



**HAL**  
open science

# Search for Higgs boson production in association with a top quark pair in two same-sign lepton final states with the ATLAS detector at LHC

Asma Hadeif

► **To cite this version:**

Asma Hadeif. Search for Higgs boson production in association with a top quark pair in two same-sign lepton final states with the ATLAS detector at LHC. High Energy Physics - Experiment [hep-ex]. CPPM; Aix-Marseille Université, 2017. English. NNT: . tel-01636554

**HAL Id: tel-01636554**

**<https://theses.hal.science/tel-01636554v1>**

Submitted on 16 Nov 2017

**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.

AIX-MARSEILLE UNIVERSITÉ  
École Doctorale 352  
ECOLE DOCTORALE

Faculté des Sciences

Centre de Physique des Particules de Marseille

Thèse présentée pour obtenir le grade universitaire de docteur

ED 352 PHYSIQUE ET SCIENCES DE LA MATIERE  
Spécialité: Physique des Particules et Astroparticules

Asma HADEF

Recherche de la production du boson de Higgs en association avec une paire de quarks top dans les canaux avec deux leptons de même charge à partir du détecteur ATLAS au LHC

Search for Higgs boson production in association with a top quark pair in two same-sign lepton final states with the ATLAS detector at LHC

Soutenue le 20 Octobre 2017 devant le jury composé de:

Dr. Djamel	BOUMEDIENE	Rapporteur
Dr. Lydia	FAYARD	Rapporteur
Dr. Daniel	BLOCH	Examineur
Dr. Cristinel	DIACONU	Examineur
Dr. Pascal	PRALAVORIO	Directeur de thèse
Dr. Fabrice	HUBAUT	Co-Directeur de thèse

2017AIXM0230/009ED352



Cette oeuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).





## Extrait de la Thèse

# Recherche de la Production du Boson de Higgs en Association avec une Paire de Quarks Top dans les Canaux avec Deux Leptons de Même Charge a Partir du Detecteur ATLAS au LHC

Asma Hadeif

supervisée par

Pascal Pralavorio & Fabrice Hubaut

Centre de Physique des Particules de Marseille

La physique des particules, ou physique des hautes énergies, étudie la matière à l'échelle subatomique. La théorie actuelle de la physique des particules, appelée Modèle Standard (SM), fusionne la relativité restreinte avec la mécanique quantique pour obtenir une théorie relativiste des champs quantiques qui décrit les particules élémentaires et leurs interactions. Cette théorie est valide jusqu'à environ  $10^{-18}$  m, ce qui correspond à une énergie de l'ordre de 100 GeV, appelée échelle électrofaible. Cette valeur est (en principe) un endroit privilégié pour observer de la nouvelle physique.

Le but de cette thèse est de tester la validité du SM à cette frontière en énergie en mesurant le couplage entre les deux particules les plus massives du SM, le quark top et le boson de Higgs. La thèse porte sur l'analyse des collisions proton-proton recueillies auprès du grand collisionneur de hadrons (LHC) situé au CERN (Organisation européenne pour la recherche nucléaire) près de Genève, seule machine aujourd'hui capable d'atteindre cette frontière en énergie dans le monde. La première prise de données (Run1) du LHC a permis la découverte du boson de Higgs en juillet 2012 et a confirmé les prédictions

du Modèle Standard. Le LHC a redémarré au printemps 2015 avec une nouvelle énergie dans le centre de masse de 13 TeV (Run2), significativement plus élevée que celle du Run1 (8 TeV) afin de continuer à explorer de nouveaux domaines de la physique des particules jusqu'à fin 2018. Les données utilisées dans cette thèse correspondent à la première année d'exploitation de Run2 (juin 2015-juillet 2016).

ATLAS (A Toroidal LHC ApparatuS), l'une des deux expériences généralistes du LHC, est l'outil idéal pour tirer profit de cette augmentation d'énergie et poursuivre les recherches des théories au-delà du Modèle Standard comme la supersymétrie (SUSY) et les dimensions supplémentaires. Le détecteur ATLAS est situé à 100 mètres sous terre et mesure 44 mètres de long, 25 mètres de diamètre et pèse environ 7 000 tonnes. ATLAS a pour but d'identifier les particules secondaires produites lors des collisions proton-proton et pour mesurer leur position dans l'espace, leurs charge, leur vitesse, leur masse et leur énergie. Cela permet de reconstruire complètement les événements engendrés par les collisions proton-proton. Pour ce faire, le détecteur comporte plusieurs sous-détecteurs ayant une excellente herméticité, une granularité très fine et une électronique résistante aux radiations.

La première partie de cette thèse porte sur la mesure de l'efficacité de reconstruction des électrons. La reconstruction et l'identification des électrons représentent en effet une information essentielle pour comprendre précisément tous les processus électro-faibles comme la mesure des sections efficaces de production des bosons  $W$ ,  $Z$ ,  $H$  où l'incertitude associée à la reconstruction des électrons et à l'efficacité d'identification est une source importante de l'erreur systématique finale. De manière équivalente la mesure de l'efficacité des électrons joue un rôle important dans diverses recherches de nouvelle physique comme le  $Z'$ ,  $W'$  ou la supersymétrie.

Les efficacités de reconstruction des électrons, ainsi que les incertitudes associées, sont mesurées dans 200 intervalles de pseudo-rapacité et d'impulsion transverse ( $\eta \times E_T$ ) à partir des premières données du Run2 ( $3.2 \text{ fb}^{-1}$ ) en utilisant un échantillon pur d'électrons ( $Z \rightarrow ee$ ) par la méthode "tag & probe" (T&P). La partie délicate de la mesure se situe dans l'obtention des incertitudes systématiques liées à la soustraction du bruit de fond dans le lot  $Z \rightarrow ee$ . Ces mesures font partie de l'effort global pour estimer l'efficacité to-

tale des électrons, divisée en différentes composantes, à savoir l'efficacité de la reconstruction ( $\epsilon_{reco}$ ), de l'identification ( $\epsilon_{ID}$ ), de l'isolation ( $\epsilon_{iso}$ ) et du déclenchement ( $\epsilon_{trigger}$ ) – ces efficacités sont corrélées et séquentielles. Ces efficacités sont obtenues pour plusieurs points de fonctionnement pour l'identification, l'isolation et le déclenchement des électrons, étudiés pour répondre aux exigences de toutes les analyses physiques d'ATLAS. Pour l'obtention des résultats de physique, des échantillons correspondant aux sélections effectuées sur les données sont simulés pour reproduire autant que possible l'efficacité mesurée dans les données. Les rapports de ces deux efficacités (celles obtenues à partir des données et celles obtenues à partir de la simulation) sont utilisés comme facteur de correction multiplicatif pour les échantillons de simulation dans toutes les analyses ATLAS impliquant des électrons.

Les résultats sont obtenus dans des intervalles ( $\eta \times E_T$ ) choisis comme le meilleur compromis entre la statistique, la cinématique et les particularités de la géométrie du calorimètre et du détecteur interne. Dans l'ensemble, l'efficacité de la reconstruction varie de 97% à 99% avec des erreurs typiques mesurées inférieures à 0.1% pour  $E_T > 25$  GeV et entre 0.7 et 5% pour des impulsions transverses inférieures à 25 GeV (Figure 1). L'efficacité de reconstruction est comparable à l'efficacité obtenue à partir des données prises en 2012 à 8 TeV, à l'exception de la région de transition calorimétrique ( $1.37 < |\eta| < 1.52$ ) où l'efficacité diminue de 2%. La baisse de l'efficacité est liée aux inefficacités de reconstruction dans la transition entre le tonneau et le bouchon du détecteur interne et du calorimètre électromagnétique. Ces inefficacités ont été corrigées pour la suite du Run2, suite à cette observation.

En résumé, les facteurs de correction qui sont appliqués sur les lots de simulation sont très proches de l'unité, ce qui donne confiance dans la description de la réponse du détecteur aux électrons pour  $E_T > 15$  GeV. Cela montre l'excellente capacité du détecteur ATLAS de reconstruire les électrons dans l'environnement hadronique du LHC d'une part et la bonne compréhension des performances des électrons d'autre part.

La deuxième partie de la thèse est consacrée à la recherche de la production du boson de Higgs en association avec une paire de quarks top ( $t\bar{t}H$ ), seule façon d'obtenir une première mesure directe du couplage de Yukawa du top. Cette mesure est très sensible

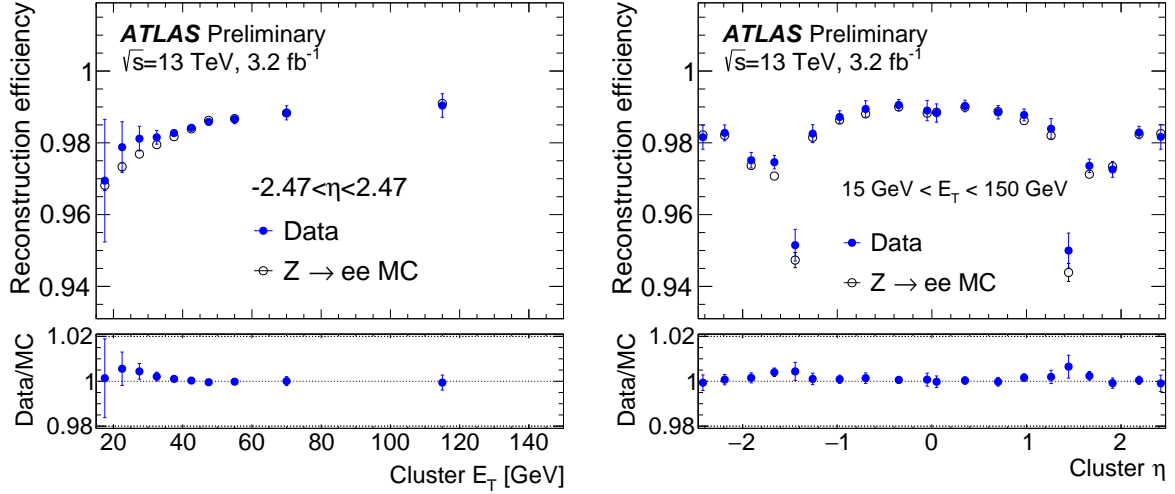


Figure 1: *Résumé des mesures de l'efficacité de la reconstruction des électrons dans les événements  $Z \rightarrow ee$  en terme d'énergie transversale  $E_T$  (à gauche) et en terme de pseudo-rapidité  $\eta$  (à droite) pour l'ensemble de données 2015 (points pleins bleus) et pour MC (points vides). Le ratio entre l'efficacité des données et celle de MC est affiché au-dessous de chaque graphe pour comparaison. Les barres d'erreur incluent les incertitudes statistiques et systématiques.*

à la présence de nouvelle physique. Les couplages du Higgs aux fermions étant proportionnels aux masses des fermions, le couplage Higgs-top est le plus grand avec une valeur proche de 1 dans le SM. Une description détaillée de l'analyse de  $t\bar{t}H$  basée sur des coupures cinématiques, optimisées pour l'analyse des  $13.2 \text{ fb}^{-1}$  de données à  $\sqrt{s} = 13 \text{ TeV}$ , est présentée dans ce document. Une attention particulière est accordée aux méthodes utilisées pour estimer le bruit de fond réductible, clé de la sensibilité de l'analyse à ce stade. Ce travail fait partie d'un énorme effort pour atteindre le but ultime: avoir l'analyse la plus sensible pour la statistique totale du Run 2 ( $130 \text{ fb}^{-1}$ ).

De nombreux canaux, correspondant aux états finals de désintégration du boson de Higgs ( $H \rightarrow bb, WW, \tau\tau, ZZ, \gamma\gamma$ ), sont étudiés. Dans cette thèse, le travail se concentre sur l'analyse  $H \rightarrow$  leptons, qui correspond aux canaux  $H \rightarrow WW, H \rightarrow ZZ$  et  $H \rightarrow \tau\tau$ . Pour optimiser le rapport "signal / bruit de fond", les événements sont classés en 4 catégories orthogonales en fonction du nombre de leptons légers et lourds: 2 leptons de même signe sans lepton  $\tau$  ( $2\ell ss$ ), 2 leptons de même signe et un lepton  $\tau$  en désintégration hadronique ( $2\ell ss + 1\tau_{\text{had}}$ ), 3 leptons ( $3\ell$ ) et 4 leptons ( $4\ell$ ). La signature  $2\ell ss$ , ciblant principalement les désintégrations  $H \rightarrow WW^*$ , est le canal d'étude de cette thèse. L'atout

principal de ce canal est sa faible contamination en bruit de fond venant du SM, ce qui en fait un des canaux  $t\bar{t}H$  les plus sensibles.

Les événements avec des faux (non-prompts) leptons représentent le principal bruit de fond du canal  $2\ell ss$ . L'estimation de ce bruit de fond, de manière la plus précise possible, est donc un enjeu central pour obtenir une bonne sensibilité pour le signal  $t\bar{t}H$ . Dans ce but, une méthode améliorée (méthode matricielle), basée sur les données, a été développée pour cette estimation. Elle est discutée en détail dans cette thèse. Alors que les leptons réels désignent les leptons isolés provenant de  $W^\pm$ ,  $Z$  ou  $\tau$ , les faux leptons sont définis comme étant des leptons non prompts produits dans des désintégrations de hadrons à saveur lourde ou un objet mal reconstruit en tant qu'un lepton issu de photons, de hadrons légers, des désintégrations en vol des kaons ou des pions aux muons, etc.

Des efficacités des vrais leptons ( $\epsilon_r$ ) et des faux leptons ( $\epsilon_f$ ) sont calculées dans des régions de contrôle orthogonales à la région de signal. Elles sont les principaux ingrédients utilisés par la méthode matricielle et sont donc cruciales pour obtenir une prédiction correcte du bruit de fond de faux leptons. Afin d'obtenir des prédictions plus fiables de faux leptons dans la région de signal,  $\epsilon_r$  et  $\epsilon_f$  sont calculés dans différents intervalles d'impulsion transverse des électrons comme le montre la Figure 2 pour les électrons (gauche) et les muons (droite) dans les données avant et après la soustraction de bruit fond. Dans la gamme d'intérêt ( $p_T > 25$  GeV), les efficacités des vrais électrons (muons) se situent entre 70 et 95% (90-100%) alors que celles des faux leptons sont d'environ 30 % pour les électrons et les muons. La taille de chaque intervalle est choisie en fonction de la statistique (statistique faible pour les grandes valeurs de  $p_T$ ) et de la dépendance supposée en  $p_T$ . Pour l'efficacité des vrais leptons sept intervalles sont choisis [10, 15, 20, 25, 30, 40, 60, 200] GeV. Pour l'efficacité des faux électrons, cinq intervalles sont utilisés [10, 15, 20, 25, 40, 200] GeV. Enfin, pour les faux muons, un seul intervalle est utilisé en raison du taux élevé de faux leptons pour  $p_T > 60$  GeV.

Après avoir validé les mesures de l'efficacité, à l'aide d'une comparaison avec les données, et la méthode, à l'aide d'un test de fermeture utilisant les lots simulés  $t\bar{t}$ , le bruit de fond des faux leptons est estimé dans la région de signal. Ces estimations sont un facteur 1.5 à 3.6 fois plus grandes que celles obtenues avec des lots d'événements simulés, ce qui

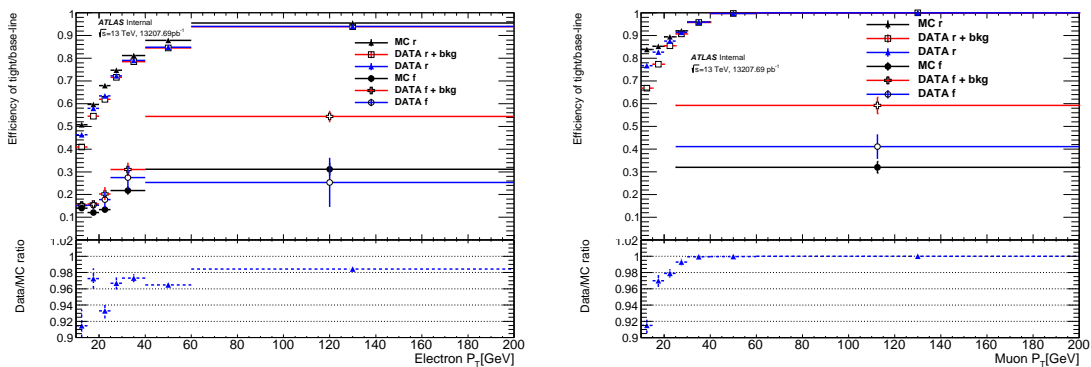


Figure 2: La répartition  $p_T$  de l'efficacité réelle et de taux de faux leptons pour les électrons (à gauche) et pour les muons (à droite). Les mesures sont effectuées pour les données et  $t\bar{t}$  MC (ligne noire). Dans les données, les efficacités sont représentées avant et après soustraction du bruit de fond par les lignes rouge et bleue respectivement. Le ratio entre les efficacités réelles des données et celles de MC est indiqué au-dessous de chaque graphe.

confirme l'impossibilité de s'appuyer sur les simulations pour obtenir une estimation des faux leptons. Le bruit de fond venant des faux leptons représente entre 32 et 48% du bruit de fond total (Figure 3) et domine l'erreur totale sur l'estimation du bruit de fond dans le canal  $2\ell ss$ . Il est suivi du bruit de fond de reconstruction incorrecte de la charge (QMisID reco), estimé à l'aide d'une méthode basée sur la probabilité orientée par les données. Pour les muons, cela se révèle négligeable. Pour les électrons, cela est principalement dû aux processus de trident  $e^\pm \rightarrow \gamma^* e^\pm \rightarrow e^\mp e^\pm e^\pm$  où un électron de haut- $p_T$  avec une charge opposée à l'électron prompt original est reconstruit (Bremsstrahlung dur suivi d'un processus de conversion de photon).

Ce premier résultat a permis d'introduire la méthode de la matrice comme une méthode alternative pour l'estimation du bruit de fond des faux leptons pour cette première analyse avec  $13.2 \text{ fb}^{-1}$  de données à  $\sqrt{s} = 13 \text{ TeV}$ . L'analyse en cours, combinant les données 2015 et 2016 ( $36 \text{ fb}^{-1}$ ), utilise maintenant cette méthode par défaut en intégrant les améliorations suivantes: une mesure des taux de faux leptons à des grands  $p_T$  et une paramétrisation multidimensionnelle (nombre de jets de  $b$ , distance entre lepton et jet le plus proche, ...).

La mesure de la force du signal  $t\bar{t}H$  comparée aux prédictions du SM ( $\mu_{t\bar{t}H}$ ) dans le canal  $2\ell ss$  est de  $4.0^{+1.2}_{-1.1}$  (stat)  $^{+1.7}_{-1.3}$  (syst). La contribution des incertitudes statistiques et systématiques est équivalente. Un léger excès est observé principalement venant du canal  $e^\pm \mu^\pm$ . Pour augmenter la sensibilité au signal, une combinaison de tous les canaux

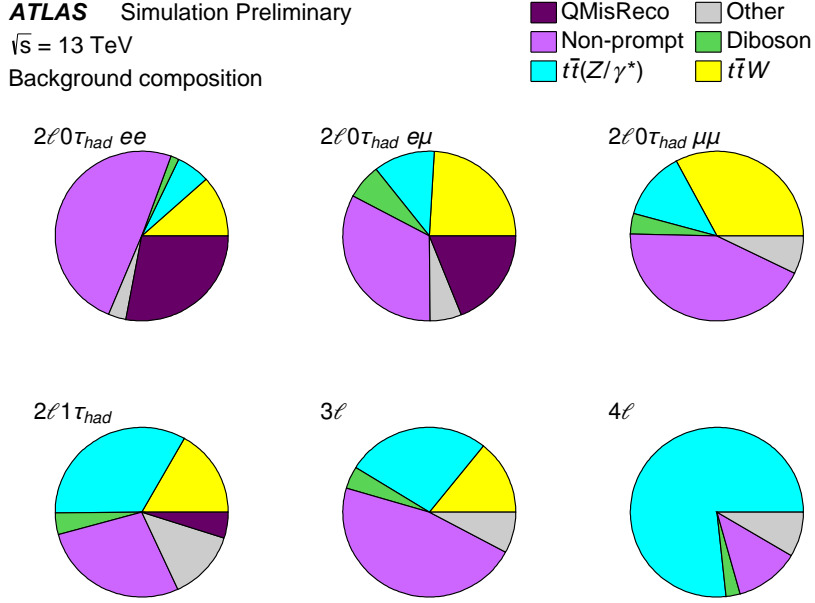
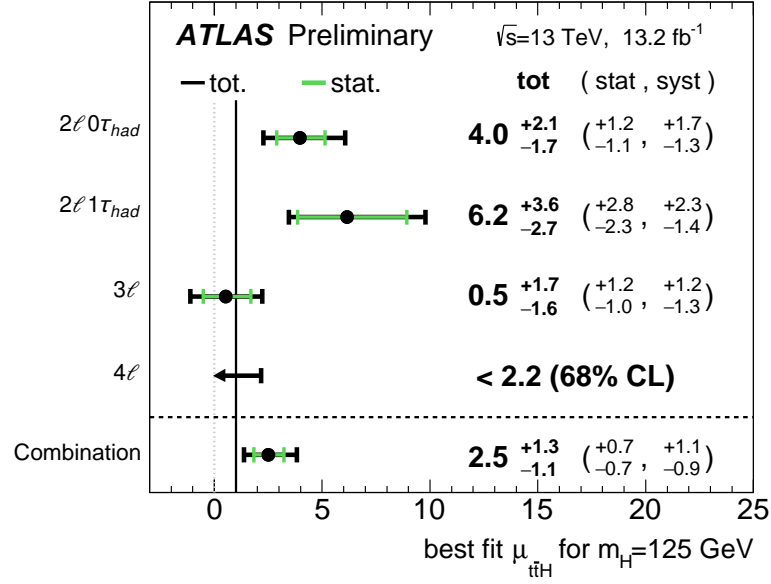


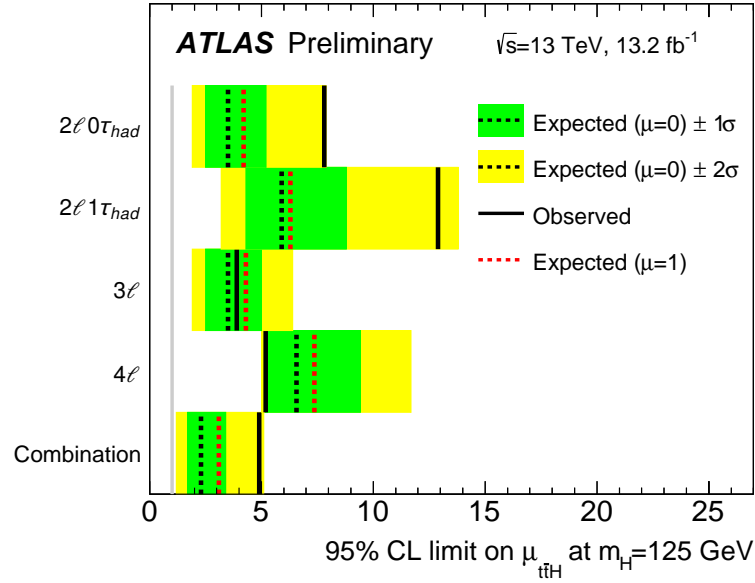
Figure 3: La contribution attendue du bruit de fond dans chaque canal à partir de diverses sources, en utilisant les valeurs des estimations de bruit de fond avant l’ajustement. Les bruits de fond de la mauvaise reconstruction de charge sont désignés par ”QMisReco”.

$t\bar{t}H \rightarrow \text{leptons}$  est effectué (Figure 4) et donne comme résultat  $2.5 \pm 0.7$  (stat)  $^{+1.1}_{-0.9}$  (syst). La limite supérieure correspondante pour l’exclusion du processus  $t\bar{t}H$  est de 4.9 (2.3 attendu) fois la valeur du SM à 95% de niveau de confiance. La valeur  $p$  associée à  $\mu_{t\bar{t}H} = 1$ , qui correspond aux prédictions du SM, est de 0.09 ( $1.3\sigma$ ).

La mesure de la force du signal  $t\bar{t}H$  en considérant tous les états finals (multilepton,  $bb$  et diphoton) est de  $1.8 \pm 0.7$ , ce qui correspond à une signification de  $2.8\sigma - 1.8\sigma$  serait attendu en présence du Modèle Standard  $t\bar{t}H$ . La sensibilité de cette combinaison dépasse la sensibilité obtenue avec les données du Run1. Aucune tension n’est observée entre les analyses  $t\bar{t}H$ . En utilisant 10% de la statistique totale attendue pendant le Run2, le processus de production total de  $t\bar{t}H$  n’a pas encore été observé dans ATLAS. Aucun écart significatif par rapport aux prédictions du SM n’est observé pour ce processus.



(a)



(b)

Figure 4: a) Valeurs de la force du signal  $\mu_{t\bar{t}H}$  par état final et après combinaison. La prédiction du Modèle Standard est  $\mu_{t\bar{t}H} = 1$ . Pour le canal  $4\ell$ , comme aucun événement n'est observé, une limite supérieure à 68% de niveau de confiance est calculé. b) Limites supérieures sur la force du signal  $t\bar{t}H$  à 95% de niveau de confiance. La prédiction du Modèle Standard est  $\mu_{t\bar{t}H} = 1$ . La valeur médiane de la limite supérieure est donnée en présence d'un signal  $t\bar{t}H$  ( $\mu_{t\bar{t}H} = 1$ ).





# Contents

<b>Introduction</b>	<b>14</b>
<b>1 Standard Model</b>	<b>16</b>
1.1 History	16
1.2 Matter particles	18
1.3 Particles interactions	20
1.3.1 Quantum Electrodynamics	20
1.3.2 Weak interactions	21
1.3.3 Quantum Chromodynamics	23
1.3.4 Electroweak interaction	24
1.4 Problems/Outlook	24
1.5 Summary	25
<b>2 Higgs Physics</b>	<b>26</b>
2.1 Brout-Englert-Higgs mechanism	26
2.2 Higgs production	28
2.2.1 Gluon-gluon fusion ( $ggF$ )	29
2.2.2 Vector boson fusion (VBF)	30
2.2.3 Higgs-strahlung	30
2.2.4 $t\bar{t}H$ production	31
2.3 Higgs decays	31
2.3.1 Fermionic tree-level decay modes	31
2.3.2 Bosonic tree-level decay modes	32
2.3.3 Loop-induced decay modes	33
2.4 Measurement of Higgs properties	33
2.4.1 Higgs mass	33
2.4.2 Higgs spin-parity	34
2.4.3 Higgs width and lifetime	35
2.4.4 Higgs couplings	35
2.5 $t\bar{t}H$ searches at LHC	37
<b>3 ATLAS at LHC</b>	<b>40</b>
3.1 The CERN accelerator complex	40
3.2 The ATLAS detector	42
3.2.1 ATLAS layout	43
3.2.1.1 The inner detector	43
3.2.1.2 Solenoid	45
3.2.1.3 Calorimetry	45
3.2.1.4 Muon spectrometer	50
3.2.1.5 Trigger system [1]	51

3.2.2	Data quality monitoring in ATLAS	51
3.2.3	Luminosity measurements in ATLAS	52
<b>4</b>	<b>Object reconstruction</b>	<b>54</b>
4.1	Tracks and primary vertices (PV)	54
4.2	Electrons	55
4.2.1	Electron trigger	55
4.2.2	Electron reconstruction	56
4.2.3	Electron identification	57
4.2.4	Electron isolation	59
4.3	Photons	60
4.3.1	Photon reconstruction	60
4.3.2	Photon identification	61
4.4	Muons	61
4.4.1	Muon reconstruction	61
4.4.2	Muon identification	62
4.5	Taus	62
4.6	Jets	62
4.6.1	Jet reconstruction	63
4.6.2	<i>b</i> -tagged jets	64
4.7	Conclusion	65
<b>5</b>	<b>Measurement of the electron reconstruction efficiency</b>	<b>66</b>
5.1	Definition	67
5.2	Method	68
5.2.1	Tag-and-Probe method with $Z \rightarrow ee$ events	68
5.2.2	Background estimation	70
5.2.2.1	Background for clusters with an associated track	70
5.2.2.2	Background for clusters with no associated track	71
5.2.3	Uncertainties	72
5.3	Dataset and simulation	74
5.4	Results	74
5.4.1	Background level at denominator and numerator	74
5.4.2	Efficiency	76
5.4.3	Scale Factors	78
5.4.4	Combination	81
5.5	Conclusion	83
<b>6</b>	<b><math>t\bar{t}H</math> Multilepton analysis with <math>13.2 \text{ fb}^{-1}</math> at <math>\sqrt{s} = 13 \text{ TeV}</math></b>	<b>85</b>
6.1	Dataset and simulation	86
6.1.1	Data sample	86
6.1.2	Signal MC sample	87
6.1.3	Background MC samples	88
6.2	Signal region	92
6.2.1	Object definition	92
6.2.1.1	Electron	92
6.2.1.2	Muon	93
6.2.1.3	Tau	94
6.2.1.4	Jets and <i>b</i> -tagged jets	94
6.2.1.5	Overlap removal	94

6.2.2	Signal region definition	94
6.3	Data-driven background estimation	95
6.3.1	Charge flip estimation	96
6.3.2	Fake lepton estimation with the Matrix Method	98
6.3.2.1	Description of Matrix Method	98
6.3.2.2	Inputs to Matrix Method	100
6.3.2.3	Validation	108
6.3.2.4	Results	111
6.3.3	Comparison with the Fake Factor method	113
6.4	Validation regions for prompt leptons background	116
6.5	Fit results in $2\ell ss$ channel	118
6.5.1	Likelihood function	118
6.5.2	Systematics	118
6.5.3	Signal strength	119
6.6	First $t\bar{t}H$ combination with 13 TeV data	121
6.6.1	$t\bar{t}H$ Multilepton	121
6.6.2	All $t\bar{t}H$ decays	126
6.7	Conclusions and prospects	128
<b>Conclusion</b>		<b>129</b>
<b>A Background estimation in electron reconstruction efficiency</b>		<b>131</b>
A.1	Description of background estimation for clusters with no associated track	131
A.2	Invariant mass distribution of the background	132
A.3	Tables of electron reconstruction efficiencies and SF	132
<b>B Background estimation for <math>t\bar{t}H</math>, Multilepton analysis</b>		<b>140</b>
B.1	The Matrix Method	140
B.1.1	Mathematical description	140
B.1.2	Di-lepton category	142
B.2	Selection of validation regions	143
B.3	Fake lepton purity	143
B.4	Charge flip	143
<b>Bibliography</b>		<b>147</b>
<b>Acknowledgement</b>		<b>159</b>
<b>Abstract</b>		<b>161</b>
<b>Résumé</b>		<b>162</b>



# Introduction

Searches in particle physics at the high energy frontier could be one step into solving the most fundamental riddles of the universe. It looks at matter in its smallest dimension at the subatomic level in order to study the nature of the "elementary" particles and answer the fundamental questions of the nature of mass, energy, and matter and their interaction. The current dominant theory explaining these fundamental particles and their interactions is called the Standard Model (SM). This thesis presents a test of the validity of the SM at the energy frontiers; by measuring the coupling between the two most massive particles of the SM, the top quark and the Higgs boson. This measurement is performed by analysing data of the proton-proton collisions collected at the Large Hadron Collider (LHC) situated at CERN (the European Organization for Nuclear Research) near Geneva. The LHC first run produced the breakthrough discovery of the Higgs boson in July 2012 (Run1) and confirmed some predictions of the SM. After shutting down for planned maintenance in early 2013, it restarted in spring 2015 at a new center-of-mass energy of 13 TeV (Run2), significantly higher compared to the 8 TeV of the Run 1, to probe new realms of particle physics over the next three years. ATLAS (A Toroidal LHC ApparatuS), one of the two general purpose experiments of the LHC, is the perfect tool to profit from this increase of energy. It has many layers with excellent hermiticity, fine granularity and highly resisted electronics to the radiation. At Run2, ATLAS can get greater precision on the measurements of the properties of the Higgs boson, especially the couplings of the Higgs boson to matter particles, to refine our understanding of the Higgs sector of the SM. The data used in this thesis are corresponding to the first year of Run2 operation (June 2015-July 2016) collected by the ATLAS experiment. It searches of one of the, relatively, rare Higgs production in association with a top quark pair and treats one of the most important properties of the Higgs boson (Higgs top Yukawa coupling).

The thesis is organised as follows. In Chapter 1, the theory of the SM is introduced, explaining how the basic building blocks of matter interact with each other. One of the interesting topics of the SM is the Higgs physics described in Chapter 2, which answer the question of the origin of the mass and link both theoretical prediction with observation. This chapter describes the motivation behind the measurements performed in my analysis. Those measurements are taken by ATLAS detector, described in Chapter 3 with its sub-detector layout in order to reconstruct the detectable objects, including electrons, photons, jets and muons. The reconstruction of each object is described in Chapter 4. In particular, the ability of the ATLAS detector to reconstruct electrons is assessed by measuring its efficiency and detailed in Chapter 5; where I spent the first year of my PHD (2014-2015) on preparing the measurements for Run2 and providing the corresponding scale factors to all ATLAS physics analyses involving electrons with 2015 data. The challenge in these measurements is to estimate the high amount of the background in fine binning (200 bins in  $(E_T, \eta)$ ) which demands the design of high precision methods. The second part of my thesis (2015-2016) is dedicated to a search for the Higgs boson production in association with a top quark pair ( $t\bar{t}H$ ), described in Chapter 6. The signature

with two same-charge light leptons (electron or muon) without a hadronically decaying tau lepton final state ( $2\ell ss$ ), targeting the decays  $H \rightarrow WW^*$ , is examined, which is almost Standard Model background free, using the first 10% of the total expected Run2 dataset. In particular, I estimate the number of events with fake (non-prompt) leptons which represent the main reducible background of this signature. The estimation of this background largely drives the signal sensitivity. I worked on introducing and improving an alternative method (matrix method), more flexible and accurate than the one used since Run-1, to be a powerful validation tool and to be the baseline method for Run-2 analysis.

# Chapter 1

## Standard Model

---

*"The study of what was not understood by scientists, or was understood wrongly, seems to me often the most interesting part of the history of science."*  
- Steven Weinberg, *The Making of the Standard Model*

---

Today, particle physics, or High Energy Physics (HEP), looks at matter in its smallest dimension (order of  $10^{-18}$  m) at the subatomic level whereas astrophysics looks at matter in its largest dimensions ( $> 10^{12}$  m) at the scale of planets stars and galaxies (Figure 1.1). Particle physics studies the nature of the "elementary" particles<sup>1</sup> that constitute matter or radiation. The current dominant theory explaining these fundamental particles and their interactions is called Standard Model (SM). The SM is, in fact, a theory that merges special relativity with quantum mechanics getting what we call today the relativistic quantum field theory where all particles could be described and generated by quantum fields. It is the theory that allows for creation and destruction of particles seen at high energy interactions.

In this chapter, the principal concepts and building blocks of elementary particle physics and how they are related are presented without going into details of mathematically rigorous derivations. First, a short soft history of the HEP and SM is introduced in Section 1.1. The current fundamental "matter" particles in nature and their partners anti-matter particles, called "Fermions", are described in Section 1.2. The interaction between those particles are hold by force carriers called "Bosons" explained in Section 1.3. Every theory has its limitation including SM which is described in Section 1.4, with the outlook solutions beyond the Standard Model. A conclusion is drawn in Section 1.5.

### 1.1 History

World seems incredibly complicated and composed by thousands of different materials. Throughout history, man was looking to describe and understand nature by the simplest way. People tried to collect, categorise and analyse different phenomena to find some underlined pattern to help simplify this seeming complicated world. They tried and are still trying to answer what is the most "fundamental" elements that build nature and what are the rules that govern them. What is the smallest body in nature that cannot be divided any more and considered the constituent of all what we see in the Universe.

---

<sup>1</sup>An elementary particle is a point undivided particle without structure that is not constructed from further fundamental entities.



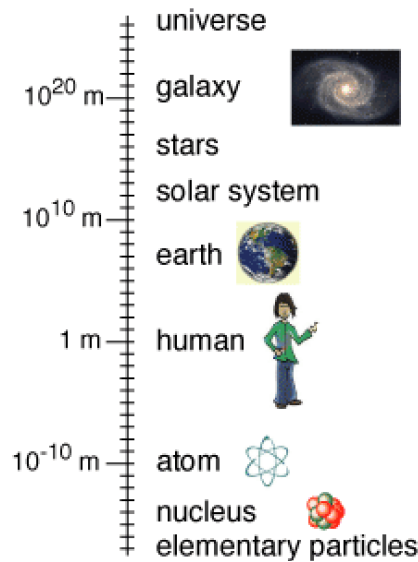


Figure 1.1: *Length scale in physics.*

In 1860, scientists thought that the world is made of 80 elements reacting in different ways or having different shapes or sizes. They are summarised in the, so called, "periodic table" of chemical elements [2, 3] where the hydrogen was considered the lightest element. In the early 1900s, most physicists believed that physics was complete, described by classical mechanics [4], thermodynamics [5], and the Maxwell theory [6]. However, experiments which measured at the beginning of the 20th century the energy distribution of the blackbody radiation and the photo-electric effect arise serious doubts and inadequacies of the classical physics.

With the use of particle accelerators, we can understand better the nature of the world around us. First by the discovery of the "electron", approximately 2000 times lighter than hydrogen, in 1897 by using a cathode ray tube (CRT) [7]; A "first" simple linear accelerator. Then the atom model is predicted followed by Rutherford's gold foil experiment, in 1911, showing that the atom is mostly empty space with a tiny, dense, positively-charged "nucleus". Based on these results, Rutherford proposed the nuclear model of the atom. Up to 1932, the protons and neutrons are observed confirming Rutherford's hypothesis. At that time, physicists claimed that everything is simply made of three particles: electron (e), proton (p) and neutron (n). But again some phenomena can not be explained by only those three particles; such as the discovery of cosmic rays and the positron, an electron with a positive charge. With the advent of particle accelerator in the 1950's in order to investigate the cosmic rays, about 80 new elementary particles were discovered, referred as a zoo particles, like the pion, delta, lambda and so on (see Figure 1.2). This has led to the need for another classification of particles. In the same time, the success of quantum electrodynamics in the late 1940s had produced a boom in elementary particle theory, and then the market crashed. It was realised that the four-fermion theory of weak interactions had infinities that could not be eliminated by the technique of renormalisation, which had worked so brilliantly in electrodynamics [8]. Therefore, 1950's was a time of frustration and confusion. In 1960, quarks idea (proton made of small things called quarks) is suggested and confirmed by experiment in the Stanford Linear Accelerator Center (SLAC). This leads to finally a successful theory called the Standard Model in the mid-1970s.

Today, SM describes all forces existing in nature except gravity and dark energy. In total, it includes sixty one fundamental particles, explained in the following.

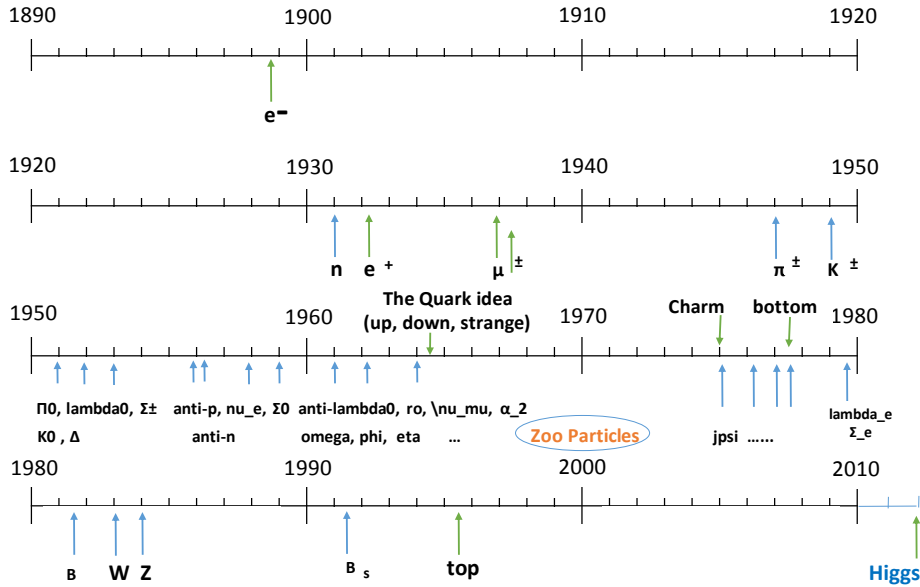


Figure 1.2: *History of particles discovery.*

## 1.2 Matter particles

A particle is defined by : a mass " $m$ ", an intrinsic spin angular momentum " $s$ ", an electric charge " $q_e$ " and a life-time " $\tau$ ". They are called observables or quantum numbers. Half-integer spin particles ( $1/2, 3/2, 5/2$  and so on) are called fermions. They all obey the Pauli exclusion principle [9] which establishes that two identical fermions can not occupy the same energy state. They are divided into two kind of particles: Quarks and leptons.

Six types or "flavours" of quarks exist namely top " $t$ ", bottom " $b$ ", charm " $c$ ", strange " $s$ ", down " $d$ " and up " $u$ ". The top is the heaviest particle among all fundamental particles. Up, charm and top share the same electric charge of  $(+2/3)$ . They are usually called the up-type particles. The remaining quarks (down-type particles) have a charge of  $(-1/3)$ . More than thirty years were needed to discover all of them in the period from 1964 to 1995. Quarks are never observed lonely. For example, a proton (neutron) is made of 3 quarks  $uud$  ( $udd$ ) giving  $+1$  ( $0$ ) total electric charge. They are always bound together to give a particle of an integer electrical charge and white colour charge<sup>2</sup> particle (colour confinement). This confinement gives rise to the 'hadronisation' process and explains how more and more pairs of quarks are created at low energies regime instead separating two quarks, similarly as someone is trying to separate 2 ends of a rubber band (the rubber band-description). The quarks, their approximate masses, their charges and mean life-time are listed in Table 1.1.

Same number of Leptons (six) are also discovered in the period 1897-2000, namely electron " $e$ " (well known particle that orbits over a nucleus in the atom) having negative electric charge  $(-1)$ , two other similar but heavier particles (muon " $\mu$ " and tau " $\tau$ ") and finally neutrinos, neutral and very low mass particles, associated to each of the three leptons during some decays ( $W$  boson, will be introduced later, decaying to an electron and a neutrino): electron neutrino  $\nu_e$ , muon neutrino  $\nu_\mu$  and tau neutrino  $\nu_\tau$ . Their

<sup>2</sup>A 'color charge' (red, blue or green) is just a quantum number characterising the particles of strong interaction (quarks and gluons).

Quark flavour	mass $m$ (MeV)	spin $s$	electric charge $q_e$	life-time $\tau$ (s)
up "u"	2	1/2	+2/3	stable
down "d"	5	1/2	-1/3	stable
strange "s"	96	1/2	-1/3	$10^{-8}$
charm "c"	1 270	1/2	+2/3	$10^{-12}$
bottom "b"	4 180	1/2	-1/3	$10^{-12}$
top "t"	173 210	1/2	+2/3	$5 \times 10^{-25}$

Table 1.1: *Quarks and their properties. The mass "m" presents the approximate value of the rest mass energy  $mc^2$  of each particle [10, 11].*

properties are summarised in Table 1.2.

Lepton	mass $m$ (MeV)	spin $s$	electric charge $q_e$	life-time $\tau$ (s)
electron "e <sup>-</sup> "	0.511	1/2	-1	stable (*)
muon "μ <sup>-</sup> "	105.7	1/2	-1	$2.2 \times 10^{-6}$
tau "τ <sup>-</sup> "	1 776.9	1/2	-1	$2.9 \times 10^{-13}$
electron neutrino "ν <sub>e</sub> "	< 0.002	1/2	0	stable
muon neutrino "ν <sub>μ</sub> "	< 0.2	1/2	0	-
tau neutrino "ν <sub>τ</sub> "	< 18.2	1/2	0	-

Table 1.2: *Leptons and their properties. (\*) The mean life-time of an electron is estimated to about  $6.6 \times 10^{28}$  years.*

Fermions are also divided into 3 generation according to the stability of the particles (life-time). The first generation includes particles that can combine to form stable atoms: electron,  $\nu_e$ , up and down quarks. Top quark belongs to the third generation. It decays before the hadronization since it has so short life time (so massive particle). Therefore, it is only known from its decay products.

**Anti-matter particles** Both neutral and charged particles have their anti-particles with same mass and opposite quantum numbers. For example, the positron "e<sup>+</sup>" is the anti-particle of an electron with the same mass but an equal and opposite charge. Anti-neutrino is the anti-particle of a neutrino with the same no charge and half spin but having opposite signs of lepton number<sup>3</sup> and chirality<sup>4</sup>. The existence of anti-particles is first predicted by Dirac in 1927 then observed later on (see Section 1.1). The anti-quarks are:

$$\bar{u}, \bar{d}, \bar{s}, \bar{c}, \bar{b}, \bar{t}$$

The anti-leptons are:

$$e^+, \mu^+, \tau^+, \bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$$

<sup>3</sup>The lepton number is one of the conserved quantum numbers in a reaction, representing the number of leptons minus the number of anti-leptons.

<sup>4</sup>The chirality of a particle is determined by whether the particle transforms in a right- or left-handed representation of the Poincaré group. It refers to how a particle's quantum mechanical wave function behaves when a particle is rotated.

Particle–anti-particle pairs can annihilate each other producing other particles giving rise to the quantum field theory (QFT).

## 1.3 Particles interactions

The fermions interact with each other by exchanging spin-1 gauge bosons (vector bosons). Basically, there are five bosons discovered so far: 4 vectors and one scalar (spin-0). The vector bosons (gauge bosons), determining the type of interaction, are: Gluon " $g$ " responsible for the strong nuclear interaction (bound the protons and neutrons together in the nucleus); Photon " $\gamma$ " for the electromagnetic (EM) interactions (give rise to light, electricity and magnetism);  $W^\pm$  and  $Z$  bosons for the weak nuclear interactions (permit a neutron to decay to a proton:  $n \rightarrow p + W^- \rightarrow p + e^- + \bar{\nu}_e$ ). The only existing scalar boson called the "Higgs" boson, the last particle discovered in 2012, is responsible to provide a mass to all massive particles. More details about this particle will be given in the next chapter. Some properties of bosons are listed in Table 1.3.

Boson	Interaction	$m$ (GeV)	spin $s$	$q_e$	$\tau$ (s)	Strength	Range (m)
photon " $\gamma$ "	EM	0	1	0	stable	$10^{-3}$	$\infty$
$W^\pm$	weak	80.4	1	$\pm 1$	$10^{-25}$	$10^{-2}$	$10^{-18}$
$Z$	weak	91.2	1	0	$10^{-25}$	$10^{-2}$	$10^{-18}$
gluon " $g$ "	strong	0	1	0	stable	1	$10^{-15}$
Higgs	Yukawa (*)	125	0	0	$10^{-22}$	-	$\infty$

Table 1.3: *Bosons and their properties. The relative "Strength", depending on the energy scale, are given at low energies of 1 to 100 MeV [12]. (\*) The "Yukawa" interaction describes the mass acquisition of massive elementary particles due to the coupling between the Higgs field and massless fermion fields.*

Unlike fermions, bosons could be described by the same quantum numbers. An overview of the theoretical framework of those interactions is presented in the following. The SM is defined by writing a mathematical function called the "Lagrangian density". As in classical mechanics, the integral of the Lagrangian density over space and time should be minimum (principle of least action). This determines how the particle, presented by a wave function, evolves in time (dynamics). Requiring a certain mathematical symmetry of the Lagrangian, called gauge symmetry, determines how particles interact by leading to the necessity that a gauge boson (namely photon,  $W^\pm$ ,  $Z$  and gluon) exist.

### 1.3.1 Quantum Electrodynamics

The part of the SM describing electromagnetic processes, such as the scattering of two electrons, is called "quantum electrodynamics" (QED). The theory of QED is perhaps the most precisely tested physics theory ever achieved. It describes the interaction of charged fermion particles by emitting photons. The most precise prediction of this very precise theory is the magnetic strength of the electron, what physicists call the magnetic moment. Prediction and measurement agree to 12 digits of precision, which is a sign that QED is correct and successfully describes the reality. QED calculation relies on two components: perturbation theory and Feynman diagrams. Feynman diagram is a simplified picture that translates one complex equation giving the probability of a certain phenomenon to

happen (i.e. the amplitudes  $\mathcal{M}$  to calculate cross section of collisions, scattering and interaction between two particles or decay width of unstable particles). Perturbation theory is an approximation theory using the first order of Feynman diagram to calculate the amplitudes since the other orders are negligible. An example of the first order of Feynman diagram of QED is shown in Figure 1.3 where a fermion  $f$  is scattering/interacting by emitting a photon  $\gamma$  ( $f \rightarrow f\gamma$ ). This diagram shows a basic coupling between a charged fermion and a photon called fermion-photon "vertex"<sup>5</sup>. The QED coupling is simply the electric charge of the fermion  $n \times e = 2n\sqrt{\pi\alpha}$ , where  $\alpha = 1/137$  [13] is the fine structure constant and "n" is 1 for the charged leptons, 1/3 and 2/3 for up- and down-type quarks and 0 for neutral leptons (therefore, neutrino-photon vertex does not exist).

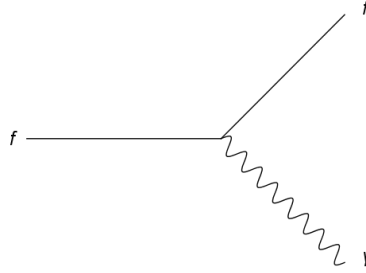


Figure 1.3: *Feynman diagram of the basic fermion-photon vertex in QED.*

For explanation of physical phenomenon in QED, Figure 1.4 shows the possible Feynman diagrams for Compton scattering where an electron absorbs a photon and deviates from its initial path ( $e^-\gamma \rightarrow e^-\gamma$ ) and a pair production where two photons interact, annihilate and give rise to two pairs of electrons  $\gamma\gamma \rightarrow e^-e^-$ . For each case, there are two photon-electron vertices. Hence the amplitude is proportional to the square of the QED coupling  $\mathcal{M} \propto e^2$ . The cross section of this process  $\sigma$  is proportional to the square of the amplitude:  $\sigma \propto |\mathcal{M}|^2 \propto e^4 \propto \alpha^2$ .

Mathematically, QED is an abelian gauge theory<sup>6</sup> with the symmetry group  $U(1)$ . The QED Lagrangian for a spin-1/2 field (particle)  $\psi$  interacting with the electromagnetic field  $A_\mu$  is given by:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.1)$$

where:  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$  is the EM field tensor.  $D_\mu = \partial_\mu + ieA_\mu + ieB_\mu$  is the gauge covariant derivative<sup>7</sup> with  $B_\mu$  the external field imposed by external source.  $\gamma^\mu$  are Dirac matrices [15].

### 1.3.2 Weak interactions

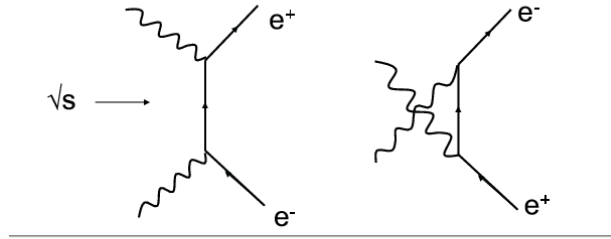
Weak interactions include the exchange of  $W^\pm$  and  $Z$  bosons. Unlike QED, the fermion could change its type or family by emitting a  $W^\pm$  boson. For example, the electron is transformed into an electron neutrino by emission of a  $W^-$ . Figure 1.5 shows the basic coupling between a fermion and a  $W^\pm / Z$  boson. For the lepton- $W$  vertices, only the type of particle is changed. However, the charged lepton and the neutrino must be of the

<sup>5</sup>One of the Feynman rules is that every vertex is associated with a "coupling" or charge, which characterises the strength of the interaction.

<sup>6</sup>An Abelian group [14] is a group satisfying the commutative law.

<sup>7</sup>Equations written using the covariant derivative preserve their physical properties under gauge transformations.

a) Pair production:



b) Compton scattering:

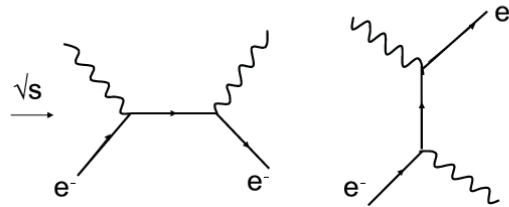


Figure 1.4: Feynman diagrams of a) an electron pair production  $\gamma\gamma \rightarrow e^-e^-$  and b) Compton scattering  $e^-\gamma \rightarrow e^-\gamma$ . Every diagram is built by the basic electron-photon vertex of QED. The arrow shows the time evolution of the process (incoming particles are on the left side; ongoing particles at the right).  $\sqrt{s}$  is the center of mass energy of the incoming particles.

same family (generation), i.e. an electron could not be transformed to a tau neutrino for example (lepton family number conservation). For the quark-W vertices, both type and family could be transformed. i.e, the vertex could contain any up-type quarks ( $u, c$  or  $t$ ) paired with any down-type ( $d, s$  or  $b$ ). For the fermion-Z vertices, no type neither family is allowed to be changed in the SM (similarly to QED). For example,  $u \rightarrow uZ$  is allowed, at tree level, but not  $u \rightarrow cZ$  despite the charge, baryon and lepton numbers are conserved. In contrast to the photon, the  $Z$  boson does couple to neutrinos. For every fermion-W

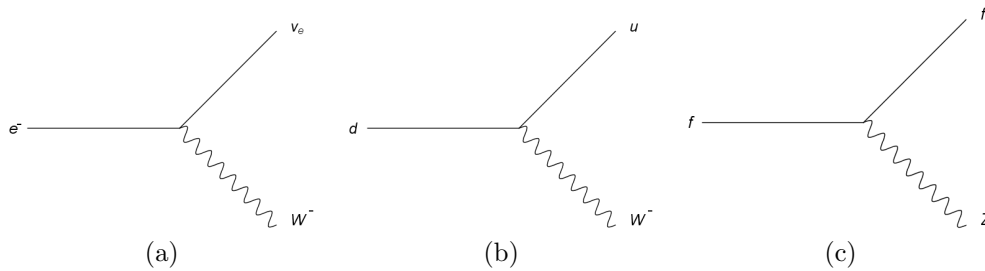


Figure 1.5: Feynman diagrams of the basic vertices in the weak interaction. (a) Lepton-W vertex (b) quark-W vertex and (c) fermion-Z vertex.

vertex, the weak coupling strength is given by  $g=0.65$  [12] at low energies (1-100 MeV). It is more precise to use the Fermi constant,  $G_F$ , the ratio of  $g$  to the mass of  $W$  boson, defined by:

$$G_F = \frac{\sqrt{2}g^2}{8M_W^2} = 1.166 \times 10^{-5} \text{ GeV}^{-2}$$

In analogy with the fine structure constant  $\alpha$ , the weak coupling constant  $\alpha_w$  is defined as:

$$\alpha_w = \frac{g^2}{4\pi} = \frac{1}{29.5}$$

One can note that the weak strength at such energies is stronger than the EM strength (1/137). However, at the macroscopic, atomic and even nuclear distances it becomes much weaker and proportional to  $e^{(-M_W r/\hbar c)}$  [16].

The fermion-Z coupling is more complicated. It is a function of  $e$  and  $g$  which depend on the fermion type and its spin state.

*CKM matrix* the coupling  $g$  is the same for lepton-W and quark-W vertices. However, the amplitude has an additional factor for the quark-W vertex called *CKM matrix element* giving the probability that a quark is transformed to another quark:

$$\begin{pmatrix} V_{ud} = 0.975 & V_{us} = 0.220 & V_{ub} = 0.003 \\ V_{cd} = 0.220 & V_{cs} = 0.975 & V_{cb} = 0.004 \\ V_{td} = 0.008 & V_{ts} = 0.040 & V_{tb} = 0.999 \end{pmatrix} \quad (1.2)$$

### 1.3.3 Quantum Chromodynamics

The part of the SM describing the strong interactions between quarks and gluons is called "quantum chromodynamics" (QCD). Figure 1.6 shows the basic coupling between a fermion and a gluon. The gluon-gluon interaction (triple gluon and four-gluon vertices) is also shown.

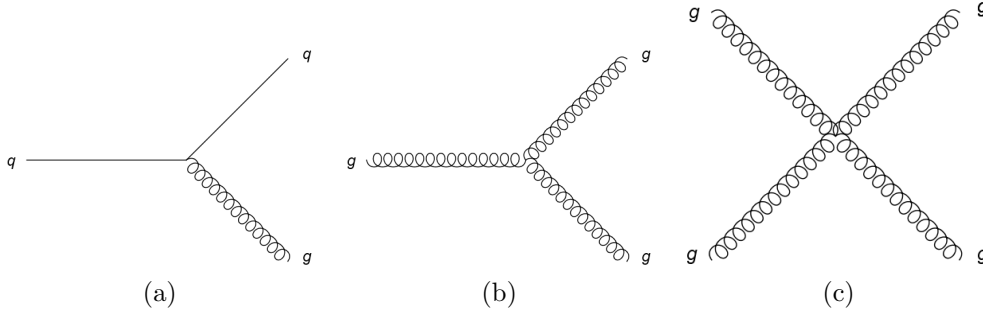


Figure 1.6: *Feynman diagrams of the basic vertices in the QCD. (a) quark-gluon vertex (b) triple gluon vertex and (c) four-gluon vertex.*

The strong coupling strength, independent of the flavour of the quark, is given by  $g_s \approx 1.2$  at an energy scale close to the mass of the  $Z$  boson [10]. The strong coupling constant  $\alpha_s$  is defined as:

$$\alpha_s = \frac{g_s^2}{4\pi} \approx 0.12$$

Mathematically, QCD is a non-abelian gauge theory with the symmetry group SU(3). The QCD Lagrangian for a spin-1/2 field (particle)  $\psi$  interacting with the strong field  $A_\mu^\alpha$  is given by:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i (i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G_{\mu\nu}^\alpha G^{\mu\nu}_\alpha \quad (1.3)$$

where:  $G_{\mu\nu}^\alpha = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha + gf^{abc}A_\mu^b A_\nu^c$  is the gauge invariant gluon field strength tensor.  $f^{abc}$  is the structure constant of SU(3).



### 1.3.4 Electroweak interaction

EM and weak forces are unified in Electroweak theory giving  $SU(2) \times U(1)$  group: Weak and EM interactions are just two aspects of the same force unified at very high temperature ( $10^{16}$  K) of  $10^{-12}$  s after the Big Bang [17]. This theory implies "no" masses for bosons. However, this contradicts the observation where the  $W^\pm$  and  $Z$  bosons are massive. To solve this, a spontaneous symmetry breaking is suggested associated with a Higgs boson discussed in the next chapter.

In electroweak theory, all triple and four gauge boson couplings in which the photon and  $Z$  interact with  $W^\pm$  are possible. However, note that no  $\gamma\gamma\gamma$  and  $ZZZ$  vertices exist since photons/ $Z$  do not couple to themselves (unlike gluons in QCD). The coupling strength for  $\gamma W^+W^-$  and  $ZW^+W^-$  are the charge  $e$  and  $\sqrt{g^2 - e^2}$  respectively. The coupling strength for  $\gamma\gamma W^+W^-$  and  $ZZW^+W^-$  are proportional to  $e^2$ . The  $WWW$  or  $WWWW$  coupling vertices are also possible but they are not enough to describe some interactions such as  $WW$ -scattering where the amplitude  $M(WW \rightarrow WW)$  is found to be proportional to the square of the center of mass energy  $s$  which violates the unitarity. One solution to this problem is suggested by introducing a new scalar particle, the Higgs boson  $H$ , where the Higgs mass should be less than 1 TeV and its coupling with any other particle  $X$  is proportional to the particle mass ( $g_{XXH} \propto M_X$ ).

## 1.4 Problems/Outlook

SM brings wonderful collaboration between experiment and theory. Observation at particle colliders of all of the fundamental particles predicted by the Standard Model has been confirmed. However, SM is not a complete theory that describes everything in the universe [18]. For example, it failed to quantize gravity and describe the most dominant components in the universe: dark energy and dark matter [19, 20]. The matter-antimatter asymmetry is another unexplained feature in the Standard Model which predicts almost equal amounts of particles and anti-particles [21, 22]. Moreover, many parameters, called free parameters, are not predicted by SM, such as the Higgs mass, and are just measured from experiment. Hence there is no explanation why the Higgs has 125 GeV or why the coupling constants have the values they actually have. On top of that, some discrepancies are seen in SM. For example, the neutrinos masses are considered to be zero which is not seen by neutrino oscillation experiment [23]. Many other open questions are also raised:

- why three lepton families? why is the first family privileged?
- are there additional (heavy) leptons and bosons?
- Why SM equations obey to a specific symmetries?
- could one unify the coupling constants?

All this calls for a more fundamental theory that solve SM problems without contradicting experimental data, called new physics. Best candidates are SUSY [24], extra-dimensions and technicolour which all predict new physics at the order of TeV scale. This needs a powerful machine to explore TeV energy range such as the Large Hadron Collider "LHC", described in chapter 3, able to reproduce the Universe energy  $10^{-10}$  s after the Big Bang.

Another way of looking at a sign of new physics is using experimental results of some phenomenon at high energy and see if they are indeed deviated from the SM prediction.



# 1.5 Summary

There are two families of fundamental particles in the SM: "Fermions" and "bosons". Any elementary particle with a half integral spin quantum number is a fermion. Any elementary particle with an integral spin quantum number is a boson. They obey different statistical rules. Fermions could not occupy the same place at the same time (Pauli exclusion principle). This could be translated in quantum world by: two fermions could not be described by the same quantum numbers. Bosons do. Fermions present matter/anti-matter particles and could be quarks or leptons. Bosons could be a vector or scalar field. The only discovered scalar boson is the Higgs (spin-0). All other bosons in the SM are vectors (W, Z, photon and gluon) with spin equals 1<sup>8</sup>.

Today, SM describes all forces existing in nature except gravity and dark energy [25]. Particles can be classified in different ways. In total, there are 61 (18) known elementary particles (particles per colour charge excluding anti-particles) in nature: 48 (12) fermions and 13 (6) bosons. Only 4 matter particles describe everything in the world around us, interacting with each other via vector bosons: Two quarks (up and down) and two leptons (electron and electron neutrino), called the first generation of elementary particles. The other 2 generations of fermions are identical to the first one but heavier and can be reproduced in high energy accelerators such as the LHC. Figure 1.7 summarises our best knowledge of all elementary particles today. Maybe tomorrow, different view or level of elementary particle classification will be discovered, more or less simplified.

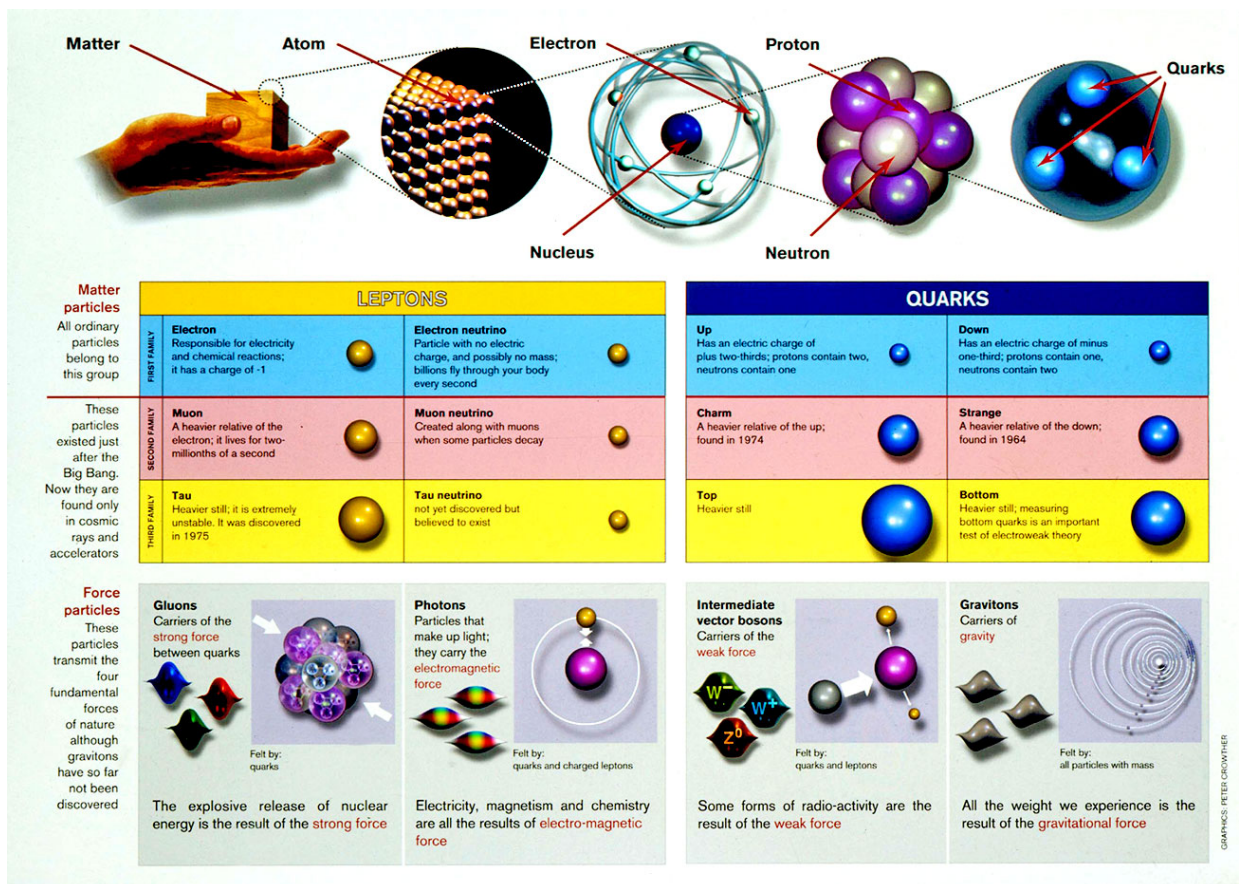


Figure 1.7: Summary of elementary particles in SM.

<sup>8</sup>A particle with spin-2 describes tensor field (that stretches and squeezes space in two directions).

# Chapter 2

## Higgs Physics

The Higgs boson is the first elementary scalar particle discovered in nature, whose non-zero vacuum expectation value provides the mass of all elementary particles [26]. There are, in fact, four Higgs fields ( $h^+, h^-, h^0, H$ ): Three of which (Goldstone bosons) are massless and eaten by the weak gauge bosons to allow them to acquire a longitudinal polarisation and therefore to become massive. The charged Higgs fields  $h^+$  and  $h^-$  give rise to the mass of  $W^+$  and  $W^-$  bosons. The third neutral Higgs  $h^0$  contributes to form the massive Z boson. The fourth particle  $H$  is responsible for the electroweak symmetry breaking, predicted by the theorists Peter Higgs, François Englert and others in 1964 [27, 28, 29]. The Higgs boson is discovered on the 4th of July 2012 with a mass around 125 GeV/ $c^2$  [30, 31] by ATLAS and CMS experiments at LHC. Determining the Higgs properties (mass, spin, life-time, width, couplings,...) is an important part of the high-energy physics programme in this decade [10]. My thesis is a part of this programme. A precise calculation of both Higgs production cross sections and decay widths with their respective uncertainties is also essential for a correct interpretation of the data.

In this chapter, a description of the Higgs mechanism is presented in Section 2.1, followed by Higgs production and Higgs decay described in Sections 2.2 and 2.3 respectively. The measurements of the Higgs properties are summarised in Section 2.4. Finally, searches at LHC of the  $t\bar{t}H$  production mode, the topic of my thesis, is introduced in Section 2.5.

### 2.1 Brout-Englert-Higgs mechanism

Without the Higgs mechanism, all bosons, including  $W^\pm$  and  $Z$ , would be massless. This mechanism is described by a mathematical procedure in which particles, both the weak force carriers and the matter particles, in a field theory can acquire mass due to spontaneous breakdown of symmetry. This is achieved by introducing just one new particle, the by now famous Higgs boson. The understanding of this phenomenon leads to the final mathematical formulation of the SM Lagrangian  $\mathcal{L}_{\text{SM}}$  which can be divided into four different parts:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{YM} + \mathcal{L}_{\text{ferm}} + \mathcal{L}_H + \mathcal{L}_{Yuk} \quad (2.1)$$

- $\mathcal{L}_{YM}$  (The Yang Mills part): describes the dynamics of the gauge fields.
- $\mathcal{L}_{\text{ferm}}$ : describes the interaction of the gauge fields with the fermions.
- $\mathcal{L}_H$  (Higgs part): describes the interaction of the gauge fields with the Higgs boson by introducing new complex scalar (doublet) field  $\phi$ .

- $\mathcal{L}_{Yuk}$ : describes the interactions of Higgs boson with all fermions via Yukawa coupling.

Before introducing the problem of the electroweak symmetry breaking (EWSB), the unbroken gauge symmetry theory describes the dynamics of gauge bosons and matter fermions, with the two first terms of SM Lagrangian in Eq.(2.1):

$$\mathcal{L}_A = \mathcal{L}_{YM} + \mathcal{L}_{ferm} \quad (2.2)$$

In such theories, applicable to both QED and QCD, gauge invariance forbids to have an explicit mass term for the gauge vector bosons in the Lagrangian. However, electroweak interaction requires massive bosons to describe weak force carriers  $W$  and  $Z$  bosons which have masses of 80 GeV and 91 GeV, respectively. The so called Higgs mechanism provides a very simple and economical solution to this problem by adding to the Lagrangian  $\mathcal{L}_A$  a complex scalar field  $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$  with Lagrangian:

$$\mathcal{L}_H = \frac{1}{2}(D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \quad (2.3)$$

where:  $V(\phi)$  is the Goldstone potential (Mexican hat potential):

$$V(\phi) = \frac{1}{2}\mu^2(\phi^\dagger\phi) + \frac{1}{4}\lambda(\phi^\dagger\phi)^2 \quad (2.4)$$

The  $(\phi^\dagger\phi)^2$  term represents a self-interaction of a strength  $\lambda$ . The parameter  $\lambda$  is positive, therefore, two cases are distinguished:

- $\mu^2 > 0$ : there is a unique minimum at  $\phi = 0$ . In this case, the Lagrangian describes the physics of a massless vector boson interacting with a massive charged scalar particle with mass  $\mu$ .
- $\mu^2 < 0$ : the potential develops a degeneracy of minima at  $\phi_0 = \pm\sqrt{-\frac{\mu^2}{\lambda}} \equiv v$ . In this case, we need to define the ground state  $\phi_0$  among the potential minima for transforming the scalar field into  $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) + \phi_0$ . Then, we can rewrite the Lagrangian, separating  $\phi_1$  and  $\phi_2$ , in such way that the original gauge symmetry is hidden (broken).

Considering the simple case of the classical abelian Yang Mills theory where:

$$\mathcal{L}_A = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.5)$$

$$D_\mu = \partial_\mu + igA_\mu \quad (2.6)$$

By choosing  $\phi_0 = +\sqrt{-\frac{\mu^2}{\lambda}}$ , the Lagrangian can be rearranged as follows:

$$\begin{aligned} \mathcal{L} = \mathcal{L}_A + \frac{g^2v^2}{2}A^\mu A_\mu + \frac{1}{2}(\partial^\mu\phi_1)^2 + \mu^2\phi_1^2 + \frac{1}{2}(\partial^\mu\phi_2)^2 + gvA_\mu\partial^\mu\phi_2 + g^2vA^\mu\phi_1A_\mu + \dots \\ - \frac{1}{6}(6\lambda v)\phi_1^3 + \frac{1}{24}(6\lambda)\phi_1^4 + \dots \end{aligned} \quad (2.7)$$

From Eq.(2.7), we can distinguish 3 fields:

- Massive vector field  $A^\mu$  with mass  $m_A^2 = g^2 v^2$ .
- Massive real scalar field  $\phi_1$ ; with mass  $m_{\phi_1}^2 = -2\mu^2$ . By putting  $\phi_1 = H$  (Higgs boson), the Lagrangian describing Higgs boson can be rewritten as:

$$\mathcal{L}_H = \frac{1}{2}(\partial^\mu H \partial_\mu H - m_H^2 H^2) \quad (2.8)$$

where:  $m_H = \sqrt{2\lambda} v$  is the Higgs mass. The minimum  $v = \frac{\mu}{\sqrt{\lambda}}$  presents the Vacuum expectation value (vev) of the Higgs field. The parameter  $\lambda$  is called the Higgs self-coupling. The coupling strength for HHH and HHHH vertices could be extracted from the factors associated to  $\phi^3$  and  $\phi^4$  terms where  $g_{HHH} = 6\lambda v = 3m_H^2/v$  and  $g_{HHHH} = 6\lambda = 3m_H^2/v^2$ . The higgs coupling to weak bosons could be extracted from the  $A^\mu \phi_1 A_\mu$  terms where  $g_{HVV} = g^2 v = 2m_V^2/v$ . It is worth to notice that the Higgs doesn't couple to photons nor to gluons (zero mass).

- Massless scalar field  $\phi_2$  (Goldstone boson) which couples to the gauge vector field  $A^\mu$ .

The reformulated Lagrangian (2.7) is not even in  $\phi_1$ ; the symmetry has been "broken". So, by using Goldstone model, we come up with generating masses for gauge bosons by combining local gauge invariance with spontaneous symmetry broken (EWSB).

As the Higgs mechanism describes how the gauge field  $A^\mu$  in the Lagrangian  $\mathcal{L}_H$  can be massive. The same mechanism could also describes how the fermions acquire their masses. Starting from the Lagrangian of a free (Dirac Fermion) field which is invariant under local gauge transformations but deals with only massless fermions and fields. This is because a standard mass term for a fermion breaks gauge invariance (only left (right) handed fields  $f_L(f_R)$  transform under SU(2)). However, both fermions and the weak gauge fields are heavy. By introducing a new scalar field, the mass terms becomes gauge invariant. Masses are then generated by requiring this new scalar field (Higgs field) to acquire a vev. Then, the corresponding fermion Lagrangian could be rewritten as:

$$\mathcal{L}_{Yuk} = -\frac{1}{\sqrt{2}}\lambda_e(H - v)\bar{e}_L e_R + \dots \quad (2.9)$$

where  $m_f = \lambda_f v$  is the fermion mass. The parameter  $\lambda_f = g_f = m_f/v$  is the fermion Yukawa coupling.

## 2.2 Higgs production

At LHC, there are four main mechanisms to produce the Higgs boson: Gluon-gluon fusion (Section 2.2.1), vector boson fusion (Section 2.2.2), Higgs-Strahlung (Section 2.2.3) and  $t\bar{t}H$  production (Section 2.2.4). The less abundant production processes  $pp \rightarrow bbH$ , the production in association with a single top quark<sup>1</sup>  $qq/qb/gb \rightarrow tHb/tHq/tHW$  (Figure 2.1a) are not described in this context. In the following, the cross sections of each production modes is given for  $m_H = 125.09$  GeV [32] at the CERN LHC pp collision center of mass energy of  $\sqrt{s} = 13$  TeV [33] (8 TeV [34]).

<sup>1</sup>The production in association with a single top quark can bring interesting information on the sign of top Yukawa coupling

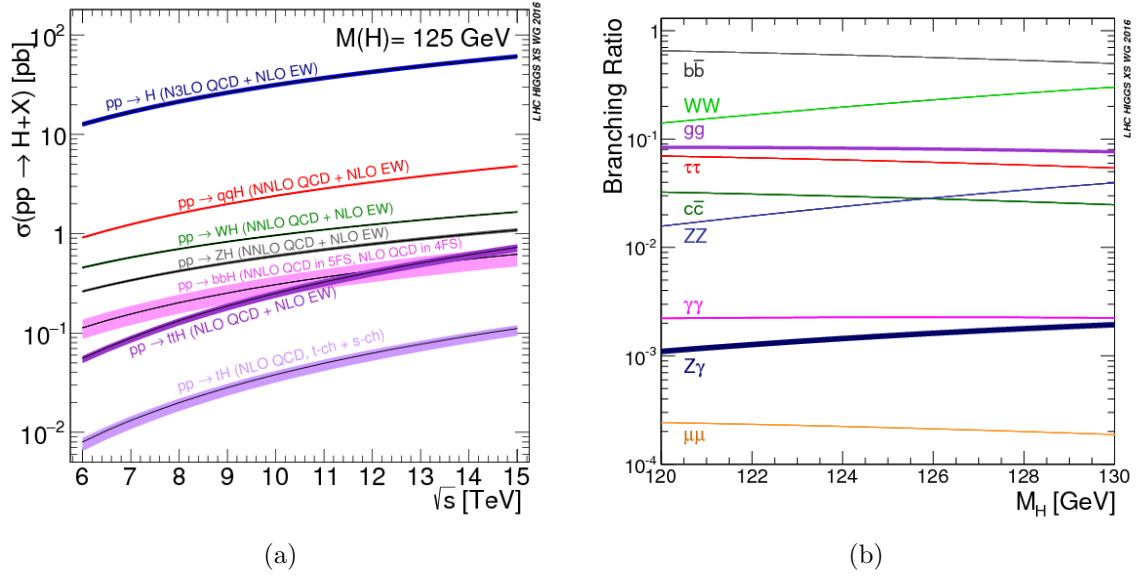


Figure 2.1: (a) SM Higgs boson production cross-sections as a function of the LHC center-of-mass energy. (b) SM Higgs boson branching ratios as a function of the Higgs boson mass.

## 2.2.1 Gluon-gluon fusion ( $ggF$ )

The dominant Higgs production mechanism at the LHC centre of mass energy is the gluon-gluon fusion ( $ggF$ ) where a virtual loop<sup>2</sup> radiates a Higgs (Figure 2.2), contributing by 88 % (87 %) of the total cross section of the four main Higgs productions. Only fermions are involved in this loop since gluons do not couple weak bosons. Where the top quark, the most massive fermion, is dominating giving the fact that the Higgs coupling is proportional to the fermion mass (see section 2.1). The  $ggF$  is the main mode used for the Higgs discovery on 2012 and it provides indirect calculation of Higgs-top coupling (top-Yukawa coupling). It is characterised by a large next-to-leading order (NLO) QCD corrections. The predicted cross section of  $ggF$  is  $48.5^{+3.7}_{-4.4}$  ( $19.2 \pm 2.0$ ) pb calculated at N<sup>3</sup>LO [33] (NNLO [34]) in QCD.

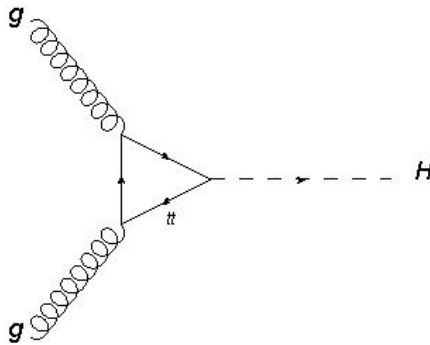


Figure 2.2: Feynman graph for gluon-gluon fusion process.



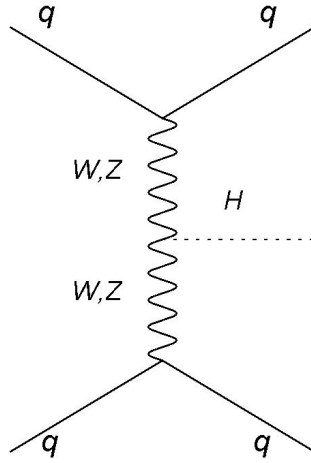


Figure 2.3: *Feynman graph for VBF process.*

### 2.2.2 Vector boson fusion (VBF)

Higgs boson production via vector-boson fusion (VBF) can be visualised as the inelastic scattering of two quarks (anti-quarks), mediated by ( $t$  channel)  $W$  or  $Z$  exchange (Figure 2.3), contributing by 6.9 % (7.1 %) of the total cross section of the four main Higgs productions. A Higgs is produced centrally, while the two jets are produced in the forward/backward direction. The latter allow to reduce the backgrounds that are present in more inclusive Higgs searches which exploit the dominant gluon-gluon fusion production process [35]. The predicted cross section of the VBF is  $3.78 \pm 0.11$  ( $1.58 \pm 0.04$ ) pb including NNLO (NLO) QCD and NLO EW corrections [33, 34].

### 2.2.3 Higgs-strahlung

It is also called VH associated production with a vector boson (Figure 2.4). Unlike VBF, the Higgs in this channel is not central and the jets coming from  $W^\pm$  or  $Z$  decay are not showered as high- $p_T$  jets in the forward regions of the detector. The predicted cross section of WH and ZH is  $1.370 \pm 0.004$  ( $0.703 \pm 0.018$ ) and  $0.882^{+0.039}_{-0.033}$  ( $0.414 \pm 0.016$ ) pb, respectively, including NNLO QCD and NLO EW corrections [33, 34]. It contributes by 4.1 % (5.1 %) of the total cross section of the four main Higgs productions.

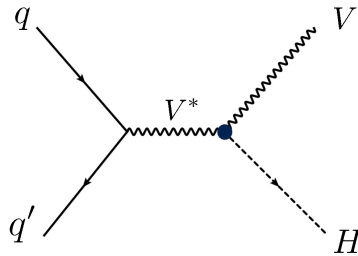


Figure 2.4: *Feynman graph for VH associated process. V stands for  $W^\pm$  or  $Z$  bosons.*

<sup>2</sup>The top quark, the most massive particle, gives the largest contribution to the loop.

### 2.2.4 $t\bar{t}H$ production

Higgs boson production via top fusion ( $t\bar{t}H$  production) can be visualised as the scattering of two gluons to a pair of top and anti-top, mediated by top quark exchange (Figure 2.5). Despite it has the smallest cross section, it is currently considered one of the most important channels at the LHC at Run2 since it provides a direct measurements of the top-Yukawa coupling. The predicted cross section of  $t\bar{t}H$  is  $0.507_{-0.006}^{+0.003}$  ( $0.133 \pm 0.014$ ) pb calculated at NLO in QCD and EW [33], contributing by only 0.9 % (0.6%) of the total cross section of the four main Higgs productions.

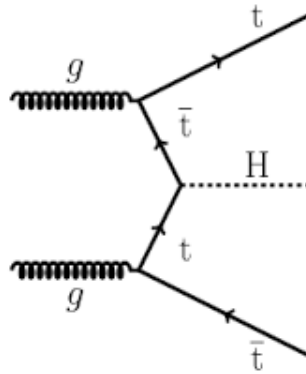


Figure 2.5: *Feynman diagram of  $t\bar{t}H$  process.*

## 2.3 Higgs decays

The Higgs boson of the SM is unstable and predominantly decays into the heaviest particle - anti-particle pair that is kinematically possible depending on the available energy [36]. In this section, a review of the main decay modes of the SM Higgs boson is presented (Figure 2.1b): Fermionic Tree-level Decay (Section 2.3.1), bosonic tree-level (Section 2.3.2) and Loop-induced Decay Modes (Section 2.3.3). In the following, the branching ratio of each SM Higgs decay mode is predicted for  $m_H = 125.09$  GeV [32, 33]. The corresponding uncertainties are the combination of the theoretical uncertainties, the parametric uncertainties from the quark masses and the strong coupling uncertainties.

### 2.3.1 Fermionic tree-level decay modes

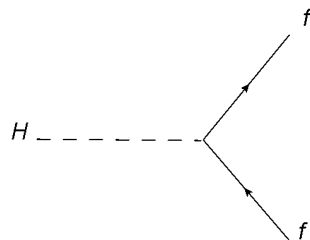


Figure 2.6: *Feynman diagram contributing to Fermionic Higgs decay.*

The Higgs boson can decay into heavy quarks and leptons (Figure 2.6):

$$H \longrightarrow c\bar{c} \tag{2.10}$$

$$H \longrightarrow b\bar{b} \quad (2.11)$$

$$H \longrightarrow \tau^+\tau^- \quad (2.12)$$

$$H \longrightarrow \mu^+\mu^- \quad (2.13)$$

The Higgs is more likely to decay into heavy fermions than light fermions, because the mass of a fermion is proportional to the strength of its interaction with the Higgs. Therefore, the most common decay is  $H \rightarrow b\bar{b}$  ( $58.1 \pm 0.7 \%$ ), followed by  $H \rightarrow \tau^+\tau^-$  ( $6.3 \pm 0.1 \%$ ),  $H \rightarrow c\bar{c}$  ( $2.9^{+0.2}_{-0.1} \%$ ) and  $H \rightarrow \mu^+\mu^-$  ( $0.02 \pm 0.00 \%$ ).

### 2.3.2 Bosonic tree-level decay modes

The diagram for this decay is given by the simple gauge vertex shown in Figure 2.7. The Higgs boson can decay into two charged weak bosons or two identical neutral bosons.

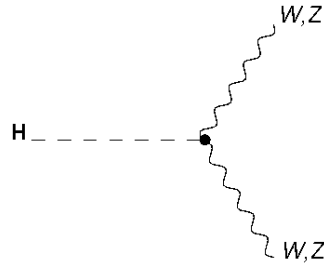


Figure 2.7: Feynman diagram of Higgs boson decay into  $W(Z)$  bosons:  $H \rightarrow W^-W^+(ZZ)$ .

$$H \longrightarrow W^-W^+ \quad (2.14)$$

$$H \longrightarrow ZZ \quad (2.15)$$

**$WW$  decay ( $21.5 \pm 0.3 \%$ )** The  $W$  bosons in the  $WW$  pair produced in the Higgs decay have opposite spin orientations, since the Higgs has spin zero. The  $W$ 's decay mainly hadronically, but this decay topology is experimentally difficult to exploit due to the high cross section of multi jet processes in pp collisions. Meanwhile, the fully leptonic decay channel of the  $W$ 's ( $WW \rightarrow l\nu l\nu$ ) can be used for Higgs search with high sensitivity. In this case, the positively charged lepton is preferably emitted in the direction of the  $W^+$  spin and the negative lepton in the opposite direction of the  $W^-$  spin. Therefore the two charged leptons are emitted close to each other.

**$ZZ$  decay ( $2.8 \pm 0.0 \%$ )** The  $ZZ$  decay mode of the SM Higgs boson has a lower branching fraction than the  $WW$  decay. However, the final state with four leptons (electrons and muons) from the  $ZZ$  decay is very clean and almost background free. Moreover, the four-lepton channel is one of the most precise final state to reconstruct the mass peak of the Higgs boson thanks to the high resolution of the lepton momentum reconstruction with the ATLAS and the CMS detectors.



### 2.3.3 Loop-induced decay modes

At lowest order, the couplings of the Higgs boson to the gauge fields are proportional to their mass. Therefore  $H$  does not couple to the photon at tree level. It is important, however, to observe that couplings that are absent at tree level may be induced at higher order in the gauge couplings by loop corrections. Particularly the couplings of the SM Higgs boson to pairs of photons, and to a photon and a Z weak boson via both a fermion loop and a W-loop as well as the coupling to pairs of gluons mediated by only a fermion loop as shown in Figure 2.8.

$$H \longrightarrow gg \quad (2.16)$$

$$H \longrightarrow \gamma\gamma \quad (2.17)$$

$$H \longrightarrow \gamma Z \quad (2.18)$$

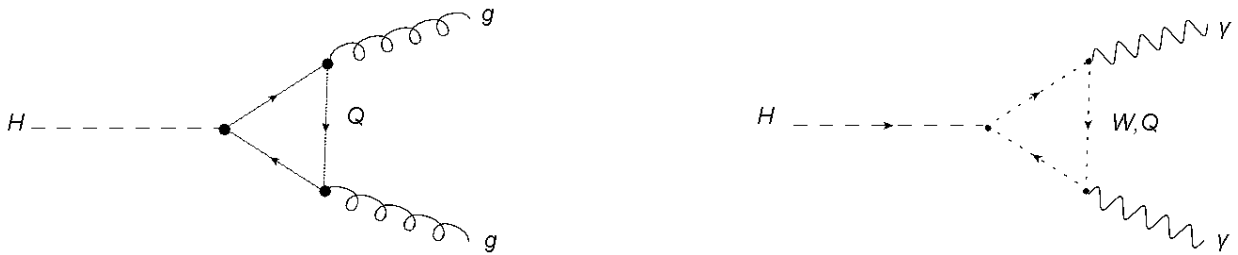


Figure 2.8: Feynman diagram contributing to loop decay modes with massless final states: (Left)  $H \longrightarrow gg$  (Right)  $H \longrightarrow \gamma\gamma(Z)$ .

**$gg$  decay** ( $8.2 \pm 0.4$  %) The most common such process is the decay into a pair of gluons through a loop of virtual heavy quarks as shown in the left of Figure 2.8.

**$\gamma\gamma$  decay** ( $0.23 \pm 0.01$ %) Despite its relatively low branching fraction and considerable reducible and irreducible backgrounds from SM QCD processes, the  $\gamma\gamma$  decay mode benefits from a clean signature, provided that a sufficiently high-resolution electromagnetic calorimeter is used.

## 2.4 Measurement of Higgs properties

The discovered SM-like Higgs boson, explored since 5 years ago, is presented in this section including the Higgs mass (Section 2.4.1), spin-parity (Section 2.4.2), width and life-time (Section 2.4.3) and couplings (Section 2.4.4). A comparison with the SM Higgs boson is also shown where 'same' properties are found when measurable.

### 2.4.1 Higgs mass

As a free parameter, the Higgs boson mass measurement is performed using the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4l$  decay modes thanks to the best mass resolution they offer. The Higgs boson resonance appears as a narrow peak in the mass spectra of its decays to two photons or to four charged leptons, as shown in Figure 2.9. The combined measured mass of the Higgs boson is  $M_H = 125.09 \pm 0.24$  GeV [32] which is in good agreement with the

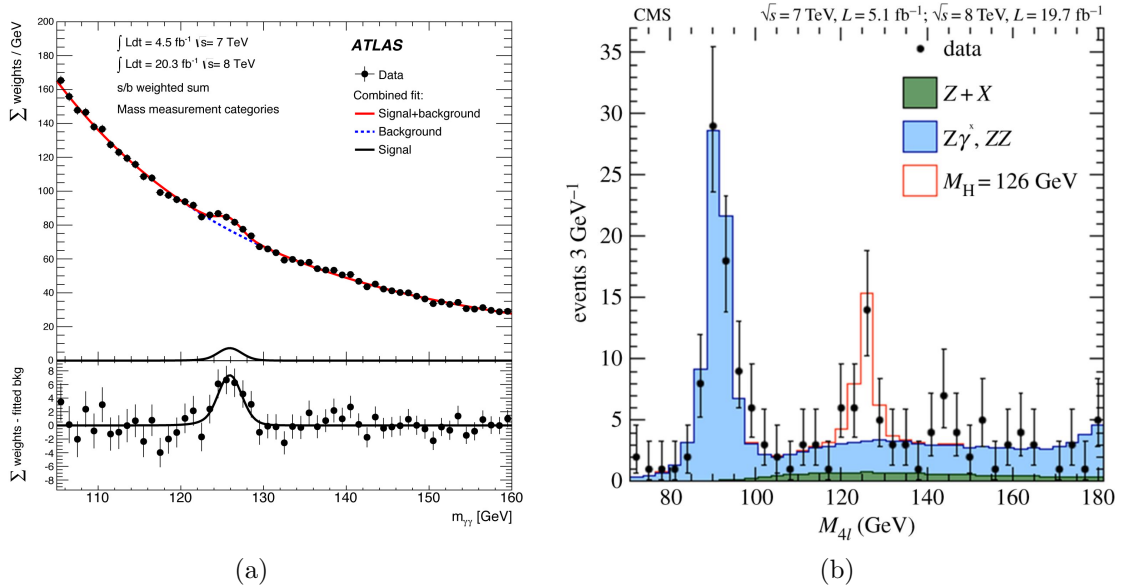


Figure 2.9: *SM Higgs boson mass measurements at a) ATLAS [37] and b) CMS [38] using  $5 \text{ fb}^{-1}$  and  $20 \text{ fb}^{-1}$  data samples respectively at  $\sqrt{s} = 7$  and  $8 \text{ TeV}$  for each experiment. a) Invariant mass distribution of di-photon candidates. The result of a fit to the background described by a polynomial and the sum of signal and background components is superimposed. The bottom inset displays the residuals of the data with respect to the fitted background component. b) The four-lepton invariant mass distribution in the CMS experiment for selected candidates relative to the background expectation. The expected signal contribution is also shown.*

global fit of the electroweak parameters [39] predictions ( $102.8 \pm 26 \text{ GeV}$ ). The dominant uncertainty in the measurement is due to statistical fluctuations mainly coming from  $H \rightarrow ZZ \rightarrow 4l$  channel, followed by systematic uncertainties related to the determination of  $\gamma$ ,  $e$  and  $\mu$  energy scales. A small instrumental uncertainty has been achieved thanks to an unprecedented understanding of the ATLAS and CMS detectors performance. The measured masses from individual channels and the two experiments are found to be consistent among themselves. In ATLAS, the combined measured resonance's Higgs mass is found to be  $125.36 \pm 0.41 \text{ GeV}$  [37], with a precision of about three parts per mille. In CMS, the combined measured resonance's Higgs mass is found to be  $125.02 \pm 0.30 \text{ GeV}$  [40], with a precision of about two per mille.

## 2.4.2 Higgs spin-parity

The SM Higgs boson is a CP-even scalar particle  $J^P = 0^+$ . All experimental studies at ATLAS and CMS indicate the compatibility of the Spin-Parity of the observed Higgs boson with the SM predictions. It is deduced from the lepton production angles ( $\theta$ ) and other discriminant observables chosen to be sensitive to the spin and parity of the signal, at  $\sqrt{s} = 7$  and  $8 \text{ TeV}$  data using  $H \rightarrow ZZ^* \rightarrow 4l$ ,  $H \rightarrow WW^* \rightarrow e\nu_e\mu\nu_\mu$  and  $H \rightarrow \gamma\gamma$  (Figure 2.10). The non-SM spin-parity hypotheses  $J^P = 0^-$  and  $J^P = 2$  with universal and non-universal couplings are excluded at 99.9% CL using a test statistic  $\tilde{q}$  [41].

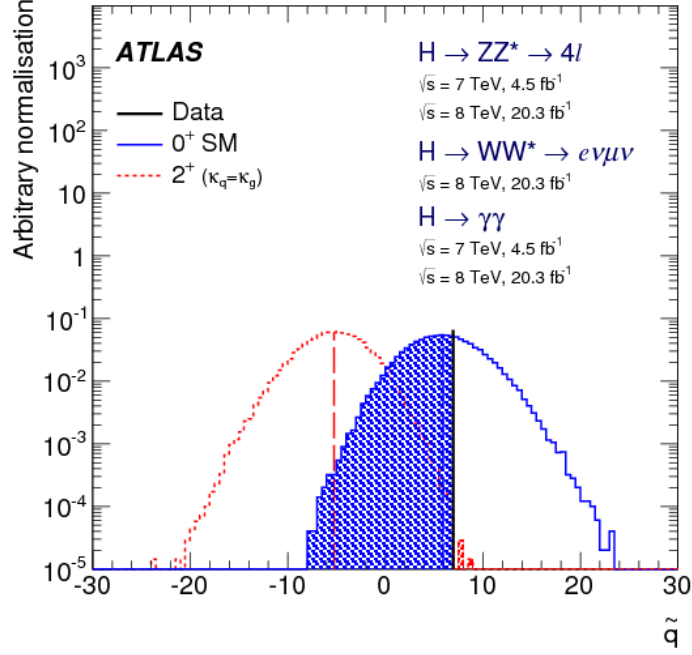


Figure 2.10: *Example of the boson spin-parity measurements of the observed Higgs boson at ATLAS experiment. The  $\tilde{q}$  refers to the test statistic defined in Ref. [41].*

### 2.4.3 Higgs width and lifetime

The SM Higgs width ( $\Gamma_H$ ) is predicted to be  $4.10 \pm 0.06$  MeV for  $M_H = 125.09$  GeV [33]. It is quite a small width, compared to the  $W^\pm$  and  $Z$  bosons (with 2 GeV and 2.5 GeV widths, respectively). Experimentally it is still impossible to measure such width because of detector resolution limits which is about 2 or 3 GeV. One method to constrain the Higgs width relies on the difference between the on-shell and off-shell cross sections of the Higgs boson production in gluon fusion and vector boson fusion processes with the Higgs boson decaying into a  $W^+W^-$  pair and the  $W$  bosons decaying leptonically [42, 43]. This allows both CMS and ATLAS to constraint the Higgs width to be less than 26 MeV at 95% CL [42]. Better observed (expected) constraints on the Higgs boson total width is obtained from the combination with the Higgs boson decaying into a  $ZZ$  pair giving  $\Gamma_H < 13$  (26) MeV at 95% CL [42], as shown in Figure 2.11a.

The SM Higgs lifetime ( $\tau_H = \hbar/\Gamma_H$ ) is  $1.6 \times 10^{-22}$  s. It is experimentally constrained using  $\sqrt{s} = 7$  and 8 TeV data with  $H \rightarrow ZZ^* \rightarrow 4l$  channel to be less than  $1.9 \times 10^{-13}$  s at 95% CL [44] (Figure 2.11b), obtained for the expected standard model Higgs boson width. The distribution of the measured lifetime is used to set a lower limit on the Higgs width:  $\Gamma_H > 3.5 \times 10^{-9}$  MeV [44].

### 2.4.4 Higgs couplings

The tree level Higgs couplings to gauge bosons and fermions (Yukawa couplings) are proportional to squares of the boson masses and proportional to the fermion masses respectively:

$$g_{Hf\bar{f}} = \frac{m_f}{v} \quad (2.19)$$

$$g_{HVV} = 2\frac{m_V^2}{v} \quad (2.20)$$

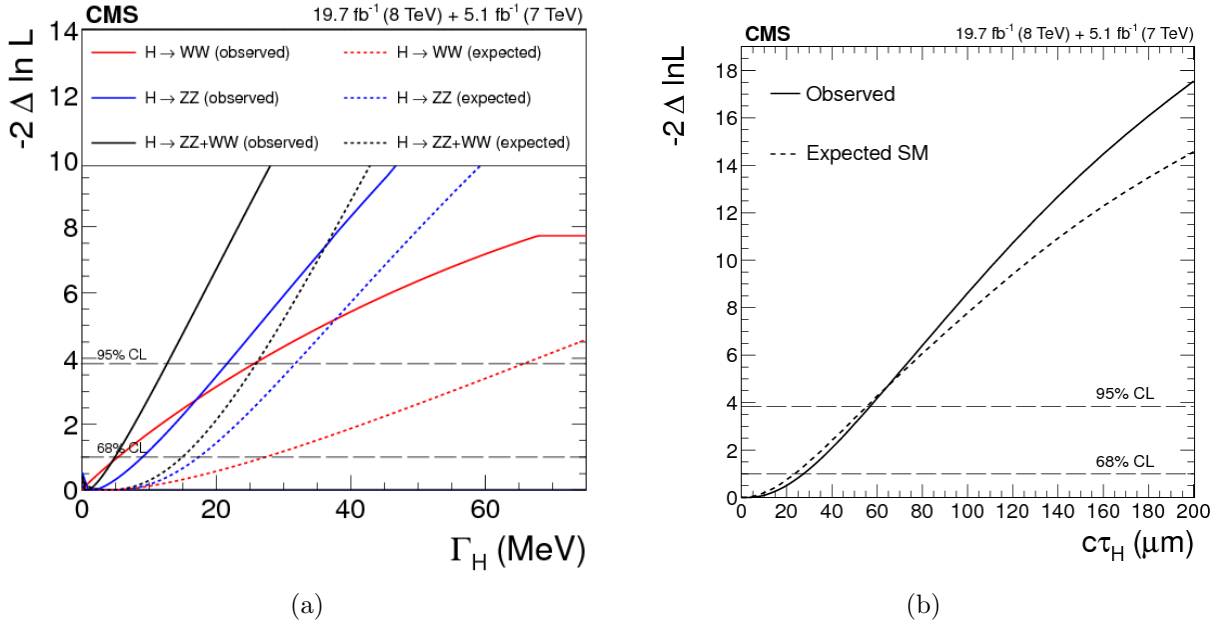


Figure 2.11: Observed (solid) and expected (dashed) distributions of  $-2\ln(L/L_{max})$  as a function of a) the Higgs boson width  $\Gamma_H$  [42] using off-shell vector boson decays and b) the Higgs boson average lifetime  $\tau_H$  [44] at CMS using 25 fb<sup>-1</sup> combined data at  $\sqrt{s} = 7$  and 8 TeV.

where  $v = 246$  GeV is the vacuum expectation value of the Higgs field (see Section 2.1). These couplings govern the Higgs production and decay rates. The Higgs also has self coupling equals to  $3\frac{m_H^2}{v}$  for  $g_{HHH}$  and  $3\frac{m_H^3}{v^2}$  for  $g_{HHHH}$  [10].

The coupling modifier  $\kappa$  is used to search for deviations from SM value Higgs couplings. It is equal to one if the observed particle is a SM Higgs boson. Six free parameters could be fitted according to the type of the particle that the Higgs couples to, namely:  $\kappa_t$ ,  $\kappa_Z$ ,  $\kappa_W$ ,  $\kappa_b$ ,  $\kappa_\tau$  and  $\kappa_\mu$ . Constraints on the observed Higgs boson couplings to vector bosons and fermions are achieved by fitting event yields in 600 exclusive channels corresponding to each 10 categories taking into account all possible configurations of five production processes, mentioned in Section 2.2, and six Higgs decay modes to  $ZZ$ ,  $WW$ ,  $\gamma\gamma$ ,  $\tau\tau$ ,  $bb$  and  $\mu\mu$  using combined ATLAS and CMS data at  $\sqrt{s} = 7$  and 8 TeV [45]. Three modes are observed ( $\kappa_W$ ,  $\kappa_Z$  and  $\kappa_\tau$ ) that behave like the SM Higgs with more than 5 standard deviation ( $\sigma$ ). Though coupling to quarks are not established yet at 5  $\sigma$ . The combined signal yield relative to the Standard Model prediction is measured to be  $1.09 \pm 0.11$ . The data are consistent with the Standard Model predictions for all parameterisations considered.

Figure 2.12 shows the best fit values of the free parameters as a function of the particle mass assuming SM structure of the loops and no BSM decays parameterisation. The parameters are defined as  $\kappa_F \times \frac{m_F}{v}$  for the fermions and as  $\sqrt{\kappa_V} \times \frac{m_V}{v}$  for the weak vector bosons [45]. Those parameters are contributed to get the best fit result to the  $[M, \epsilon]$  phenomenological model [46] with the corresponding 68% and 95% CL bands. The parameterisations  $M$  and  $\epsilon$  of the Higgs couplings to fermions and bosons are designed specifically to probe the dependence of the Higgs couplings on particle masses. This dependency reduces to the couplings of the Standard Model Higgs boson in the double limit  $\epsilon \rightarrow 0$  and  $M \rightarrow v$  [46]. The values of  $[M_{fit}, \epsilon_{fit}]$  from the fit are:  $[233 \pm 13$  GeV,  $0.023 \pm 0.028]$  which are in agreement with the SM values.

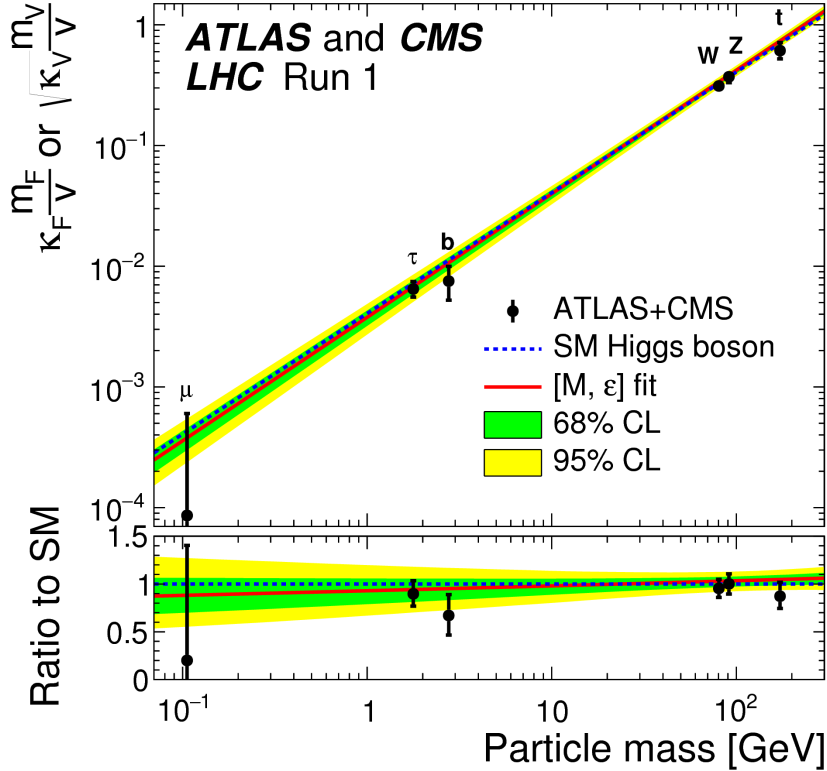


Figure 2.12: Best fit values as a function of particle mass for the combination of ATLAS and CMS data [45]. The dashed (blue) line indicates the predicted dependence on the particle mass for the SM Higgs boson. The solid (red) line indicates the best fit result to the  $[M, \epsilon]$  phenomenological model [46] with the corresponding 68% and 95% CL bands. The bottom panel shows the ratios of the reduced coupling modifiers to the SM predictions with their total uncertainties as a function of the particle mass.

## 2.5 $t\bar{t}H$ searches at LHC

The top yukawa coupling is the strongest coupling to the Higgs boson, given that the top quark is the heaviest fundamental particle in the SM, with a mass equal to 173.21 GeV (see Table 1.1). Its measurement is estimated using the  $ggF$  production where the Higgs is indirectly coupled to the top quark via a loop (see Figure 2.2). Its value is found to be compatible with the SM expectation (see Figure 2.12), where the top coupling modifier is estimated by  $\kappa_t = 0.87 \pm 0.15$ , combining ATLAS and CMS data at Run1, assuming no BSM in the loops (see Section 2.4.4). However, new physics could be hidden in the loops mediating the Higgs production via  $ggF$ . This issue could be solved by performing the measurements in another production mode involving a direct, tree-level, Higgs-top coupling. Therefore, a search for the Higgs boson production in association with a top quark pair ( $t\bar{t}H$ ) is used to allow a first direct measurement of the top quark Yukawa coupling that could reveal new physics. ATLAS [47, 48, 49] and CMS [50] experiments published the first papers on  $t\bar{t}H$  searches using 2012 data (Run1). In both ATLAS and CMS, the  $t\bar{t}H$  analysis is divided into 3 channels, shown in Figure 2.13, according to the Higgs decay modes:  $b\bar{b}$  analysis ( $H \rightarrow$  hadrons), multilepton analysis ( $H \rightarrow$  leptons) and diphoton analysis ( $H \rightarrow$  photons) contributing to 60%, 30% and 0.2% of the total Higgs boson decay width respectively. In " $H \rightarrow$  hadrons", two main Higgs boson decay modes are contributing:  $H \rightarrow b\bar{b}$  and  $H \rightarrow \tau\tau$  where both  $\tau$  leptons decay hadronically.

ATLAS only takes into account  $H \rightarrow b\bar{b}$  focusing on single-lepton +jets (having the highest significance) and dileptonic  $t\bar{t}$  final states (where one or more isolated charged leptons are coming from  $W$  boson decays from the top quarks). In "H  $\rightarrow$  leptons", the Higgs decays to the all possible multilepton final state (two, three or four leptons) mainly via  $W^+W^-$ , pioneered in Ref. [51] and subsequently in Ref. [52], and  $ZZ$  weak bosons or via  $\tau$  leptons. It provides a clean signature and low QCD activities comparing to the previous channel. Finally, in the "H  $\rightarrow$  photons", analysis where the Higgs decays to a pair of photons  $H \rightarrow \gamma\gamma$ , the high  $\gamma\gamma$  invariant mass resolution is used to separate the signal from the background.

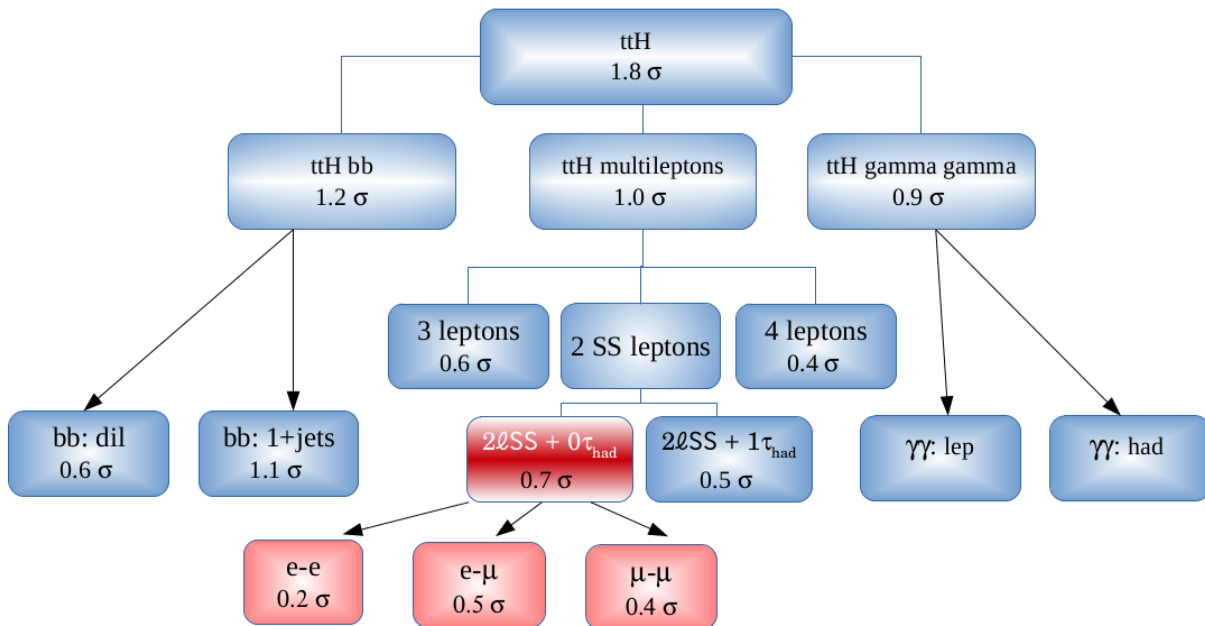


Figure 2.13: Different final states used for the direct measurements of the top-Yukawa coupling in the  $t\bar{t}H$  channel. Numbers correspond to the expected sensitivity (significance) at the beginning of Run2, using  $13.2 \text{ fb}^{-1}$  of luminosity, by the ATLAS experiment. The significance is calculated using a simple approximate formula in order to give an overview of the sensitivity in the different channels. These numbers are 'realistic', meaning that the reducible background are taken from MC and rescaled to the data output fit values (see Chapter 6 for more details). Boxes in red highlights the final state channels of my analysis in Chapter 6.

The small  $t\bar{t}H$  production cross section (1% of the Higgs production) makes its rate measurement experimentally challenging with limited available statistics. In Run1, it is found that the measured  $t\bar{t}H$  cross-section per experiment is slightly above the SM with a signal strength  $\mu_{t\bar{t}H} = 2.3^{+0.7}_{-0.6} (1 \pm 0.5)$  [53] times the SM expectation for  $m_H = 125 \text{ GeV}$  at  $\sqrt{s} = 7$  and 8 TeV combining ATLAS and CMS measurements. The observed (expected) significance of the signal is 4.4 (2) standard deviations [53]. However, no significant excess of events above the background expectation is found because of the high amount of uncertainties where the precision on the  $t\bar{t}H$  signal strength is completely dominated by systematics ( $b$ -tagging, jet energy scale and data-driven fake estimation). Efforts are still ongoing during Run2 to observe this process, given the high statistics and higher center of mass energy. The increase of the cross section and luminosity are expected to

improve the precision of the  $t\bar{t}H$  measurements overall. For instance, the uncertainties could be reduced from 155% to 60% [54] for  $2\ell ss$  channel at ATLAS. The actual analysis and measurements using 2015-2016 data, where I contribute, is presented in Chapter 6.



# Chapter 3

## ATLAS at LHC

The raised questions from SM discussed in Section 1.4 and the predicted mechanism of the generation of the particle masses (Higgs mechanism) suggested in Chapter 2 necessitate a powerful particle accelerator and detectors with high technology. The current world-wide energy frontier machine is called the Large Hadron Collider (LHC) built by the European Organization for Nuclear Research (CERN). The data collected by ATLAS and CMS at the LHC allow to probe the predictions of the SM at a prodigious energy regime at center of mass energy of 13 TeV.

In this chapter, an overview of the CERN complex accelerators ramping up the energy to the LHC is described in Section 3.1. The ATLAS detector is described in Section 3.2.

### 3.1 The CERN accelerator complex

The accelerator complex at CERN [55, 56] is an injector chain of hadron, proton or lead ion, linear and circular accelerator machines designed and arranged in order to increase gradually the particle beam energies on the one hand and squeeze<sup>1</sup> the beam to get high intensity proton bunches and low emittance  $\epsilon^2$  on the other hand.

The basic idea of each accelerator is to apply a potential difference  $V$  allowing a particle with charge  $q$  to pass through an electric field to gain energy  $E$  ( $\Delta E = qV$ ). The used technologies are mainly radio-frequency (RF) cavities with different powerful magnets (dipoles, quadrupoles, kickers). Protons are chosen to probe very wide energy spectrum. Unlike an electron-positron collider that provides a single collision energy.

Figure 3.1 shows a general layout of the proton acceleration through the accelerator complex at CERN. First at the linear accelerator<sup>3</sup> LINAC 2, bunch of protons are obtained by stripping orbiting electrons from hydrogen atoms. They are accelerated to one third of the speed of light  $c$  (an energy of 50 MeV) using RF source with small quadrupole magnets to control the tightness of the proton beam. Protons are then injected from LINAC 2 into a series of ring accelerators starting from the PS Booster (PSB) where the beam is divided into four packets, to maximise its intensity, and be accelerated to 91.6 % of the speed of light with an energy of 1.4 GeV. A magnetic field is applied to bend the beam of protons around the circle and squeeze them together. The four packets are then gathered together again and sent to the Proton Synchrotron (PS) where the proton packets

---

<sup>1</sup> The squeeze reduces the beam size at the interaction point thereby increasing the collision rate.

<sup>2</sup> A low emittance particles are confined to a small distance and have nearly the same momentum.

<sup>3</sup> A linear accelerator is a number of conducting drift tubes arranged in a line where some of them obey to a RF voltage source and others are fixed to the ground. The RF allows the particle to go from one tube to another. The frequency of the voltage is set according to the time needed for a particle in one tube to arrive to the gap in order to change the direction and allow it to achieve the next tube.



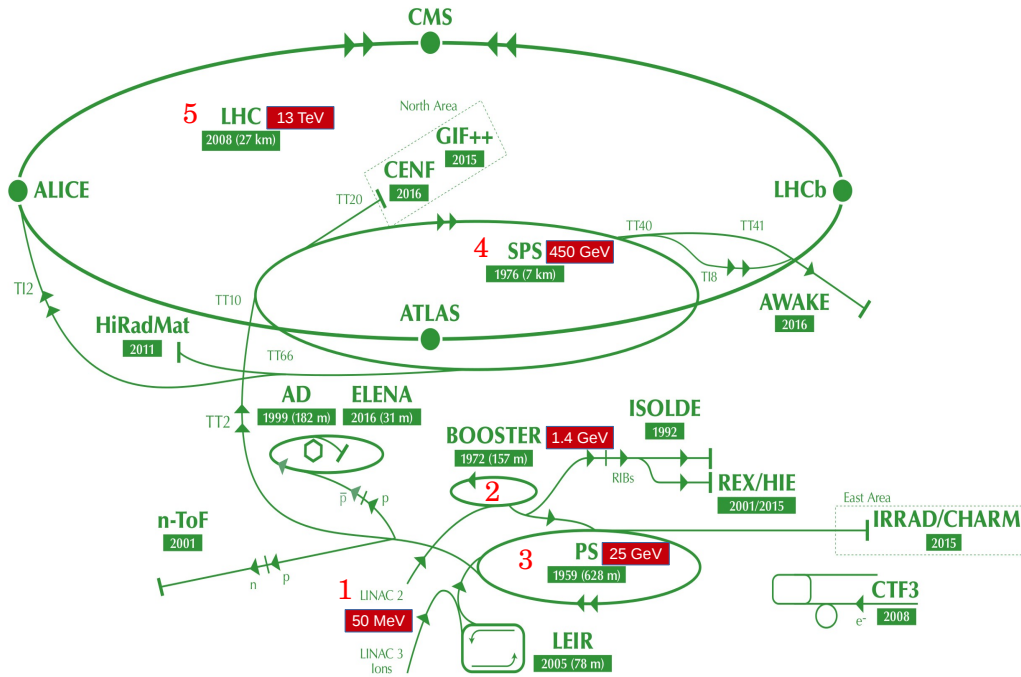


Figure 3.1: *CERN accelerator complex* [55].

achieve 99.9 % of  $c$  with 25 GeV of energy; followed by the Super Proton Synchrotron (SPS) where the beam is accelerated to 450 GeV. Finally, the beam is ready to be transferred to the Large Hadron Collider (LHC), the last piece of this chain, and split out into two opposite direction beams where they are accelerated for 20 minutes to 6.5 TeV. The center of mass energy of the collision is then equal to 13 TeV. Four different detectors are installed there, namely two general-purpose detectors, ATLAS (Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) sharing the same scientific goal (measurements of the Higgs boson properties and searches for new physics) but with some different technical solutions and design; ALICE and LHCb detectors for quark-gluon plasma, produced through heavy-ion collisions, and heavy flavour physics searches respectively. Other detectors are also installed for multi-purposes such as TOTEM, LHCf and MoEDAL. More details about the LHC machine follow.

The Large Hadron Collider LHC [56] is a two ring proton proton ( $pp$ ) collider, favored over the alternative single ring  $\bar{p}p$  collider to reach higher luminosities of the order of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . To achieve that, the central part of the LHC is designed to be the coldest place in the galaxy thanks to the cryogenic systems, made of liquid Helium, providing a temperature less than 1.9 K (-271 °C). It has the largest number of high-technology magnets ever built and the largest most complex electronic instruments. The LHC consists of two main systems: The beam acceleration system made of RF cavities and the circulation, beam orbits and dimensions are controlled by around 9000 magnets: 1232 dipole superconducting magnets defining the orbit, 392 quadrupoles acting on the beam size and many higher order multipoles for further corrections. In each dipole the current

circulates in cables of 36 twisted stands, each stand made out of 6000-9000 superconducting Nb-Ti filaments 7 micro-meters thick. Each dipole provides a magnetic field of 8.3T.

Beams at the LHC are made of “trains” of proton "bunches" moving at almost the speed of light circulating the 27 kilometer ring of the Large Hadron Collider more than 11 000 times a second. Every bunch includes  $10^{11}$  protons. Each bunch crossing could occur at 25 ns intervals with crossing frequency of 40 MHz. This allows to increase the integrated luminosity, inversely proportional to the emittance. The integrated luminosity ( $\mathcal{L} = \int L.dt$ ) is the potential number of collisions that can occur in a given amount of time. It links the number of events to the cross section of a process by:  $\langle N_{evt} \rangle = \mathcal{L} \cdot \sigma_{evt}$ . Its unit is the inverse femtobarn ( $\text{fb}^{-1}$ ).  $1 \text{ fb}^{-1}$  corresponds to around  $8 \times 10^{13}$  collisions.  $L$  is the instantaneous luminosity given by:  $L = \frac{N^2 f}{4\epsilon\beta^*}$  [57] where  $N$  is the number of particles per bunch,  $f$  is the bunches crossing frequency and  $\beta^*$  is the amplitude function at the interaction point IP.

The LHC accelerates two beams of hadrons head-on using RF cavities with alternative electric field; bends them to be adapted to the circular path using 1232 dipole magnets ( $B = 8.33 \text{ T}$ ); focusing and squeeze the beam to the interaction point using quadru/triplet pole magnets. By sending more bunches around the ring, the LHC is able to generate more collisions supplying more physics data for the experiments. This is done gradually from 368 to 1380 bunches per beam during Run1 (2010-2012) at  $\sqrt{s} = 7$  and 8 TeV. At the beginning of Run2 (with  $\sqrt{s} = 13 \text{ TeV}$ ), 2244 protons bunches per beam were collided by the end of 2015. Since 2016, the number of bunches is increased to the target of 2748 [58] providing up to 1 billion inelastic  $pp$  events per second ( $\sigma \sim 70\text{-}80 \text{ mb}$ ).

## 3.2 The ATLAS detector

The ATLAS [59, 60] detector is one of the most important and huge detectors in the world today, situated 100 meters below a small Swiss village at CERN Point-1. It is 44 meters long, 25 meters in diameter, and weighs about 7,000 tonnes. The purpose of ATLAS is to identify the secondary particles produced in collisions, and to measure their positions in space, their charges, speed, mass and energy in order to reconstruct the short life-time undetectable particles decay and probe the recorded rare events. To do so, the detector has many layers or ‘sub-detectors’ with excellent hermiticity. It is therefore possible to reconstruct the final state particles containing leptons (electrons and muons), mesons (pions and keons), baryons (protons and neutrons) and gauge bosons (photons). More details about how those particles are reconstructed in ATLAS is given in Chapter 4.

A general layout of ATLAS detector is shown in Figure 3.2. The beams direction coming from LHC and passing through ATLAS center is hold by the  $z$ -axis where the  $+z$  direction corresponds to the counterclockwise rotation when the LHC is seen from the sky, whereas, the  $+x$  direction points at the center of the LHC ring and the  $+y$  direction points at the sky vertically. In the following, the coordinate system used to describe the ATLAS detector is the cylindrical coordinate system  $(r, \phi, \theta)$  where  $\theta$  is replaced by the invariant variable "the pseudo-rapidity"  $\eta = -\ln(\tan(\theta/2))$ .

The different multi-components of ATLAS detector are presented in Section 3.2.1. The data quality monitoring recorded by ATLAS is described in Section 3.2.2. Finally, the luminosity measurement used in this document is summarised in Section 3.2.3.

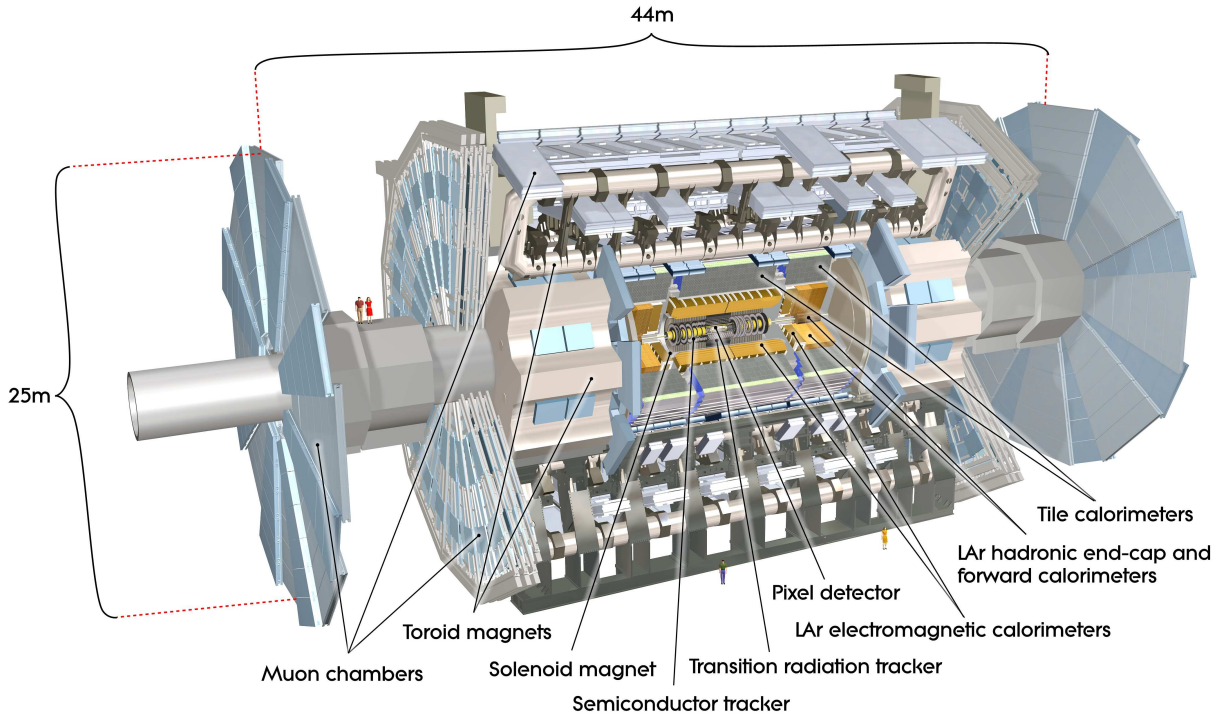


Figure 3.2: *ATLAS multi-component detector.*

### 3.2.1 ATLAS layout

The ATLAS detector consists of four main systems: Inner detector (ID) including vertex detector and tracking chamber, calorimetry system to detect the deposit energy from electrons/photons and hadrons, muon spectrometer to identify muons and a magnet system to bend the charged particles. The super-conducting solenoid covers the inner detector and the toroid magnet is set on top of all those detectors. Every system is divided into concentric cylindrical barrels, parallel to the beam direction, and end-caps (disks perpendicular to the beam direction) which offer full space coverage of the ATLAS detector.

#### 3.2.1.1 The inner detector

The inner detector consists of three subsystems in order to get precise reconstruction of charged particle tracks, interaction points and the decay position of the short-lived particles for  $|\eta| < 2.5$  on the  $y$ - $z$  plane and  $-\pi < \phi \leq +\pi$  on the  $r$ - $\phi$  plane. Figure 3.3 shows each of the major sub-detector elements with its active dimensions and envelopes. The three subsystem are: the pixel detector, the Semi Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The pixel and SCT detectors are usually gathered into one subsystem called "Silicon tracker", because both are made of silicon semi-conductor sensors. The description of each sub-system is detailed in the following.

**Silicon tracker** The silicon detector [61], consisting of a n-type substrate with p-type strips, is able to accurately measure very small distances to track the vertices of unstable particles having short flight lengths of  $100 \mu\text{m}$  or less. The semi-conductors consisting of silicon wafers with closely spaced strips connected to an amplifier and electronic read-out circuit. A voltage is applied to the  $pn$ -junction to deplete the silicon of charge carriers and establish an electric field inside the wafer. An ionising particle produces electron-hole pairs causing a current flow between the substrate and nearby strips giving signals

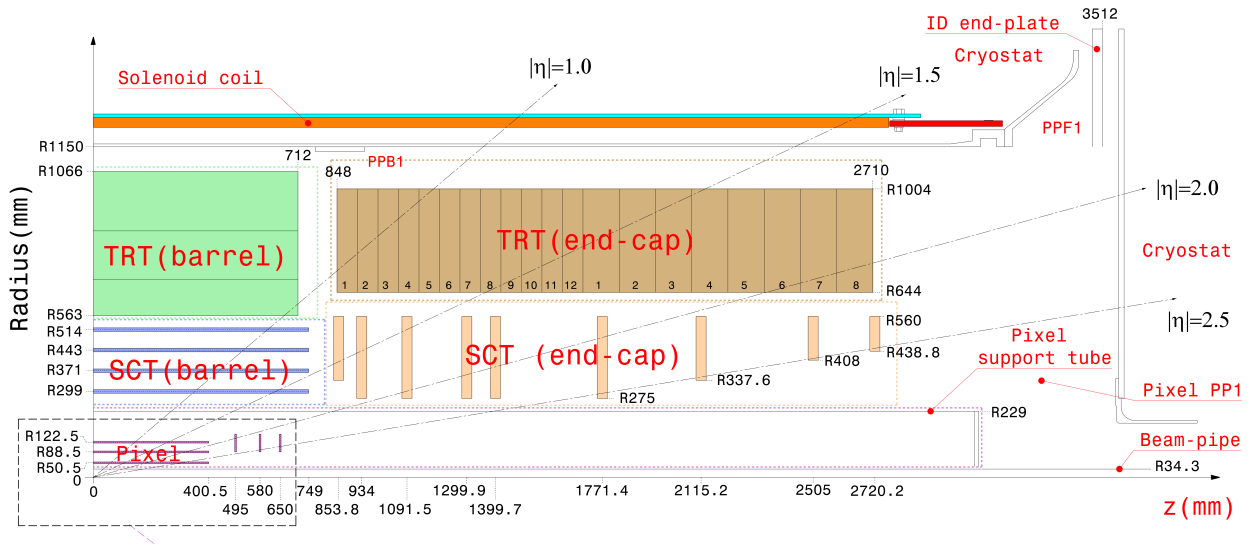


Figure 3.3: Composition and dimensions of the ATLAS inner detector in the  $r$ - $z$  plane [59]. The IBL is not shown in this scheme.

achieving a position resolution of around  $5\text{-}10\ \mu\text{m}$ .

The silicon trackers, namely the pixel detector and the SCT, work in similar way. The pixel and SCT sensors are kept at low temperature between  $-5$  to  $-10\ ^\circ\text{C}$  in order to preserve an adequate noise performance with radiation damage [62] (Type Inversion). Compensating feedback loops are installed in the detector to correct the leakage current due to the thermal free charge carriers and electron-hole pair formation due to infrared and X-ray absorption.

**Pixel detector [63]** The Pixel detector is the closest sub-detector from the interaction point since it has very high resolution needed to define the track parameters and the vertices. It consist of 92 million pixel sensors [64] included in 2024 electronic  $n$ -in- $n$  modules, arranged in four layers barrel with three disks in each end-cap side. Starting from the most inner one, the Insertable B-layer "IBL" [65] (recently included during the 2014-2015 upgrade) is installed at  $r = 33.2\ \text{mm}$  with no end-cap part since it is directly attached to the beam pipe. It is followed by the B-Layer, Layer 1 and finally Layer 2 at  $r = 50.5, 88.5,$  and  $122.5\ \text{mm}$  respectively as shown in Figure 3.4. Except IBL, every module contains 46080 pixel sensors (Every sensor is  $250\ \mu\text{m}$  thick and on the order of  $(r-\phi \times z) = (50 \times 400)\ \mu\text{m}$  in area with an expected hit resolution of  $10\ \mu\text{m}$  in  $r-\phi$  and  $115\ \mu\text{m}$  in  $z$ ) connected to 16 front-end chips (read out). Every front-end chip has 2880 channels to amplify the signals. The IBL pixel sensors are smaller in  $z$  ( $250\ \mu\text{m}$ ) and have higher intrinsic resolution on the order of  $8 \times 40\ \mu\text{m}$ .

**Semi-conductor tracker (SCT)** The SCT has silicon sensors similarly as pixel detector but with less resolution ( $17\ \mu\text{m}$  in  $r-\phi$  and  $580$  in  $z$ ) p-in-n type semiconductor. Every silicon sensor has larger size of  $6.4 \times 2\ \text{cm}$  in  $r-\phi$  (bilayer) and  $80\ \mu\text{m}$  in  $z$  with  $285\ \mu\text{m}$  thick. The barrel and end-cap parts are composed of four layers and nine disks for each side. The total number of readout channels are 6.3 million built into 2112 modules for the four barrel layers and 1976 modules for the 18 end-cap disks.

**Transition radiation tracker (TRT)** The Transition Radiation Tracker (TRT) is a drift tube gas chambers made of mixture of different gases: Xenon (Xe) to absorb the X-rays



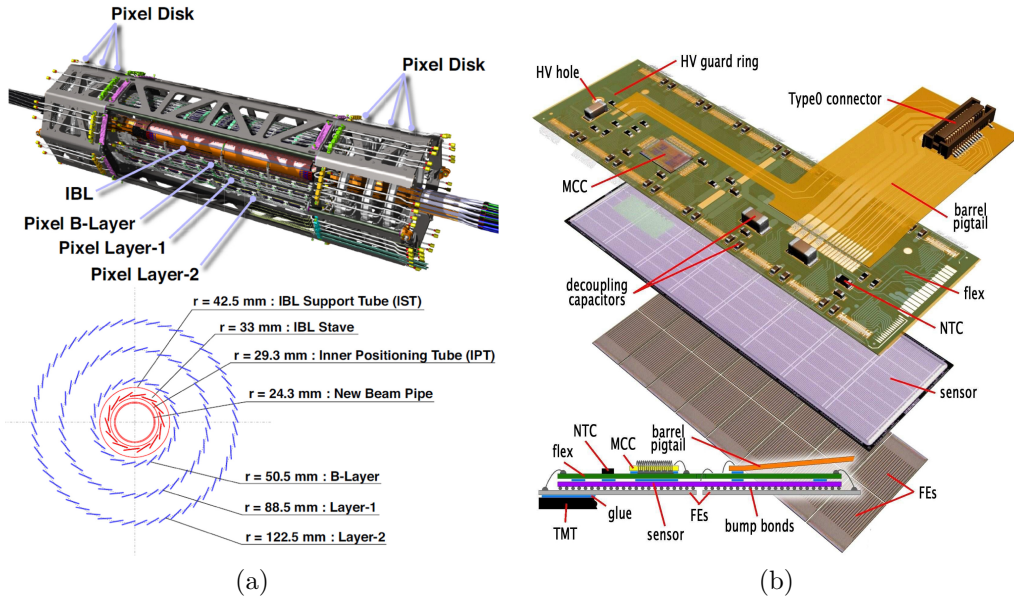


Figure 3.4: (a) Pixel detector inside the Inner Detector of ATLAS detector [66]. The beam pipe is replaced with one with a smaller radius in order to save the space for the IBL installed in 2014 for the LHC Run 2. (b) Pixel module structure of ATLAS tracker and principle of operation [67].

as the transition radiation,  $CO_2$  to stabilise the gas amplification and  $O_2$ . In addition to Xe, the Argon (Ar) gas is recently inserted because of the leakage seen with the increase of pile-up events<sup>4</sup> at Run2. The TRT also consists of a barrel ( $|\eta| < 1$ ) and end-caps ( $1 < |\eta| < 2$ ). Each barrel (end-cap) part includes 52 544 (122 880) tubes arranged in 36 (22) straws separated from each other by the transition radiation material made of polypropylene fibers with 15  $\mu\text{m}$  diameter. Despite the position resolution of TRT is less than the one of silicon tracker (130  $\mu\text{m}$  resolution per tube), the charged particle in the TRT leaves a mostly continuous track (36 hits in total) giving the advantage to improve the momentum resolution and to get more information for the identification of electrons. In particular, when the Xenon absorbs the transition radiation thanks to the polypropylene fibers, more signal is produced by electrons than by pions since the transition radiation is proportional to Lorentz factor  $\gamma = \frac{E}{m}$ .

### 3.2.1.2 Solenoid

The solenoid is a magnet installed in front of the calorimeter around the Inner Detector in order to curve the trajectory of the particle according to its charge; hence identify the charge of particles and reconstruct their momenta. It provides a 2 T axial magnetic field with a 7.73 kA nominal current. It is made of conductor with Al-stabilised NbTi material since it has low radiation length  $X_0$  (Figure 3.9) to avoid affecting the EM calorimeter performance, described in the following section.

### 3.2.1.3 Calorimetry

The calorimetry system in ATLAS is a set of devices installed beyond the Inner Detector and the solenoid magnet to measure the energy deposit as well as the direction of different types of the incident particles and distinguish between them, namely electrons,

<sup>4</sup>Pile-up events are the several separated events, associated to the event of interest, produced per bunch crossing at high-luminosity colliders.

photons, and jets. It provides an accurate measurement of the energy thanks to its fine granularity, full azimuthal angle coverage ( $|\phi| < \pi$ ) and large acceptance ( $|\eta| < 4.9$ ). The ATLAS calorimeters are all sampling calorimeters, with passive material (absorber) to produce the particle shower and stop the particle, and active material to measure the energy loss. The current, created by the displacement of the charges, is collected by electrodes inserted in the active material.

There are two kind of calorimeters, according to the type of the particle: The EM calorimeter to stop electrons and photons and measure their energy. Followed by the hadronic calorimeter to stop completely the jets and measure their deposit energy. Each system consists of a barrel, two end-caps and a forward calorimeter placed near the inner detector, to cover the forward region, as shown in Figure 3.6. This topology gives an excellent hermeticity by covering  $|\eta| < 4.9$  range except the crack regions (transition regions) situated between the barrels and end-caps cryostats. A small gap of 4 mm at  $z = 0$  separates the two EM and hadronic barrels. Liquid argon (LAr) is used as an active medium for all calorimeters except hadronic barrel calorimeter made of scintillation tiles.

Two parameters are taken into account for the calorimeter design in order to achieve precise measurements of the energy  $E$  and the transverse missing energy  $E_T^{miss}$  of the incident particle: First, the fine granularity in  $\eta$ - $\phi$  plane (Figure 3.7), in the region matched to the inner detector, gives precise localisation of the incident particle and allows to distinguish between particles (for example photon and pion). Second, the depth in  $r$ - $\phi$  plane (layers) are needed in order to stop the incident particle and cover the deep shower shapes produced by the electron and photon candidates during their interaction with the material. This allows precise particle reconstruction and identification described in Chapter 4.

The fractional calorimeter energy resolution expression is:

$$\frac{\delta E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \quad (3.1)$$

where:  $a$  and  $c$  are the sampling and noise terms respectively.  $b$  is a constant term takes into account detector uniformities and errors in the calibration. The expected values of those parameters are included in Table 3.1 for both EM and Hadronic calorimeters. A comparison between ATLAS and CMS calorimeters performances is shown in Figure 3.5 where the ATLAS hadronic resolution is three times better due to the limited hermeticity of the CMS hadronic calorimeter (no enough space) and an insufficient absorption length resulted by the strong constraints imposed by the CMS solenoid [68]. However, since the homogeneous calorimeter has better resolution than the sampling one, the CMS EM calorimeter provides smaller values of the sampling parameter " $a = 2.8\%$ " and of the constant term " $c=0.3\%$ " [69].

Detector component	Required resolution
EM calorimeter	$10\%/\sqrt{E} \oplus 0.7\%$
Hadronic calorimeter barrel and end-cap	$50\%/\sqrt{E} \oplus 3\%$
forward	$100\%/\sqrt{E} \oplus 10\%$

Table 3.1: *Expected typical energy resolution (in GeV) of different subsystems in ATLAS detector [59].*

In the following  $\Delta\eta(\Delta\phi) = 0.025$ , equivalent to 37.52 mm, is used as a unit to measure the granularity size of each layer.

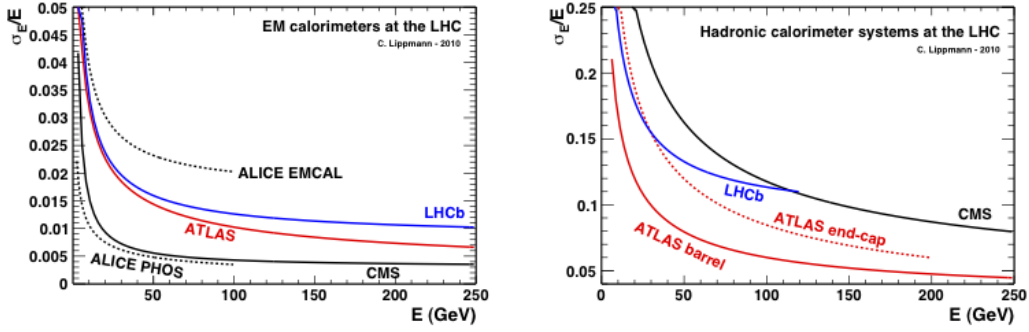


Figure 3.5: Comparison of the relative energy resolutions of the different EM calorimeters (left) and hadronic calorimeters (right) at the LHC experiments [69]. The values of the parameters  $a$ ,  $b$  and  $c$  were in all cases determined by fits to the data from beam tests.

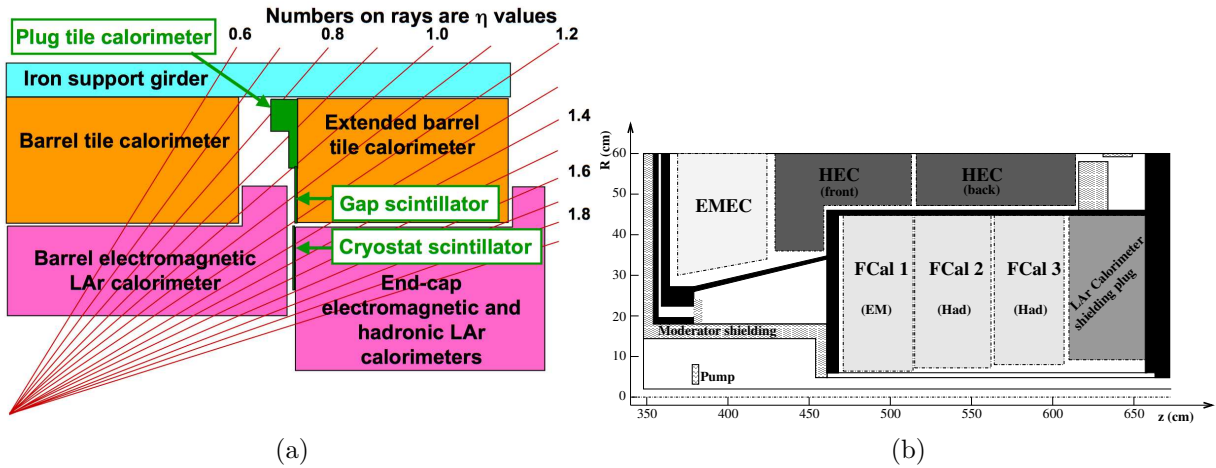


Figure 3.6: Schematic of the calorimetry system. The barrel and end-caps are shown in (a). The transition region between the barrel and end-cap liquid Argon cryostats are shown. The gap and cryostat scintillators are read out together with the other tile-calorimeter channels. the three FCal modules located in the end-cap cryostat are shown in (b). The material in front of the FCal and the shielding plug behind it are also shown. The black regions are structural parts of the cryostat. The diagram has a larger vertical scale for clarity.

**EM calorimeter** The EM calorimeter is a lead-LAr detector, designed to measure the energy of the electrons and photons. Its material is chosen to be liquid argon for the active material, lead for the absorbers and kapton electrodes. The lead causes the incident photon/electron to initiate a shower of electrons, positrons and photons through pair production and bremsstrahlung processes. The energy loss in the shower allows one to estimate the energy of the incident particle. The other heavier particles, such as muons, are weakly interacting with the EM calorimeter material (since the energy loss due to bremsstrahlung of charged particles depends on its energy, mass and the radiation length  $X_0$ <sup>5</sup>).

The barrel EM calorimeter (EMB), covering  $|\eta| < 1.475$ , is divided into two identical half-barrel, spaced by 4 mm gap at  $z = 0$ . The two end-cap components (EMEC) covering the pseudorapidity range  $1.375 < |\eta| < 3.2$ , are also divided into two coaxial wheels at

<sup>5</sup>The radiation length  $X_0$  depends on the atomic number of the crossed material.

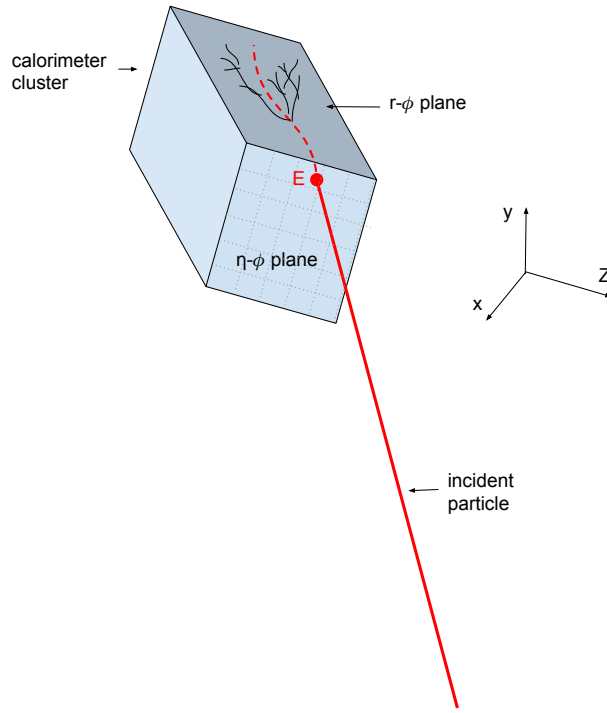


Figure 3.7: Schematic of a calorimetry cluster that detects and stops the incident particles (photon, electron and jet) coming from the interaction point. The fine granularity of the cluster in the  $\eta$ - $\phi$  plane provides precise localization and measurement of the energy  $E$  of the particle. The depth of the calorimeter, consisting of three layers (not shown), in the  $r$ - $\phi$  plane covers the deep shower shape produced by the particle.

$$|\eta| = 2.5.$$

The EM calorimeter has an accordion shape in order to ensure a complete azimuthal coverage. It consists of three layers in depth (Figure 3.8): Strip, middle and back layers. Every layer has a role in the object reconstruction (see Chapter 4) and hence different geometry. The **strip** layer has the finest granularity ( $\Delta\eta \times \Delta\phi$ ):  $\Delta\phi = 0.1$  everywhere except for  $1.40 < |\eta| < 1.475$  in the barrel with  $\Delta\phi = 0.025$ . However,  $\Delta\eta$  varies from 0.025/8 to 0.1 depending on the region of interest. The **middle** layer has coarser layers in  $\eta$  and finer in  $\phi$ . It consists of cells with the size of 0.025 in  $\phi$  except  $2.5 < |\eta| < 3.2$  in the end-cap region ( $\Delta\phi = 0.1$ ) and from 0.025 to 0.1 in  $\eta$ . The **back** layer is the coarser layer with  $(\Delta\eta \times \Delta\phi) = (0.050 \times 0.025)$  in both the barrel and the end-caps. It only covers the regions  $|\eta| < 1.35$  and  $1.5 < |\eta| < 2.5$ . A presampler LAr detector, covering  $|\eta| < 1.8$ , is installed in front to correct for energy loss in the material before reaching the calorimeter. Figure 3.9 presents the amount of material, in units of radiation length  $X_0$ , traversed by a particle as a function of  $|\eta|$  in front of the presampler detector and the EM accordion calorimeter (right), and up to the Inner Detector boundaries (left) including the services. In fact, the material distribution of the Inner Detector has slightly increased, compared to Run1, with the inclusion of the IBL and its services. The Inner Detector and the Solenoid are made with minimal substantial quantity of material in order to reduce the multiple scattering and generation of secondary particles before the calorimeter. The solenoid material contributes by  $\sim 0.66$  radiation length at nominal incidence (not shown in Figure 3.9a). The amount of material traversed by a particle in front of the presampler detector (dark blue of Figure 3.9b) is due to the Inner Detector services plugged in the space between Inner Detector, the presampler and the cryostat.



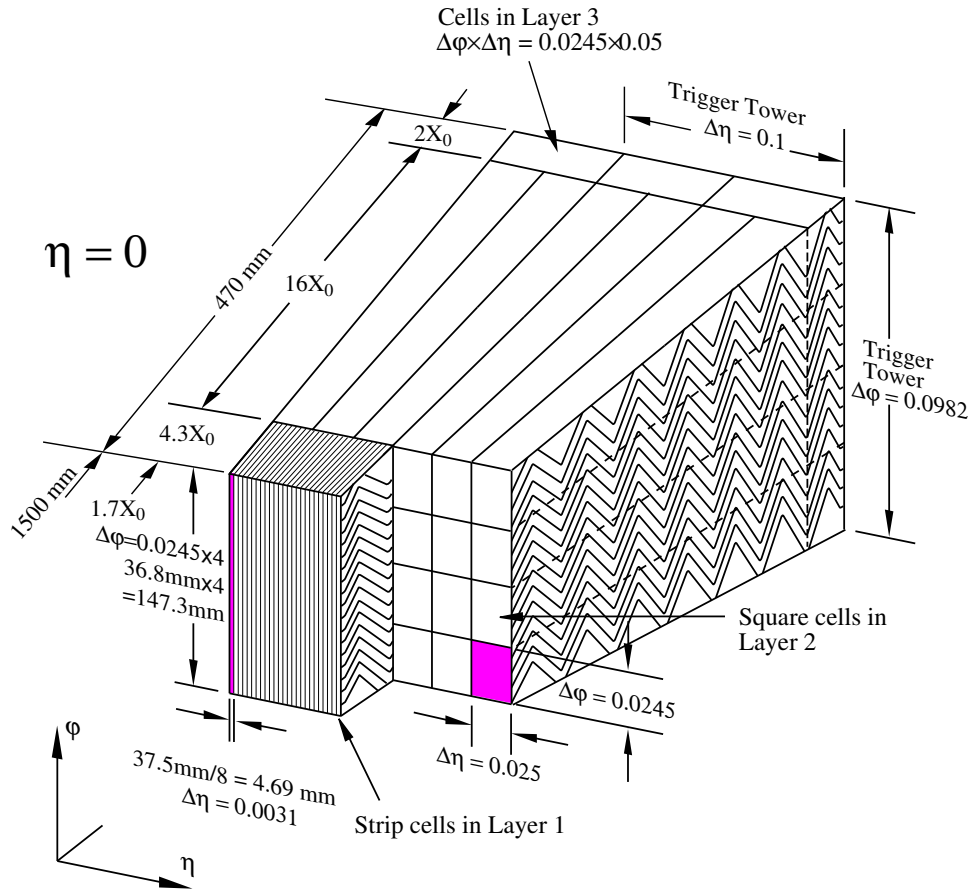


Figure 3.8: 3D Sketch of a EM calorimeter barrel module at  $|\eta| < 1.35$  where the different layers are clearly visible with the ganging of electrodes in  $\phi$ . The granularity in  $\eta$  and  $\phi$  of the cells of each of the three layers and of the trigger towers is also shown.

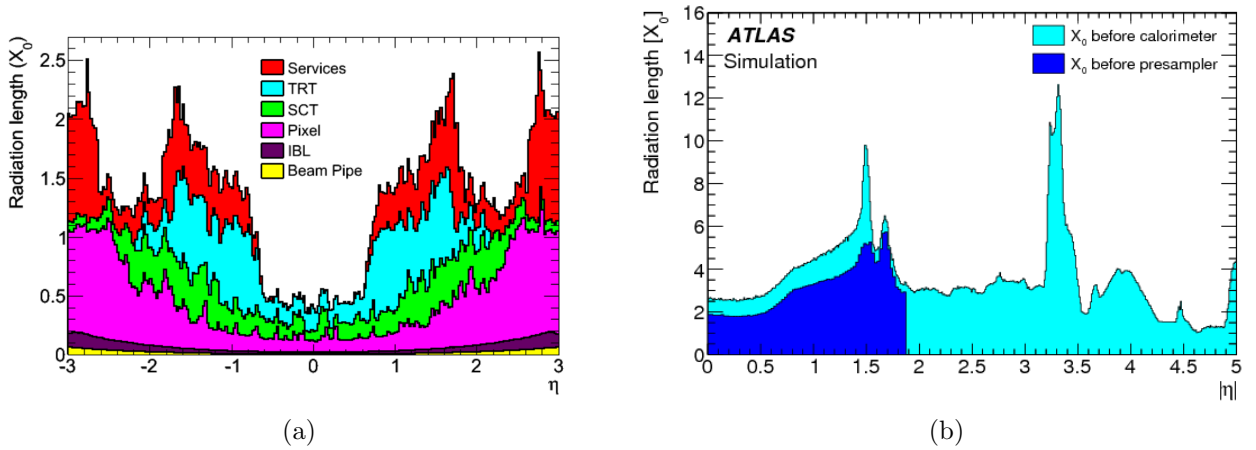


Figure 3.9: Amount of material in ATLAS detector up to the Inner Detector boundaries [65] (left) and in front of the presampler detector and in front of the EM accordion calorimeter [70] (right).

**Hadronic calorimeter** The hadronic calorimeter, located directly after the EM calorimeter, is designed to stop hadrons and estimate their energies. The incident hadron interacts strongly with the nuclei of the detector's material leading to a "hadronic shower" consisting mainly of pions and nuclear fragments. Due to the large statistical fluctuations

corresponding to the hadronic collisions, the energy resolution of hadronic calorimeter is 5 to 10 times worse than EM calorimeter resolution. The hadronic calorimeter consists of tile calorimeters (the barrel) covering the pseudo-rapidity  $|\eta| < 1.7$ , LAr hadronic end-cap calorimeters (HEC) at  $1.5 < |\eta| < 3.2$  and LAr Forward Calorimeters (FCal) installed in the forward region at  $3.1 < |\eta| < 4.9$ .

The barrel consists of two **tile calorimeters** at  $|\eta| < 1.0$  and  $0.8 < |\eta| < 1.7$  (extended barrel). Each tile calorimeter has three layers in depth with  $(\Delta\eta \times \Delta\phi) = (0.1 \times 0.1)$  except the last layer with  $(\Delta\eta \times \Delta\phi) = (0.2 \times 0.1)$ . It is made of steel as the absorber and scintillating tiles as the active material. The emitted scintillation light is collected, read out by wavelength shifting fibers into two separate photomultiplier tubes (PMT). A huge number of electrons are generated in the PMT through the photoelectric effect and collected at the anode resulting in measurable electrical signal.

The **HEC** consists of two independent wheels per end-cap at  $1.5 < |\eta| < 2.5$  and  $2.5 < |\eta| < 3.2$  respectively, made of copper plates absorber and LAr active material. It has two layers per end-cap with  $(\Delta\eta \times \Delta\phi) = (0.1 \times 0.1)$  everywhere except the region  $2.5 < |\eta| < 3.2$  with  $(\Delta\eta \times \Delta\phi) = (0.2 \times 0.2)$ .

The **FCal** consists of three components included in the same cryostat as the end-caps where the LAr is used as an active medium. The first one is made of copper absorbers and the other two with tungsten. Its geometry allows for excellent control of the gaps. Every component has three layers in depth with different coarse granularity.

### 3.2.1.4 Muon spectrometer

A spectrometer is installed over the calorimeters to identify muons and precisely measure their momentum. It consists of two sub-systems: a toroid magnet to bend the muon track and chambers for high precision tracking and triggering. Four types of muon chambers are used: the monitored drift tubes (MDT), the resistive-plate chambers (RPC), the cathode strip chambers (CSC) and the thin-gap chambers (TGC); all are shown in Figure 3.10.

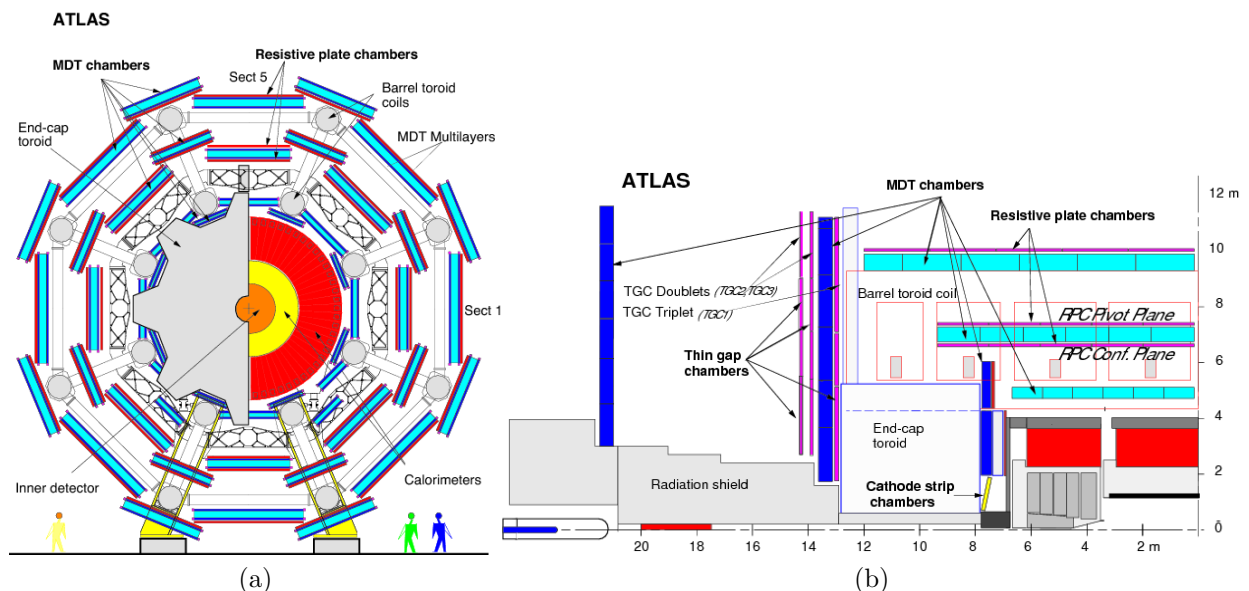


Figure 3.10: Schematic view of the muon spectrometer [71] in the (a)  $x$ - $y$  and (b)  $z$ - $y$  projections.

**The toroidal magnet** consists of a barrel (1.5-5.5 Tm) in the region  $|\eta| < 1.4$  and two end-caps (1-7.5 Tm) installed at  $1.6 < |\eta| < 2.7$ , together covering the range  $\eta < 2.7$  where  $1.4 < \eta < 1.6$  is called the transition region. The barrel toroid coils are inserted in eight individual cryostats. However, the end-cap toroid coils are housed in eight aluminum alloy where each coil is merged in one large cryostat.

**Muon chambers** are divided into barrel layers and end-cap planes, each arranged in three layers. Both MDT and CSC are used for precise measurements of the track coordinates in the central ( $|\eta| < 2$ ) and forward regions ( $2 < |\eta| < 2.7$ ) respectively. The MDT is made of two drift tubes filled by 97% of Ar and 3% of CO<sub>2</sub>. For CSC, used for getting higher counting rate, multi-wire proportional chambers (MWPC) are used which detect the passage of a particle by collecting the ionisation produced in a gas with a wire electrode and converting it into an electrical pulse. The trigger system is composed of a RPC (barrel covering  $|\eta| < 1.05$ ) and TGC (end-caps at  $1.05 < |\eta| < 2.7$ ) based on the principle of gas and MWPC chambers respectively. The RPC is divided into three components: RPC-1 and -2 used for low  $p_T$  (6-9 GeV) and RPC-3 for high  $p_T$  threshold (9-35 GeV) trigger. The TGC provides a 2D position readout with very fast response for triggering. The trigger system covers  $|\eta| < 2.4$  range and provides bunch-crossing identification, well-defined  $p_T$  thresholds and muon coordinates orthogonal to those measured by the tracking system (MDT and CSC).

### 3.2.1.5 Trigger system [1]

The data produced by the big experiments at the LHC are huge where  $10^9$  events are produced per second. The space on disk needed for an event is typically 1 MB. It is therefore not feasible to store all the data. As a consequence, events are selected by a trigger system to reduce the recording rate from 40 MHz to around 0.5-1 kHz by searching for the appropriate event to be stored. The trigger system consists of two main levels (Figure 3.11): A hardware Level-1 (L1) includes soft requirements ( $E_T$  or  $p_T$  thresholds and object multiplicity) taking into account the short time to make the trigger decision online (only 2.5  $\mu$ s). It reduces the event rate to 100 kHz based on custom-made hardware which selects the events using information from the coarse-granularity calorimeter and muon information. The L1 selects the Region Of Interest (ROI) for the next level: The "HLT"; A software-based high level trigger with average processing time less than 5s. In the HLT, the event is recommended to pass more strict requirements, including tracks information from the ID, taking into account all parts of the detector and depending on the physical object of interest (electrons, muons, photons, jets, hadronic tau-decays, and  $E_T^{miss}$ ). Once the event is selected by HLT, it is stored permanently at CERN Computing Center (called Tier-0). Then, it is cleaned (see Section 3.2.2) and treated by the full offline reconstruction software. Finally, the processed data are spread on the grid computing centers throughout the world. It is worth to mention that the data recorded by the big experiments at the LHC after being selected by the trigger system are still big and enough to fill around 500001 TB hard disks every year!

## 3.2.2 Data quality monitoring in ATLAS

The stored data at Tier-0 are checked to be free from irregularities by data quality offline monitoring. [72]. The recorded event is tagged to avoid any detector problems such as Tile corruption, LAr noise bursts, and problems in the production which lead to the same physics event being written out more than once. Also, events affected by the recovery

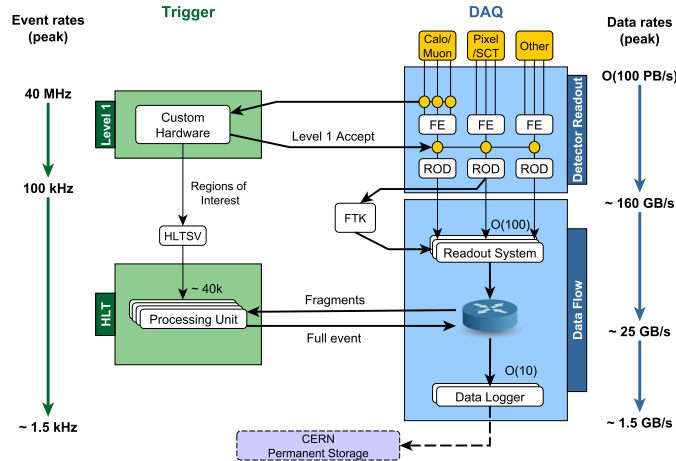


Figure 3.11: *Schematic layout of the trigger and Data Acquisition system in Run2.*

procedure for single event upsets in the SCT or information missing after the Timing, Trigger and Control (TTC [73]) system restarts are rejected [74]. They are inspected by a 24/7 shift crew, where I contributed in the period 2015-2016 of data taking, who can notify on-call experts in case of problems.

### 3.2.3 Luminosity measurements in ATLAS

Absolute luminosity measurements are performed with the Cherenkov Integrating Detector LUCID, installed in the forward region of ATLAS detector, and the Beam Conditions Monitor BCM. Their calibration is performed using data from dedicated beam-separation scans, known as van der Meer scans [75]. During Run1, 4.7 (20.7)  $\text{fb}^{-1}$  of luminosity is recorded and certified to be of good quality for the physics ATLAS analysis with  $\sqrt{s} = 7(8)$  TeV and 50 ns bunch-spacing. After the long shut-down (2012-2015), the LHC Run2 started with a center mass of energy of 13 TeV and a bunch-spacing of 25 ns. In order to cope with these changes, the ATLAS luminosity monitor LUCID and its electronics are completely rebuilt [76]. The corresponding luminosity delivered in the period 2015-2016 of Run-2 (42.7  $\text{fb}^{-1}$ ) is increased by about factor of 2 comparing to the  $\sqrt{s} = 8$  TeV measurements [75] (Figure 3.12a).

The mean number of interactions per crossing  $\mu$  during 2015 and 2016 is presented in Figure 3.12b. It is calculated as  $\mu = L_{\text{bunch}} \times \sigma_{\text{inel}} / f_r$  where  $L_{\text{bunch}}$  is the per bunch instantaneous luminosity,  $\sigma_{\text{inel}}$  is the inelastic cross section which we take to be 80 mb for 13 TeV collisions, and  $f_r$  is the LHC revolution frequency.

The preliminary delivered luminosity by LHC at 13 TeV, shown by the green histograms of Figure 3.13, is used by both ATLAS and CMS experiments. The one recorded by ATLAS, shown by yellow histograms, reflects the Data Acquisition (DAQ) inefficiency, as well as the inefficiency of the so-called ‘warm start’: when the stable beam flag is raised, the tracking detectors undergo a ramp of the high-voltage and, for the pixel system, turning on the preamplifiers [77]. Analysis presented in this document uses the good data for physics collected by ATLAS in 2015 (Figure 3.13a) for Chapter 5 and both 2015 and a part of 2016 until mid of July (Figure 3.13b) for Chapter 6.

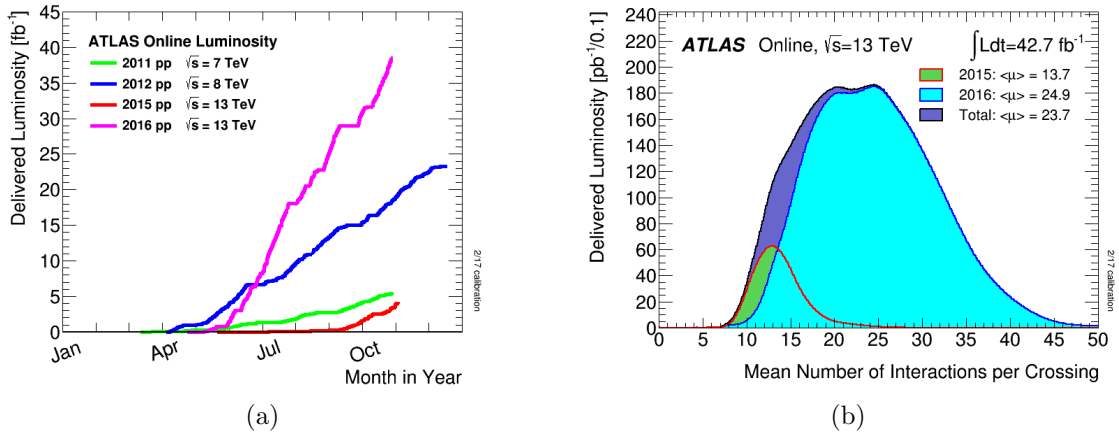


Figure 3.12: (a) Delivered Luminosity versus time, for 2011-2016, delivered to ATLAS during stable beams and for high energy  $pp$  collisions. (b) Number of Interactions per bunch crossing showing the combined 13 TeV data from 2015 and 2016.

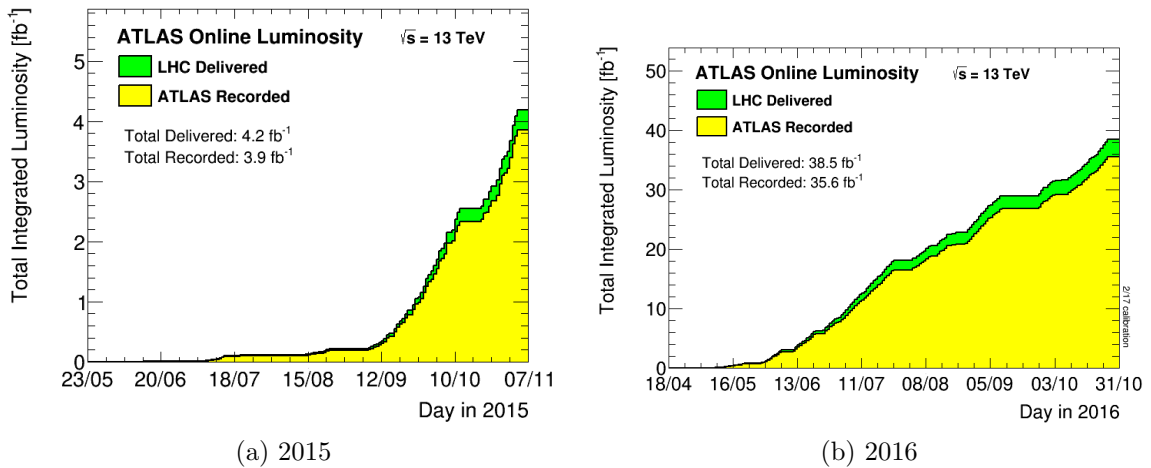


Figure 3.13: Cumulative luminosity versus time delivered to (green) and recorded by ATLAS (yellow) during stable beams for  $pp$  collisions at 13 TeV center-of-mass energy in (a) 2015 and (b) 2016.

# Chapter 4

## Object reconstruction

Each event recorded by ATLAS includes many different types of particles registered as stream of bytes of information per beam crossing; including sensor measurements, or hits, in trackers (Inner Detector and muon spectrometer) and energy measurements in the calorimeters. It is processed through the reconstruction software in order to identify every type of outgoing particle. The reconstruction of the tracks and the primary vertices, needed for each object reconstruction, is illustrated in Section 4.1. The description of the reconstruction and identification of electrons (Section 4.2), photons (Section 4.3), muons (Section 4.4), taus (Section 4.5) and jets (Section 4.6) are given which are considered later in the  $t\bar{t}H$  analysis<sup>1</sup> (Chapter 6). A conclusion closes this chapter in Section 4.7.

### 4.1 Tracks and primary vertices (PV)

Tracks provide information on the particle's origin and direction and therefore of its momentum. A track seed is defined when a track contains three hits in different layers of the silicon detectors, with a transverse momentum larger than 400 MeV [78]. A precision hit<sup>2</sup> track is a seed track with at least 4 precision hits in the silicon detector. A full track is defined by at least 7 hits; when a track seed can be successfully extended to the TRT hits.

The track reconstruction follows two steps: "pattern recognition" and "track fit" using optimised tracking algorithm to consider the increase of the pileup and make use of the new IBL at Run2 [79]. The pattern recognition is an algorithm identifying the hits that belong to a single track. Two models are used: the standard pion hypothesis and the electron hypothesis used as a second attempt in case of larger energy loss due to interactions with the detector material.

A set of algorithms is used to estimate the particle momentum 3-vector by fitting the track curvature to assess the helicoidal trajectory of charged particles in the solenoidal magnetic field. These algorithms also take into account the detailed information about the energy loss in the material along the tracks length. In total, five parameters are used: the transverse ( $d_0$ ) and longitudinal ( $z_0$ ) impact parameters, the particle direction ( $\phi$ ,  $\theta$ ) and  $q/p$ , where  $q$  is the charge and  $p$  is the track momentum in the ATLAS Global  $\chi^2$  Track Fitter [80] with pion hypothesis with  $p_T > 400$  MeV and  $|\eta| < 2.5$ , based on the scattering angle formulation of the track fit. The optimised Gaussian Sum Filter (GSF) [81] is then used for the electron hypothesis to improve the track-cluster matching and recover efficiency losses due to electrons undergoing non-linear bremsstrahlung. Thanks to its performance, the electron reconstruction efficiency is increased by roughly 5% [82].

---

<sup>1</sup>The reconstruction of missing energy  $E_T$  is not described since it is not used in this thesis.

<sup>2</sup>Precision hits refers to silicon hits in the SCT and the pixel detectors.



For the tracking efficiencies and fake rates, two working points are considered: the *Loose* and *Tight* track selection correspond to the track requirements applied during the track reconstruction with 90% and 85%, respectively, for tracks with  $p_T > 5$  GeV [83].

The extrapolation of the tracks allows to reconstruct the vertices corresponding to the interaction point of  $pp$  collisions. Due to the large number of protons per bunch crossing, multiple interaction vertices can be reconstructed in the event. The primary collision vertex (PV) is defined as the vertex with the largest  $\sum p_T^2$  computed from associated tracks, while the others are recognised as pileup vertices. Vertices incompatible with the beam collision region are considered secondary vertices or displaced vertices. The "adaptive vertex fitting" algorithm is used to suppress the impact of multiple overlapping  $pp$  collisions (pileup), by requiring the tracks to originate from the PV. The distribution of reconstructed vertices are different between data and simulation [78]. Therefore, simulation used in all ATLAS analyses, are reweighted as a function of pileup interactions for each event (pileup reweighting).

## 4.2 Electrons

The online and offline procedure to reconstruct an electron is elucidated in Section 4.2.1 and 4.2.2 respectively. To further identify electrons and reduce fake electrons, the offline identification procedure used at Run2 is explained in Section 4.2.3. Additional isolation requirements that could be applied to further reject background processes, are briefly mentioned in Section 4.2.4.

### 4.2.1 Electron trigger

The online reconstruction and identification of electrons are performed at the L1 and HLT trigger sub-systems. At L1, a trigger "towers" of  $0.4 \times 0.4 \Delta\eta \times \Delta\phi$  dimension are used to calculate the energy of the recorded signal from the calorimeter. In this stage, EM candidates are selected by requiring low  $E_T$  thresholds to get high trigger efficiencies. At HLT, a fast EM calorimeter and tracking algorithms are first applied to build EM clusters and tracks which are then matched within coarse angles ( $\Delta\eta < 0.2$ ). Tighter  $E_T$  threshold and initial requirements on  $R_{\text{had}}$ ,  $R_\eta$  and  $E_{\text{ratio}}$  (see Section 4.2.3) are applied to improve the efficiency and reduce the trigger rate. Then, offline-like algorithms are used to reconstruct and identify the HLT electron candidates. Efforts at Run2 are devoted to make the offline trigger reconstruction match as closely as possible the online trigger reconstruction [79], explained in the next section.

The Lowest unprescaled<sup>3</sup> single electron trigger operating point used in 2015 applies a 24 GeV transverse energy threshold and requires the electron to pass medium LH identification requirements [84]. To recover efficiency in the high transverse energy regime, this trigger is complemented by a trigger requiring a transverse energy above 120 GeV with loose LH identification. They are used later in Chapter 5 and 6 by requiring the offline well reconstructed and identified electron to match one, or (and) more, of the requested online electrons within  $\Delta R < 0.07$  [85]. Other set of di-electron trigger operating points exist but they are not used in this thesis.

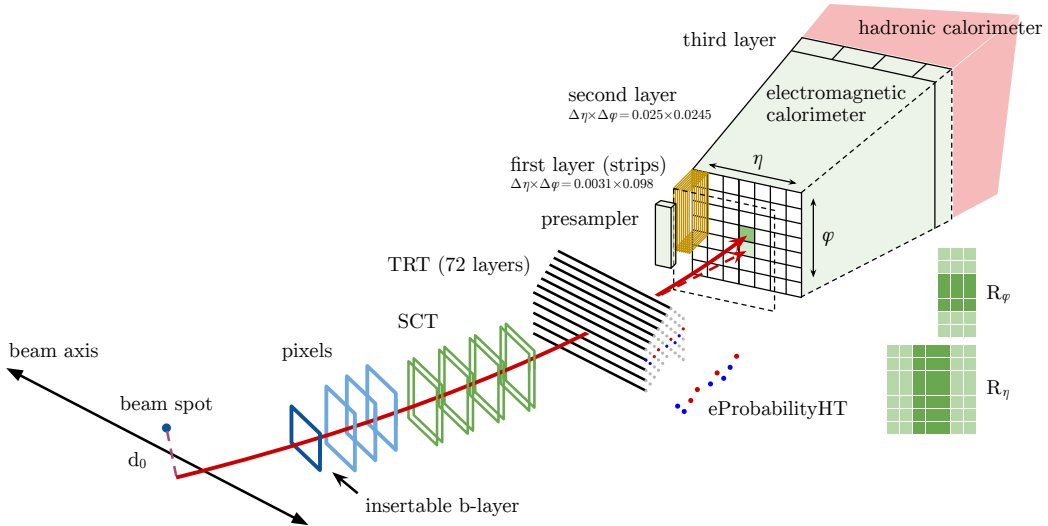


Figure 4.1: *Schematic view of the electron reconstruction and identification [85].*

## 4.2.2 Electron reconstruction

Electrons leave a track in the Inner Detector as well as an energy deposit in the EM calorimeter (Figure 4.1). Therefore, their reconstruction is performed, within EM detector acceptance  $|\eta| < 2.47$ , by the match of a seed cluster to a reconstructed track (see Section 4.1).

The seed cluster (pre-cluster) finding is performed using the "sliding-window" algorithm. First, the entire EM calorimeter is readout into  $200 \times 256$  squares of  $0.025 \times 0.025$  in the  $\eta$ - $\phi$  plane. Contributions from all layers are summed in each square, forming a 'tower'. Second, a "pre-cluster" is formed if the sum of  $E_T$  inside a "window" of  $3 \times 5$  (in units of the tower size  $0.025 \times 0.025$ ) is above 2.5 GeV [85]. In case two seed clusters close to each other are found (duplicate clusters), only the one with the highest energy is kept. The position of the pre-cluster is determined as the energy-weighted barycenter of all cells in the window size ( $3 \times 3$ ). The EM cluster region of interest (RoI) is defined as the region with a cone-size of  $\Delta R = 0.3$  around the seed cluster barycentre that passed loose shower shape requirements of  $R_\eta > 0.65$  and  $R_{had} < 0.1$ .

For the matching, at least one precision hit track is extrapolated from the last measurement hit to the second EM layer seed cluster. If the difference in  $\eta$  and  $\phi$  positions between the impact point and the seed cluster position is lower than 0.01, the matching is successful. Otherwise, the track is refitted using electron hypothesis, i.e a track seed with at least 7 hits falling within one of the EM cluster RoI. If no track is found within the EM cluster RoI, the GSF is applied to recover tracks with a large momentum loss. The resulting tracks are then corrected for bremsstrahlung. A similar procedure of cluster-track matching is repeated for the refit track with stricter conditions. If several tracks fulfill the matching condition, one track is chosen as "primary" track. However, if no associated precision hit tracks is found, the object is considered to be a photon (see Section 4.3). The efficiency of this association fulfilling the track quality cuts is measured in the data and described in details in Chapter 5.

After the matching, the electron cluster is rebuilt (EM cluster formation) using towers of second sampling cells of dimension  $3 \times 7$  and  $5 \times 5$  in the barrel and end-cap regions respectively. The energy of the clusters is calibrated to the original electron energy us-

<sup>3</sup>Unprescaled triggers means that every event passing that trigger logic is accepted.



ing multivariate techniques (MVA) based on simulated MC samples [86] to correct for the energy lost in the material upstream of the calorimeter. An additional set of corrections are applied to mitigate the non-uniformity of the detector response. Furthermore, a residual disagreement in the energy scale between data and simulation are corrected using data-driven corrections with energy scale factors " $\alpha$ " applied to data (Figure 4.2a) and additional scale factors " $c'$ " applied to MC electron energies to match the data (Figure 4.2b). The latter is applied on MC to account for the worse resolution in data. Two sets of uncertainties are assigned and used later in the analysis chapter. The uncertainty on the electron energy scale are between less than one per mil in the barrel and a few per mil in the endcaps. The uncertainty on the effective constant term added to the simulation is between two and five per mil [87].

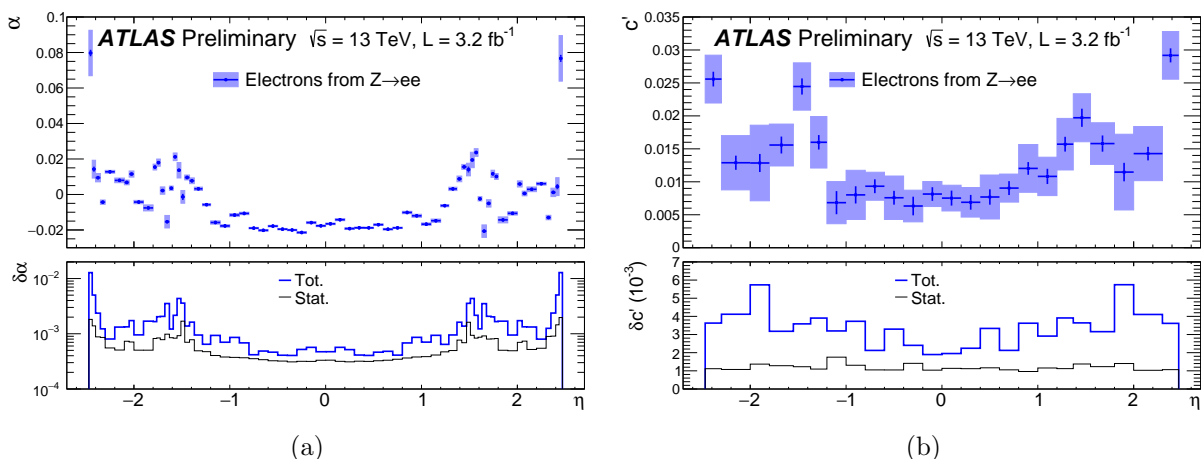


Figure 4.2: (a) Energy scale factor  $\alpha$  and (b) additional constant term  $c'$  for energy resolution from  $Z \rightarrow ee$  events as a function of  $\eta$  [87]. The uncertainty bands on the top plots represent the total uncertainties on these quantities, while the thin black (resp. thick blue) lines at the bottom represent the statistical (resp. total) uncertainties.

The last step includes the computation of the final physical parameters (e.g: four-momentum, charge) of the reconstructed electron using information from the final calibrated energy cluster and the best track matched to the original seed cluster.

### 4.2.3 Electron identification

The identification of central electrons and the rejection of background candidates is performed with a likelihood-based (LH) method, as default for all Run-2 data analyses. It is a multivariate analysis (MVA) technique that simultaneously evaluates several properties of the electron candidates when making a selection decision. The identification algorithms uses 20 discriminant variables as an input to improve electron/jet separation. They are calculated based on the energy, the number of hits or the particle position information gotten from different layers of the ATLAS detector. They could be classified into calorimeter shower shape variables, tracking variables including information from the transition radiation tracker and track properties and track-cluster matching related quantities.

- Hadronic leakage:

- $R_{\text{had1}}$  is the 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  $|\eta| > 1.37$ ).

- $R_{\text{had}}$  is the ratio of  $E_T$  in the hadronic calorimeter to  $E_T$  of the EM cluster (used over the range  $0.8 < |\eta| < 1.37$ )
- Back layer of EM calorimeter:
  - $f_3$ : is the 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.
- Middle layer of EM calorimeter:
  - $w_{\eta 2}$  is a 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.
  - $R_\phi$  is the ratio of the energy in  $3 \times 3$  cells over the energy in  $3 \times 7$  cells centered at the electron cluster position.
  - $R_\eta$  is the ratio of the energy in  $3 \times 7$  cells over the energy in  $7 \times 7$  cells centered at the electron cluster position.
- Strip layer of EM calorimeter:
  - $w_{\text{stot}}$  is a shower width,  $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$ , 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  $i_{\text{max}}$  is the index of the highest-energy strip.
  - $E_{\text{ratio}}$  is the ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies.
  - $f_1$  is the ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter.
- Track conditions:
  - $n_{\text{Blayer}}$  is the number of hits in the innermost pixel layer; discriminates against photon conversions.
  - $n_{\text{Pixel}}$  is the number of hits in the pixel detector.
  - $n_{\text{Si}}$  is the number of total hits in the pixel and SCT detectors.
  - $d_0$  is the transverse impact parameter with respect to the beam-spot
  - $d_0/\sigma_{d_0}$  is the significance of transverse impact parameter defined like the ratio of  $d_0$  and its uncertainty.
  - $\Delta p/p$  is the momentum lost by the track between the perigee and the last measurement point divided by the original momentum.
- TRT:
  - eProbabilityHT is a likelihood probability based on the TRT high-threshold hits. The probability for each TRT hit to exceed the high level threshold depends on the straw gas type, the Lorentz factor  $\gamma$  calculated from the track  $p_T$  under a particle type hypothesis, and the geometry: detector partition, straw layer, track-to-wire distance and the hit coordinates ( $z$  for the barrel and radius for the endcaps) [85].
- Track-cluster matching:

- $\Delta\eta_1$  is a  $\Delta\eta$  between the cluster position in the strip layer and the extrapolated track
- $\Delta\phi_2$  is the  $\Delta\phi$  between the cluster position in the middle layer and the track extrapolated from the perigee.
- $\Delta\phi_{\text{res}}$  is defined as  $\Delta\phi_2$ , but the track momentum is rescaled to the cluster energy before extrapolating the track from the perigee to the middle layer of the calorimeter.
- $E/p$  is the ratio of the cluster energy to the track momentum.

Three levels of identification operating points are defined in order of increasing background rejection: LooseBLayer LH, Medium LH, and Tight LH. The identification efficiency is measured in data using four methods, in  $Z \rightarrow ee$  and in  $J/\psi$  decays with different background subtraction techniques with  $E_T > 7$  GeV for the three identification working points. The impact parameter requirements are  $|d_0|/\sigma_{d_0} < 5$  and  $|\Delta z_0 \sin \theta_\ell| < 0.5$  mm. The identification efficiencies in general increase as a function of  $E_T$  (Figure 4.3a)

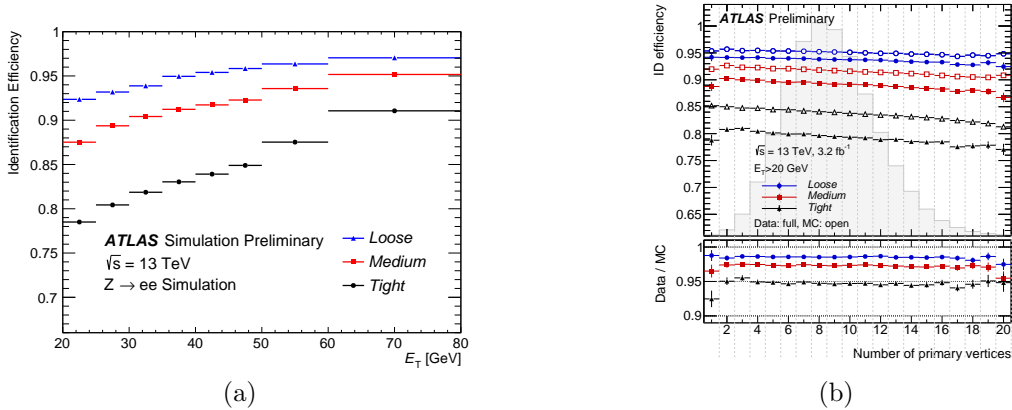


Figure 4.3: *Efficiencies for the different electron likelihood operating points obtained from  $Z \rightarrow ee$  events as a function of (a)  $p_T$  and (b) the number of reconstructed primary vertices with  $E_T > 15$  GeV using  $3.2 \text{ fb}^{-1}$  data recorded by the ATLAS experiment in 2015 at  $\sqrt{s} = 13$  TeV [85]. The distribution of the number of reconstructed primary vertices in the selected data events is overlaid in grey.*

because electrons with higher  $E_T$  are better separated from the background in many of the discriminating variables. The efficiencies for the different electron likelihood operating points are roughly 95, 91 and 83 % for LooseBLayer LH, Medium LH, and Tight LH respectively using simulated  $Z \rightarrow ee$  events. They are found to be robust with respect to the number of primary vertices in the range probed using 2015 data (Figure 4.3b). The lower efficiency in data than in MC arises from the known mismodelling of calorimeter shower shapes in the GEANT4 detector simulation.

#### 4.2.4 Electron isolation

The calorimeter and tracking isolation variables  $E_T^{\text{cone20}}$  and  $p_T^{\text{varcone20}}$  are combined to separate isolated electron candidates (prompt electrons) from non-isolated electron candidates such as electrons from heavy flavour hadron decays, photon conversion and light

hadrons mis-identified as electrons.  $E_T^{cone20}$  ( $p_T^{varcone20}$ ) is based on the sum of transverse energies (momenta) of calorimeter clusters (tracks) excluding the electron candidate cluster (track) itself within a cone of  $\Delta R < 0.2$  ( $\Delta R = \min(0.2, 10 \text{ GeV}/p_T (e))$ ) around the electron candidate. Similar variables  $E_T^{cone30}$  ( $p_T^{varcone30}$ ) and  $E_T^{cone40}$  ( $p_T^{varcone40}$ ) are defined for larger cones  $\Delta R < 0.3$  and  $\Delta R < 0.4$  which will be used in the rest of the thesis. The calculation of these variables is harmonised with the muon isolation, using energy density pile-up corrections for isolation in the calorimeter [79]. Eight operating points are defined and divided into two classes: efficiency targeted operating points and fixed requirement operating points summarised in Table 4.1. While the fixed requirement operating points have constant upper thresholds on the isolation variables, those thresholds are  $E_T$  dependent for efficiency targeted operating points, getting from different typical isolated efficiencies estimated from simulation.

Working point	Efficiency		
	calorimeter isolation	track isolation	total efficiency
LooseTrackOnly	-	99%	99%
Loose	99%	99%	~ 98%
Tight	96%	99%	~ 95%
Gradient	$0.1143\% \times E_T + 92.14\%$	$0.1143\% \times E_T + 92.14\%$	90/99% at 25/60 GeV
GradientLoose	$0.057\% \times E_T + 95.57\%$	$0.057\% \times E_T + 95.57\%$	95/99% at 25/60 GeV

Working point	Cut value	
	calorimeter isolation	track isolation
FixedCutLoose	0.20	0.15
FixedCutTightTrackOnly	-	0.06
FixedCutTight	0.06	0.06

Table 4.1: *Electron operating points definitions [85]. The upper table illustrates the efficiency targeted operating points, and the numbers expressed in percents represent the target efficiencies used in the operating point optimisation procedure. For the Gradient and GradientLoose operating points,  $p_T$  is in GeV. The fixed requirements operating points are shown in the lower table. The calorimeter and track isolation refer to the selection based on  $E_T^{cone}/p_T$  and  $p_T^{cone}/p_T$ , respectively.*

## 4.3 Photons

Section 4.3.1 and 4.3.2 are, respectively, devoted to describe the procedure to reconstruct and identify photons.

### 4.3.1 Photon reconstruction

Photons deposit an energy in the EM calorimeter but, unlike electrons, they don't leave any track in the Inner Detector unless a photon conversion happens. Therefore, two kind of photons are reconstructed: Unconverted photons, reconstructed as a cluster object without any matching track in the Inner Detector, and converted photons, reconstructed as a cluster associated to a track originating from a "conversion vertex". The conversion

<sup>4</sup> $E_T^{cone20}$  in this context refers to the isolation variable being reconstructed from the "topological" clusters  $topoE_T^{cone20}$ .

vertices are initially classified into either a double- or a single-track candidates depending on the number of assigned electron-tracks. A single-track conversion vertex is allocated in case one of the two produced electron tracks failed to be reconstructed (very soft track) or when the two tracks are very close to each other (two high- $p_T$  tracks). Converted photon (Unconverted photons) could be recovered from initially reconstructed electron if the matched track has only the TRT information and its matched track has a transverse momentum lower (greater) than 2 GeV and (or)  $E/P$  is lower (greater) than 10 [88]. This gives an efficiency of around 94 % for converted photons and close to 100 % for uncovered photons [89]. The algorithm used to reconstruct photons is the same as the one used to reconstruct electrons in order to minimise the uncertainty due to the extrapolation of the response between electrons and photons. Then, similar calibration procedure [87] to the one used for electrons and same values of correction factors are applied for photon candidates (see Section 4.2.2).

### 4.3.2 Photon identification

Prompt photons are distinguished from fake photons coming from neutral hadron decays, such as  $\pi^0 \rightarrow \gamma\gamma$  or from radiative decays of other particles. For that, two sets of cut-based photon identification criterias, loose and tight, are defined. The former is based only on the shower shape of the second layer of the EM calorimeter and on the energy deposit in the hadronic calorimeter. Wider shower shape are required compare to the later to cope with the pair of electrons arising from a photon conversion. The tight requirement makes also use of the strip layer of EM calorimeter to distinguish between neutral pions and photons thanks to its fine granularity (see Section 3.2.1.3): the shower width for three strips around strip with maximum energy deposit ( $W_{s3}$ ); the energy outside core of three central strips but within seven strips divided by energy within the three central strips ( $F_{side}$ ); the 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 ( $\Delta E$ ) in addition to  $w_{stot}$  and  $E_{ratio}$  defined in Section 4.2.3.

## 4.4 Muons

Muons leave only tracks in the Inner detector and the muon spectrometer (MS). Section 4.4.1 and 4.4.2 are, respectively, devoted to describe the procedure to reconstruct and identify muons.

### 4.4.1 Muon reconstruction

The muon track used in the most analyses combine track reconstruction in the MS with charged particles reconstruction in the Inner Detector. Four muon types are defined depending on which subdetectors are used in the reconstruction: Combined (CB) muon, Segment-tagged (ST) muons, Calorimeter-tagged (CT) muons and Extrapolated (ME) muons. A combined track for the CB muon is formed with a global refit that uses the hits from both the Inner Detector and MS subdetectors. ST muons are used when muons, reconstructed in ID, cross only one layer (local track segment in the MDT or CSC chambers) of MS chambers. The CT muons is reconstructed from an energy deposit in the calorimeter compatible with a minimum-ionising particle associated to a track in the Inner Detector. Finally, ME muon trajectory is reconstructed based only on the MS track

and loose requirement on compatibility with originating from the IP. When two muon types share the same Inner Detector track, preference is given to CB muons, then to ST, and finally to CT muons. However, track hit and fit quality features are used in case an overlap with ME muons.

In the central region of the detector, the momentum resolution is measured to be 1.7% (2.3%) for muons from  $J/\psi(Z) \rightarrow \mu\mu$  decays in the  $p_T$  range 5–20 (22–300) GeV. The momentum scale is known with an uncertainty of 0.05%. In the region  $|\eta| > 2.2$ , the  $p_T$  resolution for muons from  $Z \rightarrow \mu\mu$  decays is 2.9% while the precision of the momentum scale for low- $p_T$  muons from  $J/\psi \rightarrow \mu\mu$  decays is about 0.2% [90].

#### 4.4.2 Muon identification

To reduce fake muons coming mainly from the decay in-flight of charge pion and kaon, four muon identification selections are provided: Medium, Loose, Tight and High- $p_T$ . The latter aims to maximise the momentum resolution for tracks with transverse momentum above 100 GeV. The identification efficiencies, well reproduced by simulation, are measured to be close to 99% [90] outside the region  $|\eta| < 0.1$  (Figure 4.4).

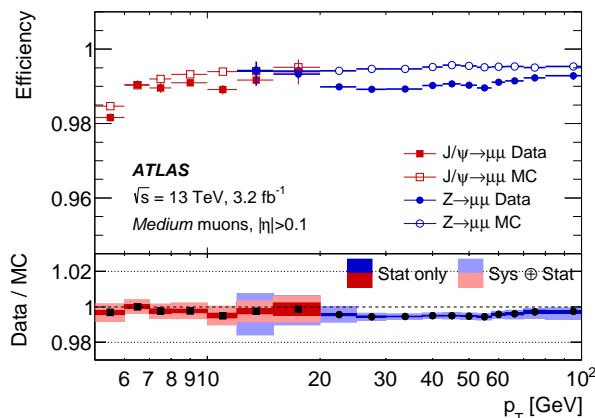


Figure 4.4: Identification efficiency for the Medium muon selection as a function of the  $p_T$  of the muon, in the region  $0.1 < |\eta| < 2.5$  [90]. The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies (scale factors), with statistical and systematic uncertainties.

## 4.5 Taus

A combination of reconstructed light leptons and jets produced from the tau decay are used to reconstruct a tau candidate. Tau lepton reconstruction involves mainly hadronically decaying tau ( $\tau_{had}$ ) representing 65% of the total decay. A hadronic tau is observed as a narrower jet compared to quark- or gluon-jets, since taus decay weakly. This helps for tau identification using discriminant shower shape variables [91]. Hadronic tau object is used only in the combination results of the analysis chapter.

## 4.6 Jets

Jets are collimated sprays of particles originating from fragmentation and hadronisation of energetic quarks or gluons. They are the most abundant products of high energy  $pp$



collisions. The reconstruction of jets is described in Section 4.6.1 including jet calibration and the associated uncertainties. The  $b$  flavour jet identification, playing a major role in the  $t\bar{t}H$  analysis, is presented in Section 4.6.2.

### 4.6.1 Jet reconstruction

Jets are reconstructed from energy deposits in the topological clusters. The topological algorithm uses the energy significance variable, defined as signal to noise ratio, to produce unfixed-size clusters. The topological algorithm includes two steps: The cluster maker and the cluster splitter. The "cluster maker" algorithm first defines a seed cell (proto-cluster) as the one that passes a typical high threshold of the energy significance, in order to suppress both electronics and pile-up noise. Then, neighboring cells are added to the seed cell to form a cluster. Those cells (adjacent proto-cluster) pass lower thresholds of significance to take into account the tails of showers. In case many proto-clusters are associated to the same adjacent proto-cluster, they are merged to one proto-cluster. In the case of overlapping showers, the "cluster splitting" algorithm is used to define a local "maximum cell" with the highest energy, not less than 500 MeV [92], surrounded by at least 4 neighboring cells.

The anti- $k_t$  algorithm with radius parameter  $R=0.4$  [93], where anti- $k_t$  stands for the inverse of the transverse momentum  $k_t$  of a particle, is used to define (partly) conical jets [94]. Jets are then calibrated with a series of simulation-based corrections and in situ techniques [93]. Initially the clusters are calibrated at the EM scale. Then, a local calibration scheme (LCW scale) is applied to reflect the real composition of the hadronic shower.

**Pileup jets** Tracks associated to hard scattering (HS) jets could be contaminated by particles originated in pileup interactions (PU jets). The jet-vertex-tagger (JVT) algorithm is used to reduce the pileup-originated jets (Figure 4.5). It is a multivariate variable based on 2D likelihood developed for Run2 to reduce the pileup-dependent jet selection efficiency and keep a robust performance against the number of primary vertices. It uses the information that most of the particles associated with the HS jet originate from the primary interaction (PV) [95]. The modeling of JVT is validated in data using  $Z \rightarrow \mu\mu$ +jets and  $t\bar{t}$  events. The jet efficiency, measured in data, for three different JVT cut values, is within 1-2 % of agreement with simulation [96].

**Jet energy scale (JES)** JES includes Pile-up corrections to account for energy offset due to in-time and out-of-time pile-up; origin corrections to point back to the PV; ratios with respect to truth jets<sup>5</sup> corrections (EM +JES or LCW + JES scheme) and residual in situ corrections to remove the remaining data-to-MC differences. Figure 4.6a represents the combined uncertainty in the JES of fully calibrated jets as a function of jet  $p_T$  at  $\eta=0$ . JES plays an important role in the source of systematics of  $t\bar{t}H$  analysis introduced in Chapter 6. The total uncertainty of the jet energy scale (JES) is less than 1% in the central calorimeter region ( $|\eta| < 1.2$ ) for jets with  $100 < p_T < 500$  GeV [93].

**Jet energy resolution (JER)** The values of the three parameters of the JER (see Equation 3.1) are estimated using 2012 data at  $\sqrt{s} = 8$  TeV:  $a=(71 \pm 7)\%$ ,  $b=(333 \pm 63)\%$  and  $c=(3 \pm 0.3)\%$  for anti- $k_t$   $R=0.4$  EM+JES jets [98]. Figure 4.6b represents the

---

<sup>5</sup>Truth jets are stable simulated particles

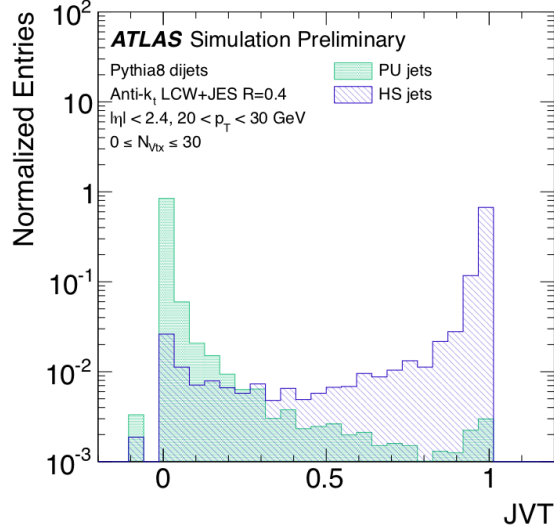


Figure 4.5: The jet-vertex-tagger (*JVT*) distribution for hard-scatter and pileup jets with  $20 < p_T < 30 \text{ GeV}$  [96].

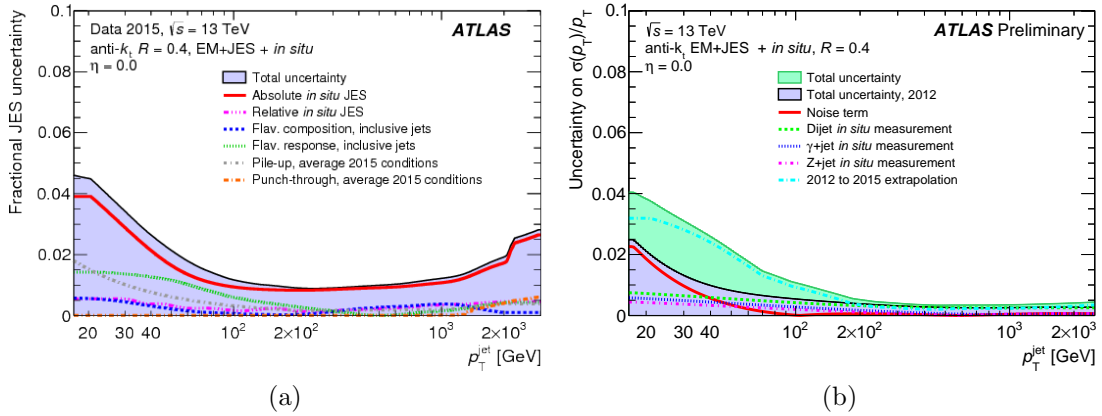


Figure 4.6: Final jet energy scale *JES* (a) and resolution *JER* (b) uncertainties of fully calibrated jets estimated for 2015 as a function of jet  $p_T$  at  $\eta = 0$  [93, 97].

combined uncertainty in the jet energy resolution (JER) of fully calibrated jets for 2015 as a function of jet  $p_T$  at  $\eta = 0$  derived from in situ studies on 2012 data. The uncertainty decreases from 4% to 0.5% going from  $E_T = 20 \text{ GeV}$  to  $E_T = 200 \text{ GeV}$  [97].

#### 4.6.2 *b*-tagged jets

The *b*-jets are jets originating from *b*-quark (*b*-hadron). The *b*-jet can be separated from the other jets since it has very distinguishable characteristics. The *b*-hadron's flight path, around  $450 \mu\text{m}$ , is long enough to be observed as a displaced vertex in the Inner Detector (secondary vertex). The *b*-tagging is performed using the features of the *b*-, *c*-, and light jets. The *c*-jets and light jets have shorter lifetimes, lower masses or lower track multiplicity in comparison with *b*-hadrons. The *b*-jet is identified using a multivariate tagger MV2, created for Run2, benefitting from the new IBL and providing 30-50 % better light jet rejection at the same *b*-tagging efficiency compared to the multivariate tagger MV1 used in Run1 [79]. The MV2c10 algorithm is used since 2016 and is illustrated



in Figure 4.7a. Its performance is found to be better than MV2c20 used during 2015 (Figure 4.7b) achieving a high  $b$ -tagging efficiency for real  $b$ -jet while retaining the  $c$  and light flavor jets misidentification efficiency at very low levels. Four operating points are used for the MV2c10  $b$ -tagging algorithm with different numbers for the  $b$ -jet efficiency, namely: 60%, 70%, 77% and 85 % corresponding to 34 (1538, 184), 12 (381, 55), 6 (134, 22) and 3.1 (33, 8.2)  $c$ -jet (light-jet,  $\tau$ ) rejections rates respectively [99]. These values have been extracted from  $t\bar{t}$  events with jet  $p_T$  above 20 GeV.

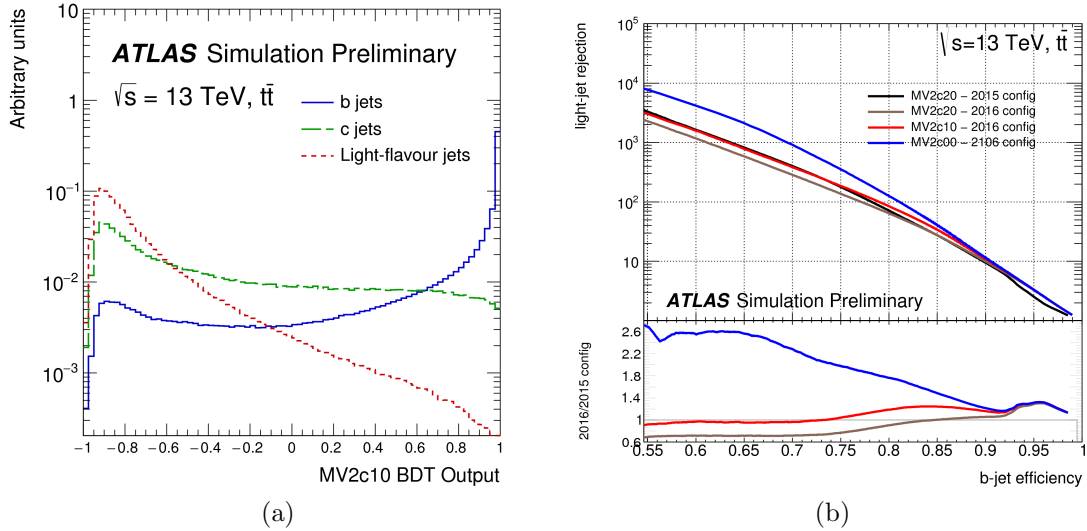


Figure 4.7: (a)  $MV2c10$  BDT output for  $b$ - (solid blue),  $c$ - (dashed green) and light-flavour (dotted red) jets evaluated with  $t\bar{t}$  events [99]. (b) Light-flavour jet rejection versus  $b$ -jet efficiency for 2015 and 2016 configuration of the MV2  $b$ -tagging algorithm evaluated on  $t\bar{t}$  events [99].  $MV2c00$  denotes the MV2 algorithm where no  $c$ -jet contribution was present in the training.  $MV2c10$  ( $MV2c20$ ) denote the MV2 outputs where a 7% (15%)  $c$ -jet fractions was present in the background sample (for the 2016 configuration).

## 4.7 Conclusion

In this chapter, electron, photon, muon, tau and jet reconstruction and identification in ATLAS at Run2 are presented. Tracks and clusters are reconstructed and matched to each other using different algorithms. They are associated to a calibration procedure in order to get the final measurements of the position, energy, mass and charge of each type of particle. The main systematic uncertainties associated to these measurements in addition to the weights, applied on simulation to make it comparable to data, were also presented. All these sources of uncertainties and weights are considered in the physics analysis chapter. The measurements of the electron reconstruction efficiency using 2015 data with  $3.2 \text{ fb}^{-1}$ , used to compute the electron reconstruction weight, will be explored in the next chapter.

# Chapter 5

## Measurement of the electron reconstruction efficiency

The remarkable discovery of the Higgs boson in the experimentally harsh LHC environment would not have been possible without a deep understanding of the detector performance. This is for instance the case of the optimisation of the electron<sup>1</sup> reconstruction and identification algorithms. The precise reconstruction and identification of electrons is also crucial to measure the SM cross sections such as  $Z/W$  and dibosons cross sections, where the associated uncertainty to the electron reconstruction and identification efficiency is an important source of systematics. Moreover, the accuracy with which the MC based detector simulation models the electron efficiency plays an important role in various searches for new physics. In order to achieve reliable physics results, the simulated samples need to be corrected to reproduce the measured data efficiencies as closely as possible. For these reasons, the efficiencies are estimated both in data and in simulation.

To study and check the performance of the electron algorithms, efficiencies are defined as the probability to detect a real electron by the sub-detectors (tracker and EM calorimeter). Electron efficiencies as well as their uncertainties are determined in fine bins of transverse momentum and rapidity  $\eta \times E_T$  from data and MC using clean and unbiased sample of electrons ( $Z \rightarrow ee$ ,  $W \rightarrow e\nu$ ,  $J/\Psi \rightarrow ee$ ) with the tag-and-probe (T&P) method. At the end of Run1, precise measurements of the electron reconstruction and identification efficiencies [82, 100] were achieved with 97% for reconstructed electrons with  $E_T = 15$  GeV and 99% at  $E_T = 50$  GeV, with sub-percent typical measured errors for  $E_T > 25$  GeV and between 0.5-2 % for lower  $E_T$  [101, 102]. The ratio between data and MC efficiencies is used as a multiplicative correction factor for MC by all ATLAS analyses involving electrons. These data-to-MC correction factors (or Scale Factors, SF) are usually rather close to unity. Deviations stem from the mis-modeling of tracking properties or shower shapes in the calorimeters. At the beginning of Run2, same measurements are done to check the new improvements on reconstruction and identification algorithms (see previous chapter). My contribution focuses on the electron reconstruction efficiency measurements using 2015 data with  $3.2 \text{ fb}^{-1}$ , made public in Ref. [85].

In ATLAS, several electron, identification, isolation and trigger working points are available to match the need of all physics analyses. Therefore, the total electron efficiency is divided into different components, namely reconstruction  $\varepsilon_{\text{reco}}$ , identification  $\varepsilon_{\text{ID}}$ , isolation  $\varepsilon_{\text{iso}}$ , and trigger  $\varepsilon_{\text{trigger}}$  efficiencies:

$$\varepsilon = \varepsilon_{\text{reco}} \times \varepsilon_{\text{ID}} \times \varepsilon_{\text{iso}} \times \varepsilon_{\text{trigger}} \quad (5.1)$$

---

<sup>1</sup>Electrons provide clear experimental signature and are useful to define triggers thanks to their low cross section compared to the dominant QCD final state.

It is worth to notice that those efficiencies are correlated and sequential, where the denominator of each efficiency is the numerator of the previous one. For example, the denominator of the identification efficiency is the numerator of the reconstruction efficiency and so on. According to the available working points in ATLAS at the beginning of Run2, 312 efficiencies (scale factors) are estimated and provided to ATLAS analyses taking into account all possible configuration of the total electron efficiency as shown in Figure 5.1.

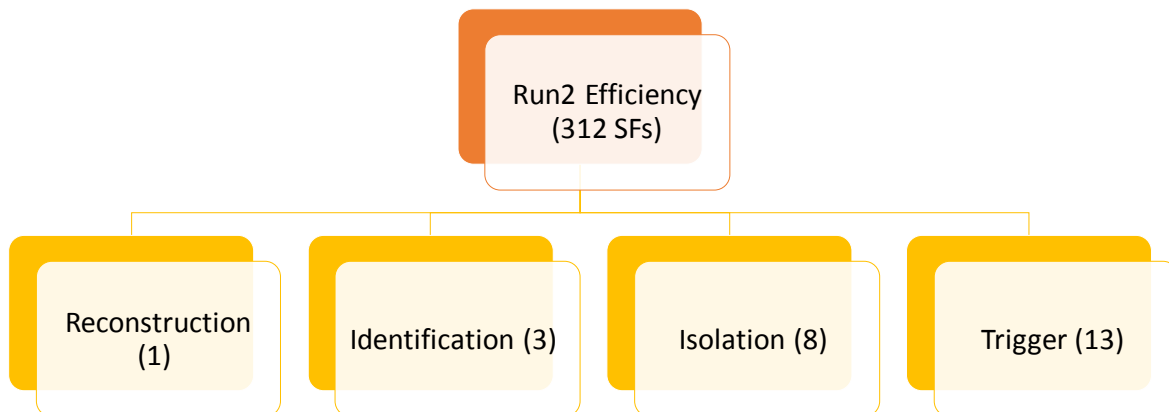


Figure 5.1: *Electron efficiency and scale factors (SFs) measurements in ATLAS at Run2 taking into account all identification, isolation and trigger WPs configuration mentioned in Section 4.2.*

Electron reconstruction efficiency, the topic of this chapter, represents the most challengeable piece in those measurements due to the high amount and difficult background to estimate in the denominator, coming from all EM clusters. The definition of the electron reconstruction efficiency is presented in Section 5.1. The methods used to measure it are detailed in Section 5.2. The input of the measurements are introduced in 5.3, while the results using 2015 data with  $3.2 \text{ fb}^{-1}$  are presented in Section 5.4. Finally, conclusions are drawn in section 5.5.

## 5.1 Definition

The efficiency to detect an energy cluster with the sliding window algorithm in the EM calorimeter is found to be greater than 99% for  $E_T > 15 \text{ GeV}$  [100]. Therefore, EM clusters are the starting point of the reconstruction efficiency measurement. The reconstruction efficiency  $\varepsilon_{\text{reco}}$  at Run2 is defined as the ratio of the number of electrons reconstructed with a matching track passing track quality requirements, with respect to

the total number of EM clusters with or without a matching track:

$$\varepsilon_{\text{reco}} = \frac{N^{\text{Quality-Track}} - B^{\text{Quality-Track}}}{(N^{\text{Quality-Track}} - B^{\text{Quality-Track}}) + (N^{\text{No-Quality}} - B^{\text{No-Quality}}) + (N^{\text{No-Track}} - B^{\text{No-Track}})} \quad (5.2)$$

where  $N$  ( $B$ ) represents the number of reconstructed signal (background) probes, “Quality-Track” represents probes associated with a good quality track having at least 7 precision hits and 1 pixel hit, “No-Quality” refers to probes associated with a track with  $\geq 4$  silicon hits but no pixel or less than 7 precision hits, and “No-Track” refers to probes with either a TRT-only associated track or no. Splitting the denominator into 3 terms allows for an optimised background determination for each case.

Since the electron efficiencies depend on  $E_T$  and  $\eta$ , the measurements are performed in two-dimensional bins in transverse energy and pseudo-rapidity ( $E_T, \eta$ ), as specified in Tables 5.1 and 5.2. Therefore, 200 central values are estimated.

Table 5.1: *Measurement bins in EM cluster transverse energy.*

Bin boundaries in $E_T$ [GeV]											
15	20	25	30	35	40	45	50	60	80	125	200

Table 5.2: *Measurement bins in EM cluster pseudo-rapidity.*

Bin boundaries in $\eta$																				
-2.47	-2.37	-2.01	-1.81	-1.52	-1.37	-1.15	-0.8	-0.6	-0.1	0	0.1	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47

## 5.2 Method

A description of the T&P method based on  $Z$  invariant mass is detailed in Section 5.2.1, followed by the background estimation methods in Section 5.2.2. The uncertainties measurements are described in Section 5.2.3.

### 5.2.1 Tag-and-Probe method with $Z \rightarrow ee$ events

The T&P method is developed to take full advantage of the precise known  $Z$  boson mass ( $91.1876 \pm 0.0021$  GeV). It is used to estimate all the terms of Equation (5.2). It requires at least one electron and one EM cluster in the event. Very strict selection criteria are applied to one of the decay electrons, called "tag", to reduce the background in the  $Z$  peak signal region. The EM cluster, the "probe", is used for the efficiency measurements after accounting for the residual background contamination. The probe is assured to come from  $Z$  decay by the requirements on the di-electron invariant mass as illustrated in Figure 5.2. In order not to bias the selected probe sample, each valid combination of tag-and-probe pairs in the event is considered, such that an electron can be the tag in one pair and the probe in another, in case two electrons in the same event fulfill the tag requirements. No requirement is made on the charge of the probe since the charge of the EM cluster is unknown. In the following, "electron" probes denote probes

selected from reconstructed electron algorithm. It could be a real or fake electron (for example, hadrons misidentified as electrons, electrons from semileptonic heavy flavour decays or from photon conversions). "Photon" probes denotes any probe selected from reconstructed photon algorithm. The latter case is taken into account to consider any electron misreconstructed as a photon (the last term in the denominator of Equation (5.2)).

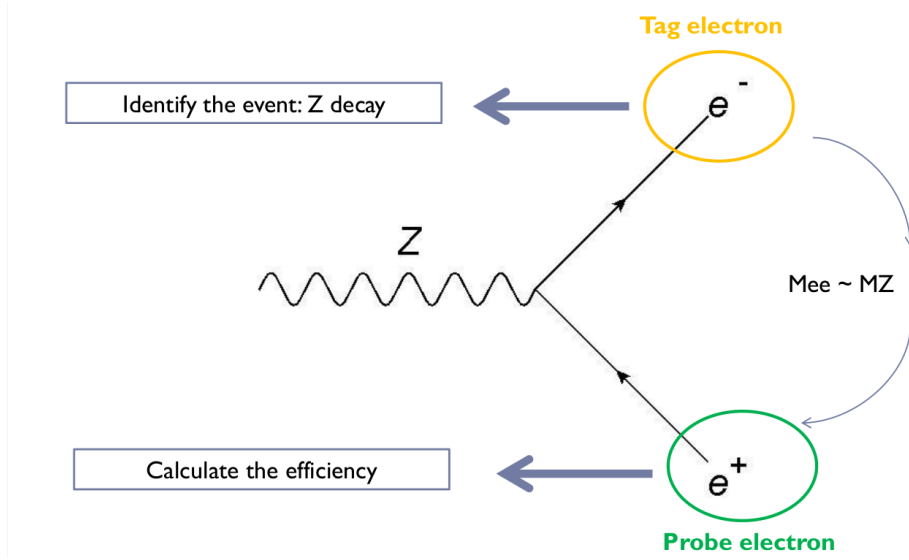


Figure 5.2: *Tag-and-probe method description using  $Z \rightarrow ee$  sample. Very strict selection criteria are applied on one of the electrons (Tag) to make sure it is coming from  $Z$  peak signal region. Loose selection is applied on the second electron (Probe) to not bias the measurements. Additional event selection criteria are applied to further reject background.*

**Event and object selection** For the efficiencies measured with both data and simulation, events are selected using a single electron trigger, with an  $E_T$  threshold of 24 GeV and *Medium* identification requirements, and one reconstructed primary vertex with at least three tracks. The tag electron is selected with  $E_T$  threshold of 25 GeV. It should be outside the crack region <sup>2</sup>, be matched to a trigger electron object within  $\Delta R < 0.07$  and passes the tight Likelihood (LH) identification criteria. All electron probes should be separated from any jet with  $E_T > 20$  GeV present in a cone with an opening angle  $\Delta R_{jet,e} = 0.4$  around the electron. This cut helps in rejecting the jets faking electrons type of background. Photon probes are required to be reconstructed by the standard egamma algorithm (see Section 4.3), and be well separated from any electron within  $\Delta R_{\gamma,e} > 0.15$ . The latter cut is rejecting photons reconstructed from a conversion vertex. The measurement is only performed for probes with  $E_T > 15$  GeV since the background contamination becomes too large at lower  $E_T$ . Finally, the invariant mass of the tag-probe system is required to be within  $\pm 15$  GeV of the  $Z$  mass.

The basic object at the *all probes* level is an electromagnetic cluster. Hence the background level is high, making the reconstruction efficiency measurement very challenging with data, especially at low  $E_T$ . After the T&P selection, aimed to reduce the background as much as possible without introducing any bias on the measurements, an enriched

<sup>2</sup>The crack region denotes the transition region between the barrel and the end-cap of the electromagnetic calorimeter,  $1.37 < |\eta| < 1.52$

sample of true electrons is selected. However still a significant background is present. The associated background, after the selection, is mainly hadrons coming from  $W$ +jets ,  $Z$ +jets and  $bb, cc \rightarrow eX$  processes, followed by  $Z \rightarrow \tau\tau \rightarrow ee, t\bar{t}$ , and di-boson processes [101]. It was shown at Run1 that the associated background is not well modeled by simulation. Therefore, data driven methods are instead used to quantify the background at high precision with fine detector granularity. The background estimation in data is described in the next section. However, no background subtraction needs to be applied on the simulated samples. Instead, the reconstructed electron is required to match the true electron during simulation<sup>3</sup>.

## 5.2.2 Background estimation

The "electron" probe samples are contaminated by background objects (for example, hadrons misidentified as electrons, non isolated electrons from semileptonic heavy flavour decays or from photon conversions) denoted by  $B^{Quality-Track}$  and  $B^{No-Quality}$  in Equation (5.2). This contamination is estimated using background template shapes of the di-electron invariant mass as detailed in Section 5.2.2.1. The "photon" probe samples are used to extract the number of electrons misreconstructed as photons (fake photons). The associated background, denoted by  $B^{No-Track}$ , is estimated using the sideband fit described in Section 5.2.2.2. In the following, for illustration purpose, Figure 5.3 shows an example of the invariant mass distribution of the tag-probe pairs for two selected bins at the *all probes* level (denominator) with  $0.80 < \eta < 1.15$  and  $15 \text{ GeV} < E_T < 20 \text{ GeV}$  (left) or  $40 < E_T < 45 \text{ GeV}$  (right) where the three background components are shown.

### 5.2.2.1 Background for clusters with an associated track

The background for clusters with an associated track, with/without track quality shown by purple/magenta markers in Figure 5.3, is subtracted using templates which are constructed with probes that fail some identification and isolation criteria in order to model the correct shape and composition of the background with minimal contribution from signal electrons. The templates are then normalised to the data using the sidebands (control regions shown in Table 5.3) of the invariant mass distribution after subtracting any residual signal contamination, estimated using simulated events. Data passing *tight* is therefore used to normalise the simulated events to data in the Z-peak signal region where the background is negligible. The number of background probes with an associated track in the Z-peak signal region could be then written as:

$$B^{Quality-Track}(B^{No-Quality}) = N_{SR}^{\text{template}} \times \underbrace{\frac{N_{CR}^{\text{baseline}} - N_{CR}^{\text{MC}} \times \overbrace{\frac{N_{SR}^{\text{DATA}, \text{tight}}}{N_{SR}^{\text{MC}, \text{tight}}}}^{\text{signal contamination}}}{N_{CR}^{\text{template}}}}_{\text{template normalization}}$$

Where:  $N^{\text{baseline}}$  and  $N^{\text{template}}$  are the total number of probes in the baseline and the number of probes in the background template, respectively, in data signal region. Indices SR and CR stand for the signal region and the control region respectively.

<sup>3</sup>Electrons matched to secondary electrons from the electron shower, or FSR photons are also considered as true electrons, essentially because they are not distinguishable experimentally from the primary electrons.

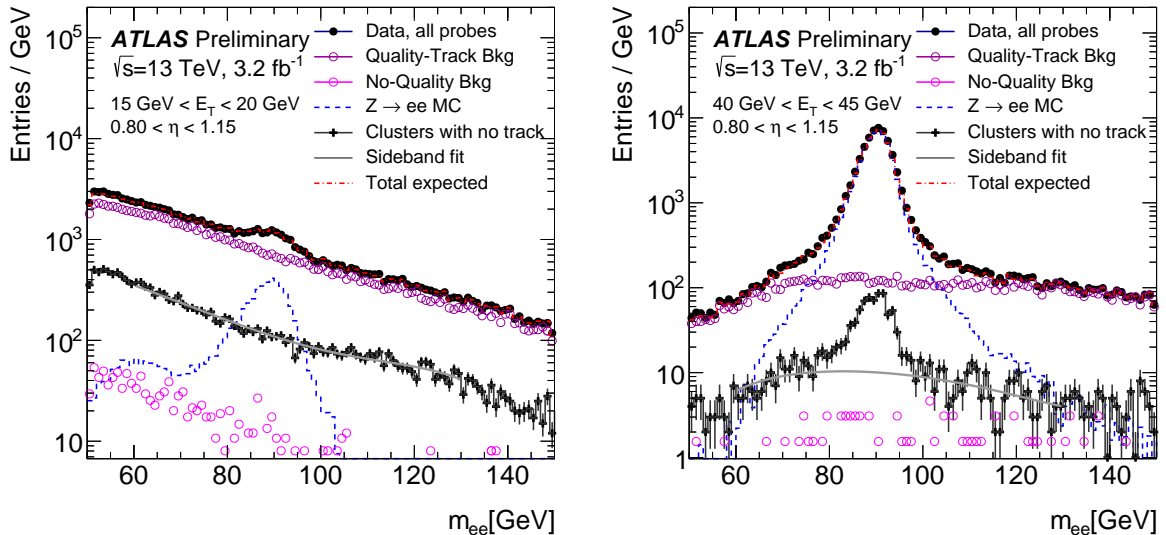


Figure 5.3: *Invariant mass distributions of the tag-and-probe pairs at the all probes level (the denominator of the electron reconstruction efficiency). Two typical bins are picked out for probes with  $0.80 < \eta < 1.15$  at  $15 \text{ GeV} < E_T < 20 \text{ GeV}$  (left) and  $40 < E_T < 45 \text{ GeV}$  (right) to show the different background compositions between high and low transverse energies. Three constituents of the background are shown using data-driven methods: Background templates for clusters with an associated track with/without track quality requirements (purple/magenta markers). In this case, the templates are normalised to the high mass control region. Background for clusters with no associated track (grey dashed line spanning from 60 to 130 GeV) estimated by a 3rd order polynomial fit. The sum of the total estimated background and the MC signal prediction (blue dashed line), shown by the red line, agrees well with the data (black dots).*

Failing at least two electron identification cuts from the *loose* cut-based criteria <sup>4</sup> is found to be the best adjustment between high background selection and high signal rejection. Further requirements on the calorimeter isolation helps to minimise the isolated electron (signal) contamination by requiring the  $E_T^{cone40}/p_T$  to be greater than 0.05 for  $E_T \geq 30 \text{ GeV}$  and 0.02 at low  $E_T$ . Those thresholds are kept similar to Run1 values since no strong motivation is raised to change them. This will be shown later in section 5.4 by comparing the results with Run1 performance.

To subtract the signal contribution from the background template,  $Z \rightarrow ee$  MC sample is used as it is known to reproduce well the data. It is shown in Run1 that the contribution of  $Z \rightarrow \tau\tau \rightarrow ee$ ,  $t\bar{t}$ , and di-boson in the base distribution are not significant (they are below 1% for  $p_T < 50 \text{ GeV}$  [101]). Any contribution from other processes is covered by systematic uncertainties by varying the fraction of the signal in the template. This is done by changing the reverted isolation criteria on the probe described in Section 5.2.3.

### 5.2.2.2 Background for clusters with no associated track

The templates do not work well with "photon" probe background due to the presence of true photons having criteria differ from other electrons (true or non-isolated) and hadrons. Therefore, the background associated to the EM clusters candidates with no associated track, shown by grey dashed line in Figure 5.3, is deduced by fitting the

<sup>4</sup>Variables related to track quality are not considered.



electron-"No-Track EM clusters" (tag-probe) invariant mass distribution using a third order polynomial function. First, a binned  $\chi^2$  fit is performed in the sideband regions of low ( $m_{ee} < 80$  GeV) and high ( $m_{ee} > 100$  GeV) masses. The size of each bin is 5 GeV. The choice of the sideband regions is based on the low amount of signal events contribution. Any residual signal contamination is subtracted using MC simulation. Then, the background estimation is taken as the integral of the fitted function in the signal region. The mathematical description of the sideband fit method is given in Appendix A.1.

The advantage of using the polynomial fit, instead of using template, comes up with the stability improvements of the results studied in details in Ref. [101]; The central values of the measured efficiencies are well defined without negative or infinite values, and the spread of the variations is quite reduced in some cases.

To check the robustness of the fit, many intervals of the sideband masses are used explained in Section 5.2.3. The impact of this variation is negligible taking into account the low contribution of the "No-Track" term comparing to the other terms shown later in Table 5.7. Hence, the precision of the method is sufficient and less crucial for this kind of background. Therefore, the same strategy is kept and no improvements are further tried.

### 5.2.3 Uncertainties

**Systematics** In order to estimate the impact of the analysis choices and potential imperfections in the background modeling, different variations of the efficiency measurement are carried out:

- Variations on the tag identification. Three different tag ID are chosen: *Tight*, *Tight* with  $E_T^{\text{cone40}} < 5$  GeV and *Medium* with  $E_T^{\text{cone40}} < 5$  GeV. This allows to modify the amount and composition of the background (proportion of events with a real isolated electron from  $W/Z$ +jets, and QCD events without any real electron).
- Variations on the  $Z$ -mass peak window: ]80,100[, ]75,105[ and ]70,110[ GeV. This allows to include low mass regions rich in bremsstrahlung electrons.
- Two background template variations, described in Table 5.3, to change the contamination of the signal proportions. The low mass control region  $60 < m_{ee} < 70$  GeV is used for  $E_T < 30$  GeV as a second variation to include more statistics. However it is not used for  $E_T > 30$  GeV because of the significant contamination of the signal coming mainly from Drell-Yan  $Z/\gamma^* \rightarrow e^+e^-$  process.
- Four sidebands variations for the fit of the *No-Track* EM clusters background: ]70, 80[ $\cup$ ]100, 110[, ]60, 80[ $\cup$ ]100, 120[, ]50, 80[ $\cup$ ]100, 130[ and ]55, 70[ $\cup$ ]110, 125[ GeV to ensure the stability of the analytic form of the third polynomial function.

Applying all of those variations results in 72 different measurements of the electron reconstruction efficiency. Additionally source of systematic uncertainty is coming from subtracting residual signal contaminations in the control regions using the MC simulation. This uncertainty is conservatively estimated by varying the MC prediction by 20% and 30% for electromagnetic clusters with and without associated tracks respectively. For the EM clusters with an associated track background evaluation uncertainty, the 20% variation covers the largest difference of *tight* efficiency between data and simulation and by estimating the signal contamination in all other regions from simulation (contributing by less than 7%). For the *No-Track* EM clusters background evaluation uncertainty, the 30%



Template		Variation 1	Variation 2
Inverted cuts		Fails at least 2 <i>cut-based Loose</i> cuts	
Isolation requirements	$E_T < 30$ GeV	$\frac{E_T^{\text{cone30}}}{p_T} > 0.02$	$\frac{E_T^{\text{cone30}}}{p_T} > 0.02$
	$E_T \geq 30$ GeV	$\frac{E_T^{\text{cone40}}}{p_T} > 0.05$	$\frac{E_T^{\text{cone40}}}{p_T} > 0.20$
Control region	$E_T < 30$ GeV	$120 < m_{ee} < 250$ GeV	$60 < m_{ee} < 70$ GeV
	$E_T \geq 30$ GeV	$120 < m_{ee} < 250$ GeV	$120 < m_{ee} < 250$ GeV

Table 5.3: *Description of the templates used to estimate the background for electrons reconstructed with an associated track. "m<sub>ee</sub>" refers to the invariant mass of the electron-cluster pair.*

variation is estimated from the largest observed difference between data and simulation for the probability of an electron is misidentified as a photon. The associated systematic uncertainty is considered fully correlated between all bins.

The total systematic uncertainty of a given efficiency measurement bin is estimated to be equal to the root mean square (RMS) of the measurements, including 72 variations and the additional conservative uncertainties, with the intention of modeling a 68% confidence interval. The central value per bin is taken to be the average value of the results from all the variations.

**Statistical uncertainty** The statistical uncertainty is taken to be the average of the statistical uncertainties over all variations of the electron reconstruction efficiency measurement. The statistical uncertainty on a single variation is given by the error propagation formula where independent variables are needed. Therefore, the central value of the reconstruction efficiency is measured using independent variables derived from equation (5.2):

$$\varepsilon_{\text{reco}} = \frac{U_{CR}(ST_{CR} - TS_{CR})}{U_{CR}(ST_{CR} - TS_{CR}) + T_{CR}(RU_{CR} - UR_{CR}) + \Gamma T_{CR}U_{CR}} \quad (5.3)$$

where:  $S$  is defined as the number of EM clusters candidates passing the track quality requirements after background subtraction ( $N^{\text{Quality-Track}} - B^{\text{Quality-Track}}$ ).  $R = N^{\text{Quality-Track}} - B^{\text{Quality-Track}}$  is the number of EM clusters candidates with associated track but failing the track quality requirements after background subtraction.  $\Gamma = N^{\text{No-Track}} - B^{\text{No-Track}}$  is the number of *No-Track* EM clusters candidates after background subtraction. The index  $\bullet_{CR}$  stands for the number of candidates in the control region.

Then, the statistical uncertainty is derived as:

$$\begin{aligned} \Delta\varepsilon^2 = & \frac{T_{CR}^2 (ST_{CR} - TS_{CR})^2}{(\text{denom})^4} [U^2 R_{CR}^2 \Delta U_{CR}^2 + U_{CR}^2 (U_{CR}^2 \Delta R^2 + U_{CR}^2 \Delta \Gamma^2 + U^2 \Delta R_{CR}^2 + R_{CR}^2 \Delta U^2)] \\ & + \frac{U_{CR}^2 (RU_{CR} - UR_{CR} + \Gamma U_{CR})^2}{(\text{denom})^4} [T^2 S_{CR}^2 \Delta T_{CR}^2 + T_{CR}^2 (T_{CR}^2 \Delta S^2 + T^2 \Delta S_{CR}^2 + S_{CR}^2 \Delta T^2)] \end{aligned}$$

Where "denom" stands for the denominator of Equation (5.3).  $\Delta X = \sqrt{X}$  is the Poisson uncertainty of the related variable  $X$  except  $\Delta \Gamma$  which is more complicated because of the term  $B^{\text{No-Track}}$  estimated from the third polynomial fit.  $\Delta B^{\text{No-Track}}$  is obtained by the error

propagation on the number of events in each bin  $n_i$  used for the fit. It reads at leading order:

$$(\Delta B^{\text{No-Track}})^2 = \sum_i \left( \frac{\partial B^{\text{No-Track}}}{\partial n_i} \right)^2 (\Delta n_i^{\text{(stat)}})^2$$

Where  $\Delta n_i^{\text{(stat)}}$  is the Poisson uncertainty on the number of events in the  $i^{\text{th}}$  bin.

## 5.3 Dataset and simulation

The reconstruction efficiency is measured using a data sample with an integrated luminosity of  $3.2 \text{ fb}^{-1}$  collected with the ATLAS experiment during the 2015 data. Only data in which all relevant detector components were fully functional are used. The spacing of subsequent LHC bunches is 25 ns.

The  $Z \rightarrow ee$  samples, used for comparison with the MC simulation, are generated using POWHEG [103, 104, 105] interfaced with PYTHIA 8 [106, 107] for parton showering. They are processed through the full ATLAS detector simulation [108] based on GEANT [109]. The simulation also includes realistic modeling of the event pile-up. Multiple overlaid proton-proton collisions are simulated with the soft QCD processes of PYTHIA 8 using tune A2 [110] and the MSTW2008LO PDF.

## 5.4 Results

### 5.4.1 Background level at denominator and numerator

Tables 5.4 and 5.5 display the estimated fraction of background in the denominator and the numerator of 3 different tag variations. The default selection indicates the *Tight* tag variation. As expected, the level of background in the loose selection (*Medium* tag variation) is higher than for the tight selection (*Tight* and isolated tag variation). The level of background is very high in the low  $E_T$  region: it even approaches 85% in some  $\eta$  bins in the denominator and 89% in the numerator. It decreases to 8% and 11% at high  $p_T$  in the denominator and the numerator respectively. The *No-Track* EM clusters background is estimated in the denominator only. It dominates at low  $E_T$ . It could even reach 100 % at some  $\eta$  bins as seen in Table 5.6. The few negative values at high  $E_T$  in the table caused by the background subtraction is due to the inaccuracy of the fitting at low statistics. This has negligible effect since the fraction of photons probes before background subtraction is of order 1% at high  $E_T$  as seen in Table 5.7.

The background components with the methods described above are depicted in Figure 5.3. The background templates are normalised in this case to the high mass control region. As demonstrated by the Figure, the measured data agrees well with the expectation, and the background subtraction procedures performs well. Photons that are mis-reconstructed as electrons are shown by the peak of the black dashed line at high  $E_T$  (the right of the Figure). At low  $E_T$ , clusters with no track are mainly background (grey dashed line spanning from 60 to 130 GeV).

More invariant mass distributions at high and low  $E_T$  for different  $\eta$  bins, together with the background estimation performed for both the numerator and the denominator, are shown in Appendix A. A good background estimation is observed for the numerator, where the expectation well matches the total background plus signal in the whole  $\eta$  range at low and high  $E_T$ . In the denominator, where two other background terms are additionally estimated (*No-Quality* and *No-Track* EM clusters background), the electron

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
(a) [-2.47,-2.37]	66.2	45.2	23.5	13.0	7.9	4.3	4.7	8.9	10.0	10.8
(a) [-2.37,-2.01]	68.4	45.4	25.3	14.1	7.6	4.4	4.4	8.6	12.3	13.4
(a) [-2.01,-1.81]	70.2	47.1	27.4	14.6	7.6	4.4	4.6	8.9	12.5	15.9
(a) [-1.81,-1.52]	74.9	52.9	30.5	15.6	8.2	4.8	5.0	8.6	13.4	15.6
(a) [-1.52,-1.37]	85.6	57.7	32.7	17.0	9.8	6.1	6.7	9.7	15.4	18.4
(a) [-1.37,-1.15]	76.0	54.2	31.9	16.4	7.2	4.5	5.1	8.7	12.8	17.0
(a) [-1.15,-0.80]	79.0	57.9	32.7	15.8	7.4	4.5	5.0	9.1	14.1	17.3
(a) [-0.80,-0.60]	78.6	59.6	34.4	15.8	7.9	4.9	5.1	9.2	14.4	17.9
(a) [-0.60,-0.10]	76.3	56.3	32.0	15.7	8.0	4.6	4.6	8.7	13.6	15.8
(a) [-0.10,0.00]	77.7	54.5	29.1	16.7	8.4	4.2	4.1	8.5	14.2	17.6
(a) [0.00,0.10]	77.1	54.1	30.8	16.4	8.4	4.7	4.4	8.7	13.5	16.5
(a) [0.10,0.60]	77.4	55.2	32.5	16.3	7.8	4.3	4.7	8.4	14.0	18.6
(a) [0.60,0.80]	75.9	60.0	34.5	15.6	7.9	4.6	4.7	9.3	14.7	17.6
(a) [0.80,1.15]	79.2	57.7	33.0	15.8	7.3	4.4	4.8	9.1	14.5	17.4
(a) [1.15,1.37]	77.1	54.0	31.9	15.8	7.5	4.3	5.3	8.7	15.8	17.3
(a) [1.37,1.52]	87.1	57.9	34.5	16.5	9.9	6.3	6.7	9.3	15.7	20.0
(a) [1.52,1.81]	74.2	53.4	30.8	16.1	8.0	4.9	4.8	8.4	14.6	14.4
(a) [1.81,2.01]	68.5	47.1	27.0	14.6	8.1	4.2	5.0	8.5	13.0	11.4
(a) [2.01,2.37]	67.5	45.2	25.7	13.9	7.5	4.7	5.0	8.2	12.7	12.5
(a) [2.37,2.47]	68.5	42.5	24.2	12.8	7.3	5.2	5.0	10.9	12.2	12.8
(b) [-2.47,-2.37]	69.6	47.6	25.5	14.2	8.9	4.8	5.2	9.6	11.2	13.7
(b) [-2.37,-2.01]	71.8	48.6	27.6	15.6	8.4	5.1	5.0	9.6	14.2	16.5
(b) [-2.01,-1.81]	73.0	50.2	30.1	16.6	8.7	5.1	5.0	9.9	14.6	20.7
(b) [-1.81,-1.52]	76.3	54.2	32.6	16.8	9.2	5.8	5.8	9.9	15.4	19.5
(b) [-1.52,-1.37]	85.0	59.5	34.7	18.4	11.4	6.9	7.1	10.9	17.9	24.5
(b) [-1.37,-1.15]	78.8	58.6	34.7	18.5	8.6	5.3	6.0	10.0	15.6	20.1
(b) [-1.15,-0.80]	81.8	61.7	36.3	18.1	8.8	5.3	6.0	10.7	17.0	22.1
(b) [-0.80,-0.60]	81.2	63.7	37.9	18.3	9.2	5.7	6.0	10.8	17.0	23.1
(b) [-0.60,-0.10]	79.0	60.2	35.4	18.0	9.4	5.4	5.6	10.2	16.0	21.0
(b) [-0.10,0.00]	80.0	58.1	32.6	19.3	10.0	5.1	4.9	10.3	16.6	21.3
(b) [0.00,0.10]	79.1	57.2	33.8	18.6	10.1	5.6	5.7	9.7	16.7	20.0
(b) [0.10,0.60]	80.2	59.4	36.0	18.6	9.3	5.2	5.7	9.9	16.5	23.1
(b) [0.60,0.80]	79.4	64.6	38.3	17.9	9.2	5.5	5.6	11.0	17.5	22.9
(b) [0.80,1.15]	81.5	61.5	36.4	18.1	8.7	5.1	5.7	10.5	17.1	22.0
(b) [1.15,1.37]	79.8	57.2	34.9	17.9	8.8	5.2	6.1	10.1	17.8	22.5
(b) [1.37,1.52]	89.1	59.3	36.3	18.3	10.9	7.0	7.4	9.9	17.2	26.2
(b) [1.52,1.81]	75.7	54.7	32.5	17.5	8.8	5.5	5.4	9.1	16.7	18.5
(b) [1.81,2.01]	71.5	50.1	29.1	16.1	9.1	4.9	5.8	9.3	15.6	15.8
(b) [2.01,2.37]	70.9	48.8	27.7	15.3	8.5	5.1	5.6	9.1	15.6	16.2
(b) [2.37,2.47]	71.2	45.8	25.7	14.3	8.3	5.3	5.5	10.4	11.8	15.0
(c) [-2.47,-2.37]	50.6	28.8	13.0	6.6	4.4	2.2	2.6	6.0	7.3	12.5
(c) [-2.37,-2.01]	52.3	29.9	14.3	7.4	3.7	2.2	1.9	4.7	7.0	7.0
(c) [-2.01,-1.81]	56.5	32.2	16.5	8.3	4.0	2.2	2.3	5.0	7.5	9.3
(c) [-1.81,-1.52]	66.9	42.3	21.0	9.4	4.8	2.5	2.7	5.2	7.6	10.4
(c) [-1.52,-1.37]	78.4	48.2	25.2	10.9	5.7	3.3	4.0	7.1	10.1	13.9
(c) [-1.37,-1.15]	65.4	40.9	21.1	9.6	4.0	2.3	2.7	4.8	7.4	9.4
(c) [-1.15,-0.80]	68.8	44.5	21.7	9.0	4.0	2.2	2.5	4.8	7.8	10.1
(c) [-0.80,-0.60]	66.7	44.5	22.1	8.9	4.2	2.4	2.7	4.6	8.0	10.5
(c) [-0.60,-0.10]	63.9	41.4	20.0	8.8	4.2	2.1	2.2	4.5	7.5	8.6
(c) [-0.10,0.00]	66.0	39.0	17.6	9.9	4.5	2.1	2.1	4.7	7.9	9.7
(c) [0.00,0.10]	66.7	39.5	19.5	10.0	4.6	2.5	2.2	4.7	7.1	8.8
(c) [0.10,0.60]	65.1	40.1	20.2	9.1	4.0	2.0	2.3	4.1	7.4	10.8
(c) [0.60,0.80]	64.2	46.0	22.9	8.8	4.1	2.3	2.4	4.6	8.0	11.3
(c) [0.80,1.15]	67.8	44.1	21.5	9.2	3.8	2.2	2.4	4.7	8.3	9.8
(c) [1.15,1.37]	66.2	39.6	20.9	9.3	3.9	2.2	2.7	4.6	8.5	9.4
(c) [1.37,1.52]	78.6	44.8	25.1	10.3	5.8	3.2	4.0	6.0	10.5	15.9
(c) [1.52,1.81]	65.0	41.4	21.8	10.5	4.6	2.5	2.6	4.4	8.6	7.9
(c) [1.81,2.01]	55.8	31.9	16.6	7.8	4.0	1.8	2.4	4.5	7.3	6.8
(c) [2.01,2.37]	51.1	29.6	14.8	7.5	3.5	2.2	2.6	4.4	7.5	8.2
(c) [2.37,2.47]	52.5	28.2	13.9	7.1	3.8	2.9	2.8	6.7	7.9	10.5

Table 5.4: *Estimated background fraction (in %) at the denominator of the electron reconstruction efficiency for three selections: default (a), loose (b) and tight (c), in different bins in  $\eta$  (rows) and  $E_T$  (columns) of the probe cluster.*

background template normalised in the control region does not perfectly match the baseline at low mass range due to the high photon contribution at low  $E_T$  (see Figure A.1). The number of the probes failing the track quality requirements is checked to be mainly dominated by the background.

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
(a) [-2.47,-2.37]	65.1	43.2	23.1	12.5	7.9	4.1	4.1	8.4	10.9	10.8
(a) [-2.37,-2.01]	66.3	44.4	24.7	13.7	7.3	4.3	4.2	8.5	11.8	12.4
(a) [-2.01,-1.81]	67.9	46.4	26.7	13.9	7.4	4.2	4.6	8.7	11.5	15.7
(a) [-1.81,-1.52]	73.1	51.0	29.4	15.2	7.7	4.6	4.8	8.7	13.3	15.8
(a) [-1.52,-1.37]	85.7	57.4	31.6	16.3	9.2	5.7	6.1	9.5	15.2	19.5
(a) [-1.37,-1.15]	74.2	52.3	30.3	15.5	6.6	4.3	4.7	8.4	12.7	16.9
(a) [-1.15,-0.80]	76.3	54.9	30.7	14.4	6.7	4.1	4.5	8.7	13.7	17.3
(a) [-0.80,-0.60]	75.6	56.7	32.4	14.3	7.1	4.5	4.7	8.6	14.7	18.0
(a) [-0.60,-0.10]	74.1	53.3	29.6	14.1	7.2	4.1	4.2	8.0	13.2	15.3
(a) [-0.10,0.00]	74.3	50.5	27.2	14.5	7.3	3.7	3.8	7.8	13.5	17.4
(a) [0.00,0.10]	73.8	50.3	28.2	14.8	7.3	4.3	3.9	7.8	13.4	16.1
(a) [0.10,0.60]	74.5	52.6	29.9	14.4	7.2	3.9	4.2	7.8	13.3	17.8
(a) [0.60,0.80]	74.7	57.2	31.7	14.6	7.2	4.4	4.5	8.8	13.9	17.2
(a) [0.80,1.15]	77.4	54.7	31.2	15.0	6.8	4.1	4.6	8.4	13.6	16.7
(a) [1.15,1.37]	74.1	52.9	29.9	14.7	6.9	4.2	4.9	8.4	15.0	16.3
(a) [1.37,1.52]	86.3	56.8	33.0	16.4	9.0	6.2	6.2	8.7	16.5	19.9
(a) [1.52,1.81]	72.9	50.5	29.8	15.3	7.6	4.7	4.4	8.1	14.3	14.4
(a) [1.81,2.01]	67.2	46.8	26.0	14.0	7.4	4.0	4.8	8.2	12.4	10.6
(a) [2.01,2.37]	65.9	43.5	24.5	13.4	7.2	4.3	4.6	8.1	12.7	12.4
(a) [2.37,2.47]	67.3	42.2	23.5	12.2	6.8	4.9	4.8	10.0	12.2	13.8
(b) [-2.47,-2.37]	68.3	45.8	25.3	13.4	8.9	4.5	4.5	9.0	12.1	13.6
(b) [-2.37,-2.01]	69.8	47.4	26.9	15.2	8.1	5.0	4.9	9.5	13.7	15.6
(b) [-2.01,-1.81]	71.3	49.5	29.3	15.7	8.4	4.9	4.9	9.7	13.6	20.3
(b) [-1.81,-1.52]	74.9	52.6	31.4	16.3	8.6	5.5	5.6	10.0	15.2	19.6
(b) [-1.52,-1.37]	84.9	58.8	33.5	17.7	10.6	6.3	6.5	10.6	18.2	25.4
(b) [-1.37,-1.15]	77.1	56.2	32.9	17.5	8.0	5.0	5.6	9.7	15.2	19.9
(b) [-1.15,-0.80]	79.2	58.5	33.9	16.5	7.9	4.8	5.3	10.2	16.6	22.0
(b) [-0.80,-0.60]	78.4	60.7	35.8	16.6	8.3	5.2	5.5	10.2	17.2	23.3
(b) [-0.60,-0.10]	76.9	56.9	32.7	16.2	8.5	4.9	5.2	9.5	15.4	20.4
(b) [-0.10,0.00]	76.8	53.8	30.5	16.7	8.5	4.6	4.6	9.5	16.1	21.1
(b) [0.00,0.10]	75.8	53.5	30.8	16.7	8.9	5.0	5.1	9.0	16.4	19.8
(b) [0.10,0.60]	77.4	56.7	33.1	16.6	8.5	4.8	5.1	9.3	15.7	22.2
(b) [0.60,0.80]	77.7	61.5	35.2	16.6	8.4	5.2	5.3	10.4	16.6	22.4
(b) [0.80,1.15]	79.6	58.7	34.3	17.0	8.0	4.9	5.5	9.8	16.2	21.0
(b) [1.15,1.37]	77.2	56.1	32.6	16.6	8.1	5.0	5.7	9.6	17.0	21.4
(b) [1.37,1.52]	88.6	58.2	34.6	18.1	10.1	7.1	7.0	9.5	17.8	25.7
(b) [1.52,1.81]	74.3	52.3	31.1	16.6	8.4	5.2	5.1	8.8	16.4	18.8
(b) [1.81,2.01]	69.9	49.3	27.8	15.2	8.3	4.8	5.6	9.1	14.7	14.9
(b) [2.01,2.37]	69.4	47.0	26.5	14.6	8.2	4.8	5.2	8.9	15.4	16.0
(b) [2.37,2.47]	70.2	44.8	24.8	13.5	7.6	5.0	5.5	9.7	12.0	16.0
(c) [-2.47,-2.37]	49.0	27.1	12.8	6.2	4.4	2.1	2.2	5.6	8.1	12.5
(c) [-2.37,-2.01]	50.0	29.1	13.9	7.2	3.6	2.1	1.8	4.6	6.6	6.5
(c) [-2.01,-1.81]	54.4	31.6	16.1	7.8	3.9	2.1	2.2	4.9	6.9	9.2
(c) [-1.81,-1.52]	65.1	40.8	20.0	9.1	4.4	2.3	2.6	5.3	7.6	10.4
(c) [-1.52,-1.37]	77.6	47.6	24.1	10.4	5.1	3.0	3.5	6.8	10.3	14.5
(c) [-1.37,-1.15]	63.0	38.7	19.9	9.0	3.6	2.2	2.5	4.7	7.2	9.4
(c) [-1.15,-0.80]	65.1	41.3	19.9	8.0	3.5	2.0	2.2	4.6	7.6	10.2
(c) [-0.80,-0.60]	62.6	41.2	20.7	8.0	3.7	2.2	2.5	4.3	8.1	10.7
(c) [-0.60,-0.10]	60.9	38.2	18.1	7.8	3.8	1.9	2.0	4.2	7.2	8.3
(c) [-0.10,0.00]	61.3	35.0	16.3	8.3	3.7	1.9	2.0	4.3	7.7	9.7
(c) [0.00,0.10]	62.9	36.1	17.5	9.0	4.0	2.2	1.9	4.3	7.0	8.7
(c) [0.10,0.60]	61.4	37.4	18.3	7.9	3.7	1.9	2.0	3.8	7.0	10.3
(c) [0.60,0.80]	62.2	42.9	20.6	8.2	3.7	2.2	2.3	4.4	7.6	11.0
(c) [0.80,1.15]	65.1	41.3	20.0	8.6	3.5	2.1	2.4	4.4	7.8	9.2
(c) [1.15,1.37]	63.0	38.8	19.3	8.6	3.5	2.1	2.5	4.4	8.1	8.7
(c) [1.37,1.52]	76.6	43.4	23.4	10.3	5.2	3.3	3.7	5.7	10.9	15.6
(c) [1.52,1.81]	63.0	39.1	20.9	9.9	4.4	2.3	2.4	4.3	8.4	8.1
(c) [1.81,2.01]	54.0	31.4	15.8	7.3	3.6	1.8	2.2	4.4	6.8	6.2
(c) [2.01,2.37]	49.6	28.0	14.1	7.2	3.4	2.0	2.4	4.3	7.5	8.1
(c) [2.37,2.47]	50.9	27.2	13.4	6.6	3.4	2.7	2.8	6.2	8.2	11.4

Table 5.5: *Estimated background fraction (in %) at the numerator of the electron reconstruction efficiency for three selections: default (a), loose (b) and tight (c) in different bins in  $\eta$  (rows) and  $E_T$  (columns) of the probe electron.*

## 5.4.2 Efficiency

The efficiency to reconstruct an electron (or positron) having a track of good quality varies from 95% to 99% going from the end-cap to the barrel region for both low  $E_T$  and high

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
(a) [-2.47,-2.37]	74.0	88.5	30.2	33.0	12.8	17.0	35.9	36.4	-52.9	-80.5
(a) [-2.37,-2.01]	93.6	63.6	41.0	22.2	19.1	11.4	13.2	7.8	55.9	98.9
(a) [-2.01,-1.81]	92.9	55.9	38.4	33.6	12.7	11.2	4.8	15.0	70.4	34.4
(a) [-1.81,-1.52]	90.5	79.0	53.1	22.8	26.9	13.5	14.8	2.4	9.8	4.4
(a) [-1.52,-1.37]	79.7	55.9	45.0	25.1	18.7	15.2	18.7	13.8	18.0	-13.4
(a) [-1.37,-1.15]	89.8	80.6	65.8	41.8	31.5	18.7	25.5	23.5	17.0	28.1
(a) [-1.15,-0.80]	99.4	100.0	73.0	61.3	39.0	29.4	37.9	27.9	35.5	19.8
(a) [-0.80,-0.60]	100.0	97.8	74.4	76.9	48.5	30.6	36.0	50.6	-12.2	6.3
(a) [-0.60,-0.10]	90.3	90.8	82.3	72.3	49.6	38.7	33.1	55.0	44.3	60.8
(a) [-0.10,0.00]	100.0	100.0	66.7	100.0	66.5	42.6	25.6	49.9	54.9	23.0
(a) [0.00,0.10]	100.0	100.0	83.4	65.7	58.3	30.3	42.5	62.5	18.5	34.8
(a) [0.10,0.60]	97.2	83.7	84.0	82.3	38.5	32.1	51.8	52.0	70.4	83.5
(a) [0.60,0.80]	83.6	94.4	96.4	50.5	50.7	20.3	23.6	44.2	84.7	57.5
(a) [0.80,1.15]	92.0	100.0	74.9	41.4	31.9	23.0	18.8	54.6	70.1	70.7
(a) [1.15,1.37]	99.9	67.0	80.4	46.0	35.5	9.7	30.8	28.7	64.7	100.0
(a) [1.37,1.52]	96.0	65.0	48.4	15.9	22.4	8.0	17.8	25.0	-3.2	21.8
(a) [1.52,1.81]	83.3	99.1	49.4	35.0	18.6	12.3	19.7	27.2	21.5	8.4
(a) [1.81,2.01]	78.4	47.4	47.0	30.5	30.7	10.4	16.4	20.6	42.3	100.0
(a) [2.01,2.37]	84.1	77.2	58.0	34.9	23.7	25.7	27.0	12.0	13.6	14.8
(a) [2.37,2.47]	79.3	33.2	45.2	30.8	32.3	22.0	23.5	65.2	-0.4	-86.1
	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
(b) [-2.47,-2.37]	79.0	87.8	28.3	44.3	10.9	25.9	43.5	38.8	-47.7	-108.6
(b) [-2.37,-2.01]	92.7	69.2	44.1	26.2	18.3	12.7	15.7	16.2	57.7	81.2
(b) [-2.01,-1.81]	86.0	58.5	42.7	39.9	16.9	14.8	7.4	16.1	60.1	49.8
(b) [-1.81,-1.52]	87.7	76.1	56.8	27.2	29.7	15.8	16.8	0.5	21.5	18.5
(b) [-1.52,-1.37]	81.3	62.0	47.5	25.6	24.0	17.7	20.2	19.1	7.0	-1.0
(b) [-1.37,-1.15]	91.8	89.6	68.6	47.2	34.1	19.7	30.3	26.8	34.8	33.4
(b) [-1.15,-0.80]	100.0	100.0	79.2	65.2	44.7	31.5	44.0	37.7	40.7	23.9
(b) [-0.80,-0.60]	100.0	99.6	76.4	73.5	56.1	36.6	41.0	48.9	-2.8	3.6
(b) [-0.60,-0.10]	92.3	95.0	85.6	74.4	56.0	42.0	35.6	58.7	52.7	68.1
(b) [-0.10,0.00]	100.0	100.0	72.3	100.0	73.7	43.3	21.3	57.4	40.3	31.2
(b) [0.00,0.10]	100.0	99.8	86.9	70.8	63.0	41.1	50.0	51.7	38.0	29.5
(b) [0.10,0.60]	98.3	87.7	86.4	80.4	47.7	34.3	54.9	52.3	76.0	93.2
(b) [0.60,0.80]	90.0	98.8	94.5	54.9	51.5	26.1	28.2	46.4	78.2	70.8
(b) [0.80,1.15]	93.7	96.6	79.9	50.3	36.6	22.0	19.9	57.3	68.0	84.5
(b) [1.15,1.37]	99.1	68.4	82.8	50.3	39.7	14.9	30.9	33.5	62.0	97.4
(b) [1.37,1.52]	94.1	65.7	52.7	19.3	21.8	4.4	16.7	20.9	-0.7	37.7
(b) [1.52,1.81]	86.0	90.7	57.7	38.0	22.4	14.0	22.2	22.2	30.6	-11.1
(b) [1.81,2.01]	85.3	59.7	56.0	36.4	31.8	9.7	18.3	13.1	57.2	100.0
(b) [2.01,2.37]	85.0	83.3	58.2	42.2	20.5	25.2	28.7	14.1	27.2	25.7
(b) [2.37,2.47]	79.3	54.9	45.4	42.8	41.6	27.6	5.5	54.5	-22.6	-143.7
	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
(c) [-2.47,-2.37]	70.3	69.5	17.3	28.7	6.5	13.4	27.8	31.3	-44.1	-32.8
(c) [-2.37,-2.01]	81.3	46.8	24.5	11.8	8.6	5.4	7.2	8.8	39.9	46.5
(c) [-2.01,-1.81]	76.3	38.9	22.7	25.0	7.9	6.6	2.4	8.2	40.9	21.7
(c) [-1.81,-1.52]	82.5	61.5	39.5	16.0	17.4	7.2	8.0	-3.5	10.9	9.1
(c) [-1.52,-1.37]	79.9	51.0	40.4	18.0	15.9	9.9	13.8	14.6	2.0	-6.9
(c) [-1.37,-1.15]	88.2	75.4	54.2	31.2	20.2	9.0	17.6	10.5	16.3	11.0
(c) [-1.15,-0.80]	100.0	99.5	67.5	49.7	28.1	16.3	26.0	21.0	20.5	1.3
(c) [-0.80,-0.60]	100.0	99.2	64.6	57.5	38.0	20.1	22.0	26.5	-15.7	-12.1
(c) [-0.60,-0.10]	88.3	90.0	75.7	57.8	37.2	22.9	18.3	36.9	36.9	42.3
(c) [-0.10,0.00]	100.0	100.0	52.5	84.9	53.1	25.1	6.9	37.7	19.5	15.5
(c) [0.00,0.10]	100.0	89.2	69.6	56.5	44.5	23.3	28.8	35.0	12.2	13.4
(c) [0.10,0.60]	94.8	80.2	74.0	67.0	28.7	16.8	34.4	32.3	52.7	76.9
(c) [0.60,0.80]	78.0	98.0	91.0	34.3	31.8	13.7	14.1	25.7	51.0	70.5
(c) [0.80,1.15]	89.9	91.0	71.1	33.7	20.6	9.7	8.1	36.7	49.5	69.5
(c) [1.15,1.37]	92.3	48.8	69.8	34.0	22.9	5.4	17.3	17.3	39.7	80.8
(c) [1.37,1.52]	100.0	58.1	43.2	10.7	13.5	1.3	11.2	16.6	-2.6	24.8
(c) [1.52,1.81]	81.5	80.1	42.8	26.2	12.2	6.3	11.4	10.7	16.0	-17.4
(c) [1.81,2.01]	73.0	35.4	35.9	20.0	17.6	3.9	10.0	6.1	33.2	76.7
(c) [2.01,2.37]	66.5	69.2	37.8	24.1	9.5	13.4	16.0	8.0	10.7	9.2
(c) [2.37,2.47]	67.9	39.6	35.1	28.9	27.3	16.3	5.1	38.9	-32.7	-

Table 5.6: Estimated background fraction (in %) in the No-Track EM cluster term for three selections: default (a), loose (b) and tight (c), in different bins in  $\eta$  (rows) and  $p_T$  (columns) of the probe cluster.

$E_T$  probe. The efficiencies in data and MC are displayed in Figure 5.4 in terms of  $\eta$  in 10  $E_T$  bins. The efficiency in data varies from 97% to 99% between the end-cap and barrel regions for electrons with  $E_T > 15$  GeV. Lower efficiencies (95%) are obtained in the calorimeter transition region. For very energetic electrons ( $E_T > 80$  GeV) the efficiency

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
[-2.47,-2.37]	5.8	3.3	2.8	2.0	1.5	1.4	1.7	1.6	1.6	0.3
[-2.37,-2.01]	6.5	3.9	2.9	2.2	2.0	1.5	1.2	1.4	1.0	1.1
[-2.01,-1.81]	7.4	4.5	3.4	2.9	2.5	2.1	1.8	1.5	1.6	1.0
[-1.81,-1.52]	8.3	4.9	3.8	3.0	2.5	2.1	1.9	1.8	1.4	1.4
[-1.52,-1.37]	9.3	8.4	6.9	6.1	4.9	4.0	4.0	4.0	3.2	3.4
[-1.37,-1.15]	9.6	5.9	3.8	2.9	2.1	1.6	1.6	1.3	1.5	1.1
[-1.15,-0.80]	11.0	5.9	4.0	2.6	1.9	1.5	1.3	1.4	1.4	1.0
[-0.80,-0.60]	11.3	6.5	4.1	2.3	1.7	1.3	1.2	1.3	1.2	1.0
[-0.60,-0.10]	12.1	7.3	4.2	2.5	1.7	1.2	1.1	1.2	1.4	1.0
[-0.10,0.00]	13.0	7.7	4.5	2.5	1.9	1.3	1.3	1.4	1.7	1.5
[0.00,0.10]	11.8	7.3	4.4	2.9	1.9	1.4	1.3	1.4	1.2	1.9
[0.10,0.60]	12.0	7.4	4.5	2.5	1.7	1.3	1.0	1.2	1.1	1.1
[0.60,0.80]	11.4	6.8	4.1	2.8	1.7	1.3	1.1	1.2	1.2	1.1
[0.80,1.15]	10.5	6.0	3.7	2.7	1.9	1.4	1.3	1.3	1.4	1.3
[1.15,1.37]	9.5	5.9	3.6	3.0	2.0	1.7	1.5	1.7	1.4	1.1
[1.37,1.52]	9.5	7.9	7.1	5.8	5.0	4.4	4.3	3.6	4.0	3.3
[1.52,1.81]	8.6	5.3	3.8	3.0	2.5	2.2	1.9	1.6	1.8	1.4
[1.81,2.01]	7.7	4.7	3.8	3.2	2.9	2.3	1.9	1.9	1.6	0.7
[2.01,2.37]	6.6	4.0	2.9	2.0	1.8	1.5	1.4	1.5	1.2	1.0
[2.37,2.47]	5.6	3.7	2.5	1.9	1.7	1.5	1.3	1.4	1.2	0.9

Table 5.7: *Fraction of the No-Track EM clusters at the denominator (in %), before background subtraction, in different bins in  $\eta$  (rows) and  $E_T$  (columns) of the probe cluster.*

is  $\sim 99\%$  over the whole  $\eta$  range. For each point, the uncertainties are statistical (inner bars) and statistical + systematic (outer bars). Statistical errors are dominating almost in all bins, where 0.1- 1.3% (0.0-0.6%) statistical (systematic) errors are estimated at  $E_T > 25$  GeV and 0.3- 2.7% (0.1-3.7)% at lower  $E_T$ . The summary of those results are shown in Figure 5.5 inclusively in  $\eta$  (left) and  $E_T$  (right).

The reconstruction efficiency is comparable to the efficiency observed in data taken in 2012 at 8 TeV, except for the calorimeter transition region ( $1.37 < |\eta| < 1.52$ ) where the efficiency degrades by 2%. This difference is further investigated by looking at every term of the reconstruction efficiency. A higher number of mis-reconstructed electrons as photons (photon term in the denominator), by O(2%) in the crack bins compared to other eta-bins, is seen in both data and MC as shown in Figure 5.6 where the reconstruction efficiency is compared with the electron efficiency excluding the photon term. From those plots, one can see that the unexpected drop of efficiency appears when we include the photon term where more electrons in the crack region are reconstructed as photons. This could be explained by the fact that the track to the middle of the second layer of the EM calo extrapolation has changed, comparing to Run-1, for electrons which have energies in both barrel and end-cap. This hypothesis is confirmed by checking both the single electron sample and the  $Z \rightarrow ee$  MC where few percents inefficiency loss in the crack is also seen when a true electron is matched to a reconstructed electron with  $\Delta R < 0.1$  with  $E_T > 7$  GeV (Figure 5.7). Therefore, the drop in efficiency is found to be caused by inefficiencies in the matching of the track to the calorimeter section (barrel/end-cap). Thanks to this study, this issue has been fixed in the next release.

### 5.4.3 Scale Factors

Electron reconstruction scale factors are calculated by taking the ratio MC to data of the determined efficiencies:

$$\text{SF} = \frac{\varepsilon_{\text{reco}}(\text{data})}{\varepsilon_{\text{reco}}(\text{MC})} \quad (5.4)$$

The measurements of the electron reconstruction scale factors with their uncertainties in 200 bins in  $(E_T, \eta)$  are summarised in Figure 5.8. They are determined for electrons from  $Z$  that are separated from jets fulfilling  $\Delta R(\text{electron}, \text{jet}) > 0.4$ . The SF are in

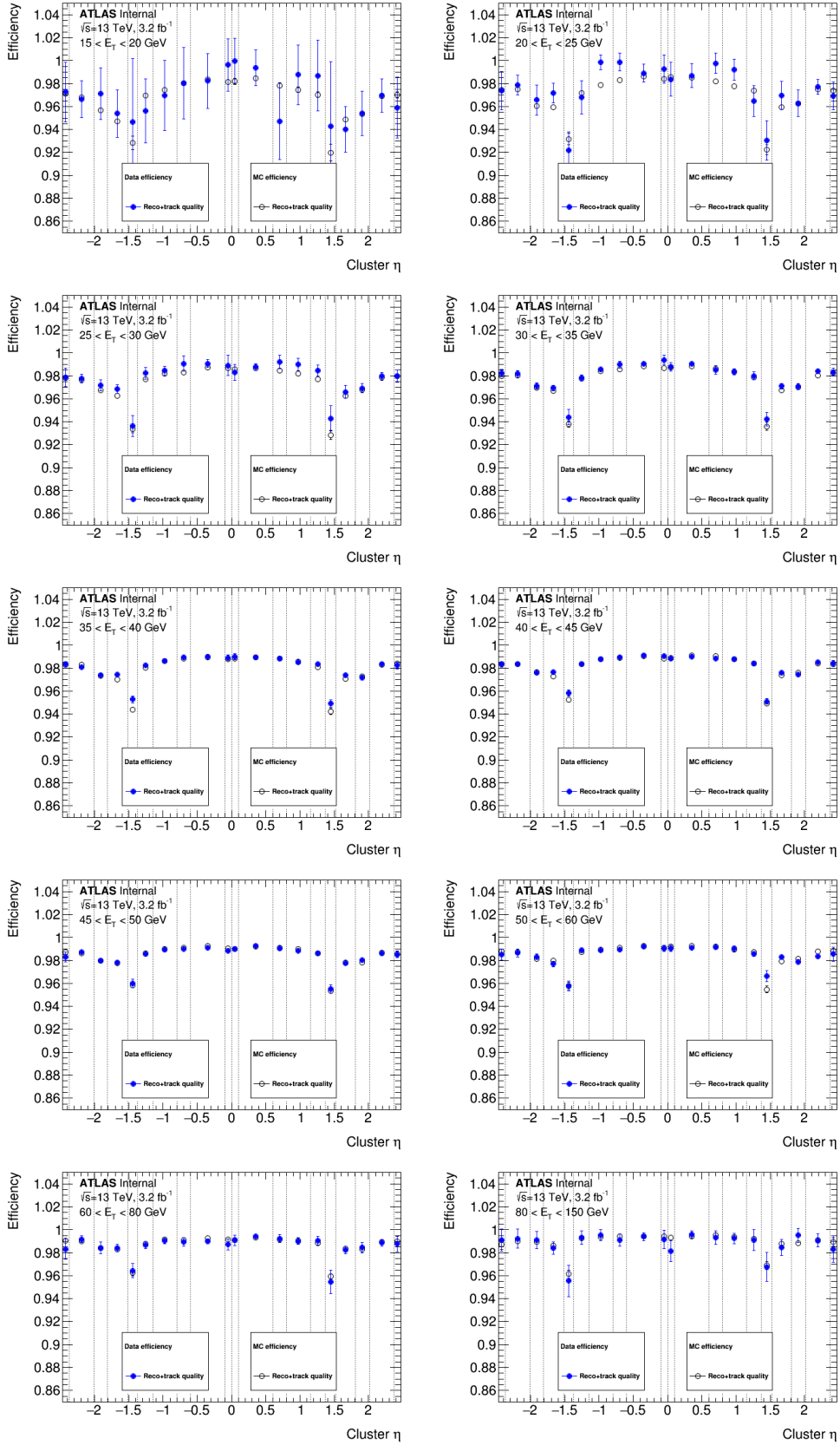


Figure 5.4: Measured electron reconstruction efficiencies as a function of  $\eta$  for different  $E_T$  bins for the 2015 datasets. The vertical lines indicate the bins in which the efficiencies are calculated. the shown uncertainties are statistical plus systematic.



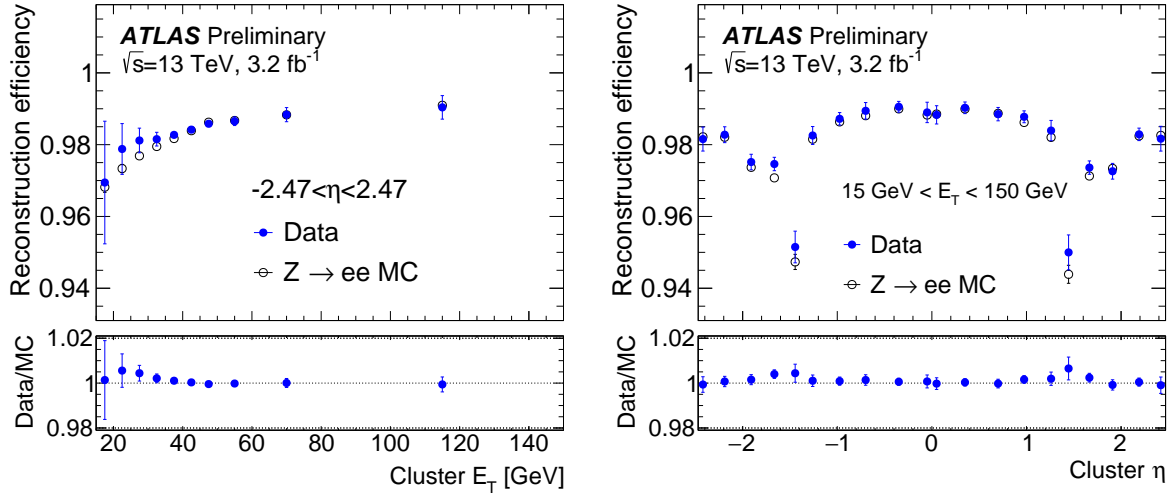


Figure 5.5: Summary of electron reconstruction efficiency measurements in  $Z \rightarrow ee$  events in terms of transverse energy  $E_T$  (left) and in terms of pseudo-rapidity  $\eta$  (right) for 2015 dataset (blue full dots) and MC (open dots). The ratio of data to MC efficiencies is shown in the bottom of each plot for comparison. The error bars include both statistical and systematic uncertainties.

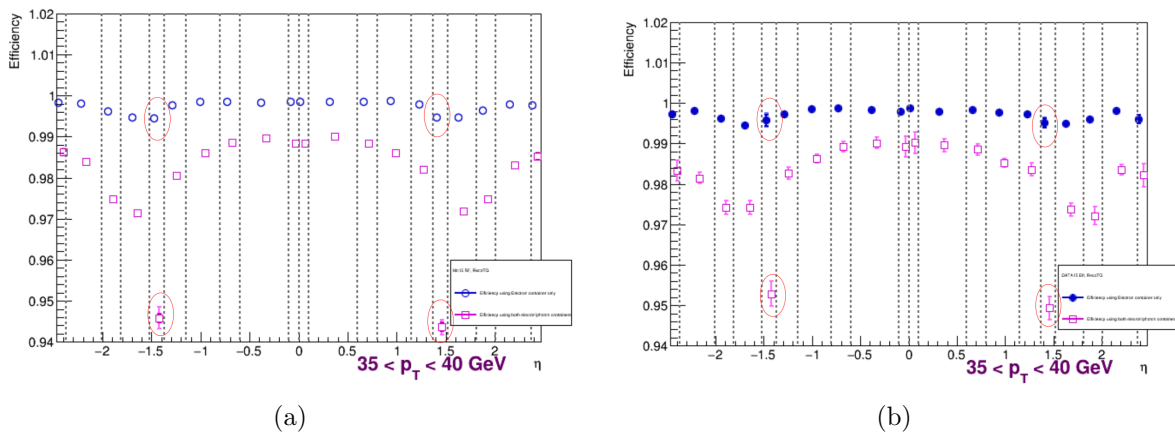


Figure 5.6: 13 TeV electron reconstruction efficiency excluding/including the photon term of the denominator (blue/magenta) in (a) simulated  $Z \rightarrow ee$  process and (b)  $3.2 \text{ fb}^{-1}$  data. The error bars include both statistical and systematic uncertainties.



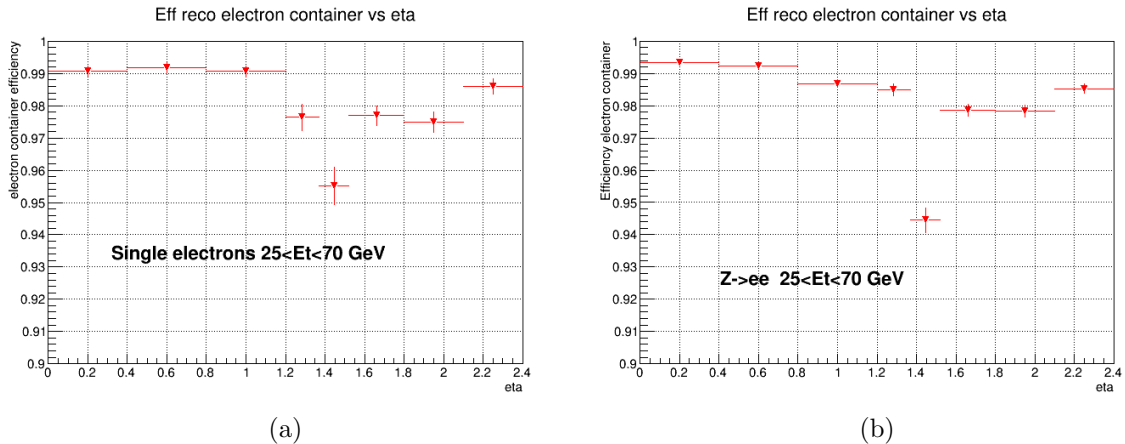


Figure 5.7: Matching a true electron to a reconstructed electron with  $\Delta R < 0.1$  with  $E_T > 7$  GeV using (a) single electron and (b)  $Z \rightarrow ee$  samples.

general very close to 1. The uncertainties, shared between statistics and systematics, are below 0.8% for electrons probes with  $E_T$  inside the [25,150] GeV interval. For the lowest  $E_T$  probe ( $< 25$  GeV), the SF uncertainty is higher (0.2-4.5%). The statistical uncertainty dominates in most of  $\eta$  bins. The good agreement between the reconstruction results obtained in data and in MC gives confidence in the MC description of the detector response for electrons with  $E_T < 15$  GeV. Therefore, the reconstruction SFs for  $E_T < 15$  are assumed to be one with an uncertainty of 2% in the barrel and 5% in the endcap region.

#### 5.4.4 Combination

The reconstruction efficiency is combined with the identification efficiency in order to provide the final official scale factors. The same procedure is followed to measure the identification efficiency using the T&P method with two independent and complementary samples  $Z \rightarrow ee$  at  $E_T > 15$  GeV and  $J/\Psi \rightarrow ee$  at  $7 < E_T < 20$  GeV. Four methods to estimate the electron background are used; in  $Z \rightarrow ee$  sample with background subtraction techniques based on a combined invariant mass-calorimetric isolation templates and in  $J/\Psi \rightarrow ee$  sample with combined mass-lifetime fit to separate prompt-non prompt electrons and polynomial fit in CR to subtract the background. They are ultimately combined together to give the final results. The sum of the background contribution and the MC signal, which are binned in  $E_T$  and  $\eta$ , describes well the signal region for all methods.

Three identification working points are used, *Tight*, *Medium* and *Loose* LH with the impact parameter requirements ( $|d_0|/\sigma < 5$  and  $Z_0 \times \sin\theta < 0.5$  mm), to measure the identification efficiency. The combined reconstruction and identification efficiencies are shown in Figure 5.9 for both data and MC in terms of  $E_T$  and  $\eta$ . The efficiencies increase proportionally with  $E_T$  and varies from 55 to 95 % (the tighter working point, the lower the efficiency). The *tight* selection has the stronger dependency on  $\eta$ . The observed drop of the efficiencies in the center, in the crack region and in the edges of the EM calorimeter is understood. The corresponding SFs are shown in the bottom of the Figure. Few percents lower than one come mostly from the modeling of lateral shower shapes and are similar to Run1 results. The uncertainties associated to the measurements, shown in Figure 5.10, are obtained by using the correlation information of the  $E_T \times \eta$  measurement. The statistical uncertainties are dominating and the systematic uncertainty is mainly due

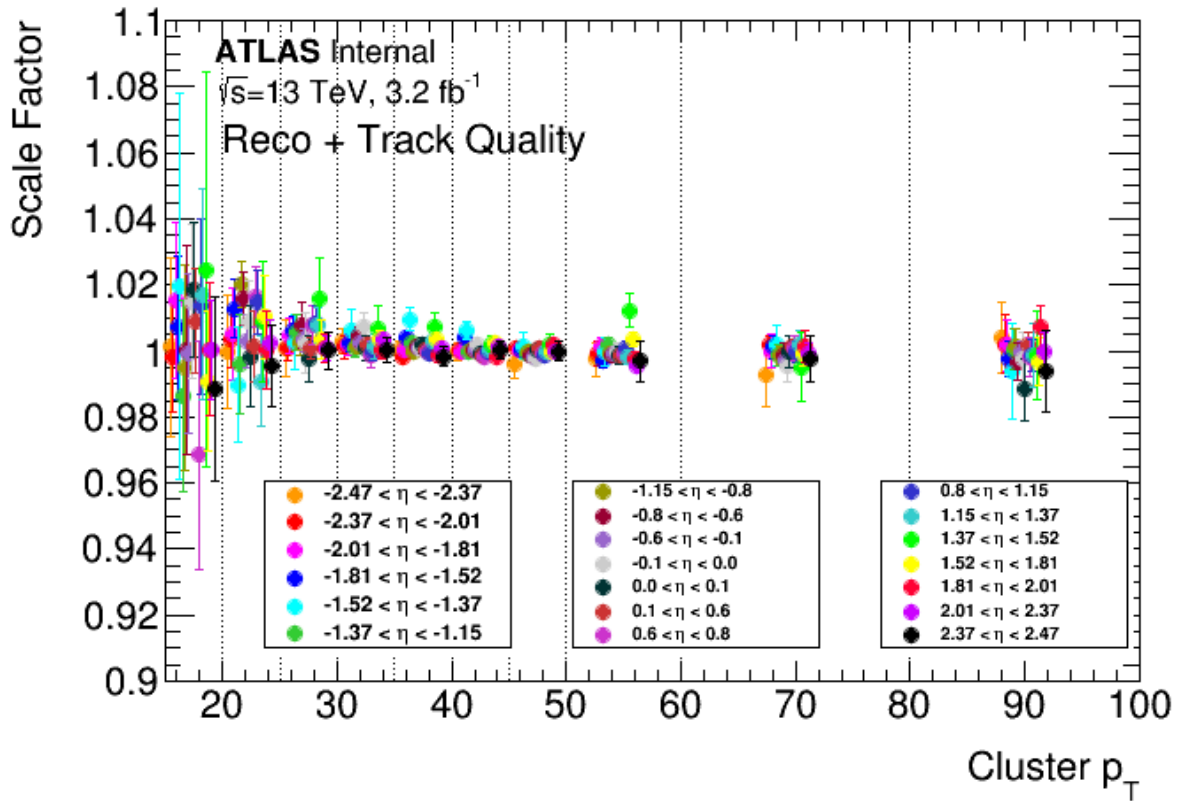


Figure 5.8: Summary of measured scale factors of reconstruction efficiency as a function of  $E_T$  and  $\eta$  for  $15 < E_T < 150 \text{ GeV}$  for the 2015 datasets. The vertical lines indicate the bins in which the efficiencies are calculated. The shown uncertainties are statistical + systematic. In a given  $E_T$  bin, the values corresponding to different  $\eta$  bins are shifted for a better reading.

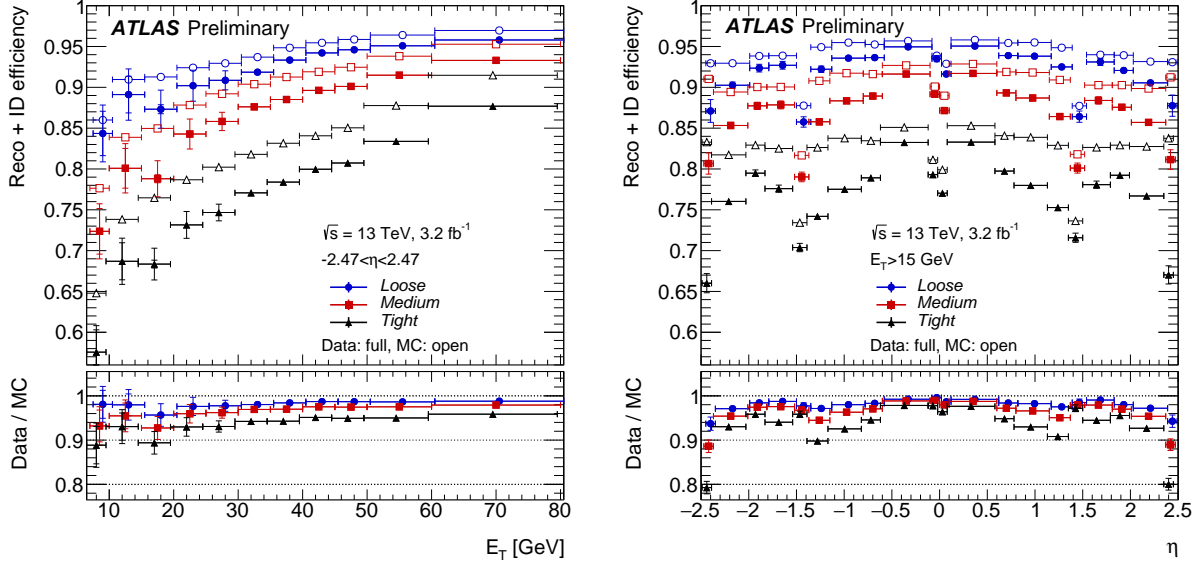


Figure 5.9: Combined electron reconstruction and identification efficiencies in terms of transverse energy  $E_T$  (left) and in terms of pseudo-rapidity  $\eta$  (right) for 2015 dataset (full dots) and MC (open dots)  $Z \rightarrow ee$  events. The measurements are performed for three identification working points: Loose, Medium and Tight LH with the impact parameter requirements. The ratio of data to MC efficiencies of each working point is shown in the bottom for comparison. The data efficiencies are derived from applying the measured reconstruction and combined identification scale factors on the  $Z \rightarrow ee$  MC prediction. The inner (outer) error bars show the statistical (total) uncertainties.

to the background subtraction procedure. The total uncertainty is less than 3.5 % for all WPs dominating at low  $E_T$  and in the edges of the EM calorimeter. For  $E_T > 30$  GeV, sub-percent level of typical errors is achieved.

## 5.5 Conclusion

Precise measurements of the electron reconstruction efficiency are performed with  $3.2 \text{ fb}^{-1}$  of data recorded by the ATLAS experiment in 2015 at a center-of-mass energy of  $\sqrt{s} = 13$  TeV. The results are obtained from the T&P method using  $Z \rightarrow ee$  sample. They are presented in  $(20 \times 10)$  ( $\eta \times E_T$ ) bins reflecting the dependency on the electron kinematics and the detector design. Overall, the reconstruction efficiencies vary from 95% to 99% with sub-percent typical measured errors for  $E_T > 25$  GeV and between 0.7-5 % for lower  $E_T$ . Data-MC scale factors are computed for electrons with  $E_T > 15$  GeV including the systematic uncertainties and used to obtain the results of the 2015 physics analyses. They are found to be very close to one. The good agreement between the reconstruction results obtained in data and in MC gives confidence in the MC description of the detector response for electrons with low energies ( $E_T < 15$  GeV).

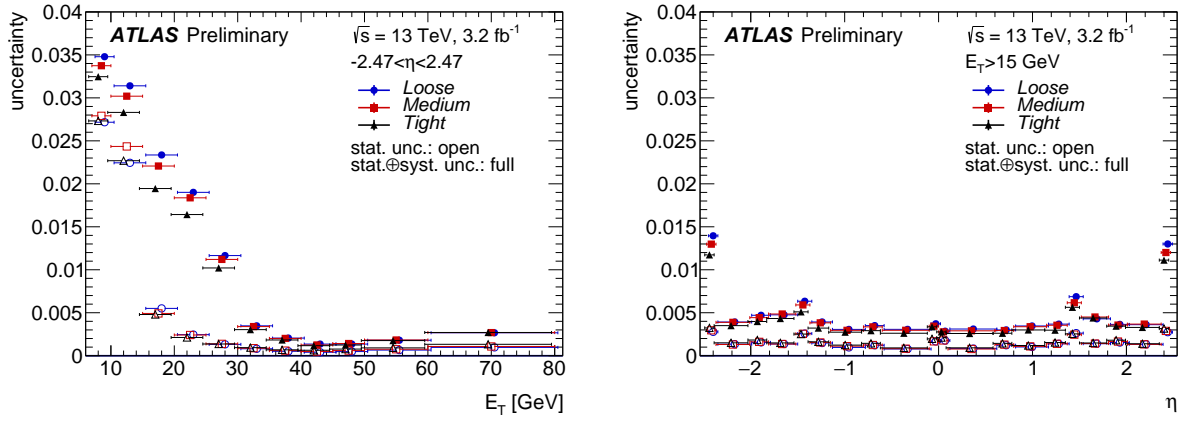


Figure 5.10: *Uncertainties on the combined electron reconstruction and identification efficiencies in terms of transverse energy  $E_T$  (left) and in terms of pseudo-rapidity  $\eta$  (right) using 2015 dataset  $Z \rightarrow ee$  events. The measurements are performed by using the correlation information of the  $E_T \times \eta$  measurement for three identification working points: Loose, Medium and Tight LH with the impact parameter requirements. Two sets of uncertainties are shown: the statistical uncertainty (open dots) and the combined statistical and systematic uncertainty (full dots).*

# Chapter 6

## $t\bar{t}H$ Multilepton analysis with $13.2 \text{ fb}^{-1}$ at $\sqrt{s} = 13 \text{ TeV}$

The observation of the associated production process of a Higgs boson and a top quark pair ( $t\bar{t}H$ ) would allow a direct measurement of the top-Yukawa coupling, as discussed in Section 2.5. In this chapter, a detailed description of the cut-based analysis designed to analyse the first  $13.2 \text{ fb}^{-1}$  of data at  $\sqrt{s} = 13 \text{ TeV}$  is presented [111]. A particular attention is given to data-driven methods used to estimate detector background, key of the analysis sensitivity at this stage. This work is part of a huge effort to achieve the ultimate goal: Establish highly sensitive analysis for the total statistics of Run2 ( $120 \text{ fb}^{-1}$ ) achievable at the end of 2018.

In multilepton analysis ( $H \rightarrow$  leptons), the leptons arise as secondary decay products from  $H \rightarrow WW$ ,  $H \rightarrow ZZ$  and  $H \rightarrow \tau\tau$  decays. To optimise the signal-to-background ratio, events are classified into 4 orthogonal channels according to the number of light and heavy leptons (Figure 6.1): 2 same sign leptons <sup>1</sup>( $2\ell ss$ ), 2 same sign leptons and one hadronically decaying  $\tau$  lepton ( $2\ell ss + 1\tau_{\text{had}}$ ), 3 leptons ( $3\ell$ ) and 4 leptons ( $4\ell$ ). My work focuses on the multilepton analysis where 2 same sign leptons without any hadronic tau object is required. This final state is dominated by the  $W^+W^-$  Higgs decay. The production of two lepton same sign processes in the SM is rare, this analysis is therefore almost SM background free. For the  $t\bar{t}H$  signal, the probability to get  $2\ell ss$  from  $t\bar{t}$  with  $H \rightarrow W^+W^-$  is found to be 4.1% (Figure 6.2) giving a total cross-section of about 20 fb.

The study of the associated reducible background to suppress non prompt leptons (fake leptons) from the  $t\bar{t}H$  signal using the matrix method (MM) is discussed. This reducible background represents the dominant background in  $2\ell ss$  and drives the systematic uncertainties on the total background. The estimation of fake leptons is performed using different data-driven methods since simulation is not reliable enough to estimate fakes. It is found that fakes from data are 1.5-3.6 times the fakes from simulation and represent 32-48% of the total background. In the first public results, fake factor method is used as the default approach to estimate fakes. Whereas, the MM (the topic of this thesis) is used as a cross check.

This chapter is organized as follows. Section 6.1 gives an overview of the data set and the Monte-Carlo simulation used for our measurements. In Section 6.2, object selection and signal region definition are explained. In Section 6.3, the details of the estimation of the reducible backgrounds using dedicated data-driven techniques are given. In Section 6.4, validation regions of prompt lepton background are described followed by the results to estimate the Yukawa coupling in  $t\bar{t}H$   $2\ell ss$  analysis in Section 6.5. Combination with

---

<sup>1</sup>leptons indicate light leptons only in this context, namely electrons and muons.

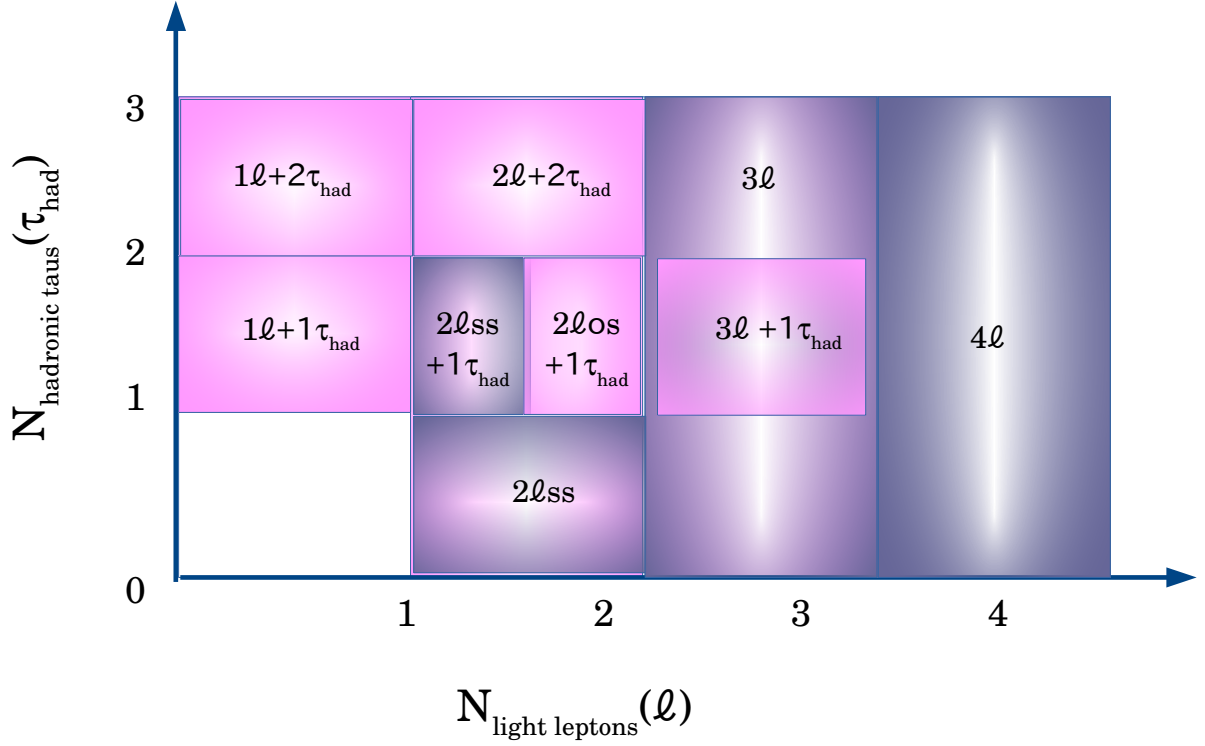


Figure 6.1: *Signal regions in  $t\bar{t}H$  multilepton analysis based on the number of light leptons, hadronic taus and the charge of light leptons. Purple channels indicate the channels that used in Ref. [111]. The pink ones are the new channels that will be included in the future iteration of this analysis.*

other  $t\bar{t}H$  channels are discussed in Section 6.6. Finally, conclusions and prospects are presented in Section 6.7.

## 6.1 Dataset and simulation

### 6.1.1 Data sample

The data set used in this analysis has been recorded by the ATLAS experiment during 2015 and the first part of 2016 at  $\sqrt{s} = 13$  TeV. It corresponds to a total integrated luminosity of  $13.2 \text{ fb}^{-1}$  ( $3.2 \text{ fb}^{-1}$  in 2015 and  $10 \text{ fb}^{-1}$  in 2016). The selected events satisfy high data quality criteria with a bunch crossing of 25 ns and IBL on (section 3.2). Such criteria require the event to belong to the recommended Good Run List (GRL) for Run2 [113] in order to remove any luminosity block with poor detector quality. Furthermore, the event is cleaned from any detector problems (section 3.2.2).

Data set has been collected by asking any of unpre-scaled single lepton triggers to be fired. The trigger selection is dependent on the number of interactions per collision (pile-up), hence different triggers are applied between 2015 and 2016 data set. The full list of single lepton triggers is shown in Table 6.1. For 2015 data set [114], the electron is required to have  $p_T$  higher than 60 GeV (or 24 GeV with isolated electrons) and passes

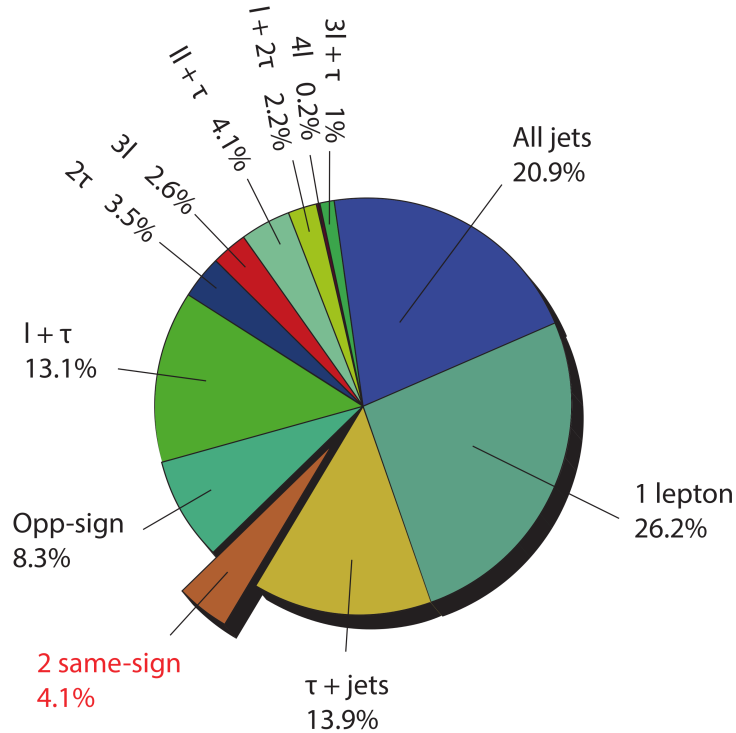


Figure 6.2: Branching ratio of different final states used for the direct measurements of the top Yukawa coupling in the  $t\bar{t}H$  channel with Higgs decaying to  $W^+W^-$  bosons. [112]

medium identification criteria (section 5) or it should have  $p_T$  higher than 120 GeV and passes loose identification criteria. The muons should pass either  $p_T > 50$  GeV or a loose identification criteria with lower  $p_T$  threshold (20 GeV). To cope with the pile-up increase in 2016, an additional requirements are added on leptons either by tightening identification criteria and  $p_T$  threshold and/or by adding  $d_0$  requirement.

Data Set (L)	2015 ( $3.2 \text{ fb}^{-1}$ )	2016 ( $10.0 \text{ fb}^{-1}$ )
Single Electron Trigger	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0
Single Muon Trigger	HLT_mu20_iloose_L1MU15 HLT_mu50	HLT_mu26_ivarmedium HLT_mu50

Table 6.1: Menu of single lepton trigger used with "or" logic for both 2015 and 2016 data analysis. HLT stands for High Level Trigger.

### 6.1.2 Signal MC sample

The analysis is designed to have two same-sign leptons in the final state, where one lepton originates from one of the top and the other lepton from mainly  $H \rightarrow W^+W^-$  decay. Figure 6.3 presents the final state under consideration consisting of two identified same sign leptons (electrons or muons excluding taus), missing energy and 6 jets, two of which are  $b$  jets. As the top quark decays with nearly 100% probability to a  $W$  boson and a  $b$



quark, the experimental signatures for top-quark pair production are determined by the decay of the  $W$  boson.

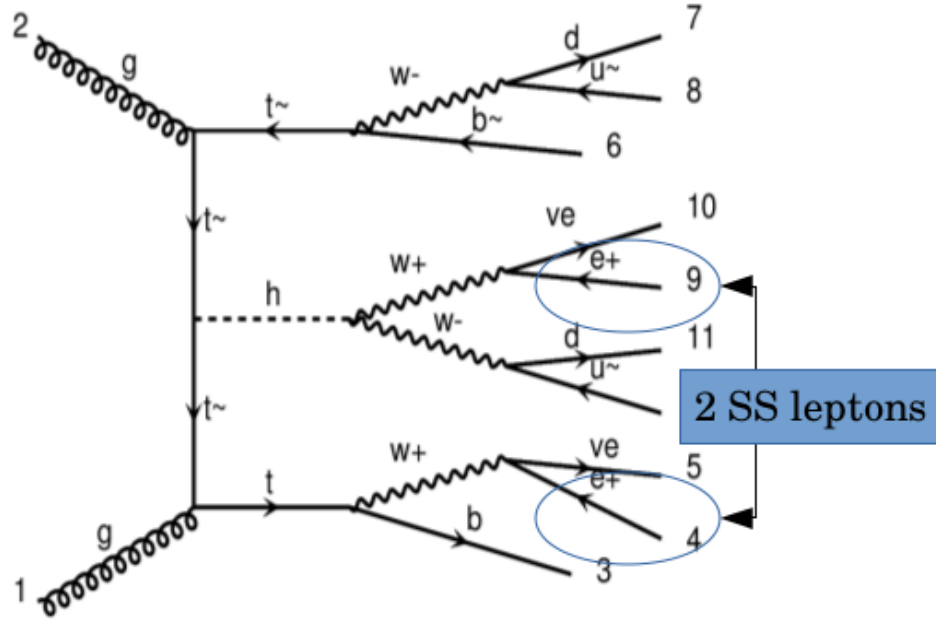


Figure 6.3: *Example of Feynman diagram giving 2lss  $t\bar{t}H$  signal event.*

The  $t\bar{t}H$  signal is simulated with a next-to-leading order (NLO) matrix element generated by MadGraph5 MG5\_AMC [115] and interfaced with PYTHIA 8 [107] for the parton showering and fragmentation. as shown in the first row of Table 6.2. It is designed using 3 decay modes: all hadronic, semileptonic and dileptonic  $t\bar{t}H$  samples where the Higgs mass is set to 125 GeV. This sample is used for all Higgs-boson production in association with a top-quark pair ( $t\bar{t}H$ ) including the subsequent semileptonic decay of the top-quark pair and the decay of the Higgs boson into a  $W$ -boson pair (contributing by 46-79%),  $\tau$  pair (17-51%) or  $Z$ -boson pair (2-9%) summarised in Table 6.3. The total cross section of  $t\bar{t}H$  Higgs production mode, 507 fb [33], is computed at NLO in QCD and electroweak couplings [116, 117, 118, 119, 120, 121, 122] as compiled in Refs. [34, 123]. It has uncertainties of  $^{+5.8\%}_{-9.2\%}$  from QCD renormalisation/factorisation scale choice and  $\pm 3.6\%$  from parton distribution function uncertainties (including  $\alpha_s$  uncertainties). It is worth to mention that the cross section of  $t\bar{t}H$  at  $\sqrt{s} = 13$  TeV is 3.9 higher than the cross section calculated at  $\sqrt{s} = 8$  TeV [34].

In order to optimise the sensitivity of the 2lss analysis, the signal region is categorised into three channels according to the combinations of light lepton flavours (same and opposite flavour categories) : electron-electron channel  $e^\pm e^\pm$ , muon-muon channel  $\mu^\pm \mu^\pm$  and electron-muon channel  $e^\pm \mu^\pm$  (note that the lepton could be the leading or sub-leading lepton  $p_T$  until it is explicitly specified in the context).

### 6.1.3 Background MC samples

Background events gather all SM processes that give the same final state as the signal. They are splitted in two types: First, the "reducible" background, where particles are mis-identified (e.g jets are reconstructed as electrons) or their charge is mis-reconstructed due



Process	Type	$\sigma$ (fb)	Generator + Hadronisation	PDF	Tune
$t\bar{t}H$	Signal	507	MG5_AMC[115] +PYTHIA 8 [107]	NNPDF 3.0 NLO [124]	A14 [125]
$t\bar{t}$	red. bckgd	$8.3 \cdot 10^5$	POWHEG-BOX[126]+PYTHIA 6	NNPDF 3.0 NLO/2.3 LO	A14
$W$ +jets	red. bckgd	$\mathcal{O}(10^7)$ [127]	SHERPA	CT10	SHERPA default
$Z$ +jets	red. bckgd	$\mathcal{O}(10^6)$ [127]	SHERPA	CT10	SHERPA default
$t\bar{t}W$	irr. bckgd	600.8	MG5_AMC +PYTHIA 8	NNPDF 3.0 NLO/2.3 LO	A14
$t\bar{t}Z$	irr. bckgd	839.3	MG5_AMC +PYTHIA 8	NNPDF 3.0 NLO/2.3 LO	A14
di-boson	irr. bckgd	$1.7 \cdot 10^5$	SHERPA [128]	CT10 [129]	SHERPA default
Single top	red. bckgd	$3.0 \cdot 10^5$ [130]	POWHEG-BOX [131, 132]	CT10/CTEQ6L1[133, 134]	Perugia2012 [135]
Rare	irr. bckgd	$\mathcal{O}(10^3)$	multiple	multiple	multiple

Table 6.2: Configurations used for event generation of 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 production of an electroweak boson ( $W$  or  $Z/\gamma^*$ ). “Tune” refers to the underlying-event tune of the parton shower generator. “MG5\_AMC” refers to MADGRAPH5\_AMC@NLO 2.2.1; “PYTHIA 6” refers to version 6.427; “PYTHIA 8” refers to version 8.2. Samples using PYTHIA 6 and PYTHIA 8 have heavy flavour hadron decays modeled by EVTGEN 1.2.0 [136]. All samples include leading-logarithm photon emission, either modeled by the parton shower generator or by PHOTOS [137].  $\sigma$  represents the total production cross section of each process. “red. bckgd” (“irr. bckgd”) stands for reducible (irreducible) background.

	$e^\pm e^\pm$	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$
$H \rightarrow WW$	76%	77%	79%
$H \rightarrow ZZ$	2%	3%	3%
$H \rightarrow \tau\tau$	17%	17%	17%
$H \rightarrow bb$	4%	3%	1%
$H \rightarrow \mu\mu$	< 1%	< 1%	< 1%
$H \rightarrow other$	< 1%	< 1%	< 1%

Table 6.3: Higgs decay mode in the three channels of  $2\ell ss$  signal region.

to Bremsstrahlung, or non-prompt leptons<sup>2</sup> decaying from heavy flavours (mainly from b-decay). Illustration of these background processes are shown in Figure 6.4. This kind of background is usually not well described by the simulation hence data driven methods are needed to estimate it. The second type of background, also called the “irreducible” background, is providing prompt leptons with  $2\ell ss$  from standard model processes. It is reasonably well described by the simulation. Two leptons with same sign can be produced by:  $t\bar{t}W$ ,  $t\bar{t}Z$ , diboson ( $W^+W^-$ ,  $ZZ$  and  $W^\pm Z$ ) and other rare processes such as VH,  $tZ$ , tri-boson and 4-top. The Feynman diagram of the most dominant ones,  $t\bar{t}W$ ,  $t\bar{t}Z$  and diboson, are presented in Figure 6.5. Figure 6.6 summarises the background estimation followed in this analysis.

Even if the reducible backgrounds are estimated by data-driven methods (section 6.3), it is useful to have a sense of the ‘performance’ of MC simulation. For  $t\bar{t}$  background, fully dominating the reducible background, two filters, “non all hadronic” and “dileptonic”, are used to get at least one lepton in the final state generated by POWHEG-BOX v2.0 and interfaced by PYTHIA 6 for the parton showering and fragmentation. The total cross

<sup>2</sup>Prompt leptons are defined as any lepton decaying from  $W^\pm$ ,  $Z$  or  $\tau$ .

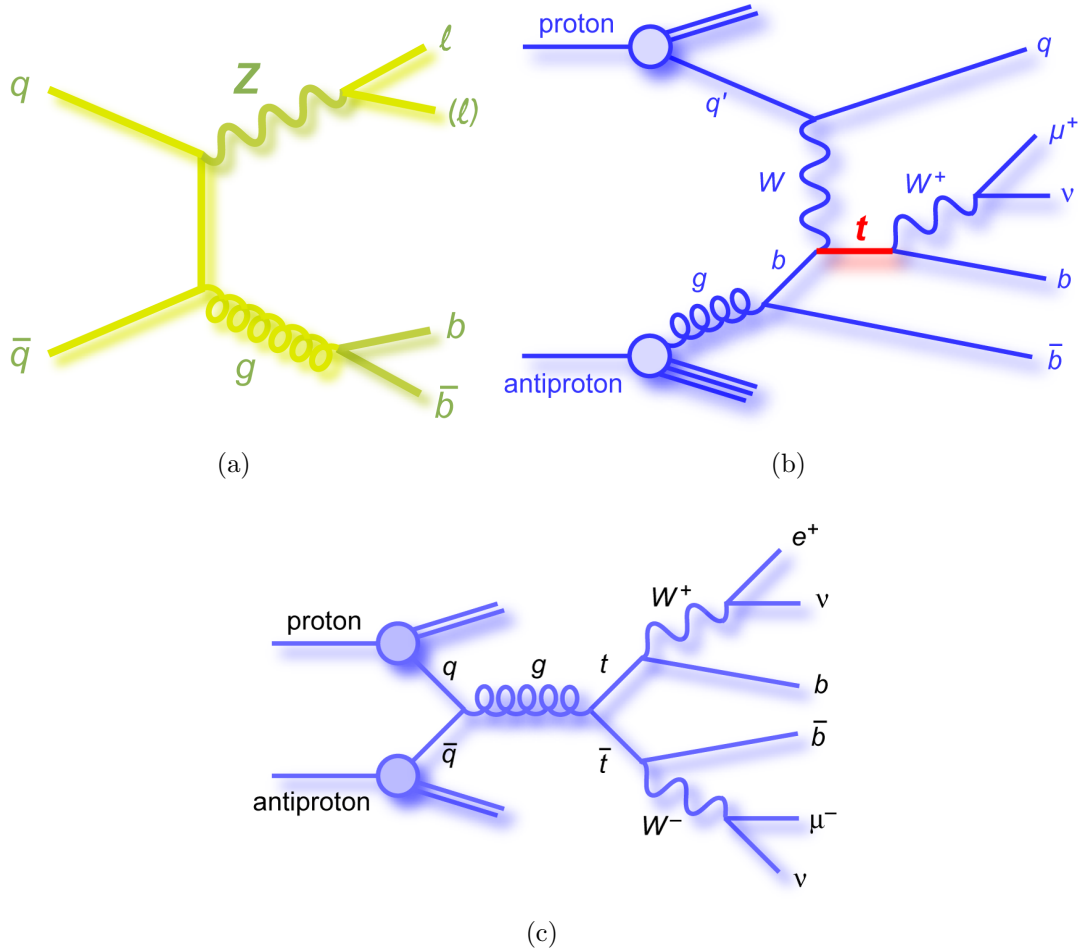


Figure 6.4: Example of Feynman diagrams of reducible background: a)  $Z$ +jets decaying to opposite sign prompt leptons. It is counted as background of  $2lss$  signal when the charge of one of the leptons is mis-reconstructed (charge flip). b) single top decaying to one prompt lepton (muon in this case). It is counted as background of  $2lss$  signal when the anti- $b$  quark is decaying leptonically (non-prompt lepton) or one of the jets is mis-reconstructed as a prompt lepton (true fake lepton). c)  $t\bar{t}$  decaying to opposite sign prompt leptons (dileptonic decay). It is counted as background of  $2lss$  signal when the charge of one of the leptons is mis-reconstructed (charge flip). Note that if one of the  $W$  boson decays hadronically,  $2lss$  final state could be gotten if one of this jets fakes the lepton (true fake lepton), or one of the  $b$  quark is decaying leptonically (non-prompt lepton).

section of  $t\bar{t}$  is 832 pb [138]<sup>3</sup>, where the top mass is set to 172.5 GeV, calculated by the program *Top++v2.0* [139] at next-to-next-to leading order NNLO and by including soft-gluon re-summation for the hadronic cross-section with full next-to-next-to-leading logarithmic (NNLL) accuracy. Since we are excluding fully hadronic decay, we only use 45% of the total cross section in this analysis. Then, it is three order of magnitude higher than the signal.

For  $t\bar{t}W$  and  $t\bar{t}Z$  background, the total cross sections are 600.8 and 839.3 fb [33] respectively<sup>4</sup> Only leptonic decays are taken into account and the corresponding cross section, shown in the Table 6.2, are computed at NLO in QCD and electroweak couplings using the configuration of Refs. [115, 122]. In the case of  $t\bar{t}Z$  background, the inclusive

<sup>3</sup>It is 3.3 higher than the cross section calculated at  $\sqrt{s} = 8$  TeV.

<sup>4</sup>They are 3.0 and 4.1 higher than the cross section calculated at  $\sqrt{s} = 8$  TeV respectively [140].

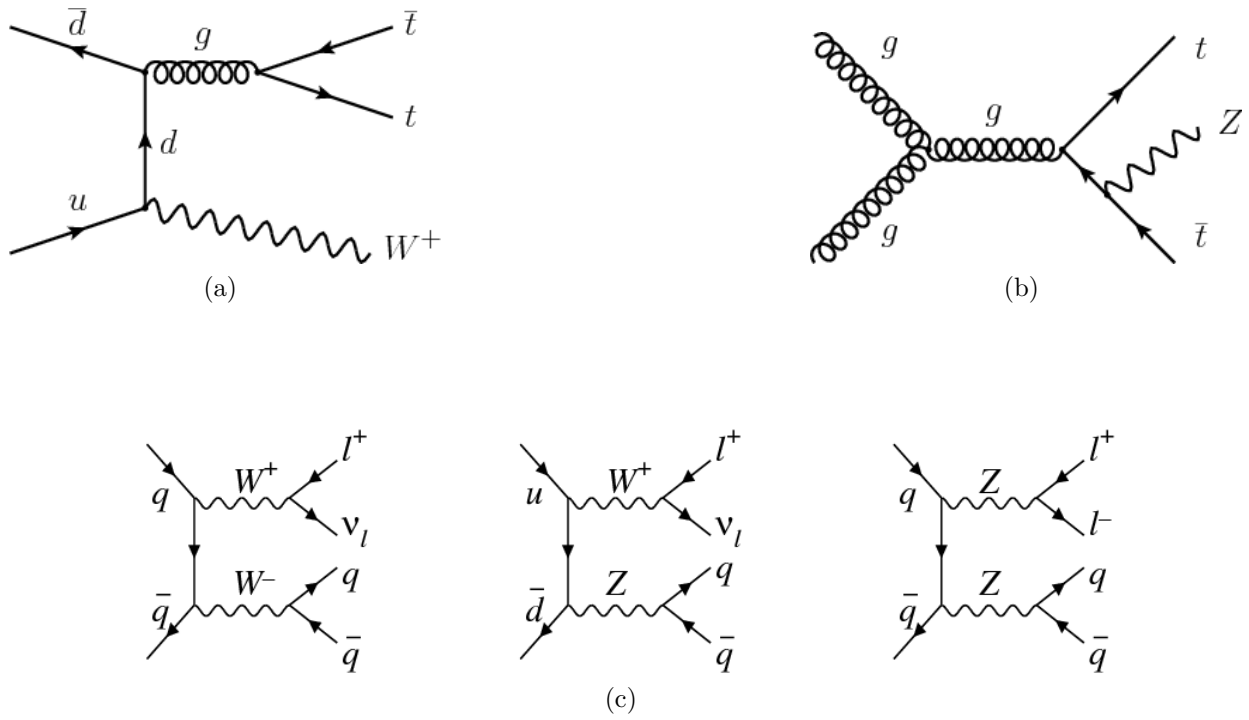


Figure 6.5: Feynman diagrams of the dominant irreducible background of  $t\bar{t}H$  signal. a)  $t\bar{t}W$ , b)  $t\bar{t}Z$  and c) diboson ( $WW$ ,  $WZ$ ,  $ZZ$ ) production processes.

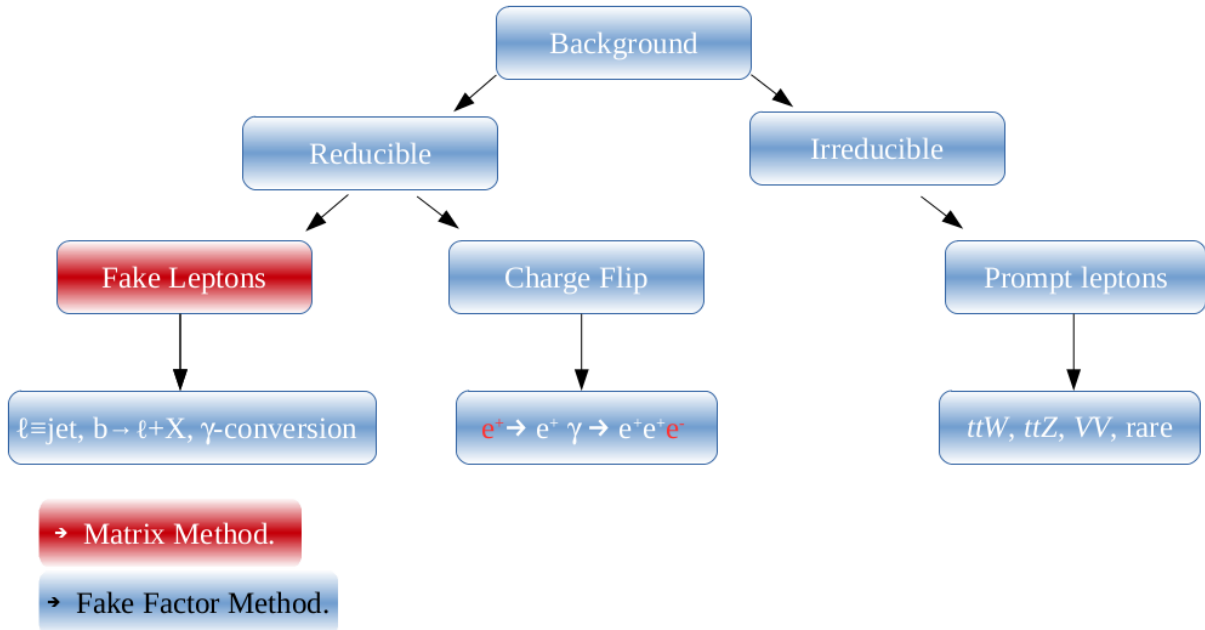


Figure 6.6: Overall view of the background categorization in the  $2lss t\bar{t}H$  analysis.

$t\bar{t}l\bar{l}$  matrix element is computed, including off-shell  $Z$  and  $\gamma^*$  contributions with  $m(\ell\bar{\ell}) > 5$  GeV.

For diboson, samples are used generated by SHERPA version 2.1.1 [128]. It is worth to notice that the diboson background is about 300 time the signal. Since there is no  $b$  jet in diboson sample, this characteristic will be used later to reduce this background. The

cross section of  $WW$ ,  $ZZ$  and  $WZ$  are 111.1 pb [141], 15.6 pb [142] and 48.2 pb [143] respectively.

For  $W$ +jets and  $Z$ +jets, samples used SHERPA taking into account all possible leptonic decay modes with different  $p_T$  thresholds and 'c' and 'b' filter/veto configurations (systematics). For low mass  $Z$ +jets background, samples are used with  $10 < m_{ll} < 40$  GeV and with the same systematic configurations used for  $Z$ +jets and  $W$ +jets backgrounds.

For rare, the following processes are included: 4-top,  $t\bar{t}W^+W^-$  designed by MG5\_AMC + PYTHIA 8,  $tZ$  designed by MG5\_AMC + PYTHIA,  $tHbj$  designed by MG5\_AMC + HERWIG++ [144], tribosons ( $WWW, ZZZ, WZZ$  with different leptonic decays) designed by SHERPA and  $tWZ$  designed by PYTHIA 8 + EvtGen [136].

Event generator programs and configurations used for simulating all background processes are summarised in Table 6.2. Generated events are passed through a full GEANT4 [109] simulation of the ATLAS detector. Additional minimum-bias  $pp$  interactions (pile-up) are modeled with the PYTHIA 8.1 generator with the MSTW2008LO parton distribution function (PDF) set [145] and the A2 tune [146], and are added to simulated events according to the luminosity profile of the recorded data.

All MC processes are weighted event by event taking into account the pre-selection, pile-up, b-tag and jet vertex tagger (JVT) event weights in addition to applying *tight*, *loose* and *trigger* object scale factors (see Chapter 4), and finally normalised to data taking into account the available luminosity of  $13.2 \text{ fb}^{-1}$  and the cross section of each process.

## 6.2 Signal region

After recalling the object definition in section 6.2.1, section 6.2.2 gives the definition of the signal region.

### 6.2.1 Object definition

Experimentally, in order to distinguish between different objects involved in the final state of a given process (electrons, muons, taus and jets<sup>5</sup>) primary cuts should be applied on individual objects before event selection. For leptons, two levels are considered: *loose* (baseline) and *tight* (signal) definitions.

#### 6.2.1.1 Electron

To distinguish real electrons from hadronic showers (true fake electrons) and reduce/suppress as much as possible the main background (the non-prompt leptons and charge misidentification in  $e^\pm e^\pm$  and  $e^\pm \mu^\pm$ ) keeping a reasonably high signal efficiency, loose cuts on electron kinematics, isolation, identification and impact parameters are applied and described in the left column of Table 6.4.

First, the electron candidates are restricted to be in the barrel region ( $|\eta| < 1.37$ ) in order to reduce the contamination from charge mis-identification background (section 6.3.1) because of the big amount of material in the end-cap tracker ( $\eta > 1.52$ ) shown in Figure 3.9. Only those electron candidates with transverse momentum greater than 10 GeV are considered. A loose electron discriminant (The *looseAndBLayer* working point)

---

<sup>5</sup>The missing transverse energy ( $E_T^{miss}$ ) is not used in signal region definition since it is not considered as one of discriminative variables that help to distinguish signal from background.

is used which gives approximate 95% electron efficiency using  $Z \rightarrow ee$  events (see previous chapter for more details). The electron candidate is required to be isolated in the calorimeter and the tracker:  $E_T^{cone20}/p_T$  ( $p_T^{varcone20}/p_T$ ) are required to be less than 0.06. Flat efficiencies of 99% in  $\eta - p_T$  plane are achieved for electrons by applying cuts on those two isolation variables [85]. Finally, the track is required to be originating from the primary vertex. Therefore, the longitudinal impact parameter of the electron track with respect to the selected event primary vertex, multiplied by the sine of the polar angle,  $|z_0 \sin \theta|$ , is required to be less than 0.5 mm. The transverse impact parameter divided by the estimated uncertainty on its measurement,  $|d_0/\sigma(d_0)|$ , must be less than 5. These values were common to all ATLAS analyses made for the first  $13.2 \text{ fb}^{-1}$  results (because of the lack of time to reoptimise those parameters).

The *tight* electron, defined in the right of Table 6.4, are used to define the signal region. Electron candidate has to pass the *TightLH* working point of likelihood based discriminant. Moreover, both  $E_T^{cone20}/p_T$  and  $p_T^{varcone20}/p_T$  are required to be less than 0.06 with additional “gradient” isolation working point for fake/non-prompt lepton rejection, corresponding to the *FixedCutTight* isolation WP. Also, lepton transverse momenta have to be greater than 25 GeV in order to avoid trigger bias for the leading lepton and since the sub-leading lepton has better signal sensitivity at  $p_T > 25 \text{ GeV}$  for almost all channels.

### 6.2.1.2 Muon

The muon selection is also detailed in Table 6.4. The *loose* muon candidates are selected within  $|\eta| < 2.5$  (muon detector acceptance). However, unlike electrons, no further cut on  $\eta$  is needed to suppress charge miss-identification since it is negligible for muons. Calorimetric and track isolation variables are defined for muon candidates similarly as for electron candidates, except that the track isolation uses a larger cone size at low  $p_T$  ( $\Delta R = \min(0.3, 10 \text{ GeV}/p_T (\mu))$ ) corresponding to  $p_T^{varcone30}$ . Finally, the transverse impact parameter requirement for muon candidates is slightly tighter than for electrons ( $|d_0/\sigma(d_0)|$ , must be less than 3). For *tight* muon candidates, the track based isolation  $p_T^{varcone30}/p_T$  has to be less than 0.06 with additional “gradient” isolation working point for fake/non-prompt lepton rejection. This corresponds to the *FixedCutTightTrackOnly* isolation WP). Similarly to *tight* electrons, lepton transverse momenta have to be greater than 25 GeV.

Cut	Loose (L)		Tight (T)	
	$e$	$\mu$	$e$	$\mu$
$ \eta $	$< 1.37$	$< 2.5$	$< 1.37$	$< 2.5$
$p_T$	$> 10 \text{ GeV}$	$> 10 \text{ GeV}$	$> 25 \text{ GeV}$	$> 25 \text{ GeV}$
$p_{T,e}^{varcone20}, p_{T,\mu}^{varcone30}$	$< 0.06 \times p_T$	$< 0.06 \times p_T$	$< 0.06 \times p_T$ (*)	$< 0.06 \times p_T$ (*)
$E_T^{cone20}$	$< 0.06 \times p_T$	$< 0.06 \times p_T$	$< 0.06 \times p_T$ (*)	$< 0.06 \times p_T$ (*)
Identification working point	<i>looseAndBLayer</i>	Loose	Tight LH	Loose
Transverse impact parameter $ d_0 /\sigma_{d_0}$	$< 5$	$< 3$	$< 5$	$< 3$
$z$ impact parameter $ \Delta z_0 \sin \theta_\ell $	$< 0.5 \text{ mm}$	$< 0.5 \text{ mm}$	$< 0.5 \text{ mm}$	$< 0.5 \text{ mm}$

Table 6.4: *Tight and loose lepton definitions. The lepton identification working points are described in section 4.2 and section 4.4 for electrons and muons respectively. Selections for tight leptons are applied on top of the selections for loose leptons. (\*) An additional “gradient” isolation working point is defined with efficiency and fake/non-prompt lepton rejection intermediate between the loose and tight isolation selections.*

### 6.2.1.3 Tau

Hadronically decaying  $\tau$  lepton candidates are required to have either one or three associated tracks, with a total charge of  $\pm 1$ . Candidates with  $p_T > 25$  GeV and  $|\eta| < 2.5$ , excluding the electromagnetic calorimeter transition region, are considered. A boosted decision tree discriminant using calorimeter and tracking-based variables is used to identify  $\tau_{\text{had}}$  candidates and reject generic jet backgrounds. The chosen working point has an efficiency of 75% (59%) for one- (three-)prong  $\tau_{\text{had}}$  decays. Electrons which are reconstructed as one-prong  $\tau_{\text{had}}$  candidates are removed using a sliding cut on the electron likelihood ID variable; the rejection factor (inverse efficiency) for electrons is  $\approx 30$ –100 depending on  $\eta$  and  $p_T$ . It is worth to notice that no  $\tau$  candidate is considered in  $2\ell\text{ss}$  analysis but it is used later for combination with the other multilepton channels.

### 6.2.1.4 Jets and b-tagged jets

“Bad” jets, defined as object coming from the beam background or falling in the regions with hot noisy calorimeter cells, are removed. Jets are accepted within the fiducial region  $p_T > 25$  GeV and  $|\eta| < 2.5$ . To discriminate between hard scattered jet and pile-up, jet with  $p_T < 60$  GeV and  $|\eta| < 2.4$  is required to satisfy the criteria on the JVT (see Section 4.6) to be less than 0.59, which gives about 92% efficiency and about 2% fake rate.

Jets containing b-hadrons are identified as b-jets using the MV2c10 algorithm (section 4.6.2), a multivariate discriminant making use of the long lifetime and high decay multiplicity and hard fragmentation and high mass of b-hadrons. The working point used for this analysis corresponds to approximately 70% efficiency to tag b-hadron jet.

### 6.2.1.5 Overlap removal

Further treatment is required to remove duplication that may happen during the reconstruction of one physical object as two different objects or to define well isolated objects for two separate but close-by objects. Hence an overlap removal procedure (OvR) is performed among objects, on top of above object baseline selection.  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$  is used as a measure of the distance of two objects. The removal procedure is in order and it is summarised in Table 6.5. Tau is removed when it is close to any electrons or muons within  $\Delta R < 0.2$ . Then electron is also removed if it shares track with muon. Jet is removed if it is close to any electron within  $\Delta R < 0.2$ , then electron close to any jet within  $\Delta R < 0.4$  is removed. Within  $\Delta R < 0.4$ , jet is removed if it is close to any muon and associated with less than three tracks, otherwise muon is removed. Finally, jet is removed if it is close to any tau within  $\Delta R < 0.2$ .

## 6.2.2 Signal region definition

The selection for the signal region is summarised in Table 6.6 inspired by Run1 analysis [49], given the low available luminosity for the first public results at  $\sqrt{s} = 13$  TeV. There must be at least one offline leptons with  $p_T > 25$  GeV (or 21 GeV for muon in 2015), to be matched to any of the online lepton trigger objects. In order to suppress  $t\bar{t}$ ,  $t\bar{t}V$  and di-boson backgrounds, signal events are required to have at least one b-tagged



Keep	Remove	Cone size ( $\Delta R$ ) or track
electron	tau	0.2
muon	tau	0.2
electron	CaloTagged muon	shared track
muon	electron	shared track
electron	jet	0.2
jet	electron	0.4
muon	jet	(0.2 or ghost-matched to muon) and (numJetTrk $\leq 2$ )
jet	muon	0.4
tau	jet	0.2

Table 6.5: Summary of overlap removal between electrons, muons, taus and jets

jets<sup>6</sup>; at least two leptons<sup>7</sup> should have same-signed charge. To ensure orthogonality with the  $2lss+1\tau_{had}$  channel, events having hadronic tau object are vetoed. No further requirement is applied on either of the top quarks to be reconstructed, since the principal backgrounds also contain top quarks. Moreover, despite the signal could have additional neutrino comparing to the background, no missing transverse momentum cut is required since the experimental resolution smears this difference. The Higgs boson reconstruction, difficult due to the loss of neutrinos from leptonic  $W$  decays, is also not attempted. Lepton candidates must pass *tight* requirements, summarised on the right of Table 6.4,

Channel	Leptons	Taus	Jets	B-Tags	Lepton flavour	Trigger match
$2lss$	$== 2$ <i>tight</i> $\Sigma Q_{lep} = \pm 2$ $p_T^{lead} > 25$ GeV $p_T^{lead} > 25$ GeV $ \eta^{elec}  < 1.37$	$== 0$	$\geq 5$	$\geq 1$	$ee$ $e\mu$ $\mu\mu$	at least one trigger matched lepton with $p_T > 25$ GeV (21 GeV for muon in 2015)

Table 6.6: Signal region definition for  $2lss$  channel. "tight" is defined in Table 6.4.

to increase the sensitivity of the signal. Finally, signal events are required to have at least five jets instead of four because of very unfavorable signal to background ratio. It is found that the 4-jet multiplicity decreases the sensitivity due to the high amount of  $t\bar{t}$ ,  $t\bar{t}V$  and di-boson backgrounds in this region. It also allows to have an orthogonal regions to the signal region ( $N_{jets} \leq 4$ ), to increase the statistics in the control regions used for fake and charge flip estimation and secondly, it is quite close to the kinematics of this signal region so we are able to reduce the bias when extrapolating from the control region to the signal region.

### 6.3 Data-driven background estimation

The methods to estimate the two reducible backgrounds are treated in this section: First, electron charge flip (QmisId) using likelihood based method (section 6.3.1) and, sec-

<sup>6</sup>Since the b jet identification efficiency is about 70%, we practically require at least one b jet instead 2 b jets. On top of that, events with no b jets are not included in order to reduce the diboson background contribution and any background that does not contain top quarks.

<sup>7</sup>The leading  $p_T$  lepton is ordered as lepton index "1" and the sub-leading  $p_T$  lepton to be lepton index "2".



ond, fake leptons using Matrix method (MM) described in section 6.3.2. Section 6.3.3 is devoted for comparison with the method chosen for the first publication [111].

### 6.3.1 Charge flip estimation

Charge misreconstruction, commonly referred to as "charge flip", is defined when the sign of the electric charge of a lepton is mis-assigned, making an event with an opposite-charge lepton pair to appear as same-charge pair. For electrons, this happens in two cases: Either from the trident process  $e^\pm \rightarrow \gamma^* e^\pm \rightarrow e^\mp e^\pm e^\pm$  where a high- $p_T$  electron with opposite charge to the original prompt electron is reconstructed (hard Bremsstrahlung followed by photon conversion processes as explained in Figure 6.7). The misreconstructed electron could also arise from a slightly curved track that induces a measurement error of the charge by the magnet. This effect is important at very high transverse momentum. Thus, a small dependence on electron  $p_T$  is also expected. The contribution of hard Bremsstrahlung represents the main contribution to this background. The fraction of trident electron depends on the amount of material (Figure 3.9). Therefore, a strong dependence on  $\eta$  is expected. For muons, this background is found to be negligible at high  $p_T$ , thanks to the charge measurements in both the inner detector and muon spectrometer, and at low  $p_T$  since the probability that a bremsstrahlung happens for muons is much weaker than for electrons.

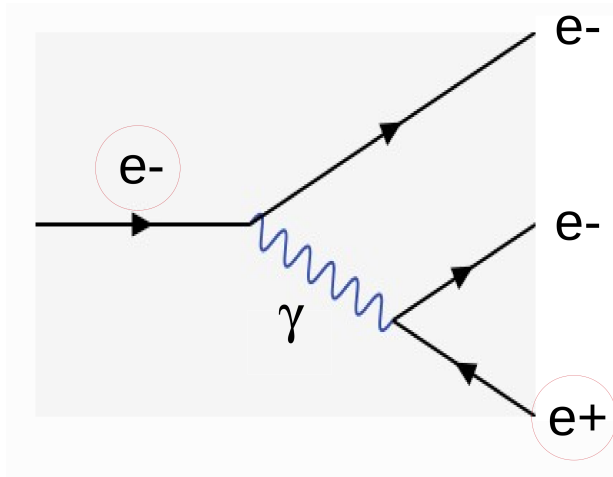


Figure 6.7: *Electron charge misreconstruction due to hard Bremsstrahlung followed by photon conversion process.*

In order to estimate the number of same sign events corresponding to charge flip, the mis-identification rates are defined as the probability that a lepton charge from OS event is mis-reconstructed to give SS event. The rate of charge mis-identification  $\epsilon$  is extracted from a two-dimensional Likelihood fit in terms of electron  $\eta$  and  $p_T$  using electrons selected from pure  $Z \rightarrow ee$  events. This method is chosen for its good statistical precision since the charge flip rate is quite small [147]. The background subtraction below the  $Z$  mass peak is done by using side-band method. The measured charge mis-identification rates from data are shown in Figure 6.8. The dependency of those rates as a function of  $\eta$  and  $p_T$  follows the expectations (higher at high  $\eta$  and high  $p_T$ ). The charge misreconstruction rates for *tight* electron vary from 0.02% for low- $p_T$  *tight* electrons at small  $|\eta|$  to 10% at high  $p_T$  in the end-cap. For that reason, the region  $|\eta| > 1.37$  is removed for the  $2\ell ss$  analysis. For electrons with  $p_T$  within [130, 1000] GeV, rates are extrapolated

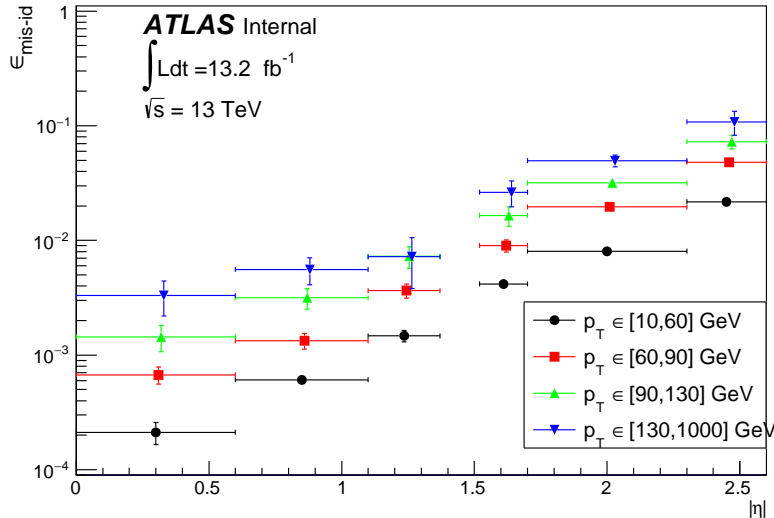
from [90, 130] GeV bin, due to limited statistics at high  $p_T$  region, by using a  $p_T$  dependent factor extracted from simulated  $t\bar{t}$  events. Relative uncertainties vary from 10 to 40 %. The main sources of the uncertainties are coming from the statistics of the electron sample used for the method and of the extracted factor for doing high  $p_T$  extrapolation; The non-closure of the likelihood method with the truth-matching method; Background subtraction in selecting  $Z$  events.

Knowing these charge mis-identification rates  $\epsilon$ , it is possible to compute the estimated number of same-sign events  $N^{ss}$  from the measured number of opposite-sign events  $N^{os}$  in the signal region (defined in section 6.2) using charge mis-identification weights  $w_{TT,ee}^{QMisID}$  and  $w_{TT,e\mu}^{QMisID}$  measured for *tight* electrons in  $ee$  and  $e\mu$  channels respectively. Those QmisID weights are written in terms of the charge mis-identification rates for the two different electrons  $i$  and  $j$  as shown in Equations (6.1) and (6.2) derived in Appendix B.4:

$$w_{TT,ee}^{QMisID} = \frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1 - \epsilon_i - \epsilon_j + 2\epsilon_i\epsilon_j} \quad (6.1)$$

$$w_{TT,e\mu}^{QMisID} = \frac{\epsilon}{1 - \epsilon} \quad (6.2)$$

The same measurements is performed on *anti-tight* electrons (*loose* electrons failing *tight* selection) for the matrix method (see section 6.3.2). For that, the expression of Likelihood is slightly changed to account for the asymmetry between the two electrons forming the pair (following SUSY  $2\ell ss$  analysis [147]).



(a) *Tight*

Figure 6.8: *Electron tight charge mis-identification rates measured in data with the likelihood method on  $Z$  events (black points, red squares and blue triangles) as a function of  $|\eta|$  and  $p_T$ . The 2015 dataset has been used to estimate the rates below 130 GeV. Above this value, the charge flip rates have been estimated by extrapolating the rates in the region where the  $p_T \in [90, 130]$  GeV with a  $p_T$  dependent factor extracted from simulated  $t\bar{t}$  events (green triangles). Statistical and systematic uncertainties have been included in this plot.*

### 6.3.2 Fake lepton estimation with the Matrix Method

Whereas real leptons denote prompt and isolated leptons coming from the primary interaction vertex from  $W^\pm$ ,  $Z$  or  $\tau$ , fake leptons are defined to be any non-prompt lepton produced in heavy flavour meson decays or misreconstructed object as a lepton coming from converted photons from various origins, light hadrons faking the electron shower, in-flight decays of kaons or pions to muons, etc. In this thesis, the estimation of fakes are performed using the matrix method (MM) explained in details in section 6.3.2.1. Section 6.3.2.2 presents the inputs of the method (real lepton efficiency and fake lepton rate). The validation of this method, an important and challengeable piece, is shown in section 6.3.2.3. The final results in data are presented in section 6.3.2.4 and compared with the Fake factor method, used as the baseline method for the  $13.2 \text{ fb}^{-1}$  analysis, in section 6.3.3. In the following, the fake event indicates any event including at least one fake lepton (one or two fake leptons). The real event is any event that both leptons are real.

#### 6.3.2.1 Description of Matrix Method

The number of fake leptons in the SR, as a tight region, denoted by  $f_{SR}$  could be linked to the number of fakes in the baseline<sup>8</sup> " $f_{base}$ ", as a loose region (including both *tight* and *anti-tight* leptons  $L = T \cup \bar{\mathcal{X}}$ ), by the fake rate  $\varepsilon_f$  which is defined as the probability of *loose* fake lepton to pass *tight* selection.

$$f_{SR} = \varepsilon_f \times f_{base} \quad (6.3)$$

To estimate the number of fakes in the baseline region  $f_{base}$ , where both real and fake leptons are present, the matrix method is used which links the unknown number of real and fake events in the baseline region,  $r_{base}$  and  $f_{base}$ , to the observed *tight* and *loose* events ( $T$  and  $L$ ) by the real efficiency  $\varepsilon_r$  and fake rate  $\varepsilon_f$ , defined as the probability of a baseline real and fake lepton, respectively, to pass *tight* selection. The  $4 \times 4$  matrix equation can be written as:

$$\begin{pmatrix} TT \\ T\bar{\mathcal{X}} \\ \bar{\mathcal{X}}T \\ \bar{\mathcal{X}}\bar{\mathcal{X}} \end{pmatrix}_i = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\bar{\varepsilon}_{r,2} & \varepsilon_{r,1}\bar{\varepsilon}_{f,2} & \varepsilon_{f,1}\bar{\varepsilon}_{r,2} & \varepsilon_{f,1}\bar{\varepsilon}_{f,2} \\ \bar{\varepsilon}_{r,1}\varepsilon_{r,2} & \bar{\varepsilon}_{r,1}\varepsilon_{f,2} & \bar{\varepsilon}_{f,1}\varepsilon_{r,2} & \bar{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{f,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2} \end{pmatrix}_i \begin{pmatrix} rr \\ rf \\ fr \\ ff \end{pmatrix}_i, \quad (6.4)$$

where:  $TT_i$  denotes the event  $i$  with both leptons passing *tight* selection.  $T\bar{\mathcal{X}}_i$  ( $\bar{\mathcal{X}}T_i$ ) is the event  $i$  with leading lepton passing (failing) tight cut and sub-leading lepton failing (passing) *tight* cut and  $\bar{\mathcal{X}}\bar{\mathcal{X}}_i$  is the event  $i$  with both leptons failing *tight* selection.  $rr_i$  denotes the event  $i$  with both leptons being real.  $rf_i$  ( $fr_i$ ) is the event  $i$  with leading lepton being real (fake) and sub-leading lepton being fake (real) and  $ff_i$  is the event  $i$  with both leptons being fake.  $\bar{\varepsilon}_r \equiv (1 - \varepsilon_r)$  ( $\bar{\varepsilon}_f \equiv (1 - \varepsilon_f)$ ) represent the probability for a real (fake) lepton to fail *tight* but to pass baseline selection. In this analysis the indexes for  $\varepsilon_r$  and  $\varepsilon_f$  are parameterised in terms of lepton  $p_T$  ordering.

Finally, the final number of fakes in the signal region, corresponding to the total number of events with at least one fake lepton, can be obtained by summing on all events:

<sup>8</sup>The baseline region includes *anti-tight* leptons in addition to the selections defined in Table 6.6.

$$f_{SR} = \sum_i (\varepsilon_{r,1}\varepsilon_{f,2}(rf))_i + (\varepsilon_{r,2}\varepsilon_{f,1}(fr))_i + (\varepsilon_{f,1}\varepsilon_{f,2}(ff))_i =$$

$$\sum_i w_{TT}^{MM} TT_i + \sum_i w_{T\mathcal{F}}^{MM} T\mathcal{F}_i + \sum_i w_{\mathcal{F}T}^{MM} \mathcal{F}T_i + \sum_i w_{\mathcal{F}\mathcal{F}}^{MM} \mathcal{F}\mathcal{F}_i \quad (6.5)$$

where the matrix method weights are defined as following:

$$w_{TT}^{MM} = (1 - \alpha\varepsilon_{r,1}\varepsilon_{r,2}\bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2})_i \quad (6.6)$$

$$w_{T\mathcal{F}}^{MM} = (\alpha\varepsilon_{r,1}\varepsilon_{r,2}\varepsilon_{f,2}\bar{\varepsilon}_{f,1})_i \quad (6.7)$$

$$w_{\mathcal{F}T}^{MM} = (\alpha\varepsilon_{r,1}\varepsilon_{r,2}\varepsilon_{f,1}\bar{\varepsilon}_{f,2})_i \quad (6.8)$$

$$w_{\mathcal{F}\mathcal{F}}^{MM} = -(\alpha\varepsilon_{r,1}\varepsilon_{r,2}\varepsilon_{f,1}\varepsilon_{f,2})_i \quad (6.9)$$

$$\alpha_i = \frac{1}{(\varepsilon_{r,1i} - \varepsilon_{f,1i})(\varepsilon_{r,2i} - \varepsilon_{f,2i})} \quad (6.10)$$

It is worth to note that each event contributes to only one of the four factors on the right side of the Equation (6.5). Real and fake efficiencies ( $\varepsilon_r$  and  $\varepsilon_f$ ) can be measured directly in data using *tag-and-probe* method, as described in detail in section 6.3.2.2. Both efficiencies are estimated in other regions, called control regions "CR", defined in such way to get efficiencies as "same" values as in the SR but considered as real and fake-enriched regions respectively, where the signal is negligible. To get similar fake rate and real efficiency in both SR and CR, it is important that those CRs have similar background composition to the SR. For illustration, five control regions, in addition to the SR, are defined in Figure 6.9. Three discriminative variables are used to distinguish between SR and other control regions, namely the number of jets  $N_{jets}$ , assuring the orthogonality between signal and CR where efficiencies are estimated, lepton selection (*Tight T* or *Loose L* defined in Table 6.4), assuring enough statistics for fakes, and finally Lepton charge (opposite-sign *OS* or same-sign *SS*) giving the orthogonality between real or fake enriched-regions respectively.

In practice, the assumption of the MM "same efficiencies in CR and SR" is not ideal. Hence systematics should be added as described later.

**Charge mis-identification veto for matrix method** The baseline selection must contain only real and fake leptons. Hence any event including electrons with mis-reconstructed charge should be removed from the *loose* selection. As a consequence, regions with large charge flip electrons background - namely  $ee$  and  $e\mu$  - must be treated differently than  $\mu\mu$  in the context of the MM. Starting from the charge flip rates computed in section 6.3.1, the contamination from charge flip events is removed from the four same-sign sideband regions. The following modification to Equation (6.5) is introduced:

$$f_{SR} = \sum_{XX=TT,T\mathcal{F},\mathcal{F}T,\mathcal{F}\mathcal{F}} \left( \sum_i w_{XX}^{MM,SS} XX^{SS}_i - \sum_j w_{XX}^{MM,OS} w_{XX}^{Qflip} XX^{OS}_j \right) \quad (6.11)$$

where  $w^{Qflip}$  stands for the charge flip weights (see section 6.3.1).  $w^{MM,SS}$  and  $w^{MM,OS}$  stand for the matrix method weights applied in SS and OS control regions respectively. They are both equivalent and calculated using fake rate  $\varepsilon_f$  estimated in SS control region because a study performed on simulation shows that the fake electron origin in more than 90% of the cases is the same in the baseline ( $CR_3$ ) for both OS and SS charge dielectron events as explained in Table 6.7. This allows exploiting a unique measurement of  $\varepsilon_f$  for re-weighting both SS and OS data events.

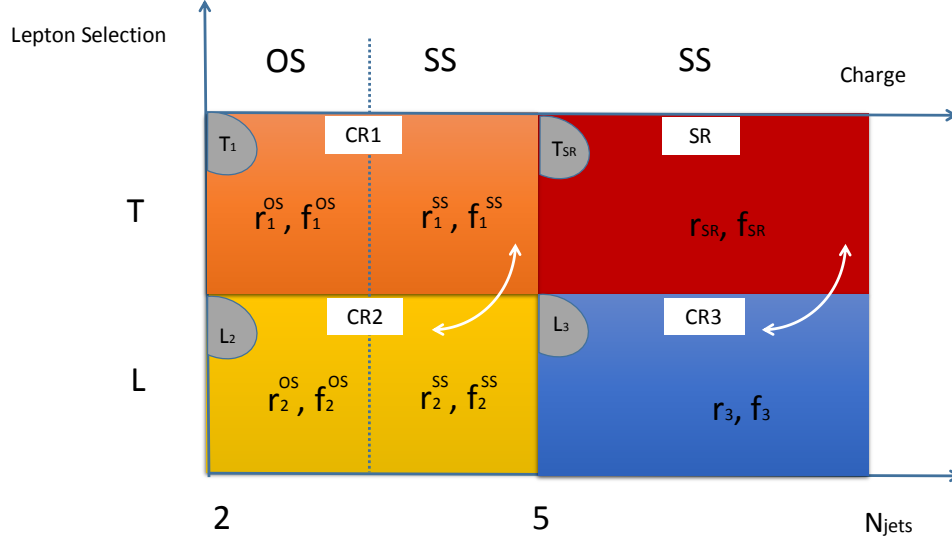


Figure 6.9: Basics of MM to estimate fakes and control regions definition kinematically close to  $2\ell ss$  signal region.  $CR_1^{OS(SS)}$  denotes events with 2-3-4 jet multiplicity and two tight opposite-sign leptons (same-sign leptons).  $CR_2^{OS(SS)}$  denotes events with 2-3-4 jet multiplicity and two loose opposite-sign leptons (same-sign leptons). "baseline" ==  $CR_3$  denotes events with at least five jets and two loose same-sign leptons. Note that  $CR_2$  region includes  $CR_1$  and  $CR_3$  includes  $SR$  since loose lepton could be Tight or anti-tight.  $f_{base}$  is estimated in  $CR_3$ . The fake rate is estimated in  $CR_1^{SS}$  and  $CR_2^{SS}$ . The real efficiency is estimated in  $CR_1^{OS}$  and  $CR_2^{OS}$ .

### 6.3.2.2 Inputs to Matrix Method

Looking at Equation (6.5), three inputs should be estimated in the MM to compute the number of fakes in the SR: real efficiency  $\varepsilon_r$ , fake rate  $\varepsilon_f$  and sideband yields ( $TT$ ,  $T\bar{\mathcal{X}}$ ,  $\bar{\mathcal{X}}T$  and  $\bar{\mathcal{X}}\bar{\mathcal{X}}$ ). The measurements of real efficiency and fake rate are now discussed.

**Real efficiency ( $\varepsilon_r$ )** Real efficiency is measured using the T&P method, similar to the one discussed in section 5.2.1, by taking the ratio *loose* to *tight* probe leptons in data, in  $CR_1^{OS}$  and  $CR_2^{OS}$  shown in Figure 6.9, after subtracting any non-prompt lepton background:

$$\varepsilon_r = \frac{N_T^{data\ OS} - N_T^{non-prompt\ OS}}{N_L^{data\ OS} - N_L^{non-prompt\ OS}} \quad (6.12)$$

The *real* control region is designed to be enriched in prompt leptons from  $t\bar{t}$  dileptonic decays<sup>9</sup> by requiring the presence of two opposite-sign and opposite flavour leptons (to remove  $Z$ +jets) and at least one  $b$ -jet. The complete definition is given in Table 6.8. The *tag* in the *real* enriched CR is the lepton that passes *tight* and trigger match requirements. The *probe* is the other lepton of the OF dilepton event. Two probes could be selected in the same event in case both leptons are found to be *tight* and trigger matched.

<sup>9</sup> $t\bar{t}$  sample is chosen instead  $Z \rightarrow \ell\ell$  sample according to its closer topology to the Signal Regions with leptons close to  $b$ -jet.

	Fake OS CR	Fake SS CR
Probe e - T	95% : HF non-prompt 5% : $\gamma$ conversions	90%: HF non-prompt 10% $\gamma$ conversions
Probe e - $\cancel{T}$	90% : HF non-prompt 10% : hadrons (jets)	90% : HF non-prompt 10% : $\gamma$ conversions

Table 6.7: Comparison of fakes composition in OS and SS regions. difference in 10% cases for anti-tight electrons where OS events have fakes from mis-ID jets. Those measurements are performed in  $CR_2$ .

lepton-enriched CR	<i>real lepton</i>	<i>fake electron</i>	<i>fake muon</i>
nr. jets	2,3,4	2,3,4	2,3,4
nr. b-tagged jets	$\geq 1$ (MV2c10, 70% eff.)	$\geq 1$ (MV2c10, 70% eff.)	$\geq 1$ (MV2c10, 70% eff.)
nr. leptons	2	2	2
lepton charge	OS	SS	SS
lepton flavour	$e\mu, \mu e$	$e\mu, \mu e$	$\mu\mu$
lepton offline sel.	<i>loose</i> (Table 6.4)	<i>loose</i> (Table 6.4)	<i>loose</i> (Table 6.4)
$p_T^\ell$	$\geq 10$ GeV	$\geq 10$ GeV	$\geq 10$ GeV
$ \eta(e) $	$\leq 1.37$	$\leq 1.37$	-
$ \eta(\mu) $	$\leq 2.5$	$\leq 2.5$	$\leq 2.5$
<i>tag lepton sel.</i>	<i>Tight lepton</i> matched to lepton trigger $p_T(\ell) > 25$ GeV	<i>Tight <math>\mu</math></i> matched to $\mu$ trigger $p_T(\mu) > 25$ GeV	<i>Tight <math>\mu</math></i> matched to $\mu$ trigger $p_T(\mu) > 25$ GeV

Table 6.8: Definition of the control regions used for measuring the real and fake lepton efficiency. The same region is used to measure both  $\varepsilon_r(e)$  and  $\varepsilon_r(\mu)$ . Note that we require the leading muon to tag the event, with the sub-leading electron/muon being used as selection-unbiased probe for the fake electron/muon rate measurement.

The subtraction of non-prompt lepton background in the numerator and the denominator of the real efficiency is estimated using all available MC samples:  $t\bar{t}$ ,  $Z$ +jets,  $W$ +jets,  $t\bar{t}Z$ ,  $t\bar{t}W$ , di-boson, single top and rare samples. The charge flip contribution, mainly coming from  $t\bar{t}V$ , is found to be negligible using simulation in the real-enriched control region. The non-prompt lepton background is mainly coming from  $t\bar{t}$  by 74% (14% from  $t\bar{t}V$ , around 5% from  $V$ +jet or single top). Fake leptons from di-boson and rare are negligible. It is chosen to use the simulated non-prompt lepton background of real efficiency for two reasons: First, the associated MC systematics coming from this background are not dominant in the fake estimation. Secondly, the time constraint of the analysis prevents a careful and dedicated study to investigate a data-driven estimate of this background from  $t\bar{t}$  process. Unlike real leptons coming from the pure  $Z \rightarrow \ell\ell$  resonance, there is no obvious way to build the sidebands regions to estimate the background associated to real leptons from  $t\bar{t}$  (no resonance).

The  $p_T$  distribution of each term of the real efficiency for both the numerator and the denominator is shown in Figure 6.10a and 6.10b for electrons and Figure 6.10c and Figure 6.10d for muons. In each bin, data events are shown, as well as MC background normalised to the luminosity. The splitting of  $p_T$  bins are chosen to provide sufficient statistics of data in each bin. The contribution of non-prompt lepton background is dominated by  $t\bar{t}$  background in all  $p_T$  bins. The non-prompt lepton background is dominating at low  $p_T$  then decreases exponentially with respect to  $p_T$  and becomes negligible at  $p_T > 30$  GeV, reflecting the purity of real leptons at  $p_T > 25$  GeV selected in  $CR^{OS}$ . Note that efficiencies at  $p_T < 25$  GeV are not used in the fake estimation but just shown in the plots to get the full picture. The real purity is above 95 % in all  $p_T$  bins above 25 GeV.

**Fake rate ( $\varepsilon_f$ )** Fake rate is measured using the standard T&P method by taking the ratio *loose* to *tight* probe leptons in data in the  $CR_1^{SS}$  and  $CR_2^{SS}$  (Figure 6.9) after subtracting any prompt and charge flip lepton background as shown in Equation (6.13).

$$\varepsilon_f = \frac{N_T^{data\ SS} - N_T^{prompt\ SS} - N_T^{Qflip}}{N_L^{data\ SS} - N_L^{prompt\ SS} - N_L^{Qflip}} \quad (6.13)$$

The *fake* control region definition, shown in Table 6.8, is enriched in non-prompt leptons from  $t\bar{t}$  dileptonic decays by requiring the presence of two same-sign leptons (since all prompt leptons from  $t\bar{t}$  have opposite charge) at low-jet multiplicity where the signal is negligible. For fake electron efficiency, the measurement is performed using opposite flavour pairs ( $e^\pm\mu^\pm$  events) in which the muon is considered to be the tag lepton. For fake muon efficiency, the measurement is performed using  $\mu^\pm\mu^\pm$  events. The choice of lepton flavour is done after a study of the fake purity (see Appendix B.3, Figure B.1, for more details).

The subtraction of prompt lepton background in the numerator and the denominator of the fake rate measurement is performed using MC samples ( $t\bar{t}Z$ ,  $t\bar{t}W$ , di-boson and rare). The subtraction of charge flip is performed using data-driven charge flip estimation described in section 6.3.1.

The  $p_T$  distribution of each term of the fake rate for both the numerator and the denominator is shown in Figure 6.11a and 6.11b for electrons and Figure 6.11c and 6.11d for muons. Data events are shown, alongside the contribution from processes that needs to be subtracted from data. Unlike real efficiency, the charge flip contribution is significant in the fake-enriched control region. It represents 44% of the total background and mainly comes from  $t\bar{t}$  process (at 99%). The prompt lepton background, representing 56% of the total background, is mainly coming from  $t\bar{t}W$ , di-boson by 35%, 28% of the



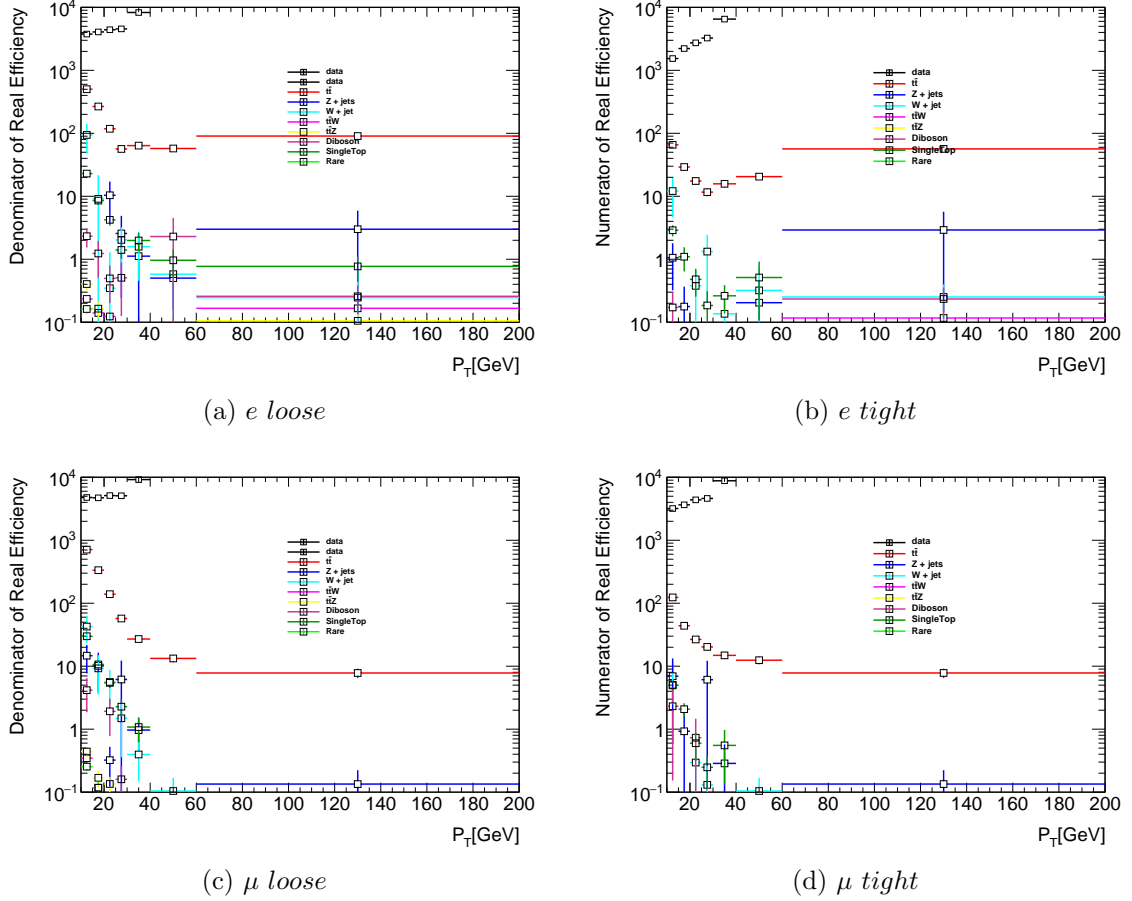
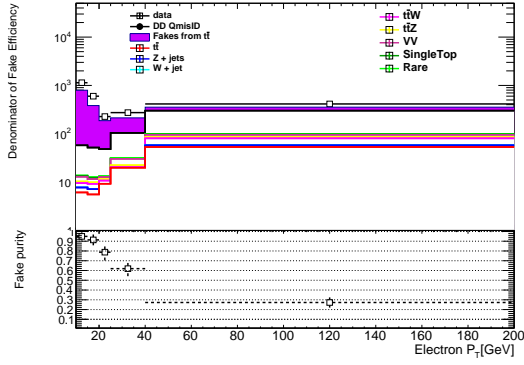
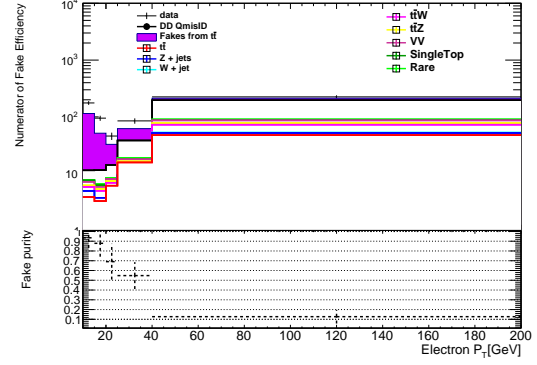


Figure 6.10: *Electron and muon  $p_T$  distribution in the OS control region  $CR_1^{OS}/CR_2^{OS}$  when the lepton passes/fails tight selection in the Numerator/denominator using data and the associated non-prompt background. Error bars account for statistical uncertainty only.*

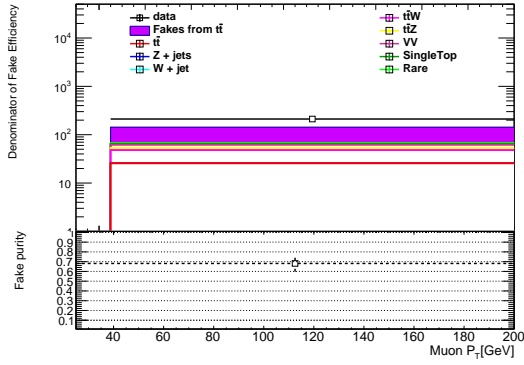
total prompt lepton background respectively. The  $t\bar{t}$  process is also found to be a source of prompt lepton background by 20% because of the mis-assignment of fake lepton probe in real-fake events. The choice of  $p_T$  bins follows the same principle as for real efficiency. In the muon case, it has been observed that the baseline selection adopted is already close to the tight selection. This results in very large fake lepton efficiencies at high  $p_T$  which eventually leads to method instability (as can be deduced from the Matrix Method equations). In order to avoid this, a single  $p_T$  bin for [25,200] GeV range has been adopted for measuring the muon fake rate. Note that efficiencies at  $p_T < 25$  GeV are not used in the fake estimation but just shown in the plots to get the full picture of  $p_T$  dependency.



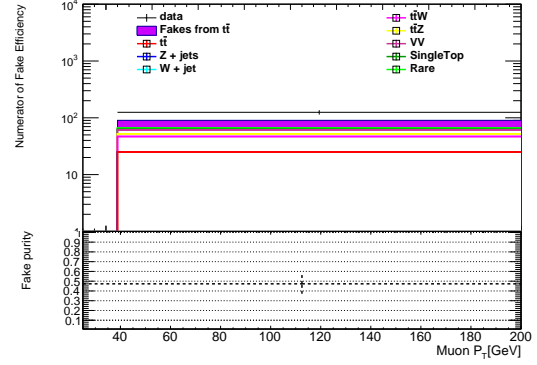
(a)  $e$  loose



(b)  $e$  Tight



(c)  $\mu$  loose



(d)  $\mu$  Tight

Figure 6.11: Electron and muon  $p_T$  distribution in the SS control region (2-3-4 jet multiplicity) when the lepton passes/fails tight selection in the Numerator/denominator using data and the associated prompt/charge flip background. Error bars account for statistical uncertainty only.

**Fake purity** A study carried out on simulation shows that the major source of fakes is the  $t\bar{t}$  semileptonic process, by more than 90% in the SR and more than 80% in the baseline region( $CR_3$ ), with a smaller contribution of  $W$ +jets and - mostly for low jet multiplicity events - single  $top$  and  $tW$  processes. For both electrons and muons, the fake lepton are mainly coming from heavy flavours (HF) decay which indicates they are mainly non-prompt leptons coming from top decay to b-quarks. This study also shows that the fake composition is mostly similar in the SR (*tight* fake lepton at  $N_{jets} \geq 5$ ) and other CRs as shown in Figure 6.12 where the fake lepton truth origin fraction - defined as the ratio of fake leptons coming from a specific source over the total number of fakes - is observed to be stable against the jet multiplicity for both electrons and muons passing *tight* and *loose* selections.

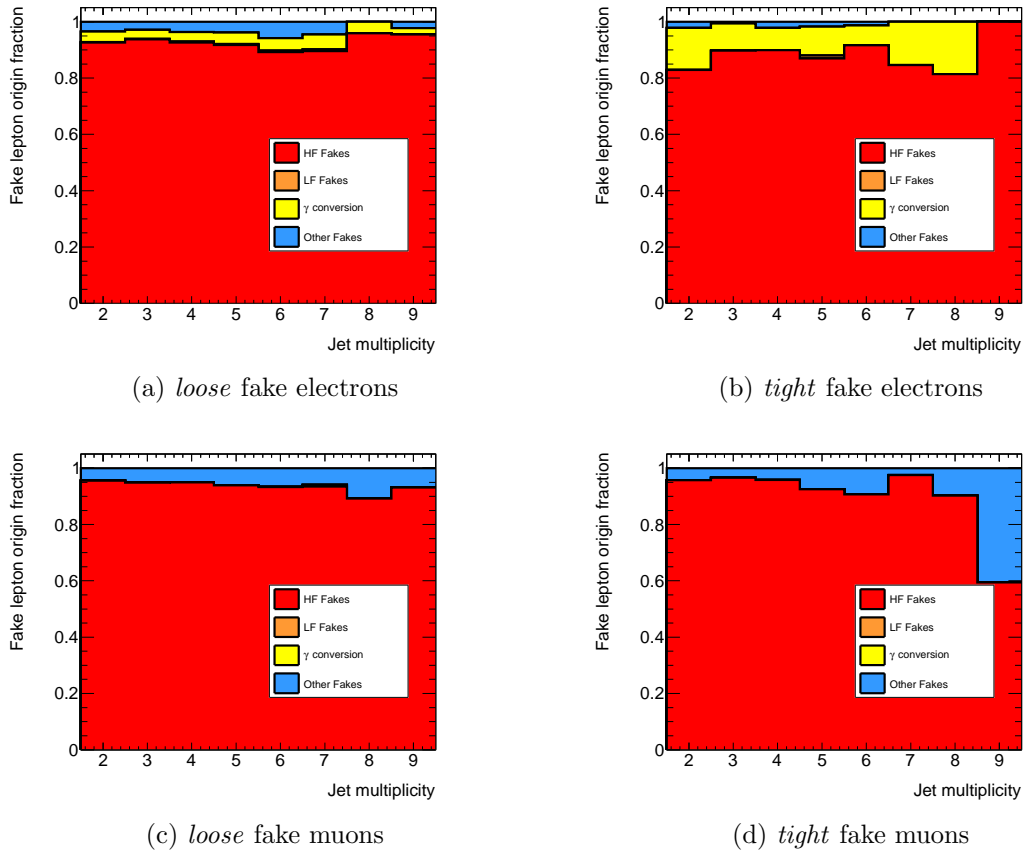


Figure 6.12: Fake lepton truth origin fraction for  $t\bar{t}$  events, as a function of jet multiplicity, for loose and tight offline selection.

**Summary** In order to get a fake prediction dependent on the lepton kinematics, the efficiencies have been factorised in bins of  $p_T$ . No sizeable dependency on other variables has been observed (such as  $\eta$  or  $\Delta R$ ). Figure 6.13 shows the results of real efficiency and fake rate in terms of  $p_T$  for both electron and muon in data before and after background subtraction. In the range of interest ( $p_T > 25$  GeV), electron (muon) real efficiencies are in the range of 70-95 % (90-100 %). Meanwhile, the fake rate values are around 30 % for both electrons and muons. The size of each bin is chosen depending on statistics (low stats at high  $p_T$  for example so the size of the bin becomes bigger at high  $p_T$ ) and  $p_T$  dependency of real efficiency (more flat at high  $p_T$ ). For real efficiency, seven bins are chosen [10,15,20,25,30,40,60,200] GeV. For electron fake rate, five bins are used [10, 15, 20, 25, 40, 200] GeV. However, for muon fake rate, one bin is only used because of the high fake rate for  $p_T > 60$  GeV.

The efficiencies measured from  $t\bar{t}$  MC is also shown for comparison with data after background subtraction. Overall, fair agreement between data and MC is seen for fake rates and more than 96% agreement is gotten for real efficiency for both electron and muon probes at  $p_T > 25$  GeV.

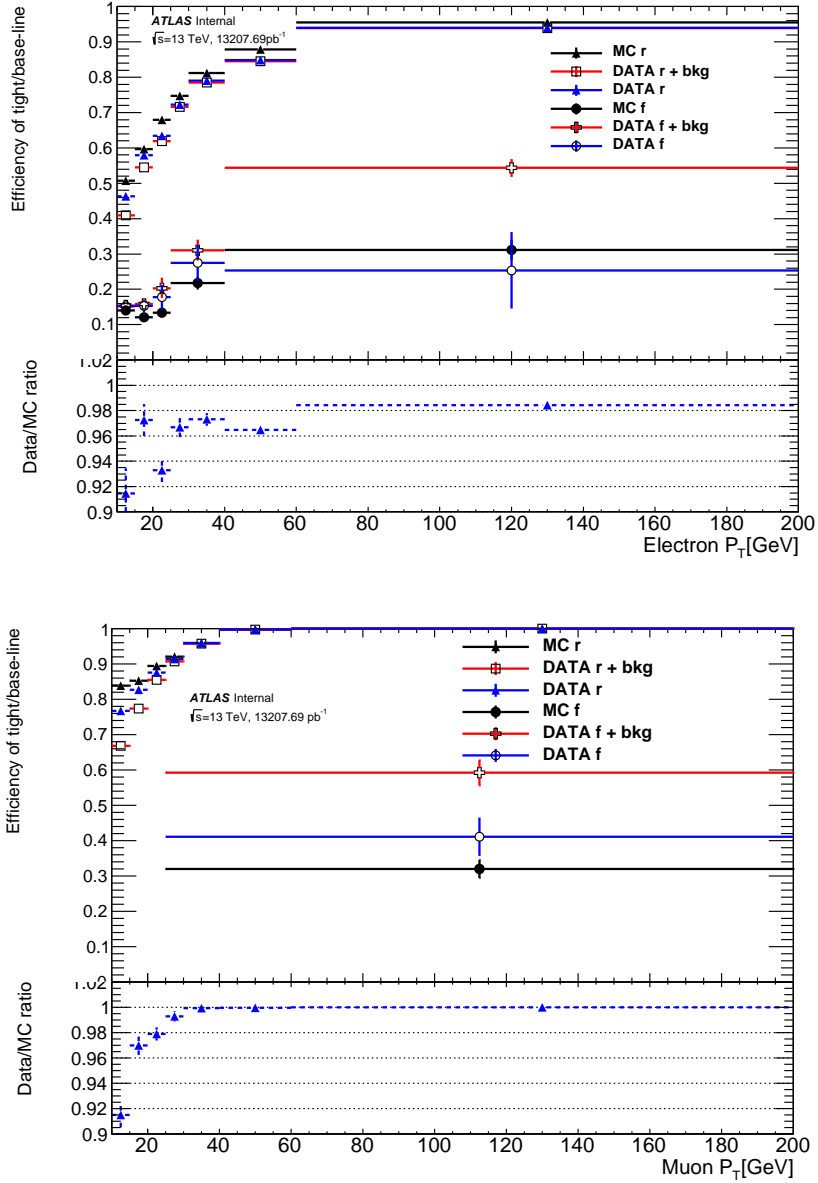


Figure 6.13:  $p_T$  distribution of real efficiency and fake rate for electrons (top) and muons (bottom). The measurements are done for both data and  $t\bar{t}$  MC (black line). In data, the efficiencies are shown before and after the background subtraction by red and blue lines respectively. Data to MC ratio for real efficiencies is shown in the bottom of each plot.

### 6.3.2.3 Validation

After describing the methods used to estimate the fakes and extracting the inputs of the MM, it is mandatory to validate the Matrix Method results with  $t\bar{t}$  Monte-Carlo simulation and on validation region with data.

**Closure test in CR** The closure test is an entirely simulation-based comparison of the number of fake leptons from the MM calculation with the true fake leptons. The latter are obtained by checking the origin and the type of the lepton to not being matched to any prompt or charge flip truth lepton. As highlighted previously, the major source of non-prompt fake leptons are semileptonic  $t\bar{t}$  events, therefore a non-allhadronic  $t\bar{t}$  sample is used. The closure test is performed separately in the three final state categories:  $e^\pm e^\pm$ ,  $\mu^\pm \mu^\pm$  and  $e^\pm \mu^\pm$ . The common event selection are applied in  $CR_2^{SS}$ . It corresponds to the event selection of Table 6.6 but with 2-4 jet multiplicity.

Table 6.9 shows the results of the "inclusive" closure test in the CR. The associated uncertainties are statistical only. Errors on the Matrix Method estimate are statistical uncertainty driven by the size of the CR and the uncertainty on the measured efficiencies which gets propagated to the final event weights. Any non closure observed in the low jet multiplicity regions - which correspond to the Real/Fake control region selection - will account for residual biases in the method.

$t\bar{t}$ MC in CR	$ee$	$\mu\mu$	$e^\pm \mu^\pm$
Simulated fake lepton	$23.0 \pm 2.4$	$30.5 \pm 2.8$	$51.4 \pm 3.4$
Matrix Method fake lepton	$19.4 \pm 1.4$	$27.9 \pm 1.8$	$52.8 \pm 2.4$
Non-closure: $\left(\frac{MM-t\bar{t} MC}{t\bar{t} MC}\right)$	$15.4 \pm 2.5$ [%]	$8.8 \pm 2.2$ [%]	$2.6 \pm 2.1$ [%]

Table 6.9: Total event yields at  $\sqrt{s} = 13$  TeV and at  $L=13.2$  fb $^{-1}$  in the control region (CR) from  $t\bar{t}$  events of simulated fakes, estimated fakes from Matrix Method and the non-closure ratio. Errors for the Matrix Method fakes and on the non-closure ratio are statistical only. The non-closure test is considered as the actual non-closure systematics.

The discrepancy between the central prediction of the Matrix Method and of plain Monte-Carlo is to be quoted as the actual non-closure systematic, and is below 10% for both  $\mu^\pm \mu^\pm$  and  $e^\pm \mu^\pm$  channels, which gives us confidence in the validity of the fake estimation strategy. However, the non-closure reaches 15% in  $e^\pm e^\pm$  channel mainly because of low fake purity at high  $p_T$  electron fake rate (30% (10%) in *loose* (*tight*) region) due to possible mis-assignment of the real tag in real-fake events. This is shown by a study where the number of probes gotten from the T&P method are compared to the number of fake electrons gotten from the truth information in  $CR_2^{SS}$ . A good agreement is seen at  $p_T < 40$  GeV reflecting the reliability of the method at low  $p_T$ . However, at  $p_T > 40$  GeV, a 46 % under-estimation of fake probes is seen which reflects that the probability of picking a real probe electron is still sizeable.

**Closure test in SR** The non closure at high jet multiplicity region will cover up for the extrapolation effects from CR to SR, such as different fake lepton origin fractions as detailed in Figure 6.12. Table 6.10 shows the inclusive closure in the SR for the three  $2\ell ss$  channels. The discrepancy between the central prediction of the Matrix Method and of plain Monte-Carlo is less than 16 %. This gives a good sign of the fake composition

similarity between the CR and SR. In other words, it justifies the hypothesis of similar efficiencies in both CR and SR. This is confirmed by a study of the real efficiency and fake rate dependency of the number of jets where stable efficiencies are seen in both regions for both electron and muon probes as shown in Figure 6.14.

$t\bar{t}$ MC in SR	$ee$	$\mu\mu$	$e^\pm\mu^\pm$
Simulated fake lepton	$4.3 \pm 0.9$	$4.2 \pm 1.1$	$9.4 \pm 1.6$
Matrix Method fake lepton	$4.8 \pm 0.7$	$4.8 \pm 0.7$	$8.8 \pm 0.9$
Non-closure: $\left(\frac{MM-t\bar{t} MC}{t\bar{t} MC}\right)$	$11.9 \pm 7.9$ [%]	$15.5 \pm 13.2$ [%]	$6.7 \pm 5.4$ [%]

Table 6.10: Total event yields at  $\sqrt{s} = 13$  TeV and at  $L=13.2$  fb $^{-1}$  in the signal region (SR) from  $t\bar{t}$  events of simulated fakes, estimated fakes from Matrix Method and the non-closure ratio. Errors for the Matrix Method fakes and on the non-closure ratio are statistical only. The non-closure test is considered as one of the systematics of CR-SR extrapolation.

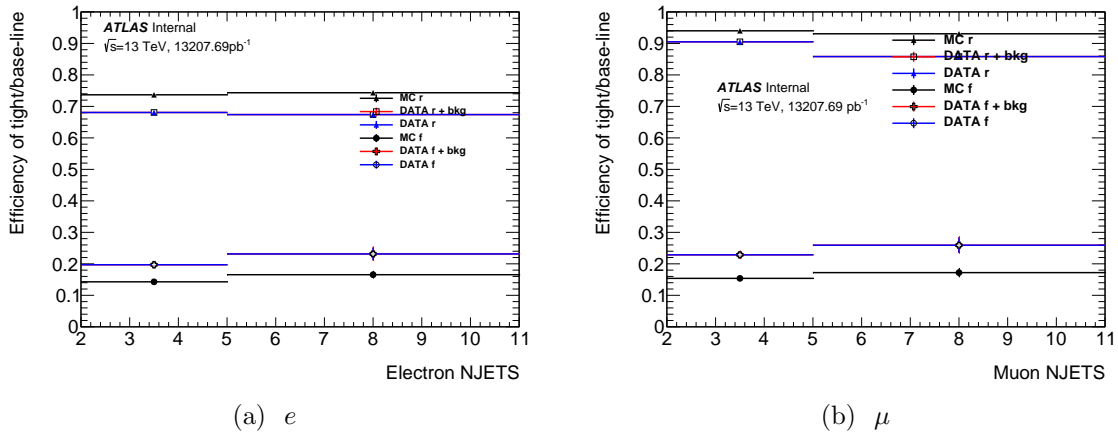


Figure 6.14: Stability check of electron and muon fake efficiencies in low and high jet multiplicity. Error bars account for statistical uncertainty only.

**Validation with data** Figure 6.15 presents the validation regions (2-3-4 jet multiplicity) in the three  $2\ell ss$  channels using data. Prompt leptons are estimated using Monte-Carlo simulation. Charge flip and fake lepton backgrounds are estimated using the *tight* charge flip rate and Matrix Method on data. Error bars account for statistical uncertainty only. A very good agreement between prediction and data is observed in all  $2\ell ss$  channels which give us a confidence of the followed process to estimate fakes in  $2\ell ss$  channel using the Matrix Method.



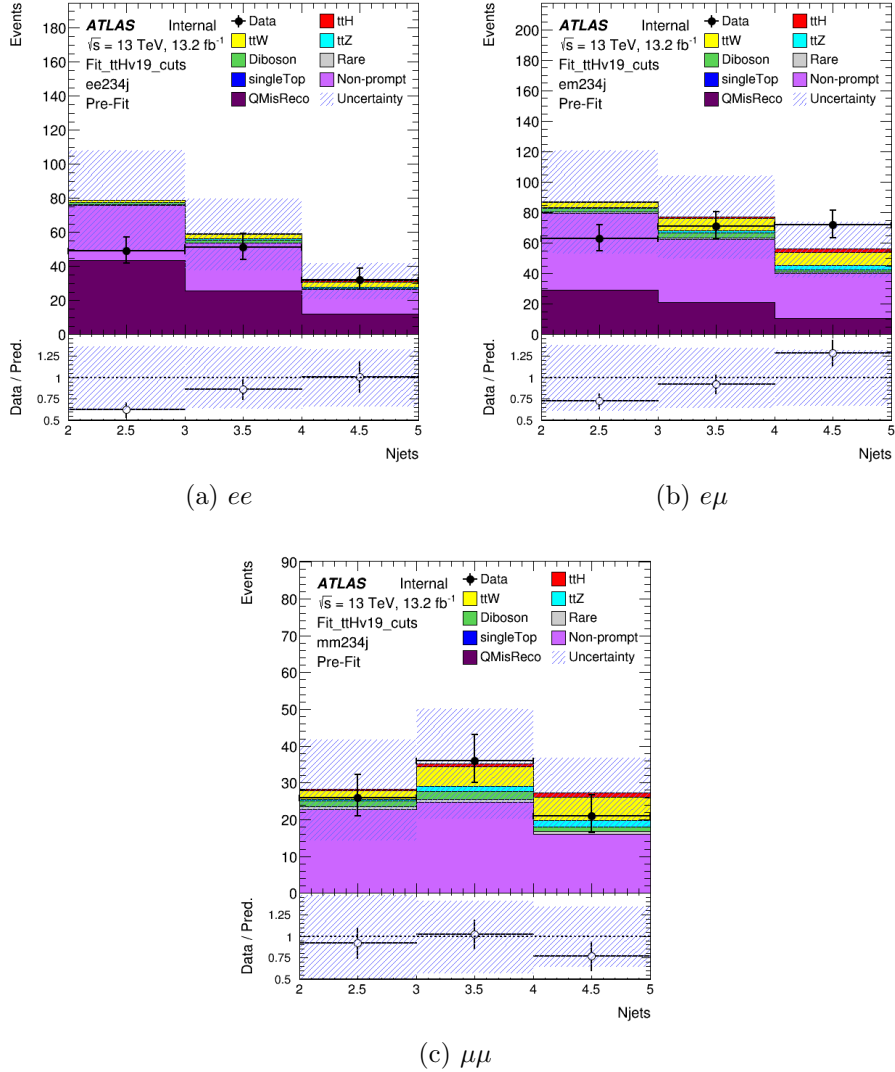


Figure 6.15: Validation plots in the  $2lss$  channels in the low number of jet validation region (2-3-4 jet multiplicity) using data. Prompt leptons are estimated using simulation. Charge flip and fake lepton background are estimated using data-driven methods. Error bars account for statistical uncertainty only.

### 6.3.2.4 Results

Now that the method has been validated on  $t\bar{t}$  MC samples and in validation regions, fake events can be estimated in the signal region.

Table 6.11 summarises all the yields contributing in the matrix method equations.  $TT_i$  region defines the SR. It is worth to notice that the yields presented in the table are counted after data-driven charge flip subtraction (estimated from data in section 6.3.1).

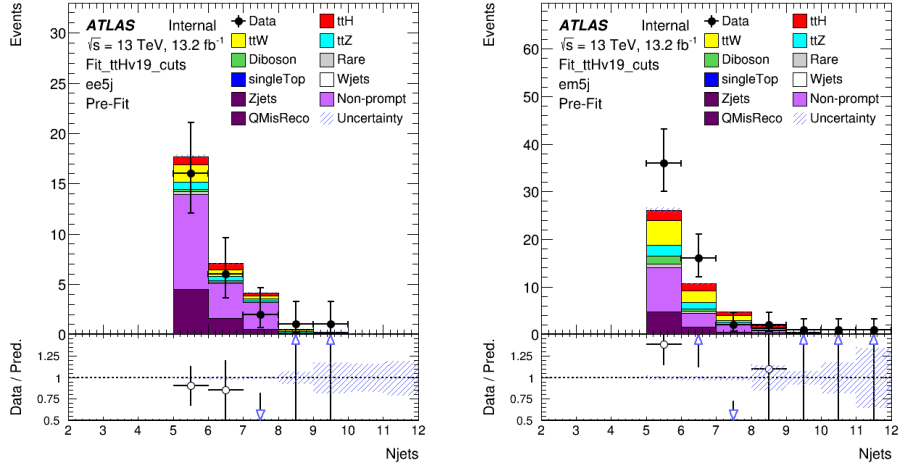
Sideband	Nb of events		
	$e^\pm e^\pm$ - Qflip	$e^\pm \mu^\pm$ - Qflip	$\mu^\pm \mu^\pm$
$TT$	19.1	51.9	31
$T\cancel{T} + \cancel{T}T$	34.0	33.6	14
$\cancel{T}\cancel{T}$	-0.2	2.8	0

Table 6.11: Number of events in the signal region  $TT$  and three sideband regions  $T\cancel{T}$ ,  $\cancel{T}T$  and  $\cancel{T}\cancel{T}$  after charge flip data-driven subtraction.

Figure 6.16 shows the  $N_{\text{jets}}$  distribution in the signal region with the data-driven fake lepton background estimation in the three  $2\ell$ ss channels. The reducible background estimation in the three  $2\ell$ ss channels including both fake lepton and charge flip data driven backgrounds is shown in Table 6.12. The total background includes prompt lepton background, from  $t\bar{t}W$ ,  $t\bar{t}Z$ , di-boson and others, in addition to the reducible background. One can note that the reducible background dominates in all  $2\ell$ ss channels, especially in the  $e^\pm e^\pm$  channels.

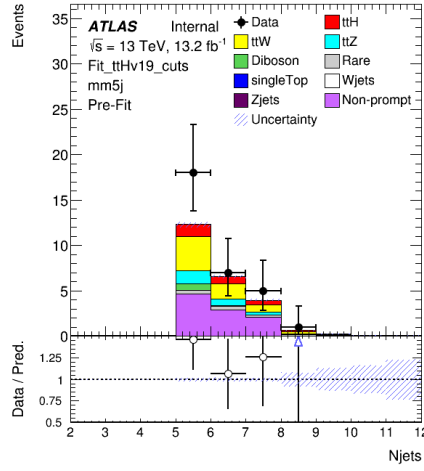
Channel	$e^\pm e^\pm$	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$
Fakes	$15.5 \pm 3.3$	$14.0 \pm 3.3$	$9.7 \pm 2.8$
Charge flip	$6.9 \pm 1.3$	$7.1 \pm 1.7$	$0.0 \pm 0.0$
Total background	28.5	40	21
$B_{\text{red}}/B_{\text{tot}}$	54 %	35 %	46 %

Table 6.12: Reducible background estimation in the three  $2\ell$ ss channels including both fake and charge flip data driven backgrounds. The total background include prompt lepton background, from  $t\bar{t}W$ ,  $t\bar{t}Z$ , di-boson and others, in addition to the reducible background. Errors on fakes are statistical only. However, errors on the charge flip includes both statistical and systematical uncertainties.



(a)  $ee$

(b)  $e\mu$



(c)  $\mu\mu$

Figure 6.16: Fake estimation in the  $2lss$  channels in the high number of jet signal region (at least 5 jets) using matrix method. Prompt leptons are estimated using MC simulation, while charge flip and fakes are estimated using charge flip rate and Matrix Method on data. Error bars account for statistical uncertainty only.

### 6.3.3 Comparison with the Fake Factor method

Fake Factor method (FF) is used as baseline for the first public results with  $13.2 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$  [111]. It is a simplification of the MM under some assumptions:

- Inclusive kinematic measurements of the fake factor. Hence, only one  $(p_T, \eta)$  bin is used.
- All *loose* leptons pass *tight* requirement in the real-enriched CR. Hence, real efficiencies is set to one ( $\varepsilon_r = 1$ ).
- events that fail *tight* requirements  $\mathcal{X}\mathcal{X}$  in the baseline  $CR_3$  are neglected.

The use of a fake factor method is motivated by the fact that the final sensitivity is extracted using a limited number of signal regions and where there is no need to predict the background shapes (use of counting experiment). It was also chosen for its robustness in a context of high luminosity rising (see Figure 3.13).

Assuming that  $\varepsilon_r = 1$ , Equation (6.5) is simplified to:

$$f_{SR} = \frac{\varepsilon_f}{1 - \varepsilon_f} (T\mathcal{X} + \mathcal{X}T) - \frac{\varepsilon_f^2}{(1 - \varepsilon_f)^2} \mathcal{X}\mathcal{X} \quad (6.14)$$

The fake factor is defined as the ratio between the number of same-sign events with *tight* leptons only ( $TT$ ) and events with one *anti-tight* lepton ( $T\mathcal{X}$ ).

$$\theta = \frac{N_{TT}^{SS}}{N_{T\mathcal{X}}^{SS}} = \frac{N_{TT}^{SS}}{N_{T\mathcal{L}}^{SS} - N_{TT}^{SS}} = \frac{\varepsilon_f}{1 - \varepsilon_f} \quad (6.15)$$

It is measured in the fake-enriched control region  $CR^{SS}$  based on the assumption that the fake factor is stable with respect to the jet multiplicity.

By using the equivalence between the fake factor  $\theta$  and the fake rate  $\varepsilon_f$ , the number of fakes could be written in terms of the fake factor as shown in Equation (6.16).

$$f_{SR} = \theta(T\mathcal{X} + \mathcal{X}T) - \underbrace{\theta^2 \mathcal{X}\mathcal{X}}_{\text{neglected in FF}} \quad (6.16)$$

The term including the second order of  $\theta$  is neglected in the fake factor method since the number of events where both of leptons fail the *tight* selection is very small in the  $2\ell_{SS}$  channel (see Table 6.11).

The measurements of the fake factor is dependent on di-lepton flavours. Hence two values of  $\theta_\ell$  where  $\ell$  denotes electrons or muons is computed using the  $e^\pm e^\pm$  and  $\mu^\pm \mu^\pm$  channels respectively after prompt and charge flip background subtraction:

$$\theta_e = \frac{N_{T_e T_e}^{Data} - N_{T_e T_e}^{\text{Prompt SS}} - N_{T_e T_e}^{\text{Qflip}}}{N_{T_e \mathcal{X}_e}^{Data} - N_{T_e \mathcal{X}_e}^{\text{Prompt SS}} - N_{T_e \mathcal{X}_e}^{\text{Qflip MC}}} \quad (6.17)$$

$$\theta_\mu = \frac{N_{T_\mu T_\mu}^{Data} - N_{T_\mu T_\mu}^{\text{Prompt SS}}}{N_{T_\mu \mathcal{X}_\mu}^{Data} - N_{T_\mu \mathcal{X}_\mu}^{\text{Prompt SS}}} \quad (6.18)$$

The charge flip background in  $TT$  events is estimated using the data, as done in MM, and noted  $N_{T_e T_e}^{\text{Qflip}}$ . However, the charge flip background in  $T\mathcal{X}$  events is estimated with the simulation where the electrons (in  $e^\pm \mu^\pm$ ) or the two electrons (in  $e^\pm e^\pm$ ) match real prompt leptons (using a  $\Delta R < 0.2$  distance) with one mis-measured lepton charge are counted and noted  $N_{T_e \mathcal{X}_e}^{\text{Qflip MC}}$ .

The results are presented in Table 6.13 with the associated systematic errors  $\Delta\theta$ . The same Table shows how one could compute the fake rate and the associated uncertainties from Equation (6.15). Values are very close to the fake rate estimation by the MM presented in Figure 6.13 by averaging along all  $p_T$  bins. The non-closure measured with simulated  $t\bar{t}$  events is within 25% or better as shown in Table 6.14.

Lepton flavour	$\theta_\ell \rightarrow \varepsilon_f$	$\Delta\theta \rightarrow \Delta\varepsilon_f$
Electron	0.36 $\rightarrow$ 0.26	24.79% $\rightarrow$ 18.3%
Muon	0.64 $\rightarrow$ 0.39	9.76 % $\rightarrow$ 5.95%

Table 6.13: Fake rate derivation from the fake factors computation using the FF in one inclusive  $p_T$  bin in the  $2\ell ss$  channel.  $\Delta\theta$  and  $\Delta\varepsilon_f$  denote the systematics error associated to the fake factor  $\theta_\ell$  and fake rate  $\Delta\varepsilon_f$  measurements respectively.

Channel	$e^\pm e^\pm$	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$
MM non-closure	12%	7%	15%
FF non-closure	25%	19%	10%

Table 6.14: Comparison of the non-closure in the SR between Matrix and Fake Factor methods to estimate the extrapolation systematics associated to the fake estimate of  $2\ell ss$  channel.

In order to estimate fake events in the SR using FF, charge flip and SS prompt lepton background are excluded from the sideband yields ( $T\mathcal{X}$  and  $\mathcal{X}T$  events) using simulation for the  $e^\pm e^\pm$ ,  $\mu^\pm \mu^\pm$  and  $e^\pm \mu^\pm$ :

$$f_{SR}^{ee} = (N_{T_e \mathcal{X}_e}^{Data} - N_{T_e \mathcal{X}_e}^{\text{Prompt SS}} - N_{T_e \mathcal{X}_e}^{\text{Qflip MC}})(\geq 5\text{jets}) \times \theta_e \quad (6.19)$$

$$f_{SR}^{\mu\mu} = (N_{T_\mu \mathcal{X}_\mu}^{Data} - N_{T_\mu \mathcal{X}_\mu}^{\text{Prompt SS}})(\geq 5\text{jets}) \times \theta_\mu \quad (6.20)$$

$$f_{SR}^{e\mu} = N_{T_e \mathcal{X}_\mu}(\geq 5\text{jets}) \times \theta_\mu + N_{T_\mu \mathcal{X}_e}(\geq 5\text{jets}) \times \theta_e \quad (6.21)$$

The fake estimation using FF in  $2\ell ss$  is presented in Table 6.15 for the three channels  $e^\pm e^\pm$ ,  $e^\pm \mu^\pm$  and  $\mu^\pm \mu^\pm$ . The uncertainties include all possible sources from:

- statistical uncertainty on  $\theta$  due to the limited size of the fake enriched sample in low jet multiplicity regions ( $\theta_\ell$  stat).
- validity of the extrapolation from low jet multiplicity to high jet multiplicity region tested on simulated  $t\bar{t}$  events ( $\theta_\ell$  syst). A comparison with MM is proposed in Table 6.14. Overall, the closure test in MM is better or similar.
- prompt and charge flip background subtraction.
- Other processes producing fake leptons that could contribute (eg.  $W$ +jets) and be underestimated (Fake sample composition syst). This is checked by a stability test of the non-prompt lepton background estimate in the signal region by varying the minimum  $p_T$  of the selected  $b$ -jet, restricting to two- or three-jet events only, or imposing a minimum  $E_T^{miss}$  requirement.

Table 6.15 also shows that the fake estimation using FF is compatible with the MM method within the uncertainties. The relative statistical uncertainties contributes by around 20 % for all channels in FF and compatible with the MM in the  $e^\pm e^\pm$  and  $e^\pm \mu^\pm$  channels. However, it is much lower in the  $\mu^\pm \mu^\pm$  for FF by factor of 50% comparing to MM statistical error.

On top of that, a comparison between MM and simulated fake lepton estimation is also shown in the same Table. A significant mis-modeling is seen from the simulated  $t\bar{t}$  sample. This reflects the importance of using Data-driven methods to estimate the fake lepton background.

Channel	$e^\pm e^\pm$	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$
MM Fakes	$15.5 \pm 3.3$	$14.0 \pm 3.3$	$9.7 \pm 2.8$
FF Fakes	$12.1 \pm 2.9 \pm 5$	$12.4 \pm 2.3 \pm 4.5$	$8.7 \pm 1.6 \pm 3.0$
$t\bar{t}$ -PP6 Fakes	$4.3 \pm 0.9$	$9.4 \pm 1.5$	$4.2 \pm 1.1$
DD/MC ratio	3.6	1.5	2.3

Table 6.15: Comparison between Matrix and Fake Factor methods to estimate fake lepton events in  $2\ell ss$  channel. A comparison with MM and simulated fake lepton estimation is also shown by " DD/MC ratio". "  $t\bar{t}$ -PP6" refers to the simulated  $t\bar{t}$  MC sample generated by POWHEG+PYTHIA 6. Errors are statistical only for MM and  $t\bar{t}$ -PP6 and statistical + systematics for FF.

As a conclusion, MM results for fake lepton background estimation provide an excellent cross-check and validation tool that gives confidence in the fake factor method, the baseline method used for the first public results at  $13.2 \text{ fb}^{-1}$ . This is particularly important since fake leptons are the dominant background in this analysis. Currently, the MM is chosen to be the baseline method with the full 2015-2016 data ( $36.1 \text{ fb}^{-1}$ ) thanks to its flexibility and accuracy. In the following, only fake factor method results is used.

## 6.4 Validation regions for prompt leptons background

The number of prompt leptons background, namely  $t\bar{t}W$ ,  $t\bar{t}Z$ , diboson and other, are predicted using MC after validating those processes in different regions described in Appendix Table B.1. Some examples of validation prompt background in the validation regions is shown in Figure 6.17. A validation region for  $t\bar{t}Z$  is defined by inverting the  $Z$  veto of the  $3\ell$  signal region, and in addition tightening the jet selection by requiring  $\geq 4$  jets of which  $\geq 2$  are  $b$ -tagged. To gain additional events for comparison, a looser selection ( $\geq 4$  jets of which at least one is  $b$ -tagged, or 3 jets of which at least 2 are  $b$ -tagged) is also used; the latter includes a larger fraction of  $WZ$  events. Invariant mass plots for these VRs are shown in Figure 6.17a and 6.17b where a good agreement between data and expectation are shown in both *tight* and *loose* regions where  $t\bar{t}Z$  purity achieves 68 and 58 % respectively. A validation region for  $t\bar{t}W$  production is defined by using the  $2\ell 0\tau_{\text{had}}$  same-charge lepton selection, requiring either two or three jets of which at least two are  $b$ -tagged, and requiring the scalar sum of jet transverse momenta to exceed 220 GeV for  $ee$  and  $e\mu$  events. For  $ee$  events, missing transverse energy  $E_{\text{T}}^{\text{miss}} > 50$  GeV is required, and  $M(ee)$  cannot lie in the range  $[75, 105]$  GeV. Even with these stringent requirements, the purity of  $t\bar{t}W$  in this region is only 23%. The number of electrons and the number of jets of  $t\bar{t}W$  validation region are presented in Figures 6.17c and 6.17d respectively. Some discrepancies are seen in this validation region in particular for  $\mu\mu$  channel with 4 jets in the final state.



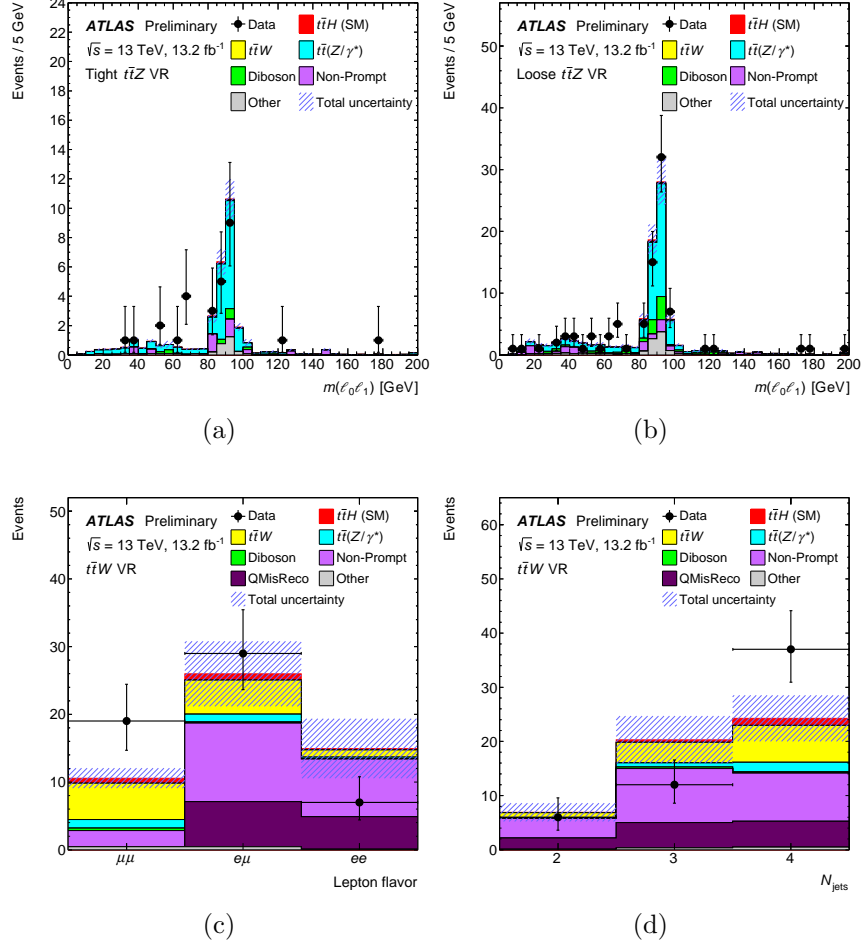


Figure 6.17:  $t\bar{t}Z$  and  $t\bar{t}W$  validation regions: Invariant mass of leptons 0 and 1 for the a) tight and b) loose  $t\bar{t}Z$  validation regions described in Table B.1. The leptons are labeled in the same way as for the  $3\ell$  signal region. Events away from the Z peak are those satisfying the Z selection with leptons 0 and 2. The number of electrons c) and the number of jets d) of  $t\bar{t}W$  validation region described in Table B.1. Non-prompt lepton backgrounds are estimated using data with Fake Factor method as described in Section 6.3.3. Charge misreconstruction backgrounds (indicated as “QMisReco”) are estimated using data as described in Section 6.3.1.

## 6.5 Fit results in $2\ell ss$ channel

### 6.5.1 Likelihood function

The signal strength parameter  $\mu_{t\bar{t}H}^{2\ell ss}$  is defined as the ratio between the observed number of events and the expected number of signal events from the Standard Model (SM) Higgs boson produced in association with a pair of top quarks ( $t\bar{t}H$ ) decaying to a multi-lepton final state. It is obtained by a likelihood function built from a Poisson term for the total expected event yield in the signal regions. A set of Gaussian terms are used for each of the sources of systematic uncertainties considered to be fully correlated between different sources of background and/or signal. However, Poisson terms are instead used for any sources of uncertainties that are of statistical nature. Hypothesis testing is performed with the corresponding one-sided profile likelihood ratio [148]. The 95% confidence level (CL) upper limits on  $\mu_{t\bar{t}H}$  using the  $CL_s$  method [149, 148] is obtained using a likelihood function as well, where the test statistic (the profile likelihood ratio) is computed against the  $\mu_{t\bar{t}H} = 0$  hypothesis.

### 6.5.2 Systematics

Systematic uncertainties on the expected yields are called nuisance parameters (NP or  $\theta$ ). Several nuisance parameters are considered in the fit. The preliminary uncertainty on the combined 2015+2016 integrated luminosity is 2.9%. It is derived, following a methodology similar to that detailed in Refs. [150] and [151], from a preliminary calibration of the luminosity scale using  $x$ - $y$  beam-separation scans performed in August 2015 and May 2016. Because this impacts both the estimation of the prompt backgrounds and the conversion from observed yields to an effective  $t\bar{t}H$  cross section, it has an impact on  $\mu_{t\bar{t}H}$  of nearly 20%.

The most important detector-related systematic uncertainty arises from the efficiency of the jet-to-vertex association method (section 4.6) of approximately 2.5% per jet with  $p_T < 60$  GeV, which becomes important in high-jet-multiplicity final states such as the ones considered here. The uncertainties in the modeling of pile-up interactions and the jet energy scale also contribute significantly. The effect of uncertainties in the modeling of pile-up are determined by varying the assumed inelastic cross section by  ${}^{+16}_{-6}$ %. Uncertainties in lepton reconstruction and trigger efficiencies have negligible impact.

The most important uncertainty arising from theoretical predictions is on the modeling of the acceptance for  $t\bar{t}W$  events with high jet multiplicity requirements. Systematic uncertainties on the acceptance for the  $t\bar{t}V$  ( $V = W$  or  $Z/\gamma^*$ ) backgrounds are derived using MC event simulation. Variations of hard process renormalisation/factorisation scale and PDF uncertainties are considered, as are uncertainties in the A14 parton shower tune for the  $t\bar{t}V$  samples. For  $t\bar{t}W$  only, a comparison between NLO event generation and a leading-order merged calculation is done to test possible matrix element-parton shower matching effects. Due to the small contributions of the  $t\bar{t}t\bar{t}$ ,  $t\bar{t}W^+W^-$ , and  $t b j Z$  processes, only overall cross section uncertainties are considered. The  $\mu_{t\bar{t}H}$  also depends on the assumed SM cross section and acceptance for  $t\bar{t}H$  production. The systematic uncertainties on the  $t\bar{t}H$  signal process are obtained similarly to those for the  $t\bar{t}V$  backgrounds. The uncertainties in these quantities have an impact on  $\mu_{t\bar{t}H}$  comparable to the equivalent uncertainties in  $t\bar{t}W$  and  $t\bar{t}Z$  production.

The systematic uncertainties on the fake lepton background estimate, using the fake factor method, is described in section 6.3.3. The systematic uncertainties on the charge flip background estimate is mentioned in section 6.3.1.

### 6.5.3 Signal strength

The expected backgrounds,  $t\bar{t}H$  signal, and observed data yields in each category, presented in Table 6.16, are needed to measure  $\mu_{t\bar{t}H}^{2\ell ss}$ . In each  $t\bar{t}H$  signal category, the fake (non-prompt) lepton and QmisId background are measured using data-driven method explained in section 6.3. The number of prompt lepton background, namely  $t\bar{t}W$ ,  $t\bar{t}Z$ , diboson and other, are predicted using MC. The fake lepton background is dominating in the three channels followed by charge flip and prompt lepton background from  $t\bar{t}W$  process. The errors of the total background is driven by the fake lepton background uncertainty. The  $t\bar{t}H$  events represent 8%, 13% and 15% of the total SM prediction in  $e^\pm e^\pm$ ,  $e^\pm \mu^\pm$  and  $\mu^\pm \mu^\pm$  channels respectively. Prediction is in good agreement with data for  $e^\pm e^\pm$  channel. However, some deviation between data and prediction is seen in the other two channels by around 30%.

Table 6.16: *Expected and observed yields in the three signal region categories. the fake (non-prompt) and charge flip background are measured using data-driven method explained in section 6.3. The number of prompt background, namely  $t\bar{t}W$ ,  $t\bar{t}Z$ , diboson and other, are predicted using MC. Uncertainties in the background expectations due to systematic effects and limited MC statistics are shown. Values are obtained pre-fit, i.e., using the initial values of background systematic uncertainty nuisance parameters.*

	$2\ell 0\tau ee$	$2\ell 0\tau e\mu$	$2\ell 0\tau \mu\mu$
$t\bar{t}W$	$2.9 \pm 0.7$	$9.1 \pm 2.5$	$6.6 \pm 1.6$
$t\bar{t}(Z/\gamma^*)$	$1.55 \pm 0.29$	$4.3 \pm 0.9$	$2.6 \pm 0.6$
Diboson	$0.38 \pm 0.25$	$2.5 \pm 1.4$	$0.8 \pm 0.5$
Non-prompt leptons	$12 \pm 6$	$12 \pm 5$	$8.7 \pm 3.4$
Charge misreconstruction	$6.9 \pm 1.3$	$7.1 \pm 1.7$	—
Other	$0.81 \pm 0.22$	$2.2 \pm 0.6$	$1.4 \pm 0.4$
Total background	$25 \pm 6$	$38 \pm 6$	$20 \pm 4$
$t\bar{t}H$ (SM)	$2.0 \pm 0.5$	$4.8 \pm 1.0$	$2.9 \pm 0.6$
Data	26	59	31

A global fit to the observed and expected yields in all signal regions is performed to extract  $\mu_{t\bar{t}H}^{2\ell ss}$ . All systematic uncertainties are included in this fit with appropriate correlations between signal regions. The impact of the most important systematic uncertainties on  $\mu_{t\bar{t}H}^{2\ell ss}$ , evaluated after the global fit, is shown in Figure 6.18. Due to correlations between the different sources of uncertainties, the total systematic uncertainty can be different from the sum in quadrature of the individual sources.

The best-fit value of  $\mu_{t\bar{t}H}^{2\ell ss}$ , combining all channels, is  $4.0 \pm_{-1.1}^{+1.2}$  (stat)  $\pm_{-1.3}^{+1.7}$  (syst). A slight excess is seen in  $2\ell ss$  channel mainly from  $e^\pm \mu^\pm$  channel. The uncertainties on  $\mu_{t\bar{t}H}$  are more than 100 % in  $2\ell ss$  channel. The contribution of statistical and systematical uncertainties are equivalent giving the available luminosity. In order to increase the sensitivity on  $\mu_{t\bar{t}H}$ , a combination of all available channels is needed as will be shown in the next section.

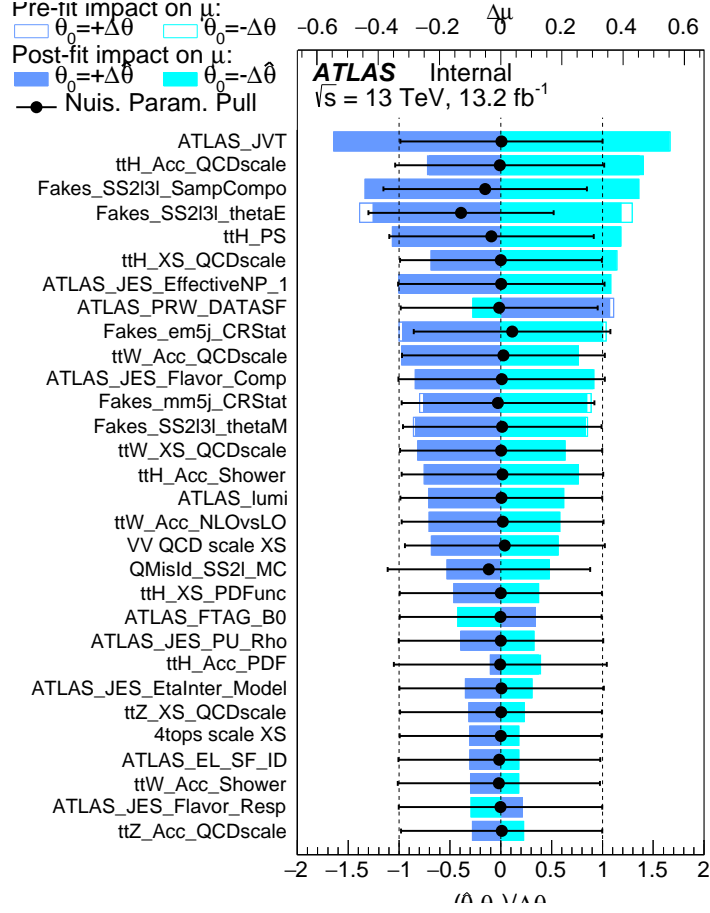


Figure 6.18: *Effect of the thirty most important systematic uncertainty nuisance parameters  $\theta$  on the signal strength  $\mu = \mu_{ttH}^{2lss}$  and constraints on the nuisance parameters from the fit. The blue and cyan bars show the  $\pm 1\sigma$  impact of the nuisance parameter on the signal strength (shown on the top axis). The points and associated error bars show the best-fit values of the nuisance parameters and post-fit uncertainties on the nuisance parameters (shown on the bottom axis). The open bars show the effect of the systematic uncertainties on  $\mu_{ttH}^{2lss}$  before the fit, and the solid bars show the effect after the fit. The nuisance parameters are initially normalised to  $0 \pm 1$ . The dotted vertical lines show  $\pm 1\sigma$  excursions of the nuisance parameters from their initial values. The systematics associated to the fake lepton background estimation in the  $2lss$  and  $3l$  channels originating from non-prompt sample variation and electron/muon transfer factors estimation are indicated by "Fakes\_SS2l3l\_SampCompo" and "Fakes\_SS2l3l\_thetaE/M" respectively. Fake lepton systematics coming from the control region size of the  $mm$  and  $em$  channels are indicated by "Fakes\_mm5j" and "Fakes\_em5j". "XS" indicates the cross section. "Acc" indicates the acceptance. "PS": the parton shower. "NP\_1": first ranked nuisance parameter. "PRW\_DATASF": pre-reweighted data scale factors. "FTAG\_B0": First  $b$  tag parameter. "ATLAS\_JES\_Flavor\_Resp" reflects the jet energy scale response to separate quarks from gluon by measuring the quark-gluon jets proportional.*

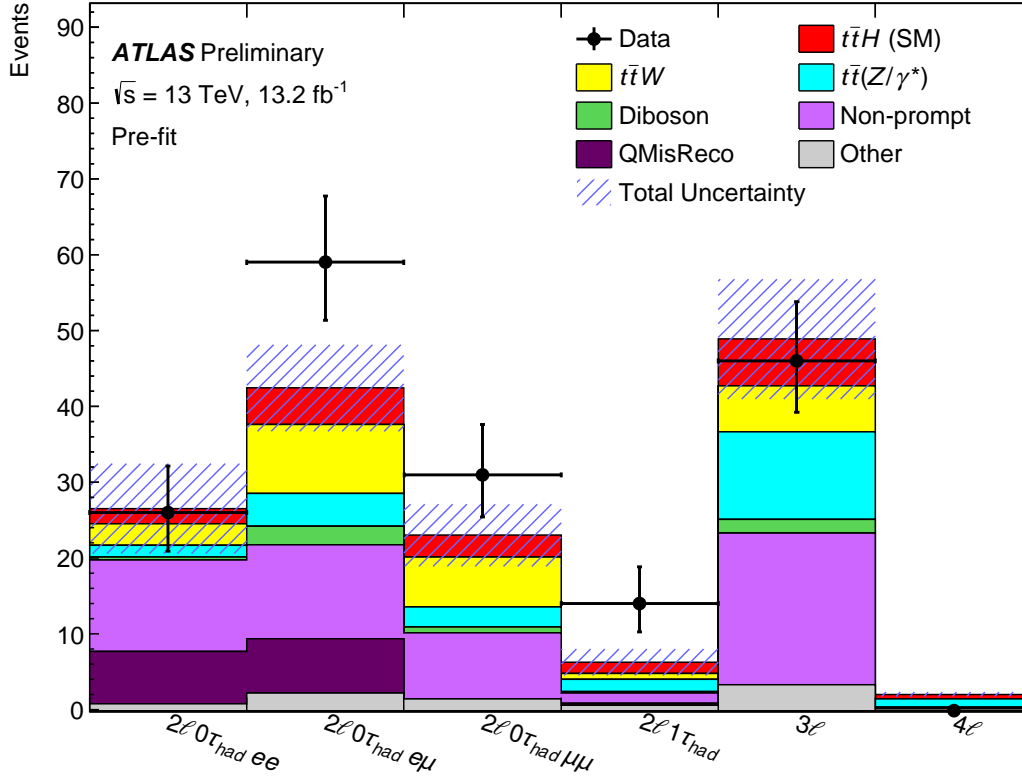


Figure 6.19: Pre-fit background and signal predictions and observed data yields for each signal region. The  $t\bar{t}H$  prediction corresponds to the SM expectation ( $\mu_{t\bar{t}H} = 1$ ). Charge mis-reconstruction backgrounds are indicated as “QMisReco.”

## 6.6 First $t\bar{t}H$ combination with 13 TeV data

### 6.6.1 $t\bar{t}H$ Multilepton

The signal strength parameter  $\mu_{t\bar{t}H}$  is obtained by combining six measurements of the signal strength per channel. A global fit to the observed and expected yields in all signal regions, shown in Figure 6.19, is performed to extract  $\mu_{t\bar{t}H}$ . All systematic uncertainties are included in this fit with appropriate correlations between signal regions. The background composition in each region is represented on the Figure 6.20. In each  $t\bar{t}H$  signal category, the fake (non-prompt) and charge flip background are measured using data-driven method similarly to  $2\ell ss$  channel explained in section 6.3 except for  $4\ell$  channel where MC simulation corrected by non-prompt scale factors (NPSFs), determined from data-MC comparison in trilepton control regions, is used to estimate fakes instead FF/MM methods because of the constrained statistics observed in this channel. The fake and charge flip background are dominating in  $e^\pm e^\pm$ ,  $e^\pm \mu^\pm$  and  $3\ell$  channels. The charge flip background is negligible and not estimated in  $\mu^\pm \mu^\pm$ ,  $3\ell$  and  $4\ell$  channels.  $t\bar{t}Z$  background increases proportionally to the number of leptons and dominates in  $4\ell$ .

The best-fit value of  $\mu_{t\bar{t}H}$ , combining all channels, is  $2.5 \pm 0.7$  (stat)  ${}^{+1.1}_{-0.9}$  (syst). For the  $4\ell$  channel, the observation of zero events makes it difficult to quote a best-fit result with meaningful uncertainties, hence a 68% confidence level  $CL_s$  upper limit is shown instead.

The combined observed (expected) value of 95% CL upper limit on  $\mu_{t\bar{t}H}$  is estimated

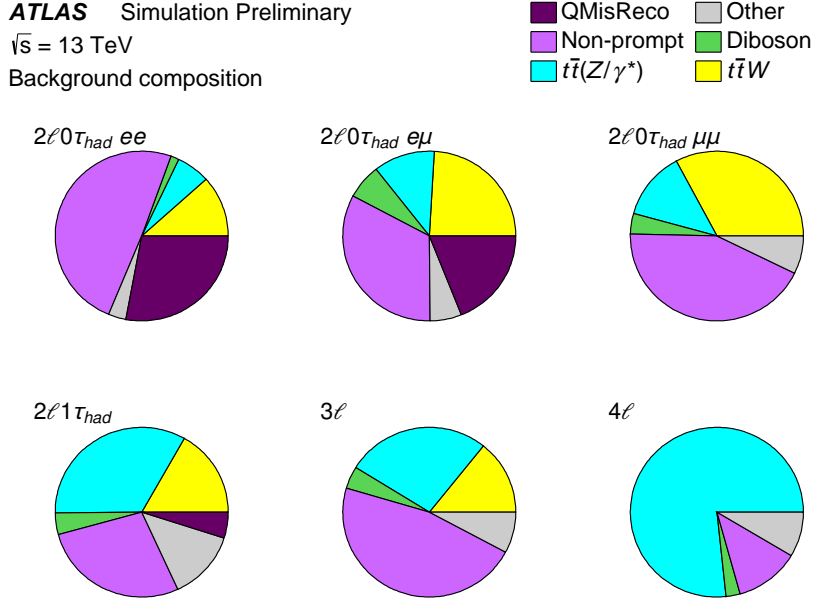
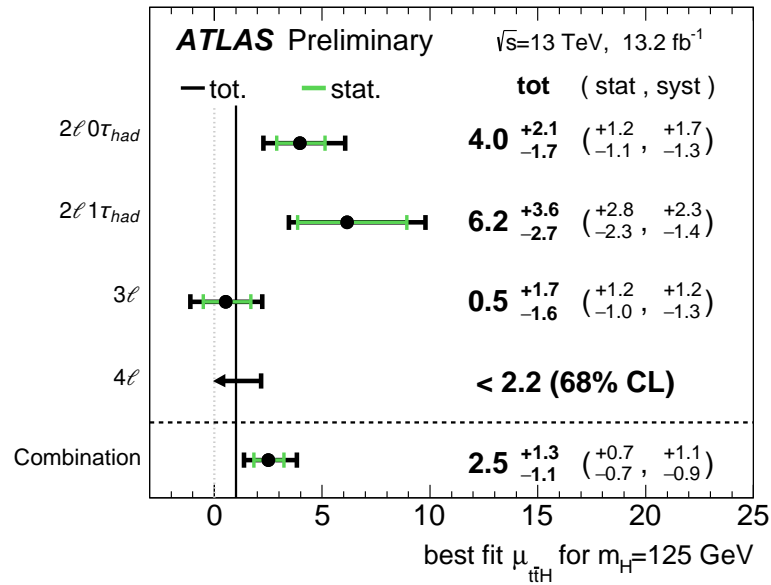


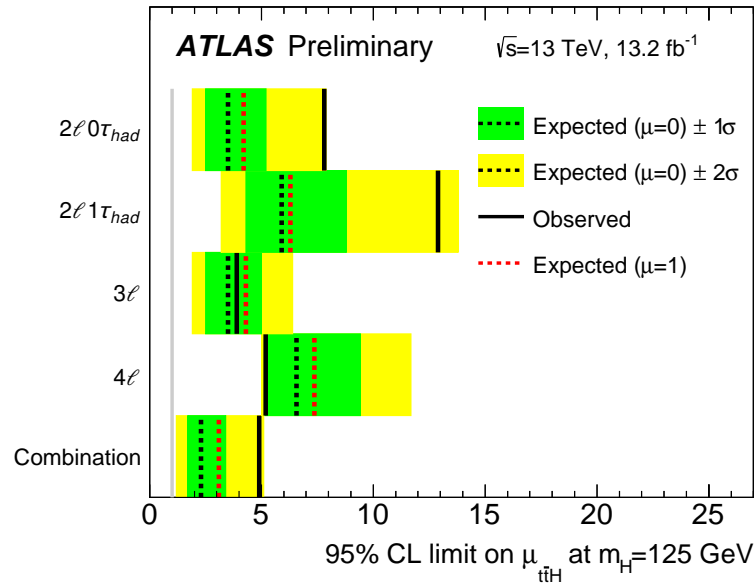
Figure 6.20: *Expected contribution to the background in each channel from various sources, using values of the background estimates before the fit. Charge mis-reconstruction backgrounds are indicated as “QMisReco.”*

by 4.9 ( $2.3^{+1.1}_{-0.6}$ ). The best-fit value of and 95% CL upper limit on  $\mu_{t\bar{t}H}$  for each individual channel and the combination of all channels are shown in Figures 6.21. For the  $4\ell$  channel the limits are obtained with the use of pseudo-Monte Carlo events, while for other channels and the combination the limits are obtained using asymptotic approximations for the  $CL_s$  value [148].

In the presence of the SM  $t\bar{t}H$  signal, the fit is expected to return  $\mu_{t\bar{t}H} = 1.0^{+0.7}_{-0.6}$  (stat)  $^{+0.9}_{-0.8}$  (syst). The  $p$ -value associated with the no- $t\bar{t}H$  hypothesis is 0.015 ( $2.2\sigma$ ), and the  $p$ -value associated with the SM expectation  $\mu_{t\bar{t}H} = 1$  is 0.09 ( $1.3\sigma$ ). The Nuisance Parameters of the fit correspond to the theoretical and experimental systematics described in Section 6.5.2, and uncertainties linked to the data-driven fake estimates methods detailed in Section 6.3. The impact of the most important systematic uncertainties on  $\mu_{t\bar{t}H}$ , evaluated after the global fit, is shown in Table 6.17. The ranking of the most important systematic uncertainty on the signal strength is presented in Figure 6.22. One can see that the systematics associated to the data-driven fake estimates is dominating in addition to JVT uncertainty. The changes in the central values and uncertainties of the nuisance parameters after the fit are small and not presented in this thesis.



(a)



(b)

Figure 6.21: a) Best fit values of the  $t\bar{t}H$  signal strength  $\mu_{t\bar{t}H}$  by final state category and combined. The SM prediction is  $\mu_{t\bar{t}H} = 1$ . For the  $4\ell$  category, as zero events are observed, a 68%  $CL_s$  upper limit is shown instead. b) Upper limits on the  $t\bar{t}H$  signal strength  $\mu_{t\bar{t}H}$  at 95% CL by final state category and combined. The SM prediction is  $\mu_{t\bar{t}H} = 1$ . The median upper limit that would be set in the presence of a SM  $t\bar{t}H$  signal ( $\mu = 1$ ) is also shown.



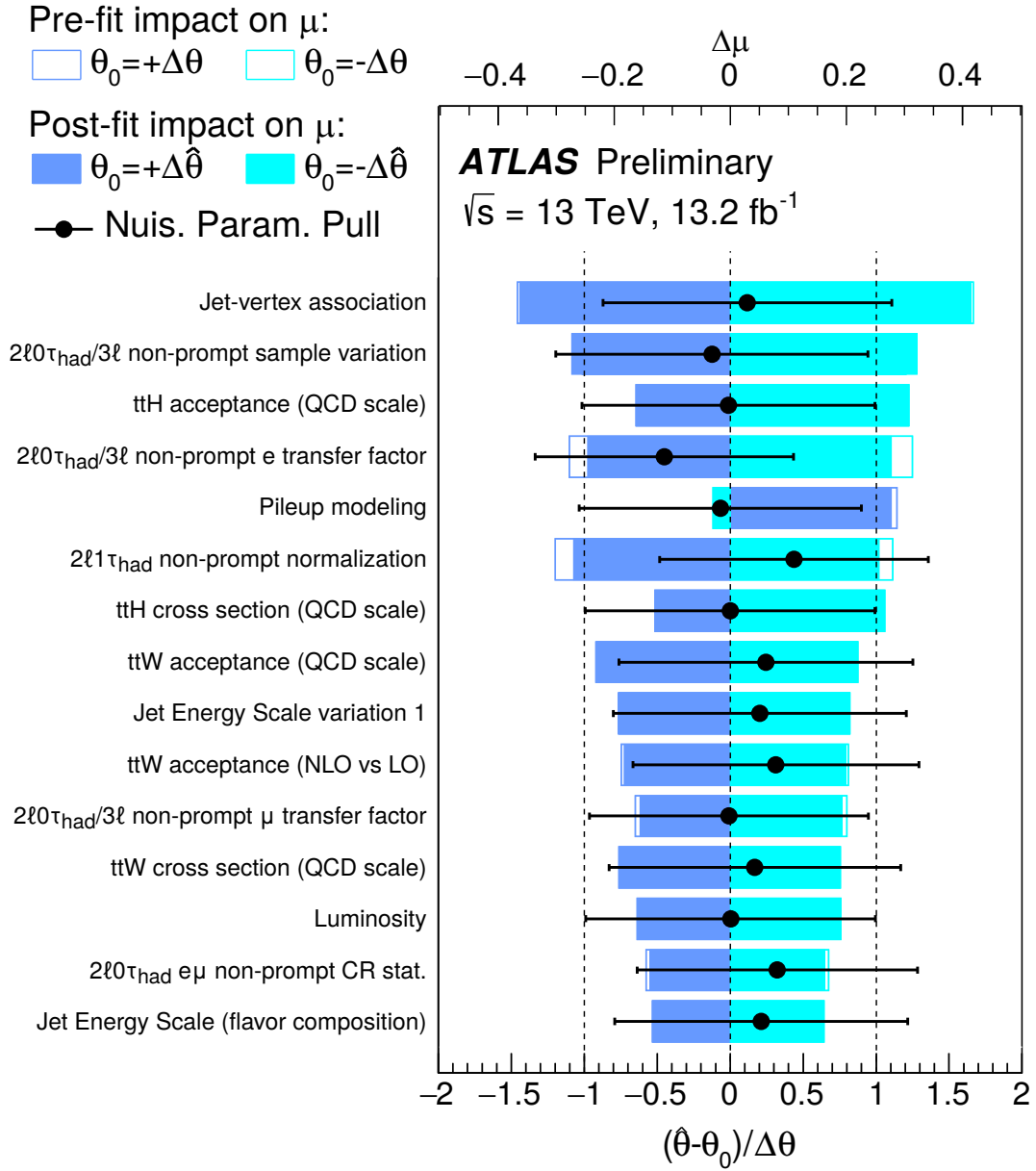


Figure 6.22: Effect of the fifteen most important systematic uncertainty nuisance parameters  $\theta$  on the signal strength  $\mu = \mu_{t\bar{t}H}$  and constraints on the nuisance parameters from the fit. The blue and cyan bars show the  $\pm 1\sigma$  impact of the nuisance parameter on the signal strength (shown on the top axis). The points and associated error bars show the best-fit values of the nuisance parameters and post-fit uncertainties on the nuisance parameters (shown on the bottom axis). The open bars show the effect of the systematic uncertainties on  $\mu_{t\bar{t}H}$  before the fit, and the solid bars show the effect after the fit. The nuisance parameters are initially normalised to  $0 \pm 1$ . The dotted vertical lines show  $\pm 1\sigma$  excursions of the nuisance parameters from their initial values.

Table 6.17: *Summary of the effects of the systematic uncertainties on  $\mu$ . Due to correlations between the different sources of uncertainties, the total systematic uncertainty can be different from the sum in quadrature of the individual sources. The impact of the systematic uncertainties is evaluated after the fit described in Section 6.5.*

Uncertainty Source	$\Delta\mu$	
Non-prompt leptons and charge mis-reconstruction	+0.56	-0.64
Jet-vertex association, pile-up modeling	+0.48	-0.36
$t\bar{t}W$ modeling	+0.29	-0.31
$t\bar{t}H$ modeling	+0.31	-0.15
Jet energy scale and resolution	+0.22	-0.18
$t\bar{t}Z$ modeling	+0.19	-0.19
Luminosity	+0.19	-0.15
Diboson modeling	+0.15	-0.14
Jet flavour tagging	+0.15	-0.12
Light lepton ( $e, \mu$ ) and $\tau_{\text{had}}$ ID, isolation, trigger	+0.12	-0.10
Other background modeling	+0.11	-0.11
Total systematic uncertainty	+1.1	-0.9

## 6.6.2 All $t\bar{t}H$ decays

A combination of the  $t\bar{t}H$  searches in the  $\gamma\gamma$ , multi-lepton, and  $b\bar{b}$  decay channels is performed using up to  $13.3 \text{ fb}^{-1}$  of  $pp$  collisions data at  $\sqrt{s}=13 \text{ TeV}$  [152]. 200 nuisance parameters are present in this combination, those from the same source in each of the independent analyses are generally treated as correlated. The largest systematic uncertainty contribution is related to the  $t\bar{t}+ \geq 1b$  modeling uncertainties affecting the  $H \rightarrow b\bar{b}$  analysis. The impact of the systematic uncertainties on  $\mu_{t\bar{t}H}$  after the fit to data is displayed in Table 6.18.

Uncertainty Source	$\Delta\mu$	
$t\bar{t}+ \geq 1b$ modeling	+0.34	-0.33
Jet flavour tagging	+0.19	-0.19
Background model statistics	+0.18	-0.18
$t\bar{t}+ \geq 1c$ modeling	+0.17	-0.17
Jet energy scale and resolution	+0.18	-0.18
$t\bar{t}H$ modeling	+0.20	-0.13
$t\bar{t}$ +light modeling	+0.14	-0.14
Other background modeling	+0.16	-0.15
Fake lepton uncertainties	+0.11	-0.12
Jet-vertex association, pile-up modeling	+0.09	-0.09
Luminosity	+0.09	-0.09
$t\bar{t}Z$ modeling	+0.08	-0.07
Light lepton ( $e, \mu$ ), photon, and $\tau$ ID, isolation, trigger	+0.04	-0.04
Total systematic uncertainty	+0.57	-0.54
$t\bar{t}+ \geq 1b$ normalisation	+0.24	-0.24
$t\bar{t}+ \geq 1c$ normalisation	+0.11	-0.11
Statistical uncertainty	+0.38	-0.38
Total uncertainty	+0.69	-0.66

Table 6.18: Summary of uncertainties on  $\mu_{t\bar{t}H}$ . Due to correlations between the different sources of uncertainties, the total systematic uncertainty can be different from the sum in quadrature of the individual sources. The normalisation factors for both  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ \geq 1c$  are included in the statistical component.

Figure 6.23 summarises the observed fitted signal strength  $\mu_{t\bar{t}H}$  and the corresponding upper limit, at 95% CL, of the individual channels, and their combination.

The combined  $t\bar{t}H$  signal strength ( $\sigma/\sigma_{\text{SM}}$ ) is found to be  $1.8_{-0.7}^{+0.7}$  which corresponds to an observed significance of  $2.8 \sigma$ , where  $1.8 \sigma$  would be expected in the presence of Standard Model  $t\bar{t}H$ . The sensitivity of this combination exceeds the Run-1  $t\bar{t}H$  expected significance of  $1.5 \sigma$ . All three analyses are within  $1.5 \sigma$  of the central value. Therefore, no tension is found between the  $t\bar{t}H$  analyses. The combined observed value of 95% CL upper limit on  $\mu_{t\bar{t}H}$  is estimated to 3.0 which is similar to Run1 observed value. Moreover, the couplings of the Higgs boson to fermions and bosons ( $\kappa_F$  and  $\kappa_V$ ) for the  $t\bar{t}H$  production mode are found to be compatible with the Standard Model expectation.

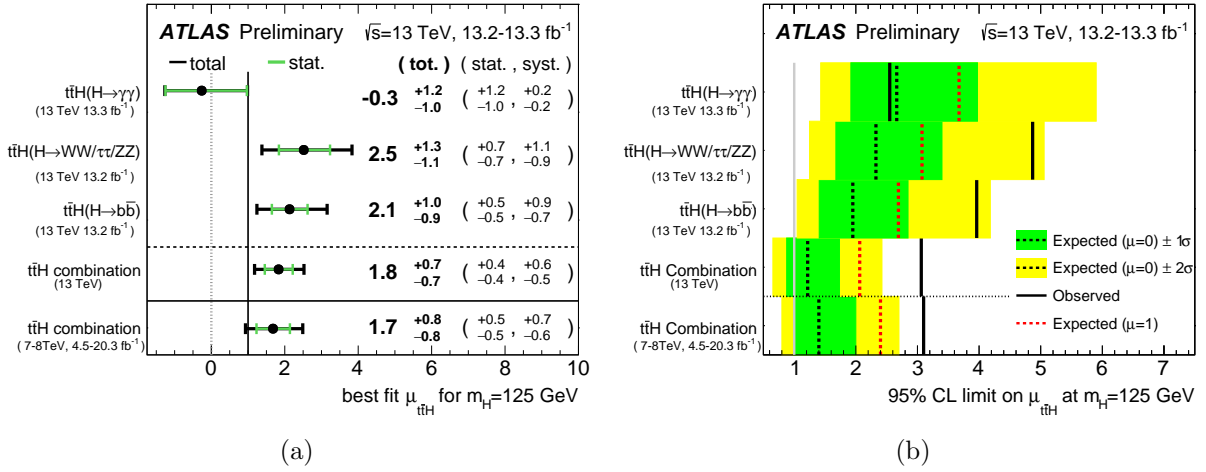


Figure 6.23: a) Summary of the observed  $\mu_{t\bar{t}H}$  signal strength measurements from the individual analyses and for their combination, assuming  $m_H = 125$  GeV. The total (tot.), statistical (stat.), and systematic (syst.) uncertainties on  $\mu_{t\bar{t}H}$  are shown. The SM  $\mu_{t\bar{t}H} = 1$  (0) expectation is shown as the black (grey) vertical line. The observed  $\mu_{t\bar{t}H}$  signal strength measurement obtained from the Run-1 combination is also shown for comparison (bottom). b) The upper limits on the  $t\bar{t}H$  signal strength for the individual analyses as well as their combination at 95% CL. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis (black dashed lines) and under the signal-plus-background hypothesis assuming the SM prediction for the  $t\bar{t}H$  process (red dashed lines). The surrounding shaded bands correspond to the  $\pm 1\sigma$  and  $\pm 2\sigma$  ranges around the expected limits under the background-only hypothesis. The vertical grey line at  $\mu_{t\bar{t}H} = 1$  represents the point below which the SM  $t\bar{t}H$  production would be excluded. The observed and expected limits obtained from the Run-1 combination are also shown for comparison (bottom).

## 6.7 Conclusions and prospects

The study of the associated reducible background to suppress non prompt leptons (fake leptons) from the  $t\bar{t}H$  signal using the matrix method (MM) is discussed. It represents the most dominant background in  $2\ell ss$  and one of the most important parameters driving systematic uncertainties. The estimation of fake leptons is performed using different data-driven methods since simulation is not reliable enough to estimate fakes. It is found that fakes from data are 1.5-3.6 times the fakes from simulation and represent 32-48% of the total background. In the first public results, fake factor method is used as the main approach to estimate fakes and the MM (the topic of this thesis) as a cross check and validation tool. For the next iteration of the analysis with full 2015/2016 data set, new technics are planned to be used in order to reduce fakes and charge flip contribution in the  $t\bar{t}H$  signal region, out of scope of this thesis, such as replacing *Cut&Count* by multivariate analysis techniques for lepton isolation and event selection. On top of that, I managed with other groups in ATLAS to introduce the MM as the baseline method to estimate fake leptons for the next iteration of this analysis.

In the matrix method, further improvements are suggested. An improvement of the closure test is still possible by improving the fake rate measurements using different methods at high  $p_T$  than the T&P method. Also, a multi-dimension parametrisation will be useful to take into account all the efficiency dependency on kinematics and hence assign more accurate values for the MM inputs. Finally, better  $p_T$  binning could also have an impact on the closure test. All those tests are planned to be performed for the next iteration of this analysis.

So far, the  $t\bar{t}H$  production process has not been seen yet at  $\sqrt{s} = 13$  TeV with  $13.2 \text{ fb}^{-1}$  in both ATLAS and CMS. On top of that, no significant deviation from the SM (no new physics) is seen for this process.

# Conclusion

Since resuming operation for Run 2, the LHC has been producing roughly 20,000 Higgs bosons per day in its 13 TeV proton–proton collisions. At the end of 2015, the data collected by the ATLAS and CMS collaborations were already enough to re-observe the Higgs boson at the new collision energy. The ATLAS collaboration has released a new preliminary measurement of the Higgs boson production in association with a pair of top quark using 2015 and 2016 LHC data. An improved precision in the measurement of the Higgs boson production has been made possible by both the increased collision energy of 13 TeV and improved collision rate.

The first part of the thesis belongs to the ATLAS detector performance tests at the beginning of Run2. It represents a precise measurement of electron reconstruction efficiency in ATLAS with  $Z \rightarrow ee$  data sample using  $3.2 \text{ fb}^{-1}$  of data recorded in 2015. The high amount and difficult background, coming from all EM clusters, is the most challengeable piece in those measurements. To get reliable results, the efficiencies are estimated in  $(20 \times 10)$   $(\eta \times E_T)$  bins reflecting the dependency on the electron kinematics and the detector design. Overall, the reconstruction efficiencies vary from 95% to 99% with sub-percent typical measured errors for  $E_T > 25 \text{ GeV}$  and between 0.7-5 % for lower  $E_T$ . The reconstruction efficiency is comparable to the efficiency observed in data taken in 2012 at 8 TeV, except for the calorimeter transition region ( $1.37 < |\eta| < 1.52$ ) where the efficiency degrades by 2%, caused by inefficiencies in the matching of the track to the calorimeter section, recovered during 2016 data operation. Data-MC scale factors are computed including the systematic uncertainties and used in the 2015 ATLAS physics analyses involving electrons. They are found to be very close to the unity. The good agreement between the reconstruction results obtained in data and in MC shows the high ability of the ATLAS detector to reconstruct electrons from one hand and the good understanding of its performance on the other hand.

The second part of the thesis is dedicated to a search for the Higgs boson production in association with a top quark pair ( $t\bar{t}H$ ). The signature with two same-charge light leptons (electron or muon) without a hadronically decaying tau lepton final state ( $2\ell ss$ ) is examined using the first 10% of the total expected Run2 dataset. An improved method (matrix method) has been developed to estimate the fake (non-prompt) leptons background for three categories depending on the lepton flavour final state: two same sign electrons  $e^\pm e^\pm$ , two same sign muons  $\mu^\pm \mu^\pm$  and two same sign opposite flavour leptons  $e^\pm \mu^\pm$ . Real lepton efficiencies ( $\epsilon_r$ ) and fake lepton rates ( $\epsilon_f$ ) are calculated in an orthogonal region to the signal region. In order to get more reliable predictions, the efficiencies are factorised in bins of the transverse momentum  $p_T$ . In the range of interest ( $p_T > 25 \text{ GeV}$ ), electron (muon) real efficiencies are in the range of 70-95 % (90-100 %). Meanwhile, the fake rate values are around 30 % for both electrons and muons. After validating the efficiency measurements and the method, by performing a closure test using  $t\bar{t}$  simulation, fake lepton backgrounds are estimated in the signal region for the three categories. Those estimation are found to be 1.5 to 3.6 times higher than in

simulation and represent between 32 and 48% of the total background. It represents the most dominant background in  $2\ell ss$  and one of the most important parameters driving systematic uncertainties.

After estimating the different sources of background, the best-fit value of the ratio of observed and Standard Model cross sections of the  $t\bar{t}H$   $2\ell ss$  production process ( $\mu_{t\bar{t}H}$ ) is:  $4.0 \pm_{-1.1}^{+1.2}$  (stat)  $_{-1.3}^{+1.7}$  (syst). The contribution of statistical and systematic uncertainties are equivalent. The slight excess with respect to the SM is mainly from  $e^\pm\mu^\pm$  channel. To increase the sensitivity to the  $t\bar{t}H$  signal, a combination of all multileptonic  $t\bar{t}H$  channels is performed and gives  $2.5 \pm 0.7$  (stat)  $_{-0.9}^{+1.1}$  (syst). The  $t\bar{t}H$  process is excluded at 4.9 (2.3 expected) times the SM prediction at 95% confidence level. The  $p$ -value associated with the SM expectation  $\mu_{t\bar{t}H} = 1$  is 0.09 ( $1.3\sigma$ ). The value of  $t\bar{t}H$  signal strength considering all final states, with  $b\bar{b}$  and diphoton final states channels, is found to be  $1.8_{-0.7}^{+0.7}$  which corresponds to an observed significance of  $2.8\sigma$  ( $1.8\sigma$  would be expected in the presence of Standard Model  $t\bar{t}H$ ). The sensitivity of this combination exceeds the Run1  $t\bar{t}H$  expected significance of  $1.5\sigma$ .

Those first results allow to introduce the matrix method as a validation tool and an alternative method to estimate the fake lepton background for the preliminary analysis with  $13.2\text{ fb}^{-1}$  data at 13 TeV. Thanks to its flexibility and accuracy, the matrix method is now considered as the baseline method for the current analysis, combining the 2015 with 2016 data ( $36.1\text{ fb}^{-1}$ ), with the following improvements: The measurements of fake lepton background at high  $p_T$  and a multi-dimensional parametrisation (in terms of the number of b-jet, the distance between the lepton and the closest jet,...).

Using 10% of the total expected Run2 dataset, the total  $t\bar{t}H$  production process has not been seen yet in ATLAS. No significant deviation from the prediction is seen for this process. ATLAS will continue to study the Higgs boson properties for the rest of Run2, isolating its rare production modes and measuring its more elusive properties. Uncovering these secrets will either further cement the Standard Model, or give us insight of what lies beyond.



# Appendix A

## Background estimation in electron reconstruction efficiency

### A.1 Description of background estimation for clusters with no associated track

This section includes the mathematical description of the sideband fit method inherited from Ref. [101]. The background associated to the *No-Track* EM clusters candidates is deduced by fitting a third order polynomial function in the sideband of the electron-*No-Track* EM clusters invariant mass distribution. The background estimation is given by taking the integral of a binned  $\chi^2$  fit function in the signal region. The polynomial function can be written as following:

$$f(\mu) = \sum_{i=0}^3 k_i \mu^i$$

And the corresponding  $\chi^2$  reads:

$$\chi^2 = \sum_{i=1}^N (n_i - f(\mu_i))^2$$

Where  $n_i$  is the number of events and  $\mu_i$  the invariant mass in the center of the bin  $i$ . Bins have  $\Delta\mu = 5$  GeV width. Signal contamination is subtracted for each bin like following:

$$n_i = n_i^{\text{data}} - n_i^{\text{MC}}$$

The MC estimation is computed the same way as for data and normalised to the number of *tight-tight* electron-positron pairs. Minimising  $\chi^2$  with respect to its four parameters leads to the following system of linear equations:

$$\sum_{i=1}^N \left( n_i - \sum_{j=0}^3 k_j \mu_i^j \right) \mu^q = 0$$

Where  $N$  is the number of bins. This can be reorganised as:

$$\forall q \in \{0, 1, 2, 3\}, \quad \sum_{i=1}^N \sum_{j=0}^3 k_j \mu_i^{j+q} = \sum_{i=1}^N \mu_i^q n_i \iff MA = B \quad \text{with} \quad \begin{cases} M_{qj} A_j = B_q \\ M_{jq} = M_{qj} = \sum_{i=1}^N \mu_i^{q+j} \\ A_j = k_j \\ B_q = \sum_{i=1}^N \mu_i^q n_i \end{cases}$$

And finally read:

$$\begin{pmatrix} S_{00} & S_{10} & S_{20} & S_{30} \\ S_{10} & S_{20} & S_{30} & S_{40} \\ S_{20} & S_{30} & S_{40} & S_{50} \\ S_{30} & S_{40} & S_{50} & S_{60} \end{pmatrix} \begin{pmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} S_{01} \\ S_{11} \\ S_{21} \\ S_{31} \end{pmatrix} \quad \text{with} \quad S_{\alpha\beta} = \sum_{i=1}^N \mu_i^\alpha n_i^\beta$$

Which allows to write:

$$\begin{pmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{pmatrix} = M^{-1} \begin{pmatrix} S_{01} \\ S_{11} \\ S_{21} \\ S_{31} \end{pmatrix}$$

Then the number of background events is obtained by integrating the fitted function in the peak region  $[\mu_{\text{peak low}}, \mu_{\text{peak up}}]$ :

$$B^{\text{No-Track}} = k_0 \delta^{(1)} + \frac{k_1}{2} \delta^{(2)} + \frac{k_2}{3} \delta^{(3)} + \frac{k_3}{4} \delta^{(4)} \quad \text{with} \quad \delta^{(i)} = \mu_{\text{peak up}}^i - \mu_{\text{peak low}}^i$$

## A.2 Invariant mass distribution of the background

Figures A.1 - A.6 show the invariant mass distribution in two bins at high and low  $E_T$ , together with the background estimation performed with the methods described above. A good background estimation is observed in the whole  $\eta$  range for the numerator (Figure A.2 and A.5) at low and high  $E_T$  respectively. In the denominator, where two other background terms are additionally estimated (fail TQ and *No-Track* EM clusters background), the electron background template normalized in the tail, does not perfectly match the baseline at low mass range due to the high photon contribution at low  $p_T$  as shown in Figure A.1 and A.4. Figure A.3 and A.6 show that the number of the probes failing the track quality requirements is mainly dominated by background where the signal contribution is negligible.

## A.3 Tables of electron reconstruction efficiencies and SF

This section contains a table of the electron reconstruction efficiencies and scale factors with their uncertainties in 200 bins in  $(E_T, \eta)$ , determined for electrons from  $Z$  that are separated from jets fulfilling  $\Delta R(\text{electron}, \text{jet}) > 0.4$ . The efficiencies are calculated for both 2015 data and MC at  $\sqrt{s} = 13$  TeV presented in Table A.1 and A.2 respectively. The SF are in general very close to one (Table A.3). The uncertainties, shared between statistics and systematics, are below 0.8% for electrons probes with  $E_T$  inside the [25,150] GeV interval. For the lowest  $E_T$  probe ( $< 25$  GeV), the SF uncertainty is higher, 0.2-4.5%. The statistical uncertainty dominates in most of  $\eta$  bins.

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
[-2.47,-2.37]	97.3 ± 2.2 ± 1.4	97.4 ± 1.0 ± 1.4	97.9 ± 0.5 ± 0.7	98.3 ± 0.3 ± 0.1	98.3 ± 0.2 ± 0.1	98.3 ± 0.2 ± 0.1	98.3 ± 0.3 ± 0.3	98.6 ± 0.4 ± 0.1	98.3 ± 0.6 ± 0.6	99.2 ± 0.8 ± 0.4
[-2.37,-2.01]	96.6 ± 1.3 ± 1.0	97.9 ± 0.6 ± 0.7	97.8 ± 0.3 ± 0.2	98.2 ± 0.2 ± 0.3	98.1 ± 0.1 ± 0.1	98.4 ± 0.1 ± 0.0	98.7 ± 0.1 ± 0.1	98.6 ± 0.3 ± 0.1	99.2 ± 0.3 ± 0.1	99.2 ± 0.8 ± 0.3
[-2.01,-1.81]	97.1 ± 2.0 ± 1.2	96.6 ± 0.8 ± 1.1	97.2 ± 0.4 ± 0.3	97.1 ± 0.2 ± 0.2	97.4 ± 0.2 ± 0.1	97.6 ± 0.1 ± 0.1	98.0 ± 0.2 ± 0.1	98.3 ± 0.2 ± 0.1	98.4 ± 0.4 ± 0.3	99.1 ± 0.6 ± 0.3
[-1.81,-1.52]	95.4 ± 2.1 ± 0.6	97.2 ± 0.8 ± 0.4	96.9 ± 0.4 ± 0.2	97.0 ± 0.2 ± 0.2	97.4 ± 0.1 ± 0.1	97.6 ± 0.1 ± 0.0	97.8 ± 0.1 ± 0.0	97.7 ± 0.2 ± 0.2	98.4 ± 0.3 ± 0.1	98.4 ± 0.5 ± 0.3
[-1.52,-1.37]	94.7 ± 3.9 ± 3.7	92.2 ± 1.5 ± 0.7	93.7 ± 0.7 ± 0.6	94.5 ± 0.4 ± 0.5	95.3 ± 0.3 ± 0.1	95.8 ± 0.2 ± 0.1	96.0 ± 0.3 ± 0.3	95.7 ± 0.3 ± 0.3	96.4 ± 0.5 ± 0.3	95.5 ± 1.1 ± 0.7
[-1.37,-1.15]	95.6 ± 2.7 ± 1.1	96.8 ± 1.0 ± 1.1	98.2 ± 0.4 ± 0.3	97.8 ± 0.2 ± 0.2	98.2 ± 0.1 ± 0.1	98.3 ± 0.1 ± 0.0	98.6 ± 0.1 ± 0.1	98.9 ± 0.2 ± 0.1	98.7 ± 0.3 ± 0.1	99.3 ± 0.4 ± 0.3
[-1.15,-0.80]	97.0 ± 2.2 ± 2.3	99.8 ± 0.6 ± 0.3	98.5 ± 0.4 ± 0.2	98.6 ± 0.2 ± 0.1	98.6 ± 0.1 ± 0.0	98.8 ± 0.1 ± 0.0	99.0 ± 0.1 ± 0.1	98.9 ± 0.1 ± 0.1	99.0 ± 0.2 ± 0.1	99.5 ± 0.3 ± 0.3
[-0.80,-0.60]	98.1 ± 2.6 ± 2.1	99.9 ± 0.8 ± 0.2	99.1 ± 0.5 ± 0.5	99.0 ± 0.2 ± 0.2	98.9 ± 0.1 ± 0.1	98.9 ± 0.1 ± 0.1	99.0 ± 0.1 ± 0.1	99.0 ± 0.2 ± 0.1	99.0 ± 0.2 ± 0.3	99.1 ± 0.5 ± 0.3
[-0.60,-0.10]	98.3 ± 1.3 ± 1.9	98.9 ± 0.6 ± 0.5	99.1 ± 0.3 ± 0.2	99.1 ± 0.1 ± 0.1	99.0 ± 0.1 ± 0.1	99.1 ± 0.1 ± 0.0	99.1 ± 0.1 ± 0.1	99.2 ± 0.1 ± 0.1	99.0 ± 0.2 ± 0.1	99.4 ± 0.3 ± 0.1
[-0.10,0.00]	99.6 ± 2.2 ± 0.4	99.3 ± 1.1 ± 0.7	98.9 ± 0.5 ± 0.7	99.4 ± 0.3 ± 0.3	98.9 ± 0.2 ± 0.1	99.1 ± 0.1 ± 0.0	98.8 ± 0.2 ± 0.1	99.1 ± 0.2 ± 0.1	98.7 ± 0.4 ± 0.2	99.2 ± 0.6 ± 0.4
[0.00,0.10]	100.0 ± 2.0 ± 0.1	98.4 ± 1.1 ± 1.0	98.3 ± 0.6 ± 0.4	98.8 ± 0.3 ± 0.2	99.0 ± 0.2 ± 0.1	98.8 ± 0.1 ± 0.1	99.0 ± 0.2 ± 0.1	99.1 ± 0.2 ± 0.1	99.1 ± 0.4 ± 0.2	98.1 ± 0.7 ± 0.5
[0.10,0.60]	99.4 ± 1.3 ± 0.7	98.7 ± 0.6 ± 0.8	98.7 ± 0.3 ± 0.1	99.0 ± 0.1 ± 0.1	98.9 ± 0.1 ± 0.1	99.0 ± 0.1 ± 0.0	99.3 ± 0.1 ± 0.1	99.1 ± 0.1 ± 0.0	99.4 ± 0.2 ± 0.1	99.6 ± 0.3 ± 0.2
[0.60,0.80]	94.7 ± 2.9 ± 2.0	99.8 ± 0.8 ± 0.3	99.2 ± 0.5 ± 0.4	98.5 ± 0.2 ± 0.3	98.8 ± 0.1 ± 0.1	98.8 ± 0.1 ± 0.1	99.0 ± 0.1 ± 0.1	99.1 ± 0.2 ± 0.1	99.2 ± 0.3 ± 0.2	99.3 ± 0.5 ± 0.3
[0.80,1.15]	98.8 ± 2.1 ± 1.5	99.2 ± 0.9 ± 0.5	99.0 ± 0.4 ± 0.4	98.3 ± 0.2 ± 0.2	98.5 ± 0.1 ± 0.0	98.8 ± 0.1 ± 0.0	98.9 ± 0.1 ± 0.1	99.0 ± 0.1 ± 0.2	99.0 ± 0.2 ± 0.2	99.2 ± 0.4 ± 0.2
[1.15,1.37]	98.7 ± 2.7 ± 1.9	96.5 ± 1.0 ± 1.0	98.5 ± 0.4 ± 0.3	98.0 ± 0.2 ± 0.3	98.3 ± 0.1 ± 0.1	98.4 ± 0.1 ± 0.1	98.6 ± 0.1 ± 0.1	98.6 ± 0.2 ± 0.1	99.1 ± 0.3 ± 0.2	99.1 ± 0.5 ± 0.8
[1.37,1.52]	94.1 ± 3.7 ± 4.3	93.1 ± 1.5 ± 1.0	94.4 ± 0.7 ± 0.8	94.3 ± 0.4 ± 0.5	95.0 ± 0.3 ± 0.2	95.1 ± 0.2 ± 0.2	95.5 ± 0.3 ± 0.2	96.0 ± 0.4 ± 0.3	95.5 ± 1.1 ± 0.4	96.1 ± 0.4 ± 0.2
[1.52,1.81]	94.0 ± 2.0 ± 0.7	96.9 ± 0.8 ± 0.9	96.6 ± 0.4 ± 0.4	97.1 ± 0.2 ± 0.1	97.4 ± 0.1 ± 0.1	97.6 ± 0.1 ± 0.0	97.8 ± 0.1 ± 0.1	98.3 ± 0.2 ± 0.1	98.3 ± 0.3 ± 0.1	98.5 ± 0.6 ± 0.3
[1.81,2.01]	95.4 ± 2.0 ± 0.4	96.3 ± 0.8 ± 0.9	96.9 ± 0.4 ± 0.1	97.1 ± 0.2 ± 0.1	97.2 ± 0.2 ± 0.1	97.4 ± 0.1 ± 0.1	98.0 ± 0.2 ± 0.1	97.9 ± 0.2 ± 0.1	98.4 ± 0.4 ± 0.2	99.5 ± 0.5 ± 0.3
[2.01,2.37]	96.9 ± 1.3 ± 0.7	97.7 ± 0.6 ± 0.4	98.0 ± 0.3 ± 0.2	98.4 ± 0.2 ± 0.1	98.3 ± 0.1 ± 0.1	98.5 ± 0.1 ± 0.1	98.6 ± 0.1 ± 0.1	98.3 ± 0.2 ± 0.1	98.9 ± 0.3 ± 0.1	99.1 ± 0.5 ± 0.3
[2.37,2.47]	95.9 ± 2.3 ± 1.4	96.9 ± 1.0 ± 0.8	98.0 ± 0.5 ± 0.3	98.3 ± 0.3 ± 0.1	98.2 ± 0.3 ± 0.1	98.4 ± 0.2 ± 0.1	98.5 ± 0.2 ± 0.1	98.5 ± 0.4 ± 0.4	98.8 ± 0.6 ± 0.3	98.3 ± 1.0 ± 0.5

Table A.1: Reconstruction data efficiencies (in %), stat. and syst. errors.

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
[-2.47,-2.37]	97.1 ± 0.2 ± 0.0	97.4 ± 0.2 ± 0.0	97.8 ± 0.1 ± 0.1	98.0 ± 0.1 ± 0.0	98.3 ± 0.1 ± 0.0	98.4 ± 0.1 ± 0.0	98.7 ± 0.1 ± 0.0	98.9 ± 0.1 ± 0.0	99.0 ± 0.2 ± 0.0	98.9 ± 0.4 ± 0.0
[-2.37,-2.01]	96.8 ± 0.1 ± 0.1	97.5 ± 0.1 ± 0.0	97.7 ± 0.1 ± 0.0	98.0 ± 0.1 ± 0.0	98.3 ± 0.0 ± 0.0	98.4 ± 0.0 ± 0.0	98.6 ± 0.0 ± 0.0	98.8 ± 0.1 ± 0.0	99.0 ± 0.1 ± 0.0	98.9 ± 0.2 ± 0.1
[-2.01,-1.81]	95.7 ± 0.2 ± 0.0	96.1 ± 0.1 ± 0.1	96.7 ± 0.1 ± 0.0	96.9 ± 0.1 ± 0.0	97.4 ± 0.1 ± 0.0	97.7 ± 0.1 ± 0.0	98.0 ± 0.1 ± 0.0	98.2 ± 0.1 ± 0.0	98.4 ± 0.1 ± 0.0	98.9 ± 0.2 ± 0.0
[-1.81,-1.52]	94.7 ± 0.2 ± 0.1	95.9 ± 0.1 ± 0.1	96.2 ± 0.1 ± 0.1	96.7 ± 0.1 ± 0.1	97.0 ± 0.1 ± 0.0	97.2 ± 0.1 ± 0.0	97.8 ± 0.1 ± 0.0	98.0 ± 0.1 ± 0.0	98.3 ± 0.1 ± 0.0	98.7 ± 0.2 ± 0.0
[-1.52,-1.37]	92.8 ± 0.3 ± 0.5	93.2 ± 0.2 ± 0.5	93.4 ± 0.2 ± 0.3	93.8 ± 0.1 ± 0.3	94.3 ± 0.1 ± 0.2	95.3 ± 0.1 ± 0.1	95.8 ± 0.1 ± 0.1	95.7 ± 0.1 ± 0.3	96.2 ± 0.2 ± 0.2	96.1 ± 0.4 ± 0.1
[-1.37,-1.15]	96.9 ± 0.2 ± 0.1	97.2 ± 0.1 ± 0.1	97.7 ± 0.1 ± 0.1	97.8 ± 0.1 ± 0.1	98.0 ± 0.0 ± 0.0	98.3 ± 0.0 ± 0.0	98.6 ± 0.0 ± 0.0	98.7 ± 0.1 ± 0.0	98.8 ± 0.1 ± 0.0	99.2 ± 0.1 ± 0.0
[-1.15,-0.80]	97.4 ± 0.1 ± 0.1	97.9 ± 0.1 ± 0.2	98.2 ± 0.1 ± 0.1	98.4 ± 0.0 ± 0.1	98.6 ± 0.0 ± 0.0	98.8 ± 0.0 ± 0.0	99.0 ± 0.0 ± 0.0	98.9 ± 0.0 ± 0.0	99.2 ± 0.1 ± 0.0	99.3 ± 0.1 ± 0.0
[-0.80,-0.60]	98.0 ± 0.1 ± 0.1	98.3 ± 0.1 ± 0.1	98.3 ± 0.1 ± 0.1	98.5 ± 0.0 ± 0.1	98.8 ± 0.0 ± 0.0	98.9 ± 0.0 ± 0.0	99.1 ± 0.0 ± 0.0	99.1 ± 0.0 ± 0.0	99.1 ± 0.1 ± 0.0	99.4 ± 0.1 ± 0.0
[-0.60,-0.10]	98.4 ± 0.1 ± 0.1	98.6 ± 0.0 ± 0.1	98.7 ± 0.0 ± 0.1	98.8 ± 0.0 ± 0.0	98.9 ± 0.0 ± 0.0	99.1 ± 0.0 ± 0.0	99.3 ± 0.0 ± 0.0	99.3 ± 0.0 ± 0.0	99.3 ± 0.0 ± 0.0	99.4 ± 0.1 ± 0.0
[-0.10,0.00]	98.1 ± 0.2 ± 0.1	98.4 ± 0.1 ± 0.1	98.7 ± 0.1 ± 0.1	98.6 ± 0.1 ± 0.0	98.8 ± 0.1 ± 0.0	98.9 ± 0.0 ± 0.0	99.1 ± 0.1 ± 0.0	99.1 ± 0.1 ± 0.0	99.1 ± 0.1 ± 0.0	99.4 ± 0.1 ± 0.0
[0.00,0.10]	98.2 ± 0.2 ± 0.2	98.6 ± 0.1 ± 0.1	98.5 ± 0.1 ± 0.1	98.7 ± 0.1 ± 0.0	98.8 ± 0.0 ± 0.0	98.9 ± 0.0 ± 0.0	99.0 ± 0.1 ± 0.0	99.2 ± 0.1 ± 0.0	99.1 ± 0.1 ± 0.0	99.3 ± 0.2 ± 0.0
[0.10,0.60]	98.5 ± 0.1 ± 0.1	98.5 ± 0.1 ± 0.1	98.7 ± 0.0 ± 0.1	98.9 ± 0.0 ± 0.1	98.9 ± 0.0 ± 0.0	99.1 ± 0.0 ± 0.0	99.2 ± 0.0 ± 0.0	99.2 ± 0.0 ± 0.0	99.3 ± 0.0 ± 0.0	99.4 ± 0.1 ± 0.0
[0.60,0.80]	97.8 ± 0.1 ± 0.2	98.2 ± 0.1 ± 0.1	98.4 ± 0.1 ± 0.1	98.6 ± 0.0 ± 0.1	98.9 ± 0.0 ± 0.0	99.0 ± 0.0 ± 0.0	99.1 ± 0.0 ± 0.0	99.2 ± 0.0 ± 0.0	99.1 ± 0.1 ± 0.0	99.5 ± 0.1 ± 0.0
[0.80,1.15]	97.5 ± 0.1 ± 0.2	97.8 ± 0.1 ± 0.1	98.2 ± 0.1 ± 0.1	98.3 ± 0.0 ± 0.1	98.6 ± 0.0 ± 0.0	98.8 ± 0.0 ± 0.0	99.0 ± 0.0 ± 0.0	98.9 ± 0.0 ± 0.0	99.0 ± 0.1 ± 0.0	99.2 ± 0.1 ± 0.0
[1.15,1.37]	97.0 ± 0.2 ± 0.2	97.3 ± 0.1 ± 0.1	97.7 ± 0.1 ± 0.1	97.9 ± 0.1 ± 0.1	98.1 ± 0.0 ± 0.0	98.4 ± 0.0 ± 0.0	98.6 ± 0.0 ± 0.0	98.7 ± 0.1 ± 0.0	98.9 ± 0.1 ± 0.0	99.4 ± 0.1 ± 0.0
[1.37,1.52]	91.8 ± 0.3 ± 0.7	92.3 ± 0.2 ± 0.3	92.9 ± 0.2 ± 0.3	93.5 ± 0.1 ± 0.3	94.2 ± 0.1 ± 0.2	94.9 ± 0.1 ± 0.1	95.4 ± 0.1 ± 0.2	95.5 ± 0.1 ± 0.3	96.0 ± 0.2 ± 0.1	96.9 ± 0.3 ± 0.2
[1.52,1.81]	94.9 ± 0.2 ± 0.1	96.0 ± 0.1 ± 0.1	96.3 ± 0.1 ± 0.1	96.7 ± 0.1 ± 0.0	97.0 ± 0.1 ± 0.0	97.4 ± 0.0 ± 0.0	97.8 ± 0.1 ± 0.0	98.0 ± 0.1 ± 0.0	98.4 ± 0.1 ± 0.0	98.8 ± 0.2 ± 0.0
[1.81,2.01]	95.4 ± 0.2 ± 0.0	96.3 ± 0.1 ± 0.0	96.8 ± 0.1 ± 0.0	97.0 ± 0.1 ± 0.0	97.3 ± 0.1 ± 0.0	97.6 ± 0.1 ± 0.0	97.8 ± 0.1 ± 0.0	98.1 ± 0.1 ± 0.0	98.3 ± 0.1 ± 0.0	98.8 ± 0.2 ± 0.0
[2.01,2.37]	96.9 ± 0.1 ± 0.0	97.5 ± 0.1 ± 0.0	97.9 ± 0.1 ± 0.0	98.0 ± 0.1 ± 0.0	98.3 ± 0.0 ± 0.0	98.4 ± 0.0 ± 0.0	98.7 ± 0.0 ± 0.0	98.8 ± 0.1 ± 0.0	99.0 ± 0.1 ± 0.0	99.1 ± 0.1 ± 0.1
[2.37,2.47]	97.0 ± 0.3 ± 0.1	97.3 ± 0.2 ± 0.1	98.0 ± 0.1 ± 0.0	98.3 ± 0.1 ± 0.1	98.4 ± 0.1 ± 0.0	98.4 ± 0.1 ± 0.0	98.6 ± 0.1 ± 0.0	98.9 ± 0.1 ± 0.0	99.0 ± 0.2 ± 0.0	98.9 ± 0.4 ± 0.1

Table A.2: Reconstruction MC efficiencies (in %), stat. and syst. errors.

	15-20 GeV	20-25 GeV	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	50-60 GeV	60-80 GeV	80-150 GeV
[-2.47,-2.37]	100.1 ± 2.3 ± 1.5	99.9 ± 1.0 ± 1.4	100.1 ± 0.5 ± 0.7	100.3 ± 0.3 ± 0.1	100.0 ± 0.2 ± 0.1	99.9 ± 0.2 ± 0.1	99.6 ± 0.3 ± 0.3	99.7 ± 0.5 ± 0.1	99.3 ± 0.7 ± 0.6	100.3 ± 0.9 ± 0.4
[-2.37,-2.01]	99.8 ± 1.4 ± 1.1	100.4 ± 0.6 ± 0.7	100.1 ± 0.3 ± 0.3	100.1 ± 0.2 ± 0.3	99.8 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.0	100.1 ± 0.1 ± 0.1	99.9 ± 0.3 ± 0.1	100.2 ± 0.3 ± 0.1	100.3 ± 0.8 ± 0.3
[-2.01,-1.81]	101.5 ± 2.1 ± 1.3	100.5 ± 0.8 ± 1.1	100.5 ± 0.4 ± 0.3	100.2 ± 0.2 ± 0.2	100.0 ± 0.2 ± 0.1	100.0 ± 0.1 ± 0.1	100.0 ± 0.2 ± 0.1	100.2 ± 0.2 ± 0.1	100.0 ± 0.4 ± 0.3	100.2 ± 0.7 ± 0.2
[-1.81,-1.52]	100.7 ± 2.2 ± 0.7	101.3 ± 0.9 ± 0.4	100.6 ± 0.4 ± 0.2	100.2 ± 0.2 ± 0.2	100.4 ± 0.2 ± 0.1	100.4 ± 0.1 ± 0.0	100.1 ± 0.1 ± 0.0	99.7 ± 0.2 ± 0.2	100.1 ± 0.3 ± 0.2	99.7 ± 0.5 ± 0.3
[-1.52,-1.37]	102.0 ± 4.2 ± 3.9	99.0 ± 1.6 ± 0.9	100.3 ± 0.7 ± 0.4	100.7 ± 0.4 ± 0.1	101.0 ± 0.3 ± 0.2	100.6 ± 0.2 ± 0.1	100.2 ± 0.3 ± 0.3	100.0 ± 0.4 ± 0.2	100.1 ± 0.6 ± 0.2	99.4 ± 1.2 ± 0.7
[-1.37,-1.15]	98.7 ± 2.8 ± 1.1	99.6 ± 1.0 ± 1.2	100.5 ± 0.5 ± 0.3	100.1 ± 0.2 ± 0.2	100.2 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.0	99.9 ± 0.1 ± 0.1	100.2 ± 0.2 ± 0.1	99.9 ± 0.3 ± 0.1	100.1 ± 0.5 ± 0.3
[-1.15,-0.80]	99.5 ± 2.2 ± 2.4	102.0 ± 0.6 ± 0.3	100.3 ± 0.4 ± 0.2	100.1 ± 0.2 ± 0.1	100.0 ± 0.1 ± 0.0	100.0 ± 0.1 ± 0.0	100.0 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.1	99.9 ± 0.2 ± 0.1	100.2 ± 0.3 ± 0.3
[-0.80,-0.60]	100.1 ± 2.6 ± 2.1	101.6 ± 0.8 ± 0.2	100.8 ± 0.5 ± 0.5	100.4 ± 0.2 ± 0.2	100.1 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.1	99.9 ± 0.1 ± 0.1	99.8 ± 0.2 ± 0.1	99.9 ± 0.2 ± 0.3	99.7 ± 0.5 ± 0.3
[-0.60,-0.10]	99.9 ± 1.4 ± 2.0	100.3 ± 0.6 ± 0.5	100.3 ± 0.3 ± 0.2	100.3 ± 0.1 ± 0.1	100.1 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.0	99.8 ± 0.1 ± 0.1	99.9 ± 0.1 ± 0.1	99.7 ± 0.2 ± 0.1	100.0 ± 0.3 ± 0.1
[-0.10,0.00]	101.6 ± 2.2 ± 0.4	101.0 ± 1.1 ± 0.7	100.2 ± 0.5 ± 0.7	100.8 ± 0.3 ± 0.3	100.1 ± 0.2 ± 0.1	100.2 ± 0.1 ± 0.0	99.8 ± 0.2 ± 0.1	100.0 ± 0.2 ± 0.1	99.6 ± 0.4 ± 0.2	99.7 ± 0.7 ± 0.4
[0.00,0.10]	101.9 ± 2.0 ± 0.2	99.8 ± 1.2 ± 1.0	99.8 ± 0.6 ± 0.4	100.1 ± 0.3 ± 0.2	100.2 ± 0.2 ± 0.1	99.9 ± 0.1 ± 0.1	100.1 ± 0.2 ± 0.1	99.9 ± 0.2 ± 0.1	100.0 ± 0.4 ± 0.2	98.9 ± 0.7 ± 0.5
[0.10,0.60]	100.9 ± 1.4 ± 0.7	100.1 ± 0.6 ± 0.9	100.1 ± 0.3 ± 0.1	100.2 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.1	99.9 ± 0.1 ± 0.0	100.1 ± 0.1 ± 0.1	99.9 ± 0.1 ± 0.0	100.1 ± 0.2 ± 0.1	100.2 ± 0.3 ± 0.2
[0.60,0.80]	96.8 ± 3.0 ± 2.1	101.6 ± 0.8 ± 0.3	100.8 ± 0.5 ± 0.3	99.9 ± 0.2 ± 0.3	100.0 ± 0.1 ± 0.1	99.8 ± 0.1 ± 0.1	99.9 ± 0.2 ± 0.1	100.1 ± 0.3 ± 0.2	99.8 ± 0.5 ± 0.3	
[0.80,1.15]	101.4 ± 2.2 ± 1.6	101.5 ± 0.9 ± 0.5	100.8 ± 0.4 ± 0.4	100.0 ± 0.2 ± 0.2	99.9 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.0	99.9 ± 0.1 ± 0.1	100.1 ± 0.1 ± 0.2	100.0 ± 0.2 ± 0.2	99.8 ± 0.4 ± 0.2
[1.15,1.37]	101.7 ± 2.7 ± 2.0	99.1 ± 1.0 ± 1.0	100.8 ± 0.5 ± 0.3	100.2 ± 0.2 ± 0.3	100.3 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.1	100.0 ± 0.1 ± 0.1	99.8 ± 0.2 ± 0.1	100.2 ± 0.3 ± 0.2	99.9 ± 0.5 ± 0.8
[1.37,1.52]	102.5 ± 4.0 ± 4.5	100.9 ± 1.7 ± 0.9	101.6 ± 0.8 ± 0.9							

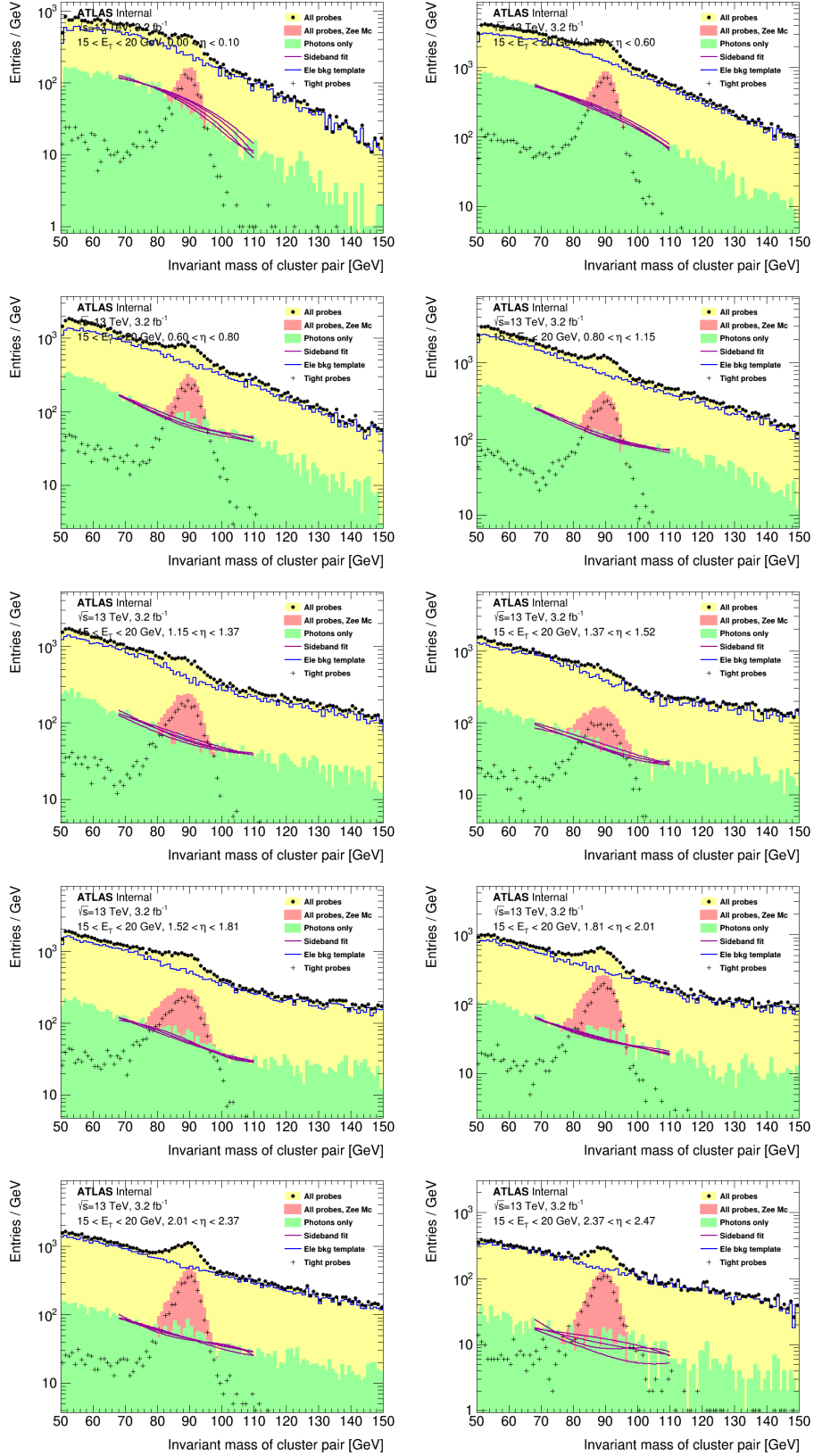


Figure A.1: Invariant mass distributions at the denominator level in the bin  $15 < E_T < 20$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all reconstructed electrons and photons are displayed. Photon background estimation is shown for the different fit ranges used as systematics.

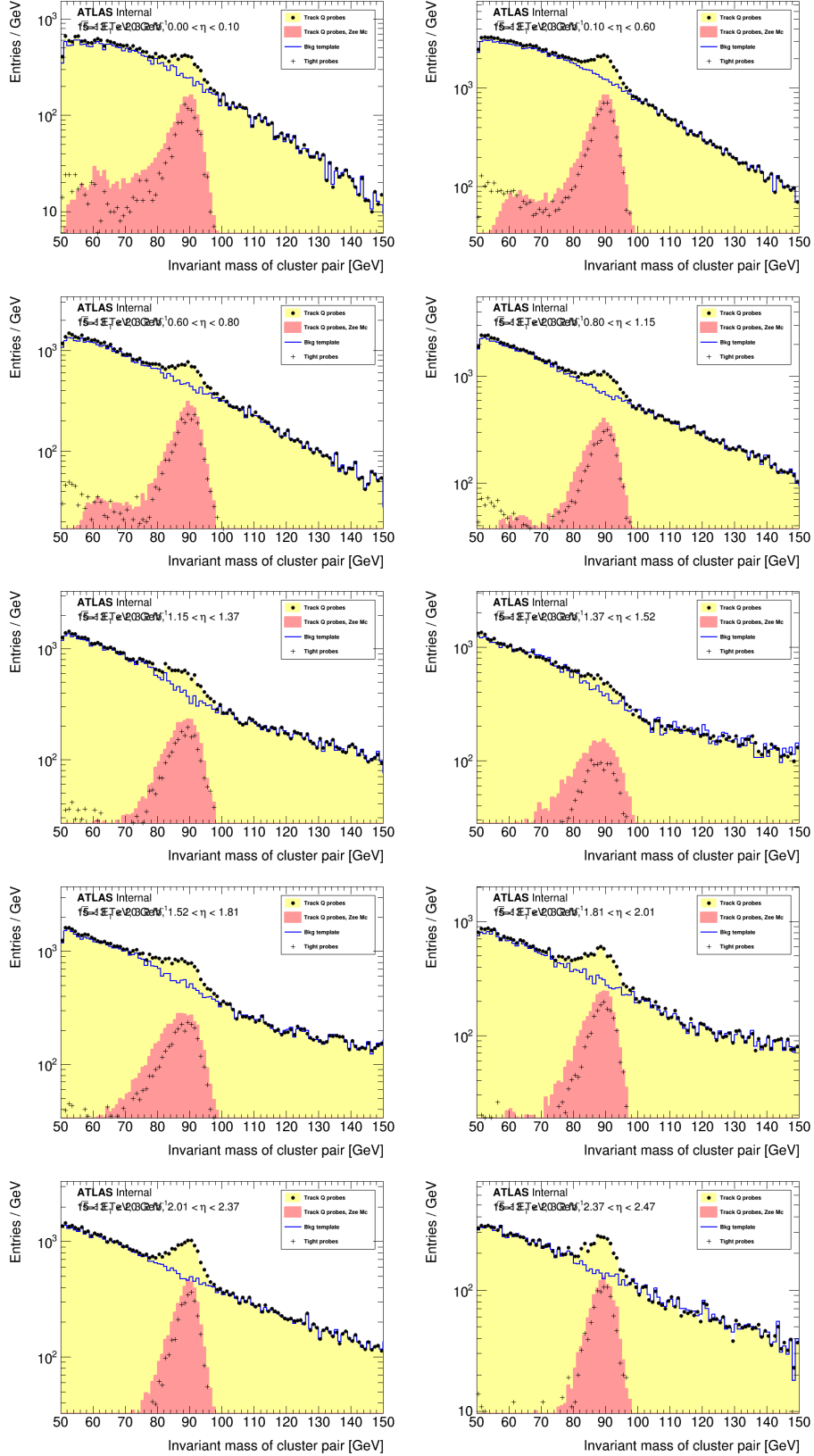


Figure A.2: Invariant mass distributions at the numerator level in the bin  $15 < E_T < 20$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all electrons passing the track quality requirement and the corresponding background estimation are shown.

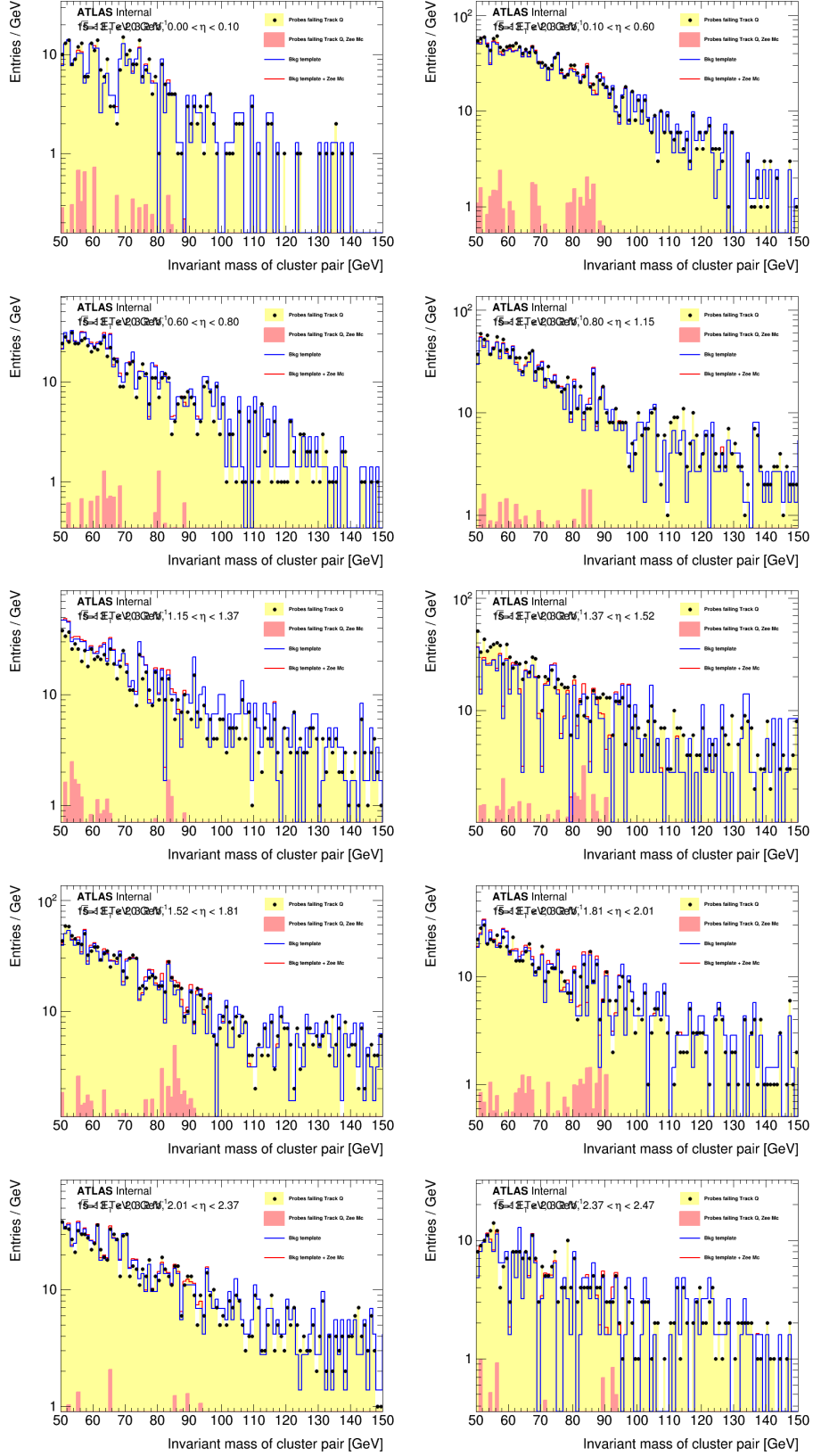


Figure A.3: Invariant mass distributions in the bin  $15 < E_T < 20$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all electrons failing the track quality requirement and the corresponding background estimation are shown.

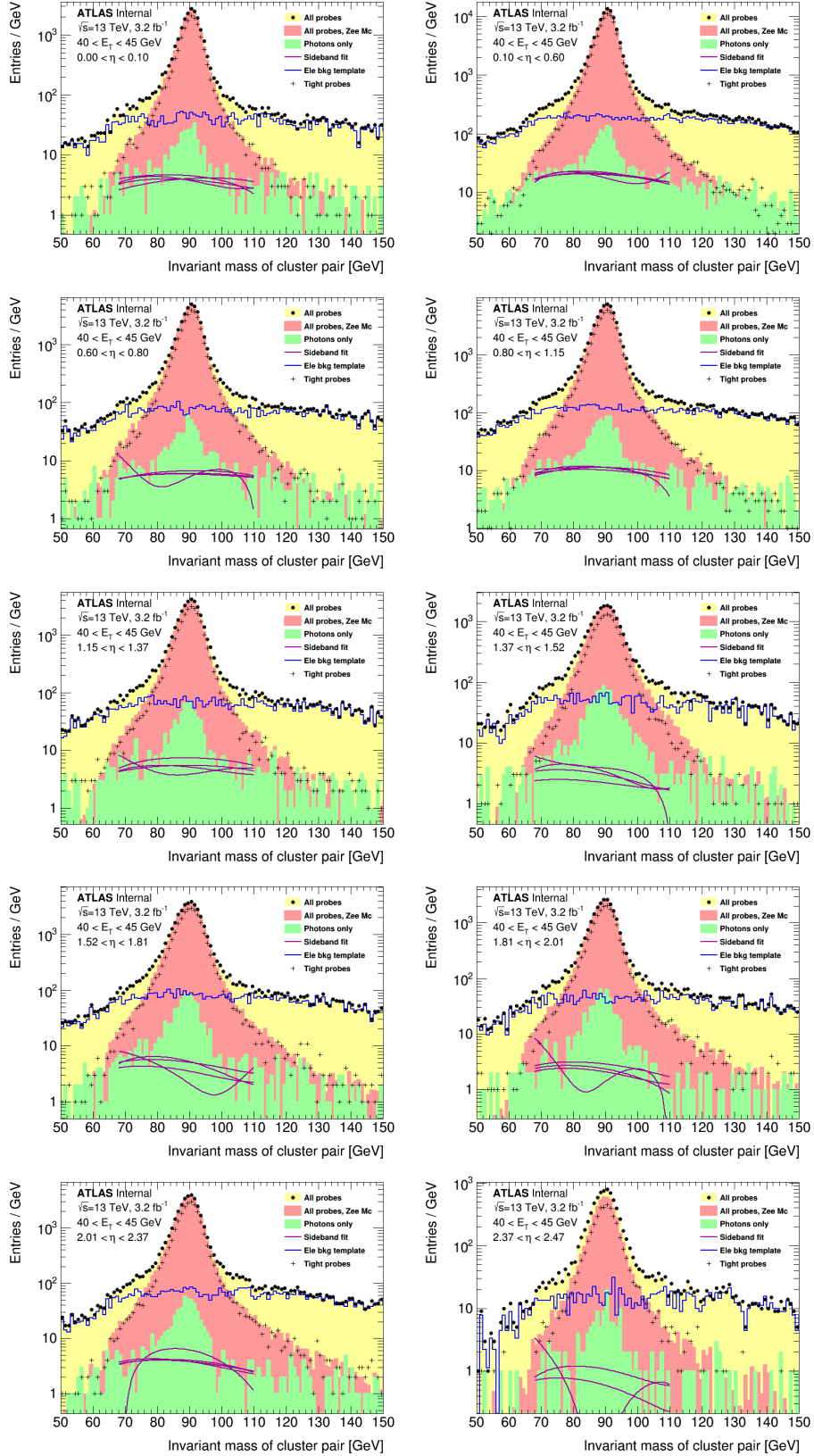


Figure A.4: *Invariant mass distributions at the denominator level in the bin  $40 < E_T < 45$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all reconstructed electrons and photons are displayed. The photon background estimation is shown for the different fit ranges used as systematics.*

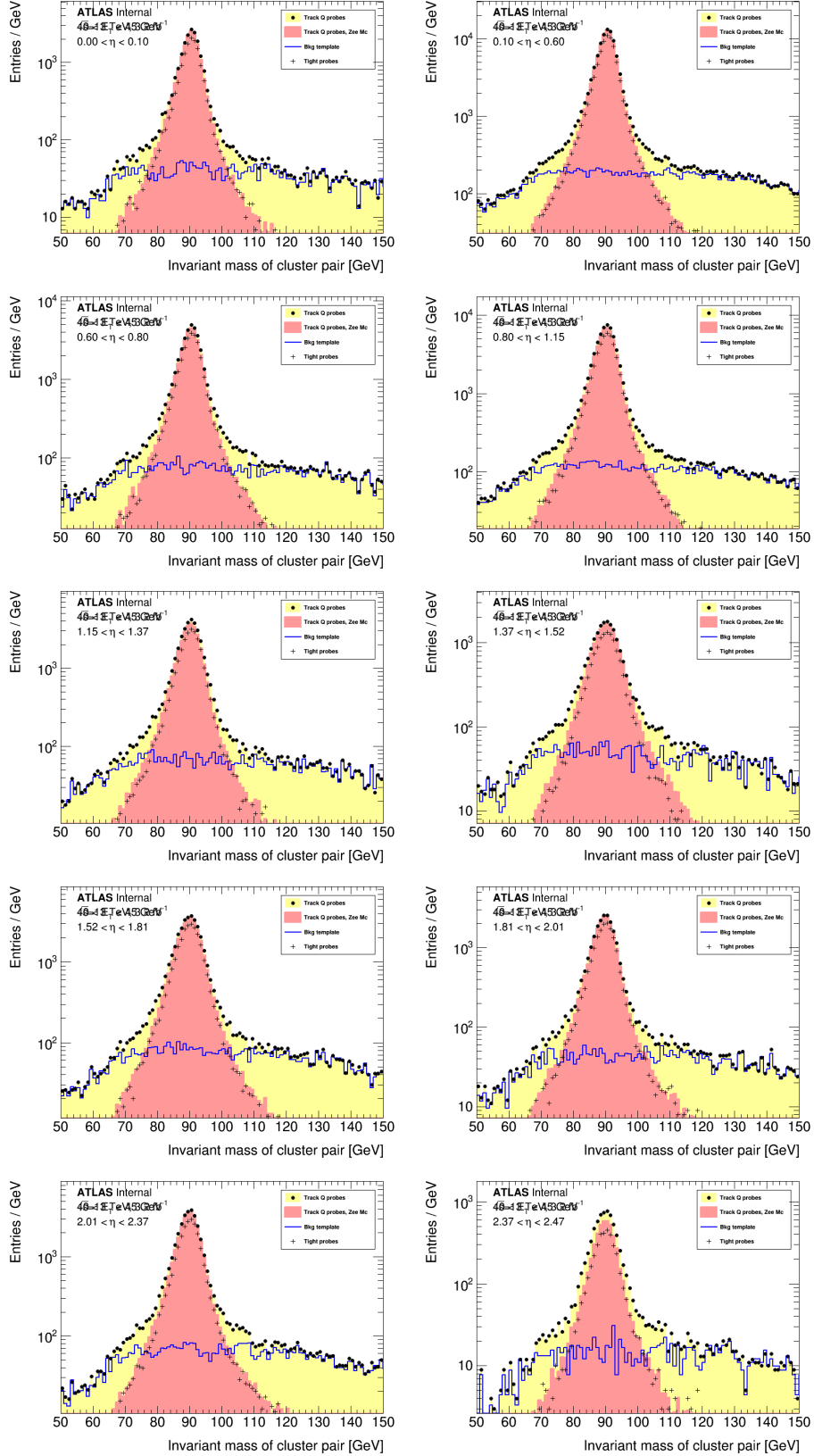


Figure A.5: Invariant mass distributions at the numerator level in the bin  $40 < E_T < 45$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all electrons passing the track quality requirement and the corresponding background estimation are shown.



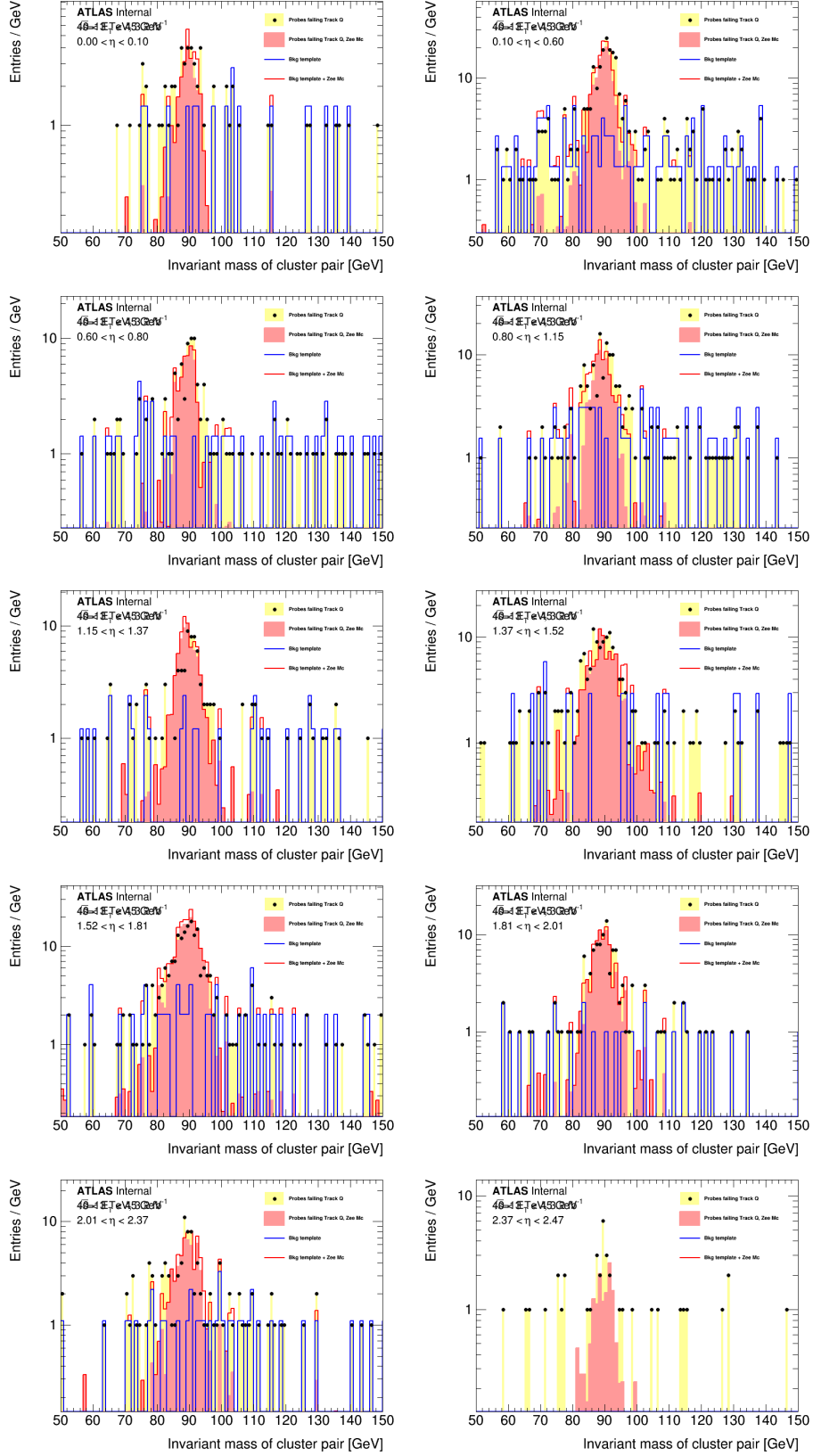


Figure A.6: Invariant mass distributions in the bin  $40 < E_T < 45$  GeV displayed for  $\eta > 0$  (similar plots, results for  $\eta < 0$ ): all electrons failing the track quality requirement and the corresponding background estimation are shown.

# Appendix B

## Background estimation for $t\bar{t}H$ , Multilepton analysis

### B.1 The Matrix Method

Many heuristic approaches are used in ATLAS analyses to set limits on new physics signals having backgrounds from misidentified objects; for example jets misidentified as leptons, b-jets or photons. This appendix discusses the "matrix method" (MM) for estimating the background contribution from such sources. The MM is a data driven technique for estimating the contamination of fake objects (electrons and muons) passing a given selection which corresponds to the object selection used in the SR determination - referred as *tight* selection. The MM makes use of the baseline *loose* lepton selection to estimate for each event its contribution to the fake background.

#### B.1.1 Mathematical description

What is the number of fake leptons in the SR? To answer this question, a data-driven technique is used described in the following. The number of fake leptons in the SR, as a tight region, denoted by  $f_{SR}$  could be linked to the number of fakes in the baseline " $f_{base}$ ", as a loose region (including both tight and *anti-tight* leptons  $L = T \cup \bar{\mathcal{X}}$ ), by the fake rate  $\varepsilon_f$  which is defined as the probability of *loose* fake lepton to pass *tight* selection. So if  $\varepsilon_f$  and  $f_{base}$  are known, then  $f_{SR}$  could be directly retrieved as shown in Equation (B.1).

$$f_{SR} = \varepsilon_f \times f_{base} \quad (\text{B.1})$$

To estimate the number of fakes in the baseline region, where the signal and fake leptons are not negligible, the matrix method, is used which links the unknown number of real and fake events in the baseline region,  $r_{base}$  and  $f_{base}$ , to the observed *tight* and *loose* events ( $T$  and  $L$ ) by the real efficiency  $\varepsilon_r$  and fake rate  $\varepsilon_f$ , defined as the probability of a baseline real and fake lepton, respectively, to pass *tight* selection. Equation (B.2) shows the matrix method formula applied on 'leptons' where  $f_{base}$  could be written as a function of  $f(\varepsilon_f, \varepsilon_r, T, L)$ . Instead *Loose*, we can use *anti-tight*  $\bar{\mathcal{X}}$  in the matrix method as shown later. The corresponding formula applied on dilepton 'events' is described in B.11 where *loose* could be written as:  $L = TT + T\bar{\mathcal{X}} + \bar{\mathcal{X}}T + \bar{\mathcal{X}}\bar{\mathcal{X}}$  and fake events can be defined in four dileptonic regions characterised by different real and fake lepton composition:  $f_{base} = rf + fr + ff$ .

$$\begin{pmatrix} T \\ L \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ 1 & 1 \end{pmatrix} \begin{pmatrix} r_{base} \\ f_{base} \end{pmatrix}, \quad (\text{B.2})$$

real and fake efficiencies ( $\varepsilon_r$  and  $\varepsilon_f$ ) are estimated in other regions, called control regions "CR", defined in such way to get efficiencies as "same" values as in the SR but considered as real and fake-enriched regions respectively, where the signal is negligible. For illustration, five control regions, in addition to the SR, are defined in Figure 6.9 to make use of the matrix method. Three discriminative variables are used to distinguish between SR and other control regions, namely the number of jets  $N_{jets}$ , assuring the orthogonality between signal and CR where efficiencies are estimated, lepton selection (*Tight* T or *Loose* L defined in Table 6.4), assuring enough statistics for fakes, and finally Lepton charge (opposite-sign *OS* or same-sign *SS*) giving the orthogonality between real or fake enriched-regions respectively.

- $CR_1^{OS}$  : event with 2-3-4 jet multiplicity and two *tight* opposite-sign leptons.
- $CR_1^{SS}$  : event with 2-3-4 jet multiplicity and two *tight* same-sign leptons.
- $CR_2^{OS}$  : event with 2-3-4 jet multiplicity and two *loose* opposite-sign leptons.
- $CR_2^{SS}$  : event with 2-3-4 jet multiplicity and two *loose* same-sign leptons.
- $CR_3$  : event with at least five jets and two *loose* same-sign leptons. It is also called "baseline" region as defined in Table 6.6.
- $SR$  : event with at least five jets and two *tight* same-sign leptons.

$f_{base}$  is estimated in  $CR_3$ . The fake rate is estimated in  $CR_1^{SS}$  and  $CR_2^{SS}$ . The real efficiency, estimated in  $CR_1^{OS}$  and  $CR_2^{OS}$ . Equations B.3 and B.4 show how both real and fake efficiencies could be estimated from measurable observables ( $T_i$  and  $L_i$ ) instead unknown variables ( $r_i$  and  $f_i$ ).

$$\varepsilon_f = \frac{f_{SR}}{f_3} = \frac{f_1}{f_2} = \frac{f_1^{SS}}{f_2^{SS}} = \frac{T_1^{SS}}{L_2^{SS}} \quad (B.3)$$

$$\varepsilon_r = \frac{r_{SR}}{r_3} = \frac{r_1}{r_2} = \frac{r_1^{OS}}{r_2^{OS}} = \frac{T_1^{OS}}{L_2^{OS}} \quad (B.4)$$

where:

- $T_i$  : Number of leptons that pass *tight* in  $CR_i$ .
- $L_i$  : Number of leptons that pass *loose* in  $CR_i$ .
- $f_i$  : Number of fakes in  $CR_i$ .
- $r_i$  : Number of reals in  $CR_i$ .

$f_1$  and  $f_2$  are equivalent to  $T_1^{SS}$  and  $L_2^{SS}$  respectively.  $r_1$  and  $r_2$  are equivalent to  $T_1^{OS}$  and  $L_2^{OS}$  respectively.  $f_3 = f_{base}$ .

At any region, the number of events could be written as the sum of signal, prompt background, fakes and charge flip (*QmisId*) as shown in Equation (B.5).

$$N = S + B^{prompt} + f + QmisId \quad (B.5)$$

However, in the following we subtract any *QmisId* event in order to come up with only real and fake events where real leptons  $r_{SR}$  could be either signal or prompt background

$r_{SR} = S + B^{prompt}$ . Then, the number of events in the SR is a set of real and fake events only as shown in Equation (B.6).

$$N_{SR} = r_{SR} + f_{SR} \quad (\text{B.6})$$

To estimate the fake rate, all selected leptons that pass/fail *tight* selection in the fake-enriched region  $CR_i^{SS}$  should be fakes. For real efficiency, all selected leptons that pass/fail *tight* selection in the real-enriched region  $CR_i^{OS}$  should be real. Therefore, a background subtraction is needed to remove real and charge flip contamination from  $CR_i^{SS}$  and fake contamination from  $CR_i^{OS}$ . Any residual contamination should be added as systematics.

## B.1.2 Di-lepton category

Depending on whether or not each lepton passes the tight selection, each  $i$ -th event can be categorised into any of four orthogonal (*sidebands*) regions:

- $TT_i$  : event with both leptons passing tight selection
- $T\bar{T}_i$  : event with leading lepton passing tight cut and subleading lepton failing tight cut
- $\bar{T}T_i$  : event with leading lepton failing tight cut and subleading lepton passing tight cut
- $\bar{T}\bar{T}_i$  : event with both leptons failing tight selection

, where  $T$  denotes leptons passing the tight cuts,  $\bar{T}$  denotes leptons failing tight but passing baseline selection. In the latter case, they will be referred to as *anti-tight* leptons.

Baseline leptons will be also identified with  $L$  in the context of this chapter which includes both tight and anti-tight leptons:  $L = T \cup \bar{T}$ . *loose* events can be defined in four dileptonic regions characterised by different real and fake lepton composition, namely:

- $rr_i$  : event with both leptons being real
- $rf_i$  : event with leading lepton being real and subleading lepton being fake
- $fr_i$  : event with leading lepton being fake and subleading lepton being real
- $ff_i$  : event with both leptons being fake

Then, the total number of events including 2 same sign leptons pass *loose* selection and  $N_{\text{jets}} \geq 5$  is just the sum of events with different combination of real and fake leptons that could pass or not tight selection as described in equation B.7.

$$LL = \sum_i (rr_i + rf_i + fr_i + ff_i) = \sum_i (TT_i + T\bar{T}_i + \bar{T}T_i + \bar{T}\bar{T}_i) \quad (\text{B.7})$$

The number of events in the signal region could be written as:

$$TT_i = N_{rr}^{TT} + N_{rf}^{TT} + N_{fr}^{TT} + N_{ff}^{TT} = N_{rr}^{TT} + N_f^{TT} \quad (\text{B.8})$$

The number of Tight events are linked to Loose events using real and fake efficiencies (since efficiency is defined as the number of tight to loose events) as following:

$$N_{rr}^{TT} = \sum_i (\varepsilon_{r,1} \varepsilon_{r,2} rr)_i \quad (\text{B.9})$$

$$N_f^{TT} = N_{rf}^{TT} + N_{fr}^{TT} + N_{ff}^{TT} = \sum_i (\varepsilon_{f,1} \varepsilon_{f,2} ff)_i + (\varepsilon_{r,1} \varepsilon_{f,2} rf)_i + (\varepsilon_{f,1} \varepsilon_{r,2} fr)_i \quad (\text{B.10})$$

The  $4 \times 4$  matrix equation can be written as:

$$\begin{pmatrix} TT \\ T\bar{T} \\ \bar{T}T \\ \bar{T}\bar{T} \end{pmatrix}_i = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\bar{\varepsilon}_{r,2} & \varepsilon_{r,1}\bar{\varepsilon}_{f,2} & \varepsilon_{f,1}\bar{\varepsilon}_{r,2} & \varepsilon_{f,1}\bar{\varepsilon}_{f,2} \\ \bar{\varepsilon}_{r,1}\varepsilon_{r,2} & \bar{\varepsilon}_{r,1}\varepsilon_{f,2} & \bar{\varepsilon}_{f,1}\varepsilon_{r,2} & \bar{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{f,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2} \end{pmatrix}_i \begin{pmatrix} rr \\ rf \\ fr \\ ff \end{pmatrix}_i, \quad (\text{B.11})$$

where  $\varepsilon_r$  ( $\varepsilon_f$ ) represent the efficiency for a real (fake) lepton to pass tight selection, and  $\bar{\varepsilon}_r \equiv (1 - \varepsilon_r)$  ( $\bar{\varepsilon}_f \equiv (1 - \varepsilon_f)$ ) represent the probability for a real (fake) lepton to fail tight but to pass baseline selection. In this analysis the indexes for  $\varepsilon_r$  and  $\varepsilon_f$  are parameterised in terms of lepton  $p_T$  ordering.

To obtain the number of fakes in signal region, the  $4 \times 4$  matrix can be inverted to:

$$\begin{pmatrix} rr \\ rf \\ fr \\ ff \end{pmatrix}_i = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\bar{\varepsilon}_{r,2} & \varepsilon_{r,1}\bar{\varepsilon}_{f,2} & \varepsilon_{f,1}\bar{\varepsilon}_{r,2} & \varepsilon_{f,1}\bar{\varepsilon}_{f,2} \\ \bar{\varepsilon}_{r,1}\varepsilon_{r,2} & \bar{\varepsilon}_{r,1}\varepsilon_{f,2} & \bar{\varepsilon}_{f,1}\varepsilon_{r,2} & \bar{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{r,1}\bar{\varepsilon}_{f,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{r,2} & \bar{\varepsilon}_{f,1}\bar{\varepsilon}_{f,2} \end{pmatrix}_i^{-1} \begin{pmatrix} TT \\ T\bar{T} \\ \bar{T}T \\ \bar{T}\bar{T} \end{pmatrix}_i, \quad (\text{B.12})$$

We assume that Loose event could be either real or fake. In case of two lepton per event, It is real if both leptons are real. However, it is fake if at least one of the two leptons is fake (the leading lepton is fake or the subleading lepton is fake or both of them are fakes).

## B.2 Selection of validation regions

Table B.1 shows the selections used to define the validation regions (VR) for  $t\bar{t}Z$ ,  $t\bar{t}W$  and diboson simulated samples used to estimate the prompt background in  $t\bar{t}H$  multilepton analysis. The signal regions selections are also shown for all categories:  $2\ell 0\tau_{\text{had}}$ ,  $2\ell 1\tau_{\text{had}}$ ,  $3\ell$  and  $4\ell$  channels.

## B.3 Fake lepton purity

A study on the lepton flavour choice in the SS control region of the  $2\ell\text{ss}$  channel is shown in Figure B.1. The best compromise is corresponding to the highest fake purity shown in the bottom of each plot. The fake purity is tested in both the *tight* and *loose* selections (Numerator and denominator). The fake lepton background from  $t\bar{t}$  (purple area) is just depicted for comparison with the data-driven fake estimation.

## B.4 Charge flip

Assuming that  $\epsilon$  is the rate of charge mis-identification for a single electron, there are three possibilities for this event to be reconstructed:

1.  $e^+e^- + X$  without any charge mis-identification, with a probability of  $(1 - \epsilon)^2$ ,
2.  $e^+e^- + X$  with the two electrons having a charge flip, with a probability of  $\epsilon^2$ ,

Table B.1: Selections for the signal regions (SR) and validation regions (VR). The variable  $H_{T,\text{jets}}$  is the scalar sum of transverse momenta for the considered jets. Same-flavour, opposite-charge lepton pairs are referred to as SFOC pairs. Trigger-matched leptons correspond to an object reconstructed by the trigger, and must have  $p_T > 25$  GeV (21 GeV for muons in 2015 data). In all regions at least one selected light lepton is required to be trigger-matched.

SR/VR	Channel	Selection criteria
SR	$2\ell 0\tau_{\text{had}}$	Two tight light leptons with $p_T > 25, 25$ GeV Sum of light lepton charges $\pm 2$ Any electrons must have $ \eta_e  < 1.37$ Zero $\tau_{\text{had}}$ candidates $N_{\text{jets}} \geq 5$ and $N_{b\text{-jets}} \geq 1$
SR	$2\ell 1\tau_{\text{had}}$	Two tight light leptons, with $p_T > 25, 15$ GeV Sum of light lepton charges $\pm 2$ Exactly one $\tau_{\text{had}}$ candidate, of opposite charge to the light leptons $ m(ee) - 91.2 \text{ GeV}  > 10$ GeV for $ee$ events $N_{\text{jets}} \geq 4$ and $N_{b\text{-jets}} \geq 1$
SR	$3\ell$	Three light leptons; sum of light lepton charges $\pm 1$ Two same-charge leptons must be tight and have $p_T > 20$ GeV $m(\ell^+\ell^-) > 12$ GeV and $ m(\ell^+\ell^-) - 91.2 \text{ GeV}  > 10$ GeV for all SFOC pairs $ m(3\ell) - 91.2 \text{ GeV}  > 10$ GeV $N_{\text{jets}} \geq 4$ and $N_{b\text{-jets}} \geq 1$ , or $N_{\text{jets}} = 3$ and $N_{b\text{-jets}} \geq 2$
SR	$4\ell$	Four light leptons; sum of light lepton charges 0 All leptons pass “gradient” isolation selection $m(\ell^+\ell^-) > 12$ GeV and $ m(\ell^+\ell^-) - 91.2 \text{ GeV}  > 10$ GeV for all SFOC pairs $100 \text{ GeV} < m(4\ell) < 350 \text{ GeV}$ and $ m(4\ell) - 125 \text{ GeV}  > 5$ GeV $N_{\text{jets}} \geq 2$ and $N_{b\text{-jets}} \geq 1$
VR	Tight $t\bar{t}Z$	$3\ell$ lepton selection At least one $\ell^+\ell^-$ pair with $ m(\ell^+\ell^-) - 91.2 \text{ GeV}  < 10$ GeV $N_{\text{jets}} \geq 4$ and $N_{b\text{-jets}} \geq 2$
VR	Loose $t\bar{t}Z$	$3\ell$ lepton selection At least one $\ell^+\ell^-$ pair with $ m(\ell^+\ell^-) - 91.2 \text{ GeV}  < 10$ GeV $N_{\text{jets}} \geq 4$ and $N_{b\text{-jets}} \geq 1$ , or $N_{\text{jets}} = 3$ and $N_{b\text{-jets}} \geq 2$
VR	$WZ + 1 b\text{-tag}$	$3\ell$ lepton selection At least one $\ell^+\ell^-$ pair with $ m(\ell^+\ell^-) - 91.2 \text{ GeV}  < 10$ GeV $N_{\text{jets}} \geq 1$ and $N_{b\text{-jets}} = 1$
VR	$t\bar{t}W$	$2\ell 0\tau_{\text{had}}$ lepton selection $2 \leq N_{\text{jets}} \leq 4$ and $N_{b\text{-jets}} \geq 2$ $H_{T,\text{jets}} > 220$ GeV for $ee$ and $e\mu$ events $E_T^{\text{miss}} > 50$ GeV and $(m(ee) < 75$ or $m(ee) > 105 \text{ GeV})$ for $ee$ events

3.  $e^\pm e^\pm + X$  when only one of the two electrons is mis-identified, with a probability of  $2\epsilon(1 - \epsilon)$ .

Therefore, if there are  $N$  true opposite-sign events, the reconstructed events will be:

- $N^{os} = (1 - 2\epsilon + 2\epsilon^2)N$  opposite-sign events,
- $N^{ss} = 2\epsilon(1 - \epsilon)N$  same-sign events,

For the  $e^\pm e^\pm$  channel:

$$N_{ee}^{ss} = \frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1 - \epsilon_i - \epsilon_j + 2\epsilon_i\epsilon_j} N^{os} = w_{TT,ee}^{QMisID} N^{os} = 6.9 \pm 1.3$$

For the  $e^\pm \mu^\pm$  channel:

$$N_{e\mu}^{ss} = \frac{\epsilon}{1 - \epsilon} N^{os} = w_{TT,e\mu}^{QMisID} N^{os} = 7.1 \pm 1.7$$

where  $\epsilon_i$  and  $\epsilon_j$  are the charge mis-identification rates for the two different electrons.  $w_{TT}^{QMisID,ee}$  and  $w_{TT}^{QMisID,e\mu}$  are charge mis-identification weights (QmisId weights) measured for *tight* electrons in  $ee$  and  $e\mu$  channels respectively.

For completeness, Figure B.2 shows the errors on the charge flip rates.

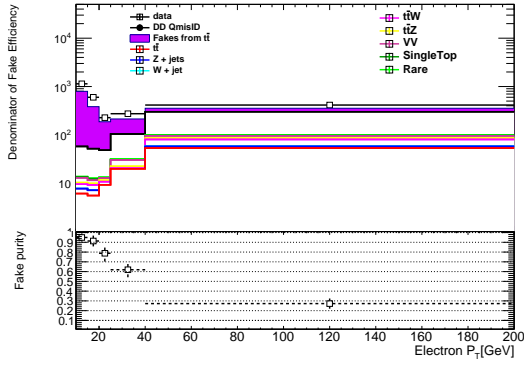
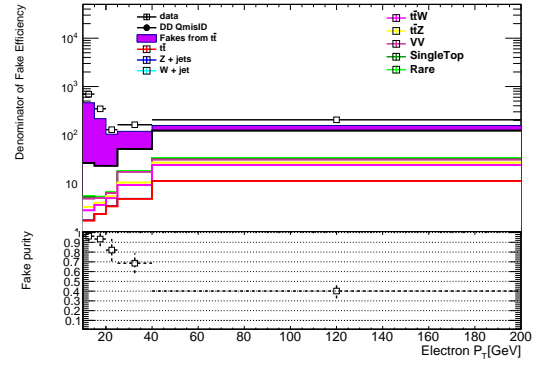
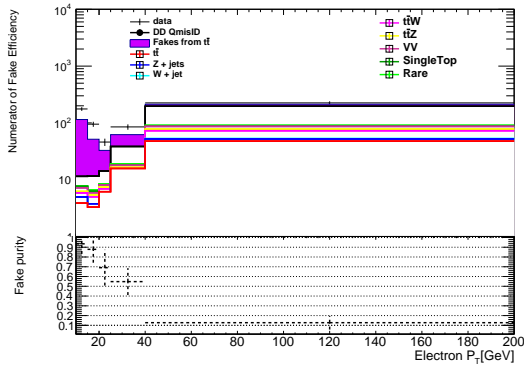
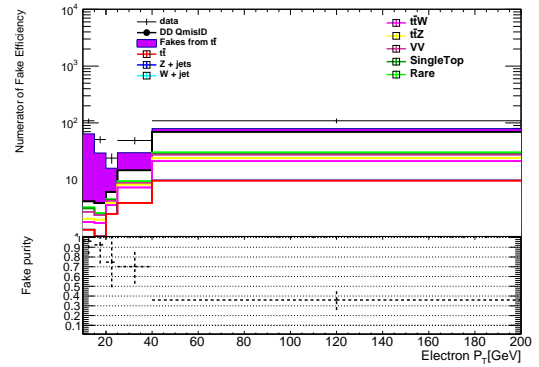
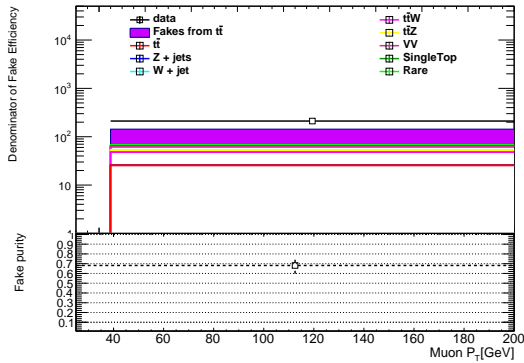
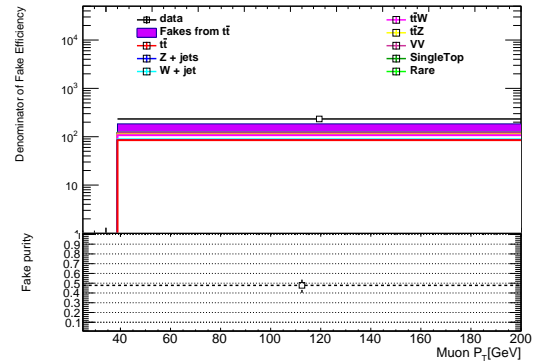
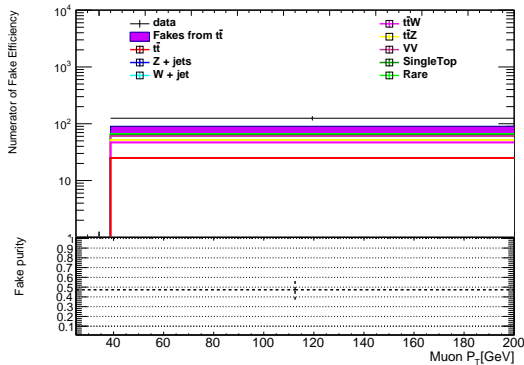
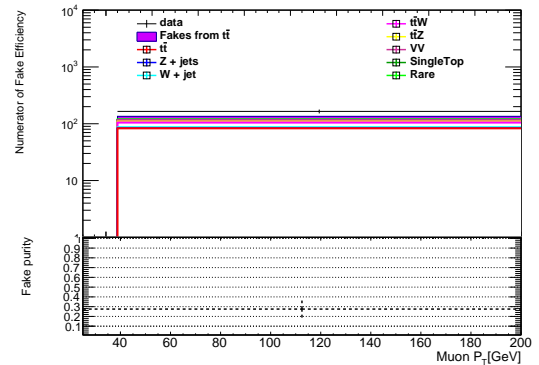
(a) OF+SF  $e$  loose(b) OF  $e$  loose(c) OF+SF  $e$  tight(d) OF  $e$  tight(e) SF  $\mu$  loose(f) OF  $\mu$  loose(g) SF  $\mu$  tight(h) OF  $\mu$  tight

Figure B.1:  $p_T$  distribution of the electron and muon fake purity vs. Lepton flavour (opposit flavour (OF), same flavour (SF)) in the SS control region (2-3-4 jet multiplicity) of the  $2l_{SS}$  channel when the lepton passes tight/loose selection in the Numerator/denominator using data and the associated prompt/data-driven QmisID background. Error bars account for statistical uncertainty only.



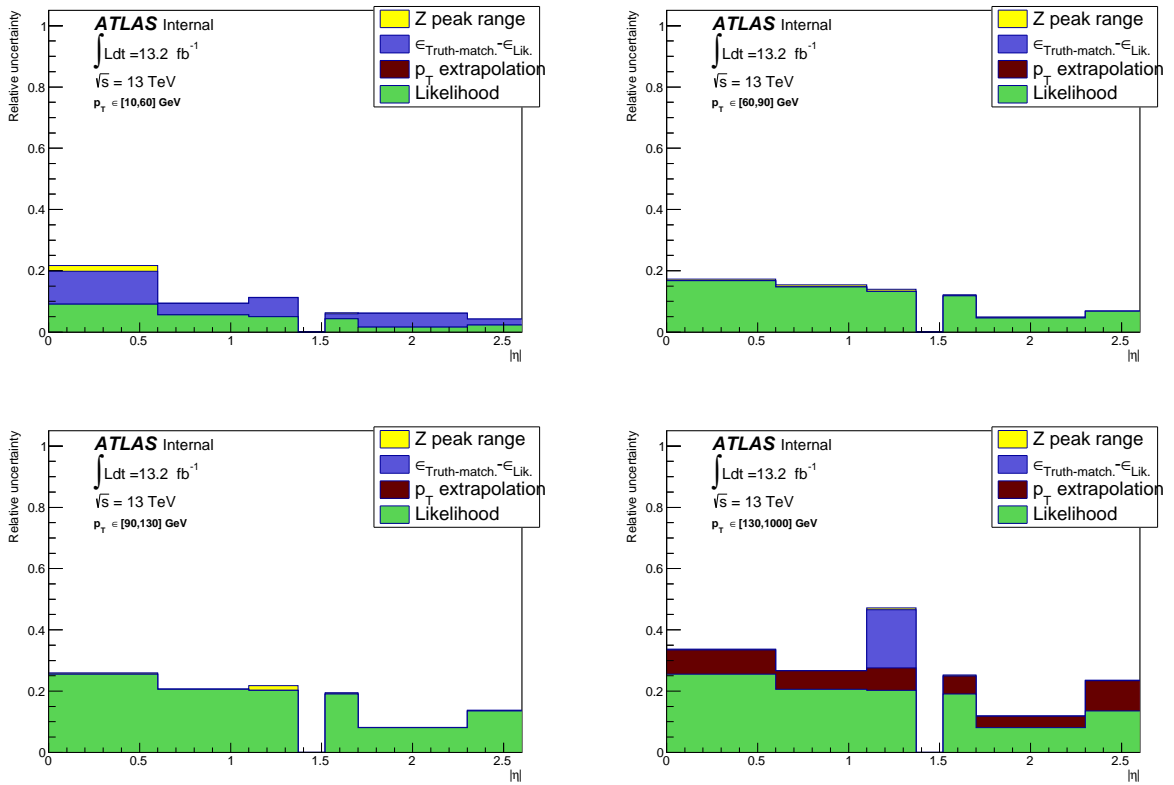


Figure B.2: Quadratic sum of systematic uncertainty contributions on the charge mis-identification rate, for different bins in  $p_T$  and  $|\eta|$ . Each systematic uncertainty is stacked to the others.



# Bibliography

- [1] Y. Nakahama, “The ATLAS Trigger System: Ready for Run-2,” *Journal of Physics: Conference Series* **664** no. 8, (2015) 082037.  
<http://stacks.iop.org/1742-6596/664/i=8/a=082037>.
- [2] R. Loss, “Atomic Weights of the Elements 2001,” tech. rep., IUPAC, Australia, 2009.
- [3] T. Renner, *Quantities, Units and Symbols in Physical Chemistry*. The Royal Society of Chemistry, 2007.
- [4] L. D. Landau and E. M. Lifshitz, *Mechanics*. Butterworth-Heinemann, 3 edition, 1976.
- [5] M. W. Zemansky, *Heat and Thermodynamics*. McGraw-Hill Book Company Inc, 1968.
- [6] J. C. Maxwell and W. D. Niven, *The Scientific Papers of James Clerk Maxwell*. Cambridge University Press, 2011.
- [7] J. J. Thomson, “Cathode Rays,” *Philosophical Magazine* **90** no. sup1, (2010) 25–29.
- [8] S. Weinberg, “The Making of the standard model,” *Eur. Phys. J.* **C34** (2004) 5–13, [arXiv:hep-ph/0401010](https://arxiv.org/abs/hep-ph/0401010) [hep-ph]. [,99(2005)].
- [9] W. Pauli, *Exclusion Principle and Quantum Mechanics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1994.
- [10] **Particle Data Group** Collaboration, “Review of Particle Physics,” *Chin. Phys.* **C40** no. 10, (2016) 100001.
- [11] J. W. Rohlf, *Modern Physics from A to Z*. John Wiley and Sons, New York, 1994.
- [12] J. Mohit, *The Mysterious World of Fundamental Particles : Cosmic Beginnings*. CreateSpace Independent Publishing Platform, 2015.
- [13] D. Hanneke, S. Fogwell, and G. Gabrielse, “New Measurement of the Electron Magnetic Moment and the Fine Structure Constant,” *Physical Review Letters* **100** no. 12, (Mar., 2008) 120801, [arXiv:0801.1134](https://arxiv.org/abs/0801.1134) [physics.atom-ph].
- [14] L. Fuchs, *Abelian Groups*. Springer International Publishing, 2015.
- [15] B. De Wit and J. Smith, *Field Theory in Particle Physics, Volume 1*. North-Holland, 1986.

- [16] H. Yukawa, “On the Interaction of Elementary Particles,” *Proc. Phys.-Math. Soc. Jpn.* no. 17, (1935) 48–57.
- [17] B. Ryden, *Introduction to cosmology*. Addison-Wesley, 2003.
- [18] S. Weinberg, *Dreams of a Final Theory*. Vintage, 1994.
- [19] N. Iain, *Dark Side of the Universe: Dark Matter, Dark Energy, and the Fate of the Cosmos*. Johns Hopkins University Press, 2007.
- [20] K. Freeman and G. McNamara, *In Search of Dark Matter*. Springer-Verlag New York, 2007.
- [21] D. Griffiths, *Introduction to Elementary Particles, 2nd, Revised Edition*. Wiley-VCH, 2008.
- [22] I. Baldes, *Early Universe Cosmology and the Matter-antimatter Asymmetry*. 2015.
- [23] F. Suekane, *Neutrino Oscillations: A Practical Guide to Basics and Applications*. Springer Japan, 2015.
- [24] I. Melzer-Pellmann and P. Pralavorio, “Lessons for SUSY from the LHC after the first run,” *Eur. Phys. J.* **C74** (2014) 2801, [arXiv:1404.7191 \[hep-ex\]](#).
- [25] J. L. Feng *et al.*, “Protophobic Fifth-Force Interpretation of the Observed Anomaly in  $^8\text{Be}$  Nuclear Transitions,” *Phys. Rev. Lett.* **117** no. 7, (2016) 071803, [arXiv:1604.07411 \[hep-ph\]](#).
- [26] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. 1995.
- [27] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.* **13** (1964) 508–509.
- [28] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.
- [29] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles,” *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.
- [30] **ATLAS** Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](#).
- [31] **CMS** Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](#).
- [32] **ATLAS, CMS** Collaboration, “Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments,” *Phys. Rev. Lett.* **114** (2015) 191803, [arXiv:1503.07589 \[hep-ex\]](#).
- [33] **LHC Higgs Cross Section Working Group** Collaboration, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,” [arXiv:1610.07922 \[hep-ph\]](#).

- [34] **LHC Higgs Cross Section Working Group** Collaboration, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties,” [arXiv:1307.1347 \[hep-ph\]](#).
- [35] D. Green, “Vector boson fusion Higgs production at the LHC: Mass variables,” [arXiv:hep-ex/0501027 \[hep-ex\]](#).
- [36] S. Dittmaier and M. Schumacher, “The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate,” *Prog. Part. Nucl. Phys.* **70** (2013) 1–54, [arXiv:1211.4828 \[hep-ph\]](#).
- [37] **ATLAS** Collaboration, “Measurement of the Higgs boson mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  channels with the ATLAS detector using  $25 \text{ fb}^{-1}$  of  $pp$  collision data,” *Phys. Rev.* **D90** no. 5, (2014) 052004, [arXiv:1406.3827 \[hep-ex\]](#).
- [38] **CMS** Collaboration, “Measurement of the properties of a Higgs boson in the four-lepton final state,” *Phys. Rev.* **D89** no. 9, (2014) 092007, [arXiv:1312.5353 \[hep-ex\]](#).
- [39] **ATLAS** Collaboration, “Measurement of the  $W$ -boson mass in  $pp$  collisions at  $\sqrt{s} = 7 \text{ TeV}$  with the ATLAS detector,” [arXiv:1701.07240 \[hep-ex\]](#).
- [40] **CMS** Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV,” *Eur. Phys. J.* **C75** no. 5, (2015) 212, [arXiv:1412.8662 \[hep-ex\]](#).
- [41] **ATLAS** Collaboration, “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector,” *Eur. Phys. J.* **C75** no. 10, (2015) 476, [arXiv:1506.05669 \[hep-ex\]](#). [Erratum: *Eur. Phys. J.*C76,no.3,152(2016)].
- [42] **CMS** Collaboration, “Search for Higgs boson off-shell production in proton-proton collisions at 7 and 8 TeV and derivation of constraints on its total decay width,” *JHEP* **09** (2016) 051, [arXiv:1605.02329 \[hep-ex\]](#).
- [43] A. Hadeef, “Off-shell Higgs boson production in vector boson fusion at the large hadron collider,”. Master thesis.
- [44] **CMS** Collaboration, “Limits on the Higgs boson lifetime and width from its decay to four charged leptons,” *Phys. Rev.* **D92** no. 7, (2015) 072010, [arXiv:1507.06656 \[hep-ex\]](#).
- [45] **ATLAS, CMS** Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC  $pp$  collision data at  $\sqrt{s} = 7$  and 8 TeV,” *JHEP* **08** (2016) 045, [arXiv:1606.02266 \[hep-ex\]](#).
- [46] J. Ellis and T. You, “Updated Global Analysis of Higgs Couplings,” *JHEP* **06** (2013) 103, [arXiv:1303.3879 \[hep-ph\]](#).
- [47] **ATLAS** Collaboration, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into  $b\bar{b}$  in  $pp$  collisions at  $\sqrt{s} = 8 \text{ TeV}$  with the ATLAS detector,” *Eur. Phys. J.* **C75** no. 7, (2015) 349, [arXiv:1503.05066 \[hep-ex\]](#).

- [48] **ATLAS** Collaboration, “Search for  $H \rightarrow \gamma\gamma$  produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector,” *Phys. Lett. B* **740** (2015) 222–242, [arXiv:1409.3122](https://arxiv.org/abs/1409.3122) [hep-ex].
- [49] **ATLAS** Collaboration, “Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector,” *Phys. Lett. B* **749** (2015) 519–541, [arXiv:1506.05988](https://arxiv.org/abs/1506.05988) [hep-ex].
- [50] **CMS** Collaboration, “Search for the associated production of the Higgs boson with a top-quark pair,” *JHEP* **09** (2014) 087, [arXiv:1408.1682](https://arxiv.org/abs/1408.1682) [hep-ex].
- [51] J. Lévêque and A. Rozanov, *Recherche d’un boson de Higgs léger produit en association avec une paire de quarks top dans l’expérience ATLAS*. PhD thesis, CPPM, Aix-Marseille U., 2003. <https://cds.cern.ch/record/722063>. Presented on 30 Jun 2003.
- [52] F. Maltoni, D. L. Rainwater, and S. Willenbrock, “Measuring the top quark Yukawa coupling at hadron colliders via  $t\bar{t}H, H \rightarrow W^+W^-$ ,” *Phys. Rev. D* **66** (2002) 034022, [arXiv:hep-ph/0202205](https://arxiv.org/abs/hep-ph/0202205) [hep-ph].
- [53] **ATLAS** Collaboration, “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at  $\sqrt{s} = 7$  and 8 TeV in the ATLAS experiment,” Tech. Rep. ATLAS-CONF-2015-007, CERN, Geneva, Mar, 2015. <https://cds.cern.ch/record/2002212>.
- [54] K. Liu *et al.*, “Search for the Associated Production of a Higgs Boson and a Top Quark Pair in multilepton final states in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector,” Tech. Rep. ATL-COM-PHYS-2017-101, CERN, Geneva, Feb, 2017. <https://cds.cern.ch/record/2244753>.
- [55] “LHC Guide.” CERN-Brochure-2017-002-Eng, Mar, 2017.
- [56] L. Evans and P. Bryant, “LHC Machine,” *JINST* **3** (2008) S08001.
- [57] “Luminosity: Taking a closer look at LHC.” [https://www.lhc-closer.es/taking\\_a\\_closer\\_look\\_at\\_lhc/0.luminosity](https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.luminosity).
- [58] H. K. Jarlett, “Power to the LHC,”. <https://cds.cern.ch/record/2143302>.
- [59] **ATLAS** Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST* **3** (2008) S08003.
- [60] **ATLAS** Collaboration, “Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics,” [arXiv:0901.0512](https://arxiv.org/abs/0901.0512) [hep-ex].
- [61] N. Wermes, L. Rossi, P. Fischer, and T. Rohe, *Pixel Detectors, From Fundamentals to Applications*. Springer-Verlag, 2006.
- [62] K. Mochizuki, *Search for the Higgs boson in the  $WH \rightarrow \ell v b\bar{b}$  channel with the ATLAS detector - Development of high performance b-jet identification algorithms*. PhD thesis, Aix-Marseille U., Oct, 2015. <http://cds.cern.ch/record/2227038>. Presented 27 Oct 2015.

- [63] **ATLAS** Collaboration, “ATLAS Inner Detector Alignment Performance with February 2015 Cosmic Rays Data,” Tech. Rep. ATL-PHYS-PUB-2015-009, CERN, Geneva, Apr, 2015. <https://cds.cern.ch/record/2008724>.
- [64] **ATLAS** Collaboration, “Search for metastable heavy charged particles with large ionization energy loss in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS experiment,” *Phys. Rev.* **D93** no. 11, (2016) 112015, [arXiv:1604.04520](https://arxiv.org/abs/1604.04520) [hep-ex].
- [65] M. Capeans *et al.*, “ATLAS Insertable B-Layer Technical Design Report,” Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, Sep, 2010. <https://cds.cern.ch/record/1291633>.
- [66] G. Mullier, “The upgraded Pixel Detector of the ATLAS experiment for Run-2 at the Large Hadron Collider,” *Journal of Instrumentation* **11** no. 02, (2016) C02061. <http://stacks.iop.org/1748-0221/11/i=02/a=C02061>.
- [67] **ATLAS** Collaboration, “ATLAS pixel detector electronics and sensors,” *Journal of Instrumentation* **3** no. 07, (2008) P07007. <http://stacks.iop.org/1748-0221/3/i=07/a=P07007>.
- [68] D. Froidevaux and V. A. Mitsou, “Experimental prospects at the Large Hadron Collider,” *J. Phys. Conf. Ser.* **171** (2009) 012021, [arXiv:0905.0258](https://arxiv.org/abs/0905.0258) [hep-ex].
- [69] C. Lippmann, “Particle identification,” *Nucl. Instrum. Meth.* **A666** (2012) 148–172, [arXiv:1101.3276](https://arxiv.org/abs/1101.3276) [hep-ex].
- [70] **ATLAS** Collaboration, “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data,” *Eur. Phys. J. C* **77** no. CERN-EP-2016-262. 3, (Dec, 2016) 195. 64 p. <https://cds.cern.ch/record/2237544>.
- [71] **ATLAS** Collaboration, “Commissioning of the atlas muon spectrometer with cosmic rays,” *The European Physical Journal C* **70** no. 3, (2010) 875–916. <http://dx.doi.org/10.1140/epjc/s10052-010-1415-2>.
- [72] J. Adelman *et al.*, “Atlas offline data quality monitoring,” *Journal of Physics: Conference Series* **219** no. 4, (2010) 042018. <http://stacks.iop.org/1742-6596/219/i=4/a=042018>.
- [73] S. Ask *et al.*, “The ATLAS central level-1 trigger logic and TTC system,” *Journal of Instrumentation* **3** no. 08, (2008) P08002. <http://stacks.iop.org/1748-0221/3/i=08/a=P08002>.
- [74] “TWiki: Combined Performance Tools for Physics Analysis.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PhysicsAnalysisWorkBookRel20CPRec>.
- [75] **ATLAS** Collaboration, “Luminosity determination in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC,” *Eur. Phys. J.* **C76** no. 12, (2016) 653, [arXiv:1608.03953](https://arxiv.org/abs/1608.03953) [hep-ex].
- [76] G. Cabras, “Luminosity measurements of the ATLAS experiment in LHC Run2 with LUCID,” *Nuovo Cim.* **C40** no. 1, (2017) 17.

- [77] “TWiki: Luminosity Public Results Run2.”  
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [78] **ATLAS** Collaboration, “Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment,” Tech. Rep. ATLAS-CONF-2012-042, CERN, Geneva, Mar, 2012.  
<https://cds.cern.ch/record/1435196>.
- [79] J. Mitrevski, “Preparing ATLAS reconstruction software for LHC’s Run 2,” *Journal of Physics: Conference Series* **664** no. 7, (2015) 072034.  
<http://stacks.iop.org/1742-6596/664/i=7/a=072034>.
- [80] T. G. Cornelissen *et al.*, “The global  $\chi^2$  track fitter in atlas,” *Journal of Physics: Conference Series* **119** no. 3, (2008) 032013.  
<http://stacks.iop.org/1742-6596/119/i=3/a=032013>.
- [81] **ATLAS** Collaboration, “Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung,” Tech. Rep. ATLAS-CONF-2012-047, CERN, Geneva, May, 2012.  
<https://cds.cern.ch/record/1449796>.
- [82] **ATLAS** Collaboration, “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data,” *Eur. Phys. J.* **C77** no. 3, (2017) 195, [arXiv:1612.01456](https://arxiv.org/abs/1612.01456) [hep-ex].
- [83] **ATLAS** Collaboration, “Early Inner Detector Tracking Performance in the 2015 data at  $\sqrt{s} = 13$  TeV,” Tech. Rep. ATL-PHYS-PUB-2015-051, CERN, Geneva, Dec, 2015. <https://cds.cern.ch/record/2110140>.
- [84] **ATLAS** Collaboration, “Performance of the ATLAS Trigger System in 2015,” *Eur. Phys. J.* **C77** no. 5, (2017) 317, [arXiv:1611.09661](https://arxiv.org/abs/1611.09661) [hep-ex].
- [85] **ATLAS** Collaboration, “Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data,”  
<https://cds.cern.ch/record/2157687>. ATLAS-CONF-2016-024.
- [86] **ATLAS** Collaboration, “Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data,” *Eur. Phys. J.* **C74** no. 10, (2014) 3071, [arXiv:1407.5063](https://arxiv.org/abs/1407.5063) [hep-ex].
- [87] **ATLAS** Collaboration, “Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at  $\sqrt{s} = 13$  TeV,” Tech. Rep. ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug, 2016.  
<https://cds.cern.ch/record/2203514>.
- [88] **ATLAS** Collaboration, “Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data,” *Eur. Phys. J.* **C76** no. 12, (2016) 666, [arXiv:1606.01813](https://arxiv.org/abs/1606.01813) [hep-ex].
- [89] **ATLAS** Collaboration, “Expected photon performance in the ATLAS experiment,” Tech. Rep. ATL-PHYS-PUB-2011-007, CERN, Geneva, Apr, 2011.  
<https://cds.cern.ch/record/1345329>.



- [90] **ATLAS** Collaboration, “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV,” *Eur. Phys. J.* **C76** no. 5, (2016) 292, [arXiv:1603.05598](https://arxiv.org/abs/1603.05598) [hep-ex].
- [91] **ATLAS** Collaboration, “Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in  $pp$  collisions at  $\sqrt{s}=8$  TeV,” *Eur. Phys. J.* **C75** no. 7, (2015) 303, [arXiv:1412.7086](https://arxiv.org/abs/1412.7086) [hep-ex].
- [92] W. Lampl *et al.*, “Calorimeter Clustering Algorithms: Description and Performance,” Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008. <https://cds.cern.ch/record/1099735>.
- [93] **ATLAS** Collaboration, “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” [arXiv:1703.09665](https://arxiv.org/abs/1703.09665) [hep-ex].
- [94] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,” *Journal of High Energy Physics* **2008** no. 04, (2008) 063. <http://stacks.iop.org/1126-6708/2008/i=04/a=063>.
- [95] **ATLAS** Collaboration, “Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector,” [arXiv:1705.02211](https://arxiv.org/abs/1705.02211) [hep-ex].
- [96] **ATLAS** Collaboration, “Tagging and suppression of pileup jets with the ATLAS detector,” Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May, 2014. <https://cds.cern.ch/record/1700870>.
- [97] **ATLAS** Collaboration, “Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at  $\sqrt{s} = 13$  TeV,” Tech. Rep. ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul, 2015. <https://cds.cern.ch/record/2037613>.
- [98] S. J. Batista and all, “Determination of jet calibration and energy resolution in proton-proton collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector,” Tech. Rep. ATL-COM-PHYS-2015-1086, CERN, Geneva, Sep, 2015. <https://cds.cern.ch/record/2048678>.
- [99] **ATLAS** Collaboration, “Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run,” Tech. Rep. ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun, 2016. <https://cds.cern.ch/record/2160731>.
- [100] **ATLAS** Collaboration, “Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data,” Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014. <https://cds.cern.ch/record/1706245>.
- [101] J. Maurer, *Mesure des performances de reconstruction des électrons et recherche de Supersymétrie dans les canaux avec deux leptons de même charge dans les données du détecteur ATLAS*. PhD thesis, Marseille, CPPM, Oct, 2013. <https://cds.cern.ch/record/1606806>. Presented 20 Sep 2013.
- [102] O. Ducu, *Search for new physics in events with same sign leptons and missing energy with ATLAS at LHC*. PhD thesis, Aix-Marseille U. and Bucharest U., 2015. <https://cds.cern.ch/record/2060995>. Presented 15 Oct 2015.

- [103] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP* **0411** (2004) 040, [arXiv:hep-ph/0409146 \[hep-ph\]](#).
- [104] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP* **0711** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](#).
- [105] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP* **1006** (2010) 043, [arXiv:1002.2581 \[hep-ph\]](#).
- [106] T. Sjostrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP* **05** (2006) 026, [arXiv:hep-ph/0603175 \[hep-ph\]](#).
- [107] T. Sjostrand, S. Mrenna, and P. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820 \[hep-ph\]](#).
- [108] ATLAS Collaboration, “The ATLAS Simulation Infrastructure,” *Eur. Phys. J.* **C70** (2010) 823–874, [arXiv:1005.4568 \[physics.ins-det\]](#).
- [109] GEANT4 Collaboration, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.* **A506** (2003) 250–303.
- [110] ATLAS Collaboration, “Summary of ATLAS Pythia 8 tunes,” Tech. Rep. ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug, 2012. <http://cds.cern.ch/record/1474107>.
- [111] ATLAS Collaboration, “Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector,” Tech. Rep. ATLAS-CONF-2016-058, CERN, Geneva, Aug, 2016. <http://cds.cern.ch/record/2206153>.
- [112] K. de Vasconcelos, “Etude du couplage du boson de Higgs avec le quark top dans les canaux avec deux leptons de meme signe avec l’experience ATLAS au LHC,” Master thesis.
- [113] “Good Run Lists For Analysis Run2.” <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/GoodRunListsForAnalysisRun2>.
- [114] ATLAS Collaboration, “2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data,” <https://cds.cern.ch/record/2136007>. ATL-DAQ-PUB-2016-001.
- [115] J. Alwall *et al.*, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP* **07** (2014) 079, [arXiv:1405.0301 \[hep-ph\]](#).
- [116] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, “Higgs radiation off top quarks at the Tevatron and the LHC,” *Phys. Rev. Lett.* **87** (2001) 201805, [arXiv:hep-ph/0107081 \[hep-ph\]](#).
- [117] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, “NLO QCD corrections to t anti-t H production in hadron collisions,” *Nucl. Phys. B* **653** (2003) 151, [arXiv:hep-ph/0211352 \[hep-ph\]](#).

- [118] S. Dawson, L. H. Orr, L. Reina, and D. Wackerroth, “Associated top quark Higgs boson production at the LHC,” *Phys. Rev. D* **67** (2003) 071503, [arXiv:hep-ph/0211438](https://arxiv.org/abs/hep-ph/0211438) [hep-ph].
- [119] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackerroth, “Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections,” *Phys. Rev. D* **68** (2003) 034022, [arXiv:hep-ph/0305087](https://arxiv.org/abs/hep-ph/0305087) [hep-ph].
- [120] Y. Zhang, W.-G. Ma, R.-Y. Zhang, C. Chen, and L. Guo, “QCD NLO and EW NLO corrections to  $t\bar{t}H$  production with top quark decays at hadron collider,” *Phys. Lett. B* **738** (2014) 1, [arXiv:1407.1110](https://arxiv.org/abs/1407.1110) [hep-ph].
- [121] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, “Weak corrections to Higgs hadroproduction in association with a top-quark pair,” *JHEP* **09** (2014) 065, [arXiv:1407.0823](https://arxiv.org/abs/1407.0823) [hep-ph].
- [122] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro, “Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons,” *JHEP* **06** (2015) 184, [arXiv:1504.03446](https://arxiv.org/abs/1504.03446) [hep-ph].
- [123] “SM Higgs production cross sections at  $\sqrt{s} = 13$  TeV.” <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV>.
- [124] NNPDF Collaboration, “Parton distributions for the LHC Run II,” *JHEP* **04** (2015) 040, [arXiv:1410.8849](https://arxiv.org/abs/1410.8849) [hep-ph].
- [125] ATLAS Collaboration, “ATLAS Run 1 Pythia8 tunes,” Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov, 2014. <http://cds.cern.ch/record/1966419>.
- [126] S. Frixione, P. Nason, and G. Ridolfi, “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction,” *JHEP* **09** (2007) 126, [arXiv:0707.3088](https://arxiv.org/abs/0707.3088) [hep-ph].
- [127] ATLAS Collaboration, “Measurement of W and Z Boson Production Cross Sections in pp Collisions at  $\sqrt{s} = 13$  TeV in the ATLAS Detector,” Tech. Rep. ATLAS-CONF-2015-039, CERN, Geneva, Aug, 2015. <http://cds.cern.ch/record/2045487>.
- [128] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, “Event generation with SHERPA 1.1,” *JHEP* **02** (2009) 007, [arXiv:0811.4622](https://arxiv.org/abs/0811.4622) [hep-ph].
- [129] H. Lai, M. Guzzi, J. Huston, Z. Li, P. Nadolsky, J. Pumplin, and C. P. Yuan, “New parton distributions for collider physics,” *Phys. Rev.* **D82** (2010) 074024, [arXiv:1007.2241](https://arxiv.org/abs/1007.2241) [hep-ph].
- [130] “NLO single-top channel cross sections.” <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [131] E. Re, “Single-top  $Wt$ -channel production matched with parton showers using the POWHEG method,” *Eur. Phys. J.* **C71** (2011) 1547, [arXiv:1009.2450](https://arxiv.org/abs/1009.2450) [hep-ph].

- [132] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions,” *JHEP* **09** (2009) 111, [arXiv:0907.4076 \[hep-ph\]](#). [Erratum: JHEP02,011(2010)].
- [133] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, “New generation of parton distributions with uncertainties from global QCD analysis,” *JHEP* **07** (2002) 012, [arXiv:hep-ph/0201195 \[hep-ph\]](#).
- [134] P. Nadolsky, H. Lai, Q. Cao, J. Huston, J. Pumplin, D. Stump, W. Tung, and C. P. Yuan, “Implications of CTEQ global analysis for collider observables,” *Phys. Rev.* **D78** (2008) 013004, [arXiv:0802.0007 \[hep-ph\]](#).
- [135] P. Skands, “Tuning Monte Carlo Generators: The Perugia Tunes,” *Phys. Rev.* **D82** (2010) 074018, [arXiv:1005.3457 \[hep-ph\]](#).
- [136] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth. Phys. Res. A* **462** (2001) 152.
- [137] P. Golonka and Z. Was, “PHOTOS Monte Carlo: A Precision tool for QED corrections in  $Z$  and  $W$  decays,” *Eur. Phys. J.* **C45** (2006) 97–107, [arXiv:hep-ph/0506026 \[hep-ph\]](#).
- [138] O. Ducu *et al.*, “Run-2 prospects for top Yukawa coupling measurement in the  $t\bar{t}H$  channel with two same-sign leptons,” Tech. Rep. ATL-COM-PHYS-2015-479, CERN, Geneva, Jun, 2015. <https://cds.cern.ch/record/2021084>.
- [139] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders,” *Comput. Phys. Commun.* **185** (2014) 2930, [arXiv:1112.5675 \[hep-ph\]](#).
- [140] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, “ $t\bar{t}W$  and  $t\bar{t}Z$  Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects,” *JHEP* **11** (2012) 056, [arXiv:1208.2665 \[hep-ph\]](#).
- [141] **ATLAS** Collaboration, “Measurement of the  $W^+W^-$  production cross section in  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 13$  TeV with the ATLAS experiment,” [arXiv:1702.04519 \[hep-ex\]](#).
- [142] **ATLAS** Collaboration, “Measurement of the  $ZZ$  Production Cross Section in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector,” *Phys. Rev. Lett.* **116** no. 10, (2016) 101801, [arXiv:1512.05314 \[hep-ex\]](#).
- [143] **ATLAS** Collaboration, “Measurement of the  $W^\pm Z$  boson pair-production cross section in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector,” *Phys. Lett.* **B762** (2016) 1–22, [arXiv:1606.04017 \[hep-ex\]](#).
- [144] M. Bahr *et al.*, “Herwig++ physics and manual,” *Eur. Phys. J. C* **58** (2008) 639, [arXiv:0803.0883 \[hep-ph\]](#).
- [145] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC,” *Eur. Phys. J.* **C63** (2009) 189–285, [arXiv:0901.0002 \[hep-ph\]](#).
- [146] **ATLAS** Collaboration, “Further ATLAS tunes of PYTHIA6 and Pythia 8,” Tech. Rep. ATL-PHYS-PUB-2011-014, CERN, Geneva, Nov, 2011. <http://cds.cern.ch/record/1400677>.

- [147] **ATLAS** Collaboration, “Search for strongly-produced superpartners in final states with same-sign or three leptons and jets in 2015+2016  $pp$  collisions data at  $\sqrt{s} = 13$  TeV (Supporting note),” Tech. Rep. ATL-PHYS-INT-2016-011, CERN, Geneva, Oct, 2016. <https://cds.cern.ch/record/2224598>.
- [148] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J.* **C71** (2011) 1554, [arXiv:1007.1727](https://arxiv.org/abs/1007.1727) [[physics.data-an](https://arxiv.org/archive/physics)]. [Erratum: *Eur. Phys. J.*C73,2501(2013)].
- [149] A. L. Read, “Presentation of search results: The CL(s) technique,” *J. Phys.* **G28** (2002) 2693–2704.
- [150] **ATLAS** Collaboration, “Improved luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC,” *Eur. Phys. J.* **C73** no. 8, (2013) 2518, [arXiv:1302.4393](https://arxiv.org/abs/1302.4393) [[hep-ex](https://arxiv.org/archive/hep)].
- [151] **ATLAS** Collaboration, “Luminosity determination in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC,” *Eur. Phys. J.* **C76** no. 12, (2016) 653, [arXiv:1608.03953](https://arxiv.org/abs/1608.03953) [[hep-ex](https://arxiv.org/archive/hep)].
- [152] **ATLAS** Collaboration, “Combination of the searches for Higgs boson production in association with top quarks in the  $\gamma\gamma$ , multilepton, and  $b\bar{b}$  decay channels at  $\sqrt{s}=13$  TeV with the ATLAS Detector,” Tech. Rep. ATLAS-CONF-2016-068, CERN, Geneva, Aug, 2016. <http://cds.cern.ch/record/2206211>.



# Acknowledgement

First and foremost, I would like to thank God for his never-ending grace, mercy and provision. For giving back my health and life. To give me another chance and inspire me with ambitious to look for the truth by science. For bringing all those nice and kind people who contribute in my academic or personal life.

My greatest debt is to my supervisors and advisors Pascal Pralavorio and Fabrice Hubaut for their support and excellent supervision to achieve this moment. To Pascal, whose vision steered this project from day one, for his valuable and constructive suggestions during the planning and development of this research work. For his patient guidance and useful critiques and advices to keep my progress on schedule. His willingness to give his time so generously has been very much appreciated. To Fabrice, whom I am humbled by his breadth of expertise and his depth knowledge as well as his elegance and kindness, for his care to get the most possible reliable and accurate results. For his thorough and detailed checks and efficient debugging. To his clever diagnosis and intelligence that encouraged me to enjoyably probe and hunt every issue we face. In few words, he just added a flavor to my academic research.

I would like to thank my committee members, Dr. Djamel Boumediene and Dr. Lydia Farayd , Dr. Daniel Bloch and Dr. Cristinel Diaconu for their useful comments and for their time to read and give professional feedback to make this document in high quality and good shape form.

I would like to thank my colleagues for their wonderful collaboration in both CPPM and ATLAS groups. You supported me greatly and were always willing to help me. In particular, my thanks are due to Julien Maurer, an expert and the 1st author of the electron reconstruction efficiency, (during 2010-2011) for his prompt and efficient help to answer all my questions in details and solve the most challenged technical problems during my first year of PHD. I am grateful to Otilia Oducu as well, who added some improvements to the reconstruction efficiency measurements (during 2012), for transferring her expertise to me in the subject and for being such kind pure person and a good friend. I would not forget my friend and colleague Kevin devasconcelos, for translating the French version of the abstract and for his technical support to automatize my code that estimate fake lepton background using the matrix method. Thanks to his work, I was able to get the results in a short time with much less effort which led CPPM 2ssl group to achieve the results firstly (among four groups) in so high challenging environment in ATLAS ttH Multilepton group. I am also grateful for the fruitful discussions with my colleague Venugopal Ellajosyula which always reminds me that I am not just a programmer but also a physicist...

This project would not be possible without the support of Aix-Marseille University scholarship. Thanks also to CERN and ATLAS committee to accept being a member of its collaboration.

My deep gratitude is due to my father for his continuously financial and emotional support to join the particle physics center at UK in the prestigious university RHUL which

opens the doors for me to continue the research in the same field. I am simultaneously grateful to my mother for her wisdom and standing by me through thick and thin. Special thanks to my sister Imane to be really a support to me during the three years of my PHD. And in particular for her contribution to translate the French version of “Extrait de la these” with the help of her husband Taki in addition to my supervisor Pascal. Not only Imane but I am extremely thankful to my brother Mourad, my sisters Djohra and Radhia and all my family for their warmth and unwavering support, encouragement and solidarity at each and every step of my academic and personal life.

Last but not least, my thanks go to all members of Physics Department for their help and kindness during the three years of my PHD in France. Their warm welcome was one of the main reasons that boost me to take the decision to join CPPM group for my PHD.



# Abstract

The Large Hadron Collider (LHC) at CERN restarted in spring 2015 for three years (Run2) at an unexplored center-of-mass energy of 13 TeV; An ideal place to search for physics beyond the Standard Model such as supersymmetry (SUSY) and extra dimensions. A precise measurement of electron reconstruction efficiency in ATLAS, one of the two general purpose experiments of the LHC, is presented in the first part of this thesis with  $Z \rightarrow ee$  data sample using  $3.2 \text{ fb}^{-1}$  of data recorded in 2015. This allows to extract scale factors between data and simulation that are used by all ATLAS physics analyses involving electrons. The results show the high ability of the ATLAS detector to reconstruct electrons from one hand and the good understanding of its performance on the other hand. The second part of the thesis is dedicated to a search for the Higgs boson production in association with a top quark pair ( $t\bar{t}H$ ), which could allow a first direct measurement of the top quark Yukawa coupling and could reveal new physics. The signature with two same-charge light leptons (electron or muon) without a hadronically decaying tau lepton final state, targeting the decays  $H \rightarrow WW^*$ , is examined using the first 10% of the total expected Run2 dataset. Events with fake (non-prompt) leptons represent the main reducible background of this signature. The estimation of this background largely drives the signal sensitivity. An improved method to estimate it has been developed and is discussed in details in this thesis. Driving the total error, fake leptons background is found to be 1.5 to 3.6 times higher than in simulation and represent between 32 and 48% of the total background. The best-fit value of the ratio of observed and Standard Model cross sections of  $t\bar{t}H$  production process, combining with other multilepton channels, is  $2.5 \pm 0.7$  (stat)  $^{+1.1}_{-0.9}$  (syst), and an upper limit on this ratio of 4.9 (2.3 expected) is found at 95% confidence level.

# Résumé

Le Grand Collisionneur de Hadrons (*Large Hadron Collider, LHC*) du CERN a redémarré au printemps 2015 pour trois ans (Run 2) avec une énergie dans le centre de masse de 13 TeV, idéale pour la recherche de physique au delà du Modèle Standard, comme la Supersymétrie ou des dimensions supplémentaires. Une mesure précise de l'efficacité de reconstruction des électrons avec le détecteur ATLAS, une des deux expériences généralistes présentes au LHC, est présentée dans la première partie de cette thèse en utilisant les  $3.2 \text{ fb}^{-1}$  de données récoltées au cours de l'année 2015 en étudiant le canal de désintégration  $Z \rightarrow ee$ . Cela a permis d'extraire les rapports d'efficacité de reconstruction des électrons obtenus avec les données et la simulation, lesquels sont utilisés par toutes les analyses d'ATLAS impliquant des électrons. Les résultats montrent d'une part une grande efficacité de reconstruction des électrons avec le détecteur ATLAS, et une bonne compréhension de ses performances d'autre part. La seconde partie de ce manuscrit est dédiée à la recherche de la production associée du boson de Higgs avec une paire de quarks top ( $t\bar{t}H$ ), qui pourrait permettre une première mesure directe du couplage de Yukawa entre le boson de Higgs et le quark top. Une déviation dans la mesure par rapport aux prédictions du Modèle Standard serait une preuve manifeste de Nouvelle Physique. La signature de l'état final avec deux leptons de même charge électrique (électrons ou muons) est examinée en utilisant les premiers 10% du total de données attendues pour le Run 2, portant l'attention sur la désintégration  $H \rightarrow WW^*$ . Les événements avec de faux (non-prompt) leptons représentent le bruit de fond dominant, mais réductible, de l'état final considéré. L'estimation de ce bruit de fond conditionne largement la sensibilité de l'analyse. Une méthode améliorant cette estimation a été développée et est discutée en détails dans ce rapport. Dominant l'erreur totale de la mesure, ce bruit de fond instrumental est mesuré de 1.5 à 3.6 fois supérieur à la prédiction des simulations et représente entre 32 et 48% du bruit de fond total. Le meilleur ajustement du rapport entre nombre d'événements observés et prédits par les valeurs de section efficace du Modèle Standard pour la production  $t\bar{t}H$ , mesuré conjointement avec les autres canaux multi-leptoniques, est de  $2.5 \pm 0.7$  (stat)  $^{+1.1}_{-0.9}$  (syst), avec une limite supérieure observée de 4.9 (2.3 attendue) à 95% de niveau de confiance.