



**HAL**  
open science

# Search for the production of the Higgs boson associated with a pair of top quarks with the Atlas detector at the LHC

Wang Chao

► **To cite this version:**

Wang Chao. Search for the production of the Higgs boson associated with a pair of top quarks with the Atlas detector at the LHC. High Energy Physics - Experiment [hep-ex]. AMU - Aix Marseille Université, 2017. English. NNT: . tel-02276363

**HAL Id: tel-02276363**

**<https://in2p3.hal.science/tel-02276363v1>**

Submitted on 2 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE D'AIX-MARSEILLE

SHANDONG UNIVERSITY

ECOLE DOCTORALE 352

School of Physics

Centre de Physique des Particules de Marseille

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Physique et Science de la Matière

Spécialité : Physique des Particules et Astroparticules

**Chao WANG**

Recherche de la production du boson de Higgs associé à une  
paire de quark top avec le détecteur ATLAS auprès du LHC  
Search for the production of the Higgs boson associated with a pair of  
top quarks with the ATLAS detector at the LHC

Soutenue le 6/12/2017 devant le jury :

Djamel BOUMEDIENNE	LPC	Examineur
Cristinel DIACONU	CPPM	Directeur de thèse
Cunfeng FENG	SDU	Examineur
Eric KAJFASZ	CPPM	Examineur invité
Yanwen LIU	USTC	Rapporteur
Lianliang MA	SDU	Directeur de thèse
Emmanuel MONNIER	CPPM	Directeur de thèse
Zhiqing ZHANG	LAL	Rapporteur
Xuai ZHUANG	IHEP	Examineur



SHANDONG UNIVERSITY

AIX MARSEILLE UNIVERSITY

DOCTORALE SCHOOL 352

School of Physics

Centre de Physique des Particules de Marseille

A dissertation to obtain the doctoral degree

Domain : Physics and Matter Science

Specialty : Particle Physics and Astroparticles

**Chao WANG**

Search for the production of the Higgs boson associated with  
a pair of top quarks with the ATLAS detector at the LHC

Recherche de la production du boson de Higgs associé à une paire de  
quark top avec le détecteur ATLAS auprès du LHC

Sustained on 6/12/2017 in front of a jury composed of:

Djamel BOUMEDIENNE	LPC	Examineur
Cristinel DIACONU	CPPM	Directeur de thèse
Cunfeng FENG	SDU	Examineur
Eric KAJFASZ	CPPM	Examineur invité
Yanwen LIU	USTC	Rapporteur
Lianliang MA	SDU	Directeur de thèse
Emmanuel MONNIER	CPPM	Directeur de thèse
Zhiqing ZHANG	LAL	Rapporteur
Xuai ZHUANG	IHEP	Examineur



# 5 Abstract

6 The production of the Higgs boson associated with a pair of top quarks is still not  
7 observed and is one of the most important Higgs boson production modes. Hence,  
8 the discovery of this production mode is a very attracting search: not only will it  
9 be the first time we can observe this Higgs production mode but also we will be  
10 able to measure its Yukawa coupling to the top quark. The measured results can  
11 answer the basic question of the Standard Model (SM) and can also search for any  
12 hints of new physics by comparing them to the SM prediction.

13 An analysis of searching for the production of the Higgs boson associated with  
14 a pair of top quarks in three leptons final state is presented in this thesis. This  
15 analysis is performed with the collected data by the ATLAS detector in 2015  
16 and 2016 during the so-called Run 2 campaign corresponding to an integrated  
17 luminosity of  $36.1 \text{ fb}^{-1}$  at a center of mass energy of  $\sqrt{s} = 13 \text{ TeV}$ . It uses a  
18 boosted decision tree algorithm to discriminate between signal and background.  
19 The dominant background of fake leptons is estimated with the data-driven method  
20 (Matrix Method).

21 For a Standard Model Higgs boson of a 125 GeV mass, an excess of events  
22 over the expected background from other SM processes is found with an observed  
23 significance of 2.2 standard deviations, compared to an expectation of 1.5 standard  
24 deviations. The best fit for the  $t\bar{t}H$  production cross section is  $1.5_{-0.7}^{+0.8}$  times the  
25 SM expectation, consistent with the value of the Yukawa coupling to top quarks  
26 in the Standard Model.

27 Keywords : LHC, ATLAS, Higgs,  $t\bar{t}H$ , Matrix Method, Multivariate analysis

# Résumé

## 29 Introduction

30 La production du boson de Higgs associée à une paire de quarks top est l'un des  
31 modes de production de boson de Higgs les plus importants bien que toujours pas  
32 encore observé. Par conséquent, découvrir ce mode de production est l'une des  
33 recherches les plus attrayantes après la découverte Higgs: non seulement cela sera  
34 la première fois que nous pourrons observer l'existence de ce mode de production  
35 du Higgs mais nous pourrons également en mesurer le couplage de Yukawa. Les  
36 résultats de ces mesures peuvent répondre aux questions fondamentales du Modèle  
37 Standard (SM) et peuvent également donner des indices de nouvelle physique en  
38 les comparant à la prédiction SM. Cela fait de l'étude de la recherche et de la  
39 mesure de la production l'une des analyses les plus importantes de l'ère post-Higgs.

40 Plusieurs canaux de désintégration du boson de Higgs ont été considérés pour  
41 mesurer le couplage de Yukawa dans la production de  $t\bar{t}H$  avec des états finaux  
42 multileptons. Des lots de données de collision proton-proton produits par le grand  
43 collisionneur de hadrons, ou Large Hadron Collider (LHC), installé au CERN  
44 (Organisation européenne pour la recherche nucléaire) à Genève ont été utilisés.  
45 Ils correspondent à des luminosités intégrées par expérience de  $5 \text{ fb}^{-1}$  à une énergie  
46 dans le centre de masse de 7 TeV (2011) et de  $20 \text{ fb}^{-1}$  à 8 TeV (2012) et représentent  
47 le Run 1. Les collaborations ATLAS et CMS, opérant les détecteurs éponymes  
48 auprès du LHC, ont chacune produit des résultats de combinaisons basées sur  
49 ces données du Run 1. Les significances combinées de deux expériences pour  
50 l'observation du processus  $t\bar{t}H$  sont de  $4.4\sigma$ , alors que seulement  $2.0\sigma$  est attendu,  
51 correspondant à un excès mesuré de  $2.3\sigma$  par rapport à la prédiction SM. Le  
52 LHC produit depuis 2015 des collisions à  $\sqrt{s} = 13 \text{ TeV}$  et la section efficace de  
53 production  $t\bar{t}H$  augmente d'un facteur quatre par rapport à 8 TeV. La mise à niveau  
54 pendant deux ans du détecteur ATLAS a permis d'améliorer ses performances et la  
55 précision des analyses physiques, en particulier pour la mesure SM et la recherche  
56 de nouvelle physique.

57 Une analyse de la recherche de la production de boson de Higgs associée à une  
58 paire de quarks top dans des états finaux à trois leptons est présentée dans cette  
59 thèse. L'analyse est réalisée avec des données collectées par le détecteur ATLAS  
60 en 2015 et 2016 à 13 TeV pendant la campagne dite Run 2 et correspondant à une  
61 luminosité intégrée de  $36.1 \text{ fb}^{-1}$ . L'analyse multivariée, dans sa version en arbre  
62 de décision renforcé, est utilisée pour distinguer le signal du bruit de fond. Les  
63 faux-leptons et les processus  $t\bar{t}V$  sont les bruits de fond principaux, en particulier  
64 le fond de faux leptons est estimé avec une méthode matricielle (MM) pilotée par  
65 les données (Méthode de la Matrice) dans le canal à trois leptons. Un ajustement  
66 simultané est utilisé pour estimer les résultats finaux et principalement la force du  
67 signal à valider la prédiction SM.

## 69 **Modèle standard et le détecteur ATLAS**

70 La théorie actuellement en cours, dominante et réussie, qui explique ces particules  
71 et champs fondamentaux, ainsi que leur dynamique, s'appelle le Modèle Standard  
72 (SM).

73 Le SM est devenu à une théorie complète décrivant les interactions de base entre  
74 les particules et les forces. La gravité, qui devrait probablement être médiée par  
75 une particule: le graviton, reste à observer.

76 Les particules fondamentales du SM sont des fermions et des bosons, avec res-  
77 spectivement des spins demi-entiers ou un entier (y compris zéro). Le spin est un  
78 nombre quantique signé intrinsèque à la particule avec une direction. Les fermions  
79 constituent la matière conventionnelle dans la nature. Les bosons sont des partic-  
80 ules de spin entières et sont les médiateurs des forces fondamentales.

81 Quatre forces fondamentales sont connues aujourd'hui: les forces forte, faible,  
82 électromagnétique et gravitationnelle. Bien qu'il y ait des particules massives, la  
83 force gravitationnelle n'a aucune influence au niveau des particules et est générale-  
84 ment ignorée dans le cadre SM. Les forces électromagnétique et gravitationnelle  
85 ont une portée infinie et obéissent à une loi carrée inverse. La portée des forces  
86 forte et faible peut être estimée par le principe d'incertitude de Heisenberg et les  
87 masses des particules porteuses de force, qui sont les gluons de la force nucléaire  
88 forte et les bosons  $W$  et  $Z$  de la force faible.

89 Les théories modernes décrivent les forces physiques en termes de champs,  
90 par exemple le champ électromagnétique, le champ gravitationnel et les champs  
91 qui décrivent les forces entre les particules élémentaires. Dans les théories de  
92 champs, différentes configurations des champs non observables peuvent donner des  
93 quantités observables identiques. Une transformation d'une telle configuration de  
94 champ à une autre est appelée une transformation de jauge. L'absence de change-  
95 ment dans les grandeurs mesurables, malgré la transformation du champ, est une  
96 propriété appelée invariance jauge et parfois aussi appelée symétrie jauge. Dans  
97 SM, la symétrie joue un rôle très important et SM est une théorie de jauge qui est  
98 construite en utilisant la symétrie. Un groupe de transformations qui peuvent être  
99 effectuées sur des champs est la base de la symétrie et laisse l'invariant lagrangien.  
100 Deux types de symétries sont présentés: les symétries globales et les symétries lo-  
101 cales. La signification de la symétrie est que les lois de conservation sont obtenues  
102 naturellement et les lois de conservation ont les symétries sous-jacentes.

103 Basé sur l'invariance de jauge locale, un lagrangien de symétrie peut être con-  
104 struit pour décrire l'électrodynamique quantique (QED) et la chromodynamique  
105 quantique (QCD). La symétrie de jauge peut garantir que la théorie est re-normalisée  
106 et que la théorie de perturbation peut être utilisée pour faire les prédictions. Mais  
107 elle interdit le terme de masse supplémentaire pour les bosons  $W$  et  $Z$ . Pour  
108 obtenir la masse, la symétrie de jauge doit être brisée tout en gardant le lagrangien  
109 symétrique. Cela peut se faire par le biais de la rupture de symétrie spontanée  
110 (SSB) en conduisant naturellement au boson de Higgs.



111 Les dernières mesures des propriétés du boson de Higgs, découvert avec succès  
112 par ATLAS et CMS dans les données du Run1 montrent une compatibilité remar-  
113 quable avec les prédictions SM, et aucune des collaborations n'a observé d'écarts  
114 significatifs par rapport à cette théorie. Le SM est la théorie, qui peut décrire  
115 la relation entre les particules et les interactions fondamentales, de loin la plus  
116 réussie. L'accord couvre plusieurs ordres de grandeur en sections efficaces et sur  
117 une grande variété de processus de référence du SM.

118 Malgré l'énorme succès du SM, il reste de nombreuses questions non résolues  
119 en physique, comme la matière noire, l'énergie noire, les problèmes de hiérarchie,  
120 l'asymétrie matière-antimatière ou l'oscillation des neutrinos. De plus, le SM ne  
121 tient pas compte de la gravité et il n'y a pas encore de compréhension confirmée  
122 expérimentalement de cette force au niveau quantique. Beaucoup d'efforts ont été  
123 faits pour combiner la gravité avec le SM pour former une théorie de tout, mais  
124 jusqu'ici les tentatives n'ont qu'un succès limité. Des modèles théoriques au-delà  
125 du Modèle Standard (BSM) sont nécessaires pour traiter ces problèmes, comme le  
126 modèle SUPER-SYmétrique dit SUSY.

127 Le Large Hadron Collider (LHC) est l'accélérateur et collisionneur d'hadrons  
128 le plus grand et le plus puissant du monde. Il s'agit de l'ajout le plus récent au  
129 complexe d'accélérateurs du CERN situé près de Genève à la frontière entre la  
130 France et la Suisse. Le LHC est situé dans un tunnel circulaire de 27 km à une  
131 profondeur variant entre 45 m et 170 m sous terre. Le LHC est conçu pour produire  
132 des collisions proton-proton jusqu'à une énergie dans le centre de masse de 14 TeV.  
133 La première collision proton-proton au LHC s'est produite en novembre 2009 et  
134 les premières collisions de 7 TeV ont débuté en 2010. Les collisions de 13 TeV ont  
135 débuté en 2015 après deux ans de fermeture pour améliorer le détecteur et le LHC.  
136 Quatre expériences principales ont été installées et développées au LHC. ATLAS et  
137 le CMS sont les plus grandes expériences et sont conçus pour comprendre le modèle  
138 standard avec précision. La physique du quark b est principalement étudiée par  
139 l'expérience LHCb et ALICE étudie le mécanisme de confinement des quarks.

140 ATLAS (A Toroidal LHC Apparatus) est une expérience à usage général conçue  
141 pour sonder un large éventail de données physiques allant de la QCD à basse én-  
142 ergie, des mesures de précision électrofaible et sur les quarks top, à la recherche du  
143 boson de Higgs et de la nouvelle physique à l'échelle du TeV. ATLAS mesure 44 m  
144 de long, 25 m de haut et pèse environ 7 000 tonnes. En commençant par son cœur  
145 et en se déplaçant vers l'extérieur, on trouve d'abord le détecteur interne (ID),  
146 structuré en trois couches et responsable de la mesure précise de l'impulsion des  
147 particules chargées dans un champ magnétique solénoïdal de 2 T. On a ensuite les  
148 calorimètres électromagnétiques et hadroniques, qui mesurent les dépôts d'énergie  
149 des électrons, des photons et des hadrons et déduisent également l'énergie transver-  
150 sale manquante des neutrinos et d'autres particules non détectées. Enfin on a un  
151 spectromètre à muons de précision (MS) qui permet le suivi et le déclenchement  
152 des muons dans un champ magnétique toroïdal. La collaboration ATLAS compte  
153 environ 3000 personnes provenant de 178 institutions dans 38 pays. ATLAS a

154 enregistré ses premières collisions en 2009, après un arrêt d'un an dû à un accident  
155 sur un des aimants supraconducteurs du LHC, et a ensuite fonctionné avec succès  
156 jusqu'au début 2013 où il a été arrêté pour une période de mise à niveau de deux  
157 ans. ATLAS a commencé à prendre des données à 13 TeV en 2015 et continuera  
158 à fonctionner jusqu'à la prochaine mise à niveau fin 2018.

159

## 160 **Recherche de la production du boson de Higgs associé à une paire de** 161 **quarks top**

162 Le couplage du boson de Higgs à d'autres particules peut être mesuré en étudiant  
163 les différents processus de production et de modes de désintégration du boson de  
164 Higgs. La production associée d'un boson de Higgs et d'une paire de quarks top  
165 peut fournir une information directe du couplage de Yukawa Higgs-Top, probable-  
166 ment le couplage le plus crucial aux fermions. Ce processus implique deux gluons  
167 en collision, qui se décomposent en une paire de quark-antiquark. Un quark et  
168 un antiquark de chaque paire peuvent ensuite se combiner pour former une par-  
169 ticule de Higgs. La mesure du couplage de Yukawa relie deux des particules les  
170 plus lourdes du modèle standard, le quark top (173 GeV) et le boson de Higgs  
171 (125 GeV). Le couplage Yukawa du boson de Higgs au quark top est un paramètre  
172 clé du SM. La comparaison de ces mesures aux prédictions du SM a le potentiel  
173 d'identifier sans ambiguïté de nouveaux effets physiques qui peuvent modifier la  
174 section efficace de la production de  $t\bar{t}H$  par rapport à l'espérance SM.

175 La production de  $t\bar{t}H$  peut être observée à travers différentes topologies selon les  
176 désintégrations de Higgs et de quarks supérieurs. L'analyse présentée dans cette  
177 thèse est conçue pour rechercher des désintégrations de Higgs en paires de bosons  
178  $W$ , le quark top et le quark anti-top se désintégrant en  $Wb$ . Chaque boson  $W$   
179 se désintègre, soit leptoniquement ( $\ell =$  électron, muon, tau) avec la production  
180 d'énergie manquante, soit hadroniquement, dans de nombreuses topologies. Selon  
181 le nombre de leptons dans l'état final des événements de signal, les topologies  
182 pourraient être avec  $2\ell SS$  (exactement deux leptons légers de mêmes signe de  
183 charges et avec un veto sur les taus hadroniques); avec  $3\ell$  (exactement 3 leptons  
184 légers); avec  $4\ell$  (exactement 4 leptons), et. L'orthogonalité des événements entrant  
185 dans les différents canaux est assurée dans la définition des canaux. La recherche  
186 de  $t\bar{t}H$  dans l'état final avec exactement 3 leptons est présentée dans cette thèse,  
187 c'est mon travail principal pendant mon doctorat.

188 L'analyse présentée dans cette thèse utilise les  $36.1 \text{ fb}^{-1}$  de données, dites "Run  
189 2", collectées à partir de collision proton-proton enregistrée par le détecteur AT-  
190 LAS à  $\sqrt{s} = 13 \text{ TeV}$  entre 2015 et 2016. Les données sont ensuite préparées en  
191 utilisant un cadre dédié de réduction et compactification dédié, appelé dérivation.  
192 Cette dérivation fournit une réduction spécifique pour les événements  $t\bar{t}H$  avec des  
193 multileptons dans les états finaux. La taille totale de l'ensemble de données a été  
194 réduite à 3,6% du total disponible pour les événements  $t\bar{t}$  simulés et à 0.1%  
195 pour les ensembles de données de collision.

196 Tous les lots de données simulées par Monte Carlo sont traités par une simu-

197 lation complète de la réponse du détecteur ATLAS basée sur Geant4. Tous les  
198 événements simulés ont ensuite été traités en utilisant les mêmes algorithmes de  
199 reconstruction et la même chaîne d'analyse que les données réelles de collisions.  
200 Les événements simulés sont corrigés afin que les efficacités de reconstruction et  
201 d'identification des objets, les échelles d'énergie et les résolutions d'énergie corre-  
202 spondent à celles déterminées à partir des données.

203 Après les sélections de base décrites ci-dessus, au niveau de l'événement, le  
204 vertex primaire avec l'exigence d'être celui ayant les traces associées de plus haut  
205  $p_T$  est requis. Les événements avec un bruit important dans les calorimètres ou  
206 avec de données corrompues sont supprimés et une procédure appelée suppression  
207 de chevauchement est conçue pour éviter le double comptage d'objets et pour  
208 supprimer les leptons ayant une forte probabilité de provenir des désintégrations  
209 de hadrons. Deux stratégies de déclenchement différentes ont été adoptées en 2015  
210 et 2016 pour tenir compte des différents schémas de collision pour chaque période.

211 Tous les critères de sélections des candidats leptons sont regroupés et appelés  
212 objets de niveau lâche. Les objets de niveau serré doivent, sous certaines con-  
213 ditions, satisfaire au point de fonctionnement de l'algorithme de maximum de  
214 vraisemblance (LH) utilisé pour les électrons afin de réduire les électrons faux ou  
215 non direct. Un algorithme dédié pour l'isolation, par arbre de décision renforcé et  
216 produisant la variable `PromptLeptonIso`, a été construit pour améliorer la réduction  
217 du bruit de fond de leptons non prompts produits dans les désintégrations de  
218 hadrons.

219 Les jets sont reconstruits à partir d'amas topologiques calibrés construits à partir  
220 des dépôts d'énergie dans les calorimètres, en utilisant l'algorithme anti-kt. Les  
221 jets ayant des contributions énergétiques susceptibles d'être générés par les effets  
222 de bruit ou de détecteur sont supprimés et seuls les jets satisfaisant une impulsion  
223 transverse,  $p_T$ , supérieure à 25 GeV sont conservés. Les jets contenant des  $b$ -  
224 hadrons sont identifiés par une méthode discriminante multivariée. Le point de  
225 fonctionnement utilisé dans cette recherche correspond à un rendement moyen de  
226 70% pour  $b$ -tagged jets de  $p_T > 20$  GeV.

227 Dans l'état final à trois leptons, la valeur absolue de la charge totale des trois  
228 leptons dans les événements doit être 1, aucun candidat tau n'est autorisé. Parmi  
229 ces trois leptons, celui de charge opposée à celles des deux autres est appelée "lep-  
230 ton 0". Le bruit de fond résultant de l'addition d'un lepton faux ou non prompt à  
231 un événement di-lepton de charges opposées aura un lepton supplémentaire comme  
232 "Lepton 1" ou "Lepton 2" parce que ce sera la même charge qu'un prompt lepton.  
233 En plus des exigences strictes sur "Lepton 1" et "Lepton 2", un  $p_T > 15$  GeV pour  
234 les deux leptons est également requis. L'exigence supplémentaire de  $p_T > 10$  GeV  
235 pour le "Lepton 0" est plus faible que "Lepton 1" et "Lepton 2", ainsi que l'isolation  
236 et l'identification du lepton faites avec un point de fonctionnement lâche. D'autre  
237 recommandations sont également appliquées pour éliminer les leptons de désinté-  
238 gration de quarks et les bruits de fond potentiels contenant des désintégrations de  
239  $Z$  en leptons où un des leptons a une impulsion très faible et n'est pas reconstruit.

240 Les jets et les jets  $b$ -étiquetés sont contraint d'avoir au moins 2 jets et au moins 1  
241 jets  $b$ -étiqueté.

242

### 243 **Estimation des faux leptons**

244 Plusieurs processus physiques peuvent produire la même signature que le signal.  
245 Une bonne estimation de ces bruits de fond est un des aspects les plus importants  
246 pour cette analyse. Les bruits de fond dans cette analyse sont classés en deux  
247 catégories:

248 Irréductible : Les événements qui peuvent conduire au même état final que le  
249 signal et l'état final contiennent trois leptons chargés. Les processus  $t\bar{t}W$ ,  $t\bar{t}Z$  et  
250 di-boson sont dans cette catégorie. L'estimation de ces fonds irréductibles s'appuie  
251 sur les simulations Monte Carlo.

252 Réductible : La plupart du temps, les événements pour lesquels un lepton non  
253 prompt ou un lepton faux est sélectionné comme lepton prompt sont appelés bruit  
254 de fond réductible. Les processus peuvent conduire à un état final compatible avec  
255 le signal avec l'objet mal reconstruit. Les principaux processus sont  $t\bar{t}$ ,  $Z$  + jets.

256 La méthode basée sur les données (la méthode matricielle dans cette analyse)  
257 est utilisée pour obtenir l'estimation du fond réductible. A titre de référence, les  
258 résultats du Run 1 montrent que les bruits de fond dominant sont les processus  
259 non-prompt (faux), le processus  $t\bar{t}V$  et dans une moindre mesure les contributions  
260 di-bosons  $VV$ .

261 En raison de la résolution limitée du détecteur, les objets ne sont pas reconstruits  
262 parfaitement. L'un des objets mal reconstruits, les faux leptons, est un problème  
263 commun dans les analyses. Notamment pour l'analyse qui implique des objets  
264 leptoniques comme l'étude présentée dans cette thèse. Les principales origines de  
265 ces faux leptons sont: la conversion des photons, et les jets de saveurs légères et  
266 lourdes.

267 Une étude spécifique de l'origine des faux leptons a été réalisée dans l'analyse  
268  $3\ell$  pour les faux électrons et les muons en utilisant les informations de vérité de  
269 lot d'événements  $t\bar{t}$  MC. En ce qui concerne la composition de l'origine des faux  
270 leptons dans la région de présélection de  $3\ell$ , les résultats montrent que les faux  
271 issus de saveurs lourdes (faux leptons qui proviennent des mésons  $b$  et  $c$ ) dominent  
272 à la fois les faux électrons et les faux muons; une contribution importante des  
273 conversions de photons produisant des faux électrons, mais aucune contribution  
274 de ce type dans le cas des faux muons comme prévu.

275 La méthode matricielle (MM) est une technique basée sur les données pour avoir  
276 l'estimation de la fausse contamination dans l'analyse. L'estimation ne peut pas  
277 décrire les faux objets en utilisant les échantillons MC seulement car le MC n'est  
278 pas fiable dans tous les cas et il y a encore beaucoup de cas non prévus dans le jeu de  
279 données réelles. Cette technique basée sur les données peut utiliser les informations  
280 provenant d'un jeu de données de collision réel et élimine la dépendance au MC.  
281 La MM est une méthode basée sur les données largement utilisée dans les analyses  
282 d'ATLAS. Pour avoir une bonne estimation du bruit de fond de faux, la MM est

283 donc également utilisé dans l'analyse à trois leptons dans l'état final. La MM y est  
284 utilisé pour estimer les faux électrons et les muons dans une région à contrainte  
285 serrée en utilisant l'information d'objet lâches dans une région à contrainte anti-  
286 serrée.

287 Les faux leptons sont estimés avec une matrice  $2 \times 2$  dans l'analyse  $2\ell SS$ . Une  
288 méthode matricielle similaire a donc été également développée pour l'analyse à  
289  $3\ell$  mais avec une matrice  $8 \times 8$ . La méthode matricielle simplifiée dans  $3\ell$  est  
290 introduite avec une assumption: le "Lepton 0", qui est de signe opposé, a une très  
291 faible possibilité d'être le faux lepton. La simulation MC  $t\bar{t}$  est utilisée pour vérifier  
292 cela avec une pré-sélection de  $3\ell$ . Il a été trouvé que la possibilité pour le "Lepton  
293 0" d'être le faux dans la région de signal pré-MVA  $3\ell$  est seulement de 1%.

294 L'échantillon  $t\bar{t}b\bar{a}$  simulé est utilisé pour la mesure réelle et l'extraction des  
295 fausse pendant le test de cohérence. Un échantillon de  $t\bar{t}$  dédié généré avec un  
296 filtre dileptonique est utilisé pour le cas  $3\ell$  pour améliorer les statistiques. En  
297 ce qui concerne le rayonnement de photon supplémentaire, un autre échantillon  
298 ciblant spécifiquement la production de  $t\bar{t}$  avec un photon rapide supplémentaire  
299 rayonné par l'un des quarks top a été généré et utilisé.

300 Deux régions de contrôle, enrichies en vraies et faux leptons, ont été conçues  
301 pour mesurer l'efficacité des leptons réels et faux à passer les exigences de sélection  
302 serrées. Ces régions ont des statistiques suffisamment grandes et peuvent  
303 représenter la cinématique et la composition du bruit de fond dans la région de  
304 signal. Les critères de sélection des di-leptons assurent l'orthogonalité à la région  
305 de signal  $3\ell$ .

306 Un test de fermeture dédié est effectué dans l'analyse  $3\ell$  pour vérifier la perfor-  
307 mance de la MM avec la sélection pré-MVA. Les résultats montrent que l'accord est  
308 bon entre la prédiction MC et l'estimation MM eut égard aux statistiques faibles  
309 et les grandes incertitudes.

310

### 311 **Méthode d'analyse multivariée**

312 La technique d'analyse multi-variable a été utilisée dans l'expérience ATLAS pour  
313 de nombreuses analyses. En particulier pour l'analyse présentée dans cette thèse,  
314 les arbres de décision renforcés (BDT) sont optimisés pour rejeter davantage le faux  
315 fond réductible qui est principalement du  $t\bar{t}b\bar{a}$  et le fond irréductible qui sont des  
316  $t\bar{t}W$  et  $t\bar{t}Z$ . La sélection de base pré-MVA est requise et la contribution dominante  
317 provient du bruit de fond de faux et des  $t\bar{t}V$ . Pour les supprimer, une stratégie  
318 d'analyse BDT à deux dimensions a été est conçue pour augmenter le nombre  
319 d'événements de signal dans la région pré-MVA. Un ensemble de variables choisies  
320 est utilisé comme entrée du BDT. L'accord entre les données et les échantillons  
321 MC pour ces entrées a été mesuré comme bon et montrent le pouvoir discriminant  
322 prometteur entre les événements de signal et de fond.

323 Un algorithme d'arbre de décision renforcé par gradient (BDTG) est entraîné sur  
324 un lot de données simulées pour le fond irréductible et pour l'estimation basée sur  
325 les données avec un total de 10 variables. Le bruit de fond de faux estimé avec la

326 méthode s'appuyant sur les données a été utilisé pendant l'entraînement du BDTG  
 327 au lieu du MC. Les études ont été faites pour trouver une bonne combinaison de  
 328 ces deux BDTG. Enfin le BDTG combiné discriminant a été ensuite divisé en six  
 329 zones pour optimiser le signal et le fond en utilisant la fonction TransfoD appelée  
 330 auto-binning. Généralement, le BDT combiné a un bon accord entre les données  
 331 et le MC et aussi une bonne séparation pour les événements de signal.

332

### 333 Résultats

334 La distribution de sortie BDT combiné et segmentée, "binée" est utilisé comme  
 335 entrée de l'ajustement. Toutes les incertitudes systématiques associées (systé-  
 336 matique expérimentale, systématique théorique et incertitudes statistiques) sont  
 337 considérées. Les bruits de fond de faux issu des données,  $t\bar{t}V$  et dibosons sont  
 338 contraints simultanément par l'ajustement.

339 Les résultats attendus sont d'abord réalisés avec l'ensemble de données Asimov.  
 340 Puis, l'ensemble total des données est utilisé pour les résultats finaux. La valeur  
 341 attendue de l'intensité du signal  $t\bar{t}H$  qui a été obtenue pour le canal  $3\ell$ , en ajustant  
 342 les données d'Asimov, est de:  $1_{-0.70}^{+0.77}$ , la signification correspondante est de 1.45  
 343 écart-type avec 1.21 d'incertitude. Les paramètres de nuisance (NP) correspondant  
 344 aux incertitudes systématiques sont tous centrés sur zéro et les facteurs d'échelle  
 345 de normalisation sont tous centrés autour de 1 comme prévu. Surtout en  $3\ell$ , les  
 346 NPs de haut rang sont ceux qui sont liés à l'estimation de faux, comme la non-  
 347 fermeture des faux en  $3\ell$ , la statistiques de faux muon, et aussi la différence entre  
 348  $2\ell$  et  $3\ell$  à cause des fausses conversions de photons. Généralement, il n'y a pas de  
 349 problème de sur-contrainte dans le canal  $3\ell$ .

350 L'ajustement a été effectué pour les données dans une région de signal et les  
 351 résultats sont: la force du signal observée est de  $1.5_{-0.7}^{+0.8}$ , la signification correspon-  
 352 dante a atteint 2.18 écarts-types avec l'incertitude 1.48. L'amélioration est grande  
 353 par rapport aux résultats du Run 1 en raison de la section efficace plus élevée et  
 354 une stratégie d'analyse avancée et optimisée.

355 ATLAS a publié les résultats approuvés avec  $36.1 \text{ fb}^{-1}$  dans les canaux multilep-  
 356 tons incluant celui à  $3\ell$ . Plus particulièrement dans le canal  $3\ell$ , le résultat publié  
 357 utilise une stratégie d'analyse multi-classe, aussi basée sur une analyse multi-variée.  
 358 La plus grande différence entre la méthode utilisée dans cette thèse, utilisée comme  
 359 vérification dans le résultat publié, et cette version multi-classe retenue pour ce  
 360 résultat est la procédure de catégorisation appliquée sur la sortie BDT finale pour  
 361 avoir une région de signal et de contrôle pure. L'analyse multi-classes a une région  
 362 de signal et quatre autres régions de contrôle qui couvrent les processus  $t\bar{t}W$ ,  $t\bar{t}Z$ ,  
 363 fakes et diboson. Toutes ces régions sont basées sur la sortie de l'entraînement  
 364 d'un BDT 5D.

365 La valeur la mieux ajustée de la force du signal observé (attendu), combinant  
 366 tous les canaux multileptons, est de  $1.54_{-0.42}^{+0.49}(1.00_{-0.39}^{+0.43})$ . La signification observée  
 367 (attendue) dans l'hypothèse de bruit de fond seul est de 4.1 écarts-types (2.7 écarts-  
 368 types). Cet excès correspond à l'évidence de la production de  $t\bar{t}H$  dans les états

369 finaux à plusieurs leptons. Les résultats présentés par ATLAS sont compatibles  
370 avec les résultats présentés dans cette thèse avec même une légère amélioration  
371 obtenue avec l'analyse multivariée utilisée.

372

## 373 **Conclusion**

374 Le boson de Higgs a été découvert en 2012 par ATLAS et CMS. Cette dernière  
375 pièce du modèle standard trouvée, la structure du modèle standard est presque  
376 terminée. Mais encore beaucoup de questions reste à répondre, le couplage de  
377 Higgs-Top Yukawa en est l'une d'entre elle.

378 Des mesures précises sur ses propriétés confirmeront la nature du boson de Higgs,  
379 et tout écart par rapport à la prédiction du modèle standard représentera un signe  
380 clair d'une nouvelle physique. La production associée du boson de Higgs avec une  
381 paire de quarks top en permet une mesure directe et est le sujet principal de cette  
382 dissertation.

383 Le travail présenté dans cette thèse porte sur la recherche de la production de  
384 bosons de Higgs associée à une paire de quarks top avec un état final à trois leptons.

385 Une technique multidimensionnelle (BDT) est employée pour avoir un bon pou-  
386 voir de discrimination entre le signal et le bruit de fond. L'un des principaux bruits  
387 de fond du aux faux leptons est estimé avec une méthode matricielle s'appuyant sur  
388 les données. Une bonne modélisation de ce fond et une meilleure séparation avec  
389 la sortie BDT sont obtenues. Les résultats finaux sont obtenus avec un ajustement  
390 simultané sur un BDT 2D combiné.

391 Pour une masse du boson de Higgs de 125 GeV, un excès d'événements par rap-  
392 port au bruit de fond attendu provenant d'autres processus du modèle standard est  
393 observé avec une signification observée de 2.18 écarts-types, contre une espérance  
394 de 1.45 écart-type. Le meilleur ajustement pour la section efficace de production  
395 de  $t\bar{t}H$ , en supposant une masse de boson de 125 GeV de Higgs, est  $1.5_{-0.7}^{+0.8}$  fois  
396 la valeur SM, et est compatible avec la valeur du couplage Yukawa au quark top  
397 dans le Modèle standard. Les résultats sont compatibles avec les résultats pub-  
398 liés d'ATLAS en utilisant le même ensemble de données mais avec une stratégie  
399 d'analyse différente pour les états finaux à trois leptons et des améliorations peu-  
400 vent être introduites dans l'analyse présentée dans cette thèse.

401 Comparé aux résultats de Run 1, une grande amélioration a été obtenue à la  
402 fois sur la signification et sur l'incertitude sur la mesure de l'intensité du signal.  
403 Cela est dû entre autre à l'énergie élevée dans le centre de masse qui apporte  
404 une plus grande section efficace, aux améliorations apportées au détecteur AT-  
405 LAS, ainsi qu'à la stratégie d'analyse optimisée. Des incertitudes toujours im-  
406 portantes sur les mesures finales sont tout de même observées et nécessitent une  
407 méthode ré-optimisée pour réduire l'impact de ces sources d'incertitude. Le résul-  
408 tat prometteur montre l'évidence de la production de  $t\bar{t}H$  mais plus de données  
409 seront nécessaires pour une mesure plus précise.

410 Après le redémarrage du LHC en 2015, le LHC a fourni davantage de données  
411 et plus de possibilités pour valider les prédictions du SM et rechercher de nou-

412 velles physiques. Des mises à niveau du détecteur ATLAS et de ses sous-systèmes  
413 ont été apportées de façon limitées. En 2018, le LHC sera arrêté pour un nou-  
414 veau cycle d'amélioration et d'ATLAS fera de même en apportant entre autre des  
415 améliorations au système de déclenchement afin de faire face à des taux de dé-  
416 clenchement plus élevés du à l'accroissement de la luminosité instantanée qu'aura  
417 le LHC. Après un redémarrage en 2019, le LHC fonctionnera pendant trois ans à  
418 un débit de 14 TeV et recueillera  $300 \text{ fb}^{-1}$  de données.

419 L'arrêt définitif du LHC se fera en 2023, période au cours de laquelle ATLAS  
420 installera un nouveau système de trajectographie intérieur plus résistant aux ray-  
421 onnements et plus granulaire. Un nouveau schéma de déclenchement, ainsi que des  
422 améliorations de l'électronique du spectromètre à muon et du calorimètre, seront  
423 réalisés. Le LHC à haute luminosité amélioré (HL-LHC) débutera en 2025, per-  
424 mettant au collisionneur de produire  $3000 \text{ fb}^{-1}$  de données. Les nouveaux jeux de  
425 données permettront alors des mesures des propriétés du boson de Higgs beaucoup  
426 plus précises afin de valider les prédictions par SM.



# Acknowledgements

428 The thesis writing is almost finished when I am wiring this acknowledgements and  
429 many previous memories come to eyes. I enjoyed the three years PhD studying  
430 life in Marseille which is a beautiful coastal city with lots of friends. I like the  
431 sea and enjoy the sunshine here. The three years of PhD studying is the most  
432 unforgettable memory in my whole life. I spent almost the whole time of PhD  
433 studying in CPPM as a co-PhD student between Aix-Marseille University and  
434 Shandong University.

435 Firstly, I should say thank to my supervisor from Shandong University, Prof.  
436 Lianliang Ma and the supervisor from CPPM, Dr. Emmanuel Monnier. They  
437 provide me the opportunity of studying the projects I am interested in and give me  
438 the chance to do this research in CPPM which has lots of experience on particle  
439 physics. As the leader of the ATLAS analysis team in Shandong University, Prof.  
440 Lianliang Ma taught me many things on data analysis. Dr. Emmanuel Monnier  
441 is as a coordinator of the LAr detector team and helped me a lot understand the  
442 knowledge on hardware. Besides, Dr. Emmanuel is a so kind person who offers a  
443 lot helps in both studying and ordinary life.

444 I should say thanks to Dr. Cristineal Diaconu who is my co-supervisor in CPPM.  
445 He guided me into the exact data analysis work during the master stage and we  
446 worked together when I did my qualification task. I learnt a lot from him and am  
447 benefit a lot from his working style.

448 Sincerely thanks to Prof. Cunfeng Feng who leads me into the experimental  
449 high energy physics and guides me the whole master stage. He let me know this  
450 beautiful field and provided me the chance to access the LHC as well as the ATLAS  
451 experiment.

452 Many thanks to the Dr. Kun Liu who is the convenor of the analysis team and  
453 colleague at the same time. His charming ideas and helpful advice helped me when  
454 I am confused.

455 Thanks to Jian Liu and Ruiqi Zhang who are my friends in Marseille. I got lots  
456 of experience on hardware and analysis from them. Also we helped each other a  
457 lot in ordinary life.

458 Special thanks to Bo Liu, Xin Xia, Yongke Zhao, Mengting Yang who are my  
459 friends and colleagues in Shandong University and they are so kind friends who I  
460 met.

461 Thanks to Venugopal Ellajosyula and Robert Wolff who are my colleagues in  
462 CPPM and cooperators of the analysis.

463 Finally, I should thank to my parents who support me all the time.

464 Many people and friends helped me a lot and encouraged me during those three  
465 years and I could not express my thanks any more to them who I do not mention  
466 here.

# 467 Contents

468	<b>Abstract</b>	<b>5</b>
469	<b>Résumé</b>	<b>15</b>
470	<b>Acknowledgments</b>	<b>16</b>
471	<b>List of Figures</b>	<b>20</b>
472	<b>List of Tables</b>	<b>23</b>
473	<b>Introduction</b>	<b>26</b>
474	<b>1 The Standard Model</b>	<b>27</b>
475	1.1 The Standard Model particles	27
476	1.1.1 Fundamental forces	29
477	1.2 Gauge symmetry	30
478	1.2.1 Quantum electrodynamics	30
479	1.2.2 Quantum chromodynamics	32
480	1.2.3 Weak interaction	33
481	1.3 Brout-Englert-Higgs mechanism	34
482	1.4 Current status and remaining issues	36
483	1.5 The production of the Higgs boson associated with a pair of top	
484	quarks ( $t\bar{t}H$ )	41
485	1.6 $t\bar{t}H$ searching at the LHC in Run 1	42
486	<b>2 Experimental apparatus</b>	<b>45</b>
487	2.1 The Large Hadron Collider	45
488	2.2 The ATLAS detector	50
489	2.2.1 The ATLAS coordinate system	51
490	2.2.2 Inner tracking detector	53
491	2.2.3 The ATLAS calorimeter	54
492	2.2.4 Muon spectrum system	56
493	2.2.5 Trigger system	56
494	2.2.6 The ATLAS simulation and computing	57
495	2.2.7 The LHC and ATLAS in Run 2	59
496	<b>3 Object reconstruction in the ATLAS</b>	<b>62</b>
497	3.1 Electrons	62
498	3.1.1 Reconstruction	63
499	3.1.2 Calibration	64

500	3.1.3 Identification	65
501	3.2 Photons	69
502	3.3 Muons	71
503	3.4 Hadronic taus	73
504	3.5 Jets	73
505	3.6 Missing transverse energy	76
506	<b>4 Search for the production of the Higgs boson associated with a pair of top quarks</b>	<b>78</b>
507	4.1 Introduction	78
508	4.2 Data and Monte Carlo samples	79
509	4.2.1 Collision data	79
510	4.2.2 Monte Carlo samples	80
511	4.3 Object definition and basic selection	82
512	4.3.1 Trigger	82
513	4.3.2 Electrons	83
514	4.3.3 Muons	83
515	4.3.4 Jets and $b$ -tagged jets	84
516	4.3.5 Overlap removal	85
517	4.3.6 Event pre-selection (The $3\ell$ pre-MVA selection)	85
518	4.4 Background estimation	87
519	4.4.1 Fake leptons	87
520	4.4.2 The Matrix Method	88
521	4.5 Background suppressing using the multi-variate analysis technique	106
522	4.5.1 Modelling of the MVA input variables	106
523	4.5.2 Performance of the BDT	106
524	4.6 Validation and control regions	112
525	4.6.1 $t\bar{t}W$	112
526	4.6.2 $t\bar{t}Z$	112
527	4.7 Uncertainties	114
528	4.7.1 Experimental systematics	114
529	4.7.2 Signal and background modelling theoretical systematics	116
530	4.8 Statistical treatment	118
531	4.8.1 Likelihood function	118
532	4.8.2 Hypothesis testing	119
533		
534	<b>5 Results with 13 TeV Run 2 data</b>	<b>122</b>
535	5.1 Asimov fit and expected significance	122
536	5.2 Data fit	122
537	5.3 Interpretation	130
538	<b>Conclusion</b>	<b>133</b>
539	<b>Bibliography</b>	<b>134</b>

540	<b>Appendix</b>	<b>141</b>
541	A The optimization of the cut-based electron identification in the AT-	
542	LAS towards Run 2	141
543	A.1 Introduction and motivation	141
544	A.2 Monte Carlo samples	141
545	A.3 Performance of LikelihoodPCA	142
546	A.4 Cut-based optimisation methodology	142
547	A.5 Coping with the changes in the TRT	150
548	A.6 Results	155
549	A.7 Conclusion	157

# 550 List of Figures

551	1.1	Standard Model of elementary particles	28
552	1.2	Feynman diagrams of the leading four Higgs boson production mechanisms at the LHC: ggF, VBF, VH and $t\bar{t}H$	37
553			
554	1.3	The SM Higgs boson decay branching ratios	37
555	1.4	The signal strength measured for the different production processes	39
556	1.5	Summary of several Standard Model total production cross section measurements	40
557			
558	1.6	SM Higgs boson cross sections of different productions versus the centre mass of energy	42
559			
560	1.7	Best-fit values of the signal strength parameter of $t\bar{t}H$ multilepton final state in Run 1 at the ATLAS	43
561			
562	2.1	The accelerator complex at CERN	46
563	2.2	Cross section of the LHC dipole	48
564	2.3	Delivered Luminosity versus time for 2011-2016 and total integrated luminosity and data quality in 2011 and 2012	49
565			
566	2.4	Number of interactions per crossing in 2011 and 2012	49
567	2.5	The overview of the ATLAS detector	50
568	2.6	The coordinate system in the ATLAS detector	52
569	2.7	View of the ATLAS inner detector	53
570	2.8	Overview of the ATLAS calorimeter	54
571	2.9	Overview of particles passage through the ATLAS's sub detector	55
572	2.10	Schematic view of the muon spectrometer in the x-y projections	57
573	2.11	The ATLAS trigger system and its integration	58
574	2.12	The main 2013-2014 LHC consolidations toward Run 2 data taking	59
575	2.13	Total integrated luminosity in 2016 and number of interactions per crossing with the combined 13 TeV data for 2015 and 2016	60
576			
577	3.1	The efficiency to identify electrons from $Z \rightarrow ee$ decays and the efficiency to identify hadrons as electrons estimated using simulated dijet samples	68
578			
579			
580	3.2	Illustration of the production of a $b$ -tagged jet	75
581	3.3	The performance of the MV2c10 algorithm	76
582	3.4	The resolution of the combined distribution of $E_x^{miss}$ and $E_y^{miss}$ for the TST $E_T^{miss}$	77
583			
584	4.1	Feynman diagram of the $t\bar{t}H$ signature with 3 leptons final state	79
585	4.2	The raw number of event and the corresponding percentage for the real and the fake lepton events in $t\bar{t}$ MC	90
586			
587	4.3	Real lepton enriched control region $t\bar{t}$ composition	93

588	4.4	Real efficiencies for electrons and muons from CR in data	94
589	4.5	Fake electron origin fraction for $t\bar{t} + t\bar{t}\gamma$ events	95
590	4.6	Origin fraction of fake electrons for $t\bar{t} + t\bar{t}\gamma$ events in the dileptonic CR, 2 $\ell$ SS SR and 3 $\ell$ SR	96
591			
592	4.7	Origin fraction of fake electrons and fake muons for $t\bar{t} + t\bar{t}\gamma$ events in the 3 $\ell$ by flavoured channel	98
593			
594	4.8	Fake rate for electrons from CR in data	99
595	4.9	Origin fraction of fake muons for $t\bar{t} + t\bar{t}\gamma$ events in the dileptonic CR, 2 $\ell$ SS SR and 3 $\ell$ SR	100
596			
597	4.10	Fake rate for muons from CR in data	101
598	4.11	Real and fake lepton efficiencies measured in $t\bar{t} + t\bar{t}\gamma$ simulated events	102
599	4.12	Closure of the $p_T$ of leptons, distance between leptons (DRll) and number of jets and $b$ -tagged jets	104
600			
601	4.13	Distribution of the BDTG non-prompt response and the BDTG $t\bar{t}V$ response and the background rejection versus signal efficiency (ROC) for the both BDTGs	108
602			
603			
604	4.14	Pre-fit distributions of the MVA input and output variables in the pre-MVA region	110
605			
606	4.15	Pre-fit distributions of the MVA input variables in the $t\bar{t}V$ control region	111
607	4.16	Distributions of the number of electrons and number of jets for the $t\bar{t}W$ cut-based validation region	113
608			
609	4.17	Distributions of the number of electrons, number of jets and invariant mass of Lep0 and Lep1 for $t\bar{t}Z$ cut-based validation region	113
610			
611	5.1	Ranking plot of the nuisance parameters from the Asimov dataset fitting	124
612	5.2	Correlation matrix of the nuisance parameters for Asimov dataset fitting	125
613	5.3	The combined BDT distribution after data fitting	126
614	5.4	Ranking plot of the nuisance parameters to the fit	128
615	5.5	Observed best fit values of the $t\bar{t}H$ signal strength $\mu_{t\bar{t}H}$	131
616	A.1	Variables distribution for signal and background	143
617	A.2	Correlation Matrix of input variables for signal and background	143
618	A.3	LikelihoodPCA response	144
619	A.4	Optimization work flow	146
620	A.5	Efficiency lost because of smoothing	148
621	A.6	Cuts value comparison of <i>Loose</i> menu with (Red)/without (Black) smoothing	148
622			
623	A.7	Cuts value comparison of <i>Medium</i> menu with (Red)/without (Black) smoothing	149
624			
625	A.8	Cuts value comparison of <i>Tight</i> menu with (Red)/without (Black) smoothing	149
626			
627	A.9	The effect of changing the TRT gas on the fraction of the high threshold TRT hits	150
628			
629	A.10	Distribution of eProbabilityHT for the signal and background	151

630	A.11 $T_{ratio}$ and eProbabilityHT comparison with different scenarios in signal events	151
631		
632	A.12 $T_{ratio}$ and eProbabilityHT comparison with different scenarios in signal events as function of pile-up	152
633		
634	A.13 $ \eta  < 2.47$ and $E_T > 5$ GeV: Efficiency vs $\mu$ for signal (left) and background (right)	152
635		
636	A.14 $ \eta  < 2$ : Efficiency vs $\mu$ for signal (left) and background (right)	153
637	A.15 $ \eta  < 1$ (left) and $ \eta  > 1$ (right): Efficiency vs $\mu$ for background	153
638	A.16 The bin-wise cuts for eProbabilityHT and the efficiency comparison	154
639	A.17 Singal (left) and background (right): Efficiency vs $\eta$ comparison	155
640	A.18 Efficiency vs pile-up of <i>Loose</i> menu	155
641	A.19 Efficiency vs pile-up of <i>Medium</i> menu	156
642	A.20 Efficiency vs pile-up of <i>Tight</i> menu	156
643	A.21 <i>Loose</i> menu performance comparison between EID 2012 menu and new re-tuned menu	157
644		
645	A.22 <i>Medium</i> menu performance comparison between EID 2012 menu and new re-tuned menu	158
646		
647	A.23 <i>Tight</i> menu performance comparison between EID 2012 menu and new re-tuned menu	158
648		
649	A.24 Efficiency versus $E_T$ for signal and background	159

# List of Tables

650

651	1.1	The four fundamental interactions known in the nature	29
652	1.2	Measured signal strengths $\mu$ and their total uncertainties for $t\bar{t}H$ production process	43
653			
654	3.1	Definitions of electron discriminating variables	66
655	3.2	Variables used for <i>Loose</i> and <i>Tight</i> photon identification cuts	71
656	3.3	<i>b</i> -tagging benchmarks of MV2 algorithm	75
657	4.1	Contributions from different Higgs decay modes for the $2\ell$ SS, $3\ell$ and $4\ell$ channel	79
658			
659	4.2	Configurations of event generations used for signal and background processes	81
660			
661	4.3	List of lowest $p_T$ -threshold, unrescaled triggers used for the whole 2015 data taking	82
662			
663	4.4	List of lowest $p_T$ -threshold, unrescaled triggers used for the whole 2016 data taking	83
664			
665	4.5	Loose and Tight cut definitions for leptons	84
666	4.6	Summary of the overlap removal procedure	85
667	4.7	The $3\ell$ pre-MVA selection criteria applied in analysis	86
668	4.8	The expected event yields for signal and background processes after the $3\ell$ pre-MVA selection	86
669			
670	4.9	Expected and observed event yields in the $3\ell$ with the $20.3 \text{ fb}^{-1}$ dataset at 8 TeV	87
671			
672	4.10	Definition of the control region used for measuring the real lepton efficiency	93
673			
674	4.11	Definition of the control region used for measuring the fake electron efficiency	95
675			
676	4.12	The impact of varying fake fraction on correction factor $\alpha$	98
677	4.13	Definition of the control region used for measuring the fake muon efficiency	101
678			
679	4.14	Expected fake event yields from MC and from the Matrix Method with $t\bar{t} + t\bar{t}\gamma$ in the $3\ell$ pre-MVA region	104
680			
681	4.15	Expected fake event yields from MC and from the Matrix Method with data in the $3\ell$ pre-MVA region	105
682			
683	4.16	List of input variables for the BDT	106
684	4.17	Description of the validation regions being designed for $t\bar{t}Z$ and $t\bar{t}W$ background	112
685			
686	4.18	Summary of experimental systematics in the analysis for the leptons	115



687	4.19 Summary of theoretical uncertainties for $t\bar{t}H$ , $t\bar{t}V$ and diboson MC predictions	116
688		
689	4.20 Summary of experimental systematics for the $b$ -tagged jets in the analysis	116
690		
691	4.21 Sources of systematic uncertainties considered in the analysis	117
692	5.1 Pre-fit expected event yields of the $3\ell$ channel	123
693	5.2 Summary on expected signal strength with all uncertainties and expected significance from the Asimov fit	125
694		
695	5.3 Post-fit event yields of the $3\ell$ channel	127
696	5.4 Summary on observed signal strength with all uncertainties and the corresponding significance with respect to the null hypothesis from data	129
697		
698	5.5 Observed and expected best-fit values of the signal strength $\mu_{t\bar{t}H}$ and associated significance with respect to the null hypothesis	130
699		
700	A.1 The input variables of each group	145
701	A.2 Target efficiency for each menu in different $p_T$ bins	146

# 702 Introduction

703 The production of the Higgs boson associated with a pair of top quarks is one  
704 of the most important Higgs boson production modes and it is still not observed  
705 so far. Therefore, to discover this production mode is one of the most charming  
706 searches after the discovery of the Higgs boson: not only for the first time we can  
707 observe the existence of this Higgs production mode but also we can measure the  
708 Yukawa coupling of the Higgs boson to the top quark. The measured results can  
709 answer the basic question of the Standard Model (SM) and can also search for  
710 any hints of the new physics by comparing to the SM prediction. This makes the  
711 study of production searching and measurement to be one of the most important  
712 analyses in Run 2 operation.

713 In the ATLAS and CMS experiments several Higgs boson decay channels were  
714 considered in Run 1 to measure the top Yukawa coupling within  $t\bar{t}H$  production:  
715  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow bb$ , and  $H \rightarrow$  multileptons final states. Collision datasets cor-  
716 responding to integrated luminosities per experiment of  $5 \text{ fb}^{-1}$  at  $\sqrt{s} = 7 \text{ TeV}$   
717 (recorded in 2011) and  $20 \text{ fb}^{-1}$  at  $\sqrt{s} = 8 \text{ TeV}$  (recorded in 2012) were employed.  
718 The ATLAS and CMS both have the individual results based on these Run 1 data.  
719 The combined significance of two experiments for the observation of the  $t\bar{t}H$  pro-  
720 cess is  $4.4\sigma$ , whereas only  $2.0\sigma$  is expected, corresponding to a measured excess of  
721  $2.3\sigma$  with respect to the SM prediction.

722 The Large Hardon Collider (LHC) has entered the  $\sqrt{s} = 13 \text{ TeV}$  energy scale  
723 since 2015 and the cross section of  $t\bar{t}H$  increases by a factor of 4 compared to 8 TeV.  
724 The two years upgrade of the ATLAS detector brings better improvements to the  
725 physics analysis, especially for the SM measurement and new physics searching.

726 The analysis of searching for the production of Higgs boson associated with a  
727 pair of top quarks in the three leptons final state is presented in this thesis. The  
728 collected data by the ATLAS detector corresponding to an integrated luminosity of  
729  $36.1 \text{ fb}^{-1}$  at a centre of mass energy of  $\sqrt{s} = 13 \text{ TeV}$  is employed. The multi-variate  
730 analysis (Boosted Decision Tree) is employed to discriminate between signal and  
731 background. Fake leptons and  $t\bar{t}V$  processes are main backgrounds, especially  
732 the fake leptons background is estimated with the data-driven method (Matrix  
733 Method). A simultaneous fit is used to estimate the final result which is mainly  
734 the signal strength ( $\mu_{t\bar{t}H}$ ) to validate the SM prediction.

735 The overview of the particle physics and theoretical framework of the Standard

736 Model are presented in Chapter 1. Besides, the measured results of  $t\bar{t}H$  production  
737 from the ATLAS and CMS with full Run 1 data are also discussed in this chapter.

738 The basic knowledge of the LHC and ATLAS experiment is introduced in Chap-  
739 ter 2. Since the importance of the physical objects which appear in the collision  
740 are reconstructed by the ATLAS detector, the definition of physical objects and  
741 the corresponding reconstruction algorithms are detailed in Chapter 3.

742 The analysis of searching for the Higgs boson associated with a pair of top quarks  
743 in the three leptons final state is presented in Chapter 4, where the detailed physical  
744 objects, the definition of the signal region, the estimation of fakes as well as the  
745 multi-variate optimization study are introduced.

746 Chapter 5 shows the results and the interpretations.

747 The Chapter Conclusion summaries the analysis results and gives the prospects  
748 of this analysis and the particle physics.

# 1 The Standard Model

750 Human beings are around with different kinds of matter, and keep thinking about  
751 the nature of those matter all the time. The idea that all matter is composed  
752 of elementary particles can date back to at least the 6th century BC. But those  
753 early ideas are simple and most of them were founded with only imaginations.  
754 Many different theories were proposed and tested over the centuries. Electron was  
755 discovered between 1879 and 1897 and became the first discovered elementary,  
756 truly fundamental particle. Since then particle physics as one of the basic part of  
757 the physics came to the public and affects us from time to time. The currently  
758 dominant and successful theory explaining these fundamental particles and fields,  
759 along with their dynamics, is called the Standard Model (SM). This model was  
760 built with many generation physicists' effort. Modern particle physics generally in-  
761 vestigates the Standard Model and its various possible extensions, from the newest  
762 "known" particle, the Higgs boson, even to the oldest known force field, gravity.  
763 The following sections are the introduction to the Standard Model.

## 1.1 The Standard Model particles

765 Since the first particle of the SM, electron, is discovered, more and more elementary  
766 particles are found during the recent 100 years. The SM is come to a complete  
767 theory of describing the basic interactions between particles and forces. The SM  
768 describes the fundamental particles which are listed in Figure 1.1.

769 The fundamental particles of the SM are fermions and bosons, with half-integer  
770 and integer (including zero) spin, respectively. Spin is an intrinsic quantum num-  
771 ber of a particle, and it has a sign (a directional quantity). Fermions make up  
772 conventional matter in the nature. Bosons are particles with integer spin and they  
773 are the mediators of the fundamental forces. The fermions are naturally orga-  
774 nized into three families/generations of increasing mass. All known stable matter  
775 is actually made up of only first generation fermions, electrons, up-quarks and  
776 down-quarks. The second and third generation fermions are unstable and decay  
777 into lighter particles.

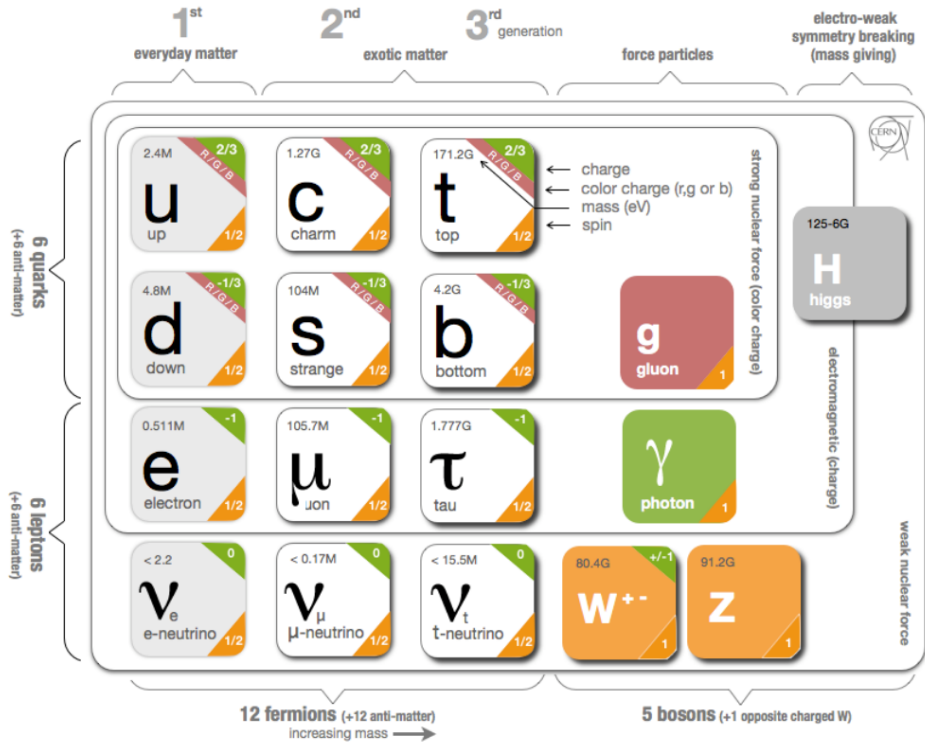


Figure 1.1: Standard Model of elementary particles.

778 Leptons are the elementary, half spin particle which do not undergo the strong  
 779 interactions. Two sets of leptons exist: charged leptons (electron, muon, tau( $\tau$ ))  
 780 and neutral leptons (neutrinos). There are six types of leptons divided by the  
 781 flavours, forming three generations. The first generation is the electronic leptons,  
 782 including the electron and electron neutrino; the second is the muonic leptons,  
 783 comprising the muon and muon neutrino; and the third is the tauonic leptons,  
 784 as the  $\tau$  and the  $\tau$  neutrino. Leptons have various intrinsic properties, including  
 785 electric charge, spin, and mass. Leptons are not subject to the strong interaction,  
 786 but they are subject to the other three fundamental interactions: gravitation, elec-  
 787 tromagnetism (excluding neutrinos, which are electrically neutral) and the weak  
 788 interaction. Leptons are an important part of the Standard Model, e.g., electrons  
 789 are one of the components of atoms, alongside protons and neutrons.

790 Quarks are spin- $\frac{1}{2}$  particles, which are fermions, and the constituent of the  
 791 proton and neutron. They carry fractional charges and there exist six different  
 792 types of quarks (up ( $u$ ), down ( $d$ ), charm ( $c$ ), strange ( $s$ ), top ( $t$ ), bottom ( $b$ ))  
 793 which are commonly referred to as flavours. Quarks have an intrinsic property  
 794 named color, that can be viewed as a charge that is conserved in the SM, just like  
 795 the electric charge. Each flavour of quark comes in three different colors: red (R),  
 796 green (G) and blue (B), tripling the actual number of quarks in the SM. Quarks

797 can experience all four fundamental interactions. Due to a phenomenon known as  
 798 color confinement, quarks are never directly observed or found in isolation; they  
 799 can be found only within hadrons, such as baryons and mesons.

800 Gauge bosons are force carriers. All known gauge bosons have a spin of 1.  
 801 Therefore, all known gauge bosons are vector bosons. Different interactions are  
 802 with different gauge bosons as the force medium. Such as photons carry the  
 803 electromagnetic interaction,  $W$  and  $Z$  bosons carry the weak interaction and gluons  
 804 carry the strong interaction. Isolated gluons do not appear because they are color  
 805 charged and subject to color confinement. Gluons are bicolored, carrying one unit  
 806 of color and one unit of a different anticolor. A total of eight gluons exists with  
 807 different colors. All the particles except photon (neutrino does have mass as the  
 808 phenomenon of neutrino oscillations) interact with Higgs field and get the mass  
 809 because of the Brout-Englert-Higgs mechanism. The Higgs boson is one of the  
 810 quantum excitations of the Higgs field and discovered by ATLAS and CMS in  
 811 2012. More details are in Section 1.3.

### 812 1.1.1 Fundamental forces

813 Forces affect our daily life, like gravitational and electromagnetic forces. They  
 814 affect the nature and life since the Universe was born. Four fundamental forces  
 815 are known so far: strong, weak, electromagnetic and the gravitational force. These  
 816 forces together with their approximate range of interaction are listed in Table 1.1.  
 817 The electromagnetic and gravitational forces have an infinite range and obey an  
 818 inverse square law. The range of the strong and weak forces can be estimated  
 819 through Heisenberg’s uncertainty principle and the masses of the force-carrying  
 820 particles, which are the gluons for the strong nuclear force and the  $W$  and  $Z$   
 821 bosons for the weak force.

Force	Relative strength	Range (m)
Strong	1	$10^{-15}$
Electromagnetic	1/137	$\infty$
Weak	$10^{-6}$	$10^{-18}$
Gravity	$10^{-39}$	$\infty$

Table 1.1: The four fundamental interactions known in the nature.

822 The strong interaction is “felt” by quarks, and is mediated through massless  
 823 particles called gluon. Quarks can bind with each other because they have color  
 824 charge. A combination of all three colors creates a colorless particle, such as a  
 825 baryon (proton for instance), which is a quark triplet state. A quark doublet is  
 826 meson, which contains a quark and anti-quark pair and is color neutral. Gluons  
 827 exist in eight independent color states. The model which describes the interactions  
 828 of colored particles through the exchange of gluons is called Quantum Chromo  
 829 Dynamics (QCD). C, P and CP-symmetry are conserved in strong interactions.

830 The Electromagnetic (EM) interaction is mediated by photons. Charged par-  
831 ticles interact through the exchange of photon. Photons are massless and with  
832 no charge. The model which describes this interact is known as Quantum Elec-  
833 trodynamics (QED). C, P and CP-symmetry are conserved in electromagnetic  
834 interactions.

835 The weak interaction is mediated through the  $W^-$ ,  $W^+$  and the  $Z$  gauge bosons.  
836 Flavored particles interact weakly through the exchange of one of these weak  
837 bosons. However, at high energy, the weak and the EM forces are indistinguishable,  
838 and a combined theory to describe both, known as Electroweak Theory (EWT) is  
839 formed. Weak interactions allow quarks to change flavor, by unit charge  $e$ . Only  
840 in the weak interactions, C, P and CP-symmetry are violated.

841 The interactions of the strong, weak and electromagnetic can be described by  
842  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  gauge symmetries. The subscripts C, L and Y rep-  
843 resent the color, weak and hypercharge symmetries respectively.  $SU(3)_C$  trans-  
844 formation describes the strong interaction, and  $SU(2)_L \otimes U(1)_Y$  transformations  
845 describe the electroweak interaction.

## 846 **1.2 Gauge symmetry**

847 Modern theories describe physical forces in terms of fields, e.g., the electromagnetic  
848 field, the gravitational field, and fields that describe forces between the elementary  
849 particles. These fundamental fields cannot be directly measured, however, some  
850 associated quantities can be measured, like charges, energies. In field theories,  
851 different configurations of the unobservable fields can result in identical observable  
852 quantities. A transformation from such a field configuration to another is called a  
853 gauge transformation; the lack of change in the measurable quantities, despite the  
854 field being transformed, is a property called gauge invariance. Since any kind of  
855 invariance under a field transformation is considered a symmetry, gauge invariance  
856 is sometimes called gauge symmetry. Generally, any theory that has the property  
857 of gauge invariance is considered a gauge theory. In the SM, symmetry plays a  
858 very important role and the SM is a gauge theory which is built using symmetry.  
859 A group of transformations which can be performed on fields are the base of the  
860 symmetry, and leaving the Lagrangian invariant. Two kinds of symmetries are  
861 presented: global symmetries which allow one to change a field in the same way  
862 all over space-time; local symmetries where the field can be changed differently in  
863 each space-time point. The meaning of the symmetry is that conservation laws  
864 are obtained naturally and conservation laws have the underlying symmetries.

### 865 **1.2.1 Quantum electrodynamics**

866 In the quantum field theory, particles can be described as a local field  $\phi(x)$ . The  
867 properties and interactions of the field are within a Lagrangian density, a function

868 of the field and its space-time derivatives:

$$\mathcal{L}(x) = \mathcal{L}(\phi, \partial_\mu \phi). \quad (1.1)$$

869 Then applying the principle of the least action, the Euler-Lagrange equation can  
870 be achieved:

$$\partial_\mu \left( \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \quad (1.2)$$

871 Considering a Dirac Lagrangian which can describe the free fermion of mass  $m$ :

$$\mathcal{L}_{Dirac} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x), \quad (1.3)$$

872 where  $\psi(x)$  is the spinor field representing the fermions,  $\bar{\psi} = \psi(x)^\dagger \gamma^0$  is its adjoint,  
873 and the  $\gamma^\mu$  are the Dirac matrices.

874 Dirac motion equation for a fermion is obtained with the Euler-Lagrange Equa-  
875 tion 1.2:

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0. \quad (1.4)$$

876 With the transformation of U(1) group the  $\mathcal{L}_{Dirac}$  can be invariant:

$$\psi(x) \rightarrow U\psi(x) = e^{i\theta}\psi(x), \quad (1.5)$$

877 where the  $\theta$  is an arbitrary real constant.

878 If the local gauge transformation,  $\theta$  is a function of  $x^\mu$ , the Lagrangian is not  
879 invariant any more. Thus, to replace the derivative  $\partial_\mu$ , the covariant derivative  
880  $D_\mu$  and gauge vector field  $A_\mu$  are introduced as:

$$D_\mu \psi(x) \rightarrow e^{i\theta(x)} D_\mu \psi(x), \quad (1.6)$$

$$D_\mu \psi(x) \equiv \partial_\mu + iqA_\mu(x). \quad (1.7)$$

881 Then the local invariance of the Lagrangian is obtained by replacing  $\partial_\mu$  by  $D_\mu$ ,  
882 then:

$$\begin{aligned} \mathcal{L} &= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) \\ &= \mathcal{L}_{Dirac} - q\bar{\psi}(x)\gamma^\mu A_\mu \psi(x). \end{aligned} \quad (1.8)$$

883 The second item represents the QED interaction between a fermion and charge  
884  $q$  and an external electromagnetic potential  $A_\mu(x)$ .  $A_\mu(x)$  can be tuned into a true  
885 propagating field with the Proca Lagrangian:

$$\mathcal{L}_{Kinematics} = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x), \quad (1.9)$$

886 where  $F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu$  is the usual electromagnetic field strength tensor.



887 This describes the free photon field and mass term is missing as this would violate  
 888 the gauge invariance. Thus the full QED Lagrangian is:

$$\mathcal{L}_{QED} = \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) - q\bar{\psi}(x)\gamma^\mu A_\mu\psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \quad (1.10)$$

889 with terms represent the kinetic energy and the mass of the fermion, the kinetic  
 890 energy of the photon, and the interaction item.

## 891 1.2.2 Quantum chromodynamics

892 In colored quark model, each quark carries one of the three color states: red, green,  
 893 blue. Quarks are triplets of the color under SU(3) and can be written as:

$$q_f = \begin{pmatrix} q_f^R \\ q_f^G \\ q_f^B \end{pmatrix} \quad \text{and} \quad \bar{q}_f = (\bar{q}_f^{\bar{R}}, \bar{q}_f^{\bar{G}}, \bar{q}_f^{\bar{B}}), \quad (1.11)$$

894 where  $f$  is the flavour. The Lagrangian of free quarks is presented as:

$$\mathcal{L} = \sum_f \bar{q}_f(i\gamma^\mu\partial_\mu - m)q_f. \quad (1.12)$$

895 In color space the Lagrangian is invariant under SU(3) transformations:

$$q(x) \rightarrow Uq(x) = e^{i\frac{\lambda_a}{2}\theta_a}q(x), \quad (1.13)$$

896 where  $U$  are  $3 \times 3$  matrices, with  $UU^\dagger = U^\dagger U = 1$  and  $\det U = 1$ .

897 The generator  $\frac{\lambda_a}{2}$  ( $a = 1, \dots, 8$ ) is the Gell-Mann matrices.  $\theta_a$  are a set of arbitrary  
 898 parameters. Local gauge invariance can be achieved by replacing  $\theta_a$  by  $\theta_a(x)$ . The  
 899 gluons  $G_\mu^a(x)$  are introduced to define the covariant derivative:

$$D_\mu\psi(x) \equiv \partial_\mu + ig_s\frac{\lambda_a}{2}G_\mu^a(x), \quad (1.14)$$

900 where  $g_s$  is the strong coupling constant. Thus Lagrangian for free quarks can  
 901 be rewritten with covariant derivative, as well as an introduced gauge-invariant  
 902 kinetic term for gluons:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc}G_\mu^b G_\nu^c \quad (1.15)$$

903 and the final Lagrangian is:

$$\mathcal{L}_{QCD} = \sum_f \bar{q}_f(i\gamma^\mu\partial_\mu - m)q_f - g_s \sum_f (\bar{q}_f\gamma^\mu\frac{\lambda_a}{2}q_f)G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.16)$$

904 The second term is the interaction between color currents of the quarks and the  
 905 gluons fields  $G_\mu^a$ . The  $G_{\mu\nu}^a G_a^{\mu\nu}$  represents the interactions between gluons. The SM  
 906 coupling constants show a dependence on energy, for EM interaction, the coupling  
 907 becomes stronger with higher energy, but the QCD coupling is opposite. Here,  
 908 it should be noted that, for photons, gauge invariance requires the gluons to be  
 909 massless.

### 910 1.2.3 Weak interaction

911 Starting from the fermions, especially one generation of up and down quarks, with  
 912 spin fields denotes by  $u(x)$  and  $d(x)$ ,

$$\psi_1(x) = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad \psi_2(x) = u_R, \quad \psi_3(x) = d_R. \quad (1.17)$$

913 For the lepton section:

$$\psi_1(x) = \begin{pmatrix} \nu_{lL} \\ l_L \end{pmatrix}, \quad \psi_2(x) = \nu_{lR}, \quad \psi_3(x) = l_R, \quad (1.18)$$

914 here  $L$  and  $R$  represent the left-handed and right-handed fermion. Then the La-  
 915 grangian is obtained as:

$$\begin{aligned} \mathcal{L} &= i\bar{u}(x)\gamma^\mu\partial_\mu u(x) + i\bar{d}(x)\gamma^\mu\partial_\mu d(x) \\ &= \sum_{j=1}^3 i\bar{\psi}_j(x)\gamma^\mu\partial_\mu\psi_j(x). \end{aligned} \quad (1.19)$$

916 Under the global gauge transformations of the  $SU(2)_L \otimes U(1)_Y$  group, the La-  
 917 grangian is invariant. To fulfil the requirement of local invariant, the covariant  
 918 derivatives is defined as:

$$D_\mu \equiv \partial_\mu + ig' \frac{Y}{2} B_\mu(x) + ig \frac{\sigma_a}{2} W_\mu^a(x), \quad (1.20)$$

919 where gauge  $B_\mu(s)$  is associated to  $U(1)_Y$ ,  $W_\mu^a(x)$  is the three gauge fields,  $g$  and  
 920  $g'$  are the coupling constant of the  $SU(2)_L$  and  $U(1)_Y$ .

921 After introducing the strength tensor:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\varepsilon_{abc} W_\mu^b W_\nu^c, \quad (1.21)$$

922 the final EW interaction is:

$$\mathcal{L}_{EW} = \sum_{j=1}^3 i\bar{\psi}(x)\gamma^\mu D_\mu\psi(x) - \frac{1}{4} B_{\mu\nu} B_{\nu\mu} - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu}. \quad (1.22)$$

923 The local gauge invariance assures the Lagrangian to describe the EM and weak  
 924 interaction in a unified way, the physical interactions between fermions and pho-

925 tons,  $W$  and  $Z$  bosons. However, the gauge invariance requires massless terms for  
 926 the gauge bosons or the fermions. Even we know that fermions and bosons have  
 927 the mass from any experiment results.

## 928 1.3 Brout-Englert-Higgs mechanism

929 Based on local gauge invariance, a symmetry Lagrangian can be built to describe  
 930 the QED and QCD. Gauge symmetry can guarantee that theory is renormalized  
 931 and perturbation theory can be used to make predictions. But it forbids the  
 932 additional mass term for the  $W$  and  $Z$  bosons. To get the mass the gauge symmetry  
 933 must be broken while keeping the Lagrangian symmetric. This could be done  
 934 through the spontaneous symmetry breaking (SSB).

935 Considering a complex scalar field  $\phi(x)$ :

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - V(\phi), \quad (1.23)$$

936 where

$$V(\phi) = \mu^2 \phi^* \phi + \lambda (\phi^* \phi)^2. \quad (1.24)$$

937 Here  $\lambda < 0$  is to ensure the potential to have the bounded ground state. The  
 938 case of  $\mu^2 > 0$  and minimum  $V(\phi)$  at ground state represents the spin zero particle,  
 939 mass of  $\mu$  and quartic coupling  $\lambda$ . If  $\mu^2 < 0$ , the potential has a minimum value  
 940 at ground state:

$$|\phi_0| = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{\nu}{\sqrt{2}}, \quad (1.25)$$

941 the quantity  $\nu$  ( $\nu^2 = -\mu^2/\lambda$ ) is called vacuum expectation value. Because of the  
 942  $U(1)$  phase invariance, the  $\phi_0$  is written as:

$$\phi_0 = \frac{\nu}{\sqrt{2}} e^{i\theta}. \quad (1.26)$$

943 The ground state solution,  $\theta = 0$ , causes the symmetry to be spontaneously  
 944 broken. Now try to let  $W$  and  $Z$  bosons have the mass meanwhile photon is  
 945 massless. Considering an  $SU(2)$  doublet of complex fields:

$$\Phi(x) = \begin{pmatrix} \phi^\dagger(x) \\ \phi^0(x) \end{pmatrix}, \quad (1.27)$$

946 the Lagrangian of Higgs fields can be written as:

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger D^\mu \Phi - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (1.28)$$

947 where  $D_\mu$  is the covariant derivative which is defined in Equation 1.20 to describe  
 948 the field and it is invariant under the  $SU(2)_L \otimes U(1)_Y$  group transformations.  $\Phi$

949 can be parametrized as a perturbation of the ground state.  $H(x)$  and  $\theta_a(x)$  ( $a =$   
950  $1, 2, 3$ ) and local  $SU(2)$  invariance are introduced to allow the elimination of the  
951 dependence on  $\theta_a(x)$ . The field  $\theta_a(x)$  is the massless Nambu-Goldstone bosons  
952 associated with the SSB. By using the unitary gauge, these fields will be eaten by  
953 the gauge bosons and get the mass. So the Lagrangian will not have the terms  
954 related to  $\theta_a(x)$ . Rewriting the covariant derivative as:

$$D_\mu = \begin{pmatrix} \partial_\mu - \frac{i}{2}(gW_\mu^3 + g'B_\mu) & -\frac{ig}{2}(W_\mu^1 - iW_\mu^2) \\ -\frac{ig}{2}(W_\mu^1 + iW_\mu^2) & \partial_\mu + \frac{i}{2}(gW_\mu^3 - g'B_\mu) \end{pmatrix}, \quad (1.29)$$

955 then get the kinetic part:

$$(D_\mu\Phi)^\dagger D^\mu\Phi = \frac{1}{2}\partial_\mu H\partial^\mu H + \frac{1}{8}g^2(\nu + H)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(\nu + H)^2|gW_\mu^3 - g'B_\mu|^2, \quad (1.30)$$

956 the final Higgs Lagrangian would be:

$$\mathcal{L}_{Higgs} = \frac{1}{2}\partial_\mu H\partial^\mu H + (\nu + H)^2\left(\frac{g^2}{4}W_\mu^\dagger W^\mu + \frac{g^2}{8\cos^2\theta_W}Z_\mu Z^\mu\right) - \lambda\nu^2 H^2 - \lambda\nu H^3 - \frac{\lambda}{4}H^4, \quad (1.31)$$

957 where  $W_\mu = \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}}$ ,  $Z_\mu = \frac{g'W_\mu^3 - gB_\mu}{\sqrt{g^2 + g'^2}}$  and  $\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}$ .

958 Taking a look at the quadratic term of Equation 1.30 which is related to gauge  
959 bosons, the  $W$  and  $Z$  bosons mass is derived:

$$m_W = \frac{1}{2}\nu g, \quad m_Z = \frac{1}{2}\nu\sqrt{g^2 + g'^2} \quad (1.32)$$

960 and the relation between them is:  $m_W = m_Z \cos\theta_W$ . The Higgs boson (H) has  
961 been added as the generation of masses for the gauge bosons. And the SM Higgs  
962 is spin-0 scalar boson with no electric charge, and the mass is:

$$m_H = \sqrt{-2\mu^2} = \nu\sqrt{2\lambda}. \quad (1.33)$$

963 Although the SM gives the mass of the Higgs, the quartic coupling  $\lambda$  is a free  
964 parameter. The mass of the SM Higgs is still unknown until the measurements  
965 provided by the experiments. Through the spontaneous breaking of EW symmetry,  
966 the gauge bosons get the mass, however, fermions are still massless. Fermions can  
967 obtain the mass by adding interaction term between fermions and the scalar field,  
968 namely Higgs field. This term is called the Yukawa term and is given by:

$$\mathcal{L}_{Yukawa} = -(\Gamma_{i,j}^l \bar{\Psi}_L^i \Phi \psi_R^j + \Gamma_{i,j}^D \bar{\Psi}_L^i \Phi \psi_R^j + \Gamma_{i,j}^U \bar{\Psi}_L^i \Phi^\dagger \psi_R^j) + h.c., \quad (1.34)$$

969 the field doublets and singlets are split into their lepton and quark components.

970 The lepton doublet and singlet are represented by  $\Psi_L^l$  and  $\Psi_R^l$ , the left-handed  
971 up- and down-type quark doublets are represented by  $\Psi_L^Q$ , and the right-handed

972 up- and down-type quark singlets are represented by  $\phi_R^U$  and  $\phi_R^D$  respectively. The  
 973  $\Gamma_{i,j}$  terms are Yukawa coupling matrices for leptons and up and down quarks.  
 974 These matrices contain terms of the Yukawa couplings  $y_{i,j}$  for each fermion flavour  
 975  $i, j$ . The Yukawa couplings are not predicted by theory and must be obtained by  
 976 experiment. After symmetry breaking in the unitary gauge the Yukawa term takes  
 977 the form:

$$\begin{aligned}
 \mathcal{L}'_{Yukawa} &= -\Gamma_{i,j}^l \bar{\Psi}'_L \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix} \psi_R^l - \Gamma_{i,j}^D \bar{\Psi}'_L \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix} \psi_R^U \\
 &\quad - \Gamma_{i,j}^U \bar{\Psi}'_L \frac{1}{\sqrt{2}} \begin{pmatrix} \nu + H \\ 0 \end{pmatrix} \psi_R^D \\
 &= \frac{\nu}{\sqrt{2}} \Gamma_{i,j}^l \bar{\Psi}'_L \psi_R^l - \frac{1}{\sqrt{2}} \Gamma_{i,j}^l \bar{\Psi}'_L \psi_R^l H - \frac{\nu}{\sqrt{2}} \Gamma_{i,j}^D \bar{\Psi}'_L \psi_R^D - \frac{1}{\sqrt{2}} \Gamma_{i,j}^D \bar{\Psi}'_L \psi_R^D H \\
 &\quad - \frac{\nu}{\sqrt{2}} \Gamma_{i,j}^U \bar{\Psi}'_L \psi_R^U - \frac{1}{\sqrt{2}} \Gamma_{i,j}^U \bar{\Psi}'_L \psi_R^U H + h.c.
 \end{aligned} \tag{1.35}$$

978 The fermion mass terms can be interpreted as  $m = \frac{y_{ij}\nu}{\sqrt{2}}$ . The primed fields  
 979 represent the fermion fields in the unitary gauge. The terms containing H represent  
 980 the fermion couplings to the Higgs boson.

981 Now the final SM Lagrangian can be written as the sum of the Lagrangian  
 982 of QCD (Equation 1.16), EW (Equation 1.22), Higgs (Equation 1.31), Yukawa  
 983 (Equation 1.35):

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \tag{1.36}$$

## 984 1.4 Current status and remaining issues

985 The Large Hadron Collider (described in Section 2.1) is a proton-proton collider  
 986 which provides several production modes of the Higgs boson. The dominant pro-  
 987 duction mechanisms for the SM Higgs boson are through gluon-gluon fusion (ggF),  
 988 vector boson fusion (VBF), associated production involving a  $W$  or  $Z$  boson (VH),  
 989 and associated production with a  $t\bar{t}$  pair ( $t\bar{t}H$ ). Figure 1.2 shows a set of Feynman  
 990 diagrams for these processes. The ggF and VBF production make the two greatest  
 991 contributions to the total production cross section, and account for around 87%  
 992 and 7% of the total for a Higgs boson of mass 125 GeV.

993 The Higgs boson is an unstable particle with a predicted mean lifetime of about  
 994  $1.6 \times 10^{-22}$  s for a mass of 125 GeV, and decays primarily into pairs of fermions or  
 995 gauge bosons. The SM Higgs boson decay branching ratios are shown in Figure 1.3.

996 As shown in Figure 1.3  $H \rightarrow b\bar{b}$  decay has the largest branching fraction, about  
 997 57% for a Higgs boson of mass 125 GeV, but the ggF process is dominated by a  
 998 background from the SM di-jet production that contaminates the signal and makes

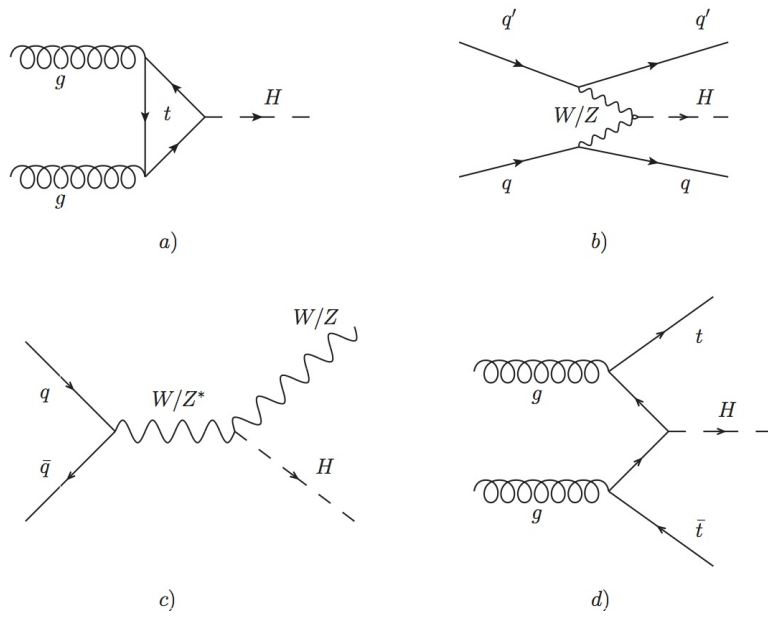


Figure 1.2: Summary of Higgs boson production mechanisms at the LHC (a) ggF, (b) VBF, (c) VH, (d)  $t\bar{t}H$ .

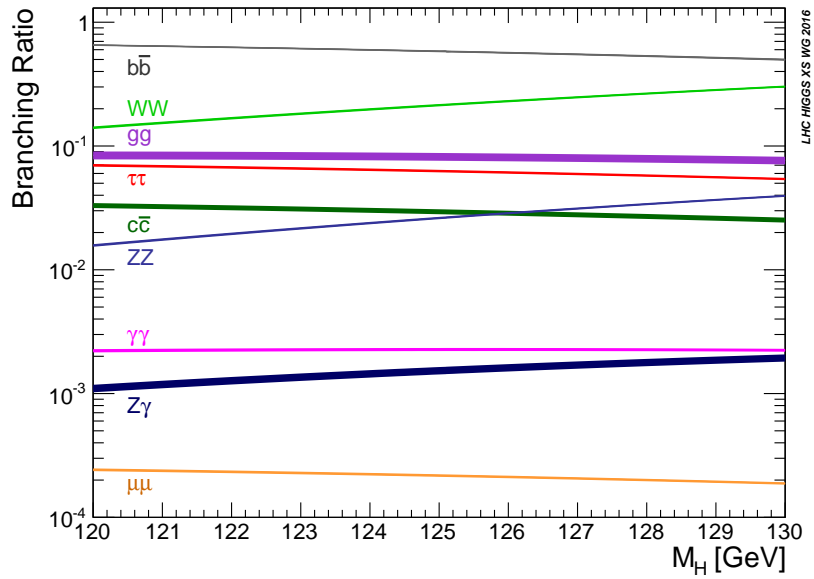


Figure 1.3: The SM Higgs boson decay branching ratios.

999 analysis challenging. Preferred channels for study are those with good signal to  
1000 background ratios,  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ^*$  and  $H \rightarrow WW^*$  channels are chosen to  
1001 search for and study Higgs currently although some of them is with small branching  
1002 ratio. The  $H \rightarrow WW^*$  channel is the main studied decay mode especially in this  
1003 thesis.

1004 On the 4th of July 2012, using only a fraction of the full dataset provided by  
1005 the LHC, the ATLAS and CMS collaborations at CERN announced the discovery  
1006 of a new particle, when searching for the SM Higgs boson. The excess of events  
1007 observed was compatible with the production and decay of the SM Higgs boson,  
1008 with a mass of approximately 125 GeV [1, 2].

1009 Combining the ATLAS and CMS data for the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4\ell$   
1010 channels, the mass of the Higgs boson is determined to be [3]

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.}) \text{ GeV.} \quad (1.37)$$

1011 Measurements of the Higgs boson cross section in the Higgs boson diphoton  
1012 decay channel are performed using  $pp$  collision data recorded by the ATLAS ex-  
1013 periment taken at a centre of mass energy of  $\sqrt{s} = 13$  TeV and correspond to an in-  
1014 tegrated luminosity of  $13.3 \text{ fb}^{-1}$ . The corresponding signal strengths measured for  
1015 the different production processes, and globally (i.e. assuming one common signal  
1016 strength parameter for all production processes), are summarised in Figure 1.4,  
1017 which also shows the global signal strength measured in Run 1. The presented  
1018 measurements are dominated by the statistical uncertainties. The measurements  
1019 agree with the SM expectations within 1 to  $2\sigma$ , and no significant deviation from  
1020 the Standard Model expectations is observed [4].

1021 The latest measurements of the Higgs boson show remarkable compatibility  
1022 with the SM expectations, and neither of the collaborations has observed any  
1023 significant deviations from this theory. The SM is the most successful theory,  
1024 which can describe the relationship between particles and fundamental interactions  
1025 so far. The latest results with ATLAS experiment are shown in Figure 1.5 [5]. The  
1026 agreement spans several orders of magnitude in cross-section and a variety of the  
1027 SM benchmark processes.

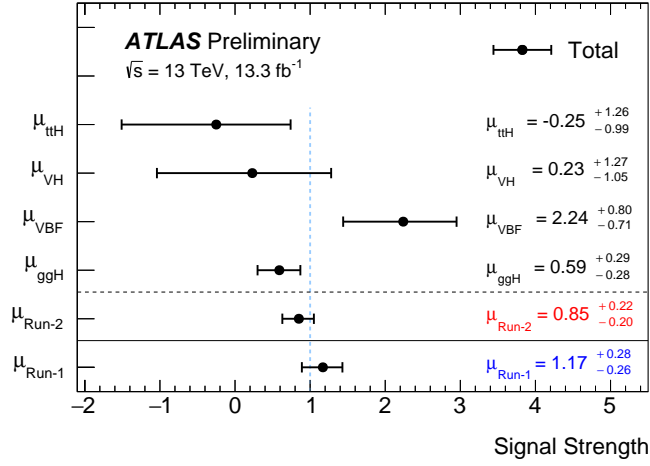


Figure 1.4: The signal strength measured for the different production processes ( $ggH$ ,  $VBF$ ,  $VH$  and  $ttH$ ) and globally ( $\mu_{Run2}$ ), compared to the global signal strength measured at 7 and 8 TeV ( $\mu_{Run1}$ ). The error bar shows the total uncertainty. The gluon-gluon fusion production cross section is larger by approximately 10%.

1028 Despite the tremendous success of the SM, there are still many unsolved issues  
 1029 in physics, from dark matter, dark energy, hierarchy problem, matter-antimatter  
 1030 asymmetry to the neutrino oscillation. The absence of gravity and implicitly of its  
 1031 mediator in the SM, and there is no experimentally confirmed understanding of  
 1032 this force at the quantum level yet. Much effort has been done to combine gravity  
 1033 with the SM to form a theory of everything, but so far attempts have only limited  
 1034 progresses. The Beyond Standard Model (BSM) is needed to deal with those  
 1035 issues, like SUPERSymmetry (SUSY). Generally the BSM physics is expected at  
 1036 the TeV scale. For example, SUSY introduces a new symmetry relating fermions  
 1037 and bosons at the TeV scale, is presented as an elegant solution to the hierarchy  
 1038 problem, as it cancels out the quadratic divergences. Moreover, SUSY has the  
 1039 advantage of introducing a candidate for dark matter. The all 7, 8 and 13 TeV  
 1040 data recorded by ATLAS are employed to search for the SUSY particles. No  
 1041 obvious evidence of the superpartner (also sparticle) are found so far.



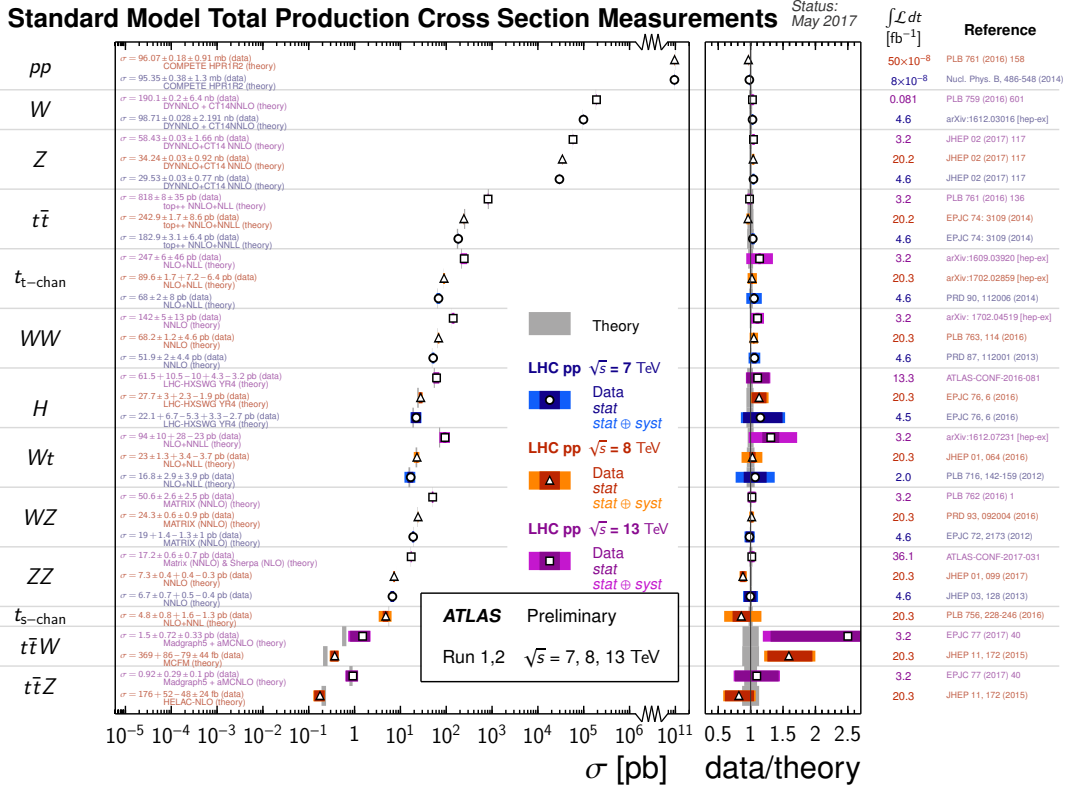


Figure 1.5: Summary of several Standard Model total production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations. All theoretical expectations were calculated at NLO or higher. The dark-color error bar represents the statistical uncertainty. The lighter-color error bar represents the full uncertainty, including systematics and luminosity uncertainties. The data/theory ratio, luminosity used and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers. They were not always evaluated using the same prescriptions for PDFs and scales. Not all measurements are statistically significant yet.

## 1.5 The production of the Higgs boson associated with a pair of top quarks ( $t\bar{t}H$ )

After the discovery of the Higgs, one of the most important studies is to measure Higgs couplings to other particles to figure out whether the couplings are consistent with the Standard Model Higgs boson, or those of a Higgs boson from an extended Higgs sector in Run 2. This section provides the basic information and meaning of the  $t\bar{t}H$  production. The following Section 1.6 is with combined results of the ATLAS and CMS by using the 7 and 8 TeV data which are collected during the Run 1. The coupling of the Higgs boson to other particles can be measured by studying the various production processes and decay modes of the Higgs boson. The associated production of a Higgs boson with a pair of top quarks ( $t\bar{t}H$ , Figure 4.1) can provide direct information of the Yukawa coupling of the Higgs boson to the top quark, probably the most crucial coupling to fermions. This process involves two colliding gluons. Those gluons through the top-Higgs coupling form a Higgs particle and a top quark pair. Top quark Yukawa coupling measurement connects two of the heaviest Standard Model particles, the top quark (173 GeV) and the Higgs boson (125 GeV) :  $\lambda_t = \frac{\sqrt{2}m_t}{\nu}$ . Its value is predicted to be equal to unity and any deviation might give a hint on the possible existence of the new physics. The absence of a deviation can constrain and even exclude the BSM theories.

The SM Higgs cross section of different production processes versus energy are shown in Figure 1.6, especially the  $t\bar{t}H$  process increases the most significantly compared to other productions around 10 TeV, which is the designed energy range for the Run 2 and the further LHC running. The Yukawa coupling of the Higgs boson to the top quark is a key parameter of the SM. It can be determined from the ratio of the top quark mass and Higgs field vacuum expectation value, from the cross section of  $gg \rightarrow H$  production through a top quark loop, or from the cross section of the process  $gg \rightarrow t\bar{t}H$ , which is a tree-level process at lowest order in perturbation theory. Comparison of these measurements has the potential to identify and disambiguate new physics effects that can modify the  $t\bar{t}H$  production cross section relative to the SM expectation.

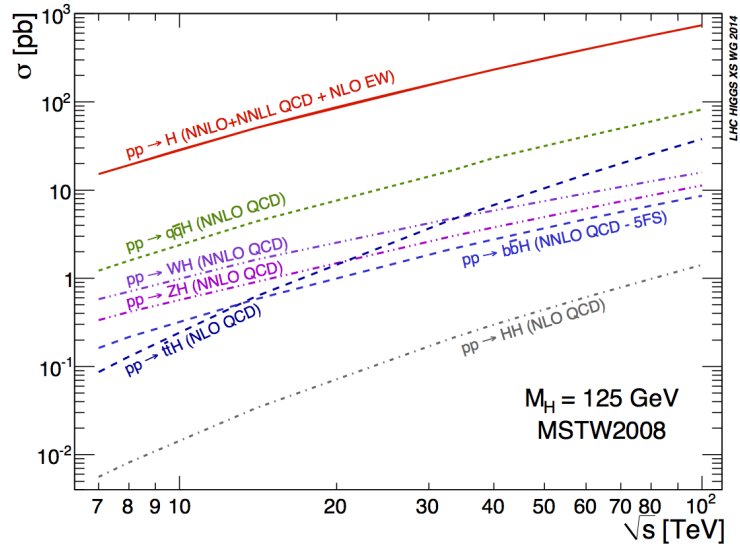


Figure 1.6: SM Higgs boson cross sections of different productions versus the centre mass of energy.

## 1072 **1.6 $t\bar{t}H$ searching at the LHC in Run 1**

1073 At the ATLAS and CMS several Higgs boson decay channels were considered in  
 1074 Run 1 to measure the top Yukawa coupling using  $t\bar{t}H$  production:  $H \rightarrow \gamma\gamma$ ,  
 1075  $H \rightarrow b\bar{b}$  and  $H \rightarrow$  multileptons final states. The combined results shown below  
 1076 are based on the complete Run 1 collision data collected by the ATLAS and CMS  
 1077 experiments. These data correspond to integrated luminosities per experiment of  
 1078 approximately  $5 \text{ fb}^{-1}$  at  $\sqrt{s} = 7 \text{ TeV}$  (recorded in 2011) and  $20 \text{ fb}^{-1}$  at  $\sqrt{s} = 8 \text{ TeV}$   
 1079 (recorded in 2012). The results of the ATLAS and CMS individual combinations  
 1080 based on the Run 1 data are reported in Refs [6, 7]. Especially the signal strength  
 1081 of  $t\bar{t}H$  production in multilepton final state by using  $20.3 \text{ fb}^{-1}$  of proton-proton  
 1082 collision data recorded by the ATLAS experiment at  $\sqrt{s} = 8 \text{ TeV}$  are shown in  
 1083 Figure 1.7, and the  $p$ -value of the likelihood ratio test of the background-only  
 1084 hypothesis corresponds to  $1.8\sigma$ ; the expectation in the presence of a Standard  
 1085 Model signal is  $0.9\sigma$  [6].

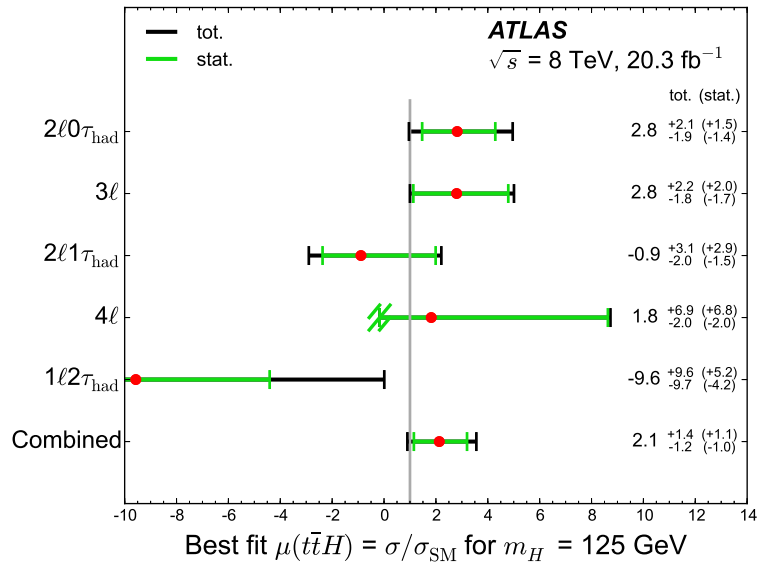


Figure 1.7: Best-fit values of the signal strength parameter  $\mu = \sigma_{t\bar{t}H,\text{obs}}/\sigma_{t\bar{t}H,\text{SM}}$ . For the  $4\ell$   $Z$ -depleted category,  $\mu < -0.17$  results in a negative expected total yield and so the lower uncertainty is truncated at this point.

1086 The global signal strength is the most precisely measured Higgs boson coupling-  
1087 related observable, but this simple parametrisation is very model dependent. All  
1088 Higgs boson production and decay measurements are combined assuming that all  
1089 their ratios are the same as in the SM. The compatibility of the measurements  
1090 with the SM can be tested in a less model-dependent way by relaxing these as-  
1091 sumptions separately for the production cross sections and the decay branching  
1092 fractions. Assuming the SM values for the Higgs boson branching fractions, the  
1093 combined results of the ATLAS and CMS data of  $t\bar{t}H$  signal strengths are listed  
1094 in Table 1.2 [8].

Production process	ATLAS + CMS	ATLAS	CMS
$\mu_{t\bar{t}H}$	$2.3^{+0.7}_{-0.6}$	$1.9^{+0.8}_{-0.7}$	$2.9^{+1.0}_{-0.9}$

Table 1.2: Measured signal strengths  $\mu$  and their total uncertainties for  $t\bar{t}H$  production process. The results are shown for the combination of the ATLAS and CMS, and separately for each experiment, for the combined  $\sqrt{s} = 7$  and 8 TeV data. These results are obtained assuming that the Higgs boson branching fractions are the same as in the SM.

1095 The combined likelihood is used to evaluate the significance for the observation  
1096 of the  $t\bar{t}H$ . The combination of the data from the two experiments corresponds  
1097 to summing their recorded integrated luminosities and consequently increases the

1098 sensitivity by approximately a factor of  $\sqrt{2}$ , since the theoretical uncertainties in  
1099 the Higgs boson signal are only weakly relevant for this evaluation and all the  
1100 other significant uncertainties are uncorrelated between the two experiments. The  
1101 combined significances for the observation of the  $t\bar{t}H$  process is  $4.4\sigma$ , whereas only  
1102  $2.0\sigma$  is expected, corresponding to a measured excess of  $2.3\sigma$  with respect to the  
1103 SM prediction [8].

1104

## 1105 2 Experimental apparatus

1106 The Large Hadron Collider (LHC) is the largest and highest energy hadron collider  
1107 in the world so far. It is the most recent addition to the CERN's accelerator and  
1108 lies in a tunnel buried around 50 to 175 m beneath the France–Switzerland border  
1109 near Geneva, Switzerland. According to the particle physics, when measuring the  
1110 properties of nature at the smallest distance scales ever recorded with a terrestrial  
1111 apparatus the highest energy particles accelerator ever built is required. The size of  
1112 a structure is related to the probe energy via the de Broglie relation  $\lambda = 1/\rho$ . The  
1113 same principle governs optical microscopes, limiting their resolution to hundreds  
1114 of nano meters. However, the LHC can produce protons with energies up to  
1115  $\sqrt{s} = 13$  TeV and A Toroidal LHC ApparatuS (ATLAS) is able to capture the  
1116 products of the proton-proton collisions to probe distance scales as small as  $10^{-20}$   
1117 m scale. Besides the ATLAS, CMS experiment has the similar physics goal as the  
1118 ATLAS, while the other two experiments, the ALICE and LHCb are designed to  
1119 have different physics goals and locate at the different parts of the LHC. In this  
1120 chapter, the design and performance of the LHC are presented in Section 2.1 and  
1121 the ATLAS detector is described in details in Section 2.2.

### 1122 2.1 The Large Hadron Collider

1123 The LHC inherited the tunnel of the Large Electron–Positron Collider (LEP) which  
1124 took place from 1984 to 1989. The ring has a circumference of 26.7 km and a gradi-  
1125 ent of 1.4%. It was decided to build the LHC in this existing tunnel to considerably  
1126 lower the construction costs, even though a hadron collider generally benefits from  
1127 a larger radius and does not suffer as much from synchrotron radiation as a cir-  
1128 cular lepton collider. The LHC is designed to produce proton-proton collisions  
1129 up to a center of mass energy of  $\sqrt{s} = 14$  TeV. To accelerate the particles to  
1130 desired energy, the beam is injected through a succession of machines which accel-  
1131 erate the particles to gradually higher energies. This accelerator chain is shown in  
1132 Figure 2.1 [9].

1133 The protons, produced by a duoplasmatron source through stripping hydrogen  
1134 atoms of their orbiting electrons, are injected in the linear accelerator in the chain,  
1135 the Linac 2, where they get accelerated to an energy of 50 MeV. The beam is then  
1136 firstly fed into the Proton Synchrotron (PS) Booster, which has a radius of 25 m

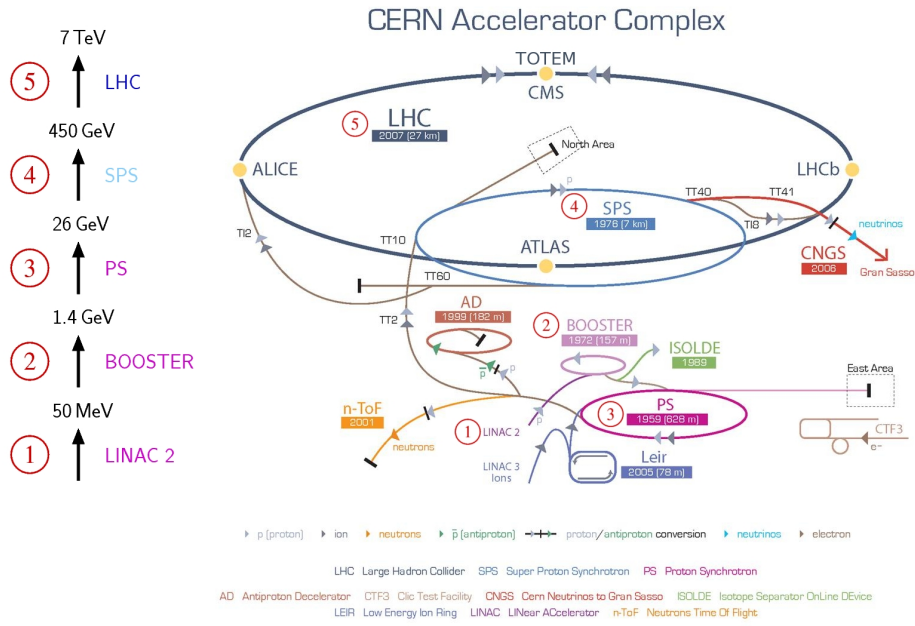


Figure 2.1: The accelerator complex at CERN.

1137 and accelerates the particles to 1.4 GeV, secondly into the PS where an energy of  
 1138 26 GeV is obtained, and finally into the Super PS (SPS), which has a circumference  
 1139 of 6.9 km, where the particles are accelerated to 450 GeV. The particles have now  
 1140 reached the LHC injection energy and are fed into the LHC ring via two transfer  
 1141 lines that circulate in opposite directions. It takes approximately 4 minutes to  
 1142 fill each of the two LHC rings. The particles are then accelerated for 20 minutes  
 1143 by electric fields in superconducting radio-frequency (RF) cavities operating at  
 1144 400 MHz with a 5 MV/m gradient to reach the design energy of 14 TeV. Once the  
 1145 particles are circulating around the ring at the nominal beam energy, they will lose  
 1146 about 7 KeV of synchrotron radiation per turn. The RF cavities provide energy  
 1147 corrections to the beams to account for this loss. The accelerator chain can also  
 1148 accelerate lead ions which are passed through a Low Energy Ion Ring and are from  
 1149 there transferred through the PS and the SPS. In the LHC ring, they get to be  
 1150 accelerated to half of the maximally collided energy per nucleon. The LHC has  
 1151 not reached its design energy yet, it has been running at 7, 8 and 13 TeV, in 2011,  
 1152 2012 and 2015-2016, respectively.

1153 The instantaneous luminosity and the center of mass energy are important pa-  
 1154 rameters for the LHC. The number of events per second,  $N_{event}$ , for a particular  
 1155 physics process is given by:

$$N_{event} = L \times \sigma_{event}, \quad (2.1)$$

1156 where  $\sigma_{event}$  is the cross section of the physics process and  $L$  is instantaneous  
 1157 luminosity. The instantaneous luminosity is defined as:

$$\mathcal{L}_{inst} = \frac{\mu n_b f_r}{\sigma_{inel}}, \quad (2.2)$$

1158 where  $\mu$  is the average number of inelastic interactions per bunch crossing,  $n_b$  is  
 1159 the number of colliding bunch pairs,  $f_r$  is the machine revolution frequency, and  
 1160  $\sigma_{inel}$  is the  $pp$  inelastic cross section. The instantaneous luminosity depends on the  
 1161 collider parameters, namely the density of protons in the colliding bunches, the  
 1162 collision transverse area (defined by the accelerator optics in the collision zone)  
 1163 as well as on the frequency of collisions and the number of bunches. The design  
 1164 luminosities for  $pp$  collisions at the four main LHC experiments are  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$   
 1165 for the ATLAS and CMS,  $10^{32} \text{ cm}^{-2}\text{s}^{-1}$  for the LHCb and  $10^{30} \text{ cm}^{-2}\text{s}^{-1}$  for the  
 1166 ALICE.

1167 An increasing challenge for the LHC data is in-time pile-up and out-of-time pile-  
 1168 up with higher center of mass energy. In-time pile-up refers to multiple protons  
 1169 from each bunch interacting in a given bunch crossing whereas out-of-time pile-  
 1170 up refers to the presence of collisions from surrounding bunch crossings in the  
 1171 read-out window of the considered bunch crossing. Pile-up increases the track  
 1172 and vertex multiplicity as well as the overall energy in an event. This renders the  
 1173 reconstruction of physics objects like tracks, vertices and jets very challenging. To  
 1174 avoid large systematic uncertainties, a precise pile-up modelling is therefore crucial  
 1175 for successful physics analyses. Generally pile-up events are modelled by using the  
 1176 minimum bias events which are based on Monte Carlo simulation.

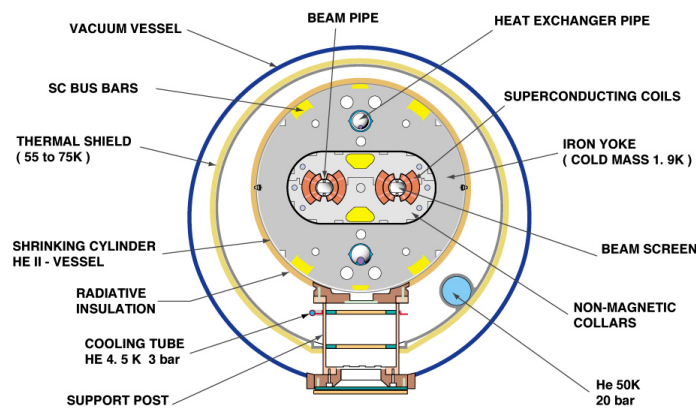
1177 Holding together the protons in a circular beam pipe, a powerful magnetic field  
 1178 is needed. To reach a high intensities, superconducting magnets are used. The  
 1179 magnet coils are based on niobiumtitan superconductor cables, which are able to  
 1180 conduct the electricity without resistance or energy loss. With the superconductiv-  
 1181 ity facilities, the LHC magnetic system is able to be operated at required intensities  
 1182 of up to 8.4 Tesla, given an electric current of 11850 A. Highest constraints on the  
 1183 cryostat system design are from the LEP tunnel and its facilities to construct the  
 1184 LHC collider. A total of eight refrigeration plants, filled with liquid Helium, are  
 1185 supplying the required cooling superfluid.

1186 The proton beam bending is ensured by 1232 dipole magnets, each piece weight-  
 1187 ing 27.5 tons, a length of 16.5 m and a diameter of 570 mm. Each magnet has  
 1188 two apertures, for the two proton beams tubes (two-in-one magnet design, Fig-  
 1189 ure 2.2 [10]) and provides a magnetic field of 8.33 T. The required temperature is  
 1190 1.9 K most of time, but it can also go up to 4.5 K. As the protons are charged  
 1191 particles, it is important to focus the beam and retain it in the ultra high vac-  
 1192 uum chamber. Other thousands of multipoles magnets are used to improve the  
 1193 beam focusing and also to reduce the unwanted interactions between the bunch  
 1194 constituents. Finally, inner triplet magnets are used to squeeze the protons in the  
 1195 collision points.



1196 Besides the mentioned magnets, many more normal magnets are used in order to  
 1197 reach the target beam parameters, leading to more than 9000 magnets surround-  
 1198 ing the LHC ring. They represent around 90% (40000 tons) of the cold mass.  
 1199 The magnet powering is performed in eight independent and symmetric sectors  
 1200 to ensure a limitation of the stored energy in a magnet. Besides the powerful  
 1201 magnetic system, the LHC is characterized by the biggest vacuum system in the  
 1202 world. The vacuum system can reach a vacuum pressure of  $10^{-6}$  mbar at cryogenic  
 1203 temperatures (5-20 K).

#### CROSS SECTION OF LHC DIPOLE



CERN AC\_HE107A\_V02/02/98

Figure 2.2: Cross section of the LHC dipole.

1204 The first proton-proton collision at LHC occurred in November 2009, at energy  
 1205 of 900 GeV in the centre of mass system and increased to 2.37 TeV within few  
 1206 days. The first 7 TeV collisions started at 2010 with validation tests after test pe-  
 1207 riod. Delivered luminosity increased rapidly in following years in Figure 2.3 (left),  
 1208 especially in the year 2012 and year 2016. The situation in 2016 is much different  
 1209 because of the two years update from 2013 to 2014 (Section 2.2.7). The last Run  
 1210 1 data taking started in May 2012, with an increase in energy in the center mass  
 1211 to 8 TeV. In Figure 2.3 (right) a total integrated luminosity and data quality are  
 1212 presented. With the good performance of the detector, data recorded efficiently by  
 1213 the ATLAS. Comparing to 7 TeV, the simultaneous interactions increased consid-  
 1214 erably (Figure 2.4) in 8 TeV. Many Standard Model measurements are published  
 1215 with better precise results by using the 7 TeV and 8 TeV data recorded in 2011  
 1216 and 2012. Especially in 2012, Higgs boson is announced to be discovered by the  
 1217 ATLAS and CMS experiments separately.

1218 In the LHC, four main experiments were installed and developed. The ATLAS  
 1219 and CMS are the biggest experiments and are designed to understand the Standard

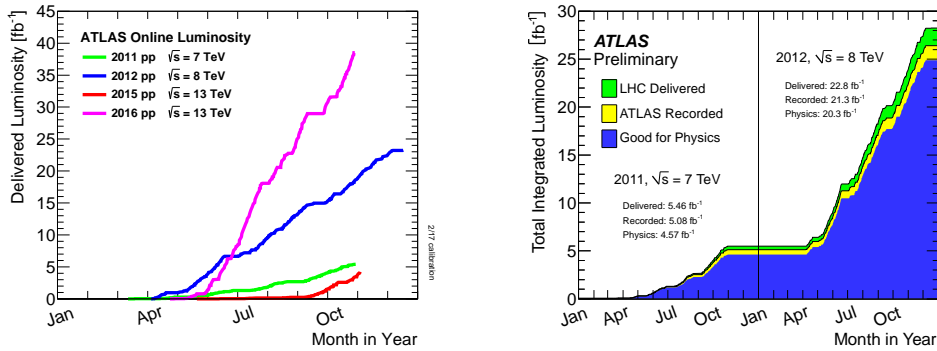


Figure 2.3: Delivered Luminosity versus time for 2011-2016 ( $pp$  data only) (left) and total integrated luminosity and data quality in 2011 and 2012 [11].

1220 Model precisely, especially for the Higgs physics and also to discover the physics  
 1221 implying unknown particles (like supersymmetric particles predicted by SUSY)  
 1222 around the TeV scale. The b physics is primarily studied at the LHCb experiment,  
 1223 which covers the main subjects of heavy flavour and electroweak physics. It allows  
 1224 the study of  $B_s \rightarrow \mu^+ \mu^-$  the rare SM processes (i.e. flavour changing neutral  
 1225 current) and the CP violation in different processes involving  $B_s$ ,  $D_0$  and Kaon  
 1226 meson states. It also led to the discovery of new baryon resonance,  $\Xi_b^-$ . A Large  
 1227 Ion Collider Experiment (ALICE) is studying the quark confining mechanism, by  
 1228 creating a quark-gluon plasma. The considered collisions are lead-lead or lead-  
 1229 proton.

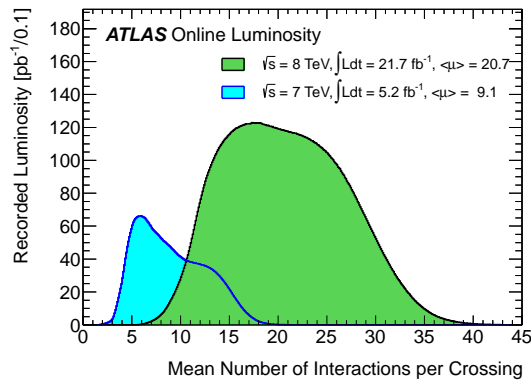


Figure 2.4: Number of interactions per crossing in 2011 and 2012 [11].

## 2.2 The ATLAS detector

1230

1231 The ATLAS is a general purpose experiment that is designed to probe a large range  
1232 of physics from soft QCD, precision electroweak and top quark measurements to  
1233 searches for the Higgs boson and new physics at TeV scale. To this point, the AT-  
1234 LAS is built in a layered structure with cylindrical geometry and almost complete  
1235 hermetic coverage. A schematic of the detector can be seen in Figure 2.5 [12].

1236 The ATLAS detector is 44 m long, 25 m tall and weighs about 7000 tons.  
1237 Starting from its core and moving outwards, first is the inner detector (ID) that is  
1238 structured in three layers and is responsible for precisely measuring the momentum  
1239 of charged particles moving in a 2 T solenoidal magnetic field, followed by the  
1240 electromagnetic and hadronic calorimeters which measure the energy deposition  
1241 of electrons, photons and hadrons and also infer the missing transverse energy from  
1242 neutrinos and other undetected particles, and finally a precise muon spectrometer  
1243 (MS) which provides tracking and triggering of muons in a toroidal magnetic field.  
1244 The detector took fifteen years to design, build and install.

1245 The ATLAS collaboration counts about 3000 people from 178 institutions in 38  
1246 countries. The ATLAS recorded its first collisions in 2009, after a one year shut  
1247 down due to an LHC magnet quenching accident, and has been running successfully  
1248 until the beginning of 2013 when it shut down for a two years upgrade period. The  
1249 ATLAS begun taking 13 TeV data in 2015 and will continue full running till the  
1250 end of 2018 for the next upgrade.

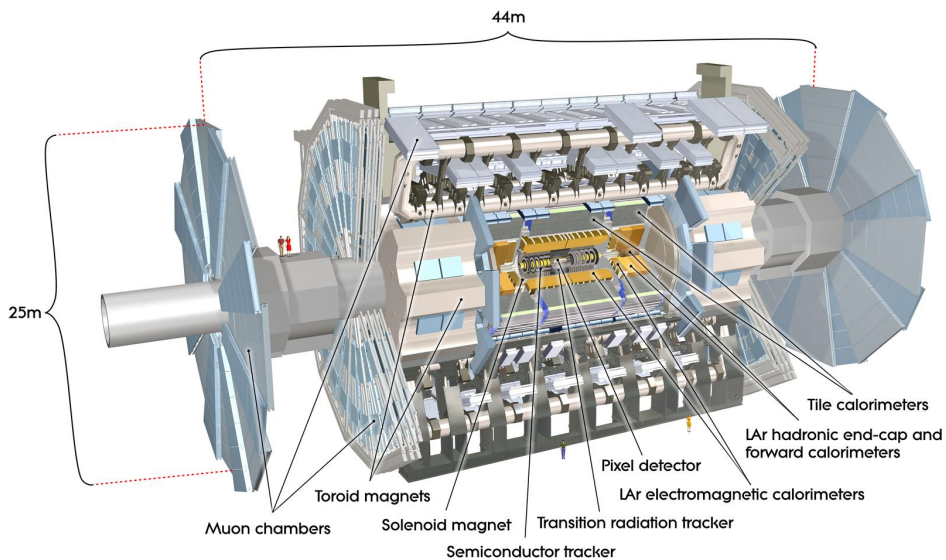


Figure 2.5: The overview of the ATLAS detector.

1251

The detector is designed to test the Standard Model predictions, including the

1252 existence of the SM Higgs boson, and to search for new physics phenomena beyond  
1253 the SM. These searches require accurate particle identification and precise mea-  
1254 surements close to the interaction point. The physics goals and the high energy  
1255 and luminosity of the LHC can be expressed in the following general requirements:  
1256 1. The high energy and flux radiation environment require the use of fast and  
1257 radiation-hard electronics and sensors. Besides, a high sub-detector granularity is  
1258 crucial to handle the high particle multiplicity and to cope with the pile-up arising  
1259 from high luminosity.  
1260 2. An almost full coverage in the polar  $\theta$ , and azimuthal  $\phi$ , angles is required to  
1261 include the entire collision event to enable the calculation of the missing transverse  
1262 energy (MET) of the undetected particles.  
1263 3. Very good charged particle momentum resolution and high track reconstruction  
1264 efficiency in the inner detector are essential for the precise measurements which  
1265 are close to the interaction point.  
1266 4. Very good electromagnetic calorimeter to identify and measure the energy of  
1267 both electrons and photons is needed.  
1268 5. Good hadronic calorimeter with full coverage is needed to measure the jet en-  
1269 ergy and calculate the MET with good resolution.  
1270 6. Muon identification and the ability to measure momentum with good resolution  
1271 over a wide range of momenta are necessary.  
1272 7. Stable magnetic field with strong bending power is fundamental for momentum  
1273 measurements of charged particles in the inner detector and muon spectrometer.  
1274 8. High efficiency triggering is needed to achieve a high and stable data taking  
1275 rate with sufficient background rejection.

1276  
1277 The ATLAS satisfies these requirements to a large extent. Its extended length  
1278 provides a large acceptance in pseudorapidity and the high granularity of the  
1279 inner detector allows an efficient track reconstruction. Both the electromagnetic  
1280 and hadronic calorimeters provide good energy resolution. The ATLAS muon  
1281 spectrometer yields a good muon identification and the trigger system has a very  
1282 high efficiency. The ATLAS coordinate system, sub-detectors and trigger system  
1283 will be described as follows.

## 1284 **2.2.1 The ATLAS coordinate system**

1285 The coordinate system defined here is used to describe tracks of particles in the  
1286 ATLAS detector. The origin of this coordinate system is defined to be at the  
1287 interaction point (IP), where the collision occurs. The positive x-axis points from  
1288 the IP towards the center of the LHC ring, while the y-axis points upwards from  
1289 the IP. The beam direction, which is transverse to the x-y plane, defines the z-  
1290 axis with the positive direction pointing towards LHCb. Since the ATLAS is  
1291 cylindrical shaped, it is also convenient to define a cylindrical coordinate system.  
1292 In this system the polar angle  $\theta$  is measured between the z-axis and the x-y plane

1293 and the azimuthal angle  $\phi$  is measured from the x-axis around the beam in the  
 1294 x-y plane.

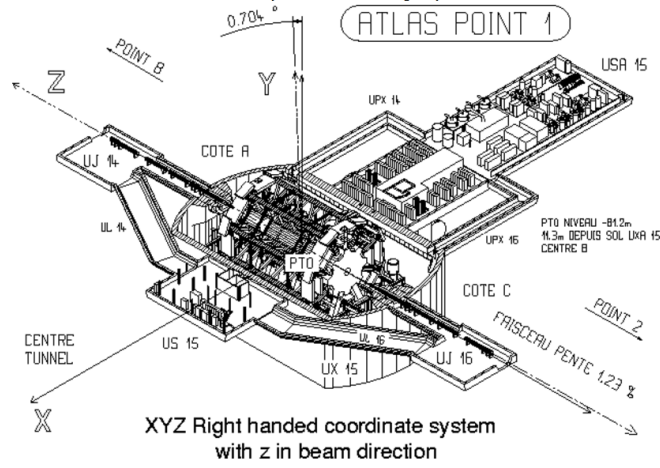


Figure 2.6: The coordinate system in the ATLAS detector. The general tilt of the LEP/LHC tunnel causes the y-axis to be slightly different from vertical [13].

1295 The pseudorapidity variable  $\eta$  defined by Equation 2.3 is used to describe the  
 1296 object's trajectory in the detector,

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \quad (2.3)$$

1297 It depends only on the angle  $\theta$ , but it can also be defined in terms of the particle's  
 1298 momentum as Equation 2.4:

$$\eta = \frac{1}{2} \ln\left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}\right). \quad (2.4)$$

1299 It should be noted that the rapidity of a particle is defined as Equation 2.5:

$$\eta = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right). \quad (2.5)$$

1300 The angular distance  $\Delta R$  between two objects in the detector is usually defined  
 1301 in the  $\eta - \phi$  space as Equation 2.6:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.6)$$

1302 The transverse momentum,  $p_T$ , is defined in the x-y plane as Equation 2.7:

$$p_T = |\mathbf{p}| \sin \theta. \quad (2.7)$$

### 1303 2.2.2 Inner tracking detector

1304 The Inner Detector (ID), shown in Figure 2.7 [14], is contained within a cylinder  
 1305 of 7 m length and 1.15 m radius and is fully immersed within the 2 T magnetic  
 1306 field provided by the solenoid. The large track density at the LHC needs a very  
 1307 good momentum resolution and precise vertex reconstruction, both for primary  
 1308 vertices from the hard collision and for secondary vertices from long-lived particles  
 1309 such as kaons,  $\tau$  leptons or jets produced by heavy flavour quarks. The ID detects  
 1310 particles by measuring the interaction of a particle with the surrounding material  
 1311 at discrete space points up to  $|\eta| < 2.5$ . It can only detect electromagnetically  
 1312 charged particles, neutral particles are not detected in the ID. The direction of  
 1313 the curvature gives the charge of the particle, the degree of curvature gives the  
 1314 momentum.

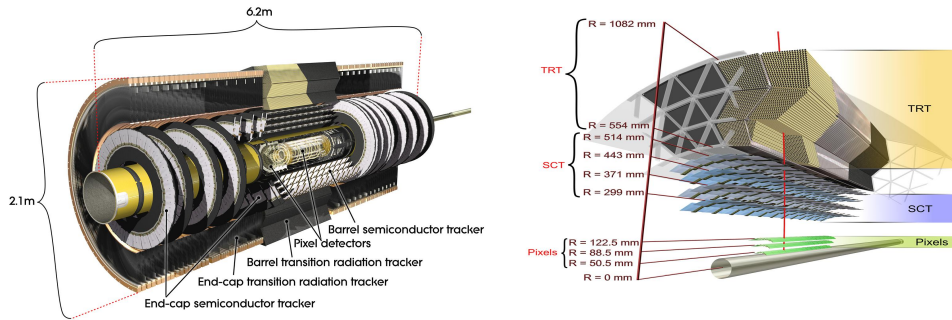


Figure 2.7: View of the ATLAS inner detector (left) and zoomed view on the layered structure of the pixel, SCT and TRT subsystems.

1315 The ID is composed of three subsystems, the pixel detector, the semiconductor  
 1316 tracker (SCT) and the transition radiation tracker (TRT). It combines high res-  
 1317 olution silicon pixel layers and silicon microstrip detectors in the inner part and  
 1318 continuous straw-tube detectors in the outer part. Out of the three ID subsystems,  
 1319 the pixel system is the one located closest to the beam line. Its innermost layer,  
 1320 often called the B-layer, is positioned at 5 cm from the IP. The pixel detector  
 1321 has the highest granularity and gives very precise measurements of track impact  
 1322 parameters which help with the identification of displaced vertices from  $b$ -hadron  
 1323 or  $\tau$  lepton decays. The TRT offers substantial discrimination between electrons  
 1324 and charged hadrons over a wide energy range through the detection of transition  
 1325 radiation photons. During Run 1, several leaks in the TRT exhaust system led to

1326 a large loss of highly expensive Xenon gas. To reduce the operation costs, a few  
1327 parts of the TRT (internal layer of the barrel TRT and one end-cap wheel on each  
1328 side of the detector) were running with Xenon-Argon gas mixture since 2015.

### 1329 **2.2.3 The ATLAS calorimeter**

1330 The ATLAS calorimeter (Figure 2.8 [15]) is separated in one barrel ( $|\eta| < 1.475$ ),  
1331 and two end-cap regions. The former is divided in the LAr electromagnetic barrel  
1332 and the Tile barrel sub-systems. Each end-cap calorimeter includes the LAr  
1333 electromagnetic end-cap (EMEC), the LAr hadronic end-cap (HEC) and the LAr  
1334 forward (FCal) subsystems. The HEC calorimeter is situated behind the EMEC,  
1335 while the FCal is placed near the inner detector. Generally, different particles  
1336 have the different behaviours in each detector (Figure 2.9 [16]). Electrons can  
1337 leave a full track in the inner detector, then they deposit its energy in the LAr  
1338 electromagnetic calorimeter by interacting with the materials. But muons can go  
1339 through the whole detector and a clear full track can be found.

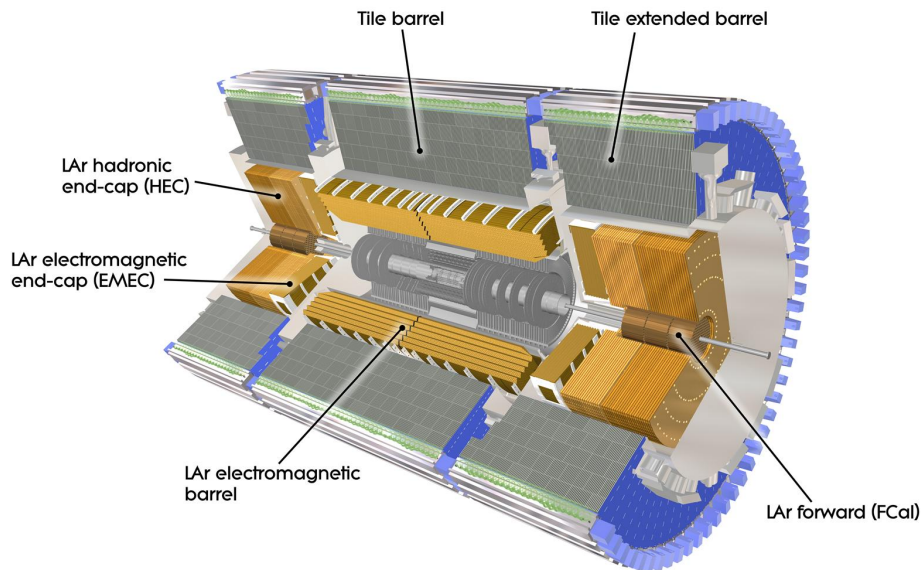


Figure 2.8: Overview of the ATLAS calorimeter. Near the beam-pipe the tracker is visible, surrounding it is the EM calorimeter and beyond the hadronic calorimeter. Both barrel and end-caps elements are displayed.

#### 1340 **2.2.3.1 Electromagnetic calorimeter**

1341 The Electromagnetic (EM) calorimeter is divided into a barrel component up to  
1342  $|\eta| < 1.475$  and two end-cap sections covering  $1.375 < |\eta| < 3.2$ . All three of

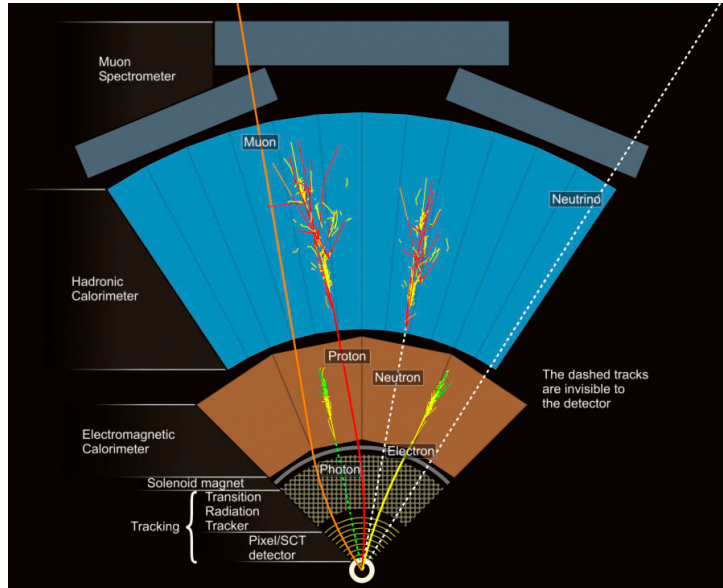


Figure 2.9: Overview of particles passage through the ATLAS tracker, electromagnetic calorimeter, hadronic calorimeter and eventually through the first layers of the muon spectrometer.

1343 these sections have their own cryostat. The barrel part comprises two half-barrels  
 1344 divided by a 4 mm gap at  $z = 0$ . Each half-barrel is 3.2 m long, weighs 57 tons  
 1345 and the inner (outer) radius measures 2.8 m (4 m). Each of the two end-cap parts  
 1346 is made of two coaxial wheels where the inner and outer wheels cover the ranges  
 1347  $1.375 < |\eta| < 2.5$  and  $2.5 < |\eta| < 3.2$ , respectively. A wheel has a width of 63 cm  
 1348 and weighs 27 tons.

1349 The EM calorimeter is a Liquid Argon (LAr) detector with accordion-shaped  
 1350 electrodes made of Kapton, a polyimide, and lead absorber plates whose thickness  
 1351 is optimized for energy resolution. The accordion-shaped arrangement allows for  
 1352 a full coverage in without any cracks and ensures that each particle travelling  
 1353 through the detector will cross approximately the same amount of material. As  
 1354 charged particles hit the absorber plates, they produce EM showers of electrons and  
 1355 photons, the latter maybe turn into pair-produce electrons. These showers ionize  
 1356 the LAr and the ionized electrons drift to the read-out electrodes. The latter are  
 1357 installed between the absorbers and are surrounded by copper plates that are held  
 1358 at a potential of 2000 V, leading to a drift time of 450 ns. The signal size on  
 1359 the electrode, which is roughly proportional to the number of electrons reaching  
 1360 the electrode, determines the energy measurement and so directly in influence  
 1361 the resolution. It is therefore important to prevent large resolution from leakage  
 1362 fluctuations. The additional electrons from the containment of the full shower  
 1363 in the longitudinal direction also improve the sampling resolution. A presampler  
 1364 detector is installed within  $|\eta| < 1.8$  to correct for inhomogeneous energy losses



1365 of electrons and photons travelling through the ID and support structures. It is  
1366 made of a thin, active LAr layer of 1.1 cm (0.5 cm) width in the barrel (end-cap).

### 1367 **2.2.3.2 Hadronic calorimeter**

1368 The purpose of the hadronic calorimeter is to measure the position and energy  
1369 of hadronic jets. Hadronic jets are narrow cones of protons, neutrons, mesons  
1370 and other particles produced by the fragmentation and hadronization of quarks  
1371 and gluons. These jets pass through the ID and ECAL without significant loss of  
1372 energy. The HCAL is located outside of the ECAL and comprises a barrel and two  
1373 end-cap regions. In the barrel region the HCAL uses steel layers as an absorber  
1374 and plastic scintillator sampling sheets, called tiles, as an active material. The  
1375 barrel itself is divided into a central section which covers  $|\eta| < 1$  and two extended  
1376 barrels covering  $0.8 < |\eta| < 1.7$ . The tile calorimeter extends radially from 2.28 m  
1377 to an outer radius of 4.25 m making a total thickness of 9.7 interaction lengths ( $\lambda$ )  
1378 at  $\eta = 0$ . Hadrons interact with absorber nuclei producing a shower of particles.  
1379 Light is emitted when the shower particles pass through the scintillator. Fibers at  
1380 the end of each tile collect this light and carry it to photomultiplier tubes which  
1381 convert it into an electric signal. This signal is used to determine the energy of the  
1382 incident particles. The Hadronic End-cap Calorimeter (HEC) uses the same LAr  
1383 technology as the ECAL, but with copper plates as an absorber. Each end-cap  
1384 region consists of two wheels fixed behind the end-cap ECAL. With a small overlap  
1385 with the tile calorimeter, the HEC covers  $1.5 < |\eta| < 3.2$ .

### 1386 **2.2.4 Muon spectrum system**

1387 The Muon Spectrometer (MS) as shown in Figure 2.10 [17] is the outer-most  
1388 layer of the detector and has the capability of accurately measuring the muon  
1389 momentum independent of the inner detector tracking system and provide an  
1390 independent muon trigger. The spectrometer consists of two precision detectors  
1391 and two triggering detectors, which are embedded in a toroidal magnetic field. The  
1392 two precision detectors are the Monitored Drift Tubes (MDT) and the Cathode  
1393 Strip Chambers (CSC) and the two triggering detectors: the Thin Gap Chambers  
1394 (TGC) and the Resistive Plate Chambers (RPC). With the use of all these sub-  
1395 detectors the muon momentum resolution is designed to be 3% for 10-200 GeV  
1396 and about 10% for 1 TeV muons.

### 1397 **2.2.5 Trigger system**

1398 The designed luminosity of the LHC is  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  and the bunch crossing rate is  
1399 about 40 MHz. Only maximum amount of data that can be stored from collision  
1400 is  $300 \text{ Mbs}^{-1}$ , which corresponds to an event rate of 200 Hz. The event rate of  
1401 200 Hz is limited by the offline processing capacity. In 2012, the offline processing

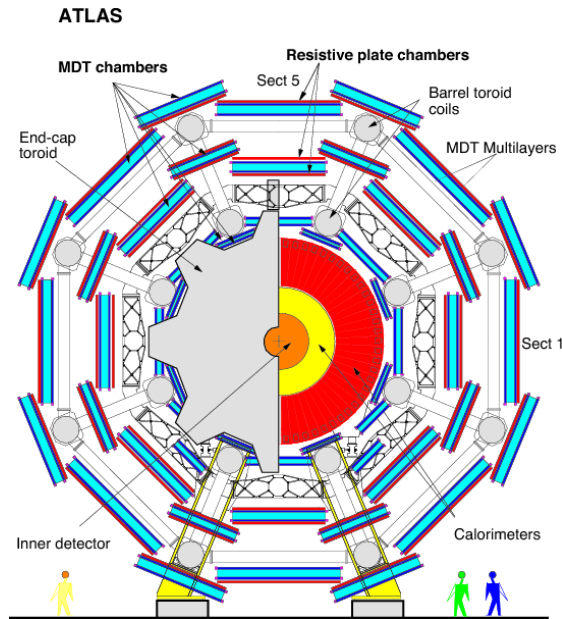


Figure 2.10: Schematic view of the muon spectrometer in the x-y projections. Inner, Middle and Outer chamber stations are denoted BI, BM, BO in the barrel and EI, EM, EO in the end-cap.

1402 capacity was raised to almost 400 Hz, by having a delayed processing of some  
 1403 fraction of the data which won't be available till after the run stops. In order to  
 1404 achieve this challenging goal, the ATLAS has a three level trigger system, shown  
 1405 in Figure 2.11 [18], designed to reduce the event rate to a manageable level. The  
 1406 level 1 trigger consists of hardware while the subsequent two triggers, level 2 and  
 1407 event filter are software based. The combination of the level 2 and event filter are  
 1408 known as the high level trigger (HLT). Each subsequent level refines the decisions  
 1409 made in the previous levels.

## 1410 2.2.6 The ATLAS simulation and computing

1411 The ATLAS experiment has developed a computing model to allow members of  
 1412 the collaboration all over the world to access the ATLAS data. The main blocks  
 1413 of the model are the Athena [19] software framework, which operates on top of  
 1414 a hierarchical model of computing - the GRID [20]. GRID is used to created a  
 1415 distributed computing framework throughout several facilities in remote locations,  
 1416 that are able to communicate with each other and share tasks. Such facilities are  
 1417 referred to as Tiers. Tier-0 is located at CERN and handles the most unrefined  
 1418 data, referred to as RAW data. Ten worldwide facilities constitute Tier-1, that  
 1419 deals primarily with event reconstruction. Approximately 35 more facilities form  
 1420 Tier-2 which provides the analysis abilities for the ATLAS collaboration. Athena

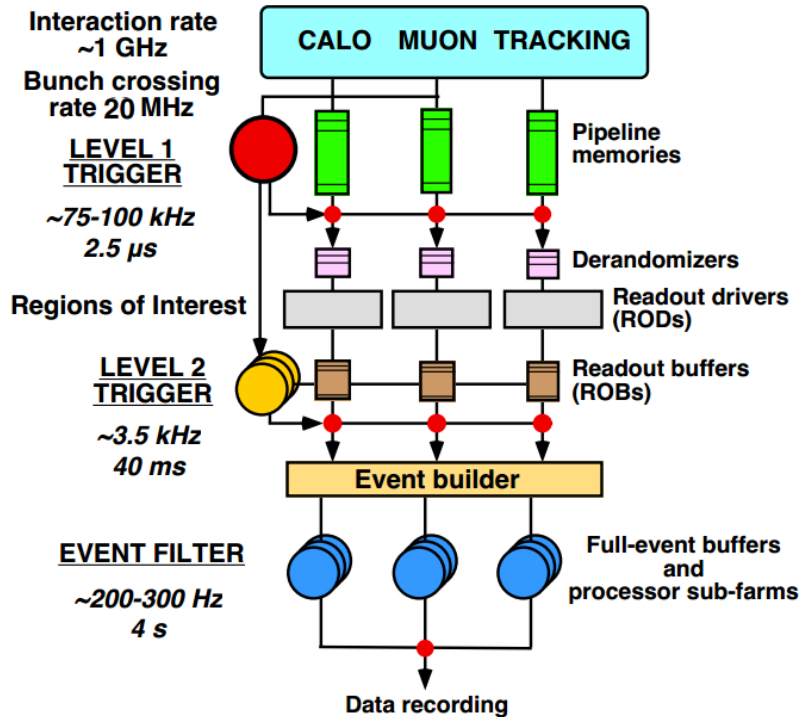


Figure 2.11: The ATLAS trigger system and its integration.

1421 includes software for event simulation, event trigger, event reconstruction and  
 1422 physics analysis tools.

1423 Event simulation [21] is fundamental in an experiment such as the ATLAS. It  
 1424 carries the events from generation to output, in a format which is identical to that  
 1425 of the true detector, allowing for a direct comparison between the real data and  
 1426 the theoretical models. The simulation software chain is generally divided into  
 1427 three steps. The first is the generation of events using a Monte Carlo program.  
 1428 The second step is the simulation of the detector and physics interactions.

1429 In the ATLAS, the simulation is integrated into Athena and uses the Geant4  
 1430 simulation toolkit [22]. Geant4 basically simulates the entire the ATLAS detector  
 1431 (material, geometry and subsystems, including trigger) and its response to traversing  
 1432 particles. Digitization is the last step, and consists in converting the energy  
 1433 deposited in the sensitive regions of the detector into voltages and currents, for  
 1434 comparison to the readout of the ATLAS detector. The output of this process  
 1435 is referred to as RAW data and is identical to the output of the ATLAS TDAQ  
 1436 system during real data taking.

1437 The complexity of the full simulation (FullSim) of the ATLAS detector has led to  
 1438 the development of fast simulation strategies. These allow for a faster production  
 1439 of the high simulated event statistics needed for physics analyses. ATLFAST-II is a  
 1440 fast simulation framework developed for the ATLAS experiment. It is made of two

1441 components: the Fast ATLAS Tracking system (FAtlas), for ID and MS simulation,  
 1442 and the Fast Calorimeter Simulation (FastCaloSim), for calorimeter simulation.  
 1443 Optionally, any of these sub-detectors can be simulated with the nominal Geant4,  
 1444 providing flexibility to suit the needs of different physics analyses.

## 1445 2.2.7 The LHC and ATLAS in Run 2



Figure 2.12: The main 2013-2014 LHC consolidations toward Run 2 data taking.

1446 The LHC Run 2 collisions started in June 2015 with the energy in the center of  
 1447 mass of 13 TeV, the largest energy scale in the world ever before. This high energy  
 1448 leads to a large increase in the production cross sections, ensure the observation  
 1449 and the study of rare process, the higher precision Standard Model measurements  
 1450 and improving the discovery potential for beyond the Standard Model physics like  
 1451 Supersymmetry, dark matter, etc. To adjust this so high collision energy, several  
 1452 upgrades were brought to the LHC and detector machines, and improvements to  
 1453 the online and offline reconstruction and analysis software during the first two years  
 1454 long shut down (2013-2015). Some details of the improvements will be presented  
 1455 in this section.

1456 During the two years long Shut down-1 (LS1) the LHC machine was highly  
 1457 improved to target an energy of 6.5 TeV per proton beam and a bunch spacing  
 1458 of 25 ns. The main consolidations between 2013 and 2014 are illustrated in Fig-  
 1459 ure 2.12 [23]. The Run 2 operation started with the initial beam commissioning in

1460 April 2015. It lasted about two months firstly with test beams with an excellent  
 1461 and improved system performance, such as the beam instrumentation and colli-  
 1462 mation, injection and beam dump systems, vacuum and magnetic systems and the  
 1463 machine protection. Software and the analysis tools are also highly improved in  
 1464 the meanwhile.

1465 Starting on April 5th, 2015 the first stable beams at 6.5 TeV were circulated in  
 1466 the LHC ring. Just after, a period of scrubbing, or of electron bombardment of  
 1467 the beam pipe surface, was mandatory to highly reduce the electron cloud around  
 1468 the beams. Finally, on June 3rd, 2015 the LHC started the first collisions at  
 1469  $\sqrt{s} = 13$  TeV. The LHC Run 2 physics proton-proton period started with around  
 1470 30 days of collisions with 50 ns bunch spacing, and ATLAS collected an integrated  
 1471 luminosity about  $0.1 \text{ fb}^{-1}$ . Following stable runs are with 25 ns bunch spacing after  
 1472 August and the expected integrated luminosity to be collected is about  $100 \text{ fb}^{-1}$   
 1473 totally before the next long shut down from 2018 to 2020. Total integral lumi-  
 1474 nosity during the whole 2016 year is shown in Figure 2.13 (left) and the number  
 1475 of interactions per crossing with Run 2 condition is also presented. The average  
 1476 value is about 25 ns as expected and increased considerably compared to the Run  
 1477 1 condition (Figure 2.4).

1478

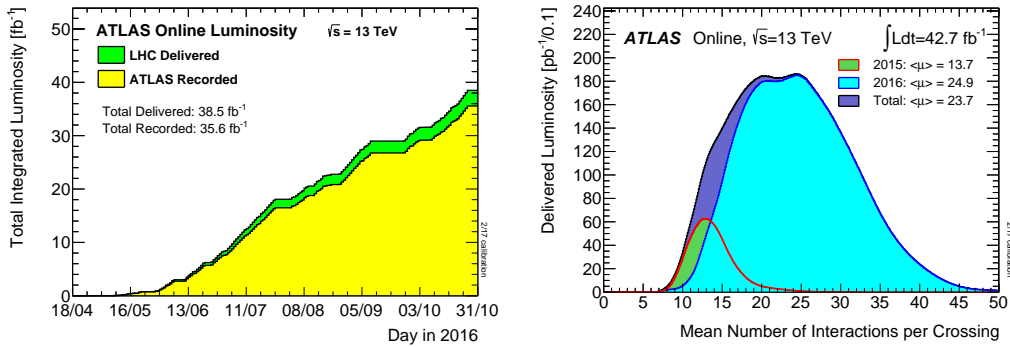


Figure 2.13: Total integrated luminosity in 2016 (left) and number of interactions per crossing with the combined 13 TeV data for 2015 and 2016 [24].

1479 The ATLAS detector was also consolidated and several upgrades were performed  
 1480 during the LS1. It has a new beam-pipe, improved magnet and cryogenic systems.  
 1481 Compared to Run 1, the muon chambers were completed in the  $[1.1, 1.3] \eta$  region  
 1482 and repaired. The dead pixel modules and the calorimeter electronics were also  
 1483 reconditioned. To improve the tracking and the vertex reconstruction performance  
 1484 at high luminosities, a fourth pixel layer or the "Insertable B-layer" (IBL) was  
 1485 added. It is located just near the beam pipe, at a distance  $R = 3.3$  cm, and  
 1486 presents a 3D silicon pixel technology. The purpose of this layer is to ensure a  
 1487 high quality tracking performance and an improved  $b$ -tagging at high luminosities

<sup>1488</sup> for the LHC Run 2 operation.

# 1489 **3 Object reconstruction in the AT-** 1490 **LAS**

1491 In the ATLAS, all physical objects are reconstructed with dedicated algorithms  
1492 using the collision information recorded by the detector. All the ATLAS physics  
1493 analyses always start with the reconstructed and identified objects representing the  
1494 observed characteristics of the particles coming from the proton-proton collisions  
1495 and travelling through the detector volume. The algorithms for reconstruction and  
1496 identification of physics objects have been developed initially on simulated samples  
1497 or early data, and then commissioned and optimized during the ATLAS running,  
1498 keeping up with the changing energy and pile-up conditions of the collisions. The  
1499 three leptons final state of the  $t\bar{t}H$  includes leptons, jets and missing transverse en-  
1500 ergy. It is important to have good view of the details of the objects reconstruction  
1501 in the ATLAS. Electrons (Section 3.1), muons (Section 3.3), photons (Section 3.2  
1502 ), tau( $\tau$ )s (Section 3.4 ), jets (Section 3.5) and missing transverse energy (Sec-  
1503 tion 3.6) are presented in the following sections. The results described here are  
1504 clearly the outcome of a well operated the ATLAS detector and a dedicated and  
1505 committed work of the ATLAS performance and physics groups.

## 1506 **3.1 Electrons**

1507 Electrons are employed widely in the ATLAS physical analysis as the main pri-  
1508 mary signature for many physical processes. They are used in a wide range of  
1509 physics analyses: from precision standard model measurements to the search for  
1510 exotic new physics. The electron candidates then can be further selected against  
1511 background, such as hadrons and background (non-prompt) electrons originating  
1512 predominantly from photon conversions and heavy flavour hadron decays, using  
1513 several sets of identification criteria with different levels of background rejection  
1514 and signal efficiency. Many aspects of the overall design of the ATLAS were driven  
1515 by the requirements that electrons be well-reconstructed and efficiently identified.

### 1516 3.1.1 Reconstruction

1517 The signature of an electron in the ATLAS is a reconstructed track in the Inner  
1518 Detector (ID), associated to a narrow, localized cluster of energy in the electro-  
1519 magnetic (EM) calorimeter. Electron reconstruction in the central region of the  
1520 ATLAS detector ( $|\eta| < 2.47$ ) is with several steps as following:

- 1521 • **Seed-cluster reconstruction**

1522 A sliding window with a size of  $3 \times 5$  in units of  $0.025 \times 0.025$ , corresponding  
1523 to the granularity of the EM calorimeter middle layer, in  $\eta \times \phi$  space is used  
1524 to search for electron cluster "seeds" as longitudinal towers with total cluster  
1525 transverse energy above 2.5 GeV. The clusters are then formed around the  
1526 seeds using a clustering algorithm [25] which allows for duplicated objects to  
1527 be removed. The cluster kinematics are reconstructed by using an extended  
1528 window depending on the cluster position in the calorimeter. The efficiency  
1529 of this cluster searching ranges from 95% at  $E_T = 7$  GeV to more than 99%  
1530 above  $E_T = 15$  GeV.

- 1531 • **Track reconstruction**

1532 Track reconstruction is proceeding in two steps: pattern recognition and  
1533 track fit. The ATLAS pattern recognition uses the pion hypothesis for energy  
1534 loss because of the interactions with the detector material. This has been  
1535 complemented with a modified pattern recognition algorithm which allows up  
1536 to 30% energy loss at each intersection of the track with the detector material  
1537 to account for possible bremsstrahlung. If a track seed (consisting of three  
1538 hits in different layers of the silicon detectors) with a transverse momentum  
1539 larger than 1 GeV can not be successfully extended to a full track of at  
1540 least seven hits using the pion hypothesis and it falls within one of the EM  
1541 cluster region of interest, a second attempt is performed with a new pattern  
1542 recognition using an electron hypothesis that allows for larger energy loss.  
1543 Track candidates are then fit either with the pion hypothesis or the electron  
1544 hypothesis (according to the hypothesis used in the pattern recognition),  
1545 using the ATLAS Global  $\chi^2$  Track Fitter [26]. If a track candidate fails  
1546 the pion hypothesis track fit (for example, due to large energy losses), it  
1547 is refitted with the electron hypothesis. In this way, a specific electron-oriented  
1548 algorithm has been integrated into the standard track reconstruction. It  
1549 improves the performance for electrons and has minimal interference with  
1550 the main track reconstruction.

- 1551 • **Electron specific track fit**

1552 The obtained tracks are then loosely matched to EM clusters by using the  
1553 distance in  $\eta$  and  $\phi$  between the position of the track, after extrapolation,  
1554 in the calorimeter middle layer and the cluster barycentre. The matching  
1555 condition account for energy loss dues to bremsstrahlung and the number of



1556 precision hits in the silicon detector. Tracks which have larger than 4 signif-  
1557 icant number of precision hits and are loosely associated to electron clusters  
1558 are refitted by using an optimised Gaussian Sum Filter (GSF) [27], which  
1559 takes into account the non-linear bremsstrahlung effects.  
1560

- 1561 • **Electron candidate reconstruction**

1562 The matching of the track candidate to the cluster seed is the final stage  
1563 of the electron reconstruction procedure. A similar matching as the one de-  
1564 scribed above is repeated for the refit track with more stricter conditions. If  
1565 several tracks fulfil the matching condition, one track is chosen as a "primary"  
1566 track. The choice is based on an algorithm using the cluster-track distance  
1567  $R$  calculated using different momentum hypotheses, the number of pixel hits  
1568 and the presence of a hit in the first silicon layer [28]. Electron candidates  
1569 without any associated precision hit tracks are removed and considered to be  
1570 photons. The electron cluster is then re-formed using  $3 \times 7$  ( $5 \times 5$ ) longi-  
1571 tudinal towers of cells in the barrel (end-caps) of the EM calorimeter. The  
1572 energy of the clusters is calibrated to the original electron energy using the  
1573 multivariate techniques [29] based on simulated MC samples.  
1574

### 1575 3.1.2 Calibration

1576 Due to the detector geometry and material distribution, the correction are needed  
1577 for the position measurement and energy measurement based on the MC samples.  
1578 The position  $(\eta, \phi)$  of the electron cluster is calculated as the energy-barycenter  
1579 of all cells in the cluster, using the information from the first and second layers of  
1580 the EM calorimeter. Corrections are considered as the LAr calorimeter accordion  
1581 geometry and finite granularity. These are applied separately for each layer and  
1582 for the barrel and end-cap regions of the detector. The  $\eta$  position is estimated in  
1583 each cell of the cluster as a function of the impact point of the incident particle and  
1584 is biased toward the center of the cell. As cells can have a finite granularity and  
1585 the shower energy, which is not fully inside the cell, the  $\eta$  position is computed  
1586 with respect to the cell center with an  $\eta$  correction. Considering the detector  
1587 geometry, the correction is computed for several  $\eta$  regions and depends on the  
1588 distance from the interaction point of the the incident particle to the center of the  
1589 cell. The measurements are performed for each layer and then combined results  
1590 are obtained to have the final position  $(\eta, \phi)$  of the electron cluster.

1591 Although the deposited energy of the electrons are recorded by the ATLAS  
1592 efficiently, still a small energy is deposited in the non-instrumented section, as  
1593 the inner detector, cryostat, solenoid, cables, etc. To account for these effects,  
1594 the energy cells are calibrated. The EM cluster energy is calibrated to the initial  
1595 electron energy using the special MC samples. A multivariate analysis is used by

1596 training on the MC as well as the inter-calibration of the scale of the first and  
1597 second longitudinal EM calorimeter layers. The overall electron energy resolution  
1598 measured in the data is adjusted as a function of  $\eta$ , using  $Z \rightarrow ee$  events. Besides,  
1599 to account for the worse resolution in data, additional scale factors are applied to  
1600 MC electron energies to match the data (scale calibration).

1601 The four-momentum of the electrons is computed using information from both  
1602 the final calibrated energy cluster and the best track matched to the original  
1603 seed cluster. The energy is given by the final calibrated cluster, while the  $\eta$  and  
1604  $\phi$  directions are taken from the corresponding track parameters with respect to  
1605 the beam-line. For Run 2 analyses, the electron measurements are performed by  
1606 requiring the track associated with the electron to be compatible with the primary  
1607 interaction vertex of the hard collision, and the aim is to reduce the background  
1608 from conversions and secondary particles. The track parameters are calculated in  
1609 a reference frame where the z-axis is taken along the measured beam-line position.  
1610 The following conditions are applied together with all the identification operating  
1611 points considered in this thesis:  $d_0/\sigma_{d_0} < 5$  and  $\Delta z_0 \sin \theta < 0.5$  mm, where the  
1612 impact parameter  $d_0$  is the distance of closest approach of the track to the measured  
1613 beam-line,  $z_0$  is the distance along the beam-line between the point where  $d_0$  is  
1614 measured and the beam-spot position, and  $\theta$  is the polar angle of the track. To  
1615 achieve the compatibility with the primary vertex of the hard collision the  $\Delta z_0$   
1616 between the track and the primary vertex is employed. This vertex is selected  
1617 from the reconstructed primary vertices (compatible with the beam-line) as the  
1618 one with the highest sum of transverse momenta of the associated tracks. The  $\sigma_{d_0}$   
1619 represents the estimated uncertainty of the  $d_0$  parameter, and  $\theta$  is the polar angle  
1620 of the track. The efficiency of these requirements in data and MC is estimated  
1621 together with the efficiency of the various identification operating points.

### 1622 3.1.3 Identification

1623 The electrons from  $W$  or  $Z$  decays are typically isolated, hadron fake if it does  
1624 not match a true electron or tau or muon, non-isolated if it matches a true elec-  
1625 tron originating from  $b$  or  $c$  mesons and background electron if it matches a true  
1626 electron coming from Dalitz decays or conversions. To determine whether the  
1627 reconstructed electron candidates are signal-like objects or background-like ob-  
1628 jects such as hadronic jets or converted photons, electron identification (EID) is  
1629 applied. The EID uses quantities related to the electron cluster and track mea-  
1630 surements including calorimeter shower shapes, information from the transition  
1631 radiation tracker, track-cluster matching related quantities, track properties, and  
1632 variables measuring the bremsstrahlung effects for the distinguishing signal from  
1633 background. The quantities are summarised in Table 3.1 [30].

1634 Towards Run 2, several changes to the input variables used for EID have been  
1635 introduced. Taking advantage of the IBL, the number of hits in this innermost  
1636 pixel layer is used for discriminating between electrons and converted photon and

Table 3.1: Definitions of electron discriminating variables.

Type	Description	Name
Hadronic leakage	Ratio of $E_T$ in the first layer of the hadronic calorimeter to $E_T$ of the EM cluster (used over the range $ \eta  < 0.8$ or $ \eta  > 1.37$ )	$R_{had1}$
	Ratio of $E_T$ in the hadronic calorimeter to $E_T$ of the EM cluster (used over the range $0.8 <  \eta  < 1.37$ )	$R_{had}$
Back layer of EM calorimeter	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter. This variable is only used below 100 GeV because it is known to be inefficient at high energies.	$f_3$
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$ , where $E_i$ is the energy and $\eta_i$ is the pseudorapidity of cell $i$ and the sum is calculated within a window of $3 \times 5$ cells	$\omega_{stot}$
	Ratio of the energy in $3 \times 3$ cells over the energy in $3 \times 7$ cells centered at the electron cluster position	$R_\phi$
	Ratio of the energy in $3 \times 7$ cells over the energy in $7 \times 7$ cells centered at the electron cluster position	$R_\eta$
Strip layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i (i - i_{max})^2)/(\sum E_i)}$ , where $i$ runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$ , corresponding typically to 20 strips in $\eta$ , and $i_{max}$ is the index of the highest-energy strip	$w_{stot}$
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	$E_{ratio}$
	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter	$f_1$
Track conditions	Number of hits in the innermost pixel layer; discriminates against photon conversions	$n_{Blayer}$
	Number of hits in the pixel detector	$n_{Pixel}$
	Number of total hits in the pixel and SCT detectors	$n_{Si}$
	Transverse impact parameter with respect to the beam-line	$d_0$
	Significance of transverse impact parameter defined as the ratio of $d_0$ and its uncertainty	$d_0/\sigma_{d_0}$
	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum	$\Delta p/p$
TRT	Likelihood probability based on transition radiation in the TRT	eProbabilityHT
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the middle layer and the track extrapolated from the perigee	$\Delta\phi_2$
	Defined as $\delta\phi$ , but the track momentum is rescaled to the cluster energy before extrapolating the track from the perigee to the middle layer of the calorimeter	$\Delta\phi_{res}$
	Ratio of the cluster energy to the track momentum	$E/p$

1637 now is the second-to-innermost pixel layer. The change in the TRT gas led to  
 1638 modifications in the detector response and prompted the introduction of a new  
 1639 discriminating variable in the electron identification algorithms. In Run 1, only  
 1640 the fraction of high-threshold hits was used from the TRT as a signature of tran-  
 1641 sition radiation to distinguish electrons from hadrons. In Run 2, a likelihood  
 1642 method based on the TRT high-threshold hits is introduced to compensate for the  
 1643 lower transition radiation absorption probability of the argon. The TRT likeli-  
 1644 hood method uses the high-threshold probability of each TRT hit to construct a  
 1645 discriminant variable, referred to here as eProbabilityHT. Cut-based method and  
 1646 Likelihood-based (LH) method are used with those discriminate variables to reject  
 1647 the background efficiently, especially for Likelihood-based method is the baseline  
 1648 in Run 2 analysis. However, higher collision energy and more complex pile-up  
 1649 situation compared to Run 1, both methods need re-optimize to adjust to the new  
 1650 condition.

1651

1652 **Cut-based method**

1653 In Run 1 cut-based method using a set of rectangular cuts on the EID discrim-  
 1654 inating variables was used. Three sets of cut-based methodology identification  
 1655 menus are defined with increasing background rejection power: *Loose*, *Medium*  
 1656 and *Tight*. Towards Run 2 re-optimize work is done for cut-based method by  
 1657 using the  $Z \rightarrow ee$ ,  $J/\psi \rightarrow ee$  MC samples and  $Z \rightarrow ee$  events of the first data  
 1658 recorded in 2015. With re-optimized cut-based method, similar performance with  
 1659 Run 2 condition comparing to Run 1 can be achieved. Even under the more com-  
 1660 plex pile-up condition, the re-optimized menus give a good separation power for  
 1661 the signal-like electron from the background (details in Appendix A, and this is  
 1662 my qualification task done in 2015). The re-optimized cut-based method is used  
 1663 as for cross-checks during the 2015 data taking.

1664

1665 **Likelihood based method**

1666 Likelihood-based method (LH) is a multi-variate analysis (MVA) technique that  
 1667 simultaneously evaluates several properties of the electron candidates when making  
 1668 a selection decision. The LH method uses the signal and background probability  
 1669 density functions (PDFs) of the discriminating variables (shown in Table 3.1).  
 1670 Based on these PDFs, an overall probability is calculated for the object to be signal  
 1671 or background. The signal and background probabilities for a given electron are  
 1672 then combined into a discriminant  $d_{\mathcal{L}}$  on which a requirement is applied:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \quad \mathcal{L}_{S(B)}(\vec{x}) = \prod_{i=1}^n P_{s(b),i}(x_i), \quad (3.1)$$

1673 where  $\vec{x}$  is the vector of discriminating variable values and  $P_{s,i}(x_i)$  is the value of  
 1674 the signal probability density function of the  $i^{\text{th}}$  variable evaluated at  $x_i$ . In the  
 1675 same way,  $P_{b,i}(x_i)$  refers to the background probability function. This allows for a  
 1676 better background rejection for a given signal efficiency than a cut-based algorithm  
 1677 that would use selection criteria sequentially on each variable. In addition to the  
 1678 variables used as input to the LH discriminant, simple selection criteria are used  
 1679 for the variables counting the number of hits on the track.

1680 Three levels of identification operating points are typically provided for EID.  
 1681 In order of increasing background rejection they are *Loose*, *Medium*, and *Tight*.  
 1682 The *Loose*, *Medium*, and *Tight* operating points are defined such that the samples  
 1683 selected by them are subsets of one another. Each operating point uses the same  
 1684 variables to define the LH discriminant, but the selection on this discriminant is  
 1685 different for each operating point. Thus, electrons selected by *Medium* are all  
 1686 selected by *Loose*, and *Tight* electrons are all selected by *Medium*. The distribu-  
 1687 tions of electron shower shapes depend on the amount of material the electrons  
 1688 pass through, and therefore vary with the pseudorapidity of the electron candi-  
 1689 dates. In addition, significant changes to the shower shapes and track properties  
 1690 are expected with increasing energy. The EID operating points were consequently

1691 optimised in several bins in  $\eta$  and  $E_T$ . The performance of the LH identification  
 1692 algorithm is illustrated in Figure 3.1 [30]. Depending on the operating point, the  
 1693 signal (background) efficiencies for electron candidates with  $E_T = 25$  GeV are in  
 1694 the range from 78 to 90% (0.3 to 0.8%) and increase (decrease) with  $E_T$ .

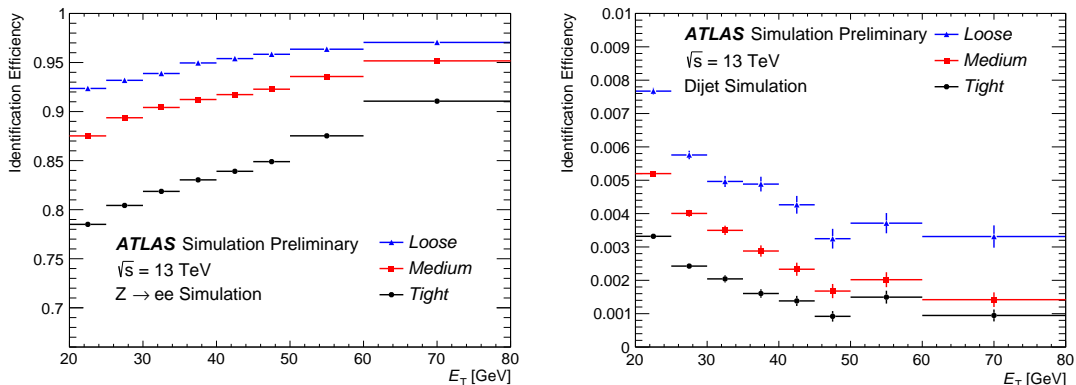


Figure 3.1: The efficiency to identify electrons from  $Z \rightarrow ee$  decays (left) and the efficiency to identify hadrons as electrons (background rejection, right) estimated using simulated dijet samples. The efficiencies are obtained using MC, and are measured with reconstructed electrons. The candidates are matched to true electron candidates for  $Z \rightarrow ee$  events. For background rejection studies the electrons matched to true electron candidates are not included. The last bin used for the optimisation of the ID is 45-50 GeV, which explains the signal efficiency increases slightly more in the 50 GeV bin than in others, and the background efficiency increases in this bin as well.

1695 The electron identification performance may be influenced by the parasitic col-  
 1696 lisions taking place in the same beam crossing (in-time pile-up) or a consecutive  
 1697 bunch crossing (out-of-time pile-up) as the hard  $pp$  collision producing the elec-  
 1698 tron candidates. The number of reconstructed primary vertices is indicative of the  
 1699 level of pile-up in each event, with the average number of primary vertices (eight  
 1700 per event) corresponding to an average pile-up of 13.7. Since some shower shape  
 1701 distributions depend on the number of pile-up collisions per bunch crossing, the  
 1702 cut on the LH discriminant value is loosened as a function of the number of pri-  
 1703 mary vertices. This is done to ensure that the LH identification remains efficient  
 1704 at high pile-up condition without drastically increasing the amount of background  
 1705 accepted by the LH selection. The optimisation included simulations with a num-  
 1706 ber of pile-up collisions of is up to 40, covering the range of the pile-up observed  
 1707 in 2015.

1708 Some of the calorimeter variable distributions are different from the typical dis-  
 1709 tributions obtained with  $Z \rightarrow ee$  at high  $E_T$ . Higher energy electrons tend to  
 1710 deposit relatively smaller fractions of their energy in the early layers of the EM  
 1711 calorimeter, and more in the later layers of the EM calorimeter or even in the

1712 hadronic calorimeter. *Loose* and *Medium* of LH method were deemed to be loose  
1713 enough to be robust against these  $E_T$  dependent changes. For electron candidates  
1714 with  $E_T$  above 125 GeV, LH *Tight* uses the same discriminant selection as LH  
1715 *Medium* but adding rectangular cuts on  $\omega_{stot}$  and  $E/P$ , which were found to be  
1716 particularly effective at discriminating signal from background at high  $E_T$ .  
1717

## 1718 **3.2 Photons**

### 1719 **Reconstruction**

1720 The reconstruction of photon is performed separately for converted and unconverted  
1721 photons. The former is characterized by the existence of at least one track  
1722 originating from a vertex inside the tracker matched to the electromagnetic cluster,  
1723 while the later is not. The conversion vertices reconstructed in the inner detector  
1724 are classified according to the associated number of tracks. If a track is matched in  
1725  $\eta$  and  $\phi$  to a reconstructed EM cluster, the object is added to the converted photon  
1726 collection. The matching is considered to be successful if the impact parameter  
1727 associated to the track, after extrapolation from its last measurement point to the  
1728 second EM layer, is inside an  $(\eta, \phi)$  window of radius 0.05 from the cluster centre.  
1729 The track is extrapolated to the position corresponding to the expected maximum  
1730 energy deposit for EM showers. Finally, if a reconstructed electromagnetic cluster  
1731 can not be associated to a track, then it is considered to be an unconverted photon  
1732 candidate.

1733 Almost all converted photons are also regarded as electrons. The existence of  
1734 a track associated to a conversion vertex is used to separate those two categories.  
1735 If the track associated to the EM cluster coincides with a track originating from  
1736 a conversion vertex and the electron is treated as an converted photon object. If  
1737 the track can not be associated to a conversion vertex, the object is classified as  
1738 an electron. In the case of a initially reconstructed electron with a matched track  
1739 having only the TRT information (and the usual requirements on energy and track  
1740  $p_T$  are fulfilled), the candidate is considered to be a converted photon even if no  
1741 conversion vertex is associated to the EM cluster. Unconverted photons are also  
1742 recovered: if the reconstructed electron has the best track candidate with only the  
1743 TRT information or the converted photon candidate condition is not passed and  
1744 the track  $p_T$  is smaller than a typical value of 2 GeV, the candidate is considered  
1745 to be an unconverted photon. The photon calibration is similar to electrons, and  
1746 the most important sources of experimental systematic uncertainties are: photon  
1747 energy scale and photon energy scale resolution.

1748

### 1749 **Identification**

1750 Photon identification with high signal efficiency and high background rejection is  
1751 required for transverse momenta from 10 GeV to the TeV scale to distinguish prompt

1752 photons from background photons. Photon identification in the ATLAS is based  
 1753 on a set of cuts on several discriminating variables listed in Table 3.2, characterise  
 1754 the lateral and longitudinal shower development in the electromagnetic calorimeter  
 1755 and the shower leakage fraction in the hadronic calorimeter. Prompt photons  
 1756 produce narrower energy deposits in the electromagnetic calorimeter and have  
 1757 smaller leakage to the hadronic one compared to background photons from jets, due  
 1758 to the presence of additional hadrons near the photon candidate in the latter case.  
 1759 Background photons from isolated  $\pi^0 \rightarrow \gamma\gamma$  decays, unlike prompt photons, are  
 1760 often characterised by two separate local energy maxima in the finely segmented  
 1761 strips of the first layer, because of the small separation between the two photons.  
 1762 Due to the pile-up in presence of low- $E_T$  activity in the detector this tends to  
 1763 broaden the distributions of the discriminating variables and thus to reduce the  
 1764 separation between prompt and background photon candidates.

1765 A *Loose* and a *Tight* selection are defined. The *Loose* selection is based only on  
 1766 shower shapes in the second layer of the electromagnetic calorimeter and on the  
 1767 energy deposited in the hadronic calorimeter, and is used by the photon triggers.  
 1768 The *Loose* requirements are designed to provide a high prompt-photon identifica-  
 1769 tion efficiency with respect to reconstruction. Their efficiency rises from 97% at  
 1770  $E_T^\gamma = 20$  GeV to above 99% for  $E_T^\gamma > 40$  GeV for both the converted and uncon-  
 1771 verted photons. The *Tight* selection adds information from the finely segmented  
 1772 strip layer of the calorimeter, which provides good rejection of hadronic jets where  
 1773 a neutral meson carries most of the jet energy. The *Tight* criteria are separately  
 1774 optimised for unconverted and converted photons to provide a photon identifica-  
 1775 tion efficiency of about 85% for photon candidates with transverse energy  $E_T >$   
 1776 40 GeV. The selection criteria are different in seven intervals of the reconstructed  
 1777 photon pseudorapidity (0.0–0.6, 0.6–0.8, 0.8–1.15, 1.15–1.37, 1.52–1.81, 1.81–2.01,  
 1778 2.01–2.37) to account for the calorimeter geometry and for different effects on the  
 1779 shower shapes from the material upstream of the calorimeter.

1780 The photon identification criteria were first optimised prior to the start of the  
 1781 data taking in 2010, using simulated samples of prompt photons from  $\gamma$ +jet, dipho-  
 1782 ton and  $H \rightarrow \gamma\gamma$  events and samples of background photons in QCD multi-jet  
 1783 events [31]. To adjust the 8 TeV run in 2012, the identification were re-optimised  
 1784 based on improved simulations in which the values of the shower shape variables  
 1785 are slightly shifted to improve the agreement with the data shower shapes. To cope  
 1786 with the higher pile-up, the criteria on the shower shapes which is more sensitive  
 1787 to pile-up were relaxed while the others were tightened.

1788 For the data taken in 2011,  $4.9 \text{ fb}^{-1}$  at  $\sqrt{s} = 7$  TeV, the efficiency of the cut-based  
 1789 identification algorithm increases from 60–70% at  $E_T = 20$  GeV up to 87–95%  
 1790 (90–99%) at  $E_T > 100$  GeV for unconverted (converted) photons. With an opti-  
 1791 mised neural network this efficiency increases from 85–90% at  $E_T = 20$  GeV to  
 1792 about 97% (99%) at  $E_T > 100$  GeV for unconverted (converted) photon candidates  
 1793 for a similar background rejection. For the data taken in 2012,  $20.3 \text{ fb}^{-1}$  at  $\sqrt{s} =$   
 1794 8 TeV, the efficiency of a re-optimised cut-based photon identification algorithm

1795 increases from 50–65% (45–55%) for unconverted (converted) photons at  $E_T = 10$   
 1796 GeV to 95–100% at  $E_T > 100$  GeV, being larger than 90% for  $E_T > 40$  GeV [32].

Category	Description	Name	Loose	Tight
Acceptance	$ \eta  < 2.37$ , $1.37 <  \eta  < 1.52$ excluded	–	✓	✓
Hadronic leakage	Ratio of $E_T$ in the first sampling of the hadronic calorimeter to $E_T$ of the EM cluster (used over the range $ \eta  < 0.8$ and $ \eta  > 1.37$ )	$R_{had_1}$	✓	✓
	Ratio of $E_T$ in all the hadronic calorimeter to $E_T$ of the EM cluster (used over the range $0.8 <  \eta  < 1.37$ )	$R_{had}$	✓	✓
EM Middle layer	Ratio in $\eta$ of cell energies in $3 \times 7$ versus $7 \times 7$ cells	$R_\eta$	✓	✓
	Lateral width of the shower	$\omega_{\eta_2}$	✓	✓
	Ratio in $\phi$ of cell energies in $3 \times 3$ and $3 \times 7$ cells	$R_\phi$		✓
EM Strip layer	Shower width for three strips around maximum strip	$\omega_{s3}$		✓
	Total lateral shower width	$\omega_{tot}$		✓
	Fraction of energy outside core of three central strips but within seven strips	$F_{side}$		✓
	Difference between the energy associated with the second maximum in the strip layer, and the energy reconstructed in the strip with the minimal value found between the first and second maxima	$\Delta E$		✓
	Ratio of the energy difference associated with the largest and second largest energy deposits over the sum of these energies	$E_{ratio}$		✓

Table 3.2: Variables used for *Loose* and *Tight* photon identification cuts.

### 1797 3.3 Muons

#### 1798 Reconstruction

1799 Muon reconstruction is first performed independently in the ID and MS. The in-  
 1800 formation from individual sub-detectors is then combined to form the muon tracks  
 1801 that are used in physics analyses. Muon track candidates are then built by fit-  
 1802 ting together hits from segments (MDT, RPC, CSC and TGC) in different layers.  
 1803 The algorithm used for this task performs a segment-seeded combinatorial search  
 1804 that starts by using as seeds the segments generated in the middle layers of the



1805 detector where more trigger hits are available. At least two matching segments  
1806 are required to build a track, except in the barrel–endcap transition region where  
1807 a single high-quality segment with  $\eta$  and  $\phi$  information can be used to build a  
1808 track. An overlap removal algorithm selects the best assignment to a single track,  
1809 or allows for the segment to be shared between two tracks. The hits associated  
1810 with each track candidate are fitted using a global  $\chi^2$  fit [33]. A track candidate is  
1811 accepted if the  $\chi^2$  of the fit satisfies the selection criteria. The calculation of the  
1812 energy loss in the calorimeter was also improved. An analytic parametrization of  
1813 the average energy loss is derived from a detailed description of the detector geom-  
1814 etry. The final estimate of the energy loss is obtained by combining the analytic  
1815 parametrization with the energy measured in the calorimeter. This method yields  
1816 a precision on the mean energy loss of about 30 MeV for 50 GeV muons.

1817

### 1818 Identification

1819 Muon identification is performed by applying quality requirements that suppress  
1820 background, mainly from pion and kaon decays, while selecting prompt muons  
1821 with high efficiency and/or guaranteeing a robust momentum measurement. Sev-  
1822 eral variables offering good discrimination between prompt muons and background  
1823 muon candidates are studied in simulated  $t\bar{t}$  events. Muons from  $W$  decays are  
1824 categorized as signal muons while muon candidates from light-hadron decays are  
1825 categorized as background. Four muon identification selections (*Loose*, *Medium*,  
1826 *Tight*, and *High- $p_T$* ) are provided to address the specific needs of different physics  
1827 analyses. *Loose*, *Medium*, *Tight* are inclusive categories so that muons identified  
1828 with tighter requirements are also included in the looser categories. Four identifi-  
1829 cation selection criteria can separate low ( $4 < p_T < 20$  GeV) and high ( $20 < p_T <$   
1830  $100$  GeV) transverse momentum muon candidates.

1831 Muon reconstruction performance has been measured by using  $3.2 \text{ fb}^{-1}$  of data  
1832 at  $\sqrt{s} = 13$  TeV recorded during the 25 ns run at the LHC in 2015. The muon  
1833 reconstruction efficiency is close to 99% over most of the pseudorapidity range of  
1834  $|\eta| < 2.5$  for  $p_T > 5$  GeV. The  $Z \rightarrow \mu\mu$  sample enables the measurement of the  
1835 efficiency with a precision at the 0.2% level for  $p_T > 20$  GeV. The  $J/\psi \rightarrow \mu\mu$   
1836 sample provides a measurement of the reconstruction efficiency between 5 and 20  
1837 GeV with a precision better than 1%. The  $Z \rightarrow \mu\mu$  sample is also used to measure  
1838 the isolation efficiency for seven isolation working points in the momentum range  
1839 10–120 GeV. The isolation efficiency varies between 93% and 100% depending on  
1840 the selections and on the momentum of the particle, and is well reproduced in  
1841 the simulation. The muon momentum scale and resolution have been studied in  
1842 details using  $J/\psi \rightarrow \mu\mu$  and  $Z \rightarrow \mu\mu$  decays. These studies are used to correct the  
1843 simulation to improve the agreement with data and to minimise the systematic  
1844 uncertainties in physics analyses [33].

## 1845 **3.4 Hadronic taus**

1846 In the ATLAS,  $\tau$  reconstruction and identification concentrates on the hadronic  
1847 decay modes of a  $\tau$  lepton. The majority of  $\tau$  hadronic decays are character-  
1848 ized by one or three charged pions accompanied by neutral pions and neutrinos.  
1849  $\tau$  candidates are seeded by jets formed using the procedure described in Sec-  
1850 tion 3.5. A  $\tau$ -specific energy calibration, baseline calibration or boosted regression  
1851 tree (BRT) [34] is applied to the  $\tau$  candidate in order to correct the energy depo-  
1852 sition measured in the detector to the average value of the energy carried by the  
1853 measured decay products at the generator level.

1854 The  $\tau$  identification algorithm is designed to reject backgrounds from quark-  
1855 initiated and gluon-initiated jets. The identification uses Boosted Decision Tree  
1856 (BDT) based method. Three working points labelled *Loose*, *Medium* and *Tight*  
1857 are provided, and corresponding to different  $\tau$  identification efficiency values, with  
1858 the efficiency designed to be independent of  $p_T$ . The input variables of BDT are  
1859 corrected such that the mean of their distribution for signal samples is constant as  
1860 a function of pile-up. This ensures that the efficiency for each working point does  
1861 not depend strongly on the pile-up conditions.

## 1862 **3.5 Jets**

1863 Jets are essential objects in the ATLAS physics analysis. They play a fundamental  
1864 role in many physics measurements. In the searching of the  $t\bar{t}H$  production analysis  
1865 with the multi-lepton final states, jets can be employed to discriminate the signal  
1866 events from diboson process for instance. Reconstruction of the jets starts with  
1867 3D topological clusters built in the ATLAS calorimeters, which are matched to  
1868 tracks of charged particles measured in the inner detector. The algorithm used is  
1869 anti- $k_t$  [35] and the energy calibration is performed using MC simulations.

1870 The inputs of jet reconstruction in the ATLAS are locally-calibrated 3D topo-  
1871 logical clusters (topo-clusters), built from calorimeter cells. Topo-clustering starts  
1872 by identifying seed cells with energy significance  $4\sigma$  above noise. The noise here is  
1873 defined as the sum in quadrature of electronic and pile-up noise. Neighbour cells  
1874 with energy significance higher than 2 are then added to form the seed clusters.  
1875 An extra ring of direct neighbour cells is added to the final clusters. After topo-  
1876 clusters finding, a splitting algorithm is used to further separate clusters, based on  
1877 local energy maxima within clusters. Individual clusters are calibrated using local  
1878 properties such as energy density, calorimeter depth, and isolation with respect to  
1879 nearby clusters. This local cluster weighting calibration (LCW) allows clusters to  
1880 be classified as electromagnetic or hadronic and uses a dedicated cluster calibration  
1881 derived from single pion Monte Carlo simulations. Jets are built using the anti- $k_t$   
1882 algorithm with radius parameters  $R=0.4$ . Jets are calibrated to the particle level  
1883 in dijet events: a Monte Carlo pile-up offset correction, a Monte Carlo jet energy

1884 response correction, and a residual calibration are applied to jets in data only, to  
1885 account for the differences in response between data and Monte Carlo.

1886 Main background jets come from proton beam collisions with the residual gas in  
1887 the beam pipe, interactions in the tertiary collimators, cosmic muons overlapping  
1888 in-time with collision events, calorimeter noise, etc. Four selections are defined  
1889 with different levels of fake-jets rejection: *Looser*, *Loose*, *Medium* and *Tight*. The  
1890 most loose one has the highest jet efficiency while the tight one has the highest  
1891 background rejection. Their definitions are based on the reconstructed energy at  
1892 the cell level, on the jet energy deposited in the direction of the shower develop-  
1893 ment and on the number of reconstructed tracks matched to the jets. The pile-up  
1894 offset correction aims at subtracting the extra energy added to jet by additional  
1895  $pp$  interactions overlapping with the physics events of interest. Another effect of  
1896 pile-up is to generate the additional fake jets. Such fake jets originating from pile-  
1897 up fluctuations after the application of the offset correction are rejected by using  
1898 the Jet Vertex Fraction (JVF) algorithm [36]. JVF calculates the fraction of total  
1899 track  $p_T$  matched to a jet that originates from the hard scatter vertex. Pile-up  
1900 jets have very small JVF values as most of their tracks originate from additional  
1901 pile-up vertices.

1902

### 1903 **$b$ -jets tagging**

1904 The identification of jets containing  $b$  hadrons is an important tool in precision  
1905 measurements in the top quark sector as well as in the search for the Higgs boson  
1906 and new phenomena, the suppression of background processes that contain pre-  
1907 dominantly light-flavour jets using  $b$ -tagging is of great use. It may also become  
1908 critical to achieve an understanding of the flavour structure of any new physics  
1909 (e.g. supersymmetry) revealed at the LHC. An illustration of the production  
1910 of a  $b$ -jet is shown in Figure 3.2. The decay of a  $b$  hadron is suppressed by a  
1911 Cabibbo–Kobayashi–Maskawa (CKM) [37] factor which results in a longer flight  
1912 path before decaying in the inner detector compared to charm and light hadrons.

1913 The rate at which a true  $b$ -jet is identified defines the  $b$ -tagging efficiency for a  
1914 particular  $b$ -tagging algorithm. The true flavour of a jet is defined in simulated  
1915 data using a spatial  $\Delta R$  matching between stable hadrons and reconstructed jets.  
1916 A hierarchy matching is performed, first checking whether a  $b$  hadron can be  
1917 matched, followed by a charm hadron and followed by a  $\tau$ -lepton. This matching  
1918 procedure results in a jet being classified respectively as either a  $b$ -jet,  $c$ -jet,  $\tau$ -jet  
1919 or a light-jet.

1920 There are three main categories of  $b$ -tagging algorithms commonly used in the  
1921 ATLAS. There are impact parameter based algorithms (IP2D, IP3D) [38]; inclusive  
1922 secondary vertex reconstruction algorithms (SV); and decay chain reconstruction  
1923 algorithms (JetFitter) [39]. These algorithms contribute complementary infor-  
1924 mation and can be combined by using a multivariate function to create a single  
1925  $b$ -tagging discriminant (MV).

1926 The three types of algorithms provide input to a multivariate classifier. In Run 2

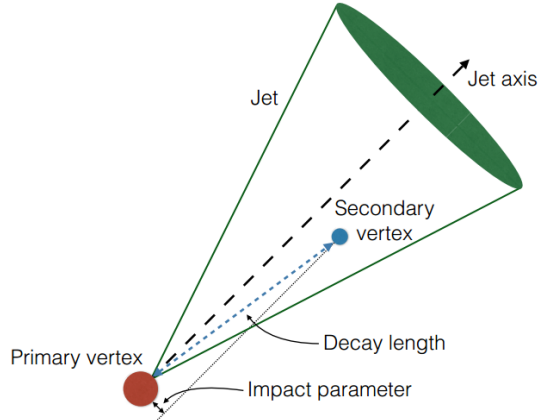


Figure 3.2: Illustration of the production of a  $b$ -tagged jet.

1927 this classifier is a boosted decision tree (BDT) trained to discriminate  $b$ -jets from  
 1928 light-jets. The background samples of light-jets can contain an admixture of charm  
 1929 and light-jets in order to improve the charm-jet rejection. The discrimination  
 1930 performance is shown in Figure 3.3(a) for the MV2 (MV2c10) algorithm with a  
 1931 10% charm-jet admixture. The performance between the Run 1 MV algorithm  
 1932 (MV1) and the Run 2 MV algorithm (MV2) has seen an improvement of 10%,  
 1933 primarily driven by the inclusion of the IBL. Between 2015 and 2016, retraining  
 1934 of the  $b$ -tagging classifier has further improved the charm-jet rejection by around  
 1935 40% at the 77% working point. The charm-jet rejection rate as a function of  $b$ -  
 1936 tagging efficiency is shown in Figure 3.3(b). Jets which have been selected using a  
 1937 top-pair event selection are shown in Figure 3.3(c) comparing the performance of  
 1938 the  $b$ -tagging weight (MV2c10) in simulation and data, where a good agreement  
 1939 can be observed [40].

1940 Benchmark values obtained with MV2 algorithm are shown in Table 3.3 [41].  
 1941 Several working points (WP) are shown together with the  $b$ -tagging efficiency,  
 1942 purity,  $c$  quark and light flavor quark rejection factor (RF) together with the tau  
 1943 lepton rejection efficiency. The 70% working point of MV2c10 is chosen in this  
 1944 thesis for the following section when  $b$ -tagged jet efficiency mentioned.

WP[%]	$b$ jet efficiency[%]	Purity[%]	$c$ RF	$\tau$ RF	light jets RF
60	60.03	99.00	34.54	183.98	1538.78
70	69.97	97.46	12.17	54.72	381.32
77	76.97	95.17	6.21	22.04	134.34

Table 3.3:  $b$ -tagging benchmarks of MV2 algorithm.

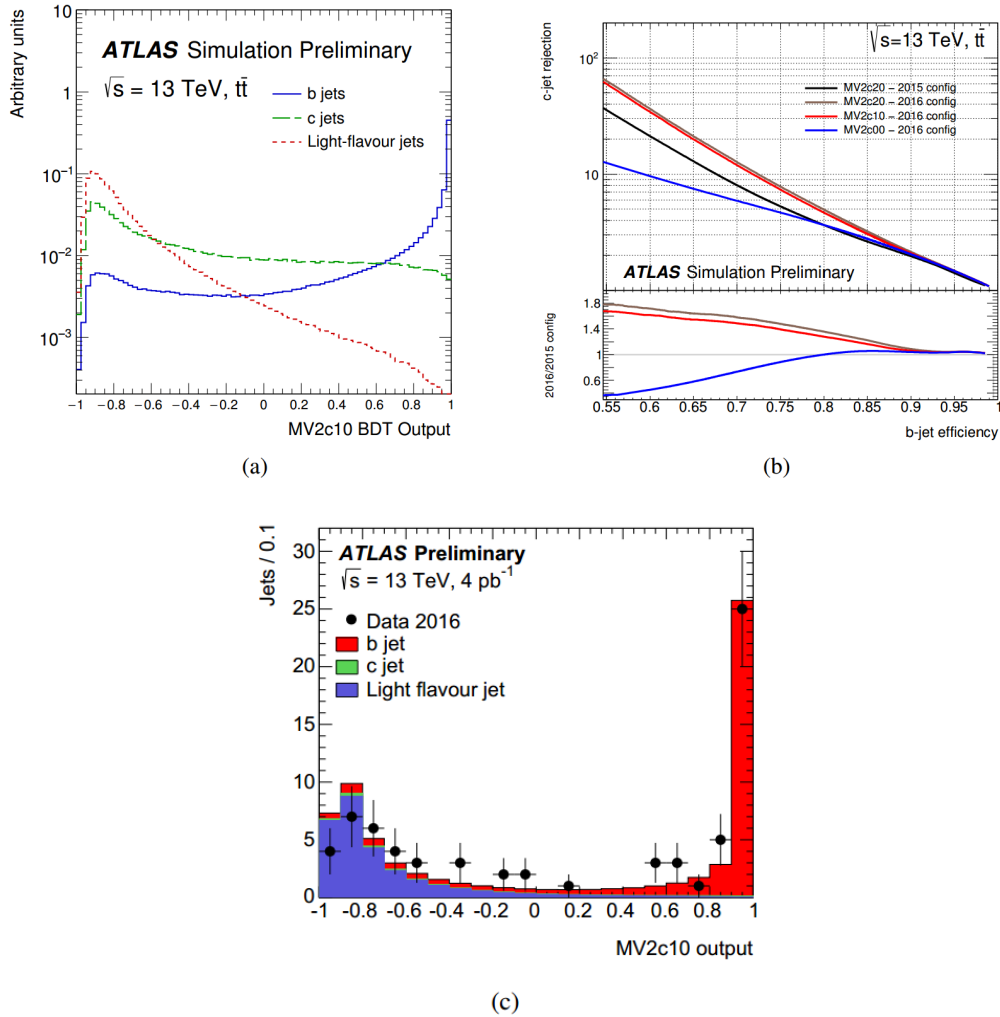


Figure 3.3: (a): The performance of the MV2c10 algorithm is shown for different jet flavours [40], (b): a comparison between the MV2c algorithms trained in 2015 and trained in 2016 [40] and (c): the performance of the MV2c10 algorithm compared to 2016 data [42].

### 1945 3.6 Missing transverse energy

1946 The Missing Transverse Energy (MET) in the ATLAS is reconstructed from cells  
 1947 belonging to topo-clusters and from reconstructed muons. Cells in topo-clusters  
 1948 are calibrated and the calibration of all physics objects in each final state is also  
 1949 propagated to the MET. The soft term of the MET, which consists of topo-clusters  
 1950 not belonging to any reconstructed physics object, is corrected for the effect of pile-  
 1951 up using a track-based technique. The Soft Term Vertex Fraction (STVF) [43] is  
 1952 defined as the ratio of the sum of  $p_T$  of all tracks unmatched to jets from the

1953 hard-scatter vertex and all tracks unmatched to jets from all vertices in a given  
 1954 event. The soft term of the MET is then rescaled by STVF, event-by-event. The  
 1955 MET performance and systematic uncertainties are established from differences  
 1956 between data and simulations of the MET distribution in  $Z \rightarrow ll$  and  $W \rightarrow l\nu$   
 1957 events. The evolution of the  $E_T^{miss}$  resolution is shown for different numbers of jets  
 1958 in Figure 3.4 with the TST  $E_T^{miss}$  algorithm as a representative example [44].

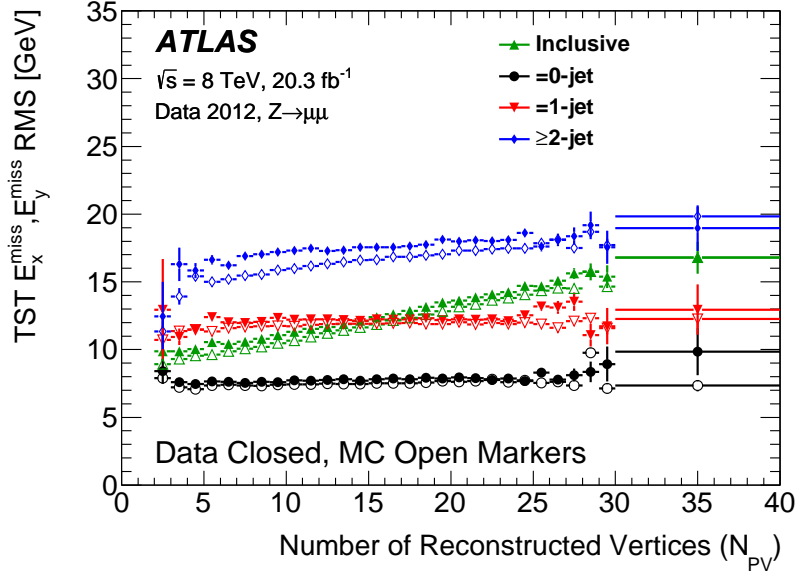


Figure 3.4: The resolution of the combined distribution of  $E_x^{miss}$  and  $E_y^{miss}$  for the TST  $E_T^{miss}$  as a function of NPV for the 0-jet, 1-jet,  $\geq 2$ -jet, and inclusive  $Z \rightarrow \mu\mu$  samples. The data (closed markers) and MC simulation (open markers) are overlaid. The jet counting uses the same JVF criterion as the TST  $E_T^{miss}$  reconstruction algorithm.

# 1959 4 Search for the production of the 1960 Higgs boson associated with a 1961 pair of top quarks

## 1962 4.1 Introduction

1963 The Higgs boson production associated with a pair of top quarks is one of the  
1964 most important measurements in Run 2 after Higgs discovery. Details of the  $t\bar{t}H$   
1965 production and results with Run 1 data have been mentioned in Chapter 1.6. The  
1966 search for  $t\bar{t}H$  production with Run 2 data will be presented in this chapter.

1967 The  $t\bar{t}H$  production can be observed through different topologies according to  
1968 the Higgs and top quark decays. The analysis presented in this thesis is to search  
1969 for Higgs decays to  $WW^*$ ,  $\tau\tau$ ,  $ZZ^*$ . The top quark and anti-top quark decays  
1970 to  $W^\pm b$ . Especially for Higgs decays to  $WW^*$ , each  $W^\pm$  boson in the final state  
1971 decays either leptonically ( $\ell = e^\pm, \mu^\pm, \tau^\pm$ ) with missing energy or hadronically,  
1972 leading to many topologies. Depending on the number of leptons in the final state  
1973 of the signal events, topologies could be the  $2\ell SS$  (exactly two light leptons with  
1974 same sign charges and with hadronic tau veto);  $3\ell$  (exactly three light leptons and  
1975 with hadronic tau veto);  $4\ell$  (exactly four light leptons), etc. The search for  $t\bar{t}H$   
1976 in the exactly three leptons final state is presented in thesis as it's my main work  
1977 during the PhD study.

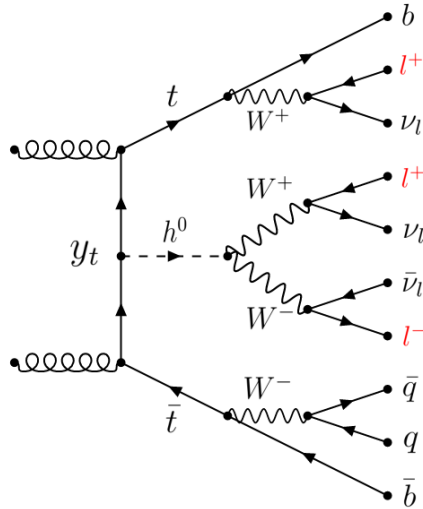


Figure 4.1: Feynman diagram of the  $t\bar{t}H$  signature with 3 leptons final state.

1978 The feynman diagram for the final state of  $3\ell$  with Higgs decays to  $WW$  for  
 1979 instance is shown in Figure 4.1. The Higgs decays to  $WW^*$  channel makes the  
 1980 dominant contribution compared to others as shown in Table 4.1. The channels  
 1981 associated with hadronic  $\tau$  is considered separately, denoted  $\tau$  channel in  $t\bar{t}H$   
 1982 searching. Object definition and the selections used in analysis are mentioned in  
 1983 the following sections.

	$3\ell$	$4\ell$	$2\ell SS$
$H \rightarrow WW^*$	74%	72%	77%
$H \rightarrow ZZ^*$	4%	9%	3%
$H \rightarrow \tau\tau$	20%	18%	17%
$H \rightarrow \text{others}$	2%	2%	3%

Table 4.1: Contributions from different Higgs decay modes for the  $2\ell SS$ ,  $3\ell$  and  $4\ell$  channel.

## 1984 4.2 Data and Monte Carlo samples

### 1985 4.2.1 Collision data

1986 The analysis presented in this thesis uses  $36.1 \text{ fb}^{-1}$  of data collected from proton-  
 1987 proton collision recorded by the ATLAS detector at  $\sqrt{s} = 13 \text{ TeV}$  during 2015 and  
 1988 2016. The dataset has been collected with a bunch crossing of 25 ns and verified  
 1989 data quality cuts namely which must be in the recommended Good Run List.



1990 The data is prepared with `xAOD` format and further produced to `DxAOD` format  
 1991 using a dedicated derivation framework. This `xAOD` to `DxAOD` derivation provides  
 1992 a reduction specifically for the  $t\bar{t}H$  events with multileptons in the final states.  
 1993 The total size of dataset is reduced to 3.6% for simulated  $t\bar{t}$  events and 0.1%  
 1994 for collision dataset. These reductions come from the effects of smart slimming  
 1995 (remove un-needed variables), thinning (remove entire objects from events) and  
 1996 additional skimming for collision dataset only. The skimming means removing the  
 1997 whole event if it is rejected by any of the following selections:

- 1998 • at least two light leptons passing loose identification criteria with leading  
 1999 lepton  $p_T > 15$  GeV and subleading lepton  $p_T > 5$  GeV, and  $|\eta| < 2.6$ .
- 2000 • at least one light lepton passing loose identification criteria with  $p_T > 15$   
 2001 GeV and  $|\eta| < 2.6$ , and at least two hadronic taus. The  $\tau$  jet has to pass  
 2002 `JetBDTSigLoose` requirement with  $p_T > 15$  GeV, and its charge must be one  
 2003 and has one or three associated tracks.

## 2004 4.2.2 Monte Carlo samples

2005 Event generator programs and configurations used for simulating the signal and  
 2006 background processes are summarized in Table 4.2. The Higgs mass is 125 GeV  
 2007 and this is assumed for all.

2008 Productions of  $t\bar{t}H$ ,  $t\bar{t}W$  and  $t\bar{t}Z$  are generated with a next-to-leading order  
 2009 (NLO) QCD matrix element computed by `MADGRAPH5_AMC@NLO`, interfaced  
 2010 to `Pythia8` for parton showering and fragmentation into particles. In the case of  
 2011  $t\bar{t}Z$ , the inclusive  $t\bar{t}\ell\ell$  matrix element is computed, including off-shell  $Z$  and  $\gamma^*$   
 2012 contributions with  $m(\ell\ell) > 5$  GeV.

2013 The overall  $t\bar{t}H$  cross section is 507.1 fb, which is computed at NLO in QCD [51,  
 2014 52, 53, 54, 55, 56, 57, 58, 59]. It has uncertainties of  $^{+5.8\%}_{-9.2\%}$  from QCD re-  
 2015 normalization/factorization scale choice and  $\pm 3.6\%$  from parton distribution func-  
 2016 tion uncertainties (including  $\alpha_s$  uncertainties).

2017 The cross sections for  $t\bar{t}V$  production, including the process  $pp \rightarrow t\bar{t}\ell^+\ell^- + X$   
 2018 over the full  $Z/\gamma^*$  mass spectrum, are computed at NLO in QCD and electroweak  
 2019 couplings. These have QCD scale uncertainties  $\approx 12\%$  and PDF+ $\alpha_s$  uncertainties  
 2020 of 3 – 4%. The total cross section used for  $t\bar{t}\ell^+\ell^-$  (with  $M(\ell^+\ell^-) > 5$  GeV) is  
 2021 123.7 fb, and 600.8 fb for  $t\bar{t}W^\pm$  [60, 57].

2022  $t\bar{t}$  events are generated with `Powheg v2.0` and interfaced with `Pythia 8` for the  
 2023 parton showering and fragmentation. A14 tune is used for showering. The overall  
 2024  $t\bar{t}$  cross section is 832 pb. `Powheg` is also used to model other top backgrounds  
 2025 such as single top t-channel, s-channel and Wt-channel.

2026 Diboson processes are generated with `Sherpa v2.1.1` at leading order, using CT10  
 2027 PDF set. The matrix elements consider for the fully leptonic diboson production  
 2028 with up to three additional partons. Processes split into  $b$ -filtered and  $b$ -vetoed

Process	ME Generator	Parton Shower	PDF	Tune
$t\bar{t}H$	MG5_AMC	PYTHIA 8	NNPDF 3.0 NLO/2.3 LO	A14[45]
$tHqb$	MG5_AMC	PYTHIA 8	CT10	A14
$tHW$	MG5_AMC	HERWIG++	CT10/CTEQ6L1	UE-EE-5[46]
$t\bar{t}W$	MG5_AMC	PYTHIA 8	NNPDF 3.0 NLO/2.3 LO	A14
$t\bar{t}(Z/\gamma^*)$	MG5_AMC	PYTHIA 8	NNPDF 3.0 NLO/2.3 LO	A14
$t(Z/\gamma^*)$	MG5_AMC	PYTHIA 6	CTEQ6L1	Perugia2012
$tW(Z/\gamma^*)$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}t, t\bar{t}\bar{t}$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}$	POWHEG-BOX	PYTHIA 8	CT10/CTEQ6L1	Perugia2012
$t\bar{t}\gamma$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$s-, t\text{-channel},$ $Wt$ single top	POWHEG-BOX[47, 48]	PYTHIA 6	CT10/CTEQ6L1	Perugia2012
$VV, qqVV, VVV$	SHERPA 2.1.1[49]	SHERPA	CT10	SHERPA
$Z \rightarrow \ell^+\ell^-$	SHERPA 2.2	SHERPA	NNPDF 3.0 NNLO	SHERPA
$W \rightarrow \ell\nu$	SHERPA 2.2	SHERPA	NNPDF 3.0 NNLO	SHERPA

Table 4.2: Configurations of event generations used for signal and background processes. If only one parton distribution function (PDF) is shown, the same one is used for both the matrix element (ME) and parton shower generators; if two are shown, the first is used for the matrix element calculation and the second for the parton shower. "V" refers to an electroweak boson ( $W$  or  $Z/\gamma^*$ ). "Tune" refers to the underlying-event tune of the parton shower generator. "MG5\_AMC" refers to Mad-Graph5\_AMC@NLO 2.2.1; "Pythia 6" refers to version 6.427; "Pythia 8" refers to version 8.2; "Herwig++" refers to version 2.7. Samples using Pythia 6 and Pythia 8 have heavy flavor hadron decays modelled by EvtGen 1.2.0. All samples include leading-logarithm photon emission, either modelled by the parton shower generator or by PHOTONS [50].

2029 samples, according to having or not truth jets containing a  $b$  hadron with transverse  
2030 momentum above 5 GeV.

2031 All Monte Carlo samples are processed through a complete simulation of the  
2032 ATLAS detector response based on Geant4. Additional simulated  $pp$  collisions  
2033 generated with Pythia 8 were overlaid to model the effects of both in- and out-of-  
2034 time pile-up, from additional  $pp$  collisions in the same and nearby bunch crossings.  
2035 The pile-up distribution is reweighed to reflect the mean number of additional  
2036 interactions observed in data. All simulated events were processed using the same  
2037 reconstruction algorithms and analysis chain as the data. Simulated events are  
2038 corrected so that the object reconstruction and identification efficiencies, energy  
2039 scales and energy resolutions match those determined from data.

## 4.3 Object definition and basic selection

Many objects like leptons, jets, neutrinos are presented in the  $3\ell$  final state as shown in the Feynman diagram 4.1. The details of the reconstruction of the common objects are described in Chapter 3. The optimizations on those are done to improve the signal acceptance and signal significance preliminarily. Following sections introduce the detailed criterion on different objects. Firstly at event level, the primary vertex with requirement of the vertex with largest  $\sum p_T^2$  of associated tracks with  $p_T > 400$  MeV is required in an event. Events with significant noise in the calorimeters or data corruption which is due to software and hardware failures are removed.

### 4.3.1 Trigger

Because of the different running conditions between 2015 and 2016 (more complex pile-up), the trigger is also different for data collected in each year. Generally, single lepton trigger (single electron, single muon) or dilepton trigger ( $ee$ ,  $e\mu$ ,  $\mu\mu$ ) is applied to the events. Lowest  $p_T$  threshold, unrescaled single lepton (SLT) and dilepton (DLT) trigger chains used in the current trigger of analysis are presented in Table 4.3<sup>a</sup> and Table 4.4 for 2015 and 2016 respectively.

Single lepton triggers (2015)	
$\mu$	HLT_mu20_iloose_L1MU15, HLT_mu50
$e$	HLT_e24_lhmedium_L1EM20VH, HLT_e60_lhmedium, HLT_e120_lhloose
Dilepton triggers (2015)	
$\mu\mu$	HLT_mu18_mu8noL1
$ee$	HLT_2e12_lhloose_L12EM10VH
$e\mu, \mu e$	HLT_e17_lhloose_mu14

Table 4.3: List of lowest  $p_T$ -threshold, unrescaled triggers used for the whole 2015 data taking.

Several studies showed that this logical combination of triggers provides the largest gain in signal acceptance of the same sign lepton pair, allowing the minimum lepton  $p_T$  threshold of the trigger-matched leading lepton to be lower than the allowed by the single lepton triggers alone. From the fake estimate side, no trigger bias originating from the online lepton trigger selection is observed to significantly

<sup>a</sup>The "HLT" means the high level trigger which consists of the Level 2 (LVL2) and the Event Filter (EF) trigger. The "mu" is muon and "e" stands for electron. The number after the muon or electron is the cut on the corresponding  $p_T$ . The "L1" means the L1 trigger level. The "iloose" is the *Loose* working point of the muon identification and "lhloose", "lhmedium" are the *Loose*, *Medium* working point of the LH electron identification. The "nod0" means there is no  $d0$  requirement which is shown in Table 4.3. The "VH" means the hadronic veto requirement.

Single lepton triggers (2016)	
$\mu$	HLT_mu26_ivarmedium, HLT_mu50
$e$	HLT_e26_lhtight_nod0_ivarlose, HLT_e60_lhmedium_nod0, HLT_e140_lhlose_nod0
Dilepton triggers (2016)	
$\mu\mu$	HLT_mu22_mu8noL1
$ee$	HLT_2e17_lhvlose_nod0
$e\mu, \mu e$	HLT_e17_lhlose_nod0_mu14

Table 4.4: List of lowest  $p_T$ -threshold, unprescaled triggers used for the whole 2016 data taking.

2062 affect the Matrix Method results when using an OR of single lepton and dilepton  
2063 triggers, hence the proposed choice does not require any additional complexity in  
2064 the fake estimate procedure. As for data-driven trigger efficiency corrections to  
2065 be applied to simulated samples, a new tool developed which can provide a per-  
2066 event correction for the chosen combination of triggers and a given offline lepton  
2067 selection is used.

### 2068 4.3.2 Electrons

2069 The electron objects used in thesis are with the requirements as follows. The can-  
2070 didates are with  $p_T > 10$  GeV. They are required to satisfy  $|\eta_{cluster}| < 2.47$ . The  
2071 electron candidates which are in the transition region between different electro-  
2072 magnetic calorimeter components,  $1.37 < |\eta_{cluster}| < 1.52$ , are rejected. The *Loose*  
2073 working point of the likelihood electron identification is employed in the object  
2074 preselection. No isolation requirement is applied at the preselection level as to  
2075 increase the statistics of the fake leptons estimate with data driven method. The  
2076 non-prompt electrons are further reduced with the track requirement of being con-  
2077 sistent with originating from the primary vertex and the requirements (shown in  
2078 Table 4.5) are imposed on the transverse impact parameter significance ( $|d_0|/\sigma_{d0}$ )  
2079 and the longitudinal impact parameter ( $|\Delta_{Z0} \sin \theta_l|$ ).

### 2080 4.3.3 Muons

2081 In the region  $|\eta| < 0.1$ , where muon spectrometer coverage is reduced, muons are  
2082 also reconstructed from the inner detector tracks matched to the isolated energy  
2083 deposits in the calorimeter consistent with the passage of a minimum-ionizing  
2084 particle. Candidates are required to satisfy  $p_T > 10$  GeV and  $|\eta| < 2.5$ . The  
2085 transverse impact parameter is slightly tighter than the electrons, while the lon-  
2086 gitudinal impact parameter selection is the same. No isolation is required in the

2087 object preselection.

2088 All the requirements above on lepton candidates are grouped and called Loose  
 2089 level objects. The Tight level objects should satisfy the Tight LH working point for  
 2090 electrons at some cases to reduce the fake and non-prompt electrons. A dedicated  
 2091 isolation variable, `PromptLeptonIso` based on Boosted Decision Tree algorithm  
 2092 is applied to improve the reduction of the background of non-prompt produced  
 2093 in hadron decays. The detailed requirements for Loose and Tight leptons are  
 2094 summarised in Table 4.5.

	Loose		Tight	
	$e$	$\mu$	$e$	$\mu$
Non-prompt lepton BDT and loose isolation	-	-	Yes	Yes
Identification	Loose	Loose	Tight	Loose
Charge misID veto BDT	-	-	Yes ( $e^\pm e^\pm$ and $e^\pm \mu^\pm$ )	-
Transverse impact parameter significance $ d_0 /\sigma_{d_0}$	$< 5$	$< 3$	$< 5$	$< 3$
Longitudinal impact parameter $ \Delta z_0 \sin \theta_\ell $ (mm)	$< 0.5$	$< 0.5$	$< 0.5$	$< 0.5$

Table 4.5: Loose and Tight cut definitions for leptons. Selections for tight leptons are applied on top of the selections for loose leptons. IsoLoose is an additional minimum "99% eff" calorimeter and track isolation requirement on top of the `PromptLeptonIso`, and refers to isolation working points designed to be 99% efficient for isolated leptons at all  $p_T$  range. The cut on the charge misID is applied in  $e^\pm e^\pm$  and  $e^\pm \mu^\pm$  channels when doing the fake estimation.

### 2095 4.3.4 Jets and $b$ -tagged jets

2096 Jets are reconstructed from calibrated topological clusters built from energy de-  
 2097 posits in the calorimeters, using the anti- $k_t$  algorithm with a radius parameter  
 2098  $R = 0.4$  which are described in Section 3.5. Jets with energy contributions arising  
 2099 from noise or detector effects are removed, and only jets satisfying  $p_T > 25$  GeV  
 2100 and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ ,  
 2101 a jet-track association algorithm is used to confirm that they originates from the  
 2102 selected primary vertex, in order to reject jets arising from pile-up collisions. The  
 2103 average efficiency of this association is 92% per jet.

2104 Jets containing  $b$ -hadrons are identified ( $b$ -tagging) via a multivariate discrim-  
 2105 inant method. The working point used in this search corresponds to an average  
 2106 efficiency of 70% (explained in Section 3.5) for  $b$ -tagged jets with  $p_T > 20$  GeV

2107 and  $|\eta| < 2.5$ . The expected rejection factors against light and  $c$ -jets are 380 and  
 2108 12, respectively.

### 2109 4.3.5 Overlap removal

2110 A procedure called overlap removal is designed to avoid double counting objects  
 2111 and to remove leptons originating from hadron decays. The overlap removal is done  
 2112 in the following order: any electron candidate within  $\Delta R = 0.1$  of another electron  
 2113 candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$  of  
 2114 a muon candidate is removed; any jet within  $\Delta R = 0.3$  of an electron candidate  
 2115 is removed; if a muon candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10$   
 2116  $[\text{GeV}]/p_T(\text{muon}))$  of each other, the jet is kept and the muon is removed; any  $\tau_{had}$   
 2117 candidate within  $\Delta R = 0.2$  of an electron or a muon candidate is removed; any  
 2118 jet within  $\Delta R = 0.3$  of a  $\tau_{had}$  is removed in events with two light leptons. The full  
 2119 overlap removal procedure are listed in Table 4.6.

Keep	Remove	Cone size ( $\Delta R$ )
electron	electron (low $p_T$ )	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	$\tau$	0.2
muon	$\tau$	0.2
$\tau$	jet	0.3

Table 4.6: Summary of the overlap removal procedure between electrons, muons, hadronically decaying  $\tau$ s, and jets.

### 2120 4.3.6 Event pre-selection (The $3\ell$ pre-MVA selection)

2121 Depending on the number of leptons in the final state,  $t\bar{t}H$  multi-lepton channel  
 2122 is complex, like the  $2\ell SS$ ,  $3\ell$ ,  $4\ell$ ,  $2\ell SS1\tau$ , etc. In this thesis, only the three leptons  
 2123 final state is considered. The orthogonality of events entering various channels  
 2124 is ensured in the definition of the channels. In the three leptons final state, the  
 2125 total charge of the three leptons in the events must be  $\pm 1$ , no  $\tau$  candidate is  
 2126 allowed. Among these three leptons, the lepton with opposite charge compared  
 2127 to other two is called "lepton 0" (Lep 0). The one in the same sign pair with  
 2128 the smallest  $\Delta R(\ell, \ell_0)$ , which means the one has closest distance to the "lepton  
 2129 0", is designed to be "lepton 1" (Lep 1) and remaining one is "lepton 2" (Lep 2).  
 2130 Besides the Tight requirement listed in Table 4.5 on "lepton 1" and "lepton 2", the  
 2131  $p_T > 15$  GeV for both leptons is also required. As "lepton 0" is always the prompt  
 2132 one, the additional requirement is looser than "lepton 1" and "lepton 2", such as  
 2133  $p_T > 10$  GeV, *Loose* isolation and *Loose* lepton identification.

---



---

The $3\ell$ pre-MVA selection
Three light leptons with $p_T > 10$ GeV; sum of light lepton charges $\pm 1$
Two same-charge leptons must be Tight and have $p_T > 15$ GeV
The opposite-charge lepton must be loose, isolated and pass the non-prompt BDT
Zero medium $\tau_{had}$ candidates
$m_{\ell^+\ell^-} > 12$ GeV and $ m_{\ell^+\ell^-} - 91.2 \text{ GeV}  > 10$ GeV for all SFOC pairs
$ m_{\ell\ell\ell} - 91.2 \text{ GeV}  > 10$ GeV
$N_{jets} \geq 2$ and $N_{b-jets} \geq 1$

---



---

Table 4.7: The  $3\ell$  pre-MVA selection criteria applied in analysis. Same-flavour, opposite-charge lepton pairs are referred to as SFOC pairs.

2134 To remove the leptons from quarkonium decays, all same flavour  $\ell^+\ell^-$  pairs  
2135 must satisfy  $m_{\ell^+\ell^-} > 12$  GeV. The potential backgrounds with  $Z$  decays to  $\ell\ell\gamma^{(*)}$   
2136 or  $\ell\ell\ell'(\ell')$ , where one lepton has very low momentum and is not reconstructed, the  
2137 three lepton invariant mass should satisfy  $|m_{\ell\ell\ell} - 91.2 \text{ GeV}| > 10$  GeV. The jets  
2138 and  $b$ -tagged jets requirement is at least 2 jets and at least 1  $b$ -tagged jets.

2139 An additional cut to reject the  $t\bar{t}Z$  process is used in the training stage of  
2140 the  $3\ell$  analysis, where all same flavour lepton pairs in the event should satisfy  
2141  $|m_{\ell^+\ell^-} - 91.2 \text{ GeV}| > 10$  GeV. There is no need to consider the charge flip issue  
2142 in the  $3\ell$  as the study shows that no charge flip events exist in this channel after  
2143 basic selection.

2144 As the selections are applied before the Multi-variate analysis stage, it's also  
2145 called the  $3\ell$  pre-MVA selection. A summary of the  $3\ell$  pre-MVA selection is shown  
2146 in Table 4.7. The expected event yields with the  $3\ell$  pre-MVA selection for different  
2147 processes are shown in Table 4.8.

Process	Expected yields
$t\bar{t}H$	$22.5 \pm 0.4$
$t\bar{t}W$	$42.9 \pm 0.6$
$t\bar{t}Z$	$40.4 \pm 0.5$
VV	$19.7 \pm 2.4$
$t\bar{t}$	$52.4 \pm 7.2$
Rare	$22.1 \pm 0.6$
Total	$200.0 \pm 7.7$

Table 4.8: The expected event yields for signal and background processes after the  $3\ell$  pre-MVA selection.

## 2148 4.4 Background estimation

2149 Several physics processes can produce the similar signal signature. Good estima-  
 2150 tion of those backgrounds is the key issue for this analysis. The background in  
 2151 this analysis are sorted into two categories:

- 2152 • **Irreducible background**

2153 Events which can lead to the same final state as the signal and the final state  
 2154 do indeed contain the three prompt charged leptons.  $t\bar{t}W$ ,  $t\bar{t}Z$  and di-boson  
 2155 processes are in this category.

- 2156 • **Reducible background**

2157 Mostly events in which a non-prompt lepton or fake lepton is selected as  
 2158 prompt lepton are called the reducible background. These processes can lead  
 2159 to a final state which is compatible with signal with the misreconstructed  
 2160 object. The main process are  $t\bar{t}$  and  $Z$ +jets. The data-driven method (Matrix  
 2161 Method) is used to estimate the reducible background in Section 4.4.2.

2162 As a reference, Table 4.9 shows the expected and observed event yields in the  
 2163  $3\ell$  channel at 8 TeV using the full Run 1 20.3 fb<sup>-1</sup> dataset [61]. The dominant  
 2164 backgrounds are the non-prompt (fake leptons) and  $t\bar{t}V$  process. The contribution  
 2165 from the  $VV$  is small. Although many conditions (energy scale, signal region  
 2166 definition, etc.) are different between Run 1 and Run 2, a brief picture of the  
 2167 background composition can be drawn according to the table.

Category	Non-prompt	$t\bar{t}W$	$t\bar{t}Z$	$VV$	Expected bkg.	$t\bar{t}H$	Observed
$3\ell$	$3.2 \pm 0.7$	$2.3 \pm 0.7$	$3.9 \pm 0.8$	$0.86 \pm 0.55$	$11.4 \pm 2.3$	$2.34 \pm 0.35$	18

Table 4.9: Expected and observed event yields in the  $3\ell$  with the 20.3 fb<sup>-1</sup> dataset at 8 TeV.

### 2168 4.4.1 Fake leptons

2169 Due to the limitation of the detector, the objects are not reconstructed perfectly.  
 2170 As one of the mis-reconstructed objects, fake leptons are common issue in the  
 2171 analysis. The main origins of the fake leptons are: photon conversion, light flavour  
 2172 jets and heavy flavour jets.

2173 Fake electrons from conversion are produced when a photon interacting with  
 2174 detector material and split into an electron-positron pair. When this happens in  
 2175 the inner detector, a track is left and the track will be combined with the EM  
 2176 cluster from the electron and positron, and this may lead a reconstructed electron.  
 2177 In principle, the fake electrons from photon conversion can be removed by requiring  
 2178 the hits in the very first layer of the tracker, but random calorimeter clusters can  
 2179 still be combined with conversion tracks to form fake electrons. According to the



2180 description of jet object reconstruction in Section 3.5, the algorithm employs the  
 2181 track information in the inner detector and energy in the calorimeter. If a jet is  
 2182 reconstructed with a large fraction of EM energy and there is a random track that  
 2183 matches the calorimeter cluster, this would lead to an electron.

2184 The fake muons from photon conversion are expected to be negligible. If a  
 2185 charged hadron with a lifetime sufficient to go through the whole detector, whilst  
 2186 the track of this hadron in the ID might not match that expected from a muon.  
 2187 But there are chances that matches can be found with unrelated hits and a fake  
 2188 muon can be reconstructed in this way.

2189 A dedicated study of the fake lepton origin is performed in  $3\ell$  for both fake  
 2190 electrons and muons by using the truth information with MC. The study of the  
 2191 origin composition of the fake leptons in the  $3\ell$  pre-MVA region shows that the  
 2192 heavy flavour fakes (fake leptons which are from  $b$  and  $c$  meson) is dominant  
 2193 for both fake electrons and fake muons; large contribution comes from photon  
 2194 conversion fakes for fake electrons but no contribution is from this for fake muons  
 2195 as expected.

2196 The data-driven method is introduced in this analysis to have a good estimation  
 2197 of the fake leptons. The following section will discuss the Matrix Method which is  
 2198 used widely in many analyses.

## 2199 4.4.2 The Matrix Method

2200 The Matrix Method (MM) is a data-driven technique to have the estimation of the  
 2201 fake contamination in the analysis. The estimation based on MC samples can not  
 2202 describe the fake objects well as the MC is not reliable all the time. Data-driven  
 2203 technique can employ the information from real collision data and get rid of the  
 2204 dependence on MC. The MM is a widely used data-driven method in the ATLAS.  
 2205 To have a good estimation of the fakes background, the MM is employed in the  
 2206 three leptons final state.

2207 Generally, the MM is used to estimate the fake electrons and muons in a tight  
 2208 (denoted with  $T$ ) region by using the loose object information in an anti-tight  
 2209 (denoted with  $\bar{T}$ ) region. The basic strategy of the MM can be explained with a  
 2210 simplified case where only one lepton is considered. The number of the events with  
 2211 a tight (referred to  $N^T$ ) lepton and the one with a lepton which fails the selection  
 2212 ( $N^{\bar{T}}$ ) can be expressed in the terms of efficiencies and inefficiencies for the baseline  
 2213 loose real ( $N^r$ ) or fake ( $N^f$ ) leptons to pass the tight selection via two equations:

$$\begin{aligned} N^T &= \varepsilon_r N^r + \varepsilon_f N^f, \\ N^{\bar{T}} &= \bar{\varepsilon}_r N^r + \bar{\varepsilon}_f N^f, \end{aligned} \tag{4.1}$$

2214 where  $\varepsilon_r$  ( $\varepsilon_f$ ) represents the efficiency for a real (fake) lepton to pass tight selection,  
 2215 and  $\bar{\varepsilon}_r \equiv (1 - \varepsilon_r)$  ( $\bar{\varepsilon}_f \equiv (1 - \varepsilon_f)$ ) represents the probability for a real (fake) lepton  
 2216 to fail tight but still pass baseline selection, or in a matrix form:

$$\begin{pmatrix} N^T \\ N^{\bar{T}} \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ \bar{\varepsilon}_r & \bar{\varepsilon}_f \end{pmatrix} \begin{pmatrix} N^r \\ N^f \end{pmatrix}. \quad (4.2)$$

2217 Then reversing the matrix can give the relation between the number of real  
 2218 and fake leptons events, given the observed number of tight and anti-tight lepton  
 2219 events:

$$\begin{pmatrix} N^r \\ N^f \end{pmatrix} = \frac{1}{\varepsilon_r - \varepsilon_f} \begin{pmatrix} \bar{\varepsilon}_f & -\varepsilon_f \\ -\bar{\varepsilon}_r & \varepsilon_r \end{pmatrix} \begin{pmatrix} N^T \\ N^{\bar{T}} \end{pmatrix}. \quad (4.3)$$

2220 Thus the number of the fake leptons in the tight region can be obtained:

$$N_f^T = \frac{\varepsilon_r \varepsilon_f}{\varepsilon_r - \varepsilon_f} N^{\bar{T}} + \frac{(-\bar{\varepsilon}_r) \varepsilon_f}{\varepsilon_r - \varepsilon_f} N^T. \quad (4.4)$$

2221 For a simplified example above, fake leptons can be estimated with a  $2 \times 2$   
 2222 matrix. A similar matrix method can be developed in the  $3\ell$  as well but with an  
 2223  $8 \times 8$  matrix. Besides building the complex matrix, the most difficulty for the  $3\ell$   
 2224 is to find a reliable definition of the real lepton and fake lepton enriched control  
 2225 region. This is not as easy as the case in the tight and anti-tight region mentioned  
 2226 above. However, a simplified Matrix Method based on a  $4 \times 4$  matrix (which  
 2227 is already employed in the  $t\bar{t}H$  2 $\ell$ SS final state) can be used and this simplified  
 2228 Matrix Method can solve those difficulties with negligible effect.

2229 To complete this simplified Matrix Method in the  $3\ell$ , a premise is needed: "Lep 0"  
 2230 which is with opposite sign is with a much lower possibility to be the fake lepton.  
 2231  $t\bar{t}$  MC is used to check this premise with the  $3\ell$  pre-MVA selection. Following  
 2232 check is to see the real and fake raw number and corresponding percentage in the  
 2233  $t\bar{t}$  MC. Figure 4.2 shows the possibility of "Lep 0" to be the fake one in the  $3\ell$   
 2234 pre-MVA region with  $t\bar{t}$  MC. Overall only 1% "Lep 0" would be the fake one in the  
 2235  $3\ell$  pre-MVA region.

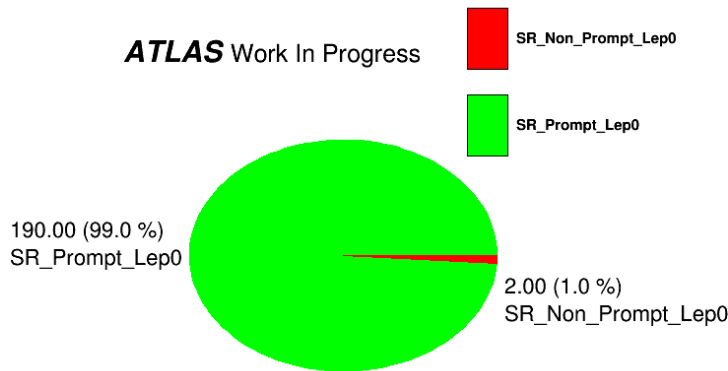


Figure 4.2: The raw number of event and the corresponding percentage for the real and the fake lepton events in  $t\bar{t}$  MC.

2236 Generally, most "Lep 0" are the prompt leptons and very few would be the fake  
 2237 objects for both electron and muon. The situation "Lep 0" to be the fake would be  
 2238 neglected. Thus three leptons final state is much similar to the  $t\bar{t}H$  two leptons  
 2239 with same sign charge channel, and the main different is to estimate the same  
 2240 sign lepton pair, "Lep 1" and "Lep 2" in the  $3\ell$ . The advantages of this simplified  
 2241 Matrix Method in the  $3\ell$  are: the simple Matrix Method which is the same one  
 2242 used in the  $2\ell$ SS; no need to redefine the real and fake enriched control region in  
 2243 the  $3\ell$ ; the real efficiency and fake rate which are used in matrix are the same as  
 2244 the measured results from the  $2\ell$ SS.

2245 Since the assumption of "Lep 0" works well in the  $3\ell$  final states, the same sign  
 2246 lepton pair which pass the baseline selection could be fakes. The real and fake  
 2247 efficiency can be measured by using the two leptons events with the  $2\ell$  selection.  
 2248 Then the Matrix Method is applied to the same sign lepton pair in three leptons  
 2249 events during the application stage with the  $3\ell$  selection.

2250 Depending on whether each lepton pass the Tight requirement (the definition  
 2251 of Loose and Tight is in Table 4.5), each  $i$ -th events can be categorised into any  
 2252 of four orthogonal (*sidebands*) region:

- 2253 •  $TT_i$ : Events with both leptons are under Tight requirement (total events is  
 2254 denoted as  $N^{TT}$ ).
- 2255
- 2256 •  $T\bar{T}_i$ : Events with leading leptons is Tight and subleading lepton is anti-Tight  
 2257 (total events is denoted as  $N^{T\bar{T}}$ ).
- 2258
- 2259 •  $\bar{T}T_i$ : Events with leading lepton is anti-Tight and subleading lepton is Tight  
 2260 (total events is denoted as  $N^{\bar{T}T}$ ).
- 2261

2262 •  $\bar{T}\bar{T}_i$ : Events with both leptons are under anti-Tight requirement (total events  
 2263 is denoted as  $N^{\bar{T}\bar{T}}$ ).  
 2264

2265 The  $4 \times 4$  matrix is defined to map the total number of such events into the  
 2266 total number of events in four di-lepton regions characterized by different real and  
 2267 fake lepton composition :

2268 •  $rr_i$ : Events with both real leptons (total events is denoted as  $N^{rr}$ ).  
 2269

2270 •  $rf_i$ : Events with leading leptons is real and subleading lepton is fake (total  
 2271 events is denoted as  $N^{rf}$ ).  
 2272

2273 •  $fr_i$ : Events with leading lepton is fake and subleading lepton is real (total  
 2274 events is denoted as  $N^{fr}$ ).  
 2275

2276 •  $ff_i$ : Events with both fake leptons (total events is denoted as  $N^{ff}$ ).  
 2277

2278 The matrix is as follows:

$$\begin{pmatrix} N^{TT} \\ N^{T\bar{T}} \\ N^{\bar{T}T} \\ N^{\bar{T}\bar{T}} \end{pmatrix} = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\bar{\varepsilon}_{r,2} & \varepsilon_{r,1}\bar{\varepsilon}_{f,2} & \varepsilon_{f,1}\bar{\varepsilon}_{r,2} & \varepsilon_{f,1}\bar{\varepsilon}_{f,2} \\ \bar{\varepsilon}_{r,1}\varepsilon_{r,2} & \bar{\varepsilon}_{r,1}\varepsilon_{f,2} & \bar{\varepsilon}_{f,1}\varepsilon_{r,2} & \bar{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{f,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2} \end{pmatrix} \begin{pmatrix} N^{rr} \\ N^{rf} \\ N^{fr} \\ N^{ff} \end{pmatrix}, \quad (4.5)$$

2279 where the indexes for  $\varepsilon_r$  and  $\varepsilon_f$  are ordered in terms of lepton  $p_T$  .

2280 The number of fakes in the signal region as a function of observables can be  
 2281 obtained by reverting the matrix:

$$\begin{pmatrix} N^{rr} \\ N^{rf} \\ N^{fr} \\ N^{ff} \end{pmatrix} = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\bar{\varepsilon}_{r,2} & \varepsilon_{r,1}\bar{\varepsilon}_{f,2} & \varepsilon_{f,1}\bar{\varepsilon}_{r,2} & \varepsilon_{f,1}\bar{\varepsilon}_{f,2} \\ \bar{\varepsilon}_{r,1}\varepsilon_{r,2} & \bar{\varepsilon}_{r,1}\varepsilon_{f,2} & \bar{\varepsilon}_{f,1}\varepsilon_{r,2} & \bar{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{f,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2} \end{pmatrix}^{-1} \begin{pmatrix} N^{TT} \\ N^{T\bar{T}} \\ N^{\bar{T}T} \\ N^{\bar{T}\bar{T}} \end{pmatrix}, \quad (4.6)$$

2282 where the final number of fakes in the signal region  $N_{TT}^f$ , namely the total number  
 2283 of  $TT$  events with at least one fake lepton, can be obtained with equation:

$$N_{TT}^f = N_{TT}^{rf} + N_{TT}^{fr} + N_{TT}^{ff} = \varepsilon_{r,1}\varepsilon_{r,2}N^{rf} + \varepsilon_{f,1}\varepsilon_{r,2}N^{fr} + \varepsilon_{f,1}\varepsilon_{f,2}N^{ff}. \quad (4.7)$$

2284 According to the equation above, the main work is to measure the real and fake  
2285 efficiency as the input of the matrix.

2286 This method deals with the  $t\bar{t}$  background because the most signal lepton of  $t\bar{t}$   
2287 are the fakes in the  $t\bar{t}H$   $3\ell$  final state. The simulated  $t\bar{t}$  sample with at least one  
2288  $W$  decaying leptonically is used for the real and fake measurement. A dedicated  
2289  $t\bar{t}$  sample generated with a dileptonic filter is used for the  $3\ell$  case to enhance the  
2290 statistics in the closure test stage. Additional photon radiation, which can give  
2291 rise to extra leptons (mostly electron) through material interactions, is simulated  
2292 inclusively in these samples. In order to improve statistics for those events, an-  
2293 other sample specifically targeting  $t\bar{t}$  production with an additional prompt photon  
2294 radiated by any of the top quarks ( $t\bar{t}\gamma$ ), has been generated and used. An overlap  
2295 removal procedure which is based on distance with respect to any lepton in the  
2296 event is applied to avoid the double counting of events between  $t\bar{t}$  and  $t\bar{t}\gamma$ .

#### 2297 **4.4.2.1 Real and fake lepton efficiency measurement**

2298 Two control regions, real and fake enriched, are designed to measure the efficiency  
2299 of the real and fake lepton to pass the Tight selection requirements (mentioned in  
2300 Table 4.5). Those regions are with sufficiently large statistics and can represent  
2301 the kinematics and background composition in the signal region. The di-lepton  
2302 requirement ensures orthogonality to the  $3\ell$  signal region. At least one  $b$ -tagged  
2303 jet is required to avoid potentially large changes in the fake composition. The  
2304 efficiencies have been factorised in bins of  $p_T$  to get a fake prediction dependent  
2305 on the lepton kinematics. A  $(N_{b\text{-tagged}}^{Jet}, p_T^\ell)$  two dimensional parametrisation is  
2306 used for electron fake rate as it can improve fakes modelling in events with high  
2307  $b$ -tagged jet multiplicity. The two dimensional parametrisation of distance of the  
2308 muon and its closet jets and  $p_T$  ( $\Delta R(\mu, jet), p_T^\ell$ ) is designed for fake muon rate to  
2309 have a good fake modelling.

2310

#### 2311 **Real lepton control region and real lepton efficiency measurement**

2312 The real lepton control region is designed to be enriched with prompt leptons from  
2313  $t\bar{t}$  di-lepton decays by requiring the presence of two opposite-sign charge, opposite  
2314 flavour leptons. Table 4.10 shows the definition of the real control region where  $\varepsilon_r$   
2315 is performed. The prompt lepton purity achieved in this region is very high and  
2316 is shown in Figure 4.3.

2317 A standard Tag and Probe method is used to measure the real lepton efficiency  
2318 for both electrons and muons. Each event in the CR is Tagged by requiring at least  
2319 one of the leptons to pass the Tight selection and be trigger-matched. The other  
2320 one, which is unbiased by the Tight and trigger selection is the Probe candidate  
2321 for the efficiency measurement. In case two tag candidates are found, both leptons  
2322 are considered as valid probes and used for the measurement. This procedure  
2323 follows the one used in CP groups to measure the electron and muon efficiencies  
2324 for calibration.

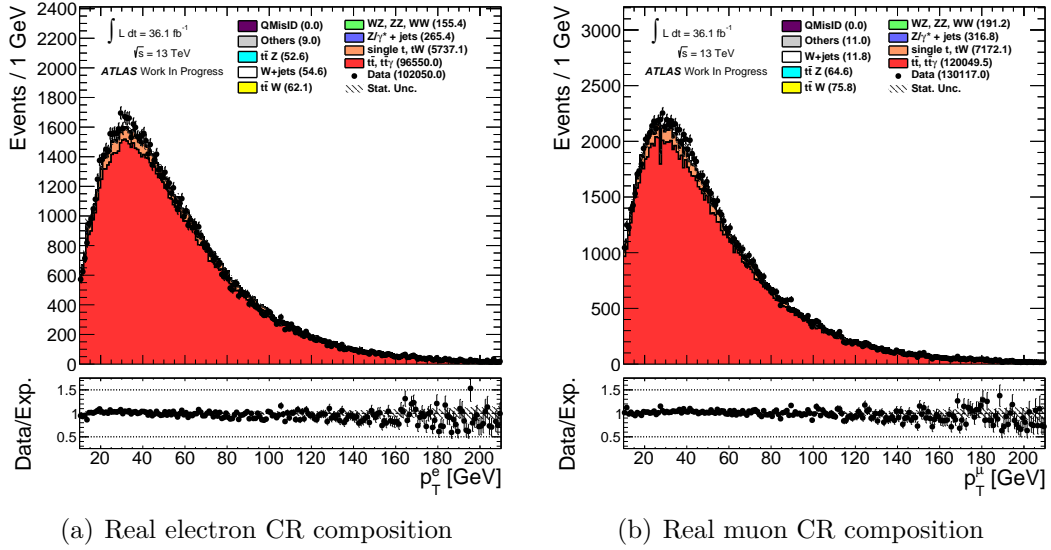


Figure 4.3: Real lepton enriched control region composition for  $p_T$  distribution of Tight electrons (left) and Tight muons.

2325 The efficiency for prompt leptons is defined on pre-event basis as the ratio of  
 2326 numerator events where the probe passes Tight selection and denominator events  
 2327 where no additional selection requirement other than the baseline Loose one (Loose  
 2328 = Tight + anti-Tight) is applied on the probe. The subtraction of background  
 2329 events is applied to both numerator and denominator:

$$\varepsilon_r(e, \mu)_i = \frac{N_i^{T(e, \mu)} - N_{bkg}^{T(e, \mu)}_i}{N_i^{L(e, \mu)} - N_{bkg}^{L(e, \mu)}_i}, \quad (4.8)$$

2330 where the background to be subtracted accounts for the contamination of OS

	Real lepton enriched CR
$N_{jets}$	2, 3
$N_{b-tagged jets}$	$\geq 1$ (MV2c10, 70% eff.)
$N_\ell$	2
lepton charge	OS
lepton flavour	$e\mu, \mu e$
$p_T^\ell$	$\geq 10$ GeV
No. of trigger-matched $\ell$	$\geq 1$

Table 4.10: Definition of the control region used for measuring the real lepton efficiency. The same region is used to measure both  $\varepsilon_r^e$  and  $\varepsilon_r^\mu$ .

2331 events with a fake lepton, and mostly comes from the top and single-Top. The  
 2332 index  $i$  is the  $p_T$  binning chosen for parametrising the efficiency.

2333 As the contribution of those fakes are small, estimated from simulation sam-  
 2334 ples directly, and a 30% systematic uncertainty is added on their normalization.  
 2335 Figure 4.4 shows the measured real electron and muon efficiency in data with sys-  
 2336 tematic uncertainty coming from fake background subtraction in data.

2337

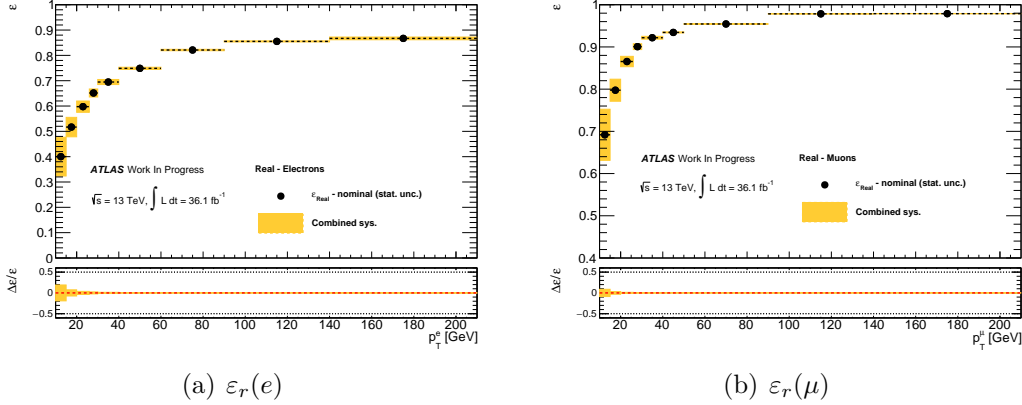


Figure 4.4: Real efficiencies for electrons and muons from CR in data. The systematic uncertainty from the OS fakes normalisation uncertainty are show in orange band.

### 2338 Fake control region and fake rate measurement

2339 It is difficult to define good control region for fake study like real lepton control  
 2340 region. Actually, only the control region dominated by  $t\bar{t}$  semileptonic events can  
 2341 be defined with one real lepton and one non-prompt lepton or converted photon. To  
 2342 address this ambiguity, a modified Tag and Probe method is designed to measure  
 2343 the fake efficiency for electrons and muons separately.

#### 2344 • Fake rate $\varepsilon_f(e)$ of electrons

2345

2346 In the opposite flavour same sign dilepton events when a tight muon firing  
 2347 the single muon trigger is found, the muon has a high probability of being the  
 2348 prompt lepton of the pair, and the remaining electron as a suitable unbiased probe  
 2349 candidate to measure the fake efficiency. Additional benefit by using opposite  
 2350 flavour events is that the reduction of the amount of the charge misID events to  
 2351 be subtracted which can be large in same-sign electron-electron region, hence lead  
 2352 to a large uncertainties on the efficiency measurement.

2353 The definition of the fake electron control region is listed in Table 4.11. The  
 2354 check on fake origin in fake CR as a function of the number of  $b$ -tagged jets  
 2355 is shown in Figure 4.5. There is a large difference in the fake electron origin

Fake electron enriched CR	
$N_{jets}$	2, 3
$N_{b-tag jets}$	$\geq 1$ (MV2c10, 70% eff.)
$N_\ell$	2
lepton charge	SS
lepton flavour	$e\mu, \mu e$
$p_T^\ell$	$\geq 10$ GeV

Table 4.11: Definition of the control region used for measuring the fake electron efficiency. Note that we require the muon to *tag* the event, with the electron being used as selection-unbiased *probe* for the efficiency measurement.

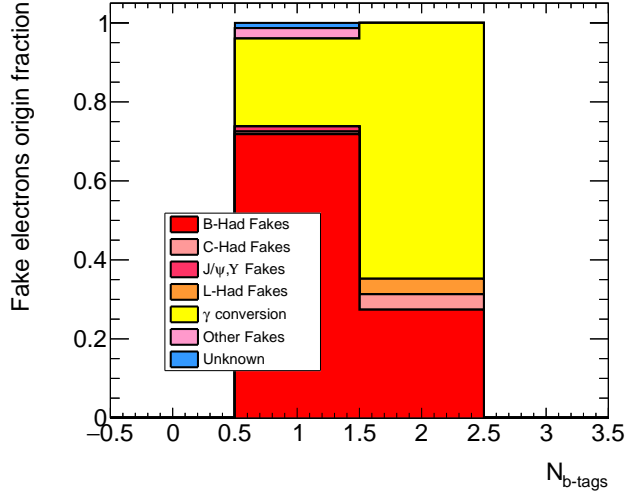


Figure 4.5: Fake electron origin fraction for  $t\bar{t} + t\bar{t}\gamma$  events, in the dileptonic  $OF$  CR where electron fake rates are measured, as a function of  $N_{b-tagged}$ .

2356 fraction between 1  $b$ -tagged jet and 2  $b$ -tagged jets. The photon conversion fakes  
 2357 in 1  $b$ -tagged jet bin is 20% compared to 60% in the 2  $b$ -tagged jets bin. One  
 2358 dimensional parametrisation of the fake rate depending on  $p_T$  only together with  
 2359 at least one  $b$ -tagged jets requirement can lead to a poor modelling of the  $b$ -tagged  
 2360 jet distribution in data. Thus, a two dimensional ( $N_{b-tagged}^{Jet}, p_T^\ell$ ) parametrisation  
 2361 of the electron fake rate is done. The efficiency of the fake electron is similar to  
 2362 the Equation 4.8:

$$\varepsilon_f(e)_{i,j} = \frac{N_{i,j}^{T(e)} - N_{bkg\ i,j}^{T(e)}}{N_{i,j}^{L(e)} - N_{bkg\ i,j}^{L(e)}}. \quad (4.9)$$



2363 The double index refers to the  $(N_{b\text{-tagged}}^{Jet}, p_T^l)$  bin. The subtracted background  
 2364 includes the event with prompt SS lepton pair, charge misID events and  $t\bar{t}$  semilep-  
 2365 tonic events with the mis-assigned probe electron. Overall the fake lepton purity  
 2366 is good, the  $t\bar{t}$  event where the probe electron happens to be incorrectly matched  
 2367 to the truth prompt lepton in the pair is overall small for only about 10% of the  
 2368 total background events to be subtracted at the numerator.

2369 Additional check on the truth origin of fake leptons is performed in the  $2\ell$ SS  
 2370 CR,  $2\ell$ SS SR and  $3\ell$  SR (with the  $3\ell$  pre-MVA selection). Figure 4.6 shows the  
 2371 truth origin of fake electrons and fake muons for those three regions. The ratio  
 2372 is defined as follows: fake leptons from specific source over the total fakes. The  
 2373 majority of fakes are the non-prompt leptons from heavy flavoured particles decays  
 2374 in  $b$ ,  $c$  quark according to Figure 4.6. The fraction of fake muons which from heavy  
 2375 flavour decays is stable in the  $2\ell$ SS CR and  $3\ell$  SR, the difference is only about  
 2376 15%.

2377 As shown in Figure 4.6, the electron fakes from heavy flavour decay dominate  
 2378 in the  $2\ell$ SS CR (about 60%), but there is a large fraction of fake electron coming  
 2379 from photon conversion (about 30%). The fraction of the photon conversion fakes  
 2380 is larger (about 50%) in the  $3\ell$  SR. Those photons can originate from the ISR/FSR,  
 2381 a  $\pi^0$  decay product or prompt photons radiated off the top or the anti-top quark.  
 2382 Then interacting with the detector material to produce a pair of opposite sign  
 2383 electrons of which only one get reconstructed. A dedicated correction is applied  
 2384 to cover this difference in the both  $2\ell$ SS SR and  $3\ell$  SR.

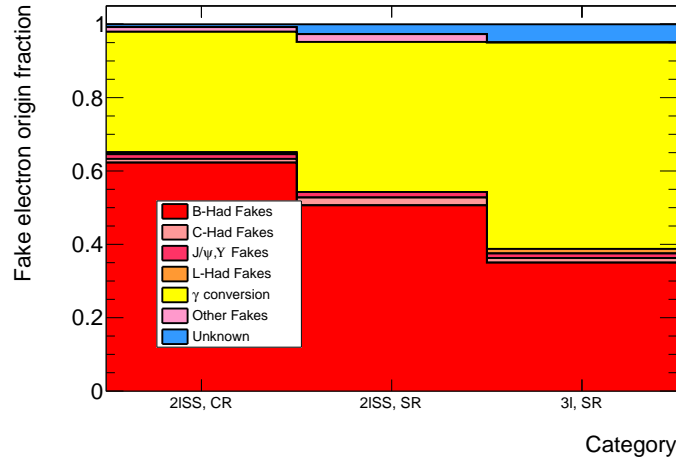


Figure 4.6: Origin fraction of fake electrons for  $t\bar{t} + t\bar{t}\gamma$  events in the dileptonic CR,  $2\ell$ SS SR and  $3\ell$  SR.

2385 As shown in Figure 4.6 in the  $3\ell$  SR, the fraction of the fake electrons which  
 2386 are from photon conversion is about 50%. The difference between the  $2\ell$ SS CR  
 2387 and  $3\ell$  SR is non-negligible. Subsequently the fake rate which is measured in the  
 2388  $2\ell$ SS CR can not describe the situation in the  $3\ell$  SR precisely. To solve this, a

2389 dedicated scale factor is designed to rescale the fraction of fake electrons from  
 2390 photon conversions in the 2ℓSS CR to be the same level in the 3ℓ SR.

2391 The dedicated factor  $\alpha$  is applied during the fake rate measurement to adjust  
 2392 the difference between the same sign fake enrich control region and the 3ℓ SR. The  
 2393 definition of factor  $\alpha$  is:

$$\alpha = \frac{(1 - f_x)\varepsilon_f^{non-PhConv} + f_x\varepsilon_f^{PhConv}}{(1 - f_{CR})\varepsilon_f^{non-PhConv} + f_{CR}\varepsilon_f^{PhConv}} - 1, \quad (4.10)$$

2394 where  $f_x$  and  $f_{CR}$  is the fraction of the fake leptons which are from photon con-  
 2395 version in dedicated region (the 3ℓ pre-MVA region in this thesis) and fake enrich  
 2396 control region,  $\varepsilon_f^{non-PhConv}$  and  $\varepsilon_f^{PhConv}$  is the fake rate of leptons for non-photon  
 2397 conversion and photon conversion origin respectively. The measurement of  $\alpha$  is  
 2398 performed for electron channel and opposite flavour channel in the fake enrich  
 2399 control region and 3ℓ pre-MVA region (only targeting at same sign lepton pair).  
 2400 The  $\varepsilon_f^{non-PhConv}$  ( $\varepsilon_f^{HF}$  in the table) and  $\varepsilon_f^{PhConv}$  are derived from fake enrich control  
 2401 region and measurements of all corresponding variables are shown in Table 4.12,  
 2402 the "X" in the table represents electrons or muons, and the impact on all related  
 2403 variables by varying fake fraction is also shown in the table to have the uncertainty  
 2404 on this correction factor. The uncertainties on conversion fraction ( $f, f_{CR}$ ) is: 40%  
 2405 taken as conservative estimate from the 2015 data/MC agreement for material de-  
 2406 scription in Ref. [62]; 50% from heavy flavour fake efficiency ( $\varepsilon_f^{non-PhConv}$ ); 50%  
 2407 from conversion fakes efficiency ( $\varepsilon_f^{PhConv}$ ). The final measured results concerning  
 2408 the uncertainties:

$$\begin{aligned} \alpha_{Xee} &= 0.57^{+0.11}_{-0.12}(f)^{+0.17}_{-0.13}(\varepsilon_f^{HF})^{+0.10}_{-0.22}(\varepsilon_f^{PhConv}), \\ \alpha_{Xe\mu(\mu e)} &= 0.07 \pm 0.01(f) \pm 0.02(\varepsilon_f^{HF})^{+0.01}_{-0.03}(\varepsilon_f^{PhConv}). \end{aligned} \quad (4.11)$$

2409 Then the fake rate in the interested region would be  $\varepsilon_f^x = (1 + \alpha)\varepsilon_f^{CR}$ . Effective  
 2410 rescaling of the fake efficiencies (symmetrizing uncertainties) are:

$$\begin{aligned} \varepsilon_f^{Xee} / \varepsilon_f^{CR} &= (1 + \alpha_{Xee}) = 1.57 \pm 0.25, \\ \varepsilon_f^{Xe\mu(\mu e)} / \varepsilon_f^{CR} &= (1 + \alpha_{Xe\mu(\mu e)}) = 1.07 \pm 0.04. \end{aligned} \quad (4.12)$$

2411 Generally this correction describes the fakes situation in the 3ℓ very well, and  
 2412 the closure test which to be introduced in Section 4.4.2.2 shows a good agreement  
 2413 between MC and the Matrix Method prediction.

2414 The fake rate of the electrons in fakes CR with data is shown in Figure 4.8. One  
 2415 dimensional efficiency represents a projection for inclusive  $N_{b\text{-tagged}}^{Jet} \geq 1$  events of  
 2416 the  $(N_{b\text{-tagged}}^{Jet}, p_T^l)$  efficiency which is also shown in the figure and used as input  
 2417 in the MM.  $p_T$  bin is set [20, 210+] GeV as to have a better statistics. For the  
 2418 systematics uncertainty on the efficiency, a theory uncertainty of 14% is set be-  
 2419 cause of the theoretical cross section, QCD scale and PDF uncertainties ( $t\bar{t}W, t\bar{t}Z$ ).  
 2420 50% uncertainty is set on the  $VV$ , and 30% uncertainty is set on the remaining

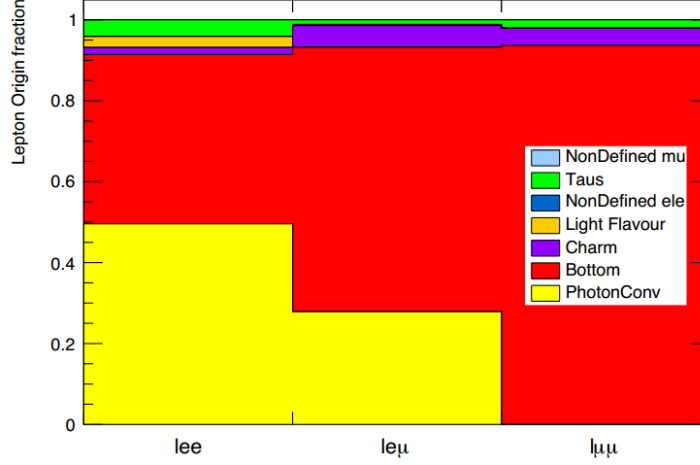


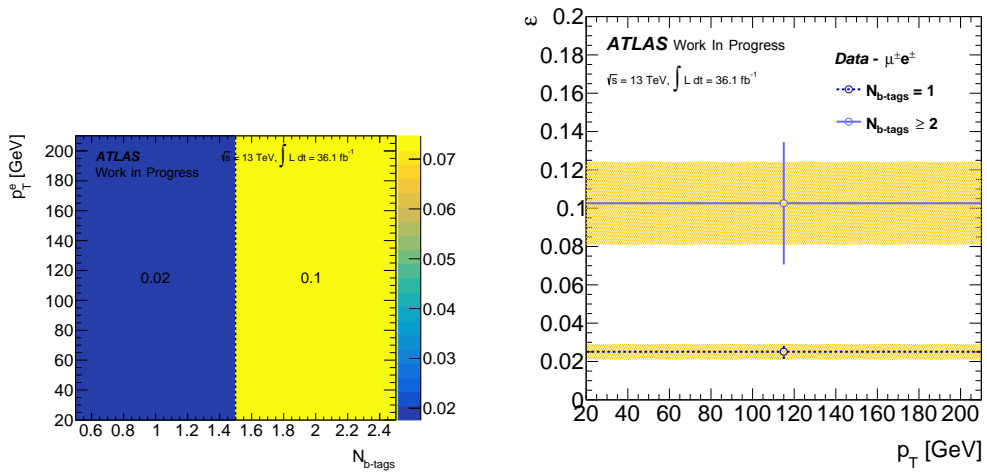
Figure 4.7: Origin fraction of fake electrons and fake muons for  $t\bar{t} + t\bar{t}\gamma$  events in the  $3\ell$  by flavoured channel.

	Nominal	$f_{CR} - 40\%$	$f_{CR} + 40\%$	$\varepsilon_f^{HF} - 50\%$	$\varepsilon_f^{HF} + 50\%$	$\varepsilon_f^{PhConv} - 50\%$	$\varepsilon_f^{PhConv} + 50\%$
$X_{ee}$							
$f_x(3\ell SR)$	0.498	0.299	0.697	0.498	0.498	0.498	0.498
$f_{CR}$	0.25	0.15	0.35	0.25	0.25	0.25	0.25
$\varepsilon_f^{HF}$	0.009	0.009	0.009	0.005	0.014	0.009	0.009
$\varepsilon_f^{PhConv}$	0.057	0.057	0.057	0.057	0.057	0.029	0.086
$\alpha_{X_{ee}}$	0.56	0.44	0.65	0.74	0.44	0.35	0.67
$X_{e\mu}$ or $X_{\mu e}$							
$f_x(3\ell SR)$	0.279	0.167	0.391	0.279	0.279	0.279	0.279
$f_{CR}$	0.25	0.15	0.35	0.25	0.25	0.25	0.25
$\varepsilon_f^{HF}$	0.009	0.009	0.009	0.005	0.014	0.009	0.009
$\varepsilon_f^{PhConv}$	0.057	0.057	0.057	0.057	0.057	0.029	0.086
$\alpha_{X_{e\mu}(\mu e)}$	0.07	0.06	0.08	0.09	0.05	0.04	0.08

Table 4.12: The impact of varying fake fraction on correction factor  $\alpha$ .

2421 processes to be subtracted. The total uncertainty of the data-driven charge misID  
 2422 is also considered. All the systematics uncertainties related to the detector are  
 2423 neglected as those uncertainties can be covered by the size of the background nor-  
 2424 malization uncertainties.

2425



(a)  $\varepsilon_f(e)$ ,  $(N_{b\text{-tagged}}^{Jet}, p_T^\ell)$  parametrisation (b)  $\varepsilon_f(e)$ ,  $p_T$  projection for  $N_{b\text{-tagged}}^{Jet}$  slices

Figure 4.8: Fake rate for electrons from CR in data. 4.8(a) shows the two-dimensional  $(N_{b\text{-tagged}}^{Jet}, p_T^\ell)$  map, and 4.8(b) represents the fake rate projection over  $p_T$  for each  $N_{b\text{-tagged}}^{Jet}$  slice. The orange bands represent the total systematic uncertainty on the fake rate. Please note that the efficiencies in 4.8(b) have been rescaled by the  $\alpha_{ee}$  factor.

2426 • Fake rate  $\varepsilon_f(\mu)$  of muons

2427

2428 Unlike using the opposite flavour SS events to measure fake electron rate, a  
 2429 Tag and Probe based on  $\mu\mu$  events is adopted in fake muon rate measurement,  
 2430 assuming the subleading muon is more likely to be the fake in the lepton pair, the  
 2431 subleading one is chosen as probe when both muons are tight and fire the trigger.  
 2432 According to Figure 4.9, the fake muons from heavy flavour decay dominate in the  
 2433  $2\ell$ SS CR (about 90%). No fake contributions from photon conversion in the both  
 2434  $2\ell$ SS CR and  $3\ell$  SR for fake muons. About 20% fake muons in the  $3\ell$  SR are from  
 2435 "Unknown", and actually those muons are the prompts which come from virtual  
 2436 photon conversions.

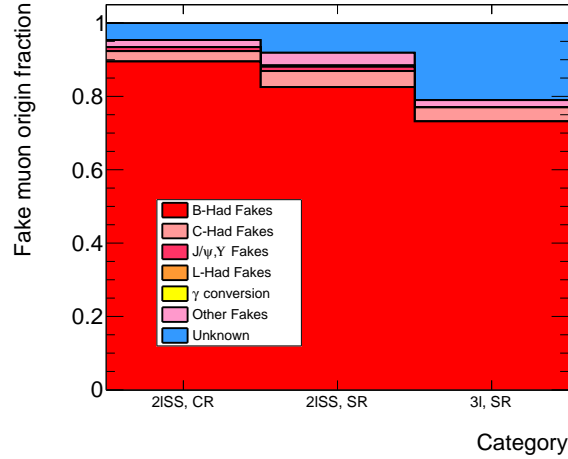


Figure 4.9: Origin fraction of fake muons for  $t\bar{t} + t\bar{t}\gamma$  events in the dileptonic CR,  $2\ell$ SS SR and  $3\ell$  SR.

2437 Poor modelling of the low  $\Delta R(\mu, jet)$  is observed when using one dimensional  
 2438  $p_T$  parametrisation of the muon fake rate. A two dimensional measurement on  
 2439  $(\Delta R(\mu, jet), p_T^\ell)$  for muon fake rate is designed to have a better modelling of the  
 2440 fakes. The binning in  $\Delta R(\mu, jet)$  is  $[0.0, 1.0, 5.0]$  and the  $p_T$  binning is  $[20.0, 50.0,$   
 2441  $210+]$ . The fake rate for muons is:

$$\varepsilon_f(\mu)_{i,j} = \frac{N_{i,j}^{T(\mu)} - N_{bkg}^{T(\mu)}_{i,j}}{N_{i,j}^{L(\mu)} - N_{bkg}^{L(\mu)}_{i,j}}, \quad (4.13)$$

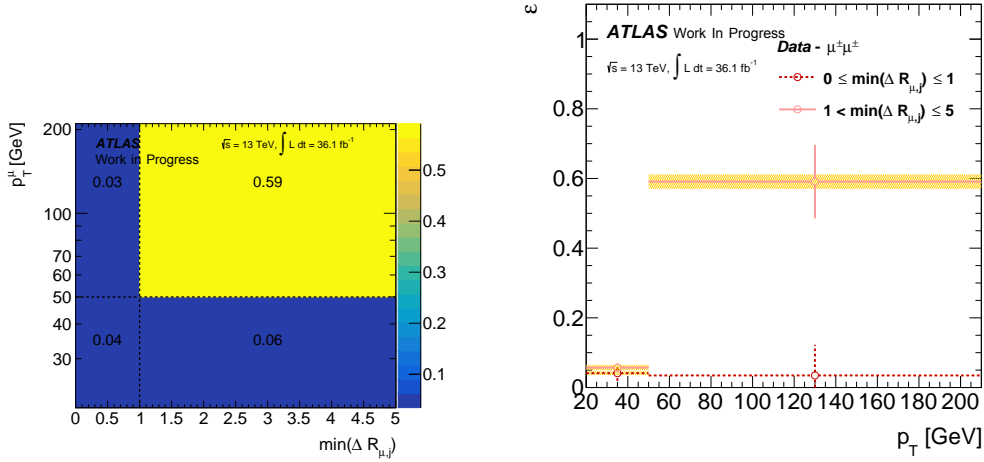
2442 the background to be subtracted comes from the prompt same sign lepton pairs  
 2443 and  $t\bar{t}$  semileptonic events with mis-assigned probe muons. Table 4.13 summaries  
 2444 the definition of the fake muons control region which is used to measure the fake  
 2445 rate of muons.

2446 The assumption of using  $p_T$  ranking as criterion to solve the ambiguity of both  
 2447 muons being tight and trigger matched does not have a significant impact on the

Fake muon enriched CR	
$N_{jets}$	2, 3
$N_{b-tag jets}$	$\geq 1$ (MV2c10, 70% eff.)
$N_\ell$	2
lepton charge	SS
lepton flavour	$\mu\mu$
$p_T^\ell$	$\geq 10$ GeV

Table 4.13: Definition of the control region used for measuring the fake muon efficiency.

2448 muon fake estimate. Overall the fake muons purity is good and  $t\bar{t}$  events where  
 2449 the probe is the prompt lepton in the pair is only 16% of the total process to be  
 2450 subtracted from data in the numerator. Figure 4.10 shows the fake rate of the  
 2451 muons after background subtracting in data. The systematics uncertainties of the  
 2452 fake rate is similar to the electron case, but uncertainties from charge misID is  
 2453 neglected.



(a)  $\varepsilon_f(\mu)$ ,  $(\Delta R(\mu, jet), p_T^\ell)$  parametrisation (b)  $\varepsilon_f(\mu)$ ,  $p_T$  projection for  $\Delta R(\mu, jet)$  slices

Figure 4.10: Fake rate for muons from CR in data. 4.10(a) shows the two-dimensional  $(\Delta R(\mu, jet), p_T^\ell)$  map, and 4.10(b) represents the fake rate projection over  $p_T$  for each  $\Delta R(\mu, jet)$  slice. The orange bands represent the total systematic uncertainty on the fake rate.

2454 **4.4.2.2 Closure test**

2455 The Matrix Method is based on the assumption that the fake composition is stable  
 2456 between CR where the real or fake efficiency measured and SR where efficiency  
 2457 are applied. In fact the fake sources are different in the CR and the SR, and also  
 2458 the efficiency of passing the Tight selection may be different. Thus it is necessary  
 2459 to check the assumption, as well as to ensure no additional bias from the method,  
 2460 and the Matrix Method performs correctly. A so-called closure test has been  
 2461 performed in the  $3\ell$  pre-MVA region. Generally, it's done by comparing the fake  
 2462 event yields from MC and the Matrix Method in the  $3\ell$  pre-MVA region, and  
 2463 events are normalized to the luminosity of  $36.1 \text{ fb}^{-1}$ . The corresponding real and  
 2464 fake efficiencies are measured by using the MC in the related CRs. Figure 4.11  
 2465 shows the real and fake efficiencies measured in  $t\bar{t} + t\bar{t}\gamma$  events, where a truth  
 2466 requirement of exactly two prompt leptons is applied in the real enriched CR. The  
 2467 probe lepton is required to be a non-prompt in fake enriched CRs. Truth charge  
 2468 flip events have been vetoed in both real and fake CRs. Such efficiencies are used  
 2469 to compute the Matrix Method weights for the same  $t\bar{t} + t\bar{t}\gamma$  sample to derive the  
 2470 fake yields.

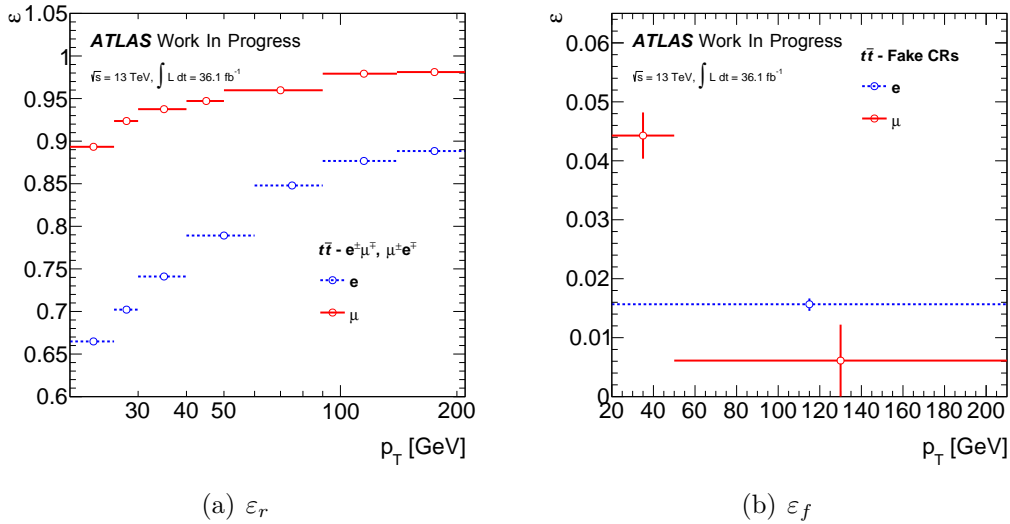


Figure 4.11: Electron, muon real (4.11(a)) and fake (4.11(b)) efficiencies as measured in  $t\bar{t} + t\bar{t}\gamma$  MC simulation. In the fake case, the plots are inclusive projections over  $p_T$  of the 2D efficiency maps previously described.

2471 To check the robustness of the method and the additional effects other than  
 2472 one cited, the closure test is done in three channels by lepton flavour ( $ee$ ,  $\mu\mu$ , and  
 2473 opposite flavour ( $OF$ )) in the SR of  $3\ell$ . Any non-closure observed will also account  
 2474 for missing non-trivial efficiency parametrisations and binning effects. Finally the  
 2475 observed non-closure is the one to be quoted as an additional nuisance parameter  
 2476 into the final fit.

2477 Due to the high fraction of the photon conversion fakes,  $t\bar{t}\gamma$  sample is added when  
 2478 doing the closure test. Besides this  $t\bar{t}\gamma$  sample, the di-lepton filter  $t\bar{t}$  sample is used  
 2479 in the  $3\ell$  case instead of the non-allhadronic one which is used for the efficiency  
 2480 measurement to increase the statistics of the  $3\ell$  events. The event selection for the  
 2481 closure test is the  $3\ell$  pre-MVA selection. The expected prediction is obtained from  
 2482  $t\bar{t} + t\bar{t}\gamma$  MC sample with two Tight leptons (the same sign lepton pair in the  $3\ell$   
 2483 channel), and then requiring at least one truth-matched fake lepton either coming  
 2484 from a photon conversion or heavy flavour decay. The truth charge flip lepton  
 2485 events are vetoed. The same truth selection is applied to the Matrix Method as  
 2486 well.

2487 The non-closure ( $\zeta$ ) is obtained from the ratio of the Matrix Method ( $N_{MM}$ )  
 2488 and pure MC ( $N_{t\bar{t},t\bar{t}\gamma}$ ) prediction, and the relative bias between the two estimates  
 2489 is defined as:

$$\zeta = \frac{N_{MM} - N_{t\bar{t},t\bar{t}\gamma}}{N_{MM}}, \quad (4.14)$$

$$\sigma_\zeta = \sigma_{SF} = \sqrt{\frac{1}{N_{MM}^2} \cdot \sigma_{N_{t\bar{t},t\bar{t}\gamma}}^2 + \frac{N_{t\bar{t},t\bar{t}\gamma}^2}{N_{MM}^4} \cdot \sigma_{N_{MM}}^2}.$$

2490 The error on the Matrix Method estimate  $\sigma_{N_{MM}}$  is a combination of the sta-  
 2491 tistical uncertainty driven by the size of the  $TT, T\bar{T}, \bar{T}T, \bar{T}\bar{T}$  sidebands, and the  
 2492 uncertainty on the measured efficiencies. The above formula for the non-closure  
 2493 uncertainty ( $\sigma_\zeta$ ) holds in case  $\sigma_{N_{MM}}$  and  $\sigma_{N_{t\bar{t},t\bar{t}\gamma}}$  are not correlated. This is true  
 2494 in the first order, since the contribution of the  $TT$  sideband, which is the only  
 2495 one not independent between the pure MC event set and the MM set has a small  
 2496 contribution to the total MM fake event yields compared to the sidebands with  
 2497 anti-tight leptons.

2498 The  $3\ell$  pre-MVA selection is applied to both MC and the Matrix Method to get  
 2499 the fake event yields for closure test. The two tight leptons with same sign and  
 2500 at least one fake from photon conversion or heavy flavour decay are required. The  
 2501 truth selection is applied to remove the charge flip and keep the pure fakes from  
 2502 MC. Figure 4.12 shows the variables' distribution comparison between simulation  
 2503 prediction and the Matrix Method predicted fakes in inclusive channel. Table 4.14  
 2504 shows the fake event yields comparison between MC and MM in flavoured chan-  
 2505 nel with the  $3\ell$  pre-MVA selection. The agreement is good considering the low  
 2506 statistics and large uncertainties.

### 2507 4.4.2.3 Fake results in data

2508 The measured real efficiency and fake rate which are mentioned in Section 4.4.2.1  
 2509 are as the input to the Matrix Method to get the fake leptons in the  $3\ell$  pre-MVA  
 2510 region by using the data. The uncertainties on the efficiencies have been propa-



Fake Yields	$\mu\mu$	$ee$	OF
MM	$16.4 \pm 4.1$	$11.1 \pm 3.3$	$24.9 \pm 5.0$
Expected number	$19.1 \pm 4.3$	$11.0 \pm 3.3$	$27.4 \pm 5.2$
Non-closure	$-0.14 \pm 33\%$	$0.01 \pm 43\%$	$-0.1 \pm 26\%$

Table 4.14: Expected fake event yields from MC and from the Matrix Method with  $t\bar{t} + t\bar{t}\gamma$  in the  $3\ell$  pre-MVA region.

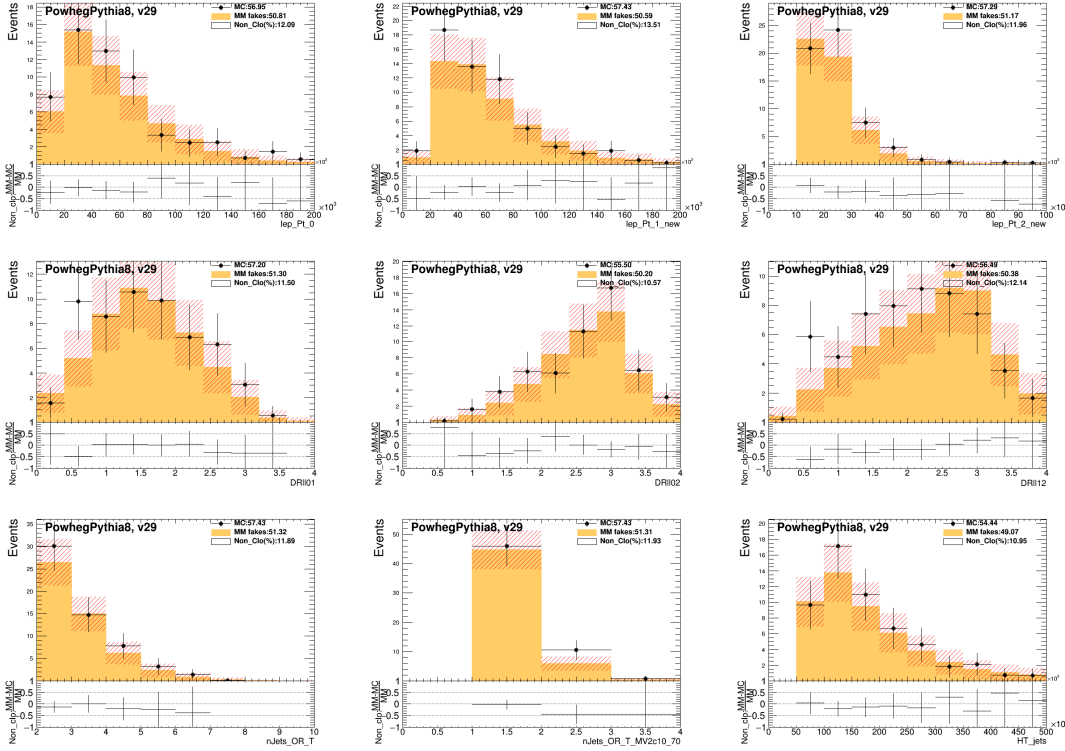


Figure 4.12: Closure of the  $p_T$  of leptons, distance between leptons (DRll) and number of jets and  $b$ -tagged jets between  $t\bar{t} + t\bar{t}\gamma$  simulated events and MM fake events.

2511 gated to the event weights. As for the statistical uncertainty on the efficiencies  
 2512 and the systematic uncertainty from charge misID subtraction, each bin of the  
 2513 efficiency parametrisation has been varied independently from others. The back-  
 2514 ground theory uncertainty variation is applied simultaneously across the efficiency  
 2515 bins. The final shape uncertainty on a variable of interest is taken as the sum in  
 2516 quadrature of all the uncertainties contributing to each bin of the distribution.

2517 Table 4.15 shows the fakes event yields by using the Matrix Method with data  
 2518 in  $ee$ ,  $\mu\mu$  and  $OF$  in the  $3\ell$  pre-MVA region with statistical and systematic un-  
 2519 certainties.

	$\mu\mu$	$ee$	OF
Fake Yields(MM)	$35.3 \pm 5.9$	$24.0 \pm 4.9$	$62.0 \pm 7.9$
MC prediction	$16.4 \pm 4.1$	$11.1 \pm 3.3$	$24.9 \pm 5.0$
MM/MC-prediction ratio	2.15	2.18	2.49

Table 4.15: Expected fake event yields from MC and from the Matrix Method with data in the  $3\ell$  pre-MVA region.

## 4.5 Background suppressing using the multi-variate analysis technique

The multi-variate analysis (MVA) technique has been widely employed in the ATLAS experiment in many analyses. For the analysis in this thesis, the Boosted Decision Trees (BDT) is optimized to further reject the "reducible" fake background which is dominantly from the  $t\bar{t}$  and irreducible background from  $t\bar{t}W$  and  $t\bar{t}Z$ . Before the MVA optimization, the  $3\ell$  pre-MVA selection shown in the Section 4.3.6 is required and corresponding events yields are shown in Table 4.8. As shown in Table 4.8, the dominant contributions are from non-prompt background and  $t\bar{t}V$ . To suppress those backgrounds, a two dimensional (2D) BDT analysis strategy is designed to enhance the signal events in the pre-MVA region.

### 4.5.1 Modelling of the MVA input variables

A set of variables are used as the input for the MVA. The definition of those variables is shown in Table 4.16. The comparison between data and MC for those input variables is shown in Figure 4.14. The agreement is good and those inputs show the promising discriminate power.

Variable	Definition
<i>ttH vs ttV</i> BTDG	
$P_T(\ell_0, \ell_1)$	Transverse momentum of the lepton pair $\ell_0, \ell_1$
$m(\ell_0, \ell_1)$	Invariant mass of the leptons $\ell_0, \ell_1$
$m(\ell_1, \ell_2)$	Invariant mass of the leptons $\ell_1, \ell_2$
$best\_Z\_m(\ell\ell)$	The closest invariant mass between two leptons to the Z boson pole mass
$\Sigma\Delta R(\ell_0, \ell_1, \ell_2)$	Scalar sum of the $\Delta R$ between $(\ell_0, \ell_1), (\ell_0, \ell_2), (\ell_1, \ell_2)$
$\Delta\eta(\ell_1, \ell_2)$	Pseudo-rapidity difference between $\ell_1, \ell_2$
$m_{eff}$	The effective mass defined as the sum of muons, electrons, jets transverse momentum and $E_T^{miss}$
<i>ttH vs Non-prompt</i> BTDG	
$n_{jets}$	Number of jets
$n_{jets} + 10 \cdot n_{b-jets}$	Multiplicity of jets and multiplied by 10 times the multiplicity of $b$ -tagged jets
$P_t^2$	Third lepton Pt
$m(\ell_1, \ell_2)$	Invariant mass of the lepton pair $\ell_1, \ell_2$
$\Sigma\Delta R(\ell_0, \ell_1, \ell_2)$	Scalar sum of the $\Delta R$ between $(\ell_0, \ell_1), (\ell_0, \ell_2), (\ell_1, \ell_2)$
$\Delta\eta(\ell_1, \ell_2)$	Pseudo-rapidity difference between $\ell_1, \ell_2$
$m_{eff}$	The effective mass defined as the sum of muons, electrons, jets transverse momentum and $E_T^{miss}$

Table 4.16: List of input variables for the BDT.

### 4.5.2 Performance of the BDT

The BDTG (Gradient boosting algorithm) is trained on sets of MC samples for the signal, irreducible background and data-riven sample for the non-prompt background with total 10 variables which are mentioned in Section 4.5.1. Especially for

2540 the reducible background, the data-driven result, the Matrix Method estimation  
 2541 is used as the good description of the non-prompt background. The closure test of  
 2542 the Matrix Method shows the good performance of the method, and the predicted  
 2543 non-prompt could be employed in the BDTG training especially for the shape  
 2544 modelling. The cross training strategy is employed as the good performance for  
 2545 the low statistics analysis. The main idea is that using the half of the total events  
 2546 for training and rests are used for testing, and events are chosen by requiring odd  
 2547 or even event number.

2548 The separation of the BDTG for signal and background events is illustrated in  
 2549 Figure 4.13. No overtraining issue is observed for both. The input variables in the  
 2550 top ranking are:

- 2551 •  $t\bar{t}V$ :  $P_T(\ell_0, \ell_1)$ ,  $m(\ell_0, \ell_1)$ ,  $\Sigma\Delta R(\ell_0, \ell_1, \ell_2)$  and  $m_{\text{eff}}$ .
- 2552 • Non-prompt:  $n_{\text{jets}}$ ,  $\Delta\eta(\ell_1, \ell_2)$ ,  $m(\ell_1, \ell_2)$ ,  $m_{\text{eff}}$  and  $\Sigma\Delta R(\ell_0, \ell_1, \ell_2)$ .

2553 Figure 4.13(c) and Figure 4.13(d) show the background rejection versus sig-  
 2554 nal efficiency (Receiver Operating Characteristic curve, the ROC curve) obtained  
 2555 with odd or even events. Similar performance are obtained for the both BDTGs  
 2556 demonstrating by using odd or even events.

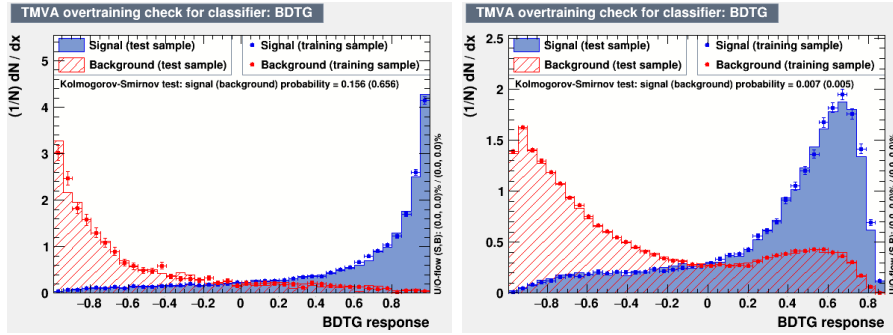
2557 The full shape of the BDTGs are used to maximize the signal sensitivity. Two  
 2558 BDTGs are combined into one final discriminant, called  $BDTG_{\text{Combination}}$  and is  
 2559 defined as:

$$BDTG_{\text{Combination}} = (\text{BDTG\_Non-prompt} + a \times \text{BDTG\_}t\bar{t}V)/(1 + a), \quad (4.15)$$

2560 here  $a$  is defined as the slope in the BDTG\_Non-prompt vs BDTG\_  $t\bar{t}V$  plane as  
 2561 shown in Figure 4.13(e).

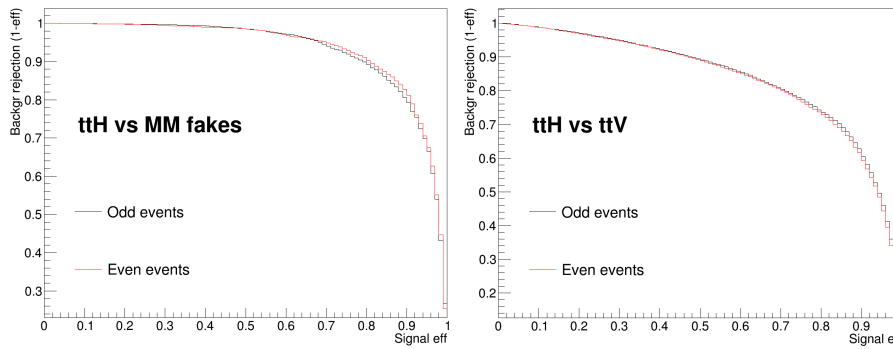
2562 The studies have been done to find a good combination of those two BDTGs,  
 2563 finally parameter  $a$  with value 1 is the best choice. The discriminant BDTG is then  
 2564 splitted in 6 bins which has been optimised to separate signal and background using  
 2565 the TransfoD function called auto-binning in ***ttHFitter***. This iterative algorithm  
 2566 defines the bins of the BDTG distribution according to two free parameters  $b$  and  
 2567  $s$ , whose sum is constrained to the number of bins. Several tries to improve on the  
 2568 signal sensitivity are attempted by scanning 2 parameters:  $b$  and  $s$ . The Asimov  
 2569 fit of the  $t\bar{t}H$  signal strength including all statistical and systematic uncertainties  
 2570 is used for this study. The best sensitivity is obtained with  $a = 1$ ,  $b = 6$  and  
 2571  $s = 0$ . This corresponds to a situation where the signal is constant in each bins  
 2572 of the BDTG discriminant. As a result the lowest bins are more populated by  
 2573 non-prompt background, the central bins more populated by  $t\bar{t}V$  and the highest  
 2574 bins more pure by  $t\bar{t}H$ . The final discriminant is therefore defined as :

$$BDTG_{\text{Combination}} = (\text{BDTG\_Non-prompt} + \text{BDTG\_}t\bar{t}V)/2 \quad (4.16)$$



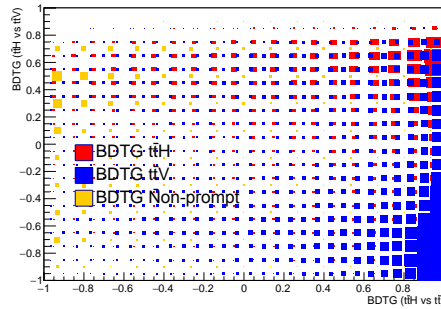
(a) BDTG response to non-prompt background

(b) BDTG response to  $t\bar{t}V$



(c) The ROC curve to non-prompt background

(d) The ROC curve for  $t\bar{t}V$



(e) BDTG\_Non-prompt vs BDTG\_ $t\bar{t}V$  plane

Figure 4.13: Distribution of the BDTG non-prompt response and the BDTG  $t\bar{t}V$  response. The background rejection versus signal efficiency (ROC) for the both BDTGs and BDTG\_Non-prompt vs BDTG\_ $t\bar{t}V$  plane.

2575 with TransfoD (6,0) configuration.

2576 The modelling of this final BDT discriminant is good as shown in Figure 4.14,  
 2577 the  $BDTG_{Combination}$ .

2578

2579 As  $t\bar{t}V$  is one of the main backgrounds, it is necessary to check its modelling  
2580 with the BDTG output. A set of cuts on the BDTG output is applied to select a  
2581 relative pure  $t\bar{t}V$  region. A detailed cut and count method is designed to find the  
2582  $t\bar{t}V$  CR and it is mentioned in Section 4.6. Figure 4.15 shows the BDTG input  
2583 variables distribution comparison between data and prediction in the  $t\bar{t}V$  region  
2584 based on the BDTG output. The purity of  $t\bar{t}V$  and variables' modelling is fine.  
2585 The validation of the BDTG output is good.

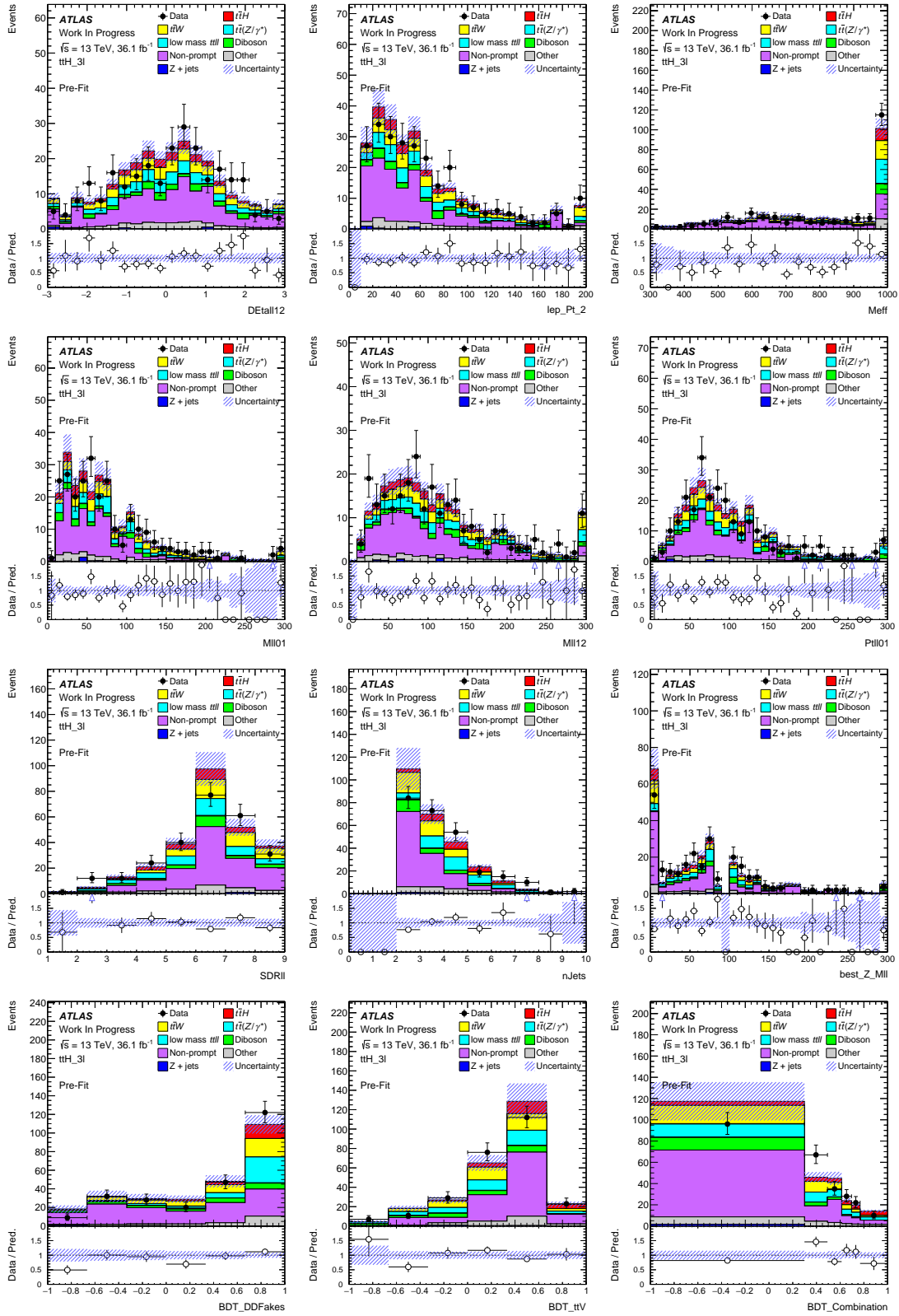


Figure 4.14: Pre-fit distributions of the MVA input and output variables in the pre-MVA region.

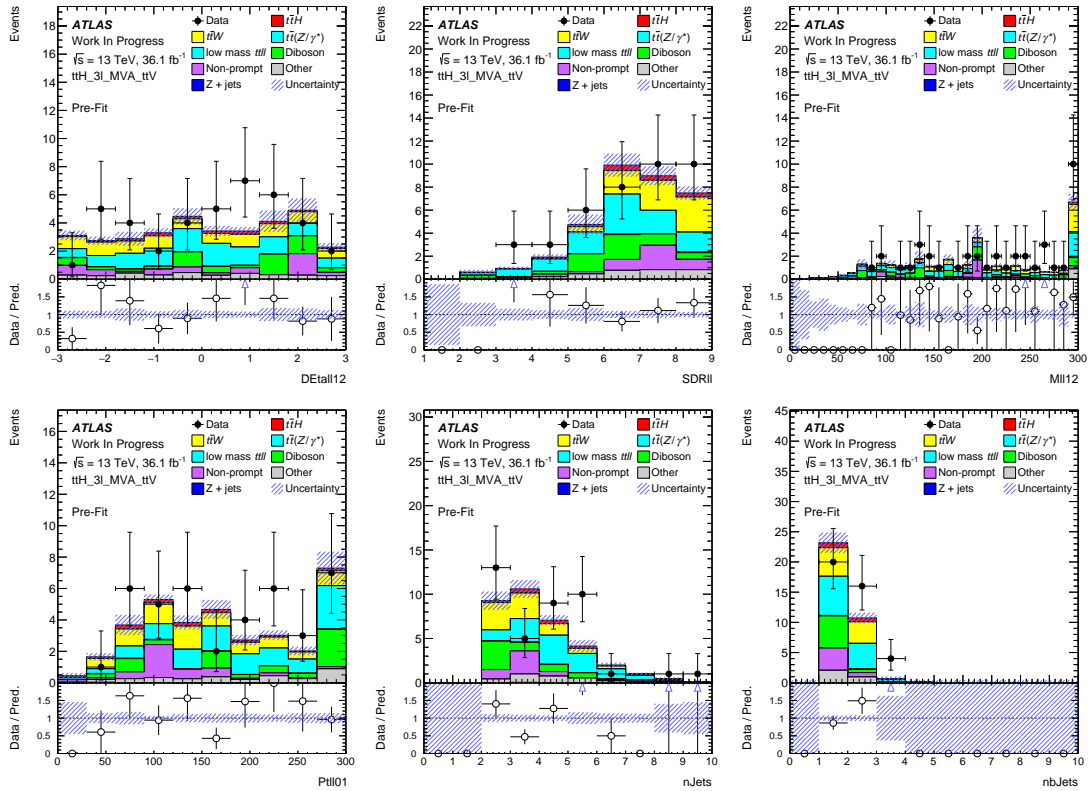


Figure 4.15: Pre-fit distributions of the MVA input variables in the  $t\bar{t}V$  control region.



## 4.6 Validation and control regions

The description of data by various Monte Carlo samples (MC) is tested in several validation regions close to the signal region in the event topology. The validation regions under study are defined in Table 4.17. A selection of those comparison plots is shown in this section. The  $t\bar{t}Z$  and  $t\bar{t}W$  backgrounds are modelled from MC simulation in the current plots and can be used as a cross check in the analysis.

VR	Selection
$t\bar{t}Z$	$3\ell$ lepton and jet selection require at least one OS SF pair within 10 GeV of $m_Z = 91.2$ GeV and requiring 4 jets and 2 $b$ -tagged jets
$t\bar{t}W$	$2\ell$ lepton selection, leading and subleading lepton $p_t > 30$ GeV require $\geq 2$ $b$ -tagged jets and 3 or 4 reconstructed jets $H_T(\text{jets}) > 220$ GeV in $ee$ and $e\mu$ channel $M_{ee}$ not within [75,105] GeV and $E_T^{\text{miss}} > 50$ GeV in $ee$ channel both leptons with positive charge

Table 4.17: Description of the validation regions being designed for  $t\bar{t}Z$  and  $t\bar{t}W$  background.

### 4.6.1 $t\bar{t}W$

As  $t\bar{t}W$  is the dominant irreducible background and it has not been observed yet in ATLAS, it is important to build a validation region to make sure that the MC prediction matches the data in  $t\bar{t}W$  enriched region. Two VRs are built based from pre-MVA region: a region defined with cut and count using simple kinematic variables (Table 4.17) and a region based on the BDTG\_Non-prompt vs BDTG\_ $t\bar{t}V$  plane (shown in Figure 4.15). In both cases, the purity is above 40% but modelling of  $t\bar{t}W$  region based on MVA output is better. In both cases, a fit to the strength of  $t\bar{t}W$  gives an agreement with the MC prediction at better than  $1\sigma$  level.

Distributions of the number of electrons and the number of jets are shown in Figure 4.16. A good agreement between prediction and data is observed and the tests reinforce the confidence that  $t\bar{t}W$  is well modelled by the Monte Carlo prediction.

### 4.6.2 $t\bar{t}Z$

The Run 1 definition uses an inverted  $Z$ -veto cut in the  $3\ell$  signal region selection, which means selecting events in  $Z$  mass window with at least 4 jets and 1  $b$ -tagged

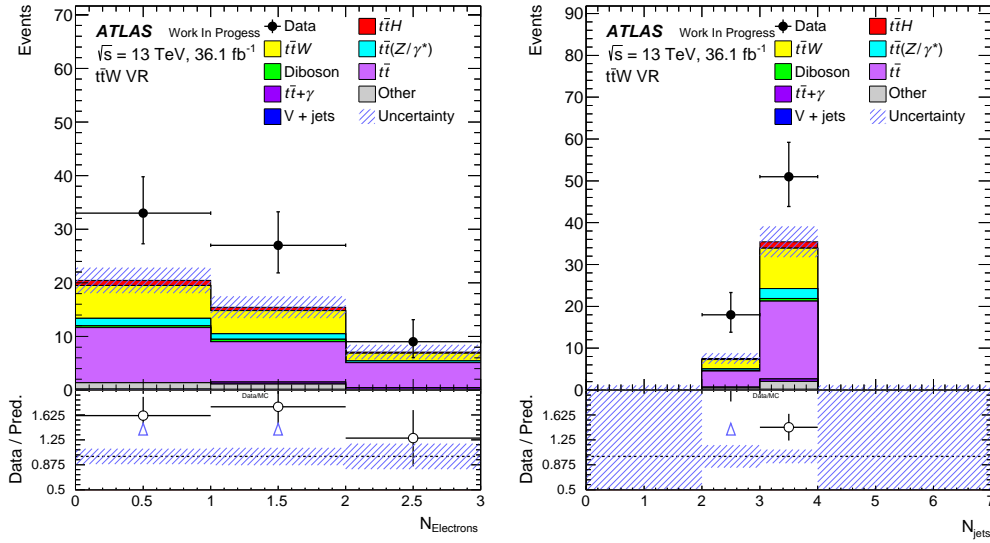


Figure 4.16: Distributions of the number of electrons and number of jets for the  $t\bar{t}W$  cut-based validation region.

2609 jet or at least 3 jets and 2  $b$ -tagged jets.  $t\bar{t}Z$  purity (especially against WZ events)  
 2610 can be improved by requiring that events must have at least 4 jets and 2  $b$ -tagged  
 2611 jets. Overall  $t\bar{t}Z$  purity is about 85%. Comparisons between data and MC are  
 2612 shown in Figure 4.17.

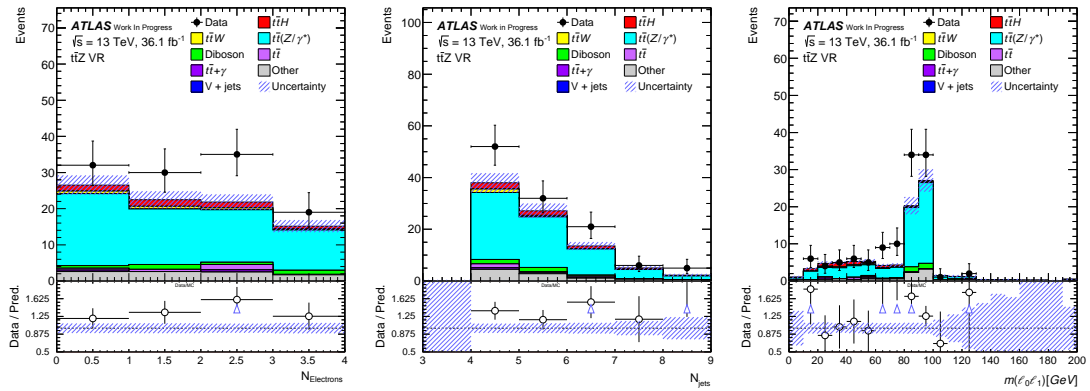


Figure 4.17: Distributions of the number of electrons, number of jets and invariant mass of Lep0 and Lep1 for  $t\bar{t}Z$  cut-based validation region and errors include statistical and systematic uncertainties.

## 4.7 Uncertainties

Sets of uncertainties which are related to the analysis in this thesis are presented. A brief summary on the uncertainties is discussed here, mainly covering experimental systematics, signal and background modelling theoretical systematics. The sources of systematic uncertainty considered in this analysis are summarised in Table 4.21. They impact the estimated signal and background rates, the migration of events between categories and/or the shape of the discriminants used in the final fit.

The uncertainty in the combined 2015+2016 integrated luminosity is 2.1%. It is derived, a methodology similar to that detailed in Ref. [63], from a calibration of the luminosity scale by using x-y beam-separation scans performed in August 2015 and May 2016.

The uncertainties in physics objects which are related to the reconstruction and identification of light leptons and hadronic  $\tau$  leptons, to the reconstruction and  $b$ -tagging of jets, and to the reconstruction of  $E_T^{miss}$  are all considered. The impact of the  $b$ -tagging efficiency uncertainty is weaker, as the  $b$ -tagged jet multiplicity in all selected events is typically one or two. The impact of the uncertainty of jets containing either charm jets or  $\tau_{had}$  is significant and, due to the calibration procedure applied, is taken as fully correlated between the two jet flavours. Uncertainties in lepton reconstruction, identification and trigger efficiencies have negligible impact.

The systematic uncertainties associated with the generation of signal and background processes are due to uncertainties in the assumed cross sections and the acceptance for each process in each category and bin used in the final fit. The most important uncertainty arising from theoretical predictions is on the assumed SM cross sections and the modelling of the acceptance for  $t\bar{t}H$ ,  $t\bar{t}Z$  and  $t\bar{t}W$  production.

### 4.7.1 Experimental systematics

Many quantities used in thesis are subject to the experimental systematic uncertainties. Each systematic effect is evaluated individually by using the given uncertainties on an event by event basis. These uncertainties are related to the trigger efficiency, lepton reconstruction and identification, jet calibration, continuous  $b$ -tagging and the global event activity. The experimental systematic treatments are evaluated by ATLAS performance groups and are used in this analysis as an overall reweighting or rescaling of the object energy and momentum. The list of the systematics uncertainties to be considered and included is summarised in Table 4.18 along with their type, description and name of systematics in the workspace and status of inclusion in the analysis.

The application row in Table 4.18 indicates the methodology of inclusion of the systematic in the analysis: overall event reweight or as a data/MC determined scale factor of the transverse momentum. If no explicit indication is provided the

2653 rescaling approach is adopted. The systematic associated to Jet Vertex Tagger,  
2654 taking into account various information to quantifying the fraction of the track  
2655 transverse momentum associated to a jet from the hard scattering interaction,  
2656 requires a particular treatment. This systematic uncertainty is determined from  
2657 the variation of the corresponding cut. The effect of the various systematics is  
2658 evaluated assuming a positive and negative  $1\sigma$  variation around the nominal value  
2659 of the interested quantity to evaluate the effects on the results yields.

2660 The summary of experimental systematics for the  $b$ -tagged jets in the analysis  
2661 is shown in Table 4.20.

Experimental Systematics on Leptons				
Type	Description	Systematics Name	Application	Analysis
<b>Trigger</b>				
Scale Factors	Trigger Efficiency	lepSFTrigTight_MU(EL)_SF_Trigger_STAT(SYST)	Event Weight	✓
<b>Muons</b>				
Efficiencies	Reconstruction and Identification	lepSFObjTight_MU_SF_ID_STAT(SYST)	Event Weight	✓
	Isolation	lepSFObjTight_MU_SF_Isol_STAT(SYST)	Event Weight	✓
	Track To Vertex Association	lepSFObjTight_MU_SF_TTVA_STAT(SYST)	Event Weight	✓
$p_T$ Scale	$p_T$ Scale	MUONS_SCALE	$p_T$ Correction	✓
Resolution	Inner Detector Energy Resolution	MUONS_ID	$p_T$ Correction	✓
	Muon Spectrometer Energy Resolution	MUONS_MS	$p_T$ Correction	✓
<b>Electrons</b>				
Efficiencies	Reconstruction	lepSFObjTight_EL_SF_ID	Event Weight	✓
	Identification	lepSFObjTight_EL_SF_Reco	Event Weight	✓
	Isolation	lepSFObjTight_EL_SF_Isol	Event Weight	✓
Scale Factor	Energy Scale	EG_SCALE_ALL	Energy Correction	✓
Resolution	Energy Resolution	EG_RESOLUTION_ALL	Energy Correction	✓
<b>Hadronic Taus</b>				
Efficiencies	Reconstruction	tauSFLoose_TAU_SF_RECO_TOTAL	Event Weight	✓
	Identification BDT	tauSFTight_TAU_SF_JETID_TOTAL	Event Weight	✓
	Electron Veto BDT	tauSFTight_TAU_SF_ELEOLR_TOTAL	Event Weight	✓
Scale Factor	$p_T$ Scale	TAUS_TRUEHADTAU_SME_TES_MODEL	$p_T$ Correction	✓
		TAUS_TRUEHADTAU_SME_TES_DETECTOR	$p_T$ Correction	✓
		TAUS_TRUEHADTAU_SME_TES_INSITU	$p_T$ Correction	✓

Table 4.18: Summary of experimental systematics in the analysis for muons, electrons and hadronic tau objects. From left: type, description, name of systematics in the code, mode of application and status of inclusion in the analysis. The mode of application indicates the systematic evaluation: overall event re-weighting (Event Weight) or rescaling (e.g.  $p_T$  Correction).

2662 **4.7.2 Signal and background modelling theoretical**  
 2663 **systematics**

2664 The uncertainties concerning on the cross section, parton shower and generator  
 2665 variation for the main background relying on the MC prediction are summarised  
 2666 in Table 4.19. The systematic uncertainties related to the data driven fakes are  
 2667 discussed in the Section 4.4.2.

Process (default)	X-section [%]	Generator (alternative)	Parton Shower (alternative)	Scale uncertainty
$t\bar{t}H$ (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF( $+\alpha_S$ ): $\pm 3.6$		(aMC@NLO+Herwig++)	event-by-event weight
$t\bar{t}Z$ (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF( $+\alpha_S$ ): $\pm 4$	(Sherpa)		event-by-event weight event-by-event weight
$t\bar{t}W$ (aMC@NLO+Pythia8)	QCD Scale: $^{+12.9}_{-11.5}$ PDF( $+\alpha_S$ ): $\pm 3.4$	(Sherpa)		event-by-event weight event-by-event weight
Diboson (Sherpa 2.2.1)	$\pm 50$			event-by-event weight event-by-event weight

Table 4.19: Summary of theoretical uncertainties for  $t\bar{t}H$ ,  $t\bar{t}V$  and diboson MC predictions.

Experimental Systematics on $b$ -tagged jets				
Type	Origin	systematics Name		Analysis
b-tags				
Scale Factors	MV2c10 b-tagger efficiency on b originated jets in bins of $\eta$	MC2c10_70_EventWeight_B0-5		✓
	MC2c10 b-tagger efficiency on c originated jets in bins of $\eta$	MC2c10_70_EventWeight_C0-3		✓
	MC2c10 b-tagger efficiency on light flavoured originated jets in bins of $\eta$ and $p_T$	MC2c10_70_EventWeight_Light0-11		✓
	MC2c10 b-tagger extrapolation efficiency	MC2c10_70_EventWeight_extrapolation MC2c10_70_EventWeight_extrapolation_from_charm		✓ ✓

Table 4.20: Summary of experimental systematics for the  $b$ -tagged jets in the analysis, using the MC2c10 tagging algorithm used at the 70% Working Point. All of the b-tagging related systematics are applied as event weights. From left: type, description, name of systematic in the code and status of inclusion in the analysis.

Table 4.21: Sources of systematic uncertainties considered in the analysis. "N" means that the uncertainty is taken as normalisation-only for all processes and channels affected, whereas "S" denotes systematics that are considered shape-only in all processes and channels. "SN" means that the uncertainty is taken on both shape and normalisation. Some of the systematic uncertainties are split into several components("Comp"), as indicated by the number in the rightmost column.

Systematic uncertainty	Type	Comp
Luminosity	N	1
Pile-Up reweighting	SN	1
<b>Physics Objects</b>		
Electron	SN	6
Muon	SN	15
Jet energy scale and resolution	SN	28
Jet vertex fraction	SN	1
Jet flavour tagging	SN	126
$E_T^{miss}$	SN	3
Total (Experimental)	–	181
<b>Data-driven non-prompt/fake leptons and charge misassignment</b>		
Control region statistics	SN	38
Light lepton efficiencies	SN	22
Non-prompt light lepton estimates: non-closure	N	5
$\gamma$ -conversion fraction	N	5
Electron charge misassignment	SN	1
Total (Data driven reducible background)	–	71
<b><math>t\bar{t}H</math> modelling</b>		
Cross section (QCD, PDF)	N	2
QCD scale	S	3
Parton shower	SN	1
Higgs branching ratio	N	4
Shower tune	SN	1
<b><math>t\bar{t}W</math> modelling</b>		
Cross section (QCD, PDF)	N	2
QCD scale	S	3
Generator	SN	1
Shower tune	SN	1
<b><math>t\bar{t}Z</math> modelling</b>		
Cross section (QCD, PDF)	N	2
QCD scale	S	3
Generator	SN	1
Shower tune	SN	1
<b>Other background modelling</b>		
Cross section	N	15
QCD Scale	SN	1
Total (Signal and background modelling)	–	41
Total (Overall)	–	293

## 4.8 Statistical treatment

The simultaneous fit is used to drive the signal significance and quantify the agreement between the observed data and MC prediction in the signal region for the  $t\bar{t}H$  three leptons final state.

A statistical analysis tool called *ttHFitter* [64] is employed. It is implemented with *HistFitter* [65] and the *RooFit/RooStat* [66]. The method is based on profile-likelihood ratio test (more details are in the Section 4.8.1). The signal and background are described by binned probability density function (PDF) defined in the statistical independent regions and implemented as histograms.

In this thesis, the combined BDTG shape is used as the input for the fitting to have shape information. In the final fit configuration, the signal strength  $\mu_{t\bar{t}H}$  is the only parameter of interest (POI) (scaling factors which are not constrained, used to adjust the relative contribution of the main background and signal components) and is extracted from the shape fit on combined BDTG with 6 autobins configuration. The considered nuisance parameter (NP)(scaling factors with external constrained used to model all statistical and systematic uncertainties) are listed in Table 4.21.

### 4.8.1 Likelihood function

The likelihood -  $\mathcal{L}(\mu, \theta)$  is a function of the signal strength  $\mu$  and of set of nuisance parameters  $\theta = \theta_a, \theta_b, \dots$ , which represent the uncertainties. The signal strength  $\mu$  ( $\mu_{t\bar{t}H} = \frac{\sigma \times BR}{(\sigma \times BR)_{SM}}$ ) is the parameter of interest that needs to be determined and is a free parameter of the fit.  $\mu = 0$  corresponds to a background only hypothesis and  $\mu = 1$  refers to the SM signal and background hypothesis. The real value of these parameters are unknown and are estimated by maximizing the likelihood in the fit. Generally the likelihood function describes the analysis with a product of the Poisson terms for numbers of events in bins and Gaussian term for systematic uncertainties as follows:

$$\begin{aligned} \mathcal{L} &= P_{SR} \times P_{CR} \times C_{Syst} \\ &= \prod_{i=1}^{N_{SR}} \prod_{j=1}^{N_{bins}} P(n_{SR}^{ij} | \lambda_{SR}(\mu_{sig}, \vec{b}, \vec{\theta})) \times \prod_{i=1}^{N_{CR}} \prod_{j=1}^{N_{bins}} P(n_{CR}^{ij} | \lambda_{CR}(\mu_{sig}, \vec{b}, \vec{\theta})) \\ &\quad \times C_{Syst}(\vec{\theta}). \end{aligned} \quad (4.17)$$

The first two items are the Poisson items for the numbers of observed events in the signal region ( $n_{SR}^{ij}$ ) and the numbers of events in each control region ( $n_{CR}^{ij}$ ). The index  $i$  and  $j$  are accounting for the number of signal region or control regions and numbers of considered bins of the discriminant variables, respectively. The expectation value ( $\lambda_{SR,CR}$ ) for Poisson is written as:

$$\lambda_{SR,CR}(\mu_{sig}, \vec{b}, \vec{\theta}) = \mu_{sig} n_{sig}^{ij}(\vec{\theta}) + n_{bkg}^{ij}(\vec{\theta}, \vec{b}), \quad (4.18)$$

2700 where  $\mu_{sig}$ : the free parameter signal strength;  $\vec{b} = \sum_i b_i$  are the background  
 2701 normalization factors for the background sources  $i$ ;  $\vec{\theta} = \theta_1, \dots, \theta_N$  are the nuisance  
 2702 parameters used to parametrize the systematic uncertainties like the luminosity,  
 2703 etc. The  $n_{sig}^{ij}$  and  $n_{bkg}^{ij}$  are the number of signal events and background events in  
 2704 the  $i$  signal (control) region and bin index  $j$ .

2705 The different sources of the systematics are included using the PDF  $C_{Syst}(\vec{\theta})$ ,  
 2706 which is the product of the probability distribution describing each source of sys-  
 2707 tematic uncertainty. Generally the PDF is a Gaussian with a width equal to 1:

$$C_{Syst}(\vec{\theta}) = \prod_{k=1}^{N_\theta} G(\theta_j). \quad (4.19)$$

2708 where  $k$  is the number of the nuisance parameters. The impact of nuisance pa-  
 2709 rameters on the expectation value is described by the  $\lambda_{SR,CR}$ .

## 2710 4.8.2 Hypothesis testing

2711 We can not get the information on the level of agreement between data and the  
 2712 background only or signal plus background hypothesis through the estimation of  
 2713 POI  $\mu_{t\bar{t}H}$ . Therefore, the hypothesis testing [67] used in the analysis is based on  
 2714 profiled likelihood ratio,  $\tilde{\lambda}(\mu)$ , defined as:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta}(\hat{\mu}))}, & \hat{\mu} \geq 0; \\ \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(0, \hat{\theta}(0))}, & \hat{\mu} < 0. \end{cases} \quad (4.20)$$

2715 The numerator maximize the likelihood for a specific value of  $\mu$  and denominator  
 2716 is the unconditional maximized likelihood when  $\mu \geq 0$ . The  $\mu < 0$  is refer to  
 2717 avoiding complications in the computations in the case of deficit of signal-like  
 2718 events. Depending on the tested hypothesis, different test statistics  $\tilde{t}_\mu$  are defined  
 2719 to test a particular of  $\mu$ . For the observed data, the test statistics is called  $\tilde{t}_\mu^{obs}$ . The  
 2720 sampling distribution of the test statistics is denoted by  $f(\tilde{t}_\mu | \mu, \hat{\theta}(\mu))$ . The  $p$ -value  
 2721 for a given observation, under a particular hypothesis, is the probability for an  
 2722 equal or more extreme outcome than the observed, under the assumed hypothesis:

$$p_\mu = \int_{\tilde{t}_\mu^{obs}}^{\infty} f(\tilde{t}_\mu | \mu, \hat{\theta}(\mu)) d\tilde{t}_\mu. \quad (4.21)$$

2723  $p$ -value with a small value presents evidence against the test hypothesis and is  
 2724 often converted into the normal significance  $Z$  value.



2725 The test statistics  $\tilde{q}_0$  is defined as follows in case of any excess in data:

$$\tilde{q}_0 = \begin{cases} -2\ln\tilde{\lambda}(0), & \hat{\mu} > 0; \\ 0, & \hat{\mu} \leq 0. \end{cases} \quad (4.22)$$

2726 To test the compatibility of the data with the background only hypothesis ( $\mu =$   
2727 0):

$$p_0 = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_0|0, \hat{\theta}(0))d\tilde{q}_0, \quad (4.23)$$

2728 where  $p_0$  is the probability the background only hypothesis leads to a test statistics  
2729 equal to or larger than observed. A discovery is declared if the background only  
2730 hypothesis is rejected at the  $5\sigma$  level.

2731 Besides a test statistics  $\tilde{q}_\mu$ :

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu), & \hat{\mu} \geq \mu; \\ 0, & \hat{\mu} < \mu. \end{cases} \quad (4.24)$$

2732 is used to for setting upper limits in the  $\mu$ . The Confidence Level of signal ( $CL_s$ )  
2733 is employed to set these limits.

$$p_\mu = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}(\mu))d\tilde{q}_\mu, \quad p_b = \int_{\infty}^{\tilde{q}_0^{obs}} f(\tilde{q}_0|0, \hat{\theta}(0))d\tilde{q}_0, \quad (4.25)$$

2734 where  $p_b$  is the  $p_\mu$  under the background-only hypothesis. The 95% CL upper  
2735 limits on  $\mu$  are given for

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \leq 5\%. \quad (4.26)$$

2736 To compute  $p$ -value related to a certain hypothesis, the full distribution of the  
2737 test statistics  $\tilde{t}_\mu$  needs to be determined. By constructing pseudo-experiments of  
2738 the hypothesis and calculating the test statistics  $\tilde{t}_\mu$ . An approximation of the  
2739  $f(\tilde{t}_\mu|\mu, \hat{\theta}(\mu))$  can be obtained in the limit of large statistics by using the Wald's  
2740 approximation [68]. A "Asimov" dataset is used to test the signal plus background  
2741 hypothesis to get the expected significance in the thesis. The "Asimov" dataset is  
2742 an artificial dataset which can replace the ensemble testing performed with MC  
2743 pseudo-experiment with single dataset. The dataset is defined in a way to return  
2744 exactly true value for each estimated parameter and is also used to get the error  
2745 bands on the median expected limit on  $\mu$ . This dataset can study the constraints  
2746 on nuisance parameters which are obtained with expected data and statistical  
2747 uncertainties.

2748 To summarise, the profiled likelihood method can give a picture of some of the  
2749 systematic uncertainties by fitting the data and a good agreement between data  
2750 and MC can be achieved through the shift (pull) of a given systematic uncertainty.  
2751 Nuisances parameters can reveal correlations among themselves during the likeli-

2752 hood maximisation including further reduction of the effect of the total systematic  
2753 uncertainties.

2754 The binning of the templates used in the fit model is a subtle parameter of  
2755 the fitting. The choice is from a compromise between the best separation for the  
2756 classifier shape and available amount of events in data and in MC. Extremely  
2757 narrow bins could cause a lack of background events in the signal enriched side  
2758 of classifier distribution. On the other hand, a large width binning could reduce  
2759 the discriminating power of the classifier by mixing the shape between signal and  
2760 background. In this thesis, an auto-binning configuration of *ttHFitter* is adopted  
2761 with the optimized parameters  $b$  and  $s$  (mentioned in Section 4.5.2) to have good  
2762 separation and balance between signal and background as well as to avoid the low  
2763 statistics bins during the fit.

## 5 Results with 13 TeV Run 2 data

### 5.1 Asimov fit and expected significance

The *ttHFitter* based on profile-likelihood is used to drive the final fit results as well as the  $\mu_{t\bar{t}H}$  evaluation which is performed at this stage of the analysis. The binned combined BDT output distributions are taken as the input to the fit. All the related systematic uncertainties mentioned in Section 4.7 are considered. Data-driven fakes,  $t\bar{t}V$  and diboson background are constrained simultaneously by the fit. Expected results are performed with Asimov dataset first and then the full dataset is used for the final fit. The pre-fit event yields are shown in Table 5.1, and Table 5.2 gives a summary on expected value on  $t\bar{t}H$  signal strength and expected significance for the  $3\ell$  channel, by fitting on Asimov dataset. The results include statistical and systematic uncertainties discussed in previous sections.

The ranking plots are shown in Figure 5.1 and the correlation between NPs is provided in Figure 5.2.

The nuisance parameters corresponding to systematic uncertainties are all centred on zero and the normalisation scale factors are all centred around 1 as expected. For what concerns the analysis sensitivity, the leading source of systematic uncertainties are those that show a high level of correlation with the signal strength.

The impact of a nuisance parameter on the fit  $\mu_{t\bar{t}H}$  is calculated by the fixing the corresponding nuisance parameter at a level of  $\theta \pm \sigma_\theta$  and perform the fit again.  $\theta$  is the fitted value of the nuisance parameter and  $\sigma_\theta$  is the post-fit uncertainty. The difference between the default and modified  $\mu$ ,  $\delta\mu$ , means the effect on the  $\mu$  of this particular systematic uncertainty.

Especially in the  $3\ell$  channel, the high ranking NPs are those which are related to the fakes estimation, such as the non-closure of the fakes in the  $3\ell$ , statistics of the fake muon and also the difference between the  $2\ell$  and  $3\ell$  because of the photon conversion fakes. Generally, there is no over-constrain issue in the  $3\ell$  channel.

### 5.2 Data fit

The fit has been performed to the data in one signal region and the results are presented in this section.

<b>Pre-fit expected event yields of the <math>3\ell</math> channel</b>	
<b>Main background</b>	
$t\bar{t}W$	$42.9 \pm 7.75$
$t\bar{t}Z$	$40.4 \pm 4.72$
$VV$	$19.7 \pm 14.3$
Non-prompt	$117 \pm 31.9$
<b>Rare background</b>	
Single top t-chan	$5 \times 10^{-5} \pm 3.51 \times 10^{-6}$
Single top s-chan	$5 \times 10^{-5} \pm 3.51 \times 10^{-6}$
$Wt$	$0.266 \pm 0.609$
Three top	$0.448 \pm 0.225$
Four top	$2.45 \pm 1.23$
$t\bar{t}WW$	$2.76 \pm 0.381$
low mass $t\bar{t}l$	$1.09 \pm 0.157$
$Z + jets$	$1.91 \pm 1.85$
$tZ$	$5 \times 10^{-5} \pm 2.51 \times 10^{-5}$
$WtZ$	$3.59 \pm 1.83$
rare top	$7.62 \pm 3.83$
$VVV$	$0.593 \pm 0.317$
$VH$	$0.691 \pm 1.15$
$tHjb (H \rightarrow WW)$	$0.367 \pm 0.0536$
$tHjb (H \rightarrow ZZ)$	$0.0456 \pm 0.0133$
$tHjb (H \rightarrow \tau\tau)$	$0.0479 \pm 0.0116$
$tHjb (H \rightarrow \text{Others})$	$0.0549 \pm 0.0163$
$WtH (H \rightarrow WW)$	$0.855 \pm 0.105$
$WtH (H \rightarrow ZZ)$	$0.0875 \pm 0.0192$
$WtH (H \rightarrow \tau\tau)$	$0.271 \pm 0.0495$
$WtH (H \rightarrow \text{Others})$	$0.0153 \pm 0.00870$
<b>Signal</b>	
$t\bar{t}H (H \rightarrow WW)$	$17.7 \pm 1.78$
$t\bar{t}H (H \rightarrow ZZ)$	$1.14 \pm 0.124$
$t\bar{t}H (H \rightarrow \tau\tau)$	$3.75 \pm 0.468$
$t\bar{t}H (H \rightarrow \text{Others})$	$0.492 \pm 0.412$
<b>Total</b>	$266 \pm 34.3$

Table 5.1: Pre-fit expected event yields of the  $3\ell$  channel.

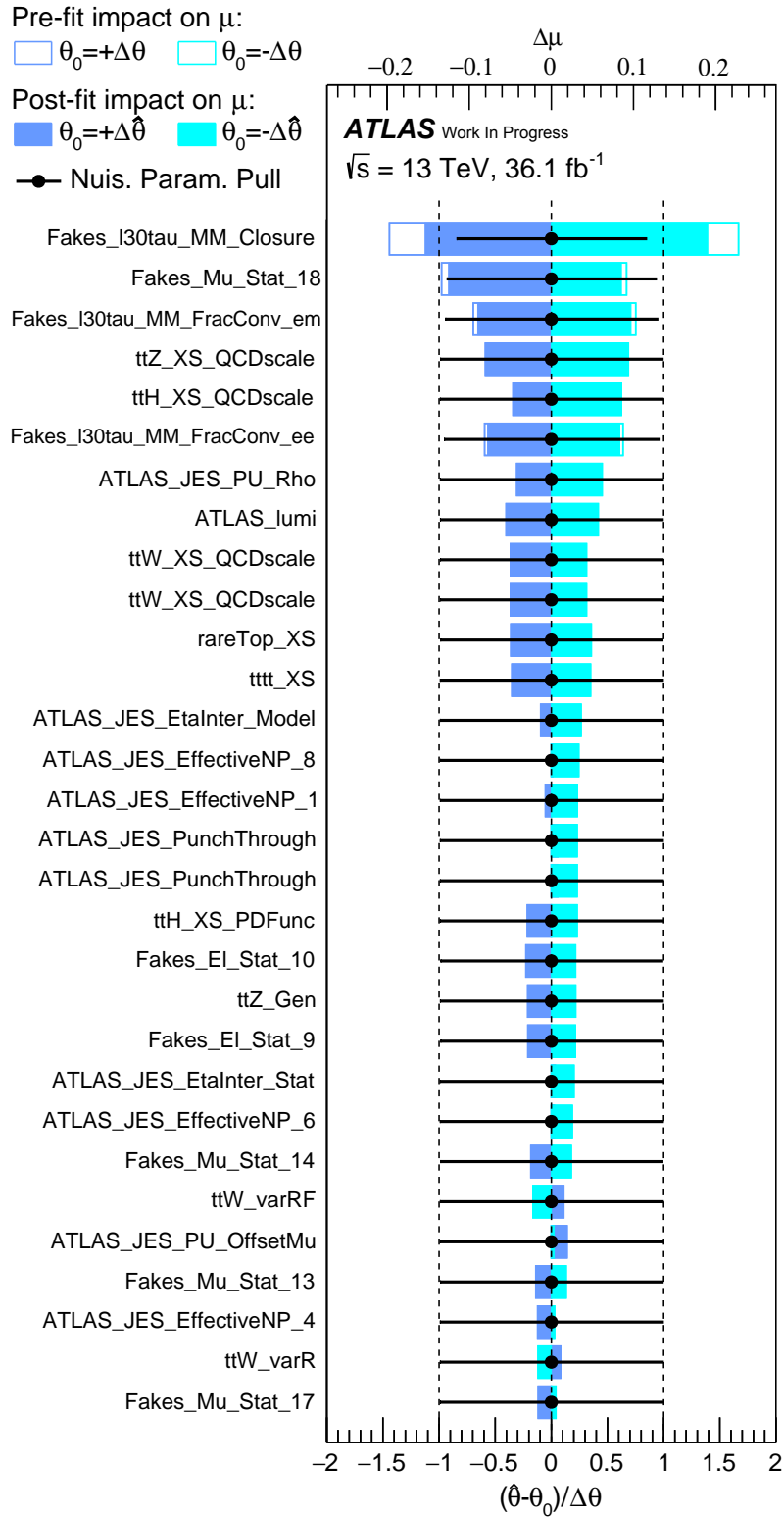


Figure 5.1: Ranking plot of the nuisance parameters from the Asimov dataset fitting.

	$\mu$ with total uncertainty	Expected significance ( $\sigma$ )
$3\ell$	$1^{+0.8}_{-0.7}$	$1.5 \pm 1.2$

Table 5.2: Summary on expected signal strength with all uncertainties and expected significance from the Asimov fit.

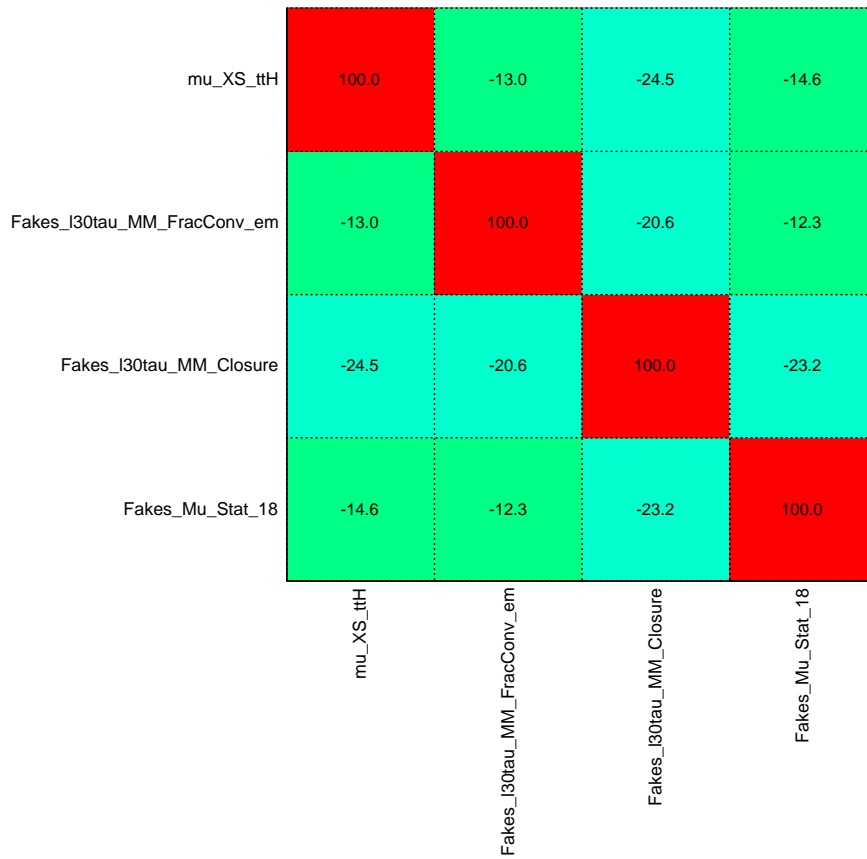


Figure 5.2: Correlation matrix of the nuisance parameters for Asimov dataset fitting.

2795 Table 5.3 shows the event yields comparison between MC and real data after  
 2796 the fit. Table 5.4 shows the observed signal strength and observed significance  
 2797 with respect to the null hypothesis by using the total Lumi  $36.1 \text{ fb}^{-1}$  data with  
 2798 the total uncertainties (which including the statistical uncertainty and systematic  
 2799 uncertainty). The results show that the observed significance with respect to the  
 2800 null hypothesis has reached 2.18 standard deviations which is much better than the  
 2801 Run 1 measurement in the  $3\ell$ , but still large uncertainties need to be considered  
 2802 at the same time.

2803 The distribution of combined BDT after fitting to the data is shown in Figure 5.3.  
 2804 The relative uncertainties decrease significantly due to the constraints provided by  
 2805 the data.

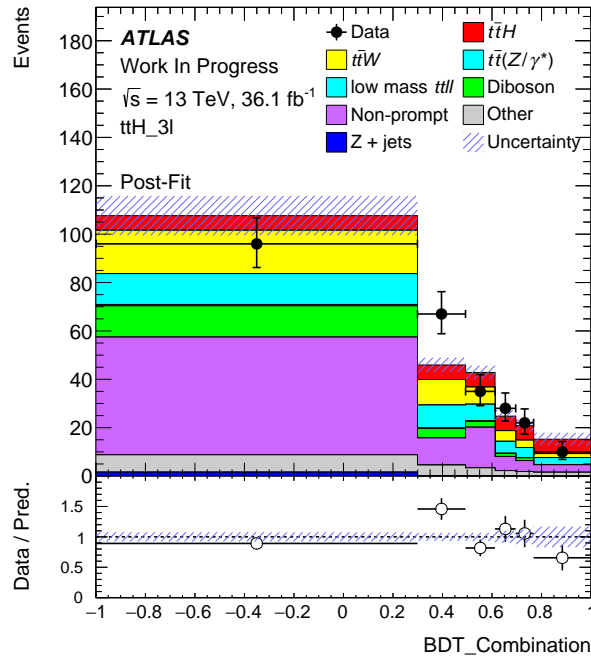


Figure 5.3: The combined BDT distribution after data fitting.

2806 The Figure 5.4 shows the ranking of the NP after data fitting. The largest impact  
 2807 on the signal strength fitting comes from the fakes estimation, the uncertainties  
 2808 due to the non-closure of the fake estimation.

<b>Post-fit event yields of the <math>3\ell</math> channel</b>	
<b>Main background</b>	
$t\bar{t}W$	$44.7 \pm 11.2$
$t\bar{t}Z$	$41.2 \pm 4.64$
$VV$	$21.6 \pm 20.5$
Non-prompt	$90.2 \pm 17.7$
<b>Rare background</b>	
Single top t-chan	$6.01 \times 10^{-5} \pm 4.06 \times 10^{-6}$
Single top s-chan	$6.01 \times 10^{-5} \pm 4.06 \times 10^{-6}$
$Wt$	$0.260 \pm 0.664$
Three top	$0.451 \pm 0.225$
Four top	$2.48 \pm 1.23$
$t\bar{t}WW$	$2.76 \pm 0.379$
low mass $tll$	$1.11 \pm 0.155$
$Z$ + jets	$1.86 \pm 1.83$
$tZ$	$6.01 \times 10^{-5} \pm 3.00 \times 10^{-5}$
$WtZ$	$3.70 \pm 1.84$
rare top	$7.94 \pm 3.88$
$VVV$	$0.603 \pm 0.318$
$VH$	$0.695 \pm 0.873$
$tHjb (H \rightarrow WW)$	$0.369 \pm 0.0534$
$tHjb (H \rightarrow ZZ)$	$0.0461 \pm 0.0134$
$tHjb (H \rightarrow \tau\tau)$	$0.0483 \pm 0.0116$
$tHjb (H \rightarrow \text{Others})$	$0.0553 \pm 0.0163$
$WtH (H \rightarrow WW)$	$0.858 \pm 0.105$
$WtH (H \rightarrow ZZ)$	$0.0878 \pm 0.0193$
$WtH (H \rightarrow \tau\tau)$	$0.271 \pm 0.0503$
$WtH (H \rightarrow \text{Others})$	$0.0154 \pm 0.00870$
<b>Signal</b>	
$t\bar{t}H (H \rightarrow WW)$	$27.5 \pm 13.3$
$t\bar{t}H (H \rightarrow ZZ)$	$1.77 \pm 0.866$
$t\bar{t}H (H \rightarrow \tau\tau)$	$5.85 \pm 2.85$
$t\bar{t}H (H \rightarrow \text{Others})$	$0.824 \pm 0.748$
Total	$257 \pm 19.4$
<b>Data</b>	<b>258</b>

Table 5.3: Post-fit event yields of the  $3\ell$  channel.



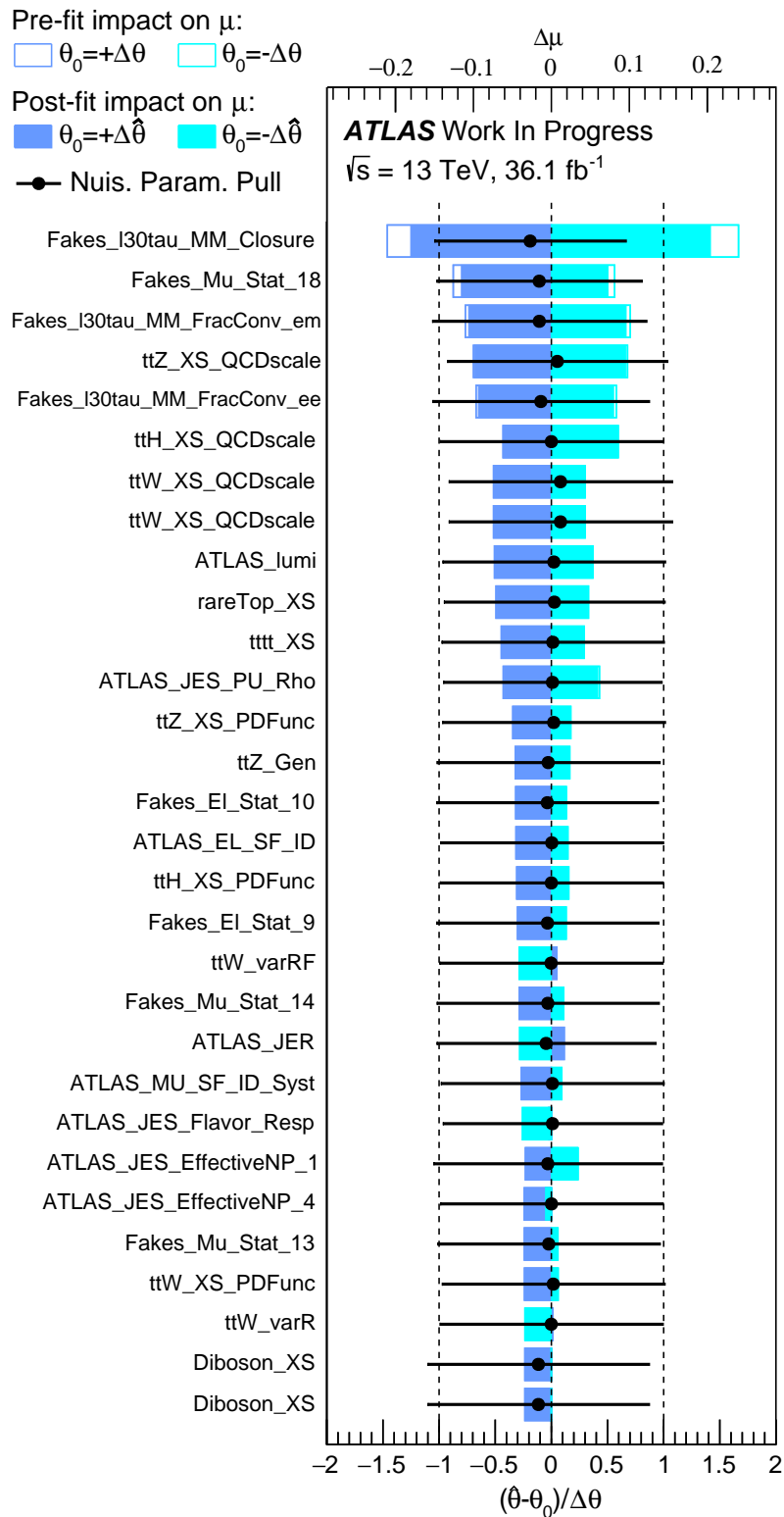


Figure 5.4: Ranking plot of the nuisance parameters to the fit.

	$\mu$ with total uncertainty	Observed Significance ( $\sigma$ )
$3\ell$	$1.5^{+0.8}_{-0.7}$	$2.2 \pm 1.4$

Table 5.4: Summary on observed signal strength with all uncertainties and the corresponding significance with respect to the null hypothesis from data.

## 5.3 Interpretation

2809

2810 The results shown in the above section with the two dimensional BDT cut in a  
 2811 single signal region show the promising outcomes compared to Run 1  $t\bar{t}H$  three  
 2812 leptons final state study. The higher cross section and advanced analysis strategy  
 2813 bring the significant improvements. Especially the optimised analysis in the  $3\ell$   
 2814 channel presented in this thesis can have a better statistics of the signal region  
 2815 and good description of both MC and the data-driven estimation.

2816 The ATLAS has released the approved results with  $36.1 \text{ fb}^{-1}$  data in the multi-  
 2817 lepton channel including the  $3\ell$ . Especially in the  $3\ell$  channel, a multi-class analysis  
 2818 strategy which is based on the multi-variate analysis as well is adopted. The most  
 2819 difference between the method used in this thesis and the multi-class method is  
 2820 the categorization procedure applied on final BDT output to have a pure signal  
 2821 and control region. The multi-class analysis has one signal region and the other  
 2822 four control regions which cover the  $t\bar{t}W$ ,  $t\bar{t}Z$ , non-prompt and diboson process.  
 2823 All those regions are based on a 5D BDT training output.

2824 The observed (expected) best-fit value of  $\mu_{t\bar{t}H}$ , combining all multilepton chan-  
 2825 nels, is  $1.56^{+0.30}_{-0.29}$  (stat)  $^{+0.39}_{-0.30}$  (syst) =  $1.56^{+0.49}_{-0.42}$  ( $1.00^{+0.29}_{-0.28}$  (stat)  $^{+0.32}_{-0.27}$  (syst) =  $1.00^{+0.43}_{-0.39}$ ).  
 2826 The best-fit value of  $\mu_{t\bar{t}H}$  for each individual channel and the combination of all  
 2827 channels are shown in Figure 5.5 and Table 5.5. The consistency of the fitted signal  
 2828 strengths between the seven channels is 34%. The observed (expected) significance  
 2829 associated with respect to the background hypothesis is  $4.1\sigma$  ( $2.7\sigma$ ). This excess  
 2830 provides evidence for  $t\bar{t}H$  production in the multilepton final states. The observed  
 2831 significance associated with the SM expectation  $\mu_{t\bar{t}H} = 1$  is  $1.4\sigma$ .

Table 5.5: Observed and expected best-fit values of the signal strength  $\mu_{t\bar{t}H}$  and associated significance with respect to the null hypothesis. The expected values are shown for the pre-fit background estimates.

Channel	Best fit $\mu_{t\bar{t}H}$ (observed)		Best fit $\mu_{t\bar{t}H}$ (expected)		Observed (expected) significance
$2\ell$	1.51	$^{+0.43}_{-0.41}$ (stat.) $^{+0.50}_{-0.41}$ (syst.)	1.00	$^{+0.43}_{-0.40}$ (stat.) $^{+0.47}_{-0.40}$ (syst.)	$2.8\sigma$ ( $1.8\sigma$ )
$3\ell$	1.76	$^{+0.61}_{-0.57}$ (stat.) $^{+0.60}_{-0.50}$ (syst.)	1.00	$^{+0.57}_{-0.53}$ (stat.) $^{+0.51}_{-0.44}$ (syst.)	$2.4\sigma$ ( $1.5\sigma$ )
$4\ell$	-0.51	$^{+1.35}_{-0.81}$ (stat.) $^{+0.29}_{-0.29}$ (syst.)	1.00	$^{+1.75}_{-1.15}$ (stat.) $^{+0.43}_{-0.22}$ (syst.)	- ( $0.8\sigma$ )
$1\ell 2\tau$	-0.58	$^{+1.07}_{-0.86}$ (stat.) $^{+1.15}_{-1.33}$ (syst.)	1.00	$^{+1.31}_{-1.10}$ (stat.) $^{+1.45}_{-1.29}$ (syst.)	- ( $0.6\sigma$ )
$2\ell 1\tau$	3.68	$^{+1.47}_{-1.25}$ (stat.) $^{+0.99}_{-0.53}$ (syst.)	1.00	$^{+1.13}_{-0.91}$ (stat.) $^{+0.52}_{-0.31}$ (syst.)	$3.5\sigma$ ( $1.1\sigma$ )
$2\ell OS1\tau$	1.74	$^{+1.56}_{-1.45}$ (stat.) $^{+1.38}_{-1.16}$ (syst.)	1.00	$^{+1.53}_{-1.42}$ (stat.) $^{+1.29}_{-1.15}$ (syst.)	$0.9\sigma$ ( $0.5\sigma$ )
$3\ell 1\tau$	1.61	$^{+1.71}_{-1.27}$ (stat.) $^{+0.64}_{-0.19}$ (syst.)	1.00	$^{+1.56}_{-1.12}$ (stat.) $^{+0.49}_{-0.23}$ (syst.)	$1.3\sigma$ ( $0.9\sigma$ )
Combined	1.56	$^{+0.30}_{-0.29}$ (stat.) $^{+0.39}_{-0.30}$ (syst.)	1.00	$^{+0.29}_{-0.28}$ (stat.) $^{+0.32}_{-0.27}$ (syst.)	$4.1\sigma$ ( $2.8\sigma$ )

2832 The combined fit on Asimov data with  $\mu_{t\bar{t}H} = 1$  of the no-tau channels  $2\ell SS$ ,  $3\ell$   
 2833 and  $4\ell$  gives an expected significance of 2.3 and errors on  $\mu_{t\bar{t}H}$  of  $^{+0.49}_{-0.45}$ . For the  $4\ell$ ,  
 2834  $1\ell 2\tau$ ,  $2\ell OS1\tau$  and  $3\ell 1\tau$  channels, the uncertainties on  $\mu$  are mainly statistical, while

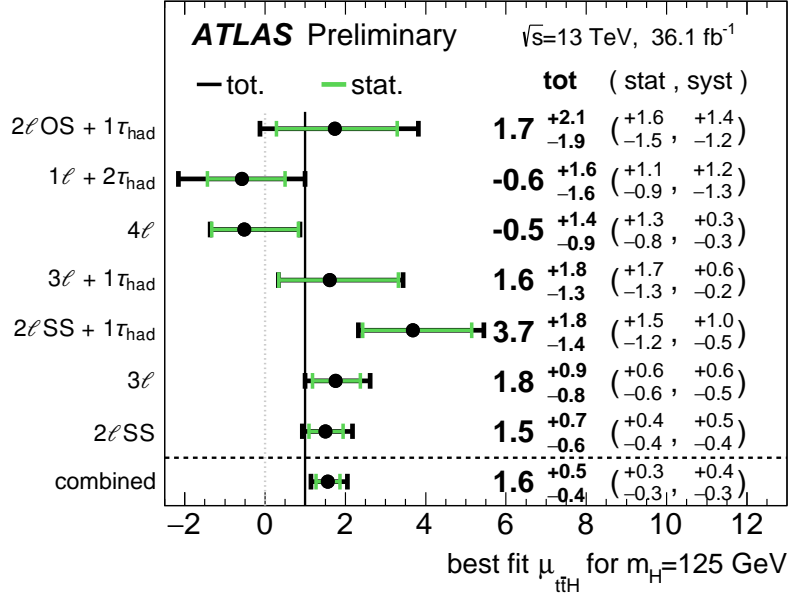


Figure 5.5: Observed best fit values of the  $t\bar{t}H$  signal strength  $\mu_{t\bar{t}H}$  and their uncertainties by final state category and combined. The SM prediction is  $\mu_{t\bar{t}H} = 1$ .

2835 the statistical and systematic uncertainties are of comparable size for the 2ℓ, 3ℓ  
 2836 and 2ℓ1τ channels. The uncertainties with highest impact on  $\Delta\mu_{t\bar{t}H}$  are currently  
 2837  $t\bar{t}H$  and  $t\bar{t}W$  cross-section QCD scale, jet energy scale, pile-up subtraction, JES  
 2838 flavour composition and several Fake rate statistical uncertainties.

2839 The results presented in the thesis are compatible with the above ATLAS ap-  
 2840 proval results. Especially in the 3ℓ, the best-fit value of  $\mu_{t\bar{t}H}$  is  $1.76^{+0.61}_{-0.57}$  (stat)  
 2841  $^{+0.60}_{-0.50}$  (syst), and the corresponding observed significance is  $2.4\sigma$  in the ATLAS ap-  
 2842 proved results. The non-prompt background estimation with the Matrix Method  
 2843 in this thesis is adopted in the ATLAS approval results. Besides, the improvement  
 2844 (MVA strategy and the treatment of the uncertainty) can be seen for the current  
 2845 multi-variate analysis used in the thesis.

# Conclusion

2847 Higgs boson is discovered in 2012 by the ATLAS and CMS. The last piece of  
2848 the Standard Model is found, and the structure of the Standard Model is almost  
2849 completed. But there are still many questions to be answered, and the Yukawa  
2850 coupling of the Higgs boson to the top quark is one of them. The top Yukawa  
2851 coupling influences the evolution of the effective Higgs potential with the energy:  
2852 any tension between the values allowed by the Standard Model and the observation  
2853 would demand for new physics to solve the inconsistency. Precise measurements  
2854 of its properties will confirm its nature, and any significant deviations from the  
2855 Standard Model prediction will represent a clear sign of new physics. The asso-  
2856 ciated production of the Higgs boson with a top quark pair allows for a direct  
2857 measurement and is the main topic of this dissertation.

2858 The work presented in this thesis focuses on a search for the Higgs boson pro-  
2859 duction associated with a pair of top quarks in three leptons final state. The  
2860 multi-variate technical (BDT) is employed to have a good discriminate power be-  
2861 tween signal and background. One of the main backgrounds, fake leptons, the  
2862 Matrix Method is used to estimate its contribution. Generally, good modelling of  
2863 the fake background and better separation with BDT output are achieved. Final  
2864 results are obtained with a simultaneous fit on a combined 2D BDT.

2865 For a Higgs boson mass of 125 GeV, an excess of events over the expected  
2866 background from the other Standard Model processes is found with an observed  
2867 significance of 2.2 standard deviations, compared to an expectation of 1.5 standard  
2868 deviations. The best fit for the  $t\bar{t}H$  production cross section, assuming a Higgs  
2869 boson mass of 125 GeV, is  $1.5^{+0.8}_{-0.7}$  times the SM expectation, and is consistent  
2870 with the value of the Yukawa coupling to top quarks in the Standard Model. The  
2871 results are compatible with the ATLAS approval results by using the same dataset  
2872 in three leptons final state and improvements can be introduced into the analysis  
2873 presented in this thesis.

2874 Compared to the Run 1 results, a great improvement is obtained for both the  
2875 significance and uncertainty on the signal strength measurement. The improve-  
2876 ments come from the high centre of mass energy which brings the higher cross  
2877 section, the upgrades of the ATLAS which bring lots of improvements, and also  
2878 the advanced analysis strategy using. The promising result shows the evidence of  
2879 the  $t\bar{t}H$  production but more data are need to have the more precise measurement.

2880 Uncertainties on the final measurements are still large, and can be re-optimized to  
2881 reduce the impact.

2882 After the restart of the LHC in 2015, larger statistics data brings further possi-  
2883 bilities to validate the predictions of the SM and to search for new physics. During  
2884 the current upgrades, the ATLAS detector is making upgrades to the limited sub-  
2885 systems. The LHC will shut down for next upgrades at the end of 2018 and  
2886 the ATLAS will make improvements to the trigger system in order to deal with  
2887 higher trigger rates. After restarting in 2019, the LHC will operate for three years  
2888 at  $\sqrt{s} = 14$  TeV and collect  $300 \text{ fb}^{-1}$  of data. The final LHC shut down will  
2889 be in 2023, during which time the ATLAS will install a new, higher granularity,  
2890 more radiation-hard inner tracker. A new triggering scheme, as well as improve-  
2891 ments to the muon and calorimeter electronics, will be made. The upgraded High-  
2892 Luminosity LHC (HL-LHC) will start in 2025, after which the collider is expected  
2893 to collect  $3000 \text{ fb}^{-1}$  of data. The new dataset will allow for more precise Higgs  
2894 measurements to have a further validation of the predictions by SM. With more  
2895 and more colliding data and advanced technical analysis methods, the picture of  
2896 the physics will be more clear or the discovery of new physics, which all need time  
2897 to search for.

# Bibliography

- 2899 [1] ATLAS Collaboration. “Observation of a new particle in the search for the  
2900 Standard Model Higgs boson with the ATLAS detector at the LHC.” In:  
2901 *Phys. Lett.* B716 (2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.](https://doi.org/10.1016/j.physletb.2012.08.020)  
2902 [020](https://arxiv.org/abs/1207.7214). arXiv: [1207.7214](https://arxiv.org/abs/1207.7214) [hep-ex] (cit. on p. 38).
- 2903 [2] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV  
2904 with the CMS experiment at the LHC.” In: *Phys. Lett.* B716 (2012), pp. 30–  
2905 61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235](https://arxiv.org/abs/1207.7235) [hep-ex]  
2906 (cit. on p. 38).
- 2907 [3] ATLAS Collaboration. “Combined Measurement of the Higgs Boson Mass  
2908 in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Ex-  
2909 periments.” In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: [10.1103/](https://doi.org/10.1103/PhysRevLett.114.191803)  
2910 [PhysRevLett.114.191803](https://arxiv.org/abs/1503.07589). arXiv: [1503.07589](https://arxiv.org/abs/1503.07589) [hep-ex] (cit. on p. 38).
- 2911 [4] ATLAS Collaboration. *Measurement of fiducial, differential and production*  
2912 *cross sections in the  $H \rightarrow \gamma\gamma$  decay channel with  $13.3 \text{ fb}^{-1}$  of 13 TeV proton-*  
2913 *proton collision data with the ATLAS detector.* Tech. rep. ATLAS-CONF-  
2914 2016-067. Geneva: CERN, 2016. URL: [http://cds.cern.ch/record/](http://cds.cern.ch/record/2206210)  
2915 [2206210](http://cds.cern.ch/record/2206210) (cit. on p. 38).
- 2916 [5] ATLAS Collaboration. “Standard Model Total Production Cross Section  
2917 Measurement.” In: (2016). URL: [https://atlas.web.cern.ch/Atlas/](https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html)  
2918 [GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html](https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html) (cit. on p. 38).
- 2919 [6] ATLAS Collaboration. “Measurements of the Higgs boson production and  
2920 decay rates and coupling strengths using  $pp$  collision data at  $\sqrt{s} = 7$  and  
2921 8 TeV in the ATLAS experiment.” In: *Eur. Phys. J.* C76.1 (2016), p. 6. DOI:  
2922 [10.1140/epjc/s10052-015-3769-y](https://doi.org/10.1140/epjc/s10052-015-3769-y) (cit. on p. 42).
- 2923 [7] CMS Collaboration. “Precise determination of the mass of the Higgs bo-  
2924 son and tests of compatibility of its couplings with the standard model  
2925 predictions using proton collisions at 7 and 8 TeV.” In: *Eur. Phys. J.* C75.5  
2926 (2015), p. 212. DOI: [10.1140/epjc/s10052-015-3351-7](https://doi.org/10.1140/epjc/s10052-015-3351-7). arXiv: [1412.](https://arxiv.org/abs/1412.8662)  
2927 [8662](https://arxiv.org/abs/1412.8662) [hep-ex] (cit. on p. 42).

## Bibliography

- 2928 [8] ATLAS Collaboration. “Measurements of the Higgs boson production and  
2929 decay rates and constraints on its couplings from a combined ATLAS and  
2930 CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV.” In: *JHEP*  
2931 08 (2016), p. 045. DOI: [10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045). arXiv: [1606.02266](https://arxiv.org/abs/1606.02266)  
2932 [[hep-ex](https://arxiv.org/abs/1606.02266)] (cit. on pp. 43, 44).
- 2933 [9] De Melis and Cinzia. “The CERN accelerator complex. Complexe des ac-  
2934 célérateurs du CERN.” In: (July 2016). URL: [https://cds.cern.ch/  
2935 record/2197559](https://cds.cern.ch/record/2197559) (cit. on p. 45).
- 2936 [10] Jean-Luc Caron. “Cross section of LHC dipole.. Dipole LHC: coupe transver-  
2937 sale.” May 1998. URL: <https://cds.cern.ch/record/841539> (cit. on  
2938 p. 47).
- 2939 [11] LHC. “LHC Luminosity in 2011, 2012.” In: (2017). URL: [https://twiki.  
2940 cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults) (cit.  
2941 on p. 49).
- 2942 [12] Joao Pequena. “Computer generated image of the whole ATLAS detec-  
2943 tor.” In: (Mar. 2008). URL: <https://cds.cern.ch/record/1095924> (cit.  
2944 on p. 50).
- 2945 [13] D. Parchet. “Axes Detecteur, Machine, Genie Cevil. technical drawing.”  
2946 In: (Dec. 1998). URL: [http://www.hep.lu.se/atlas/thesis/egede/  
2947 thesis-node39.html](http://www.hep.lu.se/atlas/thesis/egede/thesis-node39.html) (cit. on p. 52).
- 2948 [14] Joao Pequena. “Computer generated image of the ATLAS inner detector.”  
2949 In: (Mar. 2008). URL: <http://cds.cern.ch/record/1095926> (cit. on  
2950 p. 53).
- 2951 [15] Joao Pequena. “Computer generated image of the ATLAS Colarimeter.”  
2952 In: (Mar. 2008). URL: <http://cds.cern.ch/record/1095927> (cit. on  
2953 p. 54).
- 2954 [16] Joao Pequena. “Event Cross Section in a computer generated image of  
2955 the ATLAS detector.” In: (Mar. 2008). URL: [https://cds.cern.ch/  
2956 record/1096081](https://cds.cern.ch/record/1096081) (cit. on p. 54).
- 2957 [17] ATLAS Collaboration. “Commissioning of the ATLAS Muon Spectrometer  
2958 with Cosmic Rays.” In: *Eur. Phys. J. C* 70.arXiv:1006.4384, CERN-PH-EP-  
2959 2010-045 (Aug. 2010), 875–916. 58 p. URL: [http://cds.cern.ch/  
2960 record/1275998](http://cds.cern.ch/record/1275998) (cit. on p. 56).
- 2961 [18] ATLAS Collaboration. “A Fast Hardware Tracker for the ATLAS Trigger  
2962 System.” In: *Particles and fields. Proceedings, Meeting of the Division of  
2963 the American Physical Society, DPF 2011, Providence, USA, August 9-13,  
2964 2011*. 2011. arXiv: [1110.1910](https://arxiv.org/abs/1110.1910) [[hep-ex](https://arxiv.org/abs/1110.1910)]. URL: [http://inspirehep.net/  
2965 record/931205/files/arXiv:1110.1910.pdf](http://inspirehep.net/record/931205/files/arXiv:1110.1910.pdf) (cit. on p. 57).



## Bibliography

- 2966 [19] ATLAS Collaboration. *ATLAS Computing: technical design report*. Technical  
2967 Design Report ATLAS. Geneva: CERN, 2005. URL: [https://cds.cern.ch/  
2968 record/837738](https://cds.cern.ch/record/837738) (cit. on p. 57).
- 2969 [20] LHC. “Worldwide LHC Computing Grid.” In: (2017). URL: [http://wlcg.  
web.cern.ch/](http://wlcg.<br/>2970 web.cern.ch/) (cit. on p. 57).
- 2971 [21] ATLAS Collaboration. “The ATLAS Simulation Infrastructure.” In: *The Eu-  
2972 ropean Physical Journal C* 70.3 (2010), pp. 823–874. ISSN: 1434-6052.  
2973 DOI: [10.1140/epjc/s10052-010-1429-9](https://doi.org/10.1140/epjc/s10052-010-1429-9). URL: [http://dx.doi.org/10.  
1140/epjc/s10052-010-1429-9](http://dx.doi.org/10.<br/>2974 1140/epjc/s10052-010-1429-9) (cit. on p. 58).
- 2975 [22] S. Agostinelli and etc. “Geant4 – a simulation toolkit.” In: *Nuclear Instru-  
2976 ments and Methods in Physics Research Section A: Accelerators, Spectrom-  
2977 eters, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303.  
2978 ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)  
01368-8](https://doi.org/10.1016/S0168-9002(03)<br/>2979 01368-8). URL: [http://www.sciencedirect.com/science/article/pii/  
S0168900203013688](http://www.sciencedirect.com/science/article/pii/<br/>2980 S0168900203013688) (cit. on p. 58).
- 2981 [23] Graphic Design. “LHC consolidations 2013-14.” Feb. 2013. URL: [http://  
cds.cern.ch/record/1516031](http://<br/>2982 cds.cern.ch/record/1516031) (cit. on p. 59).
- 2983 [24] LHC. “LHC Luminosity in 2015, 2016.” In: (2017). URL: [https://twiki.  
cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2](https://twiki.<br/>2984 cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2<br/>2985)  
(cit. on p. 60).
- 2986 [25] W Lampl, S Laplace, D Lelas, et al. *Calorimeter Clustering Algorithms: De-  
2987 scription and Performance*. Tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-  
2988 LARG-2008-003. Geneva: CERN, Apr. 2008. URL: [https://cds.cern.ch/  
record/1099735](https://cds.cern.ch/<br/>2989 record/1099735) (cit. on p. 63).
- 2990 [26] T G Cornelissen, M Elsing, I Gavrilenko, et al. “The global  $\chi^2$  track  
2991 fitter in ATLAS.” In: *Journal of Physics: Conference Series* 119.3 (2008),  
2992 p. 032013. URL: <http://stacks.iop.org/1742-6596/119/i=3/a=032013>  
2993 (cit. on p. 63).
- 2994 [27] ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the  
2995 Gaussian Sum Filter-based model for bremsstrahlung*. Tech. rep. ATLAS-  
2996 CONF-2012-047. Geneva: CERN, May 2012. URL: [http://cds.cern.  
ch/record/1449796](http://cds.cern.<br/>2997 ch/record/1449796) (cit. on p. 64).
- 2998 [28] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS de-  
2999 tector using the 2012 LHC proton-proton collision data*. Tech. rep. ATLAS-  
3000 CONF-2014-032. Geneva: CERN, June 2014. URL: [http://cds.cern.ch/  
record/1706245](http://cds.cern.ch/<br/>3001 record/1706245) (cit. on p. 64).
- 3002 [29] ATLAS Collaboration. “Electron and photon energy calibration with the  
3003 ATLAS detector using LHC Run 1 data.” In: *Eur. Phys. J. C* 74.10 (2014),  
3004 p. 3071. DOI: [10.1140/epjc/s10052-014-3071-4](https://doi.org/10.1140/epjc/s10052-014-3071-4). arXiv: [1407.5063](https://arxiv.org/abs/1407.5063)  
3005 [[hep-ex](https://arxiv.org/abs/1407.5063)] (cit. on p. 64).

## Bibliography

- 3006 [30] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS de-*  
3007 *detector using the 2015 LHC proton-proton collision data.* Tech. rep. ATLAS-  
3008 COM-CONF-2016-028. Geneva: CERN, Mar. 2016. URL: <https://cds.cern.ch/record/2142831> (cit. on pp. 65, 68).  
3009
- 3010 [31] ATLAS Collaboration. *Expected photon performance in the ATLAS exper-*  
3011 *iment.* Tech. rep. ATL-PHYS-PUB-2011-007. Geneva: CERN, Apr. 2011.  
3012 URL: <http://cds.cern.ch/record/1345329> (cit. on p. 70).
- 3013 [32] ATLAS Collaboration. “Measurement of the photon identification efficien-
- 3014 cies with the ATLAS detector using LHC Run-1 data.” In: *Eur. Phys. J.*  
3015 C76.12 (2016), p. 666. DOI: [10.1140/epjc/s10052-016-4507-9](https://doi.org/10.1140/epjc/s10052-016-4507-9). arXiv:  
3016 [1606.01813](https://arxiv.org/abs/1606.01813) [hep-ex] (cit. on p. 71).
- 3017 [33] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS
- 3018 detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV.” In: *Eur. Phys. J.*  
3019 C76.5 (2016), p. 292. DOI: [10.1140/epjc/s10052-016-4120-y](https://doi.org/10.1140/epjc/s10052-016-4120-y). arXiv:  
3020 [1603.05598](https://arxiv.org/abs/1603.05598) [hep-ex] (cit. on p. 72).
- 3021 [34] ATLAS Collaboration. *Measurement of the tau lepton reconstruction and*  
3022 *identification performance in the ATLAS experiment using pp collisions at*  
3023  *$\sqrt{s} = 13$  TeV.* Tech. rep. ATLAS-CONF-2017-029. Geneva: CERN, May  
3024 2017. URL: <http://cds.cern.ch/record/2261772> (cit. on p. 73).
- 3025 [35] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti-k(t) jet
- 3026 clustering algorithm.” In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-](https://doi.org/10.1088/1126-6708/2008/04/063)  
3027 [6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189](https://arxiv.org/abs/0802.1189) [hep-ph] (cit. on p. 73).
- 3028 [36] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report.*  
3029 Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. 2010. URL: [https://](https://cds.cern.ch/record/1291633)  
3030 [cds.cern.ch/record/1291633](https://cds.cern.ch/record/1291633) (cit. on p. 74).
- 3031 [37] H. Pérez R., P. Kielanowski, and S. R. Juárez W. “Cabibbo–Kobayashi–Maskawa
- 3032 matrix: rephasing invariants and parameterizations.” In: *J. Phys. Conf. Ser.*  
3033 485 (2014), p. 012058. DOI: [10.1088/1742-6596/485/1/012058](https://doi.org/10.1088/1742-6596/485/1/012058). arXiv:  
3034 [1209.5812](https://arxiv.org/abs/1209.5812) [hep-ph] (cit. on p. 74).
- 3035 [38] ATLAS Collaboration. *Flavour tagging algorithms and performance at the*  
3036 *ATLAS experiment.* Tech. rep. ATL-PHYS-PROC-2016-105. Geneva: CERN,  
3037 2016. URL: <https://cds.cern.ch/record/2209195> (cit. on p. 74).
- 3038 [39] Ovidiu Miu. “JetFitter vertex reconstruction performance in the ATLAS
- 3039 experiment.” In: (2016). URL: <http://cds.cern.ch/record/2210576>  
3040 (cit. on p. 74).
- 3041 [40] ATLAS Collaboration. *Optimisation of the ATLAS b-tagging performance for*  
3042 *the 2016 LHC Run.* Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN,  
3043 June 2016. URL: <http://cds.cern.ch/record/2160731> (cit. on pp. 75,  
3044 76).

## Bibliography

- 3045 [41] *b*-tagging Benchmarks for taggers MV2c10 on DC14 8 TeV Samples (Feb  
3046 2014). Tech. rep. URL: [https://twiki.cern.ch/twiki/bin/view/  
3047 AtlasProtected/BTaggingBenchmarksRelease17](https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarksRelease17) (cit. on p. 75).
- 3048 [42] ATLAS Collaboration. *ATLAS Collaboration, b*-tagging performance plots in  
3049 a  $t\bar{t}$  dominated sample from early 2016 ATLAS data. Tech. rep. URL: [http:  
3050 //atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2016-001](http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2016-001)  
3051 (cit. on p. 76).
- 3052 [43] ATLAS Collaboration. “ATLAS jet and missing-ET reconstruction, calibra-  
3053 tion, and performance.” In: *Nucl. Part. Phys. Proc.* 273-275 (2016), pp. 1121–  
3054 1126. DOI: [10.1016/j.nuclphysbps.2015.09.176](https://doi.org/10.1016/j.nuclphysbps.2015.09.176) (cit. on p. 76).
- 3055 [44] ATLAS Collaboration. “Performance of algorithms that reconstruct missing  
3056 transverse momentum in  $\sqrt{s} = 8$  TeV proton-proton collisions in the AT-  
3057 LAS detector.” In: *Eur. Phys. J.* C77.4 (2017), p. 241. DOI: [10.1140/epjc/  
3058 s10052-017-4780-2](https://doi.org/10.1140/epjc/s10052-017-4780-2). arXiv: [1609.09324](https://arxiv.org/abs/1609.09324) [hep-ex] (cit. on p. 77).
- 3059 [45] Richard D. Ball et al. “Parton distributions with LHC data.” In: *Nucl. Phys.*  
3060 B867 (2013), pp. 244–289. DOI: [10.1016/j.nuclphysb.2012.10.003](https://doi.org/10.1016/j.nuclphysb.2012.10.003).  
3061 arXiv: [1207.1303](https://arxiv.org/abs/1207.1303) [hep-ph] (cit. on p. 81).
- 3062 [46] Michael H. Seymour and Andrzej Siodmok. “Constraining MPI models us-  
3063 ing  $\sigma_{eff}$  and recent Tevatron and LHC Underlying Event data.” In: *JHEP*  
3064 10 (2013), p. 113. DOI: [10.1007/JHEP10\(2013\)113](https://doi.org/10.1007/JHEP10(2013)113). arXiv: [1307.5015](https://arxiv.org/abs/1307.5015)  
3065 [hep-ph] (cit. on p. 81).
- 3066 [47] Simone Alioli, Paolo Nason, Carlo Oleari, et al. “NLO single-top production  
3067 matched with shower in POWHEG: s- and t-channel contributions.” In:  
3068 *JHEP* 09 (2009). [Erratum: *JHEP*02,011(2010)], p. 111. DOI: [10.1007/  
3069 JHEP02\(2010\)011](https://doi.org/10.1007/JHEP02(2010)011), [10.1088/1126-6708/2009/09/111](https://doi.org/10.1088/1126-6708/2009/09/111). arXiv: [0907.4076](https://arxiv.org/abs/0907.4076)  
3070 [hep-ph] (cit. on p. 81).
- 3071 [48] Emanuele Re. “Single-top  $W$ t-channel production matched with parton  
3072 showers using the POWHEG method.” In: *Eur. Phys. J.* C71 (2011), p. 1547.  
3073 DOI: [10.1140/epjc/s10052-011-1547-z](https://doi.org/10.1140/epjc/s10052-011-1547-z). arXiv: [1009.2450](https://arxiv.org/abs/1009.2450) [hep-ph] (cit.  
3074 on p. 81).
- 3075 [49] T. Gleisberg, Stefan. Hoeche, F. Krauss, et al. “Event generation with  
3076 SHERPA 1.1.” In: *JHEP* 02 (2009), p. 007. DOI: [10.1088/1126-6708/  
3077 2009/02/007](https://doi.org/10.1088/1126-6708/2009/02/007). arXiv: [0811.4622](https://arxiv.org/abs/0811.4622) [hep-ph] (cit. on p. 81).
- 3078 [50] Piotr Golonka and Zbigniew Was. “PHOTOS Monte Carlo: A Precision tool  
3079 for QED corrections in  $Z$  and  $W$  decays.” In: *Eur. Phys. J.* C45 (2006),  
3080 pp. 97–107. DOI: [10.1140/epjc/s2005-02396-4](https://doi.org/10.1140/epjc/s2005-02396-4). arXiv: [hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026)  
3081 [hep-ph] (cit. on p. 81).

## Bibliography

- 3082 [51] W. Beenakker, S. Dittmaier, M. Kramer, et al. “Higgs radiation off top  
3083 quarks at the Tevatron and the LHC.” In: *Phys. Rev. Lett.* 87 (2001), p. 201805.  
3084 DOI: [10.1103/PhysRevLett.87.201805](https://doi.org/10.1103/PhysRevLett.87.201805). arXiv: [hep-ph/0107081](https://arxiv.org/abs/hep-ph/0107081) [[hep-ph](#)]  
3085 (cit. on p. 80).
- 3086 [52] W. Beenakker, S. Dittmaier, M. Kramer, et al. “NLO QCD corrections to  
3087  $t$  anti- $t$  H production in hadron collisions.” In: *Nucl. Phys. B* 653 (2003),  
3088 p. 151. DOI: [10.1016/S0550-3213\(03\)00044-0](https://doi.org/10.1016/S0550-3213(03)00044-0). arXiv: [hep-ph/0211352](https://arxiv.org/abs/hep-ph/0211352)  
3089 [[hep-ph](#)] (cit. on p. 80).
- 3090 [53] S. Dawson, L. H. Orr, L. Reina, et al. “Associated top quark Higgs boson  
3091 production at the LHC.” In: *Phys. Rev. D* 67 (2003), p. 071503. DOI: [10.1103/PhysRevD.67.071503](https://doi.org/10.1103/PhysRevD.67.071503). arXiv: [hep-ph/0211438](https://arxiv.org/abs/hep-ph/0211438) [[hep-ph](#)] (cit. on  
3092 p. 80).  
3093
- 3094 [54] S. Dawson, C. Jackson, L. H. Orr, et al. “Associated Higgs production  
3095 with top quarks at the large hadron collider: NLO QCD corrections.” In:  
3096 *Phys. Rev. D* 68 (2003), p. 034022. DOI: [10.1103/PhysRevD.68.034022](https://doi.org/10.1103/PhysRevD.68.034022).  
3097 arXiv: [hep-ph/0305087](https://arxiv.org/abs/hep-ph/0305087) [[hep-ph](#)] (cit. on p. 80).
- 3098 [55] Yu Zhang, Wen-Gan Ma, Ren-You Zhang, et al. “QCD NLO and EW NLO  
3099 corrections to  $t\bar{t}H$  production with top quark decays at hadron collider.”  
3100 In: *Phys. Lett. B* 738 (2014), p. 1. DOI: [10.1016/j.physletb.2014.09.](https://doi.org/10.1016/j.physletb.2014.09.022)  
3101 [022](https://doi.org/10.1016/j.physletb.2014.09.022). arXiv: [1407.1110](https://arxiv.org/abs/1407.1110) [[hep-ph](#)] (cit. on p. 80).
- 3102 [56] S. Frixione, V. Hirschi, D. Pagani, et al. “Weak corrections to Higgs hadropro-  
3103 duction in association with a top-quark pair.” In: *JHEP* 09 (2014), p. 065.  
3104 DOI: [10.1007/JHEP09\(2014\)065](https://doi.org/10.1007/JHEP09(2014)065). arXiv: [1407.0823](https://arxiv.org/abs/1407.0823) [[hep-ph](#)] (cit. on  
3105 p. 80).
- 3106 [57] S. Frixione, V. Hirschi, D. Pagani, et al. “Electroweak and QCD corrections  
3107 to top-pair hadroproduction in association with heavy bosons.” In: *JHEP*  
3108 06 (2015), p. 184. DOI: [10.1007/JHEP06\(2015\)184](https://doi.org/10.1007/JHEP06(2015)184). arXiv: [1504.03446](https://arxiv.org/abs/1504.03446)  
3109 [[hep-ph](#)] (cit. on p. 80).
- 3110 [58] S. Heinemeyer et al. “Handbook of LHC Higgs Cross Sections: 3. Higgs  
3111 Properties.” In: (2013). DOI: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004). arXiv: [1307.1347](https://arxiv.org/abs/1307.1347)  
3112 [[hep-ph](#)] (cit. on p. 80).
- 3113 [59] LHC Higgs Cross Section Working Group. 2016. URL: [https://twiki.](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)  
3114 [cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV)  
3115 (cit. on p. 80).
- 3116 [60] J. Alwall et al. “The automated computation of tree-level and next-to-  
3117 leading order differential cross sections, and their matching to parton  
3118 shower simulations.” In: *JHEP* 07 (2014), p. 079. DOI: [10.1007/JHEP07\(2014\)](https://doi.org/10.1007/JHEP07(2014)079)  
3119 [079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: [1405.0301](https://arxiv.org/abs/1405.0301) [[hep-ph](#)] (cit. on p. 80).

## Bibliography

- 3120 [61] ATLAS Collaboration. “Search for the associated production of the Higgs  
3121 boson with a top quark pair in multilepton final states with the ATLAS  
3122 detector.” In: *Phys. Lett. B* 749 (2015), pp. 519–541. DOI: [10.1016/j.  
3123 physletb.2015.07.079](https://doi.org/10.1016/j.physletb.2015.07.079). arXiv: [1506.05988](https://arxiv.org/abs/1506.05988) [hep-ex] (cit. on p. 87).
- 3124 [62] ATLAS Collaboration. “Study of the material of the ATLAS inner detector  
3125 for Run 2 of the LHC.” In: (2017). arXiv: [1707.02826](https://arxiv.org/abs/1707.02826) [hep-ex] (cit. on  
3126 p. 97).
- 3127 [63] ATLAS Collaboration. “Luminosity determination in pp collisions at  $\sqrt{s}$   
3128 = 8 TeV using the ATLAS detector at the LHC.” In: *Eur. Phys. J. C* 76.12  
3129 (2016), p. 653. DOI: [10.1140/epjc/s10052-016-4466-1](https://doi.org/10.1140/epjc/s10052-016-4466-1). arXiv: [1608.  
3130 03953](https://arxiv.org/abs/1608.03953) [hep-ex] (cit. on p. 114).
- 3131 [64] ATLAS Collaboration. “ttHFitter information.” In: (2017). URL: [https://  
3132 twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter](https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter) (cit.  
3133 on p. 118).
- 3134 [65] ATLAS Collaboration. “HistFitter information.” In: (2017). URL: [http://  
3135 histfitter.web.cern.ch/histfitter/](http://histfitter.web.cern.ch/histfitter/) (cit. on p. 118).
- 3136 [66] ATLAS Collaboration. “Roofit information.” In: (2017). URL: [http://  
3137 roofit.sourceforge.net/intro.html](http://roofit.sourceforge.net/intro.html) (cit. on p. 118).
- 3138 [67] Glen Cowan, Kyle Cranmer, Eilam Gross, et al. “Asymptotic formulae for  
3139 likelihood-based tests of new physics.” In: *Eur. Phys. J. C* 71 (2011). [Erra-  
3140 tum: *Eur. Phys. J.*C73,2501(2013)], p. 1554. DOI: [10.1140/epjc/s10052-  
3141 011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), [10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z). arXiv: [1007.1727](https://arxiv.org/abs/1007.1727)  
3142 [physics.data-an] (cit. on p. 119).
- 3143 [68] Abraham Wald. “Tests of Statistical Hypotheses Concerning Several Pa-  
3144 rameters When the Number of Observations is Large.” In: *Transactions  
3145 of the American Mathematical Society* 54.3 (1943), pp. 426–482. ISSN:  
3146 00029947. URL: <http://www.jstor.org/stable/1990256> (cit. on  
3147 p. 120).

# 3148 Appendix

## 3149 **A The optimization of the cut-based electron** 3150 **identification in the ATLAS towards Run 2**

### 3151 **A.1 Introduction and motivation**

3152 The following part will introduce a optimization study of the cut-based electron  
3153 identification towards Run 2 condition at low transverse energy (5-20 GeV). The  
3154 background at lower energy is much higher due to photons and light pions and  
3155 this makes it necessary to optimize cut-based electron identification in this low  
3156 energy region because of the new situation in Run 2 (higher centre mass of energy,  
3157 more complex pile-up condition and etc.). Signal, background objects and Monte  
3158 Carlo samples are presented in Section A.2. Section A.3 shows the performance  
3159 of the LikelihoodPCA method while following sections are dedicated to the cut-  
3160 based optimization, changes in the TRT and their effects on cut-based electron  
3161 identification and the pile-up dependence. The last sections are the results and  
3162 conclusions.

### 3163 **A.2 Monte Carlo samples**

3164 The process of optimization begins with monitoring of different signal and back-  
3165 ground signatures. The proper signal samples are selected to study "good" quality  
3166 electrons, and a comparison is made with suitable background samples as well.  
3167 Variables are ranked depending on their signal-background discrimination power.  
3168 The signal MC samples used for the studies are  $J/\psi \rightarrow ee$  and  $Z \rightarrow ee$  events  
3169 with only good electrons. A variety of  $J/\psi \rightarrow ee$  with different truth  $p_T$  cuts are  
3170 available for the studies.  $J/\psi e3e3$ ,  $J/\psi e3e8$  and  $J/\psi e3e13$  have truth  $p_T$  cuts of  
3171 3 GeV on the first electron and 3 GeV, 8 GeV, and 13 GeV on the second electron  
3172 respectively. The  $Z \rightarrow ee$  has a dilepton filter which means that there is a  $p_T$  cut  
3173 of 15 GeV on both electrons.

3174 The minimum bias MC sample is used as background. The statistics in the  $p_T$   
3175 bin 15-20 GeV is very low for this sample. The JF17 sample is added in this bin  
3176 to cover this issue. The JF17 sample has a truth  $p_T$  cut at 17 GeV and hence  
3177 cannot be used for  $p_T < 17$  GeV. Any true electrons in these background samples

3178 are removed for the optimization studies. To further select good quality electrons,  
3179 a loose isolation cut is applied. This causes an efficiency changes about 5%.

3180 The kinematics of the electron,  $p_T$  and  $\eta$ , determine the distributions of the  
3181 variables included in the cut-based menu. This justifies a binning in those two  
3182 kinematic variables. The menu would be more effective if this binning is finer.  
3183 But a lack of statistics in finer bins introduces statistical artefacts and affects the  
3184 quality of the optimization. This also makes the truth cuts in the MC samples more  
3185 dangerous since the kinematics, and hence the variables themselves are biased. For  
3186 example, as mentioned above, the  $Z \rightarrow ee$  dilepton filter sample has truth  $p_T$  cuts  
3187 at 15 GeV each. Thus, these cannot be used for bins with  $p_T$  less than 15 GeV.  
3188 Beside the statistical reasons, the shapes of variables, like  $R_\eta$ , and the isolation  
3189 variables, are significantly different at high and low  $p_T$ . This makes it necessary  
3190 to optimize the low  $p_T$  range independently.

### 3191 **A.3 Performance of LikelihoodPCA**

3192 Performance of the LikelihoodPCA method with MC samples is presented in this  
3193 section. LikelihoodPCA is selected due to its good performance as a multi-variate  
3194 analysis method. Thus it can provide a baseline for the cut-based method. MVA  
3195 LikelihoodPCA method is as a reference here and main goal is to check the input  
3196 variables' performance and get a overview of MVA method, and one can also get a  
3197 better understanding of cut-based method by comparing different MVA methods'  
3198 performance.

3199 Electron ID uses many discriminating variables to select good electrons generally.  
3200 So the MVA can take advantage of the characteristics of these variables to employ  
3201 them to distinguish between signal and background more efficiently. This test will  
3202 be a cross check for the cut-based method and can be a good reference for choosing  
3203 the working point at the same time.

3204 Following are the variables which are used as input of the MVA (definitions are  
3205 listed in Table 3.1):

- 3206 •  $f_3, w_{stot}, T_{ratio}(DEmaxs1), R_{Had}, R_\eta, R_\phi, W_{\eta 2}, eOverP, \Delta\Phi.$

3207 Distributions of the variables are shown in Figure A.1.

3208 According to Figure A.2, a high correlation can be seen between  $DEmaxs1$  and  
3209  $w_{stot}$ , and this will be considered for forming groups of variables which are needed  
3210 to be optimized independently during the optimization process. Figure A.3 shows  
3211 clearly that the LikelihoodPCA can give a very good performance by using those  
3212 variables.

### 3213 **A.4 Cut-based optimisation methodology**

#### 3214 **Preparation**

3215

## Bibliography

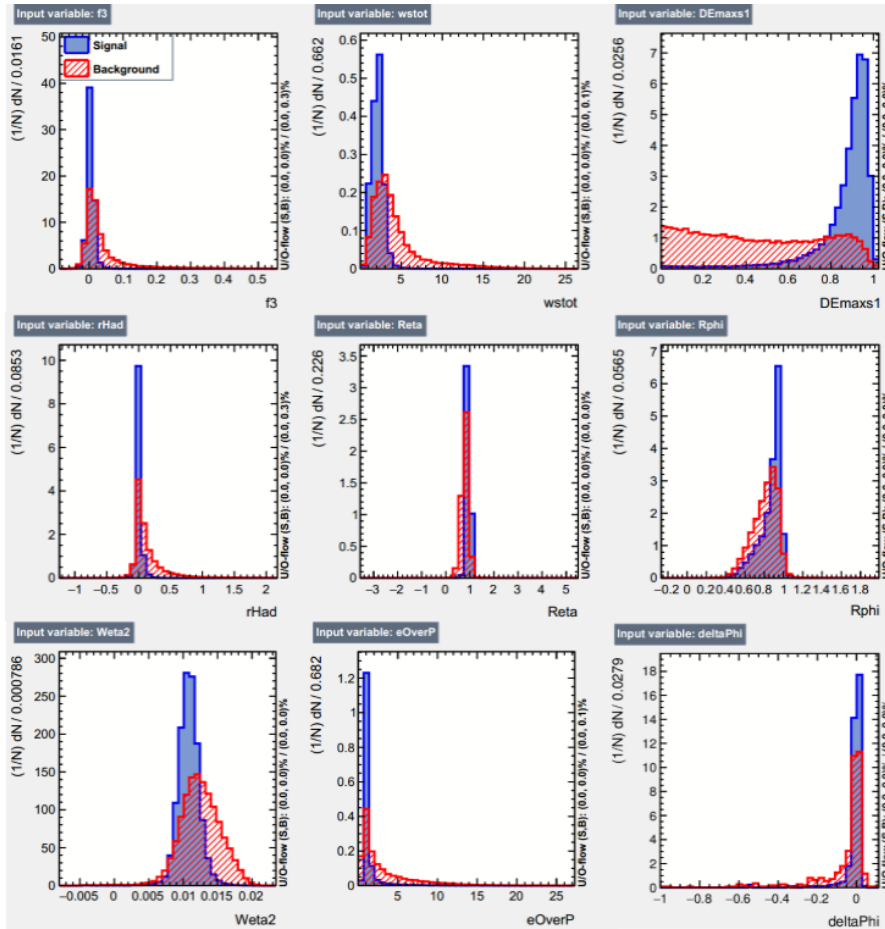
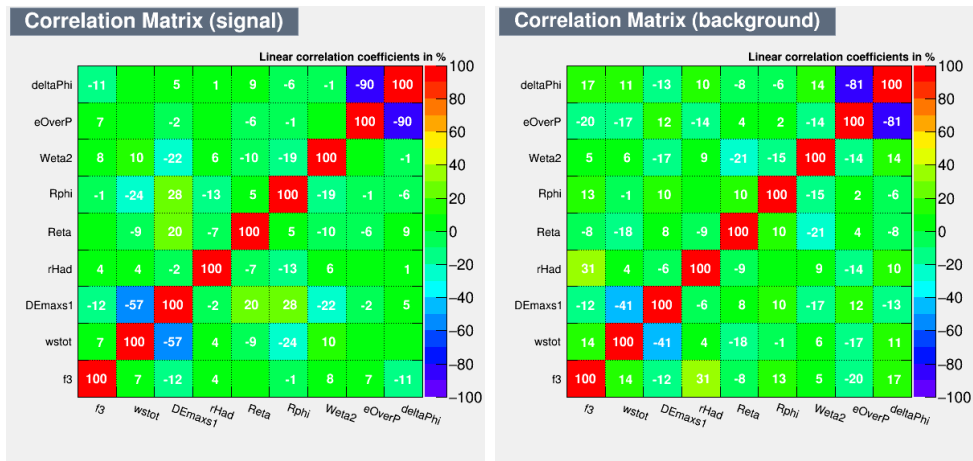


Figure A.1: Variables distribution for signal and background.



(a) Correlation for signal.

(b) Correlation background.

Figure A.2: Correlation Matrix of input variables for signal and background.



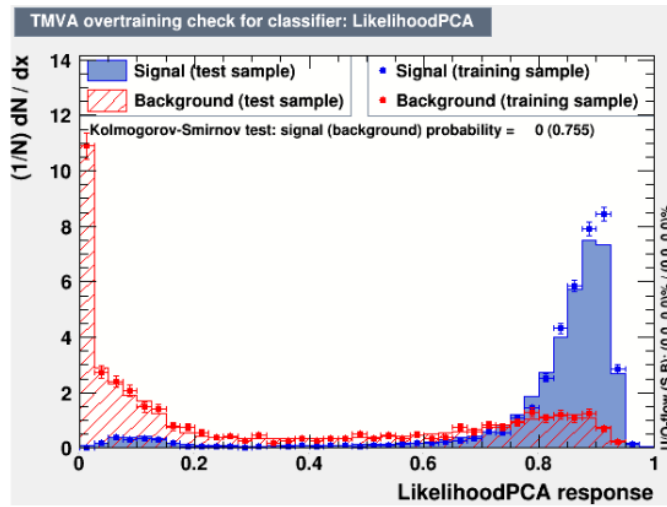


Figure A.3: LikelihoodPCA response.

3216 Nine variables mentioned above are considered in the cut-based menu due to  
 3217 their separation power. Especially for the menus used in the cut-based electron  
 3218 identification, the cuts for different discriminate variables are processed in  $p_T$  and  
 3219  $\eta$  bins. To simplify the optimization, the  $p_T$  and  $\eta$  binning used in this study is  
 3220 the same as the cut-based EID menus of 2012. So optimization is processed in ten  
 3221  $\eta$  bins: 0-0.1, 0.1-0.6, 0.6-0.8, 0.8-1.15, 1.15-1.37, 1.37-1.52, 1.52-1.81, 1.81-2.01,  
 3222 2.01-2.47 and three  $p_T$  bins: 5-10 GeV, 10-15 GeV, and 15-20 GeV.

3223 Results from the bin 20-30 GeV are used as a reference. One thing to mention,  
 3224 for the  $\eta$  bin 2.01-2.37 and 2.37-2.47 which used in the old menu, these two bins  
 3225 are merged into one due to the lack of statistics. Eventually, the same cut is used  
 3226 in both bins.

3227

### 3228 Pile-up dependence and pre-cleaning

3229

3230 When choosing the proper discriminate variables, one should consider not only  
 3231 the correlation between the variables (shown in Figure A.2), but also the pile-up  
 3232 dependence of each cut used in electron ID menu.

3233 Because of the complex pile-up ( $\mu$ ) condition in Run 2 compared to Run 1 it's  
 3234 necessary to avoid the pile-up dependence of input variables which leads to the  
 3235 bias in the final results as function of pile-up. The check of the pile-up dependence  
 3236 for those input variables is processed and this can be studied in two ways:

- 3237 • Applying only one cut, and measuring the slope of the efficiency vs  $\mu$  curve.
- 3238 • Applying all but one cut, measuring the efficiency, and taking the difference  
 3239 between the slopes of efficiency vs  $\mu$  when all cuts are applied and when all  
 3240 cuts but one are applied.

## Bibliography

Group1	$f_3, w_{stot}, T_{ratio}(DEmaxs1), R_{Had}$
Group2	$R_\eta, R_\phi, W_{\eta 2}$
Group3	$eOverP, \Delta\Phi$

Table A.1: The input variables of each group.

3241 The above studies are performed in three  $p_T$  bins (5-10 GeV, 10-15 GeV, 15-20  
 3242 GeV) of interest while also the bin (20-30 GeV) as reference. The results reflect  
 3243 the fact that different variables could be sensitive in difference  $p_T$  ranges. The  
 3244 variables deemed sensitive here are treated with extra care. Once the pile-up  
 3245 sensitive variables are studied, a pre-cleaning process is applied on some variables  
 3246 to avoid having cuts in the tails<sup>a</sup>. An attempt is made to remove a lot of the  
 3247 background events and remove as few of the signal events at the same time.

3248 For each variable preliminary cuts are applied such that efficiency of the back-  
 3249 ground decreases by 10% for each cut while keeping the decrease in signal efficiency  
 3250 less than 0.1%. After this is done, the global decrease in the signal efficiency is  
 3251 under 1% – 2%. After applying the tail cuts, 10% background events are rejected  
 3252 and only 1% signal are lost totally.

3253

### 3254 Variables in groups

3255

3256 Concerning the variable correlations which are show in Figure A.2, the input  
 3257 variables are sorted into groups shown in Table A.1.

3258 The variables used for the Loose, Medium, and Tight working points are the  
 3259 same as those used in the high  $p_T$  regime. The TRT PID variable, eProbabilityHT,  
 3260 is treated separately and will be mentioned later; this is justified by the fact that  
 3261 its linear correlation with other variables is negligible.

3262

### 3263 Efficiency target

3264

3265 The performance of the cut-based EID menus of 2012 is used as a benchmark  
 3266 here. For each  $p_T$  bin, the same signal efficiency as the 2012 menu is used for the  
 3267 three working points, Loose, Medium and Tight. In addition, a flatness in the  
 3268 efficiency as a function of  $\eta$  is forced by hand. Table A.2 shows the target signal  
 3269 efficiency ( $\epsilon_0$ ) used in the optimization process.

3270

### 3271 Optimization flow

3272

3273 Signal and background events are pre-selected firstly. Furthermore, all cuts from  
 3274 the 2012 menu are applied except the variables which need to be optimized (eg.,

---

<sup>a</sup>Cut values could jump a lot with negligible changes in efficiency if this is not taken care of as unexpected.

## Bibliography

Menus	5-10(GeV)	10-15(GeV)	15-20(GeV)
Loose	0.88	0.88	0.88
Medium	0.75	0.75	0.80
Tight	0.60	0.65	0.75

Table A.2: Target efficiency for each menu in different  $p_T$  bins.

3275 group 1 variables). A primary selection efficiency ( $\epsilon_1$ ) can be obtained after this  
 3276 step. Then those optimized variables will be the input variables for cut-based  
 3277 method. The CutBased method which is implemented in the TMVA toolkit is  
 3278 chosen for the optimization work. The general optimization work flow is shown in  
 3279 Figure A.4.

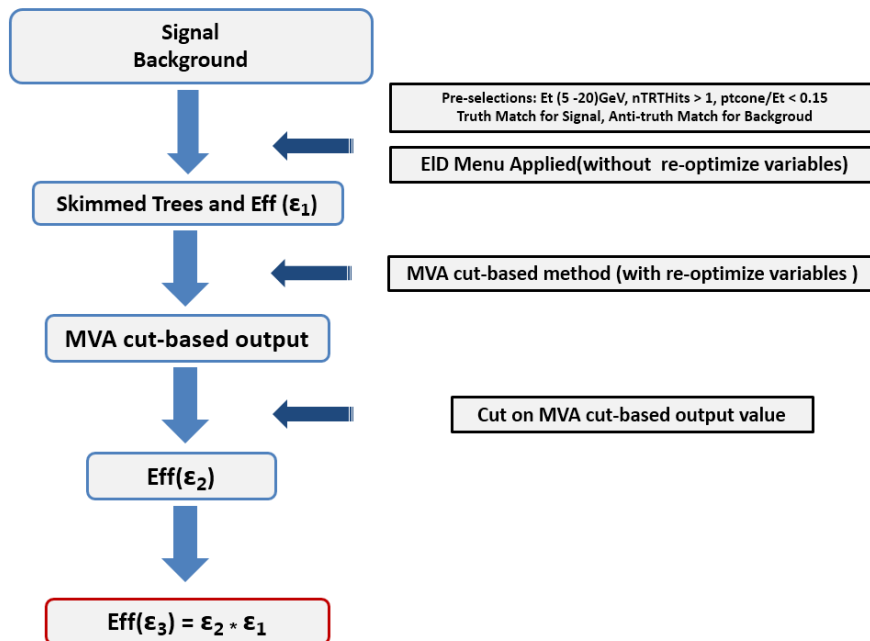


Figure A.4: Optimization work flow.

3280  
 3281 After MVA cut-based method training and testing, a new signal efficiency ( $\epsilon_2$ )  
 3282 can be obtained. The total signal efficiency ( $\epsilon_3$ ) can be calculated with equation  
 3283  $\epsilon_3 = \epsilon_1 * \epsilon_2$  which is the target efficiency in the table above. The values of all cuts  
 3284 corresponding to the signal efficiency ( $\epsilon_2$ ) can be found in the MVA output.

3285 The variables are grouped into three and the optimization chain can be ex-  
 3286 plained like Group-1  $\rightarrow$  Group-2  $\rightarrow$  Group-3. It means that during the Group-2  
 3287 optimization process, cut-based method will use the new first Group variables' cut  
 3288 value which are obtained from Group-1 process and so on. To give more opti-  
 3289 mization room for each group, a proper efficiency target tuning is needed for each

## Bibliography

group. Signal efficiencies are set differently for Groups to leave a balanced optimization room. And after whole optimizing process, targeted signal efficiency will be achieved as expected. An example of Tight menu optimization which contains all three group variables:

- 1). Efficiency with pre-cleaning:  $\epsilon_C$ .
- 2). G1 optimization: preparing samples with pre-cleaning and target efficiency:  $\epsilon_{G1} = \epsilon_0 + 0.66 * (\epsilon_C - \epsilon_0)$  ( $\epsilon_0$  is the target efficiency above).
- 3). G2 optimization: preparing samples with pre-cleaning+G1 and target efficiency:  $\epsilon_{G2} = \epsilon_0 + 0.33 * (\epsilon_C - \epsilon_{G1})$ .
- 4). G3 optimization: preparing samples with pre-cleaning+G1+G2 and target efficiency:  $\epsilon_{G3} = \epsilon_0$ .

The CutBased method in TMVA is chosen for *Loose* and *Medium* menu optimization when doing the training and testing. The CutsGA method in TMVA is chosen for *Tight* menu as variables' cut value with two sides can be obtained. Variables' cut value corresponding to the signal efficiency will be chosen as new cuts value of the menu. Such as G1 variables' cuts value are the values corresponding to  $\epsilon_{G1}$  in MVA cut-based method, etc.

### Smoothing and monotonicity process

After inspecting the cut values in different  $\eta$  bins in any given  $p_T$  bin, it is found that some cut values changed a lot from one  $\eta$ -bin to the next. This is not desired, so a post-optimization cuts smoothing procedure is applied to avoid these jumps without loss in efficiency or the flatness. The procedure is as follows:

- Move cuts in small steps. The step value is defined using the maximum and minimum values of the cut among all  $\eta$  bins in a given  $p_T$  bin (call this  $C_{max}$  and  $C_{min}$ ). The step size is then defined as  $\frac{C_{max}-C_{min}}{200}$ .
- This change in cut values is stopped as soon as the efficiency reduces by 1%.

But it is observed that this resulted in a global decrease of about 4%. To get around this, an extra condition is applied on the efficiency by forcing it to depend on the working point efficiency by requiring the following relation:

$$\epsilon_T = \epsilon_{T,MAX} \left(1 - C\epsilon_{WP}^2\right), \quad (\text{A.1})$$

where  $\epsilon_{T,MAX} = 0.01$  and  $C = 0.05$ . The results are as desired which are shown in Figure A.5.

Figures A.6, A.7, A.8 show the comparison of *Loose*, *Medium*, *Tight menu* with/without the cut smoothing. And also this process affects very little for the whole performance with many check tests.

Besides this cut smoothing process, inclusiveness of menus should be ensured which means *Loose* menu is looser than *Medium* and *Medium* is looser than *Tight*.

Bibliography

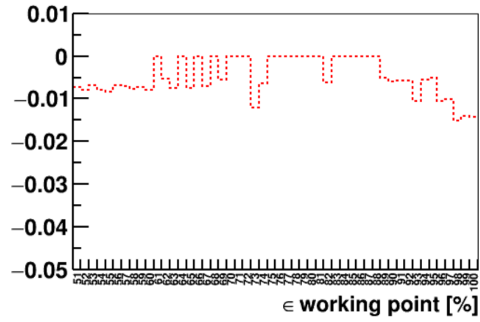


Figure A.5: Efficiency lost because of smoothing.

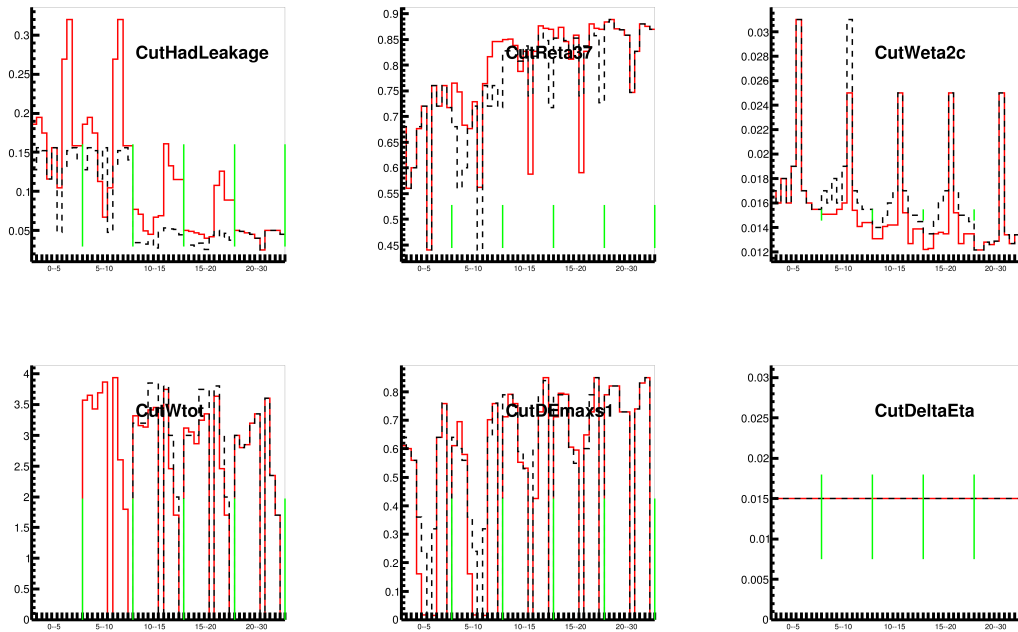


Figure A.6: Cuts value comparison of *Loose* menu with (Red)/without (Black) smoothing.

## Bibliography

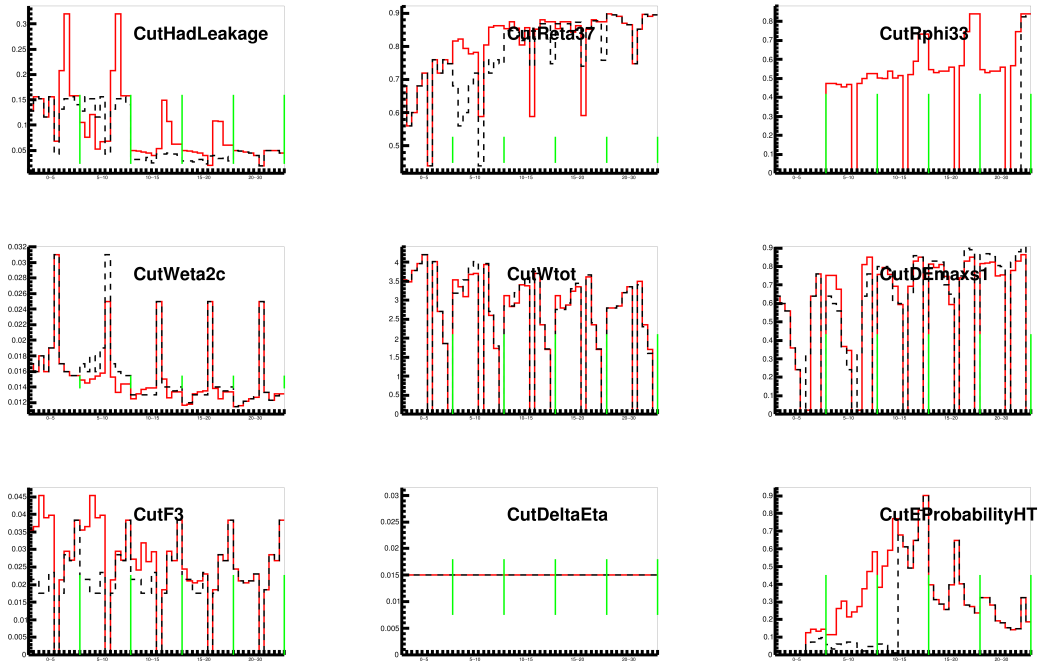


Figure A.7: Cuts value comparison of *Medium* menu with (Red)/without (Black) smoothing.

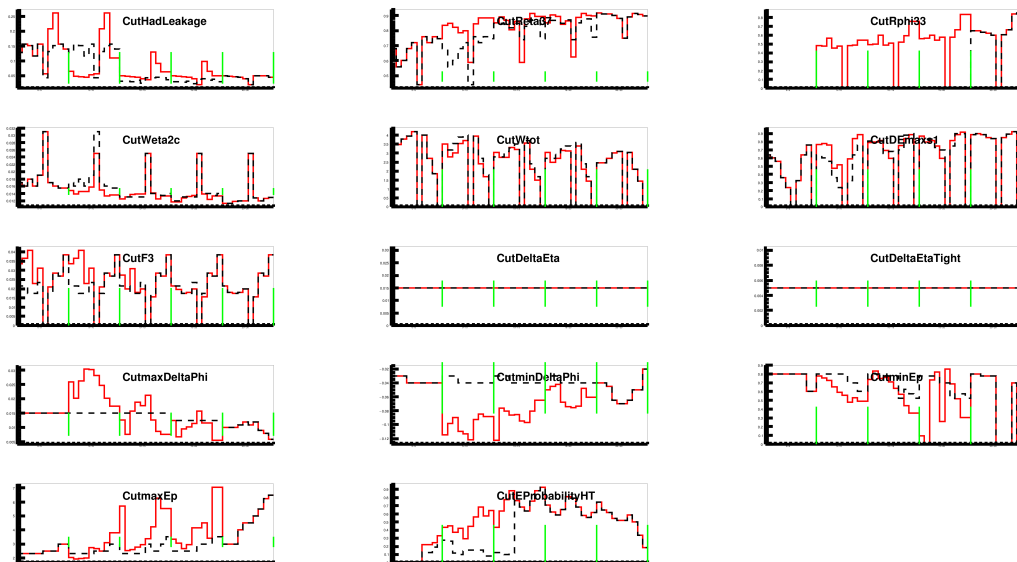


Figure A.8: Cuts value comparison of *Tight* menu with (Red)/without (Black) smoothing.

3329 This is necessary after doing smooth process as this will affect final performance a  
 3330 lot. Monotonicity of cut values as a function of  $p_T$  is also taken into consideration.

### 3331 **A.5 Coping with the changes in the TRT**

3332 During Run 1, the gas in the TRT used to be only Xenon, which is very expensive.  
 3333 This was problematic because there were major leaks in the TRT. A cheaper  
 3334 solution was to use a Xenon-Argon mixture instead. This had major effects on  
 3335 electron identification because of the TRT-related variables. Two new scenarios  
 3336 called the Baseline scenario and the Pessimistic scenario are simulated based on  
 3337 the amount of Argon being used in the Xe-Ar mixture. The electron ID includes  
 3338 cuts on the total number of TRT hits, and the fraction of high threshold Xenon  
 3339 hits (defined as  $f_{HT} = \frac{HT \text{ Xe hits}}{\text{Total of TRT hits}}$ ). The comparison between different scenarios  
 is shown in Figure A.9.

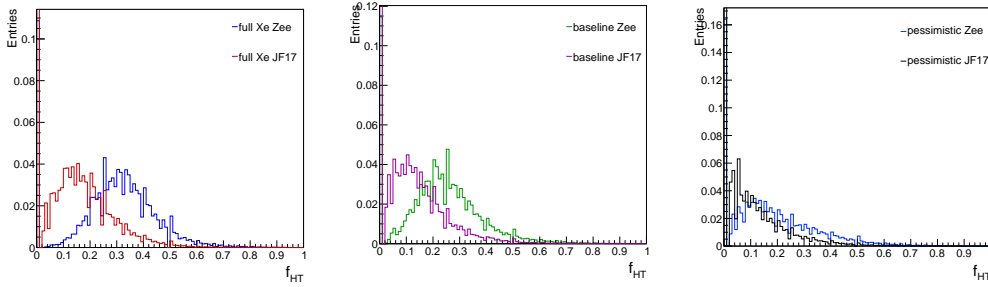


Figure A.9: The effect of changing the TRT gas on the fraction of the high threshold TRT hits (left: full Xe, middle: baseline, right: pessimistic).

3340  
 3341 It is clear that the discrimination power of the variable is significantly reduced  
 3342 in shifting from full Xe to the baseline Xe-Ar scenario or the pessimistic Xe-Ar  
 3343 scenario. To cope with this, a new variable, called eProbabilityHT, which also  
 3344 uses the high threshold hits information is introduced. The eProbabilityHT is a  
 3345 likelihood-type variable defined as,

$$p^{e,\pi} = \Pi p_{HT}^{e,\pi} \times \Pi (1 - p_{HT}^{e,\pi}), \quad (\text{A.2})$$

$$p_{final}^{e,\pi} = \frac{p^{e,\pi}}{p^e + p^\pi}. \quad (\text{A.3})$$

3346 The main purpose of this variable is to distinguish electrons from hadrons, espe-  
 3347 cially pions and hence, plays an important role in electron ID. The cut-based menu  
 3348 has to be re-tuned to replace  $f_{HT}$  by eProbabilityHT. To do this, it is important  
 3349 to understand the distribution for the signal and background. This is shown in  
 3350 Figure A.10 for the baseline and pessimistic scenarios. These are inclusive in  $p_T$   
 3351 and  $\eta$ . The true electrons tend to have values closer to 1, which is clear in the plots.

## Bibliography

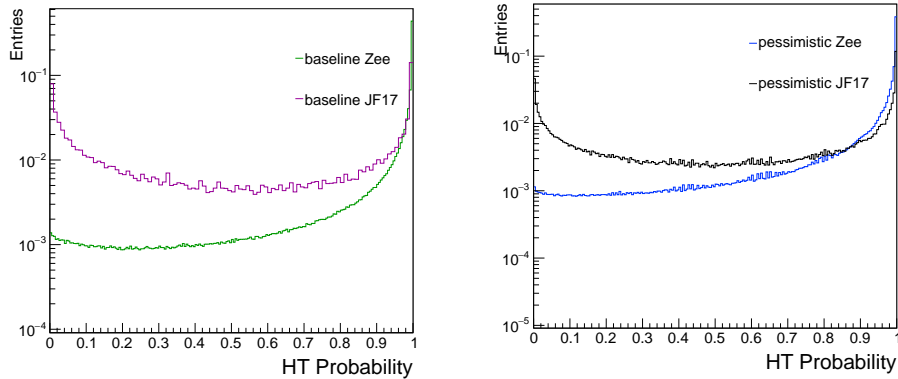


Figure A.10: Distribution of eProbabilityHT for the signal and background.

3352 After checking the distribution, it is necessary to replace  $f_{HT}$  by eProbabilityHT.  
 3353 This is done in two steps:

- 3354 • The effect of removing the  $f_{HT}$  cut on the efficiency is evaluated in all the  
 3355  $p_T$  and  $\eta$  bins in which electron ID is optimized.
- 3356 • The eProbabilityHT distributions in these bins are scanned and cuts are  
 3357 proposed to cause the same effect in the efficiency as  $f_{HT}$ .

3358 The fact that the procedure works to give the same efficiencies is shown in  
 Figure A.11.

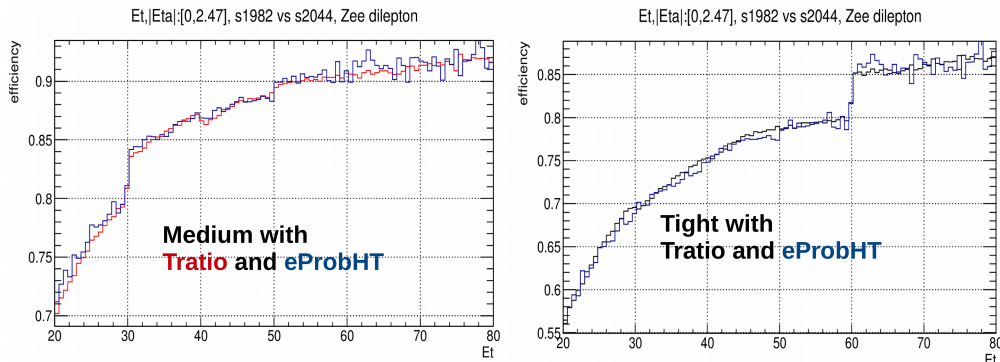


Figure A.11:  $T_{ratio}$  and eProbabilityHT comparison with different scenarios in signal events.

3359  
 3360 It should be noted that  $T_{ratio}$  is used with a sample in which the TRT only  
 3361 contained Xe whereas eProbabilityHT is used with the sample which contained  
 3362 the Xe-Ar mixture. This is distinguished by the s-tags; s1982 corresponds to the  
 3363 full Xe samples while s2044 corresponds to the baseline Xe-Ar mixture. A major  
 3364 feature of the TRT variables is that they are very pile-up sensitive and need to be



## Bibliography

3365 treated with special care. A positive aspect of the swap of variables is that the  
 3366 efficiency is more stable with respect to  $\mu$  and is shown in Figure ??.

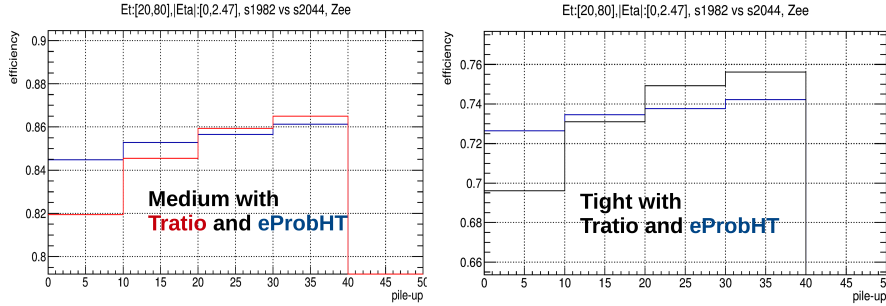


Figure A.12:  $T_{ratio}$  and eProbabilityHT comparison with different scenarios in signal events as function of pile-up.

3366  
 3367 A comparison of the performance of the menus is made between the baseline and  
 3368 pessimistic scenarios. The performance is similar, which is not expected given the  
 3369 difference in the eProbabilityHT distribution in the two scenarios. The following  
 3370 plots show the efficiencies as a function of the  $\mu$  for signal and background for the  
 3371 two scenarios. A major feature of the TRT variables is that they are very pile-up  
 3372 sensitive and need to be treated with special care. A positive aspect of the swap  
 3373 of variables is that the efficiency is more stable with respect to  $\mu$ . This is shown  
 in Figure A.13.

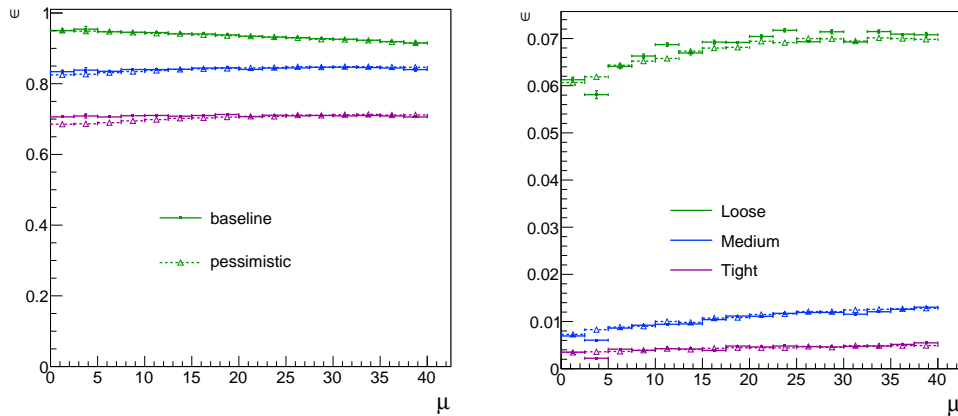


Figure A.13:  $|\eta| < 2.47$  and  $E_T > 5$  GeV: Efficiency vs  $\mu$  for signal (left) and background (right).

3374  
 3375 The efficiencies of the eProbabilityHT cut alone in the  $-2 < \eta < 2$  range for  
 3376 signal and background are shown in Figure A.14<sup>b</sup>.

<sup>b</sup>The loose menu does not have the eProbabilityHT cut and the plots corresponding to the loose menu represent the complete loose menu applied.

## Bibliography

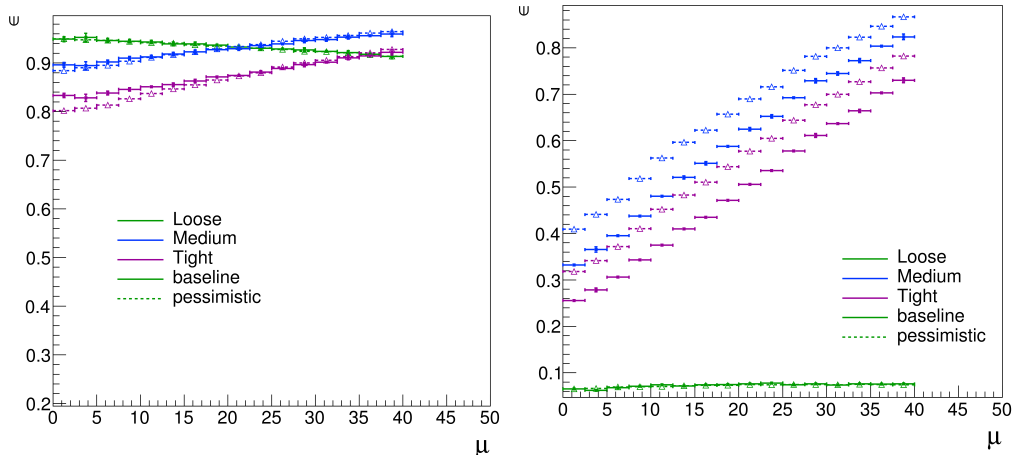


Figure A.14:  $|\eta| < 2$ : Efficiency vs  $\mu$  for signal (left) and background (right).

3377 The differences can be further attributed mostly to the  $-1 < \eta < 1$  region as shown in Figure A.15.

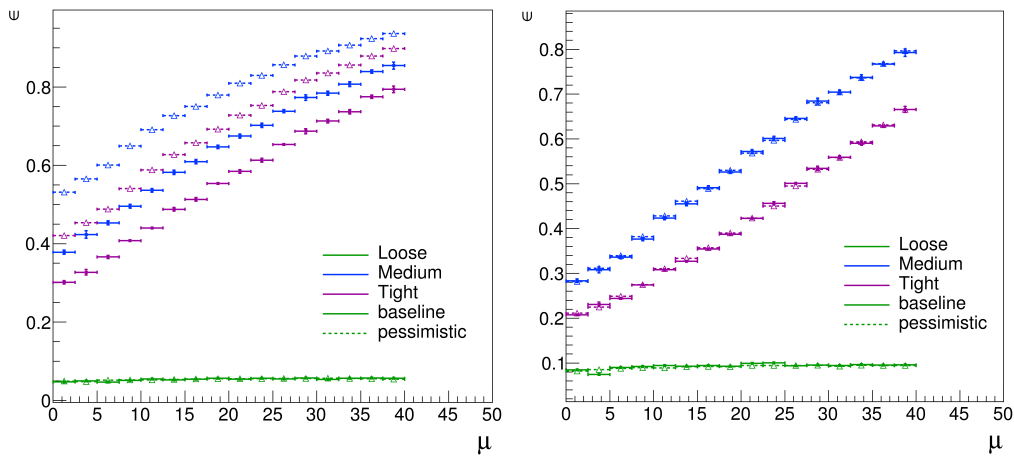


Figure A.15:  $|\eta| < 1$  (left) and  $|\eta| > 1$  (right): Efficiency vs  $\mu$  for background.

3378  
 3379 The bin-wise cuts for eProbabilityHT and the efficiency comparison between the  
 3380  $T_{\text{Ratio}}$  cut and the eProbabilityHT cut can be seen in Figure A.16.  
 3381 The monotonicity that is enforced while tuning with  $f_{HT}$  is destroyed during  
 3382 this swap of variables. The modelling of eProbabilityHT has slightly changed.  
 3383 This causes effects on the pile-up sensitivities.

# Bibliography

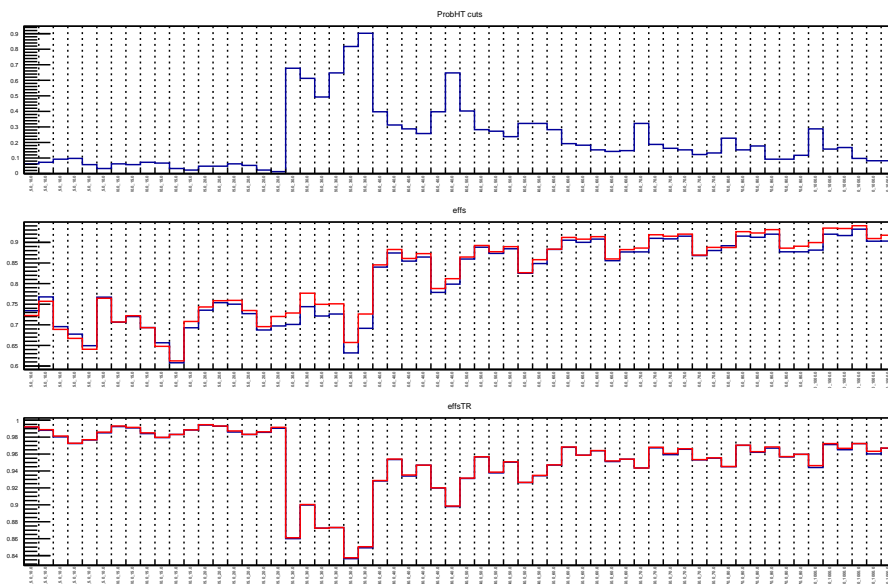


Figure A.16: Top: cut values for eProbabilityHT; middle: total efficiency comparison; bottom: comparison of efficiency of  $T_{\text{Ratio}}$  and eProbabilityHT cuts.

3384 **A.6 Results**

3385 Figure A.17 shows the efficiency comparison between cut-based EID menus of 2012  
 3386 (EID 2012) (benchmark) and cut-based EID menus of 2015 (EID 2015) as function  
 3387 of  $\eta$  for the good electrons and background electrons. The re-optimised menu EID  
 3388 2015 can have the good robustness with respect to the high pile-up situation.  
 3389 Figures A.18, A.19 and A.20 show efficiency comparison between EID 2012 and  
 3390 EID 2015 as function of pile-up for the background in full  $\eta$  region and  $0 < p_T < 20$   
 3391 GeV for *Loose*, *Medium* and *Tight* menu. Still EID 2015 menu can have higher  
 3392 background rejection and flat efficiency as function of pile-up compared to EID  
 3393 2012.

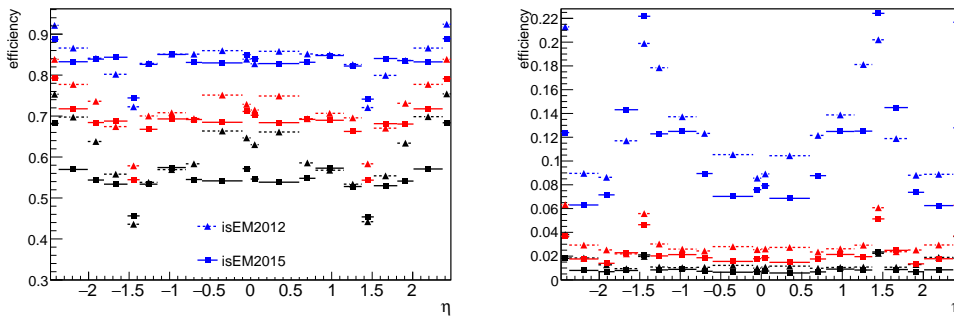


Figure A.17: Signal (left) and background (right): Efficiency vs  $\eta$  comparison between EID 2012 menu and new tuned EID 2015 menu.

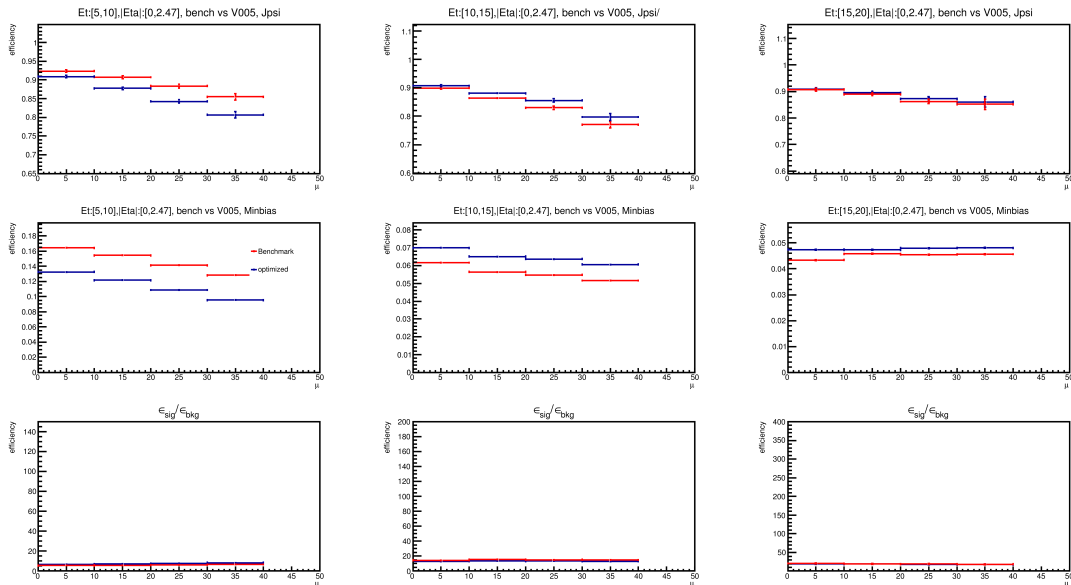


Figure A.18: Efficiency vs pile-up of *Loose* menu.

# Bibliography

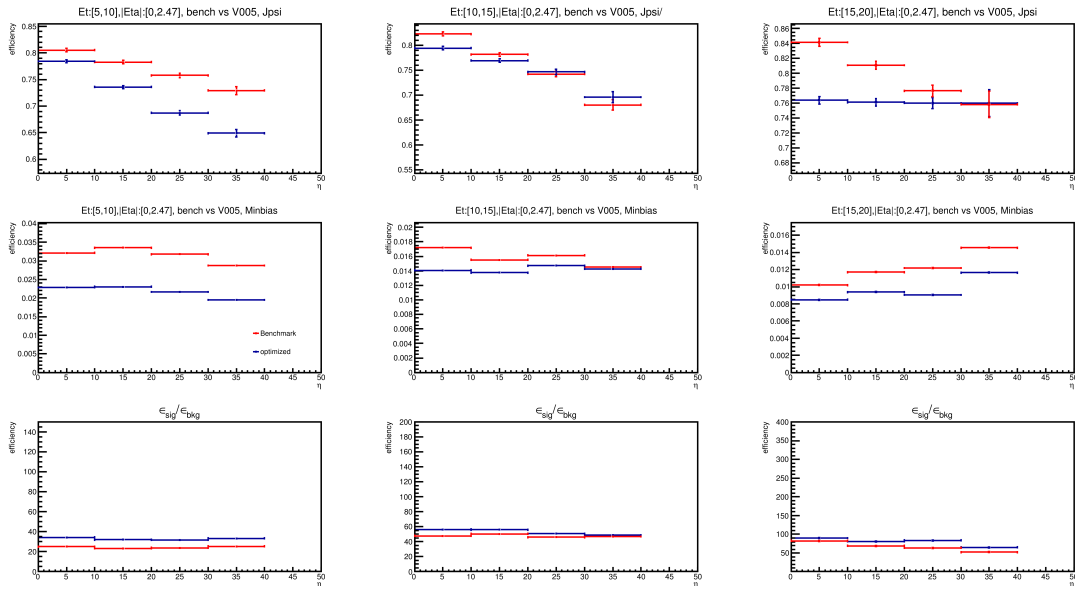


Figure A.19: Efficiency vs pile-up of *Medium* menu.

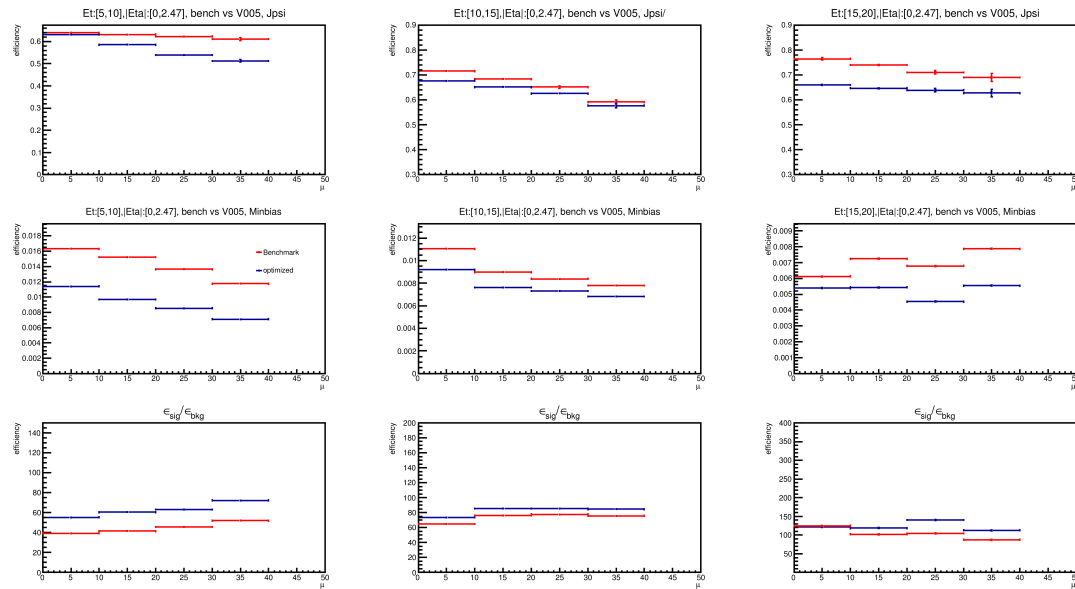


Figure A.20: Efficiency vs pile-up of *Tight* menu.

## Bibliography

3394 Figures A.21, A.22 and A.23 show the menu performance comparison between  
 3395 EID 2012 menu and new re-tuned menu for *Loose*, *Medium* and *Tight*. Overall,  
 3396 the re-tuned menu can have similar signal efficiency but with a better background  
 3397 rejection and keep a flatness of the efficiency as function of pile-up in the mean-  
 3398 while.

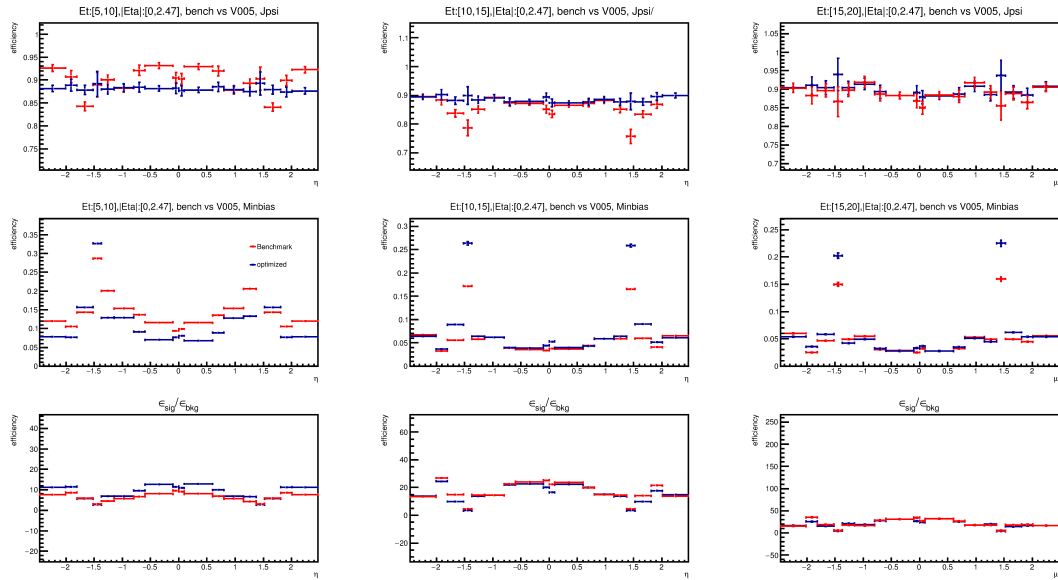


Figure A.21: *Loose* menu performance comparison between EID 2012 menu and new re-tuned menu.

3399 Additional check on the efficiency as function of  $E_T$  is shown in Figure A.24.  
 3400 No strange things and obvious bias can be seen.

### 3401 A.7 Conclusion

3402 The new re-tuned menus towards Run 2 condition keep the good performance  
 3403 as 2012 menus towards Run 1 and have many improvements at the same time,  
 3404 such as flat signal efficiency in full eta range and dealing with more complex pile-  
 3405 up condition well, higher background rejection compared to the EID 2012 menu.  
 3406 The re-tuned cut-based electron identification menu achieves the goals as expected  
 3407 overall.

## Bibliography

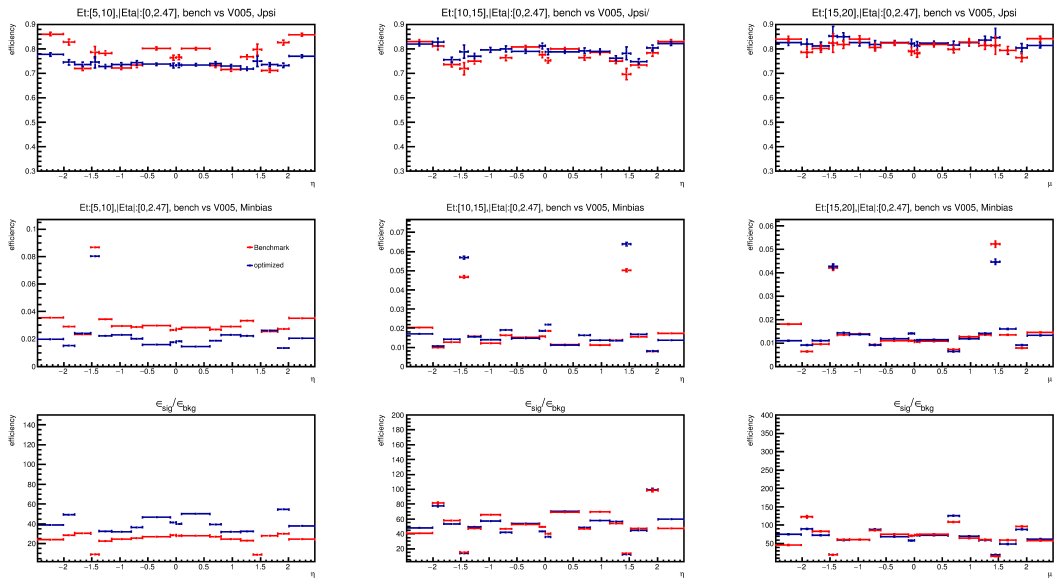


Figure A.22: *Medium* menu performance comparison between EID 2012 menu and new re-tuned menu.

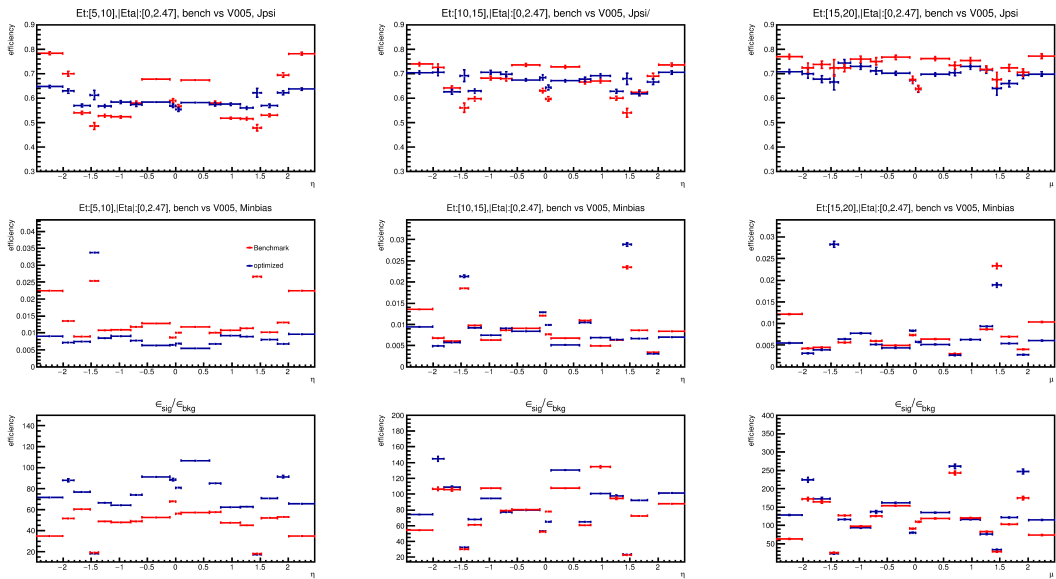


Figure A.23: *Tight* menu performance comparison between EID 2012 menu and new re-tuned menu.

Bibliography

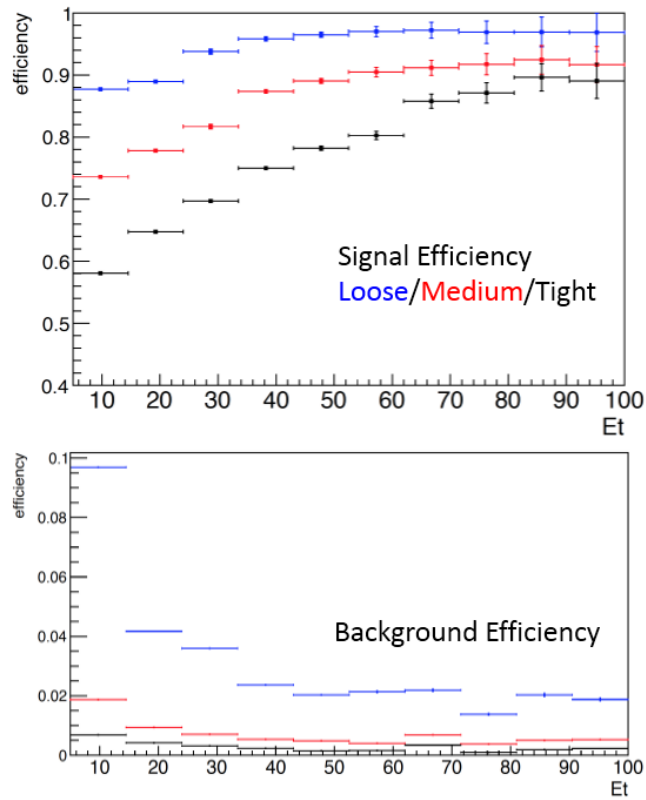


Figure A.24: Efficiency versus  $E_T$  for signal and background.



3408 **Abstract**

3409 **English version**

3410 The production of the Higgs boson associated with a pair of top quarks is one of the most im-  
 3411 portant Higgs boson productions still not observed by far. Therefore, its discovery is one of the  
 3412 most challenging searches after the Higgs discovery: not only it will be the first time we observe  
 3413 the existence of this Higgs production mode but moreover it will allow to measure the Yukawa  
 3414 coupling of top to the Higgs. This measurement can probe the basics of the Standard Model  
 3415 (SM) but also can search for any hints of new physics beyond the SM prediction. This makes this  
 3416 production search and measurement one of the most important analyses in post Higgs discovery  
 3417 era.  
 3418

3419 The ATLAS (A Toroidal LHC Apparatus) collaboration is operating since 2009 at the Large  
 3420 Hadron Collider (LHC), at CERN, in Geneva. During the first data taking campaign, so-called  
 3421 Run 1, and with corresponding to an integrated luminosities per experiment of  $5 \text{ fb}^{-1}$  at 7 TeV  
 3422 center of mass energy (in 2011) and  $20 \text{ fb}^{-1}$  at 8 TeV (in 2012), several Higgs boson decay  
 3423 channels were considered to measure the Higgs Yukawa coupling to the top quark within  $t\bar{t}H$   
 3424 production through multileptons final states. The ATLAS experiment produced individual com-  
 3425 binations which were then converted into a combined significances with the CMS collaboration  
 3426 combinations to get a  $t\bar{t}H$  observation significance, statistical of  $4.4\sigma$ , for a  $2.0\sigma$  expected from  
 3427 the Standard Model, corresponding to a measured excess of  $2.3\sigma$ .

3428 After a two years shutdown period, the LHC is now running at 13 TeV since 2015 and the cross  
 3429 section of  $t\bar{t}H$  increases by a factor of four compared to the 8 TeV run. The two years upgrade  
 3430 of the ATLAS detector brought substantial improvements to the physics analysis, especially for  
 3431 the SM measurement such as  $t\bar{t}H$  production and new physics searching.

3432 After having introduced the theoretical and experimental framework, the work presented in  
 3433 this thesis focuses on the search for the Higgs boson production associated with a pair of top  
 3434 quarks in multilepton final states and more particularly those with three leptons. Multi-variable  
 3435 analysis techniques are introduced and employed to have a good discriminate power between  
 3436 signal and background. For one of the main backgrounds, fake leptons, a Matrix Method is  
 3437 used to estimate its contribution. Generally, good modelling of the fake leptons background and  
 3438 better separation are achieved with a Boosted Decision Tree (BDT) distribution. Final results  
 3439 are obtained with a fit on a combined 2D BDT distribution.

3440 For a Higgs boson mass of 125 GeV with  $36.1 \text{ fb}^{-1}$  at 13 TeV data recorded by ATLAS, an  
 3441 excess of events over the expected background from other Standard Model processes is found  
 3442 with an observed significance of 2.18 standard deviations, compared to an expectation of 1.45  
 3443 standard deviations. The best fit for the  $t\bar{t}H$  production cross section is  $1.54^{+0.81}_{-0.74}$  times the  
 3444 SM expectation, and is consistent with the value of the Yukawa coupling to top quarks in the  
 3445 Standard Model. The results are compatible with the latest ATLAS approved results obtained  
 3446 with a further optimized background treatment.

3447 Comparing to the Run 1 results, a great improvement has been obtained on the significance  
 3448 and the uncertainty of the signal strength measurement. This results from a higher centre of  
 3449 mass energy hence a higher cross section, but also from the ATLAS performance upgrades as  
 3450 well as a more advanced analysis strategy. Still large uncertainties on the final measurements  
 3451 remains and further optimization of the analysis and higher statistics will be needed to get a  
 3452 more precise results in the future.

## Bibliography

### 3453 Français version

3454 La production du boson de Higgs associée à une paire de quarks top est une des plus importantes  
3455 productions non encore observée à ce jour. Sa découverte est donc une des recherches les plus  
3456 ambitieuses après la découverte du Higgs : non seulement cela sera la première fois que ce  
3457 mode sera observé mais surtout il permettra la mesure du couplage de Yukawa du Higgs avec  
3458 le quark top. Le résultat de cette mesure pourra sonder les bases du Modèle Standard (SM)  
3459 mais aussi rechercher des traces de nouvelle physique au-delà du Modèle Standard. Cela fait  
3460 de cette recherche et de la mesure de cette production une des analyses les plus importantes  
3461 depuis la découverte du Higgs. La collaboration ATLAS (A Toroidal LHC Apparatus) opère  
3462 depuis 2009 le détecteur généraliste ATLAS auprès du grand collisionneur de Hadron (LHC),  
3463 à Genève. Durant la première campagne de prise de données appelée Run 1 et correspondant  
3464 à une luminosité intégrée de  $5 \text{ fb}^{-1}$  à 7 TeV d'énergie dans le centre de masse (2011) et de 20  
3465  $\text{fb}^{-1}$  à 8 TeV (2012), plusieurs canaux de désintégration du boson de Higgs ont été considérés  
3466 pour mesurer le couplage de Yukawa du Higgs avec le quark top dans les productions  $t\bar{t}H$  avec  
3467 des états finaux multi-leptoniques. L'expérience ATLAS a ainsi produit une mesure combinée  
3468 qui a ensuite été associée avec la mesure combinée obtenue par l'expérience CMS pour produire  
3469 une signification globale observée de  $4.4\sigma$ , pour  $2.0\sigma$  prédite correspondant à un excès mesuré de  
3470  $2.3\sigma$ .

3471 Après une période d'arrêt de deux ans, le LHC tourne maintenant à 13 TeV depuis 2015 et  
3472 la section efficace  $t\bar{t}H$  a augmentée d'un facteur quatre par rapport au run à 8 TeV permettant  
3473 des mesures plus précises voire des découvertes. Les deux années de mise à niveau du détecteur  
3474 ATLAS ont apportées des améliorations substantielles pour les analyses de physique, et plus  
3475 particulièrement pour les mesures du SM comme la production  $t\bar{t}H$  et la recherche de nouvelle  
3476 physique.

3477 Après avoir introduit le cadre théorique et expérimental, le travail présenté dans cette thèse  
3478 se concentre ensuite sur la recherche et l'étude de la production du boson de Higgs associée à une  
3479 paire de quarks top dans des états finaux semi-leptoniques et plus particulièrement ceux à trois  
3480 leptons. Des techniques d'analyse multi-variables sont introduites et employées pour obtenir une  
3481 bonne discrimination entre signal et bruit de fond. Pour un des principaux bruits de fond, celui  
3482 de faux leptons, une méthode matricielle est présentée et utilisée pour estimer sa contribution.  
3483 Une bonne modélisation du bruit de fond de faux leptons et une meilleure séparation du signal  
3484 est obtenue avec une distribution d'arbre de décision stimulé (BDT). Les résultats finaux sont  
3485 obtenus par un ajustement sur les distributions BDT combinées à deux dimensions.

3486 Pour un boson de Higgs de 125 GeV de masse, un excès d'événements par rapport au bruit  
3487 de fond attendu des autres processus du Modèle Standard est obtenu avec une signification  
3488 observée de 2.18 déviations standard, à comparer à une estimation attendue de  $1.45\sigma$ . Le meilleur  
3489 ajustement pour la section efficace de production  $t\bar{t}H$  est de  $1.54^{+0.81}_{-0.74}$  fois l'estimation du SM  
3490 et est en accord avec la valeur du couplage de Yukawa du Higgs au quark top dans le Modèle  
3491 Standard. Ces résultats sont compatibles avec les derniers résultats officiels d'ATLAS obtenues  
3492 avec une optimisation plus poussée du traitement des bruits de fond.

3493 En comparant avec les résultats du Run 1, une amélioration importante a été obtenue sur  
3494 la signification et l'incertitude sur la mesure de la force du signal. Cela provient non seulement  
3495 de l'augmentation de l'énergie de collision et donc de la section efficace mais aussi de perfor-  
3496 mances améliorées du détecteur ainsi que d'une stratégie d'analyse optimisée. Néanmoins, les  
3497 incertitudes sur la mesure finale demeurent importantes et une optimisation plus poussée avec  
3498 une augmentation de la statistique sera nécessaire pour obtenir un résultat plus précis dans le  
3499 futur.