel wheels and 6 endcap disks made
 1267 of iron, serves not only as the media for magnetic flux return but also provides the
 1268 house for the muon detector system and structural stability to the full detector.

1269 3.3.7 Muon system

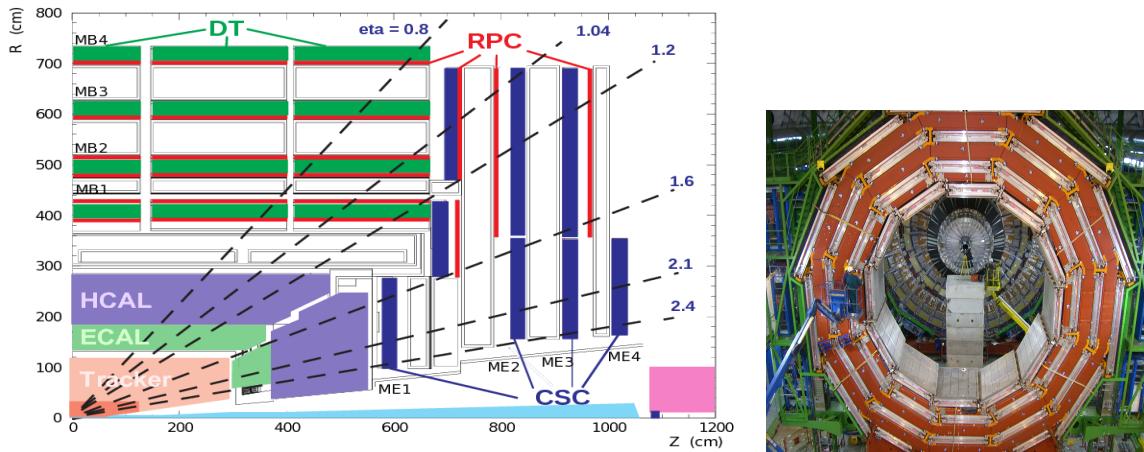


Figure 3.16: Left: CMS muon system schematic view; Right: one of the yoke rings with the muon DTs and RPCs installed; in the back it is possible to see the muon endcap [76].

1270 Muons are the only charged particles able to pass through all the CMS detector due
 1271 to their low ionization energy loss; thus, muons can be separated easily from the
 1272 high amount of particles produced in a pp collision. Also, muons are expected to be
 1273 produced in the decay of several new particles; therefore, a good detection of muons

1274 was on the leading principles when designing the CMS detector.

1275

1276 The CMS muon detection system (muon spectrometer) is embedded in the return
1277 yoke as seen in figure 3.16. It is composed of three different detector types, the drift
1278 tube chambers (DT), Cathode strip chambers (CSC), and resistive plate chambers
1279 (RPC); DT are located in the central region $\eta < 1.2$ arranged in four layers of drift
1280 chambers filled with an Ar/CO₂ gas mixture.

1281

1282 The muon endcaps are made of CSCs covering the region $\eta < 2.4$ and filled with a
1283 mixture of Ar/CO₂/CF₄. The reason behind using a different detector type lies on
1284 the different conditions in the forward region like the higher muon rate and higher
1285 residual magnetic field compared to the central region.

1286

1287 The third type of detector used in the muon system is a set of four disks of RPCs
1288 working in avalanche mode. The RPCs provide good spatial and time resolutions.
1289 The track of $high - p_T$ muon candidates is built combining information from the
1290 tracking system and the signal from up to six RPCs and four DT chambers.

1291 The muon tracks are reconstructed from the hits in the several layers of the muon
1292 system.

1293 **3.3.8 CMS trigger system**

1294 Under normal conditions, CMS expects pp collisions every 25 ns, i.e., an interaction
1295 rate of 40 MHz for which it is not possible to store the recorded data in full. In order
1296 to handle this high event rate data, an online event selection, known as triggering, is
1297 performed; triggering reduce the event rate to 100 Hz for storage and further offline

1298 analysis.

1299

1300 The trigger system starts with a reduction of the event rate to 100 kHz in the so-called
 1301 “level 1 trigger (L1)”. L1 is based on dedicated programmable hardware like Field
 1302 Programmable Gate Arrays (FPGAs) and Application Specific Integrated Circuits
 1303 (ASICs), partly located in the detector itself; another portion is located in the CMS
 1304 under-ground cavern. Hit patterns information from the muon chambers and the en-
 1305 ergy deposits in the calorimeter are used to decide if an event is accepted or rejected,
 1306 according to selection requirements previously defined, which reflect the interesting
 1307 physics processes. Figure 3.17 shows the L1 trigger architecture.

1308

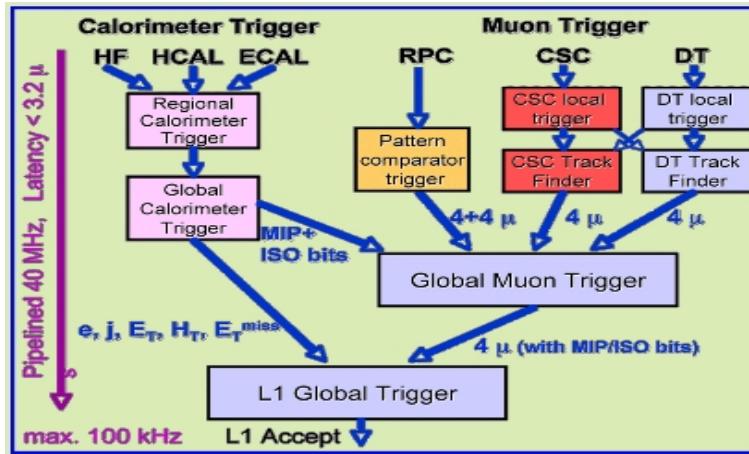


Figure 3.17: CMS Level-1 trigger architecture [77].

1309 The second stage in the trigger system is called “high-level trigger (HLT)”; events
 1310 accepted by L1 are passed to HLT in order to make an initial reconstruction of them.
 1311 HLT is software based and runs on a dedicated server farm, using selection algo-
 1312 rithms and high-level object definitions; the event rate at HLT is reduced to 100 Hz.
 1313 The first HLT stage takes information from the muon detectors and the calorimeters
 1314 to make the initial object reconstruction; in the next HLT stage, information from

1315 the pixel and strip detectors is used to do first fast-tracking and then full tracking
1316 online. This initial object reconstruction is used in further steps of the trigger system.

1317

1318 Events and preliminary reconstructed physics objects from HLT are sent to be fully
1319 reconstructed at the CERN computing center. Again, the pixel detector information
1320 provides high-quality seeds for the track reconstruction algorithm offline, primary ver-
1321 tex reconstruction, electron and photon identification, muon reconstruction, τ iden-
1322 tification, and b-tagging. After full reconstruction, data sets are made available for
1323 offline analyses.

1324

1325 During the 2016-2017 technical stop, the L1 system was updated in order to improve
1326 the physics object identification by improving the algorithms and accounting for the
1327 increasing pile-up scenario.

1328 **3.3.9 CMS computing**

1329 After the data, coming from the experiment, are processed at several levels, they have
1330 to be stored and made available for further analysis; in order to cope all the tasks
1331 implied in the offline data processing, like transfer, simulation, reconstruction and
1332 reprocessing, among others, a big computing power is required. The CMS computing
1333 system is based on the distributed architecture concept, where users of the system
1334 and physical computer centers are distributed worldwide and interconnected by high-
1335 speed networks.

1336 The worldwide LHC computing grid (WLCG) is the mechanism used to provide that
1337 distributed environment. WLCG is a tiered structure connecting computing centers
1338 around the world, which provides the necessary storage and computing facilities. The

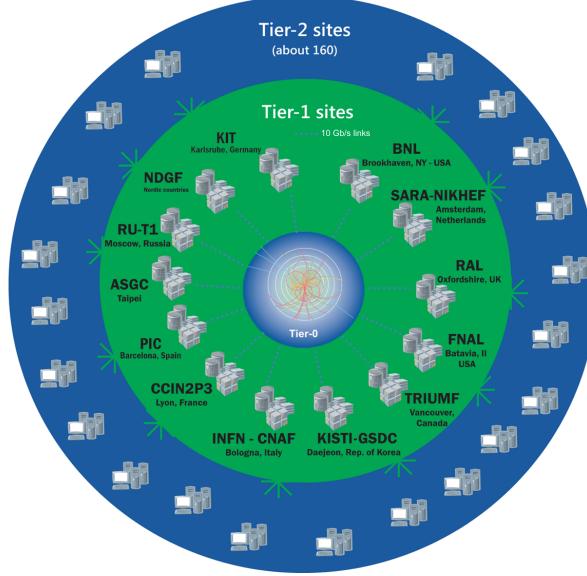


Figure 3.18: WLCG structure. The primary computer centers (Tier-0) are located at CERN (data center) and at the Wigner datacenter in Budapest. Tier-1 is composed of 13 centers and Tier-2 is composed of about 160 centers. [78].

1339 primary computing centers of the WLCG are located at the CERN and the Wigner
 1340 datacenter in Budapest and are known as Tier-0 as shown in figure 3.18. The main
 1341 responsibilities for each tier level are [78]

- 1342 • **Tier-0:** initial reconstruction of recorded events and storage of the resulting
 1343 datasets, the distribution of raw data to the Tier-1 centers.
- 1344 • **Tier-1:** provide storage capacity, support for the Grid, safe-keeping of a pro-
 1345 portional share of raw and reconstructed data, large-scale reprocessing and safe-
 1346 keeping of corresponding output, generation of simulated events, distribution
 1347 of data to Tier 2s, safe-keeping of a share of simulated data produced at these
 1348 Tier 2s.
- 1349 • **Tier-2:** store sufficient data and provide adequate computing power for specific
 1350 analysis tasks, provide analysis requirements and proportional share of simu-
 1351 lated event production and reconstruction.

1352 Aside from the general computing strategy to manage the huge amount of data pro-
1353 duced by experiments, CMS uses a framework to perform a variety of processing,
1354 selection and analysis tasks. The central concept of the CMS data model referred to
1355 as “event data model” (EDM) is the “Event”; therefore, an event is the unit that con-
1356 tains the information from a single bunch crossing as well as any data derived from
1357 that information like the reconstructed objects, the details under which additional
1358 data are derived.

1359

1360 Events are passed as the input to the “physics modules” that obtain information from
1361 them and create new one; for instance, “event data producers” add new data into the
1362 events, “analyzers” produce an information summary from an event set, “filters” per-
1363 form selection and triggering.

1364

1365 CMS uses several event formats with different levels of detail and precision

1366 • **Raw format:** events in this format contain the full recorded information from
1367 the detector as well as trigger decision and other metadata. An extended version
1368 of raw data is used to store information from the CMS Monte Carlo simulation
1369 tools. Raw data are stored permanently, occupying about 2MB/event

1370 • **RECO format:** events in this format correspond to raw data that have been
1371 submitted to reconstruction algorithms like primary and secondary vertex re-
1372 construction, particle ID, track-finding. RECO events contain physical objects
1373 and all the information used to reconstruct them; average size is about 0.5
1374 MB/event.

1375 • **AOD format:** Analysis Object Data (AOD) is the data format used in the
1376 physics analyses given that it contains the parameters describing the high-level

1377 physics objects in addition to enough information to allow a kinematic refitting if
 1378 needed. AOD events are filtered versions of the RECO events to which skimming
 1379 or other kind processes have been applied. Requires about 100 kB/event.

1380 • **Non-event data** are data needed to interpret and reconstruct events. Some
 1381 of the non-event data used by CMS contains information about the detector
 1382 contraction and condition data like calibrations, alignment, and detector status.

1383 Figure 3.19 shows the data flow scheme between CMS detector and hardware tiers.

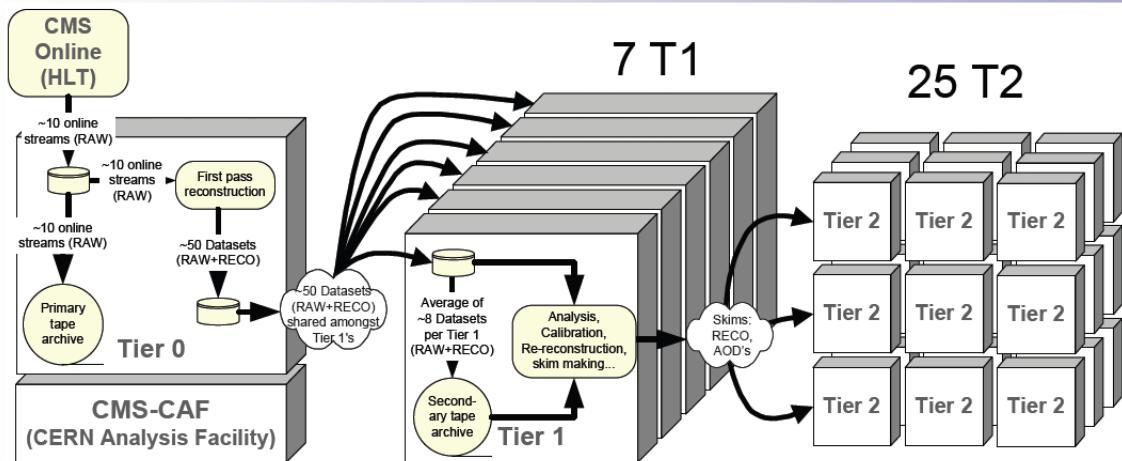


Figure 3.19: Data flow from CMS detector through hardware Tiers.

1384 The whole collection of software built as a framework is referred to as “CMSSW”. This
 1385 framework provides the services needed by the simulation, calibration and alignment,
 1386 and reconstruction modules that process event data, so that physicists can perform
 1387 analysis. The CMSSW event processing model is composed of one executable, called
 1388 cmsRun, and several plug-in modules which contains all the tools (calibration, recon-
 1389 struction algorithms) needed to process an event. The same executable is used for
 1390 both detector and Monte Carlo data [79].

1391 **Chapter 4**

1392 **Event generation, simulation and
1393 reconstruction**

1394 The process of analyzing the data recorded by the CMS experiment involves several
1395 stages where the data are processed in order to interpret the information provided by
1396 all the detection systems; in those stages, the particles produced after the pp collision
1397 are identified by reconstructing their trajectories and measuring their features. In
1398 addition, the SM provides a set of predictions that have to be compared with the
1399 experimental results; however, in most of the cases, theoretical predictions are not
1400 directly comparable to experimental results due to the diverse source of uncertainties
1401 introduced by the experimental setup and theoretical approximations among others.

1402

1403 The strategy to face these conditions consist in using statistical methods implemented
1404 in computational algorithms to produce numerical results that can be contrasted with
1405 the experimental results. These computational algorithms are commonly known as
1406 Monte Carlo (MC) methods and, in the case of particle physics, they are designed to
1407 apply the SM rules and produce predictions about the physical observables measured

in the experiments. Since particle physics is governed by quantum mechanics principles, predictions are not allowed for 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.

1413

1414 This chapter presents a description of the event generation strategy and the tools
 1415 used to perform the detector simulation and physics objects reconstruction. A comprehensive review of event generators for LHC physics can be found in reference [80]
 1416 on which this chapter is based.
 1417

1418 4.1 Event generation

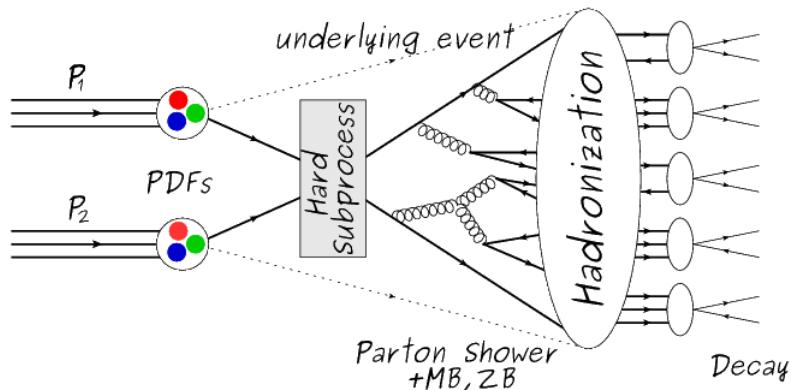


Figure 4.1: Event generation process. In the first step, the PDF of the colliding particles is considered so the specific interaction is described. The actual interaction is generated in the hard subprocess; the cross-section of the process is calculated from the matrix element connecting the initial and final states. The parton shower describes the evolution of the partons from the hard subprocess according to the DGLAP equations. At this step, the underlying event and PU effects are included in the generation. The resulting partons from the parton shower are recombined to form hadrons in the hadronization step; most of them are unstable, therefore, their decays are also generated in agreement to the known branching ratios. Modified from reference [81].

1419 The event generation is intended to create events that mimic the behavior of actual
 1420 events produced in the collisions; they obey a sequence of steps from the particles colli-
 1421 sion hard process to the decay process into the final state particles. Figure 4.1 shows
 1422 a schematic view of the event generation process; the fact that the full process can
 1423 be treated as several independent steps is based on the QCD factorization theorem.

1424

1425 Generation starts by taking into account the PDFs of the incoming particles. Event
 1426 generators offer the option to choose from several PDF sets depending on the partic-
 1427 ular process under simulation¹; in the following pp collisions will be considered. The
 1428 *hard subprocess* describes the actual interaction between partons from the incoming
 1429 protons; it is represented by the matrix element connecting the initial and final states
 1430 of the interaction. Normally, the matrix element can be written as a sum over Feyn-
 1431 man diagrams and consider interferences between terms in the summation. During
 1432 the generation of the hard subprocess, the production cross section is calculated.

1433

1434 The order to which the cross section is calculated depends on the order of the Feyn-
 1435 man diagrams involved in the calculation; therefore, radiative corrections are included
 1436 by considering a higher order Feynman diagrams where QCD radiation dominates.
 1437 Currently, cross sections calculated to LO do not offer a satisfactory description of the
 1438 processes, i.e., the results are only reliable for the shape of distributions; therefore,
 1439 NLO calculations have to be performed with the implication that the computing time
 1440 needed is highly increased.

1441

1442 The final parton content of the hard subprocess is subjected to the *parton shower*
 1443 which generates the gluon radiation. Parton shower evolves the partons; i.e., gluons

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

1444 split into quark-antiquark pairs and quarks of enough energy radiate gluons giving rise
 1445 to further parton multiplication, following the DGLAP (Dokshitzer-Gribov-Lipatov-
 1446 Altarelli-Parisi) equations. Showering continues until the energy scale is low enough
 1447 to reach the non-perturbative limit.

1448

1449 In the simulation of LHC processes that involve b quarks like the single top quark or
 1450 Higgs associated production, it is needed to consider that the b quark is heavier than
 1451 the proton; in this sense, the QCD interaction description is made in two different
 1452 schemes [83]

1453 • four-flavor (4F) scheme. b quarks appear only in the final state because they
 1454 are heavier than the proton and therefore they can be produced only from the
 1455 splitting of a gluon into pairs or singly in association with a t quark in high
 1456 energy-scale interactions. During the simulation, the b -PDFs are set to zero
 1457 because it cannot be part of the proton. Calculations in this scheme are more
 1458 complicated due to the presence of the second b quark but the full kinematics is
 1459 considered already at LO and therefore the accuracy of the description is better.

1460 • five-flavor (5F) scheme. b quarks are considered massless, therefore they can
 1461 appear in both initial and final states since it can now be part of the proton; thus,
 1462 during the simulation b -PDFs are not set to zero. In this scheme, calculations
 1463 are simpler than in the 4F scheme and possible logarithmic divergences are
 1464 absorbed by the PDFs through the DGLAP evolution.

1465 In this thesis, the tHq events are generated using the 4F scheme in order to reduce
 1466 uncertainties, while the tHW events are generated using the 5F scheme to eliminate
 1467 LO interference with the $t\bar{t}H$ process [48].

1468

1469 Partons involved in the pp collision are the focus of the simulation, however, the rest
 1470 of the partons inside the incoming protons are also affected because the remnants are
 1471 colored objects; also, multiple parton interactions can occur. The hadronization of
 1472 the remnants and multiple parton interactions are known as “underlying event” and
 1473 it has to be included in the simulation. In addition, multiple pp collisions in the same
 1474 bunch crossing (pile-up mentioned in 3.2) occurs, actually in two forms

- 1475 • *in-time PU* which refers to multiple pp collision in the bunch crossing but that
 1476 are not considered as primary vertices.
- 1477 • *Out-of-time PU* which refers to overlapping pp collisions from consecutive bunch
 1478 crossings; this can occurs due to the time-delays in the detection systems where
 1479 information from one bunch crossing is assigned to the next or previous one.

1480 While the underlying event effects are included in generation using generator-specific
 1481 tools, PU effects are added to the generation by overlaying Minimum-bias (MB) and
 1482 Zero-bias (ZB) events to the generated events. MB events are inelastic events selected
 1483 by using a loose (minimum bias) trigger with as little bias as possible, therefore ac-
 1484 cepting a large fraction of the overall inelastic event; ZB events correspond to random
 1485 events recorded by the detector when collisions are likely. MB model in-time PU and
 1486 ZB model out-of-time PU.

1487

1488 The next step in the generation process is called “hadronization”. Since particles
 1489 with a net color charge are not allowed to exits isolated, they have to recombine
 1490 to form bound states. This is precisely the process by which the partons resulting
 1491 from the parton shower arrange themselves as color singlets to form hadrons. At
 1492 this step, the energy-scale is low and the strong coupling constant is large, there-
 1493 fore hadronization process is non-perturbative and the evolution of the partons is

1494 described using phenomenological models. Most of the baryons and mesons produced
 1495 in the hadronization are unstable and hence they will decay in the detector.

1496

1497 The last step in the generation process corresponds to the decay of the unstable
 1498 particles generated during hadronization; it is also simulated in the hadronization
 1499 step, based on the known branching ratios.

1500 4.2 Monte Carlo Event Generators.

1501 The event generation described in the previous section has been implemented in
 1502 several software packages for which a brief description is given.

- 1503 • **PYTHIA 8.** It is a program designed to perform the generation of high en-
 1504 ergy physics events which describe the collisions between particles such as elec-
 1505 trons, protons. Several theories and models are implemented in it, in order to
 1506 describe physical aspects like hard and soft interaction, parton distributions,
 1507 initial and final-state parton showers, multiple parton interactions, beam rem-
 1508 nants, hadronization² and particle decay. Thanks to extensive testing, several
 1509 optimized parametrizations, known as “tunings”, have been defined in order
 1510 to improve the description of actual collisions to a high degree of precision; for
 1511 analysis at $\sqrt{s} = 13$ TeV, the underline event CUETP8M1 tune is employed [85].
 1512 The calculation of the matrix element is performed at LO which is not enough
 1513 for the current required level of precision; therefore, pythia is often used for
 1514 parton shower, hadronization and decays, while other event generators are used
 1515 to generate the matrix element at NLO.

² based in the Lund string model [84]

1516 • **MadGraph5_aMC@NLO.** MadGraph is a matrix element generator which
 1517 calculates the amplitudes for all contributing Feynman diagrams of a given pro-
 1518 cess but does not provide a parton shower while MC@NLO incorporate NLO
 1519 QCD matrix elements consistently into a parton shower framework; thus, Mad-
 1520 Graph5_aMC@NLO, as a merger of the two event generators MadGraph5 and
 1521 aMC@NLO, is an event generator capable to calculate tree-level and NLO cross
 1522 sections and perform the matching of those with the parton shower. It is one of
 1523 the most frequently used matrix element generators; however, it has the partic-
 1524 ular feature of the presence of negative event weights which reduce the number
 1525 of events used to reproduce the properties of the objects generated [86].

1526

1527 • **POWHEG.** It is an NLO matrix element generator where the hardest emis-
 1528 sion of color charged particles is generated in such a way that the negative event
 1529 weights issue of MadGraph5_aMC@NLO is overcome; however, the method re-
 1530 quires an interface with p_T -ordered parton shower or a parton shower generator
 1531 where this highest emission can be vetoed in order to avoid double counting of
 1532 this highest-energetic emission. PYTHIA is a commonly matched to POWHEG
 1533 event generator [87].

1534 Events resulting from the whole generation process are known as MC events.

1535 4.3 CMS detector simulation.

1536 After generation, MC events contain the physics of the collisions but they are not
 1537 ready to be compared to the events recorded by the experiment since these recorded
 1538 events correspond to the response of the detection systems to the interaction with

1539 the particles traversing them. The simulation of the CMS detector has to be applied
1540 on top of the event generation; it is simulated with a MC toolkit for the simulation
1541 of particles passing through matter called Geant4 which is also able to simulate the
1542 electronic signals that would be measured by all detectors inside CMS.

1543

1544 The simulation takes the generated particles contained in the MC events as input,
1545 makes them pass through the simulated geometry, and models physics processes that
1546 particles experience during their passage through matter. The full set of results from
1547 particle-matter interactions correspond to the simulated hit which contains informa-
1548 tion about the energy loss, momentum, position. Particles of the input event are
1549 called “primary”, while the particles originating from GEANT4-modeled interactions
1550 of a primary particle with matter are called a “secondary”. Simulated hits are the in-
1551 put of subsequent modules that emulate the response of the detector readout system
1552 and triggers. The output from the emulated detection systems and triggers is known
1553 as digitization [88, 89].

1554

1555 The modeling of the CMS detector corresponds to the accurate modeling of the
1556 interaction among particles, the detector material, and the magnetic field. This
1557 simulation procedure includes the following standard steps

- 1558 • Modeling of the Interaction Region.
- 1559 • Modeling of the particle passage through the hierarchy of volumes that compose
1560 CMS detector and of the accompanying physics processes.
- 1561 • Modeling of the effect of multiple interactions per beam crossing and/or the
1562 effect of events overlay (Pile-Up simulation).

1563 • Modeling of the detector’s electronics response, signal shape, noise, calibration
 1564 constants (digitization).

1565 In addition to the full simulation, i.e. a detailed detector simulation, a faster simula-
 1566 tion (FastSim) have been developed, that may be used where much larger statistics
 1567 are required. In FastSim, detector material effects are parametrized and included in
 1568 the hits; those hits are used as input of the same higher-level algorithms³ used to an-
 1569 alyze the recorded events. In this way, comparisons between fast and full simulations
 1570 can be performed [91].

1571

1572 After the full detector simulation, the output events can be directly compared with
 1573 events actually recorded in the CMS detector. The collection of MC events that
 1574 reproduce the expected physics for a given process are known as MC samples.

1575 **4.4 Event reconstruction.**

1576 In contrast to MC samples for which all the particles’ information is available from
 1577 it’s identity to its mass and energy, recorded events contain the electronic signals,
 1578 provided by the CMS detection systems, encoding the interaction of physical parti-
 1579 cles with the detector matter; these electronic signals have to be combined in order
 1580 to identify these particles and measure their features i.e., particles have to be “recon-
 1581 structed” using the signals provided by the detection systems. The CMS experiment
 1582 use the “particle-flow event reconstruction algorithm (PF)” to do the reconstruction
 1583 of particles produced in pp collisions. Next sections will present a basic description

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

1584 of the *Elements* used by PF (tracker tracks, energy clusters, and muon tracks), based
 1585 in the references [92, 93] where more detailed descriptions can be found.

1586 **4.4.1 Particle-Flow Algorithm.**

1587 Each of the several sub detection systems of the CMS detector is dedicated to identi-
 1588 fying a specific type of particles, i.e., photons and electrons are absorbed by the ECAL
 1589 and their reconstruction is based on ECAL information; hadrons are reconstructed
 1590 from clusters in the HCAL while muons are reconstructed from hits in the muon
 1591 chambers. PF is designed to correlate signals from all the detector layers (tracks and
 1592 energy clusters) in order to reconstruct and identify each final state particle and its
 1593 properties as sketched in figure 4.2.

1594

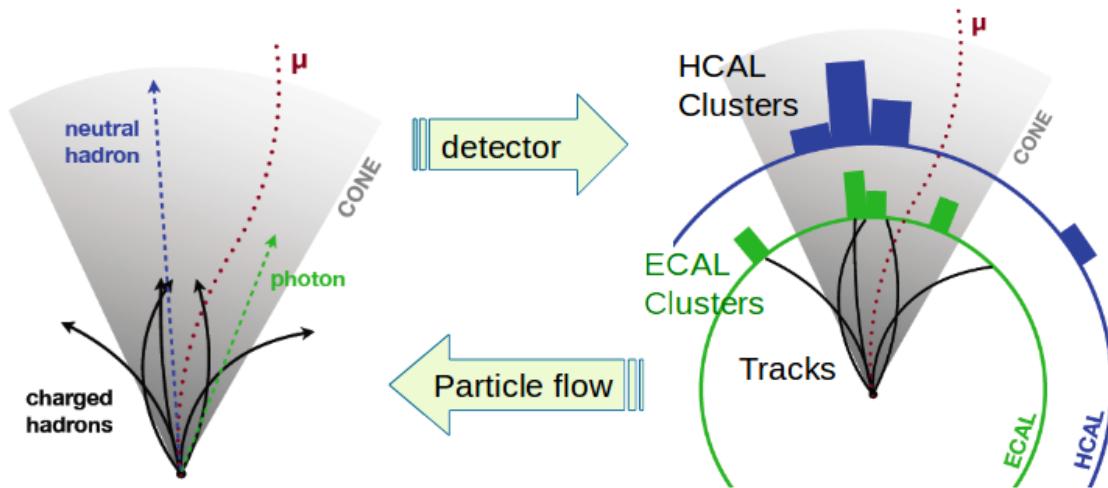


Figure 4.2: Particle flow algorithm. Information from the several CMS detection systems if provided as input to the algorithm which then combine it to identify and reconstruct all the particles in the final state and their properties. Reconstruction of simulated events is also performed by providing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

1595 For instance, a charged hadron is identified by a geometrical connection, know as *link*
 1596 between one or more calorimeter clusters and a track in the tracker provided there

1597 are no hits in the muon system; combining several measurements allows a better
 1598 determination of the energy and charge sign of the charged hadron.

1599 **Charged-particle track reconstruction.**

1600 The strategy used by PF in order to reconstruct tracks is called “Iterative Tracking”
 1601 which occurs in four steps

1602 • Seed generation where initial track candidates are found by looking for a combi-
 1603 nation of hits in the pixel detector, strip tracker, and muon chambers. In total
 1604 ten iterations are performed, each one with a different seeding requirement.
 1605 Seeds are used to estimate the trajectory parameters and uncertainties at the
 1606 time of the full track reconstruction. Seeds are also considered track candidates.

1607 • Track finding using a tracking software known as Combinatorial Track Finder
 1608 (CTF) [95]. The seed trajectories are extrapolated along the expected flight
 1609 path of a charged particle, in agreement to the trajectory parameters obtained
 1610 in the first step, in an attempt to find additional hits that can be assigned to
 1611 the track candidates.

1612 • Track-fitting where the found tracks are passed as input to a module which
 1613 provides the best estimate of the parameters of each trajectory.

1614 • Track selection where track candidates are submitted to a selection which dis-
 1615 cards those that fail a set of defined quality criteria.

1616 Iterations differ in the seeding configuration and the final track selection as elaborated
 1617 in references [92, 93]. In the first iteration, high p_T tracks and tracks produced near
 1618 to the interaction region are identified and those hits are masked thereby reducing
 1619 the combinatorial complexity. Next iterations search for more complicated tracks,

1620 like low p_T tracks and tracks from b hadron decays, which tend to be displaced from
 1621 the interaction region.

1622 **Vertex reconstruction.**

1623 During the track reconstruction, an extrapolation toward to the calorimeters is per-
 1624 formed in order to match energy deposits; that extrapolation is performed also toward
 1625 the beamline in order to find the origin of the track known as *vertex*. The vertex re-
 1626 construction is performed by selecting from the available reconstructed tracks, those
 1627 that are consistent with being originated in the interaction region where pp collisions
 1628 are produced. The selection involves a requirement on the number of tracker (pixel
 1629 and strip) hits and the goodness of the track fit.

1630

1631 Selected tracks are clustered using a “deterministic annealing algorithm (DA)”⁴. A
 1632 set of candidate vertices and their associated tracks, resulting from the DA, are then
 1633 fitted with an “adaptive vertex fitter (AVF)” to produce the best estimate of the
 1634 vertices locations.

1635

1636 The p_T of the several tracks associated to a reconstructed vertex is added, squared and
 1637 used to organize the vertices; the vertex with the highest squared sum is designated
 1638 as the *primary vertex (PV)* while the rest are designated as PU vertices.

1639 **Calorimeter clustering.**

1640 After traversing the CMS tracker system, electrons, photons and hadrons deposit their
 1641 energy in the ECAL and HCAL cells. The PF clustering algorithm aims to provide
 1642 a high detection efficiency even for low-energy particles and an efficient distinction

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

1643 between close energy deposits. The clustering runs independently in the ECAL barrel
 1644 and endcaps, HCAL barrel and endcaps, and the two preshower layers, following two
 1645 steps

- 1646 • cells with an energy larger than a given seed threshold and larger than the energy
 1647 of the neighboring cells are identified as cluster seeds. The neighbor cells are
 1648 those that either share a side with the cluster seed candidate, or the eight closest
 1649 cells including cells that only share a corner with the seed candidate.
- 1650 • cells with at least a corner in common with a cell already in the cluster seed
 1651 and with an energy above a cell threshold are grouped into topological clusters.

1652 Clusters formed in this way are known as *particle-flow clusters*. With this clustering
 1653 strategy, it is possible to detect and measure the energy and direction of photons and
 1654 neutral hadrons as well as differentiate these neutral particles from the charged hadron
 1655 energy deposits. In cases involving charged hadrons for which the track parameters
 1656 are not determined accurately, for instance, low-quality and high- p_T tracks, clustering
 1657 helps in the energy measurements.

1658 **Electron track reconstruction.**

1659 Although the charged-particle track reconstruction described above works for elec-
 1660 trons, they lose a significant fraction of their energy via bremsstrahlung photon radi-
 1661 ation before reaching the ECAL; thus, the reconstruction performance depends on the
 1662 ability to measure also the radiated energy. The reconstruction strategy, in this case,
 1663 requires information from the tracking system and from the ECAL. Bremsstrahlung
 1664 photons are emitted at similar η values to that of the electron but at different values
 1665 of ϕ ; therefore, the radiated energy can be recovered by grouping ECAL clusters in a
 1666 η window over a range of ϕ around the electron direction. The group is called ECAL

1667 supercluster.

1668

1669 Electron candidates from the track-seeding and ECAL super clustering are merged
 1670 into a single collection which is submitted to a full electron tracking fit with a
 1671 Gaussian-sum filter (GSF) [96]. The electron track and its associated ECAL su-
 1672 percluster form a *particle-flow electron*.

1673 Muon track reconstruction.

1674 Given that the CMS detector is equipped with a muon spectrometer capable to iden-
 1675 tify and measure the momentum of the muons traversing it, the muon reconstruction
 1676 is not specific to PF; therefore, three different muon types are defined

- 1677 • *Standalone muon*. A clustering on the DTs or CSCs hits is performed to form
 1678 track segments; those segments are used as seeds for the reconstruction in the
 1679 muon spectrometer. All DTs, CSCs, and RPCs hits along the muon trajectory
 1680 are combined and fitted to form the full track. The fitting output is called a
 1681 *standalone-muon track*.
- 1682 • *tracker muon*. Each track in the inner tracker with p_T larger than 0.5 GeV and
 1683 a total momentum p larger than 2.5 GeV is extrapolated to the muon system. A
 1684 *tracker muon track* corresponds to the extrapolated tracks that match at least
 1685 one muon segment.
- 1686 • *Global muon*. When tracks in the inner tracker (inner tracks) and standalone-
 1687 muon tracks are matched and turn out being compatibles, their hits are com-
 1688 bined and fitted to form a *global-muon track*.

1689 Global muons sharing the same inner track with tracker muons are merged into a
 1690 single candidate. PF muon identification uses the muon energy deposits in ECAL,
 1691 HCAL, and HO associated with the muon track to improve the muon identification.

1692 **Particle identification and reconstruction.**

1693 PF elements are connected by a linker algorithm that tests the connection between any
 1694 pair of elements; if they are found to be linked, a geometrical distance that quantifies
 1695 the quality of the link is assigned. Two elements may be linked indirectly through
 1696 common elements. Linked elements form *PF blocks* and a PF block may contain
 1697 elements originating in one or more particles. Links can be established between
 1698 tracks, between calorimeter clusters, and between tracks and calorimeter clusters.
 1699 The identification and reconstruction start with a PF block and proceeds as follows

1700 • Muons. An “isolated global muon” is identified by evaluating the presence of
 1701 inner track and energy deposits close to the global muon track in the (η, ϕ)
 1702 plane, i.e., in a particular point of the global muon track, inner tracks and
 1703 energy deposits are sought within a radius of $\Delta R = 0.3$ (see eqn. 3.7) from the
 1704 muon track; if they exist and the p_T of the found track added to the E_T of the
 1705 found energy deposit does not exceed 10% of the muon p_T then the global muon
 1706 is an isolated global muon. This isolation condition is stringent enough to reject
 1707 hadrons misidentified as muons.

1708 “Non-isolated global muons” are identified using additional selection require-
 1709 ments on the number of track segments in the muon system and energy deposits
 1710 along the muon track. Muons inside jets are identified with more stringent crite-
 1711 ria in isolation and momentum as described in reference [99]. The PF elements
 1712 associated with an identified muon are masked from the PF block.

- 1713 ● Electrons are identified and reconstructed as described above plus some addi-
 1714 tional requirements on fourteen variables like the amount of energy radiated,
 1715 the distance between the extrapolated track position at the ECAL and the po-
 1716 sition of the associated ECAL supercluster among others, which are combined
 1717 in a specialized multivariate analysis strategy that improves the electron iden-
 1718 tification. Tracks and clusters used to identify and reconstruct electrons are
 1719 masked in the PF block.
- 1720 ● Isolated photons are identified from ECAL superclusters with E_T larger than 10
 1721 GeV, for which the energy deposited at a distance of 0.15, from the supercluster
 1722 position on the (η, ϕ) plane, does not exceed 10% of the supercluster energy;
 1723 note that this is an isolation requirement. In addition, there must not be links
 1724 to tracks. Clusters involved in the identification and reconstruction are masked
 1725 in the PF block.
- 1726 ● Bremsstrahlung photons and prompt photons tend to convert to electron-positron
 1727 pairs inside the tracker, therefore, a dedicated finder algorithm is used to link
 1728 tracks that seem to originate from a photon conversion; in case those two tracks
 1729 are compatible with the direction of a bremsstrahlung photon, they are also
 1730 linked to the original electron track. Photon conversion tracks are also masked
 1731 in the PF block.
- 1732 ● The remaining elements in the PF block are used to identify hadrons. In the
 1733 region $|\eta| \leq 2.5$, neutral hadrons are identified with HCAL clusters not linked
 1734 to any track while photons from neutral pion decays are identified with ECAL
 1735 clusters without links to tracks. In the region $|\eta| > 2.5$ ECAL clusters linked to
 1736 HCAL clusters are identified with a charged or neutral hadron shower; ECAL
 1737 clusters with no links are identified with photons. HCAL clusters not used yet,

1738 are linked to one or more unlinked tracks and to an unlinked ECAL in order to
 1739 reconstruct charged-hadrons or a combination of photons and neutral hadrons
 1740 according to certain conditions on the calibrated calorimetric energy.

- 1741 • Charged-particle tracks may be liked together when they converge to a “sec-
 1742 ondary vertex (SV) ” displaced from the interaction point where the PV and
 1743 PU vertices are reconstructed; at least three tracks are needed in that case,
 1744 of which at most one has to be an incoming track with hits in tracker region
 1745 between a PV and the SV.

1746

1747 The linker algorithm, as well as the whole PF algorithm, has been validated and
 1748 commissioned; results from that validation are presented in the references [92].

1749 **Jet reconstruction.**

1750 Quarks and gluons may be produced in the pp collisions, therefore, their hadronization
 1751 will be seen in the detector as a shower of hadrons and their decay products in the
 1752 form of a “jet”. The anti- k_t algorithm [100] is used to perform the jet reconstruction
 1753 by clustering those PF particles within a cone (see figure 4.3); previously, isolated
 1754 electrons, isolated muons, and charged particles associated with other interaction
 1755 vertices are excluded from the clustering.

1756 The anti- k_t algorithm proceeds in a sequential recombination of PF particles; the
 1757 distance between particles i and j (d_{ij}) and the distance between particles and the
 1758 beam are defined as

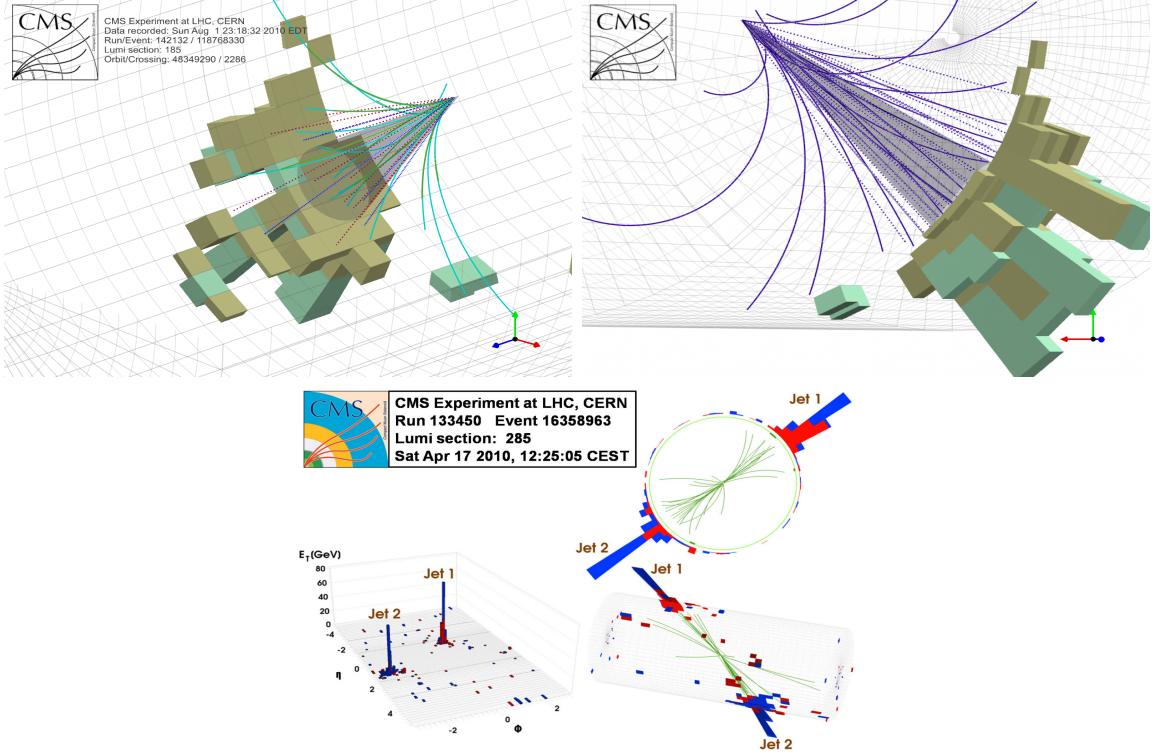


Figure 4.3: Jet reconstruction performed by the anti- k_t algorithm. Top: Two different views of a CMS recorded event are presented. Continuous lines correspond to the tracks left by charged particles in the tracker while dotted lines are the imaginary paths followed by neutral particles. The green cubes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

$$\begin{aligned}
 d_{ij} &= \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right) \frac{\Delta_{ij}^2}{R^2} \\
 d_{iB} &= \frac{1}{k_{ti}^2}
 \end{aligned} \tag{4.1}$$

1759 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-
 1760 pidity and azimuth of particle i respectively and R is the called jet radius. For all
 1761 the remaining PF particles, after removing the isolated ones, d_{ij} and d_{iB} are calcu-

lated⁵ and the smallest is identified; if it is a d_{ij} , particles i and j are replaced with a new object whose momentum is the vectorial sum of the combined particles. If the smallest distance is a d_{iB} the clustering process ends, the object i (which at this stage should be a combination of several PF particles) is declared as a *Particle-flow-jet* (PF jet) and all the associated PF particles are removed from the detector. The clustering process is repeated until no PF particles remain.

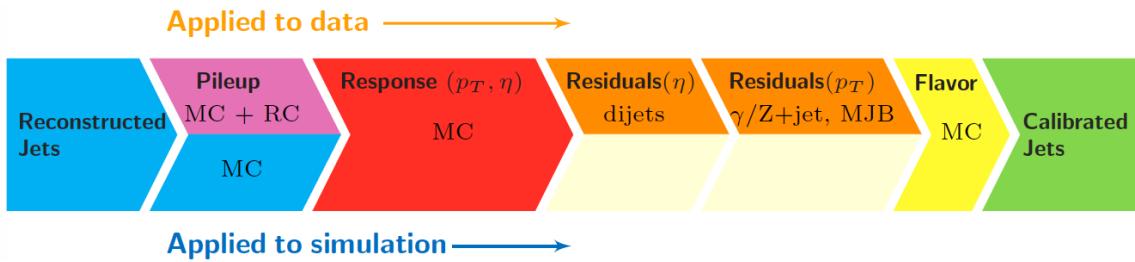


Figure 4.4: Jet energy correction diagram. Correction levels are applied sequentially in the indicated fixed order [104].

Even though jets can be reconstructed efficiently, there are some effects that are not included in the reconstruction and that lead to discrepancies between the reconstructed results and the predicted results; in order to overcome these discrepancies, a factorized model has been designed in the form of jet energy corrections (JEC) [103, 104] applied sequentially as shown in the diagram of figure 4.4.

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

⁵ Notice that this is a combinatorial calculation.

- 1780 • MC-truth correction accounts for differences between the reconstructed jet en-
- 1781 ergy and the MC particle-level energy. The correction is determined on a QCD
- 1782 dijet MC sample and is parametrized in terms of the jet p_T and η .
- 1783 • Residuals correct remaining small differences within jet response in data and
- 1784 MC. The Residuals η -dependent correction compares jets of similar p_T in the
- 1785 barrel reference region. The Residuals p_T -dependent correct the jet absolute
- 1786 scale (JES vs p_T).
- 1787 • Jet-flavor corrections are derived in the same way as MC-truth corrections but
- 1788 using QCD pure flavor samples.

1789 ***b*-tagging of jets.**

1790 A particular feature of the hadrons containing bottom quarks (b-hadrons) is that
 1791 they have a lifetime long enough to travel some distance before decaying, but it is
 1792 not as long as those of light quark hadrons; therefore, when looking at the hadrons
 1793 produced in pp collisions, b-hadrons decay typically inside the tracker rather than
 1794 reach the calorimeters as some light-hadrons do. As a result, a b-hadron decay gives
 1795 rise to a displaced vertex (secondary vertex) with respect to the primary vertex as
 1796 shown in figure 4.5; the SV displacement is in the order of a few millimeters. A jet
 1797 resulting from the decay of a b-hadron is called *b* jet; other jets are called light jets.

1798

1799 Several methods to identify *b*-jets (*b*-tagging) have been developed; the method used in
 1800 this thesis is known as “Combined Secondary Vertex” algorithm in its second version
 1801 (CSVv2) [105]. By using information of the impact parameter, the reconstructed
 1802 secondary vertices and the jet kinematics in a multivariate analysis that combines
 1803 the discrimination power of each variable in one global discriminator variable, three

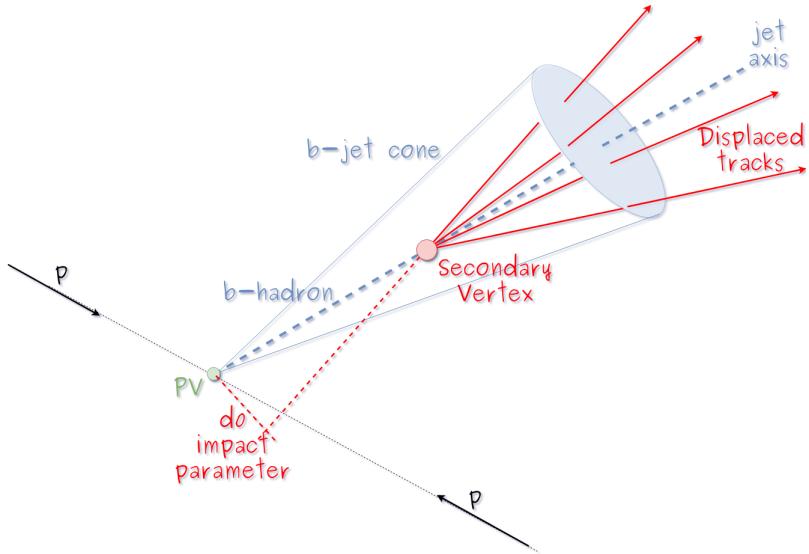


Figure 4.5: Secondary vertex in a b-hadron decay.

1804 working points (references): loose, medium and tight, are defined which quantify the
 1805 probabilities of mistag jets from light quarks as jets from b quarks; 10, 1 and 0.1 %
 1806 respectively. Although the mistagging probability decrease with the working point
 1807 strength, the efficiency to correctly tag b -jets also decrease as 83, 69 and 49 % for the
 1808 respective working point; therefore, a balance needs to be achieved according to the
 1809 specific requirements of the analysis.

1810 Missing transverse energy.

1811 The fact that proton bunches carry momentum along the z axis implies that for each
 1812 event, momentum balance in the transverse plane is expected. Imbalances are quan-
 1813 tified by the missing transverse energy (MET) and are attributed to several sources
 1814 including particles escaping undetected through the beam pipe, neutrinos produced in
 1815 weak interactions processes which do not interact with the detector and thus escaping
 1816 without leaving a sign, or even undiscovered particles predicted by models beyond
 1817 the SM.

1818

1819 The PF algorithm assign the negative sum of the momenta of all reconstructed PF
 1820 particles to the *particle-flow MET* according to

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

1821 JEC are propagated to the calculation of the \vec{E}_T as described in the reference [106].

1822

1823 4.4.2 Event reconstruction examples

1824 Figures 4.6-4.8 show the results of the reconstruction performed on 3 recorded events.
 1825 Descriptions are taken directly from the source.

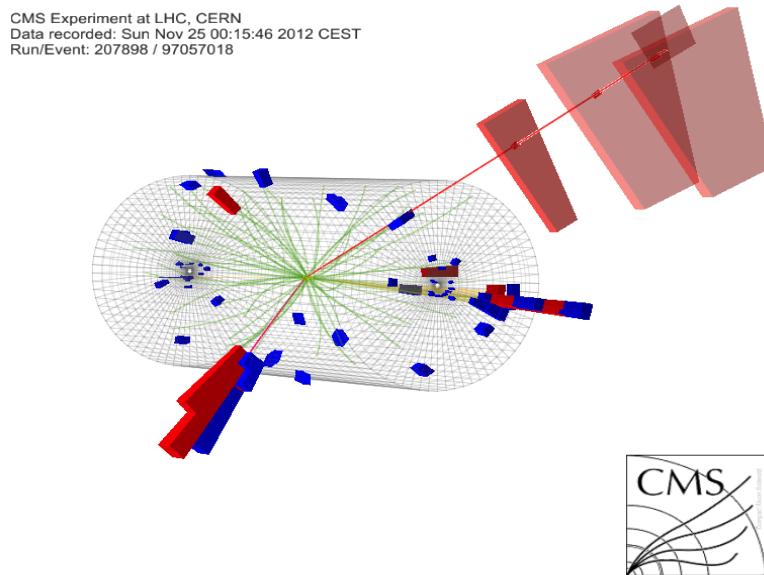


Figure 4.6: HIG-13-004 Event 1 reconstruction results; “HIG-13-004 Event 1: Event recorded with the CMS detector in 2012 at a proton-proton center-of-mass energy of 8 TeV. The event shows characteristics expected from the decay of the SM Higgs boson to a pair of τ leptons. Such an event is characterized 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.” [?].

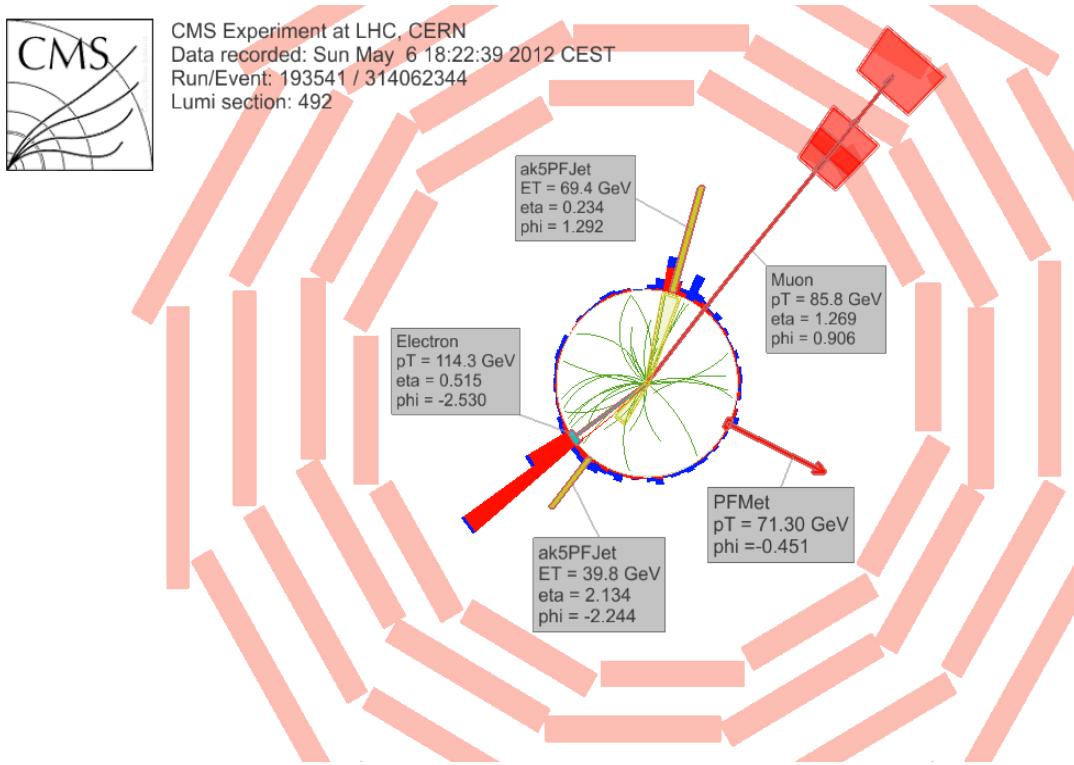


Figure 4.7: $e\mu$ event reconstruction results; “An $e\mu$ event candidate selected in 8 TeV data, as seen from the direction of the proton beams. The kinematics of the main objects used in the event selection are highlighted: two isolated leptons and two particle-flow jets. The reconstructed missing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

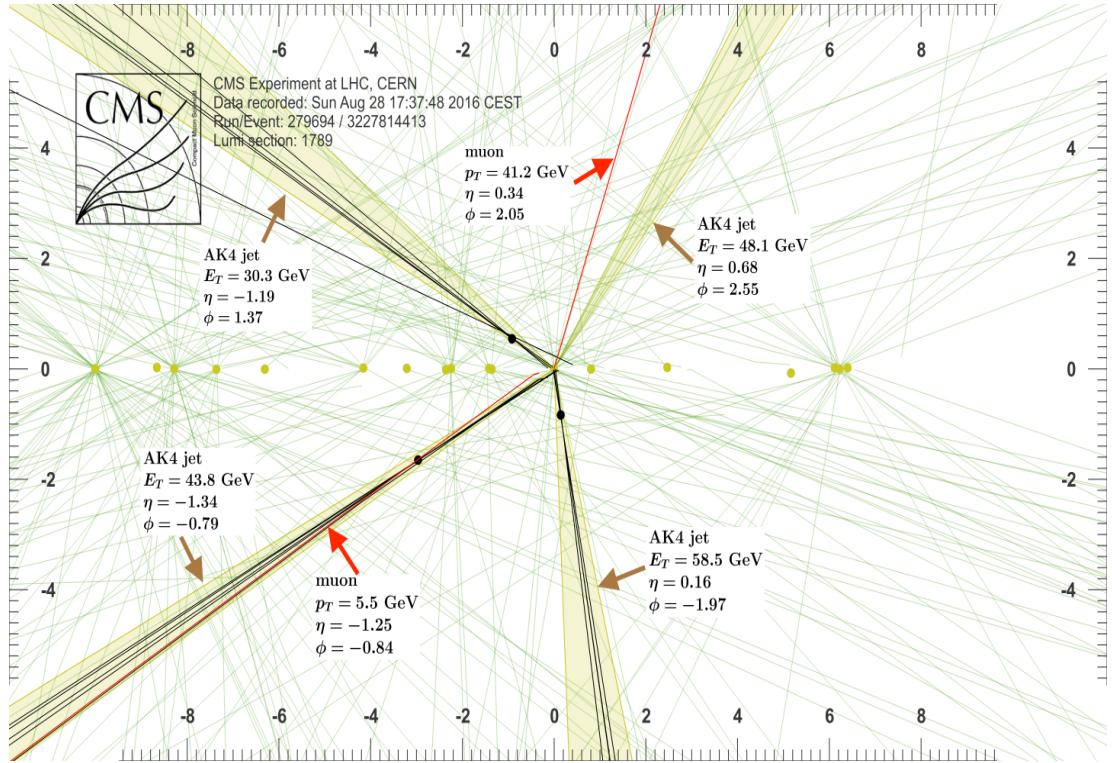


Figure 4.8: Recorded event reconstruction results; “Recorded event (ρ - z projection) with three jets with $p_T > 30$ GeV with one displaced muon track in 2016 data collected at 13 TeV. Each of the three jets has a displaced reconstructed vertex. The jet with $p_T(j) = 43.8$ GeV, $\eta(j) = -1.34$, $\phi(j) = -0.79$ 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” [107].

1826 **Chapter 5**

1827 **Statistical methods**

1828 In the course of analyzing the data sets provided by the CMS experiment and used in
1829 this thesis, several statistical tools have been employed; in this chapter, a description
1830 of these tools will be presented, starting with the general statement of the multivariate
1831 analysis method, followed by the particularities of the Boosted Decision Trees (BDT)
1832 method and its application to the classification problem. Statistical inference methods
1833 used will also be presented. This chapter is based mainly on the reference [108].

1834 **5.1 Multivariate analysis**

1835 Multivariate data analysis (MVA) makes reference to statistical techniques that an-
1836 alyze data containing information of more than one variable, commonly taking into
1837 account the effects of all variables on the response of the particular variable under
1838 investigation, i.e., considering all the correlations between variables. MVA is em-
1839 ployed in a variety of fields like consumer and market research, quality control and
1840 process optimization. From a MVA it is possible to identify the dominant patterns
1841 in the data, like groups, outliers and trends, and determine to which group a set of

1842 values belong; in the particle physics context, MVA methods are used to perform the
 1843 selection of certain type of events, from a large data set, using a potentially large
 1844 number of measurable properties for each event.

1845 Processes with small cross section, as the tHq process, normally are hidden behind
 1846 more common processes; therefore, the data set results in a subset of events with
 1847 characteristic features of interest (signal) mixed in randomly with a much larger
 1848 number of SM events that can mimic these features of interest (background) which
 1849 implies that it is not possible to say with certainty that a given event is signal or
 1850 background. In that sense, the problem can be formulated as one where a set of
 1851 events have to be classified according to some features; these features correspond to
 1852 the measurements of several parameters like energy, momentum organized in a set
 1853 of “input variables”. The measurements for each event can be written in a vector
 1854 $\mathbf{x} = (x_1, \dots, x_n)$ for which

- 1855 • Signal hypotheses $\rightarrow f(\mathbf{x}|s)$ is the probability density (“likelihood”) for \mathbf{x} given
 1856 it is a signal event
- 1857 • Background hypotheses $\rightarrow f(\mathbf{x}|b)$ is the probability density (“likelihood”) of \mathbf{x}
 1858 given it is a background event

1859 Figure 5.1 shows three ways to perform a classification of events for which mea-
 1860 surements of two properties, two input variables, have been performed; blue circles
 1861 represent signal events while red triangles represent background events. The classifi-
 1862 cation on (a) is “cut-based” requiring $x_1 < c_1$ and $x_2 < c_2$; usually the cut values are
 1863 chosen according to some knowledge about the event process. In (b), the classification
 1864 is performed by stating a cut involving a linear function of the input variables and
 1865 so the boundary, while in (c) the relationship between the input variables is not
 1866 linear thus the boundary is not linear either.

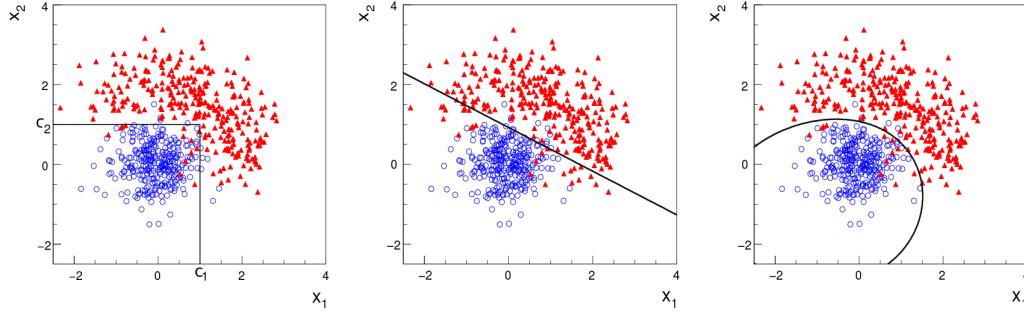


Figure 5.1: Scatter plots-MVA event classification. Distribution of two input variables x_1 and x_2 measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on (a) cuts, (b) linear boundary, and (c) nonlinear boundary [108]

1867 The boundary can be parametrized in terms of the input variables such that the
 1868 cut is set on the parametrization instead of on the variables, i.e., $y(\mathbf{x}) = y_{cut}$ with
 1869 y_{cut} a constant; thus, the acceptance or rejection of an event is based on what side
 1870 of the boundary is the event located. If $y(\mathbf{x})$ has functional form, it can be used to
 1871 determine the probability distribution functions $p(y|s)$ and $p(y|b)$ and then perform
 1872 a scalar test statistic with a single cut on the scalar variable y .

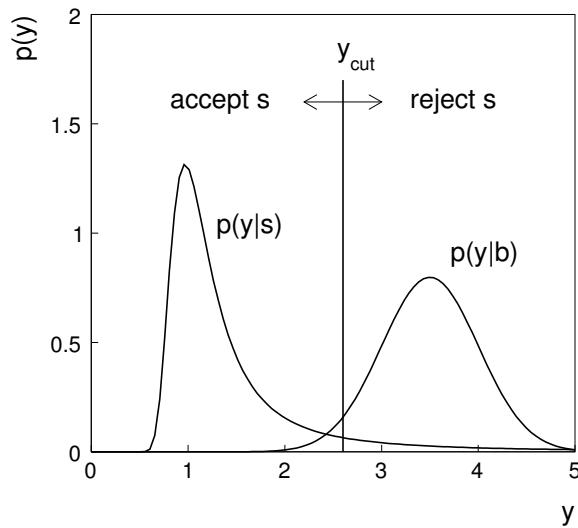


Figure 5.2: Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses. [108]

1873 Figure 5.2 illustrates what would be the probability distribution functions under
 1874 the signal and background hypotheses for a scalar test statistic with a cut on y .

1875 **5.1.1 Boosted decision trees (BDT)**

1876 For this thesis, the implementation of the MVA strategy, described above, is per-
 1877 formed through boosted decision trees (BDT). In a simple picture, a decision tree
 1878 (\mathcal{DT}) classifies events according to their input variables values by setting a cut on
 1879 each input variable and checking which events are on which side of the cut, just as
 1880 proposed in the MVA strategy, but in addition, as a machine learning algorithm, \mathcal{DT}
 1881 offer the possibility to be trained and then perform the classification efficiently.

1882 The training process consists in taking MC samples of signal and background
 1883 events and split them in two parts each; first parts form the training sample which
 1884 will be used in the \mathcal{DT} training while the second parts form the test sample which
 1885 will be used for testing the final classifier obtained from the training. Each event has
 1886 associated a set of input variables $\mathbf{x} = (x_1, \dots, x_n)$ which serve to distinguish between
 1887 signal and background events. Pick one variable, say x_1 , and for each event value of
 1888 x_1 split the training sample into two sets, then, find the splitting value that provides
 1889 the best classification, i.e., one set mostly made signal events while the other set is
 1890 mostly made of background events.

1891 Then repeat this for each variable in turn. Select the variable and splitting value
 1892 which gives the best separation. Initially there was a sample of events at a $\text{ode}\text{Jn-}$
 1893 o . Now there are two samples called Jbranches . For each branch, repeat
 1894 the process, i.e., again try each value of each variable for the events within that branch
 1895 to find the best variable and splitting point for that branch. One keeps splitting un-til
 1896 a given number of final branches, called leaves, are obtained, or until each leaf is pure

1897 signal or pure background, or has too few events to continue. This description is a
 1898 little oversimplified. In fact at each stage one picks as the next branch to split, the
 1899 branch which will give the best increase in the quality of the separation. A schematic
 1900 of a decision tree is shown in Fig.1, in which 3 variables are used for signal/background
 1901 separation: event hit multiplicity, energy, and reconstructed radial position. What
 1902 criterion is used to define the quality of separation between signal and background in
 1903 the split? Imagine the events are weighted with each event having weight W_i . Define
 1904 the purity of the sample in a branch by

1905 where P_s is the sum over signal events and P_b is the sum over background events.
 1906 Note that $P(1 - \sqrt{P})$ is 0 if the sample is pure signal or pure background. For a
 1907 given branch let

1908 A decision trees (DT) is a consist of consecutive binary questions based on a set
 1909 of input variables. Depending on the previous answer a different decision has to be
 1910 taken on the next

1911 node. After a maximal number of nodes, a final verdict is reached depending
 1912 on the majority of event class in the occupied endpoint, the leaf. In the training of
 1913 Decision Trees the best possible selection criterion for a node is defined by maximizing
 1914 the separation gain S between two consecutive nodes. In order to determine the
 1915 separation gain, the purity of each leaf has to be determined as

1916 $P_s w_s P = P_s^2 + P_b w_b P$ where s, b are the summed weights of the
 1917 signal and background events, respectively 1. The purity allows for the determination
 1918 of the Gini index G , a measure of statistical dispersion, as $n \times G = \sum_{i=1}^n w_i (1 - \sqrt{P_i})$,
 1919 where w_i are the normalized weights for each event. The best possible
 1920 question for the node can then be determined by maximizing S , given by $S = G_{\text{father}} - G_{\text{child I}} - G_{\text{child II}}$. As the Gini index has a maximum at $P = 0.5$
 1921 of $G = 0.25$ (for unweighted events) the maximal separation gain would also be $S =$

1923 0.25 for a maximally disperse starting ensemble ($G_{\text{father}} = 0.25$) and two perfectly
 1924 separated child nodes ($G_{\text{child}} = 0$). Based on the purity after the training, leafs
 1925 are either labeled as signal or background leaves. By using a multitude of slightly
 1926 altered decision trees and averaging over the predicted outcome the robustness of
 1927 the procedure can be ensured, as a badly trained tree due to aberrations does not
 1928 have enough power to sway the decision into the wrong direction. The average of all
 1929 classifier outputs is later used as a final discrimination variable, labeled as \hat{A}_{IBDT}
 1930 output.

1931 By using the boosting method the full potential of decision trees can be exploited.
 1932 By subsequently training a multitude of trees, where in each iteration falsely catego-
 1933 rized events get assigned a higher weight, the focus of the training is shifted to classify
 1934 these previously misiden- tified events correctly. In this thesis the AdaBoost [124] al-
 1935 gorithm is utilized. Misclassified events are assigned a boost weight \hat{w}_i for each tree i
 1936 which is dependent on the misclassification rate \hat{S}_i of tree i in the previous training
 1937 iteration: $1 - \hat{S}_i = \hat{w}_i$. Correctly classified events get reweighted such that
 1938 the sum of weights remains constant. The BDT output of a boosted ensemble of N
 1939 decision trees is determined as $y_{\text{Boost}}(x) = N^{-1} \sum_i \ln \hat{w}_i h_i(x)$,
 1940 where $h_i(x)$ is the binary outcome of tree i , mathematically realized as $h_i(x) = 1$
 1941 for events ending on a signal leaf and $h_i(x) = -1$ for events ending on a background
 1942 leaf.

1943 5.2

1944 Notice that the tails of the distributions indicate that some signal events fall on the
 1945 rejection region and some background events fall on the acceptance region; there-
 1946 fore, it is convenient to define the “efficiency” with which events of a given type are

1947 accepted, thus, the signal and background efficiencies are given by

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

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

1948 where A is the acceptance region. Under these conditions, the background hy-
 1949 pothesis corresponds to the “null hypothesis (H_0)”, the signal hypothesis corresponds
 1950 to the “alternative hypothesis (H_1)”, the background efficiency is the significance level
 1951 of the test, and signal efficiency is the power of the test; what is sought in an analysis
 1952 is to maximize the power of the test relative to the significance level.

1953 ; it is achieved, according to the Neyman-Pearson lemma [109],

1954 by defining the acceptance region such that, for \mathbf{x} inside the region, the likelihood
 1955 ratio, i.e., the ratio of probability distribution functions for signal and background,

- ¹⁹⁵⁶ **5.3 MVA methods, NN, BDT, boosting,**
¹⁹⁵⁷ **overtraining, variable ranking**
- ¹⁹⁵⁸ **5.4 statistical inference, likelihood**
¹⁹⁵⁹ **parametrization**
- ¹⁹⁶⁰ **5.5 nuisance parameters**
- ¹⁹⁶¹ **5.6 exclusion limits**
- ¹⁹⁶² **5.7 asymptotic limits**

¹⁹⁶³ **Chapter 6**

¹⁹⁶⁴ **Search for production of a Higgs**
¹⁹⁶⁵ **boson and a single top quark in**
¹⁹⁶⁶ **multilepton final states in pp**
¹⁹⁶⁷ **collisions at $\sqrt{s} = 13$ TeV**

¹⁹⁶⁸ **6.1 Introduction**

¹⁹⁶⁹ Dont forget to mention previous constrains to ct check reference ?? and references
¹⁹⁷⁰ <https://link.springer.com/content/pdf/10.1007%2FJHEP01>
¹⁹⁷¹ A. Azatov, R. Contino and J. Galloway, "Model-Independent Bounds on a
¹⁹⁷² Light Higgs," JHEP 1204 (2012) 127 [arXiv:1202.3415 [hep-ph]].
¹⁹⁷³ J. R. Espinosa, C. Grojean, M. Muhlleitner and M. Trott, "Fingerprinting
¹⁹⁷⁴ Higgs Suspects at the LHC," JHEP 1205 (2012) 097 [arXiv:1202.3697 [hep-ph]].
¹⁹⁷⁵ This chapter present the search for the associated production of a Higgs boson and
¹⁹⁷⁶ a single top quark events with three leptons in the final state, targeting Higgs decay

1977 modes to WW , ZZ , and $\tau\tau$. The analysis uses the 13 TeV dataset produced in 2016,
 1978 corresponding to an integrated luminosity of 35.9fb^{-1} . It is based on and expands
 1979 previous analyses at 8 TeV [?, ?] and searches for associated production of $t\bar{t}$ and
 1980 Higgs in the same channel [?], and complements searches in other decay channels
 1981 targeting $H \rightarrow b\bar{b}$ [?].
 1982 As showed in section 2.4, the cross section of the associated production of a Higgs
 1983 boson and a single top quark (tHq) process is driven by a destructive interference of
 1984 two contributions (see Figure 6.1), where the Higgs couples to either the W boson or
 1985 the top quark. Any deviation from the standard model (SM) in the Higgs coupling
 1986 structure could therefore lead to a large enhancement of the cross section, making
 1987 this analysis sensitive to such deviations. A second process, where the Higgs and
 1988 top quark are accompanied by a W boson (tHW) has similar behavior, albeit with a
 1989 weaker interference pattern.

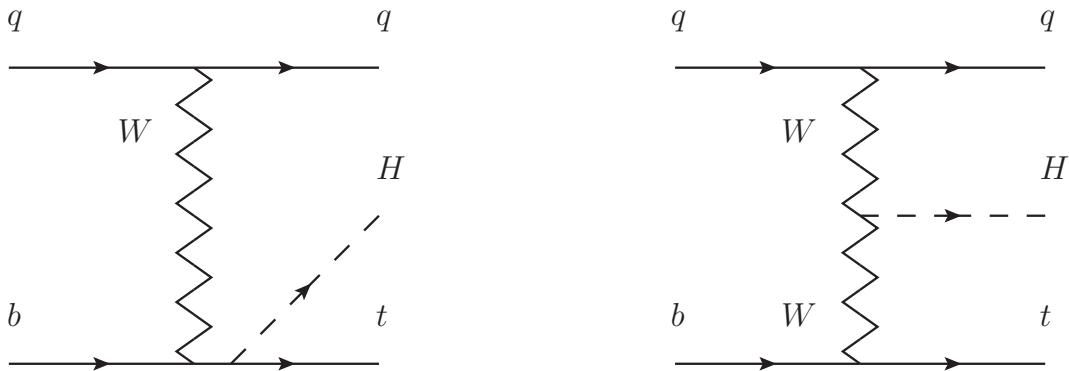


Figure 6.1: The two leading-order diagrams of tHq production.

1990 We selects events with three leptons and a b tagged jet in the final state. The tHq
 1991 signal contribution is then determined in a fit of the observed data to two multivariate
 1992 classifier outputs, each trained to discriminate against one of the two dominant back-
 1993 grounds of events with non-prompt leptons from $t\bar{t}$ and of associated production of $t\bar{t}$

1994 and vector bosons ($t\bar{t}W$, $t\bar{t}Z$). The fit result is then used to set an upper limit on the
 1995 combined $t\bar{t}H$, tHq and tHW production cross section, as a function of the relative
 1996 coupling strengths of Higgs and top quark and Higgs and W boson, respectively.

1997 6.2 Data and MC Samples

1998 The data considered in this analysis were collected by the CMS experiment dur-
 1999 ing 2016 and correspond to a total integrated luminosity of 35.9fb^{-1} . Only periods
 2000 when the CMS magnet was on were considered when selecting the data samples, that
 2001 corresponds to the 23 Sep 2016 (Run B to G) and PromptReco (Run H) versions
 2002 of the datasets. The MC samples used in this analysis correspond to the RunI-
 2003 ISummer16MiniAODv2 campaign produced with CMSSW 80X. The two signal sam-
 2004 ples (for tHq and tHW) were produced with MG5_aMC@NLO (version 5.222), in
 2005 leading-order mode, and are normalized to next-to-leading-order cross sections,
 2006 see Tab. 6.1. Each sample is generated with a set of event weights corresponding to
 2007 different values of κ_t and κ_V couplings as shown in Tab. 6.2.

2008 6.2.1 Full 2016 dataset and MC samples

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. [?] for more details.

2009 Different MC generators were used to generate the background processes. The
 2010 dominant sources ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$) were produced using AMC@NLO interfaced
 2011 to PYTHIA8, and are scaled to NLO cross sections. Other processes are simulated

		<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. [?].

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_L0_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.

2012 using POWHEG interfaced to PYTHIA, or bare PYTHIA (see table 6.3 and [?] for
2013 more details).

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_WPLoose_Gsf_v*

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

2014 6.2.2 Triggers

2015 We consider online-reconstructed events triggered by one, two, or three leptons.
 2016 Single-lepton triggers are included to boost the acceptance of events where the p_T of
 2017 the sub-leading lepton falls below the threshold of the double-lepton triggers. Ad-
 2018 ditionally, by including double-lepton triggers in the ≥ 3 lepton category, as well
 2019 as single-lepton triggers in all categories, we increase the efficiency, considering the
 2020 logical “or” of the trigger decisions of all the individual triggers in a given category.
 2021 Tab. 6.5 shows the lowest-threshold non-prescaled triggers present in the High-Level
 2022 Trigger (HLT) menus for both Monte-Carlo and data in 2016.

2023 6.2.2.1 Trigger efficiency scale factors

2024 The efficiency of events to pass the trigger is measured in simulation (trivially using
 2025 generator information) and in the data (using event collected by an uncorrelated

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.

2026 MET trigger). Small differences between the data and MC efficiencies are corrected
 2027 by applying scale factors as shown in Tab. 6.6. The exact procedure and control plots
 2028 are documented in [?] for the current analysis.

2029 6.3 Object Identification and event selection

2030 6.3.1 Jets and b tagging

2031 The analysis uses anti- k_t (0.4) particle-flow (PF) jets, corrected for charged hadrons
 2032 not coming from the primary vertex (charged hadron subtraction), and having jet
 2033 energy corrections (`Summer16_23Sep2016V3`) applied as a function of the jet E_T and
 2034 η . Jets are only considered if they have a transverse energy above 25GeV.

2035 In addition, they are required to be separated from any lepton candidates passing
 2036 the fakeable object selections (see Tables 6.7 and 6.8) by $\Delta R > 0.4$.

2037 The loose and medium working points of the CSV b-tagging algorithm are used to
 2038 identify b jets. Data/simulation differences in the b tagging performance are corrected
 2039 by applying per-jet weights to the simulation, dependent on the jet p_T , eta, b tagging
 2040 discriminator, and flavor (from simulation truth) [?]. The per-event weight is taken
 2041 as the product of the per-jet weights, including those of the jets associated to the
 2042 leptons. More details can be found in the corresponding $t\bar{t}H$ documentation [?, ?].

2043 **6.3.2 Lepton selection**

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

Table 6.7: 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. For the two p_T^{ratio} and CSV rows, the cuts marked with a † 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.

2044 The lepton reconstruction and selection is identical to that used in the $t\bar{t}H$ mul-
 2045 tilepton analysis, as documented in Refs. [?, ?]. For details on the reconstruction
 2046 algorithms, isolation, pileup mitigation, and a description of the lepton MVA dis-
 2047 criminator and validation plots thereof, we refer to that document since they are out
 2048 of the scope of this thesis. Three different selections are defined both for the electron
 2049 and muon object identification: the *Loose*, *Fakeable Object*, and *Tight* selection. As
 2050 described in more detail later, these are used for event level vetoes, the fake rate
 2051 estimation application region, and the final signal selection, respectively. The p_T of
 2052 fakeable objects is defined as $0.85 \times p_T(\text{jet})$, where the jet is the one associated to the
 2053 lepton object. This mitigates the dependence of the fake rate on the momentum of
 2054 the fakeable object and thereby improves the precision of the method.

2055 Tables 6.7 and 6.8 list the full criteria for the different selections of muons and
 2056 electrons.

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

Table 6.8: 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).

2057 6.3.3 Lepton selection efficiency

2058 Efficiencies of reconstruction and selecting loose leptons are measured both for muons
 2059 and electrons using a tag and probe method on both data and MC, using $Z \rightarrow \ell^+ \ell^-$.
 2060 Corresponding scale factors are derived from the ratio of efficiencies and applied to the
 2061 selected These. Events are produced for the leptonic SUSY analyses using equivalent
 2062 lepton selections and recycled for the $t\bar{t}H$ analysis as well as for this analysis. The
 2063 efficiencies of applying the tight selection as defined in Tables 6.7 and 6.8, on the
 2064 loose leptons are determined again by using a tag and probe method on a sample of
 2065 DY-enriched events. They are documented for the $t\bar{t}H$ analysis in Ref. [?] and are
 2066 exactly equivalent for this analysis.

2067 6.4 Background predictions

2068 The modeling of reducible and irreducible backgrounds in this analysis uses the exact
 2069 methods, analysis code, and ROOT trees used for the $t\bar{t}H$ multilepton analysis. We
 2070 give a brief description of the methods and refer to the documentation of that analysis
 2071 in Refs. [?, ?] for any details.

2072 The backgrounds in three-lepton final states can be split in two broad categories:
 2073 irreducible backgrounds with genuine prompt leptons (i.e. from on-shell W and Z
 2074 boson decays); and reducible backgrounds where at least one of the leptons is “non-
 2075 prompt”, i.e. produced within a hadronic jet, either a genuine lepton from heavy
 2076 flavor decays, or simply mis-reconstructed jets.

2077 Irreducible backgrounds can be reliably estimated directly from Monte-Carlo sim-
 2078 ulated events, using higher-order cross sections or data control regions for the overall
 2079 normalization. This is done in this analysis for all backgrounds involving prompt lep-
 2080 tons: $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, WZ , ZZ , $W^\pm W^\pm qq$, $t\bar{t}t\bar{t}$, tZq , tZW , WWW , WWZ , WZZ ,
 2081 ZZZ .

2082 Reducible backgrounds, on the other hand, are not well predicted by simulation,
 2083 and are estimated using data-driven methods. In the case of non-prompt leptons, a
 2084 fake rate method is used, where the contribution to the final selection is estimated by
 2085 extrapolating from a sideband (or “application region”) with a looser lepton definition
 2086 (the fakeable object definitions in Tabs. 6.7 and 6.8) to the signal selection. The tight-
 2087 to-loose ratios (or “fake rates”) are measured in several background dominated data
 2088 events with dedicated triggers, subtracting the residual prompt lepton contribution
 2089 using MC. Non-prompt leptons in our signal regions are predominantly produced in $t\bar{t}$
 2090 events, with a much smaller contribution, from Drell–Yan production. The systematic
 2091 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 events, limited statistics in the data control regions due to necessary prescaling of lepton triggers, and the uncertainty in the subtraction of residual prompt leptons from the control region.

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

Figures 6.2 show the distributions of some relevant kinematic variables, normalized to the cross section of the respective processes and to the integrated luminosity.

6.5 Signal discrimination

The tHq signal is separated from the main backgrounds using a boosted decision tree (BDT) classifier, trained on simulated signal and background events. A set of discriminating variables are given as input to the BDT which produces a output distribution maximizing the discrimination power. Table 6.9 lists the input variables used while Figures 6.3 show their distributions for the relevant signal and background samples, for the three lepton channel. Two BDT classifiers are trained for the two main backgrounds expected in the analysis: events with prompt leptons from $t\bar{t}W$ and $t\bar{t}Z$ (also referred to as $t\bar{t}V$), and events with non-prompt leptons from $t\bar{t}$. The datasets used in the training are the tHq signal (see Tab. 6.1), and LO MADGRAPH samples

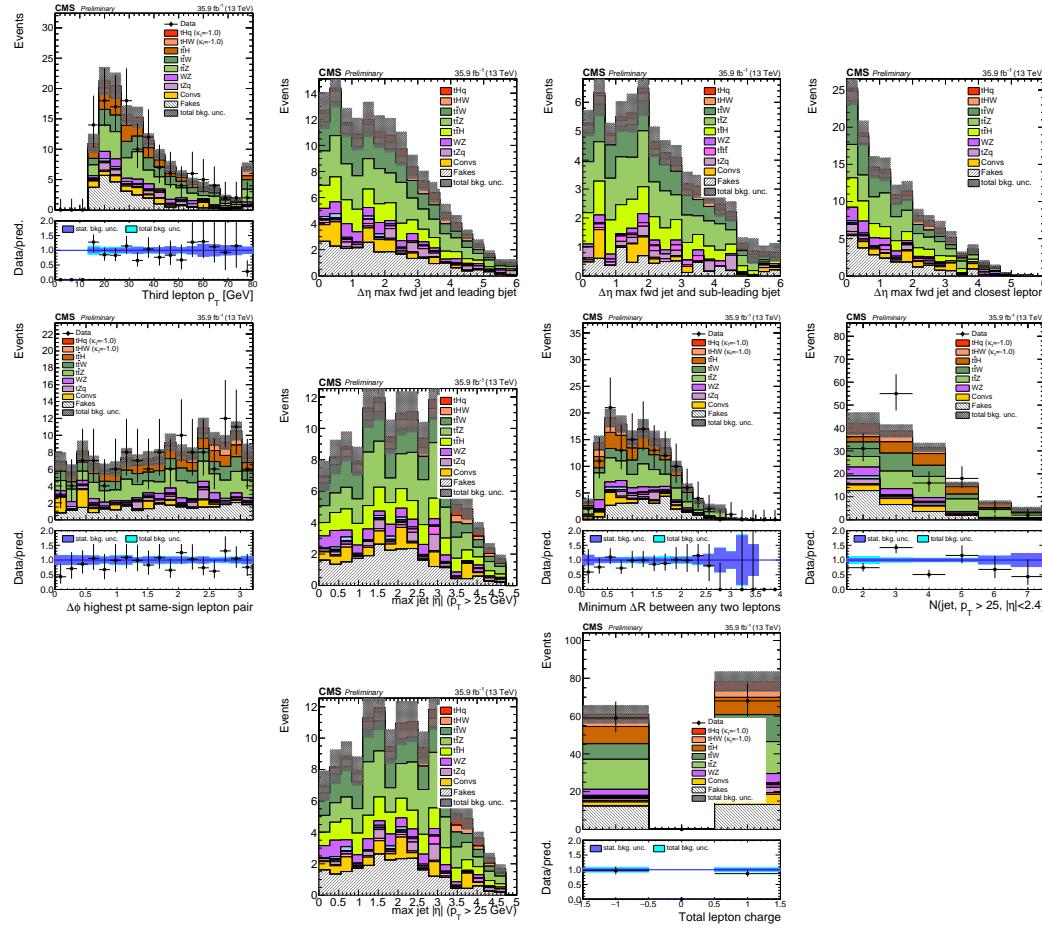


Figure 6.2: Distributions of input variables to the BDT for signal discrimination, three lepton channel, normalized to their cross section and to 35.9 fb^{-1} .

of $t\bar{t}W$ and $t\bar{t}Z$, in an admixture proportional to their respective cross sections (see Tab. 6.4).

The MVA analysis consist of two stages: first a “training” where the MVA method 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 6.4 show the input variables distributions as seen by the MVA algorithm. Note that in contrast to the distributions in Fig. 6.3 only the main backgrounds ($t\bar{t}$ from simulation, $t\bar{t}V$) are included.

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 6.9: MVA input discriminating variables

2125 Note that splitting the training in two groups reveals that some variables show
 2126 opposite behavior for the two background sources; potentially screening the discrimi-
 2127 nation power if they were to be used in a single discriminant. For some other variables
 2128 the distributions are similar in both background cases.

2129 From table 6.9, it is clear that the input variables are correlated to some extend.
 2130 These correlations play an important role for some MVA methods like the Fisher
 2131 discriminant method in which the first step consist of performing a linear transfor-
 2132 mation to an phase space where the correlations between variables are removed. In
 2133 case a boosted decision tree (BDT) method however, correlations do not affect the
 2134 performance. Figure 6.6 show the linear correlation coefficients for signal and back-
 2135 ground for the two training cases (the signal values are identical by construction). As
 2136 expected, strong correlations appears for variables related to the forward jet activity.
 2137 Same trend is seen in case of the same sign dilepton channel in Figure ??.

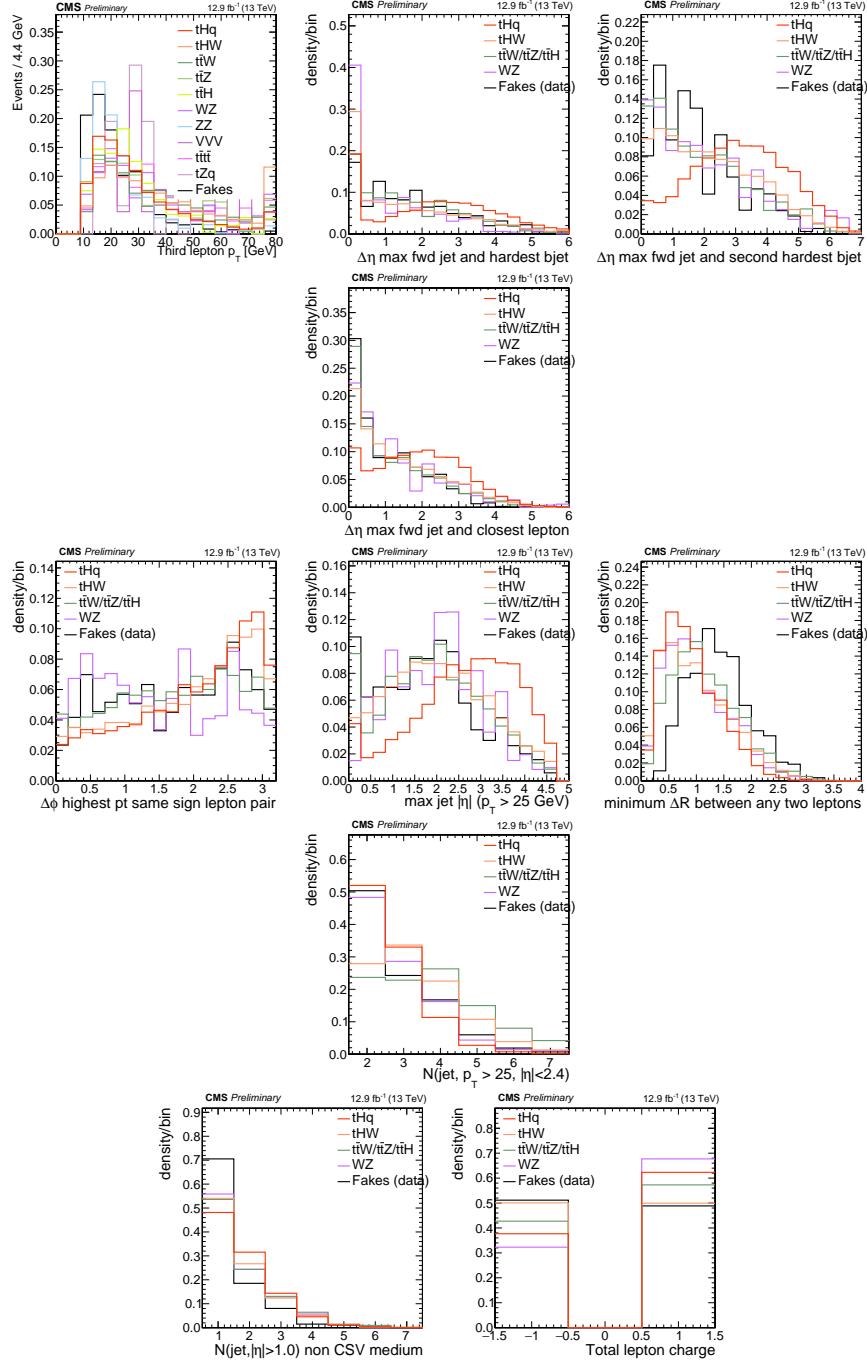


Figure 6.3: Distributions of input variables to the BDT for signal discrimination, three lepton channel.

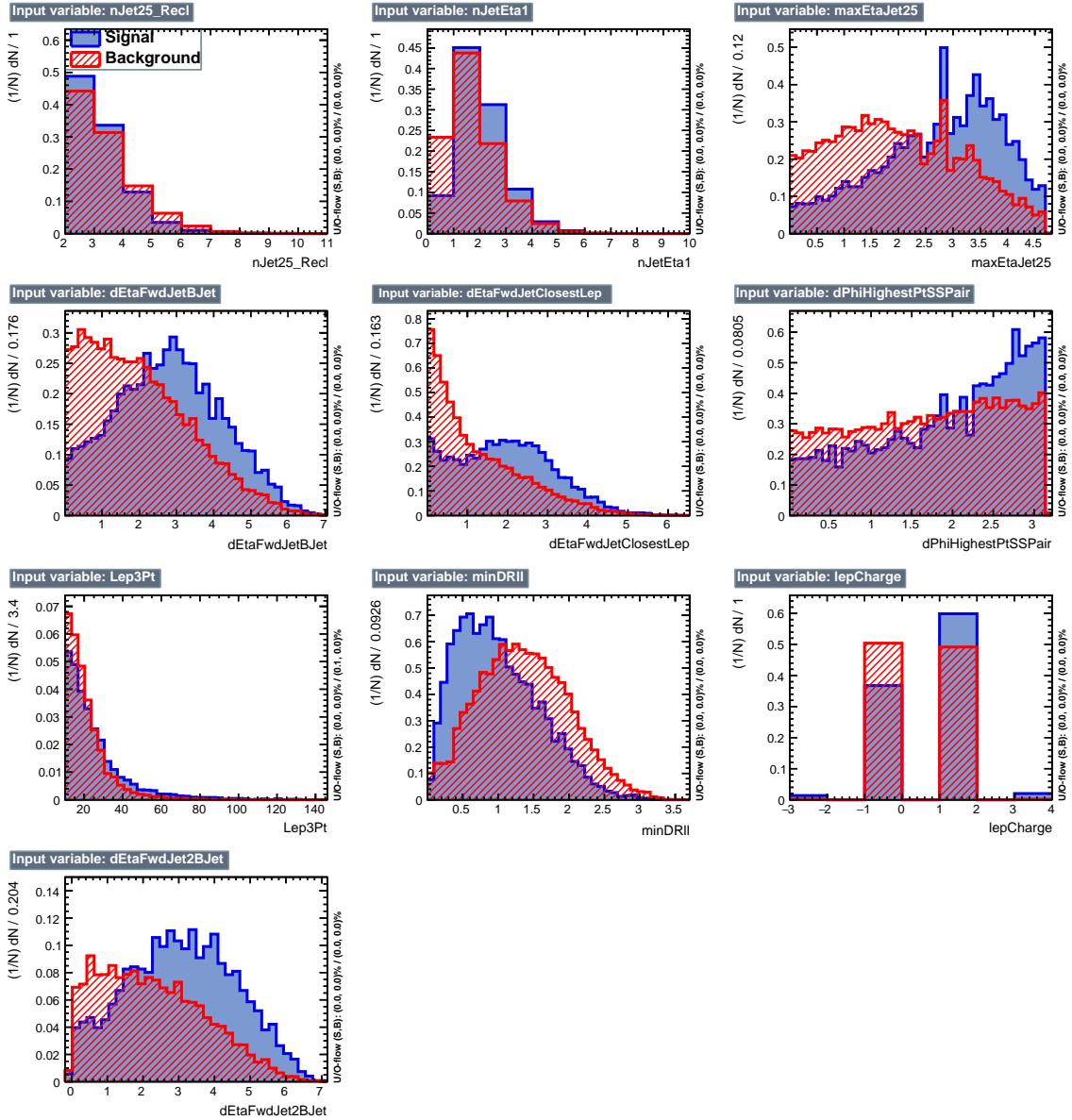


Figure 6.4: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}$) for the three lepton channel, discriminated against $t\bar{t}$ (fakes) background.

2138 6.5.1 Classifiers response

2139 Several MVA algorithms were evaluated to determine the most appropriate method
 2140 for this analysis. The plots in Fig. 6.7 (top) show the background rejection as a
 2141 function of the signal efficiency for $t\bar{t}$ and $t\bar{t}V$ trainings (ROC curves) for the different

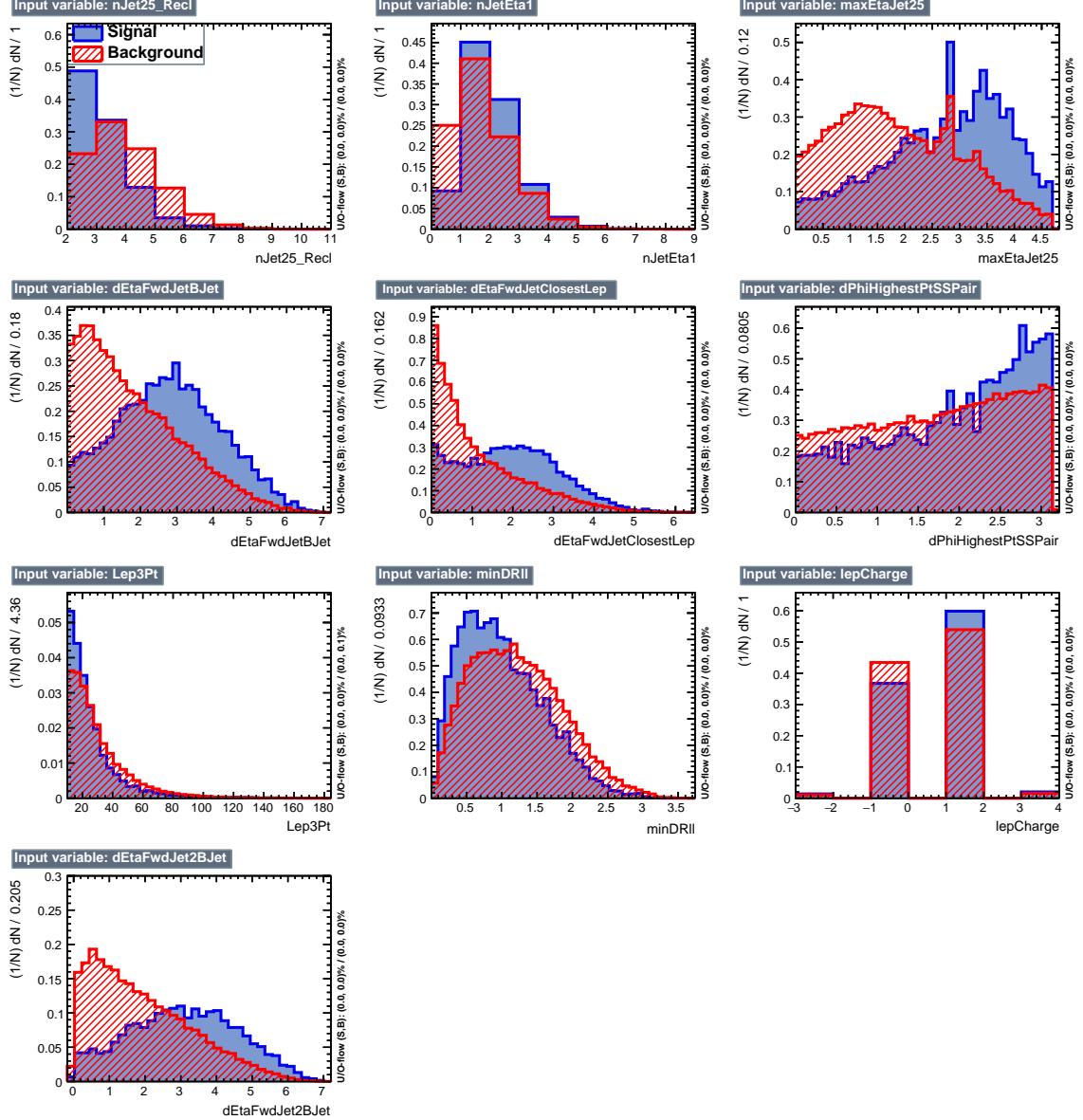


Figure 6.5: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}W + t\bar{t}Z$) for the three lepton channel, discriminated against $t\bar{t}V$ background.

algorithms that were evaluated.

In both cases the gradient boosted decision tree (“BDTA_GRAD”) classifier offers 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. 6.7. As expected, a good discrimination power is obtained using default discrim-

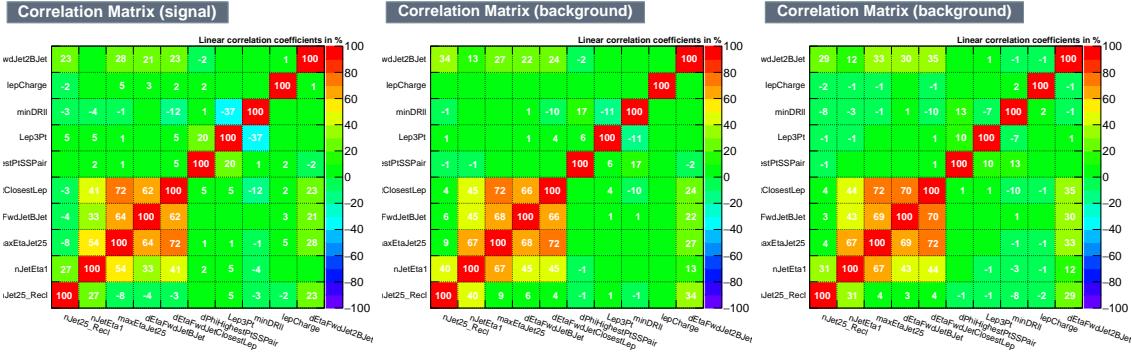


Figure 6.6: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the three lepton channel.

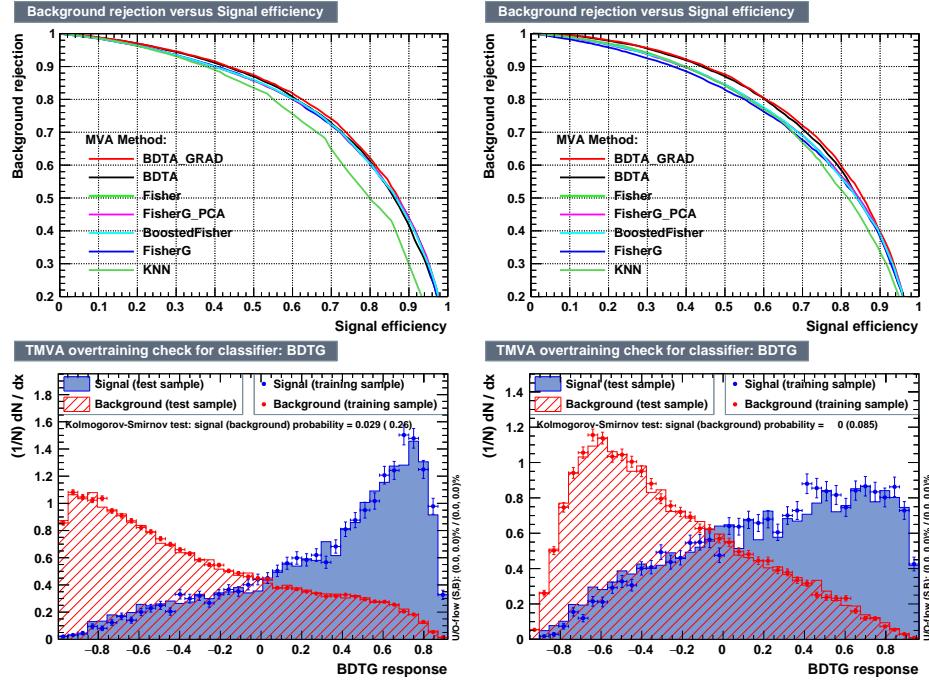


Figure 6.7: Top: background rejection vs signal efficiency (ROC curves) for various MVA classifiers (top) in the three lepton channel against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distributions for the gradient boosted decision trees, for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

2147 inator parameter values, with minimal overtraining. TMVA provides a ranking of the
 2148 input variables by their importance in the classification process, shown in Tab. 6.10.
 2149 The TMVA settings used in the BDT training are shown in Tab. 6.11.

ttbar training			ttV training		
Rank	Variable	Importance	Variable	Importance	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 6.10: 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 6.11: TMVA configuration used in the BDT training.

2150 6.6 Additional discriminating variables

2151 Two additional discriminating variables were tested considering the fact that the
 2152 forward jet in the background could come from the pileup; since we have a real
 2153 forward jet in the signal, it could give some improvement in the discriminating power.
 2154 The additional variables describe the forward jet momentum (fwdJetPt25) and the
 2155 forward jet identification(fwdJetPUID). Distributions for these variables in the three
 2156 lepton channel are shown in the figure 6.8. The forward jet identification distribution
 2157 show that for both, signal and background, jets are mostly real jets.

2158 The testing was made including in the MVA input one variable at a time, so we

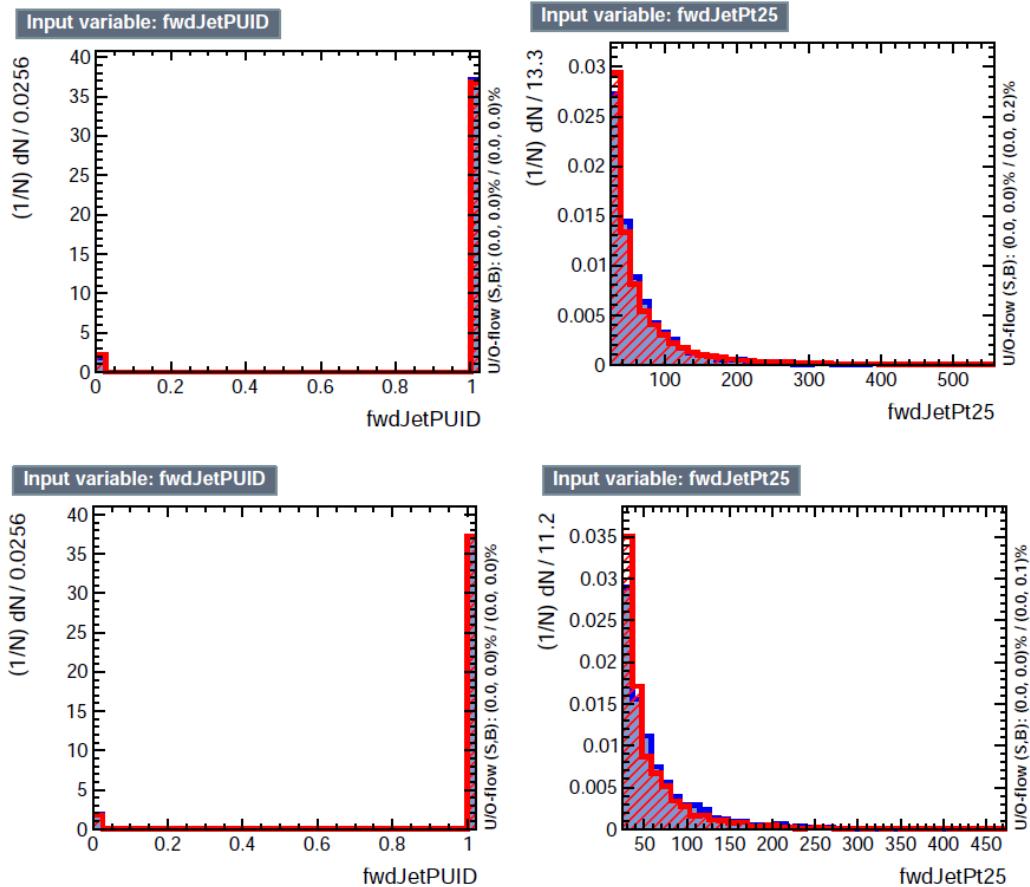


Figure 6.8: Additional discriminating variables distributions for ttV training (Top row) and tt training (bottom row) in the three lepton channel. The origin of the jets in the forward jet identification distribution is tagged as 0 for “pileup jets” while “real jets” are tagged as 1.

2159 can evaluate the discrimination power of each variable, and then both simultaneously.
 2160 fwdJetPUID was ranked in the last place in importance (11) in both training (ttV
 2161 and tt) while fwdJetPt25 was ranked 3 in the ttV training and 7 in the tt training.
 2162 When training using 12 variables, fwdJetPt25 was ranked 5 and 7 in the ttV and tt
 2163 trainings respectively, while fwdJetPUID was ranked 12 in both cases.

2164 The improvement in the discrimination performance provided by the additional
 2165 variables is about 1%, so it was decided not to include them in the procedure. Table
 2166 6.12 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 6.12: 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%

²¹⁶⁷ References

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

- 2185 [9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)
2186 and 2017 update.
- 2187 [10] M. Goldhaber, L. Grodzins, A.W. Sunyar “Helicity of Neutrinos”, Phys. Rev.
2188 109, 1015 (1958).
- 2189 [11] Palanque-Delabrouille N et al. “Neutrino masses and cosmology with Lyman-
2190 alpha forest power spectrum”, JCAP 11 011 (2015).
- 2191 [12] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. Physics Letters.
2192 8 (3): 214-215 (1964).
- 2193 [13] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking”
2194 (PDF). CERN Report No.8182/TH.401 (1964).
- 2195 [14] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking:
2196 II” (PDF). CERN Report No.8419/TH.412(1964).
- 2197 [15] M. Gell-Mann. “The Interpretation of the New Particles as Displaced Charged
2198 Multiplets”. Il Nuovo Cimento 4: 848. (1956).
- 2199 [16] T. Nakano, K, Nishijima. “Charge Independence for V-particles”. Progress of
2200 Theoretical Physics 10 (5): 581-582. (1953).
- 2201 [17] N. Cabibbo, “Unitary symmetry and leptonic decays” Physical Review Letters,
2202 vol. 10, no. 12, p. 531, (1963).
- 2203 [18] M.Kobayashi, T.Maskawa, “CP-violation in the renormalizable theory of weak
2204 interaction,” Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).
- 2205 [19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Com-
2206 mons, the free media repository. Retrieved November 27, 2017 from

- 2207 https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)
2208 .svg&oldid=247498592.
- 2209 [20] Georgia Tech University. Coupling Constants for the Fundamental
2210 Forces(2005). Retrieved January 10, 2018, from http://hyperphysics.phy-
2211 astr.gsu.edu/hbase/Forces/couple.html#c2
- 2212 [21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved Jan-
2213 uary 10, 2018, from https://profmattstrassler.com/articles-and-posts/particle-
2214 physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/
- 2215 [22] S.L. Glashow. “Partial symmetries of weak interactions”, Nucl. Phys. 22 579-
2216 588, (1961).
- 2217 [23] A. Salam, J.C. Ward. “Electromagnetic and weak interactions”, Physics Letters
2218 13 168-171, (1964).
- 2219 [24] S. Weinberg, “A model of leptons”, Physical Review Letters, vol. 19, no. 21, p.
2220 1264, (1967).
- 2221 [25] M. Peskin, D. Schroeder, “An introduction to quantum field theory”. Perseus
2222 Books Publishing L.L.C., (1995).
- 2223 [26] A. Pich. “The Standard Model of Electroweak Interactions”
2224 https://arxiv.org/abs/1201.0537
- 2225 [27] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved
2226 from: https://www.quantum-bits.org/?p=233
- 2227 [28] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255

- 2228 [29] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge
2229 Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964)
2230 doi:10.1103/PhysRevLett.13.321
- 2231 [30] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Re-
2232 view Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.
- 2233 [31] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws
2234 and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964).
2235 doi:10.1103/PhysRevLett.13.585.
- 2236 [32] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with
2237 the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012).
2238 arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021
- 2239 [33] ATLAS collaboration. “Observation of a New Particle in the Search for the Stan-
2240 dard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters
2241 B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.
- 2242 [34] ATLAS collaboration; CMS collaboration (26 March 2015). “Combined Mea-
2243 surement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with
2244 the ATLAS and CMS Experiments”. Physical Review Letters. 114 (19): 191803.
2245 arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.
- 2246 [35] LHC InternationalMasterclasses“When protons collide”. Retrieved from http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm
2247
- 2248 [36] CMS Collaboration, “SM Higgs Branching Ratios and Total Decay Widths (up-
2249 date in CERN Report4 2016)”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR> , last accessed on 17.12.2017.
- 2250

- 2251 [37] R.Grant V. “Determination of Higgs branching ratios in $H \rightarrow W^+W^- \rightarrow l\nu jj$
 2252 and $H \rightarrow ZZ \rightarrow l^+l^-jj$ channels”. Physics Department, University of Tennessee
 2253 (Dated: October 31, 2012). Retrieved from <http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf>
- 2255 [38] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al.
 2256 “Standard model Higgs-boson branching ratios with uncertainties”. Eur. Phys.
 2257 J. C (2011) 71: 1753. <https://doi.org/10.1140/epjc/s10052-011-1753-8>
- 2258 [39] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production
 2259 of Higgs and single top at hadron colliders”, Phys.Rev. D64 (2001) 094023,
 2260 [hep-ph/0106293].
- 2261 [40] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the
 2262 top-Higgs coupling from the 8 TeV LHC data,” Journal of High Energy Physics,
 2263 vol. 07, p. 073, (2013).
- 2264 [41] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting de-
 2265 generacies in Higgs couplings using single top production in association with a
 2266 Higgs boson,” Journal of High Energy Physics, vol. 05, p. 022, (2013).
- 2267 [42] T.M. Tait and C.-P. Yuan, “Single top quark production as a window to physics
 2268 beyond the standard model”, Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].
- 2269 [43] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in
 2270 association with a single top quark at the LHC,” European Physical Journal C,
 2271 vol. 75, p. 267, (2015).

- 2272 [44] CMS Collaboration, “Modelling of the single top-quark production in associa-
2273 tion with the Higgs boson at 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV)
2274 [viewauth/CMS/SingleTopHiggsGeneration13TeV](#), last accessed on 16.01.2018.
- 2275 [45] CMS Collaboration, “SM Higgs production cross sections at $\sqrt{s} =$
2276 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)
2277 [view/LHCPhysics/CERNYellowReportPageAt13TeV](#), last accessed on 16.01.2018.
- 2278 [46] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.
- 2279 [47] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [[arXiv:1211.0499 \[hep-ph\]](#)].
- 2281 [48] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated
2282 production at the LHC”, European Physical Journal C, vol. 77, p. 34, (2017).
2283 [arXiv:1607.05862](#)
- 2284 [49] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross
2285 Sections: 4.Deciphering the Nature of the Higgs Sector”, [arXiv:1610.07922](#).
- 2286 [50] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top
2287 Couplings in Associated Production”, JHEP 1404 (2014) 004, [[arXiv:1312.5736](#)].
- 2288 [51] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass
2289 of the Higgs boson and tests of compatibility of its couplings with the standard
2290 model predictions using proton collisions at 7 and 8 TeV,” [arXiv:1412.8662](#).
- 2291 [52] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the
2292 Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton
2293 collision data”, ATLAS-CONF-2014-009.

- 2294 [53] ATLAS and CMS Collaborations, “Measurements of the Higgs boson produc-
2295 tion and decay rates and constraints on its couplings from a combined ATLAS
2296 and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,” (2016).
2297 CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.
- 2298 [54] File:Cern-accelerator-complex.svg. Wikimedia Commons,
2299 the free media repository. Retrieved January, 2018 from
2300 <https://commons.wikimedia.org/wiki/File:Cern-accelerator-complex.svg>
- 2301 [55] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”,
2302 (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Re-
2303 trieval from <https://cds.cern.ch/record/841542>
- 2304 [56] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-
2305 2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762
- 2306 [57] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).
- 2307 [58] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March
2308 1983, CERN-AC-8303511. Retrieved from <https://cds.cern.ch/record/615852>.
- 2309 [59] CERN Photographic Service “Animation of CERN’s accelerator net-
2310 work”, 14 October 2013. DOI: 10.17181/cds.1610170 Retrieved from
2311 <https://videos.cern.ch/record/1610170>
- 2312 [60] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17, 2013. Re-
2313 trieval from <https://www.britannica.com/technology/particle-accelerator>.
- 2314 [61] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-
2315 AC-0007016. Retrieved from <https://cds.cern.ch/record/41567>.

- 2316 [62] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”.
 2317 March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-
 2318 PHO-1998-325. Retrieved from <https://cds.cern.ch/record/841511>
- 2319 [63] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025
 2320 retrieved from <https://cds.cern.ch/record/40524>.
- 2321 [64] CMS Collaboration “Public CMS Luminosity Information”.
 2322 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016>
 2323 _proton_proton_13_TeV_collis, last accessed 24.01.2018
- 2324 [65] J.L. Caron. “LHC Layout” AC Collection. Legacy of AC. Pictures
 2325 from 1992 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from
 2326 <https://cds.cern.ch/record/841573>.
- 2327 [66] J.A. Coarasa. “The CMS Online Cluster:Setup, Operation and Maintenance
 2328 of an Evolving Cluster”. ISGC 2012, 26 February - 2 March 2012, Academia
 2329 Sinica, Taipei, Taiwan.
- 2330 [67] CMS Collaboration. “The CMS experiment at the CERN LHC” JINST 3 S08004
 2331 (2008).
- 2332 [68] CMS Collaboration. “CMS detector drawings 2012” CMS-PHO-GEN-2012-002.
 2333 Retrieved from <http://cds.cern.ch/record/1433717>.
- 2334 [69] R. Breedon. “View through the CMS detector during the cooldown of the
 2335 solenoid on February 2006. CMS Collection”, February 2006, CMS-PHO-
 2336 OREACH-2005-004, Retrieved from <https://cds.cern.ch/record/930094>.
- 2337 [70] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector
 2338 Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.

- 2339 [71] CMS Collaboration. “Description and performance of track and primary-vertex
2340 reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no.
2341 10, p. P10009,(2014).
- 2342 [72] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner
2343 Barrel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from
2344 <https://cds.cern.ch/record/1431467>.
- 2345 [73] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm
2346 and beauty hadrons Chicago, June 28-July 3 2004.
- 2347 [74] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at
2348 the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013.
2349 arXiv:1307.7135
- 2350 [75] L. Veillet. “End assembly of HB with EB rails and rotation in-
2351 side SX ”,January 2002. CMS-PHO-HCAL-2002-002. Retrieved from
2352 <https://cds.cern.ch/record/42594>.
- 2353 [76] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the
2354 UX5 cavern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from
2355 <https://cds.cern.ch/record/1019185>
- 2356 [77] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web
2357 site: Taking a closer look at LHC. Retrieved from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger
- 2359 [78] WLCG Project Office, “Documents & Reference - Tiers - Structure,” (2014).
2360 <http://wlcg.web.cern.ch/documents-reference> , last accessed on 30.01.2018.

- 2361 [79] CMS Collaboration. “CMSSW Application Framework”,
 2362 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>,
 2363 last accesses 06.02.2018
- 2364 [80] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event gen-
 2365 erators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011
- 2366 [81] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics
 2367 of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-
 2368 3-540-71060-8 (2007)
- 2369 [82] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter.”
 2370 <http://hepdata.cedar.ac.uk/pdf/pdf3.html> , last accessed on 02.02.2018.
- 2371 [83] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a
 2372 reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).
- 2373 [84] B. Andersson, G. Gustafson, G.Ingelman and T. Sjostrand, “Parton fragmen-
 2374 tation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145,
 2375 1983.
- 2376 [85] CMS Collaboration, “Event generator tunes obtained from underlying event
 2377 and multiparton scattering measurements;” European Physical Journal C, vol.
 2378 76, no. 3, p. 155, (2016).
- 2379 [86] J. Alwall et. al., “The automated computation of tree-level and next-to-leading
 2380 order differential cross sections, and their matching to parton shower simula-
 2381 tions,” Journal of High Energy Physics, vol. 07, p. 079, (2014).

- 2382 [87] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with
2383 Parton Shower simulations: the POWHEG method,” Journal of High Energy
2384 Physics, vol. 11, p. 070, (2007).
- 2385 [88] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments
2386 and Methods in Physics, vol. A506, pp. 250–303, (2003).
- 2387 [89] J.Allison et.al.,“Recent developments in Geant4”, Nuclear Instruments and
2388 Methods in Physics Research A 835 (2016) 186-225.
- 2389 [90] CMS Collaboration “Full Simulation Offline Guide”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation>, last accessed 04.02.2018
- 2390
- 2391 [91] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf.
2392 Ser. 513 022012 (2014)
- 2393 [92] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description
2394 with the CMS detector”, JINST 12 P10003 (2017) <https://doi.org/10.1088/1748-0221/12/10/P10003>.
- 2395
- 2396 [93] The CMS Collaboration. “ Description and performance of track and pri-
2397 mary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014).
2398 doi:10.1088/1748-0221/9/10/P10009
- 2399 [94] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides.
2400 Retrieved from https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf
- 2401
- 2402 [95] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by
2403 the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).

- 2404 [96] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of
 2405 electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf
 2406 C 0303241 (2003) TULT009 [physics/0306087].
- 2407 [97] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification,
 2408 Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.
- 2409 [98] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”,
 2410 CMS Note 2007-008 (2007).
- 2411 [99] CMS collaboration, “Performance of CMS muon reconstruction in pp collision
 2412 events at $\sqrt{s} = 7$ TeV ”, JINST 7 P10002 2012, [arXiv:1206.4071].
- 2413 [100] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,”
 2414 Journal of High Energy Physics, vol. 04, p. 063, (2008).
- 2415 [101] B. Dorney. “Anatomy of a Jet in CMS”. Quantum Diaries. June
 2416 1st, 2011. Retrieved from <https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>
- 2418 [102] The CMS Collaboration.“Event Displays from the high-energy collisions at 7
 2419 TeV”, May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from <https://cds.cern.ch/record/1429614>.
- 2421 [103] The CMS collaboration. “Determination of jet energy calibration and transverse
 2422 momentum resolution in CMS”. JINST 6 P11002 (2011). <http://dx.doi.org/10.1088/1748-0221/6/11/P11002>
- 2424 [104] The CMS Collaboration, “Introduction to Jet Energy Corrections at
 2425 CMS.”. <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>, last ac-
 2426 cessed 10.02.2018.

- 2427 [105] CMS Collaboration Collaboration. “Identification of b quark jets at the CMS
2428 Experiment in the LHC Run 2”. Tech. rep. CMS-PAS-BTV-15-001. Geneva:
2429 CERN, (2016). <https://cds.cern.ch/record/2138504>.
- 2430 [106] CMS Collaboration Collaboration. “Performance of missing energy reconstruc-
2431 tion in 13 TeV pp collision data using the CMS detector”. Tech. rep. CMS-PAS-
2432 JME16-004. Geneva: CERN, 2016. <https://cds.cern.ch/record/2205284>.
- 2433 [107] K. Skovpen. “Event displays highlighting the main properties of heavy flavour
2434 jets in the CMS Experiment”, Aug 2017, CMS-PHO-EVENTS-2017-006. Re-
2435 trieved from <https://cds.cern.ch/record/2280025>.
- 2436 [108] G. Cowan. “Topics in statistical data analysis for high-energy physics”.
2437 arXiv:1012.3589v1
- 2438 [109] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests
2439 of statistical hypotheses”. Philosophical Transactions of the Royal Society of
2440 London. Series A, Containing Papers of a Mathematical or Physical Character.
2441 Vol. 231 (1933), pp. 289-337