

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

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	xi
²⁵	1 Theoretical approach	1
²⁶	1.1 Introduction	1
²⁷	1.2 Standard model of particle physics	2
²⁸	1.2.1 Fermions	4
²⁹	1.2.1.1 Leptons	5
³⁰	1.2.1.2 Quarks	7
³¹	1.2.2 Fundamental interactions	11
³²	1.2.3 Gauge invariance.	15
³³	1.2.4 Gauge bosons	17
³⁴	1.3 Electroweak unification and the Higgs mechanism	18
³⁵	1.3.1 Spontaneous symmetry breaking (SSB)	26
³⁶	1.3.2 Higgs mechanism	30
³⁷	1.3.3 Masses of the gauge bosons	33
³⁸	1.3.4 Masses of the fermions	34

39	1.3.5	The Higgs field	35
40	1.3.6	Production of Higgs bosons at LHC	36
41	1.3.7	Higgs boson decay channels	40
42	1.4	Experimental status of the anomalous Higgs-fermion coupling	42
43	1.5	Associated production of a Higgs boson and a single top quark	44
44	1.6	CP-mixing in tH processes	49
45	2	The CMS experiment at the LHC	54
46	2.1	Introduction	54
47	2.2	The LHC	55
48	2.3	The CMS experiment	65
49	2.3.1	CMS coordinate system	68
50	2.3.2	Tracking system	70
51	2.3.3	Silicon strip tracker	73
52	2.3.4	Electromagnetic calorimeter	74
53	2.3.5	Hadronic calorimeter	76
54	2.3.6	Superconducting solenoid magnet	77
55	2.3.7	Muon system	79
56	2.3.8	CMS trigger system	80
57	2.3.9	CMS computing	81
58	3	Event generation, simulation and reconstruction	86
59	3.1	Event generation	87
60	3.2	Monte Carlo Event Generators.	90
61	3.3	CMS detector simulation.	92
62	3.4	Event reconstruction.	94
63	3.4.1	Particle-Flow Algorithm.	94

64	3.4.1.1	Missing transverse energy.	108
65	3.4.2	Event reconstruction examples	109
66	5	Statistical methods	112
67	5.1	Multivariate analysis	112
68	5.1.1	Decision trees	116
69	5.1.2	Boosted decision trees (BDT)	119
70	5.1.3	Overtraining	121
71	5.1.4	Variable ranking	122
72	5.1.5	BDT output example	122
73	5.2	Statistical inference	123
74	5.2.1	Nuisance parameters	124
75	5.2.2	Maximum likelihood estimation method	125
76	5.3	Upper limits	126
77	5.4	Asymptotic limits	130
78	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	132
80	6.1	Introduction	132
81	6.2	<i>tHq</i> signature	135
82	6.3	Background processes	137
83	6.4	Data and MC Samples	139
84	6.4.1	Full 2016 data set	139
85	6.4.2	Triggers	140
86	6.4.3	Signal modeling and MC samples	143
87	6.5	Object Identification	145
88	6.5.1	Lepton reconstruction and identification	146

89	6.5.2	Lepton selection efficiency	152
90	6.5.3	Jets and b -jet tagging	156
91	6.5.4	Missing Energy MET	157
92	6.6	Event selection	158
93	6.7	Signal discrimination	159
94	6.7.1	MVA classifiers evaluation	160
95	6.7.2	Discriminating variables	160
96	6.7.3	BDTG classifiers response	165
97	6.7.4	Additional discriminating variables	166
98	6.7.5	Signal extraction procedure	169
99	6.7.6	Binning and selection optimization	170
100	6.8	Signal model	173
101	6.9	Background modeling and predictions	176
102	6.9.1	Diboson backgrounds	177
103	A Datasets and triggers		183
104	B BDTG aditional plots		187
105	B.1	BDTG input variables distributions for $2lss$ channel	187
106	B.2	Input variables distributions from BDTG classifiers	189
107	C Other binning strategies		193
108	D BDTG output variation with κ_V/κ_t		196
109	Bibliography		196
110	References		197

₁₁₁ List of Figures

112	1.1	Standard Model of particle physics.	3
113	1.2	Transformations between quarks	11
114	1.3	Fundamental interactions in nature.	12
115	1.4	SM interactions diagrams	13
116	1.5	Neutral current processes	19
117	1.6	Spontaneous symmetry breaking mechanism	27
118	1.7	SSB Potential form	28
119	1.8	Potential for complex scalar field	29
120	1.9	SSB mechanism for complex scalar field	30
121	1.10	Proton-Proton collision	36
122	1.11	Proton PDFs	37
123	1.12	Higgs boson production mechanism Feynman diagrams	38
124	1.13	Higgs boson production cross section and decay branching ratios	39
125	1.14	κ_t - κ_V plot of the coupling modifiers. ATLAS and CMS combination.	42
126	1.15	Higgs boson production in association with a top quark	45
127	1.16	Cross section for tHq process as a function of κ_t	48
128	1.17	Cross section for tHW process as a function of κ_{Htt}	48
129	1.18	NLO cross section for tX_0 and $t\bar{t}X_0$.	52

130	1.19 NLO cross section for $tWX_0, t\bar{t}X_0$	53
131	2.1 CERN accelerator complex	55
132	2.2 LHC protons source. First acceleration stage.	56
133	2.3 The LINAC2 accelerating system at CERN.	57
134	2.4 LHC layout and RF cavities module.	58
135	2.5 LHC dipole magnet.	60
136	2.6 Integrated luminosity delivered by LHC and recorded by CMS during 2016	62
137	2.7 LHC interaction points	63
138	2.8 Multiple pp collision bunch crossing at CMS.	65
139	2.9 Layout of the CMS detector	66
140	2.10 CMS detector transverse slice	67
141	2.11 CMS detector coordinate system	69
142	2.12 CMS tracking system schematic view.	70
143	2.13 CMS pixel detector	71
144	2.14 SST Schematic view.	73
145	2.15 CMS ECAL schematic view	75
146	2.16 CMS HCAL schematic view	77
147	2.17 CMS solenoid magnet	78
148	2.18 CMS Muon system schematic view	79
149	2.19 CMS Level-1 trigger architecture	81
150	2.20 WLCG structure	82
151	2.21 Data flow from CMS detector through hardware Tiers	84
152	3.1 Event generation process.	87
153	3.2 Particle flow algorithm.	95
154	3.3 Stable cones identification	102

155	3.4	Jet reconstruction.	104
156	3.5	Jet energy corrections.	106
157	3.6	Secondary vertex in a b-hadron decay.	107
158	3.7	HIG-13-004 Event 1 reconstruction.	109
159	3.8	$e\mu$ event reconstruction.	110
160	3.9	Recorded event reconstruction.	111
161	5.1	Scatter plots-MVA event classification.	114
162	5.2	Scalar test statistical.	115
163	5.3	Decision tree.	116
164	5.4	Decision tree output example.	119
165	5.5	BDT output example.	122
166	5.6	t_r p.d.f. assuming each H_0 and H_1	128
167	5.7	Illustration of the CL_s limit.	129
168	5.8	Example of Brazilian flag plot	130
169	6.1	Analysis strategy workflow	135
170	6.2	tHq event signature	136
171	6.3	$t\bar{t}$ and $t\bar{t}W$ events signature	139
172	6.4	Trigger efficiency for the same-sign $\mu\mu$ category	141
173	6.5	Trigger efficiency for the $e\mu$ category	142
174	6.6	Trigger efficiency for the $3l$ category	143
175	6.7	tHq and tHW cross section in the κ_t - κ_V phase space	144
176	6.8	Tight vs loose lepton selection efficiencies in the $2lss$ channel.	154
177	6.9	Tight vs loose lepton selection efficiencies in the $3l$ channel.	155
178	6.10	MVA classifiers performance.	161
179	6.11	BDTG classifier Input variables distributions.	163

180	6.12 BDT input variables. Discrimination against $t\bar{t}$ and $t\bar{t}V$ in $3l$ channel.	164
181	6.13 Correlation matrices for the BDT input variables.	165
182	6.14 BDTG classifier response. Default parameters.	166
183	6.15 BDTG classifier output.	167
184	6.16 Additional discriminating variables distributions.	168
185	6.17 2D BDT classifier output planes	170
186	6.18 Binning overlaid on the S/B ratio map on the plane of classifier outputs.	171
187	6.19 Binning combination scheme.	172
188	6.20 Kinematic distributions in the diboson control region.	178
189	6.21 Input variables to the BDT for signal discrimination normalized.	180
190	B.1 Distributions of input variables to the BDT for signal discrimination, two lepton same sign channel.	188
191	B.2 BDT input variables. Discrimination against $t\bar{t}$ in $2lss$ channel.	189
192	B.3 BDT input variables. Discrimination against $t\bar{t}V$ in $2lss$ channel.	190
193	B.4 BDT input variables. Discrimination against $t\bar{t}$ in $3l$ channel.	191
194	B.5 BDT input variables. Discrimination against $t\bar{t}V$ in $3l$ channel.	192
195	C.1 Binning by S/B regions for $2lss$ (left) and $3l$ (right).	193
196	C.2 Final bins (corresponding to S/B regions in the 2D plane)	194
197	C.3 Binning into geometric regions using a k -means algorithm.	195
198	C.4 Final bins using a k -means algorithm.	195
200	D.1 BDTG output variation with κ_V/κ_t	196

²⁰¹ List of Tables

202	1.1	Fermions of the SM.	4
203	1.2	Fermion masses.	5
204	1.3	Lepton properties.	7
205	1.4	Quark properties.	8
206	1.5	Fermion weak isospin and weak hypercharge multiplets.	9
207	1.6	Fundamental interactions features.	14
208	1.7	SM gauge bosons.	18
209	1.8	Higgs boson properties.	36
210	1.9	Predicted branching ratios for a SM Higgs boson with $m_H = 125 \text{ GeV}/c^2$	41
211	1.10	Predicted SM cross sections for tH production at $\sqrt{s} = 13 \text{ TeV}$	46
212	1.11	Predicted enhancement of the tHq and tHW cross sections at LHC	49
213	6.1	Trigger efficiency scale factors and associated uncertainties.	142
214	6.2	MC signal samples.	144
215	6.3	Effective areas, for electrons and muons.	150
216	6.4	Requirements on each of the three muon selections.	152
217	6.5	Criteria for each of the three electron selections.	153
218	6.6	Summary of event selection.	158
219	6.7	BDTG input variables.	162

220	6.8 Configuration used in the final BDTG training.	166
221	6.9 Input variables ranking for BDTG classifiers	167
222	6.10 ROC-integral for all the testing cases.	169
223	6.11 Selection cuts optimization.	171
224	6.12 Limit variation as a function of bin size.	172
225	6.13 Limit variation as a function of bin size in the same-sign dimuon channel.	
226	(In bold: the final bin borders used in the $2lss$ channel.)	173
227	6.14 The 33 distinct values of κ_t/κ_V and f_t as mapped by the 51 κ_t and κ_V	
228	points.	175
229	A.1 Full 2016 dataset.	183
230	A.2 HLT	184
231	A.3 κ_V and κ_t combinations.	185
232	A.4 List of background samples used in this analysis (CMSSW 80X).	186

²³³ Chapter 1

²³⁴ Theoretical approach

²³⁵ 1.1 Introduction

²³⁶ The physical description of the universe is a challenge that physicists have faced by
²³⁷ making theories that refine existing principles and proposing new ones in an attempt
²³⁸ to embrace emerging facts and phenomena.

²³⁹ At the end of 1940s Julian Schwinger [1] and Richard P. Feynman [2], based on
²⁴⁰ the work of Sin-Itiro Tomonaga [3], developed an electromagnetic theory consistent
²⁴¹ with special relativity and quantum mechanics that describes how matter and light
²⁴² interact; the so-called *quantum electrodynamics* (QED) was born.

²⁴³ QED has become the blueprint for developing theories that describe the universe.
²⁴⁴ It was the first example of a quantum field theory (QFT), which is the theoretical
²⁴⁵ framework for building quantum mechanical models that describes particles and their
²⁴⁶ interactions. QFT is composed of a set of mathematical tools that combines classical
²⁴⁷ fields, special relativity and quantum mechanics, while keeping the quantum point
²⁴⁸ particles and locality ideas.

²⁴⁹ This chapter gives an overview of the standard model of particle physics, starting

250 with a description of the particles and their interactions, followed by a description of
 251 the electroweak interaction, the Higgs boson and the associated production of Higgs
 252 boson and a single top quark (tH). The description contained in this chapter is based
 253 on References [4–6].

254 1.2 Standard model of particle physics

255 The *standard model of particle physics (SM)* describes particle physics at the funda-
 256 mental level in terms of a collection of interacting particles and fields. The full picture
 257 of the SM is composed of three fields¹ whose excitations are interpreted as particles
 258 called mediators or force-carriers, a set of fields whose excitations are interpreted as
 259 elementary particles interacting through the exchange of those mediators, and a field
 260 that gives the mass to elementary particles. Figure 1.1 shows a scheme of the SM
 261 particles’ organization. In addition, for each of the particles in the scheme there exists
 262 an antiparticle with the same mass and opposite quantum numbers. The existence of
 263 antiparticles is a prediction of the relativistic quantum mechanics from the solution
 264 of the Dirac equation for which a negative energy solution is also possible. In some
 265 cases a particle is its own anti-particle, like photon or Higgs boson.

266 The mathematical formulation of the SM is based on group theory and the use of
 267 Noether’s theorem [8] which states that for a physical system modeled by a Lagrangian
 268 that is invariant under a group of transformations a conservation law is expected. For
 269 instance, a system described by a time-independent Lagrangian is invariant (symmet-
 270 ric) under time changes (transformations) with the total energy conservation law as
 271 the expected conservation law. In QED, the charge operator (Q) is the generator of

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

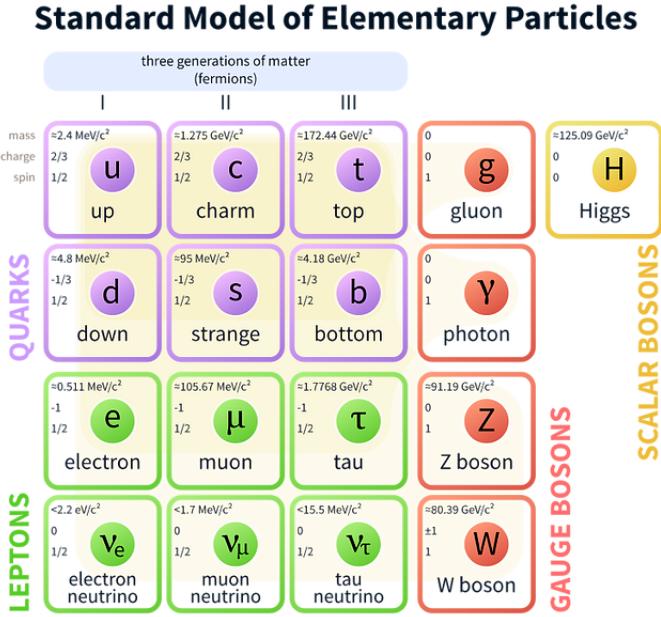


Figure 1.1: Schematic representation of the Standard Model of particle physics. The SM is a theoretical model intended to describe three of the four fundamental forces of the universe in terms of a set of particles and their interactions. [7].

272 the U(1) symmetry which according to the Noether's theorem means that there is a
273 conserved quantity; this conserved quantity is the electric charge and thus the law
274 conservation of electric charge is established.

In the SM, the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ describes three of the four fundamental interactions in nature (see Section 1.2.2): strong interaction (SI), weak interaction (WI) and electromagnetic interactions (EI) in terms of symmetries associated to physical quantities:

- 279 • Strong: $SU(3)_C$ associated to color charge

280 • Weak: $SU(2)_L$ associated to weak isospin and chirality

281 • Electromagnetic: $U(1)_Y$ associated to weak hypercharge and electric charge

282 It will be shown that the electromagnetic and weak interactions are combined in

283 the so-called electroweak interaction where chirality, hypercharge, weak isospin and
 284 electric charge are the central concepts.

285 **1.2.1 Fermions**

286 The basic constituents of the ordinary matter at the lowest level, which form the set
 287 of elementary particles in the SM formulation, are quarks and leptons. All of them
 288 have spin 1/2, therefore they are classified as fermions since they obey Fermi-Dirac
 289 statistics. There are six *flavors* of quarks and three of leptons organized in three
 290 generations, or families, as shown in Table 1.1.

		Generation		
Type		1st	2nd	3rd
Leptons	Charged	Electron (e)	Moun(μ)	Tau (τ)
	Neutral	Electron neutrino (ν_e)	Muon neutrino (ν_μ)	Tau neutrino (ν_τ)
Quarks	Up-type	Up (u)	Charm (c)	Top (t)
	Down-type	Down (d)	Strange (s)	Bottom (b)

Table 1.1: Fermions of the SM. There are six flavors of quarks and three of leptons, organized in three generations, or families, composed of two pairs of closely related particles. The close relationship is motivated by the fact that each pair of particles is a member of an $SU(2)_L$ doublet that has an associated invariance under isospin transformations. WI between leptons is limited to the members of the same generation; WI between quarks is not limited but greatly favored, to same generation members.

291

292 There is a mass hierarchy between generations (see Table 1.2), where the higher
 293 generation particles decays to the lower one, which can explain why the ordinary
 294 matter is made of particles from the first generation. In the SM, neutrinos are modeled
 295 as massless particles so they are not subject to this mass hierarchy; however, today it
 296 is known that neutrinos are massive so the hierarchy could be restated. The reason
 297 behind this mass hierarchy is one of the most important open questions in particle
 298 physics, and it becomes more puzzling when noticing that the mass difference between

299 first and second generation fermions is small compared to the mass difference with
 300 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 1.2: Fermion masses [9]. Generations differ by mass in a way that has been interpreted as a mass hierarchy. Approximate values with no uncertainties are used, for comparison purpose.

301

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

304 **1.2.1.1 Leptons**

305 A lepton is an elementary particle that is not subject to the SI. As seen in Table 1.1,
 306 there are two types of leptons, the charged ones (electron, muon and tau) and the
 307 neutral ones (the three neutrinos). The electric charge (Q) is the property that gives
 308 leptons the ability to participate in the EI. From the classical point of view, Q plays
 309 a central role determining, among others, the strength of the electric field through
 310 which the electromagnetic force is exerted. It is clear that neutrinos are not affected
 311 by EI because they don't carry electric charge.

312 Another feature of the leptons that is fundamental in the mathematical description
 313 of the SM is the chirality, which is closely related to spin and helicity. Helicity
 314 defines the handedness of a particle by relating its spin and momentum such that
 315 if they are parallel then the particle is right-handed; if spin and momentum are

316 antiparallel the particle is said to be left-handed. The study of parity conservation
 317 (or violation) in β -decay has shown that only left-handed electrons/neutrinos or right-
 318 handed positrons/anti-neutrinos are created [10]; the inclusion of that feature in the
 319 theory was achieved by using projection operators for helicity, however, helicity is
 320 frame dependent for massive particles which makes it not Lorentz invariant and then
 321 another related attribute has to be used: *chirality*.

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

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

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

$$\begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, l_R := \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, e_R, \mu_R, \tau_R, \nu_{eR}, \nu_{\mu R}, \nu_{\tau R} \quad (1.1)$$

337 The isospin third component refers to the eigenvalues of the weak isospin operator

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

338 which for doublets is $T_3 = \pm 1/2$, while for singlets it is $T_3 = 0$. The physical meaning
 339 of this doublet-singlet arrangement falls in that the WI couples the two particles in
 340 the doublet by exchanging the interaction mediator while the singlet member is not
 341 involved in WI. The main properties of the leptons are summarized in Table 1.3.

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

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

Table 1.3: Lepton properties [9]. Q: electric charge, T_3 : weak isospin. Only left-handed leptons and right-handed anti-leptons participate in the WI. Anti-particles with inverted T_3 , Q and lepton number complete the leptons set but are not listed. Right-handed leptons and left-handed anti-leptons, neither listed, form weak isospin singlets with $T_3 = 0$ and do not take part in the weak interaction.

349

350 1.2.1.2 Quarks

351 Quarks are the basic constituents of protons and neutrons. The way quarks join to
 352 form bound states, called *hadrons*, is through the SI. Quarks are affected by all the
 353 fundamental interactions which means that they carry all the four types of charges:
 354 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 1.4: Quark properties [9]. Q: electric charge, I_3 : isospin, T_3 : weak isospin, B: baryon number, C: charmness, S: strangeness, T: topness, B' : bottomness, Y: hypercharge. Anti-quarks posses the same mass and spin as quarks but all charges (color, flavor numbers) have opposite sign.

355

356 Table 1.4 summarizes the features of quarks, among which the most remarkable
 357 is their fractional electric charge. Note that fractional charge is not a problem, given
 358 that quarks are not found isolated, but serves to explain how composed particles are
 359 formed out of two or more valence quarks³.

360 Color charge is responsible for the SI between quarks and is the symmetry ($SU(3)_C$)
 361 that defines the formalism to describe SI. There are three colors: red (r), blue (b)
 362 and green (g) and their corresponding three anti-colors; thus each quark carries one
 363 color unit while anti-quarks carries one anti-color unit. As explained in Section 1.2.2,
 364 quarks are not allowed to be isolated due to the color confinement effect, hence, their
 365 features have been studied indirectly by observing their bound states created when

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

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

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

370 In practice, when a quark is left alone isolated a process called *hadronization* occurs
 371 where the quark emits gluons (see Section 1.2.4) which eventually will generate new
 372 quark-antiquark pairs and so on; those quarks will recombine to form hadrons that
 373 will decay into leptons. This proliferation of particles looks like a *jet* coming from
 374 the isolated quark. More details about the hadronization process and jet structure
 375 will be given in chapter3.

376 In the first version of the quark model (1964), M. Gell-Mann [12] and G. Zweig
 377 [13, 14] developed a consistent way to classify hadrons according to their properties.
 378 Only three quarks (u, d, s) were involved in a scheme in which all baryons have baryon
 379 number $B=1$ and therefore quarks have $B=1/3$; non-baryons have $B=0$. Baryon
 380 number is conserved in SI and EI which means that single quarks cannot be created
 381 but in pairs $q - \bar{q}$.

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

$$Q = I_3 + \frac{Y}{2} \quad (1.2)$$

385 where $Y = B + S + C + T + B'$ are the quantum numbers listed in Table 1.4.

386 There are six quark flavors organized in three generations (see Table 1.1) fol-
 387 lowing a mass hierarchy which, again, implies that higher generations decay to first
 388 generation quarks.

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

Table 1.5: Fermion weak isospin and weak hypercharge multiplets. Weak hypercharge is calculated through the Gell-Mann-Nishijima formula 1.2 but using the weak isospin and charge for quarks.

389

390 Isospin doublets of quarks are also defined (see Table 1.5), and same as for neutrinos,
 391 the WI eigenstates are not the same as the mass eigenstates which means that
 392 members of different quark generations are connected by the WI mediator; thus, up-
 393 type quarks are coupled not to down-type quarks (the mass eigenstates) directly but
 394 to a superposition of down-type quarks (q'_d ; *the weak eigenstates*) via WI according
 395 to:

396

$$\begin{aligned} q'_d &= V_{CKM} q_d \\ \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \end{aligned} \quad (1.3)$$

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

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix}. \quad (1.4)$$

399 The weak decays of quarks are represented in the diagram of Figure 1.2; again
 400 the CKM matrix plays a central role since it contains the probabilities for the differ-
 401 ent quark decay channels, in particular, note that quark decays are greatly favored
 402 between generation members.

403 CKM matrix is a 3×3 unitary matrix parametrized by three mixing angles and
 404 the *CP-mixing phase*; the latter is the parameter responsible for the Charge-Parity

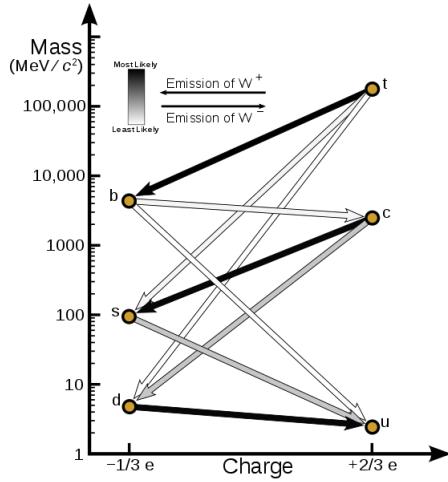


Figure 1.2: Transformations between quarks through the exchange of a WI. Higher generations quarks decay to first generation quarks by emitting a W boson. The arrow color indicates the likelihood of the transition according to the grey scale in the top left side which represent the CKM matrix parameters [19].

405 symmetry violation (CP-violation) in the SM. The fact that the top quark decays
 406 almost all the time to a bottom quark is exploited in this thesis when making the
 407 selection of the signal events by requiring the presence of a jet tagged as a jet coming
 408 from a b quark in the final state.

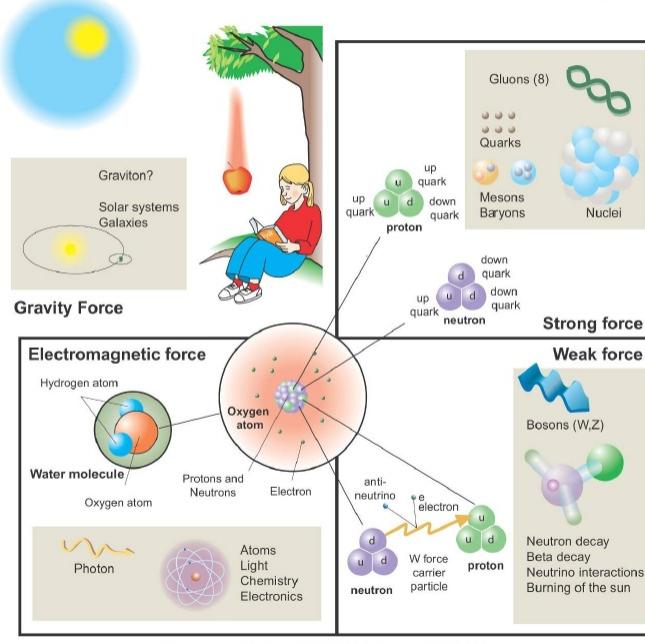
409 1.2.2 Fundamental interactions

410 Even though there are many manifestations of force in nature, like the ones repre-
 411 sented in Figure 1.3, we can classify all of them in four fundamental interactions:

- 412 • *Electromagnetic interaction (EI)* affects particles that are *electrically charged*,
 413 like electrons and protons. Figure 1.4a. shows a graphical representation, known
 414 as *Feynman diagram*, of electron-electron scattering.
- 415 • *Strong interaction (SI)* described by Quantum Chromodynamics (QCD). Hadrons
 416 like the proton and the neutron have internal structure given that they are com-

Fundamental interactions.

Illustration: Typoform



Summer School KPI 15 August 2009

Figure 1.3: 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.

417 posed of two or more valence quarks⁴. Quarks have fractional electric charge
 418 which means that they are subject to electromagnetic interaction and in the case
 419 of the proton they should break apart due to electrostatic repulsion; however,
 420 quarks are held together inside the hadrons against their electrostatic repulsion
 421 by the *Strong Force* through the exchange of *gluons*. The analog to the electric
 422 charge is the *color charge*. Electrons and photons are elementary particles as
 423 quarks but they don't carry color charge, therefore they are not subject to SI. A
 424 Feynman diagram for gluon exchange between quarks is shown in Figure 1.4b.
 425 • *Weak interaction (WI)* described by the weak theory (WT), is responsible, for
 426 instance, for the radioactive decay in atoms and the deuterium production

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

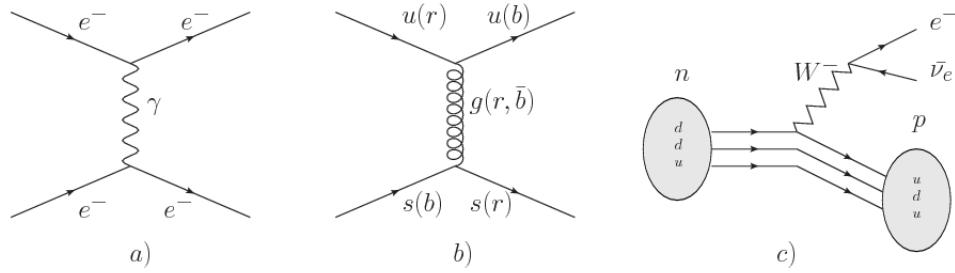


Figure 1.4: Feynman diagrams representing the interactions in SM; a) EI: e - e scattering; b) SI: gluon exchange between quarks ; c) WI: β -decay

within the sun. Quarks and leptons are the particles affected by the weak interaction; they possess a property called *flavor charge* (see 1.2.1) which can be changed by emitting or absorbing one weak force mediator. There are three mediators of the *weak force* known as Z boson in the case of electrically neutral flavor changes and W^\pm bosons in the case of electrically charged flavor changes. The *weak isospin* is the WI analog to electric charge in EI, and color charge in SI, and defines how quarks and leptons are affected by the weak force. Figure 1.4c. shows the Feynman diagram of β -decay where a neutron (n) is transformed in a proton (p) by emitting a W^- particle.

- *Gravitational interaction (GI)* described by General Theory of Relativity (GR). It is responsible for the structure of galaxies and black holes as well as the expansion of the universe. As a classical theory, in the sense that it can be formulated without even appeal to the concept of quantization, it implies that the space-time is a continuum and predictions can be made without limitation to the precision of the measurement tools. The latter represents a direct contradiction of the quantum mechanics principles. Gravity is deterministic while quantum mechanics is probabilistic; despite that, efforts to develop a quantum theory of gravity have predicted the *graviton* as mediator of the gravitational

445 force⁵.

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

Table 1.6: Fundamental interactions features [20].

446

447 Table 1.6 summarizes the main features of the fundamental interactions. The
 448 strength of the interactions is represented by the coupling constants which depend
 449 on the energy scale at which the interaction is evaluated, therefore, it is the relative
 450 strength of the fundamental forces that reveals the meaning of strong and weak; in
 451 a context where the relative strength of the SI is 1, the EI is about hundred times
 452 weaker and WI is about million times weaker than the SI. A good description on how
 453 the relative strength and range of the fundamental interactions are calculated can
 454 be found in References [20, 21]. In the everyday life, only EI and GI are explicitly
 455 experienced due to the range of these interactions; i.e., at the human scale distances
 456 only EI and GI have appreciable effects, in contrast to SI which at distances greater
 457 than 10^{-15} m become negligible. Is it important to clarify that the weakness of the
 458 WI is attributed to the fact that its mediators are highly massive which affects the
 459 propagators of the interaction, as a result, the effect of the coupling constant is
 460 reduced.

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

461 **1.2.3 Gauge invariance.**

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

- 465 • Lorentz invariance: independence on the reference frame.
- 466 • Locality: interacting fields are evaluated at the same space-time point to avoid
 467 action at a distance.
- 468 • Renormalizability: physical predictions are finite and well defined.
- 469 • Particle spectrum, symmetries and conservation laws already known must emerge
 470 from the theory.
- 471 • Local gauge invariance.

472 The gauge invariance requirement reflects the fact that the fundamental fields
 473 cannot be directly measured but associated fields which are the observables. Electric
 474 (**E**) and magnetic (**B**) fields in CED are associated with the electric scalar potential
 475 V and the vector potential **A**. In particular, **E** can be obtained by measuring the
 476 change in the space of the scalar potential (ΔV); however, two scalar potentials
 477 differing by a constant f correspond to the same electric field. The same happens
 478 in the case of the vector potential **A**; thus, different configurations of the associated
 479 fields result in the same set of values of the observables. The freedom in choosing one
 480 particular configuration is known as *gauge freedom*; the transformation law connecting
 481 two configurations is known as *gauge transformation* and the fact that the observables
 482 are not affected by a gauge transformation is called *gauge invariance*.

483 When the gauge transformation:

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} - \Delta f \\ V &\rightarrow V - \frac{\partial f}{\partial t} \end{aligned} \tag{1.5}$$

484 is applied to Maxwell equations, they are still satisfied and the fields remain invariant.
 485 Thus, CED is invariant under gauge transformations and is called a *gauge theory*.
 486 The set of all gauge transformations form the *symmetry group* of the theory, which
 487 according to the group theory, has a set of *group generators*. The number of group
 488 generators determine the number of *gauge fields* of the theory.

489 As mentioned in the first lines of Section 1.2, QED has one symmetry group ($U(1)$)
 490 with one group generator (the Q operator) and one gauge field (the electromagnetic
 491 field A^μ). In CED there is not a clear definition, beyond the historical convention,
 492 of which fields are the fundamental and which are the associated, but in QED the
 493 fundamental field is A^μ . When a gauge theory is quantized, the gauge fields are
 494 quantized and their quanta are called *gauge bosons*. The word boson characterizes
 495 particles with integer spin which obey Bose-Einstein statistics.

496 As will be detailed in Section 1.3, interactions between particles in a system can
 497 be obtained by considering first the Lagrangian density of free particles in the sys-
 498 tem, which of course is incomplete because the interaction terms have been left out,
 499 and demanding global phase transformation invariance. Global phase transforma-
 500 tion means that a gauge transformation is performed identically to every point
 501 in the space⁶ and the Lagrangian remains invariant. Then, the global transforma-
 502 tion is promoted to a local phase transformation (this time the gauge transforma-
 503 tion depends on the position in space) and again invariance is required.

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

504 Due to the space dependence of the local transformation, the Lagrangian density is
 505 not invariant anymore. In order to reinstate the gauge invariance, the gauge covariant
 506 derivative is introduced in the Lagrangian and with it the gauge field responsible for
 507 the interaction between particles in the system. The new Lagrangian density is gauge
 508 invariant, includes the interaction terms needed to account for the interactions and
 509 provides a way to explain the interaction between particles through the exchange of
 510 the gauge boson.

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

513 1.2.4 Gauge bosons

514 The importance of the gauge bosons comes from the fact that they are the force
 515 mediators or force carriers. The features of the gauge bosons reflect those of the fields
 516 they represent and they are extracted from the Lagrangian density used to describe
 517 the interactions. In Section 1.3, it will be shown how the gauge bosons of the EI and
 518 WI emerge from the electroweak Lagrangian. The SI gauge bosons features are also
 519 extracted from the SI Lagrangian but it is not detailed in this document. The main
 520 features of the SM gauge bosons will be briefly presented below and summarized in
 521 Table 1.7.

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

- 528 • **Gluon.** SI is mediated by gluons which just as photons are massless. They
 529 carry one unit of color charge and one unit of anticolor charge, hence, gluons
 530 can couple to other gluons. As a result, the range of the SI is not infinite
 531 but very short due to the attraction between gluons, giving rise to the *color*
 532 *confinement* which explains why color charged particles cannot be isolated but
 533 live within composite particles, like quarks inside protons.
- 534 • **W, Z.** W^\pm and Z, are massive which explains their short-range. Given that
 535 the WI is the only interaction that can change the flavor of the interacting
 536 particles, the W boson is the responsible for the nuclear transmutation where
 537 a neutron is converted into a proton or vice versa with the involvement of an
 538 electron and a neutrino (see Figure 1.4c). The Z boson is the responsible for the
 539 neutral weak processes like neutrino elastic scattering where no electric charge
 540 but momentum transference is involved. WI gauge bosons carry isospin charge
 541 which makes interaction between them possible.

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

Table 1.7: SM gauge bosons main features [9].

542

543

1.3 Electroweak unification and the Higgs 544 mechanism

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

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

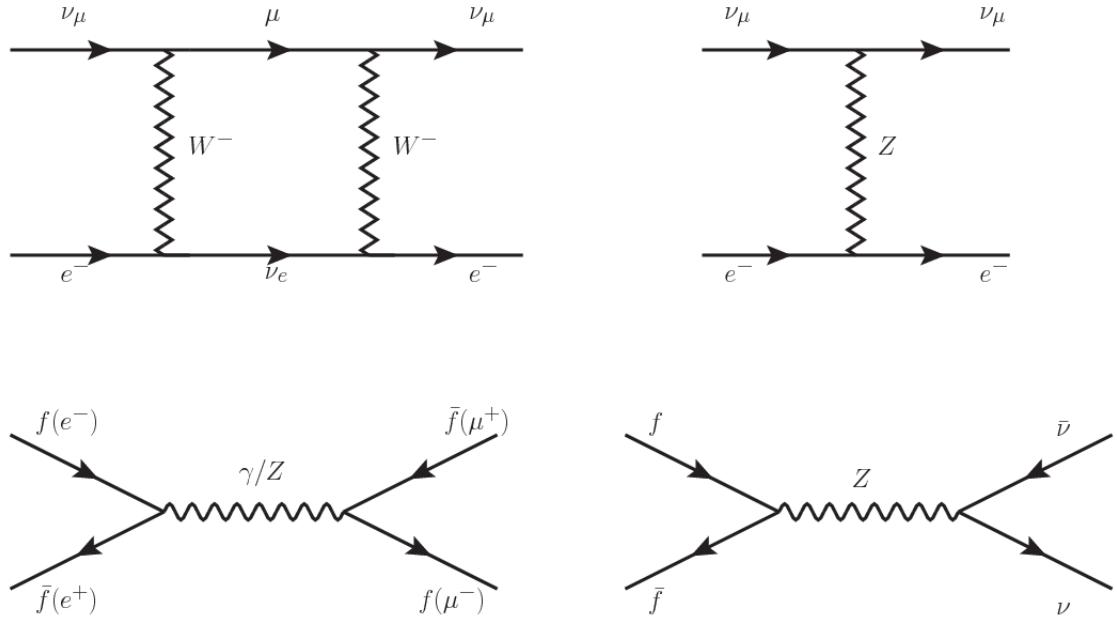


Figure 1.5: Top: $\nu_\mu - e^-$ scattering going through charged currents (left) and neutral currents (right). Bottom: neutral current processes for charged fermions (left) and involving neutrinos (right). While neutral current processes involving only charged fermions can proceed through EI or WI, those involving neutrinos can only proceed via WI.

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

559 shell. When going up in energy, the theory of weak interactions involving the W
 560 boson is capable of explaining the β -decay and in general the processes mediated by
 561 W^\pm bosons. However, there were some processes like the $\nu_\mu - e$ scattering which
 562 would require the exchange of two W bosons (see Figure 1.5 top diagrams) giving
 563 rise to divergent loop integrals and then non-finite predictions. The EWI theory, by
 564 including neutral currents involving fermions via the exchange of a neutral bosons Z,
 565 overcomes those divergences and the predictions become realistic.

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

570 The prescription to build a gauge theory of the WI consists of proposing a free
 571 field Lagrangian density that includes the particles involved; next, by requesting
 572 invariance under global phase transformations first and generalizing to local phase
 573 transformations invariance later, the conserved currents are identified and interactions
 574 are generated by introducing gauge fields. Given that the goal is to include the EI
 575 and WI in a single theory, the group symmetry considered should be a combination of
 576 $SU(2)_L$ and $U(1)_{em}$, however the latter cannot be used directly because the EI treats
 577 left and right-handed particles indistinctly in contrast to the former. Fortunately, the
 578 weak hypercharge, which is a combination of the weak isospin and the electric charge
 579 (Eqn. 1.2) is suitable to be used since it is conserved by the EI and WI. Thus, the
 580 symmetry group to be considered is

$$G \equiv SU(2)_L \otimes U(1)_Y \quad (1.6)$$

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

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

583 Given the first generation of leptons

$$\psi_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad \psi_2 = \nu_{eR}, \quad \psi_3 = e_R^- \quad (1.7)$$

584 the charged fermionic currents are given by

$$J_\mu \equiv J_\mu^+ = \bar{\nu}_{eL} \gamma_\mu e_L, \quad J_\mu^\dagger \equiv J_\mu^- = \bar{e}_L \gamma_\mu \nu_{eL} \quad (1.8)$$

585 and the free Lagrangian is given by

$$\mathcal{L}_0 = \sum_{j=1}^3 i\bar{\psi}_j(x) \gamma^\mu \partial_\mu \psi_j(x). \quad (1.9)$$

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

$$m_W^2 W_\mu^\dagger(x) W^\mu(x) + \frac{1}{2} m_Z^2 Z_\mu(x) Z^\mu(x) - m_e \bar{\psi}_e(x) \psi_e(x) \quad (1.10)$$

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

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

$$\begin{aligned}
\psi_1(x) &\xrightarrow{G} \psi'_1(x) \equiv U_Y U_L \psi_1(x), \\
\psi_2(x) &\xrightarrow{G} \psi'_2(x) \equiv U_Y \psi_2(x), \\
\psi_3(x) &\xrightarrow{G} \psi'_3(x) \equiv U_Y \psi_3(x)
\end{aligned} \tag{1.11}$$

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

$$U_L \equiv \exp\left(i \frac{\sigma_i}{2} \alpha^i\right), \quad U_Y \equiv \exp(i y_i \beta) \quad (i = 1, 2, 3) \tag{1.12}$$

599 with σ_i the Pauli matrices and y_i the weak hypercharges. In order to promote the
 600 transformations from global to local while keeping the invariance, it is required that
 601 $\alpha^i = \alpha^i(x)$, $\beta = \beta(x)$ and the replacement of the ordinary derivatives by the covariant
 602 derivatives

$$\begin{aligned}
D_\mu \psi_1(x) &\equiv \left[\partial_\mu + ig\sigma_i W_\mu^i(x)/2 + ig'y_1 B_\mu(x) \right] \psi_1(x) \\
D_\mu \psi_2(x) &\equiv \left[\partial_\mu + ig'y_2 B_\mu(x) \right] \psi_2(x) \\
D_\mu \psi_3(x) &\equiv \left[\partial_\mu + ig'y_3 B_\mu(x) \right] \psi_3(x)
\end{aligned} \tag{1.13}$$

603 introducing in this way four gauge fields, $W_\mu^i(x)$ and $B_\mu(x)$, in the process. The
 604 covariant derivatives (Eqn. 1.13) are required to transform in the same way as fermion
 605 fields $\psi_i(x)$ themselves, therefore, the gauge fields transform as:

$$\begin{aligned} B_\mu(x) &\xrightarrow{G} B'_\mu(x) \equiv B_\mu(x) - \frac{1}{g'} \partial_\mu \beta(x) \\ W_\mu^i(x) &\xrightarrow{G} W_\mu^{i\prime}(x) \equiv W_\mu^i(x) - \frac{i}{g} \partial_\mu \alpha_i(x) - \varepsilon_{ijk} \alpha_i(x) W_\mu^j(x). \end{aligned} \quad (1.14)$$

606 The G invariant version of the Lagrangian density 1.9 can be written as

$$\mathcal{L}_0 = \sum_{j=1}^3 i \bar{\psi}_j(x) \gamma^\mu D_\mu \psi_j(x) \quad (1.15)$$

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

$$B_{\mu\nu}(x) \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.16)$$

$$W_{\mu\nu}^i(x) \equiv \partial_\mu W_\nu^i(x) - \partial_\nu W_\mu^i(x) - g \varepsilon^{ijk} W_\mu^j W_\nu^k \quad (1.17)$$

610 the last term in Eqn. 1.17 is added in order to hold the gauge invariance; therefore,

$$\mathcal{L}_G = -\frac{1}{4} B_{\mu\nu}(x) B^{\mu\nu}(x) - \frac{1}{4} W_{\mu\nu}^i(x) W_i^{\mu\nu}(x) \quad (1.18)$$

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

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

$$J_\mu^i(x) = \frac{1}{2} \bar{\psi}_1(x) \gamma_\mu \sigma^i \psi_1(x) \quad (1.19)$$

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

$$J_\mu^Y = \sum_{j=1}^3 \bar{\psi}_j(x) \gamma_\mu y_j \psi_j(x) \quad (1.20)$$

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

$$\mathcal{L}_I = -g J^{i\mu}(x) W_\mu^i(x) - \frac{g'}{2} J^{Y\mu}(x) B_\mu(x) \quad (1.21)$$

621 Note that the weak isospin currents are not the same as the charged fermionic cur-
 622 rents that were used to describe the WI (Eqn. 1.8), since the weak isospin eigenstates
 623 are not the same as the mass eigenstates, but they are closely related

$$J_\mu = \frac{1}{2}(J_\mu^1 + iJ_\mu^2), \quad J_\mu^\dagger = \frac{1}{2}(J_\mu^1 - iJ_\mu^2). \quad (1.22)$$

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

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2), \quad W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2). \quad (1.23)$$

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

631 current (j_μ^{em}), indicating that there is a relation between them and resembling the
 632 Gell-Mann-Nishijima formula 1.2 adapted to electroweak interactions

$$Q = T_3 + \frac{Y_W}{2}. \quad (1.24)$$

633 Just as Q generates the $U(1)_{em}$ symmetry, the weak hypercharge generates the
 634 $U(1)_Y$ symmetry as said before. It is possible to write the relationship in terms of
 635 the currents as

$$j_\mu^{em} = J_\mu^3 + \frac{1}{2} J_\mu^Y. \quad (1.25)$$

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

$$\begin{aligned} A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \end{aligned} \quad (1.26)$$

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

$$\begin{aligned} \mathcal{L}_I = -\frac{g}{\sqrt{2}}(J^\mu W_\mu^+ + J^{\mu\dagger} W_\mu^-) - &\left(g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{J_\mu^Y}{2}\right) A^\mu \\ &- \left(g \cos \theta_W J_\mu^3 - g' \sin \theta_W \frac{J_\mu^Y}{2}\right) Z^\mu \end{aligned} \quad (1.27)$$

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

640 electromagnetic interaction under the condition

$$g \sin \theta_W = g' \cos \theta_W = e, \quad \frac{g'}{g} = \tan \theta_W \quad (1.28)$$

641 contained in the Eqn.1.25; the third term is the neutral weak current.

642

643 Note that the neutral fields transformation given by the Eqn. 1.26 can be written
 644 in terms of the coupling constants g and g' as:

$$A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + g'^2}}, \quad Z_\mu = \frac{g W_\mu^3 - g' B_\mu}{\sqrt{g^2 + g'^2}}. \quad (1.29)$$

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

$$\mathcal{L}_{nmEWI} = \mathcal{L}_0 + \mathcal{L}_G \quad (1.30)$$

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

651 1.3.1 Spontaneous symmetry breaking (SSB)

652 Figure 1.6 left shows a steel nail (top) which is subject to an external force; the form
 653 of the potential energy is also shown (bottom).

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

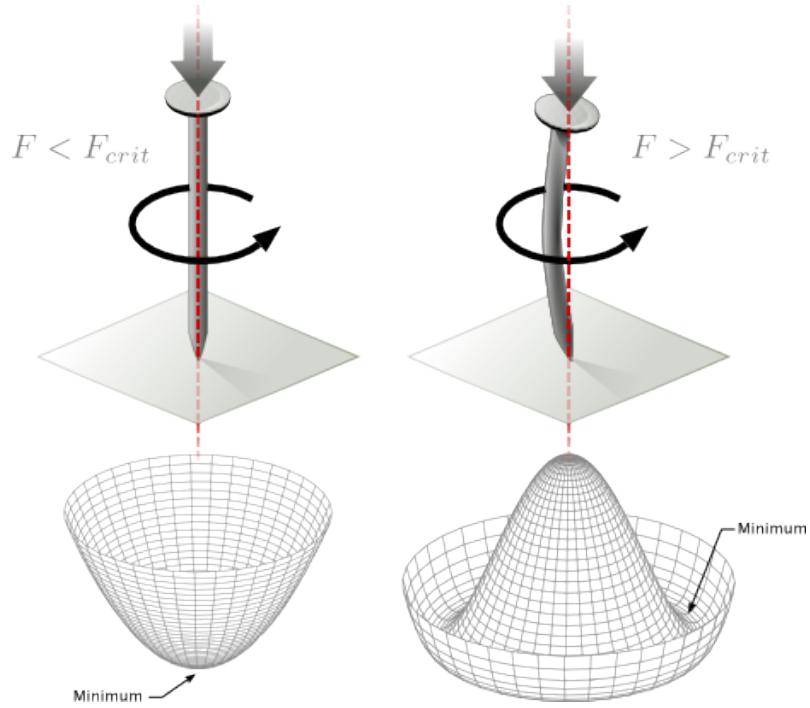


Figure 1.6: Spontaneous symmetry breaking mechanism. The steel nail, subject to an external force (top left), has rotational symmetry with respect to its axis. When the external force overcomes a critical value the nail buckles (top right) choosing a minimal energy state (ground state) and thus *breaking spontaneously the rotational symmetry*. The potential energy (bottom) changes but holds the rotational symmetry; however, an infinite number of asymmetric ground states are generated and circularly distributed in the bottom of the potential [31].

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

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

665 Figure 1.7 shows a plot of the potential $V(\phi)$ in the case of a scalar field ϕ

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.31)$$

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

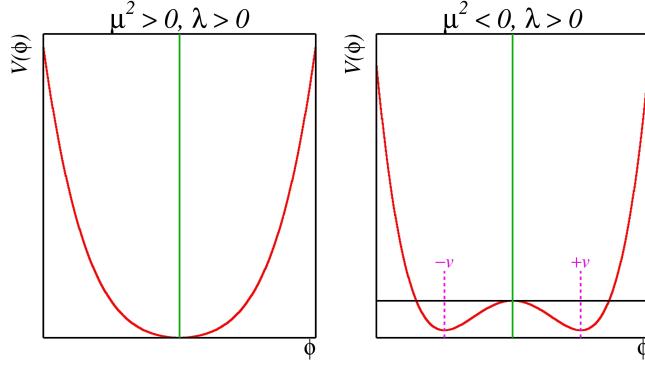


Figure 1.7: Shape of the potential $V(\phi)$ for $\lambda > 0$ and: $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The case $\mu^2 < 0$ corresponds to the potential suitable for introducing the SSB mechanism by choosing one of the two ground states which are connected via reflection symmetry. [31].

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

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (1.32)$$

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

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.33)$$

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

672 As seen in Figure 1.8, the potential has now an infinite number of minima circularly
 673 distributed along the ξ -direction which makes possible the occurrence of the SSB by
 674 choosing an arbitrary ground state; for instance, $\xi = 0$, i.e. $\phi_1 = v, \phi_2 = 0$

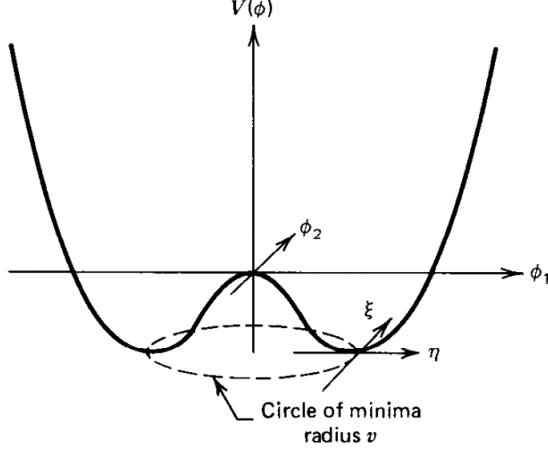


Figure 1.8: Potential for complex scalar field. There is a circle of minima of radius v along the ξ -direction [6].

$$\phi_0 = \frac{v}{\sqrt{2}} \exp(i\xi) \xrightarrow{\text{SSB}} \phi_0 = \frac{v}{\sqrt{2}} \quad (1.34)$$

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

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)) \quad (1.35)$$

677 which when substituted into Eqn. 1.33 produces a Lagrangian in terms of the new
 678 fields η and ξ

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu^2 \eta^2 - V(\phi_0) - \lambda v \eta (\eta^2 + \xi^2) - \frac{\lambda}{4}(\eta^2 + \xi^2)^2 \quad (1.36)$$

679 where the last two terms represent the interactions and self-interaction between the
 680 two fields η and ξ . The particular feature of the SSB mechanism is revealed when
 681 looking to the first three terms of \mathcal{L}' . Before the SSB, only the massless ϕ field is

682 present in the system; after the SSB there are two fields of which the η -field has
 683 acquired mass $m_\eta = \sqrt{-2\mu^2}$ while the ξ -field is still massless (see Figure 1.9).

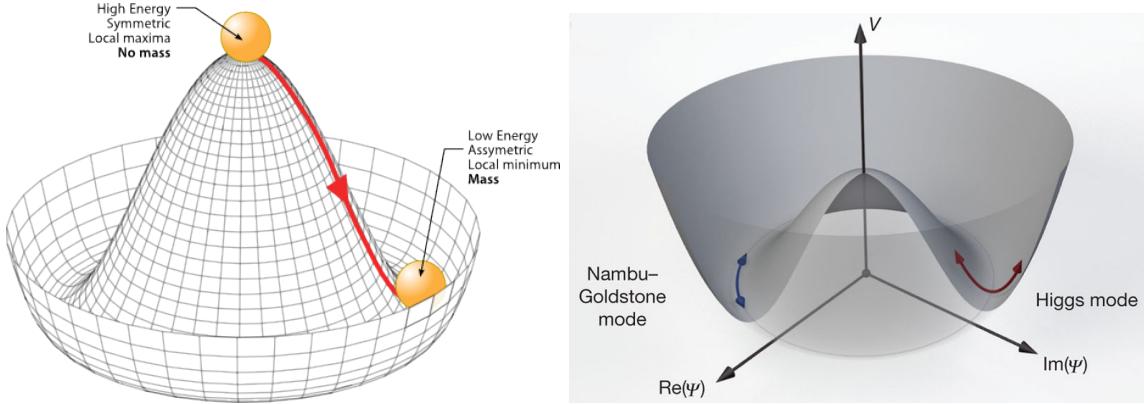


Figure 1.9: SSB mechanism for a complex scalar field [31, 32].

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

694 1.3.2 Higgs mechanism

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

698 the mass of the EW gauge bosons, a G invariant Lagrangian density (\mathcal{L}_S) has to be
 699 added to the non massive EWI Lagrangian (Eqn. 1.30)

$$\mathcal{L}_S = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0, \mu^2 < 0 \quad (1.37)$$

$$D_\mu \phi = \left(i\partial_\mu - g \frac{\sigma_i}{2} W_\mu^i - g' \frac{Y}{2} B_\mu \right) \phi \quad (1.38)$$

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

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.39)$$

702 The minima of the potential are defined by

$$\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}. \quad (1.40)$$

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

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 \equiv -\frac{\mu^2}{\lambda}. \quad (1.41)$$

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

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

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

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{i}{v} \sigma_i \theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \approx \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1(x) + i\theta_2(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \quad (1.42)$$

712 to describe fluctuations from the ground state ϕ_0 . The fields $\theta_i(x)$ represent the
 713 Nambu-Goldstone bosons while $H(x)$ is known as *Higgs field*. The fundamental fea-
 714 ture of the parametrization used is that the dependence on the $\theta_i(x)$ fields is factored
 715 out in a global phase that can be eliminated by taking the physical *unitary gauge*
 716 $\theta_i(x) = 0$. Therefore the expansion about the ground state is given by:

$$\phi(x) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.43)$$

717 which when substituted into \mathcal{L}_S (Eqn. 1.37) results in a Lagrangian containing the
 718 now massive three gauge bosons W^\pm, Z , one massless gauge boson (photon) and
 719 the new Higgs field (H). The three degrees of freedom corresponding to the Nambu-
 720 Goldstone bosons are now integrated into the massive gauge bosons as their lon-
 721 gitudinal polarizations which were not available when they were massless particles.
 722 The effect by which vector boson fields acquire mass after an spontaneous symmetry
 723 breaking, but without an explicit gauge invariance breaking is known as the *Higgs*
 724 *mechanism*.

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

731 **1.3.3 Masses of the gauge bosons**

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

$$\left| \left(\partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - i \frac{g'}{2} B_\mu \right) \phi_0 \right|^2 = \left(\frac{1}{2} v g \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (W_\mu^3, B_\mu) \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} \quad (1.44)$$

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

$$M_W = \frac{1}{2} v g. \quad (1.45)$$

The second term in the right side of the Eqn.1.44 comprises the masses of the neutral bosons, but it needs to be written in terms of the gauge fields Z_μ and A_μ in order to be compared to the typical mass terms for neutral bosons, therefore using Eqn. 1.29

$$\begin{aligned} \frac{1}{8} v^2 [g^2 (W_\mu^3)^2 - 2gg' W_\mu^3 B^\mu + g'^2 B_\mu^2] &= \frac{1}{8} v^2 [g W_\mu^3 - g' B_\mu]^2 + 0[g' W_\mu^3 + g B_\mu]^2 \quad (1.46) \\ &= \frac{1}{8} v^2 [\sqrt{g^2 + g'^2} Z_\mu]^2 + 0[\sqrt{g^2 + g'^2} A_\mu]^2 \end{aligned}$$

735 and then

$$M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}, \quad M_A = 0 \quad (1.47)$$

736 **1.3.4 Masses of the fermions**

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

$$\mathcal{L}_{Yl} = -G_l \left[(\bar{\nu}_l, \bar{l})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} l_R + \bar{l}_R (\phi^-, \bar{\phi}^0) \begin{pmatrix} \nu_l \\ l \end{pmatrix} \right], \quad l = e, \mu, \tau. \quad (1.48)$$

739 After the SSB and replacing the usual field expansion about the ground state
 740 (Eqn.1.41) into \mathcal{L}_{Yl} , the mass term arises

$$\mathcal{L}_{Yl} = -m_l (\bar{l}_L l_R + \bar{l}_R l_L) - \frac{m_l}{v} (\bar{l}_L l_R + \bar{l}_R l_L) H = -m_l \bar{l} l \left(1 + \frac{H}{v} \right) \quad (1.49)$$

741

$$m_l = \frac{G_l}{\sqrt{2}} v \quad (1.50)$$

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

$$\phi_c = -i\sigma_2 \phi^* = \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix}. \quad (1.51)$$

745 Additionally, given that the quark isospin doublets are not constructed in terms
 746 of the mass eigenstates but in terms of the flavor eigenstates, as shown in Table 1.5,
 747 the coupling parameters will be related to the CKM matrix elements; thus, the quark
 748 Lagrangian is given by:

$$\mathcal{L}_{Yq} = -G_d^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{jR} - G_u^{i,j} (\bar{u}_i, \bar{d}'_i)_L \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} u_{jR} + h.c. \quad (1.52)$$

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

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

$$\mathcal{L}_{Yq} = -m_d^i \bar{d}_i d_i \left(1 + \frac{H}{v}\right) - m_u^i \bar{u}_i u_i \left(1 + \frac{H}{v}\right) \quad (1.53)$$

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

753 1.3.5 The Higgs field

754 After the characterization of the fermions and gauge bosons as well as their interac-
755 tions, it is necessary to characterize the Higgs field itself. The Lagrangian \mathcal{L}_S in Eqn.
756 1.37 written in terms of the gauge bosons is given by

$$\mathcal{L}_S = \frac{1}{4} \lambda v^4 + \mathcal{L}_H + \mathcal{L}_{HV} \quad (1.54)$$

757

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} m_H^2 H^2 - \frac{1}{2v} m_H^2 H^3 - \frac{1}{8v^2} m_H^2 H^4 \quad (1.55)$$

758

$$\mathcal{L}_{HV} = m_H^2 W_\mu^+ W^{\mu-} \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \left(1 + \frac{2}{v} H + \frac{2}{v^2} H^2\right) \quad (1.56)$$

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

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v \quad (1.57)$$

761 however, it is not predicted by the theory either. The experimental measurement of
762 the Higgs boson mass have been performed by the *Compact Muon Solenoid (CMS)*
763 experiment and the *A Toroidal LHC Appartus (ATLAS)* experiments at the *Large
764 Hadron Collider (LHC)*, [36–38], and is presented in Table 1.8.

765

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

Table 1.8: Higgs boson properties. Higgs mass is not predicted by the theory and the value here corresponds to the experimental measurement.

766 1.3.6 Production of Higgs bosons at LHC

767 At the LHC, Higgs bosons are produced as a result of the collision of two counter-
 768 rotating protons beams. A detailed description of the LHC machine will be presented
 in chapter 2.

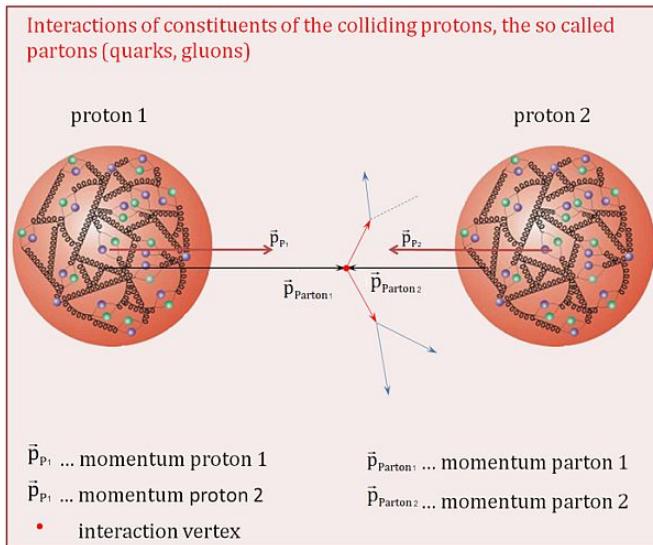


Figure 1.10: Proton-proton collision. Protons are composed of 3 valence quarks, a sea of quarks and gluons; therefore in a proton-proton collision, quarks and gluons are those who collide. [39].

769

770 Protons are composed of quarks and these quarks are bound by gluons; however,
771 what is commonly called the quark content of the proton makes reference to the
772 valence quarks. In fact, a proton is not just a rigid entity with three balls in it all
773 tied up with springs, but the gluons exchanged by the valence quarks tend to split

774 spontaneously into quark-antiquark pairs or more gluons, creating *sea of quarks and*
 775 *gluons* as represented in Figure 1.10.

776 In a proton-proton (pp) collision, the proton's constituents, quarks and gluons, are
 777 those that collide. The pp cross section depends on the momentum of the colliding
 778 particles, reason for which it is needed to know how the momentum is distributed
 779 inside the proton. Quarks and gluons are known as partons, hence, the functions
 780 that describe how the proton momentum is distributed among partons inside it are
 781 called *parton distribution functions (PDFs)*; PDFs are determined from experimental
 782 data obtained in experiments where the internal structure of hadrons is tested, and
 783 depend on the momentum transfer Q and the fraction of momentum x carried by an
 784 specific parton. Figure 1.11 shows the proton PDFs ($xf(x, Q^2)$) for two values of Q .

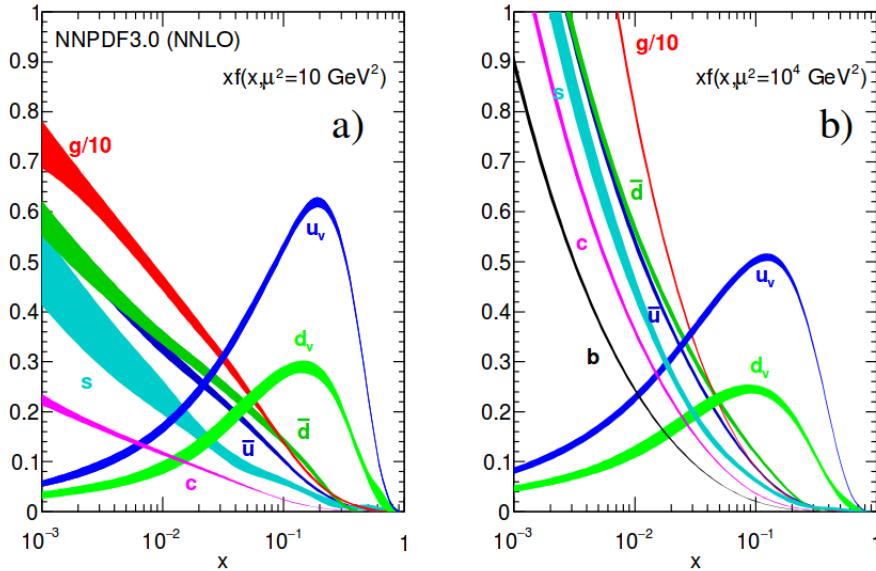


Figure 1.11: Proton PDFs for two values of Q^2 : left. $\mu^2 = Q^2 = 10 \text{ GeV}^2$, right. $\mu^2 = Q^2 = 10^4 \text{ GeV}^2$. u_v and d_v correspond to the u and d valence quarks, $s, c, b, \bar{u}, \bar{d}$ correspond to sea quarks, and g corresponds to gluons. Note that gluons carry a high fraction of the proton's momentum. [9]

785 In physics, a common approach to study complex systems consists of starting
 786 with a simpler version of them, for which a well known description is available, and

adding 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. The perturbation series corresponds to an expansion in power series of a small parameter, therefore, the more terms are considered in the series (the higher order in the perturbation series), the more precise is the description of the complex system. If the perturbation does not get progressively smaller, the strategy cannot be applied and new methods have to be employed.

High energy systems, like the Higgs production at LHC explored in this thesis, usually can be treated perturbatively with the expansion made in terms of the coupling constants. The overview presented here will be oriented specifically to the Higgs boson production mechanisms in pp collisions at LHC.

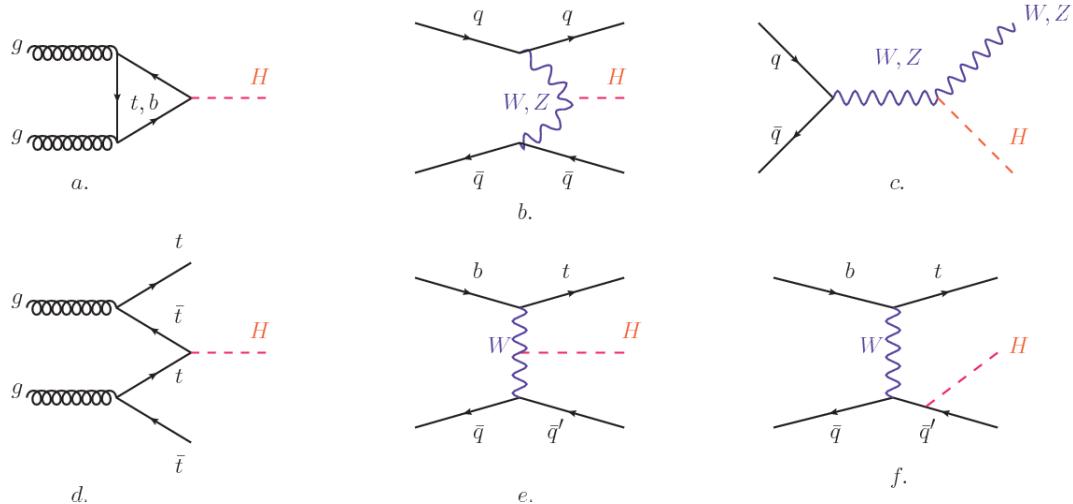


Figure 1.12: Main Higgs boson production mechanism Feynman diagrams. a. gluon-gluon fusion, b. vector boson fusion (VBF), c. Higgs-strahlung, d. Associated production with a top or bottom quark pair, e-f. associated production with a single top quark.

Figure 1.12 shows the Feynman diagrams for the leading order (first order) Higgs production processes at LHC; note that in these diagrams the incoming particles are not the protons themselves but the partons from the protons that actually participate

in the interaction, hence, theorists typically calculate the cross section for the parton interaction, and then convolute that cross section with the information from the PDFs to get a production cross section that is actually measured in experiments. The cross section for Higgs production as a function of the center of mass-energy (\sqrt{s}) for pp collisions is showed in Figure 1.13 left. The tags NLO (next to leading order), NNLO (next to next to leading order) and N3LO (next to next to next to leading order) make reference to the order at which the perturbation series have been considered while the tags QCD and EW correspond to the strong and electroweak coupling constants respectively.

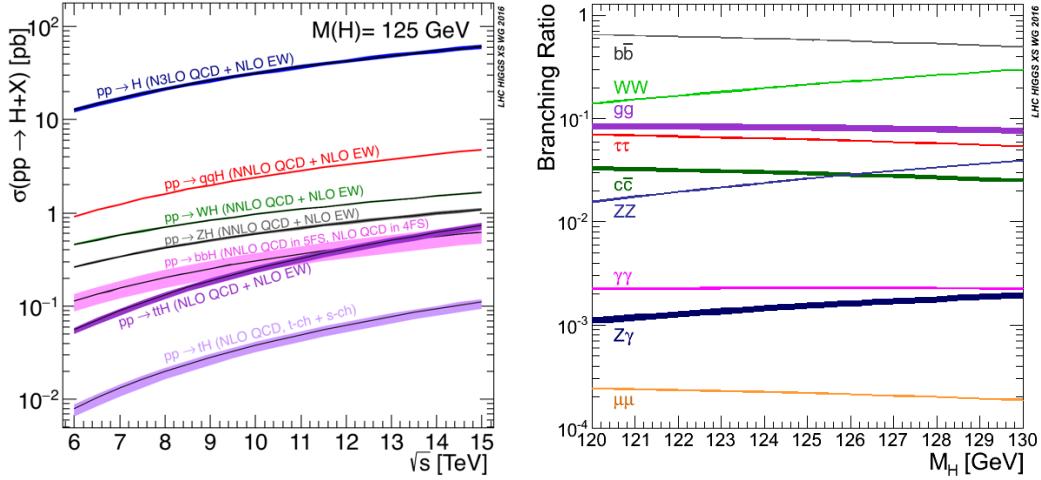


Figure 1.13: Higgs boson production cross sections (left) and decay branching ratios (right) for the main mechanisms. The VBF is indicated as qqH [40].

The main production mechanism is the gluon fusion (Figure 1.12a and $pp \rightarrow H$ in Figure 1.13) given that gluons carry the highest fraction of momentum of the protons in pp colliders (as shown in Figure ??). Since the Higgs boson does not couple to gluons, the mechanism proceeds through the exchange of a virtual top-quark loop. Note that in this process the Higgs boson is produced alone, turning out to be problematic for some Higgs decays, because such absence of anything produced in

817 association with the Higgs represent a trouble for triggering, however, this mechanism
 818 is experimentally clean when combined with the two-photon or the four-lepton decay
 819 channels (see Section 1.3.7).

820 Vector boson fusion (Figure 1.12b and $pp \rightarrow qqH$ in Figure 1.13) has the second
 821 largest production cross section. The scattering of two fermions is mediated by a weak
 822 gauge boson which later emits a Higgs boson. In the final state, the two fermions tend
 823 to be located in the central region of the detector; this kind of features are generally
 824 used as a signature when analyzing the datasets provided by the experiments⁷.

825 In the Higgs-strahlung mechanism (Figure 1.12c and $pp \rightarrow WH, pp \rightarrow ZH$ in
 826 Figure 1.13) two fermions annihilate to form a weak gauge boson. If the initial
 827 fermions have enough energy, the emergent boson might emit a Higgs boson.

828 The associated production with a top or bottom quark pair and the associated
 829 production with a single top quark (Figure 1.12d-f and $pp \rightarrow bbH, pp \rightarrow t\bar{t}H, pp \rightarrow tH$
 830 in Figure 1.13) have a smaller cross section than the main three mechanisms above,
 831 but they provide a good opportunity to test the Higgs-top coupling. The analysis
 832 reported in this thesis is developed using these production mechanisms. A detailed
 833 description of the tH mechanism will be given in Section 1.5.

834 1.3.7 Higgs boson decay channels

835 When a particle can decay through several modes, also known as channels, the prob-
 836 ability of decaying through a given channel is quantified by the *branching ratio (BR)*
 837 of the decay channel; thus, the BR is defined as the ratio of number of decays go-
 838 ing through that given channel to the total number of decays. In regard to the
 839 Higgs boson decay, the BR can be predicted with accuracy once the Higgs mass is
 840 known [41, 42]. In Figure 1.13 right, a plot of the BR as a function of the Higgs mass

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

841 is presented; the largest predicted BR corresponds to the $b\bar{b}$ pair decay channel (see
 842 Table 1.9) given that it is the heaviest particle pair whose on-shell⁸ production is
 843 kinematically allowed in the decay.

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

844

845 Decays to other lepton and quark pairs, like electron, strange, up, and down
 846 quark pairs not listed in the table, are also possible but their likelihood is too small
 847 to measure since they are very lightweight, hence, their interaction with the Higgs
 848 boson is very weak. On other hand, the decay to top quark pairs is heavily suppressed
 849 due to the top quark mass ($\approx 173\text{ GeV}/c^2$).

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

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

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

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

861 **1.4 Experimental status of the anomalous
 862 Higgs-fermion coupling**

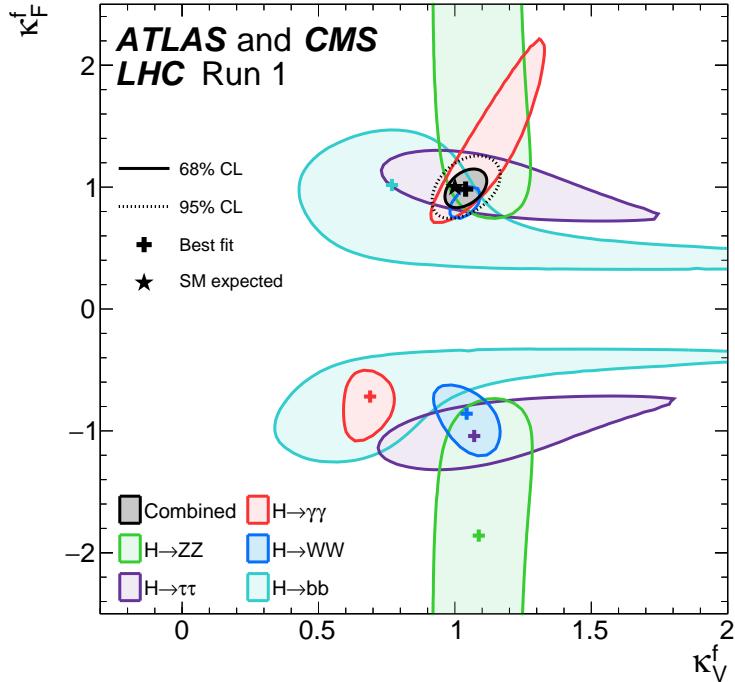


Figure 1.14: Combination of the ATLAS and CMS fits for coupling modifiers $\kappa_t - \kappa_V$; also shown the individual decay channels combination and their global combination. No assumptions have been made on the sign of the coupling modifiers [44].

863 ATLAS and CMS have performed analyses of the anomalous Higgs-fermion cou-
 864 pling by making likelihood scans for the two coupling modifiers, κ_f and κ_V , under
 865 the assumption that $\kappa_Z = \kappa_W \equiv \kappa_V$ and $\kappa_t = \kappa_\tau = \kappa_b \equiv \kappa_f$. Figure 1.14 shows the
 866 result of the combination of ATLAS and CMS fits; also the individual decay channels
 867 combination and the global combination results are shown. Note that from this plot
 868 there is limited information on the sign of the coupling since the only information
 869 available about the sign of the coupling comes from decays rather than production.

870 While all the channels are compatible for positive values of the modifiers, for
 871 negative values of κ_f there is no compatibility. The best fit for individual channels
 872 is compatible with negative values of κ_f except for the $H \rightarrow bb$ channel. The best
 873 fit for the combination yields $\kappa_f \geq 0$, in contrast to the yields from the individual
 874 channels; the reason of this yield resides in the $H \rightarrow \gamma\gamma$ coupling. $H \rightarrow \gamma\gamma$ decay
 875 proceeds through a loop of either top quarks or W bosons, hence, this channel is
 876 sensitive to κ_t thanks to the interference of these two amplitude contributions; under
 877 the assumption that no beyond SM particles take part in the loops, a flipped sign
 878 of κ_t will increase the $H \rightarrow \gamma\gamma$ branching fraction by a factor of ~ 2.4 which is not
 879 supported by measurements; thus, this large asymmetry between the positive and
 880 negative coupling ratios in the $H \rightarrow \gamma\gamma$ channel drives the yield of the global fit and
 881 would mean that the anomalous H-t coupling is excluded as stated in Reference [44],
 882 but there is a caveat, this exclusion holds only if no new particles contribute to the
 883 loop in the main diagram for that decay.

884 Although the $H \rightarrow bb$ channel is expected to be the most sensitive channel and
 885 its best fit value of κ_t is positive, and then the global fit yield is still supported,
 886 the contributions from all the other decay channels, small compared to the $H \rightarrow bb$,
 887 indicate that the anomalous H-t coupling cannot be excluded completely, motivating
 888 to look at tH processes which can help with both, the limited information on the sign

889 of the H-t coupling and the access to information from the Higgs boson production
 890 rather than from its decays.

891 It will be shown in Section 1.5 that the same interference effect enhance the
 892 tH production rate and could reveal evidence of direct production of heavy new par-
 893 ticles as predicted in composite and little Higgs models [45], or new physics related
 894 to Higgs boson mediated flavor changing neutral currents [46] as well as probes the
 895 CP-violating phase of the H-t coupling [47, 48].

896 **1.5 Associated production of a Higgs boson and a 897 single top quark**

898 The production of Higgs boson in association with a top quark has been extensively
 899 studied [47, 49–52]. While measurements of the main Higgs production mechanisms
 900 rates are sensitive to the strength of the Higgs coupling to W boson or top quark,
 901 they are not sensitive to the relative sign between the two couplings. In this thesis,
 902 the Higgs boson production mechanism explored is the associated production with a
 903 single top quark (tH) which offers sensitivity to the relative sign of the Higgs couplings
 904 to W boson and to top quark. The description given here is based on Reference [51]

A process where two incoming particles interact and produce a final state with two
 particles can proceed in three called channels (see, for instance, Figure 1.15 omitting
 the red line). The t-channel represents processes where an intermediate particle is
 emitted by one of the incoming particles and absorbed by the other. The s-channel
 represents processes where the two incoming particles merge into an intermediate par-
 ticle which eventually will split into the particles in the final state. The third channel,
 u-channel, is similar to the t-channel but the two outgoing particles interchange their

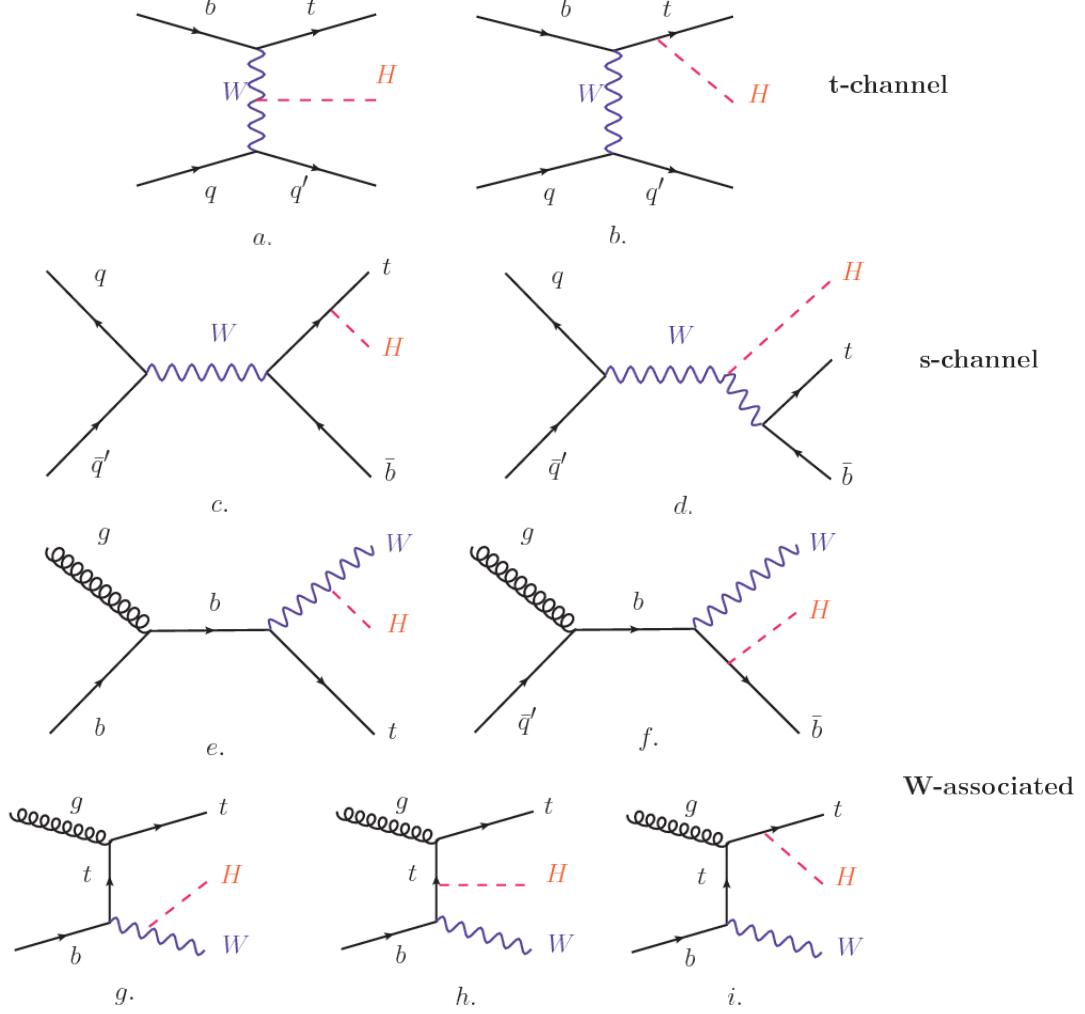


Figure 1.15: Associated Higgs boson production with a top quark mechanism Feynman diagrams. a.,b. t-channel (tHq), c.,d. s-channel (tHb), e-i. W-associated.

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

$$s = (p_1 + p_2)^2 = (p'_1 + p'_2)^2 \rightarrow \text{square of the center mass-energy.} \quad (1.58)$$

$$t = (p_1 - p'_1)^2 = (p'_2 - p_2)^2 \rightarrow \text{square of the four-momentum transfer.} \quad (1.59)$$

$$u = (p_1 - p'_2)^2 = (p'_1 - p_2)^2 \rightarrow \text{square of the crossed four-momentum transfer.} \quad (1.60)$$

$$s + t + u = m_1^2 + m_2^2 + m'_1^2 + m'_2^2 \quad (1.61)$$

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

The tH production, where Higgs boson can be radiated either from the top quark or from the W boson, is represented by the leading order Feynman diagrams in Figure 1.15. The cross section for the tH process is calculated, as usual, summing over the contributions from the different Feynman diagrams; therefore it depends on the interference between the contributions. In the SM, the interference for t-channel (tHq process) and W-associated (tHW process) production is destructive [49] resulting in the small cross sections presented in Table 1.10.

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 1.10: Predicted SM cross sections for tH production at $\sqrt{s} = 13$ TeV [53, 54].

917

The s-channel contribution can be neglected. It will be shown that a deviation from the SM destructive interference would result in an enhancement of the tH cross section compared to that in SM, which could be used to get information about the sign of the Higgs-top coupling [51, 52]. In order to describe tH production processes, Feynman diagram 1.15b will be considered; there, the W boson is radiated by a quark in the proton and eventually it will interact with the b quark. In the high energy regime, the effective W approximation [55] is used to describe the process as the

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

$$\mathcal{A} = \frac{g}{\sqrt{2}} \left[(\kappa_t - \kappa_V) \frac{m_t \sqrt{s}}{m_W v} A \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) + \left(\kappa_V \frac{2m_W s}{v t} + (2\kappa_t - \kappa_V) \frac{m_t^2}{m_W v} \right) B \left(\frac{t}{s}, \varphi; \xi_t, \xi_b \right) \right], \quad (1.62)$$

927 where $\kappa_V \equiv g_{HVV}/g_{HVV}^{SM}$ and $\kappa_t \equiv g_{Ht}/g_{Ht}^{SM} = y_t/y_t^{SM}$ are scaling factors that quantify
 928 possible deviations of the couplings from the SM values, Higgs-Vector boson (H-
 929 W) and Higgs-top (H-t) respectively, from the SM couplings; $s = (p_W + p_b)^2$, $t =$
 930 $(p_W - p_H)^2$, φ is the Higgs azimuthal angle around the z axis taken parallel to the
 931 direction of motion of the incoming W; A and B are functions describing the weak
 932 interaction in terms of the chiral states (ξ_t, ξ_b) of the quarks b and t . Terms that
 933 vanish in the high energy limit have been neglected as well as the Higgs and b quark
 934 masses⁹.

935 The scattering amplitude grows with energy like \sqrt{s} for $\kappa_V \neq \kappa_t$, in contrast to
 936 the SM ($\kappa_t = \kappa_V = 1$), where the first term in 1.62 cancels out and the amplitude
 937 is constant for large s ; therefore, a deviation from the SM predictions represents an
 938 enhancement in the tHq cross section. In particular, for a SM H-W coupling and a
 939 H-t coupling of inverted sign with respect to the SM ($\kappa_V = -\kappa_t = 1$) the tHq cross
 940 section is enhanced by a factor greater 10 as seen in the Figure 1.16 taken from
 941 Reference [51]; Reference [56] has reported similar enhancement results.

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

⁹ A detailed explanation of the structure and approximations used to derive \mathcal{A} can be found in Reference [51]

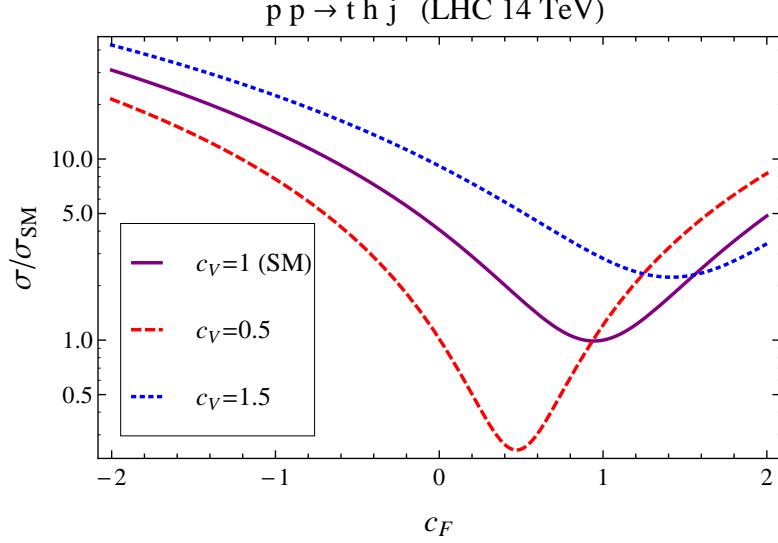


Figure 1.16: Cross section for tHq process as a function of κ_t , normalized to the SM, for three values of κ_V . In the plot c_f refers to the Higgs-fermion coupling which is dominated by the H-t coupling and represented here by κ_t . Solid, dashed and dotted lines correspond to $c_V \rightarrow \kappa_V = 1, 0.5, 1.5$ respectively. Note that for the SM ($\kappa_V = \kappa_t = 1$), the destructive effect of the interference is maximal.

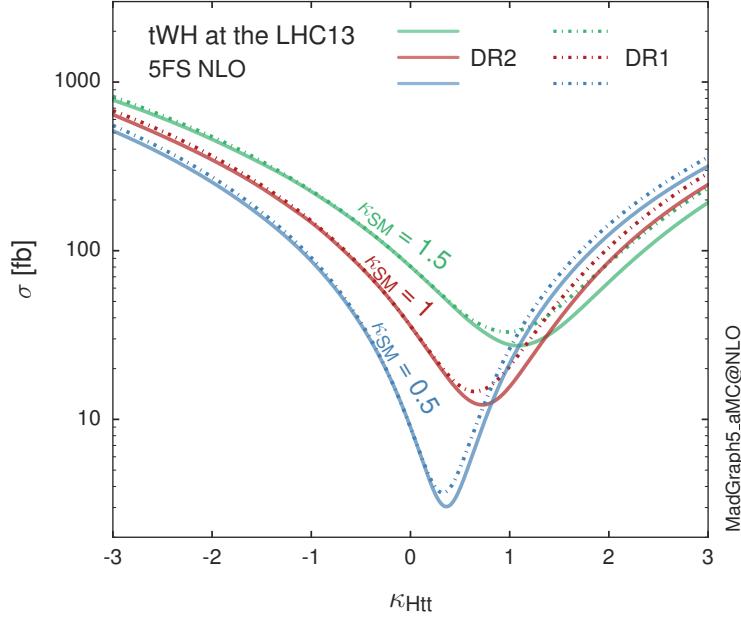


Figure 1.17: Cross section for tHW process as a function of κ_{Htt} , for three values of κ_{SM} at $\sqrt{s} = 13$ TeV. $\kappa_{Htt}^2 = \sigma_{Htt}/\sigma_{Htt}^{SM}$ is a simple re-scaling of the SM Higgs interactions.

interfering diagrams are removed (or added) from the calculations in order to evaluate the impact of the removed contributions. DR1 was defined to neglect $t\bar{t}H$ interference while DR2 was defined to take $t\bar{t}H$ interference into account [48]. As shown in Figure 1.17, the tHW cross section is enhanced from about 15 fb (SM: $\kappa_{Htt} = 1$) to about 150 fb ($\kappa_{Htt} = -1$). Differences between curves for DR1 and DR2 help to gauge the impact of the interference with $t\bar{t}H$. Results of the calculations of the tHq and tHW cross sections at $\sqrt{s} = 13$ TeV can be found in Reference [57] and a summary of the results is presented in Table 1.11.

	\sqrt{s} TeV	$\kappa_t = 1$	$\kappa_t = -1$
$\sigma^{LO}(tHq)(fb)$ [51]	8	≈ 17.4	≈ 252.7
	14	≈ 80.4	≈ 1042
$\sigma^{NLO}(tHq)(fb)$ [51]	8	$18.28^{+0.42}_{-0.38}$	$233.8^{+4.6}_{-0.0}$
	14	$88.2^{+1.7}_{-0.0}$	$982.8^{+28}_{-0.0}$
$\sigma^{LO}(tHq)(fb)$ [56]	14	≈ 71.8	≈ 893
$\sigma^{LO}(tHW)(fb)$ [56]	14	≈ 16.0	≈ 139
$\sigma^{NLO}(tHq)(fb)$ [57]	8	$18.69^{+8.62\%}_{-17.13\%}$	-
	13	$74.25^{+7.48\%}_{-15.35\%}$	$848^{+7.37\%}_{-13.70\%}$
	14	$90.10^{+7.34\%}_{-15.13\%}$	$1011^{+7.24\%}_{-13.39\%}$
$\sigma^{LO}(tHW)(fb)$ [48]	13	$15.77^{+15.91\%}_{-15.76\%}$	-
$\sigma^{NLO}DR1(tHW)(fb)$ [48]	13	$21.72^{+6.52\%}_{-5.24\%}$	≈ 150
$\sigma^{NLO}DR2(tHW)(fb)$ [48]	13	$16.28^{+7.34\%}_{-15.13\%}$	≈ 150

Table 1.11: Predicted enhancement of the tHq and tHW cross sections at LHC for $\kappa_V = 1$ and $\kappa_t = \pm 1$ at LO and NLO; the cross section enhancement of more than a factor of 10 is due to the flipping in the sign of the H-t coupling with respect to the SM one.

954

1.6 CP-mixing in tH processes

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

958 not conserve CP [47, 48, 58].

959 In this thesis, the sensitivity of tH processes to CP-mixing is also studied on the
 960 basis of References [47, 48] using the effective field theory framework where a generic
 961 particle (X_0) of spin-0 and a general CP violating interaction with the top quark
 962 (Htt coupling), can couple to scalar and pseudo-scalar fermionic densities. The H-W
 963 interaction is assumed to be SM-like. The Lagrangian modeling the H-t interaction
 964 is given by

$$\mathcal{L}_0^t = -\bar{\psi}_t (c_\alpha \kappa_{Htt} g_{Htt} + i s_\alpha \kappa_{Att} g_{Att} \gamma_5) \psi_t X_0, \quad (1.63)$$

965 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
 966 dimensionless re-scaling parameters¹⁰ used to parametrize the magnitude of the CP-
 967 violating and CP-conserving parts of the amplitude. The model defines $g_{Htt} =$
 968 $g_{Att} = m_t/v = y_t/\sqrt{2}$ with $v \sim 246$ GeV the Higgs vacuum expectation value. In
 969 this parametrization, three special cases can be recovered

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

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

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

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

$$\mathcal{L}_0^g = -\frac{1}{4} \left(c_\alpha \kappa_{Hgg} g_{Hgg} G_{\mu\nu}^a G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} \right) X_0. \quad (1.64)$$

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

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

¹⁰ analog to κ_t and κ_V

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

$$\frac{\sigma_{NLO}^{gg \rightarrow X_0}}{\sigma_{NLO,SM}^{gg \rightarrow H}} = c_\alpha^2 \kappa_{Htt}^2 + s_\alpha^2 \left(\kappa_{Att} \frac{g_{Agg}}{g_{Hgg}} \right)^2. \quad (1.65)$$

979 If the re-scaling parameters are set to

$$\kappa_{Htt} = 1, \quad \kappa_{Att} = \left| \frac{g_{Hgg}}{g_{Agg}} \right| = \frac{2}{3}. \quad (1.66)$$

980 the gluon-fusion SM cross section is reproduced for every value of the CP-mixing
 981 angle α ; therefore, by imposing that condition to the Lagrangian density 1.63, the
 982 CP-mixing angle is not constrained by current data. Figure 1.18 shows the NLO cross
 983 sections for t-channel tX_0 (blue) and $t\bar{t}X_0$ (red) associated production processes as a
 984 function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark
 985 CP-violating coupling. Re-scaling factors κ_{Htt} and κ_{Att} have been set to reproduce
 986 the SM gluon-fusion cross sections.

987 It is interesting to notice that the tX_0 cross section is enhanced, by a factor of
 988 about 10, when a continuous rotation in the scalar-pseudoscalar plane is applied; this
 989 enhancement is similar to the enhancement produced when the H-t coupling is flipped
 990 in sign with respect to the SM ($y_t = -y_{t,SM}$ in the plot), as showed in Section 1.5. In
 991 contrast, the degeneracy in the $t\bar{t}X_0$ cross section is still present given that it depends
 992 quadratically on the H-t coupling, but more interesting is to notice that $t\bar{t}X_0$ cross
 993 section is exceeded by tX_0 cross section after $\alpha \sim 60^\circ$.

994 A similar parametrization can be used to investigate the tHW process sensitivity
 995 to CP-violating H-t coupling. As said in 1.5, the interference in the W-associated
 996 channel is more complicated because there are more than two contributions and also
 997 there is interference with the $t\bar{t}H$ production process.

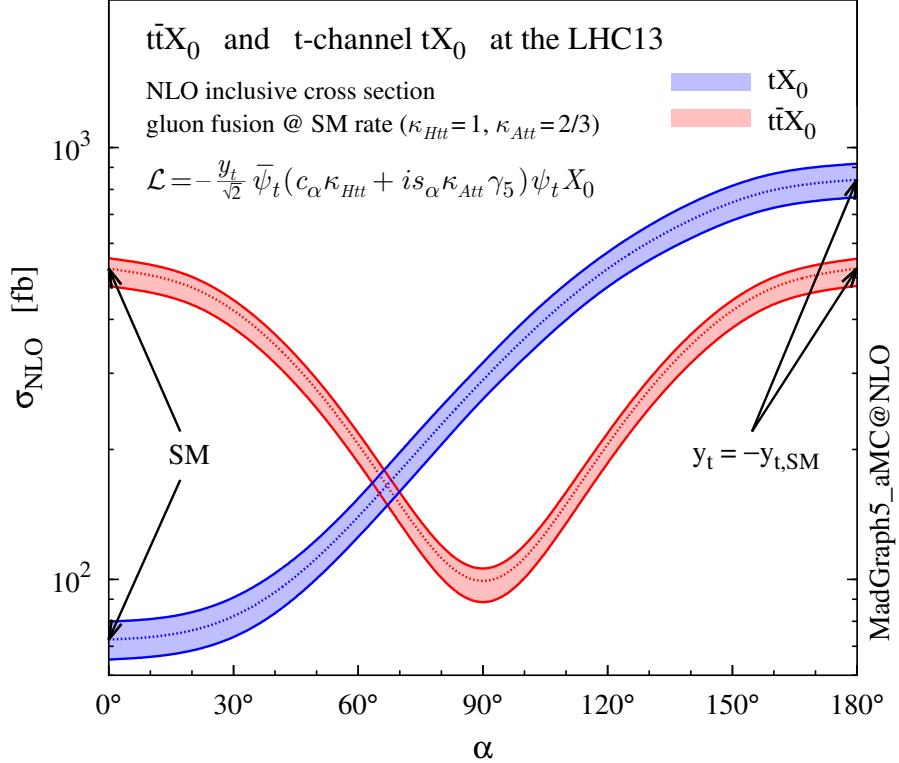


Figure 1.18: NLO cross sections for t-channel tX_0 (blue) and $t\bar{t}X_0$ (red) associated production processes as a function of the CP-mixing angle α . X_0 is a generic spin-0 particle with top quark CP-violating coupling [47].

998 Figure 1.19 shows the NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red)
999 associated production and for the combined $tWX_0 + t\bar{t}X_0 + \text{interference}$ (orange) as
1000 a function of the CP-mixing angle. It is clear that the effect of the interference in the
1001 combined case is the lifting of the degeneracy present in the $t\bar{t}X_0$ production. The
1002 constructive interference enhances the cross section from about 500 fb at SM ($\alpha = 0$)
1003 to about 600 fb ($\alpha = 180^\circ \rightarrow y_t = -y_{t,SM}$).

1004 An analysis combining tHq and tHW processes will be made in this thesis taking
1005 advantage of the sensitivity improvement.

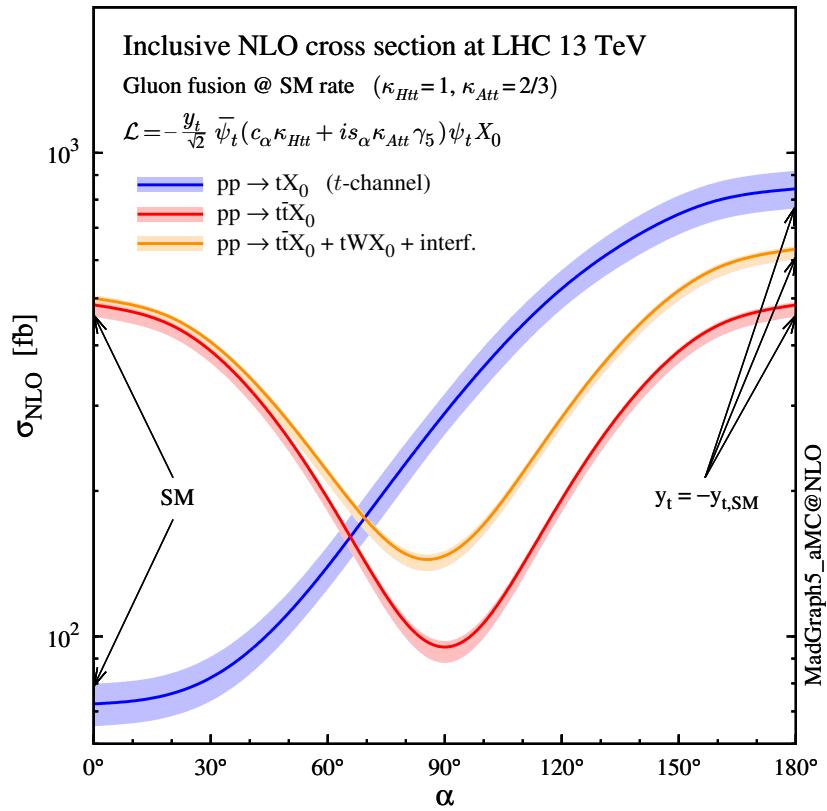


Figure 1.19: NLO cross sections for t-channel tX_0 (blue), $t\bar{t}X_0$ (red) associated production processes and combined $tWX_0 + t\bar{t}X_0$ (including interference) production as a function of the CP-mixing angle α [47].

1006 **Chapter 2**

1007 **The CMS experiment at the LHC**

1008 **2.1 Introduction**

1009 Located on the Swiss-French border, the European Council for Nuclear Research
1010 (CERN) is the largest scientific organization leading particle physics research. About
1011 13000 people in a broad range of roles including users, students, scientists, engineers,
1012 among others, contribute to the data taking and analysis, with the goal of unveiling
1013 the secrets of nature and revealing the fundamental structure of the universe. CERN
1014 is also the home of the Large Hadron Collider (LHC), the largest particle accelerator
1015 around the world, where protons (or heavy ions) traveling close to the speed of light,
1016 are made to collide. These collisions open a window to investigate how particles (and
1017 their constituents if they are composite) interact with each other, providing clues
1018 about the laws of nature. This chapter presents an overview of the LHC structure
1019 and operation. A detailed description of the CMS detector is offered, given that the
1020 data used in this thesis have been taken with this detector.

1021 2.2 The LHC

1022 With 27 km of circumference, the LHC is currently the most powerful circular accelerator
 1023 in the world. It is installed in the same tunnel where the Large Electron-Positron
 1024 (LEP) collider was located, taking advantage of the existing infrastructure. The LHC
 1025 is part of the CERN's accelerator complex composed of several successive accelerat-
 1026 ing stages before the particles are injected into the LHC ring where they reach their
 1027 maximum energy (see Figure 2.1).

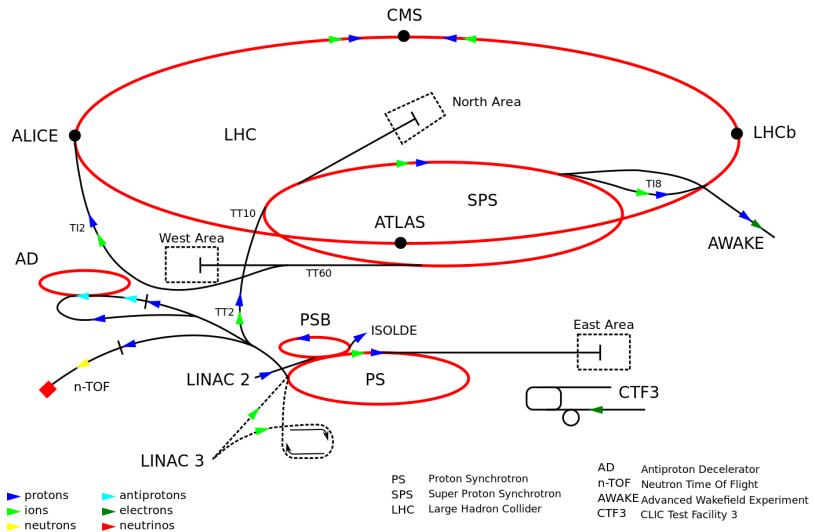


Figure 2.1: CERN accelerator complex. Blue arrows show the path followed by protons along the acceleration process [61].

1028 The LHC runs in three collision modes depending on the particles being acceler-
 1029 ated

- 1030 • Proton-Proton collisions (pp) for multiple physics experiments.
- 1031 • Lead-Lead collisions ($Pb-Pb$) for heavy ion experiments.
- 1032 • Proton-Lead collisions ($p-Pb$) for quark-gluon plasma experiments.

1033 In this thesis only pp collisions will be considered.

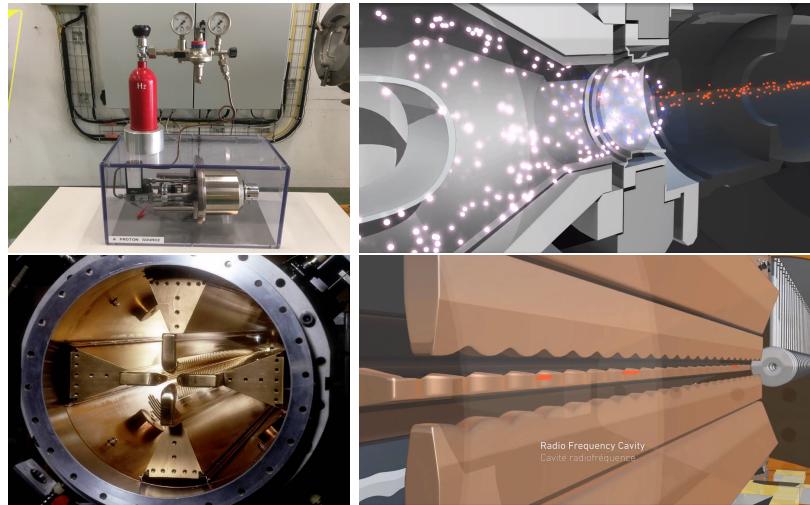


Figure 2.2: LHC protons source and the first acceleration stage. Top: the bottle contains hydrogen gas which is injected into the metal cylinder (white dots) to be broken down into electrons (blue dots) and protons (red dots); Bottom: the obtained protons are directed towards the radio frequency quadrupole which perform the first acceleration, focus the beam and create the bunches of protons. Left images are real pictures while right images are drawings [65, 66].

1034 Collection of protons starts with hydrogen atoms taken from a bottle, containing
 1035 hydrogen gas, and injecting them in a metal cylinder; hydrogen atoms are broken
 1036 down into electrons and protons by an intense electric field (see Figure 2.2 top).
 1037 The resulting protons leave the metal cylinder towards a radio frequency quadrupole
 1038 (RFQ) that focus the beam, accelerates the protons and creates the packets of protons
 1039 called bunches. In the RFQ, an electric field is generated by a RF wave at a frequency
 1040 that matches the resonance frequency of the cavity where the electrodes are contained.
 1041 The beam of protons traveling on the RFQ axis experiences an alternating electric
 1042 field gradient that generates the focusing forces.

1043 In order to accelerate the protons, a longitudinal time-varying electric field com-
 1044 ponent is added to the system; it is done by giving the electrodes a sine-like profile as
 1045 shown in Figure 2.2 bottom. By matching the speed and phase of the protons with
 1046 the longitudinal electric field the bunching is performed; protons synchronized with

1047 the RFQ (synchronous protons) do not feel an accelerating force, but those protons in
 1048 the beam that have more (or less) energy than the synchronous proton (asynchronous
 1049 protons) will feel a decelerating (accelerating) force; therefore, asynchronous protons
 1050 will oscillate around the synchronous ones forming bunches of protons [63]. From the
 1051 RFQ protons emerge with energy 750 keV in bunches of about 1.15×10^{11} protons [64].

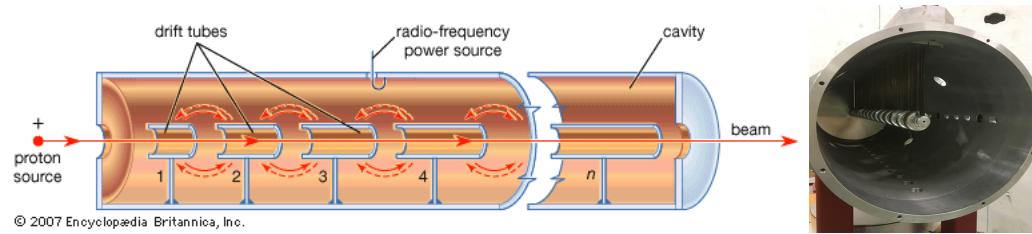


Figure 2.3: Left: drawing of the LINAC2 accelerating system at CERN. Electric fields generated by radio frequency (RF) create acceleration and deceleration zones inside the cavity; deceleration zones are blocked by drift tubes where quadrupole magnets focus the proton beam. Right: picture of a real RF cavity [67].

1052 Proton bunches coming from the RFQ go to the linear accelerator 2 (LINAC2)
 1053 where they are accelerated to reach 50 MeV energy. In the LINAC2 stage, acceleration
 1054 is performed using electric fields generated by radio frequency which create zones
 1055 of acceleration and deceleration as shown in Figure 2.3. In the deceleration zones,
 1056 the electric field is blocked using drift tubes where protons are free to drift while
 1057 quadrupole magnets focus the beam.

1058 The beam coming from LINAC2 is injected into the proton synchrotron booster
 1059 (PSB) to reach 1.4 GeV in energy. The next acceleration is provided at the proton
 1060 synchrotron (PS) up to 26 GeV, followed by the injection into the super proton
 1061 synchrotron (SPS) where protons are accelerated to 450 GeV. Finally, protons are
 1062 injected into the LHC where they are accelerated to the target energy of 6.5 TeV.

1063 PSB, PS, SPS and LHC accelerate protons using the same RF acceleration tech-
 1064 nique described before.

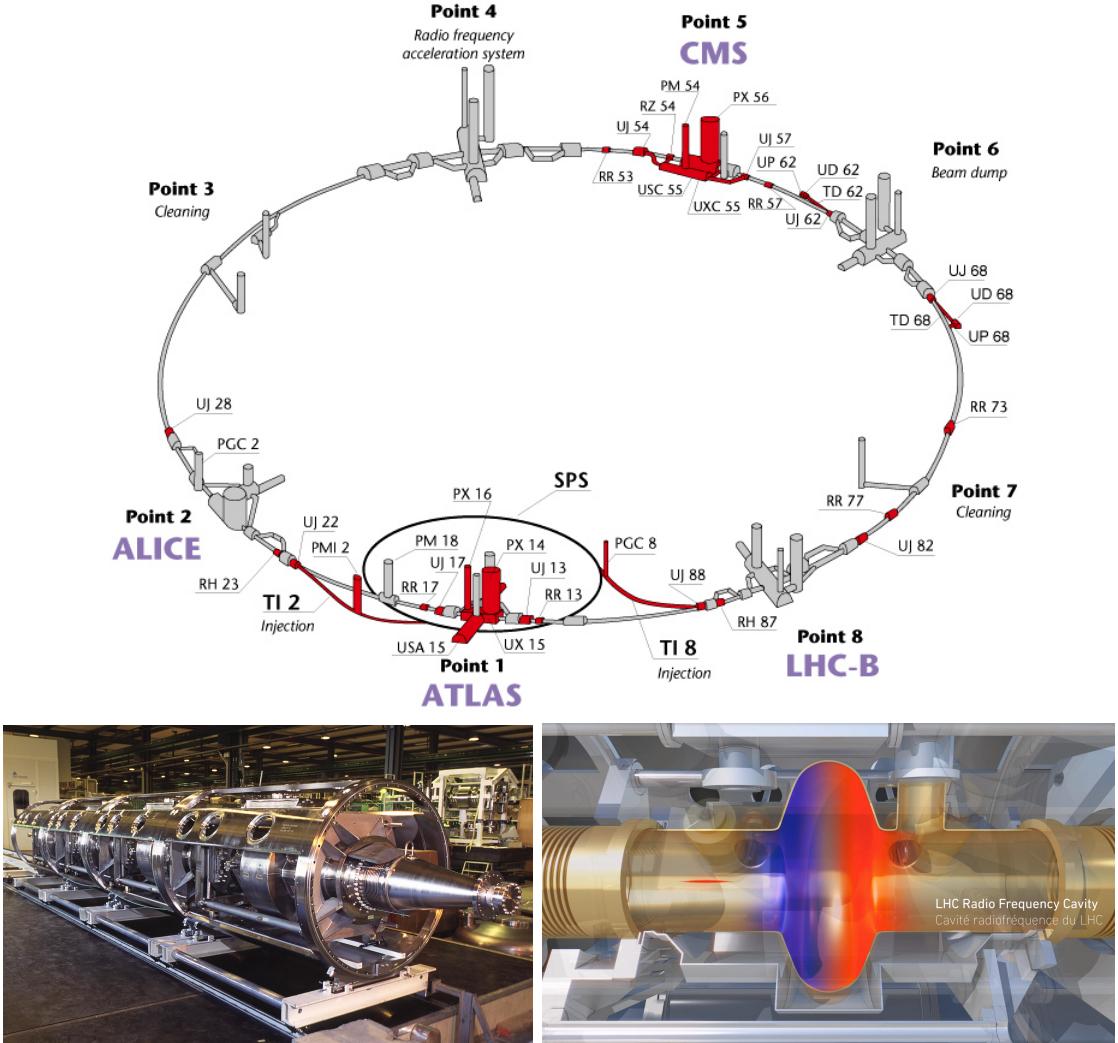


Figure 2.4: Top: LHC layout. The red zones indicate the infrastructure additions to the LEP installations, built to accommodate the ATLAS and CMS experiments which exceed the size of the former experiments located there [62]. Bottom: LHC RF cavities. A module (left picture) accommodates 4 cavities, each like the one in the right drawing, that accelerate protons and preserve the bunch structure of the beam. The color gradient pattern represents the strength of the electric field inside the cavity; red zone corresponds to the maximum of the oscillation electric field while the blue zone corresponds to the minimum. [66, 68]

1065 The LHC has a system of 16 RF cavities located in the so-called point 4, as
 1066 shown in Figure 2.4 top, tuned at a frequency of 400 MHz. The bottom side of
 1067 Figure 2.4 shows a picture of a RF module composed of 4 RF cavities working in a
 1068 superconducting state at 4.5 K; also, a representation of the accelerating electric field

1069 that accelerates the protons in the bunch is shown. The maximum of the oscillating
 1070 electric field (red region) picks the proton bunches at the entrance of the cavity
 1071 and keeps accelerating them through the whole cavity. The protons are carefully
 1072 timed so that in addition to the acceleration effect the bunch structure of the beam
 1073 is preserved.

1074 While protons are accelerated in one section of the LHC ring, where the RF cavities
 1075 are located, in the rest of their path they have to be kept in the curved trajectory
 1076 defined by the LHC ring. Technically, LHC is not a perfect circle; RF, injection, beam
 1077 dumping, beam cleaning and sections before and after the experimental points where
 1078 protons collide are all straight sections. In total, there are 8 arcs 2.45 km long each
 1079 and 8 straight sections 545 m long each. In order to curve the proton's trajectory in
 1080 the arc sections, superconducting dipole magnets are used.

1081 Inside the LHC ring, there are two proton beams traveling in opposite directions
 1082 in two separated beam pipes; the beam pipes are kept at ultra-high vacuum ($\sim 10^{-9}$
 1083 Pa) to ensure that there are no particles that interact with the proton beams. The
 1084 superconducting dipole magnets used in LHC are made of a NbTi alloy, capable of
 1085 transporting currents of about 12000 A when cooled at a temperature below 2K using
 1086 liquid helium (see Figure 2.5).

1087 Protons in the arc sections of LHC feel a centripetal force exerted by the dipole
 1088 magnets; the magnitude of magnetic field needed to keep the protons in the LHC
 1089 curved trayectomy can be found assuming that protons travel at $v \approx c$, using the
 1090 standard values for proton mass and charge and the LHC radius, as

$$F_m = \frac{mv^2}{r} = qBv \rightarrow B = 8.33T \quad (2.1)$$

1091 which is about 100000 times the Earth's magnetic field. A representation of the

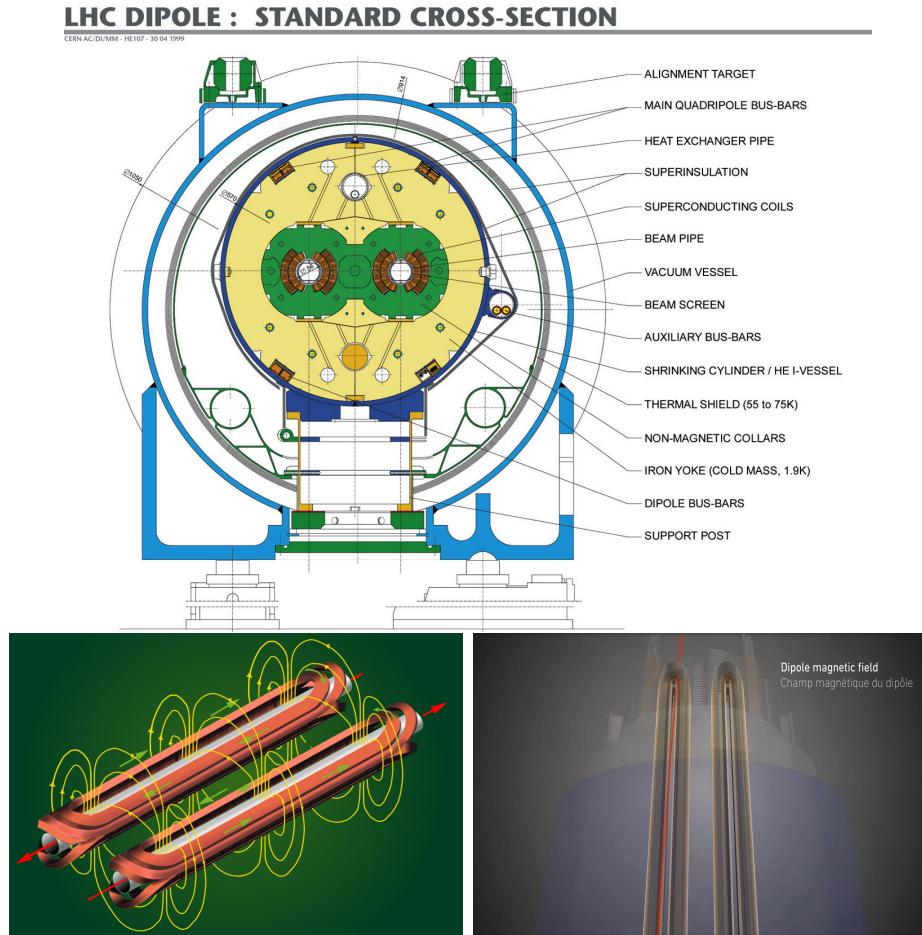


Figure 2.5: Top: LHC dipole magnet transverse view; cooling, shielding and mechanical support are indicated. Bottom left: Magnetic field generated by the dipole magnets; note that the direction of the field inside one beam pipe is opposite with respect to the other beam pipe which guarantee that both 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 [66, 69, 70].

1092 magnetic field generated by the dipole magnets is shown on the bottom left side of
 1093 Figure 2.5. The bending effect of the magnetic field on the proton beam is shown on
 1094 the bottom right side of Figure 2.5. Note that the dipole magnets are not curved;
 1095 the arc section of the LHC ring is composed of straight dipole magnets of about 15
 1096 m. In total there are 1232 dipole magnets along the LHC ring.
 1097 In addition to the bending of the beam trajectory, the beam has to be focused. The

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

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

$$L = fn \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (2.2)$$

where f is the revolution frequency, n is the number of bunches per beam, N_1 and N_2 are the numbers of protons per bunch, σ_x and σ_y are the gaussian transverse sizes of the bunches. With the expected parameters, the LHC expected luminosity is about

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

$$n = 2808$$

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

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

1113

$$L = 1.28 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1} = 1.28 \times 10^{-5} \text{ fb}^{-1} \text{ s}^{-1} \quad (2.3)$$

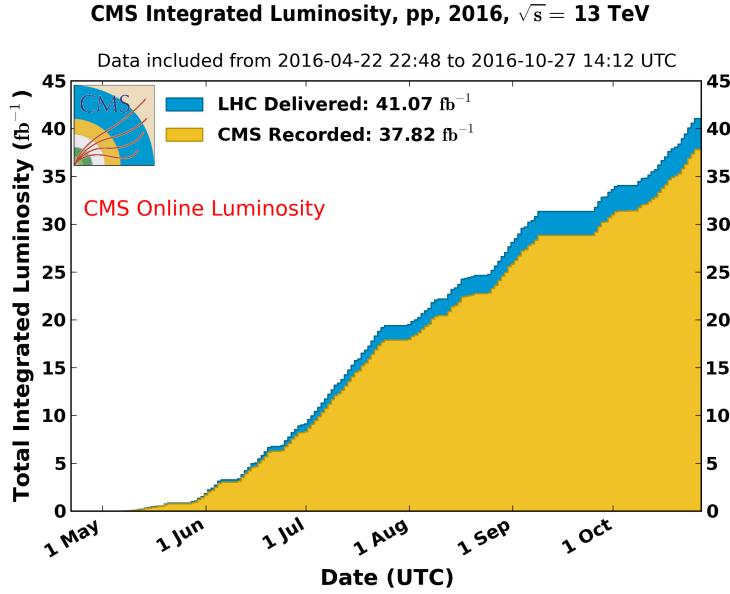


Figure 2.6: Integrated luminosity delivered by LHC and recorded by CMS during 2016. The difference between the delivered and the recorded luminosity is due to fails and issues occurred during the data taking in the CMS experiment [71].

1114 Luminosity is a fundamental aspect of LHC given that the bigger luminosity the
 1115 bigger number of collisions, which means that for processes with a very small cross
 1116 section the number of expected occurrences is increased and so the chances of being
 1117 detected. The integrated luminosity, collected by the CMS experiment during 2016
 1118 is shown in Figure 2.6; the data analyzed in this thesis corresponds to an integrated
 1119 luminosity of 35.9 fb^{-1} at a center of mass-energy $\sqrt{s} = 13 \text{ TeV}$.

1120 One way to increase L is increasing the number of bunches in the beam. Cur-
 1121 rently, the separation between two consecutive bunches in the beam is 7.5 m which
 1122 corresponds to a time separation of 25 ns. In the full LHC ring the allowed number
 1123 of bunches is $n = 27\text{km}/7.5\text{m} = 3600$; however, there are some gaps in the bunch pat-
 1124 tern intended for preparing the dumping and injection of the beam, thus, the proton
 1125 beams are composed of 2808 bunches.

1126 Once the proton beams reach the desired energy, they are brought to cross each

other producing pp collisions. The bunch crossing happens in precise places where the four LHC experiments are located, as seen in the top of Figure 2.7. In 2008 pp collisions of $\sqrt{s} = 7$ TeV were performed; the energy was increased to 8 TeV in 2012 and to 13 TeV in 2015.

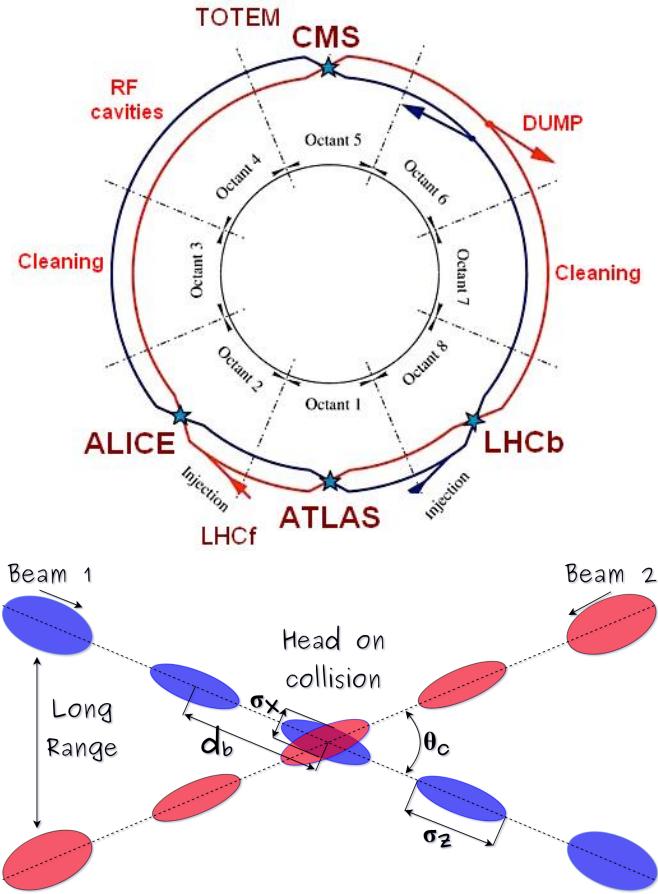


Figure 2.7: Top: LHC interaction points. Bunch crossing occurs where the LHC experiments are located [72]. Sections indicated as cleaning are dedicated to collimate the beam in order to protect the LHC ring from collisions with protons in very spreaded bunches. Bottom: bunch crossing scheme. Since the bunch crossing is not perfectly head-on, the luminosity is reduced in a factor of 17%; adapted from Reference [86].

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

1134 ion collisions searches; TOTEM and LHCf are dedicated to forward physics studies;
 1135 MoEDAL (not indicated in the Figure) is intended for monopole or massive pseudo
 1136 stable particles searches.

1137 At the IP there are two interesting details that need to be addressed. The first
 1138 one is that the bunch crossing does not occur head-on but at a small crossing angle θ_c
 1139 (280 μ rad in CMS and ATLAS) as shown in the bottom side of Figure 2.7, affecting
 1140 the overlapping between bunches; the consequence is a reduction of about 17% in
 1141 the luminosity (represented by a factor not included in eqn. 2.2). The second one
 1142 is the occurrence of multiple pp collisions in the same bunch crossing; this effect is
 1143 called *pileup* (PU). A fairly simple estimation of the PU follows from estimating the
 1144 probability of collision between two protons, one from each of the bunches in the
 1145 course of collision; it depends roughly on the ratio of proton size and the cross section
 1146 of the bunch in the IP, i.e.,

$$P(pp\text{-}collision) \sim \frac{d_{proton}^2}{\sigma_x \sigma_y} = \frac{(1\text{fm})^2}{(16\mu\text{m})^2} \sim 4 \times 10^{-21} \quad (2.4)$$

1147 however, there are $N = 1.15 \times 10^{11}$ protons in a bunch, thus the estimated number of
 1148 collisions in a bunch crossing is

$$PU = N^2 * P(pp\text{-}collision) \sim 50pp \text{ collision per bunch crossing}, \quad (2.5)$$

1149 about 20 of which are inelastic. A multiple pp collision event in a bunch crossing at
 1150 CMS is shown in Figure 2.8.

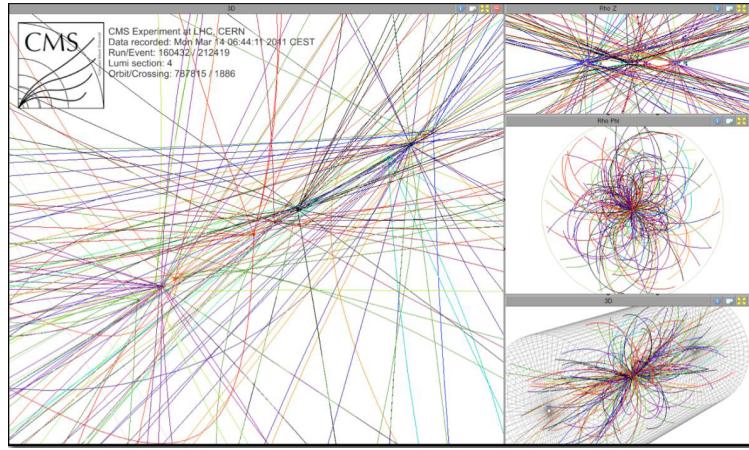


Figure 2.8: Multiple pp collision bunch crossing at CMS. [73].

1151 2.3 The CMS experiment

1152 CMS is a general-purpose detector designed to conduct research in a wide range
 1153 of physics from the standard model to new physics like extra dimensions and dark
 1154 matter. Located at Point 5 in the LHC layout as shown in Figure 2.4, CMS is
 1155 composed of several detection systems distributed in a cylindrical structure; in total,
 1156 CMS weights about 12500 tons in a very compact 21.6 m long and 14.6 m diameter
 1157 cylinder. It was built in 15 separate sections at the ground level and lowered to the
 1158 cavern individually to be assembled. A complete and detailed description of the CMS
 1159 detector and its components is given in Reference [74] on which this section is based.
 1160 Figure 2.9 shows the layout of the CMS detector. The detection system is composed
 1161 of (from the innermost to the outermost)

- 1162 • Pixel detector.
- 1163 • Silicon strip tracker.
- 1164 • Preshower detector.
- 1165 • Electromagnetic calorimeter.

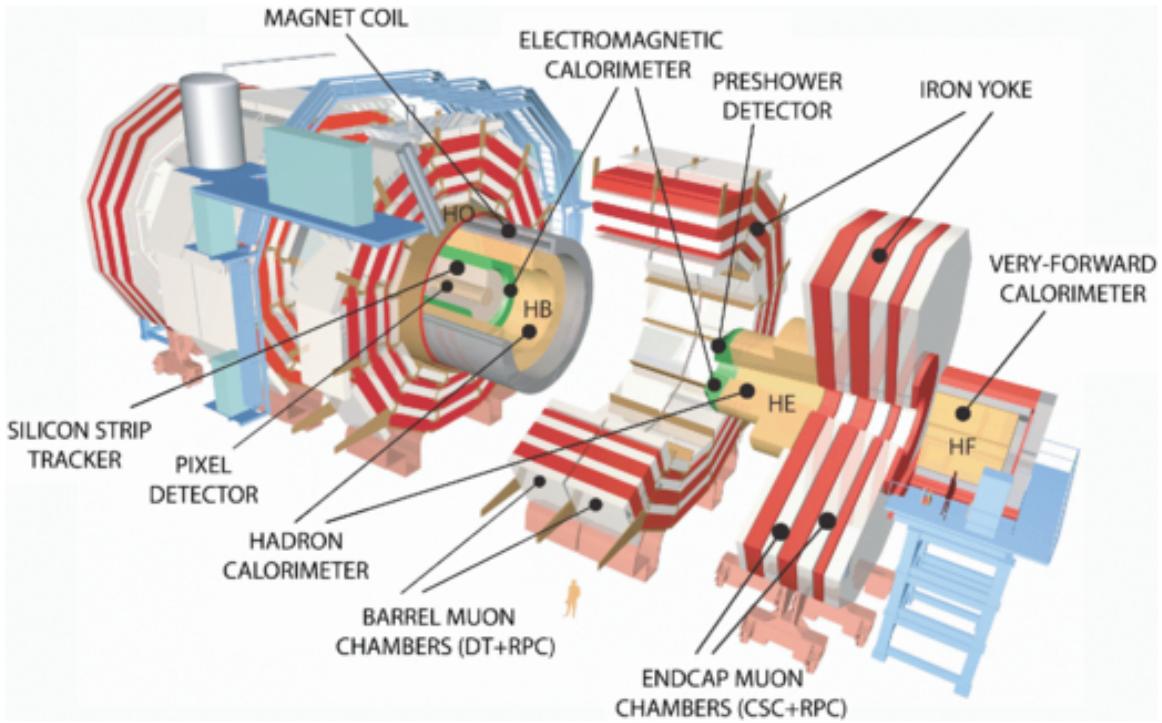


Figure 2.9: Layout of the CMS detector. The several subdetectors are indicated. The central region of the detector is referred as the barrel section while the endcaps are referred as the forward sections. [75].

1166 • Hadronic calorimeter.

1167 • Muon chambers (barrel and endcap)

1168 The central region of the detector is commonly referred as the barrel section while
 1169 the endcaps are referred as the forward sections of the detector; thus, each subdetector
 1170 is composed of a barrel section and a forward section.

1171 When a pp collision happens inside the CMS detector, many different particles are
 1172 produced, but only some of them live long enough to be detected; they are electrons,
 1173 photons, pions, kaons, protons, neutrons and muons; neutrinos are not detected by
 1174 the CMS detector. Thus, the CMS detector was designed to detect those particles and
 1175 measure their properties. Figure 2.10 shows a transverse slice of the CMS detector.
 1176 The silicon tracker (pixel detector + strip tracker) is capable to register the track of

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

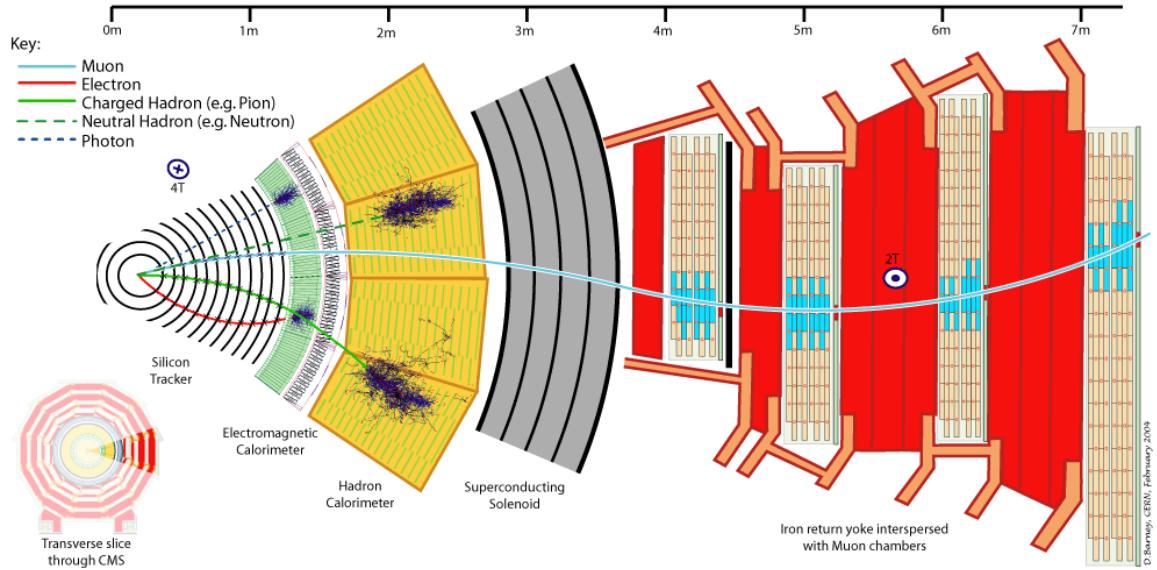


Figure 2.10: CMS detector transverse slice [76].

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

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

1192 A photon will not leave a track because it is neutral, but it will be absorbed in
 1193 the electromagnetic calorimeter.

1194 A neutral hadron, like the neutron, will not leave a track either but it will lose a
 1195 small amount of its energy during its passage through the electromagnetic calorimeter
 1196 and then it will be absorbed in the hadron calorimeter depositing the rest of its energy.

1197 A charged hadron, like the proton or π^\pm , will leave a curved track on the silicon
 1198 tracker, some of its energy in the electromagnetic calorimeter and finally will be
 1199 absorbed in the hadronic calorimeter.

1200 A more detailed description of each detection system will be presented in the
 1201 following sections.

1202 2.3.1 CMS coordinate system

1203 The coordinate system used by CMS is centered on the geometrical center of the
 1204 detector which is the nominal IP as shown in Figure 2.11¹. The z -axis is parallel
 1205 to the beam direction, while the Y -axis pointing vertically upward, and the X -axis
 1206 pointing radially inward toward the center of the LHC.

1207 In addition to the common cartesian and cylindrical coordinate systems, two co-
 1208 ordinates are of particular utility in particle physics: rapidity (y) and pseudorapidity
 1209 (η), defined in connection to the polar angle θ , energy and longitudinal momentum
 1210 component (momentum along the z -axis) according to

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad \eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (2.6)$$

1211 Rapidity is related to the angle between the XY -plane and the direction in which
 1212 the products of a collision are emitted; it has the nice property that the difference

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

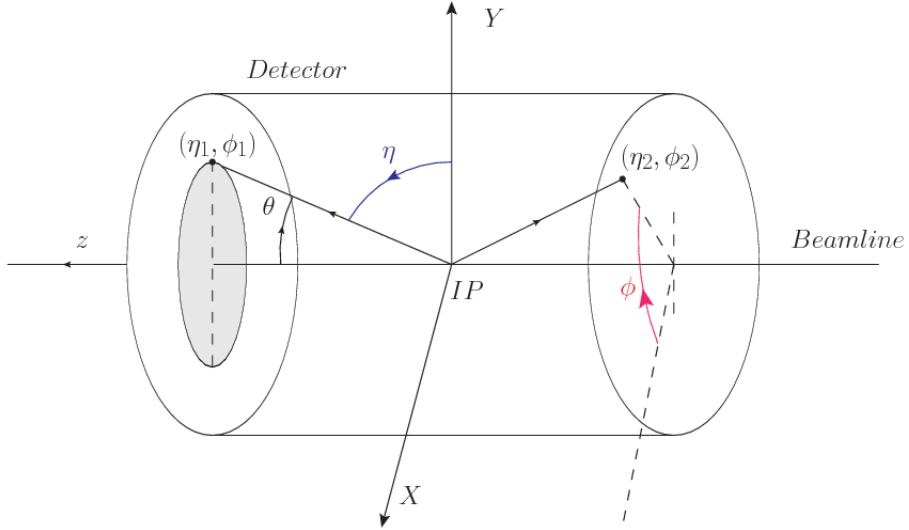


Figure 2.11: CMS detector coordinate system.

1213 between the rapidities of two particles is invariant with respect to Lorentz boosts
 1214 along the z -axis, hence, data analysis becomes more simple when based on rapid-
 1215 ity; however, it is not simple to measure the rapidity of highly relativistic particles,
 1216 as those produced after pp collisions. Under the highly relativistic motion approxi-
 1217 mation, y can be rewritten in terms of the polar angle, concluding that rapidity is
 1218 approximately equal to the pseudorapidity defined above, i.e., $y \approx \eta$. Note that η
 1219 is easier to measure than y given the direct relationship between the former and the
 1220 polar angle.

1221 The angular distance between two objects in the detector (ΔR) is commonly used
 1222 to judge the isolation of those object; it is defined in terms of their coordinates (η_1, ϕ_1) ,
 1223 (η_2, ϕ_2) as

$$\Delta R = \sqrt{(\Delta\eta)^2 - (\Delta\phi)^2} \quad (2.7)$$

1224 2.3.2 Tracking system

1225 The CMS tracking system is designed to provide a precise measurement of the trajectories (*track*) followed by the charged particles created after the pp collisions; also, the
 1226 precise reconstruction of the primary and secondary origins of the tracks (*vertices*) is
 1227 expected in an environment where, each 25 ns, the bunch crossing produces about 20
 1228 inelastic collisions and about 1000 particles.
 1229

1230 Physics requirements guiding the tracking system performance include the precise characterization of events involving gauge bosons, W and Z, and their leptonic
 1231 decays for which an efficient isolated lepton and photon reconstruction is of capital
 1232 importance, given that isolation is required to suppress background events to a level
 1233 that allows observations of interesting processes like Higgs boson decays or beyond
 1234 SM events.

1236 The ability to identify and reconstruct b -jets and B-hadrons within these jets is also
 1237 a fundamental requirement, achieved through the ability to reconstruct accurately
 1238 displaced vertices, given that b -jets are part of the signature of top quark physics, like
 1239 the one treated in this thesis.

1240 An schematic view of the CMS tracking system is shown in Figure 2.12

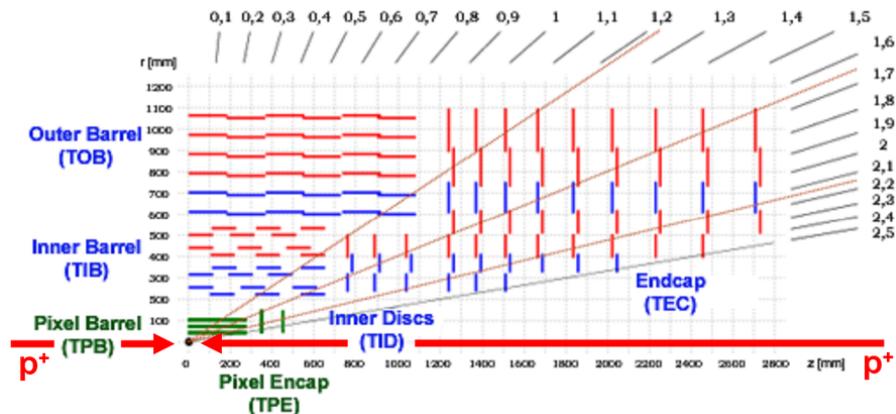


Figure 2.12: CMS tracking system schematic view [78].

1241 In order to satisfy these performance requirements, the tracking system uses two
 1242 different detector subsystems arranged in concentric cylindrical volumes, the pixel
 1243 detector and the silicon strip tracker; the pixel detector is located in the high particle
 1244 density region ($r < 20\text{cm}$) while the silicon strip tracker is located in the medium and
 1245 lower particle density regions $20\text{cm} < r < 116\text{cm}$.

1246 **Pixel detector**

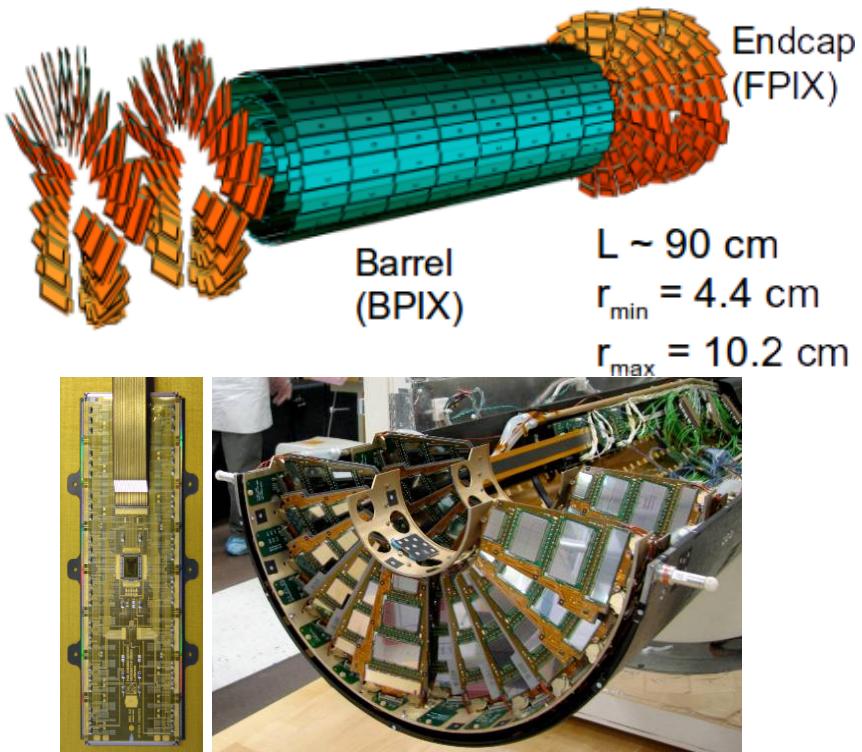


Figure 2.13: CMS pixel detector. Top: schematic view; Bottom: pictures of a barrel(BPIX) module(left) and forward modules(right) [74].

1247 The pixel detector was replaced during the 2016-2017 extended year-end technical
 1248 stop, due to the increasingly challenging operating conditions like the higher particle
 1249 flux and more radiation harsh environment, among others. The new one is responding
 1250 as expected, reinforcing its crucial role in the successful way to fulfill the new LHC

physics objectives after the discovery of the Higgs boson. Since the data sets used in this thesis were produced using the previous version of the pixel detector, it will be the subject of the description in this section. The last chapter of this thesis is dedicated to describe my contribution to the *Forward Pixel Phase 1 upgrade*.

The pixel detector was composed of 1440 silicon pixel detector modules organized in three-barrel layers in the central region (BPix) and two disks in the forward region (FPix) as shown in the top side of Figure 2.13; it was designed to record efficiently and with high precision, up to $20\mu\text{m}$ in the XY -plane and $20\mu\text{m}$ in the z -direction, the first three space-points (*hits*) nearest to the IP region in the range $|\eta| \leq 2.5$. The first barrel layer was located at a radius of 44 mm from the beamline, while the third layer was located at a radius of 102 mm, closer to the strip tracker inner barrel layer (see Section 2.3.3) in order to reduce the rate of fake tracks. The high granularity of the detector is represented in its about 66 Mpixels, each of size $100 \times 150\mu\text{m}^2$. The transverse momentum resolution of tracks can be measured with a resolution of 1-2% for muons of $p_T = 100$ GeV.

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

Commonly, the charge produced by traversing of a particle is collected by and shared among several pixels; by interpolating between pixels, the spatial resolution is improved. In the barrel section the charge sharing in the $r\phi$ -plane is due to the Lorentz effect. In the forward pixels the charge sharing is enhanced by arranging the

1277 blades in the turbine-like layout as shown in Figure 2.13 bottom left.

1278 2.3.3 Silicon strip tracker

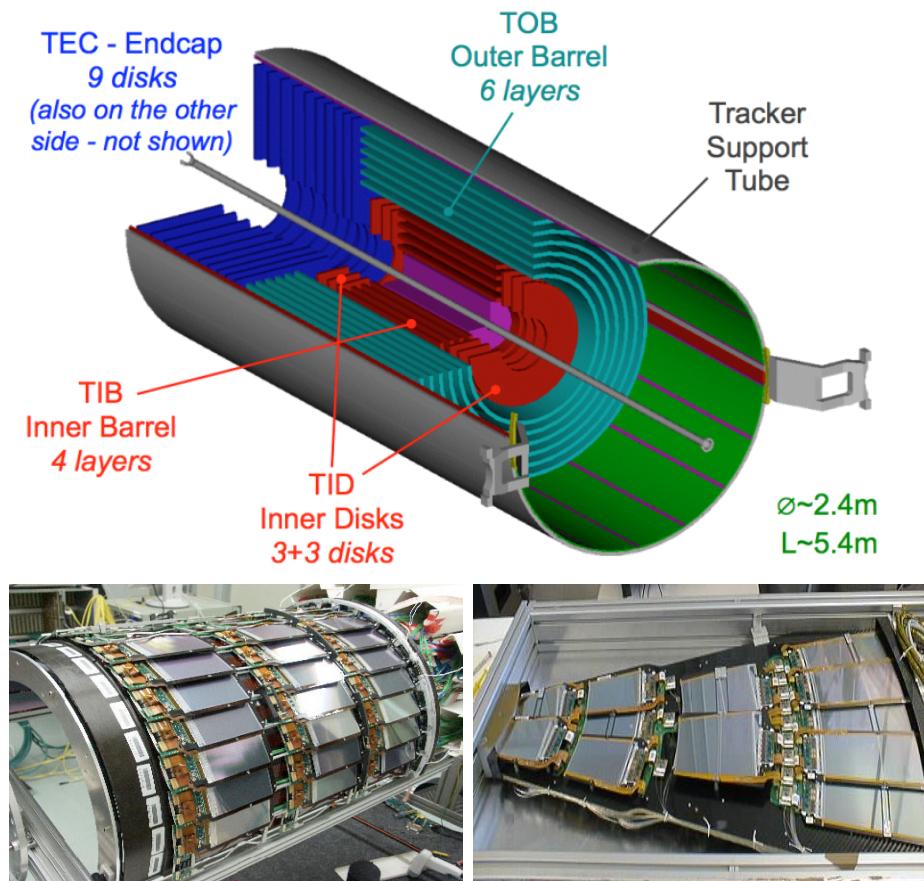


Figure 2.14: Top: CMS Silicon Strip Tracker (SST) schematic view. The SST is composed of the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). Each part is made of silicon strip modules; the modules in blue represent two modules mounted back-to-back and rotated in the plane of the module by a stereo angle of 120 mrad in order to provide a 3-D reconstruction of the hit positions. Bottom: pictures of the TIB (left) and TEC (right) modules [80–82].

1279 The silicon strip tracker (SST) is the second stage in the CMS tracking system.

1280 The top side of Figure 2.14 shows a schematic of the SST. The inner tracker region is

1281 composed of the tracker inner barrel (TIB) and the tracker inner disks (TID) covering

1282 the region $r < 55$ cm and $|z| < 118$ cm. The TIB is composed of 4 layers while the TID

1283 is composed of 3 disks at each end. The silicon sensors in the inner tracker are 320
 1284 μm thick, providing a resolution of about 13-38 μm in the $r\phi$ position measurement.

1285 The modules indicated in blue in the schematic view of Figure 2.14 are two mod-
 1286 ules mounted back-to-back and rotated in the plane of the module by a *stereo* angle
 1287 of 100 mrad; the hits from these two modules, known as *stereo hits*, are combined to
 1288 provide a measurement of the second coordinate (z in the barrel and r on the disks)
 1289 allowing the reconstruction of hit positions in 3-D.

1290 The outer tracker region is composed of the tracker outer barrel (TOB) and the
 1291 tracker endcaps (TEC). The six layers of the TOB offer coverage in the region $r > 55$
 1292 cm and $|z| < 118$ cm, while the 9 disks of the TEC cover the region $124 < |z| < 282$
 1293 cm. The resolution offered by the outer tracker is about 13-38 μm in the $r\phi$ position
 1294 measurement. The inner four TEC disks use silicon sensors 320 μm thick; those in
 1295 the TOB and the outer three TEC disks use silicon sensors of 500 μm thickness. The
 1296 silicon strips run parallel to the z -axis and the distance between strips varies from 80
 1297 μm in the inner TIB layers to 183 μm in the inner TOB layers; in the endcaps the
 1298 wedge-shaped sensors with radial strips, whose pitch range between 81 μm at small
 1299 radii and 205 μm at large radii.

1300 The whole SST has 15148 silicon modules, 9.3 million silicon strips and cover a
 1301 total active area of about 198 m^2 .

1302 2.3.4 Electromagnetic calorimeter

1303 The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of
 1304 electrons and photons. It is composed of 75848 lead tungstate crystals which have a
 1305 short radiation length (0.89 cm) and fast response, since 80% of the light is emitted
 1306 within 25 ns; however, they are combined with Avalanche photodiodes (APDs) as

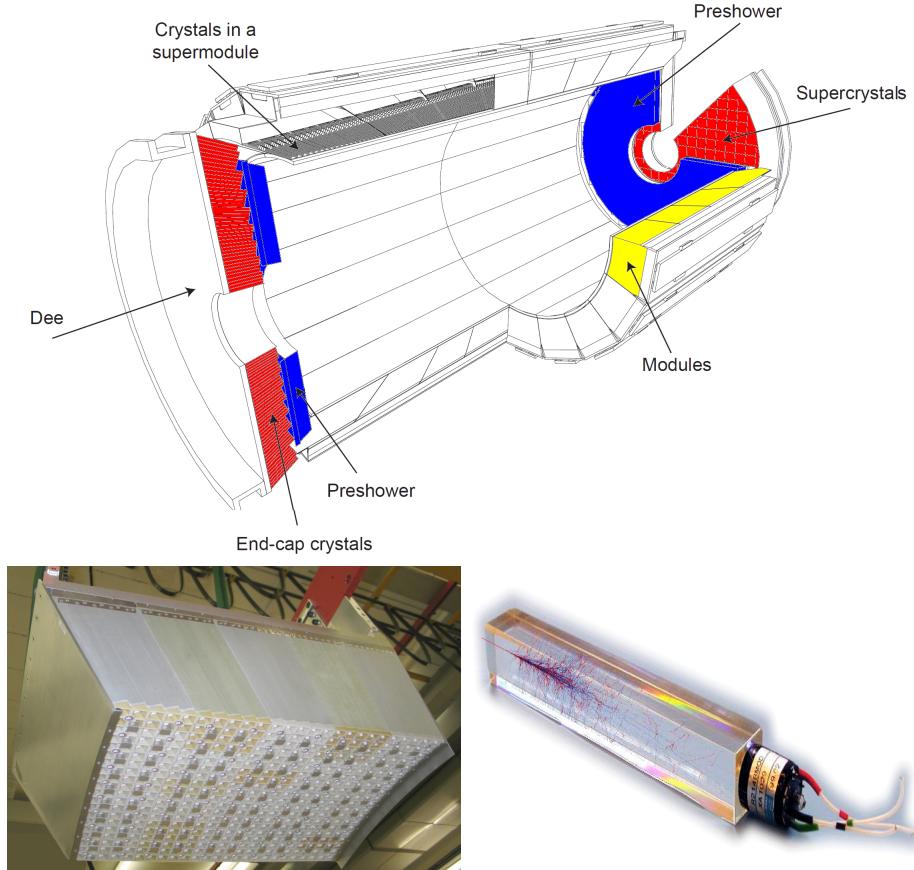


Figure 2.15: Top: CMS ECAL schematic view. Bottom: Module equipped with the crystals (left); ECAL crystal(right) with an artistic representation of an electromagnetic shower [74].

1307 photodetectors given that crystals themselves have a low light yield ($30\gamma/\text{MeV}$). A
 1308 schematic view of the ECAL is shown in Figure 2.15.

1309 Energy is measured when electrons and photons are absorbed by the crystals
 1310 which generates an electromagnetic *shower*, as seen in bottom right picture of the
 1311 Figure 2.15; the shower is seen as a *cluster* of energy which depending on the amount
 1312 of energy deposited can involve several crystals. The ECAL barrel (EB) covers the
 1313 region $|\eta| < 1.479$, using crystals of depth of 23 cm and $2.2 \times 2.2 \text{ cm}^2$ transverse
 1314 section; the ECAL endcap (EE) covers the region $1.479 < |\eta| < 3.0$ using crystals of
 1315 depth 22 cm and transverse section of $2.86 \times 2.86 \text{ cm}^2$; the photodetectors used are

1316 vacuum phototriodes (VPTs). Each EE is divided in two structures called *Dees*.

1317 The preshower detector (ES) is installed in front of the EE and covers the region
 1318 $1.653 < |\eta| < 2.6$. The ES provides a precise measurement of the position of electro-
 1319 magnetic showers, which allows to distinguish electrons and photon signals from π^0
 1320 decay signals. The ES is composed of a layer of lead radiators followed by a layer of
 1321 silicon strip sensors. The lead radiators initiate electromagnetic showers when reached
 1322 by photons and electrons, then, the strip sensors measure the deposited energy and
 1323 the transverse shower profiles. The full ES thickness is 20 cm.

1324 2.3.5 Hadronic calorimeter

1325 Hadrons are not absorbed by the ECAL² but by the hadron calorimeter (HCAL),
 1326 which is made of a combination of alternating brass absorber layers and silicon photo-
 1327 multiplier(SiPM) layers; therefore, particles passing through the scintillator material
 1328 produce showers, as in the ECAL, as a result of the inelastic scattering of the hadrons
 1329 with the detector material. Since the particles are not absorbed in the scintillator,
 1330 their energy is sampled; therefore the total energy is not measured but estimated from
 1331 the energy clusters, which reduces the resolution of the detector. Brass was chosen
 1332 as the absorber material due to its short interaction length ($\lambda_I = 16.42\text{cm}$) and its
 1333 non-magnetivity. Figure 2.16 shows a schematic view of the CMS HCAL.

1334 The HCAL is divided into four sections; the Hadron Barrel (HB), the Hadron
 1335 Outer (HO), the Hadron Endcap (HE) and the Hadron Forward (HF) sections. The
 1336 HB covers the region $0 < |\eta| < 1.4$, while the HE covers the region $1.3 < |\eta| < 3.0$.
 1337 The HF, made of quartz fiber scintillator and steel as absorption material, covers the
 1338 forward region $3.0 < |\eta| < 5.2$. Both the HB and HF are located inside the solenoid.
 1339 The HO is placed outside the magnet as an additional layer of scintillators with the

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

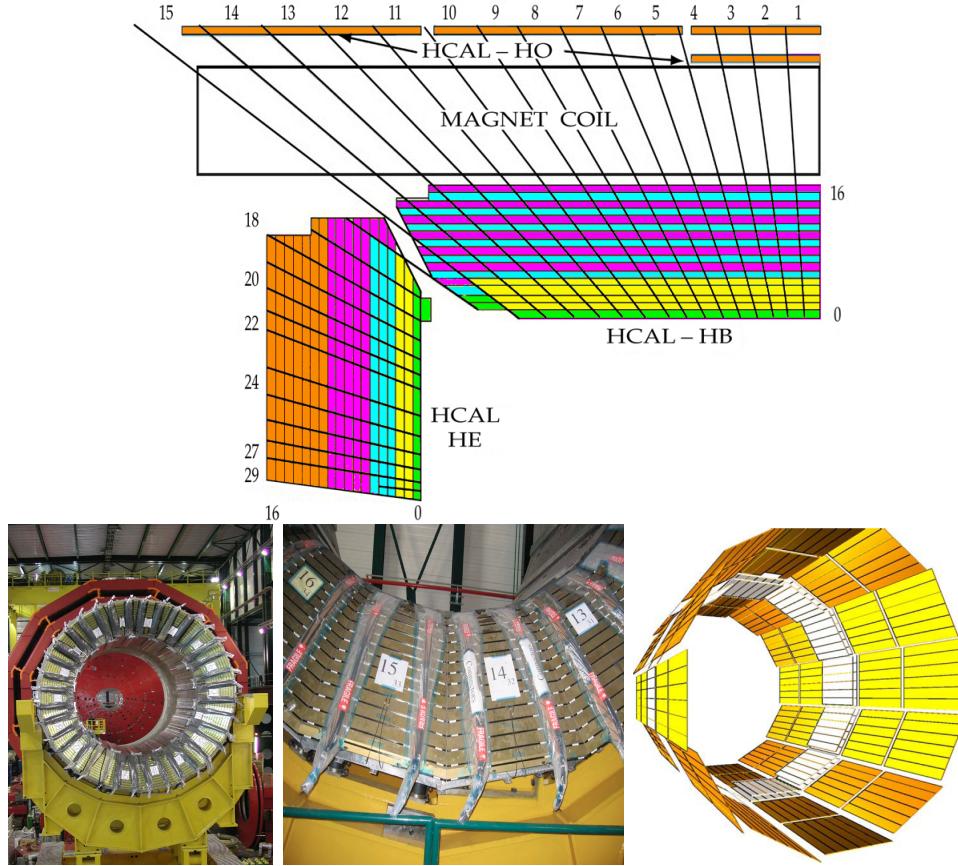


Figure 2.16: Top: CMS HCAL schematic view, the colors indicate the layers that are grouped into the same readout channels. Bottom: picture of a section of the HB; the absorber material is the golden region and scintillators are placed in between the absorber material (left and center). Schematic view of the HO (right). [83,84]

1340 purpose of measure the energy tails of particles passing through the HB and the
 1341 magnet (see Figure 2.16 top and bottom right).

1342 **2.3.6 Superconducting solenoid magnet**

1343 The superconducting magnet installed in the CMS detector is designed to provide
 1344 an intense and highly uniform magnetic field in the central part of the detector.
 1345 In fact, the tracking system takes advantage of the bending power of the magnetic
 1346 field to measure with precision the momentum of the particles that traverse it; the

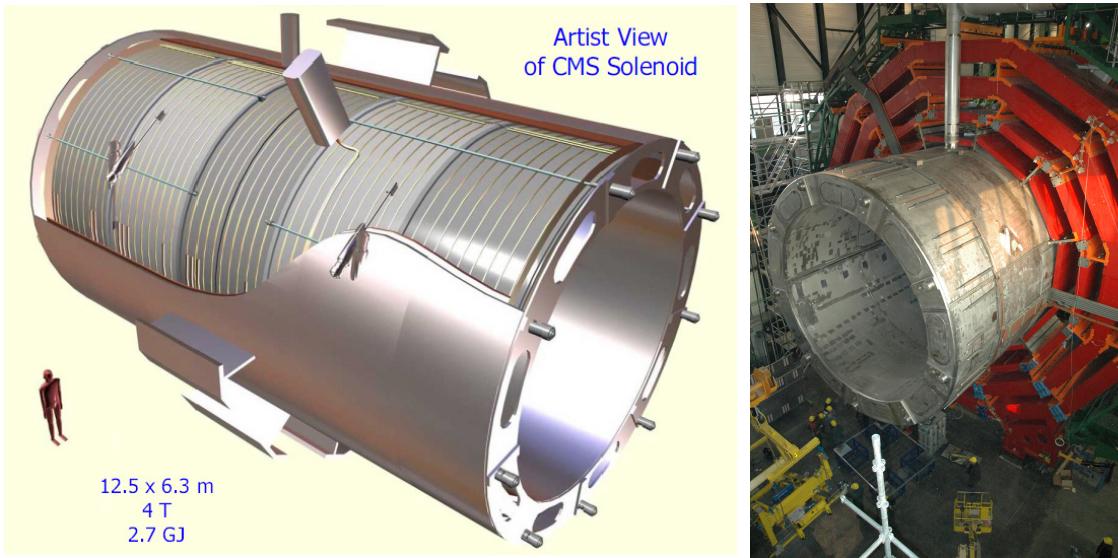


Figure 2.17: Artistic representation of the CMS solenoid magnet(left). The magnet is supported on an iron yoke (right) which also serves as the house of the muon detector and as mechanical support for the whole CMS detector [77].

1347 unambiguous determination of the sign for high momentum muons was a driving
 1348 principle during the design of the magnet. The magnet has a diameter of 6.3 m, a
 1349 length of 12.5 m and a cold mass of 220 t; the generated magnetic field reaches a
 1350 strength of 3.8T. Since it is made of Ni-Tb superconducting cable it has to operate at
 1351 a temperature of 4.7 K by using a helium cryogenic system; the current circulating in
 1352 the cables reaches 18800 A under normal running conditions. The left side of Figure
 1353 2.17 shows an artistic view of the CMS magnet, while the right side shows a transverse
 1354 view of the cold mass where the winding structure is visible.

1355 The yoke (see Figure 2.17), composed of 5 barrel wheels and 6 endcap disks made
 1356 of iron, serves not only as the media for magnetic flux return but also provides housing
 1357 for the muon detector system and structural stability to the full detector.

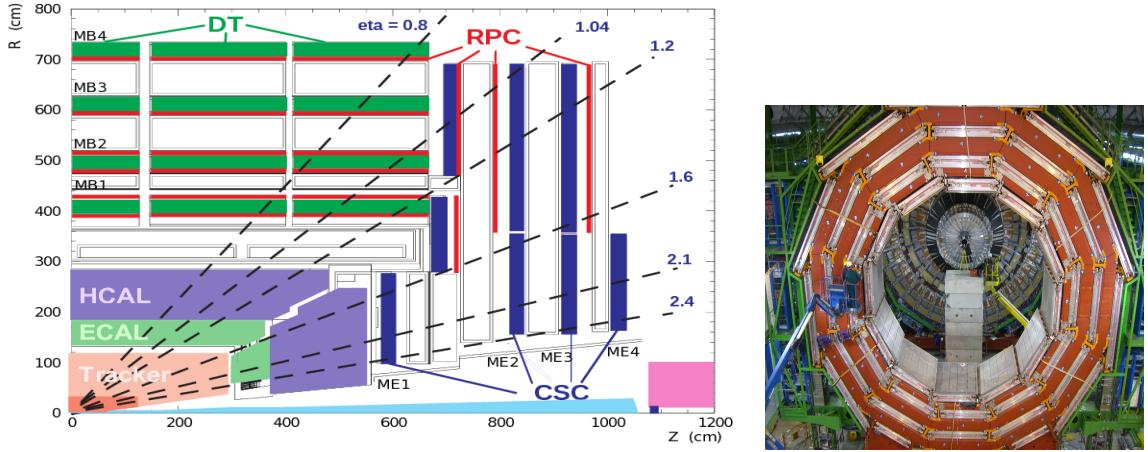


Figure 2.18: Left: CMS muon system schematic view; Right: one of the yoke rings with the muon DTs and RPCs installed; in the back it is possible to see the muon endcap [85].

1358 2.3.7 Muon system

1359 Muons are the only charged particles able to pass through all the CMS detector due
 1360 to their low ionization energy loss; thus, muons can be separated easily from the
 1361 high amount of particles produced in a pp collision. Also, muons are expected to be
 1362 produced in the decay of several new particles; therefore, good detection of muons
 1363 was one of the leading principles when designing the CMS detector.

1364 The CMS muon detection system (muon spectrometer) is embedded in the return
 1365 yoke as seen in Figure 2.18. It is composed of three different detector types, the drift
 1366 tube chambers (DT), cathode strip chambers (CSC), and resistive plate chambers
 1367 (RPC); DT are located in the central region $\eta < 1.2$ arranged in four layers of drift
 1368 chambers filled with an Ar/CO₂ gas mixture.

1369 The muon endcaps are made of CSCs covering the region $\eta < 2.4$ and filled with
 1370 a mixture of Ar/CO₂/CF₄. The reason behind using a different detector type lies on
 1371 the different conditions in the forward region like the higher muon rate and higher
 1372 residual magnetic field compared to the central region.

1373 The third type of detector used in the muon system is a set of four disks of RPCs

1374 working in avalanche mode. The RPCs provide good spatial and time resolutions. The
 1375 track of high- p_T muon candidates is built combining information from the tracking
 1376 system and the signal from up to six RPCs and four DT chambers.

1377 The muon tracks are reconstructed from the hits in the several layers of the muon
 1378 system.

1379 **2.3.8 CMS trigger system**

1380 CMS expects pp collisions every 25 ns, i.e., an interaction rate of 40 MHz for which
 1381 it is not possible to store the recorded data in full. In order to handle this high event
 1382 rate data, an online event selection, known as triggering, is performed; triggering
 1383 reduces the event rate to 100 Hz for storage and further offline analysis.

1384 The trigger system starts with a reduction of the event rate to 100 kHz in the
 1385 so-called *the level 1 trigger* (L1). L1 is based on dedicated programmable hardware
 1386 like Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated
 1387 Circuits (ASICs), partly located in the detector itself; another portion is located in
 1388 the CMS underground cavern. Hit pattern information from the muon chambers
 1389 and the energy deposits in the calorimeter are used to decide if an event is accepted
 1390 or rejected, according to selection requirements previously defined, which reflect the
 1391 interesting physics processes. Figure 2.19 shows the L1 trigger architecture.

1392 The second stage in the trigger system is called *the high-level trigger* (HLT); events
 1393 accepted by L1 are passed to HLT in order to make an initial reconstruction of them.
 1394 HLT is software based and runs on a dedicated server farm, using selection algorithms
 1395 and high-level object definitions; the event rate at HLT is reduced to 100 Hz. The
 1396 first HLT stage takes information from the muon detectors and the calorimeters to
 1397 make the initial object reconstruction; in the next HLT stage, information from the

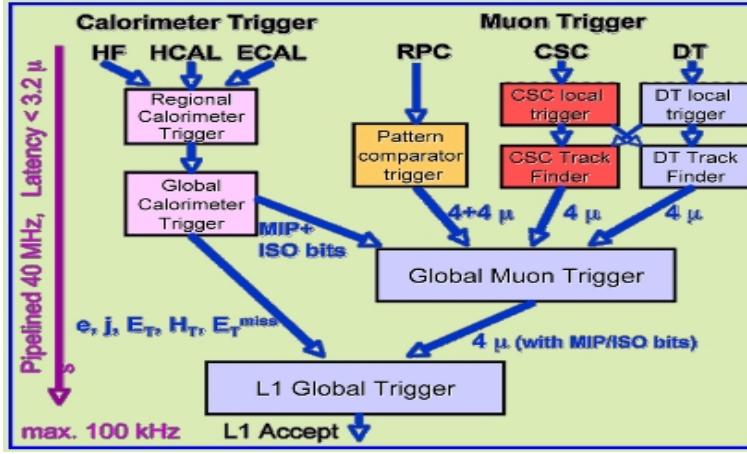


Figure 2.19: CMS Level-1 trigger architecture [86].

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

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

2.3.9 CMS computing

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

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

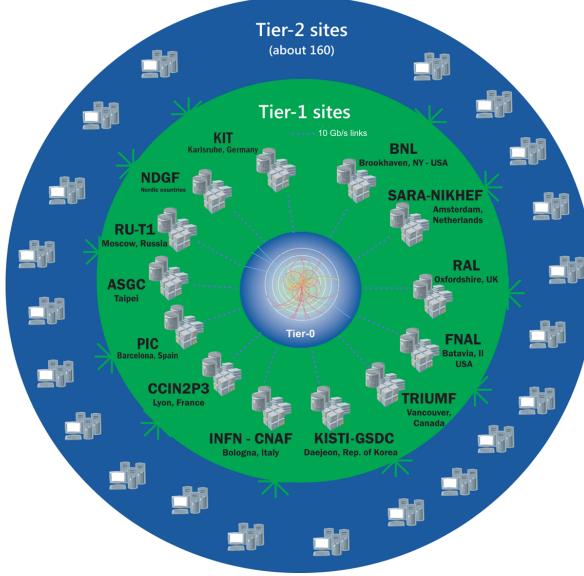


Figure 2.20: WLCG structure. The primary computer centers (Tier-0) are located at CERN (data center) and at the Wigner datacenter in Budapest. Tier-1 is composed of 13 centers and Tier-2 is composed of about 160 centers. [87].

1414 that distributed environment. WLCG is a tiered structure connecting computing
 1415 centers around the world, which provides the necessary storage and computing facil-
 1416 ties. The primary computing centers of the WLCG are located at the CERN and
 1417 the Wigner datacenter in Budapest and are known as Tier-0 as shown in Figure 2.20.
 1418 The main responsibilities for each tier level are [87]

- 1419 • **Tier-0:** initial reconstruction of recorded events and storage of the resulting
 1420 datasets, the distribution of raw data to the Tier-1 centers.
- 1421 • **Tier-1:** provide storage capacity, support for the Grid, safe-keeping of a pro-
 1422 portional share of raw and reconstructed data, large-scale reprocessing and safe-
 1423 keeping of corresponding output, generation of simulated events, distribution
 1424 of data to Tier 2s, safe-keeping of a share of simulated data produced at these
 1425 Tier 2s.
- 1426 • **Tier-2:** store sufficient data and provide adequate computing power for spe-

1427 cific analysis tasks and proportional share of simulated event production and
1428 reconstruction.

1429 Aside from the general computing strategy to manage the huge amount of data
1430 produced by experiments, CMS uses a software framework to perform a variety of
1431 processing, selection and analysis tasks. The central concept of the CMS data model
1432 referred to as *event data model* (EDM) is the *Event*; therefore, an event is the unit
1433 that contains the information from a single bunch crossing, any data derived from
1434 that information like the reconstructed objects, and the details of the derivation.

1435 Events are passed as the input to the *physics modules* that obtain information
1436 from them and create new information; for instance, *event data producers* add new
1437 data into the events, *analyzers* produce an information summary from an event set,
1438 *filters* perform selection and triggering.

1439 CMS uses several event formats with different levels of detail and precision

1440 • **Raw format:** events in this format contain the full recorded information from
1441 the detector as well as trigger decision and other metadata. An extended version
1442 of raw data is used to store information from the CMS Monte Carlo simulation
1443 tools (see Chapter 3). Raw data are stored permanently, occupying about
1444 2MB/event

1445 • **RECO format:** events in this format correspond to raw data that have been
1446 submitted to reconstruction algorithms like primary and secondary vertex re-
1447 construction, particle ID, and track finding. RECO events contain physics ob-
1448 jects and all the information used to reconstruct them; average size is about 0.5
1449 MB/event.

- 1450 • **AOD format:** Analysis Object Data (AOD) is the data format used in the
 1451 physics analyses given that it contains the parameters describing the high-level
 1452 physics objects in addition to enough information to allow a kinematic refitting if
 1453 needed. AOD events are filtered versions of the RECO events to which skimming
 1454 or other filtering have been applied, hence AOD events are subsets of RECO
 1455 events. Requires about 100 kB/event.
- 1456 • **Non-event data** are data needed to interpret and reconstruct events. Some
 1457 of the non-event data used by CMS contains information about the detector
 1458 contraction and condition data like calibrations, alignment, and detector status.

1459 Figure 2.21 shows the data flow scheme between CMS detector and tiers.

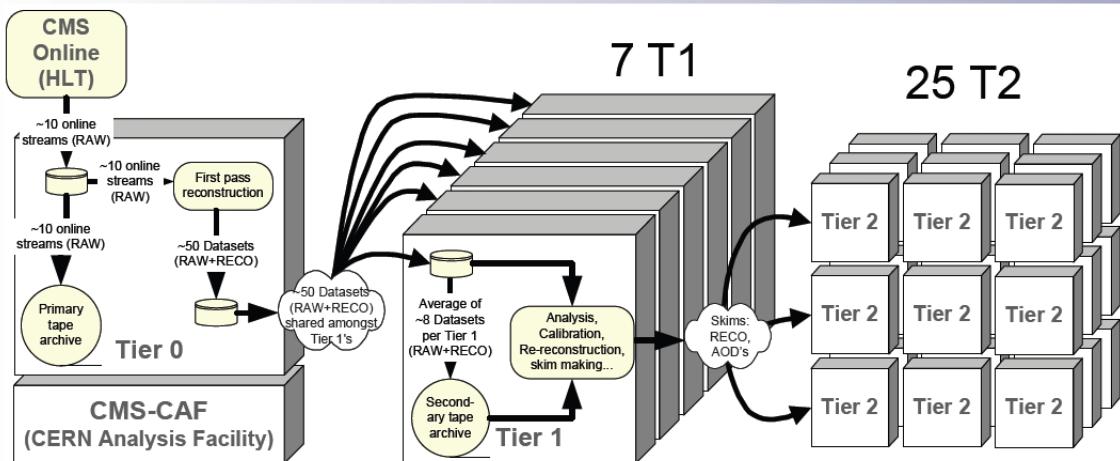


Figure 2.21: Data flow from CMS detector through tiers.

1460 The whole collection of software built as a framework is referred to as *CMSSW*. This
 1461 framework provides the services needed by the simulation, calibration and alignment,
 1462 and reconstruction modules that process event data, so that physicists can perform
 1463 analysis. The CMSSW event processing model is composed of one executable, called
 1464 `cmsRun`, and several plug-in modules which contains all the tools (calibration, recon-

1465 struction algorithms) needed to process an event. The same executable is used for
1466 both detector data and Monte Carlo simulations [88].

¹⁴⁶⁷ **Chapter 3**

¹⁴⁶⁸ **Event generation, simulation and
¹⁴⁶⁹ reconstruction**

¹⁴⁷⁰ The process of analyzing data recorded by the CMS experiment involves several stages
¹⁴⁷¹ where the data are processed in order to interpret the information provided by all
¹⁴⁷² the detection systems; in those stages, the particles produced after the pp collision
¹⁴⁷³ are identified by reconstructing their trajectories and measuring their features. In
¹⁴⁷⁴ addition, the SM provides a set of predictions that have to be compared with the
¹⁴⁷⁵ experimental results; however, in most of the cases, theoretical predictions are not
¹⁴⁷⁶ directly comparable to experimental results due to the diverse source of uncertainties
¹⁴⁷⁷ introduced by the experimental setup and theoretical approximations, among others.

¹⁴⁷⁸ The strategy to face these conditions consists in using statistical methods imple-
¹⁴⁷⁹ mented in computational algorithms to produce numerical results that can be con-
¹⁴⁸⁰ trasted with the experimental results. These computational algorithms are commonly
¹⁴⁸¹ known as Monte Carlo (MC) methods and, in the case of particle physics, they are
¹⁴⁸² designed to apply the SM rules and produce predictions about the physical observ-
¹⁴⁸³ ables measured in the experiments. Since particle physics is governed by quantum

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

This chapter presents a description of the event generation strategy and the tools used to perform the detector simulation and physics objects reconstruction. A comprehensive review of event generators for LHC physics can be found in Reference [89] on which this chapter is based.

3.1 Event generation

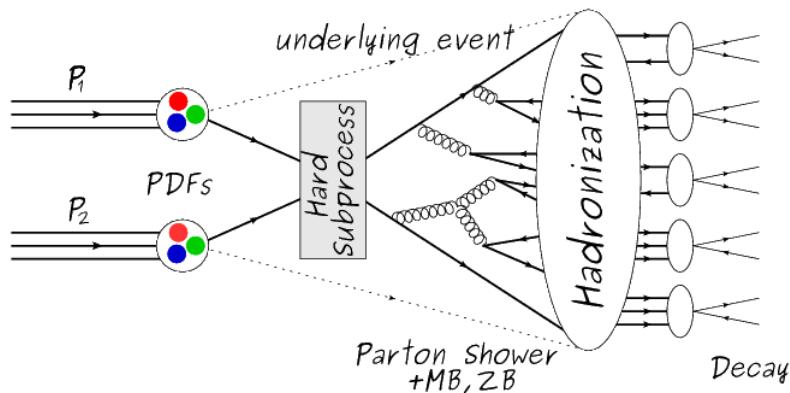


Figure 3.1: Event generation process. The actual interaction is generated in the hard subprocess. The parton shower describes the evolution of the partons from the hard subprocess. Modified from Reference [90].

The event generation is intended to create events that mimic the behavior of actual events produced in collisions; they obey a sequence of steps from the particles collision hard process to the decay process into the final state. Figure 3.1 shows a schematic view of the event generation process; the fact that the full process can be treated as several independent steps is motivated by the QCD factorization theorem.

1498 Generation starts by taking into account the PDFs of the incoming particles.
 1499 Event generators offer the option to chose from several PDF sets depending on the
 1500 particular process under simulation¹; in the following, pp collisions will be consid-
 1501 ered. The *hard subprocess* describes the actual interaction between partons from the
 1502 incoming protons; it is represented by the matrix element connecting the initial and
 1503 final states of the interaction. Normally, the matrix element can be written as a
 1504 sum over Feynman diagrams and consider interferences between terms in the sum-
 1505 mation. During the generation of the hard subprocess, the production cross section
 1506 is calculated.

1507 The order to which the cross section is calculated depends on the order of the Feyn-
 1508 man diagrams involved in the calculation; therefore, radiative corrections are included
 1509 by considering a higher order Feynman diagrams where QCD radiation dominates.
 1510 Currently, cross sections calculated to LO do not offer a satisfactory description of the
 1511 processes, i.e., the results are only reliable for the shape of distributions; therefore,
 1512 NLO calculations have to be performed with the implication that the computing time
 1513 needed is highly increased.

1514 The final parton content of the hard subprocess is subjected to the *parton shower*
 1515 which generates the gluon radiation. Parton shower evolves the partons, i.e., glouns
 1516 split into quark-antiquark pairs and quarks with enough energy radiate gluons giv-
 1517 ing rise to further parton multiplication, following the DGLAP (Dokshitzer-Gribov-
 1518 Lipatov-Altarelli-Parisi) equations. Showering continues until the energy scale is low
 1519 enough to reach the non-perturbative limit.

1520 In the simulation of LHC processes that involve b quarks, like the single top quark
 1521 or Higgs associated production, it is needed to consider that the b quark is heavier
 1522 than the proton; hence, the QCD interaction description is made in two different

¹ Tool in Reference [91] allows to plot different PDF sets under customizable conditions.

1523 schemes [95]

- 1524 • four-flavor (4F) scheme. b quarks appear only in the final state because they
 1525 are heavier than the proton and therefore they can be produced only from the
 1526 splitting of a gluon into pairs or singly in association with a t quark in high
 1527 energy-scale interactions; furthermore, during the simulation, the b -PDFs are set
 1528 to zero. Calculations in this scheme are more complicated due to the presence
 1529 of the second b quark but the full kinematics is considered already at LO and
 1530 therefore the accuracy of the description is better.

- 1531 • five-flavor (5F) scheme. b quarks are considered massless, therefore they can
 1532 appear in both initial and final states since they can now be part of the proton;
 1533 thus, during the simulation b -PDFs are not set to zero. In this scheme, calcu-
 1534 lations are simpler than in the 4F scheme and possible logarithmic divergences
 1535 are absorbed by the PDFs through the DGLAP evolution.

1536 In this thesis, the tHq events are generated using the 4F scheme in order to reduce
 1537 uncertainties, while the tHW events are generated using the 5F scheme to eliminate
 1538 LO interference with $t\bar{t}H$ process [48].

1539 Partons involved in the pp collision are the focus of the simulation, however, the
 1540 rest of the partons inside the incoming protons are also affected because the remnants
 1541 are colored objects; also, multiple parton interactions can occur. The hadronization
 1542 of the remnants and multiple parton interactions are known as *underlying event* and
 1543 it has to be included in the simulation. In addition, multiple pp collisions in the same
 1544 bunch crossing (pile-up mentioned in 2.2) occurs, actually in two forms

- 1545 • *in-time PU* which refers to multiple pp collision in the bunch crossing but that
 1546 are not considered as primary vertices.

1547 • *Out-of-time PU* which refers to overlapping pp collisions from consecutive bunch
 1548 crossings; this can occur due to the time-delays in the detection systems where
 1549 information from one bunch crossing is assigned to the next or previous one.

1550 While the underlying event effects are included in generation using generator-
 1551 specific tools, PU effects are added to the generation by overlaying Minimum-bias (MB)
 1552 and Zero-bias (ZB) events to the generated events. MB events are inelastic events
 1553 selected by using a loose trigger with as little bias as possible, therefore accepting a
 1554 large fraction of the overall inelastic event; ZB events correspond to random events
 1555 recorded by the detector when collisions are likely. MB models in-time PU and ZB
 1556 models out-of-time PU.

1557 The next step in the generation process is called *hadronization*. Since particles
 1558 with a net color charge are not allowed to exits isolated, they have to recombine
 1559 to form bound states. This is precisely the process by which the partons resulting
 1560 from the parton shower arrange themselves as color singlets to form hadrons. At
 1561 this step, the energy-scale is low and the strong coupling constant is large, therefore
 1562 hadronization process is non-perturbative and the evolution of the partons is described
 1563 using phenomenological models. Most of the baryons and mesons produced in the
 1564 hadronization are unstable and hence they will decay in the detector.

1565 The last step in the generation process corresponds to the decay of the unstable
 1566 particles generated during hadronization; it is also simulated in the hadronization
 1567 step, based on the known branching ratios.

1568 **3.2 Monte Carlo Event Generators.**

1569 The event generation described in the previous section has been implemented in
 1570 several software packages for which a brief description is given.

- 1571 • **PYTHIA 8.** It is a program designed to perform the generation of high energy
 1572 physics events which describes the collisions between particles such as electrons
 1573 and protons. Several theories and models are implemented in it, in order to
 1574 describe physical aspects like hard and soft interaction, parton distributions,
 1575 initial and final-state parton showers, multiple parton interactions, beam rem-
 1576 nants, hadronization² and particle decay. Thanks to extensive testing, several
 1577 optimized parametrizations, known as *tunings*, have been defined in order to
 1578 improve the description of actual collisions to a high degree of precision; for
 1579 analysis at $\sqrt{s} = 13$ TeV, the underline event CUETP8M1 tune is employed [97].
 1580 The calculation of the matrix element is performed at LO which is not enough
 1581 for the current required level of precision; therefore, pythia is often used for
 1582 parton shower, hadronization and decays, while other event generators are used
 1583 to generate the matrix element at NLO.
- 1584 • **MadGraph5_aMC@NLO.** MadGraph is a matrix element generator which
 1585 calculates the amplitudes for all contributing Feynman diagrams of a given
 1586 process but does not provide a parton shower while MC@NLO incorporates
 1587 NLO QCD matrix elements consistently into a parton shower framework; thus,
 1588 MadGraph5_aMC@NLO, as a merger of the two event generators MadGraph5
 1589 and aMC@NLO, is an event generator capable to calculate tree-level and NLO
 1590 cross sections and perform the matching of those with the parton shower. It is
 1591 one of the most frequently used matrix element generators; however, it has the
 1592 particular feature of the presence of negative event weights which reduce the
 1593 number of events used to reproduce the properties of the objects generated [98].
- 1594 • **POWHEG.** It is an NLO matrix element generator where the hardest emis-

² based in the Lund string model [96]

1595 sion of color charged particles is generated in such a way that the negative event
 1596 weights issue of MadGraph5_aMC@NLO is overcome; however, the method re-
 1597 quires an interface with p_T -ordered parton shower or a parton shower generator
 1598 where this highest emission can be vetoed in order to avoid double counting of
 1599 this highest-energetic emission. PYTHIA is a commonly matched to POWHEG
 1600 event generator [100].

1601 Events resulting from the whole generation process are known as MC events.

1602 **3.3 CMS detector simulation.**

1603 After generation, MC events contain the physics of the collisions but they are not
 1604 ready to be compared to the events recorded by the experiment since these recorded
 1605 events correspond to the response of the detection systems to the interaction with
 1606 the particles traversing them. The simulation of the CMS detector has to be applied
 1607 on top of the event generation; it is simulated with a MC toolkit for the simulation
 1608 of particles passing through matter called Geant4 which is also able to simulate the
 1609 electronic signals that would be measured by all detectors inside CMS.

1610 The simulation takes the generated particles contained in the MC events as input,
 1611 makes them pass through the simulated geometry, and models physics processes that
 1612 particles experience during their passage through matter. The full set of results from
 1613 particle-matter interactions corresponds to the simulated hit which contains informa-
 1614 tion about the energy loss, momentum and position. Particles of the input event are
 1615 called *primary*, while the particles originating from GEANT4-modeled interactions of
 1616 a primary particle with matter are called a *secondary*. Simulated hits are the input
 1617 of subsequent modules that emulate the response of the detector readout system and

1618 triggers. The output from the emulated detection systems and triggers is known as
 1619 digitization [101, 102].

1620 The modeling of the CMS detector corresponds to the accurate modeling of the
 1621 interaction among particles, the detector material, and the magnetic field. This
 1622 simulation procedure includes the following standard steps

1623 • Modeling of the Interaction Region.

1624 • Modeling of the particle passage through the hierarchy of volumes that compose
 1625 CMS detector and of the accompanying physics processes.

1626 • Modeling of the effect of multiple interactions per beam crossing and/or the
 1627 effect of events overlay (Pile-Up simulation).

1628 • Modeling of the detector's electronics response, signal shape, noise, calibration
 1629 constants (digitization).

1630 In addition to the full simulation, i.e., a detailed detector simulation, a faster
 1631 simulation (FastSim) have been developed, that may be used where much larger
 1632 statistics are required. In FastSim, detector material effects are parametrized and
 1633 included in the hits; those hits are used as input of the same higher-level algorithms³
 1634 used to analyze the recorded events. In this way, comparisons between fast and full
 1635 simulations can be performed [104].

1636 After the full detector simulation, the output events can be directly compared
 1637 to events actually recorded in the CMS detector. The collection of MC events that
 1638 reproduces the expected physics for a given process is known as MC sample.

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

1639 **3.4 Event reconstruction.**

1640 The CMS experiment use the *particle-flow event reconstruction algorithm (PF)* to do
1641 the reconstruction of particles produced in pp collisions. Next sections will present
1642 a basic description of the *Elements* used by PF (tracker tracks, energy clusters, and
1643 muon tracks), based in the References [105, 106] where more detailed descriptions can
1644 be found.

1645 **3.4.1 Particle-Flow Algorithm.**

1646 Each of the several sub detection systems of the CMS detector is dedicated to identify
1647 an specific type of particles, i.e., photons and electrons are absorbed by the ECAL
1648 and their reconstruction is based on ECAL information; hadrons are reconstructed
1649 from clusters in the HCAL while muons are reconstructed from hits in the muon
1650 chambers. PF is designed to correlate signals from all the detector layers (tracks and
1651 energy clusters) in order to reconstruct and identify each final state particle and its
1652 properties as sketched in Figure 3.2.

1653 For instance, a charged hadron is identified by a geometrical connection, known
1654 as *link*, between one or more calorimeter clusters and a track in the tracker, provided
1655 there are no hits in the muon system; combining several measurements allows a better
1656 determination of the energy and charge sign of the charged hadron.

1657 **Charged-particle track reconstruction.**

1658 The strategy used by PF in order to reconstruct tracks is called *Iterative Tracking*
1659 which occurs in four steps

- 1660 • Seed generation where initial track candidates are found by looking for a combi-
1661 nation of hits in the pixel detector, strip tracker, and muon chambers. In total

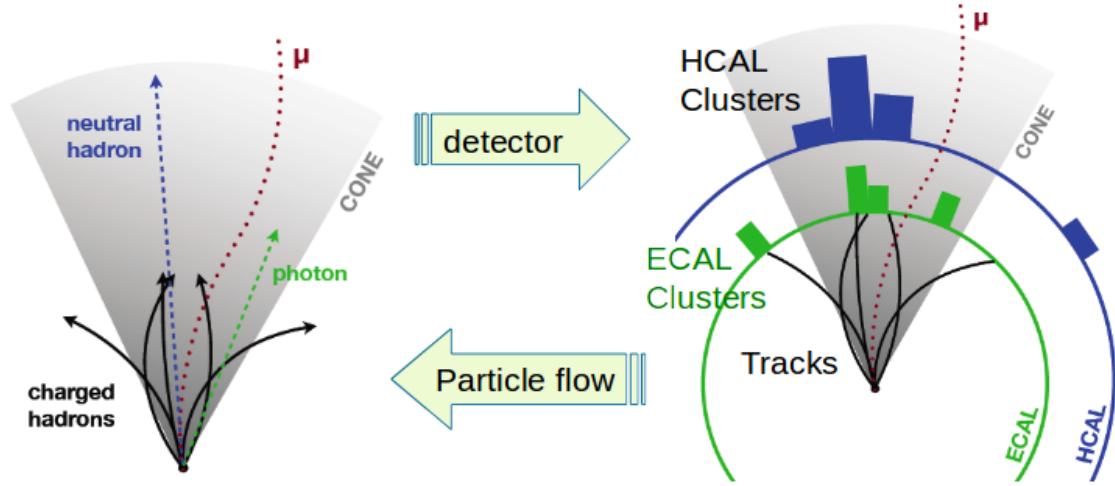


Figure 3.2: Particle flow algorithm. Information from the several CMS detection systems is provided as input to the algorithm which then combines it to identify and reconstruct all the particles in the final state and their properties. Reconstruction of simulated events is also performed by providing information from MC samples, detector and trigger simulation [107].

1662 ten iterations are performed, each one with a different seeding requirement.
 1663 Seeds are used to estimate the trajectory parameters and uncertainties at the
 1664 time of the full track reconstruction. Seeds are also considered track candidates.

- 1665 ● Track finding using a tracking software known as Combinatorial Track Finder
 1666 (CTF) [108]. The seed trajectories are extrapolated along the expected flight
 1667 path of a charged particle, in agreement to the trajectory parameters obtained
 1668 in the first step, in an attempt to find additional hits that can be assigned to
 1669 the track candidates.

- 1670 ● Track-fitting where the found tracks are passed as input to a module which
 1671 provides the best estimate of the parameters of each trajectory.

- 1672 ● Track selection where track candidates are submitted to a selection which dis-
 1673 cards those that fail a set of defined quality criteria.

1674 Iterations differ in the seeding configuration and the final track selection as elab-

1675 orated in References [105, 106]. In the first iteration, high p_T tracks and tracks pro-
 1676 duced near to the interaction region are identified and those hits are masked thereby
 1677 reducing the combinatorial complexity. Next, iterations search for more complicated
 1678 tracks, like low p_T tracks and tracks from b hadron decays, which tend to be displaced
 1679 from the interaction region.

1680 **Vertex reconstruction.**

1681 During the track reconstruction, an extrapolation toward to the calorimeters is per-
 1682 formed in order to match energy deposits; that extrapolation is performed also toward
 1683 the beamline in order to find the origin of the track known as *vertex*. The vertex re-
 1684 construction is performed by selecting from the available reconstructed tracks, those
 1685 that are consistent with being originated in the interaction region where pp collisions
 1686 are produced. The selection involves a requirement on the number of tracker (pixel
 1687 and strip) hits and the goodness of the track fit.

1688 Selected tracks are clustered using a *deterministic annealing algorithm* (DA)⁴. A
 1689 set of candidate vertices and their associated tracks, resulting from the DA, are then
 1690 fitted with an *adaptive vertex fitter* (AVF) to produce the best estimate of the vertices
 1691 locations.

1692 The p_T of the tracks associated to a reconstructed vertex is added, squared and
 1693 used to organize the vertices; the vertex with the highest squared sum is designated
 1694 as the *primary vertex* (*PV*) while the rest are designated as PU vertices.

1695 **Calorimeter clustering.**

1696 After traversing the CMS tracker system, electrons, photons and hadrons deposit their
 1697 energy in the ECAL and HCAL cells. The PF clustering algorithm aims to provide

⁴ DA algorithm and AVF are described in detail in References [110, 111]

1698 a high detection efficiency even for low-energy particles and an efficient distinction
 1699 between close energy deposits. The clustering runs independently in the ECAL barrel
 1700 and endcaps, HCAL barrel and endcaps, and the two preshower layers, following two
 1701 steps

- 1702 • cells with an energy larger than a given seed threshold and larger than the energy
 1703 of the neighboring cells are identified as cluster seeds. The neighbor cells are
 1704 those that either share a side with the cluster seed candidate, or the eight closest
 1705 cells including cells that only share a corner with the seed candidate.
- 1706 • cells with at least a corner in common with a cell already in the cluster seed
 1707 and with an energy above a cell threshold are grouped into topological clusters.

1708 Clusters formed in this way are known as *particle-flow clusters*. With this cluster-
 1709 ing strategy, it is possible to detect and measure the energy and direction of photons
 1710 and neutral hadrons as well as differentiate these neutral particles from the charged
 1711 hadron energy deposits. In cases involving charged hadrons for which the track pa-
 1712 rameters are not determined accurately, for instance, low-quality and high- p_T tracks,
 1713 clustering helps in the energy measurements.

1714 **Electron track reconstruction.**

1715 Although the charged-particle track reconstruction described above works for elec-
 1716 trons, they lose a significant fraction of their energy via bremsstrahlung photon radi-
 1717 ation before reaching the ECAL; thus, the reconstruction performance depends on the
 1718 ability to measure also the radiated energy. The reconstruction strategy, in this case,
 1719 requires information from the tracking system and from the ECAL. Bremsstrahlung
 1720 photons are emitted at similar η values to that of the electron but at different values
 1721 of ϕ ; therefore, the radiated energy can be recovered by grouping ECAL clusters in a

1722 η window over a range of ϕ around the electron direction. The group is called ECAL
 1723 supercluster.

1724 Electron candidates from the track-seeding and ECAL super clustering are merged
 1725 into a single collection which is submitted to a full electron tracking fit with a
 1726 Gaussian-sum filter (GSF) [109]. The electron track and its associated ECAL su-
 1727 percluster form a *particle-flow electron*.

1728 **Muon track reconstruction.**

1729 Given that the CMS detector is equipped with a muon spectrometer capable to iden-
 1730 tify and measure the momentum of the muons traversing it, the muon reconstruction
 1731 is not specific to PF; therefore, three different muon types are defined

- 1732 • *Standalone muon.* A clustering on the DTs or CSCs hits is performed to form
 1733 track segments; those segments are used as seeds for the reconstruction in the
 1734 muon spectrometer. All DTs, CSCs, and RPCs hits along the muon trajectory
 1735 are combined and fitted to form the full track. The fitting output is called a
 1736 *standalone-muon track*.
- 1737 • *Tracker muon.* Each track in the inner tracker with p_T larger than 0.5 GeV and
 1738 a total momentum p larger than 2.5 GeV is extrapolated to the muon system.
 1739 A *tracker muon track* corresponds to a extrapolated track that matches at least
 1740 one muon segment.
- 1741 • *Global muon.* When tracks in the inner tracker (inner tracks) and standalone-
 1742 muon tracks are matched and turn out being compatibles, their hits are com-
 1743 bined and fitted to form a *global-muon track*.

1744 Global muons sharing the same inner track with tracker muons are merged into
 1745 a single candidate. PF muon identification uses the muon energy deposits in ECAL,
 1746 HCAL, and HO associated with the muon track to improve the muon identification.

1747 **Particle identification and reconstruction.**

1748 PF elements are connected by a linker algorithm that tests the connection between any
 1749 pair of elements; if they are found to be linked, a geometrical distance that quantifies
 1750 the quality of the link is assigned. Two elements may be linked indirectly through
 1751 common elements. Linked elements form *PF blocks* and each PF block may contain
 1752 elements originating in one or more particles. Links can be established between
 1753 tracks, between calorimeter clusters, and between tracks and calorimeter clusters.
 1754 The identification and reconstruction start with a PF block and proceed as follows

- 1755 • Muons. An *isolated global muon* is identified by evaluating the presence of
 1756 inner track and energy deposits close to the global muon track in the (η, ϕ)
 1757 plane, i.e., in a particular point of the global muon track, inner tracks and
 1758 energy deposits are sought within a radius of $\Delta R = 0.3$ (see eqn. 2.7) from the
 1759 muon track; if they exist and the p_T of the found track added to the E_T of the
 1760 found energy deposit does not exceed 10% of the muon p_T then the global muon
 1761 is an isolated global muon. This isolation condition is stringent enough to reject
 1762 hadrons misidentified as muons.

1763 *Non-isolated global muons* are identified using additional selection requirements
 1764 on the number of track segments in the muon system and energy deposits along
 1765 the muon track. Muons inside jets are identified with more stringent criteria
 1766 in isolation and momentum as described in Reference [112]. The PF elements
 1767 associated with an identified muon are masked from the PF block.

- 1768 ● Electrons are identified and reconstructed as described above plus some addi-
 1769 tional requirements on fourteen variables like the amount of energy radiated,
 1770 the distance between the extrapolated track position at the ECAL and the po-
 1771 sition of the associated ECAL supercluster, among others, which are combined
 1772 in an specialized multivariate analysis strategy that improves the electron iden-
 1773 tification. Tracks and clusters used to identify and reconstruct electrons are
 1774 masked in the PF block.
- 1775 ● Isolated photons are identified from ECAL superclusters with E_T larger than 10
 1776 GeV, for which the energy deposited at a distance of 0.15, from the supercluster
 1777 position on the (η, ϕ) plane, does not exceed 10% of the supercluster energy;
 1778 note that this is an isolation requirement. In addition, there must not be links
 1779 to tracks. Clusters involved in the identification and reconstruction are masked
 1780 in the PF block.
- 1781 ● Bremsstrahlung photons and prompt photons tend to convert to electron-positron
 1782 pairs inside the tracker, therefore, a dedicated finder algorithm is used to link
 1783 tracks that seem to originate from a photon conversion; in case those two tracks
 1784 are compatible with the direction of a bremsstrahlung photon, they are also
 1785 linked to the original electron track. Photon conversion tracks are also masked
 1786 in the PF block.
- 1787 ● The remaining elements in the PF block are used to identify hadrons. In the
 1788 region $|\eta| \leq 2.5$, neutral hadrons are identified with HCAL clusters not linked
 1789 to any track while photons from neutral pion decays are identified with ECAL
 1790 clusters without links to tracks. In the region $|\eta| > 2.5$ ECAL clusters linked to
 1791 HCAL clusters are identified with a charged or neutral hadron shower; ECAL
 1792 clusters with no links are identified with photons. HCAL clusters not used yet,

1793 are linked to one or more unlinked tracks and to an unlinked ECAL in order to
 1794 reconstruct charged-hadrons or a combination of photons and neutral hadrons
 1795 according to certain conditions on the calibrated calorimetric energy.

- 1796 • Charged-particle tracks may be liked together when they converge to a *sec-*
 1797 *ondary vertex (SV)* displaced from the IP where the PV and PU vertices are
 1798 reconstructed; at least three tracks are needed in that case, of which at most
 1799 one has to be an incoming track with hits in tracker region between a PV and
 1800 the SV.

1801 The linker algorithm, as well as the whole PF algorithm, has been validated and
 1802 commissioned; results from that validation are presented in the Reference [105].

1803 **Jet reconstruction.**

1804 Quarks and gluons may be produced in the pp collisions, therefore, their hadronization
 1805 will be seen in the detector as a shower of hadrons and their decay products in the
 1806 form of a *jet*. Two classes of clustering algorithms have been developed based in
 1807 their jet definition [113]:

- 1808 • Iterative cone algorithms (IC). Jets are defined in terms of circles of fixed radius
 1809 R in the $\eta\text{-}\phi$ plane, known as *stable cones*, for which the sum of the momenta
 1810 of all the particles within the cone points in the same direction as the center
 1811 of the circle. The seed of the iteration is the hardest non-isolated particle in
 1812 the event, then, the resulting momentum direction is assigned as the new cone
 1813 direction and a new iteration starts; iteration process stops when the cone if
 1814 found to be stable.

1815 • Sequential recombination algorithms. The distance between non-isolated par-
 1816 ticles is calculated; if that distance is below a threshold, these particles are
 1817 recombined into a new object. The sequence is repeated until the separation
 1818 between the recombined object and any other particle is above certain thresh-
 1819 old; the recombined object is called a jet and the algorithm starts again with
 1820 the remaining particles.

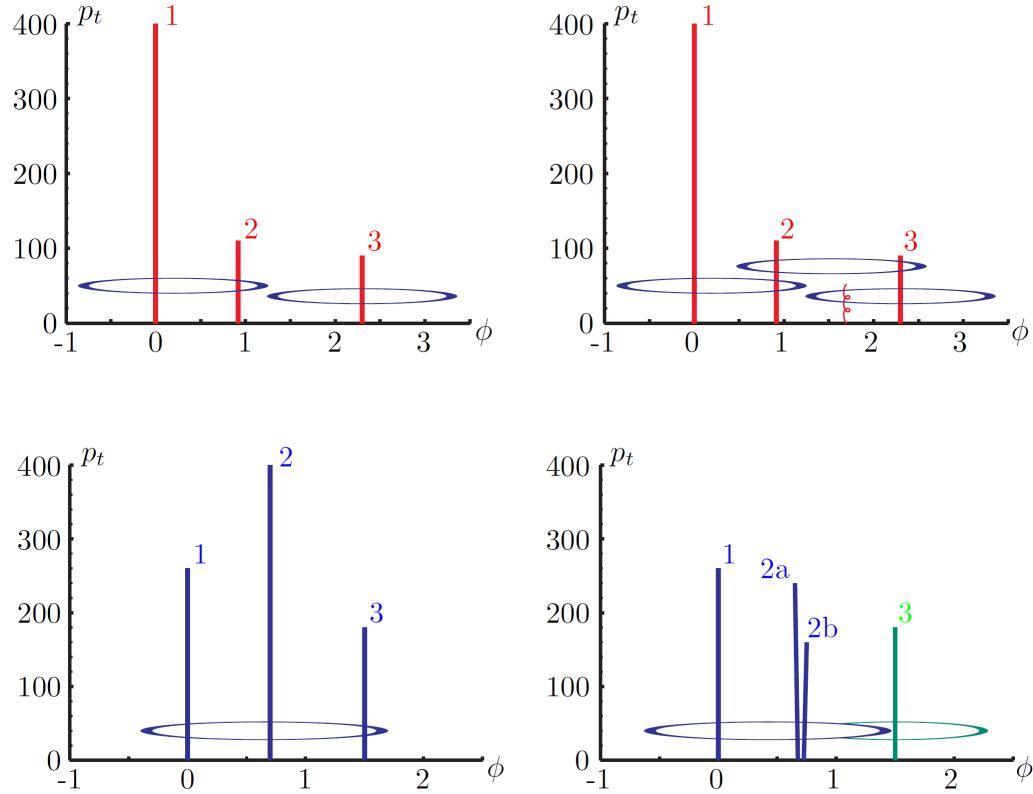


Figure 3.3: Stable cones identification using IC algorithms [113].

1821 Two conditions are of particular importance for the clustering algorithms, *infrared*
 1822 and *collinear (IRC) safety*. In order to explain the concept of infrared (IR) safety,
 1823 consider an event with three hard particles as shown in the top left side of Figure 3.3,
 1824 two stable cones are found and then two jets are identified; if a soft gluon is added, as
 1825 shown in the top right side of Figure 3.3, three stable cones are found and the three

1826 hard particles are now clustered into a single jet. If the addition of soft particles
 1827 change the outcome of the clustering, then it is said that the algorithm is IR unsafe.
 1828 Soft radiation is highly likely in perturbative QCD, which dominates the physics of
 1829 the jets, and then IR unsafe effect leads to divergences [113].

1830 The concept of collinear safety can also be explained considering a three hard
 1831 particles event, as shown in the bottom left side of Figure 3.3, where one stable cone
 1832 containing all three particles is found and one jet is identified; if the hardest particle
 1833 is split into two collinear particles (2a and 2b) in the bottom right side of Figure 3.3,
 1834 then the clustering results in a different jet identification and the algorithm is said
 1835 to be collinear unsafe. The collinear unsafe effect leads to divergences in jet cross
 1836 section calculations [114].

1837 It has been determined that IC algorithms are IRC unsafe, and therefore, they
 1838 have to be replaced by algorithms that not only provide the finite perturbative results
 1839 from theoretical computations, but also that are not highly dependent on underlying
 1840 event and pileup effects which leads to significant corrections [113].

1841 The sequential recombination algorithms arise as the IRC safe alternative used by
 1842 the CMS experiment; in particular the anti- k_t algorithm which is a generalization of
 1843 the previously existing k_t [115] and Cambridge/Aachen [116] jet clustering algorithms.

1844 The anti- k_t algorithm is used to perform the jet reconstruction by clustering those
 1845 PF particles within a cone (see Figure 3.4); previously, isolated electrons, isolated
 1846 muons, and charged particles associated with other interaction vertices are excluded
 1847 from the clustering.

1848 The anti- k_t algorithm proceeds in a sequential recombination of PF particles; the
 1849 distance between particles i and j (d_{ij}) and the distance between particles and the
 1850 beam are defined as

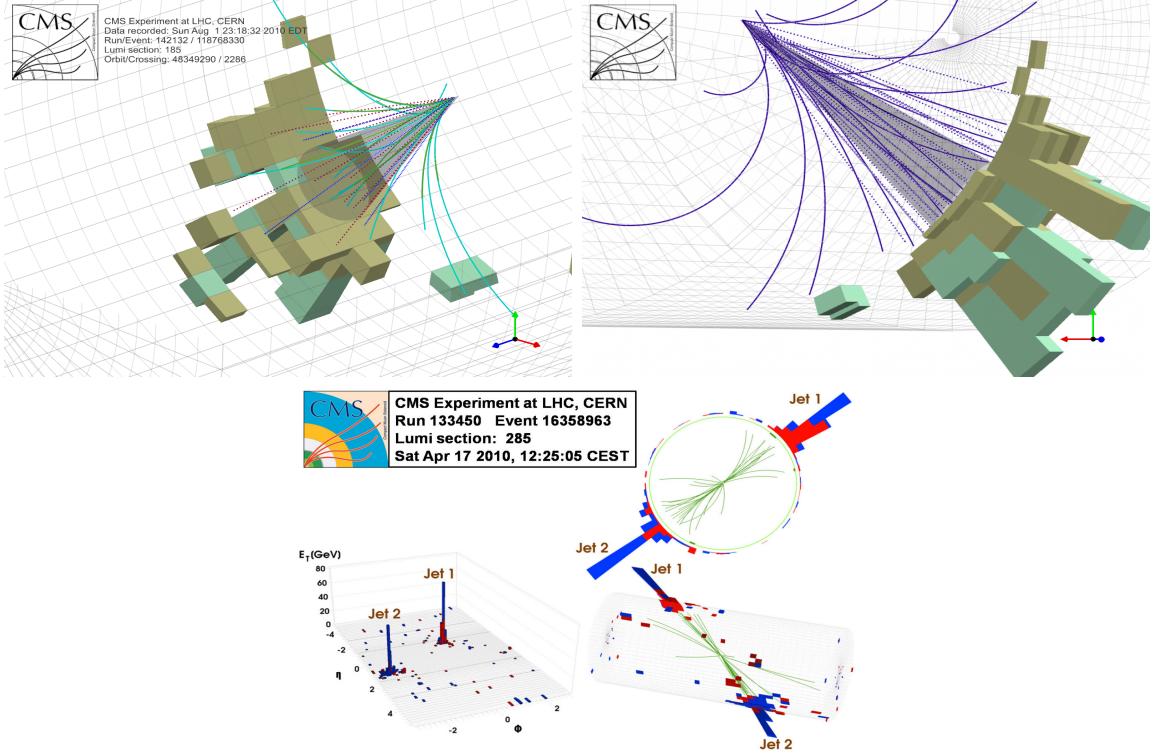


Figure 3.4: Jet reconstruction performed by the anti- k_t algorithm. Top: Two different views of a CMS recorded event are presented. Continuous lines correspond to tracks left by charged particles in the tracker while dotted lines are the imaginary paths followed by neutral particles. The green cubes represent the ECAL cells while the blue ones represent the HCAL cells; in both cases, the height of the cube represent the amount of energy deposited in the cells [117]. Bottom: Reconstruction of a recorded event with two jets [118].

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

$$d_{iB} = \frac{1}{k_{ti}^2} \quad (3.1)$$

1851 where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} , y_i and ϕ_i are the transverse momentum, ra-
 1852 pidity and azimuth of particle i respectively and R is the called jet radius. For all
 1853 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. R is a free parameter that can be
 adjusted according to the specific analysis conditions; usually, two values are used,
 $R=0.4$ and $R=0.5$, giving the name to the so-called AK4-jet and AK5-jet respectively.

An advantage of the anti- k_t algorithm over other clustering algorithms is the regularity of the boundaries of the resulting jets. For all known IRC safe algorithms,
 soft radiation can introduce irregularities in the boundaries of the final jets; however,
 anti- k_t algorithm is soft-resilient, meaning that jets shape is not affected by soft radiation,
 which is a valuable property considering that knowing the typical shape of jets
 makes experimental calibration of jets more simple. In addition, that soft-resilience
 is expected to simplify certain theoretical calculations and reduce the momentum-
 resolution loss caused by underlying-event (UE) and pileup contamination [114].

The effect of the UE and pileup contamination over a jet identification, can be
 seen as if soft events are added to the jet; for instance, if a soft event representing UE
 or pileup is added to an event for which a set of jets J have been identified, and the
 clustering is rerun on that new extended event, the outcome will be different in two
 aspects: jets will contain some additional soft energy and the distribution of particles
 in jets may have change; that effect is called *back-reaction*. The back-reaction effect in
 the anti- k_t algorithm is suppressed not by the amount of momentum added to the jet
 but by the jet transverse momentum $p_{T,J}$, which means that this strong suppression
 leads to a smaller correction due to EU and pileup effect [114].

⁵ Notice that this is a combinatorial calculation.

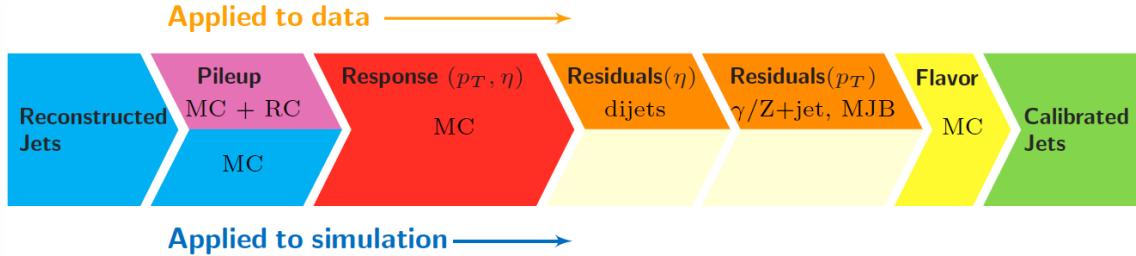
1879 **Jet energy Corrections**

Figure 3.5: Jet energy correction diagram. Correction levels are applied sequentially in the indicated fixed order [120].

1880 Even though jets can be reconstructed efficiently, there are some effects that are
 1881 not included in the reconstruction and that lead to discrepancies between the re-
 1882 constructed results and the predicted results; in order to overcome these discrep-
 1883 ancies, a factorized model has been designed in the form of jet energy corrections
 1884 (JEC) [119, 120] applied sequentially as shown in the diagram of Figure 3.5.

1885 At each level, the jet four-momentum is multiplied by a scaling factor based on
 1886 jet properties, i.e., η , flavor, etc.

- 1887 • Level 1 correction removes the energy coming from pile-up. The scale factor is
 1888 determined using a MC sample of QCD dijet (2 jets) events with and without
 1889 pileup overlay; it is parametrized in terms of the offset energy density ρ , jet
 1890 area A , jet η and jet p_T . Different corrections are applied to data and MC due
 1891 to the detector simulation.
- 1892 • MC-truth correction accounts for differences between the reconstructed jet en-
 1893 ergy and the MC particle-level energy. The correction is determined on a QCD
 1894 dijet MC sample and is parametrized in terms of the jet p_T and η .
- 1895 • Residuals correct remaining small differences within jet response in data and
 1896 MC. The Residuals η -dependent correction compares jets of similar p_T in the

1897 barrel reference region. The Residuals p_T -dependent correct the jet absolute
 1898 scale (JES vs p_T).

- 1899 • Jet-flavor corrections are derived in the same way as MC-truth corrections but
 1900 using QCD pure flavor samples.

1901 ***b*-tagging of jets.**

1902 A particular feature of the hadrons containing bottom quarks (*b*-hadrons) is that
 1903 their lifetime is long enough to travel some distance before decaying, but it is not as
 1904 long as those of light quark hadrons; therefore, when looking at the hadrons produced
 1905 in pp collisions, *b*-hadrons decay typically inside the tracker rather than reaching the
 1906 calorimeters as some light-hadrons do. As a result, a *b*-hadron decay gives rise to a
 1907 displaced vertex (secondary vertex) with respect to the primary vertex as shown in
 1908 Figure 3.6; the SV displacement is in the order of a few millimeters. A jet resulting
 1909 from the decay of a *b*-hadron is called *b* jet; other jets are called light jets.

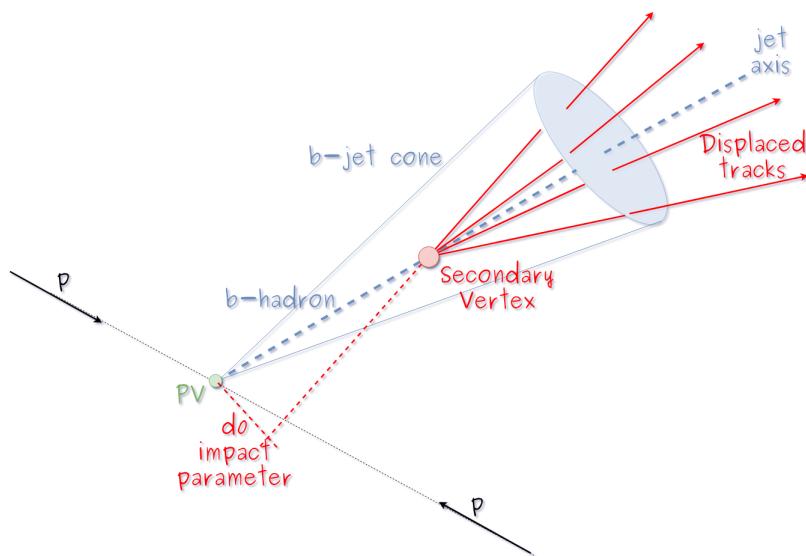


Figure 3.6: Secondary vertex in a *b*-hadron decay.

1910 Several methods to identify *b*-jets (*b*-tagging) have been developed; the method

1911 used in this thesis is known as *Combined Secondary Vertex* algorithm in its second
 1912 version (CSVv2) [121]. By using information of the impact parameter, the recon-
 1913 structed secondary vertices, and the jet kinematics as input in a multivariate analysis
 1914 that combines the discrimination power of each variable in one global discrimina-
 1915 tor variable, three working points (references): loose, medium and tight, are defined
 1916 which quantify the probabilities of mistag jets from light quarks as jets from b quarks;
 1917 10, 1 and 0.1 % respectively. Although the mistagging probability decreases with the
 1918 working point strength, the efficiency to correctly tag b -jets also decreases as 83, 69
 1919 and 49 % for the respective working point; therefore, a balance needs to be achieved
 1920 according to the specific requirements of the analysis.

1921 **3.4.1.1 Missing transverse energy.**

1922 The fact that proton bunches carry momentum along the z -axis implies that for
 1923 each event it is expected that the momentum in the transverse plane is balanced.
 1924 Imbalances are quantified by the missing transverse energy (MET) and are attributed
 1925 to several sources including particles escaping undetected through the beam pipe,
 1926 neutrinos produced in weak interactions processes which do not interact with the
 1927 detector and thus escaping without leaving a sign, or even undiscovered particles
 1928 predicted by models beyond the SM.

1929 The PF algorithm assigns the negative sum of the momenta of all reconstructed
 1930 PF particles to the *particle-flow MET* according to

$$\vec{E}_T = - \sum_i \vec{p}_{T,i} \quad (3.2)$$

1931 JEC are propagated to the calculation of the \vec{E}_T as described in the Reference [122].

1932 **3.4.2 Event reconstruction examples**

1933 Figures 3.7-3.9 show the results of the reconstruction performed on 3 recorded events.

1934 Descriptions are taken directly from the source.

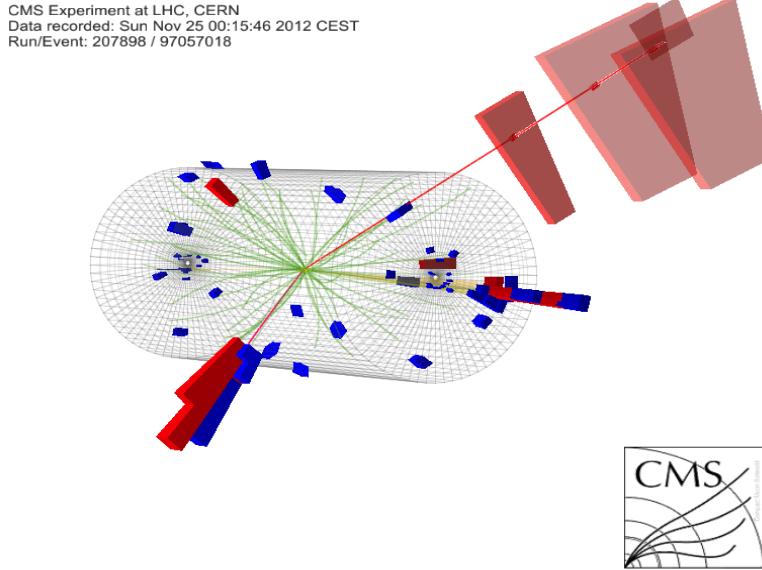


Figure 3.7: HIG-13-004 Event 1 reconstruction results; “HIG-13-004 Event 1: Event recorded with the CMS detector in 2012 at a proton-proton center-of-mass energy of 8 TeV. The event shows characteristics expected from the decay of the SM Higgs boson to a pair of τ leptons. Such an event is characterized 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.” [123].

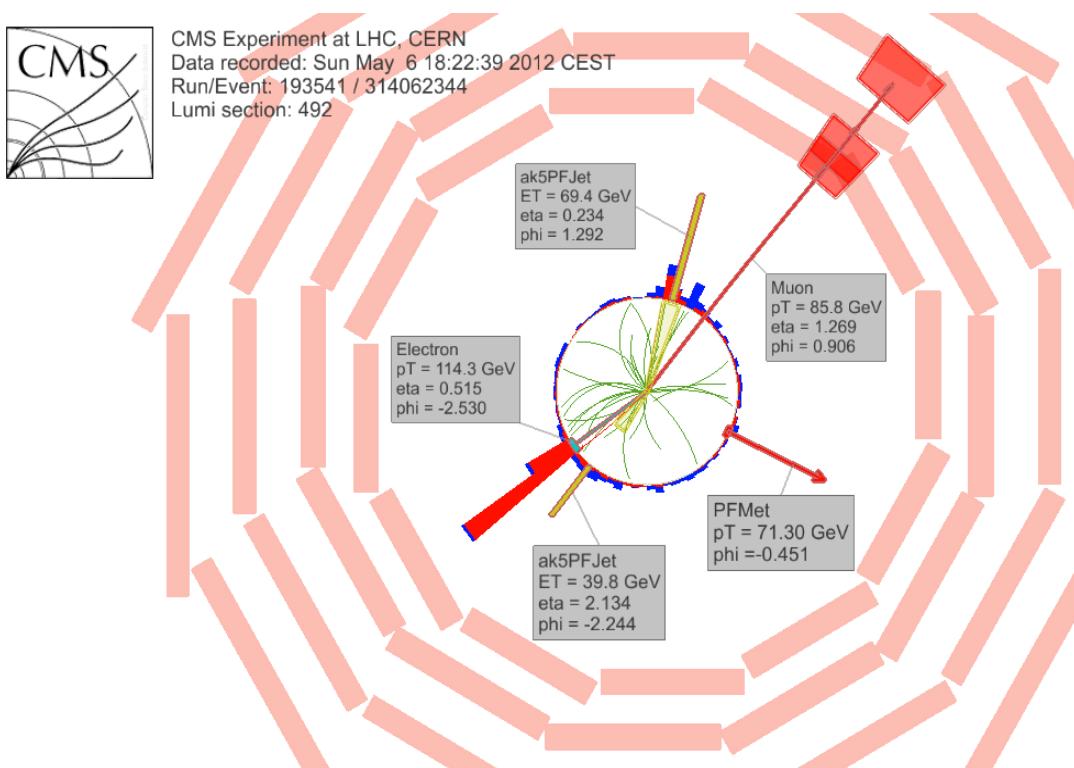


Figure 3.8: $e\mu$ event reconstruction results; “An $e\mu$ event candidate selected in 8 TeV data, as seen from the direction of the proton beams. The kinematics of the main objects used in the event selection are highlighted: two isolated leptons and two particle-flow jets. The reconstructed missing transverse energy is also displayed for reference” [124].

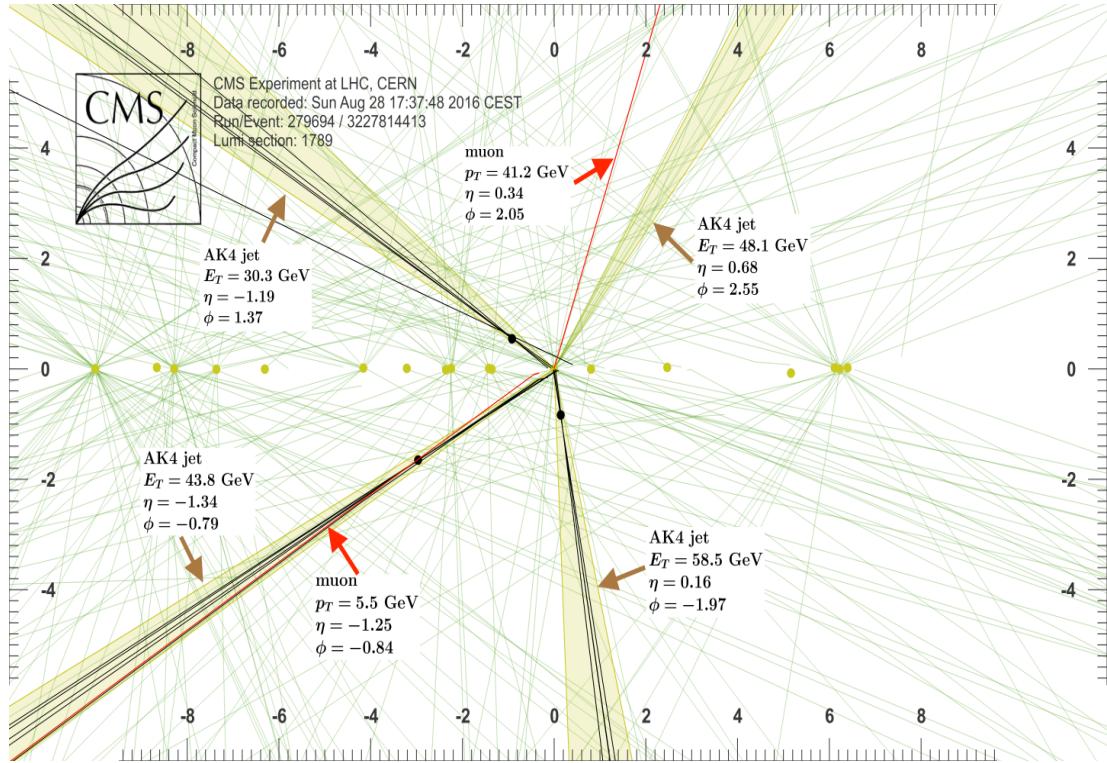


Figure 3.9: Recorded event reconstruction results; “Recorded event (ρ - z projection) with three jets with $p_T > 30$ GeV with one displaced muon track in 2016 data collected at 13 TeV. Each of the three jets has a displaced reconstructed vertex. The jet with $p_T(j) = 43.8$ GeV, $\eta(j) = -1.34$, $\phi(j) = -0.79$ 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” [125].

₁₉₃₅ **Chapter 5**

₁₉₃₆ **Statistical methods**

₁₉₃₇ In the course of analyzing the data sets provided by the CMS experiment and used in
₁₉₃₈ this thesis, several statistical tools have been employed; in this chapter, a description
₁₉₃₉ of these tools will be presented, starting with the general statement of the multivariate
₁₉₄₀ analysis methods, followed by the particularities of the Boosted Decision Trees (BDT)
₁₉₄₁ method and its application to the classification problem. Statistical inference methods
₁₉₄₂ used will also be presented. This chapter is based mainly on References [126–128].

₁₉₄₃ **5.1 Multivariate analysis**

₁₉₄₄ Multivariate data analysis (MVA) makes use of the statistical techniques developed to
₁₉₄₅ analyze more than one variable at once, taking into account all the correlations among
₁₉₄₆ variables. MVA is employed in a variety of fields like consumer and market research,
₁₉₄₇ quality control and process optimization. Using MVA it is possible to identify the
₁₉₄₈ dominant patterns in a data sample, like groups, outliers and trends, and determine
₁₉₄₉ to which group a set of values belong; in the particle physics context, MVA methods
₁₉₅₀ are used to perform the selection of certain type of events from a large data set.

1951 Processes with small cross section, such as the tHq process ($\sigma_{SM}(\sqrt{s} = 13\text{TeV}) =$
 1952 70.96 fb), are hard to detect in the presence of the processes with larger cross sections,
 1953 $\sigma_{SM}^{t\bar{t}}(\sqrt{s} = 13\text{TeV}) = 823.44$ fb for instance; therefore, only a small fraction of the data
 1954 contains events of interest (signal), the major part is signal-like events, which mimic
 1955 signal characteristics but belong to different processes, so they are a background to
 1956 the process of interest. This implies that it is not possible to say with certainty
 1957 that a given event is a signal or a background and statistical methods should be
 1958 involved. In that sense, the challenge can be formulated as one where a set of events
 1959 have to be classified according to certain special features; these features correspond
 1960 to the measurements of several parameters like energy or momentum, organized in a
 1961 set of *input variables*. The measurements for each event can be written in a vector
 1962 $\mathbf{x} = (x_1, \dots, x_n)$ for which

- 1963 • $f(\mathbf{x}|s)$ is the probability density (*likelihood function*) that \mathbf{x} is the set of mea-
 1964 sured values given that the event is a signal event (signal hypothesis).
- 1965 • $f(\mathbf{x}|b)$ is the probability density (*likelihood function*) that \mathbf{x} is the set of mea-
 1966 sured values given that the event is a background event (background hypothe-
 1967 sis).

1968 Figure 5.1 shows three ways to perform a classification of events for which mea-
 1969 surements of two properties, i.e., two input variables x_1 and x_2 , have been performed;
 1970 blue circles represent signal events while red triangles represent background events.
 1971 The classification on the left is *cut-based* requiring $x_1 < c_1$ and $x_2 < c_2$; usually the
 1972 cut values (c_1 and c_2) are chosen according to some knowledge about the event pro-
 1973 cess. In the middle plot, the classification is performed using a linear function of
 1974 the input variables, hence the boundary is a straight line, while in the right plot the

1975 the relationship between input variables is not linear thus the boundary is not linear
 1976 either.

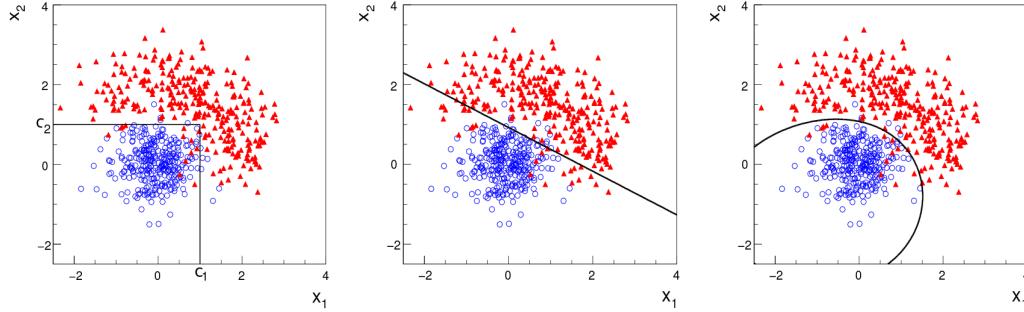


Figure 5.1: Scatter plots-MVA event classification. Distribution of two input variables x_1 and x_2 measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on cuts (left), linear boundary (center), and nonlinear boundary (right) [126]

1977 In general, the boundary can be parametrized in terms of the input variables such
 1978 that the cut is set on the parametrization instead of on the variables, i.e., $y(\mathbf{x}) = y_{cut}$
 1979 with y_{cut} being a constant; thus, the acceptance or rejection of an event is based on
 1980 which side of the boundary the event is located. If $y(\mathbf{x})$, usually called *test statistic*,
 1981 has functional form, it can be used to determine the probability distribution functions
 1982 $p(y|s)$ and $p(y|b)$ and then perform a test statistic with a single cut on the scalar
 1983 variable y .

1984 Figure 5.2 shows an example of what would be the probability distribution func-
 1985 tions under the signal and background hypotheses for a scalar test statistic with a cut
 1986 on the classifier y . Note that the tails of the distributions indicate that some signal
 1987 events fall in the rejection region and some background events fall on the acceptance
 1988 region; therefore, it is convenient to define the *efficiency* with which events of a given
 1989 type are accepted. The signal and background efficiencies are given by

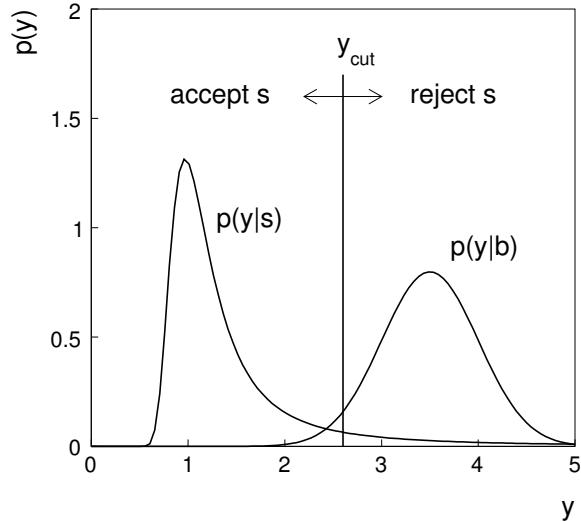


Figure 5.2: Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses. [126]

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

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

1990 where A is the acceptance region. If the background hypothesis is the *null hypothesis*
 1991 (H_0), the signal hypothesis would be *alternative hypothesis* (H_1); in this context, the
 1992 background efficiency corresponds to the significance level of the test (α) and describes
 1993 the misidentification probability, while the signal efficiency corresponds to the power
 1994 of the test ($1-\beta$)¹ and describes the probability of rejecting the background hypothesis
 1995 if the signal hypothesis is true. What is sought in an analysis is to maximize the power
 1996 of the test relative to the significance level, i.e., set a selection with the largest possible
 1997 selection efficiency and the smallest possible misidentification probability.

¹ β is the fraction of signal events that fall out of the acceptance region

1998 **5.1.1 Decision trees**

1999 For this thesis, the implementation of the MVA strategy, described above, is per-
 2000 formed through decision trees by using the TMVA software package [127] included
 2001 in the ROOT analysis framework [129]. In a simple picture, a decision tree classifies
 2002 events according to their input variables values by setting a cut on each input variable
 2003 and checking which events are on which side of the cut, just as proposed in the MVA
 2004 strategy, but in addition, as a machine learning algorithm, decision trees offer the
 2005 possibility to be trained and then perform the classification efficiently.

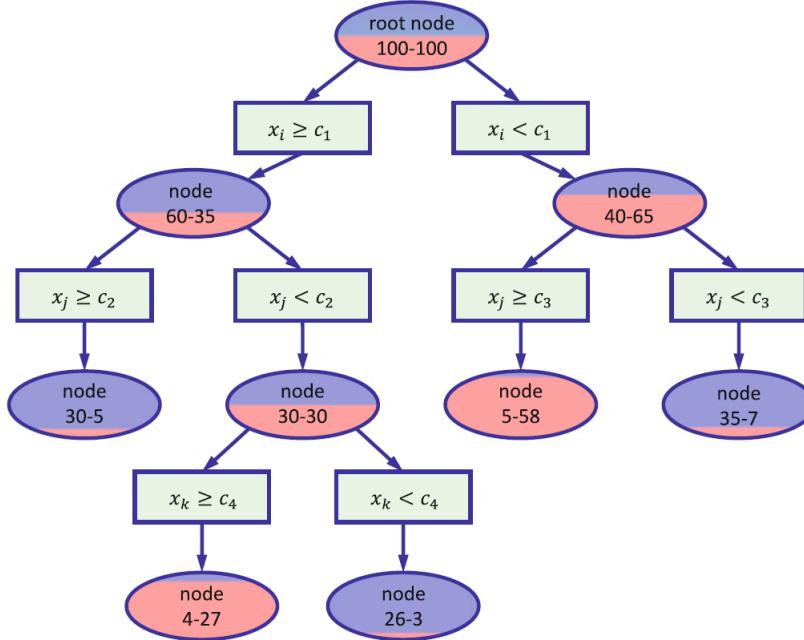


Figure 5.3: Example of a decision tree. Each node is fed with a MC sample mixing signal and background events (left-right numbers); nodes colors represent the relative number of signal/background events [128].

2006 The training or growing of a decision tree is the process where the rules for clas-
 2007 sifying events are defined; this process is represented in Figure 5.3 and consists of
 2008 several steps:

- 2009 • take MC samples of signal and background events and split them into two parts

2010 each; the first parts will be used in the decision tree training, while the second
 2011 parts will be used for testing the final classifier obtained from the training.
 2012 Each event has associated a set of input variables $\mathbf{x} = (x_1, \dots, x_n)$ which serve
 2013 to distinguish between signal and background events. The training sample is
 2014 taken in at the *root node*.

- 2015 • Pick one variable, say x_i .
- 2016 • Pick one value of x_i , each event has its own value of x_i , and split the training
 2017 sample into two subsamples B_1 and B_2 ; B_1 contains events for which $x_i < c_1$
 2018 while B_2 contains the rest of the training events;
- 2019 • scan all possible values of x_i and find the splitting value that provides the *best*
 2020 classification², i.e., B_1 is mostly made of signal events while B_2 is mostly made
 2021 of background events.
- 2022 • It is possible that variables other than the picked one produce a better classi-
 2023 fication, hence, all the variables have to be evaluated. Pick the next variable,
 2024 say x_j , and repeat the scan over its possible values.
- 2025 • At the end, all the variables and their values will have been scanned, the *best*
 2026 variable and splitting value will have been identified, say x_1, c_1 , and there will
 2027 be two nodes fed with the subsamples B_1 and B_2 .

2028 Nodes are further split by repeating the decision process until a given number of
 2029 final nodes is obtained, nodes are largely dominated by either signal or background
 2030 events, or nodes have too few events to continue. Final nodes are called *leaves* and
 2031 they are classified as signal or background leaves according to the class of the majority
 2032 of events in them. Each *branch* in the tree corresponds to a sequence of cuts.

² Quality of the classification will be treated in the next paragraph.

2033 The quality of the classification at each node is evaluated through a separation
 2034 criteria; there are several of them but the *Gini Index* (G) is the one used in the
 2035 decision trees trained for the analysis in this thesis. G is written in terms of the
 2036 purity (P), i.e., the fraction of signal events in the samples after the separation is
 2037 made; it is given by

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

2038 note that $P=0.5$ at the root node while $G=0$ for pure leaves. For a node A split into
 2039 two nodes B_1 and B_2 the G gain is

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

2040 The *best* classification corresponds to that for which the gain of G is maximized;
 2041 hence, the scanning over all the variables in an event and their values is of great
 2042 importance.

2043 In order to provide a numerical output for the classification, events in a sig-
 2044 nal(background) leaf are assigned an score of 1(-1) each, defining in this way the
 2045 decision tree *classifier/weak learner* as

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

2046 Figure 5.4 shows an example of the classification of a sample of events, containing
 2047 two variables, performed by a decision tree.

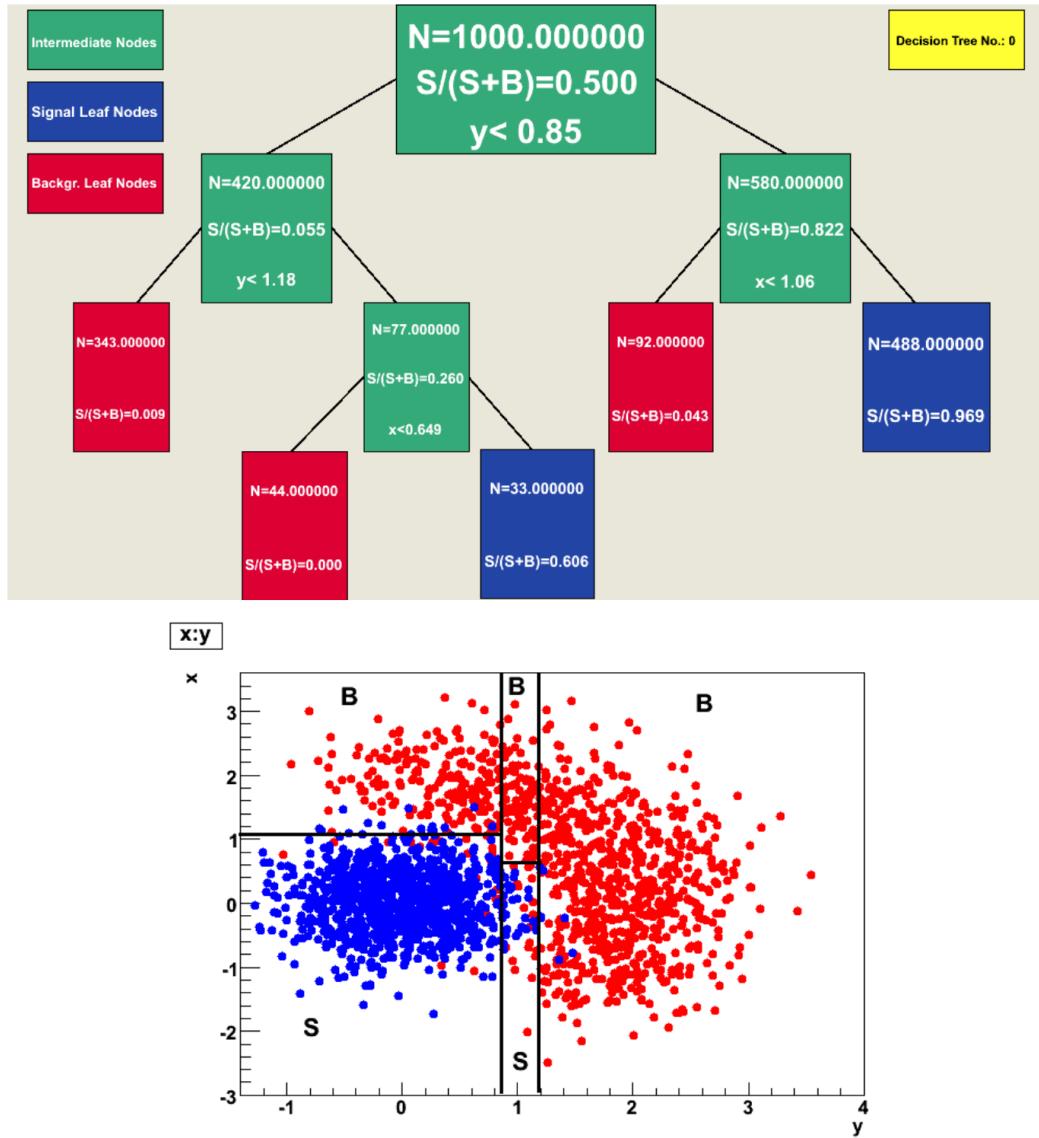


Figure 5.4: Example of a decision tree output. Each leaf, blue for signal events and red for background events, is represented by a region in the variables phase space [130].

2048 5.1.2 Boosted decision trees (BDT).

2049 Event misclassification occurs when a training event ends up in the wrong leaf, i.e., a
 2050 signal event ends up in a background leaf or a background event ends up in a signal
 2051 leaf. A way to correct it is to assign a weight to the misclassified events and train
 2052 a second tree using the reweighted events; the event reweighting is performed by a

2053 boosting algorithm in such a way that when used in the training of a new decision
 2054 tree the *boosted events* get correctly classified. The process is repeated iteratively
 2055 adding a new tree to the forest and creating a set of classifiers, which are combined
 2056 to create the next classifier; the final classifier offers more stability³ and has a smaller
 2057 misclassification rate than any individual ones. The resulting tree collection is known
 2058 as a *boosted decision tree (BDT)*.

2059 Thus, purity of the sample is generalized to

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

2060 where w_s and w_b are the weights of the signal and background events respectively;
 2061 the Gini index is also generalized

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

2062 with n the number of events in the node. The final score of an event, after pass-
 2063 ing through the forest, is calculated as the renormalized sum of all the individual
 2064 (possibly weighted) scores; thus, high(low) score implies that the event is most likely
 2065 signal(background).

2066 The boosting procedure, implemented in the *Gradient boosting* algorithm used in
 2067 this thesis, produces a classifier $F(\mathbf{x})$ which is the weighted sum of the individual
 2068 classifiers obtained after each iteration, i.e.,

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

2069 where M is the number of trees in the forest. The *loss function* $L(F, y)$ represents the

³ Decision trees suffer from sensitivity to statistical fluctuations in the training sample which may lead to very different results with a small change in the training samples.

2070 deviation between the classifier $F(\mathbf{x})$ response and the true value y obtained from the
 2071 training sample (1 for signal events and -1 for background event), according to

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

2072 thus, the reweighting is employed to ensure the minimization of the loss function; a
 2073 more detailed description of the minimization procedure can be found in Reference
 2074 [131]. The final classifier output is later used as a final discrimination variable, labeled
 2075 as *BDT output/response*.

2076 5.1.3 Overtraining

2077 Decision trees offer the possibility to have as many nodes as desired in order to
 2078 reduce the misclassification to zero (in theory); however, when a classifier is too much
 2079 adjusted to a particular training sample, the classifier's response to a slightly different
 2080 sample may leads to a completely different classification results; this effect is known
 2081 as *overtraining*.

2082 An alternative to reduce the overtraining in BDTs consists in pruning the tree
 2083 by removing statistically insignificant nodes after the tree growing is completed but
 2084 this option is not available for BDTs with gradient boosting in the TMVA-toolkit,
 2085 therefore, the overtraining has to be reduced by tuning the algorithm, number of
 2086 nodes, minimum number of events in the leaves, etc. The overtraining can be evaluated
 2087 by comparing the responses of the classifier when running over the training and
 2088 test samples.

2089 **5.1.4 Variable ranking**

2090 BDTs have a couple of particular advantages related to the input variables; they are
 2091 relatively insensitive to the number of input variables used in the vector \mathbf{x} . The
 2092 ranking of the BDT input variables is determined by counting the number of times a
 2093 variable is used to split decision tree nodes; in addition, the separation gain-squared
 2094 achieved in the splitting and the number of events in the node are accounted by
 2095 applying a weighting to that number. Thus, those variables with small or no power
 2096 to separate signal and background events are rarely chosen to split the nodes, i.e., are
 2097 effectively ignored.

2098 In addition, variables correlations play an important role for some MVA methods
 2099 like the Fisher discriminant algorithm in which the first step consist of performing a
 2100 linear transformation to a phase space where the correlations between variables are
 2101 removed; in the case of BDT algorithm, correlations do not affect the performance.

2102 **5.1.5 BDT output example**

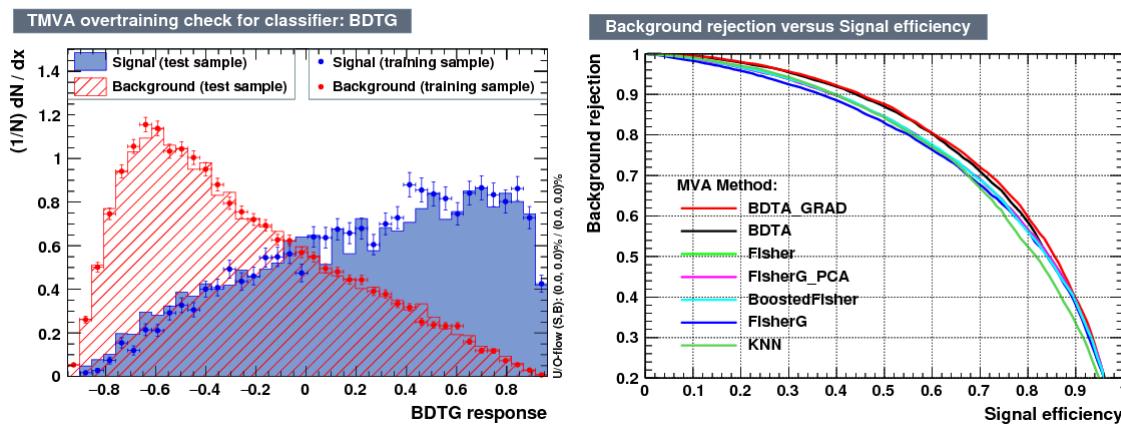


Figure 5.5: Left: Output distributions for the gradient boosted decision tree (BDTG) classifier using a sample of signal ($pp \rightarrow tHq$) and background ($pp \rightarrow tt$) events. Right: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers running over the same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

2103 The left side of figure 5.5 shows the BDT output distributions for signal ($pp \rightarrow$
 2104 tHq) and background ($pp \rightarrow t\bar{t}$) events; this plot is the equivalent to the one showed
 2105 in Figure 5.2. A forest with 800 trees, maximum depth per tree = 3, and gradient
 2106 boosting have been used as training parameters. The BDTG classifier offers a good
 2107 separation power. There is a small overtraining in the signal distribution, while the
 2108 background distribution is very well predicted which might indicate that the sample
 2109 is composed of more background than signal events.

2110 The right side of figure 5.5 shows the background rejection vs signal efficiency
 2111 curves for several combinations of MVA classifiers-boosting algorithms running over
 2112 the same MC sample; these curves are known as ROC curves and give an indication
 2113 of the performance of the classifier. In this particular example, the best performance
 2114 is achieved with the BDTG classifier (BDTA_GRAD), which motivate its use in this
 2115 thesis.

2116 **5.2 Statistical inference**

2117 Once events are classified, the next step consists of finding the parameters that define
 2118 the likelihood functions $f(\mathbf{x}|s)$, $f(\mathbf{x}|b)$ for signal and background events respectively.
 2119 In general, likelihood functions depend not only on the measurements but also on
 2120 parameters (θ_m) that define their shapes; the process of estimating these *unknown*
 2121 *parameters* and their uncertainties from the experimental data is called *inference*.

2122 The statistical inference tools used in this analysis are implemented in the RooFit
 2123 toolkit [132] and COMBINE package [133] included in the CMSSW software frame-
 2124 work.

2125 **5.2.1 Nuisance parameters**

2126 The unknown parameter vector θ is made of two types of parameters: those pa-
 2127 rameters that provide information about the physical observables of interest for the
 2128 experiment or *parameters of interest*, and the *nuisance parameters* that are not of
 2129 direct interest for the experiment but that need to be included in the analysis in
 2130 order to achieve a satisfactory description of the data; they represent effects of the
 2131 detector response like the finite resolutions of the detection systems, miscalibrations,
 2132 and in general any source of uncertainty introduced in the analysis.

2133 Nuisance parameters can be estimated from experimental data; for instance, data
 2134 samples from a test beam are usually employed for calibration purposes. In cases
 2135 where experimental samples are not availables, the estimation of nuisance parameters
 2136 makes use of dedicated simulation programs to provide the required samples.

2137 The estimation of the unknown parameters involves certain deviations from their
 2138 true values, hence, the measurement of the nuisance parameter is written in terms
 2139 of an estimated value, also called central value, $\hat{\theta}$ and its uncertainty $\delta\theta$ using the
 2140 notation

$$\theta = \hat{\theta} \pm \delta\theta \quad (5.9)$$

2141 where the interval $[\hat{\theta} - \delta\theta, \hat{\theta} + \delta\theta]$ is called *confidence interval*; it is usually interpreted,
 2142 in the limit of infinite number of experiments, as the interval where the true value
 2143 of the unknown parameter θ is contained with a probability of 0.6827 (if no other
 2144 convention is stated); this interval represents the area under a Gaussian distribution
 2145 in the interval $\pm 1\sigma$.

2146 The uncertainties associated with nuisance parameters produce *systematic uncer-*
 2147 *tainties* in the final measurement, while the uncertainties related only to fluctuations

2148 in data and that affect the determination of parameters of interest produce *statistical*
 2149 *uncertainties*.

2150 **5.2.2 Maximum likelihood estimation method**

2151 The estimation of the unknown parameters that are in best agreement with the ob-
 2152 served data is performed through a function of the data sample that returns the
 2153 estimate of those parameters; that function is called an *estimator*. Estimators are
 2154 usually constructed using mathematical expressions encoded in algorithms.

2155 In this thesis, the estimator used is the likelihood function $f(\mathbf{x}|\boldsymbol{\theta})$ ⁴ which depends
 2156 on a set of measured variables \mathbf{x} and a set of unknown parameters $\boldsymbol{\theta}$. The likelihood
 2157 function for N events in a sample is the combination of all the individual likelihood
 2158 functions, i.e.,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N f(\mathbf{x}^i|\boldsymbol{\theta}) = \prod_{i=1}^N f(x_1^i, \dots, x_n^i; \theta_1, \dots, \theta_m) \quad (5.10)$$

2159 and the estimation method used is the *Maximum Likelihood Estimation* method
 2160 (MLE); it is based on the combined likelihood function defined by eqn. 5.10 and
 2161 the procedure seeks for the parameter set that corresponds to the maximum value of
 2162 the combined likelihood function, i.e., the *maximum likelihood estimator* of the un-
 2163 known parameter vector $\boldsymbol{\theta}$ is the function that produces the vector of *best estimators*
 2164 $\hat{\boldsymbol{\theta}}$ for which the likelihood function $L(\boldsymbol{\theta})$ evaluated at the measured \mathbf{x} is maximum.

2165 Usually, the logarithm of the likelihood function is used in numerical algorithm
 2166 implementations in order to avoid underflow the numerical precision of the computers
 2167 due to the product of low likelihoods. In addition, it is common to minimize the
 2168 negative logarithm of the likelihood function, therefore, the negative log-likelihood

⁴ analogue to the likelihood functions described in previous sections

2169 function is

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

2170 The minimization process is performed by the software MINUIT [134] implemented in the ROOT analysis framework. In case of data samples with large number 2171 of measurements, the computational resources necessary to calculate the likelihood 2172 function are too big; therefore, the parameter estimation is performed using binned 2173 distributions of the variables of interest for which the *binned likelihood function* is 2174 given by

$$L(\mathbf{x}|r, \boldsymbol{\theta}) = \prod_{i=1} \frac{(r \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-r \cdot s_i(\boldsymbol{\theta}) - b_i(\boldsymbol{\theta})} \prod_{j=1} \frac{1}{\sqrt{2\pi}\sigma_{\theta_j}^2} e^{-(\theta_j - \theta_{0,j})^2/2\sigma_{\theta_j}^2}, \quad (5.12)$$

2176 with s_i and b_i the expected number of signal and background yields for the bin i , n_i
 2177 is the observed number of events in the bin i and $r = \sigma/\sigma_{SM}$ is the signal strength.
 2178 Note that the number of entries per bin follows a Poisson distribution. The effect
 2179 of the nuisance parameters have been included in the likelihood function through
 2180 the multiplication by a Gaussian distribution that models the nuisance. The three
 2181 parameters, r , s_i and b_i are jointly fitted to estimate the value of r .

2182 5.3 Upper limits

2183 In this analysis, two hypotheses are considered; the background only hypothesis
 2184 ($H_0(b)$) and the signal plus background hypothesis ($H_1(s+b)$), i.e., the sample of
 2185 events is composed of background only events ($r=0$) or it is a mixture of signal plus
 2186 background events ($r=1$). The exclusion of one hypothesis against the other means
 2187 that the observed data sample better agrees with H_0 or rather with H_1 . In order
 2188 to discriminate these hypotheses, a test statistic is constructed on the basis of the

2189 likelihood function evaluated for each of the hypothesis.

2190 The *Neyman-Pearson* lemma [135] states that the test statistic that provides the
 2191 maximum power for H_1 for a given significance level (background misidentification
 2192 probability α), is given by the ratio of the likelihood functions $L(\mathbf{x}|H_1)$ and $L(\mathbf{x}|H_0)$;
 2193 however, in order to use that definition it is necessary to know the true likelihood
 2194 functions, which in practice is not always possible. Approximate functions obtained
 2195 by numerical methods, like the BDT method described above, have to be used, so
 2196 that the *profile likelihood* test statistic is defined by

$$\lambda(\mathbf{r}) = \frac{L(\mathbf{x}|r, \hat{\boldsymbol{\theta}}(r))}{L(\mathbf{x}|\hat{r}, \hat{\boldsymbol{\theta}})}, \quad (5.13)$$

2197 where, \hat{r} and $\hat{\boldsymbol{\theta}}$ maximize the likelihood function, and $\hat{\boldsymbol{\theta}}$ maximizes the likelihood
 2198 function for a given value of the signal strength modifier r . In practice, the test
 2199 statistic t_r

$$t_r = -2\ln\lambda(r) \quad (5.14)$$

2200 is used to evaluate the presence of signal in the sample, since the minimum of t_r at
 2201 $r = \hat{r}$ suggests the presence of signal with signal strength \hat{r} . The uncertainty interval
 2202 for r is determined by the values of r for which $t_r = +1$.

2203 The expected probability density function (p.d.f) $f(t_r|r, \boldsymbol{\theta})$ of the test statistic t_r
 2204 can be obtained numerically by generating MC samples where one hypothesis, $H_0(b)$
 2205 or $H_1(s+b)$, is assumed; thus, MC samples contain the possible values of t_r obtained
 2206 from *pseudo-experiments* as shown in Figure 5.6. The probability that t_r takes a value
 2207 equal or greater than the observed value ($t_{r,obs}$) when a signal with a signal modifier
 2208 r is present in the data sample, is called the *p-value* of the observation; it can be
 2209 calculated using

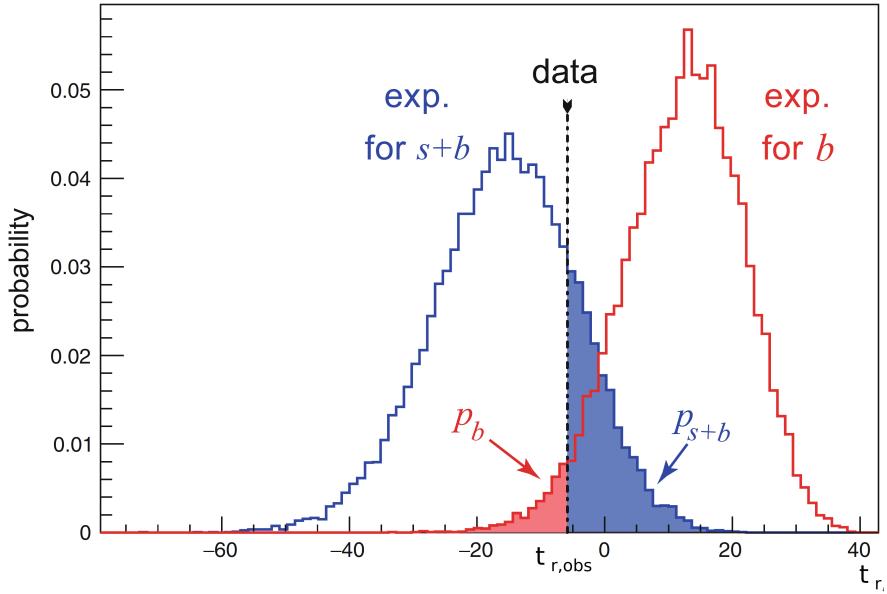


Figure 5.6: t_r p.d.f. from MC pseudo experiments assuming H_0 (red) and H_1 (blue). The black dashed line shows the value of the test statistic as measured from data. Adapted from Reference [128].

$$p_r = \int_{t_{r,obs}}^{\infty} f(t'_r | r, \boldsymbol{\theta}) dt'_r, \quad (5.15)$$

thus, $p_r < 0.05$ means that, for that particular value of r , H_1 could be excluded at 95% Confidence Level (CL). The corresponding background-only p-value is given by

$$1 - p_b = \int_{t_{r,obs}}^{\infty} f(t'_r | 0, \boldsymbol{\theta}) dt'_r, \quad (5.16)$$

If the t_r p.d.f.s for both hypotheses are well separated, as shown in the top side of Figure 5.7, the experiment is sensitive to the presence of signal in the sample. If the signal presence is small, both p.d.f.s will be largely overlapped (bottom of Figure ??) and either the signal hypothesis could be rejected with not enough justification because the experiment is not sensitive to the signal or a fluctuation of the background could be misinterpreted as presence of signal with the corresponding rejection of the

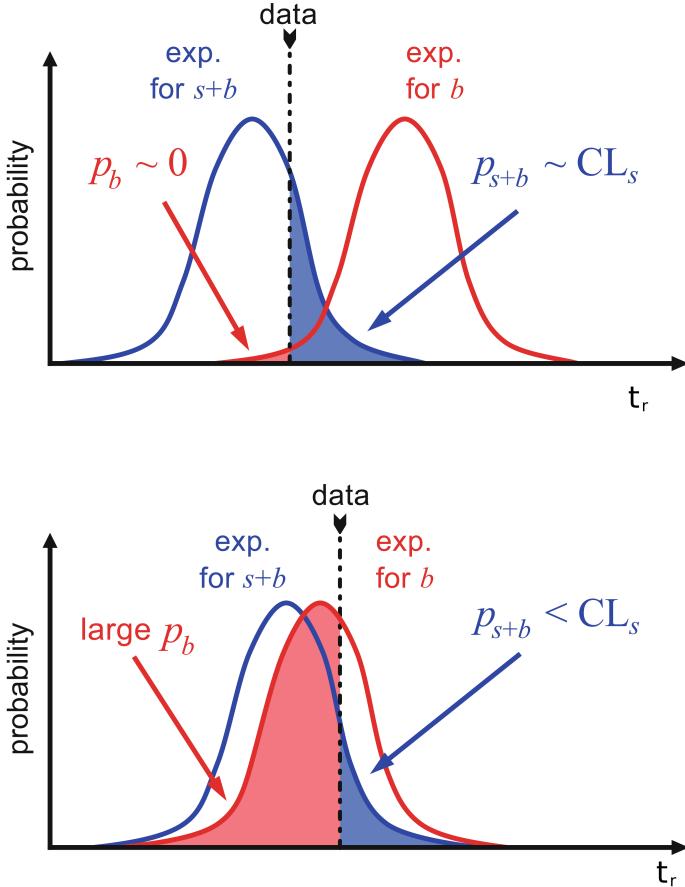


Figure 5.7: CL_s limit illustration. When the test statistic p.d.f. for the two hypotheses H_0 and H_1 are well separated (top) and when they are largely overlapped (bottom). Adapted from Reference [128].

2218 background-only hypothesis. These issues are corrected by using the modified p-
2219 value [136]

$$p'_r = \frac{p_r}{1 - p_b} \equiv CL_s. \quad (5.17)$$

2220 If H_1 is true, then p_b is small, $CL_s \simeq p_r$ and H_0 is rejected; if there is large
2221 overlap and an statistical fluctuation cause that p_b is large, then both numerator and
2222 denominator in Eqn. 5.17 become small but CL_s would allow the rejection of H_1
2223 even if there is poor sensitivity to signal.

2224 The upper limit of the parameter of interest r^{up} is determined by excluding the
 2225 range of values of r for which $CL_s(r, \theta)$ is lower than the confidence level desired,
 2226 normally 90% or 95%, e.g, scanning over r and finding the value for which $p_r'^{up} =$
 2227 0.05. The expected upper limit can be calculated using pseudo-experiments based on
 2228 the background-only hypothesis and obtaining a distribution for r_{ps}^{up} ; the median of
 2229 that distribution corresponds to the expected upper limit, while the $\pm 1\sigma$ and $\pm 2\sigma$
 2230 deviations correspond to the values of the distribution that defines the 68% and 95%
 2231 of the area under the distribution centered in the median. It is usual to present all
 2232 the information about the expected and observed limits in the so-called *Brazilian-flag*
 2233 *plot* as the one showed in Figure 5.8. The solid line represent the observed CL_s

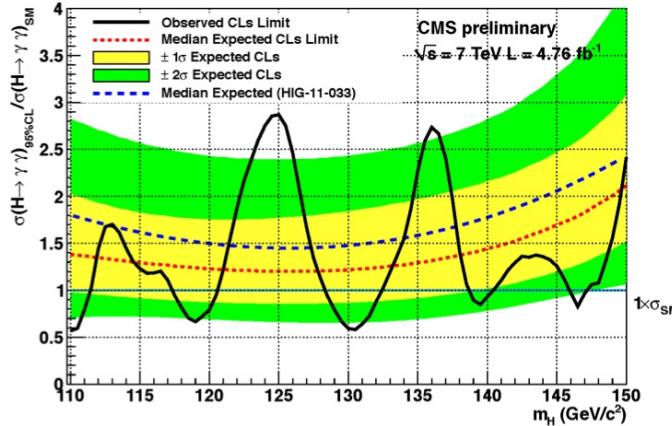


Figure 5.8: Brazilian flag plot of CMS experiment limits for Higgs boson decaying to photons [137].

2234 5.4 Asymptotic limits

2235 As said before, the complexity of the likelihood functions, the construction of test
 2236 statistics, and the calculation of the limits and their uncertainties is not always man-
 2237 ageable and requires extensive computational resources; in order to overcome those
 2238 issues, asymptotic approximations for likelihood-based test statistics, like the ones

described in previous sections, have been developed [138, 139] using Wilks' theorem.
Asymptotic approximations replace the construction of the test statistics p.d.f.s using
MC pseudo-experiments, with the approximate calculation of the test statistics p.d.f.s
by employing the so-called *Asimov dataset*.

The Asimov dataset is defined as the dataset that produce the true values of the
nuisance parameters when it is used to evaluate the estimators for all the parameters;
it is obtained by setting the values of the variables in the dataset to their expected
values [139].

Limits calculated by using the asymptotic approximation and the Asimov dataset
are know as *asymptotic limits*.

2249 **Chapter 6**

2250 **Search for production of a Higgs**

2251 **boson and a single top quark in**

2252 **multilepton final states in pp**

2253 **collisions at $\sqrt{s} = 13$ TeV**

2254 **6.1 Introduction**

2255 The Higgs boson discovery, supported on experimental observations and theoretical
2256 predictions made about the SM, gives the clue of the way in that elementary particles
2257 acquire mass through the Higgs mechanism; therefore, knowing the Higgs mass, the
2258 Higgs-vector boson and Higgs-fermion couplings can be determined. In order to test
2259 the Higgs-top coupling, several measurements have been performed, as stated in the
2260 chapter 1, but they are limited in sensitivity to measure the square of the coupling.
2261 The production of a Higgs boson in association with a single top quark (tH) not
2262 only offers access to the sign of the coupling, but also, to the CP phase of the Higgs

2263 couplings.

2264 This chapter presents the search for the associated production of a Higgs boson
 2265 and a single top quark (tHq) events, focusing on leptonic signatures provided by the
 2266 Higgs decay modes to WW , ZZ , and $\tau\tau$; the 13 TeV dataset produced in 2016, which
 2267 corresponds to an integrated luminosity of 35.9fb^{-1} , is used.

2268 As shown in Section 1.5, the SM cross section of tHq process is driven by a
 2269 destructive interference between two contributions (see Figure 1.15), where the Higgs
 2270 couples to either the W boson or the top quark; however, if the sign of the Higgs-
 2271 top coupling is flipped with respect to the SM prediction, a large enhancement of
 2272 the cross section occurs, making this analysis sensitive to such deviation. A second
 2273 process, where the Higgs boson and top quark are accompanied by a W boson (tHW)
 2274 has similar behavior, albeit with a weaker interference pattern and lower contribution
 2275 to the cross section, therefore, a combination of both processes would increase the
 2276 sensitivity to the sign of the coupling; in this analysis both contributions are combined
 2277 and referred as tH channel. A third contribution comes from $t\bar{t}H$ process. The purpose
 2278 of this analysis is to investigate the exclusion of the presence of the $tH + t\bar{t}H$ processes
 2279 under the assumption of the anomalous Higgs-top coupling modifier ($\kappa_t = -1$). The
 2280 analysis exploits signatures with two leptons of the same sign ($2lss$) channel and three
 2281 leptons ($3l$) channel in the final state.

2282 Constraints on the sign of the Higgs-top coupling (y_t) have been derived from the
 2283 decay rate of Higgs boson to photon pairs [50] and from the cross section for associated
 2284 production of Higgs and Z bosons via gluon fusion [141], with recent results disfavoring
 2285 negative signs of the coupling [44, 59, 142], although the negative sign coupling have
 2286 not been completely excluded.

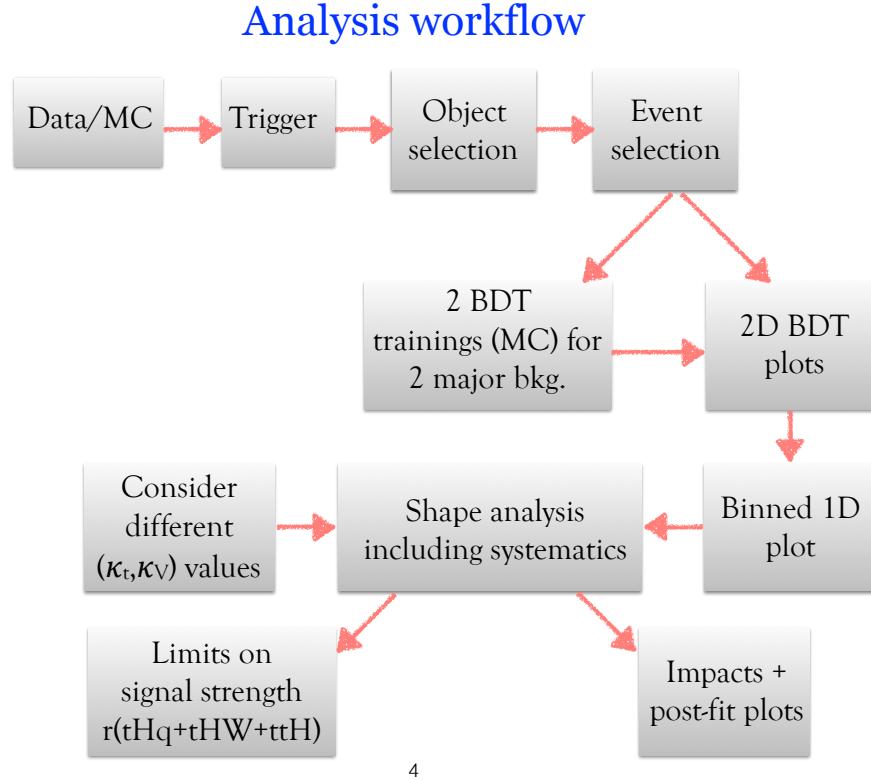
2287 The analysis presented here, expands previous analyses performed at 8 TeV [143,
 2288 144] and searches for associated production of $t\bar{t}$ pair and a Higgs boson in the multi-

lepton final state channel [145]; it also complements searches in other decay channels targeting $H \rightarrow b\bar{b}$ [146].

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 coupling κ_V .

The analysis is designed to efficiently identify and select prompt leptons from on-shell W and Z boson decays and to reject non-prompt leptons from b quark decays and spurious lepton signatures from hadronic jets. Events are then selected in the $2lss$ and $3l$ channels, and are required to contain hadronic jets, some of which must be consistent with b quark hadronization. Finally, the signal yield is extracted by simultaneously fitting the output of two dedicated multivariate discriminants, trained to separate the tHq signal from the two dominant backgrounds, in all categories. The fit result is then used to set an upper limit on the combined $t\bar{t}H + tH$ production cross section, as a function of the relative coupling strengths of Higgs-top quark and Higgs-Vector boson. Figure 6.1 shows an schematic overview of the analysis strategy workflow.

With respect to the 8 TeV analysis, the object selections have been adjusted for the updated LHC running conditions at 13 TeV, the lepton identification has been improved, and more powerful multivariate analysis techniques are used for the signal extraction.



4

Figure 6.1: A schematic overview of the analysis workflow. Based on sets of optimized physics object definitions and selection criteria, signal and background events in a data sample are discriminated. The discrimination is performed by a BDT, previously trained using MC samples of the dominant backgrounds, using discriminant variables based on the b -jet multiplicity, the activity in the forward region of the detector, and the kinematic properties of leptons. The CL_s limits on the combined $t\bar{t}H + tH$ production cross section, as a function of the relative coupling strengths are calculated.

2313 6.2 tHq signature

2314 In order to select events of tHq process, its features are translated into a set of
 2315 selection rules; Figure 6.2 shows the Feynman diagram and an schematic view of the
 2316 tHq process from the pp collision to the final state configuration. A single top quark
 2317 is produced accompanied by a light quark, denoted as q ; this light quark is produced
 2318 predominantly in the forward region of the detector. The Higgs boson can be either
 2319 emitted by the exchanged W boson or directly by the singly produced top quark.

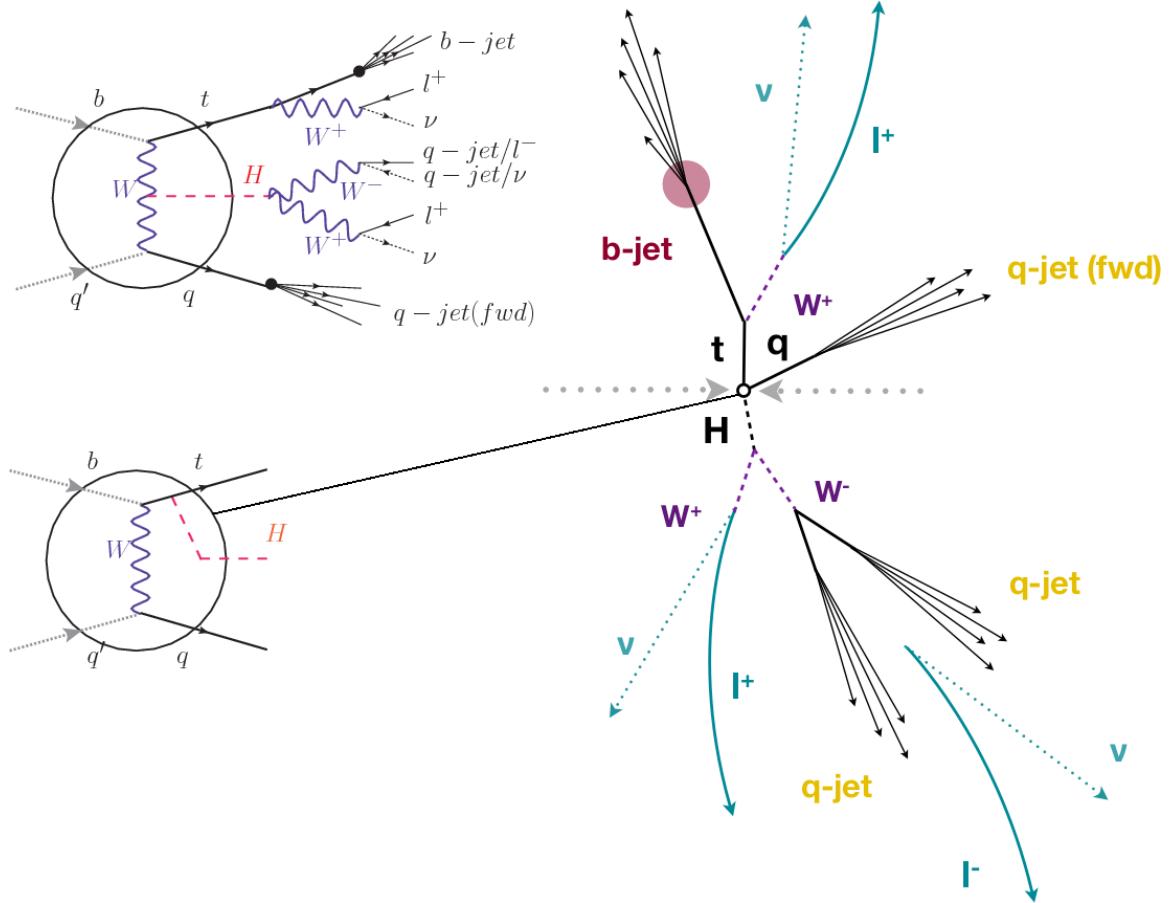


Figure 6.2: tHq event signature. Left: Feynman diagram including the whole evolution up to the final state for the case of the Higgs boson emitted by the W boson (top); Feynman diagram for the case where the Higgs boson is emitted by the top quark. Right: Schematic view as it would be seen in the detector; the circle in the Feynman diagrams on the left corresponds to the circle in the center of the schematic view as indicated by the line connecting them. In the $2lss$ channel, one of the W bosons from the Higgs boson decays to two light-quark jets while in the $3l$ channel both W bosons decay to leptons.

2320 Due to their high masses/short lifetimes, top quark and Higgs boson decay after
 2321 their production within the detector. The Higgs boson is required to decay into a W
 2322 boson pair¹. The top quark almost always decays into a bottom quark and a W boson,
 2323 as encoded in the CMK matrix. The W bosons are required to decay leptonically
 2324 either all the three in the $3l$ channel or the pair with equal electrical charge in the

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

2325 $2lss$ channel case; τ leptons are not reconstructed separately and only their leptonic
 2326 decays into either electrons or muons are considered in this analysis.

2327 In summary, the signal process is characterized by a the final state with

2328 • one light-flavored forward jet,

2329 • one central b-jet,

2330 • $2lss$ channel \rightarrow two leptons of the same sign, two neutrinos and two light (often
 2331 soft) jets,

2332 • $3l$ channel \rightarrow three leptons, three neutrinos and no central light-flavored jets,

2333 The presence of neutrinos is inferred from the presence of MET. The analysis has
 2334 been made public by CMS as a Physics Analysis Summary [147] combining the result
 2335 for the three lepton and two lepton same-sign channels. Currently, an effort to turn
 2336 the analysis into a paper is ongoing.

2337 6.3 Background processes

2338 The background processes are those that can mimic the signal signature or at least
 2339 can be reconstructed as that as a result of certain circumstances. The backgrounds
 2340 can be classified as

2341 • irreducible backgrounds: where genuine prompt leptons are produced in on-
 2342 shell W and Z boson decays; they can be reliably estimated directly from MC
 2343 simulated events, using higher-order cross sections or data control regions for
 2344 the overall normalization.

2345 • reducible backgrounds: where at least one of the leptons is *non-prompt*, i.e.,
 2346 produced within a hadronic jet; genuine leptons from heavy flavor decays and
 2347 misreconstructed jets, also known as *mis-ID leptons* or *fake leptons*, are consid-
 2348 ered non-prompt leptons. These non-prompt leptons leave tracks and hits in
 2349 the detection systems as would a prompt lepton, but correlating those hits with
 2350 nearby jets could be a way of removing them. The misassignment of electron
 2351 charge in processes like $t\bar{t}$ or Drell-Yan, represent an additional source of back-
 2352 ground, but it is relevant only for the $2lss$ channel. Reducible backgrounds are
 2353 not well predicted by simulation, hence, they are estimated using data-driven
 2354 methods.

2355 The main sources of background events for tHq process are $t\bar{t}$ process and $t\bar{t} +$
 2356 X ($X = W, Z, \gamma$) processes, the latter regarded together as $t\bar{t}V$ process. Figure 6.3
 2357 shows the signature for $t\bar{t}$ and $t\bar{t}W$ processes.

2358 The largest contribution to irreducible backgrounds comes from $t\bar{t}W$ and $t\bar{t}Z$ processes
 2359 for which the number of ($b-$)jets (($b-$)jet multiplicity) is higher than that of the sig-
 2360 nal events, while for other contributing background events, WZ , ZZ , and rare SM
 2361 processes like $W^\pm W^\pm qq$, $t\bar{t}t\bar{t}$, tZq , tZW , WWW , WWZ , WZZ , ZZZ , the ($b-$)jet
 2362 multiplicity is lower compared to that of the signal events. None of the irreducible
 2363 backgrounds present activity in the forward region of the detector.

2364 On the side of the reducible backgrounds, the largest contribution comes from the
 2365 $t\bar{t}$ events which have a very similar signature to the signal events but does no present
 2366 activity in the forward region of the detector either; A particular feature of the $t\bar{t}$
 2367 events is their charge-symmetry, which is also a difference with respect to the signal
 2368 events.

2369 The charge misidentification plays an important role in the the $2lss$ channel since

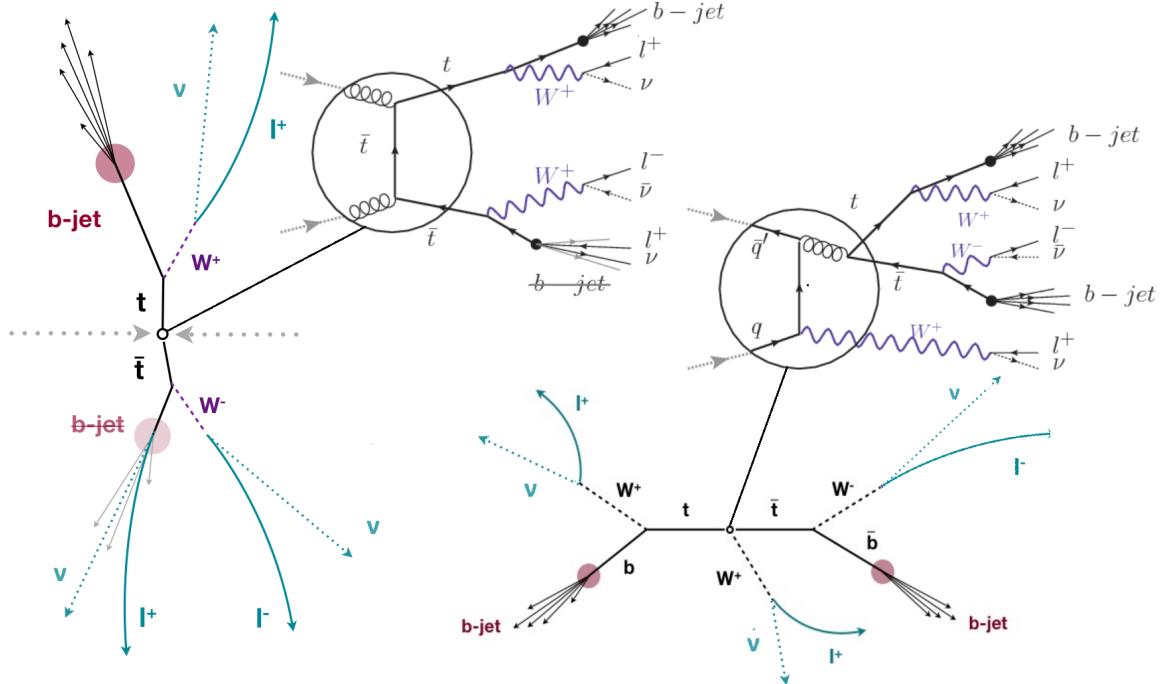


Figure 6.3: $t\bar{t}$ (left) and $t\bar{t}W$ (right) events signature as they would be seen in the detector; the Feynman diagrams including the whole evolution up to the final state are also showed. The $t\bar{t}$ process signature is very similar to that of the signal process with one fake lepton and no forward activity. The $t\bar{t}W$ process present a higher b-jet multiplicity compared to the signal process, a prompt lepton and no forward activity.

2370 leptons in processes like $t\bar{t}$ + jets or Z + jets can be charge misidentified, leading to
 2371 backgrounds increments. An identification variable have been designed in order to
 2372 reject this type of background events.

2373 6.4 Data and MC Samples

2374 6.4.1 Full 2016 data set

2375 The data set used in this analysis was collected by the CMS experiment during 2016
 2376 at while running at $\sqrt{s} = 13\text{TeV}$ and corresponds to a total integrated luminosity
 2377 of 35.9fb^{-1} . Only periods when the CMS magnet was on were considered when

2378 selecting the data samples; that corresponds to the 23Sep2016 (Run B to G) and
 2379 PromptReco (Run H) versions of the datasets.

2380 Multilepton final states with either two same-sign leptons or three leptons tar-
 2381 get the case where the Higgs boson decays to a pair of W bosons, τ leptons, or Z
 2382 bosons, and where the top quark decays leptonically, hence, the SingleElectron,
 2383 SingleMuon, DoubleEG, MuonEG, DoubleMuon dataset (see Table A.1) compose the
 2384 full dataset. The certified luminosity sections are selected using the golden JSON file
 2385 defined by the CMS experiment [148].

2386 6.4.2 Triggers

2387 The events considered are those online-reconstructed events triggered by one, two, or
 2388 three leptons. Single-lepton triggers are included in order to boost the acceptance
 2389 of events where the p_T of the sub-leading lepton falls below the threshold of the
 2390 double-lepton triggers. The trigger efficiency is increased by including double-lepton
 2391 triggers in the $3l$ category, and single-lepton triggers in all categories; it is possible
 2392 given the logical “or” of the trigger decisions of all the individual triggers in a given
 2393 category. Table A.2 shows the lowest-threshold non-prescaled triggers present in the
 2394 High-Level Trigger (HLT) menus for both Monte-Carlo and data in 2016.

2395 Trigger efficiency scale factors

2396 Trigger efficiency describes the ability of events to pass the trigger requirements. It
 2397 is measured in simulated events using generator information given that there is no
 2398 trigger bias with the MC sample. Measuring the trigger efficiency in data requires a
 2399 more elaborated procedure; first, select a set of events collected by a trigger that is
 2400 uncorrelated with the lepton triggers such that the selected events form an unbiased

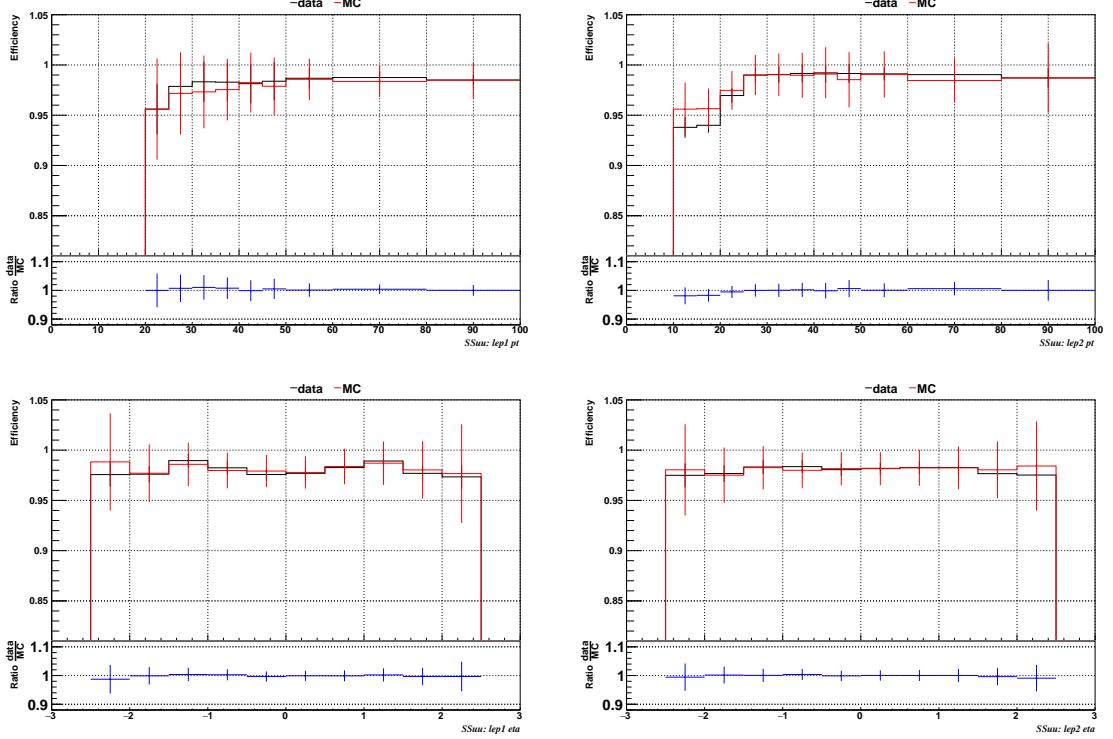


Figure 6.4: Comparison between data and MC trigger efficiencies in the same-sign $\mu\mu$ category, as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right) [149].

sample. In this analysis, that uncorrelated trigger is a MET trigger. Second step is looking for candidate events with exactly two good leptons (exactly three good leptons for the $3l$ channel). Finally, measure the efficiency for the candidate events to pass the logical “or” of triggers being considered in a given event category as defined in Table A.2.

Comparisons between the data and MC efficiencies for each category, showed in Figures 6.4, 6.5, and 6.6, reveal that they are in good agreement; the difference is corrected by applying scale factors derived from the ratio between both efficiencies.

Applied flat scale factors in each category are shown in Tab. 6.1; they have been inherited from Reference [149].

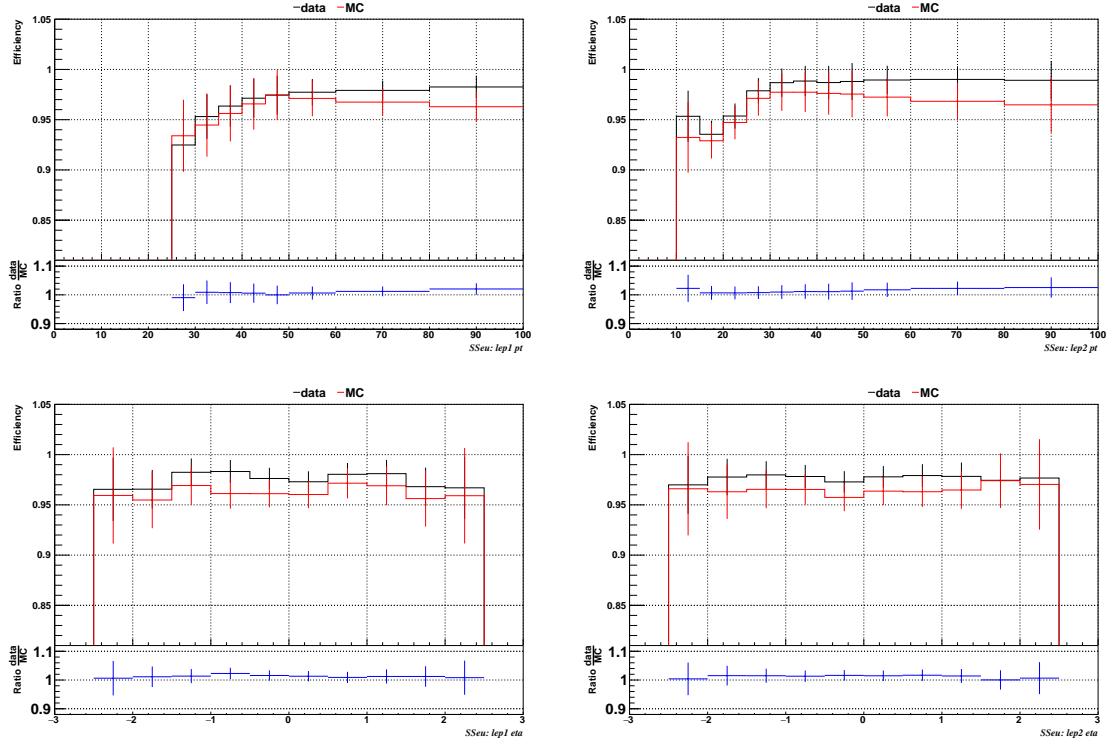


Figure 6.5: Comparison between data and MC trigger efficiencies in the same-sign $e\mu$ category as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right) [149].

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

Table 6.1: Trigger efficiency scale factors and associated uncertainties, shown here rounded to the nearest percent.

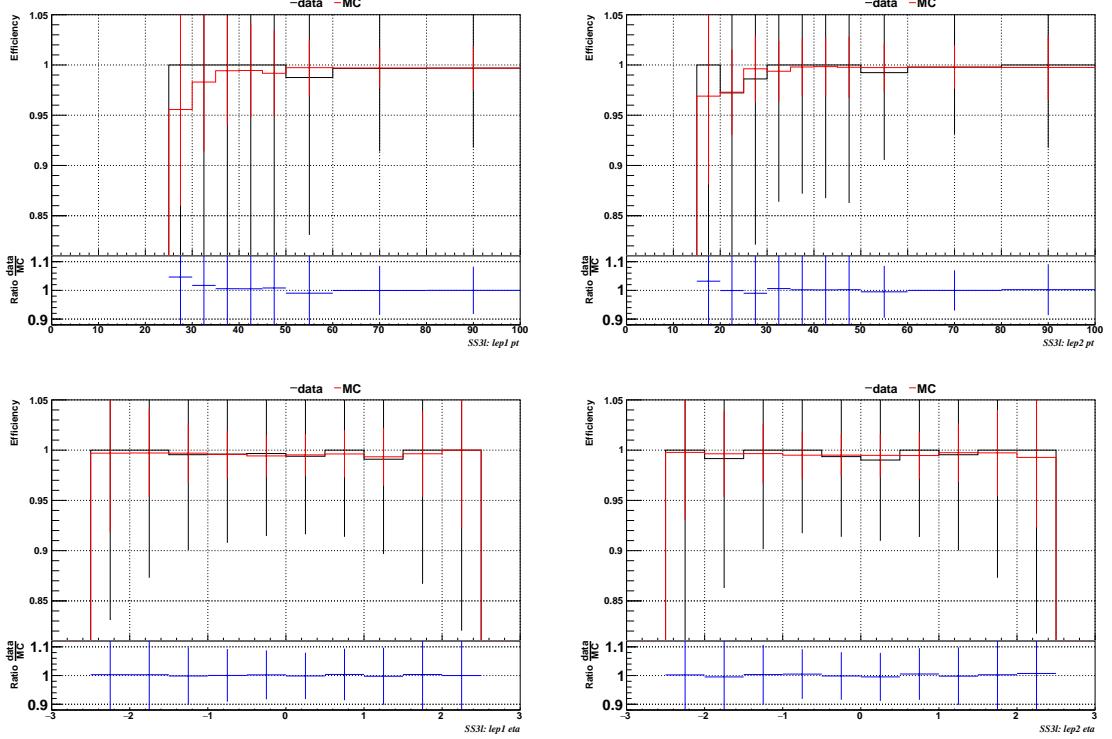


Figure 6.6: Comparison between data and MC trigger efficiencies in the $3l$ category, as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right) [149].

2411 6.4.3 Signal modeling and MC samples

2412 Current event generators allow for adjusting the kinematics of the generated events,
 2413 based on an event-wise reweighting; in this way, several generation parameters phase
 2414 spaces can be explored according to the experimental interests. The signal samples
 2415 used in this analysis were generated in such a way that not only the case $\kappa_t = -1$, but
 2416 an extended range of κ_t and κ_V values may be investigated.

2417 tHq and tHW cross section in the κ_t - κ_V phase space are shown in Figure 6.7. As
 2418 said in section 3.1, the tHq sample was generated using the 4F scheme which provides
 2419 a better description of the additional b quark from the initial gluon splitting, while the
 2420 tHW sample was generated using the 5F scheme in order to remove its interference
 2421 with $t\bar{t}H$ at LO.

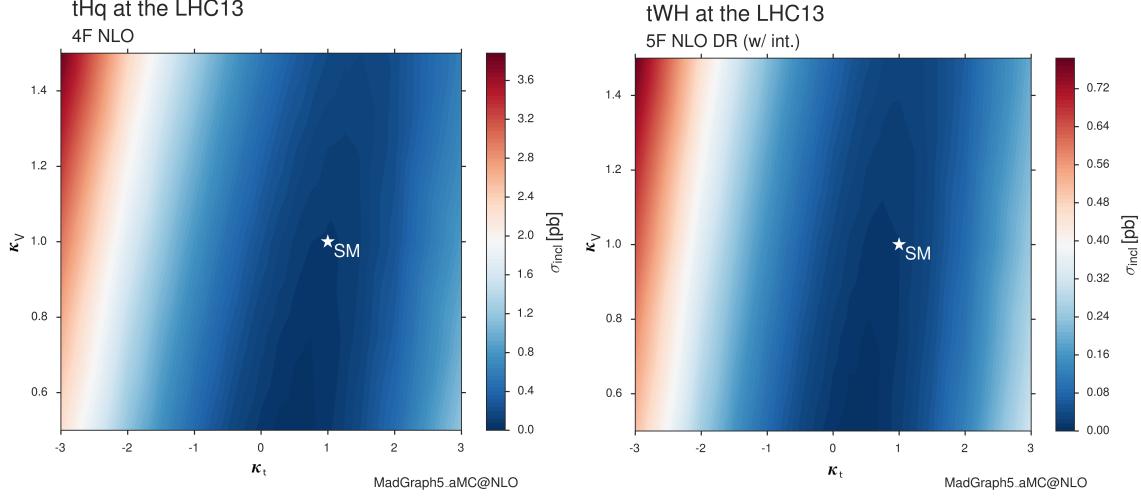


Figure 6.7: tHq and tHW cross section in the κ_t - κ_V phase space [150].

2422 MC signal samples

2423 The two signal samples, tHq and tHW , correspond to the `RunIISummer16MiniAODv2`
 2424 campaign produced with `CMSSW_80X`; they were produced with `MG5_aMC@NLO`
 2425 (version 5.2.2.3), in LO order mode at $\sqrt{s} = 13$ TeV, and are normalized to NLO cross
 2426 sections (see Table 6.2). The Higgs boson is assumed to be SM-like except for the
 2427 values of its couplings to the top quark and W boson. Each sample was generated
 2428 with a set of event weights corresponding to 51 different values of (κ_t, κ_V) couplings,
 2429 accessible in terms of LHE event weights as shown in Table A.3; however, the main
 2430 interest is the $(\kappa_t = -1, \kappa_V = 1)$ case.

Sample	σ [pb]	BF
<code>/THQ_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1/</code>	0.7927	0.324
<code>/THW_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1/</code>	0.1472	1.0
<code>/tthJetToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8_mWCutfix/</code>	0.2151	1.0

Table 6.2: MC signal samples used in this analysis; cross section and branching fraction are also listed [150].

2431 The $t\bar{t}H$ sample was produced using `AMC@NLO` interfaced to `PYTHIA 8` for
 2432 the parton shower, and is scaled to NLO cross sections. The $t\bar{t}H$ cross section depends

2433 quadratically on κ_t ; however, in contrast to the tHq and tHW samples, the scaling
 2434 is not performed during the sample generation process but in the analysis code since
 2435 it was decided to include the $t\bar{t}H$ process as part of the signal in the course of the
 2436 analysis.

2437 **MC background samples**

2438 Several MC generators were used to generate the samples of the background processes.
 2439 The dominant background sources ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$) were produced using AMC@NLO
 2440 interfaced to PYTHIA8, and are scaled to NLO cross sections. Other minor back-
 2441 ground processes are simulated using POWHEG interfaced to PYTHIA, or bare
 2442 PYTHIA as stated in the sample names in Table A.4. Pileup interactions are in-
 2443 cluded in the simulation in order to reflect the observed multiplicity in data; the
 2444 simulated events are weighted according to the actual pileup in data, estimated from
 2445 the measured bunch-to-bunch instantaneous luminosity and the total inelastic cross
 2446 section, 69.2 mb. All events are finally passed through a full simulation of the CMS
 2447 detector based on GEANT4, and reconstructed using the same algorithms as used for
 2448 the data.

2449 **6.5 Object Identification**

2450 In this section, the specific definitions of the physical objects in terms of the numerical
 2451 values assigned to the reconstruction parameters are presented; thus, the provided
 2452 details summarize and complement the descriptions presented in previous chapters.
 2453 The object reconstruction and selection strategy used in this thesis is inherited from
 2454 the analyses in References [145, 149], thus, the information in this section is extracted
 2455 from those documents unless other References are stated.

2456 **6.5.1 Lepton reconstruction and identification**

2457 Two types of leptons are defined in this analysis: *signal leptons* are those coming from
 2458 W, Z and τ decays which usually are isolated from other particles; *background leptons*
 2459 are defined as leptons produced in b -jet hadron decays, light-jets misidentification,
 2460 and photon conversions.

2461 The process of reconstruction and identification of electron and muon candidates
 2462 was described in chapter3, hence, the identification variables used in order to retain
 2463 the highest possible efficiency for signal leptons while maximizing the rejection of
 2464 background leptons are listed and described in the following sections ².

2465 The identification variables include not only observables related directly to the re-
 2466 constructed leptons themselves, but also to the clustered energy deposits and charged
 2467 particles in a cone around the lepton direction (jet-related variables); an initial loose
 2468 preselection of leptons candidates is performed and then an MVA discriminator, re-
 2469 ferred to as *lepton MVA* discriminator, is used to distinguish signal leptons from
 2470 background leptons.

2471 **Muons**

2472 The Physics Objects Groups (POG) at CMS, are in charge of studying and defining
 2473 the set of selection criteria applied on the course of reconstruction and identification
 2474 of particles. These selection criteria are implemented in the CMS framework in the
 2475 form of several object identification working points according to the strength of the
 2476 requirements.

2477 The muon candidates are reconstructed by combining information from the tracker
 2478 system and the muon detection system of CMS detector and the POG defined three

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

2479 working points for muon identification *MuonID* [153];

- 2480 • *POG Loose Muon ID* is a particle identified as a muon by the PF event re-
 2481 construction and also reconstructed either as a global-muon or as an arbitrated
 2482 tracker-muon. This identification criteria is designed to be highly efficient for
 2483 prompt muons and for muons from heavy and light quark decays; it can be com-
 2484 plemented by applying impact parameter cuts in analyses with prompt muon
 2485 signals.
- 2486 • *POG Medium Muon ID* is a Loose muon with additional track-quality and
 2487 muon-quality (spatial matching between the individual measurements in the
 2488 tracker and the muon system) requirements. This identification criteria is de-
 2489 signed to be highly efficient in the separation of the muons coming from decay
 2490 in flight of heavy quarks and muons coming from B meson decays as well as
 2491 prompt muons. An additional category *MVA Prompt ID* is defined in this iden-
 2492 tification criteria directed to discriminated muons from B mesons and prompt
 2493 muons (from W,Z and τ decays). The Medium ID provides the same fake rate as
 2494 the Tight Muon ID but a higher efficiency on prompt and B-decays muons. [154]
- 2495 • *POG Tight Muon ID* is a global muon with additional muon-quality require-
 2496 ments Tight Muon ID selects a subset of the PF muons.

2497 Only muons within the muon system acceptance $|\eta| < 2.4$ and minimum p_T of 5
 2498 GeV are considered.

2499 **Electrons**

2500 Electrons are reconstructed using information from the tracker and from the electro-
 2501 magnetic calorimeter and identified by an MVA algorithm (*MVA eID* discriminant)

2502 using the shape of the calorimetric shower variables like the shape in η and ϕ , the clus-
 2503 ter circularity, widths along η and ϕ ; track-cluster matching variables like E_{tot}/p_{in} ,
 2504 E_{Ele}/p_{out} , $\Delta\eta_{in}$, $\Delta\eta_{out}$, $\Delta\phi_{in}$, $1/E - 1/p$; and track quality variables like χ^2 of the
 2505 GSF tracks, the number of hits used by the GSF filter [155].

2506 A loose selection based on η -dependent cuts on this discriminant is used to prese-
 2507 lect electron candidates, the full shape of the discriminant is used in the lepton MVA
 2508 selection to separate signal leptons from background leptons (described in Section
 2509 6.5.1).

2510 In order to reject electrons from photon conversions, electron candidates with
 2511 missing hits in the pixel tracker layers or matched to a conversion secondary vertex
 2512 are discarded. Electrons are selected for the analysis if they have $p_T > 7$ GeV and
 2513 are located within the tracker system acceptance region ($|\eta| < 2.5$).

2514 **Lepton vertexing and pile-up rejection**

2515 The impact parameter in the transverse plane d_0 , impact parameter along the z -
 2516 axis d_z , and the impact parameter significance in the detector space SIP_{3D} , are
 2517 considered to perform the identification and rejection of pile-up, misreconstructed
 2518 tracks, and background leptons from b-hadron decays; pile-up and misreconstructed
 2519 track mitigation is achieved by imposing loose cuts on the impact parameter variables.
 2520 The full shape of the those variables is used in a lepton MVA classifier to achieve the
 2521 best separation between the signal and the background leptons.

2522 **Lepton isolation**

2523 PF is able to recognize leptons from two different sources: on one side, leptons from
 2524 the decays of heavy particles, such as W and Z bosons, which are normally isolated
 2525 in space from the hadronic activity in the event; on the other side, leptons from the

2526 decays of hadrons and jets misidentified as leptons, which are not isolated as the
 2527 former. For highly boosted systems, like the lepton and the b -jet generated in the
 2528 semileptonic decay of a boosted top, the decay products tend to be more closer and
 2529 sometimes they even overlap; thus, the PF standard definition of isolation in terms of
 2530 the separation between the lepton candidates and other PF objects in the η - ϕ plane,

$$\Delta R = \sqrt{(\eta^l - \eta^i)^2 + (\phi^l - \phi^i)^2} < 0.3 \quad (6.1)$$

2531 which considers all the neutral, charged hadrons and photons in a cone around the
 2532 leptons, is refocused to the local isolation of the leptons through the mini-isolation
 2533 I_{mini} [156] defined as the sum of particle flow candidates p_T within a cone around
 2534 the lepton, corrected for the effects of pileup and divided by the lepton p_T

$$I_{mini} = \frac{\sum_R p_T(h^\pm) - \max\left(0, \sum_R p_T(h^0) + p_T(\gamma) - \rho \mathcal{A}\left(\frac{R}{0.3}\right)^2\right)}{p_T(l)} \quad (6.2)$$

2535 where ρ is the pileup energy density, h^\pm, h^0, γ, l , represent the charged hadron, neutral
 2536 hadrons, photons, and the lepton, respectively. The radius R of the cone depends on
 2537 the p_T of the lepton according to

$$R = \frac{10\text{GeV}}{\min(\max(p_T(l), 50\text{GeV}), 200\text{GeV})}, \quad (6.3)$$

2538 The p_T dependence of the cone size allows for greater signal efficiency. Setting a
 2539 cut on I_{mini} below a given threshold ensures that the lepton is locally isolated, even
 2540 in boosted systems. The effect of pileup is mitigated using the so-called effective area
 2541 correction \mathcal{A} listed in Table 6.3.

2542 A loose cut on I_{mini} is applied to pre-select the muon and electron candidates;

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

Table 6.3: Effective areas, for electrons and muons used to mitigate the effect of pileup by using the so-called effective area correction.

however, the full shape is used in the lepton MVA discriminator when performing the signal lepton selection.

Jet-related variables

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

$$p_T^{rel} = \frac{(\vec{p}_{jet} - \vec{p}_l) \cdot \vec{p}_l}{||\vec{p}_{jet} - \vec{p}_l||}. \quad (6.4)$$

LeptonMVA discriminator

Electrons and muons passing the basic selection process described above are referred to as *loose leptons*. Additional discrimination between signal leptons and background leptons is crucial considering that the rate of $t\bar{t}$ production is much larger than the signal, hence, an overwhelming background from $t\bar{t}$ production. To maximally ex-

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

2558 ploit the available information in each event to that end, the dedicated lepton MVA
 2559 discriminator, based on a boosted decision tree (BDT) algorithm, has been built so
 2560 that all the identification variables can be used together.

2561 The lepton MVA discriminator training is performed using simulated signal Loose
 2562 leptons from the $t\bar{t}H$ MC sample and fake leptons from the $t\bar{t}$ +jets MC sample,
 2563 separately for muons and electrons. The input variables used include vertexing, iso-
 2564 lation and jet-related variables, the p_T and η of the lepton, the electron MVA eID
 2565 discriminator and the muon segment-compatibility variables. An additional require-
 2566 ment known as *tight-charge* requirement, is imposed by comparing two independent
 2567 measurement of the charge, one from the ECAL supercluster and the other from the
 2568 tracker; thus, the consistency in the measurements of the electron charge is ensured
 2569 so that events with a wrong electron charge assignment are rejected; this variable is
 2570 particularly used in the $2lss$ channel to suppress opposite-sign events for which the
 2571 charge of one of the leptons has been mismeasured. The tight-charge requirement for
 2572 muons is represented by the requirement of a consistently well measured track trans-
 2573 verse momentum given by $\Delta p_T/p_T < 0.2$. Leptons are selected for the final analysis
 2574 if they pass a given threshold of the BDT output, and are referred to as *tight leptons*
 2575 in the following.

2576 The validation of the lepton MVA algorithm and the lepton identification variables
 2577 is performed using data in various control regions; the details about that validation
 2578 are not discussed here but can be found in Reference [149].

2579 Selection definitions

2580 Electron and muon object identification is defined in three different sets of selections
 2581 criteria; the *Loose*, *Fakeable Object*, and *Tight* selection. These three levels of selection
 2582 are designed to serve for event level vetoes, the fake rate estimation application region,

and the final signal selection, respectively. The p_T of fakeable objects is defined as $0.85 \times p_T(\text{jet})$, where the jet is the one associated to the lepton object. This mitigates the dependence of the fake rate on the momentum of the fakeable object and thereby improves the precision of the method.

Tables 6.4 and 6.5 list the full criteria for the different selections of muons and electrons.

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

Table 6.4: 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.

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

6.5.2 Lepton selection efficiency

Efficiencies of reconstruction and selecting loose leptons are measured both for muons and electrons using a tag and probe method on both data and MC, using $Z \rightarrow \ell^+ \ell^-$ [157]. The scale factors are derived from the ratio of efficiencies $\varepsilon_i(p_T, \eta)$ measured

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

Table 6.5: Criteria for each of the three electron selections. In cases where the cut values change between selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. In some cases, the cut values change for different η ranges. These ranges are $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $1.479 < |\eta| < 2.5$ and the respective cut values are given in the form (value₁, value₂, value₃). For the two p_T^{ratio} and CSV rows, the cuts marked with a \dagger are applied to leptons that fail the lepton MVA cut, while the loose cut value is applied to those that pass the lepton MVA cut.

2596 for a given lepton in data/MC, according to

$$\rho(p_T, \eta) = \frac{\varepsilon_{\text{data}}(p_T, \eta)}{\varepsilon_{\text{MC}}(p_T, \eta)}. \quad (6.5)$$

2597 The scale factor for each event is used to correct the weight of the event in the
 2598 full sample; therefore, the full simulation correction is given by the product of all
 2599 the individual scale factors. The scale factors used in this thesis are inherited from
 2600 Reference [149] which in turns inherited them from leptonic SUSY analyses using
 2601 equivalent lepton selections.

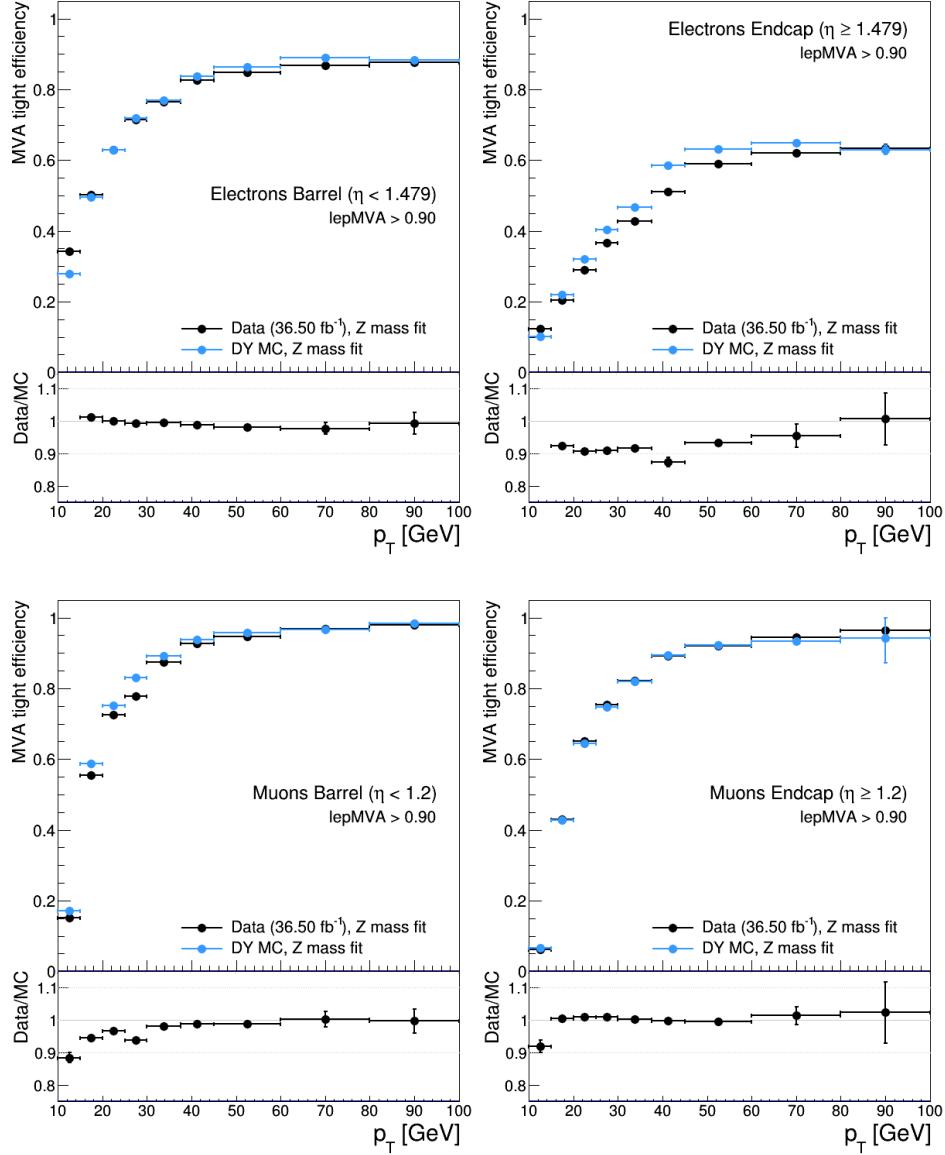


Figure 6.8: Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the $2lss$ definition, i.e., including the tight-charge requirement.

2602 The efficiency of applying the tight selection as defined in Tables 6.4 and 6.5, on the
 2603 loose leptons are determined by using a tag and probe method on a sample of Drell-
 2604 Yan enriched events. Figures 6.8 and 6.9 show the efficiencies for the $2lss$ channel and
 2605 $3l$ channel respectively. Efficiencies in the $2lss$ channel have been produced including
 2606 the tight-charge requirement, while for the $3l$ channel it is not included. Number

of passed and failed probes are determined from a fit to the invariant mass of the dilepton system. Simulation is corrected using these scale factors; note that they depends on η and p_T .

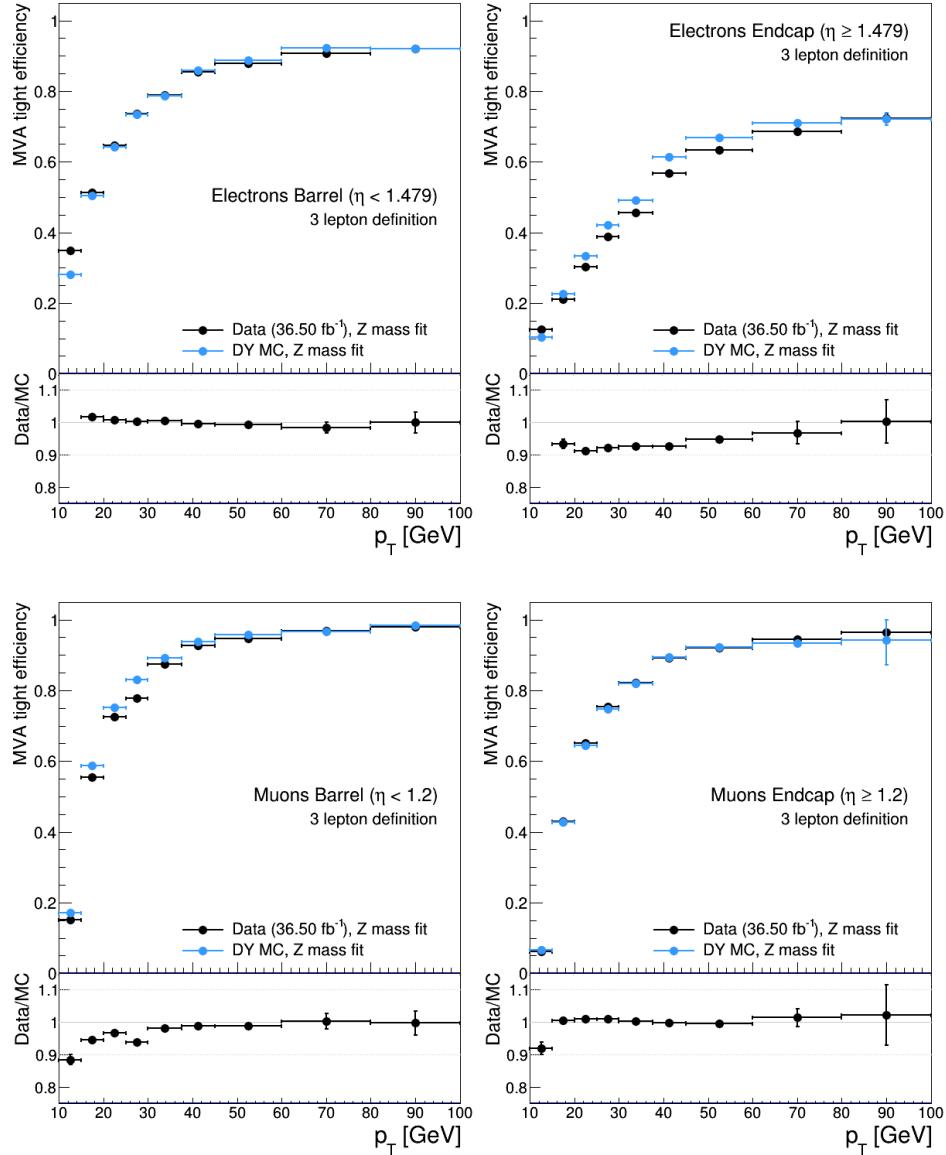


Figure 6.9: Tight vs loose selection efficiencies for electrons (top), and muons (bottom), for the $3l$ channel not including the tight-charge requirement.

2610 **6.5.3 Jets and b -jet tagging**

2611 In this analysis, jets are reconstructed by clustering PF candidates using the anti- k_t
 2612 algorithm with parameter distance $\Delta R = 0.4$; those charged hadrons that are not
 2613 consistent with the selected primary vertex are discarded from the clustering. The
 2614 jet energy is then corrected for the varying response of the detector as a function
 2615 of transverse momentum p_T and pseudorapidity η . Jets are selected for use in the
 2616 analysis only if they have $p_T > 25$ GeV and are separated from any selected leptons
 2617 by $\Delta R > 0.4$.

2618 Jets coming from the primary vertex and jets coming from pile-up vertices are
 2619 distinguished using a MVA discriminator based on the differences in the jet shapes,
 2620 in the relative multiplicity of charged and neutral components, and in the different
 2621 fraction of transverse momentum which is carried by the hardest components. Jet
 2622 tracks are also required to be compatible with the primary vertex.

2623 Jets originated from the hadronization of a b quark are selected using a MVA
 2624 likelihood discriminant which uses track-based lifetime information and reconstructed
 2625 secondary vertices (CSV algorithm). Only jets within the CMS tracker acceptance
 2626 ($\eta < 2.4$) are identified with this tool. Data samples are used to measure the efficiency
 2627 of the b -jet tagging and the probability to misidentify jets from light quarks or gluons;
 2628 in both cases the measurements are parametrized as a function of the jet p_T and η
 2629 and later used to correct differences between the data and MC simulation in the b
 2630 tagging performance, by applying per-jet weights to the simulation, dependent on
 2631 the jet p_T , η , b tagging discriminator, and flavor (from simulation truth) [151]. The
 2632 per-event weight is taken as the product of the per-jet weights, including those of the
 2633 jets associated to the leptons. The weights are derived on $t\bar{t}$ and Z+jets events.

2634 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% [152]. 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.

6.5.4 Missing Energy MET

As stated in Section 3.4.1.1, 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

$$E_T^{\text{miss}} \text{LD} = 0.00397 * E_T^{\text{miss}} + 0.00265 * H_T^{\text{miss}} \quad (6.6)$$

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.

2655 6.6 Event selection

2656 Events are selected considering the features of the signal process and the decay sig-
 2657 nature as described in Section 6.2. At the trigger level, events are selected to contain
 2658 either one, two, or three leptons with minimal p_T thresholds:

- 2659 • single-lepton trigger → 24 GeV for muons and at 27 GeV for electrons
- 2660 • double-lepton triggers → leading and sub-leading leptons: 17 and 8 GeV for
 2661 muons and 23 and 12 GeV for electrons.
- 2662 • three-lepton triggers → threshold on the third hardest lepton in the event: 5
 2663 and 9 GeV for muons and electrons, respectively.

2664 The offline event selection level targets the specific topology of the tHq signal
 2665 with $H \rightarrow WW$ and $t \rightarrow Wb \rightarrow l\nu b$; therefore, the resulting state is composed of three
 2666 W bosons, one b quark, and a light spectator quark at high rapidity. The selection
 2667 criteria for the two channels exploited in this analysis are summarized in Table 6.6.

Same-sign $\ell\ell$ channel	$\ell\ell\ell$ channel
have fired one of the corresponding trigger paths	
No loose leptons with $m_{\ell\ell} < 12\text{GeV}$	
One or more b tagged jets (CSV medium) $ \eta < 2.4$	
One or more non-tagged jets: central → $p_T > 25\text{ GeV}$, $\eta < 2.4$ forward → $p_T > 40\text{ GeV}$, $\eta > 2.4$ $E_T^{\text{miss}} \text{LD} > 0.2$	
Exactly two tight same-sign leptons	Exactly three tight leptons
Lepton $p_T > 25/15\text{GeV}$	Lepton $p_T > 25/15/15\text{GeV}$
Electrons are triple-charge consistent.	No OSSF lepton pair with $ m_{\ell\ell} - m_Z < 15\text{GeV}$
Muon p_T resolution: $\Delta p_T/p_T < 0.2$.	
No ee pair with $ m_{ee} - m_Z < 10\text{GeV}$	

Table 6.6: Summary of event selection.

2668 In the $2lss$ channel, events with additional tight leptons are vetoed as well as those
 2669 for which a loose lepton pair has an invariant mass below 12 GeV. A threshold in p_T of

2670 the leading and sub-leading leptons is also required. events where the two electrons
 2671 have invariant mass within 10 GeV of the Z boson mass (Z -*veto*) are discarded in
 2672 order to reject events from DY+jets production with charge misidentified electrons.

2673 In the $3l$ lepton channel, leptons are required to have respectively $p_T > 25\text{GeV}$, $>$
 2674 15 GeV, and > 15 GeV. Events with an opposite-sign, same-flavor lepton combina-
 2675 tion (OSSF) with invariant mass within 15 GeV of the Z boson mass are discarded in
 2676 order to reject events from $WZ + \text{jets}$ production.

2677 The selection criteria in Table 6.6 represent a relatively loose selection that allows
 2678 to maintain a large signal efficiency while suppressing the main backgrounds. This
 2679 selection includes contributions from $H \rightarrow \tau\tau$ and $H \rightarrow ZZ$ as well. The events
 2680 obtained from the selection are then used to extract the signal contribution in a
 2681 second analysis step, using BDT discriminators against the main backgrounds of
 2682 $t\bar{t}W/t\bar{t}Z$ and non-prompt leptons from $t\bar{t}$. The shape of the discriminator variables
 2683 is then fit to the observed data distribution to estimate the signal and background
 2684 yields, simultaneously for all channels.

2685 6.7 Signal discrimination

2686 The production cross section for the signal processes tHq , tHW , and $t\bar{t}H$ is only
 2687 about 600 fb (the enhancement provided by inverted couplings, $\kappa_t = -1$ almost double
 2688 it), resulting in a small signal to background ratio even for a tight selection. A
 2689 multivariate method is hence employed to train a discriminator to separate tH signal
 2690 events from the dominant background events ($t\bar{t}$ and $t\bar{t}V$).

2691 **6.7.1 MVA classifiers evaluation**

2692 Several MVA classifier algorithms were evaluated in order to determine the most
 2693 appropriate method for this analysis⁴. The comparison is based on the performance
 2694 of the classifiers, encoded in the plot of the background rejection as a function of the
 2695 signal efficiency (ROC curve). The top row of Figure 6.10 shows the ROC curves
 2696 for the several methods evaluated; two separated training were performed in the $3l$
 2697 channel: against $t\bar{t}$ (right) and $t\bar{t}V$ (left) processes.

2698 In both cases the gradient boosted decision tree *BDTG* (*BDTA_GRAD* in the
 2699 plot) classifier offers the best results, followed by the adaptive BDT classifier (*BDTA*);
 2700 the several Fisher classifiers tested, which differ in their parameters and/or boosting
 2701 method, they offer similar performance among them, while the k-Nearest Neighbour
 2702 (kNN) classifier performance is below the rest of the classifiers. The corresponding
 2703 ROC curves and in the $2lss$ channel for trainings against $t\bar{t}V$ (left) and $t\bar{t}$ (right)
 2704 processes are shown in the bottom row of Figure 6.10; the BDTG performance is
 2705 similar to that in the $3l$ channel.

2706 **6.7.2 Discriminating variables**

2707 The classifier chosen to separate the tHq signal from the main backgrounds is the
 2708 *BDTG* classifier, trained on simulated signal and background events. The samples
 2709 used in the training are the tHq sample in Table 6.2, the samples in the third section
 2710 of table A.4 and the samples marked with an * in the same table.

2711 As explained in Section 5.1.1, a set of discriminating variables are given as input to
 2712 the *BDTG* which combines the individual discrimination power of each input variable

⁴ The choice of the tested algorithms was based on the experience from previous analyses and considering the expertise of the members of the tHq and $t\bar{t}H$ analyses groups. Only the BDT classifier is described in this thesis and a detailed description of all available methods can be found in Reference [127]

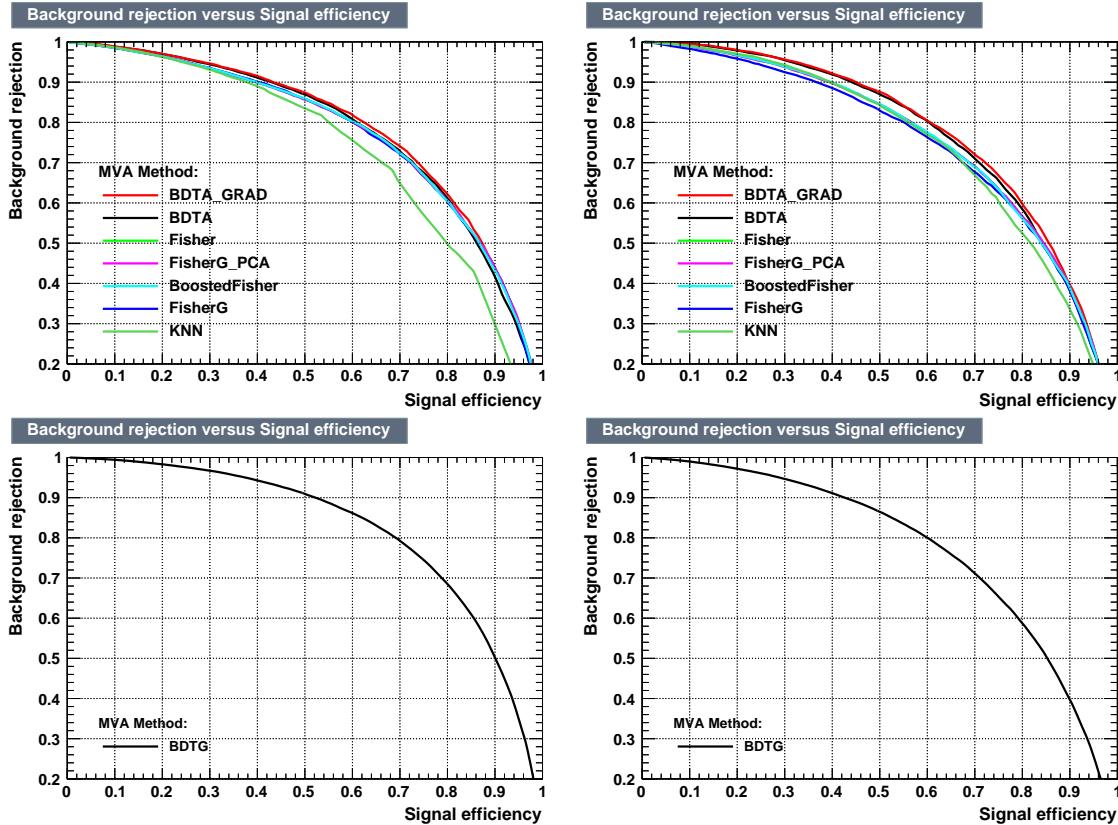


Figure 6.10: Top: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers in the $3l$ channel for training against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: background rejection vs signal efficiency (ROC curve) in the $2lss$ channel for a single discriminator: BDTG, against $t\bar{t}V$ (left) and $t\bar{t}$ (right).

2713 to produce a discriminator with the maximum discrimination power. Table 6.7 lists
 2714 the input variables used in the BDTG trainings for this analysis.

2715 The same set of input variables was used to produce the plots for MVA classifiers
 2716 evaluation.

2717 Plots in Figure 6.11 shows the BDTG input variables distributions for the signal
 2718 and background samples, in the $3l$ channels.

2719 All the input variables have some discrimination power, however, that power is
 2720 bigger for some of them; for instance, the third lepton p_T plot (top left in Figure 6.11)
 2721 shows some discrimination power against WZ and VVV backgrounds for which there

Variable name	Description
nJet25	Number of jets with $p_T > 25$ GeV, $ \eta < 2.4$
nJetEta1	Number of jets with $ \eta > 1.0$, non-CSV-loose
MaxEtaJet25	Max. $ \eta $ of any (non-CSV-loose) jet with $p_T > 25$ GeV
deltaFwdJetClosestLep	$\Delta\eta$ forward light jet and closest lepton
deltaFwdJetBJet	$\Delta\eta$ forward light jet and hardest CSV loose jet
deltaFwdJet2BJet	$\Delta\eta$ forward light jet and second hardest CSV loose jet
Lep3Pt/Lep2Pt	p_T of the 3 rd lepton (2 nd for ss2l)
totCharge	Sum of lepton charges
minDRll	Min ΔR any two leptons
dphiHighestPtSSPair	$\Delta\phi$ of highest p_T same-sign lepton pair

Table 6.7: BDTG input variables. First section lists variables related to jet multiplicities; second section lists variables related to forward jet activity, and third section lists variables related to lepton kinematics.

is a peak around 30 GeV while tHq peak around 18 GeV; although the discrimination power does not cover all the backgrounds, it counts for the final discriminator. A similar situation can be seen in the plot for the number of jets (row three, column two); $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ processes tend to have more jets compared to the tHq process. The discrimination power is more evident in other plots like in the plot of the maximum $|\eta|$ of the jets in the event (row two, column three). The same or equivalent input variables are found to be performing well for both $3l$ and $2lss$ channels. Figure B.1 shows the corresponding input variables distribution plots for the $2lss$ channel.

Discrimination power from BDTG classifier

The Discrimination power of the input variables can also be evaluated from the BDTG training, exclusively for the training samples, i.e., dominant backgrounds ($t\bar{t}$ and $t\bar{t}V$); the training samples are submitted to the selection cuts on Table 6.6.

Figure 6.12 shows the comparison between input variables for the two trainings in the $3l$ channel; it reveals that some variables show opposite behavior for the two background sources, which results in potentially screening the discrimination power

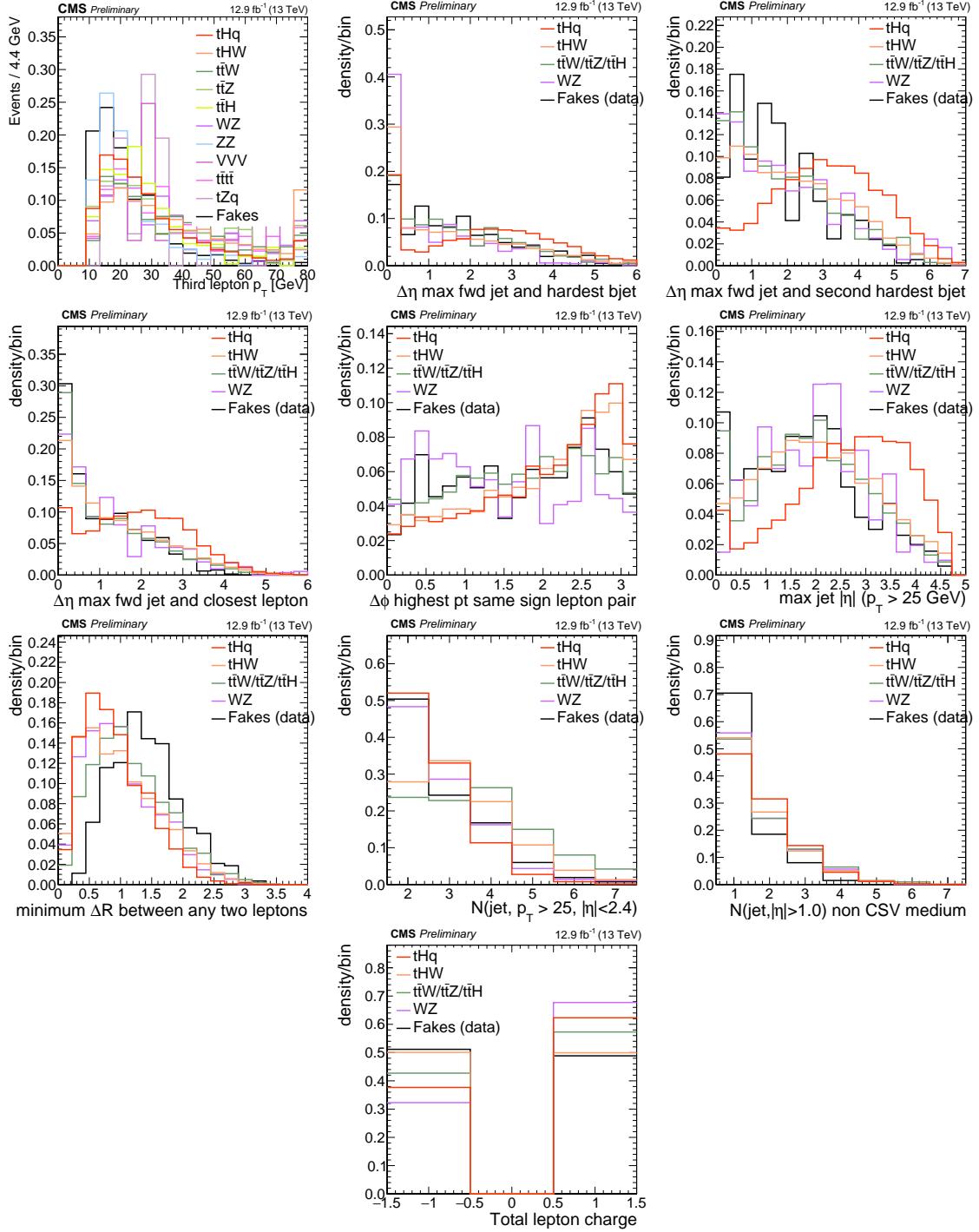


Figure 6.11: Distributions of the BDTG classifier input variables (not normalized) for signal discrimination in the $3l$ channel.

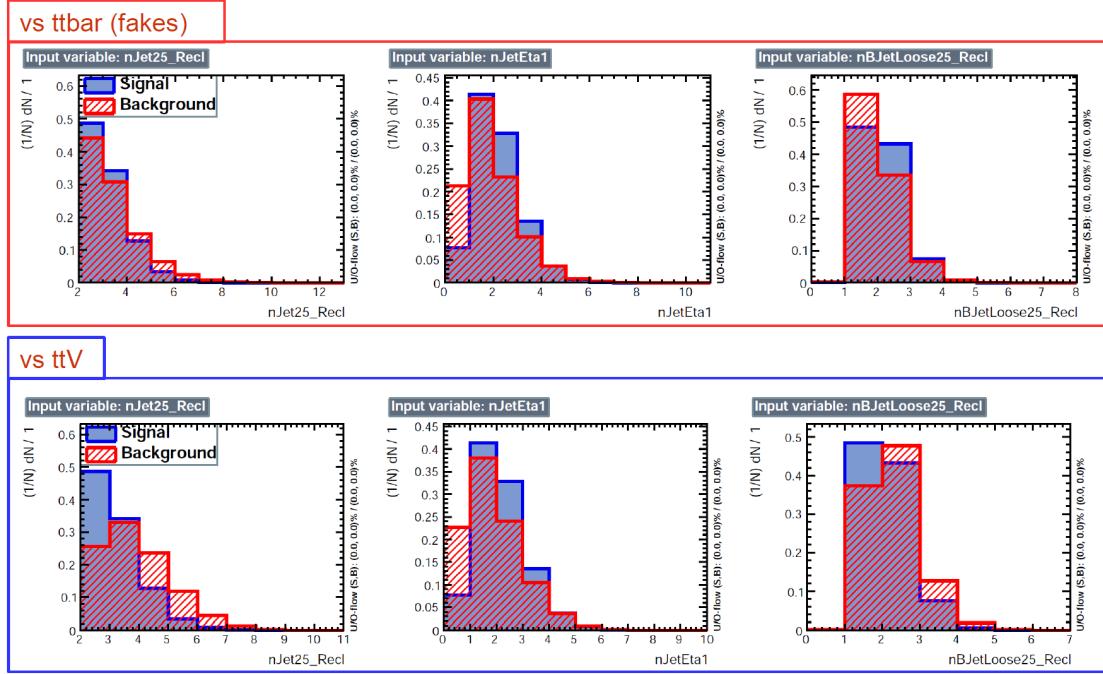


Figure 6.12: BDT input variables as seen by BDTG classifier for the $3l$ channel, tHq signal(blue) discriminated against $t\bar{t}V$ background (red).

if they were to be used in a single discriminant, i.e., if the training would join $t\bar{t}$ and $t\bar{t}V$. For some other variables the distributions are similar in both background cases. In contrast to the distributions in Figure 6.11 only the dominant backgrounds are included; however, the discrimination power agrees among plots.

Figures in the Appendix B.2, B.3, B.4, and B.5 show the input variables distributions for the $2lss$ and $3l$ channel as seen by the BDTG classifier.

Input variables correlations

From Table 6.7, it is clear that the input variables are correlated to some extend. These correlations play an important role for some MVA methods like the Fisher discriminant method in which the first step consist of performing a linear transformation to an phase space where the correlations between variables are removed. In

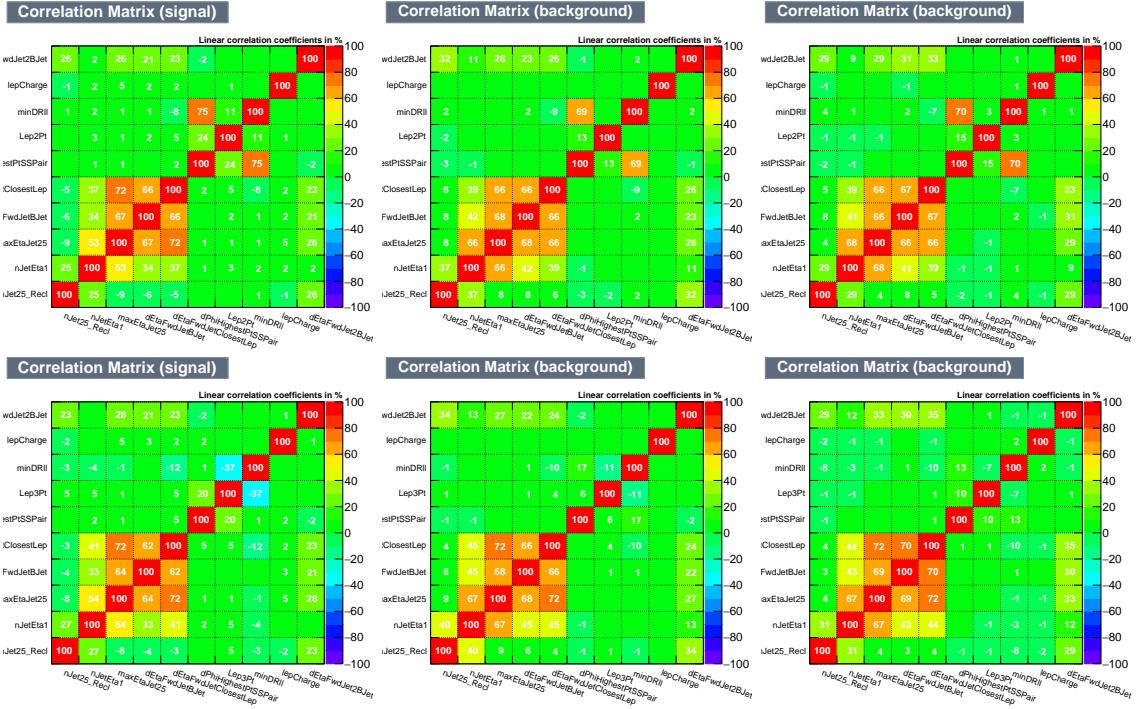


Figure 6.13: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the BDTG classifier for the $2lss$ (top) and $3l$ (bottom) channels.

the case of BDT, correlations do not affect the performance. Figure 6.13 shows the linear correlation coefficients for signal and background for the two training cases (the signal values are identical by construction). As expected, strong correlations appear for variables related to the forward jet activity.

2752 6.7.3 BDTG classifiers response

After the training stage, the BDTG classifier is tested to ensure its ability to discriminate between simulated signal and background events. The BDTG classifier output distributions for signal and backgrounds in the $3l$ channel are shown in Figure 6.14. As expected, a good discrimination power is obtained using default discriminator parameter values; some overtraining is also visible.

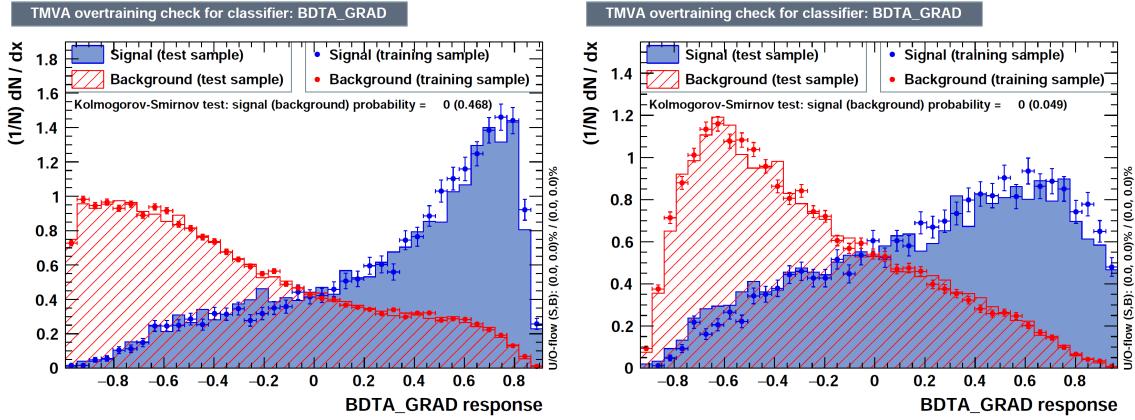


Figure 6.14: BDTG classifier output for trainings against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Default BDTG parameters have been used.

2758 In order to explore further optimization in the BDTG performance, several changes
 2759 from the default BDTG parameters were tested; Table 6.8 list the set of parameters
 2760 found to be most discriminant with minimal overtraining as shown in Figure 6.15.

TMVA.Types.kBDT		
Option	Default	Used
NTrees	200	800
BoostType	AdaBoost	Grad
Shrinkage	1	0.1
nCuts	20	50
MaxDepth	3	

Table 6.8: Configuration used in the final BDTG training. Parameters not listed were not tested.

2761 The ranking of the input variables by their importance in the classification process
 2762 is shown in Table 6.9; for both trainings the rankings show almost the same 5 variables
 2763 in the first places.

2764 6.7.4 Additional discriminating variables

2765 Given that the forward jet in background processes could be originated from pileup,
 2766 two additional discriminating variables accounting for that were tested. These addi-

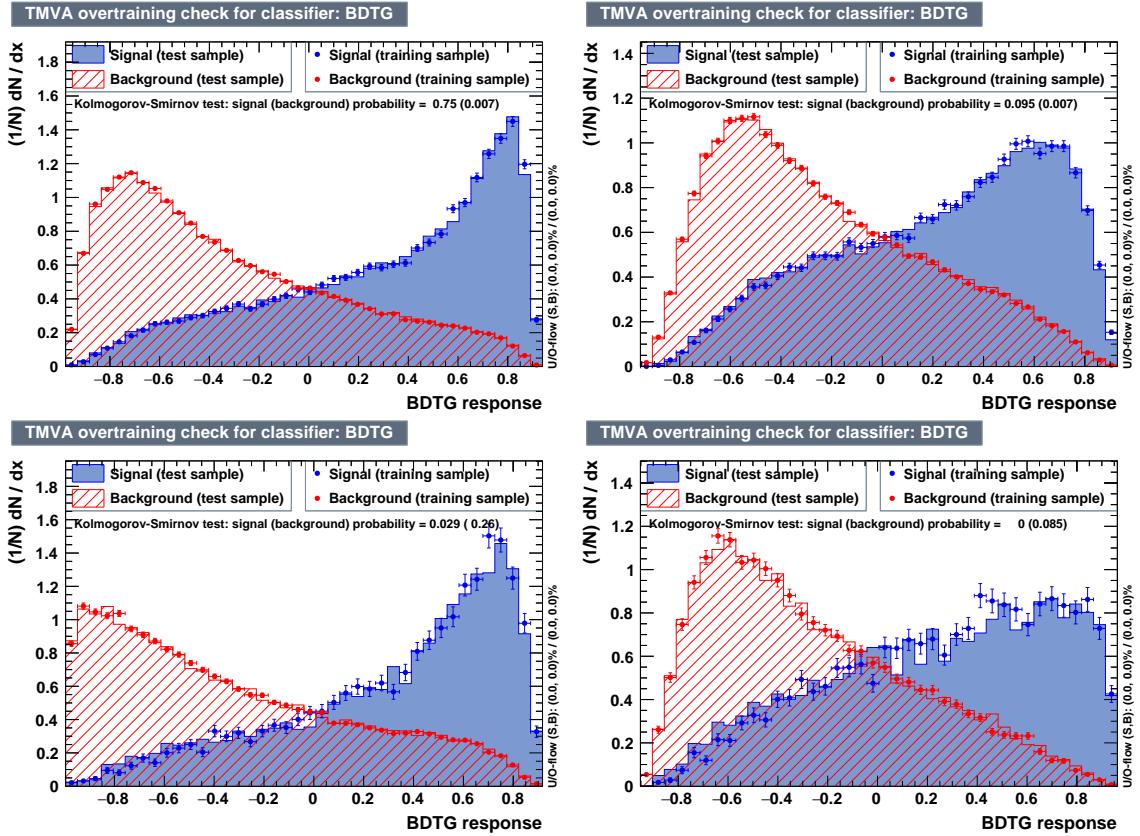


Figure 6.15: BDTG classifiers output for training against $t\bar{t}V$ (left) and $t\bar{t}$ (right) for $2lss$ channel(top) and $3l$ channel (bottom) .

2lss channel		3l channel		
Rank	$t\bar{t}$ training Variable	$t\bar{t}V$ training Variable	$t\bar{t}$ training Variable	
1	minDRll	dEtaFwdJetBJet	dEtaFwdJetClosestLep	$t\bar{t}V$ training Variable
2	dEtaFwdJetClosestLep	Lep3Pt	minDRll	maxEtaJet25
3	dEtaFwdJetBJet	maxEtaJet25	maxEtaJet25	dEtaFwdJet2BJet
4	dPhiHighestPtSSPair	dEtaFwdJet2BJet	dPhiHighestPtSSPair	dEtaFwdJetBJet
5	Lep3Pt	dEtaFwdJetClosestLep	Lep2Pt	Lep2Pt
6	maxEtaJet25	minDRll	dEtaFwdJetBJet	dEtaFwdJetClosestLep
7	dEtaFwdJet2BJet	dPhiHighestPtSSPair	dEtaFwdJet2BJet	minDRll
8	nJetEta1	nJet25	nJetEta1	nJet25
9	nJet25	nJetEta1	nJet25	dPhiHighestPtSSPair
10	lepCharge	lepCharge	lepCharge	nJetEta1

Table 6.9: Input variables ranking for BDTG classifiers for the trainings in the $3l$ channel and $2lss$ channel. In both trainings the rankings show almost the same 5 variables in the first places.

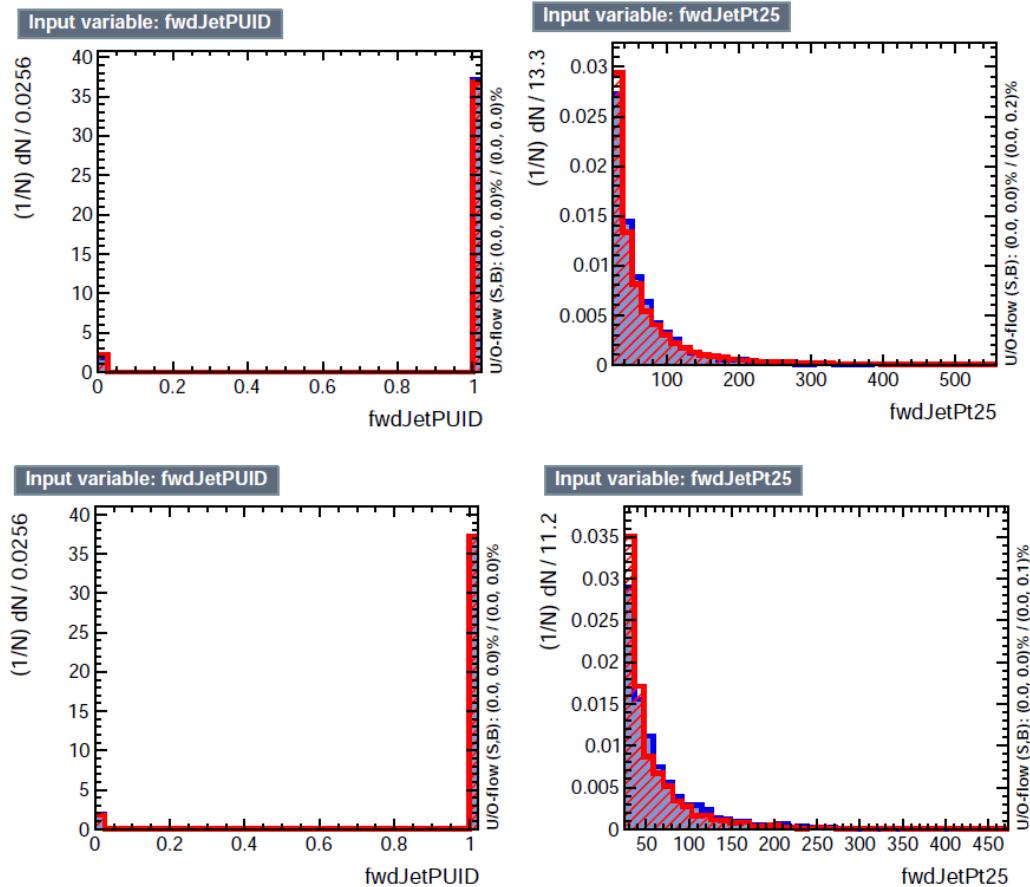


Figure 6.16: Additional discriminating variables distributions for $t\bar{t}V$ training (top row) and $t\bar{t}$ training (bottom row) in the $3l$ channel. The origin of the jets in the forward jet identification distribution is tagged as 0 for *pileup jets* while *primary vertex jets* are tagged as 1.

tional variables describe the forward jet momentum (fwdJetPt25) and the forward jet identification(fwdJetPUID); their distributions in the $3l$ channel are shown in Figure 6.16. The forward jet identification distribution show that for both, signal and background, jets are mostly originated in the primary vertex.

The testing was performed by including in the BDTG input one variable at a time, so the discrimination power of each variable can be evaluated individually, and then both simultaneously. fwdJetPUID was ranked in the last place in importance (11) in both training ($t\bar{t}V$ and $t\bar{t}$) while fwdJetPt25 was ranked 3 in the $t\bar{t}V$ training and 7 in

2775 the $t\bar{t}$ training. When training using 12 variables, fwdJetPt25 was ranked 5 and 7 in
 2776 the $t\bar{V}$ and $t\bar{t}$ trainings respectively, while fwdJetPUID was ranked 12 in both cases.

	ROC-integral	
	$t\bar{V}$	$t\bar{t}$
base 10 var	0.848	0.777
+ fwdJetPUID	0.849	0.777
+ fwdJetPt25	0.856	0.787
12 var	0.856	0.787

Table 6.10: ROC-integral for all the testing cases performed in the evaluation of the additional variables discriminating power. The improvement in the discrimination performance provided by the additional variables is about 1% .

2777 The improvement in the discrimination performance provided by the additional
 2778 variables is about 1%, so it was decided not to include them in the procedure. Table
 2779 6.10 show the ROC-integral for all the testing cases performed.

2780 6.7.5 Signal extraction procedure

2781 Once the two BDTG classifiers, introduced in the previous section, are trained against
 2782 the dominant backgrounds in each channel, they are used to classify the events in the
 2783 samples; their outputs are then used to evaluate the signal cross section limits in a
 2784 fit to the classifier shape. Figure 6.17 shows the expected output distributions in a
 2785 2D plane of one training vs. the other, i.e., $t\bar{V}$ vs. $t\bar{t}$. Top row shows the 2D planes
 2786 for tHq and tHW signals, while the bottom left plot shows the corresponding 2D
 2787 plane for the combined backgrounds, which are evaluated as in the final background
 2788 prediction, i.e., these are not the samples used in the BDTG training and this includes
 2789 data-driven backgrounds. The signal (combining of tHq and tHW) to background
 2790 ratio (S/B) is showed in the bottom right plot of Figure 6.17.

2791 Each event is now classified into one of ten 2D-bins according to its position in the
 2792 plane, as shown in Figure 6.18. The number of bins is chosen such that no bins are

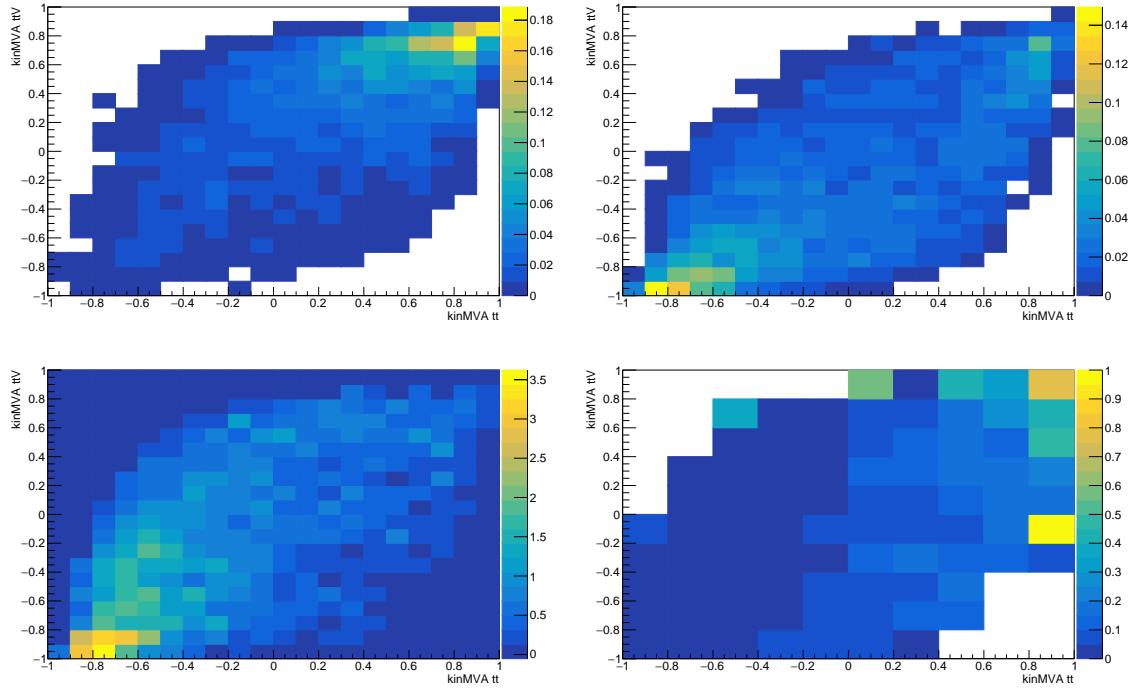


Figure 6.17: BDT classifier output planes (training vs $t\bar{t}$ on x-axis and vs $t\bar{t}V$ on y-axis) for the tHq and tHW signals (top row), and for the combined backgrounds (bottom left). Bottom right: S/B ratio (combining tHq and tHW) in the same plane. Plots are for $3l$ channel.

entirely empty for any process. The bin boundary positions and number of bins have been studied and optimized with respect to the expected limit on the signal strength (see Sec. 6.7.6).

From this event categorization, a 1D histogram of expected distribution is produced for each signal and background process, and fit to the observed data (or the Asimov dataset for expected limits).

6.7.6 Binning and selection optimization

The effect of the choice of pre-selection cuts and the number of bins of the 1D histogram on the cross section limit is evaluated by varying the most important cuts and re-calculating the limit in each case. In this analysis, the optimization was performed

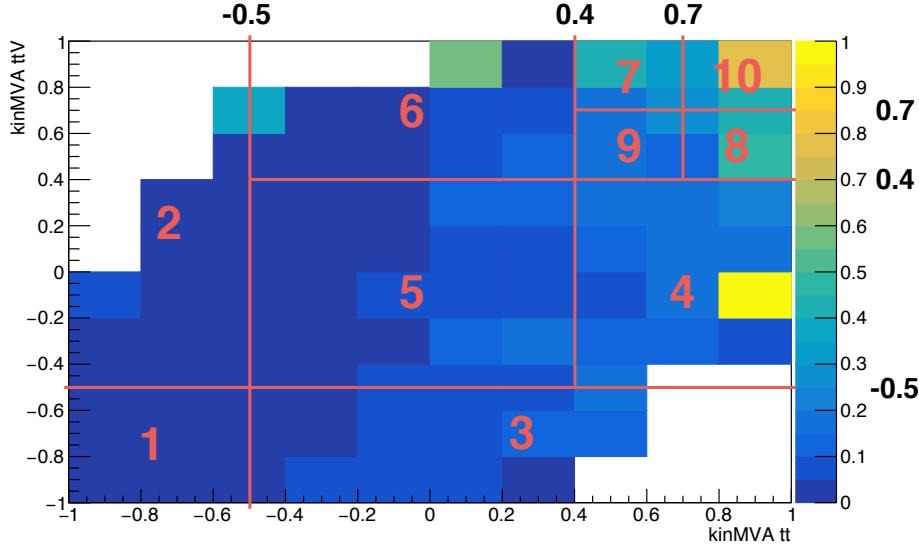


Figure 6.18: Binning overlaid on the S/B ratio map on the plane of classifier outputs.

in the $3l$ channel, by evaluating the upper limits on the $tHq + tHW$ expected signal strength only (without $t\bar{t}H$ component), always evaluated at $\kappa_t = -1.0$, $\kappa_V = 1.0$.
Table 6.11 shows the several variations explored, compared with a baseline; the baseline is similar to the selection reported in Table 6.6 but only a loose CSV jet and a Z veto of ± 10 GeV are required.

Selection	Variation	Expected limit
Baseline		< 2.93
Loose CSV tags	$\geq 1 \rightarrow \geq 2$	< 3.81
Medium CSV tags	$\geq 0 \rightarrow \geq 1$	< 2.76
Light forward jet η	$\geq 0 \rightarrow \geq 1$	< 2.94
Light forward jet η	$\geq 0 \rightarrow \geq 1.5$	< 3.00
MET>30 GeV		< 2.91
Z veto ($ m_{\ell\ell} - m_Z $)	$> 10\text{GeV} \rightarrow > 15\text{ GeV}$	< 2.79
One medium CSV + 15 GeV Z veto	combined	< 2.62

Table 6.11: Signal strength limit variation as a function of tighter cuts. The baseline selection corresponds to a looser selection compared to the one reported in Tab. 6.6 where only a CSV-loose b -jet is required, and the Z veto is loosened to ± 10 GeV. The optimal selection determined here corresponds to the baseline plus the two variations in the last row.

The optimal limit is found when requiring a slightly tighter selection with respect

2809 to the baseline. The optimal selection is reported in Table 6.6.

2810 The signal strength limit also depends on the chosen binning in the 2D plane as
 2811 the S/B ratio varies across the plane, hence, several sizes and binning combinations
 2812 were tested in order to improve the limit. Figure 6.19 shows some of the binning
 2813 combinations tested; in the default combination all the bins have the same size, while
 2814 the best limit was found for a set of 10 bins. The bin borders and the resulting limits
 2815 are shown in Table 6.12.

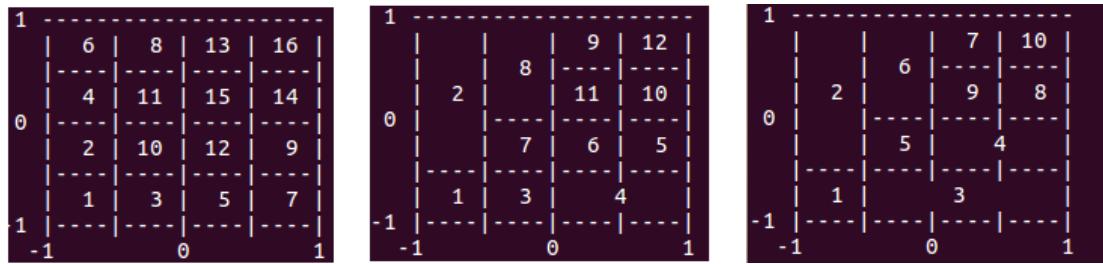


Figure 6.19: Binning combination scheme.

Number of bins	Bin borders						Expected limit
	x_1	x_2	x_3	y_1	y_2	y_3	
16 (default)	-0.5	0.0	0.5	-0.5	0.0	0.5	< 2.91
16	-0.5	0.3	0.7	-0.5	0.3	0.7	< 2.83
10	-0.5	0.0	0.5	-0.5	0.0	0.5	< 2.93
10	-0.5	0.0	0.7	-0.5	0.0	0.7	< 2.86
10	-0.5	0.0	0.7	-0.5	0.0	0.5	< 2.84
10	-0.5	0.0	0.5	-0.5	0.0	0.7	< 2.87
10	-0.5	0.4	0.7	-0.5	0.4	0.7	< 2.81

Table 6.12: Limit variation as a function of bin size. The final bin borders used in the $3l$ channel are indicated in bold.

2816 Combining the optimization of binning and using the tighter pre-selection cuts,
 2817 the expected limit in the $3l$ channel alone reaches **$r < 2.59$** .

2818 A similar binning optimization was made for $2lss$ channel, including other binning
 2819 combinations. First, the $3l$ channel binning was used to estimate the expected limit,

2820 then, bin borders were varied to obtain the best possible expected limit. The bin
 2821 borders and the resulting signal strength limits for the same-sign dimuon channel are
 2822 shown in Table 6.13:

Number of bins	Bin borders						Expected limit
	x_1	x_2	x_3	y_1	y_2	y_3	
16	-0.5	0.4	0.7	-0.5	0.4	0.7	< 1.72
12	-0.5	0.4	0.7	-0.5	0.4	0.7	< 1.72
12	-0.3	0.4	0.7	-0.5	0.4	0.7	< 1.71
12	-0.3	0.3	0.7	-0.5	0.4	0.7	< 1.71
12	-0.3	0.3	0.7	-0.4	0.4	0.7	< 1.70
12	-0.3	0.3	0.7	-0.3	0.4	0.7	< 1.70
12	-0.3	0.3	0.7	-0.3	0.2	0.7	< 1.68
12	-0.3	0.3	0.7	-0.3	0.1	0.7	< 1.70
12	-0.3	0.3	0.7	-0.3	0.2	0.6	< 1.70
10	-0.5	0.4	0.7	-0.5	0.4	0.7	< 1.75
10	-0.3	0.3	0.7	-0.3	0.2	0.6	< 1.69

Table 6.13: Limit variation as a function of bin size in the same-sign dimuon channel. (In bold: the final bin borders used in the $2lss$ channel.)

2823 The expected limit was found to be $\mathbf{r < 1.69}$ for optimized bin borders in 10 bins
 2824 and optimized pre-selection cuts.

2825 Two additional binning strategies were tested, however, the obtained limits are
 2826 degraded; they are documented in Appendix C.

2827 6.8 Signal model

2828 The goal of this analysis is to test the compatibility of points in the parameter space of
 2829 Higgs-to-vector boson and Higgs-to-top quark couplings. The simulated tHq , tHW ,
 2830 and $t\bar{t}H$ signal events are used with event-by-event weights to reflect the impact of the
 2831 couplings on kinematic distributions, and together with different predictions of the
 2832 respective production cross sections and branching ratios, we can produce limits for
 2833 different values of κ_V and κ_t . (See Tab. A.3 for the set of κ_t and κ_V values generated.)

2834 The slight shape-dependence of the BDT outputs as a function of the couplings is
 2835 documented in Appendix D.

2836 Apart from the κ_t/κ_V interference of the tHq and tHW production cross sections,
 2837 the cross section of $t\bar{t}H$ scales as κ_t^2 . Furthermore, the Higgs branching fractions to
 2838 vector bosons depend on κ_V , and the overall Higgs decay width depend both on κ_t
 2839 and κ_V when considering resolved top-quark loops in the $H \rightarrow \gamma\gamma$, $H \rightarrow Z\gamma$, and
 2840 $H \rightarrow gg$ decays. The relative contributions from $H \rightarrow WW$, $H \rightarrow ZZ$, and $H \rightarrow \tau\tau$
 2841 changes with changing κ_V .

2842 We hence set an upper limit on the combined cross section times branching ratio
 2843 of tHq , tHW , and $t\bar{t}H$.

2844 If we assume a modifier for the Higgs-to-tau coupling (κ_τ) to be equal to κ_t , the
 2845 relative fractions of WW , ZZ , and $\tau\tau$ in our selection will only depend on the ratio
 2846 of κ_t/κ_V . Any limit set at any given value of κ_t/κ_V is thus valid for all values of
 2847 κ_t and κ_V with that ratio, and could then be compared with theoretical predictions
 2848 of cross sections at different values of either modifier. Rather than as a function of
 2849 the κ_t/κ_V ratio, limits could (equivalently) be reported as a function of the relative
 2850 strength of Higgs-top and Higgs-vector-boson couplings, multiplied by the relative
 2851 sign. Such a parameter, further referred to as f_t , as defined in Equation 6.7, spans
 2852 the entire possible parameter space between -1.0 and 1.0 , with the SM expectation
 2853 at 0.5 . Absolute values of 1.0 or 0.0 would then correspond to purely Higgs-top and
 2854 purely Higgs-V couplings, respectively.

$$f_t = \text{sign}(\kappa_t/\kappa_V) \times \frac{\kappa_t^2}{\kappa_t^2 + \kappa_V^2}. \quad (6.7)$$

2855 Table 6.14 shows the points in the κ_t/κ_V and f_t parameter space that are mapped
 2856 by the 51 individual κ_t and κ_V points.

f_t	κ_t/κ_V	$\kappa_V = 0.5$	$\kappa_V = 1.0$	$\kappa_V = 1.5$
-0.973	-6.000	-3.00		
-0.941	-4.000	-2.00		
-0.900	-3.000	-1.50	-3.00	
-0.862	-2.500	-1.25		
-0.800	-2.000	-1.00	-2.00	-3.00
-0.692	-1.500	-0.75	-1.50	
-0.640	-1.333			-2.00
-0.610	-1.250		-1.25	
-0.500	-1.000	-0.50	-1.00	-1.50
-0.410	-0.833			-1.25
-0.360	-0.750		-0.75	
-0.308	-0.667			-1.00
-0.200	-0.500	-0.25	-0.50	-0.75
-0.100	-0.333			-0.50
-0.059	-0.250		-0.25	
-0.027	-0.167			-0.25
0.000	0.000	0.00	0.00	0.00
0.027	0.167			0.25
0.059	0.250		0.25	
0.100	0.333			0.50
0.200	0.500	0.25	0.50	0.75
0.308	0.667			1.00
0.360	0.750		0.75	
0.410	0.833			1.25
0.500	1.000	0.50	1.00	1.50
0.610	1.250		1.25	
0.640	1.333			2.00
0.692	1.500	0.75	1.50	
0.800	2.000	1.00	2.00	3.00
0.862	2.500	1.25		
0.900	3.000	1.50	3.00	
0.941	4.000	2.00		
0.973	6.000	3.00		

Table 6.14: The 33 distinct values of κ_t/κ_V and f_t as mapped by the 51 κ_t and κ_V points.

2857 The overall higgs decay width (modified by both κ_t and κ_V) becomes irrelevant
 2858 if limits are quoted as absolute cross sections rather than multiples of the expected
 2859 cross section (which depends on the overall Higgs decay width).

2860 The 1D histograms of events as categorized in regions of the 2D BDT plane is
 2861 then used in a maximum likelihood fit of signal and background shapes, where the
 2862 tHq , tHW , and $t\bar{t}H$ signals are floating with a common signal strength modifier r ,
 2863 producing a 95% C.L. upper limit the observed cross section of $tHq + tHW + t\bar{t}H$.

2864 This is done separately for each point of κ_t and κ_V , where the cross sections and
 2865 branching fractions are scaled accordingly in each point. Limits at fixed values of
 2866 κ_t/κ_V are by construction identical. Tables ??–?? and ??–?? in Appendix ?? show
 2867 the scalings of cross section times branching fraction, as well as branching fractions
 2868 alone for each of the Higgs decay modes and each of the signal components.

2869 6.9 Background modeling and predictions

2870 Irreducible backgrounds are reliably estimated from MC simulated events; there-
 2871 fore, in this analysis all backgrounds involving prompt leptons are estimated in this
 2872 way. Reducible backgrounds, are not well predicted by simulation, hence, they are
 2873 estimated using data-driven methods; in the case of non-prompt leptons, a fake rate
 2874 method is used.

2875 Drell-Yan contribution is reduced in the $3l$ channel due to the requirement of an
 2876 additional lepton, the jet counting and the Z-veto. Three boson contributions are
 2877 reduced by rejecting events with extra leptons and by requiring forward jets or b -jets

2878 The $2lss$ channel receives contribution from the associated production of two W
 2879 bosons of equal charge and two light jets $W^\pm W^\pm qq$ and from same-sign W boson
 2880 pairs can also be produced in double parton scattering (DPS) processes, where each
 2881 of the colliding protons gives two partons, resulting in two hard interactions.

2882 Backgrounds from $t\bar{t}W$ and $t\bar{t}Z$ processes are estimated using simulated events,
 2883 corrected for data/MC differences and inefficiencies (trigger and lepton selection) in

2884 the same way as signal events. Their production cross sections are calculated at
 2885 NLO order of QCD and EWK, considering theoretical uncertainties from unknown
 2886 higher orders of 12% for $t\bar{t}W$ and 10% for $t\bar{t}Z$. Additional uncertainties arise from the
 2887 knowledge of PDFs and \hat{S} of about 4% each for $t\bar{t}W$ and $t\bar{t}Z$.

2888 6.9.1 Diboson backgrounds

2889 Background contamination from diboson processes is strongly suppressed by imposing
 2890 the Z-veto, vetoing additional leptons and requiring b -jets in the event. The diboson
 2891 contribution is also estimated from simulated events; however, the overall normal-
 2892 ization of this process is obtained from a dedicated control region. The motivation
 2893 behind that strategy is that even though the measured inclusive cross section for
 2894 diboson processes (WZ,ZZ) is in good agreement with the NLO calculations [149],
 2895 that agreement is perturbed when leptonic Z decays and hadronic jets in the final
 2896 state are required; those requirements are precisely the ones that make the diboson
 2897 production a background for the tHq signal. Thus, by using a dedicated control
 2898 region dominated by WZ production⁵, the overall normalization is constrained.

2899 The control region is defined by the presence of at least three leptons, of which
 2900 one opposite-sign pair must be compatible with a Z boson decay, i.e., invert the Z-
 2901 veto which makes the control region orthogonal to signal region; the b-jet tagging
 2902 requirements is also inverted with respect to the signal region, i.e., require two not
 2903 b -jets. A scale factor is extracted from the predicted distribution of WZ events in the
 2904 control region, and the observed data, while keeping other processes fixed; this factor
 2905 is used to scale the diboson prediction in the signal selection region. More details
 2906 about the procedure used can be found in Reference [149] from where the scale factor
 2907 is taken.

⁵ ZZ background is strongly reduced by the cut on MET.

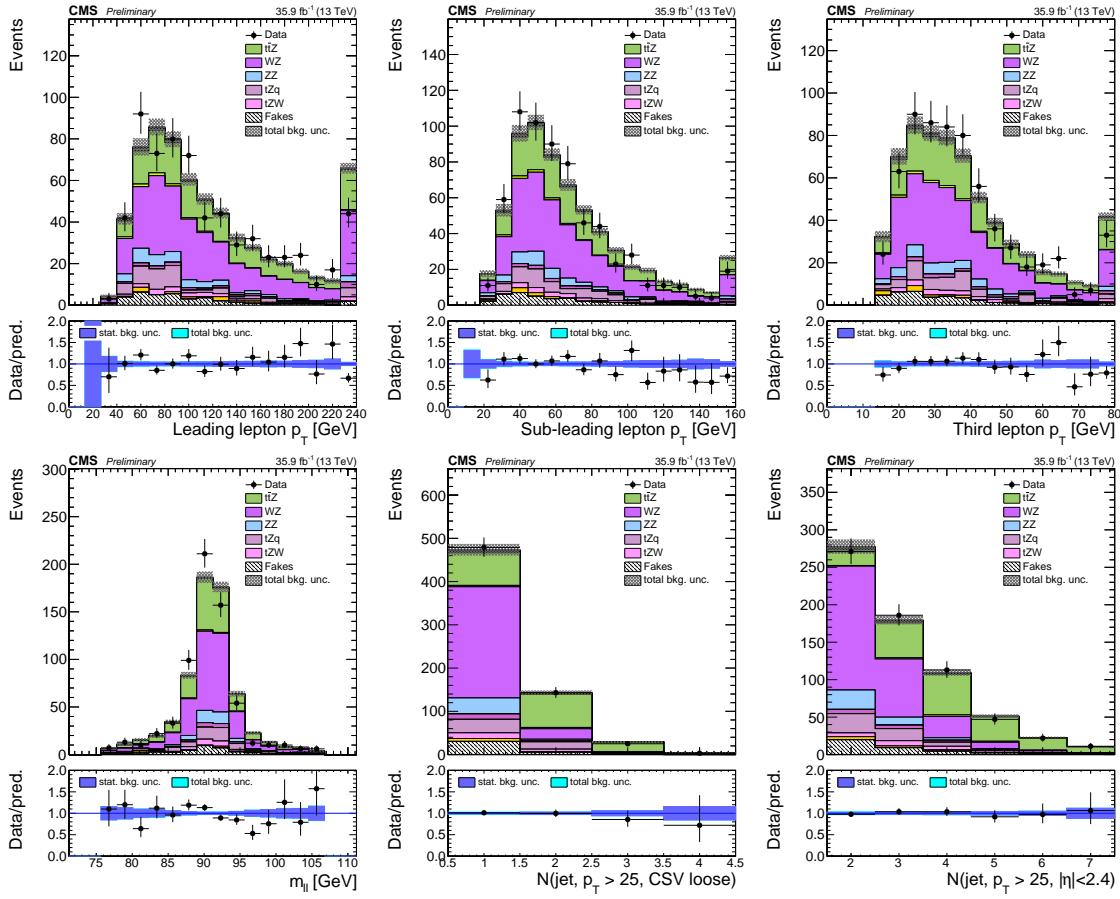


Figure 6.20: Kinematic distributions in the diboson control region.

2908 In order to test the usability of the diboson background scale factor in this analysis,
 2909 a Z-enriched control region⁶ was defined by inverting the Z-veto and requiring exactly
 2910 three tight leptons with $p_T > 25/15/15$ GeV, one or more jets passing the CSVv2 loose
 2911 working point and less than four central jets. Figure 6.20 shows the distribution of
 2912 three variables in the diboson control region; the good agreement between MC and
 2913 data motivates the adoption of the diboson background scale factor.

2914 . Their production cross sections are calculated at NLO of QCD and EWK, with
 2915 theoretical uncertainties from unknown higher orders of 12% for $t\bar{t}W$ and 10% for $t\bar{t}Z$.
 2916 Further uncertainties arise from the knowledge of PDFs and α_s of about 4% each for

⁶ This control region is different to the one used to find the scale factor.

2917 $t\bar{t}W$ and $t\bar{t}Z$.

2918 In the $2lss$ channel case, additional background arises when the charge of a lepton
 2919 in events with an originally opposite-sign pair is misidentified; usually this happens
 2920 due to strongly asymmetric conversions of hard bremsstrahlung photons emitted from
 2921 the initial lepton, therefore, it is more likely to happen for electrons than for muons.

2922 Additional identification criteria are applied for electrons with p_T greater than 30
 2923 GeV to mimic the identification applied at trigger level in order to ensure consistency
 2924 between the measurement region and application region of the fake-rate.

2925 where the contribution to the final selection is estimated by extrapolating from a
 2926 sideband (or *application region*) with a looser lepton definition (the fakeable object
 2927 definitions in Tabs. 6.4 and 6.5) to the signal selection. The tight-to-loose ratios (or
 2928 *fake rates*) are measured in several background dominated data events with dedicated
 2929 triggers, subtracting the residual prompt lepton contribution using MC. Non-prompt
 2930 leptons in our signal regions are predominantly produced in $t\bar{t}$ events, with a much
 2931 smaller contribution, from Drell–Yan production. The systematic uncertainty on the
 2932 normalization of the non-prompt background estimation is on the order of 50%, and
 2933 thereby one of the dominant limitations on the performance of multilepton analyses in
 2934 general and this analysis in particular. It consists of several individual sources, such
 2935 as the result of closure tests of the method using simulated events, limited statistics
 2936 in the data control regions due to necessary prescaling of lepton triggers, and the
 2937 uncertainty in the subtraction of residual prompt leptons from the control region.

2938 The fake background where the leptons pass the looser selection are weighted
 2939 according to how many of them fail the tight criteria. Events with a single failing
 2940 lepton are weighted with the factor $f/(1-f)$ for the estimate to the tight selection
 2941 region, where f is the fake rate. Events with two failing leptons are given the negative
 2942 weight $-f_i f_j / (1-f_i)(1-f_j)$, and for three leptons the weight is positive and equal

2943 to the product of $f/(1-f)$ factor evaluated for each failing lepton.

2944 Figures 6.21 show the distributions of some relevant kinematic variables, normal-
2945 ized to the cross section of the respective processes and to the integrated luminosity.

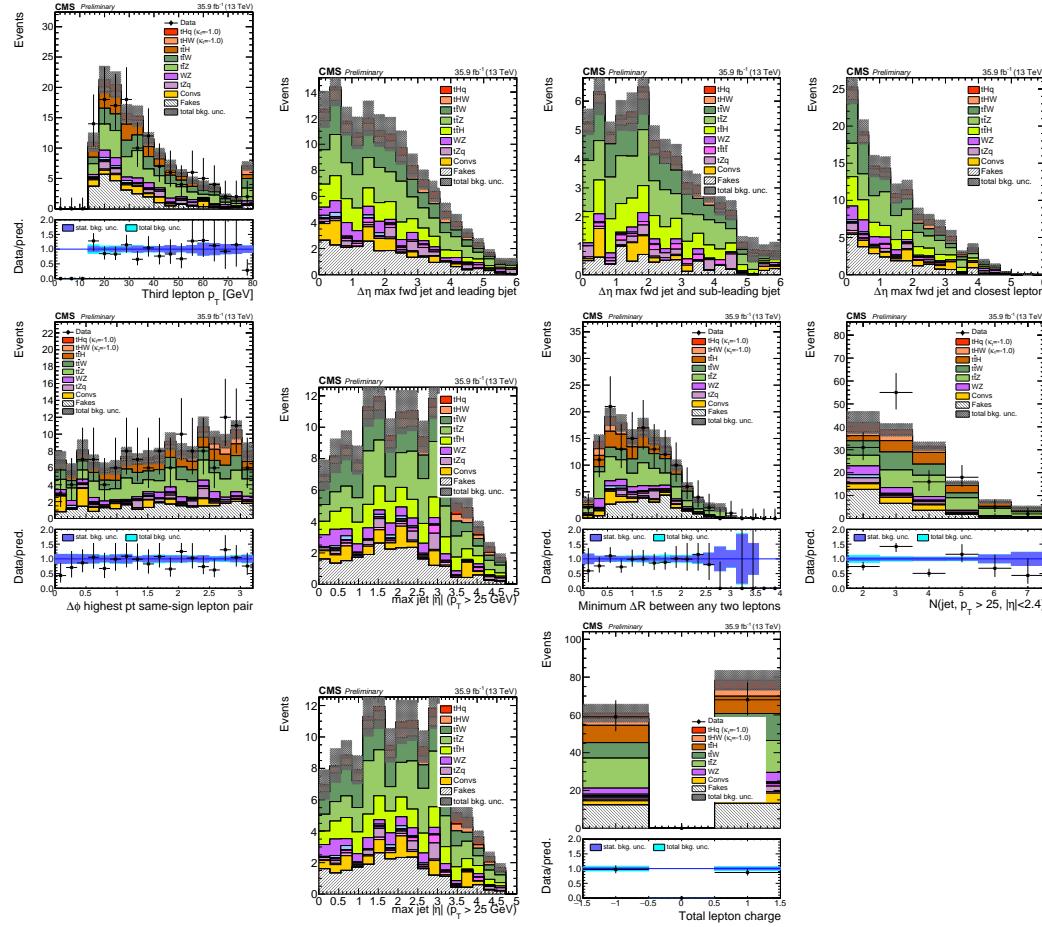


Figure 6.21: Distributions of input variables to the BDT for signal discrimination, three lepton channel, normalized to their cross section and to 35.9 fb^{-1} .

2946

2947 The modeling of reducible and irreducible backgrounds in this analysis uses the
2948 exact methods, analysis code, and ROOT trees used for the $t\bar{t}H$ multilepton analysis.
2949 We give a brief description of the methods and refer to the documentation of that
2950 analysis in Refs. [145, 149] for any details.

2951 Once the events are selected in agreement with the signal characteristics, the

2952 extraction of the signal contribution is performed using multivariate discriminators
 2953 against the main backgrounds of $t\bar{t}W/t\bar{t}Z$ and non-prompt leptons from $t\bar{t}$. The shape
 2954 of the discriminator variables is then fit to the observed data distribution to estimate
 2955 the signal and background yields, simultaneously for all channels.

2956 A significant fraction of selected data events (about 50% in the dilepton channels,
 2957 and about 80% in the trilepton channel) also passes the selection used in the dedicated
 2958 search for ttH in multilepton channels [17]. The expected and observed event yields
 2959 of this selection are shown in Tab. 2. For the tH and $t\bar{t}H$ processes, the largest
 2960 contribution comes from Higgs decays to WW (about 75%), followed by $t\bar{t}H$ (about
 2961 20%) and ZZ (about 5%). Other Higgs production modes contribute negligible event
 2962 yields (< 5% of the tH +ttH yield).

2963 .

2964 .

2965 .

2966 .

2967 .

2968 .

2969 .

2970 .

2971 .

2972 . . .

2973 . Multivariate techniques are used to discriminate the signal from the dominant
 2974 backgrounds. The analysis yields a 95% confidence level (C.L.) upper limit on the
 2975 combined tH + ttH production cross section times branching ratio of 0.64 pb, with
 2976 an expected limit of 0.32 pb, for a scenario with $kt = \sqrt{1.0}$ and $kV = 1.0$. Values
 2977 of kt outside the range of $\sqrt{1.25}$ to $\sqrt{1.60}$ are excluded at 95% C.L., assuming kV

2978 = 1.0.

2979 Dont forget to mention previous constrains to ct check Reference ?? and Refer-
2980 ences <https://link.springer.com/content/pdf/10.1007%2FJHEP01%282013%29088.pdf>
2981 (paragraph after eq 2)

²⁹⁸² **Appendix A**

²⁹⁸³ **Datasets and triggers**

Dataset name
/JetHT/Run2016X-23Sep2016-vY/MINIAOD
/HTMHT/Run2016X-23Sep2016-vY/MINIAOD
/MET/Run2016X-23Sep2016-vY/MINIAOD
/SingleElectron/Run2016X-23Sep2016-vY/MINIAOD
/SingleMuon/Run2016X-23Sep2016-vY/MINIAOD
/SinglePhoton/Run2016X-23Sep2016-vY/MINIAOD
/DoubleEG/Run2016X-23Sep2016-vY/MINIAOD
/MuonEG/Run2016X-23Sep2016-vY/MINIAOD
/DoubleMuon/Run2016X-23Sep2016-vY/MINIAOD
/Tau/Run2016B-23Sep2016-v3/MINIAOD
/JetHT/Run2016H-PromptReco-v3/MINIAOD
/HTMHT/Run2016H-PromptReco-v3/MINIAOD
/MET/Run2016H-PromptReco-v3/MINIAOD
/SingleElectron/Run2016H-PromptReco-v3/MINIAOD
/SingleMuon/Run2016H-PromptReco-v3/MINIAOD
/SinglePhoton/Run2016H-PromptReco-v3/MINIAOD
/DoubleEG/Run2016H-PromptReco-v3/MINIAOD
/MuonEG/Run2016H-PromptReco-v3/MINIAOD
/DoubleMuon/Run2016H-PromptReco-v3/MINIAOD

Table A.1: Full 2016 dataset used in the analysis. In the first section of the table are listed the 23Sep2016 samples; in the Run2016X-23Sep-vY label, X:B-G tag the run while Y:1,3 tag the version of the data sample. Second section list the PromptReco version of the dataset.

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

Table A.2: Table of high-level triggers considered in the analysis.

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

Table A.3: 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) [150].

Sample	σ [pb]	*
TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043	*
TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529	*
/store/cmst3/group/susy/gpetrucc/13TeV/u/TTLL_m1to10_L0_NoMS_for76X/	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_tW1l_5f_L0_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	
ttWJets_13TeV_madgraphMLM	0.6105	
ttZJets_13TeV_madgraphMLM	0.5297/0.692	

Table A.4: List of background samples used in this analysis (CMSSW 80X). The first section of the table lists the samples used in simulation to extract the final yields and shapes; the second section lists the samples of the processes for which the yields are estimated from data. The MC simulation is used to design the data driven methods and in the derivation of the associated systematic uncertainties. The third section list the leading order $t\bar{t}W$ and $t\bar{t}Z$ samples, which in addition to the ones marked with a *, where used in the BDT training.

²⁹⁸⁴ **Appendix B**

²⁹⁸⁵ **BDTG aditional plots**

²⁹⁸⁶ **B.1 BDTG input variables distributions for $2lss$**

²⁹⁸⁷ **channel**

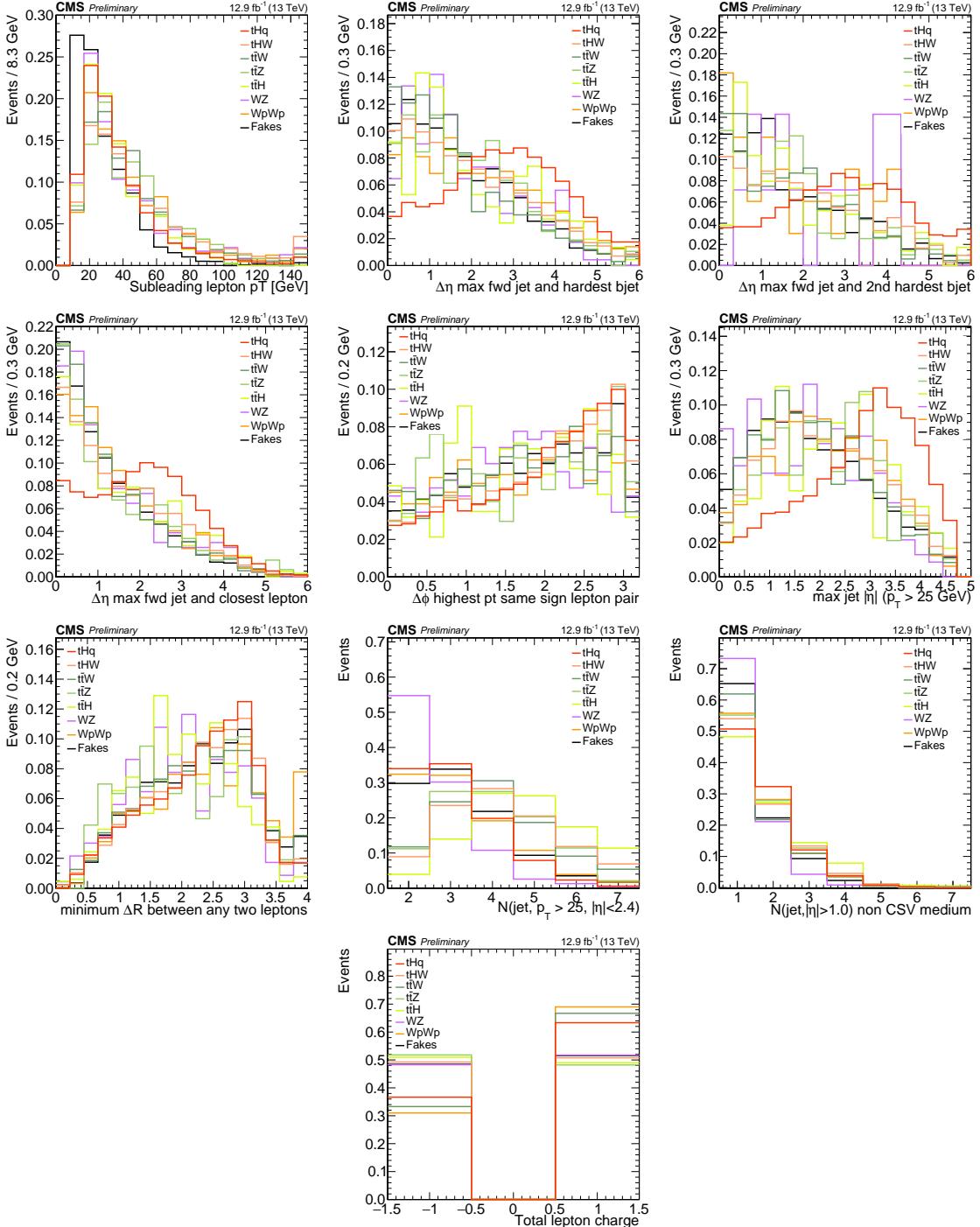


Figure B.1: Distributions of input variables to the BDT for signal discrimination, two lepton same sign channel.

2988 **B.2 Input variables distributions from BDTG**
 2989 classifiers

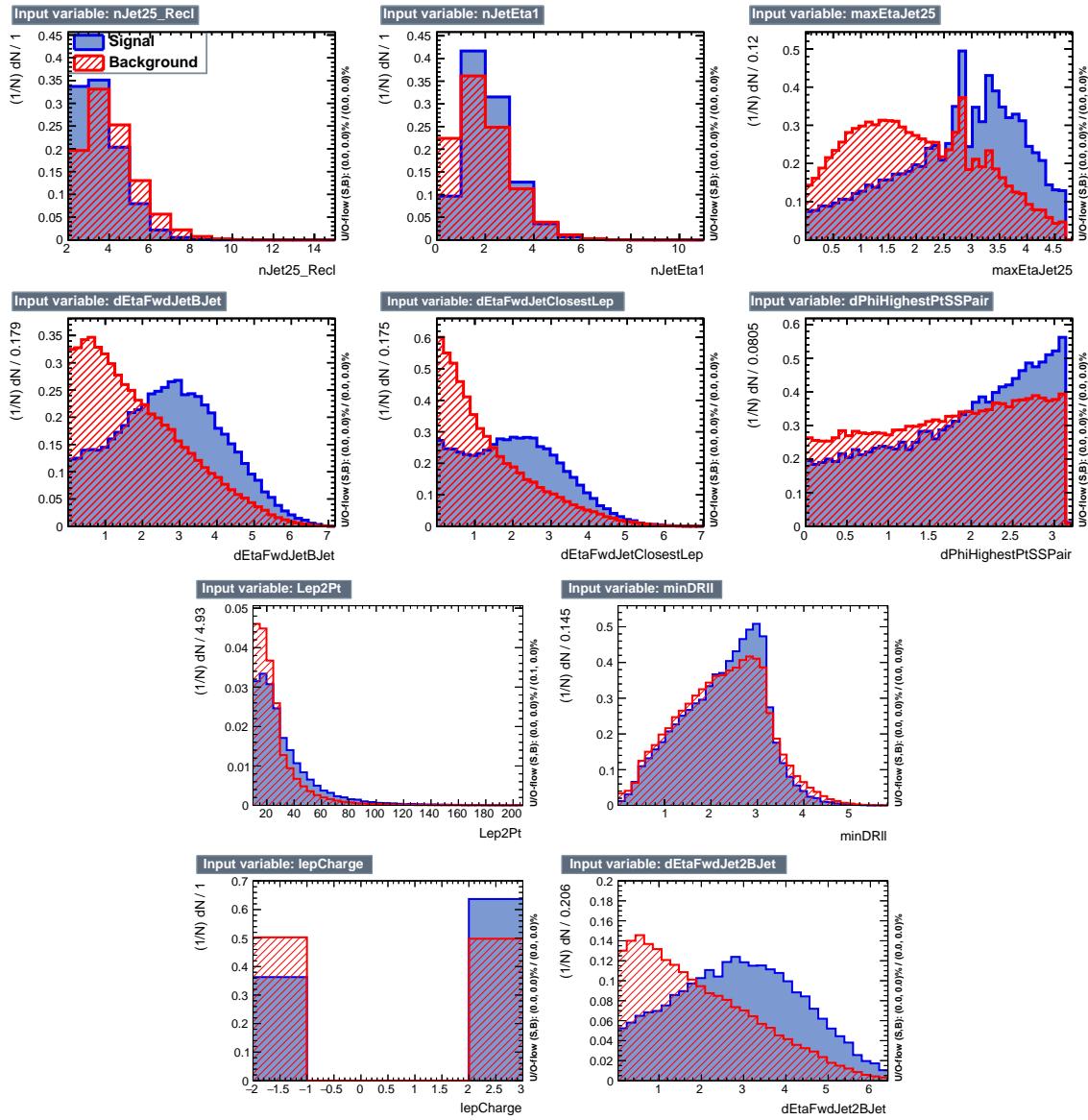


Figure B.2: BDT input variables as seen by BDTG classifier for the $2lss$ channel, tHq signal (blue) discriminated against $t\bar{t}$ background (red).

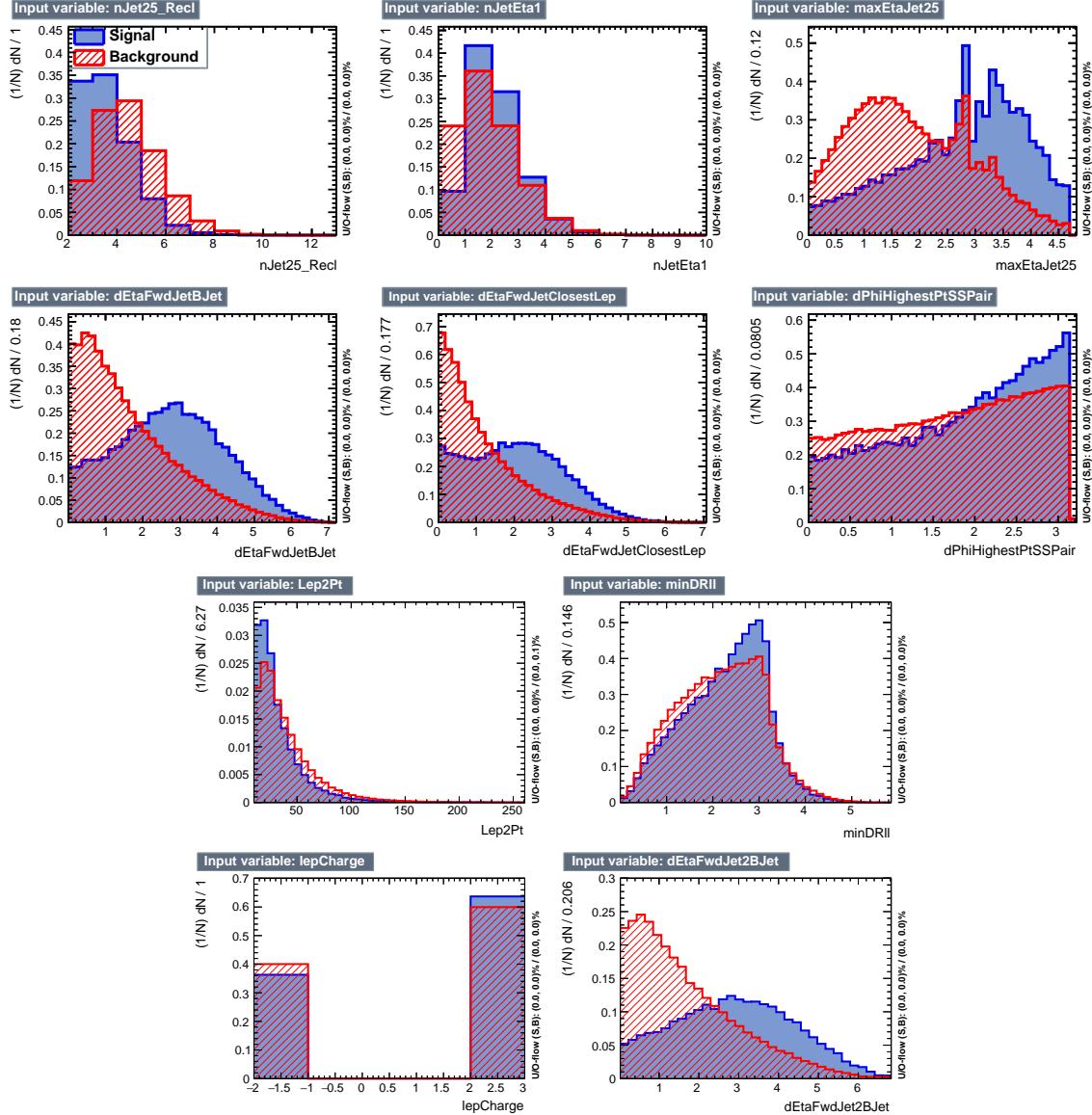


Figure B.3: BDT input variables as seen by BDTG classifier for the $2lss$ channel, tHq signal(blue) discriminated against $t\bar{t}V$ background (red).

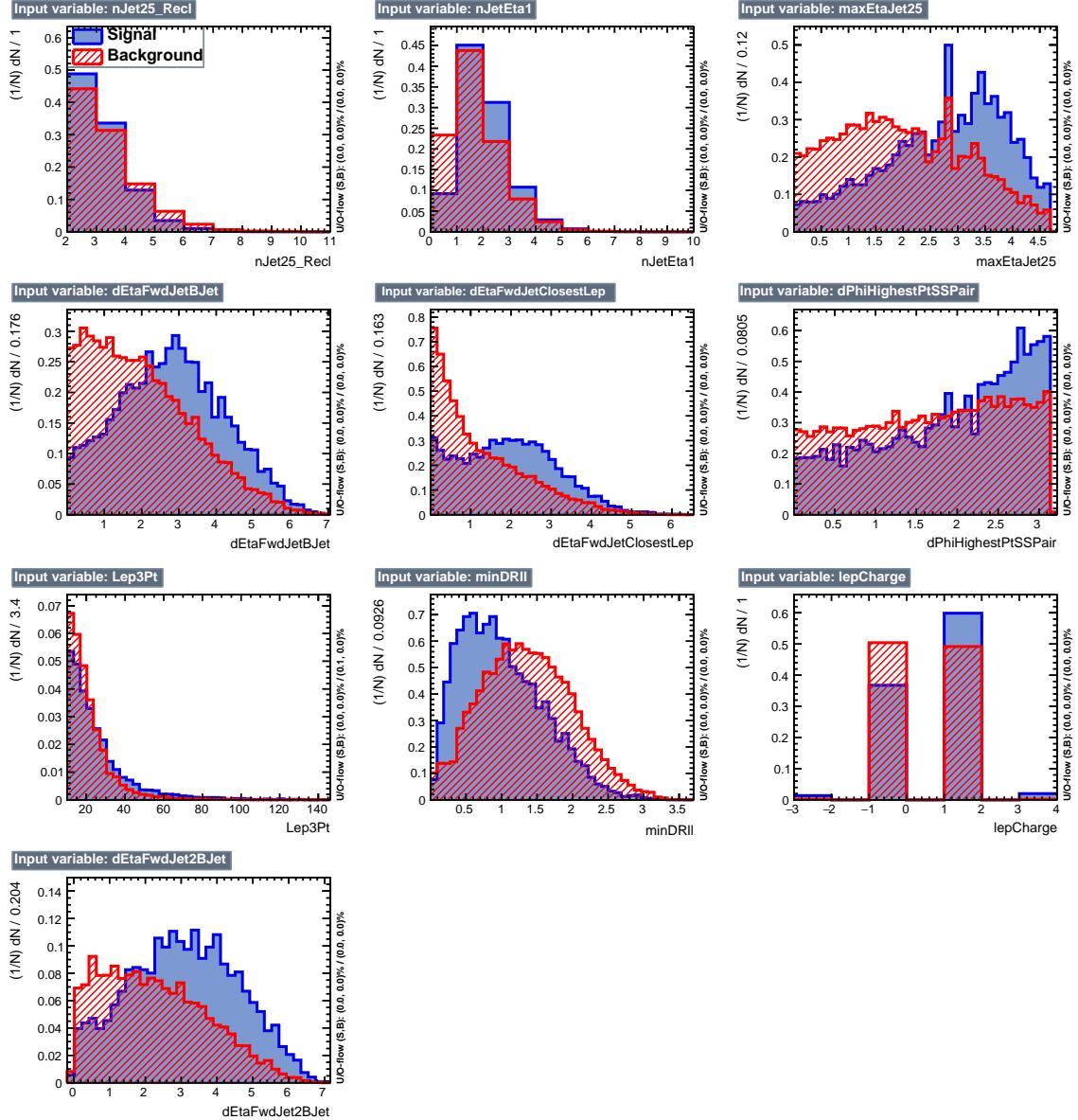


Figure B.4: BDT input variables as seen by BDTG classifier for the $3l$ channel, tHq signal (blue) discriminated against $t\bar{t}$ background (red).

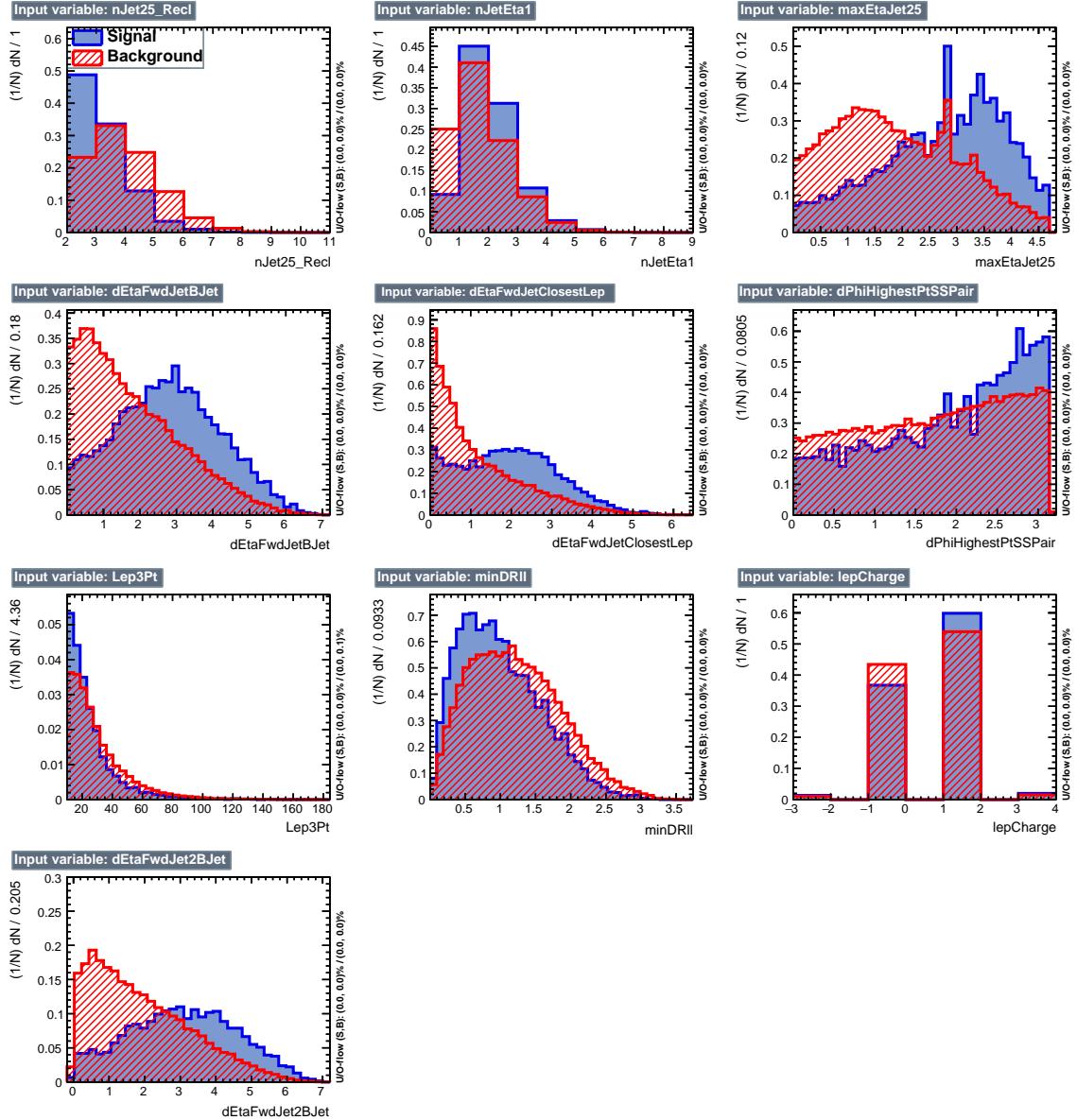


Figure B.5: BDT input variables as seen by BDTG classifier for the $3l$ channel, tHq signal (blue) discriminated against $t\bar{t}V$ background (red).

2990 **Appendix C**

2991 **Other binning strategies**

2992 Two additional strategies of clustering regions in the 2D plane of $BDTG_{tt}$ vs $BDTG_{ttV}$
 2993 into bins were attempted, following studies done and documented in great detail in
 2994 Reference [149]. A brief description is provided in the following.

2995 **Clustering by S/B ratio** In this method, the 2D plane is clustered into a given
 2996 number of bins corresponding to regions where S/B is within a certain range. The
 2997 bin borders are determined such that the number of background events in each bin is
 2998 approximately equal. The resulting regions for $2lss$ and $3l$ events are shown in Figure
 2999 C.1, while the expected distribution of signal and dominant backgrounds are shown
 3000 in Figure C.2.

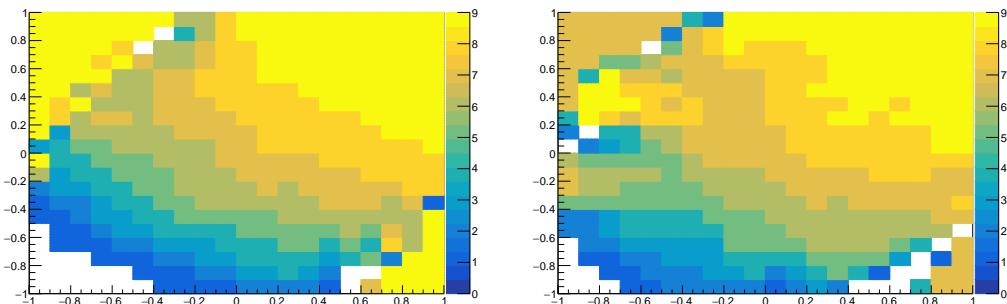


Figure C.1: Binning by S/B regions for $2lss$ (left) and $3l$ (right).

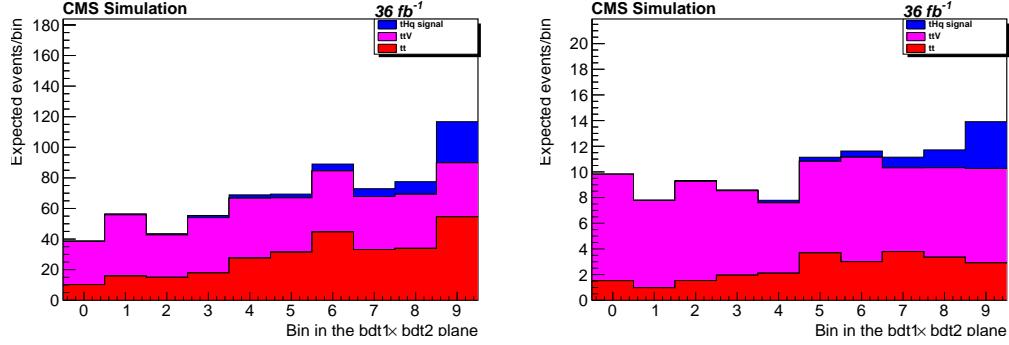


Figure C.2: Final bins (corresponding to S/B regions in the 2D plane) for $2lss$ and $3l$ (right).

Using this method, the resulting limits (for the $\kappa_t = -1, \kappa_V = 1$ scenario) are about 20% worse than with the binning in Section 6.7.6: $\mu^\pm\mu^\pm$ changed from 1.82 to 2.15, $3l$ changed from 1.52 to 1.75.

***k*-Means geometric clustering** This method employs a recursive application of the *k*-means algorithm (see Appendix D in Reference [149]) to separate the 2D plane into geometric regions. The resulting clustering (using the $t\bar{t}H$ multilepton code on tHq signal and $t\bar{t}$ and $t\bar{t}V$ background events) are shown in Figure C.3. The expected distribution of events for the signal and dominant backgrounds in these bins is shown in Fig. C.4.

Similarly to the S/B ratio binning, the limits using the *k*-means clustering are significantly worse than those of the bins described before. In the $\mu^\pm\mu^\pm$ channel, the limit deteriorates from 1.82 to 2.05, whereas in $3l$ it changes from 1.58 to 1.78.

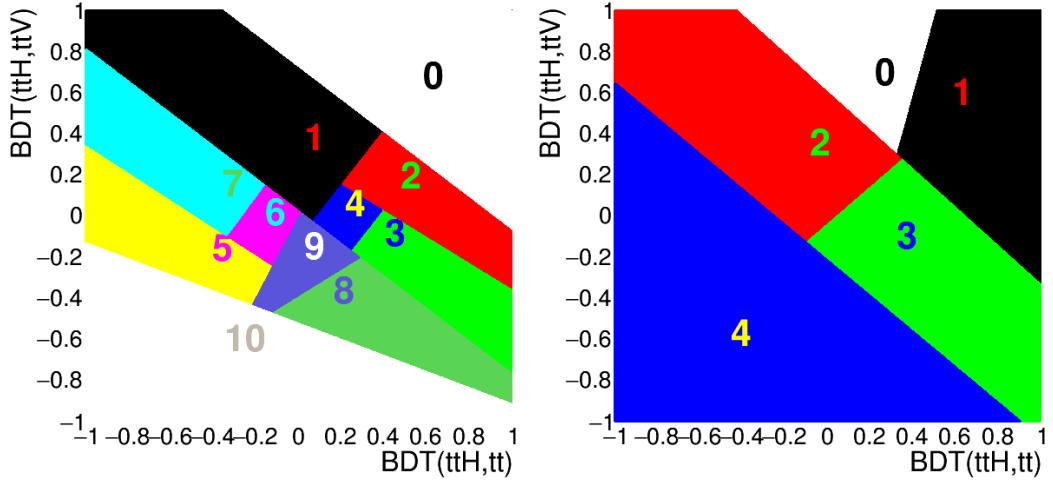


Figure C.3: Binning into geometric regions using a k -means algorithm for $2lss$ (left) and $3l$ (right).

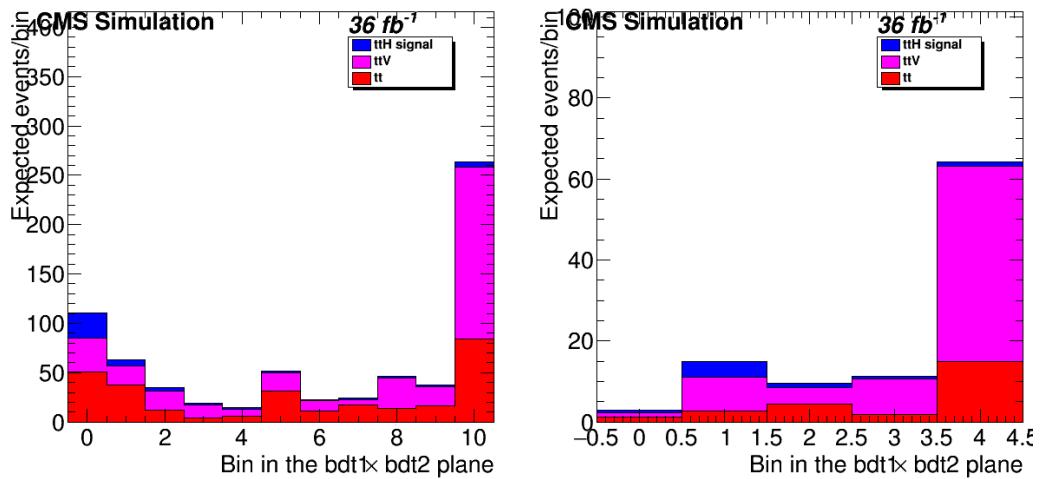


Figure C.4: Final bins using a k -means algorithm for $2lss$ (left) and $3l$ (right). Note that the bin numbering here is such that signal-like bins are lower.

³⁰¹³ **Appendix D**

³⁰¹⁴ **BDTG output variation with κ_V/κ_t**

³⁰¹⁵ The BDTG classifier output was described in Section in the $\kappa_t = -1, \kappa_V = 1$ escenario;
³⁰¹⁶ the change of BDTG classifiers output shape when varying the κ_V/κ_t coupling sce-
³⁰¹⁷ nario is shown in Figure D.1 in the $3l$ channel for five different values of κ_t , with κ_V fixed at 1.0.

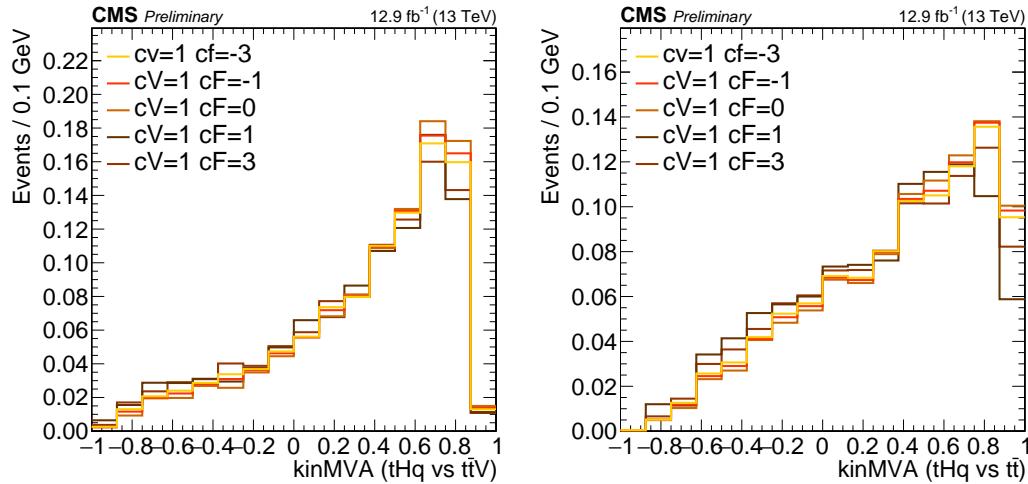


Figure D.1: Change of the BDTG classifiers output when varying κ_t coupling (κ_V is fixed at 1.0). Training vs. $tt\bar{V}$ (right) and vs. $tt\bar{t}$ (left).

3019 **References**

- 3020 [1] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". Phys-
 3021 ical Review. 74 (10): 1439-61, (1948). <https://doi.org/10.1103/PhysRev.74.1439>
- 3023 [2] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". Physical
 3024 Review. 76 (6): 769-89, (1949). <https://doi.org/10.1103/PhysRev.76.769>
- 3025 [3] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum
 3026 Theory of Wave Fields". Progress of Theoretical Physics. 1 (2): 27-42, (1946).
 3027 <https://doi.org/10.1143/PTP.1.27>
- 3028 [4] D.J. Griffiths, "Introduction to electrodynamics". 4th ed. Pearson, (2013).
- 3029 [5] F. Mandl, G. Shaw. "Quantum field theory." Chichester, Wiley (2009).
- 3030 [6] F. Halzen, and A.D. Martin, "Quarks and leptons: An introductory course in
 3031 modern particle physics". New York: Wiley, (1984) .
- 3032 [7] File: Standard_Model_of_Elementary_Particle_dark.svg. (2017, June 12)
 3033 Wikimedia Commons, the free media repository. Retrieved November 27, 2017
 3034 from <https://www.collegiate-advanced-electricity.com/single-post/2017/04/10/The-Standard-Model-of-Particle-Physics>.

- 3036 [8] E. Noether, "Invariante Variationsprobleme", Nachrichten von der Gesellschaft
3037 der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, vol. 1918,
3038 pp. 235-257, (1918).
- 3039 [9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)
3040 and 2017 update.
- 3041 [10] M. Goldhaber, L. Grodzins, A.W. Sunyar "Helicity of Neutrinos", Phys. Rev.
3042 109, 1015 (1958).
- 3043 [11] Palanque-Delabrouille N et al. "Neutrino masses and cosmology with Lyman-
3044 alpha forest power spectrum", JCAP 11 011 (2015).
- 3045 [12] M. Gell-Mann. "A Schematic Model of Baryons and Mesons". Physics Letters.
3046 8 (3): 214-215 (1964).
- 3047 [13] G. Zweig. "An SU(3) Model for Strong Interaction Symmetry and its Breaking"
3048 (PDF). CERN Report No.8182/TH.401 (1964).
- 3049 [14] G. Zweig. "An SU(3) Model for Strong Interaction Symmetry and its Breaking:
3050 II" (PDF). CERN Report No.8419/TH.412(1964).
- 3051 [15] M. Gell-Mann. "The Interpretation of the New Particles as Displaced Charged
3052 Multiplets". Il Nuovo Cimento 4: 848. (1956).
- 3053 [16] T. Nakano, K. Nishijima. "Charge Independence for V-particles". Progress of
3054 Theoretical Physics 10 (5): 581-582. (1953).
- 3055 [17] N. Cabibbo, "Unitary symmetry and leptonic decays" Physical Review Letters,
3056 vol. 10, no. 12, p. 531, (1963).

- 3057 [18] M.Kobayashi, T.Maskawa, “CP-violation in the renormalizable theory of weak
3058 interaction,” Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).
- 3059 [19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Com-
3060 mons, the free media repository. Retrieved November 27, 2017
3061 from [https://commons.wikimedia.org/w/index.php?title=File:
3062 Weak_Decay_\(flipped\)\.svg&oldid=247498592](https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)\.svg&oldid=247498592).
- 3063 [20] Georgia Tech University. Coupling Constants for the Fundamental Forces(2005).
3064 Retrieved January 10, 2018, from [http://hyperphysics.phy-astr.gsu.edu/
3065 hbase/Forces/couple.html#c2](http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/couple.html#c2)
- 3066 [21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved Jan-
3067 uary 10, 2018, from [https://profmattstrassler.com/articles-and-posts/
3068 particle-physics-basics/the-known-forces-of-nature/
3069 the-strength-of-the-known-forces/](https://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/)
- 3070 [22] S.L. Glashow. “Partial symmetries of weak interactions”, Nucl. Phys. 22 579-
3071 588, (1961).
- 3072 [23] A. Salam, J.C. Ward. “Electromagnetic and weak interactions”, Physics Letters
3073 13 168-171, (1964).
- 3074 [24] S. Weinberg, “A model of leptons”, Physical Review Letters, vol. 19, no. 21, p.
3075 1264, (1967).
- 3076 [25] M. Peskin, D. Schroeder, “An introduction to quantum field theory”. Perseus
3077 Books Publishing L.L.C., (1995).
- 3078 [26] A. Pich. “The Standard Model of Electroweak Interactions” <https://arxiv.org/abs/1201.0537>
- 3079

- 3080 [27] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 122, 103 (1983).
- 3081 [28] M. Banner et al. (UA2 Collaboration), Phys. Lett. B 122, 476 (1983).
- 3082 [29] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 126, 398 (1983).
- 3083 [30] P. Bagnaia et al. (UA2 Collaboration), Phys. Lett. B 129, 130 (1983).
- 3084 [31] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved
3085 from: <https://www.quantum-bits.org/?p=233>
- 3086 [32] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255
- 3087 [33] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge
3088 Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964)
3089 doi:10.1103/PhysRevLett.13.321
- 3090 [34] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Re-
3091 view Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.
- 3092 [35] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws
3093 and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964).
3094 doi:10.1103/PhysRevLett.13.585.
- 3095 [36] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with
3096 the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012).
3097 arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021
- 3098 [37] ATLAS collaboration. “Observation of a New Particle in the Search for the Stan-
3099 dard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters
3100 B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

- 3101 [38] ATLAS collaboration; CMS collaboration (26 March 2015). “Combined Mea-
 3102 surement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with
 3103 the ATLAS and CMS Experiments”. Physical Review Letters. 114 (19): 191803.
 3104 arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.
- 3105 [39] LHC InternationalMasterclasses“When protons collide”. Retrieved from http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm
- 3107 [40] CMS Collaboration, “SM Higgs Branching Ratios and Total Decay Widths (up-
 3108 date in CERN Report4 2016)”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR> , last accessed on 17.12.2017.
- 3110 [41] R.Grant V. “Determination of Higgs branching ratios in $H \rightarrow W^+W^- \rightarrow l\nu jj$
 3111 and $H \rightarrow ZZ \rightarrow l^+l^-jj$ channels”. Physics Department, University of Ten-
 3112 nessee (Dated: October 31, 2012). Retrieved from <http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf>
- 3114 [42] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al.
 3115 “Standard model Higgs-boson branching ratios with uncertainties”. Eur. Phys.
 3116 J. C (2011) 71: 1753. <https://doi.org/10.1140/epjc/s10052-011-1753-8>
- 3117 [43] D. de Florian et al., LHC Higgs Cross Section Working Group,
 3118 CERNâš2017âš002-M, arXiv:1610.07922[hep-ph] (2016).
- 3119 [44] ATLAS and CMS Collaborations, “Measurements of the Higgs boson produc-
 3120 tion and decay rates and constraints on its couplings from a combined ATLAS
 3121 and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,” (2016).
 3122 CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.

- 3123 [45] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Perez-Victoria,
 3124 “Handbook of vector-like quarks: Mixing and single production”, Phys. Rev. D
 3125 88 (2013) 094010, doi:10.1103/PhysRevD.88.094010, arXiv:1306.0572.
- 3126 [46] A. Greljo, J. F. Kamenik, and J. Kopp, “Disentangling flavor vio-
 3127 lation in the top-Higgs sector at the LHC”, JHEP 07 (2014) 046,
 3128 doi:10.1007/JHEP07(2014)046, arXiv:1404.1278.
- 3129 [47] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in
 3130 association with a single top quark at the LHC,” European Physical Journal C,
 3131 vol. 75, p. 267, (2015). doi:10.1140/epjc/s10052-015-3475-9, arXiv:1504.00611.
- 3132 [48] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated
 3133 production at the LHC”, European Physical Journal C, vol. 77, p. 34, (2017).
 3134 arXiv:1607.05862
- 3135 [49] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production
 3136 of Higgs and single top at hadron colliders”, Phys.Rev. D64 (2001) 094023,
 3137 [hep-ph/0106293].
- 3138 [50] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the
 3139 top-Higgs coupling from the 8 TeV LHC data,” Journal of High Energy Physics,
 3140 vol. 07, p. 073, (2013).
- 3141 [51] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting de-
 3142 generacies in Higgs couplings using single top production in association with a
 3143 Higgs boson,” Journal of High Energy Physics, vol. 05, p. 022, (2013).
- 3144 [52] T.M. Tait and C.-P. Yuan, “Single top quark production as a window to physics
 3145 beyond the standard model”, Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].

- 3146 [53] CMS Collaboration, “Modelling of the single top-quark production in associa-
3147 tion with the Higgs boson at 13 TeV.” [https://twiki.cern.ch/twiki/bin/](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV)
3148 [viewauth/CMS/SingleTopHiggsGeneration13TeV](#), last accessed on 16.01.2018.
- 3149 [54] CMS Collaboration, “SM Higgs production cross sections at $\sqrt{s} =$
3150 13 TeV.” [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)
3151 [CERNYellowReportPageAt13TeV](#), last accessed on 16.01.2018.
- 3152 [55] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.
- 3153 [56] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [[arXiv:1211.0499](https://arxiv.org/abs/1211.0499)
3154 [hep-ph]].
- 3155 [57] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross
3156 Sections: 4.Deciphering the Nature of the Higgs Sector”, [arXiv:1610.07922](https://arxiv.org/abs/1610.07922).
- 3157 [58] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top
3158 Couplings in Associated Production”, JHEP 1404 (2014) 004, [[arXiv:1312.5736](https://arxiv.org/abs/1312.5736)].
- 3159 [59] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass
3160 of the Higgs boson and tests of compatibility of its couplings with the standard
3161 model predictions using proton collisions at 7 and 8 TeV,” [arXiv:1412.8662](https://arxiv.org/abs/1412.8662).
- 3162 [60] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the
3163 Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton
3164 collision data”, ATLAS-CONF-2014-009.
- 3165 [61] File:Cern-accelerator-complex.svg. Wikimedia Commons, the free media repos-
3166 itory. Retrieved January, 2018 from <https://commons.wikimedia.org/wiki/>
3167 [File:Cern-accelerator-complex.svg](#)

- 3168 [62] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”,
3169 (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Re-
3170 trieved from <https://cds.cern.ch/record/841542>
- 3171 [63] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-
3172 2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762
- 3173 [64] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).
- 3174 [65] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March
3175 1983, CERN-AC-8303511. Retrieved from <https://cds.cern.ch/record/615852>.
- 3177 [66] CERN Photographic Service “Animation of CERN’s accelerator network”, 14
3178 October 2013. DOI: 10.17181/cds.1610170 Retrieved from <https://videos.cern.ch/record/1610170>
- 3180 [67] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17,
3181 2013. Retrieved from <https://www.britannica.com/technology/particle-accelerator>.
- 3183 [68] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-
3184 AC-0007016. Retrieved from <https://cds.cern.ch/record/41567>.
- 3185 [69] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”.
3186 March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-
3187 PHO-1998-325. Retrieved from <https://cds.cern.ch/record/841511>.
- 3188 [70] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025
3189 retrieved from <https://cds.cern.ch/record/40524>.

- 3190 [71] CMS Collaboration “Public CMS Luminosity Information”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016__proton_proton_13_TeV_collis, last accessed 24.01.2018
- 3193 [72] J.L Caron. “LHC Layout” AC Collection. Legacy of AC. Pictures from 1992
3194 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from <https://cds.cern.ch/record/841573>.
- 3196 [73] J.A. Coarasa. “The CMS Online Cluster:Setup, Operation and Maintenance
3197 of an Evolving Cluster”. ISGC 2012, 26 February - 2 March 2012, Academia
3198 Sinica, Taipei, Taiwan.
- 3199 [74] CMS Collaboration. “The CMS experiment at the CERN LHC” JINST 3 S08004
3200 (2008).
- 3201 [75] CMS Collaboration. “CMS detector drawings 2012” CMS-PHO-GEN-2012-002.
3202 Retrieved from <http://cds.cern.ch/record/1433717>.
- 3203 [76] Davis, Siona Ruth. “Interactive Slice of the CMS detector”, Aug. 2016,
3204 CMS-OUTREACH-2016-027, retrieved from <https://cds.cern.ch/record/2205172>
- 3206 [77] R. Breedon. “View through the CMS detector during the cooldown of the
3207 solenoid on February 2006. CMS Collection”, February 2006, CMS-PHO-
3208 OREACH-2005-004, Retrieved from <https://cds.cern.ch/record/930094>.
- 3209 [78] Halyo, V. and LeGresley, P. and Lujan, P. “Massively Parallel Computing and
3210 the Search for Jets and Black Holes at the LHC”, Nucl.Instrum.Meth. A744
3211 (2014) 54-60, DOI: 10.1016/j.nima.2014.01.038”

- 3212 [79] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector
3213 Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.
- 3214 [80] CMS Collaboration. “Description and performance of track and primary-vertex
3215 reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no.
3216 10, p. P10009,(2014).
- 3217 [81] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner Bar-
3218 rel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from <https://cds.cern.ch/record/1431467>.
- 3220 [82] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm
3221 and beauty hadrons Chicago, June 28-July 3 2004.
- 3222 [83] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at
3223 the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013.
3224 arXiv:1307.7135
- 3225 [84] L. Veillet. “End assembly of HB with EB rails and rotation inside SX ”,Jan-
3226 uary 2002. CMS-PHO-HCAL-2002-002. Retrieved from <https://cds.cern.ch/record/42594>.
- 3228 [85] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the UX5 cav-
3229 ern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from <https://cds.cern.ch/record/1019185>
- 3231 [86] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web site:
3232 Taking a closer look at LHC. Retrieved from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger

- 3234 [87] WLCG Project Office, “Documents & Reference - Tiers - Structure,”
 3235 (2014). <http://wlcg.web.cern.ch/documents-reference> , last accessed on
 3236 30.01.2018.
- 3237 [88] CMS Collaboration. “CMSSW Application Framework”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>,
 3238 last accesses 06.02.2018
- 3240 [89] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event gen-
 3241 erators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011
- 3242 [90] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics
 3243 of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-
 3244 3-540-71060-8 (2007)
- 3245 [91] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter.”
 3246 <http://hepdata.cedar.ac.uk/pdf/pdf3.html> , last accessed on 02.02.2018.
- 3247 [92] G. Altarelli and G. Parisi. “ASYMPTOTIC FREEDOM IN PARTON LAN-
 3248 GUAGE”, Nucl.Phys. B126:298 (1977).
- 3249 [93] Yu.L. Dokshitzer. Sov.Phys. JETP 46:641 (1977)
- 3250 [94] V.N. Gribov, L.N. Lipatov. “Deep inelastic e p scattering in perturbation the-
 3251 ory”, Sov.J.Nucl.Phys. 15:438 (1972)
- 3252 [95] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a
 3253 reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).
- 3254 [96] B. Andersson, G. Gustafson, G.Ingelman and T. Sjostrand, “Parton fragmen-
 3255 tation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145,
 3256 1983.

- 3257 [97] CMS Collaboration, “Event generator tunes obtained from underlying event
3258 and multiparton scattering measurements;” European Physical Journal C, vol.
3259 76, no. 3, p. 155, (2016).
- 3260 [98] J. Alwall et. al., “The automated computation of tree-level and next-to-leading
3261 order differential cross sections, and their matching to parton shower simula-
3262 tions,” Journal of High Energy Physics, vol. 07, p. 079, (2014).
- 3263 [99] T. Sjöstrand and P. Z. Skands, “Transverse-momentum-ordered showers and
3264 interleaved multiple interactions,” European Physical Journal C, vol. 39, pp.
3265 129–154, (2005).
- 3266 [100] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with
3267 Parton Shower simulations: the POWHEG method,” Journal of High Energy
3268 Physics, vol. 11, p. 070, (2007).
- 3269 [101] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments
3270 and Methods in Physics, vol. A506, pp. 250–303, (2003).
- 3271 [102] J.Allison et.al.,“Recent developments in Geant4”, Nuclear Instruments and
3272 Methods in Physics Research A 835 (2016) 186-225.
- 3273 [103] CMS Collaboration “Full Simulation Offline Guide”, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation>, last accessed 04.02.2018
- 3275 [104] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf.
3276 Ser. 513 022012 (2014)
- 3277 [105] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description
3278 with the CMS detector”, JINST 12 P10003 (2017) <https://doi.org/10.1088/1748-0221/12/10/P10003>.

- 3280 [106] The CMS Collaboration. “ Description and performance of track and pri-
3281 mary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014).
3282 doi:10.1088/1748-0221/9/10/P10009
- 3283 [107] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides.
3284 Retrieved from [https://indico.cern.ch/event/197461/contributions/](https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf)
3285 [1478917/attachments/290954/406673/CMS_4July2012_Final.pdf](https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf)
- 3286 [108] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by
3287 the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).
- 3288 [109] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of
3289 electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf
3290 C 0303241 (2003) TULT009 [physics/0306087].
- 3291 [110] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification,
3292 Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.
- 3293 [111] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”,
3294 CMS Note 2007-008 (2007).
- 3295 [112] CMS collaboration, “Performance of CMS muon reconstruction in pp collision
3296 events at $\sqrt{s} = 7 \text{ TeV}$ ”, JINST 7 P10002 2012, [arXiv:1206.4071].
- 3297 [113] Coco, Victor and Delsart, Pierre-Antoine and Rojo-Chacon, Juan and Soyez,
3298 Gregory and Sander, Christian, “Jets and jet algorithms”, Proceedings,
3299 HERA and the LHC Workshop Series on the implications of HERA for LHC
3300 physics: 2006-2008, pag. 182-204. <http://inspirehep.net/record/866539/files/access.pdf>, (2009), doi:10.3204/DESY-PROC-2009-02/54

- 3302 [114] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,”
3303 Journal of High Energy Physics, vol. 04, p. 063, (2008).
- 3304 [115] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, “Longitudi-
3305 nally invariant K_t clustering algorithms for hadron hadron collisions”, Nuclear
3306 Physics B, vol. 406, pp. 187–224, (1993).
- 3307 [116] Y.L. Dokshitzer, G.D. Leder, S.Moretti, and B.R. Webber, “Better jet clustering
3308 algorithms,” Journal of High Energy Physics, vol. 08, p. 001, (1997).
- 3309 [117] B. Dorney. “Anatomy of a Jet in CMS”. Quantum Diaries. June
3310 1st, 2011. Retrieved from [https://www.quantumdiaries.org/2011/06/01/
3311 anatomy-of-a-jet-in-cms/](https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/)
- 3312 [118] The CMS Collaboration.“Event Displays from the high-energy collisions at 7
3313 TeV”, May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from [https://cds.
3314 cern.ch/record/1429614.](https://cds.cern.ch/record/1429614)
- 3315 [119] The CMS collaboration. “Determination of jet energy calibration and transverse
3316 momentum resolution in CMS”. JINST 6 P11002 (2011). [http://dx.doi.org/
3317 10.1088/1748-0221/6/11/P11002](http://dx.doi.org/10.1088/1748-0221/6/11/P11002)
- 3318 [120] The CMS Collaboration, “Introduction to Jet Energy Corrections at
3319 CMS.”. <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>, last ac-
3320 cessed 10.02.2018.
- 3321 [121] CMS Collaboration Collaboration. “Identification of b quark jets at the CMS
3322 Experiment in the LHC Run 2”. Tech. rep. CMS-PAS-BTV-15-001. Geneva:
3323 CERN, (2016). [https://cds.cern.ch/record/2138504.](https://cds.cern.ch/record/2138504)

- 3324 [122] CMS Collaboration Collaboration. “Performance of missing energy reconstruc-
3325 tion in 13 TeV pp collision data using the CMS detector”. Tech. rep. CMS-PAS-
3326 JME16-004. Geneva: CERN, 2016. <https://cds.cern.ch/record/2205284>.
- 3327 [123] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,
3328 Retrieved from <http://cms.web.cern.ch/sites/cms.web.cern.ch/files/>
3329 [styles/large/public/field/image/HIG13004_Event01_0.png?itok=LAWZzPHR](#)
- 3331 [124] CMS Collaboration, “New CMS results at Moriond (Electroweak) 2013”,
3332 Retrieved from <http://cms.web.cern.ch/sites/cms.web.cern.ch/>
3333 [files/styles/large/public/field/image/TOP12035_Event01.png?itok=uMdnSqzC](#)
- 3335 [125] K. Skovpen. “Event displays highlighting the main properties of heavy flavour
3336 jets in the CMS Experiment”, Aug 2017, CMS-PHO-EVENTS-2017-006. Re-
3337 trieval from <https://cds.cern.ch/record/2280025>.
- 3338 [126] G. Cowan. “Topics in statistical data analysis for high-energy physics”.
3339 arXiv:1012.3589v1
- 3340 [127] A. Hoecker et al., “TMVA-Toolkit for multivariate data analysis”
3341 arXiv:physics/0703039v5 (2009)
- 3342 [128] L. Lista. “Statistical Methods for Data Analysis in Particle Physics”, 2nd
3343 ed. Springer International Publishing. (2017) <https://dx.doi.org/10.1007/978-3-319-62840-0>

- 3345 [129] I. Antcheva et al., “ROOT-A C++ framework for petabyte data storage, sta-
 3346 tistical analysis and visualization ,” Computer Physics Communications, vol.
 3347 182, no. 6, pp. 1384â€¢1385, (2011).
- 3348 [130] Y. Coadou. “Boosted decision trees”, ESIPAP, Archamps, 9 Febru-
 3349 ary 2016. Lecture. Retrieved from https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf
- 3352 [131] J.H. Friedman. “Greedy function approximation: A gradient boosting ma-
 3353 chine”. Ann. Statist. Volume 29, Number 5 (2001), 1189-1232. https://projecteuclid.org/download/pdf_1/euclid-aos/1013203451.
- 3355 [132] W. Verkerke and D. Kirkby, “The RooFit toolkit for data modeling,” arXiv
 3356 preprint physics, (2003).
- 3357 [133] CMS Collaboration, “Documentation of the RooStats-based statistics
 3358 tools for Higgs PAG”. <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit>, last accessed on 08.04.2018.
- 3360 [134] F. James, M. Roos, “MINUIT: Function minimization and error analysis”. Cern
 3361 Computer Centre Program Library, Geneve Long Write-up No. D506, 1989
- 3362 [135] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of
 3363 statistical hypotheses”. Springer-Verlag, (1992).
- 3364 [136] A.L. Read. “Modified frequentist analysis of search results (the CL_s method),”
 3365 (2000). CERN-OPEN-2000-205.
- 3366 [137] C. Palmer. “Searches for a Light Higgs with CMS”, CMS-CR-2012-215. <https://cds.cern.ch/record/1560435>.

- 3368 [138] A. Wald, “Tests of statistical hypotheses concerning several parameters when
 3369 the number of observations is large”, Transactions of the American Mathematical
 3370 society, vol. 54, no. 3, pp. 426–482, (1943).
- 3371 [139] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for
 3372 likelihood-based tests of new physics”, European Physical Journal C, vol. 71,
 3373 p. 1554, (2011).
- 3374 [140] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing
 3375 Composite Hypotheses”, Annals of Mathematical Statistics, vol. 9, pp. 60–62,
 3376 (03, 1938).
- 3377 [141] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated pro-
 3378 duction via gluon fusion in the SM and the 2HDM”, JHEP 06 (2015) 065,
 3379 [https://dx.doi.org/10.1007/JHEP06\(2015\)065](https://dx.doi.org/10.1007/JHEP06(2015)065), arXiv:1503.01656.
- 3380 [142] ATLAS Collaboration, “Measurements of Higgs boson pro-
 3381 duction and couplings in diboson final states with the AT-
 3382 LAS detector at the LHC”, Phys. Lett. B726 (2013) 88–119,
 3383 doi:10.1016/j.physletb.2014.05.011, 10.1016/j.physletb.2013.08.010,
 3384 arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].
- 3385 [143] CMS Collaboration, “Search for the associated production of a Higgs boson
 3386 with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV”, JHEP 06
 3387 (2016) 177, doi:10.1007/JHEP06(2016)177, arXiv:1509.08159.
- 3388 [144] B. Stieger, C. Jorda Lope et al., “Search for Associated Production of a Single
 3389 Top Quark and a Higgs Boson in Leptonic Channels”, CMS Analysis Note CMS
 3390 AN-14-140, 2014.

- 3391 [145] M. Peruzzi, C. Mueller, B. Stieger et al., “Search for ttH in multilepton final
3392 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-16-211, 2016.
- 3393 [146] CMS Collaboration, “Search for H to bbar in association with a single top quark
3394 as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis
3395 Summary CMS-PAS-HIG-16-019, 2016.
- 3396 [147] CMS Collaboration, “Search for production of a Higgs boson and a single top
3397 quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV”, CMS
3398 Physics Analysis Summary CMS-PAS-HIG-17-005, 2016.
- 3399 [148] CMS Collaboration, “PdmV2016Analysis,” (2016). <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2016Analysis#DATA>, last accessed 11.04.2016.
- 3401 [149] M. Peruzzi, F. Romeo, B. Stieger et al., “Search for ttH in multilepton final1
3402 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-17-029, 2017.
- 3403 [150] B. Maier, “SingleTopHiggProduction13TeV”, February, 2016. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV>.
- 3405 [151] B. WG, “BtagRecommendation80XReReco”, February, 2017. <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>.
- 3407 [152] CMS Collaboration, “Identification of b quark jets at the CMS Experiment
3408 in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001,
3409 2016.
- 3410 [153] CMS Collaboration, “Baseline muon selections for Run-II.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonIdRun2>, last accessed on
3411 24.02.2018.

- 3413 [154] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January,
3414 2015. [https://indico.cern.ch/event/368007/contribution/2/material/
3415 slides/0.pdf](https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf).
- 3416 [155] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”,
3417 October, 2014. [https://indico.cern.ch/event/298249/contribution/3/
3418 material/slides/0.pdf](https://indico.cern.ch/event/298249/contribution/3/material/slides/0.pdf).
- 3419 [156] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic
3420 Top Quarks at the LHC”, JHEP 03 (2011) 059, [https://dx.doi:10.1007/
3421 JHEP03\(2011\)059](https://dx.doi.org/10.1007/JHEP03(2011)059), arXiv:1007.2221.
- 3422 [157] CMS Collaboration. “Tag and Probe”, [https://twiki.cern.ch/twiki/bin/
3423 view/CMS/TagAndProbe](https://twiki.cern.ch/twiki/bin/view/CMS/TagAndProbe), last accessed on 02.03.2018.