

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	4.4.1.1	Missing transverse energy.	103
65	4.4.2	Event reconstruction examples	104
66	5 Statistical methods		107
67	5.1	Multivariate analysis	107
68	5.1.1	Decision trees	110
69	5.1.2	Boosted decision trees (BDT).	113
70	5.1.3	Overtraining.	116
71	5.1.4	Variable ranking.	116
72	5.1.5	BDT output example.	117
73	5.2	Statistical inference.	118
74	5.2.1	Nuisance parameters.	118
75	5.2.2	Maximum likelihood estimation method	119
76	5.2.3	Hypothesis test	120
77	5.3	exclusion limits	121
78	5.4	asymptotic limits	121
79	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		122
81	6.1	Introduction	122
82	6.2	tHq signature	124
83	6.3	Background processes	125
84	6.4	Object Identification and event selection	127
85	6.4.1	Jets and b tagging.	128
86	6.4.2	Missing Energy MET.	129
87	6.4.3	Lepton reconstruction and identification	130

88	6.4.3.1 Muons.	130
89	6.4.3.2 Electrons.	131
90	6.4.3.3 Lepton vertexing and pile-up rejection.	132
91	6.4.3.4 Lepton MVA discriminator.	134
92	6.4.4 Lepton selection efficiency	134
93	6.5 Data and MC Samples	136
94	6.5.1 Full 2016 dataset and MC samples	137
95	6.5.2 Triggers	137
96	6.5.2.1 Trigger efficiency scale factors	140
97	6.6 Background predictions	140
98	6.7 Signal discrimination	142
99	6.7.1 Classifiers response	147
100	6.8 Additional discriminating variables	149
101	Bibliography	151
102	References	152

¹⁰³ List of Figures

104	2.1 Standard model of particle physics.	4
105	2.2 Transformations between quarks	12
106	2.3 Fundamental interactions in nature.	13
107	2.4 SM interactions diagrams	14
108	2.5 Neutral current processes	21
109	2.6 Spontaneous symmetry breaking mechanism	29
110	2.7 SSB Potential form	30
111	2.8 Potential for complex scalar field	31
112	2.9 SSB mechanism for complex scalar field	32
113	2.10 Proton-Proton collision	39
114	2.11 Higgs boson production mechanism Feynman diagrams	40
115	2.12 Higgs boson production cross section and decay branching ratios	41
116	2.13 Associated Higgs boson production mechanism Feynman diagrams	43
117	2.14 Cross section for tHq process as a function of κ_t	46
118	2.15 Cross section for tHW process as a function of κ_{Htt}	47
119	2.16 NLO cross section for tX_0 and $t\bar{t}X_0$	50
120	2.17 NLO cross section for tWX_0 , $t\bar{t}X_0$	51

121	2.18 Two dimentional κ_t - κ_V plot of the coupling modifiers. ATLAS and CMS 122 combination.	52
123	3.1 CERN accelerator complex	55
124	3.2 LHC protons source. First acceleration stage.	56
125	3.3 The LINAC2 accelerating system at CERN.	57
126	3.4 LHC layout and RF cavities module.	58
127	3.5 LHC dipole magnet.	60
128	3.6 Integrated luminosity delivered by LHC and recorded by CMS during 2016	62
129	3.7 LHC interaction points	63
130	3.8 Multiple pp collision bunch crossing at CMS.	65
131	3.9 Layout of the CMS detector	66
132	3.10 CMS detector coordinate system	67
133	3.11 CMS pixel detector schematic view.	70
134	3.12 SST Schematic view.	71
135	3.13 CMS ECAL schematic view	72
136	3.14 CMS HCAL schematic view	74
137	3.15 CMS solenoid magnet	75
138	3.16 CMS Muon system schematic view	76
139	3.17 CMS Level-1 trigger architecture	78
140	3.18 WLCG structure	80
141	3.19 Data flow from CMS detector through hardware Tiers	82
142	4.1 Event generation process.	84
143	4.2 Particle flow algorithm.	92
144	4.3 Jet reconstruction.	100
145	4.4 Jet energy corrections.	101

146	4.5	Secondary vertex in a b-hadron decay.	103
147	4.6	HIG-13-004 Event 1 reconstruction.	104
148	4.7	$e\mu$ event reconstruction.	105
149	4.8	Recorded event reconstruction.	106
150	5.1	Scatter plots-MVA event classification.	109
151	5.2	Scalar test statistical.	109
152	5.3	Decision tree.	111
153	5.4	Decision tree output example.	114
154	5.5	BDT output example.	117
155	6.1	tHq event signature.	124
156	6.2	$t\bar{t}$ and $t\bar{t}W$ events signature.	127
157	6.3	Input variables to the BDT for signal discrimination normalized.	143
158	6.4	Input variables to the BDT for signal discrimination not normalized.	145
159	6.5	BDT inputs as seen by TMVA against $t\bar{t}$.	146
160	6.6	BDT inputs as seen by TMVA against $t\bar{t}V$.	147
161	6.7	Correlation matrices for the input variables in the TMVA.	148
162	6.8	MVA classifiers performance.	148
163	6.9	Additional discriminating variables distributions.	150

¹⁶⁴ List of Tables

165	2.1	Fermions of the SM.	5
166	2.2	Fermion masses.	6
167	2.3	Leptons properties.	9
168	2.4	Quarks properties.	9
169	2.5	Fermion weak isospin and weak hypercharge multiplets.	11
170	2.6	Fundamental interactions features.	15
171	2.7	SM gauge bosons.	20
172	2.8	Higgs boson properties.	38
173	2.9	Predicted branching ratios for a SM Higgs boson with $m_H = 125 \text{ GeV}/c^2$	42
174	2.10	Predicted SM cross sections for tH production at $\sqrt{s} = 13 \text{ TeV}$	44
175	2.11	Predicted enhancement of the tHq and tHW cross sections at LHC	48
176	6.1	Requirements on each of the three muon selections.	133
177	6.2	Criteria for each of the three electron selections.	133
178	6.3	Signal samples and their cross section and branching fraction.	137
179	6.4	κ_V and κ_t combinations.	138
180	6.5	List of background samples used in this analysis (CMSSW 80X).	139
181	6.6	Leading-order $t\bar{t}W$ and $t\bar{t}Z$ samples used in the signal BDT training.	139
182	6.7	Table of high-level triggers that we consider in the analysis.	140

183	6.8 Trigger efficiency scale factors and associated uncertainties.	141
184	6.9 MVA input discriminating variables	144
185	6.10 TMVA input variables ranking for BDTA_GRAD method	149
186	6.11 TMVA configuration used in the BDT training.	149
187	6.12 ROC-integral for all the testing cases.	151

¹⁸⁸ Chapter 1

¹⁸⁹ INTRODUCTION

¹⁹⁰ **Chapter 2**

¹⁹¹ **Theoretical approach**

¹⁹² **2.1 Introduction**

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

¹⁹⁶

¹⁹⁷ At the end of 1940s Julian Schwinger [1] and Richard P. Feynman [2], based in the
¹⁹⁸ work of Sin-Itiro Tomonaga [3], developed an electromagnetic theory consistent with
¹⁹⁹ special relativity and quantum mechanics that describes how matter and light inter-
²⁰⁰ act; the so-called “quantum eletrodynamics” (QED) had born.

²⁰¹

²⁰² QED has become the guide in the development of theories that describe the universe.
²⁰³ It was the first example of a quantum field theory (QFT), which is the theoretical
²⁰⁴ framework for building quantum mechanical models that describes particles and their
²⁰⁵ interactions. QFT is composed of a set of mathematical tools that combines classical
²⁰⁶ fields, special relativity and quantum mechanics, while keeping the quantum point

207 particles and locality ideas.

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

213 2.2 Standard model of particle physics

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

217

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

226

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

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

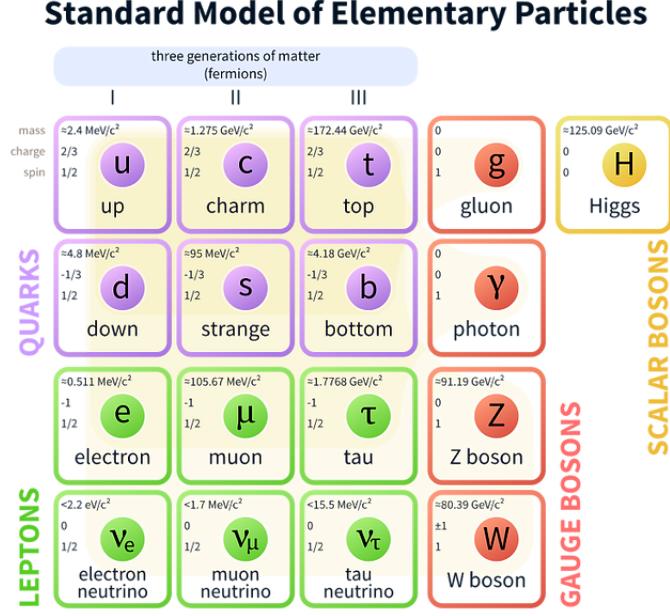


Figure 2.1: Schematic representation of the Standard model of particle physics. SM is a theoretical model intended to describe three of the four fundamental forces of the universe in terms of a set of particles and their interactions. [7].

229 that is invariant under a group of transformations a conservation law is expected. For
 230 instance, a system described by a time-independent Lagrangian is invariant (symmet-
 231 ric) under time changes (transformations) with the total energy conservation law as
 232 the expected conservation law. In QED, the charge operator (Q) is the generator of
 233 the $U(1)$ symmetry which according to the Noether's theorem means that there is a
 234 conserved quantity; this conserved quantity is the electric charge and thus the law
 235 conservation of electric charge is established.

236

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

- 241 • Strong: $SU(3)_C$ associated to color charge
- 242 • Weak: $SU(2)_L$ associated to weak isospin and chirality
- 243 • Electromagnetic: $U(1)_Y$ associated to weak hypercharge and electric charge
- 244 It will be shown that the electromagnetic and weak interactions are combined in
 245 the so-called electroweak interaction where chirality, hypercharge, weak isospin and
 246 electric charge are the central concepts.

247 2.2.1 Fermions

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

253

		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.

254

255 There is a mass hierarchy between generations (see table 2.2), where the higher gener-
 256 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

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

275

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

286

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

292

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

300

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

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

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

309

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

317

318 2.2.1.2 Quarks

319 Quarks are the basic constituents of protons and neutrons. The way quarks join to
 320 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.

321 fundamental interactions which means that they carry all the four types of charges:
 322 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.

323
 324 Table 2.4 summarizes the features of quarks, among which the most particular is
 325 their fractional electric charge. Note that fractional charge is not a problem, given
 326 that quarks are not found isolated, but serves to explain how composed particles are
 327 formed out of two or more valence quarks⁴.

328

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.

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

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

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

345 In the first version of the quark model (1964), M. Gell-Mann [12] and G. Zweig
 346 [13, 14] developed a consistent way to classify hadrons according to their properties.
 347 Only three quarks (u, d, s) were involved in a scheme in which all baryons have
 348 baryon number $B=1$ and therefore quarks have $B=1/3$; non-baryons have $B=0$. The
 349 scheme organizes baryons in a two-dimensional space ($I_3 - Y$); Y (hypercharge) and I_3
 350 (isospin) are quantum numbers related by the Gell-Mann-Nishijima formula [15, 16]:

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

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

354

355 There are six quark flavors organized in three generations (see table 2.1) following a
 356 mass hierarchy which, again, implies that higher generations decay to first generation
 357 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.

358

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

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

364

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

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

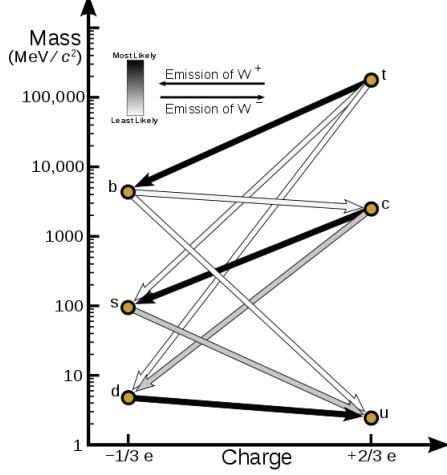


Figure 2.2: Transformations between quarks through the exchange of a WI. Higher generations quarks decay to first generation quarks by emitting a W boson. The arrow color indicates the likelihood of the transition according to the grey scale in the top left side which represent the CKM matrix parameters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

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

370

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

379 **2.2.2 Fundamental interactions**

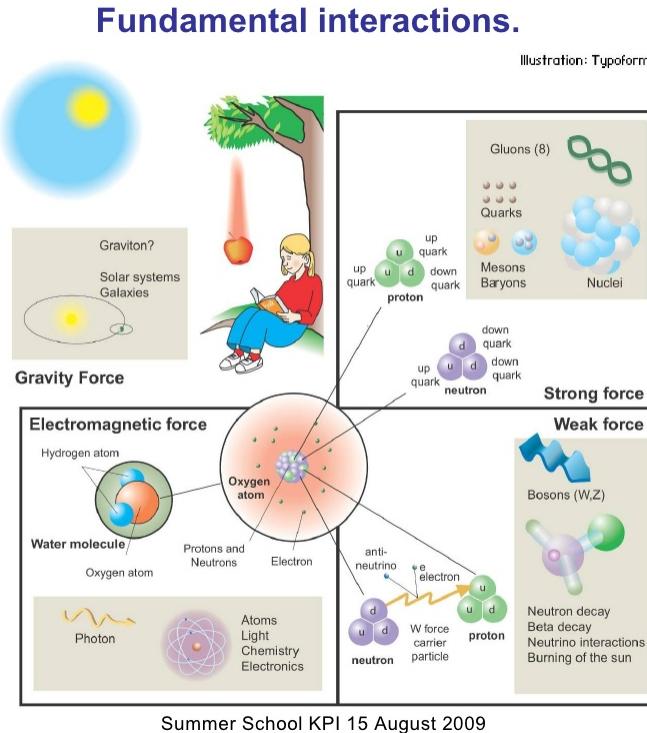


Figure 2.3: Fundamental interactions in nature. Despite the many manifestations of forces in nature, we can track all of them back to one of the fundamental interactions. The most common forces are gravity and electromagnetic given that all of us are subject and experience them in everyday life.

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

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

382 tions:

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

388 a graphical representation, known as “feynman diagram”, of electron-electron
 389 scattering.

- 390 • *Strong interaction (SI)* described by Quantum Chromodynamics (QCD). Hadrons
 391 like proton and neutron have internal structure given that they are composed
 392 of two or more valence quarks⁵. Quarks have fractional electric charge which
 393 means that they are subject to electromagnetic interaction and in the case of the
 394 proton they should break apart due to electrostatic repulsion; however, quarks
 395 are held together inside the hadrons against their electrostatic repulsion by the
 396 “Strong Force” through the exchange of “gluons.” The analog to the electric
 397 charge is the “color charge”. Electrons and photons are elementary particles
 398 as quarks but they don’t carry color charge, therefore they are not subject to
 399 SI. The feynman diagram for gluon exchange between quarks is shown in figure
 400 2.4b.

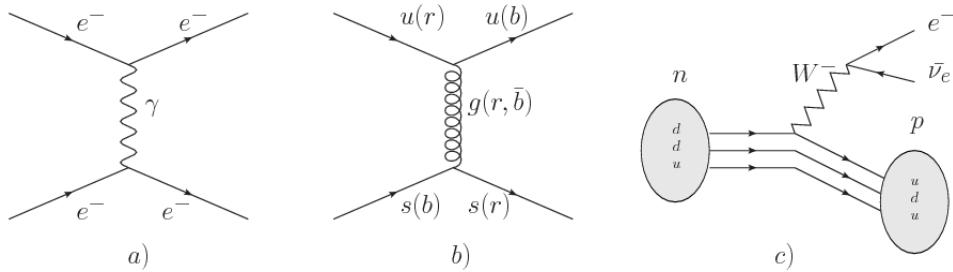


Figure 2.4: Feynman diagrams representing the interactions in SM; a) EI: e-e scattering; b) SI: gluon exchange between quarks ; c) WI: β -decay

- 401 • *Weak interaction (WI)* described by the weak theory (WT), is responsible, for
 402 instance, for the radioactive decay in atoms and proton-proton (pp) fusion
 403 within the sun. Quarks and leptons are the particles affected by the weak
 404 interaction; they possess a property called “flavor charge” (see 2.2.1) which can
 405 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.

406 mediators of the “weak force” known as “Z” boson in the case of electrically
 407 neutral changes and “ W^\pm ” bosons in the case of electrically charged changes.
 408 The “weak isospin” is the WI analog to electric charge in EI, and color charge
 409 in SI, and defines how quarks and leptons are affected by the weak force. Figure
 410 2.4c. shows the feynman diagram of β -decay where a newtron (n) is transformed
 411 in a proton (p) by emmiting a W^- particle. Since this thesis is in the frame
 412 of the electroweak interaction, a more detailed description of it will be given in
 413 section 2.3

414 • *Gravitational interaction (GI)* described by General Theory of Relativity (GR).
 415 It is responsible for the structure of galaxies and black holes as well as the
 416 expansion of the universe. As a classical theory, in the sense that it can be for-
 417 mulated without even appeal to the concept of quantization, it implies that the
 418 spacetime is a continuum and predictions can be made without limitation to the
 419 precision of the measurement tools. The latter represent a direct contradiction
 420 of the quantum mechanics principles. Gravity is deterministic while quantum
 421 mechanics is probabilistic; despite that, efforts to develop a quantum theory of
 422 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].

423

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

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

433

434 QED was built successfully on the basis of the classical electrodynamics theory (CED)
 435 of Maxwell and Lorentz, following theoretical and experimental requirements imposed
 436 by

- 437 • lorentz invariance: independence on the reference frame.
- 438 • locality: interacting fields are evaluated at the same space-time point to avoid
 439 action at a distance.
- 440 • renormalizability: physical predictions are finite and well defined
- 441 • particle spectrum, symmetries and conservation laws already known must emerge
 442 from the theory.
- 443 • gauge invariance.

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

447 “V” and the vector potential “A”. In particular, \mathbf{E} can be obtained by measuring
 448 the change in the space of the scalar potential (ΔV); however, two scalar potentials
 449 differing by a constant “f” correspond to the same electric field. The same happens in
 450 the case of the vector potential “A”; thus, different configurations of the associated
 451 fields result in the same set of values of the observables. The freedom in choosing
 452 one particular configuration is known as “gauge freedom”; the transformation law con-
 453 necting two configurations is known as “gauge transformation” and the fact that the
 454 observables are not affected by a gauge transformation is called “gauge invariance”.

455

456 When the gauge transformation:

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

457 is applied to Maxwell equations, they are still satisfied and the fields remain invariant.
 458 Thus, CED is invariant under gauge transformations and is called a “gauge theory”.
 459 The set of all gauge transformations form the “symmetry group” of the theory, which
 460 according to the group theory, has a set of “group generators”. The number of group
 461 generators determine the number of “gauge fields” of the theory.

462

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

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

470

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

479

480 Due to the space dependence of the local tranformation, the Lagrangian density is
 481 not invariant anymore. In order to restate the gauge invariance, the gauge covariant
 482 derivative is introduced in the Lagrangian and with it the gauge field responsible for
 483 the interaction between particles in the system. The new Lagrangian density is gauge
 484 invariant, includes the interaction terms needed to account for the interactions and
 485 provides a way to explain the interaction between particles through the exchange of
 486 the gauge boson.

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

489 **2.2.3 Gauge bosons**

490 The importance of the gauge bosons comes from the fact that they are the force
 491 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.

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

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

516 no electric charge but momentum transference is involved. WI gauge bosons
 517 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].

518

519 **2.3 Electroweak unification and the Higgs 520 mechanism**

521 Physicists dream of building a theory that contains all the interactions in one single
 522 interaction, i.e., showing that at some scale in energy all the four fundamental in-
 523 teractions are unified and only one interaction emerges in a “Theory of everything”.
 524 The first sign of the feasibility of such unification comes from success in the con-
 525 struction of the CED. Einstein spent years trying to reach that dream, which by
 526 1920 only involved electromagnetism and gravity, with no success; however, a new
 527 partial unification was achieved in the 1960’s, when S.Glashow [22], A.Salam [23] and
 528 S.Weinberg [24] independently proposed that electromagnetic and weak interactions
 529 are two manifestations of a more general interaction called “electroweak interaction
 530 (EWT)”. Both, QCD and EWT, were developed in parallel and following the useful
 531 prescription provided by QED and the gauge invariance principles.

532

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

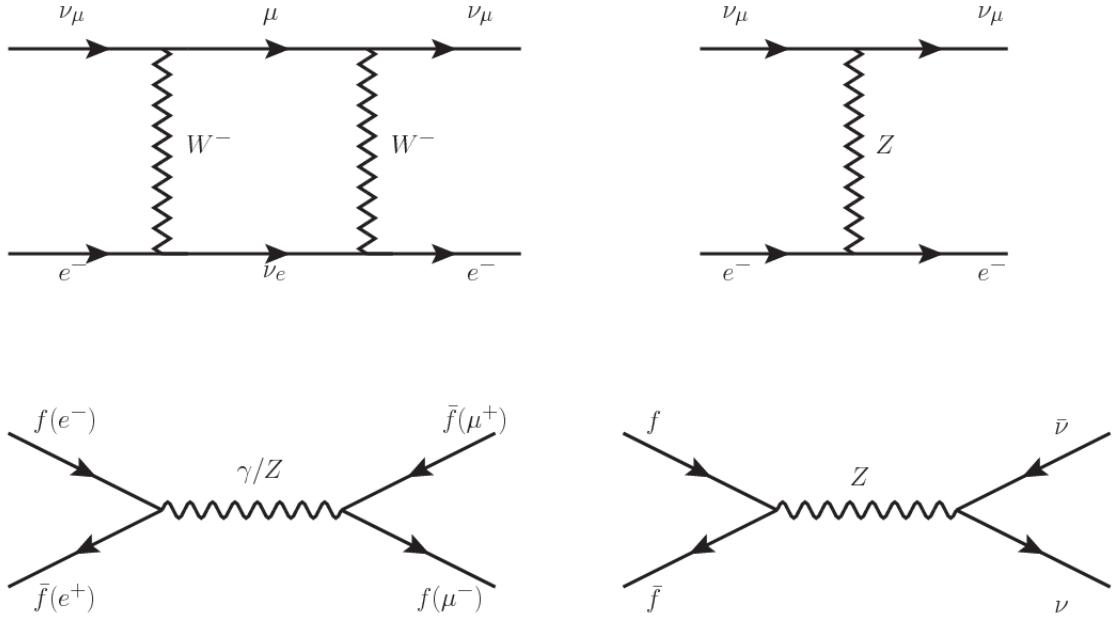


Figure 2.5: Top: $\nu_\mu - e^-$ scattering going through charged currents (left) and neutral currents (right). Bottom: neutral current processes for charged fermions (left) and involving neutrinos (right). While neutral current processes involving only charged fermions can proceed through EI or WI,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

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

539

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

544

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

547 ance under global phase transformations first and generalizing to local phase trans-
 548 formations invariance later, the conserved currents are identified and interactions are
 549 generated by introducing gauge fields. Given that the goal is to include the EI and
 550 WI in a single theory, the group symmetry considered should be a combination of
 551 $SU(2)_L$ and $U(1)_{em}$, however the latter cannot be used directly because the EI treats
 552 left and right-handed particles indistinctly in contrast to the former. Fortunately, the
 553 weak hypercharge, which is a combination of the weak isospin and the electric charge
 554 (eqn 2.2) is suitable to be used since it is conserved by the EI and WI. Thus, the
 555 symmetry group to be considered is

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

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

558

559 Given the first generation of leptons

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

560 the charged fermionic currents are given by

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

561 and the free Lagrangian is given by

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

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

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

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

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

567

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

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

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

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

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

578 $\alpha^i = \alpha^i(x)$, $\beta = \beta(x)$ and the replacement of the ordinary derivatives by the covariant
 579 derivatives

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

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

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

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

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

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

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

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

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

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

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

590

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

619 the first term is the weak charged current interaction, while the second term is the
 620 electromagnetic interaction under the condition

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

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

622

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

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

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

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

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

631 2.3.1 Spontaneous symmetry breaking (SSB)

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

634

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

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

646

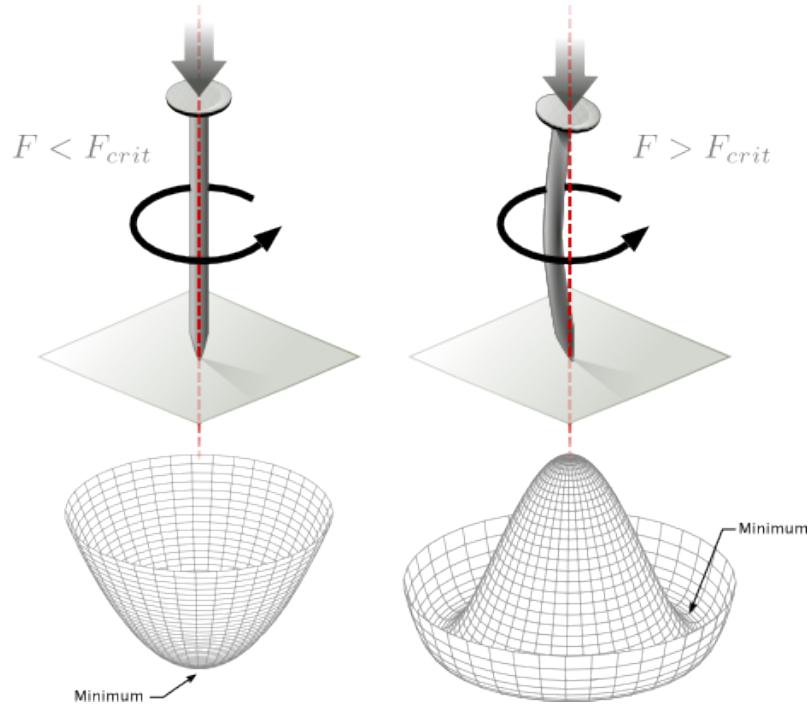


Figure 2.6: Spontaneous symmetry breaking mechanism. The steel nail, subject to an external force (top left), has rotational symmetry with respect to its axis. When the external force overcomes a critical value the nail buckles (top right) choosing a minimal energy state (ground state) and thus “*breaking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

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

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

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

651

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

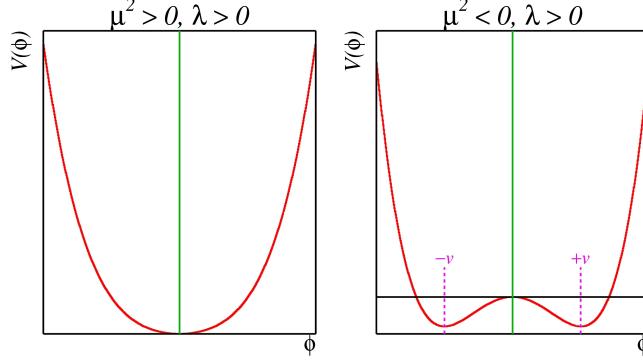


Figure 2.7: Shape of the potential $V(\phi)$ for $\lambda > 0$ and: $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The case $\mu^2 < 0$ corresponds to the potential suitable for introducing the SSB mechanism by choosing one of the two ground states which are connected via reflection symmetry. [27].

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

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

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

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

655

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

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

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

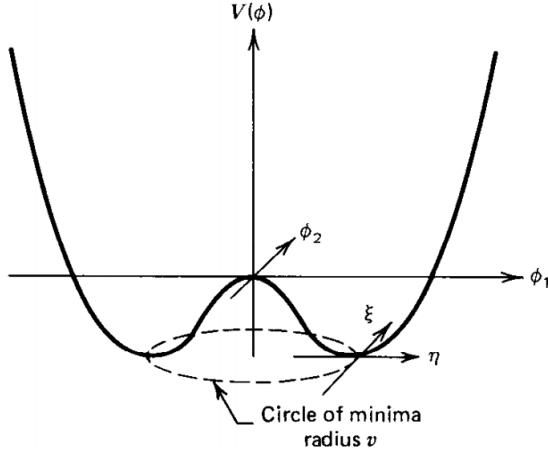


Figure 2.8: Potential for complex scalar field. There is a circle of minima of radius v along the ξ -direction [6].

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

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

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

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

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

668

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

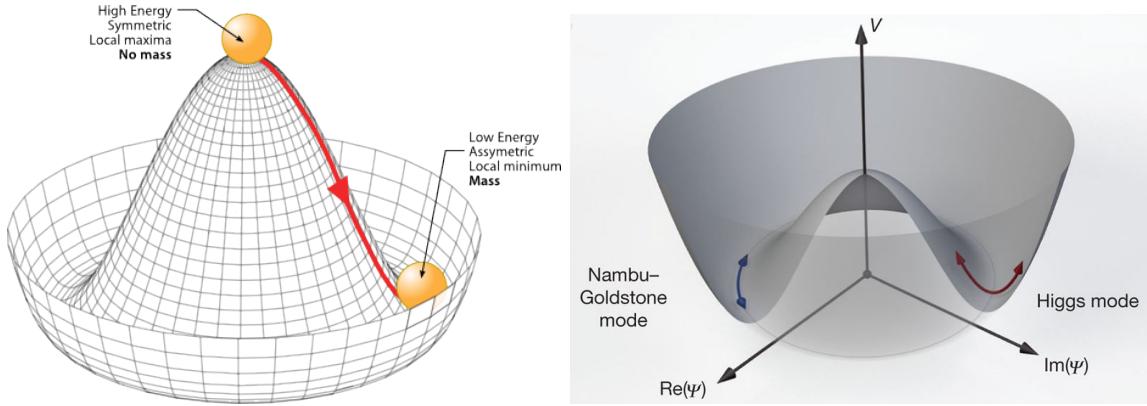


Figure 2.9: SSB mechanism for a complex scalar field [27, 28].

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

679 **2.3.2 Higgs mechanism**

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

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

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

685 ϕ has to be an isospin doublet of complex scalar fields so it preserves the G invariance;

686 thus ϕ can be defined as:

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

687 The minima of the potential are defined by

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

688 The choice of the ground state is critical. By choosing a ground state, invariant under

689 $U(1)_{em}$ gauge symmetry, the photon will remain massless and the W^\pm and Z bosons

690 masses will be generated which is exactly what is needed. In that sense, the best

691 choice corresponds to a weak isospin doublet with $T_3 = -1/2$, $Y_W = 1$ and $Q = 0$

692 which defines a ground state with $\phi_1 = \phi_2 = \phi_4$ and $\phi_3 = v$:

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

693 where the vacuum expectation value v is fixed by the Fermi coupling G_F according

694 to $v = (\sqrt{2}G_F)^{1/2} \approx 246$ GeV.

695

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

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

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

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

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

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

710

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

717 2.3.3 Masses of the gauge bosons

718 The mass of the gauge bosons is extracted by evaluating the kinetic part of Lagrangian

719 \mathcal{L}_S in the ground state (known also as the vacuum expectation value), i.e.,

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

720 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

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

721 and then

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

722 **2.3.4 Masses of the fermions**

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

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

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

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

727

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

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

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

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

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

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

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

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

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

739 2.3.5 The Higgs field

740 After the characterization of the fermions and gauge bosons as well as their interac-
741 tions, it is necessary to characterize the Higgs field itself. The Lagrangian \mathcal{L}_S in eqn.
742 2.36 written in terms of the gauge bosons is given by

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

743

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

744

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

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

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

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

754 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.

755

756 2.3.6 Production of Higgs bosons at LHC

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

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

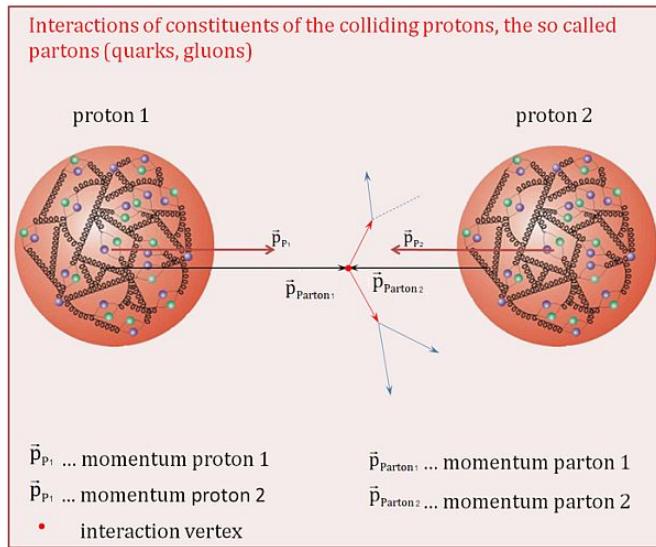


Figure 2.10: Proton-proton collision. Protons are composed of 3 valence quarks, a sea of quarks and gluons; therefore in a proton-proton collision, quarks and gluons are those who collide. [35].

772 are called “parton distribution functions (PDFs)”; PDFs are determined from experi-
 773 mental data obtained in experiments where the internal structure of hadrons is tested.

774

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 the description of the complex system.

783

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

786 LHC.

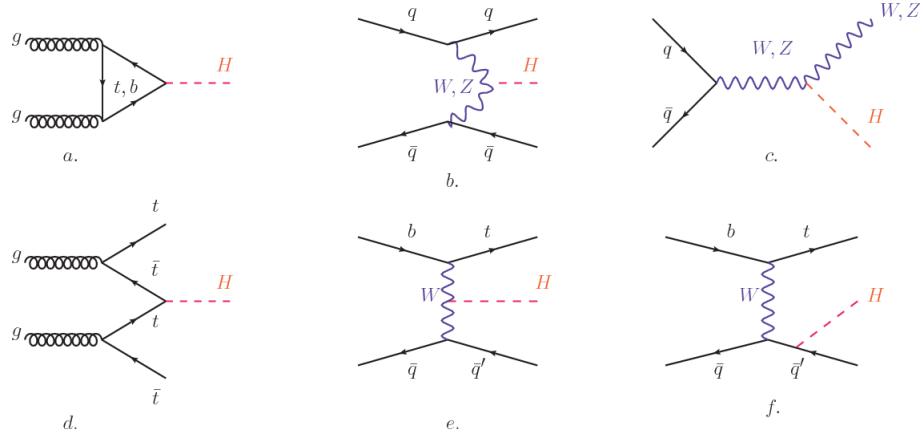


Figure 2.11: Main Higgs boson production mechanism Feynman diagrams. a. gluon-gluon fusion, b. vector boson fusion (VBF), c. Higgs-strahlung, d. Associated production with a top or bottom quark pair, e-f. associated production with a single top quark.

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

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

798 The main production mechanism is the gluon fusion (figure 2.11a and $pp \rightarrow H$ in figure
 799 2.12) given that gluons carry the highest fraction of momentum of the protons in pp
 800 colliders. Since the Higgs boson does not couple to gluons, the mechanism proceeds
 801 through the exchange of a virtual top-quark loop given that for it the coupling is

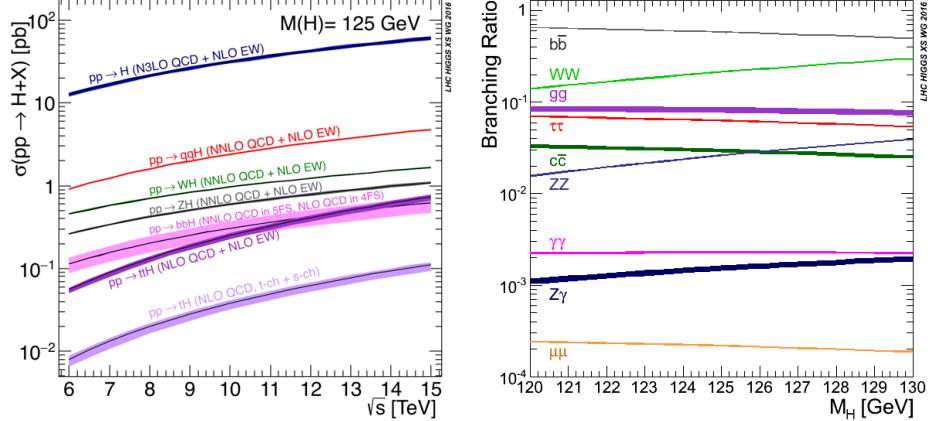


Figure 2.12: Higgs boson production cross sections (left) and decay branching ratios (right) for the main mechanisms. The VBF is indicated as qqH [36].

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,

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

821 2.3.7 Higgs boson decay channels

822 When a particle can decay through several modes, also known as channels, the
 823 probability of decaying through a given channel is quantified by the “branching ratio
 824 (BR)” of the decay channel; thus, the BR is defined as the ratio of number of decays
 825 going through that given channel to the total number of decays. In regard to the
 826 Higgs boson decay, the BR can be predicted with accuracy once the Higgs mass is
 827 known [37, 38]. In figure 2.12 right, a plot of the BR as a function of the Higgs mass
 828 is presented. The largest predicted BR corresponds to the $b\bar{b}$ pair decay channel (see
 829 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]

831 **2.4 Associated production of a Higgs boson and a
832 single Top quark.**

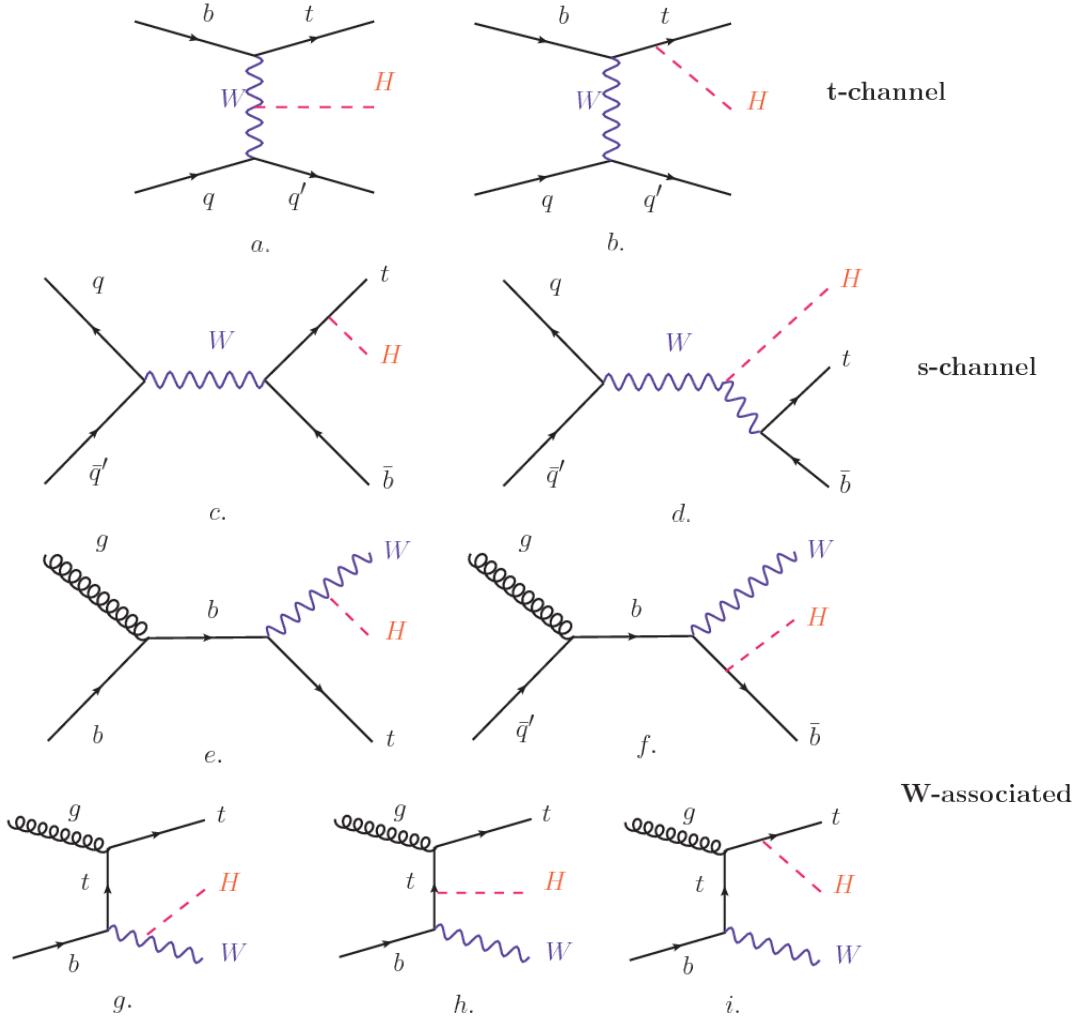


Figure 2.13: Associated Higgs boson production mechanism Feynman diagrams. a.,b. t-channel (tHq), c.,d. s-channel (tHb), e-i. W-associated.

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

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

840

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

849

850 The tH production, where Higgs boson can be radiated either from the top quark
 851 or from the W boson, is represented by the leading order Feynman diagrams in
 852 figure ???. The cross section for the tH process is calculated, as usual, summing over
 853 the contributions from the different feynman diagrams; therefore it depends on the
 854 interference between the contributions. In the SM, the interference for t-channel (tHq
 855 process) and W-associated (tHW process) production is destructive [39] resulting in
 856 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].

857

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

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

867 where $\kappa_V \equiv g_{HV}/g_{HV}^{SM}$ and $\kappa_t \equiv g_{Ht}/g_{Ht}^{SM} = y_t/y_t^{SM}$ are scaling factors that quan-
 868 tify possible deviations of the couplings, Higgs-Vector boson (H-W) and Higgs-top
 869 (H-t) respectively, from the SM couplings; $s = (p_W + p_b)^2$, $t = (p_W - p_H)^2$, φ is the
 870 Higgs azimuthal angle around the z axis taken parallel to the direction of motion of
 871 the incoming W; A and B are funtions describing the weak interaction in terms of
 872 the chiral states of the quarks b and t . Terms that vanish in the high energy limit
 873 have been neglected as well as the Higgs and b quark masses⁸.

874

875 The scattering amplitude grows with energy like \sqrt{s} for $\kappa_V \neq \kappa_t$, in contrast to
 876 the SM ($\kappa_t = \kappa_V = 1$), where the first term in 2.57 cancels out and the amplitude
 877 is constant for large s ; therefore, a deviation from the SM predictions represents an
 878 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]

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

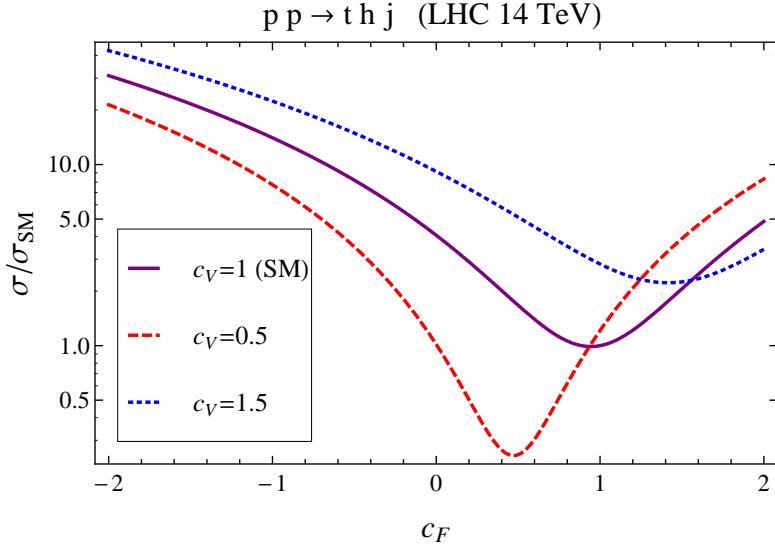


Figure 2.14: Cross section for tHq process as a function of κ_t , normalized to the SM, for three values of κ_V . In the plot c_f refers to the Higgs-fermion coupling which is dominated by the H-t coupling and represented here by κ_t . Solid, dashed and dotted lines correspond to $c_V \rightarrow \kappa_V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

882 A similar analysis is valid for the W-associated channel but, in that case, the inter-
 883 ference is more complicated since there are more than two contributions and an ad-
 884 ditional interference with the production of Higgs boson and a top pair process($t\bar{t}H$).
 885 The calculations are made using the so-called Diagram Removal (DR) technique where
 886 interfering diagrams are removed (or added) from the calculations in order to evaluate
 887 the impact of the removed contributions. DR1 was defined to neglect $t\bar{t}H$ interference
 888 while DR2 was defined to take $t\bar{t}H$ interference into account [48]. As shown in figure
 889 2.15, the tHW cross section is enhanced from about 15 fb (SM: $\kappa_{Htt} = 1$) to about
 890 150 fb ($\kappa_{Htt} = -1$). Differences between curves for DR1 and DR2 help to gauge the
 891 impact of the interference with $t\bar{t}H$.

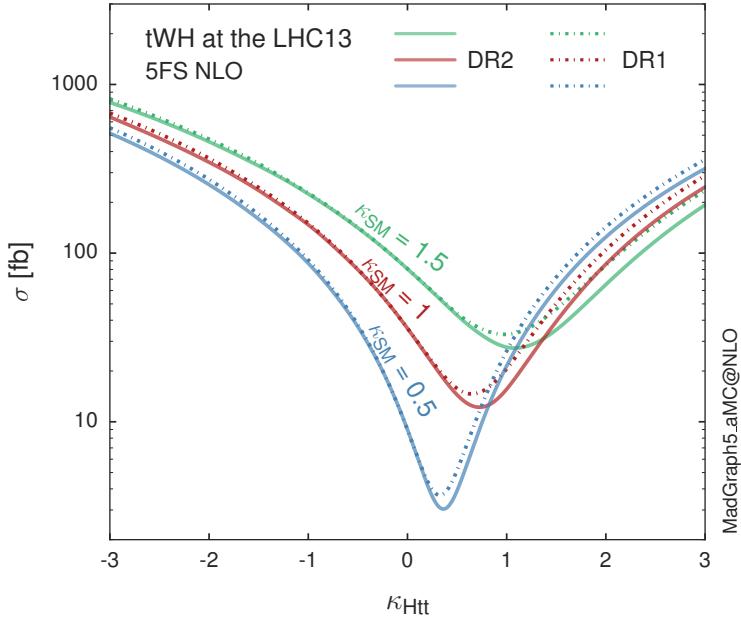


Figure 2.15: Cross section for tHW process as a function of κ_{Htt} , for three values of κ_{SM} at $\sqrt{s} = 13$ TeV. $\kappa_{Htt}^2 = \sigma_{Htt}/\sigma_{Htt}^{SM}$ is a simple rescaling of the SM Higgs interactions.

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

894

895 2.5 The CP-mixing in tH processes

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

900

901 In this thesis, the sensitivity of tH processes to CP-mixing is also studied in the
 902 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.

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

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

906 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
 907 dimensionless rescaling parameters⁹, $g_{Htt} = g_{Att} = m_t/v = y_t/\sqrt{2}$ and $v \sim 246$ GeV is
 908 the Higgs vacuum expectation value. In this parametrization, it is easy to recover
 909 three special cases

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

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

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

⁹ analog to κ_t and κ_V

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

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

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

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

919 If the rescaling parameters are set to

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

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

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

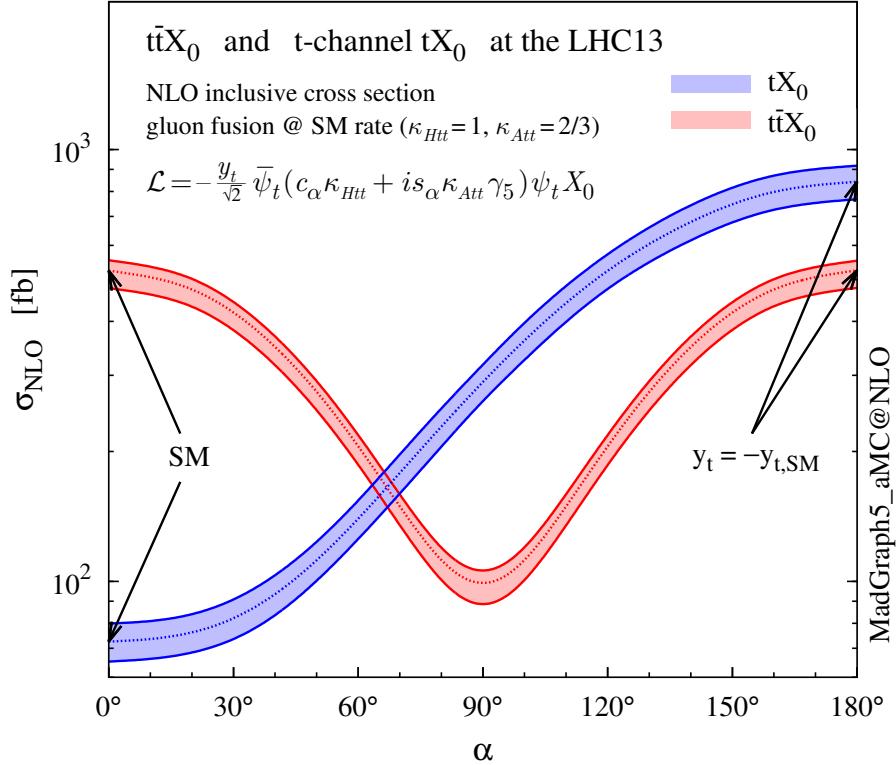


Figure 2.16: NLO cross sections for t-channel tX_0 (blue) and $t\bar{t}X_0$ (red) associated production processses as a function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark CP-violating coupling [43].

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

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

938

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

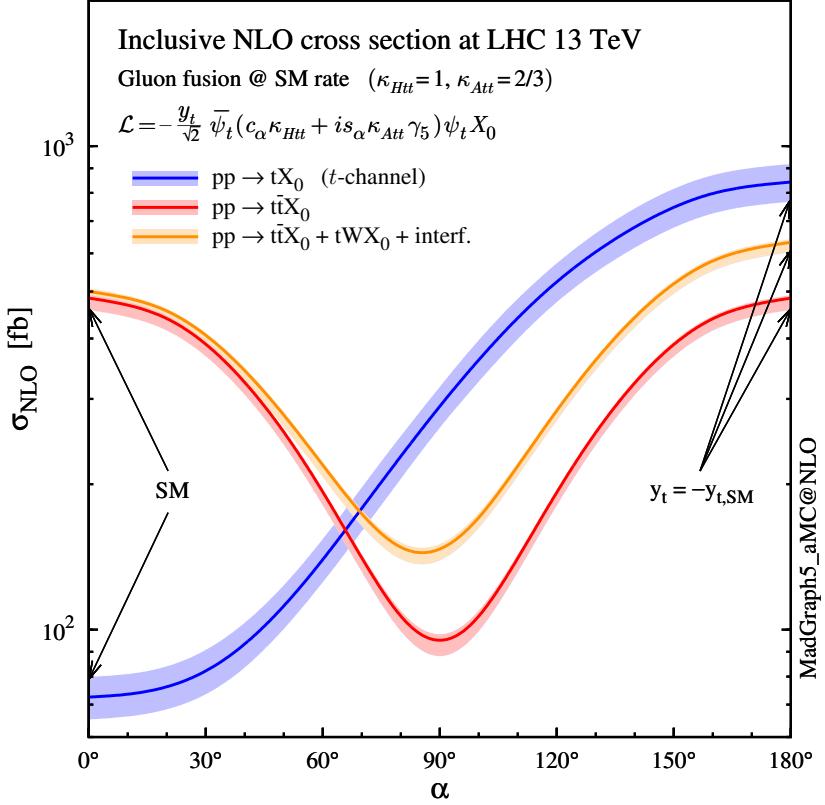


Figure 2.17: NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red) associated production processes and combined $tWX_0 + t\bar{t}X_0$ (including interference) production as a function of the CP-mixing angle α [43].

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.

947 **2.6 Experimantal status of the anomalous**
 948 **Higg-fermion coupling.**

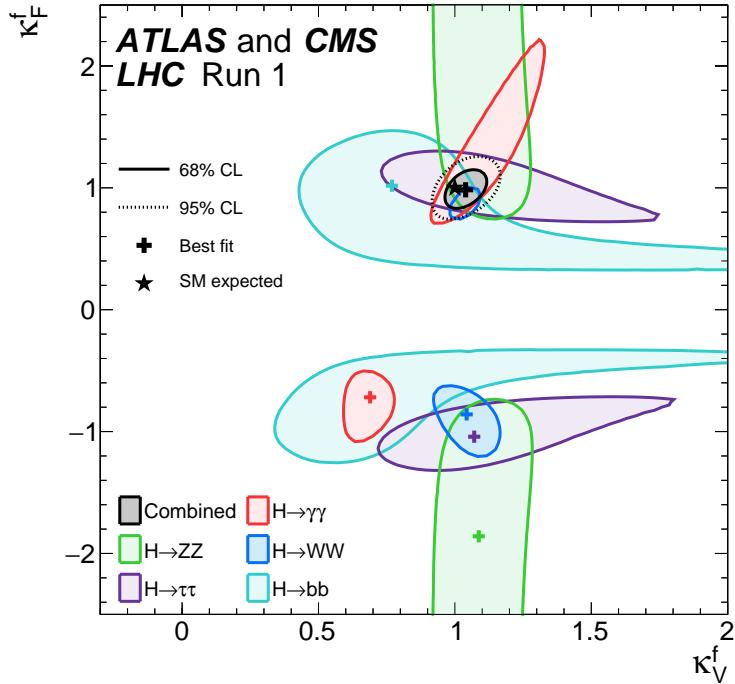


Figure 2.18: Combination of the ATLAS and CMS fits for coupling modifiers κ_t - κ_V ; also shown the individual decay channels combination and their global combination. No assumptions have been made on the sign of the coupling modifiers [53].

949 ATLAS and CMS have performed analysis of the anomalous H-f coupling by making
 950 likelihood scans for the two coupling modifiers, κ_t and κ_V , under the assumption that
 951 $\kappa_Z = \kappa_W = \kappa_V$ and $\kappa_t = \kappa_\tau = \kappa_b = \kappa_f$. Figure 2.18 shows the result of the combination
 952 of ATLAS and CMS fits; also the individual decay channels combination and the
 953 global combination results are shown.

954 While all the channels are compatible for positive values of the modifiers, for negative
 955 values of κ_t there is no compatibility. The best fit for individual channels is compatible
 956 with negative values of κ_t except for the $H \rightarrow bb$ channel which is expected to be the

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

959 **Chapter 3**

960 **The CMS experiment at the LHC**

961 **3.1 Introduction**

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

975 3.2 The LHC

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

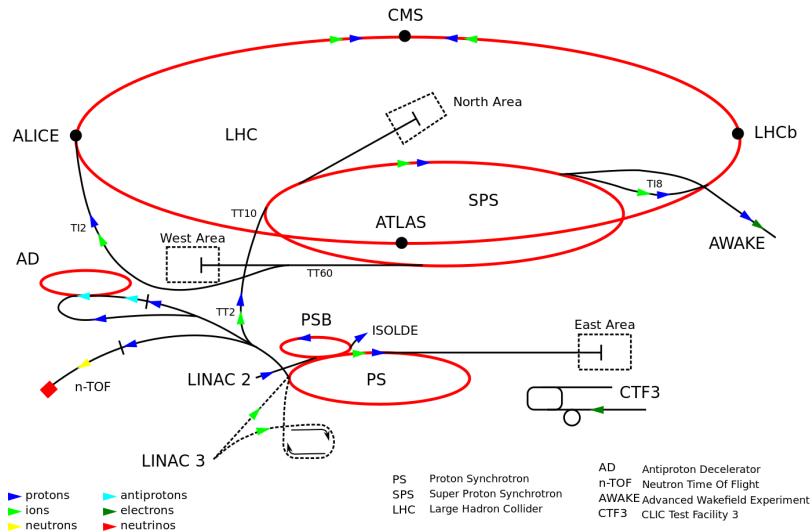


Figure 3.1: CERN accelerator complex. Blue arrows show the path followed by protons along the acceleration process [54].

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

- 983 • Proton-Proton collisions (pp) for multiple physics experiments.
- 984 • Lead-Lead collisions ($Pb-Pb$) for heavy ion experiments.
- 985 • Proton-Lead collisions ($p-Pb$) for quark-gluon plasma experiments.

986 In this thesis only pp collisions will be considered.

987

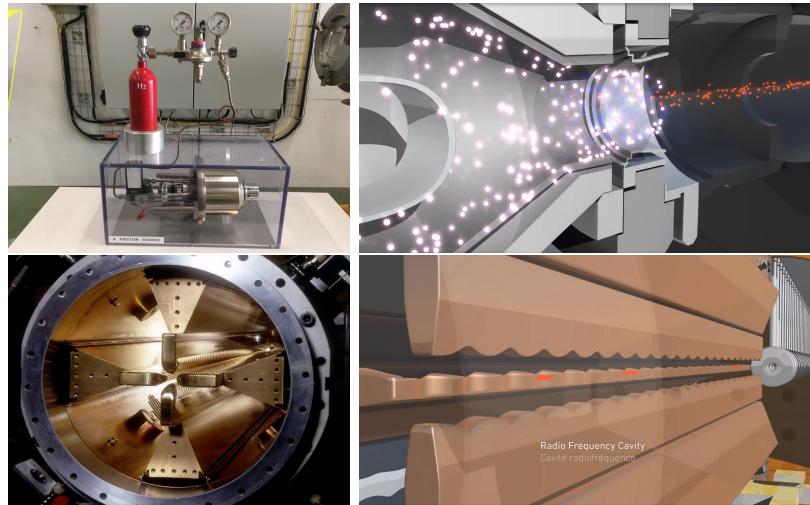


Figure 3.2: LHC protons source and the first acceleration stage. Top: the bottle contains hydrogen gas (white dots) which is injected into the metal cylinder to be broken down into electrons(blue dots) and protons(red dots); Bottom: the obtained protons are directed towards the radio frequency quadrupole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

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

997

998 In order to accelerate the protons, a longitudinal time-varying electric field component
 999 is added to the system; it is done by giving the electrodes a sinus-like profile as shown
 1000 in figure 3.2 bottom. By matching the speed and phase of the protons with the
 1001 longitudinal electric field the bunching is performed; protons synchronized with the

1002 RFQ (synchronous proton) do not feel an accelerating force, but those protons in the
 1003 beam that have more (or less) energy than the synchronous proton (asynchronous
 1004 protons) will feel a decelerating (accelerating) force; therefore, asynchronous protons
 1005 will oscillate around the synchronous ones forming bunches of protons [56]. From the
 1006 RFQ protons emerge with energy 750 keV in bunches of about 1.15×10^{11} protons [57].

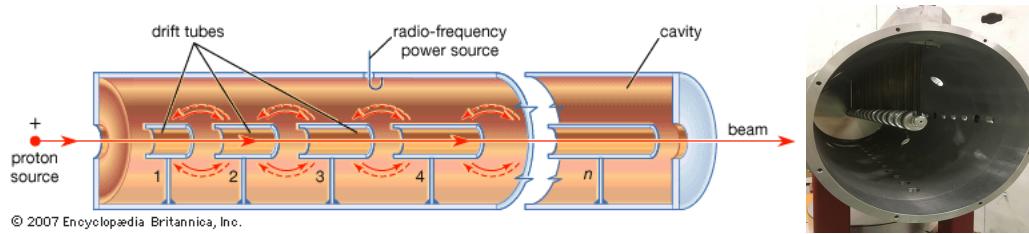


Figure 3.3: The LINAC2 accelerating system at CERN. Electric fields generated by radio frequency (RF) create acceleration and deceleration zones inside the cavity; deceleration zones are blocked by drift tubes where quadrupole magnets focus the proton beam. [60]

1007 Proton bunches coming from the RFQ go to the linear accelerator 2 (LINAC2) where
 1008 they are accelerated to reach 50 MeV energy. In the LINAC2 stage, acceleration
 1009 is performed using electric fields generated by radio frequency which create zones
 1010 of acceleration and deceleration as shown in figure 3.3. In the deceleration zones,
 1011 the electric field is blocked using drift tubes where protons are free to drift while
 1012 quadrupole magnets focus the beam.

1013

1014 The beam coming from LINAC2 is injected into the proton synchrotron booster
 1015 (PSB) to reach 1.4 GeV in energy. The next acceleration is provided at the pro-
 1016 ton synchrotron (PS) up to 26 GeV, followed by the injection into the super proton
 1017 synchrotron (SPS) where protons are accelerated to 450 GeV. Finally, protons are
 1018 injected into the LHC where they are accelerated to the target energy of 6.5 TeV.
 1019 PSB, PS, SPS and LHC accelerate protons using the same RF acceleration technique
 1020 described before.

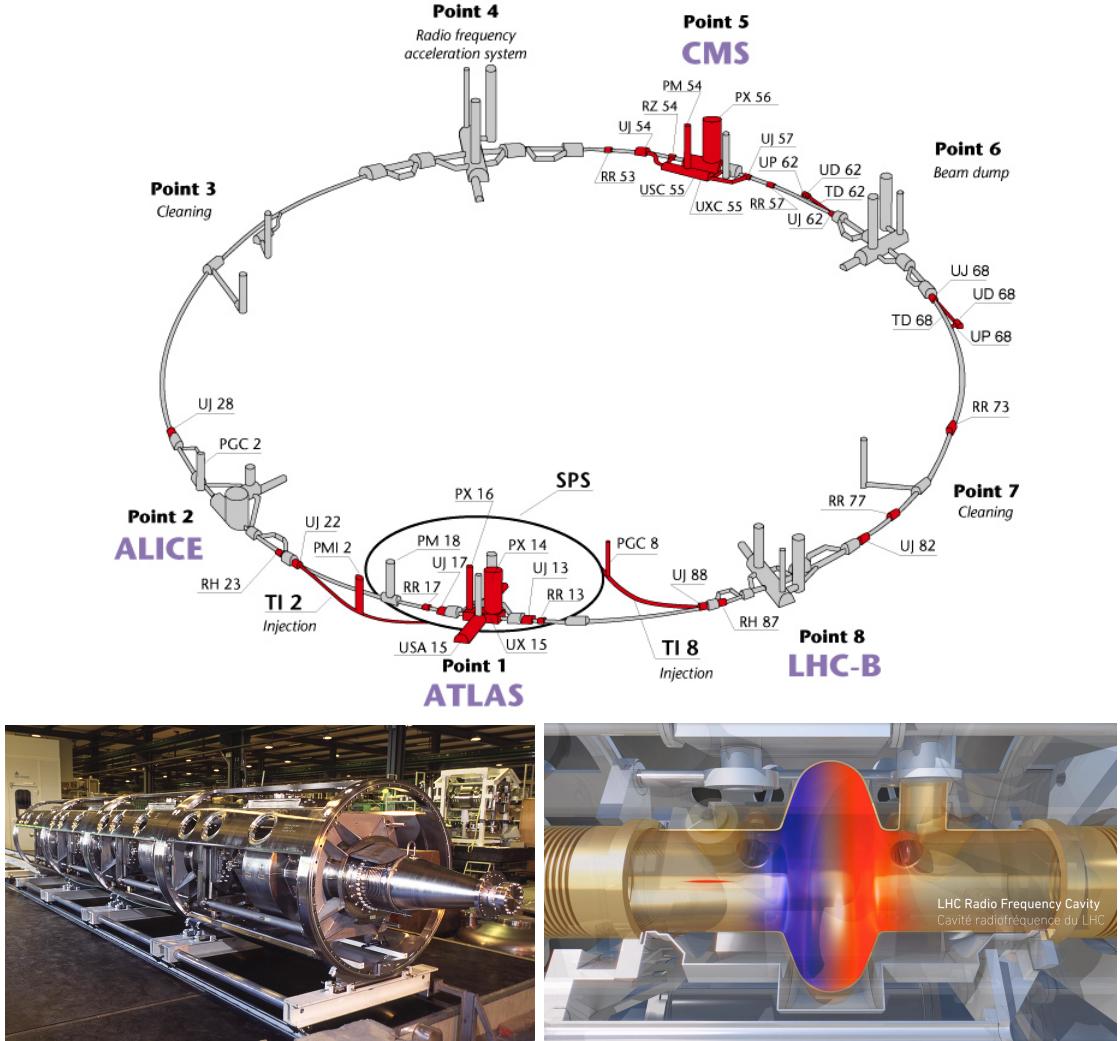


Figure 3.4: Top: LHC layout. The red zones indicate the infrastructure additions to the LEP installations, built to accommodate the ATLAS and CMS experiments which exceed the size of the former experiments located there [55]. Bottom: LHC RF cavities. A module accommodates 4 cavities that accelerate protons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

1021 LHC has a system of 16 RF cavities located in the so-called point 4, as shown in
 1022 figure 3.4 top, tuned at a frequency of 400 MHz and the protons are carefully timed,
 1023 so in addition to the acceleration effect the bunch structure of the beam is preserved.
 1024 Bottom side of figure 3.4 shows a picture of a RF module composed of 4 RF cavities
 1025 working in a superconducting state at 4.5 K; also is showed a representation of the

1026 accelerating electric field that accelerates the protons in the bunch.

1027

1028 While protons are accelerated in one section of the LHC ring, where the RF cavities
 1029 are located, in the rest of their path they have to be kept in the curved trajectory
 1030 defined by the LHC ring. Technically, LHC is not a perfect circle; RF, injection, beam
 1031 dumping, beam cleaning and sections before and after the experimental points where
 1032 protons collide are all straight sections. In total, there are 8 arcs 2.45 Km long each
 1033 and 8 straight sections 545 m long each. In order to curve the proton's trajectory in
 1034 the arc sections, superconducting dipole magnets are used.

1035

1036 Inside the LHC ring, there are two proton beams traveling in opposite directions in
 1037 two separated beam pipes; the beam pipes are kept at ultra-high vacuum ($\sim 10^{-9}$
 1038 Pa) to ensure that there are no particles that interact with the proton beams. The
 1039 superconducting dipole magnets used in LHC are made of a NbTi alloy, capable of
 1040 transporting currents of about 12000 A when cooled at a temperature below 2K using
 1041 liquid helium (see figure 3.5).

1042

1043 Protons in the arc sections of LHC feel a centripetal force exerted by the dipole
 1044 magnets which is perpendicular to the beam trajectory; The magnitude of magnetic
 1045 field needed can be found assuming that protons travel at $v \approx c$, using the standard
 1046 values for proton mass and charge and the LHC radius, as

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

1047 which is about 100000 times the Earth's magnetic field. A representation of the mag-
 1048 netic field generated by the dipole magnets is shown on the bottom left side of figure

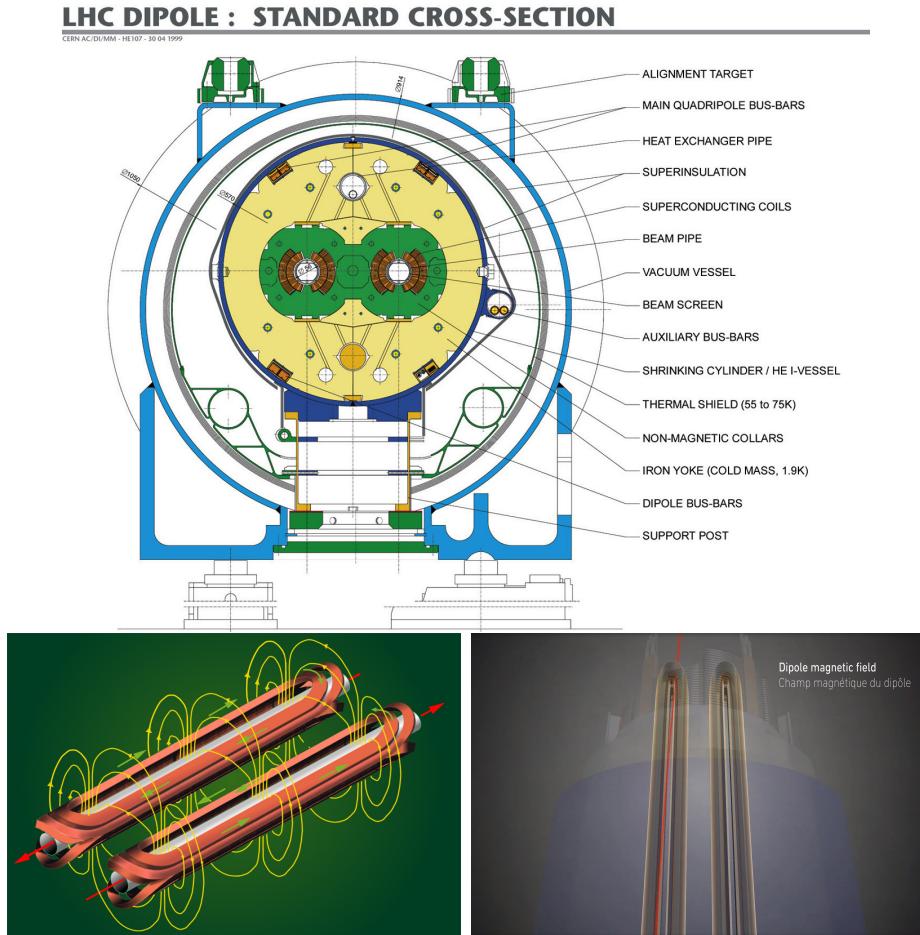


Figure 3.5: Top: LHC dipole magnet transverse view; cooling, shielding and mechanical support are indicated. Bottom left: Magnetic field generated by the dipole magnets; note that the direction of the field inside one beam pipe is opposite with respect to the other beam pipe which guarantee that both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

1049 3.5. The bending effect of the magnetic field on the proton beam is shown on the
 1050 bottom right side of figure 3.5. Note that the dipole magnets are not curved; the arc
 1051 section of the LHC ring is composed of straight dipole magnets of about 15 m. In
 1052 total there are 1232 dipole magnets along the LHC ring.

1053

1054 In addition to bending the beam trajectory, the beam has to be focused so it stays

1055 inside the beam pipe. The focusing is performed by quadrupole magnets installed in
 1056 a different straight section; in total 858 quadrupole magnets are installed along the
 1057 LHC ring. Other effects like electromagnetic interaction among bunches, interaction
 1058 with electron clouds from the beam pipe, the gravitational force on the protons, dif-
 1059 ferences in energy among protons in the same bunch, among others, are corrected
 1060 using sextupole and other magnetic multipoles.

1061

1062 The two proton beams inside the LHC ring are made of bunches with a cylindrical
 1063 shape of about 7.5 cm long and about 1 mm in diameter; when bunches are close
 1064 to the collision point (CP), the beam is focused up to a diameter of about 16 μm in
 1065 order to maximize the number of collisions per unit area and per second, known as
 1066 luminosity (L). Luminosity can be calculated using

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

1067 where f is the revolution frequency, n is the number of bunches per beam, N_1 and
 1068 N_2 are the numbers of protons per bunch (1.5×10^{11}), σ_x and σ_y are the gaussian
 1069 transverse sizes of the bunches. The expected luminosity is about

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

$$n = 2808$$

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

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

1070

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

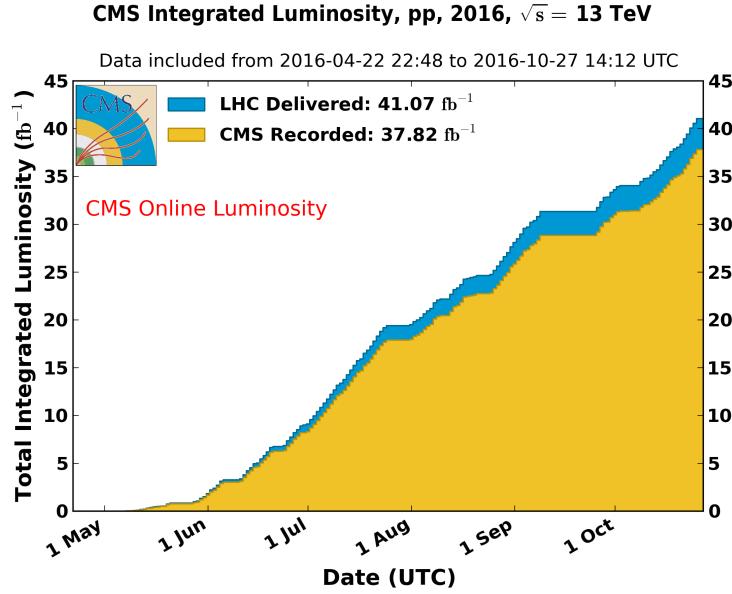


Figure 3.6: Integrated luminosity delivered by LHC and recorded by CMS during 2016. The difference between the delivered and the recorded luminosity is due to fails and issues occurred during the data taking in the CMS experiment [64].

1071 Luminosity is a fundamental aspect of LHC given that the bigger luminosity the
 1072 bigger number of collisions, which means that for processes with a very small cross
 1073 section the number of expected occurrences is increased and so the chances of being
 1074 detected. The integrated luminosity, i.e., the total luminosity, collected by the CMS
 1075 experiment during 2016 is shown in figure 3.6; the data analyzed in this thesis corre-
 1076 sponds to an integrated luminosity of 35.9 fb^{-1} at a center of mass-energy $\sqrt{s} = 13$
 1077 TeV.

1078

1079 A way to increase L is increasing the number of bunches in the beam. Currently, the
 1080 separation between two consecutive bunches in the beam is 7.5 m which corresponds
 1081 to a time separation of 25 ns. In the full LHC ring the allowed number of bunches is
 1082 $n = 27\text{km}/7.5\text{m} = 3600$; however, there are some gaps in the bunch pattern intended
 1083 for preparing the dumping and injection of the beam, thus, the proton beams are

1084 composed of 2808 bunches.

1085

1086 Once the proton beams reach the desired energy, they are brought to cross each other
 1087 producing proton-proton collisions. The bunch crossing happens in precise places
 1088 where the four LHC experiments are located, as seen in the top of figure 3.7. In 2008,
 1089 the first set of collisions involved protons with $\sqrt{s} = 7$ TeV; the energy was increased
 1090 to 8 TeV in 2012 and to 13 TeV in 2015.

1091

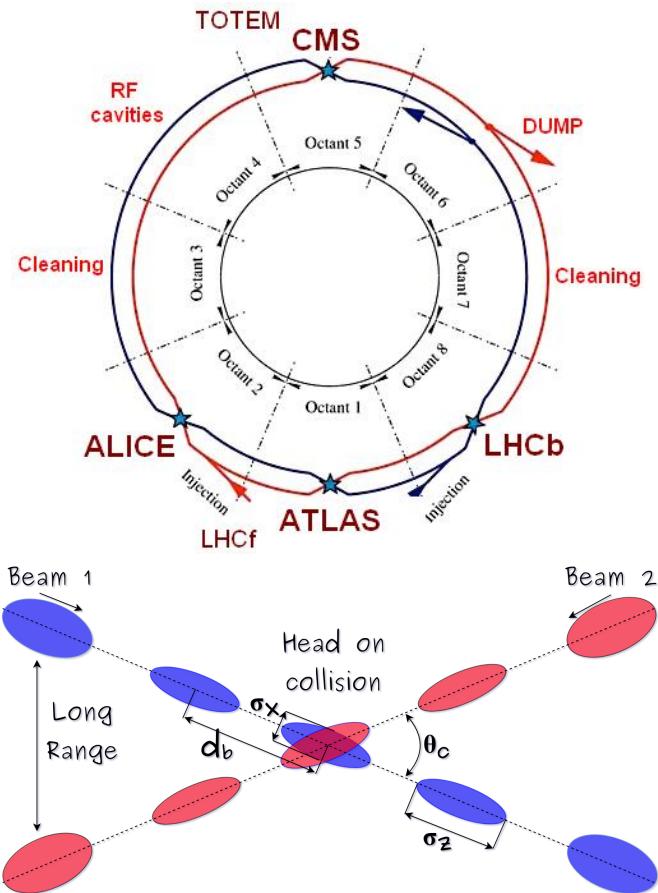


Figure 3.7: Top: LHC interaction points. Bunch crossing occurs where the LHC experiments are located [65]. Sections indicated as cleaning are dedicated to collimate the beam in order to protect the LHC ring from collisions with protons in very spreaded bunches. Bottom: bunch crossing scheme. Since the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

1092 CMS and ATLAS experiments, which are multi-purpose experiments, are enabled
 1093 to explore physics in any of the collision modes. LHCb experiment is optimized
 1094 to explore bottom quark physics, while ALICE is optimized for heavy ion collisions
 1095 searches; TOTEM and LHCf are dedicated to forward physics studies; MoEDAL (not
 1096 indicated in the figure) is intended for monopoles or massive pseudo stable particles
 1097 searches.

1098

1099 At the CP there are two interesting details that need to be addressed. The first one
 1100 is that the bunch crossing does not occur head-on but at a small crossing angle “ θ_c ”
 1101 (280 μ rad in CMS and ATLAS) as shown in the bottom side of figure 3.7, affecting
 1102 the overlapping between bunches; the consequence is a reduction of about 17% in
 1103 the luminosity (represented by a factor not included in eqn: 3.2). The second one
 1104 is the occurrence of multiple pp collisions in the same bunch crossing; this effect is
 1105 called pile-up (PU). A fairly simple estimation of the PU follows from estimating the
 1106 probability of collision between two protons, one from each of the bunches in course
 1107 of collision; it depends roughly on the ratio of proton size and the cross section of the
 1108 bunch in the interaction point, i.e.,

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

1109 however, there are $N = 1.15 \times 10^{11}$ protons in a bunch, thus the estimated number of
 1110 collisions in a bunch crossing is

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

1111 about 20 of those pp collisions are inelastic. Each collision generates a vertex, but
 1112 only the most energetic is considered as a primary vertex; the rest are considered as

1113 PU vertices. A multiple pp collision event in a bunch crossing at CMS is showed in
 1114 figure3.8. Unstable particles outgoing from the primary vertex will eventually decay;
 1115 this decay vertex is known as a secondary vertex.

1116

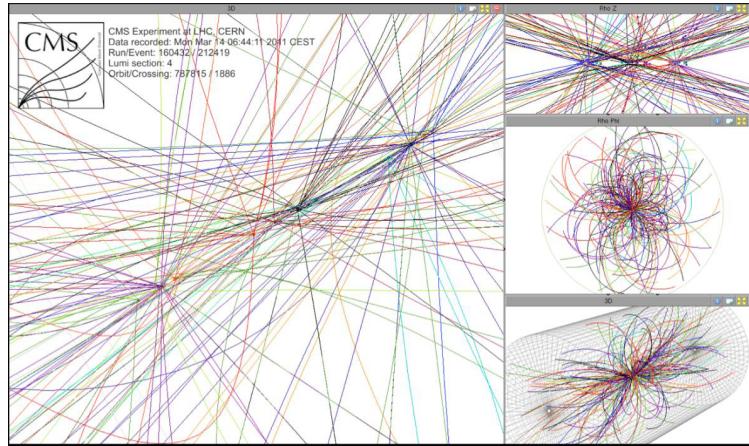


Figure 3.8: Multiple pp collision bunch crossing at CMS. Only the most energetic vertex is considered and the rest are catalogued as PU vertices [66].

1117 Next section presents a description of the CMS detector which it is the detector used
 1118 to collect the data used in this thesis.

1119 **3.3 The CMS experiment**

1120 CMS is a general-purpose detector designed to conduct research in a wide range
 1121 of physics from the standard model to new physics like extra dimensions and dark
 1122 matter. Located at the point 5 in the LHC layout as shown in figure 3.4, CMS is
 1123 composed of several detection systems distributed in a cylindrical structure; in total,
 1124 CMS weights about 12500 tons in a very compact 21.6 m long and 14.6 m diameter
 1125 cylinder. It was built in 15 separate sections at the ground level and lowered to the
 1126 cavern individually to be assembled. A complete and detailed description of the CMS

1127 detector and its components is given in reference [67] on which this section is based on.

1128

1129 Figure 3.9 shows the layout of the CMS detector. The design is driven by the require-
 1130 ments on the identification, momentum resolution and unambiguous charge determi-
 1131 nation of the muons; therefore, a large bending power is provided by the solenoid
 1132 magnet made of superconducting cable capable to generate a 3.8 T magnetic field.
 1133 The detection system is composed of (from the innermost to the outermost)

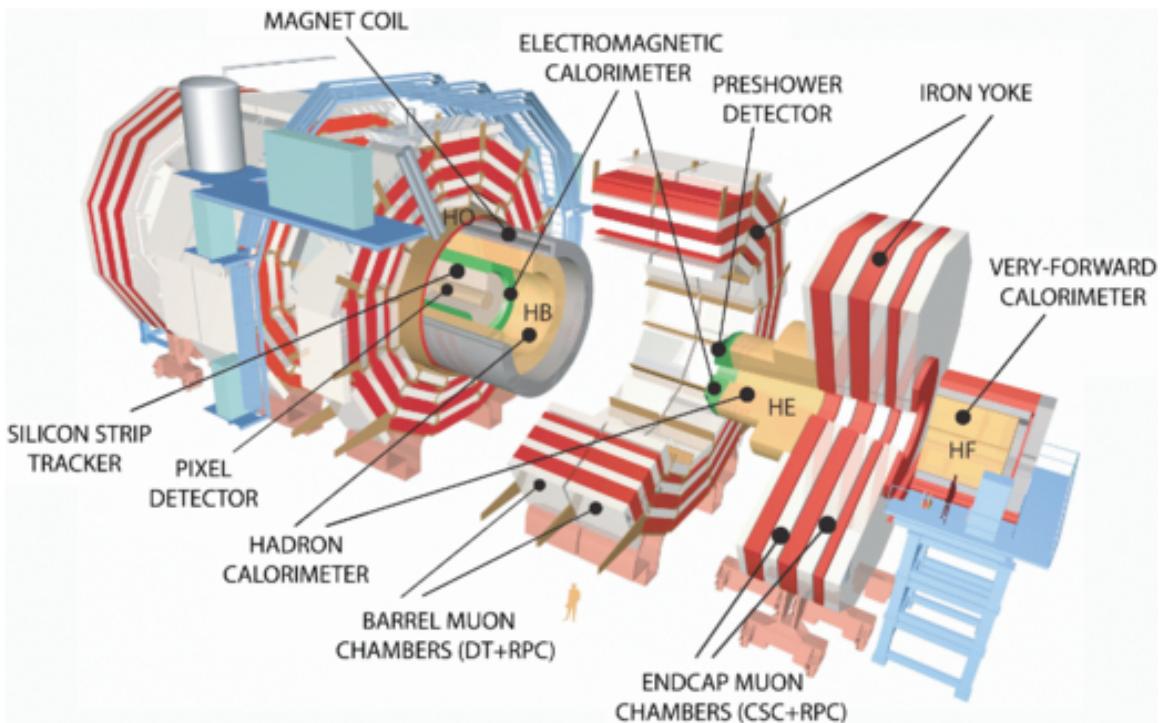


Figure 3.9: Layout of the CMS detector. The several subdetectors are indicated. The central region of the detector is referred as the Barrel section while the endcaps are referred as the forward sections. [68].

1134 • Pixel detector.

1135 • Silicon strip tracker.

1136 • Preshower detector.

1137 • Electromagnetic calorimeter.

1138 • Hadronic calorimeter.

1139 • Muon chambers (Barrel and endcap)

1140 The central region of the detector is commonly referred as the barrel section while the
 1141 endcaps are referred as the forward sections of the detector; thus, each subdetector
 1142 is composed of a barrel section and a forward section.

1143 3.3.1 Coordinate system

1144 The coordinate system used by CMS is centered in the geometrical center of the
 1145 detector which is the same as the CP as shown in figure 3.10. The z -axis is parallel
 1146 to the beam direction, while the Y -axis pointing vertically upward, and the X -axis
 1147 pointing radially inward toward the center of the LHC.

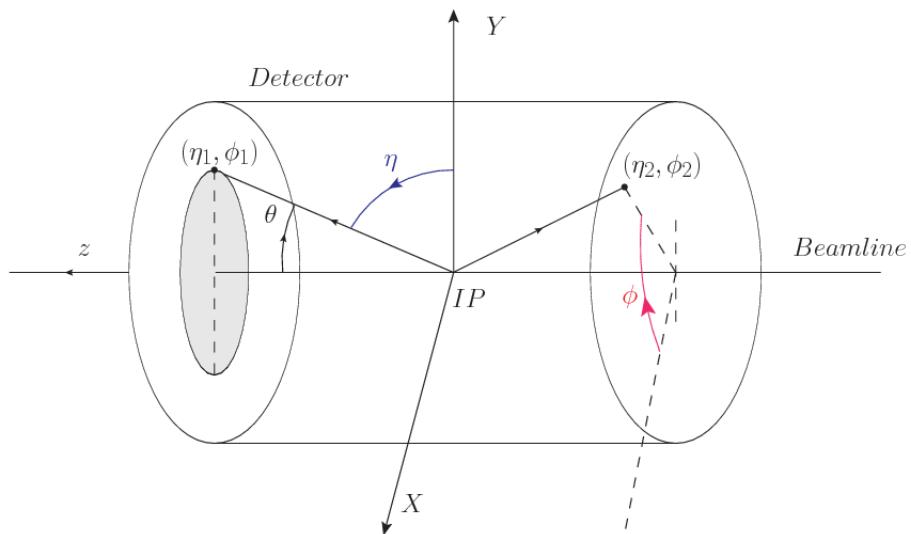


Figure 3.10: CMS detector coordinate system.

1148 In addition to the common cartesian and cylindrical coordinate systems, two coordi-
 1149 nates are of particular utility in particle physics: rapidity(y) and pseudorapidity(η),

1150 defined in connection to the polar angle θ , energy and longitudinal momentum com-
 1151 ponent (momentum along the z -axis) according to

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

1152 Rapidity is related to the angle between the XY -plane and the direction in which the
 1153 products of a collision are emitted; it has the nice property that the difference between
 1154 the rapidities of two particles is invariant with respect to Lorentz boosts along the z -
 1155 axis. Thus, data analysis becomes more simple when based on rapidity; however, it is
 1156 not simple to measure the rapidity of highly relativistic particles, as those produced
 1157 after pp collisions. Under the highly relativistic motion approximation, y can be
 1158 rewritten in terms of the polar angle, concluding that rapidity is approximately equal
 1159 to the pseudorapidity defined above, i.e., $y \approx \eta$. Note that η is easier to measure than y
 1160 given the direct relationship between the former and the polar angle. Angular distance
 1161 between two objects in the detector (ΔR) is defined in terms of their coordinates
 1162 $(\eta_1, \phi_1), (\eta_2, \phi_2)$ as

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

1163 3.3.2 Pixels detector

1164 The CMS tracking system is designed to provide a precise measurement of the tra-
 1165 jectory (*track*) followed by the charged particles created after the pp collisions; also,
 1166 the precise reconstruction of the primary and secondary origins (*vertices*) is expected
 1167 in an environment where, each 25 ns, the bunch crossing produce about 20 inelastic
 1168 collisions and about 1000 particles. An increment in the luminosity is ongoing which
 1169 implies that the PU will increase accordingly.

1171 The pixel detector was replaced during the 2016-2017 extended year-end technical
1172 stop, due to the increasingly challenging operating conditions like the higher particle
1173 flow and more radiation harsh environment, among others. The new one is respond-
1174 ing as expected, reinforcing its crucial role in the successful way to fulfill the new
1175 LHC physics objectives after the discovery of the Higgs boson. The last chapter of
1176 this thesis is dedicated to describe my contribution to the “Forward Pixel Phase 1
1177 upgrade”.

1178

1179 The current pixel detector is composed of 1856 silicon pixel detector modules orga-
1180 nized in four-barrel layers in the central region and three disks in the forward region;
1181 it is designed to record efficiently and with high precision, up to $10\mu\text{m}$ in the XY -
1182 plane and $20\mu\text{m}$ in the z -direction, the first four space-points (*hits*) near to the CP
1183 region (see figure 3.11 left side) in the range $|\eta| \leq 2.5$. The first barrel layer is located
1184 at a radius of 30 mm from the beamline, while the fourth layer is located at a radius
1185 of 160 mm closer to the strip tracker inner barrel layer (see section 3.3.3) in order to
1186 reduce the rate of fake tracks. The high granularity of the detector is represented in
1187 its about 123 Mpixels, each of size $100 \times 150\mu\text{m}^2$, which is almost twice the channels
1188 of the old detector. The transverse momentum resolution of tracks can be measured
1189 with a resolution of 1-2% for muons of $p_T = 100$ GeV.

1190

1191 Some of the improvements with respect to the previous pixel detector include a higher
1192 average tracking efficiency and lower average fake rate as well as higher track impact
1193 parameter resolution which is fundamental in order to increase the efficiency in the
1194 identification of jets originating from b quarks (b-tagging). A significant source of
1195 improvement comes from the overall reduction in the material budget of the detector
1196 which results in fewer photon conversions and less multiple scattering from charged

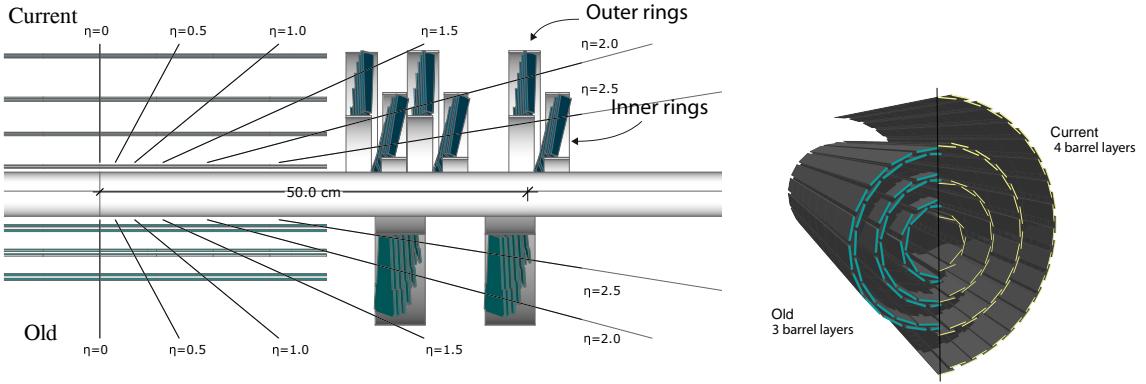


Figure 3.11: CMS pixel detector schematic view. Left: layout comparing the layers and disks in the old and current pixel detectors. Right: Transverse-oblique view comparing the pixel barrel layers in the two [70].

1197 particles.

1198 3.3.3 Silicon strip tracker

1199 The silicon strip tracker (SST) is the second stage in the CMS tracking system. The
 1200 top side of figure 3.12 shows a schematic of the SST. The inner tracker region is com-
 1201 posed of the tracker inner barrel (TIB) and the tracker inner disks (TID) covering the
 1202 region $r < 55$ cm and $|z| < 118$ cm. The TIB is composed of 4 layers while the TID
 1203 is composed of 3 disks at each end. The silicon sensors in the inner tracker are 320
 1204 μm thick, providing a resolution of about 13-38 μm in the $r\phi$ position measurement.
 1205

1206 The modules indicated in blue in the schematic view of figure 3.12 are two modules
 1207 mounted back-to-back and rotated in the plane of the module by a “stereo” angle of
 1208 100 mrad; the hits from these two modules, known as “stereo hits”, are combined to
 1209 provide a measurement of the second coordinate (z in the barrel and r on the disks)
 1210 allowing the reconstruction of hit positions in 3-D.

1211

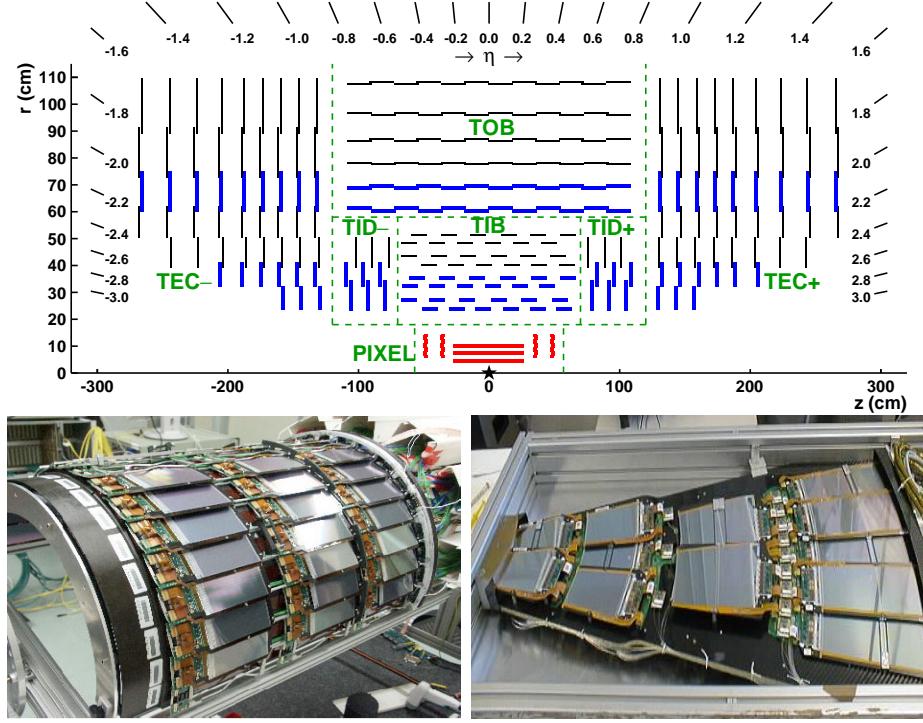


Figure 3.12: Top: CMS Silicon Strip Tracker (SST) schematic view. The SST is composed of the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). Each part is made of silicon strip modules; the modules in blue represent two modules mounted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

1212 The outer tracker region is composed of the tracker outer barrel (TOB) and the
 1213 tracker endcaps (TEC). The six layers of the TOB offer coverage in the region $r > 55$
 1214 cm and $|z| < 118$ cm, while the 9 disks of the TEC cover the region $124 < |z| < 282$
 1215 cm. The resolution offered by the outer tracker is about $13\text{--}38 \mu\text{m}$ in the $r\phi$ position
 1216 measurement. The inner four TEC disks use silicon sensors $320 \mu\text{m}$ thick; those in
 1217 the TOB and the outer three TEC disks use silicon sensors of $500 \mu\text{m}$ thickness. The
 1218 silicon strips run parallel to the z -axis and the distance between strips varies from 80
 1219 μm in the inner TIB layers to $183 \mu\text{m}$ in the inner TOB layers; in the endcaps the
 1220 wedge-shaped sensors with radial strips, whose pitch range between $81 \mu\text{m}$ at small
 1221 radii and $205 \mu\text{m}$ at large radii.

1222

1223 The whole SST has 15148 silicon modules, 9.3 million silicon strips and cover a total
 1224 active area of about 198 m^2 .

1225 **3.3.4 Electromagnetic calorimeter**

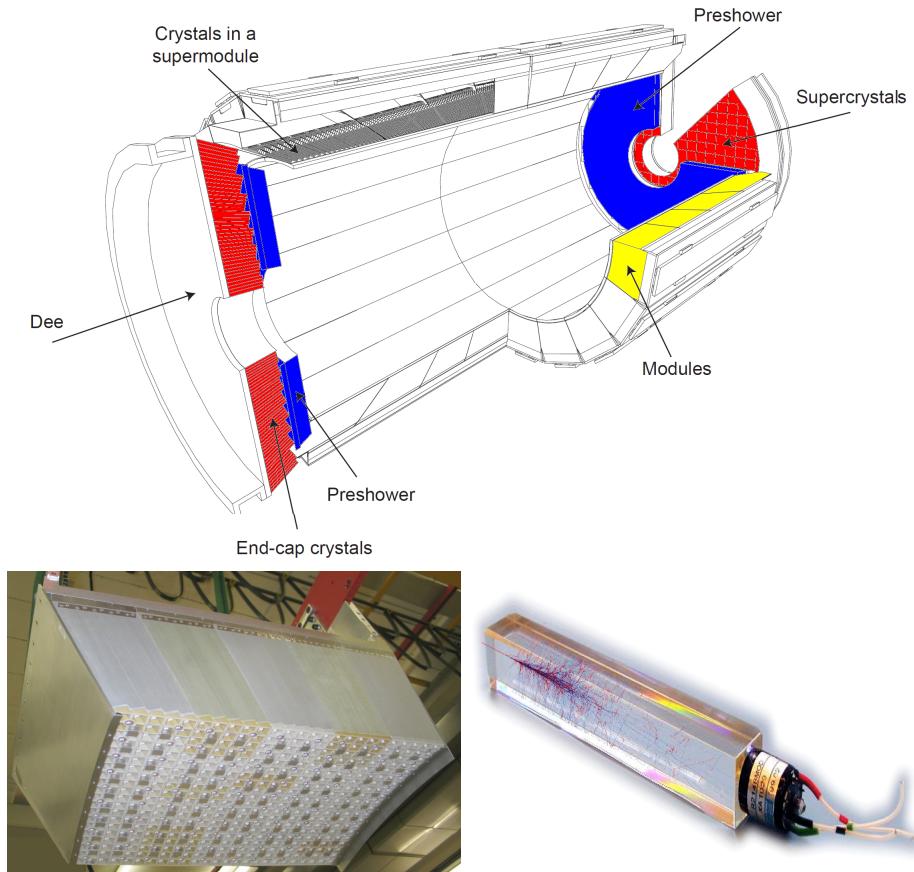


Figure 3.13: Top: CMS ECAL schematic view. Bottom: Module equipped with the crystals (left); ECAL crystal(right).

1226 The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of
 1227 electrons and photons. It is composed of 75848 lead tungstate crystals which have a
 1228 short radiation length (0.89 cm) and fast response, since 80% of the light is emitted
 1229 within 25 ns; however, they are combined with Avalanche photodiodes (APDs) as

1230 photodetectors given that crystals themselves have a low light yield ($30\gamma/\text{MeV}$). A
 1231 schematic view of the ECAL is shown in figure 3.13.

1232

1233 Energy is measured when electrons and photons are absorbed by the crystals which
 1234 generates an electromagnetic “shower”, as seen in bottom right picture of the fig-
 1235 ure3.13; the shower is seen as a *cluster* of energy which depending on the amount
 1236 of energy deposited can involve several crystals. The ECAL barrel (EB) covers the
 1237 region $|\eta| < 1.479$, using crystals of depth of 23 cm and $2.2 \times 2.2 \text{ cm}^2$ transverse
 1238 section; the ECAL endcap (EE) covers the region $1.479 < |\eta| < 3.0$ using crystals
 1239 of depth 22 cm and transverse section of $2.86 \times 2.86 \text{ cm}^2$; the photodetectors used
 1240 are vacuum phototriodes (VPTs). Each EE is divided in two structures called “Dees”.

1241

1242 In front of the EE, it is installed the preshower detector (ES) which covers the region
 1243 $1.653 < |\eta| < 2.6$. The ES provides a precise measurement of the position of electro-
 1244 magnetic showers, which allows to distinguish electrons and photons signals from π^0
 1245 decay signals. The ES is composed of a layer of lead absorber followed by a layer of
 1246 plastic scintillators

1247 3.3.5 Hadronic calorimeter

1248 Hadrons are not absorbed by the ECAL but by the hadron calorimeter (HCAL),
 1249 which is made of a combination of alternating brass absorber layers and silicon photo-
 1250 multiplier(SiPM) layers; therefore, particles passing through the scintillator material
 1251 produce showers, as in the ECAL, as a result of the inelastic scattering of the hadrons
 1252 with the detector material. Since the particles are not absorbed in the scintillator,
 1253 their energy is sampled; therefore the total energy is not measured but estimated from

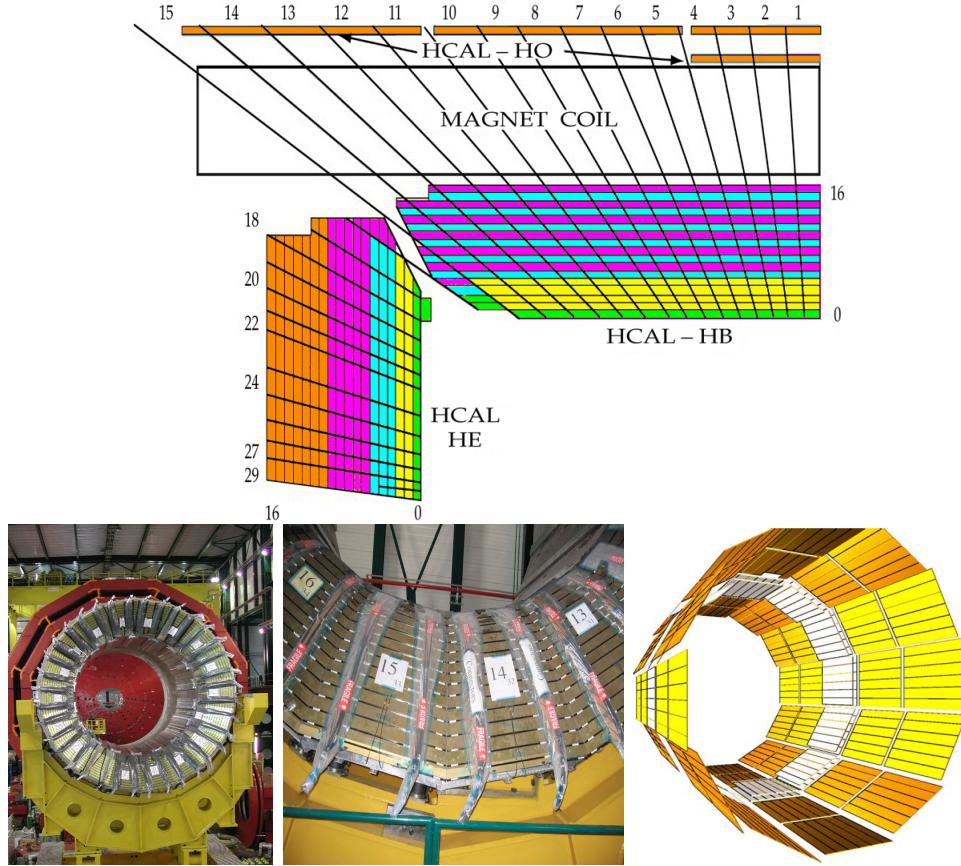


Figure 3.14: Top: CMS HCAL schematic view, the colors indicate the layers that are grouped into the same readout channels. Bottom: picture of a section of the HB; the absorber material is the golden region and scintillators are placed in between the absorber material (left and center). Schematic view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

1254 the energy clusters, which reduce the resolution of the detector. Brass was chosen
 1255 as the absorber material due to its short interaction length ($\lambda_I = 16.42\text{cm}$) and its
 1256 non-magnetivity. Figure 3.14 shows a schematic view of the CMS HCAL.

1257

1258 The HCAL is divided into four sections; the Hadron Barrel (HB), the Hadron Outer
 1259 (HO), the Hadron Endcap (HE) and the Hadron Forward (HF) sections. The HB
 1260 covers the region $0 < |\eta| < 1.4$, while the HE covers the region $1.3 < |\eta| < 3.0$. The HF,
 1261 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

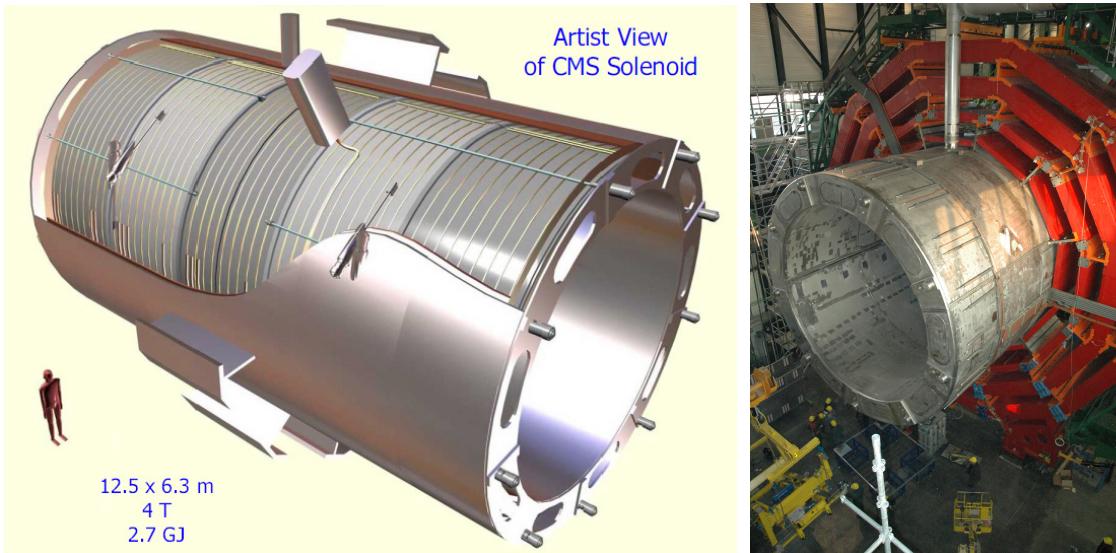


Figure 3.15: Artistic representation of the CMS solenoid magnet(left). The magnet is supported on an iron yoke (right) which also serves as the house of the muon detector and as mechanical support for the whole CMS detector [69].

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

1275 m and a cold mass of 220 t; the generated magnetic field reaches a strength of 3.8T.
 1276 Since it is made of Ni-Tb superconducting cable it has to operate at a temperature
 1277 of 4.7 K by using a helium cryogenic system; the current circulating in the cables
 1278 reaches 18800 A under normal running conditions. The left side of figure 3.15 shows
 1279 an artistic view of the CMS magnet, while the right side shows a transverse view of
 1280 the cold mass where the winding structure is visible.

1281

1282 The yoke (see figure 3.15), composed of 5 barrel wheels and 6 endcap disks made
 1283 of iron, serves not only as the media for magnetic flux return but also provides the
 1284 house for the muon detector system and structural stability to the full detector.

1285 3.3.7 Muon system

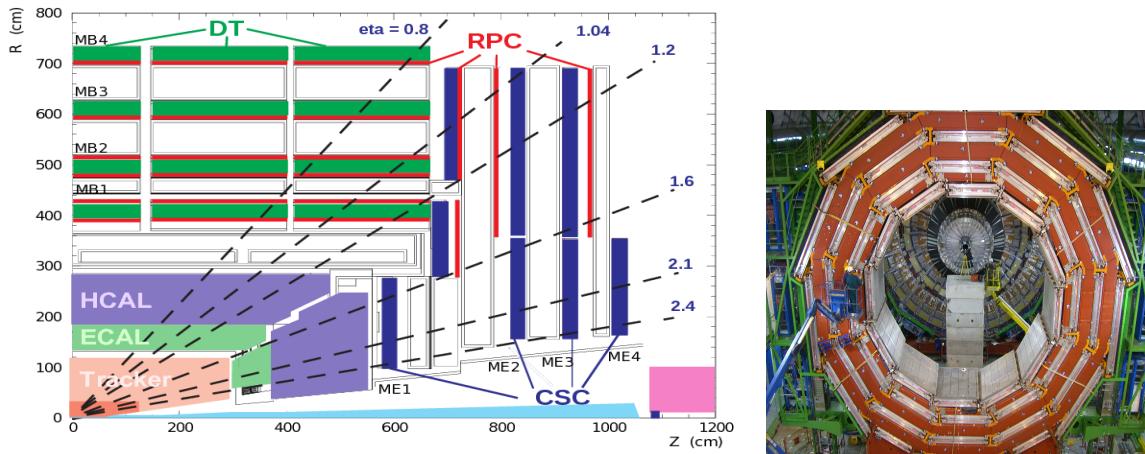


Figure 3.16: Left: CMS muon system schematic view; Right: one of the yoke rings with the muon DTs and RPCs installed; in the back it is possible to see the muon endcap [76].

1286 Muons are the only charged particles able to pass through all the CMS detector due
 1287 to their low ionization energy loss; thus, muons can be separated easily from the
 1288 high amount of particles produced in a pp collision. Also, muons are expected to be
 1289 produced in the decay of several new particles; therefore, a good detection of muons

1290 was on the leading principles when designing the CMS detector.

1291

1292 The CMS muon detection system (muon spectrometer) is embedded in the return
1293 yoke as seen in figure 3.16. It is composed of three different detector types, the drift
1294 tube chambers (DT), Cathode strip chambers (CSC), and resistive plate chambers
1295 (RPC); DT are located in the central region $\eta < 1.2$ arranged in four layers of drift
1296 chambers filled with an Ar/CO₂ gas mixture.

1297

1298 The muon endcaps are made of CSCs covering the region $\eta < 2.4$ and filled with a
1299 mixture of Ar/CO₂/CF₄. The reason behind using a different detector type lies on
1300 the different conditions in the forward region like the higher muon rate and higher
1301 residual magnetic field compared to the central region.

1302

1303 The third type of detector used in the muon system is a set of four disks of RPCs
1304 working in avalanche mode. The RPCs provide good spatial and time resolutions.
1305 The track of $high - p_T$ muon candidates is built combining information from the
1306 tracking system and the signal from up to six RPCs and four DT chambers.

1307 The muon tracks are reconstructed from the hits in the several layers of the muon
1308 system.

1309 **3.3.8 CMS trigger system**

1310 Under normal conditions, CMS expects pp collisions every 25 ns, i.e., an interaction
1311 rate of 40 MHz for which it is not possible to store the recorded data in full. In order
1312 to handle this high event rate data, an online event selection, known as triggering, is
1313 performed; triggering reduce the event rate to 100 Hz for storage and further offline

1314 analysis.

1315

1316 The trigger system starts with a reduction of the event rate to 100 kHz in the so-called
 1317 “level 1 trigger (L1)”. L1 is based on dedicated programmable hardware like Field
 1318 Programmable Gate Arrays (FPGAs) and Application Specific Integrated Circuits
 1319 (ASICs), partly located in the detector itself; another portion is located in the CMS
 1320 under-ground cavern. Hit patterns information from the muon chambers and the en-
 1321 ergy deposits in the calorimeter are used to decide if an event is accepted or rejected,
 1322 according to selection requirements previously defined, which reflect the interesting
 1323 physics processes. Figure 3.17 shows the L1 trigger architecture.

1324

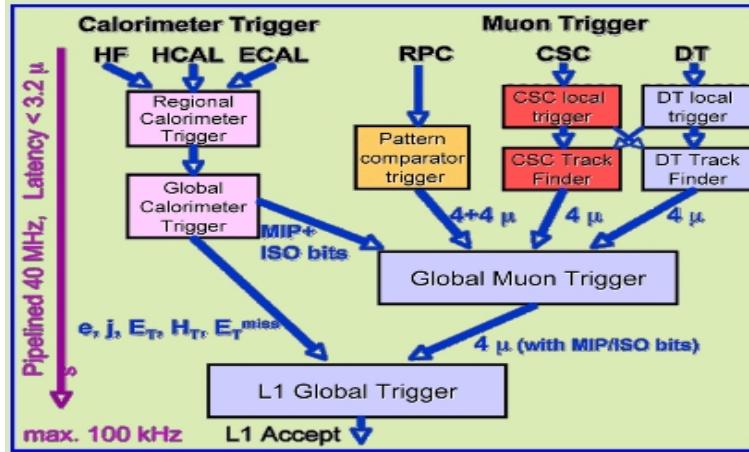


Figure 3.17: CMS Level-1 trigger architecture [77].

1325 The second stage in the trigger system is called “high-level trigger (HLT)”; events
 1326 accepted by L1 are passed to HLT in order to make an initial reconstruction of them.
 1327 HLT is software based and runs on a dedicated server farm, using selection algo-
 1328 rithms and high-level object definitions; the event rate at HLT is reduced to 100 Hz.
 1329 The first HLT stage takes information from the muon detectors and the calorimeters
 1330 to make the initial object reconstruction; in the next HLT stage, information from

1331 the pixel and strip detectors is used to do first fast-tracking and then full tracking
1332 online. This initial object reconstruction is used in further steps of the trigger system.

1333

1334 Events and preliminary reconstructed physics objects from HLT are sent to be fully
1335 reconstructed at the CERN computing center. Again, the pixel detector information
1336 provides high-quality seeds for the track reconstruction algorithm offline, primary ver-
1337 tex reconstruction, electron and photon identification, muon reconstruction, τ iden-
1338 tification, and b-tagging. After full reconstruction, data sets are made available for
1339 offline analyses.

1340

1341 During the 2016-2017 technical stop, the L1 system was updated in order to improve
1342 the physics object identification by improving the algorithms and accounting for the
1343 increasing pile-up scenario.

1344 **3.3.9 CMS computing**

1345 After the data, coming from the experiment, are processed at several levels, they have
1346 to be stored and made available for further analysis; in order to cope all the tasks
1347 implied in the offline data processing, like transfer, simulation, reconstruction and
1348 reprocessing, among others, a big computing power is required. The CMS computing
1349 system is based on the distributed architecture concept, where users of the system
1350 and physical computer centers are distributed worldwide and interconnected by high-
1351 speed networks.

1352 The worldwide LHC computing grid (WLCG) is the mechanism used to provide that
1353 distributed environment. WLCG is a tiered structure connecting computing centers
1354 around the world, which provides the necessary storage and computing facilities. The

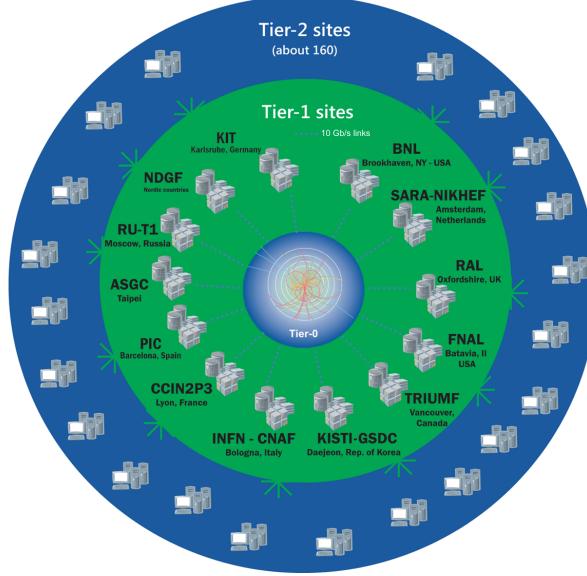


Figure 3.18: WLCG structure. The primary computer centers (Tier-0) are located at CERN (data center) and at the Wigner datacenter in Budapest. Tier-1 is composed of 13 centers and Tier-2 is composed of about 160 centers. [78].

1355 primary computing centers of the WLCG are located at the CERN and the Wigner
 1356 datacenter in Budapest and are known as Tier-0 as shown in figure 3.18. The main
 1357 responsibilities for each tier level are [78]

- 1358 • **Tier-0:** initial reconstruction of recorded events and storage of the resulting
 1359 datasets, the distribution of raw data to the Tier-1 centers.
- 1360 • **Tier-1:** provide storage capacity, support for the Grid, safe-keeping of a pro-
 1361 portional share of raw and reconstructed data, large-scale reprocessing and safe-
 1362 keeping of corresponding output, generation of simulated events, distribution
 1363 of data to Tier 2s, safe-keeping of a share of simulated data produced at these
 1364 Tier 2s.
- 1365 • **Tier-2:** store sufficient data and provide adequate computing power for specific
 1366 analysis tasks, provide analysis requirements and proportional share of simu-
 1367 lated event production and reconstruction.

1368 Aside from the general computing strategy to manage the huge amount of data pro-
1369 duced by experiments, CMS uses a framework to perform a variety of processing,
1370 selection and analysis tasks. The central concept of the CMS data model referred to
1371 as “event data model” (EDM) is the “Event”; therefore, an event is the unit that con-
1372 tains the information from a single bunch crossing as well as any data derived from
1373 that information like the reconstructed objects, the details under which additional
1374 data are derived.

1375

1376 Events are passed as the input to the “physics modules” that obtain information from
1377 them and create new one; for instance, “event data producers” add new data into the
1378 events, “analyzers” produce an information summary from an event set, “filters” per-
1379 form selection and triggering.

1380

1381 CMS uses several event formats with different levels of detail and precision

1382 • **Raw format:** events in this format contain the full recorded information from
1383 the detector as well as trigger decision and other metadata. An extended version
1384 of raw data is used to store information from the CMS Monte Carlo simulation
1385 tools. Raw data are stored permanently, occupying about 2MB/event

1386 • **RECO format:** events in this format correspond to raw data that have been
1387 submitted to reconstruction algorithms like primary and secondary vertex re-
1388 construction, particle ID, track-finding. RECO events contain physical objects
1389 and all the information used to reconstruct them; average size is about 0.5
1390 MB/event.

1391 • **AOD format:** Analysis Object Data (AOD) is the data format used in the
1392 physics analyses given that it contains the parameters describing the high-level

1393 physics objects in addition to enough information to allow a kinematic refitting if
 1394 needed. AOD events are filtered versions of the RECO events to which skimming
 1395 or other kind processes have been applied. Requires about 100 kB/event.

1396 • **Non-event data** are data needed to interpret and reconstruct events. Some
 1397 of the non-event data used by CMS contains information about the detector
 1398 contraction and condition data like calibrations, alignment, and detector status.

1399 Figure 3.19 shows the data flow scheme between CMS detector and hardware tiers.

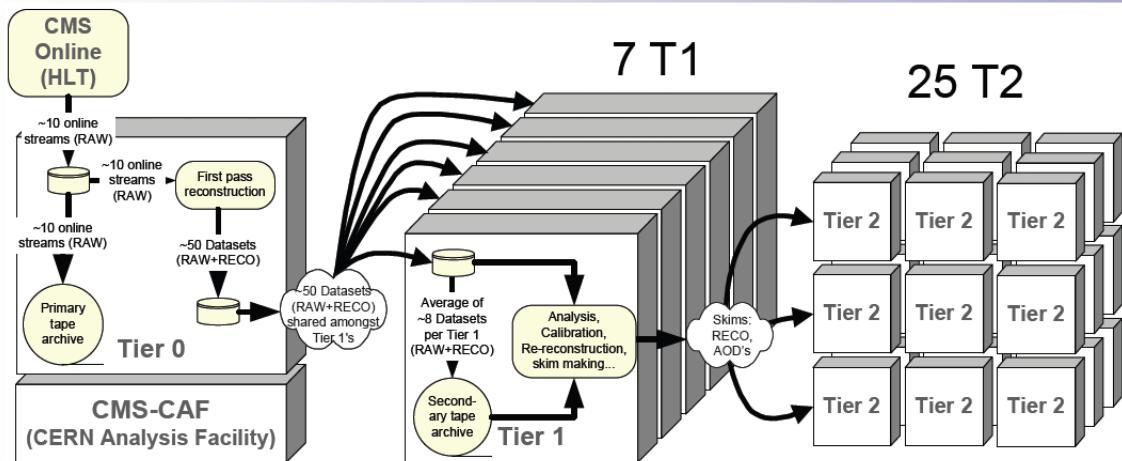


Figure 3.19: Data flow from CMS detector through hardware Tiers.

1400 The whole collection of software built as a framework is referred to as “CMSSW”. This
 1401 framework provides the services needed by the simulation, calibration and alignment,
 1402 and reconstruction modules that process event data, so that physicists can perform
 1403 analysis. The CMSSW event processing model is composed of one executable, called
 1404 cmsRun, and several plug-in modules which contains all the tools (calibration, recon-
 1405 struction algorithms) needed to process an event. The same executable is used for
 1406 both detector and Monte Carlo data [79].

1407 **Chapter 4**

1408 **Event generation, simulation and
1409 reconstruction**

1410 The process of analyzing the data recorded by the CMS experiment involves several
1411 stages where the data are processed in order to interpret the information provided by
1412 all the detection systems; in those stages, the particles produced after the pp collision
1413 are identified by reconstructing their trajectories and measuring their features. In
1414 addition, the SM provides a set of predictions that have to be compared with the
1415 experimental results; however, in most of the cases, theoretical predictions are not
1416 directly comparable to experimental results due to the diverse source of uncertainties
1417 introduced by the experimental setup and theoretical approximations among others.

1418

1419 The strategy to face these conditions consist in using statistical methods implemented
1420 in computational algorithms to produce numerical results that can be contrasted with
1421 the experimental results. These computational algorithms are commonly known as
1422 Monte Carlo (MC) methods and, in the case of particle physics, they are designed to
1423 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.

1429

1430 This chapter presents a description of the event generation strategy and the tools
 1431 used to perform the detector simulation and physics objects reconstruction. A com-
 1432 prehensive review of event generators for LHC physics can be found in reference [80]
 1433 on which this chapter is based.

1434 4.1 Event generation

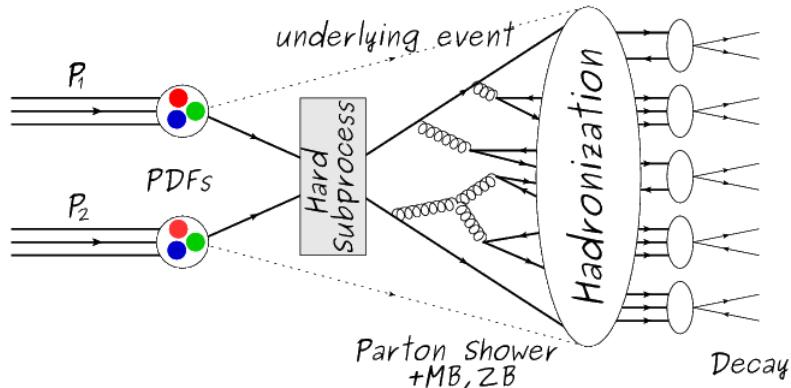


Figure 4.1: Event generation process. In the first step, the PDF of the colliding particles is considered so the specific interaction is described. The actual interaction is generated in the hard subprocess; the cross-section of the process is calculated from the matrix element connecting the initial 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].

1435 The event generation is intended to create events that mimic the behavior of actual
 1436 events produced in the collisions; they obey a sequence of steps from the particles colli-
 1437 sion hard process to the decay process into the final state particles. Figure 4.1 shows
 1438 a schematic view of the event generation process; the fact that the full process can
 1439 be treated as several independent steps is based on the QCD factorization theorem.

1440

1441 Generation starts by taking into account the PDFs of the incoming particles. Event
 1442 generators offer the option to choose from several PDF sets depending on the partic-
 1443 ular process under simulation¹; in the following pp collisions will be considered. The
 1444 *hard subprocess* describes the actual interaction between partons from the incoming
 1445 protons; it is represented by the matrix element connecting the initial and final states
 1446 of the interaction. Normally, the matrix element can be written as a sum over Feyn-
 1447 man diagrams and consider interferences between terms in the summation. During
 1448 the generation of the hard subprocess, the production cross section is calculated.

1449

1450 The order to which the cross section is calculated depends on the order of the Feyn-
 1451 man diagrams involved in the calculation; therefore, radiative corrections are included
 1452 by considering a higher order Feynman diagrams where QCD radiation dominates.
 1453 Currently, cross sections calculated to LO do not offer a satisfactory description of the
 1454 processes, i.e., the results are only reliable for the shape of distributions; therefore,
 1455 NLO calculations have to be performed with the implication that the computing time
 1456 needed is highly increased.

1457

1458 The final parton content of the hard subprocess is subjected to the *parton shower*
 1459 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.

1460 split into quark-antiquark pairs and quarks of enough energy radiate gluons giving rise
 1461 to further parton multiplication, following the DGLAP (Dokshitzer-Gribov-Lipatov-
 1462 Altarelli-Parisi) equations. Showering continues until the energy scale is low enough
 1463 to reach the non-perturbative limit.

1464

1465 In the simulation of LHC processes that involve b quarks like the single top quark or
 1466 Higgs associated production, it is needed to consider that the b quark is heavier than
 1467 the proton; in this sense, the QCD interaction description is made in two different
 1468 schemes [83]

1469 • four-flavor (4F) scheme. b quarks appear only in the final state because they
 1470 are heavier than the proton and therefore they can be produced only from the
 1471 splitting of a gluon into pairs or singly in association with a t quark in high
 1472 energy-scale interactions. During the simulation, the b -PDFs are set to zero
 1473 because it cannot be part of the proton. Calculations in this scheme are more
 1474 complicated due to the presence of the second b quark but the full kinematics is
 1475 considered already at LO and therefore the accuracy of the description is better.

1476 • five-flavor (5F) scheme. b quarks are considered massless, therefore they can
 1477 appear in both initial and final states since it can now be part of the proton; thus,
 1478 during the simulation b -PDFs are not set to zero. In this scheme, calculations
 1479 are simpler than in the 4F scheme and possible logarithmic divergences are
 1480 absorbed by the PDFs through the DGLAP evolution.

1481 In this thesis, the tHq events are generated using the 4F scheme in order to reduce
 1482 uncertainties, while the tHW events are generated using the 5F scheme to eliminate
 1483 LO interference with the $t\bar{t}H$ process [48].

1484

1485 Partons involved in the pp collision are the focus of the simulation, however, the rest
 1486 of the partons inside the incoming protons are also affected because the remnants are
 1487 colored objects; also, multiple parton interactions can occur. The hadronization of
 1488 the remnants and multiple parton interactions are known as “underlying event” and
 1489 it has to be included in the simulation. In addition, multiple pp collisions in the same
 1490 bunch crossing (pile-up mentioned in 3.2) occurs, actually in two forms

- 1491 • *in-time PU* which refers to multiple pp collision in the bunch crossing but that
 1492 are not considered as primary vertices.
- 1493 • *Out-of-time PU* which refers to overlapping pp collisions from consecutive bunch
 1494 crossings; this can occurs due to the time-delays in the detection systems where
 1495 information from one bunch crossing is assigned to the next or previous one.

1496 While the underlying event effects are included in generation using generator-specific
 1497 tools, PU effects are added to the generation by overlaying Minimum-bias (MB) and
 1498 Zero-bias (ZB) events to the generated events. MB events are inelastic events selected
 1499 by using a loose (minimum bias) trigger with as little bias as possible, therefore ac-
 1500 cepting a large fraction of the overall inelastic event; ZB events correspond to random
 1501 events recorded by the detector when collisions are likely. MB model in-time PU and
 1502 ZB model out-of-time PU.

1503

1504 The next step in the generation process is called “hadronization”. Since particles
 1505 with a net color charge are not allowed to exits isolated, they have to recombine
 1506 to form bound states. This is precisely the process by which the partons resulting
 1507 from the parton shower arrange themselves as color singlets to form hadrons. At
 1508 this step, the energy-scale is low and the strong coupling constant is large, there-
 1509 fore hadronization process is non-perturbative and the evolution of the partons is

1510 described using phenomenological models. Most of the baryons and mesons produced
 1511 in the hadronization are unstable and hence they will decay in the detector.

1512

1513 The last step in the generation process corresponds to the decay of the unstable
 1514 particles generated during hadronization; it is also simulated in the hadronization
 1515 step, based on the known branching ratios.

1516 4.2 Monte Carlo Event Generators.

1517 The event generation described in the previous section has been implemented in
 1518 several software packages for which a brief description is given.

- 1519 • **PYTHIA 8.** It is a program designed to perform the generation of high en-
 1520 ergy physics events which describe the collisions between particles such as elec-
 1521 trons, protons. Several theories and models are implemented in it, in order to
 1522 describe physical aspects like hard and soft interaction, parton distributions,
 1523 initial and final-state parton showers, multiple parton interactions, beam rem-
 1524 nants, hadronization² and particle decay. Thanks to extensive testing, several
 1525 optimized parametrizations, known as “tunings”, have been defined in order
 1526 to improve the description of actual collisions to a high degree of precision; for
 1527 analysis at $\sqrt{s} = 13$ TeV, the underline event CUETP8M1 tune is employed [85].
 1528 The calculation of the matrix element is performed at LO which is not enough
 1529 for the current required level of precision; therefore, pythia is often used for
 1530 parton shower, hadronization and decays, while other event generators are used
 1531 to generate the matrix element at NLO.

² based in the Lund string model [84]

1532 • **MadGraph5_aMC@NLO.** MadGraph is a matrix element generator which
 1533 calculates the amplitudes for all contributing Feynman diagrams of a given pro-
 1534 cess but does not provide a parton shower while MC@NLO incorporate NLO
 1535 QCD matrix elements consistently into a parton shower framework; thus, Mad-
 1536 Graph5_aMC@NLO, as a merger of the two event generators MadGraph5 and
 1537 aMC@NLO, is an event generator capable to calculate tree-level and NLO cross
 1538 sections and perform the matching of those with the parton shower. It is one of
 1539 the most frequently used matrix element generators; however, it has the partic-
 1540 ular feature of the presence of negative event weights which reduce the number
 1541 of events used to reproduce the properties of the objects generated [86].

1542

1543 • **POWHEG.** It is an NLO matrix element generator where the hardest emis-
 1544 sion of color charged particles is generated in such a way that the negative event
 1545 weights issue of MadGraph5_aMC@NLO is overcome; however, the method re-
 1546 quires an interface with p_T -ordered parton shower or a parton shower generator
 1547 where this highest emission can be vetoed in order to avoid double counting of
 1548 this highest-energetic emission. PYTHIA is a commonly matched to POWHEG
 1549 event generator [87].

1550 Events resulting from the whole generation process are known as MC events.

1551 4.3 CMS detector simulation.

1552 After generation, MC events contain the physics of the collisions but they are not
 1553 ready to be compared to the events recorded by the experiment since these recorded
 1554 events correspond to the response of the detection systems to the interaction with

1555 the particles traversing them. The simulation of the CMS detector has to be applied
1556 on top of the event generation; it is simulated with a MC toolkit for the simulation
1557 of particles passing through matter called Geant4 which is also able to simulate the
1558 electronic signals that would be measured by all detectors inside CMS.

1559

1560 The simulation takes the generated particles contained in the MC events as input,
1561 makes them pass through the simulated geometry, and models physics processes that
1562 particles experience during their passage through matter. The full set of results from
1563 particle-matter interactions correspond to the simulated hit which contains informa-
1564 tion about the energy loss, momentum, position. Particles of the input event are
1565 called “primary”, while the particles originating from GEANT4-modeled interactions
1566 of a primary particle with matter are called a “secondary”. Simulated hits are the in-
1567 put of subsequent modules that emulate the response of the detector readout system
1568 and triggers. The output from the emulated detection systems and triggers is known
1569 as digitization [88, 89].

1570

1571 The modeling of the CMS detector corresponds to the accurate modeling of the
1572 interaction among particles, the detector material, and the magnetic field. This
1573 simulation procedure includes the following standard steps

- 1574 • Modeling of the Interaction Region.
- 1575 • Modeling of the particle passage through the hierarchy of volumes that compose
1576 CMS detector and of the accompanying physics processes.
- 1577 • Modeling of the effect of multiple interactions per beam crossing and/or the
1578 effect of events overlay (Pile-Up simulation).

1579 • Modeling of the detector’s electronics response, signal shape, noise, calibration
 1580 constants (digitization).

1581 In addition to the full simulation, i.e. a detailed detector simulation, a faster simula-
 1582 tion (FastSim) have been developed, that may be used where much larger statistics
 1583 are required. In FastSim, detector material effects are parametrized and included in
 1584 the hits; those hits are used as input of the same higher-level algorithms³ used to an-
 1585 alyze the recorded events. In this way, comparisons between fast and full simulations
 1586 can be performed [91].

1587

1588 After the full detector simulation, the output events can be directly compared with
 1589 events actually recorded in the CMS detector. The collection of MC events that
 1590 reproduce the expected physics for a given process are known as MC samples.

1591 4.4 Event reconstruction.

1592 In contrast to MC samples for which all the particles’ information is available from
 1593 it’s identity to its mass and energy, recorded events contain the electronic signals,
 1594 provided by the CMS detection systems, encoding the interaction of physical parti-
 1595 cles with the detector matter; these electronic signals have to be combined in order
 1596 to identify these particles and measure their features i.e., particles have to be “recon-
 1597 structed” using the signals provided by the detection systems. The CMS experiment
 1598 use the “particle-flow event reconstruction algorithm (PF)” to do the reconstruction
 1599 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

1600 of the *Elements* used by PF (tracker tracks, energy clusters, and muon tracks), based
 1601 in the references [92, 93] where more detailed descriptions can be found.

1602 **4.4.1 Particle-Flow Algorithm.**

1603 Each of the several sub detection systems of the CMS detector is dedicated to identi-
 1604 fying a specific type of particles, i.e., photons and electrons are absorbed by the ECAL
 1605 and their reconstruction is based on ECAL information; hadrons are reconstructed
 1606 from clusters in the HCAL while muons are reconstructed from hits in the muon
 1607 chambers. PF is designed to correlate signals from all the detector layers (tracks and
 1608 energy clusters) in order to reconstruct and identify each final state particle and its
 1609 properties as sketched in figure 4.2.

1610

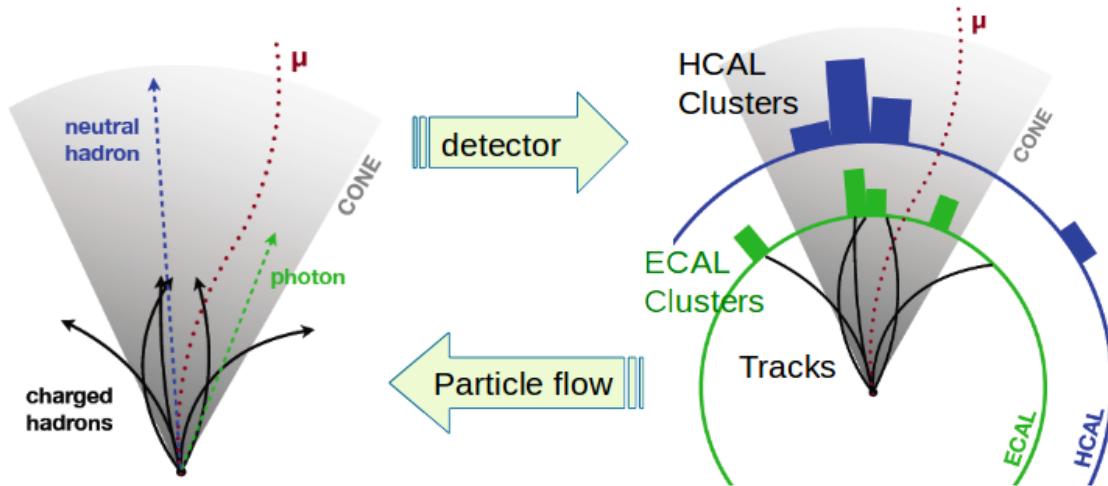


Figure 4.2: Particle flow algorithm. Information from the several CMS detection systems if provided as input to the algorithm which then combine it to identify and reconstruct all the particles in the final state and their properties. Reconstruction of simulated events is also performed by providing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

1611 For instance, a charged hadron is identified by a geometrical connection, know as *link*
 1612 between one or more calorimeter clusters and a track in the tracker provided there

1613 are no hits in the muon system; combining several measurements allows a better
 1614 determination of the energy and charge sign of the charged hadron.

1615 **Charged-particle track reconstruction.**

1616 The strategy used by PF in order to reconstruct tracks is called “Iterative Tracking”
 1617 which occurs in four steps

- 1618 • Seed generation where initial track candidates are found by looking for a combi-
 1619 nation of hits in the pixel detector, strip tracker, and muon chambers. In total
 1620 ten iterations are performed, each one with a different seeding requirement.
 1621 Seeds are used to estimate the trajectory parameters and uncertainties at the
 1622 time of the full track reconstruction. Seeds are also considered track candidates.

- 1623 • Track finding using a tracking software known as Combinatorial Track Finder
 1624 (CTF) [95]. The seed trajectories are extrapolated along the expected flight
 1625 path of a charged particle, in agreement to the trajectory parameters obtained
 1626 in the first step, in an attempt to find additional hits that can be assigned to
 1627 the track candidates.

- 1628 • Track-fitting where the found tracks are passed as input to a module which
 1629 provides the best estimate of the parameters of each trajectory.

- 1630 • Track selection where track candidates are submitted to a selection which dis-
 1631 cards those that fail a set of defined quality criteria.

1632 Iterations differ in the seeding configuration and the final track selection as elaborated
 1633 in references [92, 93]. In the first iteration, high p_T tracks and tracks produced near
 1634 to the interaction region are identified and those hits are masked thereby reducing
 1635 the combinatorial complexity. Next iterations search for more complicated tracks,

1636 like low p_T tracks and tracks from b hadron decays, which tend to be displaced from
 1637 the interaction region.

1638 **Vertex reconstruction.**

1639 During the track reconstruction, an extrapolation toward to the calorimeters is per-
 1640 formed in order to match energy deposits; that extrapolation is performed also toward
 1641 the beamline in order to find the origin of the track known as *vertex*. The vertex re-
 1642 construction is performed by selecting from the available reconstructed tracks, those
 1643 that are consistent with being originated in the interaction region where pp collisions
 1644 are produced. The selection involves a requirement on the number of tracker (pixel
 1645 and strip) hits and the goodness of the track fit.

1646

1647 Selected tracks are clustered using a “deterministic annealing algorithm (DA)”⁴. A
 1648 set of candidate vertices and their associated tracks, resulting from the DA, are then
 1649 fitted with an “adaptive vertex fitter (AVF)” to produce the best estimate of the
 1650 vertices locations.

1651

1652 The p_T of the several tracks associated to a reconstructed vertex is added, squared and
 1653 used to organize the vertices; the vertex with the highest squared sum is designated
 1654 as the *primary vertex (PV)* while the rest are designated as PU vertices.

1655 **Calorimeter clustering.**

1656 After traversing the CMS tracker system, electrons, photons and hadrons deposit their
 1657 energy in the ECAL and HCAL cells. The PF clustering algorithm aims to provide
 1658 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]

1659 between close energy deposits. The clustering runs independently in the ECAL barrel
 1660 and endcaps, HCAL barrel and endcaps, and the two preshower layers, following two
 1661 steps

- 1662 • cells with an energy larger than a given seed threshold and larger than the energy
 of the neighboring cells are identified as cluster seeds. The neighbor cells are
 those that either share a side with the cluster seed candidate, or the eight closest
 1665 cells including cells that only share a corner with the seed candidate.
- 1666 • cells with at least a corner in common with a cell already in the cluster seed
 1667 and with an energy above a cell threshold are grouped into topological clusters.

1668 Clusters formed in this way are known as *particle-flow clusters*. With this clustering
 1669 strategy, it is possible to detect and measure the energy and direction of photons and
 1670 neutral hadrons as well as differentiate these neutral particles from the charged hadron
 1671 energy deposits. In cases involving charged hadrons for which the track parameters
 1672 are not determined accurately, for instance, low-quality and high- p_T tracks, clustering
 1673 helps in the energy measurements.

1674 **Electron track reconstruction.**

1675 Although the charged-particle track reconstruction described above works for elec-
 1676 trons, they lose a significant fraction of their energy via bremsstrahlung photon radi-
 1677 ation before reaching the ECAL; thus, the reconstruction performance depends on the
 1678 ability to measure also the radiated energy. The reconstruction strategy, in this case,
 1679 requires information from the tracking system and from the ECAL. Bremsstrahlung
 1680 photons are emitted at similar η values to that of the electron but at different values
 1681 of ϕ ; therefore, the radiated energy can be recovered by grouping ECAL clusters in a
 1682 η window over a range of ϕ around the electron direction. The group is called ECAL

1683 supercluster.

1684

1685 Electron candidates from the track-seeding and ECAL super clustering are merged
 1686 into a single collection which is submitted to a full electron tracking fit with a
 1687 Gaussian-sum filter (GSF) [96]. The electron track and its associated ECAL su-
 1688 percluster form a *particle-flow electron*.

1689 **Muon track reconstruction.**

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

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

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

1708 **Particle identification and reconstruction.**

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

1716 • Muons. An “isolated global muon” is identified by evaluating the presence of
 1717 inner track and energy deposits close to the global muon track in the (η, ϕ)
 1718 plane, i.e., in a particular point of the global muon track, inner tracks and
 1719 energy deposits are sought within a radius of $\Delta R = 0.3$ (see eqn. 3.7) from the
 1720 muon track; if they exist and the p_T of the found track added to the E_T of the
 1721 found energy deposit does not exceed 10% of the muon p_T then the global muon
 1722 is an isolated global muon. This isolation condition is stringent enough to reject
 1723 hadrons misidentified as muons.

1724 “Non-isolated global muons” are identified using additional selection require-
 1725 ments on the number of track segments in the muon system and energy deposits
 1726 along the muon track. Muons inside jets are identified with more stringent crite-
 1727 ria in isolation and momentum as described in reference [99]. The PF elements
 1728 associated with an identified muon are masked from the PF block.

- 1729 ● Electrons are identified and reconstructed as described above plus some addi-
 1730 tional requirements on fourteen variables like the amount of energy radiated,
 1731 the distance between the extrapolated track position at the ECAL and the po-
 1732 sition of the associated ECAL supercluster among others, which are combined
 1733 in a specialized multivariate analysis strategy that improves the electron iden-
 1734 tification. Tracks and clusters used to identify and reconstruct electrons are
 1735 masked in the PF block.

- 1736 ● Isolated photons are identified from ECAL superclusters with E_T larger than 10
 1737 GeV, for which the energy deposited at a distance of 0.15, from the supercluster
 1738 position on the (η, ϕ) plane, does not exceed 10% of the supercluster energy;
 1739 note that this is an isolation requirement. In addition, there must not be links
 1740 to tracks. Clusters involved in the identification and reconstruction are masked
 1741 in the PF block.

- 1742 ● Bremsstrahlung photons and prompt photons tend to convert to electron-positron
 1743 pairs inside the tracker, therefore, a dedicated finder algorithm is used to link
 1744 tracks that seem to originate from a photon conversion; in case those two tracks
 1745 are compatible with the direction of a bremsstrahlung photon, they are also
 1746 linked to the original electron track. Photon conversion tracks are also masked
 1747 in the PF block.

- 1748 ● The remaining elements in the PF block are used to identify hadrons. In the
 1749 region $|\eta| \leq 2.5$, neutral hadrons are identified with HCAL clusters not linked
 1750 to any track while photons from neutral pion decays are identified with ECAL
 1751 clusters without links to tracks. In the region $|\eta| > 2.5$ ECAL clusters linked to
 1752 HCAL clusters are identified with a charged or neutral hadron shower; ECAL
 1753 clusters with no links are identified with photons. HCAL clusters not used yet,

1754 are linked to one or more unlinked tracks and to an unlinked ECAL in order to
 1755 reconstruct charged-hadrons or a combination of photons and neutral hadrons
 1756 according to certain conditions on the calibrated calorimetric energy.

- 1757 • Charged-particle tracks may be liked together when they converge to a “sec-
 1758 ondary vertex (SV) ” displaced from the interaction point where the PV and
 1759 PU vertices are reconstructed; at least three tracks are needed in that case,
 1760 of which at most one has to be an incoming track with hits in tracker region
 1761 between a PV and the SV.

1762

1763 The linker algorithm, as well as the whole PF algorithm, has been validated and
 1764 commissioned; results from that validation are presented in the references [92].

1765 **Jet reconstruction.**

1766 Quarks and gluons may be produced in the pp collisions, therefore, their hadronization
 1767 will be seen in the detector as a shower of hadrons and their decay products in the
 1768 form of a “jet”. The anti- k_t algorithm [100] is used to perform the jet reconstruction
 1769 by clustering those PF particles within a cone (see figure 4.3); previously, isolated
 1770 electrons, isolated muons, and charged particles associated with other interaction
 1771 vertices are excluded from the clustering.

1772 The anti- k_t algorithm proceeds in a sequential recombination of PF particles; the
 1773 distance between particles i and j (d_{ij}) and the distance between particles and the
 1774 beam are defined as

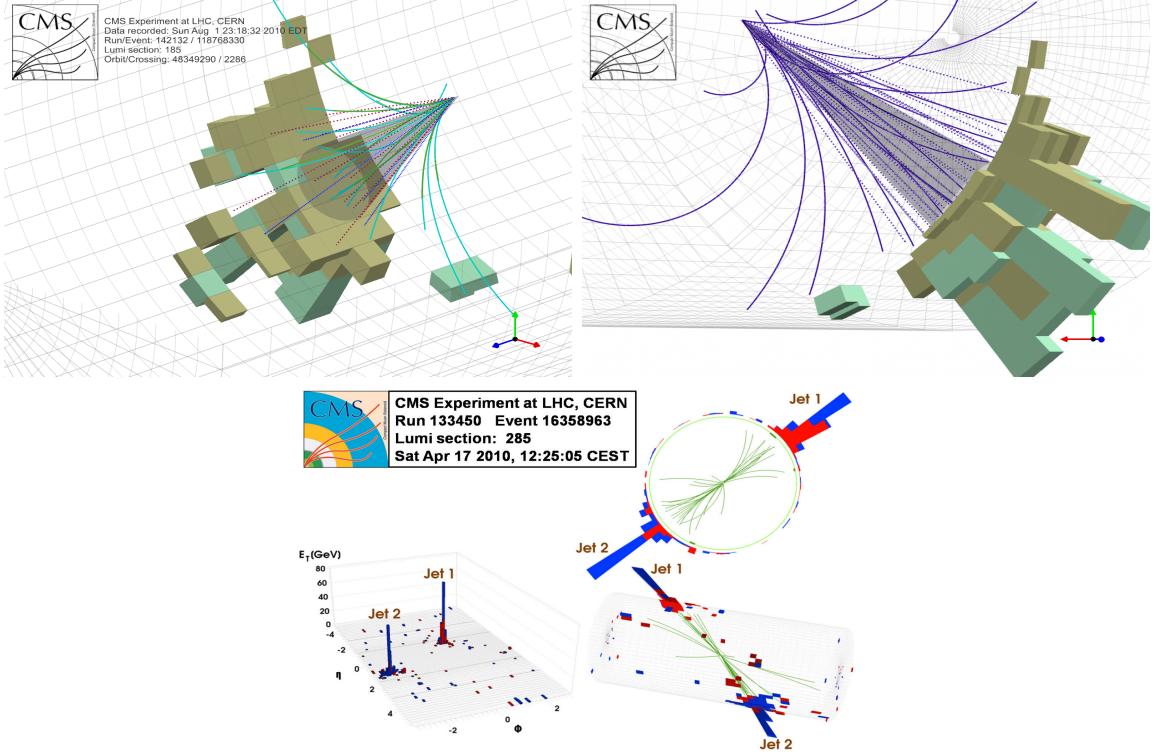


Figure 4.3: Jet reconstruction performed by the anti- k_t algorithm. Top: Two different views of a CMS recorded event are presented. Continuous lines correspond to the tracks left by charged particles in the tracker while dotted lines are the imaginary paths followed by neutral particles. The green cubes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

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

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

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

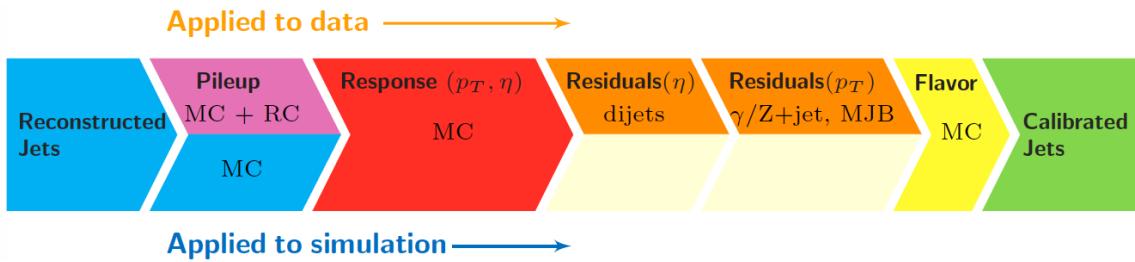


Figure 4.4: Jet energy correction diagram. Correction levels are applied sequentially in the indicated fixed order [104].

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.

- 1796 • MC-truth correction accounts for differences between the reconstructed jet en-
- 1797 ergy and the MC particle-level energy. The correction is determined on a QCD
- 1798 dijet MC sample and is parametrized in terms of the jet p_T and η .

- 1799 • Residuals correct remaining small differences within jet response in data and
- 1800 MC. The Residuals η -dependent correction compares jets of similar p_T in the
- 1801 barrel reference region. The Residuals p_T -dependent correct the jet absolute
- 1802 scale (JES vs p_T).

- 1803 • Jet-flavor corrections are derived in the same way as MC-truth corrections but
- 1804 using QCD pure flavor samples.

1805 ***b*-tagging of jets.**

1806 A particular feature of the hadrons containing bottom quarks (b-hadrons) is that

1807 they have a lifetime long enough to travel some distance before decaying, but it is

1808 not as long as those of light quark hadrons; therefore, when looking at the hadrons

1809 produced in pp collisions, b-hadrons decay typically inside the tracker rather than

1810 reach the calorimeters as some light-hadrons do. As a result, a b-hadron decay gives

1811 rise to a displaced vertex (secondary vertex) with respect to the primary vertex as

1812 shown in figure 4.5; the SV displacement is in the order of a few millimeters. A jet

1813 resulting from the decay of a b-hadron is called *b* jet; other jets are called light jets.

1814

1815 Several methods to identify *b*-jets (*b*-tagging) have been developed; the method used in

1816 this thesis is known as “Combined Secondary Vertex” algorithm in its second version

1817 (CSVv2) [105]. By using information of the impact parameter, the reconstructed

1818 secondary vertices and the jet kinematics in a multivariate analysis that combines

1819 the discrimination power of each variable in one global discriminator variable, three

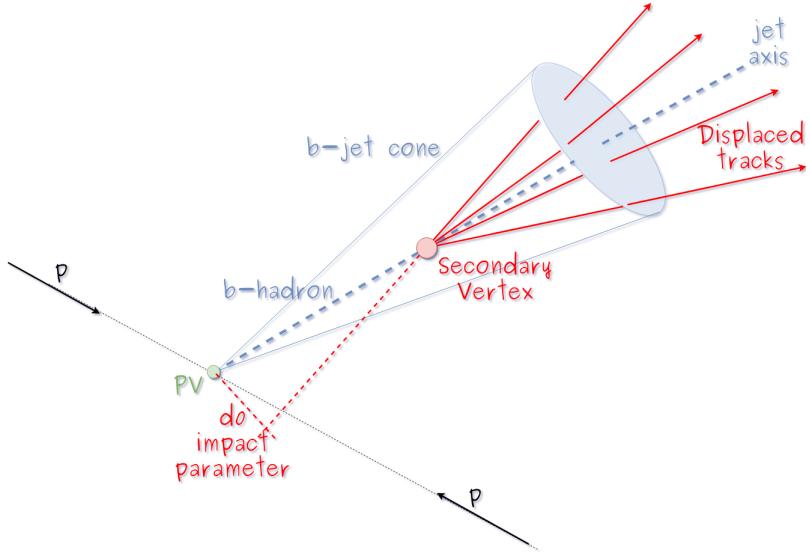


Figure 4.5: Secondary vertex in a b-hadron decay.

1820 working points (references): loose, medium and tight, are defined which quantify the
 1821 probabilities of mistag jets from light quarks as jets from b quarks; 10, 1 and 0.1 %
 1822 respectively. Although the mistagging probability decrease with the working point
 1823 strength, the efficiency to correctly tag b -jets also decrease as 83, 69 and 49 % for the
 1824 respective working point; therefore, a balance needs to be achieved according to the
 1825 specific requirements of the analysis.

1826 4.4.1.1 Missing transverse energy.

1827 The fact that proton bunches carry momentum along the z axis implies that for each
 1828 event, momentum balance in the transverse plane is expected. Imbalances are quan-
 1829 tified by the missing transverse energy (MET) and are attributed to several sources
 1830 including particles escaping undetected through the beam pipe, neutrinos produced in
 1831 weak interactions processes which do not interact with the detector and thus escaping
 1832 without leaving a sign, or even undiscovered particles predicted by models beyond
 1833 the SM.

1834

1835 The PF algorithm assign the negative sum of the momenta of all reconstructed PF
 1836 particles to the *particle-flow MET* according to

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

1837 JEC are propagated to the calculation of the \vec{E}_T as described in the reference [106].

1838

1839 4.4.2 Event reconstruction examples

1840 Figures 4.6-4.8 show the results of the reconstruction performed on 3 recorded events.
 1841 Descriptions are taken directly from the source.

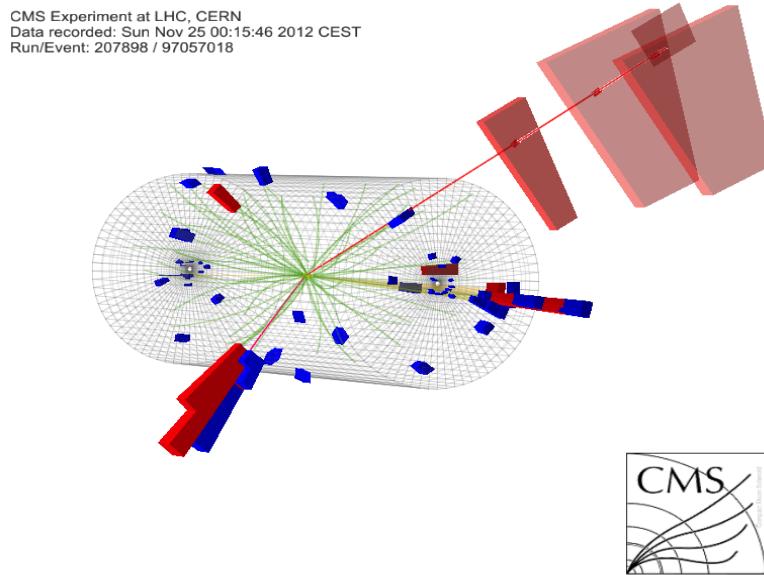


Figure 4.6: HIG-13-004 Event 1 reconstruction results; “HIG-13-004 Event 1: Event recorded with the CMS detector in 2012 at a proton-proton center-of-mass energy of 8 TeV. The event shows characteristics expected from the decay of the SM Higgs boson to a pair of τ leptons. Such an event is characterized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

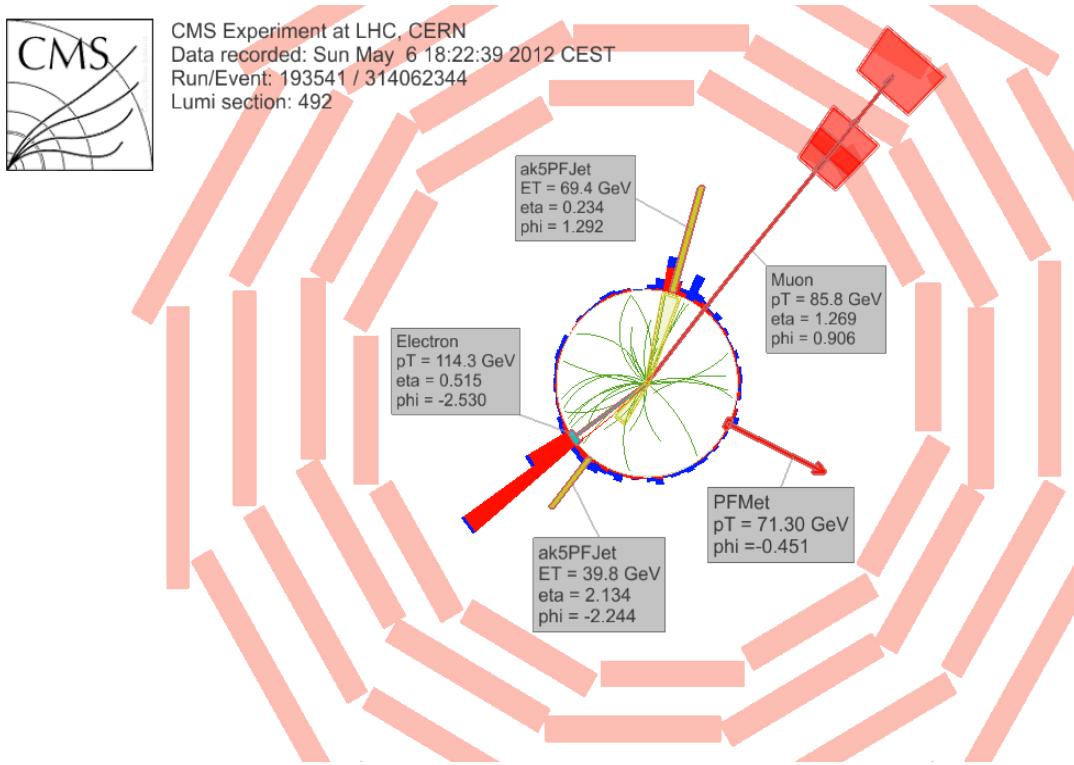


Figure 4.7: $e\mu$ event reconstruction results; “An $e\mu$ event candidate selected in 8 TeV data, as seen from the direction of the proton beams. The kinematics of the main objects used in the event selection are highlighted: two isolated leptons and two particle-flow jets. The reconstructed missing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

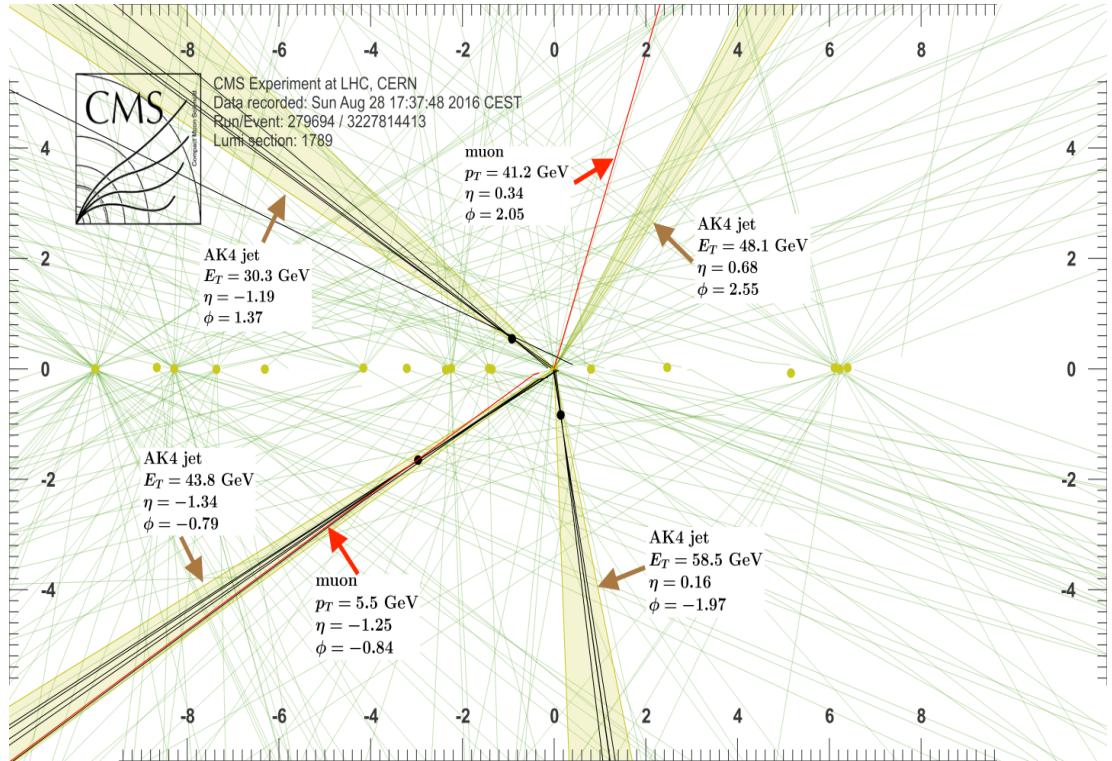


Figure 4.8: Recorded event reconstruction results; “Recorded event (ρ - z projection) with three jets with $p_T > 30$ GeV with one displaced muon track in 2016 data collected at 13 TeV. Each of the three jets has a displaced reconstructed vertex. The jet with $p_T(j) = 43.8$ GeV, $\eta(j) = -1.34$, $\phi(j)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

1842 **Chapter 5**

1843 **Statistical methods**

1844 In the course of analyzing the data sets provided by the CMS experiment and used in
1845 this thesis, several statistical tools have been employed; in this chapter, a description
1846 of these tools will be presented, starting with the general statement of the multivariate
1847 analysis method, followed by the particularities of the Boosted Decision Trees (BDT)
1848 method and its application to the classification problem. Statistical inference methods
1849 used will also be presented. This chapter is based mainly on the references [108–110].

1850 **5.1 Multivariate analysis**

1851 Multivariate data analysis (MVA) makes reference to statistical techniques that an-
1852 alyze data containing information of more than one variable, commonly taking into
1853 account the effects of all variables on the response of the particular variable under
1854 investigation, i.e., considering all the correlations between variables. MVA is em-
1855 ployed in a variety of fields like consumer and market research, quality control and
1856 process optimization. From a MVA it is possible to identify the dominant patterns
1857 in the data, like groups, outliers and trends, and determine to which group a set of

1858 values belong; in the particle physics context, MVA methods are used to perform the
 1859 selection of certain type of events, from a large data set, using a potentially large
 1860 number of measurable properties for each event.

1861 Processes with small cross section, as the tHq process, normally are hidden behind
 1862 more common processes; therefore, the data set results in a subset of events with
 1863 characteristic features of interest (signal) mixed in randomly with a much larger
 1864 number of SM events that can mimic these features of interest (background) which
 1865 implies that it is not possible to say with certainty that a given event is signal or
 1866 background. In that sense, the problem can be formulated as one where a set of
 1867 events have to be classified according to some features; these features correspond to
 1868 the measurements of several parameters like energy or momentum, organized in a
 1869 set of *input variables*. The measurements for each event can be written in a vector
 1870 $\mathbf{x} = (x_1, \dots, x_n)$ for which

- 1871 • Signal hypotheses $\rightarrow f(\mathbf{x}|s)$ is the probability density (*likelihood function*) that
 1872 \mathbf{x} is the set of measured values given that the events is a signal event.
- 1873 • Background hypotheses $\rightarrow f(\mathbf{x}|b)$ is the probability density (*likelihood function*)
 1874 that \mathbf{x} is the set of measured values given that the event is a background event.

1875 Figure 5.1 shows three ways to perform a classification of events for which mea-
 1876 surements of two properties, two input variables, have been performed; blue circles
 1877 represent signal events while red triangles represent background events. The classi-
 1878 fication on (a) is *cut-based* requiring $x_1 < c_1$ and $x_2 < c_2$; usually the cut values are
 1879 chosen according to some knowledge about the event process. In (b), the classification
 1880 is performed by stating a cut involving a linear function of the input variables and
 1881 so the boundary, while in (c) the the relationship between the input variables is not
 1882 linear thus the boundary is not linear either.

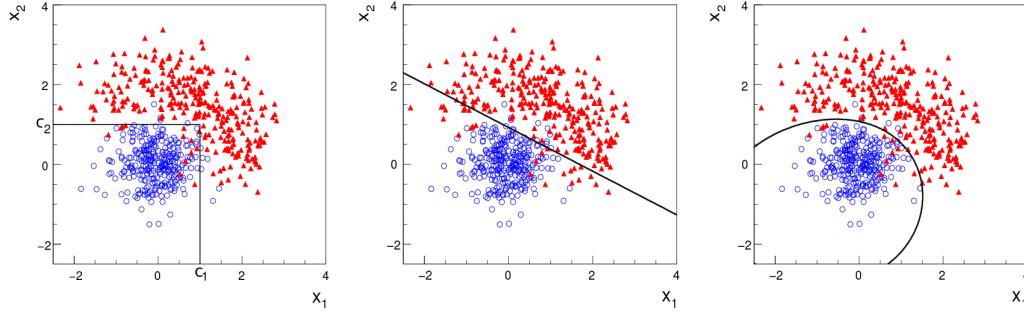


Figure 5.1: Scatter plots-MVA event classification. Distribution of two input variables x_1 and x_2 measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on (a) cuts, (b) linear boundary, and (c) nonlinear boundary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

1883 The boundary can be parametrized in terms of the input variables such that the
 1884 cut is set on the parametrization instead of on the variables, i.e., $y(\mathbf{x}) = y_{cut}$ with
 1885 y_{cut} a constant; thus, the acceptance or rejection of an event is based on what side
 1886 of the boundary is the event located. If $y(\mathbf{x})$ has functional form, it can be used to
 1887 determine the probability distribution functions $p(y|s)$ and $p(y|b)$ and then perform
 1888 a scalar test statistic with a single cut on the scalar variable y .

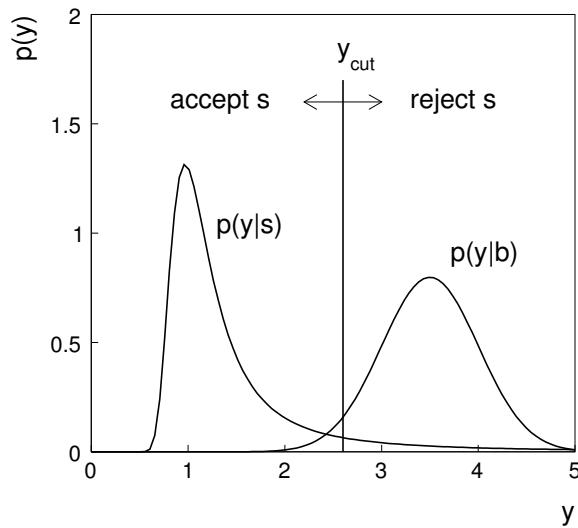


Figure 5.2: Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses. [108]

1889 Figure 5.2 illustrates what would be the probability distribution functions under
 1890 the signal and background hypotheses for a scalar test statistic with a cut on the
 1891 classifier y . Notice that the tails of the distributions indicate that some signal events
 1892 fall on the rejection region and some background events fall on the acceptance region;
 1893 therefore, it is convenient to define the *efficiency* with which events of a given type
 1894 are accepted, thus, the signal and background efficiencies are given by

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

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

1895 where A is the acceptance region. Under these conditions, the background hypothesis
 1896 corresponds to the *null hypothesis* (H_0), the signal hypothesis corresponds to the
 1897 *alternative hypothesis* (H_1), the background efficiency is the significance level of the
 1898 test, and signal efficiency is the power of the test; what is sought in an analysis is to
 1899 maximize the power of the test relative to the significance level.

1900 5.1.1 Decision trees

1901 For this thesis, the implementation of the MVA strategy, described above, is per-
 1902 formed through decision trees by using the TMVA software package [109] included in
 1903 the the ROOT analysis framework [111]. In a simple picture, a decision tree classifies
 1904 events according to their input variables values by setting a cut on each input variable
 1905 and checking which events are on which side of the cut, just as proposed in the MVA
 1906 strategy, but in addition, as a machine learning algorithm, decision trees offer the
 1907 possibility to be trained and then perform the classification efficiently.

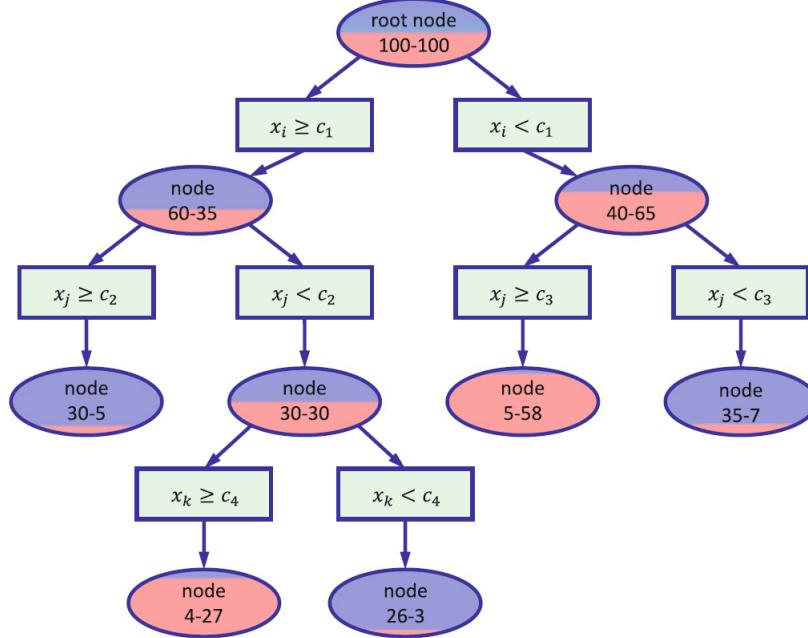


Figure 5.3: Example of a decision tree. Each node is fed with a MC sample mixing signal and background events (left-right numbers); nodes colors represent the relative amount of signal/background events [110].

1908 The training or growing of a decision tree is the process that defines the rules for
 1909 classifying events; this process is represented in figure 5.3 and consist of several steps

1910 • take MC samples of signal and background events and split them into two parts
 1911 each; first parts form the training sample which will be used in the decision tree
 1912 training, while the second parts form the test sample which will be used for
 1913 testing the final classifier obtained from the training. Each event has associated
 1914 a set of input variables $\mathbf{x} = (x_1, \dots, x_n)$ which serve to distinguish between signal
 1915 and background events. The training sample is taken in at the root *node*.

1916 • pick one variable, say x_i
 1917 • pick one value of x_i , each event has its own value of x_i , and split the training
 1918 sample into two subsamples B_1 and B_2 ; B_1 contains events for which $x_i < c_1$

1919 while B_2 contains the rest of the training events;

1920 • scan all possible values of x_i and find the splitting value that provides the *best*
 1921 classification¹, i.e., B_1 is mostly made of signal events while B_2 is mostly made
 1922 of background events.

1923 • It is possible that variables other than the picked one produce a better classi-
 1924 fication, hence, all the variables have to be evaluated. Pick the next variable,
 1925 say x_j , and repeat the scan over its possible values.

1926 • At the end, all the variables and their values will have been scanned, the *best*
 1927 variable and splitting value will have been identified, say x_1, c_1 , and there will
 1928 be two nodes fed with the subsamples B_1 and B_2 .

1929 Nodes are further split by repeating the decision process until: a given number of
 1930 final nodes is obtained, nodes are largely dominated by either signal or background
 1931 events, or nodes has too few events to continue. Final nodes are called *leaves* and they
 1932 are classified as signal or background leaves according to the class of the majority of
 1933 events in them. Each *branch* in the tree corresponds to a sequence of cuts.

1934 The quality of the classification at each node is evaluated through a separation
 1935 criteria; there are several of them but the *Gini Index* (G) is the one used in the
 1936 decision trees trained for the analysis in this thesis. G is written in terms of the
 1937 purity (P), i.e. the fraction of signal events, of the samples after the separation is
 1938 made; it is given by

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

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

1939 notice that $P=0.5$ at the root node while $G=0$ for pure leaves. For a node A split
 1940 into two nodes B_1 and B_2 the G gain is

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

1941 the *best* classification corresponds to that for which the gain of G is maximized; hence,
 1942 the scanning over all event's variables and their values is of capital importance.

1943 In order to provide a numerical output for the classification, events in a sig-
 1944 nal(background) leaf are assigned an score of 1(-1) each, defining in this way the
 1945 decision tree *classifier or weak learner* as

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

1946 Figure 5.4 shows an example of the classification of a sample of events, containing
 1947 two variables, performed by a decision tree.

1948 5.1.2 Boosted decision trees (BDT).

1949 Event misclassification occurs when a training event ends up in the wrong leaf, i.e., a
 1950 signal event ends up in a background leaf or a background event ends up in a signal
 1951 leaf. A way to correct it is to assign a weight to the misclassified events and train
 1952 a second tree using the reweighted events; the event reweighting is performed by a
 1953 boosting algorithm, events with increased weight are known as *boosted* events, in such
 1954 a way that when used in the training of a new decision tree they get correctly classified.
 1955 The process is repeated iteratively adding a new tree to a forest and creating a set
 1956 of classifiers which are combined to create the next classifier; the final classifier offers

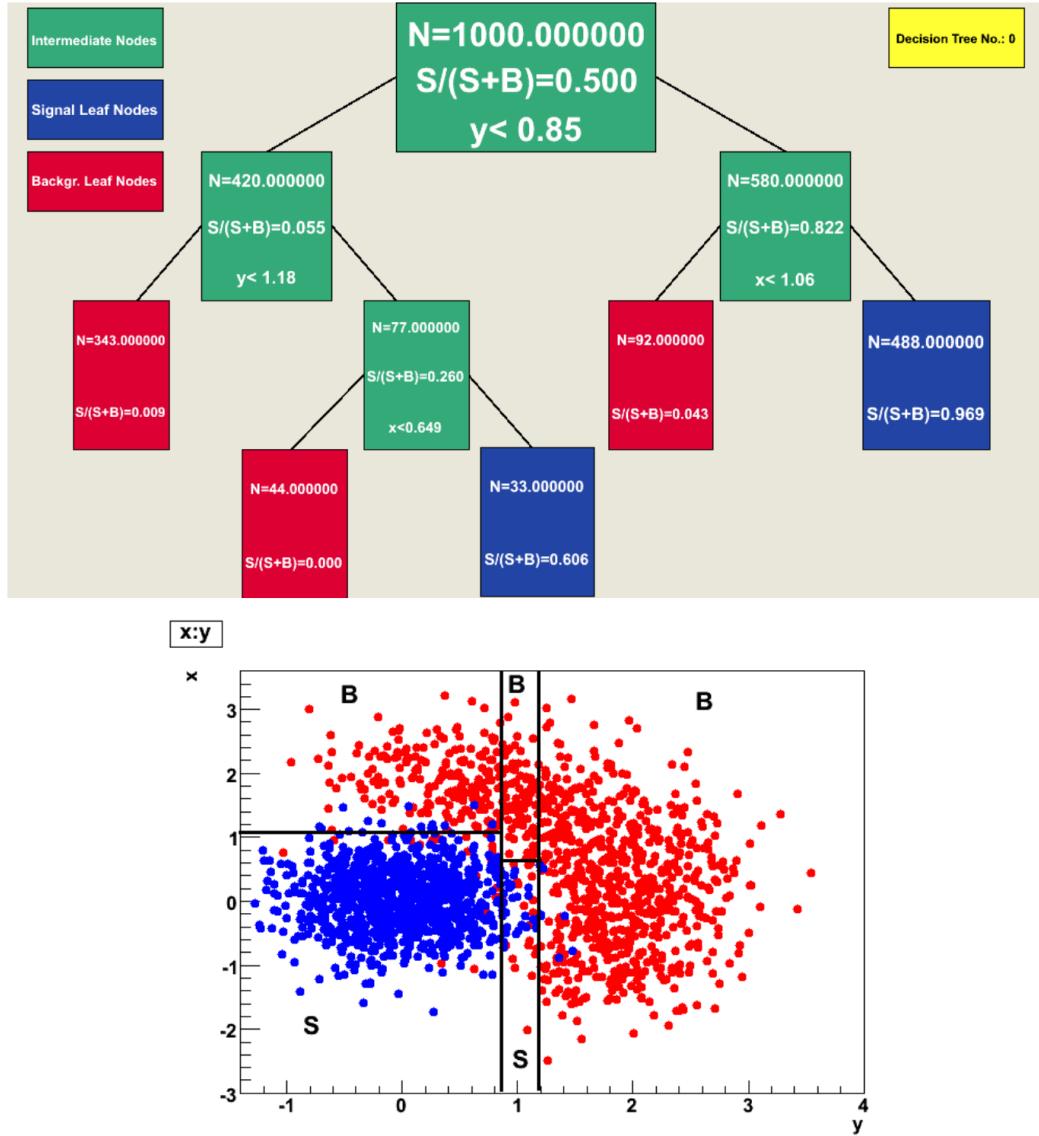


Figure 5.4: Example of a decision tree output. Each leaf, blue for signal events and red for background events, represent a region in the variables phase space [112].

1957 more stability² and has a smaller misclassification rate than any individual ones. The
 1958 resulting tree collection is known as a *boosted decision tree (BDT)*.
 1959 Thus, purity of the sample is generalized to

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

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

1960 where w_s and w_b are the weights of the events; the Gini index is also generalized

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

1961 with n the number of events in the node. The final score of an event, after pass-
 1962 ing through the forest, is calculated as the renormalized sum of all the individual
 1963 (possibly weighted) scores; thus, high(low) score implies that the event is most likely
 1964 signal(background).

1965 The boosting procedure, implemented in the *Gradient boosting* algorithm used
 1966 in this thesis, produce a classifier $F(\mathbf{x})$ which is the weighted sum of the individual
 1967 classifiers obtained after each iteration,i.e.,

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

1968 where M is the number of trees in the forest. The *loss function* $L(F, y)$ represent the
 1969 deviation between the classifier $F(\mathbf{x})$ response and the true value y obtained from the
 1970 training sample (1 for signal events and -1 for background event), according to

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

1971 thus, the reweighting is employed to ensure the minimization of the loss function;
 1972 a more detailed description of the minimization procedure can be found in reference
 1973 [113]. The final classifier output is later used as a final discrimination variable, labeled
 1974 as *BDT output/response*.

1975 **5.1.3 Overtraining.**

1976 Decision trees offer the possibility to have as many nodes as wished in order to
 1977 reduce the misclassification to zero (in theory); however, when a classifier is too much
 1978 adjusted to a particular training sample, the classifier response to a slightly different
 1979 sample may leads to a completely different classification results; this effect is known
 1980 as *overtraining*.

1981 An alternative to reduce the overtraining in BDTs consist in pruning the tree by
 1982 removing statistically insignificant nodes after the tree growing is completed but this
 1983 option is not available for BDTs with gradient boosting in the TMVA-toolkit, there-
 1984 fore, the overtraining has to be reduced by tuning the algorithm, number of nodes,
 1985 minimum number of events in the leaves, etc. The overtraining can be evaluated
 1986 by comparing the responses of the classifier when running over the training and test
 1987 samples.

1988 **5.1.4 Variable ranking.**

1989 BDTs have the couple of particular advantages related to the input variables; on one
 1990 side, they are relatively insensitive to the number of input variables used in the vector
 1991 \mathbf{x} . The ranking of the BDT input variables is determined by counting the number of
 1992 times a variable is used to split decision tree nodes; in addition, the separation gain-
 1993 squared achieved in the splitting and the number of events in the node are accounted
 1994 by applying a weighting to that number. Thus, those variables with small or no power
 1995 to separate signal and background events are rarely chosen to split the nodes,i.e., are
 1996 effectively ignored.

1997 On the other side, variables correlations play an important role for some MVA
 1998 methods like the Fisher discriminant algorithm in which the first step consist of

1999 performing a linear transformation to a phase space where the correlations between
 2000 variables are removed; in case of BDT algorithm, correlations do not affect the per-
 2001 formance.

2002 **5.1.5 BDT output example.**

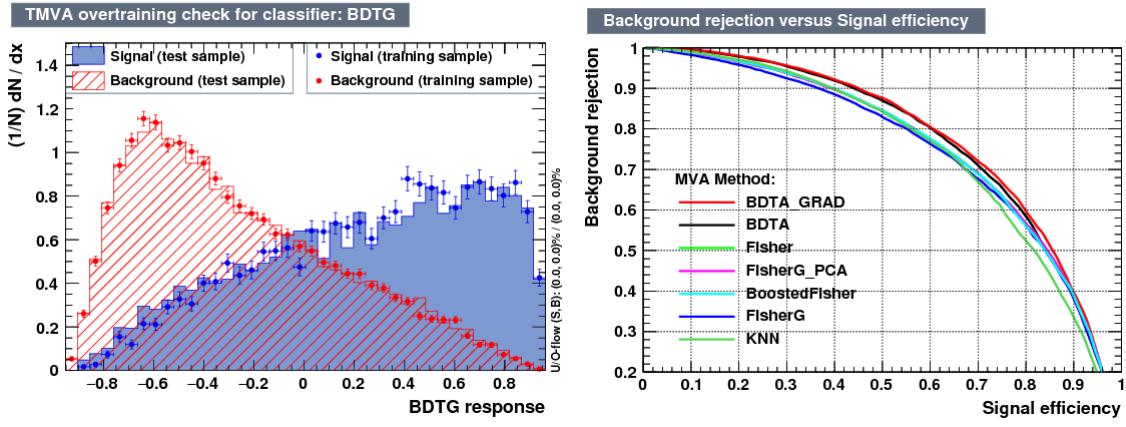


Figure 5.5: Left: Output distributions for the gradient boosted decision tree (BDTG) classifier using a sample of signal($pp \rightarrow tHq$) and background($pp \rightarrow t\bar{t}$) events. Right: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers running over the same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

2003 Left side of figure 5.5 shows the BDT output distributions for signal($pp \rightarrow tHq$)
 2004 and background($pp \rightarrow t\bar{t}$) events; this plot is the equivalent to the one showed in
 2005 figure 5.2. A forest with 800 trees, maximum depth per tree = 3, and gradient
 2006 boosting have been used as training parameters. The BDTG classifier offers a good
 2007 separation power; while there is a small overtraining in the signal distribution, the
 2008 background distribution seems to be well predicted which might indicate that the
 2009 sample is composed of more background than signal events.

2010 Right side of figure 5.5 shows the background rejection vs signal efficiency curves
 2011 for several combinations of MVA classifiers-boosting algorithms; these curves are
 2012 known as ROC curves and give an indication of the performance of the classifier. The

2013 best performance is achieved with the BDTG classifier (BDTA_GRAD).

2014 5.2 Statistical inference.

2015 Once events are classified, the next step consists in finding the parameters that define
 2016 the likelihood functions $f(\mathbf{x}|s), f(\mathbf{x}|b)$ for signal and background events respectively.
 2017 In general, likelihood functions depend not only on the measurements but also on
 2018 parameters (θ_m) that define their shapes; the process of estimating these *unknown*
 2019 *parameters* and their uncertainties from the experimental data is called *inference*.
 2020 The likelihood function for N the events the in a sample is the combination of all the
 2021 likelihoods functions

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

2022 Thus, the estimation of the unknown parameters from experimental data samples
 2023 is written in terms of a central value using the notation

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

2024 where the interval $[\hat{\theta} + \delta\theta, \hat{\theta} - \delta\theta]$ is called *confidence interval*; it is usually inter-
 2025 preted, in the limit of infinite number of experiments, as the interval where the true
 2026 value of the unknown parameter θ is contained with a probability of 0.6827 (if no
 2027 other convention is stated).

2028 5.2.1 Nuisance parameters.

2029 The unknown parameter vector $\boldsymbol{\theta}$ is made of two types of parameters: on one side,
 2030 those parameters that provide information about the physical observables of interest

2031 for the experiment (*parameters of interest*); on the other side, the *nuisance parameters*
 2032 that are not of direct interest for the experiment but that needs to be included in
 2033 the analysis in order to achieve a satisfactory description of the data. They represent
 2034 effects of the detector response like the finite resolutions of the detection systems,
 2035 miscalibrations, and in general any source of uncertainty introduced in the analysis.

2036 In some cases the nuisance parameters are estimated using dedicated data samples,
 2037 for instance data from test beams for calibration purposes, when MC samples are
 2038 not suitable. The nuisance parameter uncertainties produce *systematic uncertainties*
 2039 while the uncertainties associated to fluctuations in data and related to the estimation
 2040 of the parameters of interest produce *statistical uncertainties*.

2041 5.2.2 Maximum likelihood estimation method

2042 The function that produce the estimate of a parameter is called *estimator*, there-
 2043 fore, estimators are usually constructed using mathematical procedures encoded in
 2044 algorithms. The estimation method used in this thesis is the *Maximum Likelihood*
 2045 *Estimation* method (MLE); it is based on the combined likelihood function defined
 2046 by eqn. 5.9 and the procedure seeks for the parameter set that corresponds to the
 2047 maximum value of the combined likelihood function, i.e., the *maximum likelihood*
 2048 *estimator* of the unknown parameter vector $\boldsymbol{\theta}$ is the function that produce the vec-
 2049 tor $\hat{\boldsymbol{\theta}}$ for which the likelihood function $L(\boldsymbol{\theta})$ evaluated at the measured sample \mathbf{x} is
 2050 maximum.

2051 Usually, the logarithm of the likelihood function is used in the numerical algo-
 2052 rithms implementations in order to avoid underflow the numerical precision of the
 2053 computers due to the product of low likelihoods. In addition, it is usual minimize the
 2054 negative logarithm of the likelihood function instead of maximizing the logarithm of

2055 it because in this way the procedure consist of differentiate a sum of therms and set
 2056 the sum to zero; therefore

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

2057 The minimization process is performed by the software MINUIT [114] imple-
 2058 mented in the ROOT analysis framework. In case of large data samples the compu-
 2059 tational resources needed to calculate the likelihood function are too big; therefore,
 2060 the parameter estimation is performed using binned distributions of the variables of
 2061 interest for which the *binned likelihood function* is given by

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

2062 with s_i and b_i the expected number of signal and background yields for bin i respec-
 2063 tively, n_i is the observed number of events in the bin i and $\mu = \sigma/\sigma_{SM}$ is the signal
 2064 strength. Notice that the number of entries per bin follows a Poisson distribution.
 2065 The analysis presented in this thesis is based on the binned distribution of the ratio
 2066 signal/background obtained from the BDT outputs.

2067 5.2.3 Hypothesis test

2068 The test statistic mentioned in section 5.1 involving
 2069 ; it is achieved, according to the Neyman-Pearson lemma [115],
 2070 by defining the acceptance region such that, for \mathbf{x} inside the region, the likelihood
 2071 ratio, i.e., the ratio of probability distribution functions for signal and background,

2072 **5.3 exclusion limits**

2073 **5.4 asymptotic limits**

2074 **Chapter 6**

2075 **Search for production of a Higgs**

2076 **boson and a single top quark in**

2077 **multilepton final states in pp**

2078 **collisions at $\sqrt{s} = 13$ TeV**

2079 **6.1 Introduction**

2080 The Higgs boson discovery, supported on experimental observations and theoretical
2081 predictions made about the SM, gives the clue of the way in that elementary particles
2082 acquire mass through the Higgs mechanism; therefore, knowing the Higgs mass, the
2083 Higgs-boson and Higgs-fermions couplings can be tested. In order to test the Higgs-
2084 top coupling, several measurements have been performed, as stated in the chapter 2,
2085 but they are limited to measure the square of the coupling; however, the production
2086 of a Higgs boson in association with a single top quark (tH) not only offers access to
2087 the sign of the coupling, but also, to the CP phase of the Higgs couplings.

2088 This chapter presents the search for the associated production of a Higgs boson
 2089 and a single top quark events, focusing on leptonic signatures provided by the Higgs
 2090 decay modes to WW , ZZ , and $\tau\tau$; the 13 TeV dataset produced in 2016, which
 2091 corresponds to an integrated luminosity of 35.9fb^{-1} , is used. Constraints on the sign
 2092 of the Higgs- top coupling (y_t) have been derived from the decay rate of Higgs bosons
 2093 to photon pairs [40] and from the cross section for associated production of Higgs and
 2094 Z bosons via gluon fusion [116], with recent results disfavoring negative signs of the
 2095 coupling [51,53,117]. It expands previous analyses performed at 8 TeV [118,119] and
 2096 searches for associated production of $t\bar{t}$ pair and a Higgs boson in the multilepton final
 2097 state channel [120]; it also complements searches in other decay channels targeting
 2098 $H \rightarrow b\bar{b}$ [121].

2099 As shown in section 2.4, the SM cross section of the associated production of a
 2100 Higgs boson and a single top quark (tHq) process is driven by a destructive interfer-
 2101 ence between two contributions (see Figure 2.13), where the Higgs couples to either
 2102 the W boson or the top quark; however, if the sign of the Higgs-top coupling is flipped
 2103 with respect to the SM prediction, a large enhancement of the cross section occurs,
 2104 making this analysis sensitive to such deviation. A second process, where the Higgs
 2105 boson and top quark are accompanied by a W boson (tHW) has similar behavior,
 2106 albeit with a weaker interference pattern and lower contribution to the tH cross sec-
 2107 tion, therefore, a combination of both processes would increase the sensitivity; in
 2108 this analysis both contributions are combined and referred as tH channel. A third
 2109 contribution comes from $t\bar{t}H$ process. The purpose of this analysis is to investigate
 2110 the exclusion of the presence of the $tH + t\bar{t}H$ processes under the assumption of the
 2111 anomalous Higgs-top coupling modifier ($\kappa_t = -1$). The analysis exploits signatures with
 2112 three leptons in the final state.

2113 The first sections present the characteristic tHq signature as well as the expected

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

6.2 tHq signature

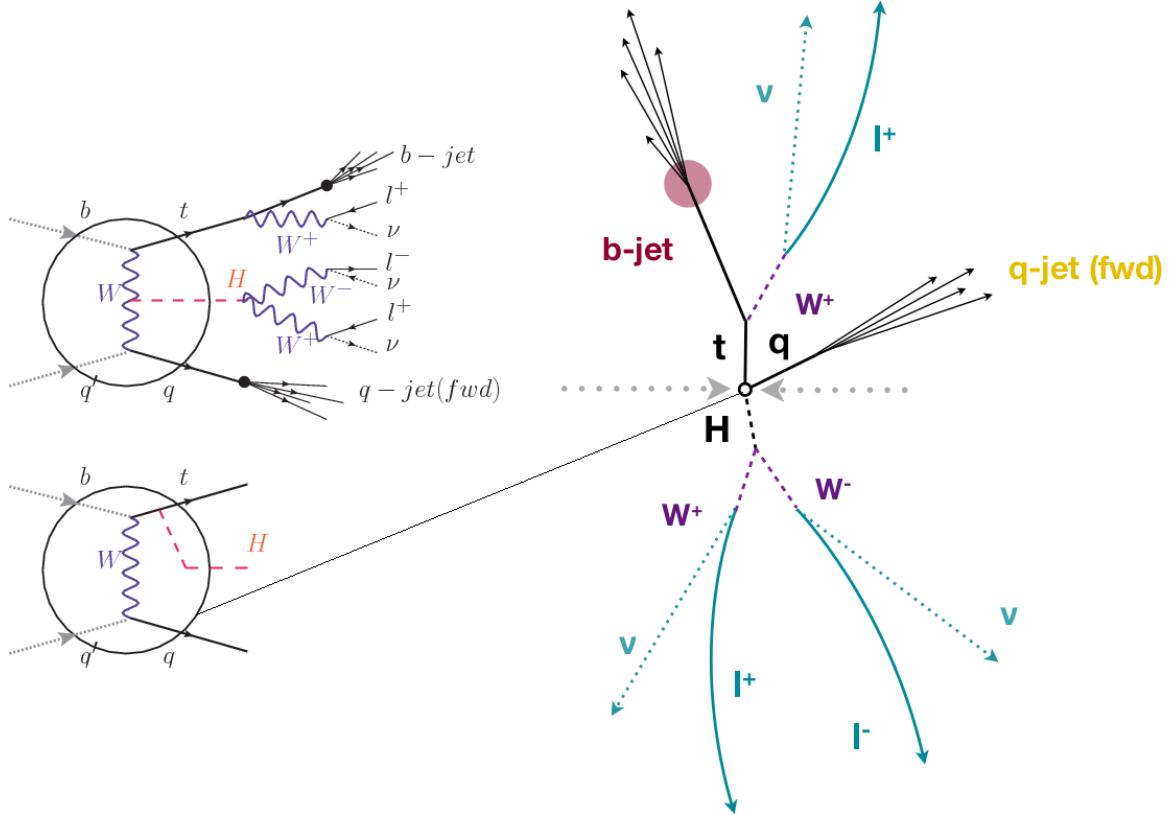


Figure 6.1: tHq event signature. Left: Feynman diagram including the whole evolution up to the final state for the case of the Higgs boson emitted by the W boson (top); feynman diagram for the case where the Higgs boson is emitted by the top quark. Right: Schematic view as it would be seen in the detector; The circle in the feynman diagrams on the left corresponds to the circle in the center of the schematic view as indicated by the line connecting them.

2120 In order to select events of tHq process, its features are translated into a set of
 2121 selection rules; figure 6.1 shows the Feynman diagram and an schematic view of the
 2122 tHq process from the pp collision to the final state configuration. A single top quark
 2123 is produced accompanied by a light quark, denoted as q ; this light quark is produced
 2124 predominantly in the forward region of the detector. The Higgs boson which can
 2125 be either emitted by the exchanged W boson or directly by the singly produced top
 2126 quark.

2127 The top quark and Higgs boson decay after their production in the detector due to
 2128 their high masses/low lifetimes. The Higgs boson is required to decay into a W boson
 2129 pair¹. The top quark almost always decays into a bottom quark and a W boson, as
 2130 encoded in the CMK matrix. The W bosons are required to decay leptonically while
 2131 τ leptons are not reconstructed separately and only their leptonic decays into either
 2132 electrons or muons are considered in this analysis.

2133 In summary, the signal process is characterized by a the final state with one
 2134 light-flavored forward jet, one central b-jet, three leptons (muons, electrons or a
 2135 combination of them), three neutrinos and no central light-flavored jets. The presence
 2136 of neutrinos is inferred from the presence of MET. The analysis has been made public
 2137 by CMS as a Physics Analysis Summary [122] combining the result for the three lepton
 2138 and two lepton same-sign channels. Currently, an effort to turn the analysis into a
 2139 paper is ongoing.

2140 6.3 Background processes

2141 The background processes are those that can mimic the signal signature or at least
 2142 can be reconstructed as that as a result of certain circumstances. The backgrounds

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

2143 can be classified as

- 2144 • Irreducible backgrounds where genuine prompt leptons are produced in on-
2145 shell W and Z boson decays; they can be reliably estimated directly from MC
2146 simulated events, using higher-order cross sections or data control regions for
2147 the overall normalization.
- 2148 • Reducible backgrounds where at least one of the leptons is *non-prompt*, i.e., pro-
2149 duced within a hadronic jet, either a genuine lepton from heavy flavor decays.
2150 Mis-reconstructed jets, also known as *mis-ID leptons* or *fake leptons*, are con-
2151 sidered non-prompt leptons as well. These non-prompt leptons leave tracks and
2152 hits in the detection systems as would a prompt lepton, but correlating those
2153 hits with nearby jets could be a way of removing them. Reducible backgrounds
2154 are not well predicted by simulation, and are estimated using data-driven meth-
2155 ods.

2156 The main sources of background events in the case of tHq process are $t\bar{t}$ process
2157 and $t\bar{t} + X (X = W, Z, \gamma)$ processes, here represented together as $t\bar{t}V$ process. Figure
2158 6.2 shows the signature for $t\bar{t}$ and $t\bar{t}W$ processes;

2159 The largest contribution to irreducible backgrounds involving prompt leptons
2160 comes from $t\bar{t}W$, $t\bar{t}Z$, processes for which the number of ($b-$)jets (($b-$)jet multipli-
2161 city) is higher than that of the signal events, while for other contributing background
2162 events, WZ , ZZ , and rare SM processes like $W^\pm W^\pm qq$, $t\bar{t}t\bar{t}$, tZq , tZW , WW ,
2163 WWZ , WZZ , ZZZ , the ($b-$)jet multiplicity is lower compared to that of the signal
2164 events. None of the irreducible backgrounds present activity in the forward region of
2165 the detector.

2166 On the side of the reducible backgrounds, the largest contribution comes from the
2167 $t\bar{t}$ events which have a very similar signature to the signal events but does no present

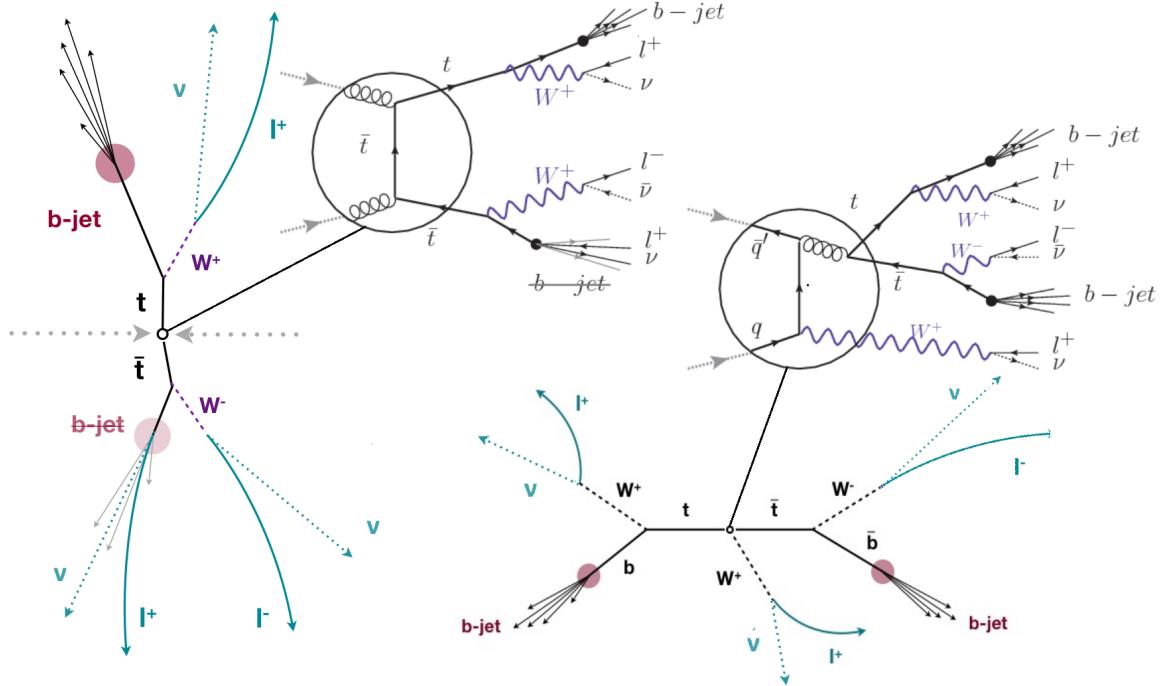


Figure 6.2: $t\bar{t}$ (left) and $t\bar{t}W$ (right) events signature as they would be seen in the detector; the feynman diagrams including the whole evolution up to the final state are also showed. The $t\bar{t}$ process signature is very similar to that of the signal process with one fake lepton and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

2168 activity in the forward region of the detector either; A particular feature of the $t\bar{t}$
 2169 events is their charge-symmetry which is also a difference with the signal events.

2170 6.4 Object Identification and event selection

2171 In this section, the specific definitions of the physical objects in terms of the numerical
 2172 values assigned to the reconstruction parameters are presented; thus, the provided
 2173 details summarize and complement the descriptions presented in previous chapters.
 2174 The object reconstruction and selection strategy used in this thesis is identical to the
 2175 one described in references [120, 125], thus, the information provided in this section
 2176 is based in those documents unless other references are stated.

²¹⁷⁷ **6.4.1 Jets and b tagging.**

²¹⁷⁸ In this analysis, jets are reconstructed by clustering PF candidates using the anti- k_t
²¹⁷⁹ algorithm with parameter distance $\Delta R = 0.4$; those charged hadrons that are not
²¹⁸⁰ consistent with the selected primary vertex are discarded from the clustering. The
²¹⁸¹ jet energy is then corrected for the varying response of the detector as a function
²¹⁸² of transverse momentum p_T and pseudorapidity η . Jets are selected for use in the
²¹⁸³ analysis only if they have $p_T > 25$ GeV and are separated from any selected leptons
²¹⁸⁴ by $\Delta R > 0.4$.

²¹⁸⁵ Jets coming from the primary vertex and jets coming from pile-up vertices are
²¹⁸⁶ distinguished using a MVA discriminator based on the differences in the jet shapes,
²¹⁸⁷ in the relative multiplicity of charged and neutral components, and in the different
²¹⁸⁸ fraction of transverse momentum which is carried by the hardest components. Jet
²¹⁸⁹ tracks are also required to be compatible with the primary vertex.

²¹⁹⁰ Jets originated from the hadronization of a b quark are selected using a MVA
²¹⁹¹ likelihood discriminant which uses track-based lifetime information and reconstructed
²¹⁹² secondary vertices (CSV algorithm). Only jets within the CMS tracker acceptance
²¹⁹³ ($\eta < 2.4$) are identified with this tool. Data samples are used to measure the efficiency
²¹⁹⁴ of the b tagging and the probability to misidentify jets from light quarks or gluons;
²¹⁹⁵ in both cases the measurements are parametrized as a function of the jet p_T and η
²¹⁹⁶ and later used to correct differences between the data and MC simulation in the b
²¹⁹⁷ tagging performance, by applying per-jet weights to the simulation, dependent on
²¹⁹⁸ the jet p_T , η , b tagging discriminator, and flavor (from simulation truth) [123]. The
²¹⁹⁹ per-event weight is taken as the product of the per-jet weights, including those of the
²²⁰⁰ jets associated to the leptons. The weights are derived on $t\bar{t}$ and Z+jets events.

²²⁰¹ Two working points are defined, based on the CSV algorithm output: ‘*loose*’ work-

ing point ($\text{CSV} > 0.46$) with a b signal tagging efficiency of about 83% and a mistagging rate of about 8%; and *medium* working point ($\text{CSV} > 0.80$) with b -tagging efficiency of about 69% and mistagging rate of order 1% [124]. Tagging of jets from charm quarks have efficiencies of about 40% and 18% for loose and medium working points respectively. Separate scale factors are applied to jets originating from bottom/charm quarks and from light quarks in simulated events to match the tagging efficiencies measured in the data.

6.4.2 Missing Energy MET.

As stated in section 4.4.1.1, the MET vector is calculated as the negative of the vector sum of transverse momenta of all PF candidates in the event and its magnitude is referred to as E_T^{miss} . Due to pile-up interactions, the performance in determining MET is degraded; in order to correct for that, the energy from the selected jets and leptons that compose the event is assigned to the variable H_T^{miss} . It is calculated in the same way as E_T^{miss} and although it has worse resolution than E_T^{miss} , it is more robust in the sense that it does not rely on the soft part of the event. The event selection uses a linear discriminator based on the two variables given by

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

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

2222 6.4.3 Lepton reconstruction and identification

2223 Two types of leptons are defined in this analysis: *signal leptons* are those coming from
 2224 W, Z and τ decays which usually are isolated from other particle; *background leptons*
 2225 are defined as leptons produced in b hadron decays, light-jets misidentification, and
 2226 photon conversions. The reconstruction and identification of electron and muon can-
 2227 didates was described in chapter4, therefore, the advanced identification criteria used
 2228 in order to retain the highest possible efficiency for signal leptons while maximizing
 2229 the rejection of background leptons are listed in this section ². The Physics Objects
 2230 Groups (POG) at CMS, are in charge of studying and defining the set of selection
 2231 criteria applied on the course of reconstruction and identification of particles. These
 2232 selection criteria are implemented in the CMS framework in the form of several object
 2233 identification working points according to the strength of the requirements.

2234 6.4.3.1 Muons.

2235 Muon candidates are reconstructed by combining information from the tracker system
 2236 and the muon detection system of CMS detector. The POG defined three working
 2237 points for muon identification *MuonID* [126];

- 2238 • *POG Loose Muon ID* is a particle identified as a muon by the PF event re-
 construction and also reconstructed either as a global-muon or as an arbitrated
 tracker-muon. This identification criteria is designed to be highly efficient for
 prompt muons and for muons from heavy and light quark decays; it can be com-
 plemented by applying impact parameter cuts in analyses with prompt muon
 signals.

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

- 2244 • *POG Medium Muon ID* is a Loose muon with additional track-quality and
 2245 muon-quality (spatial matching between the individual measurements in the
 2246 tracker and the muon system) requirements. This identification criteria is de-
 2247 signed to be highly efficient in the separation of the muons coming from decay
 2248 in flight of heavy quarks and muons coming from B meson decays as well as
 2249 prompt muons. An additional category *MVA Prompt ID* is defined in this iden-
 2250 tification criteria directed to discriminated muons from B mesons and prompt
 2251 muons (from W,Z and τ decays). The Medium ID provides the same fake rate as
 2252 the Tight Muon ID but a higher efficiency on prompt and B-decays muons. [127]

 2253 • *POG Tight Muon ID* is a global muon with additional muon-quality require-
 2254 ments Tight Muon ID selects a subset of the PF muons.

2255 Only muons within the muon system acceptance $|\eta| < 2.4$ and minimum p_T of 5
 2256 GeV are considered.

2257 **6.4.3.2 Electrons.**

2258 Electrons are reconstructed using information from the tracker and from the electro-
 2259 magnetic calorimeter and identified by an MVA algorithm (MVA eID discriminant)
 2260 using the shape of the calorimetric shower variables like the shape in η and ϕ , the clus-
 2261 ter circularity, widths along η and ϕ ; track-cluster matching variables like E_{tot}/p_{in} ,
 2262 E_{Ele}/p_{out} , $\Delta\eta_{in}$, $\Delta\eta_{out}$, $\Delta\phi_{in}$, $1/E - 1/p$; and track quality variables like ξ^2 of the
 2263 GSF tracks, the number of hits used by the GSF filter [128].

2264 A loose selection based on η -dependent cuts on this discriminant is used to pre-
 2265 select electron candidates, the full shape of the discriminant is used in the lepton
 2266 multivariate selection to separate signal leptons from background leptons (described
 2267 in section 6.4.3.4).

2268 In order to reject electrons from photon conversions, electron candidates with
 2269 missing hits in the pixel tracker layers or matched to a conversion secondary vertex
 2270 are discarded. Electrons are selected for the analysis if they have $p_T > 7$ GeV and
 2271 are located within the tracker system acceptance region ($|\eta| < 2.5$).

2272 **6.4.3.3 Lepton vertexing and pile-up rejection.**

2273 With the goal of rejecting pile-up or mis-reconstructed tracks, and more importantly
 2274 to reject background leptons from b-hadron decays, the following impact parameter
 2275 variables are also considered: impact parameter in the transverse plane d_0 , impact
 2276 parameter along the z axis d_z , and the impact parameter significance in the detector
 2277 space SIP_{3D} .

2278 Loose cuts are applied on this variables to achieve the first goal, while the full
 2279 shape of the same variables is used in a multivariate approach to reach the best
 2280 separation between the signal and the background leptons.

2281 Three different selections are defined both for the electron and muon object identi-
 2282 fication: the *Loose*, *Fakeable Object*, and *Tight* selection. As described in more detail
 2283 later, these are used for event level vetoes, the fake rate estimation application region,
 2284 and the final signal selection, respectively. The p_T of fakeable objects is defined as
 2285 $0.85 \times p_T(\text{jet})$, where the jet is the one associated to the lepton object. This mitigates
 2286 the dependence of the fake rate on the momentum of the fakeable object and thereby
 2287 improves the precision of the method.

2288 Tables 6.1 and 6.2 list the full criteria for the different selections of muons and
 2289 electrons.

2290 In addition, jets are required to be separated from any lepton candidates passing
 2291 the fakeable object selections (see Tables ?? and ??) by $\Delta R > 0.4$.

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

Table 6.1: 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.

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 $> (0.0, 0.0, 0.7)$	✓	✓	✓
$\sigma_{i\eta i\eta} < (0.011, 0.011, 0.030)$	—	✓	✓
H/E $< (0.10, 0.10, 0.07)$	—	✓	✓
$\Delta\eta_{\text{in}} < (0.01, 0.01, 0.008)$	—	✓	✓
$\Delta\phi_{\text{in}} < (0.04, 0.04, 0.07)$	—	✓	✓
$-0.05 < 1/E - 1/p < (0.010, 0.010, 0.005)$	—	✓	✓
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 6.2: 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).

2292 **6.4.3.4 Lepton MVA discriminator.**

2293 A further discrimination between prompt signal leptons (i.e. from W and Z bo-
 2294 son decays and from leptonic τ decays) and non-prompt and spurious leptons from
 2295 b hadron decays, decays-in-flight, and photon conversions is crucial in light of the
 2296 overwhelming background from $t\bar{t}$ production. The small probabilities of having the
 2297 second type of leptons results in a sizable number of background events since the rate
 2298 of $t\bar{t}$ production is much larger than the signal. To maximally exploit the available
 2299 information in each event to that end, a multivariate discriminator based on a boosted
 2300 decision tree (BDT) algorithm is built, taking as input not just observables related
 2301 directly to the reconstructed leptons themselves, but also to the clustered energy
 2302 deposits and charged particles in a cone around the lepton direction.

2303 The jet reconstruction and b-tagging algorithms are run on these, and their output
 2304 is used to train the algorithm. In particular, the ratio between the lepton p T and the
 2305 reconstructed jet p T , and the transverse momentum of the lepton with respect to
 2306 the jet axis provide good separation power in addition to more traditional observables
 2307 like the relative isolation of the lepton (calculated in a variable cone size depending
 2308 on the lepton p T [27, 28]), and the impact parameters of the lepton trajec- tory. The
 2309 BDT algorithm is trained on prompt leptons in simulated ttH signal and non-prompt
 2310 leptons in tt background events and validated using data in various control regions.
 2311 Leptons are then selected for the final analysis if they pass a given threshold of the
 2312 BDT output, and are referred to as “tight leptons” in the following.

2313 **6.4.4 Lepton selection efficiency**

2314 Efficiencies of reconstruction and selecting loose leptons are measured both for muons
 2315 and electrons using a tag and probe method on both data and MC, using $Z \rightarrow \ell^+ \ell^-$.

2316 Corresponding scale factors are derived from the ratio of efficiencies and applied to the
2317 selected These. Events are produced for the leptonic SUSY analyses using equivalent
2318 lepton selections and recycled for the $t\bar{t}H$ analysis as well as for this analysis. The
2319 efficiencies of applying the tight selection as defined in Tables ?? and ??, on the
2320 loose leptons are determined again by using a tag and probe method on a sample of
2321 DY-enriched events. They are documented for the $t\bar{t}H$ analysis in Ref. [125] and are
2322 exactly equivalent for this analysis.

2323 .

2324 .

2325 .

2326 .

2327 .

2328 .

2329 .

2330 .

2331 .

2332 . . .

2333 The analysis is designed to efficiently identify and select prompt leptons from
2334 on-shell W and

2335 Z boson decays and to reject non-prompt leptons from b quark decays and spurious
2336 lepton

2337 signatures from hadronic jets. Events are then selected in the various lepton
2338 channels, and are

2339 required to contain hadronic jets, some of which must be consistent with b quark
2340 hadronization. Finally, the signal yield is extracted by simultaneously fitting the
2341 output of two dedicated

2342 multivariate discriminants (trained to separate the tHq signal from the two dom-
 2343 inant backgrounds) in all categories

2344 . Multivariate techniques are used to discriminate the signal from the dominant
 2345 backgrounds. The analysis yields a 95% confidence level (C.L.) upper limit on the
 2346 combined $tH + ttH$ production cross section times branching ratio of 0.64 pb, with
 2347 an expected limit of 0.32 pb, for a scenario with $kt = \sqrt{1.0}$ and $kV = 1.0$. Values
 2348 of kt outside the range of $\sqrt{1.25}$ to $\sqrt{1.60}$ are excluded at 95% C.L., assuming kV
 2349 $= 1.0$.

2350 Dont forget to mention previous constrains to ct check reference ?? and references
 2351 <https://link.springer.com/content/pdf/10.1007%2FJHEP01%282013%29088.pdf> (para-
 2352 graph after eq 2)

2353 We selects events with three leptons and a b tagged jet in the final state. The tHq
 2354 signal contribution is then determined in a fit of the observed data to two multivariate
 2355 classifier outputs, each trained to discriminate against one of the two dominant back-
 2356 grounds of events with non-prompt leptons from $t\bar{t}$ and of associated production of $t\bar{t}$
 2357 and vector bosons ($t\bar{t}W$, $t\bar{t}Z$). The fit result is then used to set an upper limit on the
 2358 combined $t\bar{t}H$, tHq and tHW production cross section, as a function of the relative
 2359 coupling strengths of Higgs and top quark and Higgs and W boson, respectively.

2360 6.5 Data and MC Samples

2361 The data considered in this analysis were collected by the CMS experiment dur-
 2362 ing 2016 and correspond to a total integrated luminosity of $35.9 fb^{-1}$. Only periods
 2363 when the CMS magnet was on were considered when selecting the data samples, that
 2364 corresponds to the 23 Sep 2016 (Run B to G) and PromptReco (Run H) versions
 2365 of the datasets. The MC samples used in this analysis correspond to the RunI-

2366 ISummer16MiniAODv2 campaign produced with CMSSW 80X. The two signal sam-
 2367 ples (for tHq and tHW) were produced with MG5_aMC@NLO (version 5.222), in
 2368 leading-order order mode, and are normalized to next-to-leading-order cross sections,
 2369 see Tab. 6.3. Each sample is generated with a set of event weights corresponding to
 2370 different values of κ_t and κ_V couplings as shown in Tab. 6.4.

2371 **6.5.1 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 6.3: Signal samples and their cross section and branching fraction used in this analysis. See Ref. [129] for more details.

2372 Different MC generators were used to generate the background processes. The
 2373 dominant sources ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$) were produced using AMC@NLO interfaced
 2374 to PYTHIA8, and are scaled to NLO cross sections. Other processes are simulated
 2375 using POWHEG interfaced to PYTHIA, or bare PYTHIA (see table 6.5 and [120]
 2376 for more details).

2377 **6.5.2 Triggers**

2378 We consider online-reconstructed events triggered by one, two, or three leptons.
 2379 Single-lepton triggers are included to boost the acceptance of events where the p_T of
 2380 the sub-leading lepton falls below the threshold of the double-lepton triggers. Ad-
 2381 ditionally, by including double-lepton triggers in the ≥ 3 lepton category, as well
 2382 as single-lepton triggers in all categories, we increase the efficiency, considering the
 2383 logical “or” of the trigger decisions of all the individual triggers in a given category.

		<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 6.4: 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. [129].

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

Table 6.5: 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 6.6: Leading-order $t\bar{t}W$ and $t\bar{t}Z$ samples used in the signal BDT training.

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

2384 Tab. 6.7 shows the lowest-threshold non-preserved triggers present in the High-Level
 2385 Trigger (HLT) menus for both Monte-Carlo and data in 2016.

2386 **6.5.2.1 Trigger efficiency scale factors**

2387 The efficiency of events to pass the trigger is measured in simulation (trivially using
 2388 generator information) and in the data (using event collected by an uncorrelated
 2389 MET trigger). Small differences between the data and MC efficiencies are corrected
 2390 by applying scale factors as shown in Tab. 6.8. The exact procedure and control plots
 2391 are documented in [125] for the current analysis.

2392 **6.6 Background predictions**

2393 The modeling of reducible and irreducible backgrounds in this analysis uses the exact
 2394 methods, analysis code, and ROOT trees used for the $t\bar{t}H$ multilepton analysis. We

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

2395 give a brief description of the methods and refer to the documentation of that analysis
 2396 in Refs. [120, 125] for any details.

2397 The backgrounds in three-lepton final states can be split in two broad categories:
 2398 irreducible backgrounds with genuine prompt leptons (i.e. from on-shell W and Z
 2399 boson decays); and reducible backgrounds where at least one of the leptons is “non-
 2400 prompt”, i.e. produced within a hadronic jet, either a genuine lepton from heavy
 2401 flavor decays, or simply mis-reconstructed jets.

2402 Irreducible backgrounds can be reliably estimated directly from Monte-Carlo sim-
 2403 ulated events, using higher-order cross sections or data control regions for the overall
 2404 normalization. This is done in this analysis for all backgrounds involving prompt lep-
 2405 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 ,
 2406 ZZZ .

2407 Reducible backgrounds, on the other hand, are not well predicted by simulation,
 2408 and are estimated using data-driven methods. In the case of non-prompt leptons, a
 2409 fake rate method is used,

2410 Additional identification criteria are applied for electrons with p_T greater than 30
 2411 GeV to mimic the identification applied at trigger level in order to ensure consistency
 2412 between the measurement region and application region of the fake-rate.

2413 where the contribution to the final selection is estimated by extrapolating from
 2414 a sideband (or “application region”) with a looser lepton definition (the fakeable

2415 object definitions in Tabs. ?? and ??) to the signal selection. The tight-to-loose
 2416 ratios (or “fake rates”) are measured in several background dominated data events
 2417 with dedicated triggers, subtracting the residual prompt lepton contribution using
 2418 MC. Non-prompt leptons in our signal regions are predominantly produced in $t\bar{t}$
 2419 events, with a much smaller contribution, from Drell–Yan production. The systematic
 2420 uncertainty on the normalization of the non-prompt background estimation is on the
 2421 order of 50%, and thereby one of the dominant limitations on the performance of
 2422 multilepton analyses in general and this analysis in particular. It consists of several
 2423 individual sources, such as the result of closure tests of the method using simulated
 2424 events, limited statistics in the data control regions due to necessary prescaling of
 2425 lepton triggers, and the uncertainty in the subtraction of residual prompt leptons
 2426 from the control region.

2427 The fake background where the leptons pass the looser selection are weighted
 2428 according to how many of them fail the tight criteria. Events with a single failing
 2429 lepton are weighted with the factor $f/(1-f)$ for the estimate to the tight selection
 2430 region, where f is the fake rate. Events with two failing leptons are given the negative
 2431 weight $-f_i f_j / (1 - f_i)(1 - f_j)$, and for three leptons the weight is positive and equal
 2432 to the product of $f/(1-f)$ factor evaluated for each failing lepton.

2433 Figures 6.3 show the distributions of some relevant kinematic variables, normalized
 2434 to the cross section of the respective processes and to the integrated luminosity.

2435 **6.7 Signal discrimination**

2436 The tHq signal is separated from the main backgrounds using a boosted decision
 2437 tree (BDT) classifier, trained on simulated signal and background events. A set of
 2438 discriminating variables are given as input to the BDT which produces a output

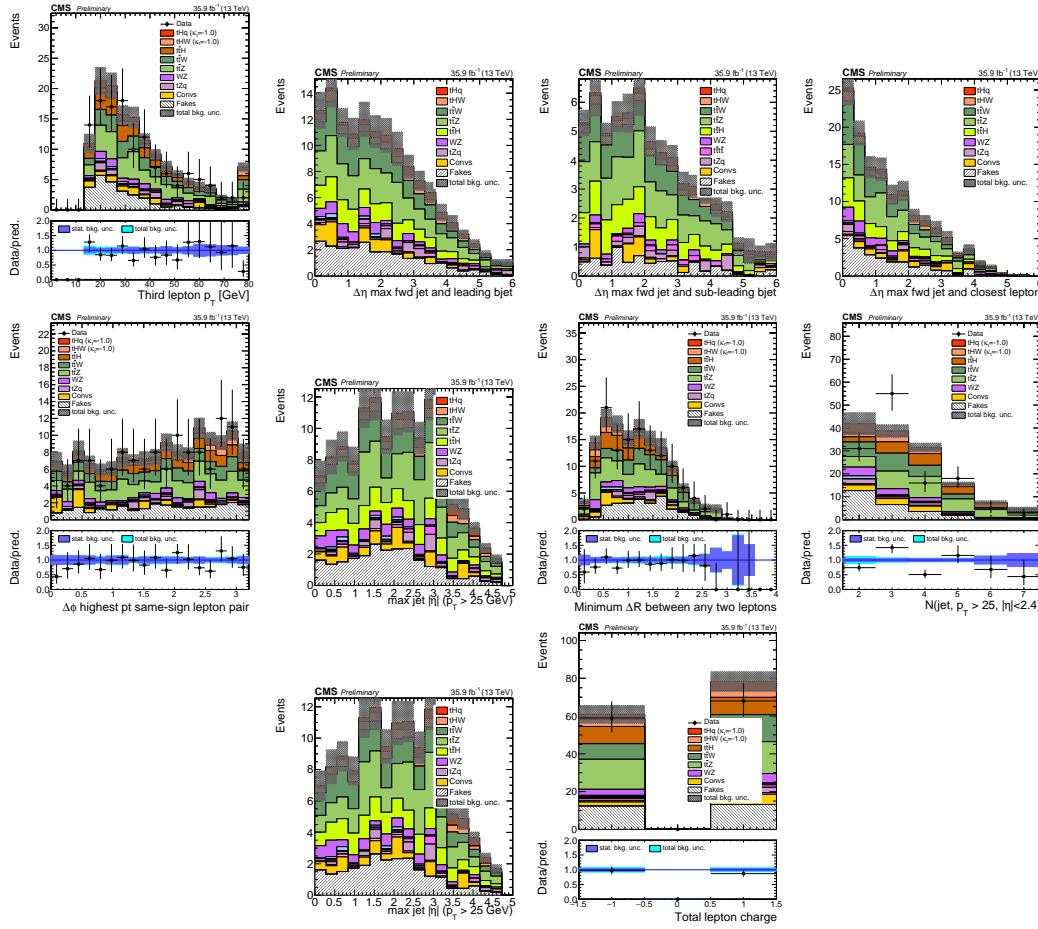


Figure 6.3: Distributions of input variables to the BDT for signal discrimination, three lepton channel, normalized to their cross section and to 35.9fb^{-1} .

distribution maximizing the discrimination power. Table 6.9 lists the input variables used while Figures 6.4 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. 6.3), and LO MADGRAPH samples of $t\bar{t}W$ and $t\bar{t}Z$, in an admixture proportional to their respective cross sections (see Tab. 6.6).

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

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

2448 is trained to discriminate between simulated signal and background events, then a
 2449 “test” stage where the trained algorithm is used to classify different events from
 2450 the samples. The sample is obtained from a pre-selection (see Tab. ?? with pre-
 2451 selection cuts). Figures 6.5 show the input variables distributions as seen by the
 2452 MVA algorithm. Note that in contrast to the distributions in Fig. 6.4 only the main
 2453 backgrounds ($t\bar{t}$ from simulation, $t\bar{t}V$) are included.

2454 Note that splitting the training in two groups reveals that some variables show
 2455 opposite behavior for the two background sources; potentially screening the discrimi-
 2456 nation power if they were to be used in a single discriminant. For some other variables
 2457 the distributions are similar in both background cases.

2458 From table 6.9, it is clear that the input variables are correlated to some extend.
 2459 These correlations play an important role for some MVA methods like the Fisher
 2460 discriminant method in which the first step consist of performing a linear transfor-
 2461 mation to an phase space where the correlations between variables are removed. In
 2462 case a boosted decision tree (BDT) method however, correlations do not affect the

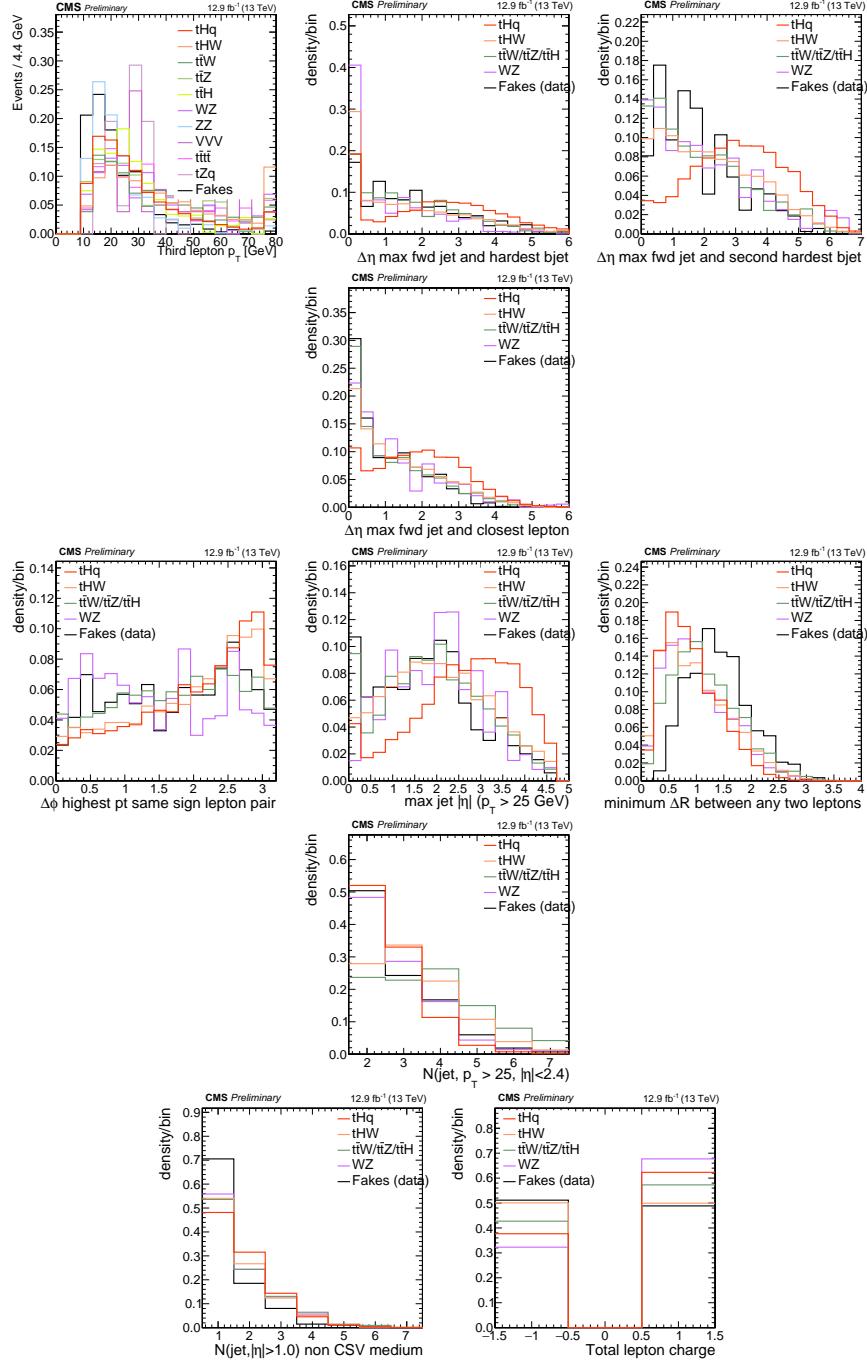


Figure 6.4: Distributions of input variables to the BDT for signal discrimination, three lepton channel.

2463 performance. Figure 6.7 show the linear correlation coefficients for signal and back-
 2464 ground for the two training cases (the signal values are identical by construction). As

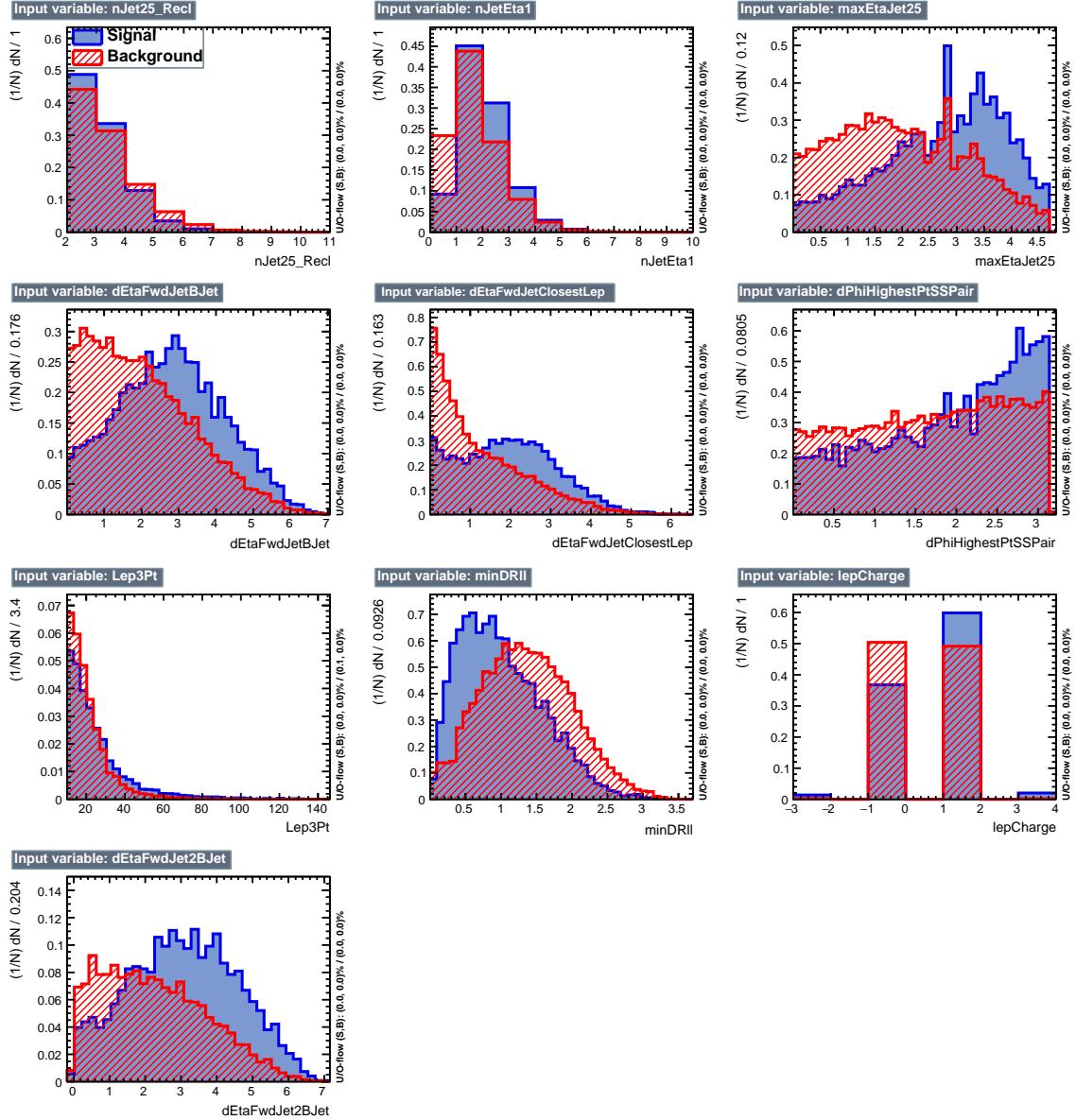


Figure 6.5: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}$) for the three lepton channel, discriminated against $t\bar{t}$ (fakes) background.

2465 expected, strong correlations appears for variables related to the forward jet activity.
 2466 Same trend is seen in case of the same sign dilepton channel in Figure ??.

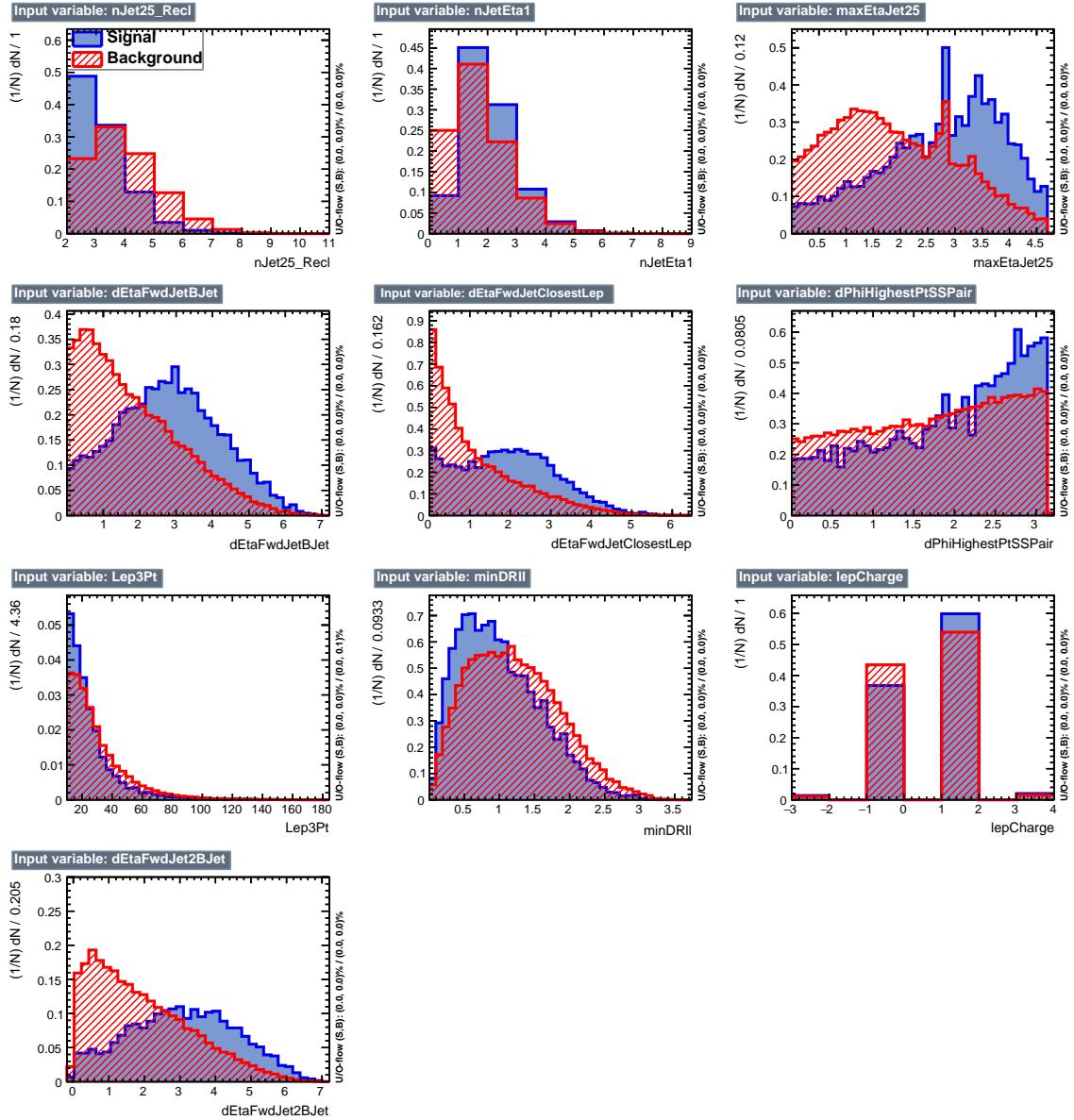


Figure 6.6: BDT inputs as seen by TMVA (signal, in blue, is tHq , background, in red, is $t\bar{t}W+t\bar{t}Z$) for the three lepton channel, discriminated against $t\bar{t}V$ background.

2467 6.7.1 Classifiers response

2468 Several MVA algorithms were evaluated to determine the most appropriate method
 2469 for this analysis. The plots in Fig. 6.8 (top) show the background rejection as a
 2470 function of the signal efficiency for $t\bar{t}$ and $t\bar{t}V$ trainings (ROC curves) for the different

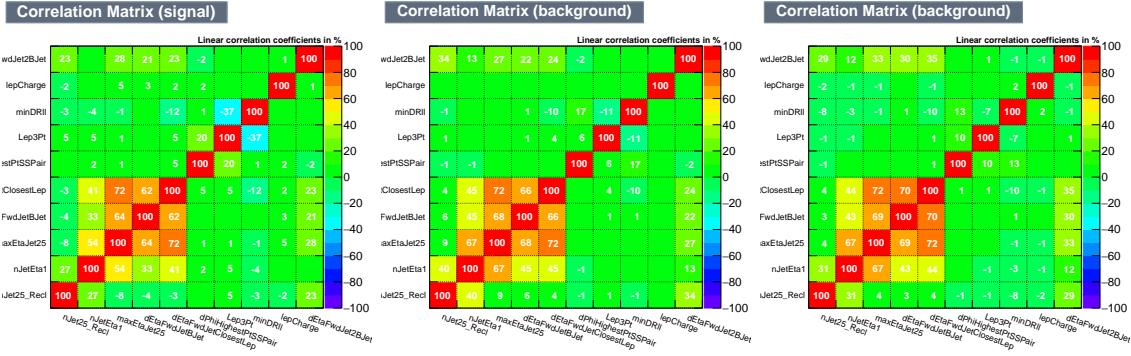


Figure 6.7: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the three lepton channel.

2471 algorithms that were evaluated.

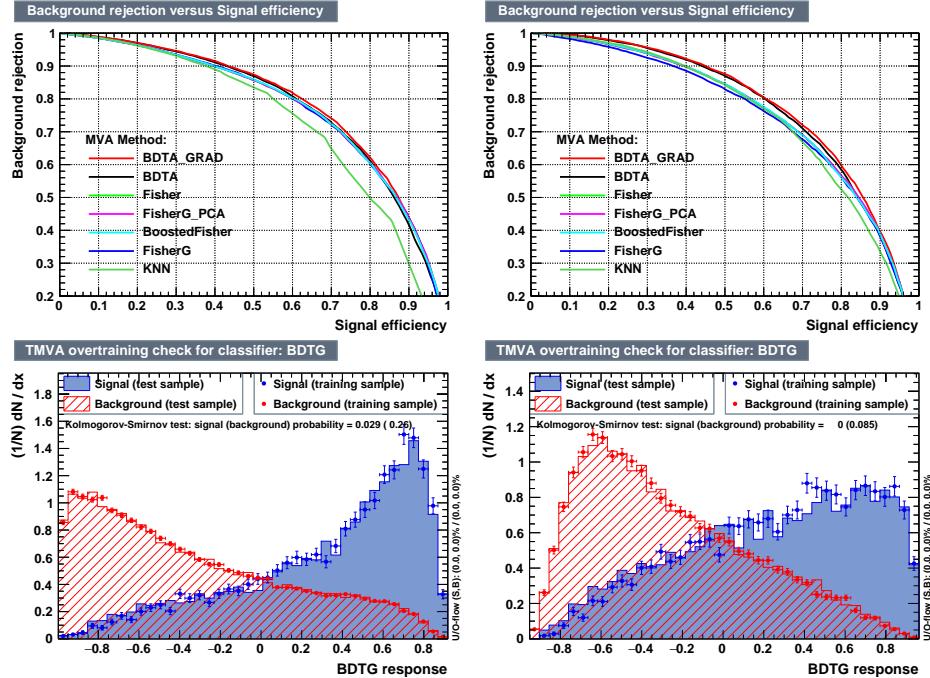


Figure 6.8: Top: background rejection vs signal efficiency (ROC curves) for various MVA classifiers (top) in the three lepton channel against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distributions for the gradient boosted decision trees, for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

2472 In both cases the gradient boosted decision tree ("BDTA_GRAD") classifier offers
 2473 the best results, followed by an adaptive BDT classifier ("BDTA"). The BDTA_GRAD
 2474 classifier output distributions for signal and backgrounds are shown on the bottom of

2475 Fig. 6.8. As expected, a good discrimination power is obtained using default discrim-
 2476 inator parameter values, with minimal overtraining. TMVA provides a ranking of the
 2477 input variables by their importance in the classification process, shown in Tab. 6.10.

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.

2478 The TMVA settings used in the BDT training are shown in Tab. 6.11.

2479 6.8 Additional discriminating variables

2480 Two additional discriminating variables were tested considering the fact that the
 2481 forward jet in the background could come from the pileup; since we have a real
 2482 forward jet in the signal, it could give some improvement in the discriminating power.

```
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.

2483 The additional variables describe the forward jet momentum (fwdJetPt25) and the
 2484 forward jet identification(fwdJetPUID). Distributions for these variables in the three
 2485 lepton channel are shown in the figure 6.9. The forward jet identification distribution
 2486 show that for both, signal and background, jets are mostly real jets.

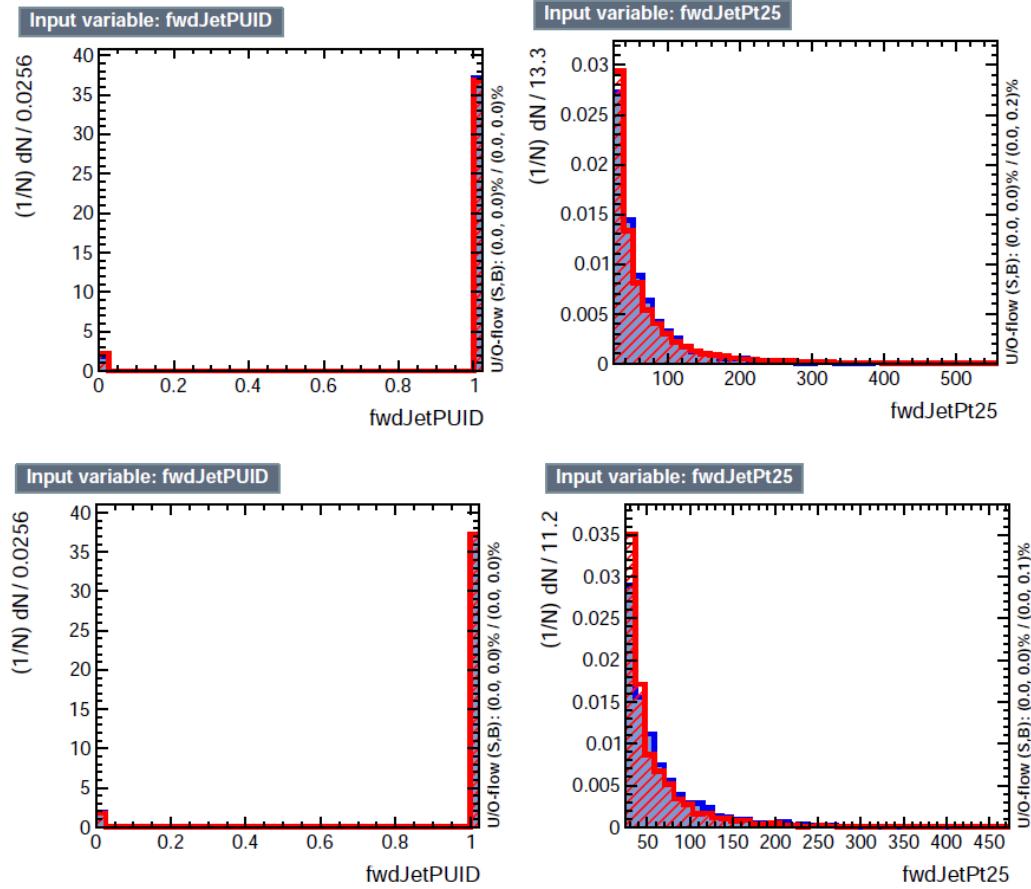


Figure 6.9: Additional discriminating variables distributions for ttv training(Top row) and tt training (bottom row) in the three lepton channel. The origin of the jets in the forward jet identification distribution is tagged as 0 for “pileup jets” while “real jets” are tagged as 1.

2487 The testing was made including in the MVA input one variable at a time, so we
 2488 can evaluate the dicrimination power of each variable, and then both simultaneously.
 2489 `fwdJetPUID` was ranked in the last place in importance (11) in both training (ttV
 2490 and tt) while `fwdJetPt25` was ranked 3 in the ttV training and 7 in the tt training.

	ROC-integral
base 10 var ttv	0.848
+ fwdJetPUID ttv	0.849
+ fwdJetPt25 ttv	0.856
12 var ttv	0.856
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%

2491 When training using 12 variables, fwdJetPt25 was ranked 5 and 7 in the ttV and tt
 2492 trainings respectively, while fwdJetPUID was ranked 12 in both cases.

2493 The improvement in the discrimination performance provided by the additional
 2494 variables is about 1%, so it was decided not to include them in the procedure. Table
 2495 6.12 show the ROC-integral for all the testing cases we made.

²⁴⁹⁶ **References**

- ²⁴⁹⁷ [1] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". Physical Review. 74 (10): 1439-61, (1948).
- ²⁴⁹⁹ [2] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". Physical Review. 76 (6): 769-89, (1949).
- ²⁵⁰¹ [3] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields". Progress of Theoretical Physics. 1 (2): 27-42, (1946).
- ²⁵⁰³ [4] D.J. Griffiths, "Introduction to electrodynamics". 4th ed. Pearson, (2013).
- ²⁵⁰⁴ [5] F. Mandl, G. Shaw. "Quantum field theory." Chichester: Wiley (2009).
- ²⁵⁰⁵ [6] F. Halzen, and A.D. Martin, "Quarks and leptons: An introductory course in modern particle physics". New York: Wiley, (1984) .
- ²⁵⁰⁷ [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>.
- ²⁵¹¹ [8] E. Noether, "Invariante Variationsprobleme", Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, vol. 1918, pp. 235-257, (1918).

- 2514 [9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)
2515 and 2017 update.
- 2516 [10] M. Goldhaber, L. Grodzins, A.W. Sunyar “Helicity of Neutrinos”, Phys. Rev.
2517 109, 1015 (1958).
- 2518 [11] Palanque-Delabrouille N et al. “Neutrino masses and cosmology with Lyman-
2519 alpha forest power spectrum”, JCAP 11 011 (2015).
- 2520 [12] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. Physics Letters.
2521 8 (3): 214-215 (1964).
- 2522 [13] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking”
2523 (PDF). CERN Report No.8182/TH.401 (1964).
- 2524 [14] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking:
2525 II” (PDF). CERN Report No.8419/TH.412(1964).
- 2526 [15] M. Gell-Mann. “The Interpretation of the New Particles as Displaced Charged
2527 Multiplets”. Il Nuovo Cimento 4: 848. (1956).
- 2528 [16] T. Nakano, K, Nishijima. “Charge Independence for V-particles”. Progress of
2529 Theoretical Physics 10 (5): 581-582. (1953).
- 2530 [17] N. Cabibbo, “Unitary symmetry and leptonic decays” Physical Review Letters,
2531 vol. 10, no. 12, p. 531, (1963).
- 2532 [18] M.Kobayashi, T.Maskawa, “CP-violation in the renormalizable theory of weak
2533 interaction,” Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).
- 2534 [19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Commons,
2535 the free media repository. Retrieved November 27, 2017 from

- 2536 https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)
2537 .svg&oldid=247498592.
- 2538 [20] Georgia Tech University. Coupling Constants for the Fundamental
2539 Forces(2005). Retrieved January 10, 2018, from http://hyperphysics.phy-
2540 astr.gsu.edu/hbase/Forces/couple.html#c2
- 2541 [21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved Jan-
2542 uary 10, 2018, from https://profmattstrassler.com/articles-and-posts/particle-
2543 physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/
- 2544 [22] S.L. Glashow. “Partial symmetries of weak interactions”, Nucl. Phys. 22 579-
2545 588, (1961).
- 2546 [23] A. Salam, J.C. Ward. “Electromagnetic and weak interactions”, Physics Letters
2547 13 168-171, (1964).
- 2548 [24] S. Weinberg, “A model of leptons”, Physical Review Letters, vol. 19, no. 21, p.
2549 1264, (1967).
- 2550 [25] M. Peskin, D. Schroeder, “An introduction to quantum field theory”. Perseus
2551 Books Publishing L.L.C., (1995).
- 2552 [26] A. Pich. “The Standard Model of Electroweak Interactions”
2553 https://arxiv.org/abs/1201.0537
- 2554 [27] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved
2555 from: https://www.quantum-bits.org/?p=233
- 2556 [28] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255

- 2557 [29] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge
 2558 Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964)
 2559 doi:10.1103/PhysRevLett.13.321
- 2560 [30] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Re-
 2561 view Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.
- 2562 [31] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws
 2563 and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964).
 2564 doi:10.1103/PhysRevLett.13.585.
- 2565 [32] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with
 2566 the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012).
 2567 arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021
- 2568 [33] ATLAS collaboration. “Observation of a New Particle in the Search for the Stan-
 2569 dard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters
 2570 B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.
- 2571 [34] ATLAS collaboration; CMS collaboration (26 March 2015). “Combined Mea-
 2572 surement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with
 2573 the ATLAS and CMS Experiments”. Physical Review Letters. 114 (19): 191803.
 2574 arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.
- 2575 [35] LHC InternationalMasterclasses“When protons collide”. Retrieved from http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm
 2576
- 2577 [36] CMS Collaboration, “SM Higgs Branching Ratios and Total Decay Widths (up-
 2578 date in CERN Report4 2016)”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR> , last accessed on 17.12.2017.
 2579

- 2580 [37] R.Grant V. “Determination of Higgs branching ratios in $H \rightarrow W^+W^- \rightarrow l\nu jj$
2581 and $H \rightarrow ZZ \rightarrow l^+l^-jj$ channels”. Physics Department, University of Tennessee
2582 (Dated: October 31, 2012). Retrieved from <http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf>
- 2584 [38] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al.
2585 “Standard model Higgs-boson branching ratios with uncertainties”. Eur. Phys.
2586 J. C (2011) 71: 1753. <https://doi.org/10.1140/epjc/s10052-011-1753-8>
- 2587 [39] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production
2588 of Higgs and single top at hadron colliders”, Phys.Rev. D64 (2001) 094023,
2589 [hep-ph/0106293].
- 2590 [40] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the
2591 top-Higgs coupling from the 8 TeV LHC data,” Journal of High Energy Physics,
2592 vol. 07, p. 073, (2013).
- 2593 [41] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting de-
2594 generacies in Higgs couplings using single top production in association with a
2595 Higgs boson,” Journal of High Energy Physics, vol. 05, p. 022, (2013).
- 2596 [42] T.M. Tait and C.-P. Yuan, “Single top quark production as a window to physics
2597 beyond the standard model”, Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].
- 2598 [43] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in
2599 association with a single top quark at the LHC,” European Physical Journal C,
2600 vol. 75, p. 267, (2015).

- 2601 [44] CMS Collaboration, “Modelling of the single top-quark production in associa-
2602 tion with the Higgs boson at 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV)
2603 [viewauth/CMS/SingleTopHiggsGeneration13TeV](#), last accessed on 16.01.2018.
- 2604 [45] CMS Collaboration, “SM Higgs production cross sections at $\sqrt{s} =$
2605 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)
2606 [view/LHCPhysics/CERNYellowReportPageAt13TeV](#), last accessed on 16.01.2018.
- 2607 [46] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.
- 2608 [47] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [[arXiv:1211.0499 \[hep-ph\]](#)].
- 2610 [48] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated
2611 production at the LHC”, European Physical Journal C, vol. 77, p. 34, (2017).
2612 [arXiv:1607.05862](#)
- 2613 [49] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross
2614 Sections: 4.Deciphering the Nature of the Higgs Sector”, [arXiv:1610.07922](#).
- 2615 [50] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top
2616 Couplings in Associated Production”, JHEP 1404 (2014) 004, [[arXiv:1312.5736](#)].
- 2617 [51] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass
2618 of the Higgs boson and tests of compatibility of its couplings with the standard
2619 model predictions using proton collisions at 7 and 8 TeV,” [arXiv:1412.8662](#).
- 2620 [52] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the
2621 Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton
2622 collision data”, ATLAS-CONF-2014-009.

- 2623 [53] ATLAS and CMS Collaborations, “Measurements of the Higgs boson produc-
2624 tion and decay rates and constraints on its couplings from a combined ATLAS
2625 and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,” (2016).
2626 CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.
- 2627 [54] File:Cern-accelerator-complex.svg. Wikimedia Commons,
2628 the free media repository. Retrieved January, 2018 from
2629 <https://commons.wikimedia.org/wiki/File:Cern-accelerator-complex.svg>
- 2630 [55] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”,
2631 (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Re-
2632 trieval from <https://cds.cern.ch/record/841542>
- 2633 [56] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-
2634 2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762
- 2635 [57] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).
- 2636 [58] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March
2637 1983, CERN-AC-8303511. Retrieved from <https://cds.cern.ch/record/615852>.
- 2638 [59] CERN Photographic Service “Animation of CERN’s accelerator net-
2639 work”, 14 October 2013. DOI: 10.17181/cds.1610170 Retrieved from
2640 <https://videos.cern.ch/record/1610170>
- 2641 [60] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17, 2013. Re-
2642 trieval from <https://www.britannica.com/technology/particle-accelerator>.
- 2643 [61] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-
2644 AC-0007016. Retrieved from <https://cds.cern.ch/record/41567>.

- 2645 [62] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”.
2646 March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-
2647 PHO-1998-325. Retrieved from <https://cds.cern.ch/record/841511>
- 2648 [63] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025
2649 retrieved from <https://cds.cern.ch/record/40524>.
- 2650 [64] CMS Collaboration “Public CMS Luminosity Information”.
2651 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016>
2652 _proton_proton_13_TeV_collis, last accessed 24.01.2018
- 2653 [65] J.L. Caron. “LHC Layout” AC Collection. Legacy of AC. Pictures
2654 from 1992 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from
2655 <https://cds.cern.ch/record/841573>.
- 2656 [66] J.A. Coarasa. “The CMS Online Cluster: Setup, Operation and Maintenance
2657 of an Evolving Cluster”. ISGC 2012, 26 February - 2 March 2012, Academia
2658 Sinica, Taipei, Taiwan.
- 2659 [67] CMS Collaboration. “The CMS experiment at the CERN LHC” JINST 3 S08004
2660 (2008).
- 2661 [68] CMS Collaboration. “CMS detector drawings 2012” CMS-PHO-GEN-2012-002.
2662 Retrieved from <http://cds.cern.ch/record/1433717>.
- 2663 [69] R. Breedon. “View through the CMS detector during the cooldown of the
2664 solenoid on February 2006. CMS Collection”, February 2006, CMS-PHO-
2665 OREACH-2005-004, Retrieved from <https://cds.cern.ch/record/930094>.
- 2666 [70] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector
2667 Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.

- 2668 [71] CMS Collaboration. “Description and performance of track and primary-vertex
2669 reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no.
2670 10, p. P10009,(2014).
- 2671 [72] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner
2672 Barrel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from
2673 <https://cds.cern.ch/record/1431467>.
- 2674 [73] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm
2675 and beauty hadrons Chicago, June 28-July 3 2004.
- 2676 [74] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at
2677 the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013.
2678 arXiv:1307.7135
- 2679 [75] L. Veillet. “End assembly of HB with EB rails and rotation in-
2680 side SX ”,January 2002. CMS-PHO-HCAL-2002-002. Retrieved from
2681 <https://cds.cern.ch/record/42594>.
- 2682 [76] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the
2683 UX5 cavern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from
2684 <https://cds.cern.ch/record/1019185>
- 2685 [77] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web
2686 site: Taking a closer look at LHC. Retrieved from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger
- 2688 [78] WLCG Project Office, “Documents & Reference - Tiers - Structure,” (2014).
2689 <http://wlcg.web.cern.ch/documents-reference> , last accessed on 30.01.2018.

- 2690 [79] CMS Collaboration. “CMSSW Application Framework”,
2691 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>,
2692 last accesses 06.02.2018
- 2693 [80] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event gen-
2694 erators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011
- 2695 [81] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics
2696 of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-
2697 3-540-71060-8 (2007)
- 2698 [82] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter.”
2699 <http://hepdata.cedar.ac.uk/pdf/pdf3.html> , last accessed on 02.02.2018.
- 2700 [83] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a
2701 reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).
- 2702 [84] B. Andersson, G. Gustafson, G.Ingelman and T. Sjostrand, “Parton fragmen-
2703 tation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145,
2704 1983.
- 2705 [85] CMS Collaboration, “Event generator tunes obtained from underlying event
2706 and multiparton scattering measurements;” European Physical Journal C, vol.
2707 76, no. 3, p. 155, (2016).
- 2708 [86] J. Alwall et. al., “The automated computation of tree-level and next-to-leading
2709 order differential cross sections, and their matching to parton shower simula-
2710 tions,” Journal of High Energy Physics, vol. 07, p. 079, (2014).

- 2711 [87] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with
2712 Parton Shower simulations: the POWHEG method,” Journal of High Energy
2713 Physics, vol. 11, p. 070, (2007).
- 2714 [88] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments
2715 and Methods in Physics, vol. A506, pp. 250–303, (2003).
- 2716 [89] J.Allison et.al.,“Recent developments in Geant4”, Nuclear Instruments and
2717 Methods in Physics Research A 835 (2016) 186-225.
- 2718 [90] CMS Collaboration “Full Simulation Offline Guide”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation>, last accessed 04.02.2018
- 2720 [91] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf.
2721 Ser. 513 022012 (2014)
- 2722 [92] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description
2723 with the CMS detector”, JINST 12 P10003 (2017) <https://doi.org/10.1088/1748-0221/12/10/P10003>.
- 2725 [93] The CMS Collaboration. “ Description and performance of track and pri-
2726 mary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014).
2727 doi:10.1088/1748-0221/9/10/P10009
- 2728 [94] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides.
2729 Retrieved from https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf
- 2731 [95] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by
2732 the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).

- 2733 [96] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of
 2734 electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf
 2735 C 0303241 (2003) TULT009 [physics/0306087].
- 2736 [97] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification,
 2737 Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.
- 2738 [98] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”,
 2739 CMS Note 2007-008 (2007).
- 2740 [99] CMS collaboration, “Performance of CMS muon reconstruction in pp collision
 2741 events at $\sqrt{s} = 7$ TeV ”, JINST 7 P10002 2012, [arXiv:1206.4071].
- 2742 [100] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,”
 2743 Journal of High Energy Physics, vol. 04, p. 063, (2008).
- 2744 [101] B. Dorney. “Anatomy of a Jet in CMS”. Quantum Diaries. June
 2745 1st, 2011. Retrieved from <https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>
- 2747 [102] The CMS Collaboration.“Event Displays from the high-energy collisions at 7
 2748 TeV”, May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from <https://cds.cern.ch/record/1429614>.
- 2750 [103] The CMS collaboration. “Determination of jet energy calibration and transverse
 2751 momentum resolution in CMS”. JINST 6 P11002 (2011). <http://dx.doi.org/10.1088/1748-0221/6/11/P11002>
- 2753 [104] The CMS Collaboration, “Introduction to Jet Energy Corrections at
 2754 CMS.”. <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>, last ac-
 2755 cessed 10.02.2018.

- 2756 [105] CMS Collaboration Collaboration. “Identification of b quark jets at the CMS
2757 Experiment in the LHC Run 2”. Tech. rep. CMS-PAS-BTV-15-001. Geneva:
2758 CERN, (2016). <https://cds.cern.ch/record/2138504>.
- 2759 [106] CMS Collaboration Collaboration. “Performance of missing energy reconstruc-
2760 tion in 13 TeV pp collision data using the CMS detector”. Tech. rep. CMS-PAS-
2761 JME16-004. Geneva: CERN, 2016. <https://cds.cern.ch/record/2205284>.
- 2762 [107] K. Skovpen. “Event displays highlighting the main properties of heavy flavour
2763 jets in the CMS Experiment”, Aug 2017, CMS-PHO-EVENTS-2017-006. Re-
2764 trieved from <https://cds.cern.ch/record/2280025>.
- 2765 [108] G. Cowan. “Topics in statistical data analysis for high-energy physics”.
2766 arXiv:1012.3589v1
- 2767 [109] A. Hoecker et al., “TMVA-Toolkit for multivariate data analysis”
2768 arXiv:physics/0703039v5 (2009)
- 2769 [110] L. Lista. “Statistical Methods for Data Analysis in Particle Physics”, 2nd
2770 ed. Springer International Publishing. (2017) <https://dx.doi.org/10.1007/978-3-319-62840-0>
- 2772 [111] I. Antcheva et al., “ROOT-A C++ framework for petabyte data storage, sta-
2773 tistical analysis and visualization ,” Computer Physics Communications, vol.
2774 182, no. 6, pp. 1384–1385, (2011).
- 2775 [112] Y. Coadou. “Boosted decision trees”, ESIPAP, Archamps, 9 Febru-
2776 ary 2016. Lecture. Retrieved from https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf

- 2779 [113] J.H. Friedman. “Greedy function approximation: A gradient boosting ma-
 2780 chine”. Ann. Statist. Volume 29, Number 5 (2001), 1189-1232. https://projecteuclid.org/download/pdf_1/euclid-aos/1013203451.
- 2782 [114] F. James, M. Roos, “MINUIT: Function minimization and error analysis”. Cern
 2783 Computer Centre Program Library, Geneve Long Write-up No. D506, 1989
- 2784 [115] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests
 2785 of statistical hypotheses”. Philosophical Transactions of the Royal Society of
 2786 London. Series A, Containing Papers of a Mathematical or Physical Character.
 2787 Vol. 231 (1933), pp. 289-337
- 2788 [116] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated pro-
 2789 duction via gluon fusion in the SM and the 2HDM”, JHEP 06 (2015) 065,
 2790 doi:10.1007/JHEP06(2015)065, arXiv:1503.01656.
- 2791 [117] ATLAS Collaboration, “Measurements of Higgs boson produc-
 2792 tion and couplings in diboson final states with the ATLAS
 2793 detector at the LHC”, Phys. Lett. B726 (2013) 88â€¢119,
 2794 doi:10.1016/j.physletb.2014.05.011,10.1016/j.physletb.2013.08.010,
 2795 arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].
- 2796 [118] CMS Collaboration, “Search for the associated production of a Higgs boson
 2797 with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV”, JHEP 06
 2798 (2016) 177,doi:10.1007/JHEP06(2016)177, arXiv:1509.08159.
- 2799 [119] B. Stieger, C. Jorda Lope et al., “Search for Associated Production of a Single
 2800 Top Quark and a Higgs Boson in Leptonic Channels”, CMS Analysis Note CMS
 2801 AN-14-140, 2014.

- 2802 [120] M. Peruzzi, C. Mueller, B. Stieger et al., “Search for ttH in multilepton final
2803 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-16-211, 2016.
- 2804 [121] CMS Collaboration, “Search for H to bbar in association with a single top quark
2805 as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis
2806 Summary CMS-PAS-HIG-16-019, 2016.
- 2807 [122] CMS Collaboration, “Search for production of a Higgs boson and a single top
2808 quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV”, CMS
2809 Physics Analysis Summary CMS-PAS-HIG-17-005, 2016.
- 2810 [123] B. WG, “BtagRecommendation80XReReco”, February, 2017. <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>.
- 2812 [124] CMS Collaboration, “Identification of b quark jets at the CMS Experiment
2813 in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001,
2814 2016.
- 2815 [125] M. Peruzzi, F. Romeo, B. Stieger et al., “Search for ttH in multilepton final1
2816 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-17-029, 2017.
- 2817 [126] CMS Collaboration, “Baseline muon selections for Run-II.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonIdRun2>, last accessed on
2818 24.02.2018.
- 2820 [127] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January,
2821 2015. <https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf>.

- 2823 [128] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”,
2824 October, 2014. [https://indico.cern.ch/event/298249/contribution/3/
2825 material/slides/0.pdf](https://indico.cern.ch/event/298249/contribution/3/material/slides/0.pdf).
- 2826 [129] B. Maier, “SingleTopHiggProduction13TeV”, February, 2016.
2827 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV>.