Recherche de la production WWW et du Higgs doublement chargé avec le détecteur ATLAS Search for the W ± W ± W production and the doubly charged Higgs with the ATLAS Detector

Ruiqi Zhang

CPPM-T-2017-05

UNIVERSITE D'AIX-MARSEILLE

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

ECOLE DOCTORALE 352

Department of Physics

Centre de Physique des Particules de Marseille

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Physique et Science de la Matière

Spécialité : Physique des Particules et Astroparticules

Ruiqi ZHANG

Recherche de la production WWW et du Higgs doublement chargé avec le détecteur ATLAS

Search for the $W^{\pm} W^{\pm} W^{\mp}$ production and the doubly charged Higgs with the ATLAS Detector

Soutenue le 1/12/2017 devant le jury :

Shenjian CHEN	NJU	Examinateur
Cristinel DIACONU	CPPM	Directeur de thèse
Suzanne GASCON	IPNL	Examinateur
Yanwen LIU	USTC	Directeur de thèse
Emmanuel MONNIER	CPPM	Directeur de thèse
Yajin MAO	PKU	Rapporteur
Emmanuel SAUVAN	LAPP	Rapporteur
Zhengguo ZHAO	USTC	Examinateur

Numéro national de thèse/suffixe local : 2017AIXM0434/049ED352

CPPM-T-2017-05

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

AIX MARSEILLE UNIVERSITY

DOCTORALE SCHOOL 352

Department of Physics

Centre de Physique des Particules de Marseille

A disertation to obtain the doctoral degree

Domain : Physics and Matter Science

Specialty : Particle Physics and Astroparticles

Ruiqi ZHANG

Search for the $W^{\pm}$ $W^{\pm}$ $W^{\mp}$ production and the doubly charged Higgs with the ATLAS Detector

Recherche de la production WWW et du Higgs doublement chargé avec le détecteur ATLAS

Sustained on 1/12/2017 in front of a jury composed of:

Shenjian CHEN	NJU	Examiner
Cristinel DIACONU	CPPM	Supervisor
Suzanne GASCON	IPNL	Examiner
Yanwen LIU	USTC	Supervisor
Emmanuel MONNIER	CPPM	Supervisor
Yajin MAO	PKU	Reviewer
Emmanuel SAUVAN	LAPP	Reviewer
Zhengguo ZHAO	USTC	Examiner

National thesis number/local suffix : 2017AIXM0434/049ED352

To my family, and those who once helped me.

Acknowledgements

It's a great opportunity to work in the ATLAS experiment which is the frontier of the high energy physics, the experience is unforgettable and I would like to express my gratitude to all those who have made this thesis possible.

I was guided into the field of high energy physics and the ATLAS experiment by Prof. Yanwen Liu when I was an undergraduate student in 2011, it has been six years from that time to the present. I was taught hand by hand from the scratch to obtain necessary ingredients in our field, coding, statistic and physics. You are ready to help me whenever and whatever difficulties I came cross not only in work but in life as well. I am confident that I could always rely on your experience and support and you taught me not only the academic skills but also ethical and social values. Your guidance and all kinds of support ensure me to complete my Ph.D. I am deeply grateful to Prof. Yanwen Liu for your instructions, support and encouragements throughout these six years and your patience for my indolence and ignorance. I became a joint Ph.D candidate between USTC and CPPM later in 2014 when I met another two supervisors: Prof. Cristi Diaconu and Prof. Emmanuel Monnier. Your instructions and supports are essential to my Ph.D research. During my stay at CPPM, Prof. Emmanuel Monnier took care of me on every aspect in work and life despite the fact that you are a very busy man, I would like to express my sincere thanks to you. Prof. Cristi Diaconu and I worked in the analysis of searching for doubly charged Higgs together, I benefit a lot from your expertise and insight on physics. I am grateful to your guidance and support.

I am also grateful to Alex Long and Louis Helary who are my colleagues in the analysis of search for $W^{\pm}W^{\pm}W^{\mp}$. This is my first physics analysis, I learned a lot from you not only on technical skills but also how to work in collaboration, thanks a lot. I wish to acknowledge the help provided by USTC and CPPM in various aspects, the technical support and great atmosphere among colleagues. My thanks also goes to the senior Ph.Ds, Hulin Wang, Shu Li and Jun Gao. They are more experienced than me in the ATLAS experiment and their help is more than valuable when I am trapped by technical issues. I benefit a lot from their expertise and I am deeply grateful for that.

Last but not least, I would like to express my deep gratitude to my family for their love and support.

Abstract

The ATLAS (A Toroidal LHC Apparatus) experiment, worldwide collaboration of more than 3000 scientists from 175 institutions in 38 countries, is conducting researches at the Large Hadron Collider (LHC), world's largest hadron collider. Designed to collide proton beams at a 14 TeV center of mass energy and a $10^{34}\ cm^{-2}s^{-1}$ peak luminosity, the LHC ran at 7 TeV in 2011 and 8 TeV in 2012, this running period is referred to as Run 1. Shut down during 2013 and 2014 for an upgrade, the LHC was restarted in 2015 at 13 TeV. The general purpose ATLAS detector, provides a rich physics potential for precise measurements of the Standard Model (SM) and search for new physics phenomena. The ATLAS experiment is a. Two physics analyses, to which I have strongly contributed, are presented in this thesis document. The first one is the search for $W^{\pm}W^{\pm}W^{\mp}$ and the study of anomalous quartic gauge couplings (aQGC). This is a search for the tri-boson $W^{\pm}W^{\pm}W^{\mp}$ production decaying in full leptonic channel and semi leptonic channel, my contribution is in the full leptonic channel. This analysis utilizes Run 1 data collected at 8 TeV and 20.3 fb^{-1} integrated luminosity. In the full leptonic channel, backgrounds can come from WZ and ZZ processes with three real leptons, this background is estimated with the Monte Carlo simulation. Events with mis-reconstructed leptons or charge mis-identified leptons are important background contributions as well, referred as reducible background and estimated with data-driven methods. The measurement of aQGC provides a sensitive probe for new physics beyond the Standard Model (SM) at high energy scale. The $WWWW$ vertex is used to conduct the aQGC study. The number of events observed is consistent with the SM prediction. The observed upper limit at 95% CL on the SM $W^{\pm}W^{\pm}W^{\mp}$ cross section is 730 fb with an expected limit of 560 fb in the absence of the $W^{\pm}W^{\pm}W^{\mp}$ production. Since no significant deviation from the SM is observed, limits on anomalous quartic gauge couplings are also derived. The second physics analysis I joined, is a search for doubly charged Higgs performed on Run 2 data collected in 2015 and 2016 at 13 TeV with a $36.1 fb^{-1}$ integrated luminosity. The doubly charged Higgs is predicted by a model that extends the SM to allow masses for neutrinos. In this model, also called Higgs Doublet Triplet Model, a triplet scalar is introduced in the Higgs sector and the electroweak symmetry breaking introduces several Higgs bosons, one of them with a doubly charged, the $H^{\pm\pm}$. Various constraints are applied to simplify the scenario and only the pair-produced $H^{\pm\pm}$ mode where all $H^{\pm\pm}$ decays to W bosons is considered in the analysis for a mass range from

200 to 700 GeV. There are three channels, named according to the different final states, in this analysis: $2\ell^{SS}$, 3ℓ and 4ℓ. I worked in the $2\ell^{SS}$ channel where background is more complex to deal with. The backgrounds can come from events with two same sign leptons such as the $W^{\pm}W^{\pm}$ process, this background is estimated with MC simulation. Events with charge mis-identified leptons or with mis-reconstructed leptons are important background contributions which are estimated with data-driven methods and their correlations are properly dealt with. In this analysis, background is found to be consistent with the data and no significant excess observed. Therefore, upper limits are derived and the model is excluded at 95% CL for $M_{H^{\pm\pm}}$ < 220 GeV.

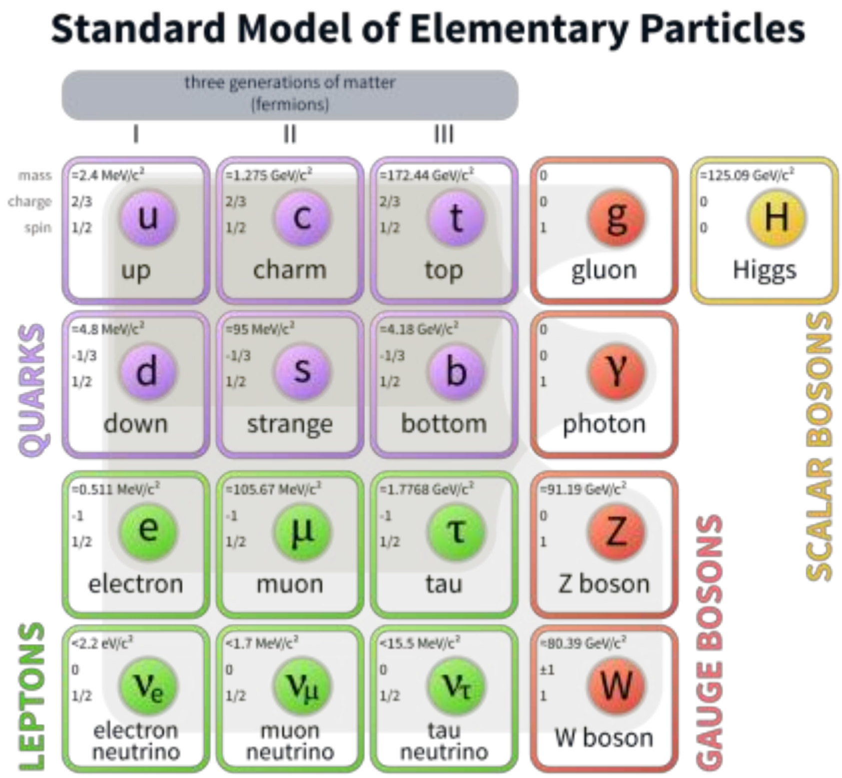
Abstract

L'expérience ATLAS (A Toroidal LHC Apparatus), collaboration mondiale de plus de 3000 scientifique provenant de 175 instituts et 38 des pays, effectue des recherches auprès du grand collisionneur de hadron (LHC), plus grand collisionneur de Hadron au monde. Conçu pour produire des collisions de faisceaux de protons à une énergie dans le centre de masse de 14 TeV et une luminosité pic de $10^{34}\ cm^{-2}s^{-1}$, le LHC a tourné à 7 TeV en 2011 et 8 TeV en 2012, cette période étant appelée Run1. Arrêté entre 2013 et 2014 pour une mise à niveau, le LHC a redémarré en 2015 à 13 TeV. Le détecteur généraliste ATLAS a un riche potentiel de mesures précises du Modèle Standard (SM) et de recherche de phénomènes de nouvel physique. Deux analyses de physiques, auxquelles j'ai beaucoup contribuées, sont présentées dans ce manuscrit de thèse. La première consiste à rechercher des $W^{\pm}W^{\pm}W^{\mp}$ et étudier le couplage de jauge quartique anormal (aQGC). C'est une recherche de la production de tri-bosons, $W^{\pm}W^{\pm}W^{\mp}$, se désintégrant totalement ou partiellement en leptons. Mon étude a été faite sur le canal contenant uniquement des leptons. Cette analyse s'appuie sur les données du Run 1 collectées à 8 TeV et 20.3 fb^{-1} de luminosité intégrée. Dans le canal tout lepton, le bruit de fond peut provenir de processus WZ ou ZZ avec trois vrais leptons et il a été estimé avec une simulation Monte Carlo. Les événements contenant des électrons improprement reconstruits ou de charge improprement identifiée ont aussi des contributions importantes aux bruits de fond, appelées bruit réductibles, et sont estimées par des méthodes s'appuyant sur les données. La mesure d'aQGC fournis une sonde sensible à de la nouvelle physique au-delà du Modèle Standard (SM) à une échelle de haute énergie. Le vertex $WWWW$ est utilisé pour effectuer ces études d'aQGC. Le nombre d' événements observés est en accord avec les prédictions du SM. La limite supérieure observée à 95% CL sur la section efficace $W^{\pm}W^{\pm}W^{\mp}$ du SM est de 730 fb avec une limite attendue de 560 fb en l'absence de production $W^{\pm}W^{\pm}W^{\mp}$. Comme aucune déviation au SM n'a été observée, les limites sur le couplage de jauge quartique anormal ont aussi été extraites. La deuxième analyse de physique à laquelle j'ai participée est la recherche du boson de Higgs doublement chargé effectuée sur les données du Run 2 collectées en 2015 et 2016 à 13 TeV avec 36.1 fb^{-1} de luminosité intégrée. Le Higgs doublement chargé est prédit par un modèle qui prolonge le SM pour permettre des neutrinos massif. Dans ce modèle, aussi appelé modèle Higgs Doublet Triplet, un triplet de scalaire est introduit dans le secteur du Higgs et la brisure de la symétrie électrofaible introduit

plusieurs bosons de Higgs, l'un d'eux étant doublement chargé, le $H^{\pm\pm}$. Plusieurs contraintes ont été appliquées pour simplifier le scenario et seul le mode de désintégration en paire $H^{\pm\pm}$ où tous les Higgs se désintègrent en bosons W a été considéré dans cette analyse pour une gamme de masses allant de 200 GeV à 700 GeV. Il y a trois canaux, nommés par rapport à leurs différents états finaux, dans cette analyse : $2\ell^{SS}$, 3ℓ and 4ℓ. J'ai travaillé sur le canal $2\ell^{SS}$ où le bruit est plus complexe à estimer et traiter. Les bruits peuvent provenir d'événements avec deux leptons de même charge comme les processus $W^{\pm}W^{\pm}$, ce bruit de fond étant estimé par simulation MC. Les événements contenant des électrons improprement reconstruits ou de charges improprement identifiées ont aussi des contributions importantes aux bruits de fond et sont estimées par des méthodes orientées données et où les corrélations sont proprement prises en compte. Dans cette analyse, le bruit de fond estimé est en accord avec les données et aucun excès significatif n'est observé. Des limites supérieures sont donc déduites et le modèle considéré est exclus à 95% CL pour $M_{H^{\pm\pm}} < 220$ GeV.

Résumé de la thèse

Le modèle standard (SM) de la physique des particules, décrit schématiquement ci-dessous, est la théorie décrivant les particules élémentaires et leurs interactions – électromagnétique, faible et forte – basées sur le cadre mathématique de la théorie quantique des champs et des principes de jauge. La théorie fournit une description unifiée des interactions électromagnétiques, faibles et fortes basées sur le groupe SU (3) × SU (2) × U (1) dans lequel le groupe SU (3) décrit l'interaction forte et SU (2) × U (1) décrit les interactions électrofaibles. La symétrie SU (2) × U (1) est rompue spontanément via le mécanisme de Higgs pour générer des masses pour les bosons de jauge faibles et les fermions tandis qu'un boson de Higgs est prédit. Sur la base d'observations et de mesures effectuées depuis des décennies par de nombreuses expériences, le SM est considéré comme le modèle le plus réussi jamais construit à ce jour, mais il est imparfait.



Certaines de ses prédictions sont en contraste avec les observations expérimentales alors que certains phénomènes dans l'univers n'y sont mêmes pas inclus comme :

- Masses des neutrinos: les neutrinos sont censés être sans masse dans le SM tandis que les expériences d'oscillation de neutrinos ont montré qu'ils ont une petite masse.

- Asymétrie matière-antimatière: la dominance de la matière observée dans l'univers ne peut pas être naturellement expliquée par la théorie du Big Bang avec le SM.

- Gravité: Trois des quatre forces fondamentales sont décrites dans le SM mais pas la gravité.

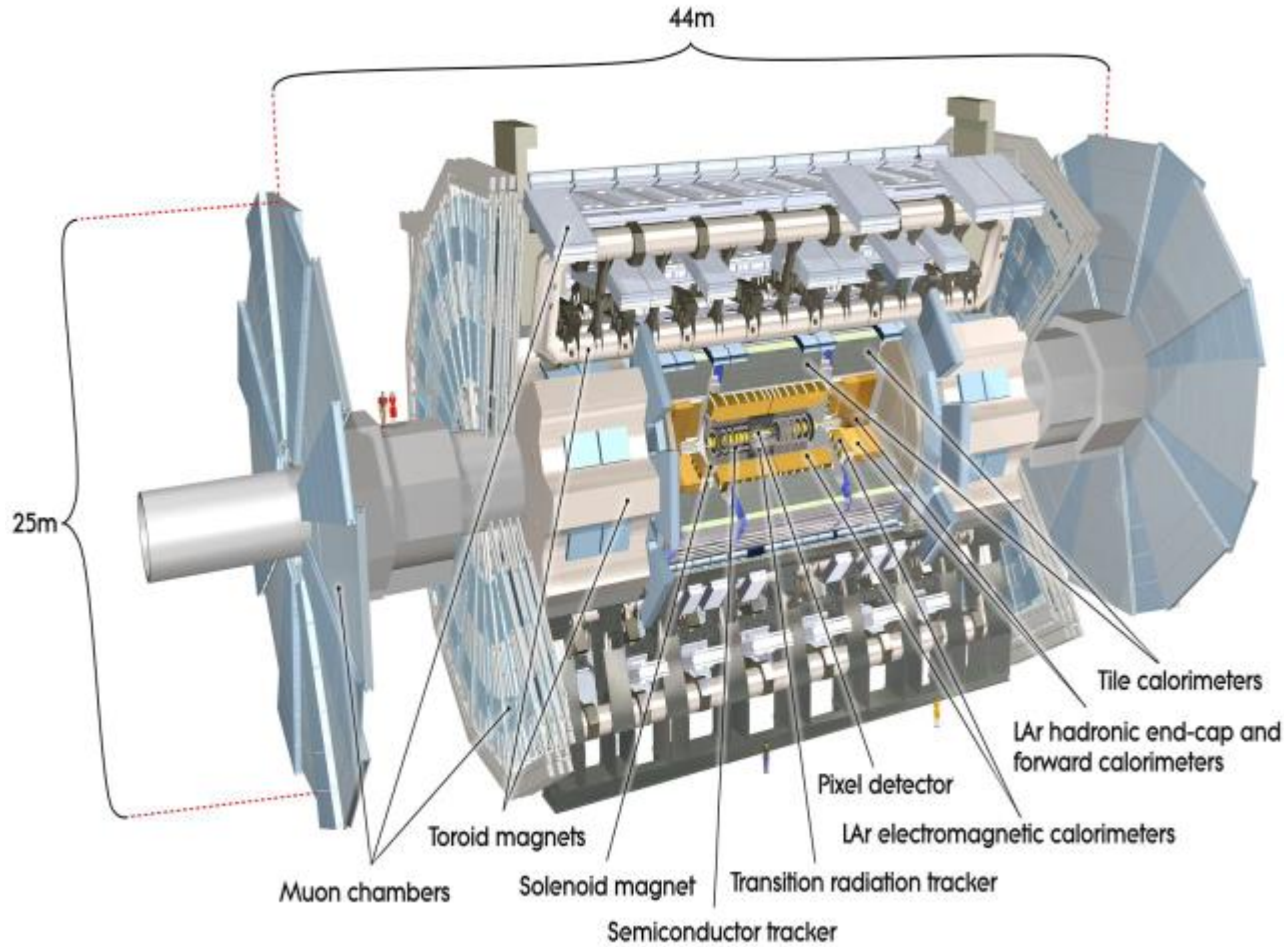
- La matière noire et l'énergie sombre: le SM ne fournit aucune explication pour la matière et l'énergie noire.

Différents modèles ont été développés pour expliquer ces phénomènes mais malheureusement aucun d'entre eux n'a été vérifié par les expériences. Les physiciens sont donc toujours à la recherche d'une nouvelle physique au-delà du SM.

Il y a deux approches complémentaires pour la recherche d'une nouvelle physique au-delà du SM. La première approche consiste à rechercher des couplages anormaux, ce qui n'introduit pas explicitement de nouvelles particules. Les nouvelles particules sont supposées trop lourdes pour être directement observées au LHC, mais elles modifient indirectement les interactions des particules du SM. Dans la deuxième approche, un modèle complet est spécifié, et les nouvelles particules sont directement recherchées.

Deux analyses de physiques, auxquelles j'ai beaucoup contribuées, sont présentées dans ce manuscrit de thèse. La première consiste à rechercher des événements contenant trois $W^{\pm}W^{\pm}W^{\mp}$ et étudier le couplage de jauge quartique anormal (aQGC). Cette analyse appartient à la première classe d'approche, tandis que la recherche du double Higgs chargé est un exemple de la deuxième catégorie.

Le grand collisionneur de hadrons (LHC) est l'accélérateur de particules le plus grand et le plus puissant au monde. Il est situé dans un tunnel de 27km de circonférence entre 45 et 170 m sous la frontière franco-suisse à Genève. Il est hébergé par l'Organisation européenne pour la recherche nucléaire (CERN). Conçu pour produire des collisions de faisceaux de protons accélérés jusqu'à une énergie dans le centre de masse de 14 TeV et une luminosité pic de 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, le LHC a tourné et produit des collisions à 7 TeV en 2011 et 8 TeV en 2012, cette période étant appelée « Run 1 ». Arrêté entre 2013 et 2014 pour une mise à niveau, le LHC a ensuite redémarré en 2015 à 13 TeV pour une nouvelle campagne appelée « Run 2 ». L'expérience ATLAS (A Toroidal LHC Apparatus) est une collaboration mondiale de plus de 3000 scientifiques provenant de 175 instituts et 38 pays. Elle effectue des recherches auprès de cet accélérateur en opérant, sur l'un de ses points d'interaction, un détecteur éponyme présentant un très riche potentiel de physique, allant de la mesure précise du modèle standard à la recherche de nouveaux phénomènes. La figure ci-dessous montre une vue en coupe du détecteur ATLAS.



Avec ce détecteur ATLAS, les études de nombreux processus de physique intéressants utilisent les informations provenant de divers objets de physique, à savoir les électrons, les muons, les jets et l'énergie transversale manquante. Ces objets physiques sont reconstruits sur la base des informations provenant de divers détecteurs et déclencheurs. Le tableau ci-dessous (page suivante) résume les variables utilisées pour identifier les électrons.

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

Deux analyses sont incluses dans la thèse, l'une est la recherche de production WWW tandis que l'autre est la recherche du scalaire de Higgs doublement chargé. La première analyse utilise les données collectées pendant le Run 1 par le détecteur ATLAS avec une énergie dans le centre de masse de 8 TeV et une luminosité intégrée de 20.3 fb^{-1}, la seconde analyse utilise les données collectées pendant le Run 2 avec une énergie dans le centre de masse de 13 TeV et une luminosité de 36.1 fb^{-1}. Outre les données, divers lots de données de simulation Monte Carlo (MC) sont utilisés pour les analyses. Ces échantillons MC sont produits avec plusieurs générateurs différents, puis la numérisation et la simulation des réponses du détecteur sont effectuées avec le logiciel Geant4. Les événements MC étant alors dans le même format que les données, ils sont ensuite reconstruits en utilisant le même logiciel hors ligne.

La production de tri-bosons, $W^{\pm}W^{\pm}W^{\mp}$, est un processus électrofaible rare autorisé par le modèle standard et sa section efficace à l’ordre « Next-to-Leading-Order (NLO) » est déjà connue.

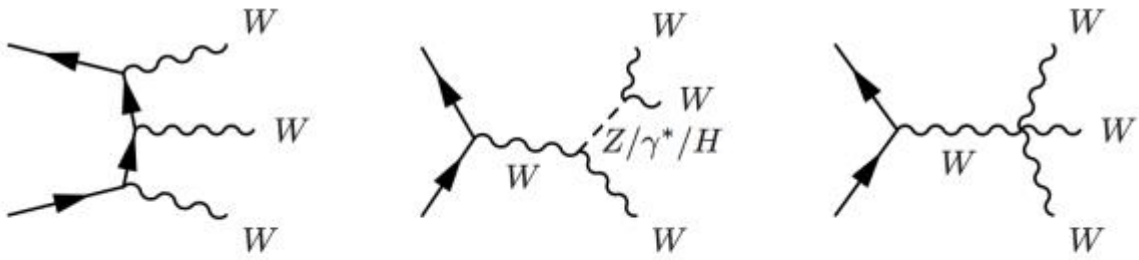


Figure 1.2 The Feynman diagrams for the $W^{\pm}W^{\pm}W^{\mp}$ production.

L'observation d'un écart significatif par rapport à la prédiction impliquera une nouvelle physique au-delà du modèle standard. La mesure de la production WWW peut être utilisée pour sonder les couplages de jauge, et en particulier, le processus est sensible aux couplages de jauge quartic (aQGC). Une liste d'opérateurs de dimension 8 qui paramètrent les effets d'une nouvelle physique à une échelle d'énergie hors de portée du LHC est utilisée pour étudier les couplages de jauges anomales.

Dans l'analyse WWW, deux canaux de désintégration sont considérés, l'un est le canal complètement leptonique tandis que l'autre est le canal semi-leptonique. Ma contribution est dans le canal complètement leptonique.

Dans cette analyse, le bruit de fond peut être dû à de faux leptons issus de désintégrations hadroniques ou à une mauvaise identification des charges de lepton. Cette partie des bruits de fond est estimée avec des techniques basées sur les données. Les autres bruits de fonds irréductibles sont estimés avec une simulation de Monte Carlo.

La figure ci-dessous montre la probabilité mesurée d'un électron avec une charge mal identifiée.

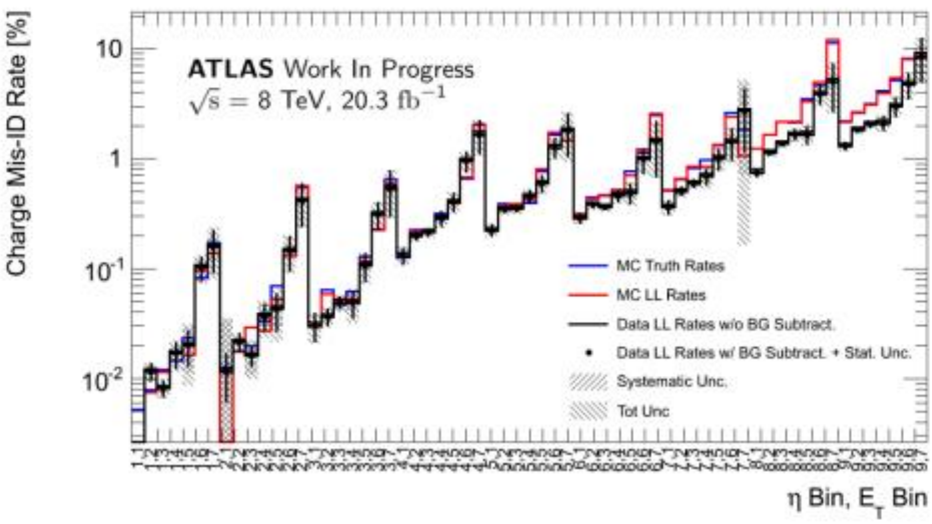


Figure 5.11 Electron charge misID rates obtained from data with the likelihood method. All the errors are now shown. The x axis label is the $|\eta|$, p_T bin index.

La figure ci-dessous montre la comparaison entre les données et le fond estimé, le bon accord indique que le fond est bien contrôlé.

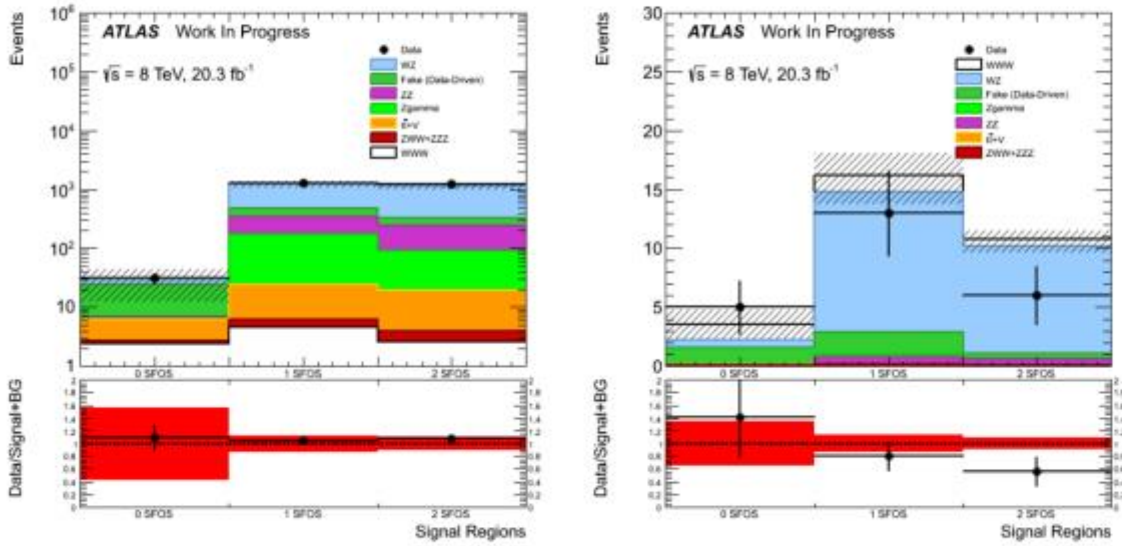


Figure 5.17 Comparison between the data and the background estimation in pre-selection region (left) and signal region (right).

Diverses incertitudes systématiques sont prises en compte dans cette analyse, y compris des incertitudes théoriques et expérimentales. La méthode du rapport de vraisemblance profilé est utilisée dans cette analyse pour l'interprétation statistique. Dans le canal tout leptonique, la section transversale estimée est de 309,2 ab alors que le nombre observé est :

$$\sigma^{\text{Observed}} = 313.5^{+348}_{-332}(\text{stat})^{+322}_{-346}(\text{sys})\text{ab}$$

En combinant les résultats de l'analyse tout-leptonique et de l'analyse semi-leptonique, la limite supérieure observée à 95% CL sur la section efficace SM de production WWW est de 730 fb avec une limite attendue de 560 fb en l'absence de production WWW. En plus de la mesure de la section efficace SM, l'étude du couplage de jauges quartiques anormales est également réalisée. Des limites sont définies pour les opérateurs $f_{S,0} / \Lambda^4$ et $f_{S,1} / \Lambda^4$ de dimension 8 de la théorie des champs effectifs. La figure ci-dessous montre les limites combinées.

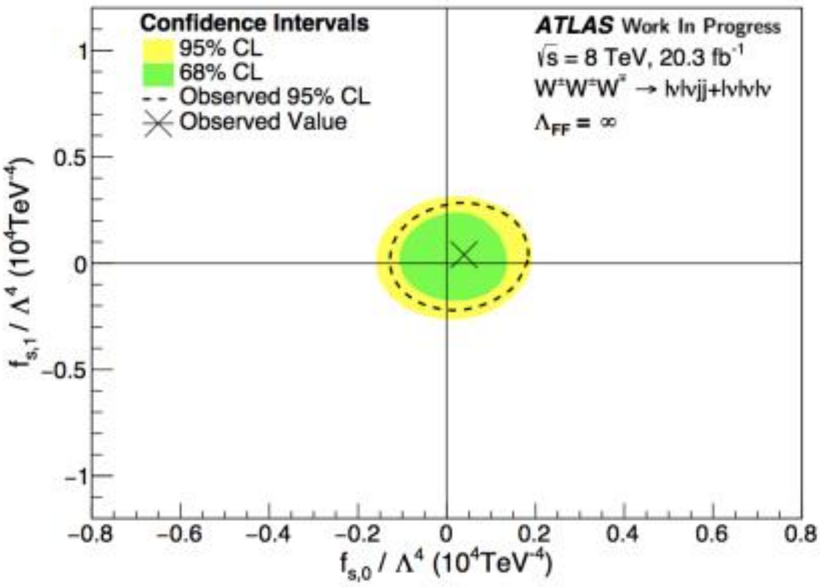


Figure 5.23 aQGC limits without form factor combining the full leptonic analysis and the semi-leptonic analysis.

Les résultats de la recherche sur la production WWW ont été publiés dans le journal de physique européen C, (EPJC).

La recherche du boson scalaire de Higgs doublement chargé utilise un modèle au-delà du SM qui est également appelé HDTM (Higgs Doublet Triplet Model). Les neutrinos sont sans masse dans le SM, mais dans la nature, la masse du neutrino a été trouvée faible mais pas nulle par diverses expériences. Pour permettre aux neutrinos d’avoir une masse, un triplet scalaire avec hypercharge Y = 2 est ajouté au secteur scalaire de Higgs. La rupture de la symétrie électrofaible produira alors cinq bosons Higgs dont l'un d'entre eux sera doublement chargé. Diverses contraintes sont appliquées à l'espace des paramètres. Dans cette analyse, le mode de production par paire est choisi et les Higgs doublement chargés se désintègrent en bosons W.

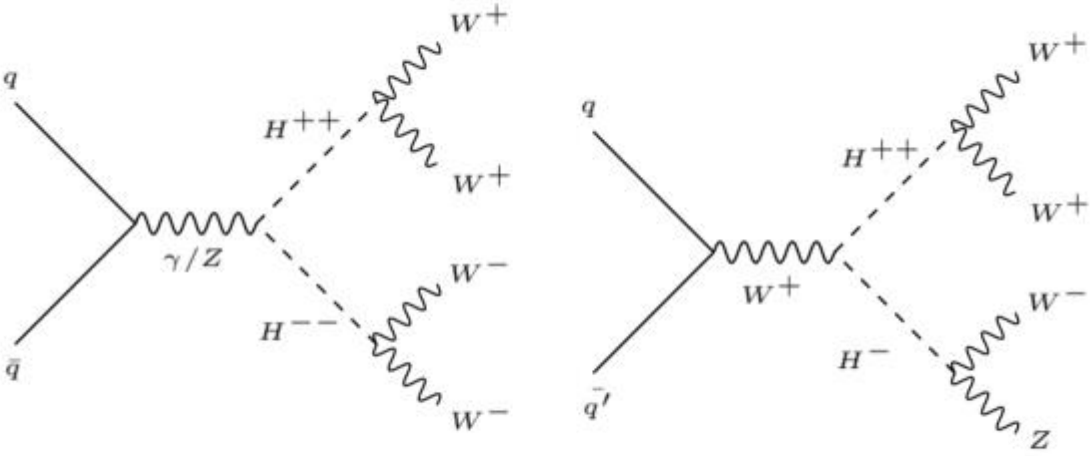


Figure 1.4 Feynman diagrams of the pair production mode (left) and the associated production mode (right).

Dans cette analyse, les données dites du "Run 2" avec une énergie dans le centre de masse de 13 TeV et une luminosité de 36,1 fb^{-1} sont utilisées et trois canaux de désintégration sont considérés. L'analyse que j'ai effectuée et présentée ici est sur le canal avec deux leptons de même signe dans l'état final.

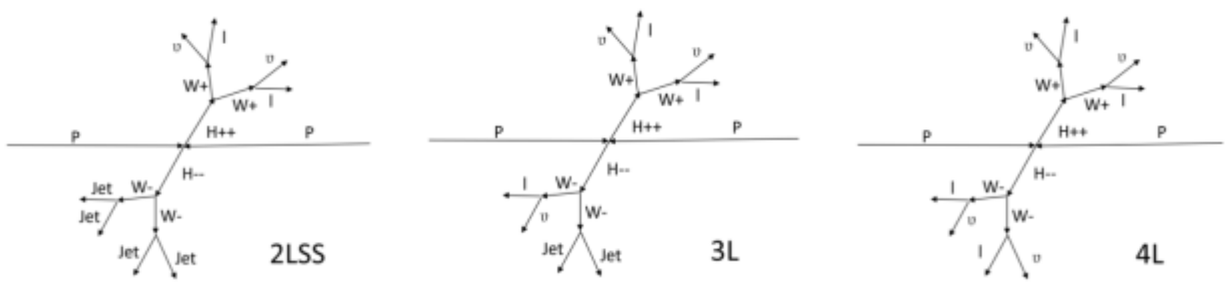


Figure 6.1 Illustrations of event topologies of the signal process for the three channels, $2\ell^{SS}$, 3ℓ and 4ℓ from the left to the right.

D'une façon similaire à la recherche de la production WWW, le bruit de fond de cette analyse peut être dû à des événements avec des leptons prompts, des électrons de charge mal identifiée ou de faux leptons. Le bruit de fond dû à la mauvaise identification des charges d'électrons et à de faux leptons est estimé à l'aide de techniques s'appuyant sur les données.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 90$	$90 < p_T/\text{GeV} < 130$	$130 < p_T/\text{GeV} < 1000$
$0< \|\eta\| <0.6$	0.021±0.001	0.065±0.008	0.150±0.028	0.324±0.068
$0.6< \|\eta\| <1.1$	0.063±0.002	0.142±0.013	0.307±0.046	0.768±0.100
$1.1< \|\eta\| <1.37$	0.147±0.005	0.348±0.030	0.703±0.102	1.359±0.224
$1.52< \|\eta\| <1.7$	0.422±0.011	0.898±0.067	1.779±0.222	3.450±0.494
$1.7< \|\eta\| <2.3$	0.837±0.008	1.972±0.057	3.246±0.178	5.830±0.376
$2.3< \|\eta\| <2.47$	2.225±0.032	4.626±0.214	7.350±0.616	9.921±1.305

Table 6.7 Charge mis-identification rates as a function of p_T and $|\eta|$ for tight electrons measured from the data using the likelihood method. The values are in % and the errors are statistical only.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 1000$
$0< \|\eta\| <1.37$	0.68±0.02	3.84±0.38
$1.52< \|\eta\| <2.47$	5.37±0.04	12.18±0.47

Table 6.8 Charge mis-identification rates as a function of p_T and $|\eta|$ for looseNotTight electrons measured from the data using the likelihood method. The values are in % and the errors are statistical only.

Le tableau ci-dessus montre les probabilités de mauvaise identification des charges d'électrons mesurées.

En ce qui concerne les bruits de fond dus à de faux leptons, une technique d'extraction de taux de faux leptons basée sur les données a été développée et utilisée. Le « facteur de faux » est défini comme le rapport entre le nombre d'événements avec deux leptons précis de celui avec un lepton précis et un lepton lâche, ils sont mesurés dans la région d'énergie transverse manquante (MET) faible et utilisés dans la région MET élevée comme illustré dans les formules ci-dessous:

$$\theta_\mu = \frac{N_{\mu\mu}}{N_{\mu\not\mu}}(E_T^{miss} < 70\text{ GeV}) = \frac{N_{\mu\mu}^{Data} - N_{\mu\mu}^{\text{Prompt }SS}}{N_{\mu\not\mu}^{Data} - N_{\mu\not\mu}^{\text{Prompt }SS}},$$

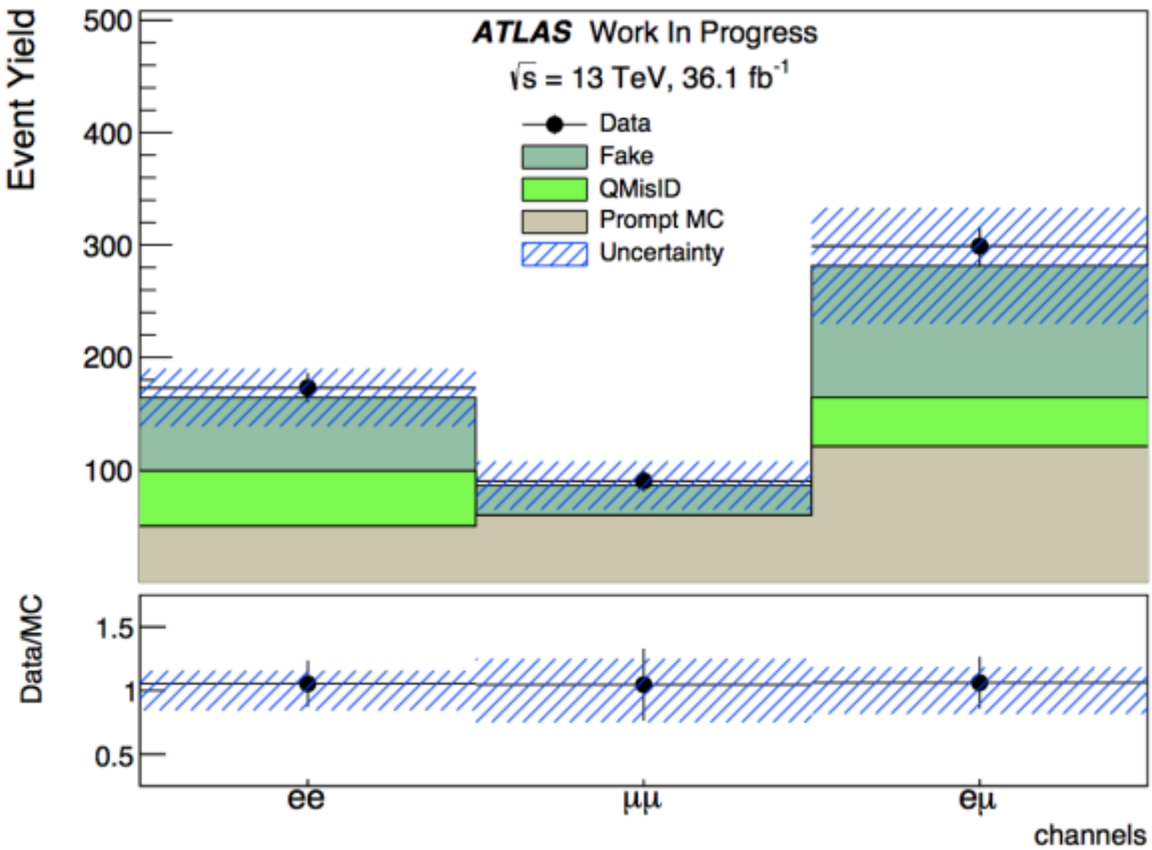
$$\theta_e = \frac{N_{\mu e}}{N_{\mu\not e}}(E_T^{miss} < 70\text{ GeV}) = \frac{N_{\mu e}^{Data} - N_{\mu e}^{\text{Prompt }SS} - N_{\mu e}^{\text{QMisId}} - N_{\mu e}^{\text{FakeMuon}}}{N_{\mu\not e}^{Data} - N_{\mu\not e}^{\text{Prompt }SS} - N_{\mu\not e}^{\text{QMisId}}}, \tag{6.4}$$

$$N_{ee}^{fakes}(\text{E}_T^{miss} \geqslant 70\text{ GeV}) = (N_{e\not e} - N_{e\not e}^{\text{Prompt }SS} - N_{e\not e}^{\text{QMisId}}) \times \theta_e,$$

$$N_{\mu\mu}^{fakes}(\mathrm{E}_T^{miss} \geqslant 70\,\mathrm{GeV}) = (N_{\mu\not\mu}^{Data} - N_{\mu\not\mu}^{\mathrm{Prompt}\ SS}) \times \theta_\mu,$$

$$N_{e\mu}^{fakes}(\mathrm{E}_T^{miss} \geqslant 70\,\mathrm{GeV}) = (N_{e\not\mu} - N_{e\not\mu}^{\mathrm{Prompt}\ SS} - N_{e\not\mu}^{\mathrm{QMisID}}) \times \theta_\mu + (N_{\mu\not e} - N_{\mu\not e}^{\mathrm{Prompt}\ SS} - N_{\mu\not e}^{\mathrm{QMisID}}) \times \theta_e.$$

La figure ci-dessous montre la comparaison entre les données et le fond estimé dans la région de pré-sélection d'événement, le fond est bien contrôlé.



Les incertitudes statistiques et systématiques sont prises en compte dans l'interprétation statistique finale. La méthode du rapport de vraisemblance profilé est exploitée et les trois canaux sont combinés. Comme aucun excès significatif n'est observé par rapport au SM, des limites sont dérivées sur le modèle.

Les significances de signal observées et attendues et les limites sont indiquées ci-dessous. Le modèle peut être exclu avec 95% de CL pour $M_{H^{\pm\pm}} < 260$ GeV avec des limites attendues combinant tous les canaux. Les limites observées excluent le modèle à $M_{H^{\pm\pm}} < 220$ GeV.

Dans cette thèse, la mesure de la production SM avec les données du Run 1 et une recherche de la nouvelle particule avec les données du Run 2 ont été présentées, mais aucun signe de nouvelle physique au-delà du SM n'a été trouvé.

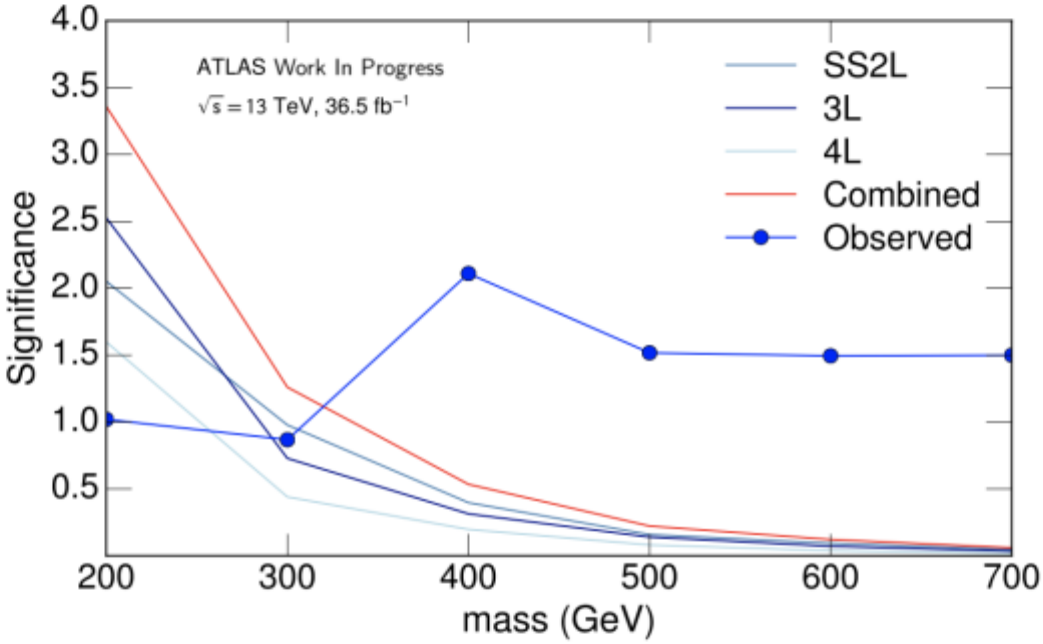


Figure 6.32 Expected and observed significances as a function of $M_{H^{\pm\pm}}$.

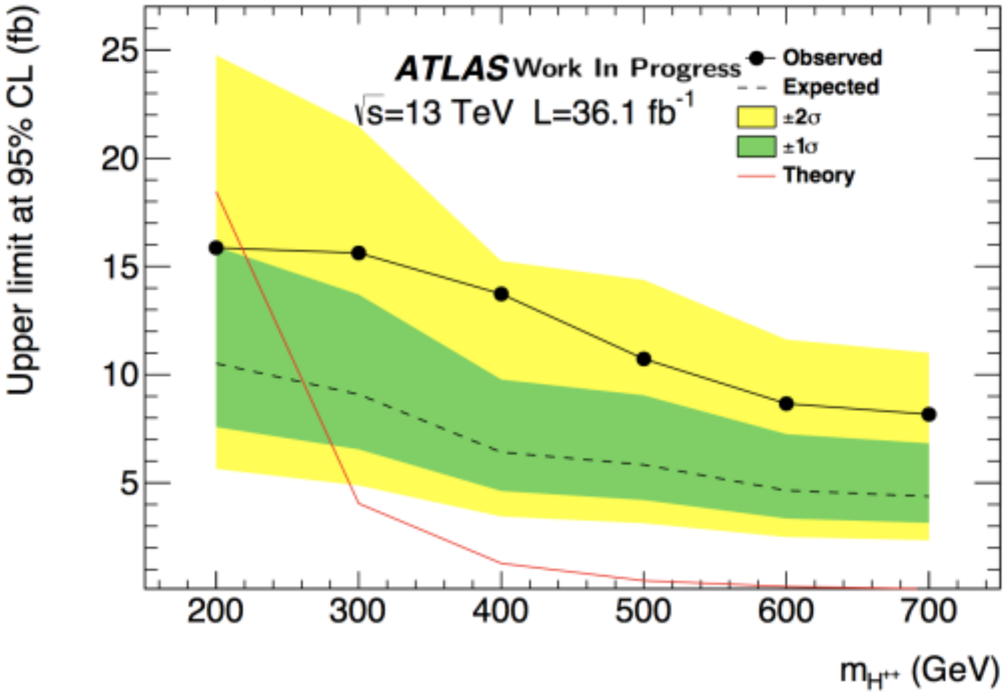


Figure 6.33 Expected and observed upper limits at 95% CL for the combination of $2\ell^{SS}$, 3ℓ and 4ℓ.

Contents

Chapter 1 Theory

1.1 The Standard Model

The Standard Model (SM) of particle physics is the theory describing elementary particles and their interactions (electromagnetic, weak and strong) based on the mathematical framework of Quantum Field Theory and gauge principles. It was developed throughout the 20^{th} century by many scientists around the world. The theory provides a unified description for the electromagnetic, weak and strong interactions based on the a $SU(3) \times SU(2) \times U(1)$ group in which $SU(3)$ group describes the strong interaction and $SU(2) \times U(1)$ describes the electroweak interactions. The $SU(2) \times U(1)$ symmetry is spontaneous broken via the Higgs mechanism to generate masses for weak gauge bosons and fermions meanwhile a Higgs boson is predicted.

The SM is theoretically self-consistent and extremely accurate, it has been verified by various experiments during the past several decades. Therefore it is considered as the most successful theory ever built but it's not perfect. There are several flaws in the theory such as neutrinos are assumed as massless which is in contrast to the experimental observations, gravitation which is one of the fundamental forces is not described in the theory. Thus various models have been developed by theorists to complete the SM. Unfortunately none of them have been verified by experiments yet.

1.1.1 Elementary Particles

The SM incorporates all known elementary particles and three of the four fundamental forces (electromagnetic, weak and strong). The 17 elementary particles are summarized in Figure 1.1. Every elementary particle is associated with a spin quantum number s which can be any integer or half-integer. Particles associated with different types of spin follow different statistical rules. They can be divided into two groups: fundamental constituent of matter called fermions and mediators of interactions called bosons. Fermions represent particles associated with half-integer spin and they follow a statistical rule called Pauli exclusion principle, i.e no two fermions can be described by the same quantum numbers. 12 of the elementary particles are fermions. Each of them has a corresponding anti-particle and these fermions can be further divided into two groups according to how they interact.

- Leptons: elementary fermions including electron neutrino, muon, muon neutrino,

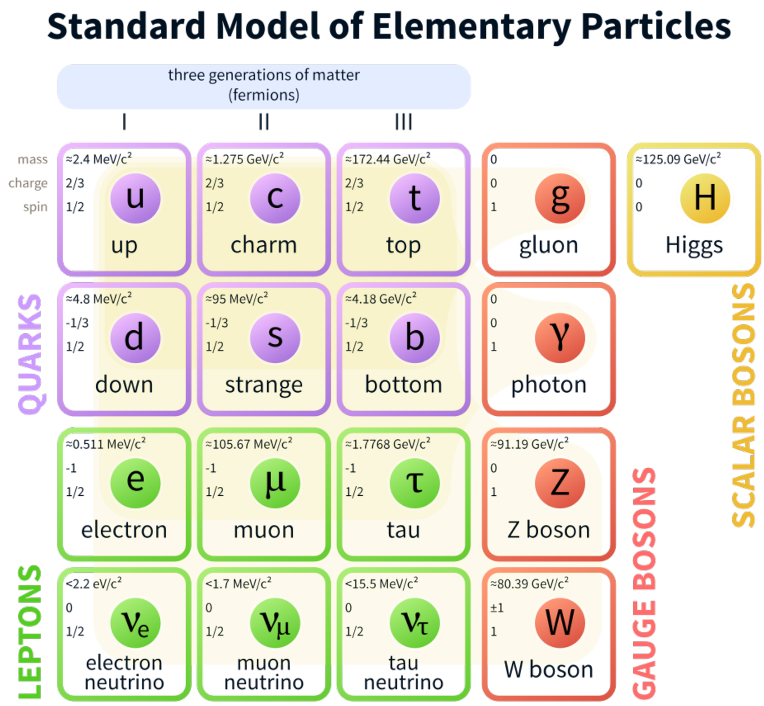


Figure 1.1 Elementary particles in the SM model.[1]

tau and tau neutrino.These six fermions are grouped to form three generations. Leptons have electric charge and weak isospin which means they can interact with other fermions via electromagnetic and weak interactions. Neutrinos do not carry any electric charge therefore they are only affected by the weak force.

- Quarks: elementary fermions which are bind into triplets and doublets to form baryons and mesons. They are also grouped into three generations according to their flavors. In addition to electric charge, quarks have a special property called color charge: R (Red), G (Green) and B (Blue). They can interact with other fermions via electromagnetic, weak and strong interactions.

Bosons refer to particles associated with integer spin numbers and they obey the Bose-Einstein statistic rule. Five of the elementary particles are bosons:

- Gauge bosons:
  - $W^{\pm}$ and Z bosons with spin 1 which are carriers of weak and electromagnetic forces.
  - Massless photon with spin 1 which is the mediator of electromagnetic interaction between electrically charged particles.
  - Eight massless gluons with spin1 which are mediators of strong interaction between color charged particles.

Forces	Strength	Range(m)	Mediating Particle
strong	1	10^{-15}	gluons
electromagnetic	$\frac{1}{137}$	Infinite	photon
weak	10^{-6}	10^{-18}	$W^{\pm}$, Z
gravity (not in the SM)	10^{-38}	Infinite	graviton?

Table 1.1 Properties of the four fundamental interactions.

- Massive scalar boson with spin 0: Higgs. Higgs boson is a unique particle in the SM since it explains the origin of mass for weak gauge bosons and fermions.

Properties of the four fundamental interactions are summarized in Table 1.1.

1.1.2 Fundamental Interactions

The SM is formulated with relativistic quantum field theory (QFT) and the gauge principles. The Lagrangian formalism is adopted to represent the quantum field theory. The elementary particles are considered as excitations of underlying fields and these fields are operators on the quantum mechanical Hilbert space of the particle states. The interactions between the particles are described via the gauge theory based on the $SU(3) \times SU(2) \times U(1)$ group.

In the SM, particles are described by different classes of fields according to their spin:

- spin 0 particles are described by scalar fields $\phi(x)$.
- spin 1 particles are described by vector fields $A_\mu(x)$.
- spin $\frac{1}{2}$ particles are described by spinor fields $\psi(x)$.

The dynamics of the physical system involving a set of fields is determined by the $\mathcal{L}$ which yields the action:

$$S[\phi] = \int d^4x \mathcal{L}(\phi(x)), \tag{1.1}$$

following the Euler-Lagrange equations from Hamilton's principle.

$$\delta S = S[\phi + \delta\phi] - S[\phi] = 0, \tag{1.2}$$

therefore, the equation of motion is obtained using the filed theory:

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{1.3}$$

A full Lagrangian can be considered as the sum of the free part and the interaction part. The Lagrangian of free fields is constructed using the knowledge of classical physics such as electrodynamics. The interacting terms are constructed from the gauge principles. Gauge invariance is a powerful principle to dictate the structure of interactions between the elementary particles. The construction of the Lagrangian of the interactions also indicates that the three fundamental forces originate from some internal symmetries of the world.

1.1.2.1 Electroweak Theory

The quantum electrodynamics (QED) is used to describe the electromagnetic interaction. It's derived by requiring a global $U(1)$ symmetry of the Lagrangian for the free charged fermion field. Lagrangian of the free charged fermion field:

$$\mathcal{L}_0 = \overline{\psi}(\gamma^\mu \partial_\mu - m)\psi, \tag{1.4}$$

is symmetric under the phase transformation:

$$\psi(x) \to \psi'(x) = e^{i\alpha}\psi(x), \tag{1.5}$$

for any real number α. This can be extended to a symmetry under local transformation where $\alpha \to \alpha(x)$ is an arbitrary real function, the partial derivative has to be replaced by a covariant derivative via minimal substitution to preserve invariance, it is defined as:

$$\partial_\mu \to D_\mu = \partial_\mu - ieA_\mu, \tag{1.6}$$

where A_μ is a vector field. Thus the local gauge transformations defined as the electromagnetic gauge group $U(1)$ can be written as:

$$\begin{aligned} \psi(x) &\to \psi'(x) = e^{i\alpha(x)}\psi(x), \\ A_\mu(x) &\to A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu \alpha(x), \end{aligned} \tag{1.7}$$

the invariant Lagrangian can be also be expressed as:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi = \mathcal{L}_0 + e\overline{\psi}\gamma^\mu \psi A_\mu = \mathcal{L}_0 + \mathcal{L}_{int}. \tag{1.8}$$

The vector field A_μ itself is not a dynamical field since the kinematic term is absent. This term can be taken from the classical electrodynamics:

$$\mathcal{L}_A = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \text{ where } F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{1.9}$$

In conclusion, there are three steps to describe the electromagnetic interaction via gauge theory:

- identify the global symmetry of the free Lagrangian.
- replace the partial derivative by the covariant derivative with a vector field.
- add a kinematic term for the vector field.

the success of QED based on the Abelian $U(1)$ group encourage people to seek similar theory to describe other fundamental interactions. Therefore the Abelian $U(1)$ symmetry is extended to the non-Abelian $SU(2) \times U(1)$ symmetry to generalize the theory. Consider a non-interacting system describing by a multiplet of fermion fields with mass m, $\Psi = (\psi_1, \psi_2, ..., \psi_n)^T$, its free Lagrangian can be written as

$$\mathcal{L}_0 = \overline{\Psi}(\gamma^\mu \partial_\mu - m)\Psi. \tag{1.10}$$

It's invariant under global transformation:

$$\Psi(x) \to U(\alpha^1, ..., \alpha^N)\Psi(x), \tag{1.11}$$

where U is unitary matrices from a non-Abelian Lie group G of rank N, depending on N real numbers $\alpha^1, ..., \alpha^N$, it is written as follows:

$$U(\alpha^1, ..., \alpha^N)\Psi(x) = e^{i(\alpha^1 T_1 + \cdots + \alpha^N T_N)}. \tag{1.12}$$

$T_1, \ldots, T_N$ are the generators of the Lie group. The global symmetry can then be extended to a local symmetry through replacing the constants α^a by real functions $\alpha^a(x)$, the covariant derivative is then introduced to replace the partial derivative:

$$\begin{aligned} \partial_\mu \to D_\mu &= \partial_\mu - ig\mathbf{W}_\mu, \\ \mathbf{W}_\mu(x) &= T_a W^a_\mu(x) \;\; \text{(summation over } a = 1, \ldots, N), \end{aligned} \tag{1.13}$$

where $\mathbf{W}_\mu$ is a vector field. Therefore the local gauge transformation that keep the Lagrangian invariant is:

$$\begin{aligned} \Psi &\to \Psi' = U\Psi, \\ \mathbf{W}_\mu \to \mathbf{W}'_\mu &= U\mathbf{W}_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}. \end{aligned} \tag{1.14}$$

The kinematic term of the vector field $\mathbf{W}$ is obtained from a generalization of the electromagnetic field strength tensor $F_{\mu\nu}$:

$$\mathbf{F}_{\mu\nu} = T_a F^a_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu - ig[\mathbf{W}_\mu, \mathbf{W}_\nu]. \tag{1.15}$$

The trace Tr($\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}$) is found to be gauge invariant, therefore the kinematic term of the field W_μ^a can be written as:

$$\mathcal{L}_W = -\frac{1}{2}Tr(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}) = -\frac{1}{4}F_{\mu\nu}^a F^{a,\mu\nu}. \tag{1.16}$$

The quadratic part of $\mathcal{L}_W$ describes the free propagation of the W fields while the cubic and quartic terms describes the self-interactions of the vector fields. By far, there is no mass term for the vector fields because any terms like $\frac{m^2}{2}W_\mu^a W_\nu^a$ will break the local gauge symmetry. The mass term is given by the Higgs boson which is discussed later. The electromagnetic and weak interactions are unified in the SM. The elementary fermions (leptons, neutrinos and quarks) can be classified into two categories: left-handed doublets and right-handed singlets. And this is just the fundamental representation of the $SU(2) \times U(1)$ group. Quantum numbers of weak isospin I, I_3 and weak hypercharge Y are used for the classification, left-handed fields have $I = \frac{1}{2}$ and thus form doublets while right-handed fields have $I = 0$. The relation between I_3, Y and the electric charge is found to be:

$$Q = I_3 + \frac{Y}{2}. \tag{1.17}$$

Therefore the $SU(2) \times U(1)$ group has four generators: I_1, I_2, I_3 and Y and each of them is associated with a vector field. The field strength tensor is then constructed as:

$$\begin{aligned} W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2\epsilon_{abc}W_\mu^b W_\nu^c, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned} \tag{1.18}$$

where $W_\mu^{1,2,3}$ are vector fields with $I_{1,2,3}$ and B_μ is singlet vector field with Y. There are two independent gauge coupling constants: g_2 for the non-Abelian factor $SU(2)$ and g_1 for the Abelian factor $U(1)$. The gauge field Lagrangian is written down as:

$$\mathcal{L}_G = -\frac{1}{4}W_{\mu\nu}^a W^{a,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \tag{1.19}$$

This Lagrangian is invariant under gauge transformations. The mass terms are not included since they violate the symmetry. Since the left-handed and right-handed fields are represented in different groups, they can also be denoted as:

$$\psi_L = \frac{1-\gamma^5}{2}\psi,\ \psi_R = \frac{1+\gamma^5}{2}\psi, \tag{1.20}$$

and these fields are grouped into doublets and singlets:

$$\psi_L^j = \begin{pmatrix} \psi_{L+}^j \\ \psi_{L-}^j \end{pmatrix},\ \psi_{R\sigma}^j, \tag{1.21}$$

where the index $\sigma = \pm$ stands for up type fermions (+) and down type fermions (-). The covariant derivative and the modified Lagrangian can be expressed as:

$$\begin{aligned} D_\mu^{L,R} &= \partial_\mu - ig_2 I_a^{L,R} W_\mu^a + ig_1 \frac{Y}{2} B_\mu \text{ with } I_a^L = \frac{1}{2}\sigma_a,\ I_a^R = 0, \\ \mathcal{L}_F &= \sum_j \overline{\psi}_L^j i\gamma^\mu D_\mu^L \psi_L^j + \sum_{j,\sigma} \overline{\psi}_{R\sigma}^j i\gamma^\mu D_\mu^R \psi_{R\sigma}^j, \end{aligned} \tag{1.22}$$

where index j indicates the three generations of lepton and quark.

1.1.2.2 QCD

The electroweak interaction is described by QED while the strong interaction is described by the quantum chromodynamics (QCD). QCD is formulated using the gauge theory with $SU(3)$ symmetry group. As mentioned in previous section, quarks have three different color quantum numbers (RGB). They are bind into mesons and baryons in doublet or triplet, but the hadrons are color neutral. Strong interaction occurs between quarks via exchanging gluons. Therefore the three color states are the foundation of the gauge theory. The fermion fields are described as $\Psi = (q_1, q_2, q_3)^T$ for each quark flavor $u, d, \ldots$. The color group $SU(3)$ has eight generators $T_a = \frac{1}{2}\lambda_a (a = 1, \ldots, 8)$, the eight generators are expressed in a 3×3 matrices, the Gell-Mann matrices λ_a. The covariant derivative acting on Ψ and the field strength are written as:

$$\begin{aligned} D_\mu &= \partial_\mu - ig_s \frac{\lambda_a}{2} G_\mu^a, \\ G_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f_{abc} G_\mu^b G_\nu^c, \end{aligned} \tag{1.23}$$

where g_s is the dimensionless coupling constant of QCD which can also be expressed in terms of fine structure constant of the strong interaction: $\alpha_s = \frac{g_s^2}{4\pi}$. The Lagrangian of QCD can be written down as:

$$\begin{aligned} \mathcal{L}_{QCD} &= \overline{\Psi}(i\gamma^\mu D_\mu - m)\Psi + \mathcal{L}_G, \\ \mathcal{L}_G &= g_s \overline{\Psi} \gamma^\mu \frac{\lambda_a}{2} \Psi G_\mu^a - \frac{1}{4} G_{\mu\nu}^a G^{a,\mu\nu}. \end{aligned} \tag{1.24}$$

This Lagrangian consists of the interaction between the quarks and the gluons as well as the gluon self interactions. The mass of the quark m is a free parameter in the Lagrangian for a given color triplet, it's different for different quark flavors.

1.1.2.3 Higgs Mechanism

The origin of mass is explained by the spontaneous breaking of $SU(2) \times U(1)$ symmetry following the Higgs mechanism. Higgs field is a doublet of scalar fields with hypercharge $Y = 1$:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \tag{1.25}$$

couple this doublet to the gauge fields via minimal substitution, the covariant derivative and the Lagrangian are constructed as:

$$\begin{aligned} D_\mu &= \partial_\mu - ig_2\frac{\sigma_a}{2}W_\mu^a + i\frac{g_1}{2}B_\mu, \\ \mathcal{L}_H &= (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi). \end{aligned} \tag{1.26}$$

The self-interaction of the Higgs field is included through the Higgs potential with constants μ^2 and λ.

$$V(\Phi) = -\mu^2\Phi^\dagger\Phi + \frac{\lambda}{4}(\Phi^\dagger\Phi)^2 \quad \mu^2, \lambda > 0. \tag{1.27}$$

The potential is naturally minimum in the ground state. V is minimized with $\Phi^\dagger\Phi = 2\mu^2/\lambda$ and the one selected is:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix} \quad \text{with } \nu = \frac{2\mu}{\sqrt{\lambda}}. \tag{1.28}$$

The Lagrangian is invariant under gauge transformations of the full $SU(2)\times U(1)$ group however the symmetry of vacuum expectation is spontaneously broken. $\langle\Phi\rangle$ is still invariant under the transformations of the electromagnetic group $U(1)$ and preserves the electromagnetic gauge symmetry. The Higgs field can also be expressed as:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ (\nu + H(x) + iK(x))/\sqrt{2} \end{pmatrix}, \tag{1.29}$$

where ϕ^+, H, K have vacuum expectation value zero. ϕ^+ and K can be eliminated by exploiting the invariance of the Lagrangian, this particular gauge where ϕ^+ and K are 0 is denoted as the unitary gauge, then Higgs doublet and the potential are then simplified as:

$$\begin{aligned} \Phi(x) &= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix}, \\ V = \mu^2H^2 + \frac{\mu^2}{\nu}H^3 + \frac{\mu^2}{4\nu^2}H^4 &= \frac{M_H^2}{2}H^2 + \frac{M_H^2}{2\nu}H^3 + \frac{M_H^2}{8\nu^2}H^4. \end{aligned} \tag{1.30}$$

Therefore, the real field $H(x)$ describes a neutral scalar particle: the Higgs boson with mass $M_H = \mu\sqrt{2}$. The mass of bosons are generated from the couplings between the

Higgs field and the gauge boson fields. The calculation is performed by replacing the Φ in Eq 1.26 by $\langle\Phi\rangle$, the mass terms obtained are:

$$\frac{1}{2}(\frac{g_2}{2}\nu)^2(W_1^2+W_2^2)+\frac{1}{2}(\frac{\nu}{2})^2(W_\mu^3,B_\mu)\begin{pmatrix} g_2^2 & g_1g_2 \\ g_1g_2 & g_1^2 \end{pmatrix}\begin{pmatrix} W^{3,\mu} \\ B^\mu \end{pmatrix}. \tag{1.31}$$

It's then transformed to physical fields W_μ^a, B_μ in terms of which the symmetry is manifest:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\ \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}. \end{aligned} \tag{1.32}$$

The mass of vector bosons and the mixing angle are:

$$\begin{aligned} M_W &= \frac{1}{2}g_2\nu, \quad M_Z = \frac{1}{2}\sqrt{g_1^2+g_2^2}\nu, \\ \cos\theta_W &= \frac{g_2}{\sqrt{g_1^2+g_2^2}} = \frac{M_W}{M_Z}. \end{aligned} \tag{1.33}$$

The relation between the electric charge e and the coupling constants g_1, g_2 can be expressed as:

$$e = \frac{g_1g_2}{\sqrt{g_1^2+g_2^2}}, \quad g_2 = \frac{e}{\sin\theta_W}, \quad g_1 = \frac{e}{\cos\theta_W}. \tag{1.34}$$

The relations illustrated above reveal the fact that there are two massive vector bosons and one massless vector boson in the electroweak theory and the mass of the massive vector bosons is determined by the coupling between the Higgs field and the gauge boson fields. The θ_W is known as the weak mixing angle which is also an important parameter for precise measurement in experiments.

To further allow mass for fermions, the Yukawa interaction between the Higgs field and the fermion fields is introduced. The Yukawa Lagrangian in the unitary gauge is written as:

$$\mathcal{L}_Y = -\sum_f m_f\overline{\psi}_f\psi_f - \sum_f \frac{m_f}{\nu}\overline{\psi}_f\psi_f H, \tag{1.35}$$

where m_f stands for the mass of the fermion. Coupling constants of the Yukawa interactions between the massive fermions and the Higgs field are proportional to the mass of the fermions, $m_f = G_f\frac{\nu}{\sqrt{2}}$ where G_f is the Yukawa coupling constant. Considering that quarks are in three generations, the flavor mixing has to be taken into account for the quark sector. Yukawa couplings are now in matrices as:

$$\mathcal{L}_Y^{quarks} = -G_{ij}^d\overline{Q}_L^i\Phi d_R^j - G_{ij}^u\overline{Q}_L^i\Phi^c u_R^j + h.c., \tag{1.36}$$

where $Q_L^i = (u_L^i, d_L^i)^T$ is for the three left-handed doublets $[u^i = u, c, t$ and $d^i = d, s, b]$. The mass term is obtained through replacing Φ by $\langle\Phi\rangle$. Unitary matrices $V_{L,R}^q$ are adopted to diagonalize the mass terms:

$$\hat{u}_{L,R}^i = (V_{L,R}^u)_{i,k} u_{L,R}^k, \quad \hat{d}_{L,R}^i = (V_{L,R}^d)_{ik} d_{L,R}^k, \tag{1.37}$$

where u and d stand for the index of up type quarks and down type quarks. The mass term can be diagonalized as:

$$diag(m_q) = \frac{\nu}{\sqrt{2}} V_L^q G_q V_R^{q+}, \ \ q = u, d. \tag{1.38}$$

Because the unitarity of the transformation, the kinematic and interaction terms with the gauge bosons are not changed, also the Yukawa interaction between quarks and the Higgs stays invariant under this transformation. The only modification occurs in the flavor-changing quark interaction via interacting with the vector bosons where the insertion of the mass eigenstates yields the unitary CKM matrix:

$$V_L^u V_L^{d\dagger} \equiv V_{CKM}. \tag{1.39}$$

V_{CKM} has four independent physical parameters due to the constraint from unitarity: three real angles and one complex phase.

No generation mixing occurs for leptons due to the massless neutrinos. However, it's observed from the neutrino oscillation experiments that neutrinos do have non-zero masses which is in contrast to the SM. Therefore various models were developed to explain the mass for neutrinos meanwhile accommodate to the SM. None of these models were verified by experiment so far.

In conclusion, the SM is formulated using the quantum field theory and the gauge theory. Different kinds of fields are adopted to describe particles with different spin, the Lagrangian of these fields are constructed using the knowledge of classical physics while the Lagrangian of the interactions are constructed using the gauge theory where the three fundamental interactions (electromagnetic, weak and strong) are considered to originate from some symmetries of the world. The electromagnetic and weak interactions are unified in a $SU(2) \times U(1)$ group while the strong interaction is described by a $SU(3)$ group. The full Lagrangian in the SM can be expressed as:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{free} + \mathcal{L}_{int}, \\ \mathcal{L}_{int} &= \mathcal{L}_{EW} + \mathcal{L}_{QCD}. \end{aligned} \tag{1.40}$$

To preserve the gauge symmetry, mass terms of gauge bosons (fermions) are obtained through the coupling between the Higgs field and the gauge boson (fermion) fields. This

Higgs mechanism also yields a neutral massive scalar boson which is the last missing puzzle of the SM. The discovery of the Higgs boson is announced in 2012 by the ATLAS and the CMS experiments.

1.1.3 Imperfection

The SM is considered as the most successful model ever built on basis of experimental observations as it was verified by various experiments over decades, however it's imperfect. Some predictions of the SM are in contrast to the observation from experiments while some phenomena in the universe is even not included in the SM.

- Neutrino masses: neutrinos are predicted to be massless in the SM while the neutrino oscillation experiments have illustrated that they do have small mass.
- Matter-antimatter asymmetry: the dominance of matter observed in the universe can not be naturally explained by the Big Bang theory together with the SM.
- Gravity: Three of the four fundamental forces are described in the SM besides the gravity.
- Dark matter and dark energy: The SM does not supply any explanation for dark matter and dark energy.

Various models have been developed to explain these phenomena but unfortunately none of them has been verified by experiments, physicists are still on the way looking for new physics beyond the SM.
There are two complementary approaches in searching for new physics beyond the SM. The first approach is to search for anomalous couplings, which does not introduce explicitly new particles. The new particles are assumed to be too heavy to be directly observed at the LHC, but they alter indirectly the interactions of the SM particles. In the second approach, a full model is specified, and the new particles are directly searched for. The $W^{\pm}W^{\pm}W^{\mp}$ analysis belongs to the first class, while the search for the doubly charged Higgs is an example of the second approach.

1.2 $W^{\pm}W^{\pm}W^{\mp}$ Production and Anomalous Gauge Couplings

The $W^{\pm}W^{\pm}W^{\mp}$ is a rare Electroweak process allowed by the Standard Model and its cross section at Next-to-Leading-Order (NLO) is already known. Observation of significant deviation from the prediction will imply new physics beyond the Standard Model.

Analysis is performed for events with full leptonic final states and semi-leptonic final states. The signatures of signal process are $pp \rightarrow W^+W^+W^- + X$ and $pp \rightarrow W^+W^-W^- + X$ with the W bosons decaying to leptons and neutrinos, both on-shell and off-shell W bosons are considered. The off-shell contribution is due to the Higgs boson production, namely, $pp \rightarrow HW$ with $H \rightarrow WW^*$. Feynman diagrams of the signal process are shown in Figure 1.2.

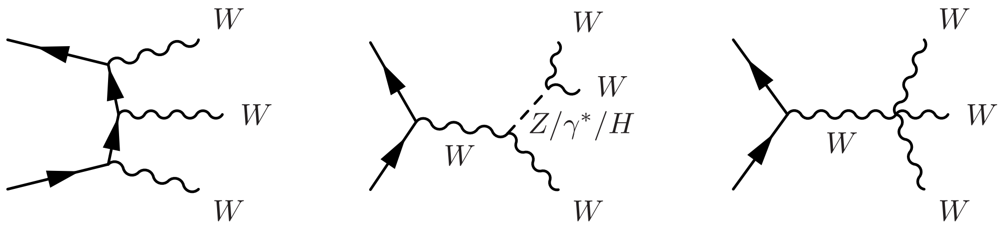


Figure 1.2 The Feynman diagrams for the $W^{\pm}W^{\pm}W^{\mp}$ production.

The cross section is calculated to NLO accuracy in QCD, cross sections without Higgs or with Higgs boson exchange and spin correlations of W boson leptons decay are both available. The signal process is simulated with a Monte Carlo package named VBFNLO[2]. VBFNLO can generate events at LO level and can compute cross section at NLO accuracy. The ratio of cross section at NLO accuracy to that at LO accuracy is defined as the k-factor which is about 1.4, detector simulation is included for the VBFNLO simulation. The signal process is also simulated with the MadGraph[3] generator which simulates both the non-resonant and resonant productions separately at NLO accuracy. Detector simulation is not implemented for the MadGraph simulation and the MadGraph samples are used to calculate the fiducial and total cross sections which can be compared to different WWW channels. Measurement of the $W^{\pm}W^{\pm}W^{\mp}$ production can be used to probe the gauge couplings, in particular, the process is sensitive to quartic gauge couplings. The VBFNLO code has implemented a list of dimension-8 operators that parameterize the effects of new physics at energy scale beyond the reach of the LHC. The effective field theory approach is widely used when there is no specific model of new physics beyond the Standard Model[4][5][6]. In this analysis, two gauge invariant dimension-8 operators are chosen:

$$\mathcal{L}_{s,0} = [(\mathrm{D}_\mu\Phi)^\dagger \mathrm{D}_\nu\Phi] \times [(\mathrm{D}^\mu\Phi)^\dagger \mathrm{D}^\nu\Phi], \tag{1.41}$$

$$\mathcal{L}_{s,1} = [(\mathrm{D}_\mu\Phi)^\dagger \mathrm{D}^\mu\Phi] \times [(\mathrm{D}_\nu\Phi)^\dagger \mathrm{D}^\nu\Phi], \tag{1.42}$$

	Cross Section [fb]	
	$W^+W^+W^-$	$W^+W^-W^-$
LO VBFNLO CTEQ6L1	3.56±0.005	1.88±0.003
NLO VBFNLO CT10NLO	4.95±0.007	2.56±0.004
VBFNLO k-factor	1.39	1.41

Table 1.2 The cross section of the SM WWW processes using VBFNLO with a center-of-mass energy of 8 TeV, only leptonic decays (e, μ, τ) of the W bosons are considered.

where Φ is the Higgs field doublet and D_μ is the covariant derivative. The Lagrangian of the effective field theory is thus:

$$\mathcal{L}_{eff} = \mathcal{L}_{SM} + \frac{f_{s0}}{\Lambda^4}\mathcal{L}_{s,0} + \frac{f_{s1}}{\Lambda^4}\mathcal{L}_{s,1}. \tag{1.43}$$

These two operators are chosen as benchmarks, more operators will be incorporated to investigate the nature of the new physics if a significant excess of events can be observed in data. In this analysis, VBFNLO is used to generate events at LO accuracy and calculate cross section at NLO accuracy for the signal processes including the SM scenario and the scenario with anomalous quartic gauge couplings.

Cross section of the SM signal at NLO accuracy from VBFNLO for $W^+W^+W^-$ and $W^+W^-W^-$ are 4.95 fb and 2.65 fb respectively, only leptonic decays of the W bosons are considered in the full leptonic channel. CTEQ6L1[7] PDF is used in the LO calculation while CT10NLO[8] PDF are used in the NLO calculation.The numbers are listed in Table 1.2. The signal process is also simulated using MadGraph at NLO accuracy, the total and fiducial cross sections are shown in Table 1.3. The MadGraph samples are produced at NLO accuracy using CTEQ6L1 PDF including all W boson decays. The samples are then reweighted to CT10NLO PDF so that the QCD order between the PDF and the generation can match. The k-factor is defined as the ratio of CT10NLO to CTEQ6L1.

Apart from the SM signal processes, signal sample with anomalous quartic gauge couplings are also required for further study. They are simulated with exactly the same settings for the SM samples, the cross sections are calculated at NLO accuracy. The cross section is very sensitive to aQGCs. It changes significantly with respect to the variation of $\frac{f_{s0}}{\Lambda^4}$ and $\frac{f_{s1}}{\Lambda^4}$, these coupling parameters are scaled with a form factor to restore the unitarity:

$$FF = \frac{1}{\left(1 + \frac{s}{\Lambda_{FF}^2}\right)^{NF}}$$ [9],

	Cross Section [fb]	
	Inclusive	Fiducial
NLO MadGraph CTEQ6L1	223.56±0.12	0.2812±0.0066
NLO MadGraph CT10NLO	241.47±0.13	0.3092±0.0072
MadGraph k-factor	1.08	1.10

Table 1.3 The cross section of the SM WWW processes using MadGraph with a center-of-mass energy of 8 TeV, all W decays are included. Table 5.17 shows the definitions of the fiducial region.

where the NF and the form factor scale Λ_{FF}^2 are arbitrary parameters. As suggested by the VBFNLO authors, NF =1 used to obtain the Λ_{FF}^2 from a VBFNLO tool[9] and NF =3 together with the obtained Λ_{FF}^2 are used for the event generation and a form factor with NF =1 can indeed control the growth of cross section at high center-mass-energy. The ratio of aQGC cross section to that of SM is stable at high center-mass-energy with proposed unitarization schema which is shown in Figure 1.3. The simulated

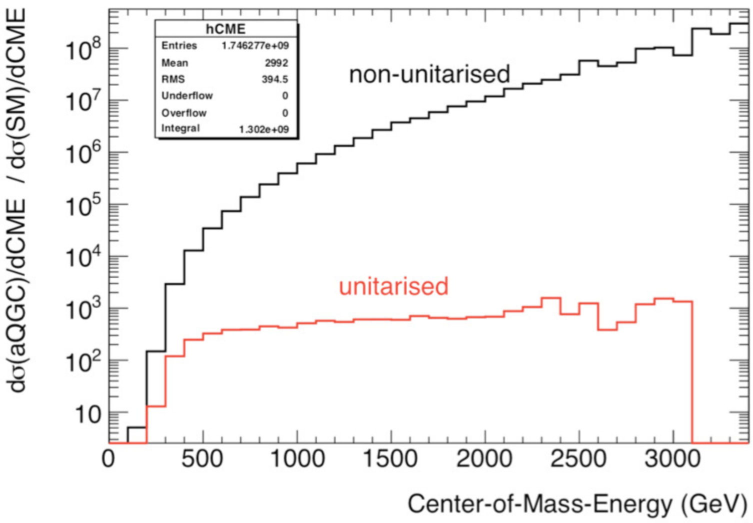


Figure 1.3 The unitarized and non-unitarized differential cross sections as a function of $\sqrt{s}$ for $f_{s0}/\Lambda^4 = 6 \times 10^{-7}\text{GeV}^{-4}$ divided by the SM values. The form factor function with $NF = 1$ and $\Lambda_{FF} = 180$ GeV is used for unitarization.

aQGC samples are generated without form factor, for each of them, two samples of very high statistic are generated with the same couplings while one is unitarized and the other is not. The non-unitarized samples are weighted to be unitarized using the ratio of the distribution of $\sqrt{s}$ of the two samples of high statistic. The measurement of the $W^{\pm}W^{\pm}W^{\mp}$ production and study of anomalous gauge couplings are elaborated in Chapter 5.

1.3 Higgs Doublet Triplet Model

Neutrinos are massless in the SM, however in nature the neutrino mass is found to be small but not zero by various experiments. To allow masses for neutrinos, a lot of models are developed and one famous model is the so called Type-2 "see-saw" model[10]. In this model, a scalar triplet with hypercharge $Y = 2$ is added to the Higgs scalar sector where the Lagrangian is expresses as:

$$\mathcal{L} = (D_\mu H)^\dagger (D^\mu H) + Tr(D_\mu \Delta)^\dagger (D^\mu \Delta) - V(H, \Delta) + \mathcal{L}_{Yukawa}, \tag{1.44}$$

where $\mathcal{L}_{Yukawa}$ contains the neutrino mass terms. $V(H, \Delta)$ is the scalar potential where H is the SM scalar doublet and Δ is the introduced scalar triplet defined as:

$$\begin{aligned} V(H,\Delta) = & - m_H^2 H^\dagger H + \frac{\lambda}{4}(H^\dagger H)^2 + \Delta^2 Tr(\Delta^\dagger \Delta) + [\mu(H^\dagger i\sigma^2 \Delta^\dagger H) + h.c.] \\ & + \lambda_1 (H^\dagger H) Tr(\Delta^\dagger \Delta) + \lambda_2 (Tr \Delta^\dagger \Delta)^2 + \lambda_3 Tr \left(\Delta^\dagger \Delta\right)^2 \\ & + \lambda_4 H^\dagger \Delta \Delta^\dagger H. \end{aligned}$$

Electroweak symmetry breaking (EWSB) can still be achieved if the vacuum expectation values of the neutral components of H (ν_d) and Δ (ν_t) are at the minimum of the potential. The mixing of the fields caused by the EWSB will result in seven scalar bosons: $H^{\pm\pm}$, $H^\pm$, A^0 (CP odd), H^0 (CP even) and h^0 (CP even). One of the neutral scalars is identical to the SM Higgs boson.

Various constraints are applied to the parameter space such as the upper bound on ν_t from other electroweak precision measurements. The SM Higgs potential is a function of $H^\dagger H$, known as custodial symmetry, implies that $\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1$ at tree level. Therefore, any extension of the SM should accommodate only small distortions from this symmetry. In this model, the modified ρ at tree level is:

$$\rho = \frac{\nu_d^2 + 2\nu_t^2}{\nu_d^2 + 4\nu_t^2} (< 1) \approx 1 - 2\frac{\nu_t^2}{\nu_d^2} (\nu_d \gg \nu_t).$$

To maintain the symmetry, finite ν_t values are allowed for exploring new physics.

There are three production modes of $H^{\pm\pm}$:

- Single production mode: $pp \to W^{\pm *} W^{\pm *} \to H^{\pm\pm}$.
- Pair production mode: $pp \to \gamma^*/Z^* \to H^{\pm\pm} H^{\mp\mp}$.
- Associated production mode: $pp \to W^{*+} \to H^{\pm\pm} H^{\mp}$.

The single production mode is negligible since it's produced via vector boson fusion and hence proportional to ν_t. Figure 1.4 shows the diagrams of the pair production mode and the associated production mode. This analysis focus on the pair production mode, the singly charged Higgs is restricted to be a few hundred GeV heavier than the doubly charged Higgs which will suppress the associated production mode. The doubly charged

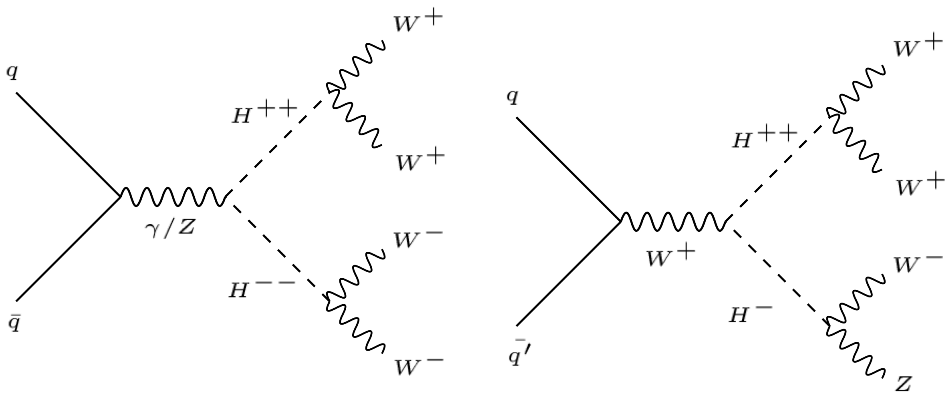


Figure 1.4 Feynman diagrams of the pair production mode (left) and the associated production mode (right).

Higgs may decay in two channels: $H^{\pm\pm} \to \ell^{\pm}\ell^{\pm}$ and $H^{\pm\pm} \to W^{\pm}W^{\pm}$, the WW decay channel is more preferable in this analysis. Figure 1.5 shows the relation between the branching ratio and the vacuum expectation value ν_t. $\nu_t = 0.1$ GeV is chosen to enlarge the branching ratio of the WW mode and the mixing between the CP-even Higgses is set to 10^{-4}, these constraints ensure the h^0 to behave like the SM Higgs. There is another group exploring the doubly charged Higgs decaying to leptons[11], but the two analyses are independent since the phase space we are exploring are totally different.

The signal process, i.e the pair production of $H^{\pm\pm}$, is simulated with the CalcHEP[12] generator using the parton distribution function (PDF) CTEQ6[13]. PYTHIA8[14] and A14 tune[15] are then used to simulate the parton shower and hadronization steps, these events are further passed to the perform the official ATLAS detector simulation (GEANT4[16]) and reconstruction. Samples are simulated with $M_{H^{\pm\pm}}$ = 200, 300, 400, 500, 600 and 700 GeV. The samples are skimmed with at least two light leptons with $p_T > 10$ GeV and $|\eta| < 10$ at truth level. Various configuration (choice of generator, PDF, etc.) are used to model the background processes, details are shown in Table 1.4.

The production cross sections and filter efficiencies for signal process with different $M_{H^{\pm\pm}}$ are summarized in Table 1.5. The search for $H^{\pm\pm}$ is discussed concretely in Chapter 6.

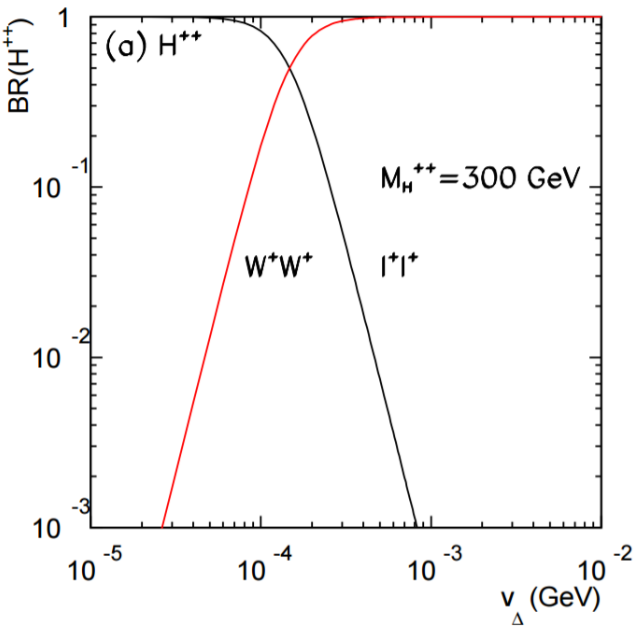


Figure 1.5 Branching ratio as a function of vacuum expectation value for $M_{H^{\pm\pm}} = 300$ GeV. Black line and red line are branching ratios for $H^{\pm\pm} \to \ell^{\pm}\ell^{\pm}$ decay channel and $H^{\pm\pm} \to W^{\pm}W^{\pm}$ decay channel respectively[10].

Process	ME Generator	Parton Shower	PDF	Tune
$t\bar{t}H$	MG5_aMC[17]	Pythia 8[14]	NNPDF 3.0 NLO[18]/ NNPDF 2.3 LO[19]	A14[15]
VH	Pyhtia 8	Pythia 8	NNPDF 2.3 LO	A14
$tHqb$	MG5_aMC	Pythia 8	CT10[8]/NNPDF 2.3 LO	A14
tHW	MG5_aMC	Herwig++[20]	CT10/CTEQ6L1[7,13]	UE-EE-5[21]
$t\bar{t}W$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO/2.3 LO	A14
$t\bar{t}(Z/\gamma^*)$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO/2.3 LO	A14
$t(Z/\gamma^*)$	MG5_aMC	Pythia 6[22]	CTEQ6L1	Perugia2012[23]
$tW(Z/\gamma^*)$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}t\bar{t}$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}$	Powheg-BOX[24]	Pythia 6	CT10/CTEQ6L1	Perugia2012
s-, t-channel, Wt single top	Powheg-BOX[25,26]	Pythia 6	CT10/CTEQ6L1	Perugia2012
VV, $qqVV$, VVV	Sherpa 2.1.1[27]	Sherpa	CT10	Sherpa default
$Z \to \ell^+\ell^-$	Sherpa 2.2	Sherpa	NNPDF 3.0 NLO	Sherpa default
$W \to \ell\nu$	Sherpa 2.2	Sherpa	NNPDF 3.0 NLO	Sherpa default

Table 1.4 Configurations used for event generation of 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/γ^*). "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; "Herwig++" refers to version 2.7. The samples have heavy flavor hadron decays modeled by EvtGen 1.2.0[28], except for samples generated with Sherpa.

$H^{\pm\pm}$ mass (GeV)	200	300	400	500	600	700
cross section (fb)	64.58	13.34	3.998	1.466	0.610	0.276
filter efficiency	0.2858	0.3031	0.3198	0.3264	0.3362	0.3451
DSID	344096	344097	344098	344364	344365	344366

Table 1.5 Cross sections, filter efficiencies and DSIDs of the signal samples.

Chapter 2 LHC and the ATLAS Detector

2.1 LHC

The Large Hadron Collider (LHC) is the world's largest and highest-energy particle accelerator which lies between 45 m and 170 m beneath the France-Switzerland border. It is hosted by the European Organization for Nuclear Research (CERN). The LHC consists of a 27-kilometer ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way. The LHC is designed to collide proton beams with a center-mass-energy of 14 TeV and an luminosity of $10^{34}\ \mathrm{cm^{-2}s^{-1}}$. It can also collide heavy (Pb) ions with an energy of 2.8 TeV per nucleon and a peak luminosity of $10^{27}\ \mathrm{cm^{-2}s^{-1}}$.

The LHC ran with $\sqrt{s} = 7$ TeV during 2011 and $\sqrt{s} = 8$ TeV during 2012, the machine was then shut down for upgrade and development between 2013 and 2014. During the upgrade, the electrical connectors between the bending magnets were upgraded to safely handle the current required for 7 TeV per beam, however the bending magnets were only trained to handle up to 6.5 TeV per beam. The second operational run of the machine started at 2015 with $\sqrt{s} = 13$ TeV.

At the LHC, protons are injected to the LHC through a complex injector chain: Linac2 $\rightarrow$ Proton Synchrotron Booster (PSB) $\rightarrow$ Proton Synchrotron (PS) $\rightarrow$ Super Proton Synchrotron (SPS) which is also shown in Figure 2.1. These accelerators are designed to produce high intensity proton bunches with small transverse and well defined longitudinal emittance which is required by the LHC. Protons are produced with 50 MeV energy in Linac2 and accelerated to about 450 GeV through these accelerators, the protons are then injected to the main ring for further acceleration and collision.

Basic layout of the LHC is shown in Figure 2.2. The LHC has eight arcs and eight straight sections. Each straight section is approximately 528 m long and can serve as an experimental or utility insertion. Six detectors are located at the LHC insertion points underground. There are two high luminosity experimental insertions located at diametrically opposite straight sections, namely, the ATLAS (A Toroidal LHC Apparatus) experiment and the CMS (Compact Muon Solenoid) experiment, aiming at peak luminosity of $10^{34}\ \mathrm{cm^{-2}s^{-1}}$ for proton beams. There are also two low luminosity experimental insertions: LHCB (Large Hadron Collider Beauty) for B-Physics aiming at peak luminosity of $10^{32}\ \mathrm{cm^{-2}s^{-1}}$ and TOTEM for the detection of protons with elastic scat-

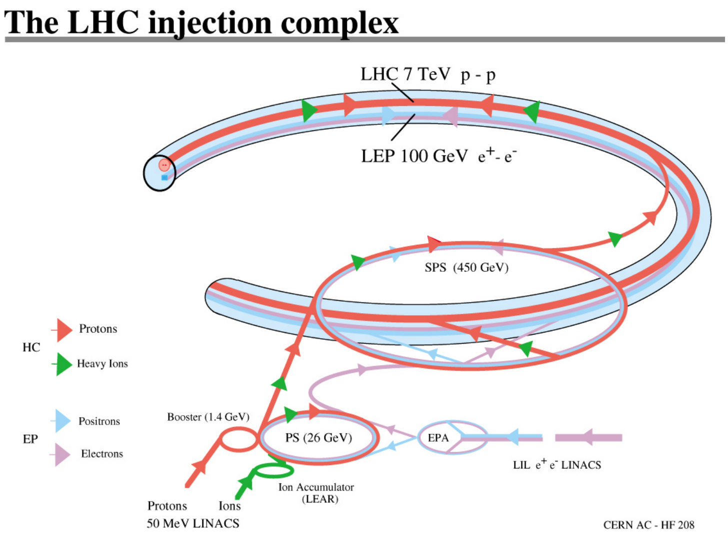


Figure 2.1 The LHC injector complex.[29]

tering at small angles, aiming at peak luminosity of $2 \times 10^{39}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. There is also an experiment called ALICE (A Large Ion Collider Experiment) dedicated to ion physics aiming at peak luminosity of $10^{27}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The Large Hadron Collider forward (LHCf) experiment uses particles thrown forward by collisions in the Large Hadron Collider as a source to simulate cosmic rays in laboratory conditions.

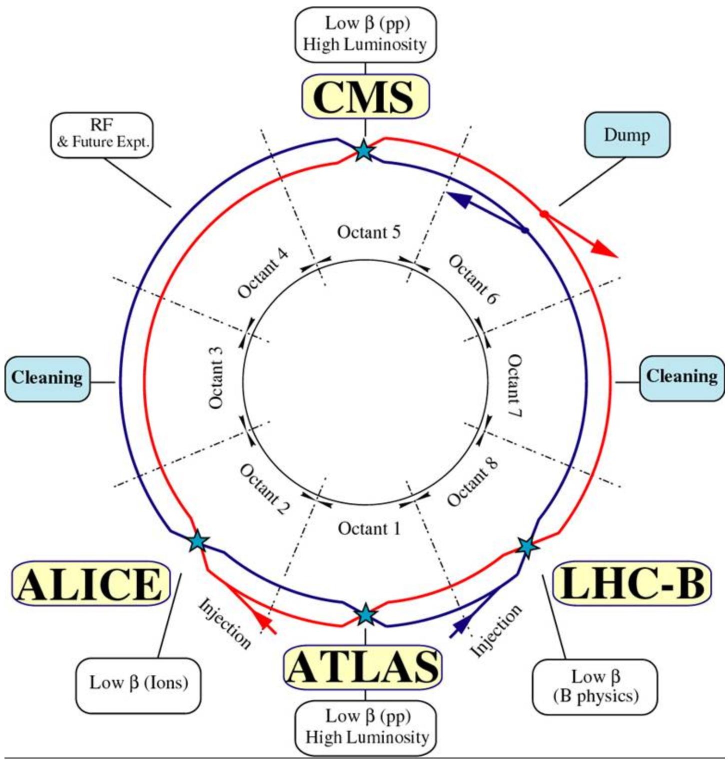


Figure 2.2 Schematic layout of the LHC.[30]

2.2 ATLAS Detector

At the LHC, bunches of up to 10^{11} protons will collide 40 million times per second. The high interaction rates, radiation doses, particle multiplicities and energies set new standards for the particle detectors. ATLAS(A Toroidal LHC ApparatuS) is one of the general purpose detectors, it provides a rich physics potential, from precise measurement of Standard Model to the search for new physics phenomena. A cut-away view of the ATLAS detector is shown in Figure 2.3. It is 46 meters long, 25 meters in diameter, weighs about 7,000 tonnes and nominally forward-backward symmetric with respect to the interaction point. ATLAS is a worldwide cooperated experiment which involves roughly 3000 physicists at 175 institutions in 38 countries.

The coordinate system and nomenclature used to described the ATLAS detector and the particles are summarized below.

The nominal interaction point is defined as the origin of the coordinate system while the beam direction defines the z-axis and the $x-y$ plane is transverse to the beam direction. The positive x-axis is defined as pointing to the centre of the LHC ring and the positive y-axis is defined as pointing upwards. The azimuthal angle ϕ is measured around the beam axis. The polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = -\ln\tan(\frac{\theta}{2})$. The transverse momentum p_T, the transverse energy E_T, and the missing transverse energy E_T^{miss} are defined in the $x-y$ plane. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ATLAS detector consists of four major parts, the magnet system, the inner detector, the calorimeters and the muon spectrometer. The magnet system consists of a thin superconducting solenoid surrounding the inner detector and three large superconducting toroids (one barrel and two end-caps) arranged with an eight-fold azimuthal symmetry around the calorimeters. The inner detector is immersed in a 2 T solenoidal field. It consists of three major parts, the pixel detector, the silicon microstrip trackers and the transition radiation tracker. Pattern recognition, momentum and vertex measurements, and electron identification are achieved through these sub-detectors. The high granularity liquid-argon electromagnetic sampling calorimeters cover the pseudorapidity range $|\eta| < 3.2$. The hadronic calorimetry in the range $\eta < 1.7$ is a scintillator-tile calorimeter separated into a large barrel and two smaller extended barrel cylinders, one on either side of the central barrel. The LAr technology is used for hadronic calorimeters in end-caps ($\eta > 1.5$) region. The LAr forward calorimeters provide both electromagnetic and hadronic energy measurements which extend the coverage to $\eta = 4.9$. The calorimeter is surrounded by the muon spectrometer. The air-core toroid system which consists of a long barrel and two end-cap magnets generates strong bending power in a large volume within a light and open structure and minimized the multiple-scattering effects. Excellent muon momentum resolution is achieved with three layers of high precision tracking chambers.

With designed luminosity of $10^{34}\ \mathrm{cm^{-2}s^{-1}}$, the proton-proton interaction rate of the LHC is about 1 GHz but the event data recording of ATLAS is limited to about 200 Hz due to technology and resource limitations. The Level-1 trigger system is designed to make a decision with a subset of the total detector information on whether or not to accept the event, reducing the data rate to about 75 kHz. The subsequent Level-2 trigger and the event filter system reduce the final data-taking rate to approximately 200 Hz.

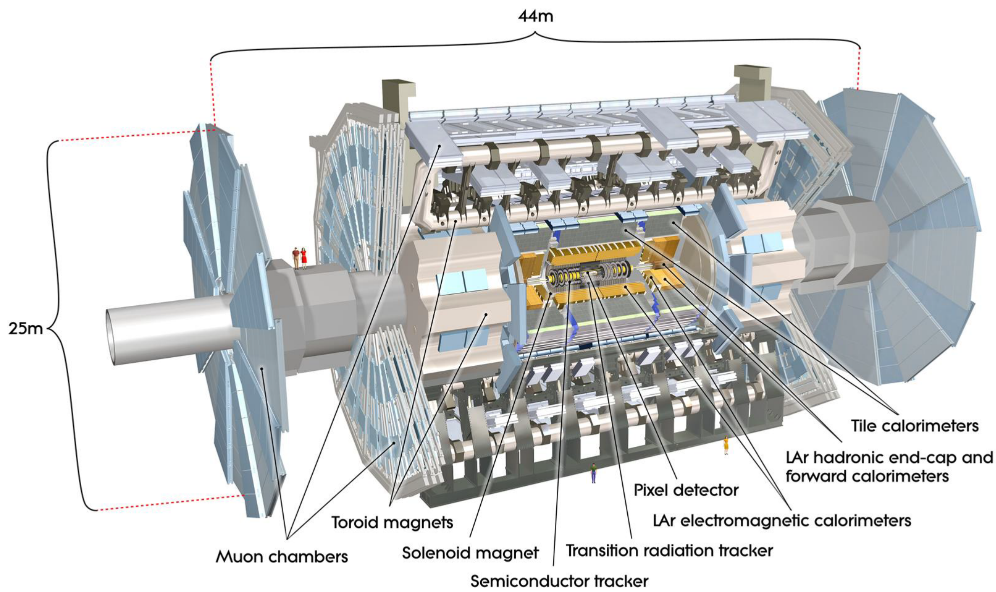


Figure 2.3 Cut-away view of the ATLAS detector.[31]

2.2.1 Inner Detector

The ATLAS Inner Detector (ID), the inner most detector of ATLAS, is designed to provide hermetic and robust pattern recognition, excellent momentum resolution and both primary and secondary vertex measurements for charged tracks. It is contained within a solenoidal magnetic field of 2 T. It can measure tracks above a given p_T threshold (nominally 0.5 GeV) and within the pseudorapidity range $|\eta| < 2.5$. It also provides electron identification over $|\eta| < 2.0$ and a wide range of energies (between 0.5 GeV and 150 GeV). It consists of three independent but complementary sub-detectors. At inner radius, silicon pixel layers (Pixels) and stereo pairs of silicon microstrip layers

(SCT) provide high-resolution pattern recognition capabilities. The pixel detector has the highest granularity in the vertex region, good intrinsic accuracy of 10 μm in $(R-\phi)$ and 100 μm in z or R direction. The pixel detector has three layers during Run 1 and one extra layer called IBL is added during the upgrade for Run 2, these layers are expected to be crossed by the tracks. For the SCT detector, there are eight strip layers which are expected to be crossed by the tracks and its resolution is about 17 μm in $(R-\phi)$ and 580 μm in z or R direction. At larger radius, the transition radiation tracker (TRT) comprises many layers of gaseous straw tube elements interleaved with transition radiation material. Accuracy of its measurement is about 130 μm per tube$(R-\phi)$. The TRT provides continuous tracking with an average of 36 hits per track which can improve the performance of pattern recognition, momentum resolution and electron identification. The layout of ID is shown in Figure 2.4.

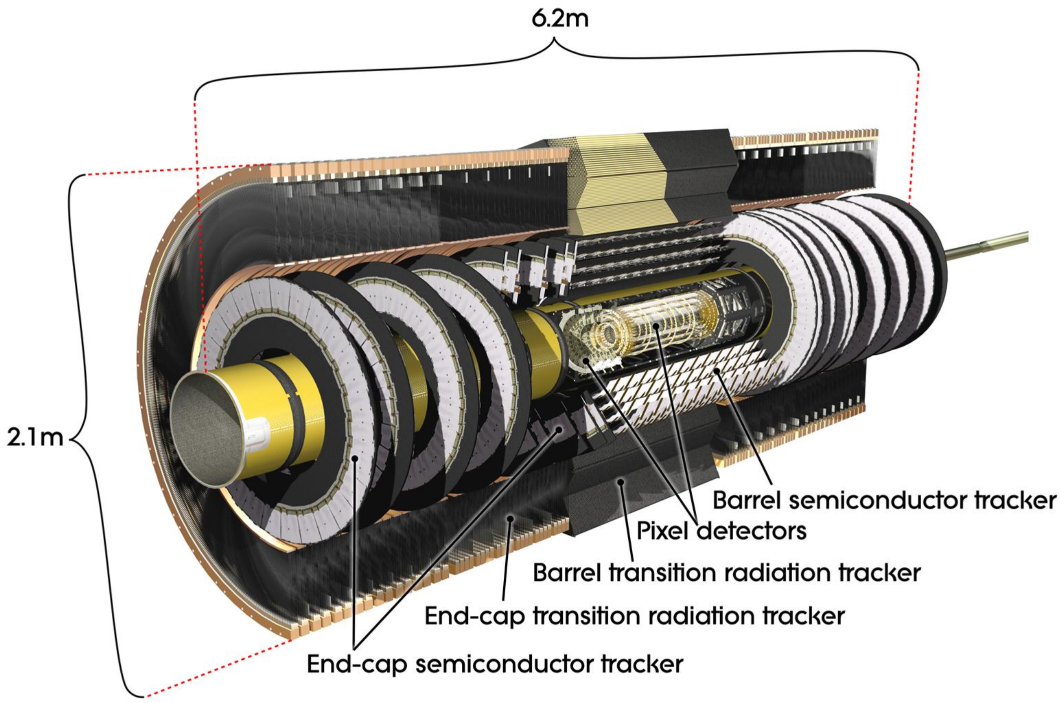


Figure 2.4 Layout of the ATLAS inner detector.[32]

2.2.2 Calorimetry

The ATLAS calorimeters are made up of a number of sampling detectors with full ϕ-symmetry and coverage around the beam axis. There are three cryostats, one barrel and two end-caps, close to the beam-axis. The barrel cryostat contains an electromagnetic calorimeter and the two end-cap cryostats each contain an electromagnetic end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC) and a forward calorimeter

(FCal). All these calorimeters choose liquid argon as the active detector medium for its intrinsic linear behavior, stability of response over time and intrinsic radiation-hardness. The electromagnetic calorimeters are lead-liquid argon detectors with accordion shape absorbers and electrodes. This geometry allows the calorimeters to have several active layers in depth, there are three layers in the precision-measurement region ($0 < |\eta| < 2.5$) and two layers in the higher η region ($2.5 < |\eta| < 3.2$). The hadronic calorimeter includes tile calorimeter, the LAr hadronic end-cap calorimeter and the LAr forward calorimeter. The tile calorimeter is composed of three parts, one central barrel and two extended barrels, the barrel part covers the range $\eta < 1.0$ and the two extended barrels cover the range $0.8 < |\eta| < 1.7$. It is a sampling calorimeter using steel as absorber and scintillating tiles as active material located behind the liquid argon electromagnetic calorimeter. The HEC is a copper/liquid-argon sampling calorimeter which covers the range $1.5 < |\eta| < 3.2$. It includes two wheels in each end-cap cryostat: a front wheel and a rear wheel. The wheels are cylindrical and each wheel consists of 32 identical wedge-shaped modules. The FCal is located in the same cryostats as the end-cap calorimeters and its coverage is $3.1 < |\eta| < 4.9$. The FCal modules are located at high η, they are exposed to high particle fluxes and this results in a design with very small liquid argon gaps. Each FCal is split into one electronmagnetic module and two hadronic modules: copper, optimized for electromagnetic measurements and tungsten, for hadronic measurement. Layout of the ATLAS calorimeter is shown in Figure 2.5.

2.2.3 Muon Spectrometer

The Muon Spectrometer is designed to detect particles exiting the barrel and end-cap calorimeters (mainly muons), it forms the outer part of the ATLAS detector. The measurement is based on the bending power of the large superconducting air-core toroid magnets, the barrel magnets (1 - 5.5 T) covers the range $|\eta| < 1.4$ while the end-cap magnets (1 - 7.5 T) covers the range $1.6 < |\eta| < 2.7$. The spectrometer includes four different types of chambers: monitored drift tube (MDT), cathode strip chambers (CSC), the resistive plate chambers (RPC) and thin gap chambers (TGC). MDT and CSC are precision-tracking chambers while RPC and TGC are trigger chambers.

The overview of the ATLAS muon spectrometer is shown in Figure 2.6. MDT provides precision momentum measurement and covers the pseudorapidity $|\eta| < 2.7$ except the innermost end-cap layer where the coverage is limited to $|\eta| < 2.0$. CSC are multi-wire proportional chambers with cathode planes segmented into strips in orthogonal directions. It covers the larger presudorapidity region and replace MDT chambers in the first

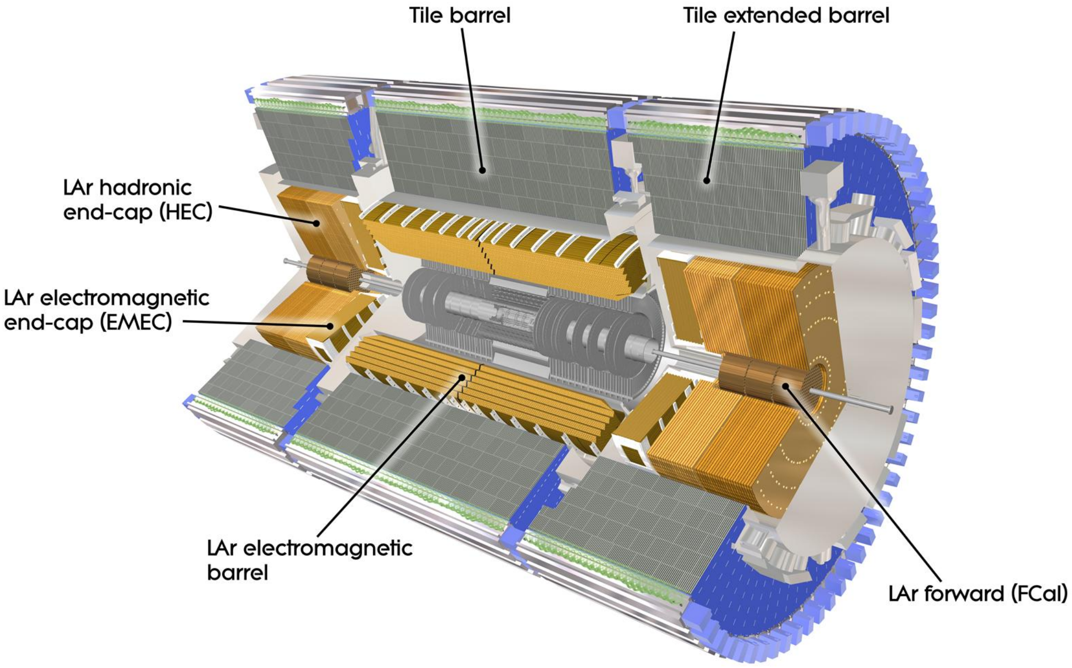


Figure 2.5 Layout of the ATLAS Calorimeter.[33]

layer due to the high particle flux. The trigger chambers cover the range $|\eta| < 2.4$ which provide the measurement of second coordinate (ϕ) as well. TGC chambers are assembled in the end-cap region while the RPC chambers are in the barrel region.

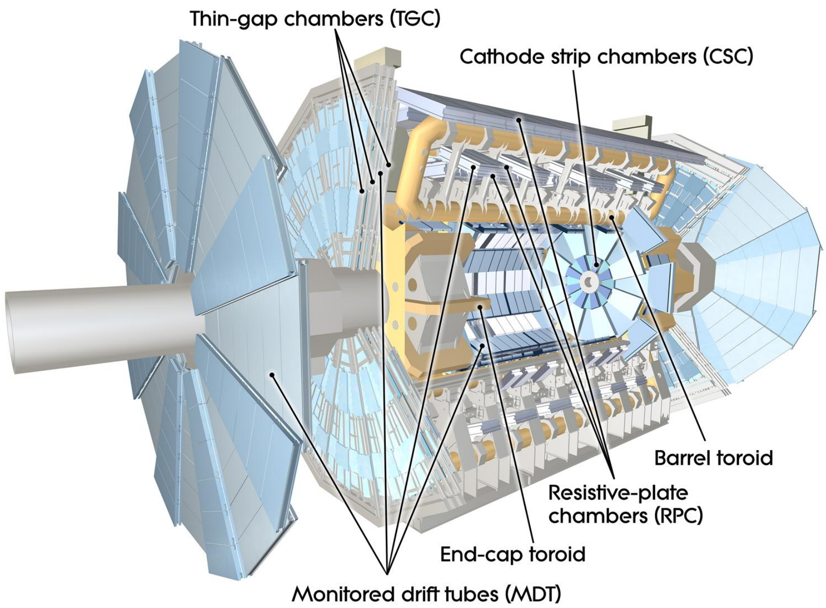


Figure 2.6 Overview of the ATLAS muon spectrometer.[34]

2.2.4 Triggers

As described in previous chapter, at the LHC, bunches of up to 10^{34} protons will collide millions of times per second which corresponds to a very high rate of 40 MHz. But the event data recording is limited to about 200 Hz due to technology and resource limitations. The ATLAS trigger system is designed to record events at approximately 200 Hz from the 40 MHz bunch crossing rate.

The trigger consists of three levels of event selection: Level-1 (L1), Level-2 (L2) and event filter (EF). The L2 and event filter together form the High-Level Trigger (HLT). The L1 trigger uses signatures of high-p_T muons, electrons, photons, jets and τ-leptons decaying into hadrons as well as large missing transverse energy (E_T^{miss}) and large transverse energy. The selection is based on information from a subset of detectors: the RPC and TGC for high-p_T muons, and all the calorimeter sub-systems for electromagnetic clusters, jets, τ-leptons, E_T^{miss} and large total transverse energy. The information from L1 muon and calorimeter triggers are processed by the Central Trigger Processor (CTP), which makes L1 accept decision. Events passing L1 trigger are transferred to the next stages of detector-specific electronics and subsequently to the data acquisition. In each event, one or more Regions-of-Interest (RoI, i.e. geographical coordinates) are defined by the L1 trigger within which the selection process has identified interesting features. The RoI data include information on the type of feature identified and the criteria passed. It is subsequently used by HLT. The maximum L1 rate is about 75 kHz (upgradeable to 100 kHz). The L2 trigger is seeded by the RoI information on coordinates, energy, and type of signatures which is provided by the L1 trigger. The L2 trigger reduces the event rate to below 3.5 kHz, with an event processing time of about 40 ms, averaged over all events. The EF uses offline analysis procedures on fully-built events to further select events down to about 200 Hz, with an average event processing time of the order of four seconds. A scheme of the ATLAS trigger system is shown in Figure 2.7.

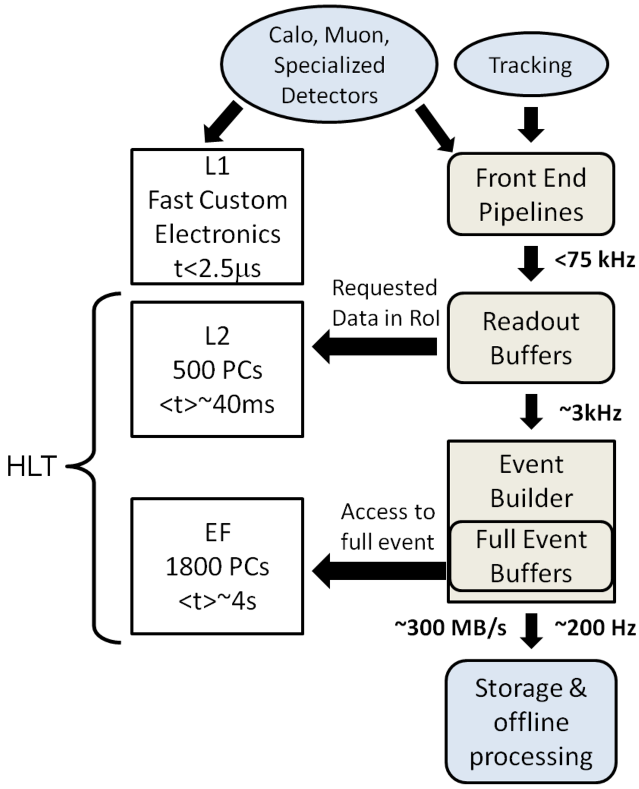


Figure 2.7 Schematics for the ATLAS trigger system.[35]

Chapter 3 Data and MC

3.1 Data

Data is the most significant ingredient for any physics analysis at the LHC. The first long run of pp collisions at LHC took place during the years of 2011 and 2012. Center-mass-energy of the pp collision is 7 TeV during 2011 and 8 TeV during 2012. The delivered and recorded integrated luminosity in the ATLAS detector during Run 1 are shown in Figure. 3.1, ATLAS record data delivered from the LHC with a quite high efficiency, all the subsystems of the ATLAS detector performed well during Run 1. The search for

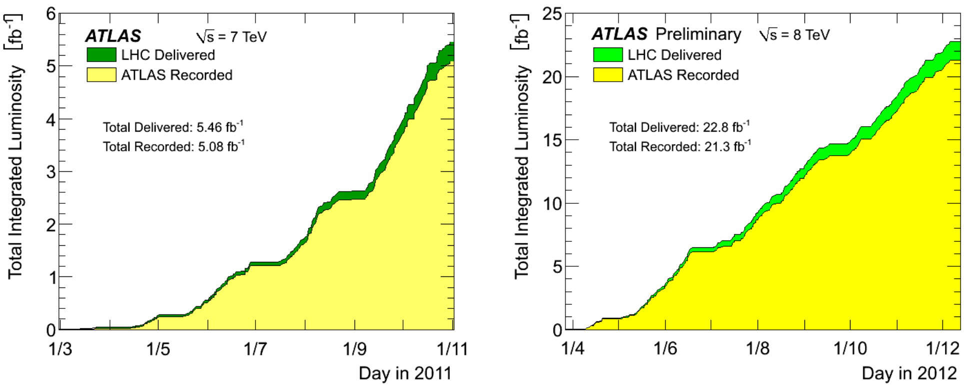


Figure 3.1 The delivered and recorded integrated luminosity in the ATLAS detector during 2011(left) and 2012(right).[36]

$W^{\pm}W^{\pm}W^{\mp}$ production and study of anomalous quartic gauge couplings which will be described later is performed with data collected during Run 1.

There is a long shut down of the LHC between 2013 and 2014, after two years of upgrade for the machine, the second long run starts from 2015, center-mass-energy of the pp collisions during Run 2 is 13 TeV. The delivered and recorded integrated luminosity in the ATLAS detector during Run 2 are shown in Figure. 3.2. The search for Doubly Charged Higgs described in Chapter 6 is performed with data collected during 2015 and 2016.

Apart from the efficiency of data taking, another important feature in the LHC collision data is the multiple interactions per bunch, the integrated luminosity as a function of mean number of interactions per crossing for Run 1 and Run 2 are also shown in Figure 3.3, the mean number of interactions per crossing corresponds the mean of the poisson distribution on the number of interactions per crossing calculated for each bunch. The protons are collided in bunches, increasing the number of protons per bunch will decrease the bunch spacing and increase the luminosity which may result in multiple

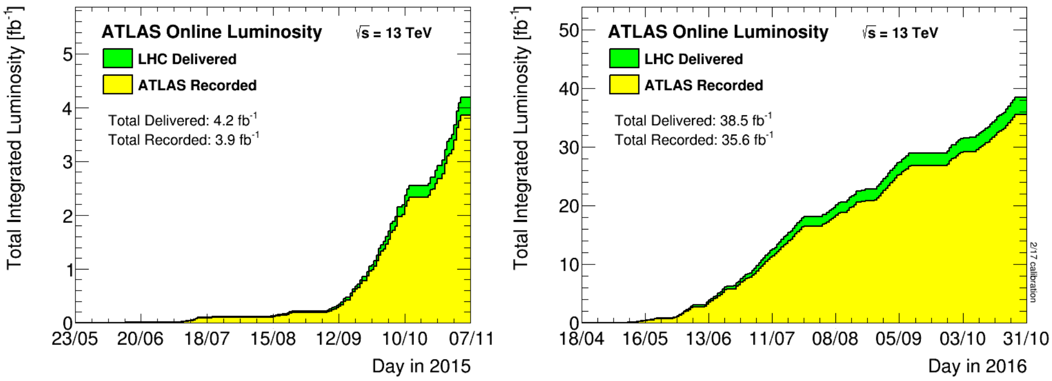


Figure 3.2 The delivered and recorded integrated luminosity in the ATLAS detector during 2015(left) and 2016(right).[37]

events recording as one, this is the so-called "pile-up" effect. Pileup affects mostly the soft terms of the E_T^{miss} and enlarge the uncertainties in many physics analysis. A pileup reweighting procedure needs to be performed on Monte Carlo samples to make the distribution of average number of interactions per bunch consistent between data and MC.

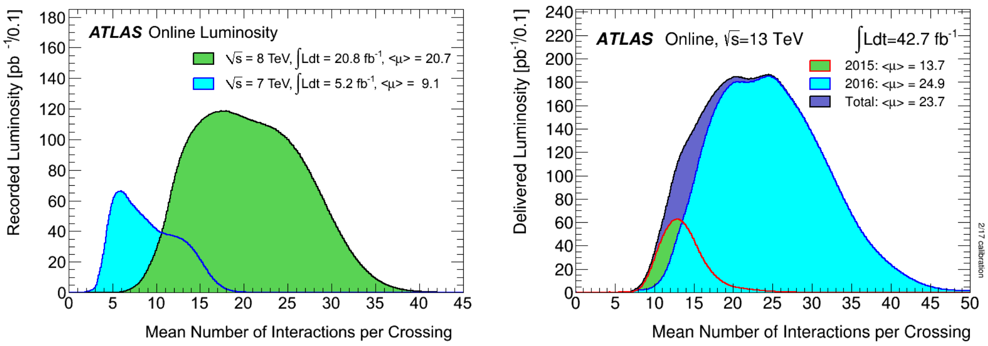


Figure 3.3 Intergated luminosity as a function of mean number of interactions per crossing for Run 1(left) and Run 2(right).[36] [37]

3.2 Monte Carlo Simulation

Monte Carlo is a calculational technique which makes use of random numbers. This technique is widely used in particle physics in various aspects such as simulation of detector response and theoretical models (SM, SUSY, etc.). In particle physics, Monte Carlo simulation starts with event generation, then the events at truth level are sequentially processed with the simulation of detector response, the emulation of the electronic read-out (digitization). Finally the simulated events are in exactly the same format as real data and are reconstructed with the same offline software. To perform a Monte Carlo simulation, Parton density functions (PDFs) evaluated with quantum field theory from experimental data are indispensable for cross section calculations in hadron colliders. In the QCD parton model, the scattering process between hadrons is described by the partons carrying a varying fractions of momenta of their parent hadrons. The partonic structure of a hadron is probed via scattering process such as the deep inelastic scattering between leptons and hadrons, the parton density functions are determined through a fit to the experimental data. There are various generators with evaluated PDFs provided by theoretical groups for Monte Carlo simulation, these generators use different techniques for event generation but later on the detector simulation and digitization as well as event reconstruction are identical for all used generators.

3.2.1 Event Generation

At the LHC, the simulation of a *pp* collision can be split into several steps:

- hard process
- parton shower
- hadronization
- underlying event
- unstable particle decays.

Most of the *pp* collisions are produced with only a few soft hadrons and the event goes out along the beam axis while only a tiny fraction of events contain a high momentum transfer of interest. It is impossible to simulate all possible collisions thereby the simulation focus on the hard process of interest. This step starts with the calculation of scattering matrix elements of a particular hard process. Then the hard process is simulated together with Parton Distribution Functions (PDFs) which describe partons coming into the process. Parton shower describes what happens to the incoming and outgoing partons. Partons involved in hard processes are colored particles, quarks and gluons, scattered electric charges radiate photons while scattered color charges radiate gluons however gluons are colored which results in new radiations from the gluons. The new radiations from gluons lead to extended shower and phase space filled up with soft gluons. The parton shower evolution starts from the hard process and works downwards to lower and lower momentum scales to a point where perturbation theory breaks down. Hadronization models take account of the confinement of a system of partons into hadrons. Initially the uncolored proton has had a colored parton taken out of it and thus the proton is left colored, there is a high probability that there will be other interactions besides the hard interaction which give rise to the underlying event. Underlying event is made up of secondary interactions between proton remnants, it produces soft hadrons everywhere in the event which will contaminate the hard process that is already simulated. Since many of the hadrons are unstable particles and go on to decay, simulation of the secondary decays is essential as well. Sketch of a hadron-hadron collision is shown in Figure 3.4.

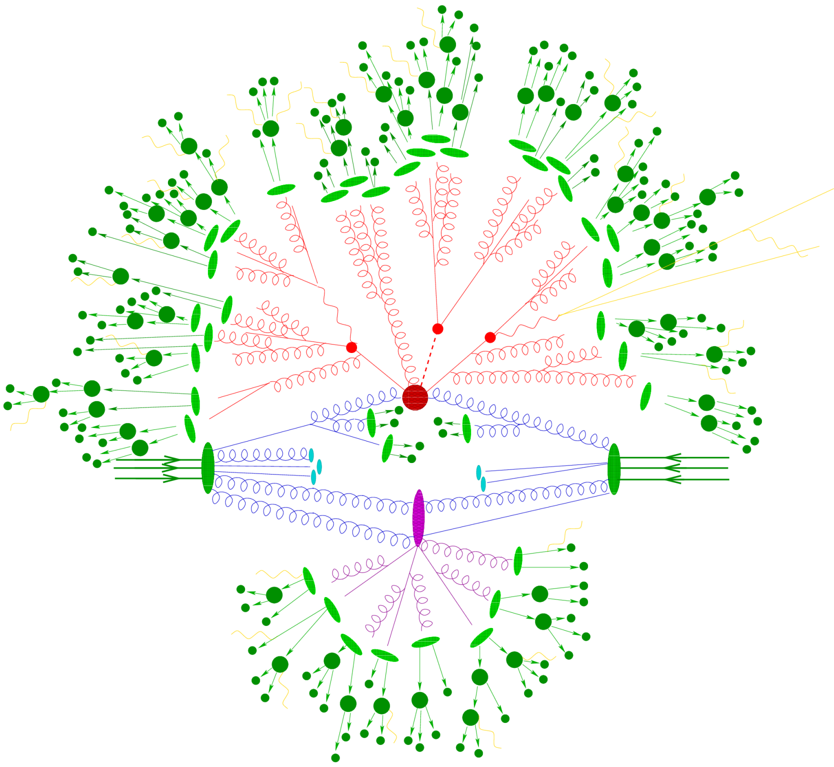


Figure 3.4 Sketch of a hadron-hadron collision simulated by MC event generator. The two large green blobs represent protons, the red blob in the center representing the collision. The tree like structure surrounded the red blob indicate Bremsstrahlung. Light green blobs represent parton to hadron transitions and the dark green blobs indicate hadron decays, the yellow lines refer to photon radiations. Partons from the initial protons radiate gluons and interact producing the parton shower which eventually hadronize and decay into final state particles.The purple blob indicate secondary decays from the renaments of protons, the gluons radiated from protons interact producing parton shower which will hadronize and decay into final state particles as well.[38]

There are three main general purpose generators: HERWIG[20], PYTHIA[22][14] and SHERPA[27] whereas there are also various generators dedicated to specific studies such as AlpGEN[39], MadGraph[3] and MC@NLO[40].

- HERWIG[20]: The HERWIG event generator was originally developed in Fortan and later updated to a C++ version, the HERWIG++. It is mainly used for simulation of lepton-lepton, lepton-hadron and hadron-hadron collisions. The parton shower approach is used to simulate initial-state and final-state QCD radiations. The underlying event is simulated with an eikonal multiple parton-parton scattering model, the formation of hadrons from quarks and gluons produced in the parton shower is described with the cluster hadronization model. Hadron decays are simulated using matrix elements which include spin correlations and off-shell effects.

- PYTHIA[22][14]: Similar to HERWIG, there are two versions of PYTHIA developed with Fortan, PYTHIA 6, and with C++, PYTHIA 8. It is a general purpose generator which can be used to simulate events in a wide range reactions, within and beyond the Standard Model. The Lund string model is used for hadronization and a highly developed multiple-interaction model is used for the underlying event. All the elements of event generation process such as hard processes, initial and final state parton showers underlying events and beam renaments as well as fragmentation and decays, this feature is used to simulate multiple collisions occurring simultaneously.

- SHERPA[41]: SHERPA is a general purpose event generator which is developed with C++ from the very beginning. It has built-in generators for the calculation of hard scattering processes within the Standard Model and various new models beyond the Standard Model. A dipole formulation is used for parton showering, and a cluster model for hadronization. A multiple interaction model, which is based on that of PYTHIA but different in some aspects, is used to account for underlying events. The fragmentation of partons into primary hadrons is described using a phenomenological cluster-hadronization model. Form factor model and matrix elements are used to allow for spin correlations.

- MadGraph[3]: MadGraph starts with the Feynman diagram of a physics process to calculate the matrix elements. The calculation can be done to any order at the tree level in principle but it turns out to be extremely difficult due to rapidly growing number of diagrams at high orders. The matrix elements are then used to simulate event of given process. Parton shower, hadronization and underlying event are carried out in other generators such as PYTHIA.

- AlpGEN[39]: AlpGEN is a tree-level matrix element calculator for a fixed number of partons in final state for hadron collisions. AlpGEN focuses on configurations with high jet multiplicities. It describes multi-partonic final states at leading order without any loops in perturbation theory and is based on exact evolution of Feynman diagrams in QCD and EW interactions. The calculation of matrix elements simulates the process with high jet multiplicities more precise than the parton shower approach.

- MC@NLO[40]: MC@NLO is developed with Fortan and it combined the Monte Carlo generators (HERWIG and HERWIG++) with Next-to-Leading-Order calculations in QCD. Various physics processes are included in the package and spin correlations are included for all processes except the ZZ production.

During actual physics analysis such as the Search for Doubly Charged Higgs which will be described in Chapter 6, various Monte Carlo event generators are used to simulate physics processes within and beyond the Standard Model.

3.2.2 Detector Simulation

The events simulated by the Monte Carlo event generators are then delivered to the detector simulation which simulate the response of particles interacting with the detector materials such as hits in the tracking detector and energy deposits in the calorimeter. The simulation software in ATLAS is based on GEANT4[16] and it is integrated into the common analysis framework of ATLAS, Athena. The detector simulation starts with geometry description of the detector, the full accurate model of the ATLAS, which contains parameters of detectors and the magnet fields, is fed to the GEANT4 for the simulation. Flexibility of GEANT also allows for simulation of single sub-detector or additional volumes added to ATLAS. The simulated information such as tracks, energy deposits etc., are then reconstructed to form physics objects which are described in Chapter 4.

Chapter 4 Physics Objects

At ATLAS, studies of many interesting physics processes utilize the information from various physics objects, i.e. electrons, muons, jets and missing transverse energy. These physics objects are reconstructed based on the information from various detectors and triggers.

4.1 Electron

Electrons and positrons, collectively referred to as electrons, give rise to tracks in the inner detector and energy deposits in the electromagnetic calorimeter. Electron candidates are reconstructed using such information and these candidates are further selected against background electrons, such as hadrons and electrons originating from heavy flavor hadron decays. Electron reconstruction in the central region ($|\eta| < 2.7$) proceeds in several steps[42] [43]:

- Seed-cluster reconstruction: Starts from the energy deposits in the EM calorimeter, a $sliding - window$ algorithm searches for seed cluster of longitudinal towers with total transverse energy greater than 2.5 GeV. The clusters are formed around the seed and duplications are removed. Kinematics of the clusters are reconstructed depending on the position of the cluster in the EM calorimeter. The efficiency of the cluster search ranges from 95% at E_T =7 GeV to more than 99% at E_T =15 GeV.

- Track Reconstruction: Track reconstruction starts with pattern recognition. The ATLAS pattern recognition uses the pion hypothesis for energy loss due to interactions with the detector material and the algorithm is modified to allow for up to 30% energy loss at each intersection of the track with the detector material. A track seed is reconstructed using the hits in the three layers of the silicon detector first. If it can not be successfully extended to a full track of at least seven hits using the pion hypothesis and it falls into one of the EM cluster region of interest, the new pattern recognition using electron hypothesis with larger energy loss will be performed. The track candidates are then fit either with the pion hypothesis or the electron hypothesis. If the fit with pion hypothesis fails, a second fit with the electron hypothesis will be performed. This electron-oriented algorithm improves the electron reconstruction and has minimum interference with the main track reconstruction.

- Electron Specific Track Fit: The obtained tracks are then matched to the EM clusters using the distance between the track and the cluster. The energy loss due to bremsstrahlung and the number of hits in silicon detector are also taken into account in the matching, tracks with more precise hits but loosely associated to clusters are refit with the non-linear bremsstrahlung effects accounted.

- Electron Candidate Reconstruction: A similar matching between tracks and EM clusters as described above but with strict conditions. If there are multiple tracks fulfilling the matching condition, one track is chosen as primary track based on an algorithm using the cluster-track distance, number of pixel hits and the presence of a hit in the first silicon layer. Electron candidates without any associated precision hit tracks are consider as photons. The energy of the electrons is calibrated using multivariate techniques based on Monte Carlo simulated samples.[43]

Reconstruction efficiency is one important feature to study the performance of the reconstruction. It is defined as the ratio of the number of reconstructed electrons with a matching track passing the track quality requirements to the total number of EM clusters from electrons. The measurement starts with EM clusters since the efficiency of detecting an energy cluster in EM calorimeter with the siding window algorithm is found to be more than 99% for $E_T > 15$ GeV[42]. Tag and probe method (a generic method to measure object efficiency by exploiting di-object resonances such as Z or J/Ψ) is applied with $Z \to ee$ events which requires the invariant mass of the tag-probe pair to be close to the mass of Z boson to separate signal from background. The tag is selected with strict requirements while the probe is loosely selected to include all EM clusters. The reconstruction efficiency is shown in Figure 4.1

The reconstructed electron candidates can be real electrons or background electrons from hadronic jets or converted photons, so electron identification (ID) is performed. Various quantities related to electrons including shower shape, number of hits in inner detector, properties of the tracks, etc are used in the electron ID to separate real electrons from background electrons. For Run 2, there are some changes to the input variable, the number of hits of the insertable B-layer (IBL) is used to separate electrons from converted photons. Change of TRT gas in Run 2 also introduced modifications to the electron ID. The ID algorithm for Run 2 is based on MC simulation samples, $Z \to ee$ samples are used for signal electrons while di-jet events are used for background electrons, in addition to $J/\psi \to ee$ and minimum bias events at low E_T electron ID[42].

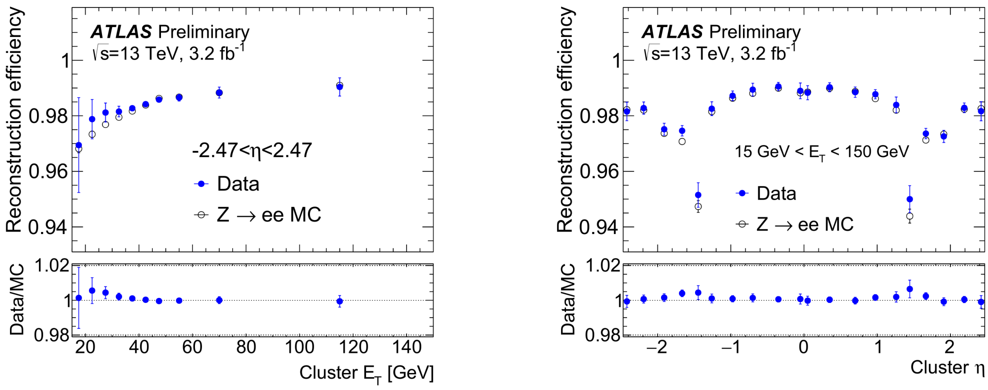


Figure 4.1 Measured reconstruction efficiencies as a function of E_T(left) and $|\eta|$(right) for the 2015 dataset. The shown uncertainties are statistical plus systematic.[42]

Baseline ID algorithm for Run 2 is Likelihood-based method using multivariate analysis (MVA) technique. The likelihood is built based on the probability density functions (PDF) of signal and background, the statistic $d_{\mathcal{L}}$ is then reconstructed and further fed to the MVA to discriminate signal from background:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \qquad \mathcal{L}_{S(B)(\vec{x})} = \prod_{i=1}^{n} P_{s(b),i(x_i)}, \tag{4.1}$$

where $\vec{x}$ is the vector of discriminating variables, $P_{s(b),i(x_i)}$ is the signal (background) PDF of the i^{th} variable and n is the number of discriminating variables. Three different levels of identification operating points are developed for the electron ID: $Loose$, $Medium$ and $Tight$. The operating points are defined such that the samples selected by one is the subset of one another which means electrons selected by $Medium$ are all selected by $Loose$ and $Tight$ electrons are all selected by $Medium$. These three working points are chosen according to the optimization for signal efficiency and background rejection where various discriminating variables are adopted. Table 4.2 summarized the variables used in the optimization. Since the electrons' shower shapes depend on amount of material the electrons passing through, thus depend on the pseudorapidity of the electrons. So the ID operating points are optimized in different $|\eta|$ and E_T bins. The combined electron identification and reconstruction efficiencies measured with $Z \to ee$ events as a function of E_T and η are shown in Figure 4.3

In addition to the LH method, a cut-based method using a set of rectangular cuts on the discriminating variables used during Run 1 is also developed as a cross check for Run 2. The cut-based method also developed three operating points: $Loose$, $Medium$ and

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

Figure 4.2 Discriminating variables for the optimization of electron identification.

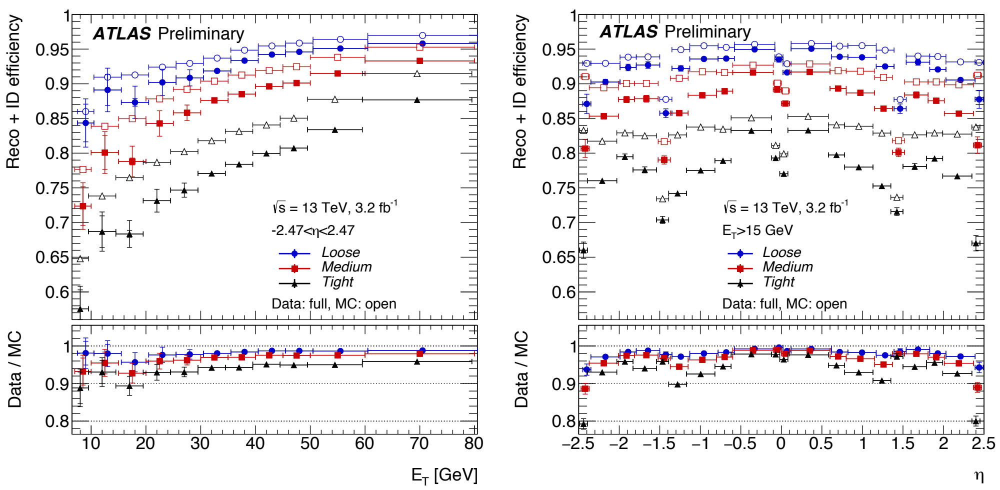


Figure 4.3 Combined electron identification and reconstruction efficiencies measured with $Z \to ee$ events as a function of E_T (left) and η (right).[42]

$Tight$. The $Loose$ operating point is based on the information from hadronic calorimeter and first two layers of the EM calorimeter. The $Medium$ adds information from TRT, transverse impact parameter and the third layer of the EM calorimeter. The $Tight$ operating point includes track matching variables such as $\frac{E}{P}$ and $\Delta\phi_2$ in addition to tighter cuts on the variables of $Medium$ operating point. The LH working points perform well during the Run 2 data-taking and are adopted in various Run 2 physics analyses.

Precise energy measurement of electrons is essential for various physics analyses. Therefore a calibration procedure is performed. $Z \rightarrow ee$ and $J/\Psi \rightarrow ee$ events are used in the study due to their large cross section and purity. Various corrections are applied to account for the response difference between data and simulation. After the initial calibration, only small mis-calibration remains which is only around 0.75% level. Therefore energy scale and resolution corrections are taken into account to deal with the mis-calibration. The energy mis-calibration is defined as the difference in response between data and simulation parameterized as:

$$E^{data} = E^{MC}(1+\alpha), \tag{4.2}$$

where α represents the deviation from optimal calibration. Electron resolution correction is derived under the assumption that the resolution curve is well modeled by the simulation up to a Gaussian constant term:

$$(\frac{\delta_E}{E})^{data} = (\frac{\delta_E}{E})^{MC} \otimes c. \tag{4.3}$$

The α and c are determined via a χ^2 minimization method using the distributions of invariant mass with scale and resolution perturbations. The measured α and c are shown in Figure 4.4.

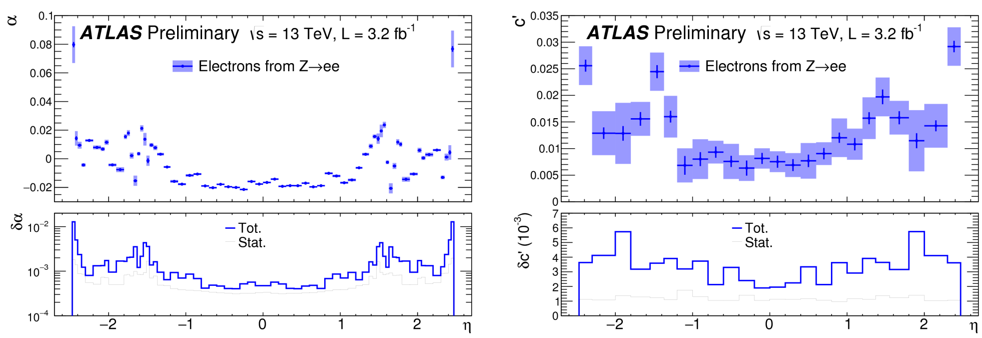


Figure 4.4 Energy scale factor α (left) and additional constant term c (right) for energy resolution from $Z \rightarrow ee$ events as a function of η. Uncertainties on the top panel are full uncertainties while the uncertainties on the bottom panel are statistical only.[43]

Figure 4.5 shows the invariant mass distribution of the Z electron pair in data and MC with the energy corrections applied. A good agreement is observed.

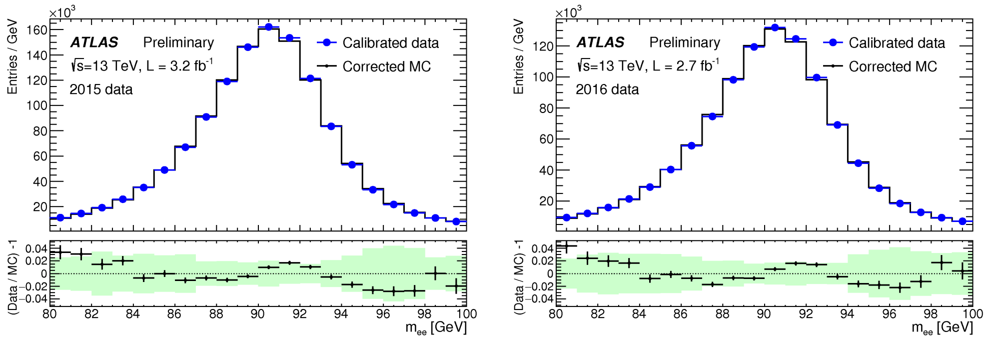


Figure 4.5 Invariant mass distribution of the Z electron pair in data and MC with the energy corrections applied.The distributions for the data are shown without applying any background subtraction. Plot on the left is with 2015 data while the one on the right is with 2016 data. Error bands in the plots stand for total uncertainties.[43]

4.2 Muon

Muons are important for the physics analyses at ATLAS experiment such as the discovery of the Higgs boson and the measurement of its properties. Muon reconstruction is performed independently in ID and MS first and then combined to form the muon tracks.

The muon reconstruction in ID is similar to the electron reconstruction in ID. The pattern recognition uses information collected from the Pixel and SCT detectors to generate track seeds and then these seeds are extended to the TRT, finally the tracks are refit with the information from all three detectors. Muon reconstruction in MS starts with a search for the hit pattern inside each chamber. The MDT segments are reconstructed through a straight-line fit to the hits found in each layer, RPC and TGC hits are used to measure the coordinate orthogonal to the bending plane, the CSC segments are reconstructed with a separate combinatorial search in the η and ϕ plane. The muon track candidates are then built by fitting together hits from segments in different layers. The hits associated with each track candidate are fitted using a global χ^2 fit. The χ^2 of the fit must satisfy certain criteria to accept a track candidate. The ID reconstruction and MS reconstruction are combined using various algorithms based on information from ID, MS and calorimeter. There are four kinds of muon candidates depending on which sub detectors are used in the reconstruction:

- Combined (CB) Muon: Muons reconstructed using tracks and hits information from both ID and MS. The MS hits may be removed or added during the global fit to improve the performance of the fit. The muons can be reconstructed with either inside-out pattern or outside-in pattern. Inside-out reconstruction indicates that the ID tracks are extrapolated and matched to the MS tracks. Outside-in refers to that the tracks are first reconstructed in MS and then extrapolated and matched to the ID tracks. Most muons are reconstructed with the Outside-in patterns.

- Segment-Tagged (ST) Muon: Muons reconstructed using ID track and MS information. An ID track is identified as a muon if the track extrapolation to MS is associated with at least one track segment in MDT or CSC chambers. ST muons are used when muons cross only one layer of MS chambers.

- Calorimeter-Tagged (CT) Muon: Muons reconstructed using ID track and information from calorimeter. An ID track is identified as muon if it can be matched to the energy deposit in the calorimeter with a minimum ionizing particle. This type of muons has the lowest purity but it covers the the region where the ATLAS MS is only partially instrumented.

- Extrapolated (ME) Muon: Muons reconstructed with only MS information and a loose requirement on the impact parameters. ME muons are mainly used to extend the acceptance of muon reconstruction into the region $2.5 < |\eta| < 2.7$ which is not covered by the ID. This type of muons are also called Stand-alone muons during Run 1.

In the muon reconstruction, if two muons share one same ID track, the preference is given to CB muons, then to ST and finally to CT muons. As for overlap with ME muons, track with better fit performance and larger number of hits is selected.
The reconstruction efficiency of muons is measured with tag and probe method for $|\eta| < 2.5$ region, a different methodology is applied for $2.5 < |\eta| < 2.7$ region where only ME muons are used. For muons in region $|\eta| < 2.5$, tag and probe method is applied based on $Z \to \mu\mu$ events which are selected by requiring two oppositely charged tracks with a di-muon invariant mass close to the Z boson. One of the tracks must be identified as a $Medium$ muon which is denoted as the "tag". The other one, the so-called "probe", is built independently with the loose criteria. There are three kinds of probes, MS tracks are used to determine the complementary efficiency of the muon reconstruction in the ID, CT muons are preferred at low transverse momentum for its higher rejection against backgrounds, ID tracks are used for measurements not directly accessible to CT muons. The reconstruction efficiency of muons as a function of $|\eta|$

and p_T is shown in Figure 4.6

Similar to electron identification, muon identification (ID) is also applied to the recon-

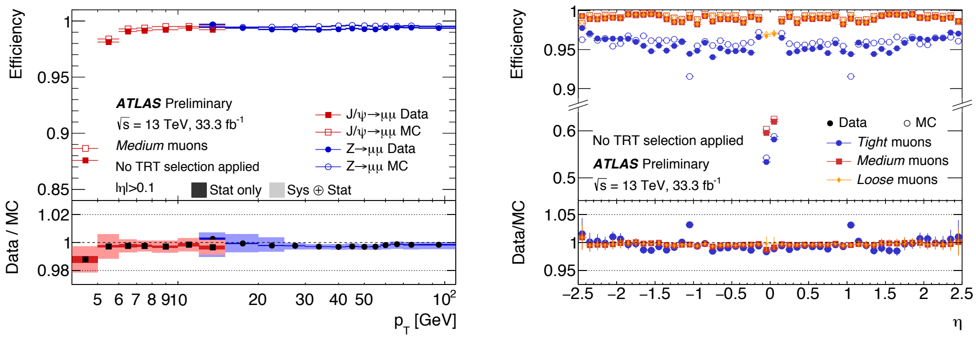


Figure 4.6 Reconstruction efficiency of muons for the $Medium$ identification algorithm as a function of $|\eta|$ (right) and E_T (left).[44]

structed muon candidates. $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ MC simulation samples are used for the muon identification study. Muon ID is performed by applying requirements on quantities that suppress background, mainly from pion and kaon decays, while selecting real muons with high efficiency. There are several variables with good discriminating power between real muons and background muons:

- $\frac{q}{p}$ significance, defined as the absolute value of the difference between the ratio of the charge and momentum of the muons measured in the ID and MS divided by the sum in quadrature of the corresponding uncertainties,
- ρ', defined as the absolute value of the difference between the transverse momentum measurements in the ID and MS divided by the p_T of the combined track.
- normalized χ^2 of the combined fit.

Four operating points are developed for the muon ID via optimization for signal efficiency and background rejection, the working points are denoted as $Loose$, $Medium$ or $Tight$ according to their signal efficiencies[45]:

- $Loose$: This working point provides good quality muons with maximum reconstruction efficiency. This is designed for the reconstruction of four-lepton final state Higgs boson. All types of muons are used, CT and ST muons are restricted to $|\eta| < 0.1$. In the region $|\eta| < 2.5$, about 97.5% of the $Loose$ muons are CB muons, about 1.5% are CT muons and the remaining 1% are ST muons.
- $Medium$: This is default muon selection in ATLAS, it minimize systematic uncertainties from reconstruction and calibration. Only CB and ME tracks are used.

- $Tight$: This selection maximize the muon purity at the cost of some efficiency. Only CB muons with hits in at least two stations of the MS and satisfying the $Medium$ criteria are used.

- $High - p_T$: This operating point aims to maximize the momentum resolution for tracks with p_T greater than 100 GeV, it is optimized for the searches for high mass Z' and W' resonances. CB muons with at least three hits in three stations of MS and satisfying the $Medium$ criteria are considered. The requirement on three stations of MS improves the p_T resolution greater than 1.5 TeV by about 30% while reducing the reconstruction efficiency by about 20%.

Various reasons can result in a deviation of the measured muon momentum such as misalignment of the muon chambers. Therefore the momentum from MC has to be smeared and shifted to match to data and the momentum scale and resolutions need to be calibrated. The distribution of the di-muon invariant mass from $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events together with a maximum likelihood fit are adopted to determine the corrections on momentum scale and resolution. Figure 4.7 shows the muon momentum scale and resolution as a function of η obtained from $Z \to \mu\mu$ events. Figure 4.8 shows

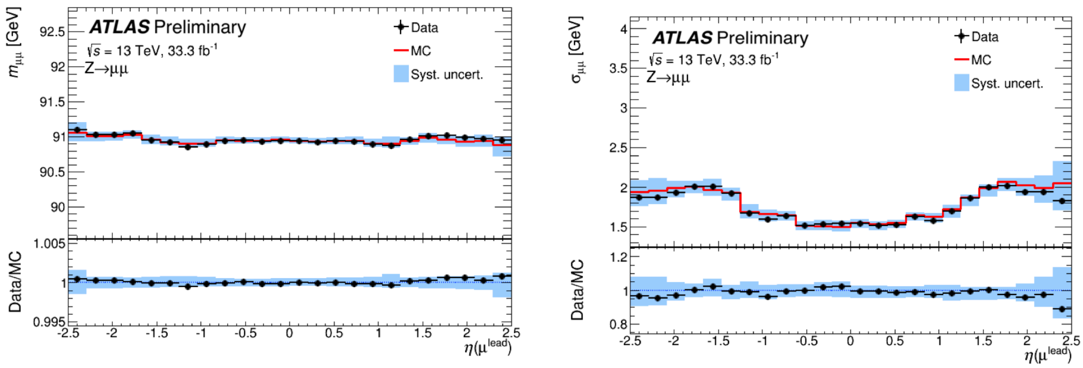


Figure 4.7 Muon momentum scale (left) and resolution (right) as a function of η obtained from $Z \to \mu\mu$ events.The systematic uncertainty is from the maximum likelihood fit.[44]

the momentum resolution of CB muons as a function of p_T.

4.3 Jet

Jets, collimated sprays of hadrons which are produced by hard process or softer interactions such as underlying events or additional pp collisions in the same proton bunch crossing, are the dominant physics objects arising in proton-proton collisions at the LHC. Jets with transverse momentum of more than a few GeV will leave significant signals in the ATLAS detector system thus good reconstruction of jets is essential for

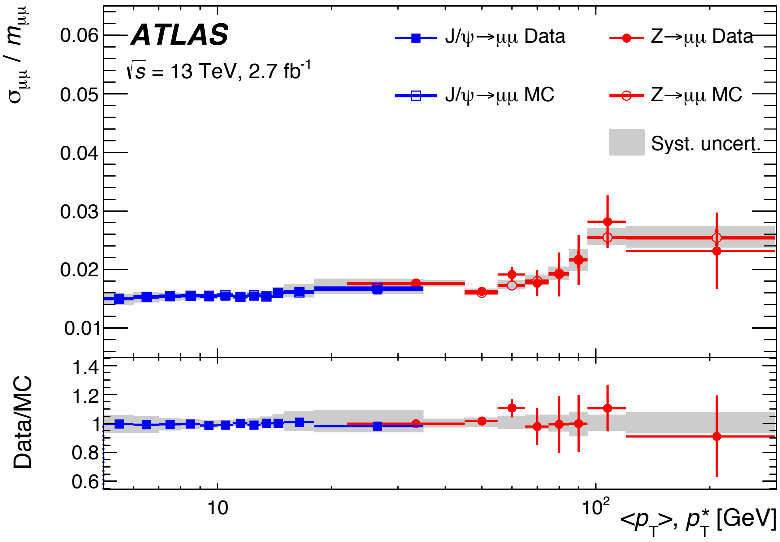


Figure 4.8 Muon momentum resolution as a function of p_T obtained from $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events.The error bars represent the statistical uncertainty while the bands show the systematic uncertainties.[44]

various studies of Standard Model processes, as well as search for new particles in the ATLAS experiment.

During Run 1, the ATLAS experiment uses either the calorimeter or the tracker to reconstruct hadronic jets and soft particles, these jets were then calibrated to the particle level using a Jet Energy Scale (JES) correction factor. Calorimeter is the most important detector of jet reconstruction for its high lateral granularity. The reconstruction of jets starts with a list of topological cell clusters, calorimeter towers and reconstructed tracks.

The topological clusters ($topo-clusters$) are three dimensional signal objects formed by calorimeter cells. They are seeded by cells whose absolute energy measurement exceed the expected noise by four times its standard deviation and then expanded both laterally longitudinally in two steps:

- First add all adjacent cells with absolute energy two standard deviation above the noise, then add all cells neighboring to the previous set.
- Separate clusters with more than two local energy maxima into separate $topo-clusters$.

A topological cluster is defined to have an energy equal to the energy sum of all included calorimeter cells, zero mass and a reconstructed direction calculated from the weighted averages (in case one cell is shared by two clusters, the weights stand for geometrical weights reflecting the distance from the cell to center of gravity of the two clusters) of the pseudorapidities and the azimuthal angles of the constituent cells. Calorimeter towers are static grid elements formed from calorimeter cells within $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. There are two types of calorimeter towers, one is with noise suppression which is built from the cells validated in a topological cluster, the other one is without noise suppression. Both types of calorimeter towers have equal energy comparing to the energy sum of all included calorimeter cells. Calorimeter cells in ATLAS are reconstructed on the basic electromagnetic energy scale. The calorimeter towers and topological clusters contain sum of the cells which means they are also reconstructed on this energy scale.

The reconstructed tracks of jets are built from charged particle tracks originated from the primary vertex of hard process, normally tracks with $p_T > 0.5$ GeV and $|\eta| < 2.5$ are considered. They provide independent detection of jet activity and measurements of jet properties. Tracks pointing to calorimeter can also be used to refine jet calibration.

Jets are reconstructed with anti-k_t algorithm[46] with radius parameter 0.4 or 0.6 using the basic objects. For each basic object ($topo - cluster$, calorimeter towers or track jets), several variables are defined as:

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta^2 R_{ij}}{R^2}; d_{iB} = k_{ti}^{2p}, \tag{4.4}$$

where k_{ti} is the transverse momentum, p is a parameter to define the relative power of the energy versus geometrical scales and i indicates the considered object while j denotes the adjacent object, d_{ij} stands for the distance between two entities i and j while d_{iB} stands for the distance between an entity and the beam. If the smallest distance is a d_{ij}, the two objects are then combined and the list of variables will be refreshed. If the smallest distance is a d_{iB}, the entity i will be treated as a complete jet and removed from the list. This algorithm normally merges close-by soft objects to the hard object and sometimes the hard objects are merged depending on the k_T and ΔR. Jets reconstructed from calorimeter towers are referred to "Tower" jets while jets built from topological clusters are denoted as "EM" jets. A brief scheme of jet reconstruction is shown in Figure 4.9

Similar to electrons or muons, because we are mainly interested in the hard scattering events, a so called "jet quality" cut is often applied to exclude the jets from backgrounds or noise (such as cosmic ray muons or large calorimeter noise). In ATLAS, two selection criteria are available: one is a loose selection with an efficiency of above 99% which is used in most physics analysis while the other one is a medium selection designed for selecting jets with high p_T.

There are several schemes for the jet energy calibration, the simplest one is called

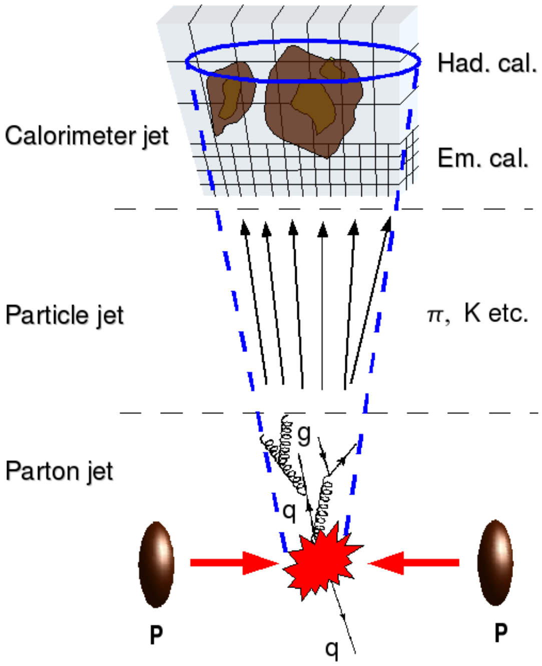


Figure 4.9 Scheme for jet reconstruction.

$EM + JES$ scheme which applies corrections as a function of the jet energy and pseudorapidity to jets reconstructed at the electromagnetic scale (EM). This scheme consists of three corrections: correction to reduce pile-up effect, correction on vertex ensuring that the jet is from the primary vertex instead of the center of the detector, corrections on jet direction and energy which is obtained through a comparison to MC.

The jet momentum resolution can be parameterized as three independent ingredients:

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \otimes \frac{S}{\sqrt{p_T}} \otimes C, \tag{4.5}$$

where N stands for effective noise, S stands for stochastic noise and C is a constant. The effective noise includes the electronic and detector noise, and the impact from pile-up. It is not (or only weakly) dependent on the jet p_T. Stochastic noise stands for the statistical fluctuations. The constant term implies the fluctuations are at a constant fraction of the jet p_T. Two methods based on di-jet events are adopted to evaluate the jet momentum resolution. One is the "di-jet balance method" which relies on the scalar balance between the transverse momenta of the two leading jets and measures the sensitivity of this balance to the presence of extra jets directly from data. The other method is called "bisector metho" which is based on a transverse balance vector defined as the vector sum of the two leading jets in di-jet events, this balance can be significantly fluctuated by any sources, therefore the sensitivity can be used to measure the jet energy resolution. Figure 4.10 shows the jet p_T resolution evaluated from di-jet events with two methods, $EM + JES$ scheme is adopted for calibration.

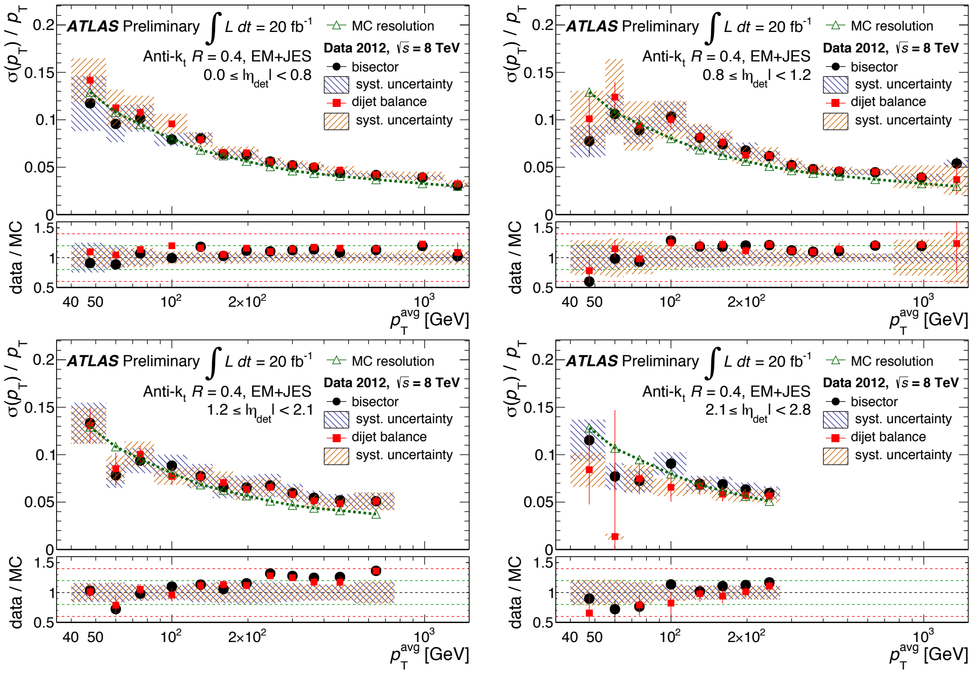


Figure 4.10 Jet energy resolution obtained for $EM + JES$ calibrated jets as a function of the jet transverse momentum in four regions of detector pseudorapidity: $|\eta| < 0.8$ (top left), $0.8 < |\eta| < 1.2$ (top right), $1.2 < |\eta| < 2.1$ (bottom left) and $2.1 < |\eta| < 2.8$ (bottom right). The green and red dashed lines indicate $1{\pm}20$ % and $1{\pm}40$ % respectively.[47]

4.4 Missing Transverse Energy

In a hadron collision event, the missing transverse momentum is defined as the event momentum imbalance in the plane transverse to the beam axis, where momentum conservation is expected. Such imbalance may imply the presence of unseen particles such as neutrinos or stable , weakly-interacting supersymmetry (SUSY) particles. The vector momentum imbalance in the transverse plane is obtained from the negative sum of the momenta of all particles detected in a pp collision and is denoted as E_T^{miss}. A good measurement of E_T^{miss} is crucial for Standard Model measurements involving neutrinos as well as searches for new particles. However, the limited detector coverage, finite resolution, dead regions and different kinds of noise are main challenges for the E_T^{miss} study. Although the ATLAS calorimeter is extended to large pseudorapidity range to minimize the impact from escaping particles, there are still inactive transition regions between calorimeters inside the ATLAS. Noise and dead read out channels as well as cosmic rays may result in fake E_T^{miss}, certain selection criteria is applied to suppress such fake E_T^{miss}[48]. The reconstruction of E_T^{miss} utilizes energy deposits in the calorimeters and muons reconstructed from the muon spectrometer. Low p_T tracks are used to cover the low p_T particles which are missing in the calorimeter whereas muons reconstructed from ID are used to cover muons in regions not covered by the MS. The E_T^{miss} reconstruction uses calorimeter cells calibrated according to the reconstructed physics objects to which they are associated in a chosen order: electrons, photons, hadronically decaying τ-leptons, jets and muons. Cells not associated to any physics objects are also took into account, denoted as $E_T^{miss,CellOut}$. The calorimeter term of E_T^{miss} is calculated as

$$\begin{aligned} E_{x(y)}^{miss,Calo} &= E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} \\ &+ E_{x(y)}^{miss,softjets} + (E_{x(y)}^{miss,Calo,\mu}) + E_{x(y)}^{miss,CellOut}, \end{aligned} \tag{4.6}$$

where each term is calculated from the negative sum of calibrated cell energies projected to x or y plane inside the corresponding objects. $E_{x(y)}^{miss,e}, E_{x(y)}^{miss,\gamma}, E_{x(y)}^{miss,\tau}$ are reconstructed from cells in clusters associated to electrons, photons and τ-jets from hadronically decaying τ-leptons, respectively. $E_{x(y)}^{miss,jets}$ is reconstructed from cells in clusters associated to jets with calibrated $p_T > 20$ GeV. $E_{x(y)}^{miss,softjets}$ is reconstructed from cells in clusters associated to jets with 7 GeV $< p_T <$ 20 GeV. $(E_{x(y)}^{miss,Calo,\mu})$ is energy loss of muons in calorimeter, it accounts for the double counting of muon energy deposited in calorimeters. $E_{x(y)}^{miss,CellOut}$ is calculated from the cells in $topo - clusters$ which are not included in the reconstructed objects. The muon term of E_T^{miss} is calculated from the momenta of muon tracks reconstructed with $|\eta| < 2.7$ as

$$E_{x(y)}^{miss,\mu} = - \sum_{muons} p_{x(y)}^{\mu}, \tag{4.7}$$

only well-reconstructed muons in the muon spectrometer with a matched track in the inner detector are considered. The E_T^{miss} and its azimuthal coordinate ϕ^{miss} are calculated as

$$\begin{aligned} E_{x(y)}^{miss} &= E_{x(y)}^{miss,Calo} + E_{x(y)}^{miss,\mu}, \\ E_T^{miss} &= \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}, \\ \phi^{miss} &= \arctan(E_y^{miss}, E_x^{miss}). \end{aligned} \tag{4.8}$$

Calibration of E_T^{miss} is done in a object-oriented way, the electrons are calibrated with the default electron calibration, photons are used as electromagnetic scale, τ jets are calibrated using LCW scheme and the jets are reconstructed with LCW scheme for soft jets($p_T <$ 20 GeV) and with LCW+JES scheme for hard jets (LCW is abbreviation for Local Cluster Weighting which calibrates the clusters before send them to the $anti - k_t$ algorithm, the jets are therefore at LCW+JES scale after the final calibration which is known as the $LCW + JES$ scheme). Performance of the reconstructed E_T^{miss} is studied with minimum biased events, di-jet events and $Z \to ll$ events. Among the three types of samples, $Z \to ll$ has clean signature and large cross section and the absence of genuine

E_T^{miss} means reconstructed E_T^{miss} in these events is a direct result of imperfection in the reconstruction process or detector response. The distribution of reconstructed E_T^{miss} is shown in Figure 4.11.

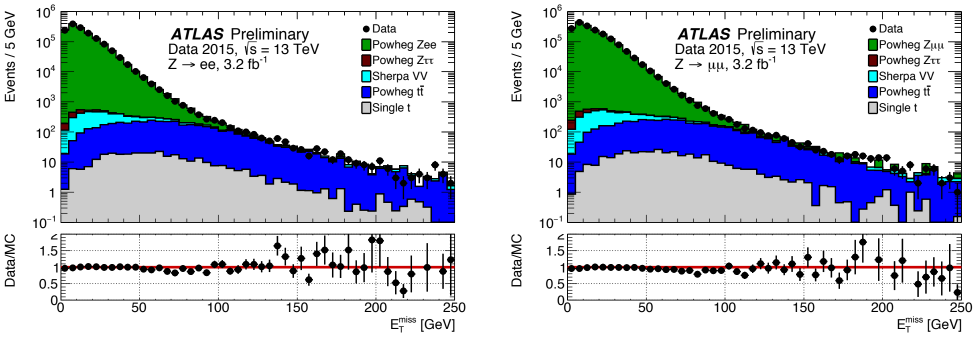


Figure 4.11 Distribution of E_T^{miss} measured in a data sample of $Z \rightarrow ee$(left) and $Z \rightarrow \mu\mu$(right), MC processes are normalized to their corresponding cross sections.[49]

The resolution of E_T^{miss} is also evaluated with $Z \rightarrow ll$ events. Since there is no genuine E_T^{miss} in $Z \rightarrow ll$ events, the true values of E_x^{miss} and E_y^{miss} are assumed to be zero. The resolution is estimated from the combined distribution of E_x^{miss} and E_y^{miss} in bins of $\sum E_T$, it is found that the E_T^{miss} resolution follows an approximately stochastic behavior as a function of the total transverse energy which can be described as

$$\sigma = k \cdot \sqrt{\sum E_T} \tag{4.9}$$

Figure 4.12 shows the resolution from data at $\sqrt{s} = 13$ TeV and MC for $Z \rightarrow \mu\mu$ as a function of the total transverse energy, data matches MC quite well.

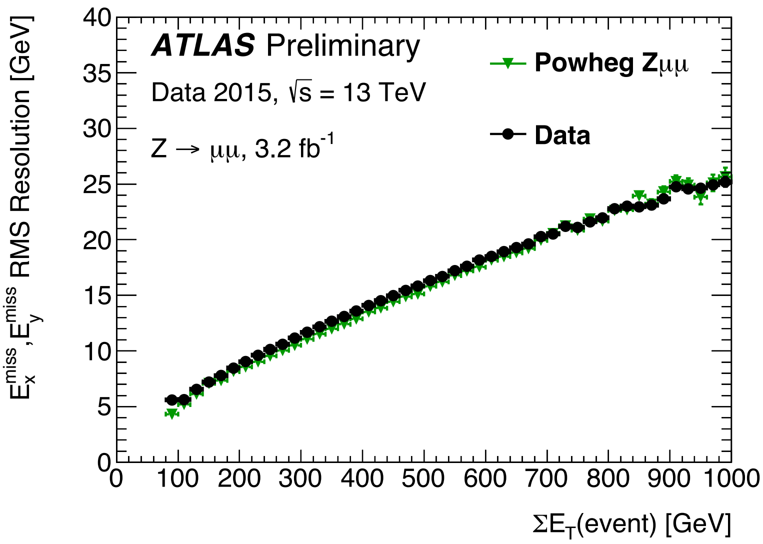


Figure 4.12 E_T^{miss} performance shown with Powheg+Pythia8 simulated $Z \rightarrow \mu\mu$ events overlaid with 2015 data as quantified by the resolution as a function of the $\sum E_T$ of the entire event.[49]

Chapter 5 Search for $W^{\pm}W^{\pm}W^{\mp}$ Production and aQGC Limits

The search for $W^{\pm}W^{\pm}W^{\mp}$ production and the study of anomalous quartic gauge couplings are performed using final state with three leptons and three neutrinos. The analysis aims at measuring the cross section of the production and setting limits on the anomalous quartic gauge couplings if no deviation from the SM observed. Data sample is full dataset of 2012 collected by the ATLAS detector with centre of mass energy of 8 TeV and corresponds to an integrated luminosity of 20.3 fb^{-1}.

Various MC samples are produced to facilitate the background estimation and aQGC study. Due to the large size of the samples, both data and MC samples are pre-selected ("Skimmed") to keep only events with three leptons and loose lepton identification criteria.

Apart from the signal process, various background processes may contaminate the signal region (a region designed to be enriched by signal events and with the best signal significance defined in Table 5.1) due to reasons like electron's charge misidentification or fake leptons originated from hadronic decays. Among the background processes, contributions with three or more leptons ($WZ, ZZ, t\bar{t}V, VVV$) or two leptons and one isolated photon ($Z+\gamma$) are estimated with MC simulation while other background processes are estimated with data which are described later in Section 5.2.

5.1 Event Selection

The selections of events are based on the WWW final states. The WWW candidates are selected by finding three good leptons (electrons or muons) in the events, applying event-level pre-selection and conducting a set of optimized WWW selections in the end. The selections of good leptons are described below.

- Muons are required to be combined muons reconstructed from the combination of an Inner Detector track and a Muon Spectrometer muon. Muons must have $p_T > 10$ GeV and $|\eta| < 2.5$. The inner detector tracks are constrained by the number of hits in each sub-detector, at least one hit in the pixel layer, at least four hits in SCT and less then three holes in all silicon layers. For muons in $0.1 < |\eta| < 1.9$, the sum of hits and outliers in TRT must be at least six and the number of outliers must be less than 0.9*(hits + outliers). For muons in $|\eta| < 0.1$ or in $|\eta| > 1.9$, the sum of hits and outliers in TRT must be at least five and the number of outliers must be less than 0.9*(hits + outliers). The tracking isolation requirement is $p_T^{\text{cone20}}/p_T < 0.04$ using the scalar sum of all tracks in a cone of $\Delta R < 0.2$. The calorimeter isolation (defined as the scalar sum of all calorimeter deposition in a cone of $\Delta R < 0.2$) is used, $E_T^{\text{cone20}}/p_T < 0.1$ for muons with $p_T > 20$ GeV and $E_T^{\text{cone20}}/p_T < 0.07$ for muons with $p_T < 20$ GeV. There are also constraints on the impact parameters, $\frac{|d_0|}{|\sigma_{d_0}|} < 3$ for the transverse impact parameter and $|z_0| \times \sin\theta < 0.5\ mm$ for the longitudinal impact parameters. In order to avoid duplicated muons, if any muons are reconstructed within $\Delta R(\mu, \mu) < 0.1$, the muon with lower p_T will be thrown away. The muon energy is corrected according to the muon energy scale measured in the data using $Z \to \mu\mu$ events. The muon momentum is also smeared to take into account of the difference between the data and the simulation. The Monte Carlo events are reweighted by the product of the reconstruction, identification, and trigger efficiency scale factors (corrections to modify the difference of efficiencies between MC simulation and data) for each muon.

- Electrons used in this analysis are reconstructed with the direction from tracks and energy deposits from calorimeter. The electrons must have $p_T > 10$ GeV, $|\eta| < 2.47$ and be outside of the EM calorimeter transition region ($1.37 < |\eta| < 1.52$). Electrons reconstructed by the standard calorimeter-based algorithm are selected (denoted as variable el_author in the samples). The electrons can not be reconstructed close to a known badly behaving calorimeter region. The tracking isolation requirement is $p_T^{\text{cone20}}/p_T < 0.04$ using the scalar sum of all tracks in a cone of $\Delta R < 0.2$. The calorimeter isolation used the scalar sum of all calorimeter deposition in a cone of $\Delta R < 0.2$, $E_T^{\text{cone20}}/p_T < 0.1$ for muons with $p_T > 20$ GeV and $E_T^{\text{cone20}}/p_T < 0.07$ for muons with $p_T < 20$ GeV, the calorimeter isolation is corrected according to the number of primary vertex of the event. $\frac{|d_0|}{|\sigma_{d_0}|} < 3$ is required for the transverse impact parameter and $|z_0| \times \sin\theta < 0.5mm$ is required for the longitudinal impact parameter. Similar to the muons, to avoid duplications, for any electrons reconstructed within $\Delta R < 0.1$, the electron with lower p_T is dropped. The electron energy is corrected according to the electron energy scale measured in the data using $Z \to ee$ events. The electron momentum in Monte Carlo is also smeared to take into account of the difference between the data and the simulation. The Monte Carlo events are reweighted by the product of the reconstruction, identification, and trigger efficiency scale factors for each electron.

- Jets are also selected with a certain criteria in this analysis. They are reconstructed with the anti-k_T algorithm with a parameter $\Delta R < 0.4$. They must have $p_T > 25$ GeV and $|\eta| < 4.5$. To suppress the jets from pile-up events, |JVF| $>$ 0.5 (Jet-Vertex Fraction) is required for jets with $p_T < 50$ GeV and $|\eta| < 2.4$. Some jets are tagged as b-jets with the MV1 classifier[50] using 85% working point. The jet energy is calibrated and the Monte Carlo events are reweighted by the product of b-tagging efficiency scale factors for jets been tagged as b-jets or by a jet tagging inefficiency scale factor.

- To avoid duplication between reconstructed objects, further overlap removal is applied in such a sequence:

  - Electron-Muon Overlap: if $\Delta R(e, \mu) < 0.1$, remove the electron and keep the muon.
  - Electron-Jet Overlap: if $\Delta R(e, j) < 0.2$, keep the electron and remove the jet.
  - Muon-Jet Overlap: if $\Delta R(\mu, j) < 0, 2$, keep the muon and remove the jet.

  Muons can frequently radiate photons which can be identified as electrons but the reverse process is heavily suppressed. Therefore muon is preferred in the electron-muon overlap removal. Both jets and electrons are reconstructed with the energy deposits in the calorimeter and electrons' reconstruction also relies on matching to a well defined inner detector track. Thus if an electron overlaps with a jet, it's more likely to be the signature of a high energy electron. If a muon overlaps a jet, the muon can originate from a heavy flavor decay, thus muon is kept in such a situation.

- The missing transverse energy (E_T^{miss}) is used in this analysis according the signature of our signal process. It's reconstructed from the calorimeter cells with $|\eta| < 4.9$ and muons. The calibrated cells are calibrated according to the reconstructed objects such as electrons, photons, etc. The calibrations and corrections applied to electrons, muons and jets are propagated in the calculation of E_T^{miss}.

With the selected good physics objects, the events will undergo a set of cuts which is called "event pre-selection", the selections are:

- Good run list: Ensure the detector and LHC conditions are good enough during the data taking.
- Event cleaning: Remove events which are recorded when LAr or Tile calorimeters are not functional normally.
- Primary vertex: The event must contain a primary vertex with at least three tracks.
- E_T^{miss} cleaning: Veto event containing jets close to badly behaving calorimeter region, this is to avoid any bias on the E_T^{miss} measurment.
- Trigger: At least one of the lowest un-prescaled single lepton triggers (EF_-e24vhi_medium1, EF_e60_medium1, EF_mu24i_tight or EF_mu36_tight, "EF" refers to "event filter", "e" or "mu" means electron or muon, the number such as "24" stands for the transverse momentum threshold on the lepton, "vh" refers to hadronic veto requirement, "i" means isolation requirement, "medium" or "tight" are different working points targeting for different signal efficiencies) must be satisfied.
- Three leptons selection: The events must have exactly three good leptons with $p_T > 20$ GeV.
- Trigger Match: At least one of the leptons have fired the trigger.

On top of the pre-selection level, events are further categorized based on the number of Same Flavor Opposite Sign (SFOS) pairs present in the event. Three separate signal regions are then defined, 0 SFOS, 1 SFOS and 2 SFOS. In the 2 SFOS region, one lepton is allowed to belong to both pair combinations. The advantage of splitting the signal region based on this classification comes when studying the background especially for backgrounds like WZ and ZZ where SFOS lepton pairs may originate from the Z boson. The 0 SFOS signal region is the purest region where the backgrounds are almost entirely reducible. It has the best sensitivity due to small background contamination. The small amount of backgrounds can originate from he effect of mis-identification of the lepton's charge and fake leptons.
The final signal selection cuts are determined through an optimization procedure which considers both the signal yield and the uncertainty on the measurement of the signal

strength. The optimization procedure starts from the event pre-selection for all three signal regions. A veto is applied on events with jets tagged as b-jet using the 85% working point of the MV1 classifier. The b-veto increases the b-jet mis-identification efficiency but remains manageable at about 1%. On top of the b-jet veto, a cut on the jet multiplicity, regardless of whether the jet is tagged or not, is applied. Only events with at most one jet are kept, the signal efficiency of this cut is about 90% while almost 50% of background is reduced. WZ and ZZ are main backgrounds in this analysis, thus a Z-veto is applied in each signal region. Mass windows are slightly different among the signal regions which are chosen by the optimization. In the 0 SFOS signal region, there is no SFOS pair while there is still a peak in the same sign electron-electron mass distribution due to charge mis-identification. A narrow symmetric window of ± 15 GeV around Z mass is chosen to suppress the Z background in 0 SFOS channel. In the 1 SFOS region, there is a large amount of background from $Z\gamma$ process which mostly show up in the lower shoulder of the Z peak, thus an asymmetric window with the boundaries being 35 GeV below Z mass and 20 GeV above is chosen. In the 2 SFOS region, both pairs are considered and the event will be vetoed if either falls into the mass window, a symmetric window of ± 20 GeV around Z mass is chosen. Since the signal events are produced with three leptons and three neutrinos, thus a cut on E_T^{miss} is also optimized and applied to the three signal regions. The direction of E_T^{miss} can also be compared to the direction of the vector sum of the three charged leptons, thus an additional variable which is defined as $|\Delta\phi(3l, E_T^{miss})| = |\phi(3l) - \phi(E_T^{miss})|$ can be used to discriminate signal from background. To some extent, the backgrounds also show such a behavior but it's much less pronounced than it is for the signal. An additional cut defined as the invariant mass of same-flavor pairs is applied in the 0 SFOS region which can remove low-mass contamination from processes like QCD. The optimized selection for each signal region is summarized in Table 5.1.

	0 SFOS	1 SFOS	2 SFOS
Pre-selection	Exactly 3 leptons with $P_T > 20$ GeV where at least one is trigger matched.		
b-tagged Jet Veto	$N_{b-jet} = 0$ (85 % b-tagging efficiency)		
Same-Flavor Mass	$m_{\mathrm{SF}} > 20$ GeV		
Z-Veto ($m_Z = 91.1876$ GeV)	$\|m_{ee} - m_Z\| > 15$ GeV	No m_{SFOS} with $m_Z - 35\mathrm{GeV} < m_{\mathrm{SFOS}} < m_Z + 20$ GeV	$\|m_{\mathrm{SFOS}} - m_Z\| > 20$ GeV
Missing E_T		$E_T^{Miss} > 45$ GeV	$E_T^{Miss} > 55$ GeV
Lepton-Missing E_T Angle	$\|\phi(3l) - \phi(E_T^{Miss})\| > 2.5$		
Inclusive Jet veto	$N_{jet} \leqslant 1$		

Table 5.1 Optimized signal selection for the three signal regions.

5.2 Background Estimation

In this analysis, background can be due to fake leptons originated from hadronic decays or lepton's charge mis-identification. This part of backgrounds are estimated with data-driven techniques. The rest irreducible backgrounds are estimated with Monte Carlo simulation.

5.2.1 MC Background

The WZ process contains irreducible background for the WWW final state. Previous studies for WWW final state or ZZ final state at the LHC reveal that the measured cross sections are usually higher than the NLO predictions and further study on NNLO cross section calculations illustrate that a large scale factor should be applied to the cross section at NLO accuracy[51][52]. However, such correction is not yet available for the WZ process, therefore the NLO predictions for this background should be checked using the data. The ABCD method is introduced to normalize this process. Events are required to pass event pre-selection and there is only one SFOS lepton pair, and one third lepton from a different flavor. To suppress the contribution from fake lepton background, all leptons must have $p_T > 25$ GeV. The definitions of the four regions are:

- Signal Region(A): Isolated and in Z peak, $|M_{ll}^{SFOS} - M_Z| < 15$ GeV, $E_T^{Iso(R<0.2)}/E_T < 0.10$ and $p_T^{Iso(R<0.2)}/p_T < 0.04$.
- Control Region(B): Isolated and off Z peak, $|M_{ll}^{SFOS} - M_Z| > 25$ GeV, $E_T^{Iso(R<0.2)}/E_T < 0.10$ and $p_T^{Iso(R<0.2)}/p_T < 0.04$.
- Control Region(C): Non-isolated and in Z peak, $|M_{ll}^{SFOS} - M_Z| < 15$ GeV, $E_T^{Iso(R<0.2)}/E_T > 0.15$ and $p_T^{Iso(R<0.2)}/p_T > 0.10$.
- Control Region(D): Non-isolated and off Z peak, $|M_{ll}^{SFOS} - M_Z| > 25$ GeV, $E_T^{Iso(R<0.2)}/E_T > 0.15$ and $p_T^{Iso(R<0.2)}/p_T > 0.10$.

The WZ events in the three control regions (B, C and D) are negligible therefore the number of observed events in these four regions can be expressed as:

$$N_A = N_A^{WZ}(\text{measured}) + N_A^{jet} + N_A^{EW} \tag{5.1}$$

$$N_B = N_B^{jet} + N_B^{EW} \tag{5.2}$$

$$N_C = N_C^{jet} + N_C^{EW} \tag{5.3}$$

$$N_D = N_D^{jet} + N_D^{EW} \tag{5.4}$$

where N^{jet} means number of events with jet faking leptons and N^{EW} indicates events containing three real leptons or two real leptons and one photon. And because the ratio of isolated to non-isolated fake lepton background should be the same in or off the Z peak. Thus the WZ yield in signal region A can be expressed as:

$$N_A^{WZ}(\text{Measured}) = (N_A - N_A^{EW}) - \frac{1}{R^{jet}}\frac{(N_B - N_B^{EW} - c_B(N_A^{WZ}(MC)))(N_C - N_C^{EW} - c_C(N_A^{WZ}(MC)))}{N_D - N_D^{EW} - c_D(N_A^{WZ}(MC))} \quad (5.5)$$

where $R^{jet} = \frac{N_B^{jet}N_C^{jet}}{N_A^{jet}N_D^{jet}}$ is defined to account for the bias on background correlation between A-B to C-D and $C_X = \frac{N_X^{WZ}(MC)}{N_A^{WZ}(MC)}$, X=(B,C,D) is defined to account for the signal leakage. The event yield of WZ in region A from MC is $N_A^{WZ}(MC) = 498 \pm 1$ while the measurement from the ABCD method is $N_A^{WZ}(measured) = 537 \pm 35$. A correction factor of $1.08 \pm 0.07(stat)$ is then derived and this correction factor is found to agree with other measurements performed by the ATLAS and the CMS collaboration[51][53]. Several systematic sources are considered such as variation on the isolation cuts, variation on the cuts of the control region definitions, systematic uncertainties from the MC and variations on R^{jet}. A total systematic uncertainty of 5.9% is assigned to the correction factor. The final correction factor on the WZ background is $k_{WZ} = 1.08 \pm 0.07(stat) \pm 0.07(syst)$. Figure 5.1 shows very good agreement between the data and the mc with WZ correction factor.

Another important background is due to the ZZ^* production where one lepton goes out of the detector acceptance or fails the lepton selection. This background is modeled with Powheg generator and $gg2ZZ$ generator for the loop included processes, a correction factor of 1.05 is adopted to scale up to the NNLO predictions. The total systematic uncertainty of the theoretical prediction is taken to be 15%. The agreement between data and the model is checked in a control region which is enriched with ZZ^* events. The control region is selected with two same flavor opposite sign lepton pairs, the p_T requirements for the four leptons (sorted by p_T in descending order) are : $p_T^1 > 25$ GeV, $p_T^2 > 15$ GeV, $p_T^3 > 15$ GeV and $p_T^4 > 10$ GeV. To suppress the contamination from fake lepton backgrounds, only the events with two on shell Z bosons are kept. The comparison between data and MC prediction in this control region is shown in Figure 5.2 and detailed numbers are listed in Table 5.2. The figures and numbers have shown a very good agreement between data and MC prediction.

The $Z\gamma$ process where the Z boson decaying to a pair of leptons (e or μ) is estimated with Monte Carlo simulation. Previous study illustrates that Sherpa generator can describe accurately the shape and normalization of data in the 7 TeV and 8 TeV datasets[54][55] thus Sherpa generator is chosen for the estimation of this background. This background

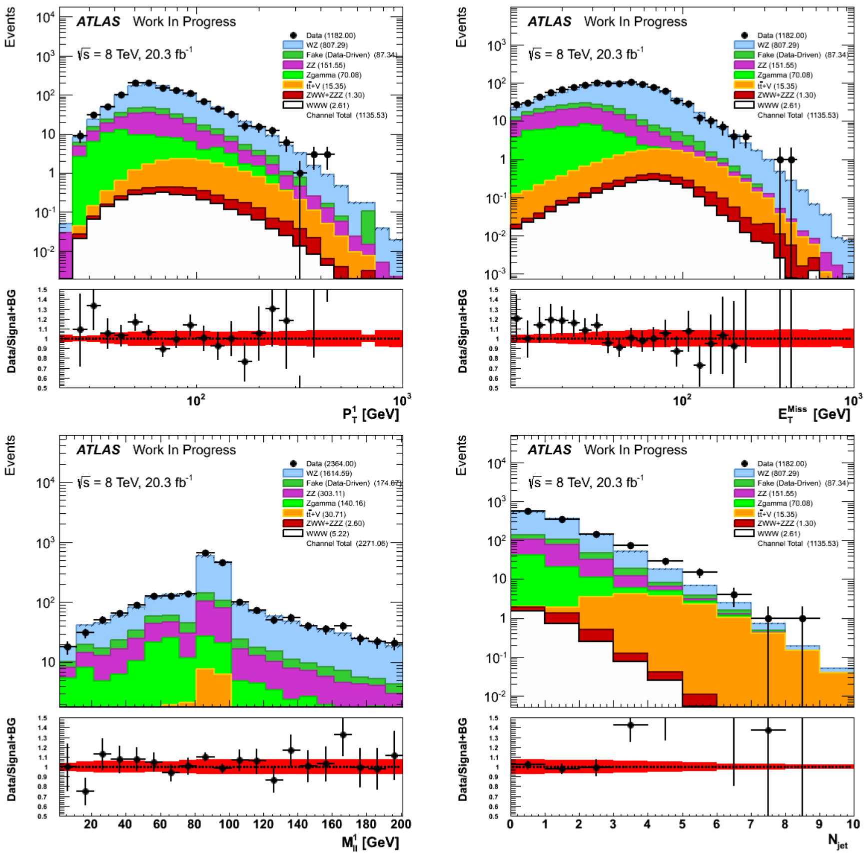


Figure 5.1 WZ 2SFOS Control regions. Distribution of leading lepton p_T, E_T^{miss}, M_{12}, and jet multiplicity. The systematic band shows the uncertainty on the WZ correction factor.

	Event Yield
WZ	0.05±0.01
ZZ	156.2±0.3(stat)±22.3(syst)
$Z\gamma$	0.0±0.0
Fake (MC)	3.6±0.2
triboson and $t\bar{t}+V$	4.1±0.2
Expected Signal + Background	164.0±0.3(stat)±22.3(syst)
Observed Data	155±12

Table 5.2 Event number of data and MC predictions in the ZZ control region. Uncertainties of the MC predictions include statistical uncertainty only except that theoretical uncertainty is also included for the ZZ MC.

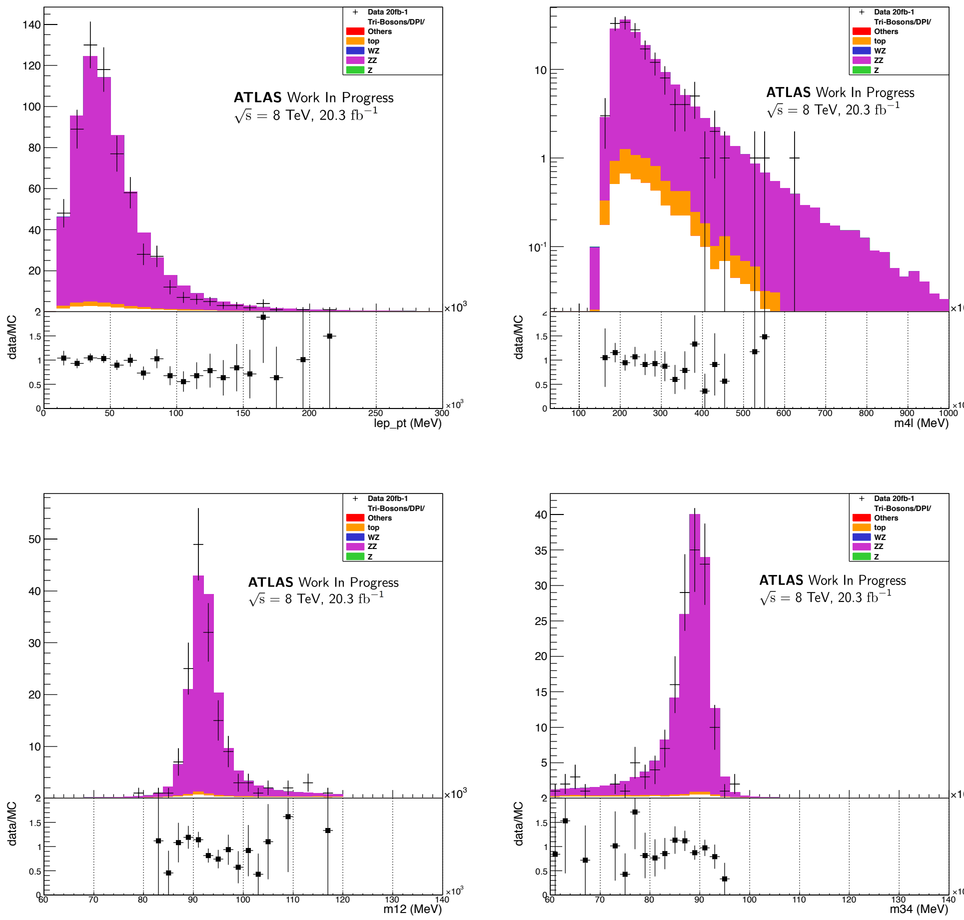


Figure 5.2 $ZZ \to 4\ell$ Control regions. Distribution of leptons p_T, M_{12}, M_{34}, M_{4l}.

may enter our selection via the conversion of one photon into a pair of electrons and then the loss of one electron in the acceptance. A delicate control region is designed to check the agreement between the MC prediction and the data. The events are selected with exactly two muons and on electron and the invariant mass of the three leptons is required to be close to the Z peak of ($|M_{\mu\mu e} - M_Z| < 15GeV$). Comparison between the data and MC prediction in the $Z\gamma$ control region is shown in Figure 5.3 and detailed numbers can be found in Table 5.3. The figures and the numbers show a very good agreement between the data and the MC prediction.

The double parton scattering (DPS) backgrounds are also studied in this analysis using the method applied in the same sign WW analysis[56]. Their contribution are found to be negligible in this analysis. Other backgrounds estimated from MC are the processes containing three real leptons: $t\bar{t} + V$, WWZ and WWZ. The $t\bar{t} + V$ process has been measured by other groups of the ATLAS collaboration and the measurement is found to be consistent with the NLO predictions[57] and the normalization uncertainty is about 30%. An equivalent uncertainty of 30% is assigned for the other VVV processes that

	Event Yield
WZ	7.47±0.11
ZZ	9.116±0.075
$Z\gamma$	80.3±2.8
$ZWW + ZZZ$	0.0285±0.0046
$t\bar{t} + V$	0.338±0.012
Fake (data-driven)	21.9±1.2
WWW	0.3199±0.0073
Expected Background	119.2±3.1
Expected Signal + Background	119.5±3.1
Observed Data	119±11

Table 5.3 Event number of data and MC predictions in the $Z\gamma$ control region. Uncertainties are statistical uncertainty only.

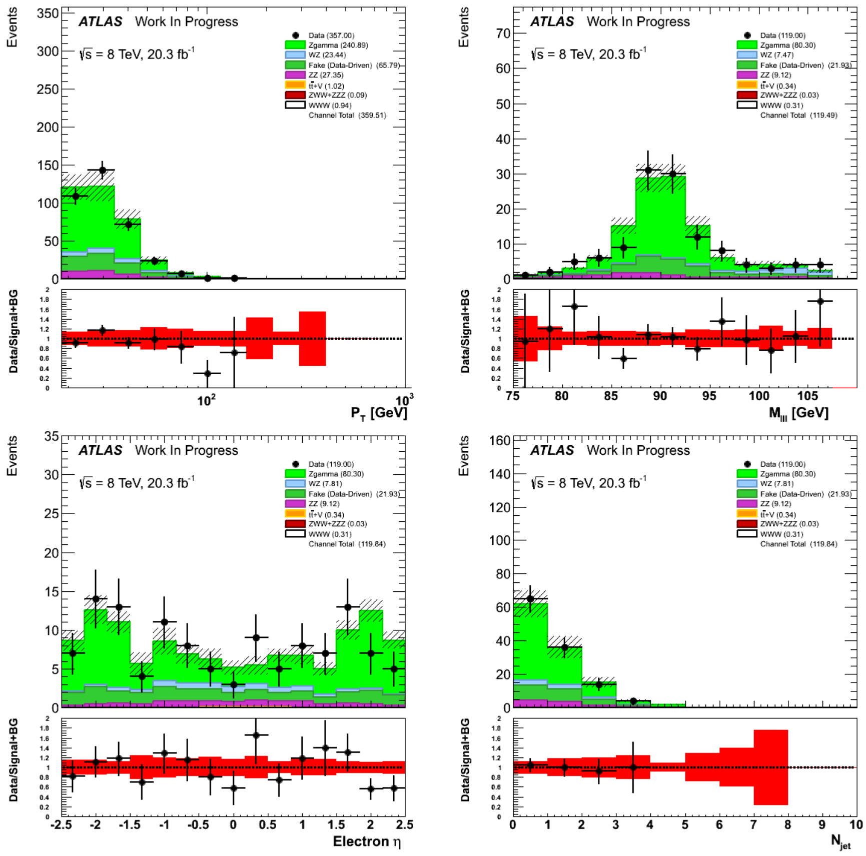


Figure 5.3 $Z\gamma$ Control region. Distribution of leptons p_T, invariant mass of the 3leptons, electron η, and jet multiplicity.

are not coming from the signal.

5.2.2 Background due to Charge Mis-Identification

In this analysis, three signal regions are divided according to the charge and flavor of the three leptons, therefore events can be mis-classified into one of the regions due to lepton's charge mis-identification. This is particularly important for the 0 SFOS region where the WZ and ZZ backgrounds are mostly due to lepton's charge mis-identification. The charge mis-identification is found to be negligible for muons and impacts mostly electrons. To estimate the background from electron's charge mis-identification, the probability of one electron with wrong charge is measured as a function of p_T and $|\eta|$ using $Z \rightarrow ee$ event selected from the data. These rates are then applied to the WZ and ZZ MC samples to evaluate the contribution from charge mis-identification in different signal regions.

The likelihood method is utilized to measure the charge mis-identification rates which are used for the measurement in data. In addition to the likelihood method, a truth matching method based on truth level information in MC is also employed as a cross check. The likelihood method assumes that for $Z \rightarrow e^+e^-$ events, the probability to reconstruct a pair of same sign electrons is $(\varepsilon_1 + \varepsilon_2)$ where ε_1 and ε_2 are the probabilities of charge mis-identification for the two electrons, respectively. The charge mis-identification rate is parametrized as a function of $|\eta|$ and p_T of the electrons. The $|\eta|$ dependence is particularly important since the amount of material the electrons traversed before entering the calorimeter is strongly dependent on the region of the detector where the electron is reconstructed. The charge mis-identification rates $(|\eta|, p_T)$ are measured from the total number of events and the number of events with a pair of same sign electrons by maximizing the following likelihood function constructed from Poisson statistics:

$$\ln \mathcal{L}(\varepsilon|N_{tot}, N_{ss}) = \sum_{i,j} \ln \left[N_{tot}^{i,j}(\varepsilon_i + \varepsilon_j)\right] N_{SS}^{i,j} - N_{tot}^{i,j}(\varepsilon_i + \varepsilon_j), \tag{5.6}$$

where $N_{tot}^{i,j}$ and $N_{SS}^{i,j}$ are the total number of candidate events and the number of events which have a same-sign electron pair, having the first and second lepton in the i-th and j-th bin respectively. The bin index i, j denotes each cell in η-p_T 2D space. The binning of p_T and $|\eta|$ are shown in Table 5.4.

The truth matching method is based on the comparison between electron's truth charge to its reconstructed charge. Two good reconstructed electrons are selected and denoted as "A" and "B", two truth electrons are selected and referred as "C" and "D". The dis-

$\lvert\eta\rvert$	[0, 0.8]	[0.8, 1.15]	[1.15, 1.6]	[1.6, 1.8]	[1.8, 2.0]		
	[2.0, 2.2]	[2.2, 2.3]	[2.3, 2.4]	[2.4, 2.5]			
p_T [GeV]	[15, 30]	[30, 40]	[40, 50]	[50, 60]	[60, 80]	[80, 120]	[120, 1000]

Table 5.4 The η and p_T binning for the measurement of charge mis-identification rate.

tance ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) between all pairs (AC, BD, AD and BC) are computed to match the reconstructed electrons to the truth electrons. If $\Delta R(AC) + \Delta R(BD) < \Delta R(AD) + \Delta R(BC)$, then A is matched to C and B is matched to D otherwise A is matched to D and B is matched to C. To avoid incorrect matching, events containing reconstructed electron matched to truth electron with $\Delta R > 0.5$ is removed. The charge between the truth electron and the reconstructed electron is then compared to determine the truth charge mis-identification rate. For the electron's charge mis-identification rate measurement, events are selected with two good electrons and the invariant mass of the two electrons should be within a Z mass window: ($M_Z - 10$ GeV, $M_Z + 10$ GeV).
A closure test is performed for the likelihood method through the comparison between the rates from the truth method and from the likelihood method using $Z \to ee$ MC samples. Figure 5.4 shows a very good agreement between the truth rates and the likelihood rates, the errors on the plot are statistical only. The slight difference between these two set of rates are taken into account as one source of systematic uncertainty.

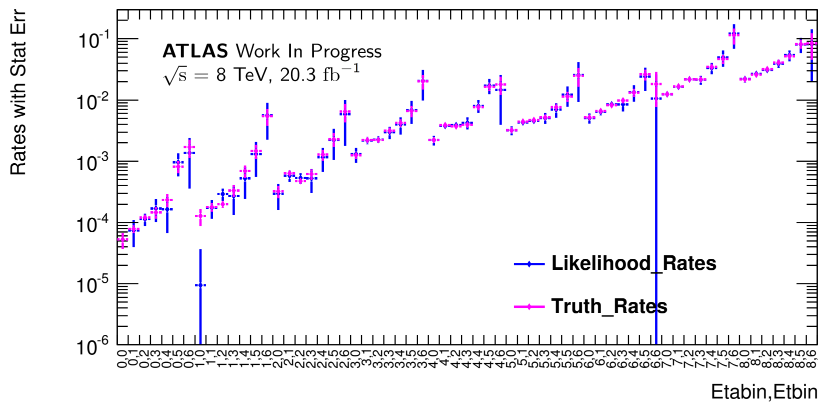


Figure 5.4 Comparison of electron's charge mis-identification rate measured with the truth method and the likelihood method using $Z \to ee$ MC sample. The errors are statistical only and the labels on x axis indicate the $\lvert\eta\rvert$ and p_T bins.

The electron's charge mis-identification rates measured in the data with the likelihood method is shown in Figure 5.5. These rates are used as central values to estimate background due to charge mis-identification in this analysis.
Several sources of systematic uncertainties are taken into account for the charge mis-identification rates.

- The contamination of non-$Z \to ee$ events in the measurement.

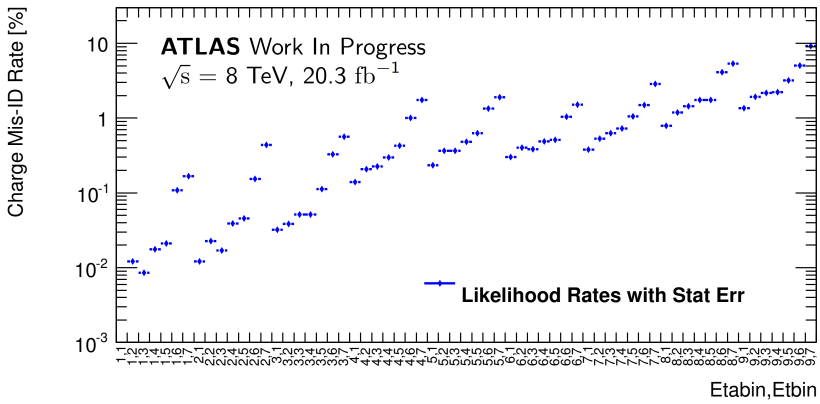


Figure 5.5 Electron's charge mis-identification rates measured in the data using the likelihood method. Errors are statistical only and the labels on the x axis indicate the $|\eta|$ and p_T bins.

- The non-closure in the closure test of the likelihood method.

The second term is directly taken from the difference between the likelihood rates and the truth rates in the closure test. The effect of contamination from background (non-$Z \rightarrow ee$) events is studied through a delicate method called Template Fit. In the template fit method, two templates are required: the signal template is obtained from the $Z \rightarrow ee$ MC sample while the background template is obtained via two steps: select a raw background template with certain criteria (one good electron and one electron fails the tight++ identification cut) from data selection which is still not purely background (contaminated by a lot of $Z \rightarrow ee$ events), this raw background is then fitted by a 4^{th} order polynomial function. The distribution of the invariant mass of the two electrons within range [60 GeV, 120 GeV] is adopted for the fit. The fitted 4^{th} order polynomial function is the background template we expected. Due to statistical constraint, the background template is obtained in each $|\eta|$ bin but p_T bins are grouped together. Table 5.5 shows the detailed selection for the signal and the raw background template.

Signal	Background
EF_e24vhi_medium1 or EF_e60_medium1	EF_e24vhi_medium1 or EF_e60_medium1
Exactly two electrons passing electron selection	Choose the leading and subleading electrons and at least one of these 2 electrons satisfied the background electron selection
Trigger Match	
$\lvert M_{ee} - M_Z \rvert < 10\text{GeV}$	

Table 5.5 Event selection for the signal and the raw background template in the template fit method.

Figure 5.6 shows the one example of the signal template obtained from $Z \rightarrow ee$ MC samples and the raw background template obtained from the data. The signal template is centralized at the Z peak as expected but looking at the background template, there is a lot of non-$Z \to ee$ events but still an obvious Z peak. Therefore, a polynomial fit is applied to the raw background template to extract the accurate background shape by subtracting the $Z \to ee$ contamination, one example is shown in Figure 5.7

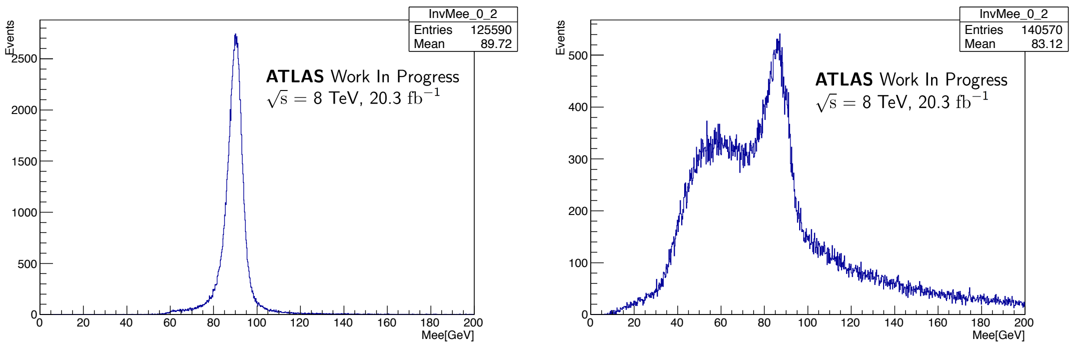


Figure 5.6 Distribution of M_{ee} for the signal template obtained from $Z \to ee$ MC sample (left) and the raw background template obtained from the data (right).

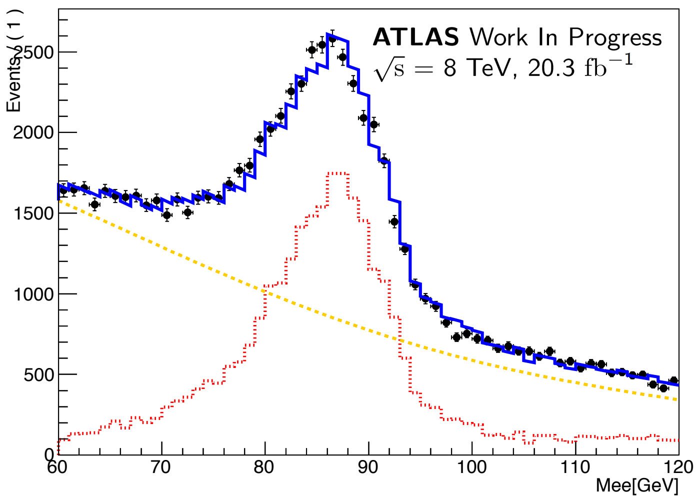


Figure 5.7 Polynomial fit for the raw background template. This is an example for events with first electron within $|\eta|$ [0,0.8] and second electron within $|\eta|$ [1.15,1.60]. The red line is the signal template obtained from $Z \to ee$ MC sample, the orange line is the polynomial function which is considered as the background component, the black solid line is the raw background template and the blue line is the fit.

The fitted polynomial functions are thus used to describe the background in the template fit method. Figure 5.8 show one example of the template fit using the signal template and the background template obtained before.

The non-$Z \to ee$ contamination in the data used for rate measurement are obtained through the template fit method, the purity of $Z \to ee$ for $N^{i,j}_{tot}$ and $N^{i,j}_{OS}$ are shown in Table 5.6 and Table 5.7. Because the statistical constraint on N_{SS}, individual fit in each

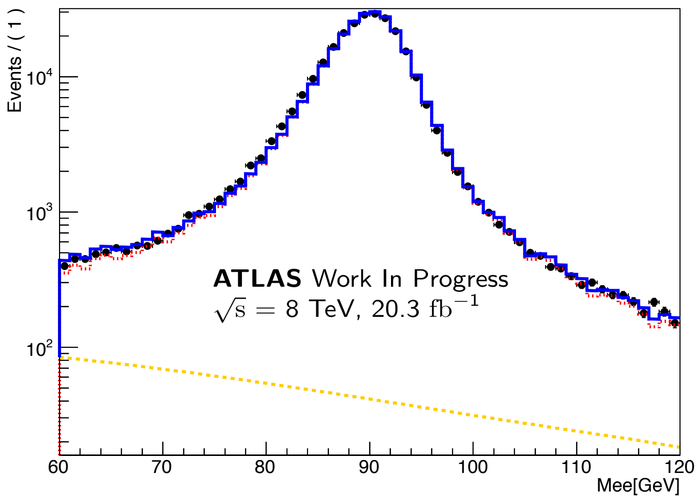


Figure 5.8 Template fit for events with first electron within $|\eta|$ [0,0.8] and second electron within $|\eta|$ [1.15,1.60]. The red line is the signal template obtained from $Z \rightarrow ee$ MC sample, the orange line is the polynomial function to describe background component, the black solid line indicates the events used for rate measurement and the blue line is the fit.

	[0,0.8]	[0.8,1.15]	[1.15,1.60]	[1.60,1.80]	[1.80,2.0]	[2.0,2.20]	[2.20,2.30]	[2.30,2.40]	[2.40,2.50]
[0,0.8]	0.9951	0.9966	0.9945	0.9956	0.9953	0.9924	0.9961	0.9898	0.9893
[0.8,1.15]	0.996	0.9982	0.9933	0.9887	0.9939	0.9953	0.992	0.9935	0.972
[1.15,1.60]	0.9904	0.9895	0.9892	0.9885	0.9895	0.9942	0.9875	0.9912	0.9802
[1.60,1.80]	0.9794	0.9766	0.9787	0.9799	0.9828	0.9822	0.9775	0.9674	0.9534
[1.80,2.0]	0.9914	0.992	0.994	0.9878	0.9927	0.9906	0.9906	0.9912	0.9546
[2.0,2.20]	0.9934	0.9978	0.9833	0.9872	0.9936	0.9847	0.9837	0.9717	0.9776
[2.20,2.30]	0.998	0.9874	0.9901	0.9743	0.992	0.9874	0.9813	0.9835	0.9509
[2.30,2.40]	0.9891	0.9883	0.9825	0.9734	0.9916	0.9846	0.9698	0.9643	0.9739
[2.40,2.50]	0.9774	0.9636	0.9794	0.9703	0.9766	0.9805	0.9804	0.9509	0.9211

Table 5.6 Signal purity for N_{tot}, different rows stand for different $|\eta|$ bins of the sub-leading electron in the event, and different columns stand for different $|\eta|$ bins of the leading electrons in the event.

bin is impossible, a global fit is perform for N_{SS} and N_{OS}, Figure 5.9 shows the global fit for N_{SS}. Signal purity is 0.9372 ± 0.0042 (stat) for N_{SS} and 0.9921 ± 0.0013 (stat) for N_{OS}, the ratio of signal purity of N_{SS} to that of N_{OS} is 0.9447 and this ratio is assumed to be independent of $|\eta|$, therefore the signal purities of $N_{SS}^{i,j}$ are obtained by scaling the purities of $N_{OS}^{i,j}$ which are also shown in Table 5.8.

With the signal purity of $N_{tot}^{i,j}$ and $N_{SS}^{i,j}$, the background contamination can be subtracted from the data used for rate measurement and a new set of likelihood rates are recomputed. The difference between this new set of rates and the central values is taken as a systematic uncertainty. This term of uncertainty is shown in Table 5.9.

Considering the electron's kinematic difference among difference processes, a test is performed using WZ MC sample. The electron's charge mis-identification rates are recomputed on WZ MC sample using the truth method and compared to the old one obtained from $Z \rightarrow ee$ MC sample. Given the limited statistic, a new set of binning

	[0,0.8]	[0.8,1.15]	[1.15,1.60]	[1.60,1.80]	[1.80,2.0]	[2.0,2.20]	[2.20,2.30]	[2.30,2.40]	[2.40,2.50]
[0,0.8]	0.9958	0.997	0.9947	0.9958	0.9951	0.9924	0.9968	0.9871	0.9863
[0.8,1.15]	0.9964	0.9982	0.9937	0.9882	0.9937	0.9951	0.9924	0.993	0.9707
[1.15,1.60]	0.9904	0.9896	0.9889	0.9883	0.9899	0.9938	0.988	0.9913	0.9792
[1.60,1.80]	0.9786	0.9765	0.9784	0.9771	0.9811	0.9815	0.9753	0.9638	0.9483
[1.80,2.0]	0.9907	0.9918	0.9944	0.9876	0.9921	0.9911	0.9893	0.9924	0.9527
[2.0,2.20]	0.9938	0.9979	0.9826	0.9882	0.9934	0.9815	0.9844	0.9735	0.9777
[2.20,2.30]	0.9982	0.9872	0.9903	0.9711	0.9921	0.987	0.9841	0.9779	0.9391
[2.30,2.40]	0.9898	0.9848	0.9826	0.9719	0.9897	0.984	0.9667	0.9562	0.9637
[2.40,2.50]	0.9775	0.9656	0.9784	0.9628	0.974	0.9783	0.9771	0.9513	0.9226

Table 5.7 Signal purity of N_{OS}, different rows stand for different $|\eta|$ bins of the sub-leading electron in the event, and different columns stand for different $|\eta|$ bins of the leading electrons in the event.

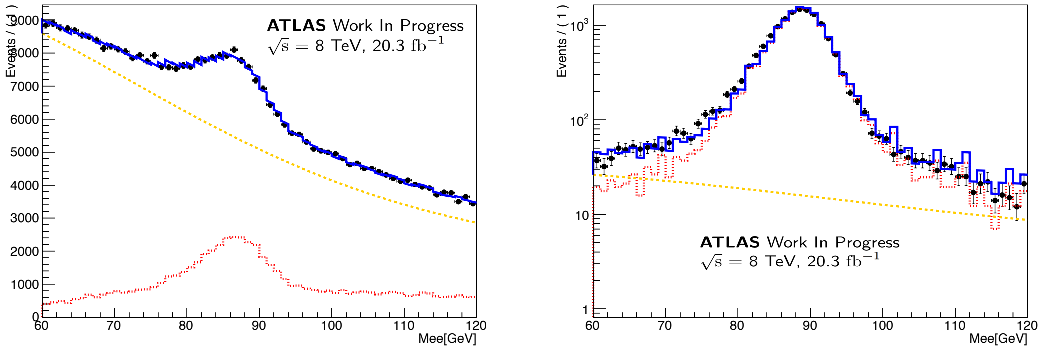


Figure 5.9 Global fit performed for N_{SS}, the polynomial fit (left) to get the background template and the template fit (right) to get the signal purity.

	[0,0.8]	[0.8,1.15]	[1.15,1.60]	[1.60,1.80]	[1.80,2.0]	[2.0,2.20]	[2.20,2.30]	[2.30,2.40]	[2.40,2.50]
[0,0.8]	0.9407	0.9419	0.9397	0.9408	0.9401	0.9375	0.9417	0.9325	0.9317
[0.8,1.15]	0.9413	0.943	0.9387	0.9335	0.9387	0.9401	0.9375	0.9381	0.917
[1.15,1.60]	0.9357	0.9349	0.9342	0.9337	0.9352	0.9389	0.9334	0.9365	0.9251
[1.60,1.80]	0.9245	0.9225	0.9243	0.9231	0.9268	0.9272	0.9214	0.9105	0.8958
[1.80,2.0]	0.9359	0.937	0.9394	0.933	0.9372	0.9363	0.9346	0.9375	0.9
[2.0,2.20]	0.9389	0.9427	0.9283	0.9336	0.9384	0.9273	0.93	0.9197	0.9237
[2.20,2.30]	0.943	0.9326	0.9355	0.9174	0.9372	0.9325	0.9297	0.9238	0.8872
[2.30,2.40]	0.935	0.9304	0.9283	0.9181	0.935	0.9296	0.9133	0.9033	0.9104
[2.40,2.50]	0.9235	0.9122	0.9243	0.9095	0.9202	0.9242	0.9231	0.8987	0.8716

Table 5.8 Signal purity of N_{SS}, different rows stand for different $|\eta|$ bins of the sub-leading electron in the event, and different columns stand for different $|\eta|$ bins of the leading electrons in the event.

$|\eta|$ / p_T[GeV]	[0,0.8]	[0.8,1.15]	[1.15,1.60]	[1.60,1.80]	[1.80,2.0]	[2.0,2.20]	[2.20,2.30]	[2.30,2.40]	[2.40,2.50]
[15,30]	8.85	5.63	5.75	5.85	5.79	5.64	5.64	5.68	5.49
[30,40]	5.73	5.71	5.83	5.97	5.75	5.78	5.72	5.81	5.59
[40,50]	5.76	5.69	5.71	5.71	5.65	5.71	5.62	5.71	5.62
[50,60]	5.74	5.55	5.64	5.53	5.61	5.65	5.41	5.49	5.65
[60,80]	5.77	5.57	5.71	5.99	5.59	5.66	5.35	5.53	5.41
[80,120]	5.79	5.63	5.73	5.71	5.74	5.77	5.36	5.74	5.89
[120,1000]	5.76	5.71	5.54	5.76	5.52	5.61	5.73	5.98	6.14

Table 5.9 Systematic uncertainties due to background contamination on the central value of charge mis-identification rates in percent.

$\|\eta\|$ bins	$\|\eta\|$ bin index	p_T bins [GeV]	p_T bin index
$[0, 1.15]$	0	$[15, 40]$	0
$[1.15, 1.8]$	1	$[40, 60]$	1
$[1.8, 2.2]$	2	$[60, 100]$	2
$[2.2, 2.5]$	3	$[100, 1000]$	3

Table 5.10 The $|\eta|$ and p_T bins used for the comparison of charge mis-identification rates obtained with MC $Z \to ee$ sample and MC WZ sample.

is used in the test which is shown in Table 5.10. Figure. 5.10 shows the comparison between two sets of rates, a good agreement is observed.

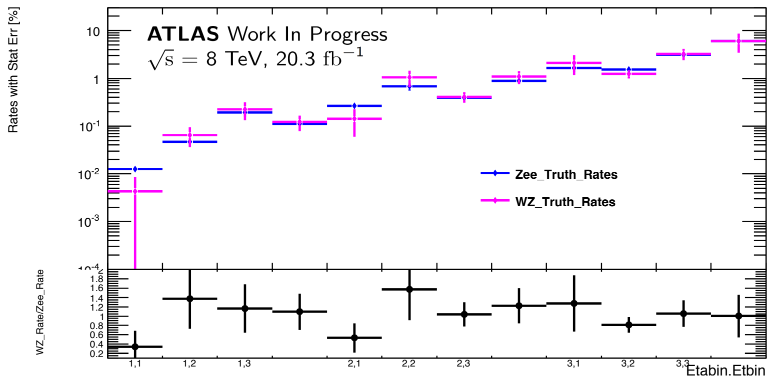


Figure 5.10 Comparison of truth rates obtained from $Z \to ee$ MC sample and WZ MC sample. The errors are statistical only and the labels on x indicate the $|\eta|$ and p_T bins.

The final electron's charge mis-identification rates with full uncertainties are shown in Figure 5.11.

Background due to lepton's charge mis-identification is primarily important in 0 SFOS region in particularly for the WZ and ZZ processes. The measured electron's charge mis-identification rates are applied to WZ and ZZ based on whether or not a charge flip can cause the event to appear in the 0 SFOS region. A weight is assigned to each event according to its final states and reconstructed electrons. The case with multiple electron charge flips is ignored since the probability is expected to be small. Only the following di-boson decays are considered:

- $WZ \to e^{\pm}\nu\, e^{+}e^{-}$
- $WZ \to \mu^{\pm}\nu\, e^{+}e^{-}$
- $WZ \to \tau^{\pm}\nu\, e^{+}e^{-}$
- $ZZ \to e^{+}e^{-}\, e^{+}e^{-}$

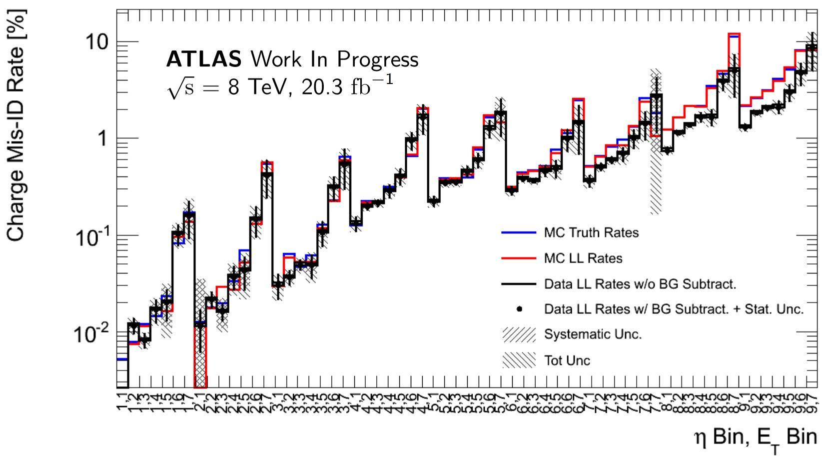


Figure 5.11 Electron charge misID rates obtained from data with the likelihood method. All the errors are now shown. The x axis label is the $|\eta|$, p_T bin index.

- $ZZ \to \mu^+\mu^-\ e^+e^-$

Take an example, a $WZ \to e^+\nu e^+e^-$ event may appear in the 0 SFOS region if and only when the positron is to flip its charge, thus the weight to be applied on this event is just the probability that this positron flip its charge. In this way, the WZ and ZZ events will be reweigthed by an additional event weight. Comparison between the reweighted yield and the MC predictions in the 0 SFOS region is shown in Figure 5.12 and Figure 5.13 for WZ and ZZ respectively. An offset between reweighted yield and MC prediction is observed which also demonstrate the necessity to perform the charge mis-identification correction.

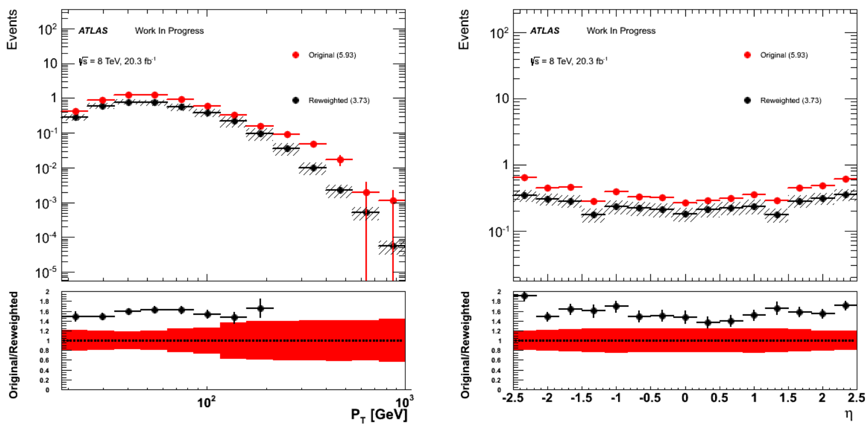


Figure 5.12 Comparison between yield reweighted with charge mis-identification rates and MC prediction for the $WZ \to \ell ee$ ($\ell = e, \mu$) process. Distribution of lepton p_T (left) and η (right).

The WZ and ZZ backgrounds are reweighted by the rates but there is no special treatment for the charge mis-identification contribution from other processes in the 0 SFOS region or from any processes in the 1 and 2 SFOS regions including diboson processes,

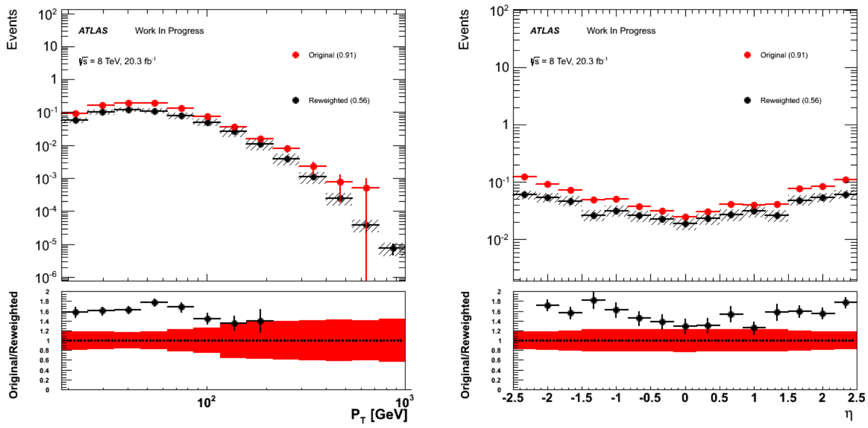


Figure 5.13 Comparison between yield reweighted with charge mis-identification rates and MC prediction for the $ZZ \rightarrow \ell\ell ee$ ($\ell = e, \mu$) process. Distribution of lepton p_T (left) and η (right).

as the effect is expected to be very small. This part of charge mis-identification background is thus estimated with MC.

5.2.3 Background due to Fake Lepton

One important source of background in this analysis is due to leptons originating from heavy flavor decays, mis-reconstructed leptons originating from hadrons in jets and electrons from photon conversion. These non-prompt leptons are referred as "fake" leptons and the background caused by fake leptons is estimated with a data-driven matrix method. Leptons are classified as loose or tight at the very beginning. Loose leptons must pass lepton preselection but fail the signal selection while tight leptons pass both the preselection and the signal selection. The preselection and signal selection are listed in Table 5.11 and Table 5.12.

Preselected electron	
Algorithm	Central Electrons
Acceptance	$p_T > 10$ GeV, $\lvert\eta\rvert < 2.47$ excluding crack region
Quality	Medium++
Impact parameter	$\lvert d_0/\sigma_{d_0}\rvert < 3.0$ $\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm
e-e isolation	$\Delta R(e,e) > 0.2$
e-μ isolation	$\Delta R(e,\mu) > 0.2$
Signal electron	
Quality	Tight++
Track isolation	$p_T^{cone20}/p_T < 0.04$
Calorimeter isolation	$E_T^{cone20}/E_T < 0.10$

Table 5.11 Summary of the electron selection criteria used for the global matrix method. The signal requirements defined in Section 5.1 are applied on top of the lepton preselection.

Preselected muon	
Algorithm	Combined
Acceptance	$p_T > 10\,\mathrm{GeV}, \lvert\eta\rvert < 2.5$
Quality	Tight
Inner detector track quality	MCP ID Hits selection
Impact parameter	$\lvert d_0/\sigma_{d_0}\rvert < 3.0$ $\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm
μ-μ isolation	$\Delta R(\mu, \mu) > 0.2$
Signal muon	
Track isolation	$p_T^{cone20}/p_T < 3.0$
Calorimeter isolation	$E_T^{cone20}/E_T < 0.10$

Table 5.12 Summary of the muon selection criteria used for the global matrix method. The signal requirements defined in Section 5.1 are applied on top of the lepton preselection.

In case of single lepton events, the equation relating the number of event with real (n_R) and fake (n_F) lepton in p_T bin i to the number of events with tight (n_T) and loose (n_L) leptons can be expressed as:

$$\begin{pmatrix} n_T \\ n_L \end{pmatrix} = \begin{pmatrix} \epsilon_i & \zeta_i \\ 1-\epsilon_i & 1-\zeta_i \end{pmatrix} \begin{pmatrix} n_R \\ n_F \end{pmatrix},$$

where ϵ_i and ζ_i are the real and fake efficiencies in p_T bin i. Therefore, n_R and n_F can be computed with n_T and n_L through an inverse matrix:

$$\begin{pmatrix} n_R \\ n_F \end{pmatrix} = \frac{1}{\epsilon_i - \zeta_i} \begin{pmatrix} 1-\zeta_i & -\zeta_i \\ \epsilon_i - 1 & \epsilon_i \end{pmatrix} \begin{pmatrix} n_T \\ n_L \end{pmatrix}.$$

Now consider the case with N preselected leptons, denote the previous notations as:

$$r = \begin{pmatrix} n_R \\ n_F \end{pmatrix}, \quad t = \begin{pmatrix} n_T \\ n_L \end{pmatrix}, \quad \phi = \begin{pmatrix} \varepsilon_i & \zeta_i \\ 1-\varepsilon_i & 1-\zeta_i \end{pmatrix} \quad \Rightarrow \quad t_\beta = \phi_\beta^{\ \alpha} r_\alpha,$$

where α takes values corresponding to R or F and β takes values corresponding to T or L. The expected number of tight leptons that are fake can be expressed as:

$$t'_\nu = \phi_\nu^\mu \omega_\mu^\beta \phi_\beta^{-1} t_\alpha, \tag{5.7}$$

where ω represents the selection of only the expected fake component, in case with N preselected leptons, this formula can be written as:

$$t'_{\nu_1 \cdots \nu_N} = \phi_{\nu_1}^{\mu_1} \cdots \phi_{\nu_N}^{\mu_N} \, \omega_{\mu_1 \cdots \mu_N}^{\beta_1 \cdots \beta_N} \, \phi^{-1}{}_{\beta_1}^{\alpha_1} \cdots \phi^{-1}{}_{\beta_N}^{\alpha_N} t_{\alpha_1 \cdots \alpha_N},$$

where ω selects the sets of indices β_i corresponding to components one wish to count as fake background:

$$\omega_{\mu_1 \cdots \mu_N}^{\beta_1 \cdots \beta_N} = \delta_{\mu_1}^{\beta_1} \cdots \delta_{\mu_N}^{\beta_N} f(\beta_1, \ldots, \beta_N).$$

δ_i^j is the Kronecker delta and f is a function of the indices taking values 1 (for a fake combination) and 0 (for a real combination).

The real lepton efficiency for real preselected leptons passing the signal lepton selection is measured with $Z \to \ell\ell$ events selected from the data with a standard tag-and-probe method. The tag must pass all signal lepton selection and is trigger matched while the probe is required to satisfy the lepton preselection only. The invariant mass of the tag and probe has to be within a Z mass window of $80 \text{ GeV} < M_{ll} < 100 \text{ GeV}$. If both leptons are tagged, they will be alternatively considered as the tag to avoid bias introduced by the selection. The real lepton efficiency is then computed as a function of p_T as:

$$\varepsilon_i = \frac{n_i^{\text{Tight}}}{n_i}.$$

The p_T binning of the efficiency is coarse due to statistical constraint. Figure 5.14 shows the real lepton efficiencies derived from the data and the MC.

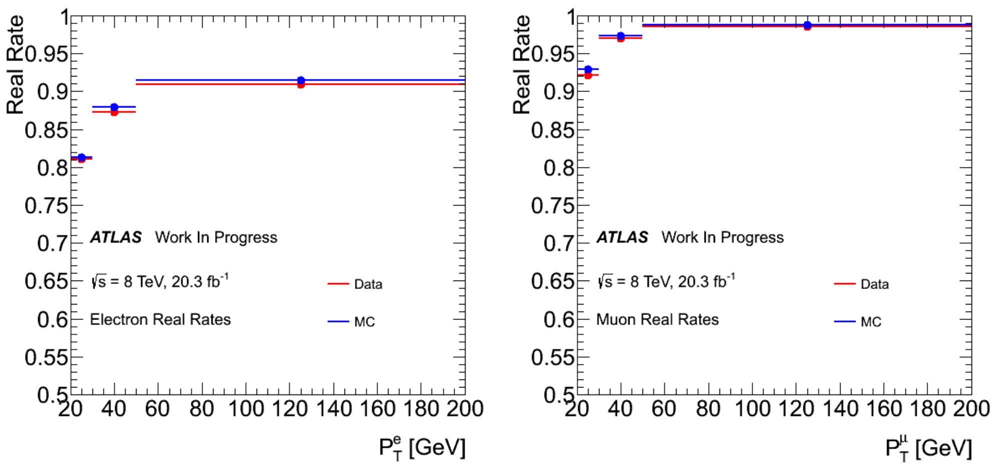


Figure 5.14 Real lepton efficiencies as a function of p_T measured from the data (red) and MC (blue) for electrons (left) and muons (right).

Two sources of systematic uncertainties are considered. One is from the choice of the 20 GeV Z mass window used in the method, the other one is from the background contamination in $Z \to \ell\ell$ events selected from the data. The first term is obtained through a 5 GeV variation on the Z mass window and this term is found to be negligible. The second term is obtained through a comparison between efficiencies derived from the data and the MC, the difference is treated as the systematic uncertainty. Table 5.13 anf Table 5.14 summarize the measured rates and corresponding uncertainties.

The fake lepton efficiency indicates the probability that a fake lepton passing the lepton preselection and the signal lepton selection. A similar tag-and-probe method is chosen

	Data		MC		
	ε	σ_{stat}	ε	σ_{stat}	σ_{sys}
$p_T \in [20, 30]$ GeV	0.8105	0.0011	0.8134	0.0013	0.0028
$p_T \in [30, 50]$ GeV	0.8732	0.0005	0.8794	0.0006	0.0062
$p_T > 50$ GeV	0.9097	0.0012	0.9150	0.0012	0.0053

Table 5.13 Real lepton efficiencies for electrons measured from the data and the MC. Errors include statistical and systematic uncertainties.

	Data		MC		
	ε	σ_{stat}	ε	σ_{stat}	σ_{sys}
$p_T \in [20, 30]$ GeV	0.9217	0.0010	0.9291	0.0012	0.0074
$p_T \in [30, 50]$ GeV	0.9700	0.0004	0.9737	0.0006	0.0038
$p_T > 50$ GeV	0.9862	0.0011	0.9878	0.0011	0.0017

Table 5.14 Real lepton efficiencies for muons measured from the data and the MC. Errors include statistical and systematic uncertainties.

for the measurement and the efficiency can be written as:

$$\zeta_i = \frac{n_i^{\mathrm{Tight}} - n_i^{\mathrm{Tight,Real}} - n_i^{\mathrm{Tight,PC}}}{n_i - n_i^{\mathrm{Real}} - n_i^{\mathrm{PC}}}, \tag{5.8}$$

where n_i^{Tight} is number of leptons passing preselection and signal selection, n_i^{Real} indicates number of real leptons while $n_i^{\mathrm{Tight,Real}}$ indicates number of real leptons in the tight leptons, n_i^{PC} stands for number of leptons from photon conversion and $n_i^{\mathrm{Tight,PC}}$ stands for number of leptons from photon conversion in the tight leptons. Subtraction of the contamination from real and photon conversion leptons is performed with truth information in MC. The measurement is performed in a fake enriched di-lepton region as a function of p_T where one lepton pass the preselection and signal selection with $p_T > 40$ GeV while the other one only pass the preselection. Additional $E_T^{miss} > 10$ GeV is required to reduce QCD background. The two leptons must have the same sign to reduce real processes such as $t\bar{t}$, WW and Z. This region is split according to the flavor of the tag and probe leptons. Electron fake efficiency is measured with one tag muon and probe electron, this is to avoid the large contamination from Z which may enter the region due to charge flip. Since muon's charge flip rate is negligible, the muon fake efficiency is measured in the muon-muon region. These regions for lepton fake efficiencies are further split based on the number of b-jets. Variation on number of b-jets will change the source of the fake leptons significantly since requiring b-jets will reduce light flavor component. The fake efficiencies measured in region with at least one b-jet (N_{b-jet}) are used as the central values since this region contains more heavy flavor contributions which is more compatible to the signal region. There are three sources of systematic uncertainties taken into account for the lepton fake efficiency:

- Subtraction of processes with two real leptons using MC introduces an uncertainty from MC cross section. It is estimated by varying the MC normalization by $\pm 20\%$ and this term is treated as "correlated" uncertainty.
- Kinematic difference between the control region where the measurement applied and the signal region. This term is obtained through varying kinematic cuts such as E_T^{miss} and p_T on the control region and is considered as "uncorrelated" uncertainty.
- Choice of number of b-jets on the control region. Apart from the nominal region with at least one b-jet, another region with no b-jet requirement is compared to the nominal region. The difference between fake lepton efficiencies measured from the two regions is considered as a systematic uncertainty. The difference in the composition in these two regions adequately covers the difference in the composition between the control region and the signal region.

The first term and the second term are combined together by adding in quadrature on an event-by-event basis. Figure 5.15 shows the measured fake lepton efficiencies with full uncertainty and detailed numbers are summarized in Table 5.15 and Table 5.16. The binning of p_T is chosen to be coarse due to statistical constraint.

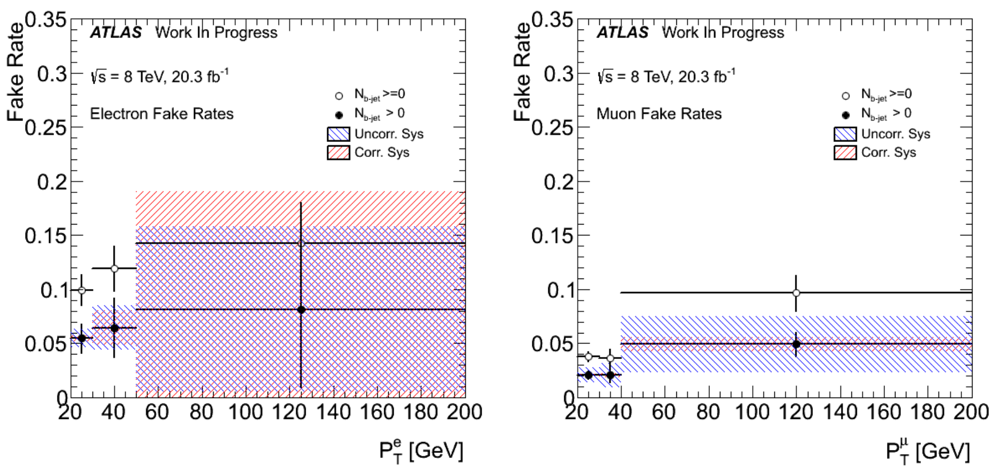


Figure 5.15 Electron (left) and muon (right) fake efficiencies as a function of p_T measued in the control regions with different requirement on N_{b-jet}. Errors include statistical and systematic uncertainties.

	ζ	σ_{stat}	σ_{sys}^{uncorr}	σ_{sys}^{corr}
	$N_{b-jet} > 0$			
$p_T \in [20, 30]$ GeV	0.0549	0.0136	0.0084	0.0032
$p_T \in [30, 50]$ GeV	0.0645	0.0272	0.0203	0.0161
$p_T > 50$ GeV	0.0816	0.0723	0.0764	0.1088
	$N_{b-jet} \geqslant 0$			
$p_T \in [20, 30]$ GeV	0.0995	0.0141	0.0270	0.0099
$p_T \in [30, 50]$ GeV	0.1192	0.0208	0.0324	0.0232
$p_T > 50$ GeV	0.1428	0.0374	0.0428	0.0674

Table 5.15 Electron fake efficiencies as a function of p_T measured in control regions with different requirement onN_{b-jet}. Errors include statical and systematic uncertainties.

	ζ	σ_{stat}	σ_{sys}^{uncorr}	σ_{sys}^{corr}
	$N_{b-jet} > 0$			
$p_T \in [20, 30]$ GeV	0.0208	0.0037	0.0067	0.0009
$p_T \in [30, 40]$ GeV	0.0207	0.0066	0.0113	0.0020
$p_T > 40$ GeV	0.0492	0.0109	0.0259	0.0068
	$N_{b-jet} \geqslant 0$			
$p_T \in [20, 30]$ GeV	0.0378	0.0046	0.0140	0.0040
$p_T \in [30, 40]$ GeV	0.0360	0.0091	0.0096	0.0089
$p_T > 40$ GeV	0.0967	0.0166	0.0252	0.0244

Table 5.16 Muon fake efficiencies as a function of p_T measured in control regions with different requirement onN_{b-jet}. Errors include statical and systematic uncertainties.

A closure test is performed for the matrix method at pre-selection region which ensures sufficient statistic. Fake estimation via matrix method is compared to the major fake background (Z+jets and $t\bar{t}$) using MC. The comparison is shown in Figure 5.16, good agreement between estimation of matrix method and MC prediction indicates that the matrix method is performing well.

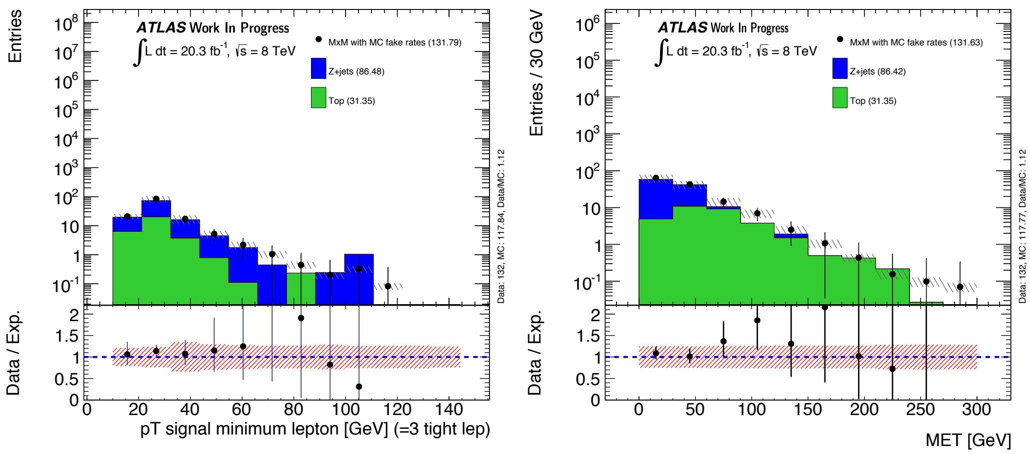


Figure 5.16 Distributions of the third leading lepton p_T and E_T^{miss} in the event pre-selection region, for Z+jets and $t\bar{t}$, compared to fake estimation using the matrix method. Good agreement is observed

5.3 Systematic Uncertainties

Various systematic uncertainties are taken into account in this analysis including both theoretical and experimental uncertainties. Uncertainties on theoretical cross sections are taken into account individually for each MC sample during normalization. The uncertainties on PDF and scale choice are also included for the signal samples.

In this analysis, the uncertainty on luminosity is found to be about 1.9%[58]. Systematic uncertainties of data-driven estimated backgrounds such as background due to lepton's charge mis-identification or fake leptons have been discussed in previous chapters. The uncertainties on WZ/ZZ correction factor describe in previous chapter is also included. Uncertainties from the reconstruction of physics objects such as the uncertainties on electron identification efficiency, E_T^{miss} soft term scale, jet energy resolution, etc. are also included. The final systematic uncertainties on total background are around 2%, 15%, and 10% for 0 SFOS, 1 SFOS and 2 SFOS signal regions respectively while on signal are around 1% for the three signal regions.

5.4 Statistical Interpretation

Definitions of pre-selection region and signal regions in this analysis are already stated in Section 5.1. With all background estimation, data-driven or MC based, performed, the comparison between the data and the background estimation is shown in Figure 5.17 for pre-selection region and signal region. Good agreement between the data and the background estimation indicates that background is well controlled in this analysis.

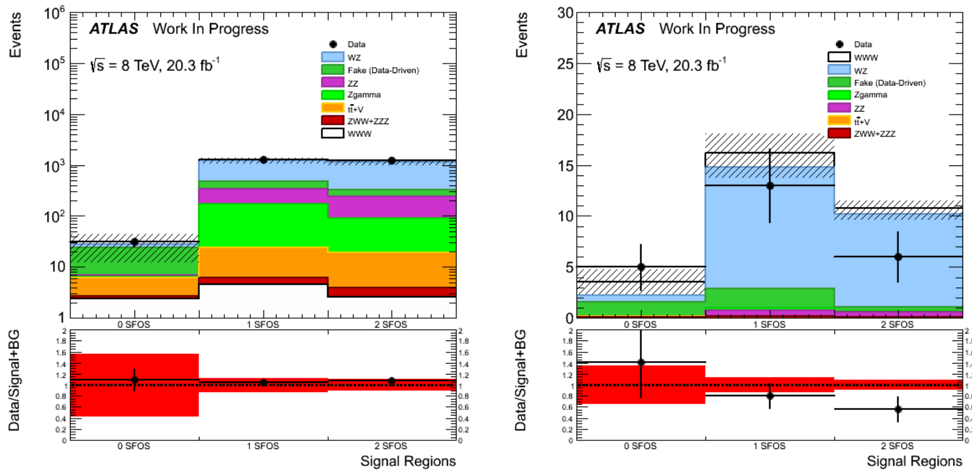


Figure 5.17 Comparison between the data and the background estimation in pre-selection region (left) and signal region (right).

5.4.1 Measurement

In this analysis we seek to measure the cross section (σ) of the WWW process in the fully leptonic channel. This total cross section σ can be expressed as:

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L} \cdot \epsilon \cdot \mathcal{A}}, \tag{5.9}$$

where N_{obs} is the number of observed data events, N_{bkg} is the number of background events estimated with either MC or data-driven methods, ϵ is the reconstruction efficiency and $\mathcal{A}$ is the detector acceptance. N_{obs} and N_{bkg} are fixed after the event selection and background estimation. ϵ and $\mathcal{A}$ are derived from the signal MC.

5.4.2 Fiducial Cross Section

Table 5.17 shows the definitions of fiducial region for the three channels:

These fiducial selections are determined by utilizing Rivet[59] at truth level. Only prompt leptons, these not originating from hadron decays, are used for lepton selections and these leptons are dressed with photons with a cone of ΔR = 0.1. Generator-level jets

	0 SFOS	1 SFOS	2 SFOS
All	All		
Tau Veto	$N_\tau < 1$		
Fiducial Leptons	Exactly 3 leptons with $p_T > 20$ GeV and $\lvert\eta\rvert < 2.5$		
Lepton Overlap Removal	$\Delta R(\ell\ell) > 0.1$		
Same-Flavor Mass	$m_{\text{SF}} > 20$ GeV		
Z-Veto ($m_Z = 91.1876$ GeV)	$\lvert m_{ee} - m_Z\rvert > 15$ GeV	No m_{SFOS} with $m_Z - 35\text{GeV} < m_{\text{SFOS}} < m_Z + 20$ GeV	$\lvert m_{\text{SFOS}} - m_Z\rvert > 20$ GeV
Missing E_T		$E_T^{Miss} > 45$ GeV	$E_T^{Miss} > 55$ GeV
Lepton-Missing E_T Angle	$\lvert\phi(3l) - \phi(E_T^{Miss})\rvert > 2.5$		
Inclusive Jet veto	$N_{jet} \leqslant 1$ with fiducial jets of $p_T > 25$ GeV and $\lvert\eta\rvert < 4.5$		

Table 5.17 Definition of fiducial regions.

are reconstructed by the anti-k_t algorithm with a radius of ΔR = 0.4. E_T^{miss} is calculated with all generator-level neutrinos. Events with τ leptons decaying from W boson are removed, therefore the $W \to \tau\nu$ branch is not included in the calculation of fiducial cross section. Table 5.18 shows the final fiducial cross section derived from the MadGraph signal samples.

	Fiducial Cross-section [ab]	
Channel	MadGraph	VBFNLO
0 SFOS	114.7 ± 4.3	126.9 ± 1.0
1 SFOS	126.6 ± 4.3	126.1 ± 1.0
2 SFOS	50.2 ± 2.7	$50.62 \pm .66$

Table 5.18 Fiducial cross section for NLO MadGraph samples with CT10 NLO PDFs.

5.4.3 Likelihood Fit

In this analysis a likelihood ratio method is used to compute the discovery significance and exclusion limits. Assume that the number of observed data and estimated background following Poisson distribution, the luminosity of the dataset and the nuisance parameters (correction factors, systematic uncertainties, etc.) following Gaussian distribution and uncertainties of the nuisance parameters are constrained to be $\pm 1\sigma$, the likelihood can be written as:

$$L(\mu, \theta) = \text{Gaus}(\mathcal{L}; \mathcal{L}_0, \Delta_{\mathcal{L}}) \prod_{i \in \text{Chan}} \text{Pois}(N_i^{obs} | N_i^{exp}(\mu, \theta)) \prod_{j \in \text{Sys}} \text{Gaus}(\theta_j; \theta_j^0, 1) \tag{5.10}$$

where μ is the so called signal strength which is defined as: $\sigma^{\text{Observed}} = \mu \cdot \sum_{i \in \text{Channels}} \sigma_i^{\text{Fiducial}}$ and θ represents the nuisance parameters. Note that the systematic uncertainties are constrained with $\pm 1\sigma$ uncertainties.

Two hypotheses are defined, one is background only hypothesis and the other one is "signal + background" hypothesis, thus a statistic is constructed using the log likelihood ratio as:

$$-2\ln\lambda(\mu) = -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}\hat{\boldsymbol{\theta}})}, \tag{5.11}$$

the denominator is an unconditional maximum likelihood evaluated at the estimators $\hat{\mu}$ and $\hat{\theta}$, the numerator is a conditional maximum likelihood which depends on μ and evaluated at the conditional maximum likelihood estimator for the set of nuisance parameters which depends on μ. Log likelihood is chosen since the logarithm is monotonic increasing function and the logarithm will make life much easier to calculate derivatives, etc. The choice of negative is due to fact that we are using the "Minuit" package to perform the maximization. There are a lot of parameters in the original likelihood, however, in this analysis, the only parameter of interest is the signal strength μ, the rest parameters such as shape uncertainties and normalization uncertainties are treated as nuisance parameters. The range of the constructed statistic is between 0 and 1, with value close to 0 showing more agreement with the background only hypothesis while value close to 1 showing more agreement with the signal hypothesis.

Denote the negative log profile likelihood of background only hypothesis ($\mu = 0$) as q_0, the probability density function can then be obtained through toy MC. The evaluated likelihood of q_0 is shown in Figure 5.18 together with the expected and observed values. The measurement of signal strength is obtained through looking for the minimum negative log likelihood for each channel and also the combination of all channels. Uncertainty on the measurement is taken by looking at the shape of the likelihood contour using Wilk's theorem. The uncertainty estimated with all systematic uncertainties as nuisance parameters is in factor a total uncertainty which includes statistical and systematic uncertainties and is considered as a quadrature combination of statistical part and systematic part. Therefore, the statistical part of the uncertainty is obtained through the evaluation without systematic uncertainties. Figure 5.19 shows the contour of negative log likelihood for the combination of all three channels. The expected fiducial cross section is:

$$\sigma^{\text{Expected}} = 309.2^{+434}_{-338}(\text{stat})^{+314}_{-342}(\text{sys})\ \text{ab}, \tag{5.12}$$

while the observed fiducial cross section is:

$$\sigma^{\text{Observed}} = 313.5^{+348}_{-332}(\text{stat})^{+322}_{-346}(\text{sys})\ \text{ab}. \tag{5.13}$$

In the absence of $W^{\pm}W^{\pm}W^{\mp}$ production, the observed (expected) upper limits on the fiducial cross section with 95% CL is 1.3 fb (1.1fb).

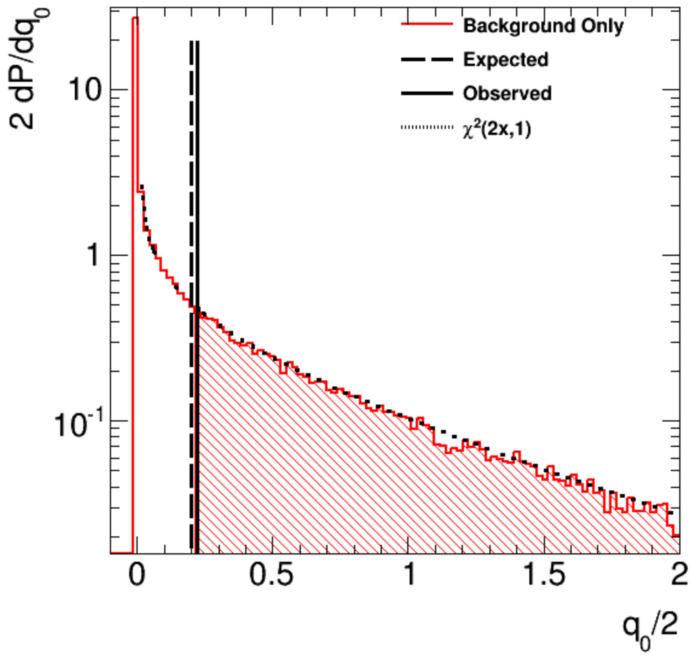


Figure 5.18 Likelihood of the background only hypothesis as a function of q_0 for the combination of all three channels. The solid black line represents the observed value of q_0 seen in the data. The shaded area above this line represents the null p-value or the integral of the background hypothesis in the signal-like region. The dotted black curve shows a χ^2 distribution for 1 degree of freedom with which it can be seen is a good approximation of the the background only PDF.

5.4.4 Anomalous Quartic Gauge Coupling

Fiducial cross section extracted in Section 5.4.2 is then used for the study of anomalous quartic gauge coupling (aQGC). Profile likelihood method is chosen to compute the aQGC limits, the likelihood is written as:

$$L(\mu, \theta) = \prod_{i=0}^{m} \mathrm{Pois}(N^i_{data}, \psi^i(\mu, \theta)) \times \left(\frac{1}{2\pi}\right)^m e^{-(\theta \cdot C^{-1} \cdot \theta)/2} \tag{5.14}$$

$$\psi^i(\mu, \theta) = N^i_{sig}(\mu) \times (1 + \theta^i) + N^i_{bg} \times (1 + \theta^{i+m}) \tag{5.15}$$

where μ is the aQGC parameter, θ stands for the nuisance parameters and C is the uncertainty matrix defined as $C_{ij} = \sum_k \sigma_{ik}\sigma_{jk}$. The observed and expected number of events together with their uncertainties are fed into the TGClim package[60] to compute the limits. The procedure to derive the limits is also based on the likelihood ratio test, upper limits of the aQGC parameters are derived using the constructed statistic (the negative log likelihood ratio). 36 aQGC samples with different parameter of cross section are produced, a 2 dimensional fit is then performed to describe the aQGC in full space:

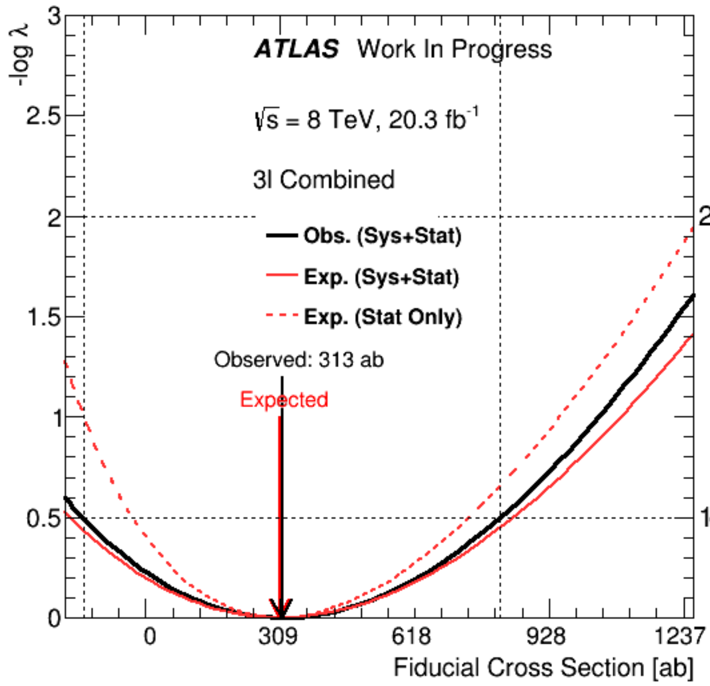


Figure 5.19 The profile likelihood contours evaluated as a function of the signal strength for the combination of all three channels. The observed (black) and expected (red) contours are shown when considering only statistical uncerataincy (dashed line) and when considering both statistical and systematic uncertainties (solid line). The dotted black lines pinpoint the location of the 1 σ and 2 σ total Gaussian uncertainties on the measurement of the signal strength which corresponds to the minimum value of the contour.

$$N_{aQGC}(f_{S,0}/\Lambda^4, f_{S,1}/\Lambda^4) = w_0 + w_1\frac{f_{S,0}^2}{\Lambda^8} + w_2\frac{f_{S,1}^2}{\Lambda^8} + 2w_3\frac{f_{S,0}}{\Lambda^4} + 2w_4\frac{f_{S,1}}{\Lambda^4} + 2w_5\frac{f_{S,0}f_{S,1}}{\Lambda^8} \tag{5.16}$$

Figure 5.20 shows the 2D fit with normalization to 20.3 fb^{-1}.

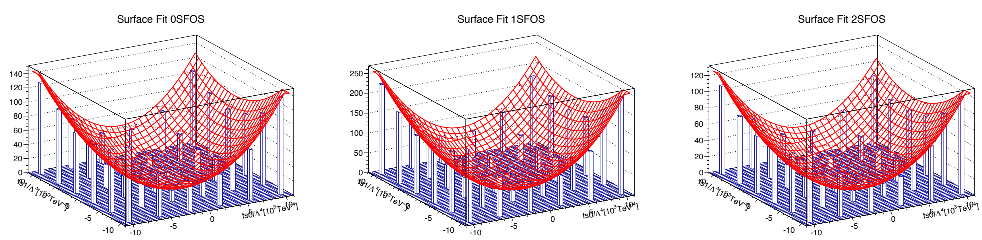


Figure 5.20 Parameterization of the signal as a function of $f_{S,0}/\Lambda^4$ and $f_{S,1}/\Lambda^4$ in the 0SFOS, 1SFOS and 2SFOS channels.

A validation procedure is performed to check the technical setup of the TGClim package, reproduced result of the $Z\gamma$ is compared to the original result of the analysis[55], Table 5.19 shows the comparison.

	H3gamma		H4gamma	
	$Z\gamma$ limits	Reproduced limits	$Z\gamma$ limits	Reproduced limits
Expected limits	(-1.8E10-3, 1.8E10-3)	(-1.81E10-3, 1.79E10-3)	(-6.1E10-6, 6.10E10-6)	(–6.11E10-6, 5.90E10-6)
Observed limits	(-8.6E10-4, 9.1E10-4)	(-8.66E10-4, 8.96E10-4)	(-3.1E10-6, 3.0E10-6)	(-3.13E10-6, 3.08E10-6)

Table 5.19 Comparison between the reproduced $Z\gamma$ limits and the original results from $Z\gamma$ group.

Channel	Expected Limit				Observed Limit					
Units: 10^3 TeV^{-4}	Limits on $f_{S,0}/\Lambda^4$		Limits on $f_{S,1}/\Lambda^4$		Limits on $f_{S,0}/\Lambda^4$			Limits on $f_{S,1}/\Lambda^4$		
	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Measured	Lower Limit	Upper Limit	Measured
Scale 500	-13.61	15.38	-17.69	21.02	-10.75	12.30	0.7 ± 7.5	-13.16	16.07	1.34 ± 8.9
Scale 1000	-6.03	7.31	-8.32	10.05	-4.57	5.63	0.5 ± 3.2	-6.09	7.66	0.7 ± 4.1
Scale 2000	-3.46	4.48	-5.04	6.27	-2.50	3.49	0.5 ± 1.8	-3.56	4.69	0.5 ± 2.5
Scale 3000	-2.82	3.83	-4.15	5.34	-1.98	2.95	0.5 ± 1.5	-2.89	3.96	0.5 ± 2.0
Un-unitarized limts	-2.18	3.14	-3.35	4.27	-1.39	2.38	0.5 ± 1.2	-2.29	3.15	0.4 ± 1.6

Table 5.20 Expected and observed limits on the aQGC Parameters.

The expected and observed limits are computed with a frequentist approach using pseudo experiments, i.e 5000 toy experiments performed with MC. Various unitarization scenarios using form factor method are considered. The 95% CL limits are derived and shown in Table 5.20, 2D limits are also shown in Figure 5.21.

5.5 Semi-Leptonic Channel

The search for $W^{\pm}W^{\pm}W^{\mp}$ production is not only conducted in the full leptonic channel but also the semi-leptonic channel where the two same sign W bosons decay to leptons (electrons or muons) and the last one decay to hadrons. No significant excess is observed over the background in this channel. The SM prediction of its fiducial cross section is 235 ab while the measured upper limit on the fiducial cross section is found to be 1149 ab. Study on anomalous quartic gauge couplings using $WWWW$ vertex is also performed. Limits are derived on the coupling parameters $f_{s,0}$ and $f_{s,1}$ as shown in Figure 5.22. Results from the two channels are combined into one paper.

5.6 Conclusion

The first search for the $W^{\pm}W^{pm}W^{\mp}$ production process is presented in the full leptonic channel. The signal plus background expectation is found to be consistent with the observation, the sensitivity is not high due to limited statistic in the final signal region. The expected fiducial cross section is 309.2 ab while the observed number is $\sigma^{\text{Observed}} = 313.5^{+348}_{-332}(\text{stat})^{+322}_{-346}(\text{sys})\text{ab}$. Combing the results from the full-leptonic analysis and the semi-leptonic analysis, the observed 95% CL upper limit on the SM

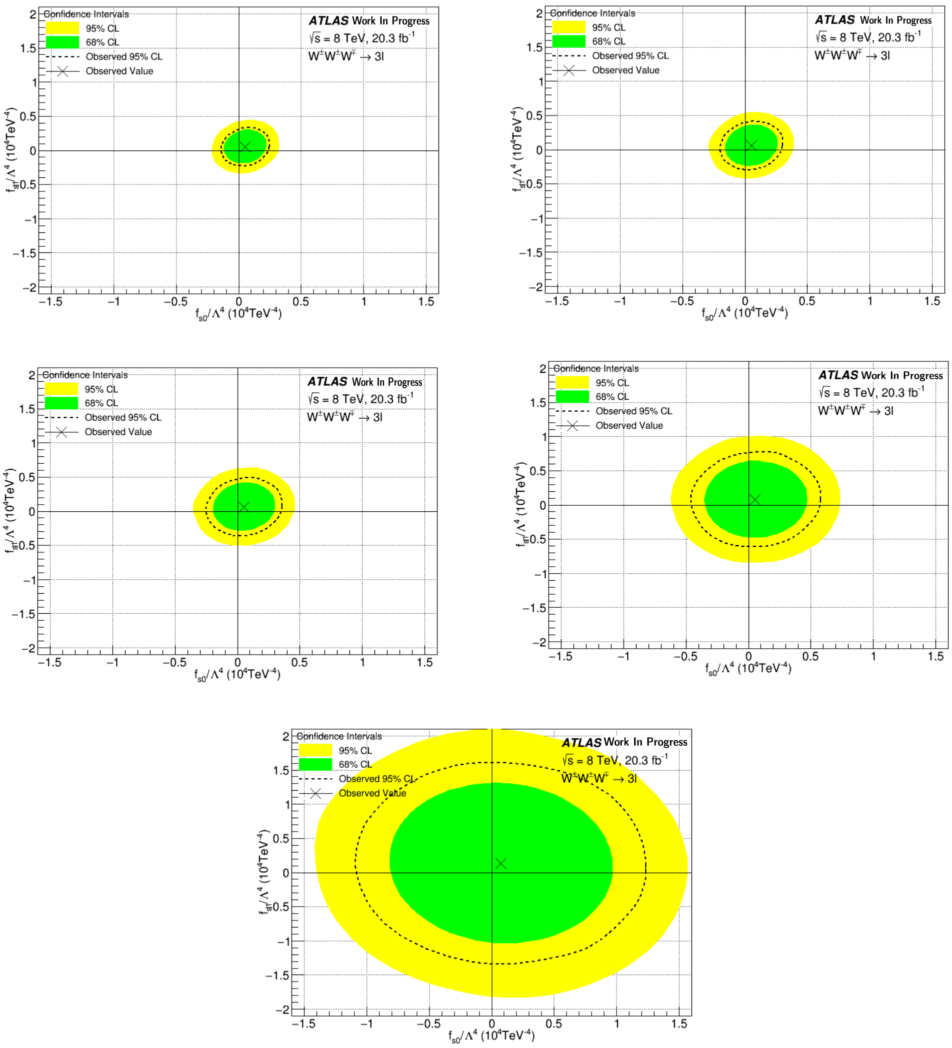


Figure 5.21 2D expected limits at 95% CL for the Un-unitarized case (top left) and three different choices of the unitarization scale, Λ: 3 TeV (top right), 2 TeV (middle left), 1 TeV (middle right), and 500 GeV (bottom). For the full leptonic channel.

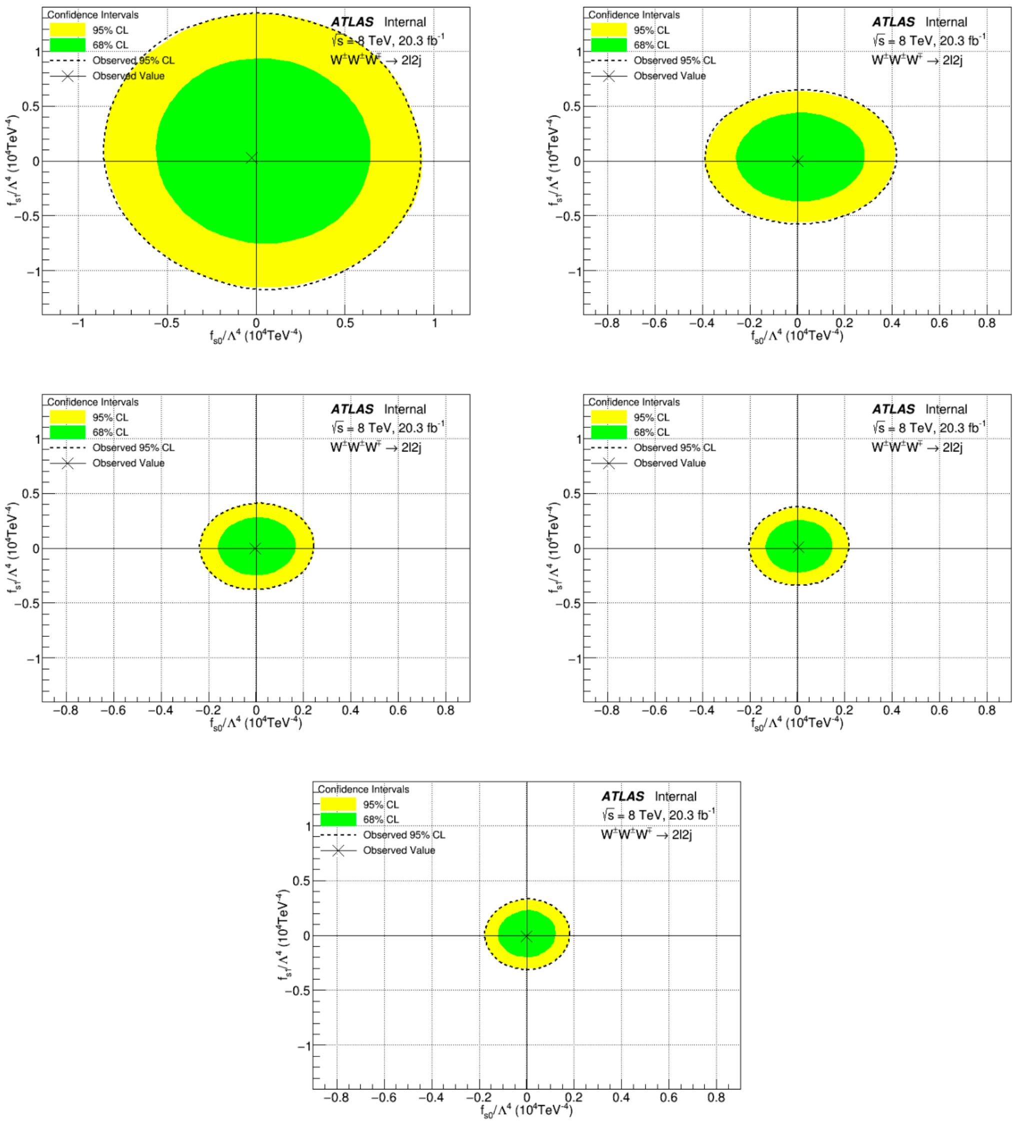


Figure 5.22 2D expected limits at 95% CL for the Un-unitarized case (bottom) and three different choices of the unitarization scale, Λ: 3 TeV (middle right), 2 TeV (middle left), 1 TeV (top right), and 500 GeV (top left). For the semi-leptonic channel.

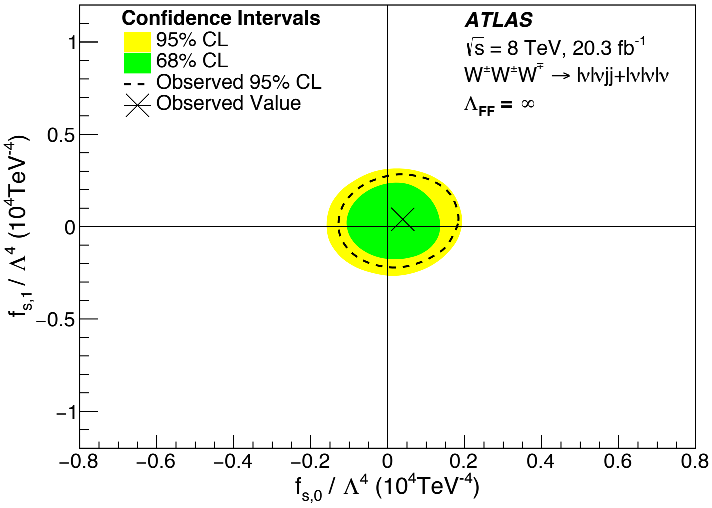


Figure 5.23 aQGC limits without form factor combining the full leptonic analysis and the semi-leptonic analysis.

$W^{\pm}W^{\pm}W^{\mp}$ cross section is found to be 730 fb with an expected limit of 560 fb in the absence of $W^{\pm}W^{\pm}W^{\mp}$ production. In addition to the measurement of the SM cross section, the study of anomalous quartic gauge coupling is also performed. Limits are set to the $f_{S,0}/\Lambda^4$ and $f_{S,1}/\Lambda^4$ dimension-8 operators of the effective field theory. Figure 5.23 shows the combined aQGC limits. The combined results are published in the EPJC[61].

Chapter 6 The Search for Doubly Charged Higgs

Neutrinos are massless in the SM, however in nature the neutrino mass is found to be small but not zero by various experiments. As an extension of the SM, the Higgs Doublet Triplet Model is introduced to allow mass for neutrinos. In this model, several new particles are predicted and the doubly charged Higgs boson we are looking for is one of them.
The search for a doubly charged scalar boson decaying to W bosons is performed using 36.1 fb^{-1} of data collected by the ATLAS detector at a center-mass-energy of 13 TeV during 2015 and 2016. A benchmarking scenario is chosen for the model in which several constraints on the parameters are applied, the doubly charged Higgs boson $H^{\pm\pm}$ are produced by pairs in the proton proton collisions and decay into W bosons. Mass range between 200 GeV to 700 GeV is explored in this analysis.

6.1 Event Selection

Data used in this analysis were collected with the un-prescaled single lepton triggers, shown in Table 6.1.

2015	**2016**
HLT_e26_lhmedium_L1EM20VH for data set	HLT_e26_lhtight_nod0_ivarloose
HLT_e60_lhmedium	HLT_e60_lhmedium_nod0
HLT_e120_lhloose	HLT_e140_lhloose_nod0
HLT_mu20_iloose_L1MU15	HLT_mu26_ivarmedium
HLT_mu50	HLT_mu50

Table 6.1 Summary of triggers used by data taking period.

Choice of event selection of this analysis is based on the final states of different channels. The event selection can be divided into two steps: pre-selection and additional selections of signal region. Pre-selection region is designed to provide sufficient statistic to perform various background estimation, on top the pre-selection region, various cuts are designed and optimized to separate signal from background to achieve a good signal significance in the signal region. Event selection starts with requirements of good physics objects (leptons, jets, E_T^{miss}) which is described below:

- Selection of the leptons are listed in Table 6.2. There are two criteria of lepton selection, one is with looser requirements denoted as “Loose”, the other one is with tighter requirements denoted as “Tight” which is a subset of “Loose”. The “Loose” criteria is mainly designed for the background estimation which will be described later. The identification algorithm of electrons is based on a likelihood which is quite different from Run 1.

- Jets are reconstructed from topological clusters using the anti-k_T algorithm with a radius of $\Delta R = 0.4$. Jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. To suppress the jets from pile-up events, jets with $p_T < 60$ GeV and $|\eta| < 2.4$ are required to have JVT (Jet-Vertex-Tagger) greater than 0.59. Jets containing $b-$hadrons are tagged as b-jet using the MV2c10 algorithm with a working point of 70%.

- To avoid duplications between the reconstructed physics objects, several overlap removal strategies are applied on the objects passing selections. Table 6.3 shows the details of overlap removal between objects.

- E_T^{miss} is calculated as the negative vector sum of the momentum of the calibrated objects and of the soft-terms. The objects are the calibrated leptons with selection and calibrated jets without selection while the soft-term refer to soft-event contribution which is reconstructed from tracks or calorimeter cell clusters not associated with the hard objects.

Lepton	Electrons		Muons	
Condition	Loose	Tight	Loose	Tight
P_T	$P_T^e > 10$ GeV		$P_T^e > 10$ GeV	
Pseudo-rapidity	$\|\eta_e\| < 2.47$, not in crack $[1.37, 1.52]$		$\|\eta_\mu\| < 2.5$	
Identification	LooseLH	TightLH	Loose	Loose
Isolation	Loose	FixedCutTight	LooseTrackOnly	FixedCutTightTrackOnly
PV longitudinal	$\|z_0 \sin\theta\| < 0.5$ mm	$\|z_0 \sin\theta\| < 0.5$ mm	$\|z_0 \sin\theta\| < 0.5$ mm	$\|z_0 \sin\theta\| < 0.5$ mm
PV transverse	$\|d_0/\sigma_{d_0}\| < 5$	$\|d_0/\sigma_{d_0}\| < 5$	$\|d_0/\sigma_{d_0}\| < 3$	$\|d_0/\sigma_{d_0}\| < 3$

Table 6.2 Selection of electron and muons used in the analysis. Likelihood electron identification is adopted, the loose and tight are different working points according to the identification power.

In the pair production mode of $H^{\pm\pm}$, four W bosons will be produced which decay to several different final states. Three channels are defined in this analysis according to three different final states, $2\ell^{SS}$ (two same sign leptons), 3ℓ and 4ℓ. Figure 6.1 shows the topology of the three different decays. Topologies between the three channels are

Keep	Remove	Cone size (Δ R) or track
electron	electron (low p_T)	0.1
muon	electon	0.1
electron	jet	0.3
jet	muon	$\Delta R < min(0.4, 0.04 + 10/P_T(\text{muon})[\text{GeV}])$

Table 6.3 Summary of overlap removal between electrons, muons and jets. The tau hadronic decays are not treated in this event final state decomposition and are part of the hadronic final state (included in jets reconstruction).

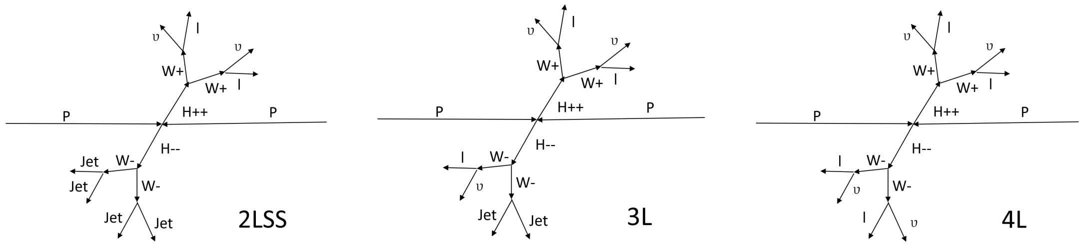


Figure 6.1 Illustrations of event topologies of the signal process for the three channels, $2\ell^{SS}$, 3ℓ and 4ℓ from the left to the right.

quite different, therefore it's necessary to design event selections for the three channels separately. For the $2\ell^{SS}$ channel, two W bosons from one doubly charged Higgs decay to leptons and neutrinos while the other two W bosons from the other doubly charged Higgs decay to hadrons, thus events are required to have a pair of same sign leptons, certain E_T^{miss}, a certain number of jets, invariant mass of the jets and constraints on several other variables defined with the angular correlations between the leptons, leptons and jets. For the 3ℓ channel, three of the W bosons decay to leptons and neutrinos while the last one decay to hadrons, therefore events are selected with exactly three leptons, E_T^{miss}, invariant mass of jets and several other variables on angular correlation. For the 4ℓ channel, all of the four W bosons decay to leptons and neutrinos, therefore, events must have four leptons and also constraints on other variables like E_T^{miss} and angular variables. Apart from the consideration on signal signature, some selections are taken into account to reduce background such as the Z window and $b-jet$ requirements. Z window is to reduce background from Z+jets background in the $2\ell^{SS}$ channel or di-boson background in the 3ℓ channel. Constraint on number of $b-jet$ is to control the contributions from production of top quarks in various control regions and signal region. In this analysis, various event level variables are explored on top of the pre-selection region to further separate signal from background:

- M_{jj}^W: the invariant mass of the two jets closest to the mass of W boson.

- M_{jets}: the invariant mass of all jets, only the leading four jets are considered is there are more than four jets.

- $M_{\ell\ell}$: the invariant mass of the two leptons.

- $\Delta R_{\ell\ell}$: the distance in $\eta - \phi$ plane between the two leptons.

- $\Delta\phi(\ell\ell, E_T^{miss})$: difference in azimuth between the di-lepton system and E_T^{miss}.

- RMS: variable used to describe the "spreads" of the azimuth angles of leptons, jets and E_T^{miss}:

$$\mathrm{RMS} = \frac{\mathrm{R.M.S.}(\phi_{\ell_1}, \phi_{\ell_2}, \phi_{E_T^{miss}}) * \mathrm{R.M.S.}(\phi_{j1}, \phi_{j2}, \cdots)}{\mathrm{R.M.S.}(\phi_{\ell_1}, , \phi_{\ell_2}, \phi_{E_T^{miss}}, \phi_{j1}, \phi_{j2}, \cdots)}, \tag{6.1}$$

  where R.M.S refers to root mean square. In the $2\ell^{SS}$ channel, the two charged leptons tend to be close in the azimuth plane due to spin correlation while the directions of E_T^{miss} and leptons should be centralized around the Higgs[62][63]. Therefore, small spread in azimuth plane of leptons and jets is expected, the ratio defined above can be used to separate signal from background.

- $\Delta R_{\ell_i - jet}$: the minimal distance in the $\eta - \phi$ plane between a lepton and leading or sub-leading jet.

- $\Delta\phi_{E_T^{miss} - jet}$: the distance between E_T^{miss} and the leading jet in the azimuth plane.

- $M_{3\ell}$: the invariant mass of the three leptons.

- $M_{4\ell}$: the invariant mass of the four leptons.

Details of pre-selection and signal region optimization will be illustrated in next sections.

6.2 Background Estimation

6.2.1 $2\ell^{SS}$ Channel

Table 6.4 summarized the selections of event pre-selection region of $2\ell^{SS}$ channel. Three sub channels are defined according to the flavor of the two same sign leptons in the $2\ell^{SS}$ channel: ee, $\mu\mu$ and $e\mu$. Comparison between the data and the MC on jet multiplicity of the three sub channels is shown in Figure 6.2.

Trigger requirement
Two tight leptons with same sign , $p_T > 30$, 20 GeV respectively.
$\lvert M_{ll}\rvert$ <80 GeV or $\lvert M_{ll}\rvert > 100$ GeV for ee channel
No b-jet
Njets $\geqslant 3$
$E_T^{miss} > 70$ GeV

Table 6.4   Definition of event pre-selection region in the $2\ell^{SS}$ channel.

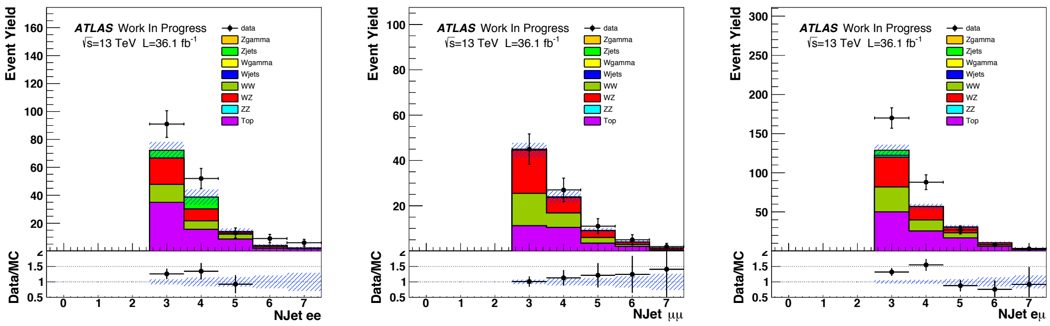


Figure 6.2   Jet multiplicities of ee, $\mu\mu$ and $e\mu$ channels from the left to the right in the event pre-selection region for the $2\ell^{SS} analysis$. The data is directly compared to the MC, errors are statistical only.

In the event preselection region, it can be observed that major backgrounds are $Z+Jets$, $t\bar{t}$, $V\gamma$ and VV processes where V stands for W or Z boson. These backgrounds can be divided into three categories:

- Background due to lepton's charge mis-identification. $Z + Jets$ and $t\bar{t}$ events can pass the same sign di-lepton requirement due to charge mis-identification of leptons.

- Background due to fake leptons: Leptons originating from hadronic decays or photon conversions and mis-reconstructed leptons from hadrons in jet are denoted as the fake leptons. $Z + Jets$, $W + Jets$ and $t\bar{t}$ events may enter pre-selection region due to this reason.

- Background with prompt same sign di-lepton events.

For background due to lepton's charge mis-identification, electron's charge mis-identification rates are measured from the data using the likelihood method, muon's charge mis-identification rate is negligible and hence ignored. For background due to fake leptons, a data-driven fake factor method is adopted to estimate this kind of background.

Prompt backgrounds with two same sign leptons such as W^+W^+ are estimated with MC simulation.

6.2.1.1 Background due to Charge Mis-Identification

In the $2\ell^{SS}$ channel, events from $Z + Jets$ and $t\bar{t}$ may enter the signal region due to electron's charge mis-identification. The electron's charge mis-identification rates are measured with the likelihood method and the truth method based on truth information in the MC which is a cross check for the likelihood rates measured from the MC. Detailed discussion of the methods will be skipped since they have been described concretely in Section 5.2.2. The rates measured in the WWW analysis is for Run 1 while this doubly charged Higgs uses Run 2 data, thus the rates need to be re-measured. The rates are measured as a function of p_T and $|\eta|$ using $Z \to ee$ events selected from the data. For further study on fake lepton background, the rates for electrons passing the loose selection but fail the tight selection (denoted as looseNotTight) need to be measured as well. Table 6.5 and Table 6.6 illustrate the binning of the rates for tight and looseNotTight scenarios.

$\|\eta\|$	[0, 0.6] [0.6, 1.1] [1.1, 1.37] [1.52, 1.7] [1.7, 2.3] [2.3, 2.47]
p_T [GeV]	[20, 60] [60, 90] [90, 130] [130, 1000]

Table 6.5 Binning of charge mis-identification rates for tight electrons.

Events for the measurement of tight rates must be within a Z mass window of 10 GeV ($80\text{GeV} < M_{\ell\ell} < 100\text{GeV}$), these events are then divided into SS region with same sign events and OS region with opposite sign events. These numbers are then passed to the likelihood and the rates are then obtained through minimizing the negative log likelihood using the Minuit package. The truth match method is applied to the Powheg $Z \to ee$ MC sample with the same event selection. Figure 6.3 shows the comparison between rates measured from Powheg $Z \to ee$ MC samples using the likelihood method and the truth method. Considering the bias from MC generators, the closure test is performed on another $Z \to ee$ MC sample generated by Sherpa, the comparison can be found in Figure 6.4, still good agreement between rates measured from two methods.

$\|\eta\|$	[0, 1.37] [1.52, 2.47]
p_T [GeV]	[20, 60] [60, 1000]

Table 6.6 Binning of charge mis-identification rates for looseNotTight electrons.

The excellent agreement observed indicates that the likelihood method is performing well.

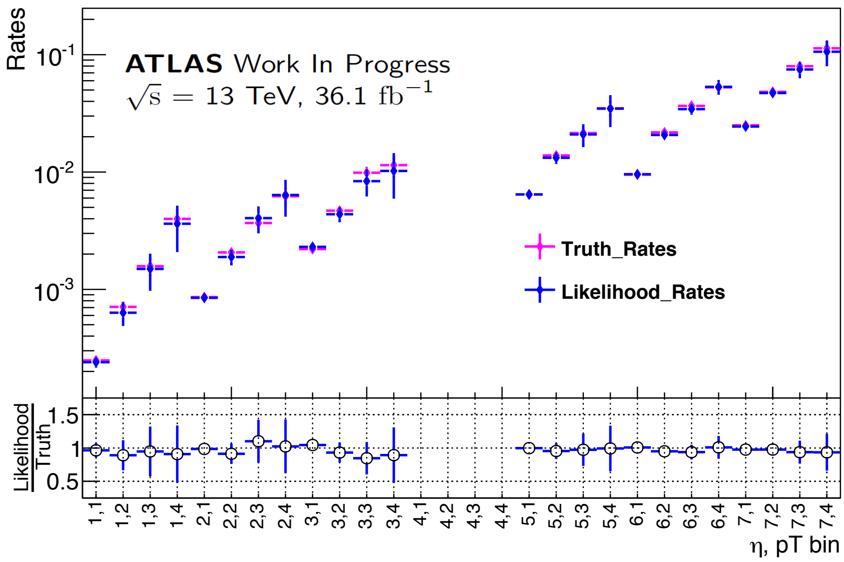


Figure 6.3 Closure test of the likelihood method: comparison of the measured electron charge mis-identification rate from the likelihood method to that determined from truth information using the same $Z \to ee$ Powheg MC sample. Note that bins 4,x are not used (they correspond to the crack).

As for the rates of looseNotTight electrons, similar likelihood method is used. Events with one tight electron and one looseNotTight electron and within a Z mass window of 10 GeV ($80\text{GeV} < M_{ee} < 100\text{GeV}$) are selected for the measurement. A coarse binning is chosen for the looseNotTight rates due to statistic constraints which is described in Table 6.6. To measure the rate of looseNotTight electron in the endcap region, the tight electron is required to be located in the barrel region. For the rate of loosNotTight electron in the barrel region, the tight electron is required to be located in the endcap region. The rates of tight electrons are fixed using the values measured before when constructing the likelihood to reduce the number of free parameters in the likelihood function.

Table 6.7 and Table 6.8 show the rates for tight and looseNonTight electrons measured from the data using the likelihood method, the errors in this table are statistical only.

Figure 6.5 shows the distributions of the invariant mass of the two electrons for both tight and looseNonTight scenarios. The rates used for background estimation in this analysis are obtained from the data which can be contaminated by background events, i.e non-$Z \to ee$ events.

The study of the impact from the background starts with looking for signal purities

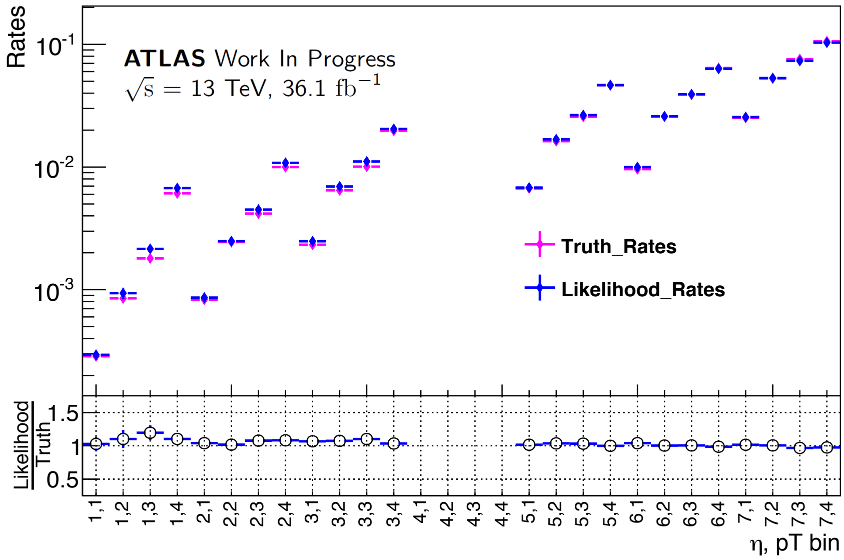


Figure 6.4 Closure test of the likelihood method: comparison of the measured electron charge mis-identification rate from the likelihood method to that determined from truth information using the same $Z \to ee$ Sherpa MC sample. Note that bins 4,x are not used (they correspond to the crack).

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 90$	$90 < p_T/\text{GeV} < 130$	$130 < p_T/\text{GeV} < 1000$
$0< \|\eta\| <0.6$	0.021±0.001	0.065±0.008	0.150±0.028	0.324±0.068
$0.6< \|\eta\| <1.1$	0.063±0.002	0.142±0.013	0.307±0.046	0.768±0.100
$1.1< \|\eta\| <1.37$	0.147±0.005	0.348±0.030	0.703±0.102	1.359±0.224
$1.52< \|\eta\| <1.7$	0.422±0.011	0.898±0.067	1.779±0.222	3.450±0.494
$1.7< \|\eta\| <2.3$	0.837±0.008	1.972±0.057	3.246±0.178	5.830±0.376
$2.3< \|\eta\| <2.47$	2.225±0.032	4.626±0.214	7.350±0.616	9.921±1.305

Table 6.7 Charge mis-identification rates as a function of p_T and $|\eta|$ for tight electrons measured from the data using the likelihood method. The values are in % and the errors are statistical only.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 1000$
$0< \|\eta\| <1.37$	0.68±0.02	3.84±0.38
$1.52< \|\eta\| <2.47$	5.37±0.04	12.18±0.47

Table 6.8 Charge mis-identification rates as a function of p_T and $|\eta|$ for looseNotTight electrons measured from the data using the likelihood method. The values are in % and the errors are statistical only.

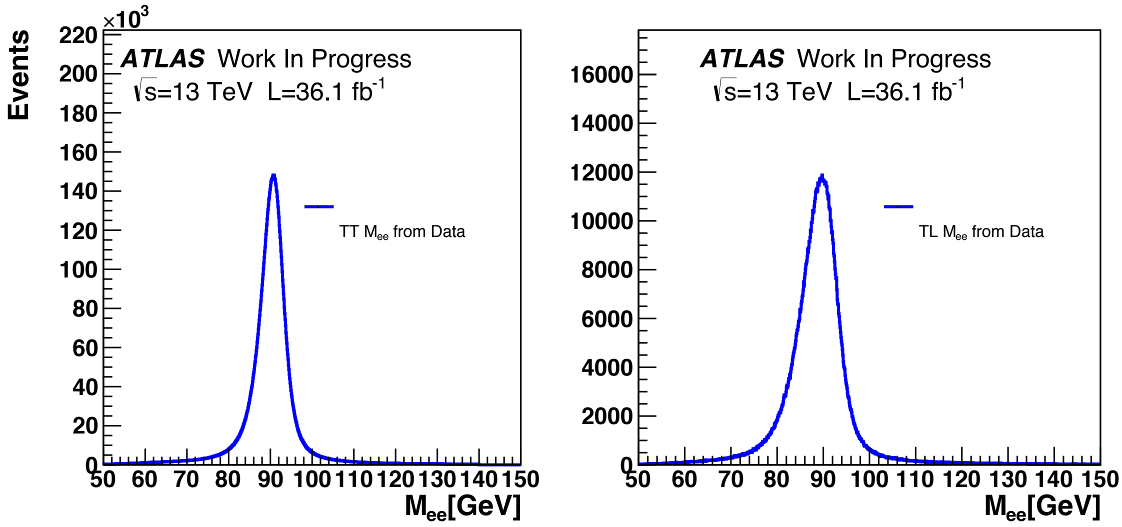


Figure 6.5 Distributions of the invariant mass of the two tight electrons (left) and one tight and one looseNotTight electron (right) selected to measure the electron charge mis-identification rate in the data (including SS and OS events)

of N_{tot} and N_{ss} (N_{tot} and N_{ss} stand for the number of total events and the number of events with two same sign electrons) which are input for the rate measurement. A Template Fit method is chosen to perform the measurement. This method has been described concretely in Section 5.2.2. The binning of tight electrons and looseNotTight electrons are tuned due to statistic limits, p_T bins are grouped together while $|\eta|$ bins stay unchanged for N_{tot} while there is only one bin for N_{ss}. The distribution of the invariant mass of the two electrons, i.e M_{ee}, is used for the fit. The key point of the template fit method is to get two templates: one signal template to describe the shape of signal and one for the background. The signal template is taken from the Powheg $Z \to ee$ MC sample while the background template is chosen to be a 2^{nd} polynomial function. An example for tight electrons is shown in Figure 6.6, fit for N_{tot} with leading electron inside $1.1 < |\eta| < 1.37$ and sub-leading electron inside $1.52 < |\eta| < 1.7$ and the global fit for N_{ss}. Example for the looseNotTight electrons is shown in Figure 6.7, fit for N_{tot} with the looseNotTight electron in the barrel region and tight electron in the end-cap region and the global fit for N_{ss}.

With the signal purities of N_{tot} and N_{ss} obtained through the template fit, the background can be subtracted, therefore, another set of rates with clean N_{tot} and N_{ss} are measured. The impact of background contamination on the charge mis-identification rates is found to be small, the difference between the two sets of rates are considered as a systematic uncertainty on the central values which are rates measured without the background subtraction. The relative uncertainties from background contamination for

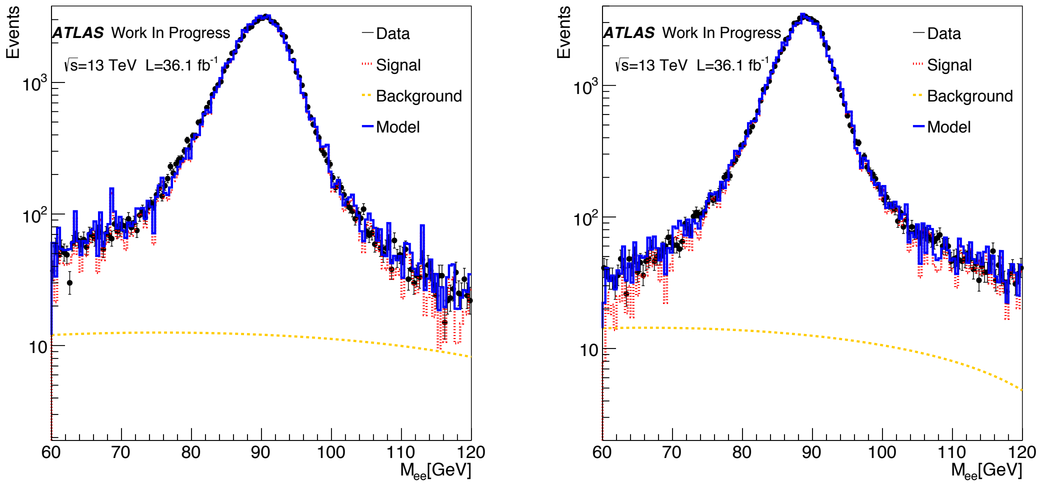


Figure 6.6 Distributions of the invariant mass of the two tight electron candidates. Left: all events with leading electron in 1.1< $|\eta|$ <1.37 and sub-leading electron in 1.52< $|\eta|$ <1.7. Signal purity obtained from the fit is 98%, χ^2/DOF is 5.7. Right: all same sign events. Signal purity obtained from the fit is 97.8%, χ^2/DOF is 3.9. Black dots show the data, the red dash line is for the signal template, the yellow dash line is for the background template and the blue dash line is the fit.

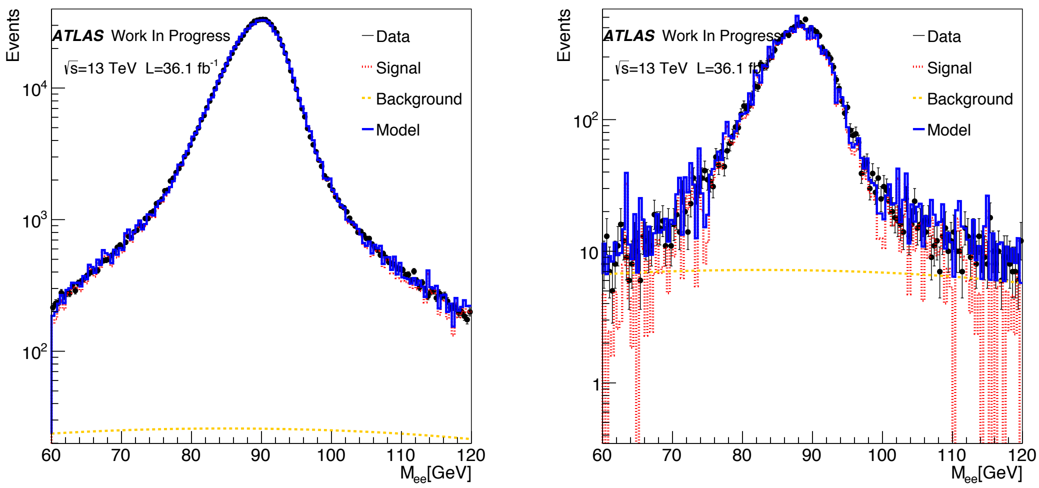


Figure 6.7 Distributions of the invariant mass of di-electron events, with one looseNotTight electron in the barrel and one tight electron in the endcap regions. Left: of all selected events (signal purity from the fit is 99.6%, χ^2/DOF is 7.9). Right: same-sign events (signal purity from the fit is 93.2%, χ^2/DOF is 4.5). Black dots show the data, the red dash line is for the signal template, the yellow dash line is for the background template and the blue dash line is the fit.

tight and looseNotTight rates can be found in Table 6.9 and Table 6.10.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 90$	$90 < p_T/\text{GeV} < 130$	$130 < p_T/\text{GeV} < 1000$
0< $\|\eta\|$ <0.6	2	2	2	2
0.6< $\|\eta\|$ <1.1	2	2	2	2
1.1< $\|\eta\|$ <1.37	1	2	2	2
1.52< $\|\eta\|$ <1.7	1	1	1	1
1.7< $\|\eta\|$ <2.3	2	2	2	2
2.3< $\|\eta\|$ <2.47	1	1	1	1

Table 6.9 Relative systematic uncertainties of tight electron's charge mis-identification rates (in %) due to background contamination.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 1000$
0< $\|\eta\|$ <1.37	15	8
1.52< $\|\eta\|$ <2.47	1	1

Table 6.10 Relative systematic uncertainties of looseNotTight electron's charge mis-identification rates (in %) due to background contamination.

In this analysis, electron's charge mis-identification rates are used to predict background in the SS region together with the yield of OS region, the probability of an OS event being identified as SS is $r = (\epsilon_1 + \epsilon_2)/(1 - \epsilon_1 - \epsilon_2)$ where ϵ_1 and ϵ_2 are the charge mis-identification rates of the two electrons. Since the prediction is performed for various physics processes not only the $Z \to ee$ process, a question then arises, can the prediction using rates measured from $Z \to ee$ process perform well for other processes? After all, the electron's kinematic distributions are quite different among difference physics processes which is shown in Figure 6.8, is the binning good of the rates fine enough to cover such difference? Therefore, another study on kinematic difference is performed. The rates measured with $Z \to ee$ MC sample are used to predict the yield of SS region for W^+W^-, $t\bar{t}$ and $Z \to ee$ processes, these three processes are the major background due to charge mis-identification in this analysis. The prediction is then compared to the number from MC simulation, the difference between prediction and MC is shown in Table 6.11.

The impact of the kinematic difference is then taken into account as an additional systematic uncertainty on the charge mis-identification rate. A systematic uncertainty of 25% is assigned for tight rates and 35% for looseNotTight rates. The total uncertainties of tight and looseNotTight rates from statistical fluctuation, background contamination and kinematic difference are shown in Table 6.12 and Table 6.13.

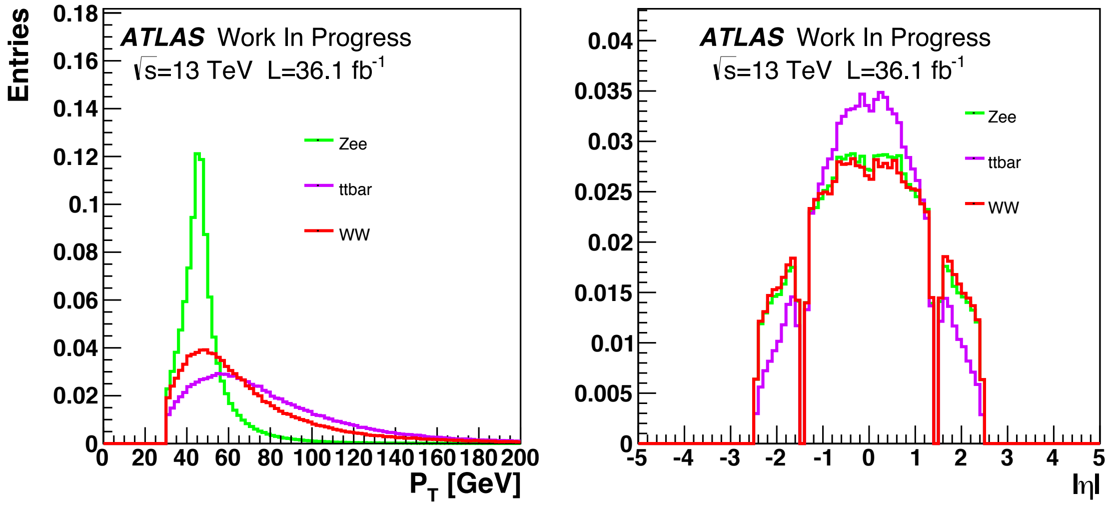


Figure 6.8 Distributions of p_T and $|\eta|$ of leading electron for $Z \to ee$, $t\bar{t}$ and W^+W^- processes. All distributions are normalized to unity to compare the shapes.

	Z+Jets	$t\bar{t}$	W^+W^-
Tight	1.1	25	25
LooseNotTight	11	35	15

Table 6.11 Difference in percentage between prediction of number of same sign events using the charge mis-identification rates measured with Z+Jets MC sample and the same sign contribution from the MC simulations. Certain requirements are applied to $Z \to ee$, $t\bar{t}$ and WW samples to get clean events. $Z \to ee$ events are required to have a pair of electrons whose invariant mass is within a Z mass window between 80 and 100 GeV, no B jet. $t\bar{t}$ events are selected with at least one B jet and E_T^{miss} above 20 GeV while WW events are selected with no B jet.

	$20 < p_T/\text{GeV} < 60$	$60 < p_T/\text{GeV} < 90$	$90 < p_T/\text{GeV} < 130$	$130 < p_T/\text{GeV} < 1000$
0< $|\eta|$ <0.6	26	28	31	33
0.6< $|\eta|$ <1.1	25	27	29	28
1.1< $|\eta|$ <1.37	25	27	29	30
1.52< $|\eta|$ <1.7	25	26	28	29
1.7< $|\eta|$ <2.3	25	25	26	26
2.3< $|\eta|$ <2.47	25	25	26	28

Table 6.12 Uncertainties on the charge mis-identification rates for tight electrons (in %). The uncertainties include statistic uncertainty, uncertainty due to background contamination and kinematic difference.

	$20 < p_T/\mathrm{GeV} < 60$	$60 < p_T/\mathrm{GeV} < 1000$
0< $\|\eta\|$ <1.37	38	37
1.52< $\|\eta\|$ <2.47	35	35

Table 6.13 Uncertainties on the charge mis-identification rates for looseNotTight electrons (in %). The uncertainties include statistic uncertainty, uncertainty due to background contamination and kinematic difference.

The background due to electron's charge flip is estimated at pre-selection level of $2\ell^{SS}$ channel using the tight rates and opposite-sign di-lepton events in the data. Prompt contribution (WZ, ZZ and $W^{\pm}W^{\pm}$) are subtracted from the opposite-sign di-lepton region using MC simulation. The estimated background contribution due to electron's charge flip are 49.20 ± 13.74 and 43.77 ± 12.04 for ee and $e\mu$ channels respectively, the uncertainties include statistical and systematic uncertainties. The looseNotTight rates are used for further study on fake lepton background which will be discussed in the next section.

6.2.1.2 Background due to Fake Lepton

Apart from the background due to electron's charge mis-identification, another important background is the background due to fake leptons where fake leptons represent non-prompt leptons originating from heavy flavor decays, mis-reconstructed leptons from hadrons in jets and electrons from photon conversion. The background due to fake lepton is estimated with a data-driven fake factor method. This method based on four disjoint regions, the four regions of $2\ell^{SS}$ channel are:

- Control region with low E_T^{miss} (< 70 GeV) and events must have one tight lepton and one looseNotTight lepton, this region is enriched by fake leptons.
- Control region with low E_T^{miss} (< 70 GeV) and events must have two tight leptons.
- Control region with high E_T^{miss} (> 70 GeV) and events must have one tight lepton and one looseNotTight lepton.
- Pre-selection region.

The two regions are defined on top of several event-level cuts, i.e trigger, trigger match, two loose same-sign leptons with p_T requirements, Z veto, b-jet veto and jet multiplicity, details can be found in Table 6.4. The control regions with one tight and one looseNotTight leptons are enriched by fake leptons. The two control regions with low E_T^{miss} are used for the measurement of the fake factors, the control region with high E_T^{miss}, one tight and one looseNotTight lepton is where the fake factor are applied to predict fake contribution in the pre-selection region. The fake factor is defined as the ratio between the number of events with two tight leptons and that with one tight and one looseNotTight lepton:

$$\theta_\ell = \frac{N_{\ell\ell}}{N_{\ell\not\ell}}, \tag{6.2}$$

where ℓ reresents tight lepton (e or μ) and $\not\ell$ represents looseNotTight lepton (e or μ). The fake factor are measured in the low E_T^{miss} control regions and then applied to the high E_T^{miss} control region. The muon fake factor is measured in $\mu\mu$ channel, the electron fake factor is measured in the $e\mu$ channel where electron's charge flip contribution is much smaller than the ee channel, the muon fake factor used to subtract fake muon contamination in the control region with one tight electron and one tight muon during the electron fake factor measurement. Therefore, the muon fake factor θ_μ is measured first in $\mu\mu$ channel as:

$$\theta_\mu = \frac{N_{\mu\mu}}{N_{\mu\not\mu}}(E_T^{miss} < 70 \text{ GeV}) = \frac{N_{\mu\mu}^{Data} - N_{\mu\mu}^{\text{Prompt } SS}}{N_{\mu\not\mu}^{Data} - N_{\mu\not\mu}^{\text{Prompt } SS}}, \tag{6.3}$$

where $N_{\mu\mu}^{Data}$ is data yield in the channel with two tight muons, $N_{\mu\not\mu}^{Data}$ is data yield in the channel with one tight muon and one looseNotTight muon, $N_{\mu\mu}^{\text{Prompt } SS}$ is the number of prompt events with two same sign tight muons and $N_{\mu\not\mu}^{\text{Prompt } SS}$ is the number of prompt events with two same sign muons where one muon is tight and one muon is looseNotTight. Then the electron fake factor θ_e is measured in $e\mu$ channel as:

$$\theta_e = \frac{N_{\mu e}}{N_{\mu\not e}}(E_T^{miss} < 70 \text{ GeV}) = \frac{N_{\mu e}^{Data} - N_{\mu e}^{\text{Prompt } SS} - N_{\mu e}^{\text{QMisId}} - N_{\mu e}^{\text{FakeMuon}}}{N_{\mu\not e}^{Data} - N_{\mu\not e}^{\text{Prompt } SS} - N_{\mu\not e}^{\text{QMisId}}}, \tag{6.4}$$

where N^{QMisID} is the contribution from electron's charge mis-identification and $N^{FakeMuon}$ is the contribution due to fake muons. During the measurement of fake factors, non-fake contributions should be subtracted. For muon, the non-fake contribution is from the prompt contribution which is estimated with MC. For electron, the situation is much more complex, the non-fake contribution can be from prompt, electron's charge flip and fake muons. The prompt part is estimated with MC simulation as well. The contamination from electron's charge flip is estimated with tight and looseNotTight charge mis-identification rates for tight+tight and tight+looseNotTight regions respectively. The tight and looseNotTight electron charge mis-identification rates are described in previous Section 6.2.1.1. The fake muon contamination in tight+tight region is estimated with the measured muon fake factor. The measured muon fake factor is 0.14±0.03 (statistic uncertainty only), the data and predictions of the control regions obtained during the measurement are listed in Table 6.14. the electron fake factor is 0.48±0.07

	data	VV_Prompt	VH	ttH+ttV	Vgamma
Numerator	139	75.2±4.5	1.5±0.6	7±0.2	0
Denominator	416	20.4±2	0	0.94±0.09	0

Table 6.14 Estimation for Prompt SS and data for muons in tight+tight and tight+looseNotTight region, uncertainty here is statistical only.

(statistic uncertainty only), the detailed components of the control regions during the measurement are shown in Table 6.15.

	data	VV_Prompt	VH	ttV	QMisID	Fake Muon
Numerator	444	135.3±5.6	4.0±1.2	11.3±0.3	47.9±1.2	71.5±3.3
Denominator	434	22.3±2.3	1±0.6	1.4±0.1	50±2.2	neglected

Table 6.15 Estimation for Prompt SS, electron charge flip, fake muon contamination and data for electrons in tight+tight and tight+anti-tight region, uncertainly here is stat-only.

The fake factors measured in low E_T^{miss} control regions are then applied in the high E_T^{miss} control region to predict fake contribution in pre-selection region:

$$N_{ee}^{fakes}(\mathrm{E}_T^{miss} \geqslant 70\ \mathrm{GeV}) = (N_{e\not{e}} - N_{e\not{e}}^{\mathrm{Prompt}\ SS} - N_{e\not{e}}^{\mathrm{QMisId}}) \times \theta_e, \tag{6.5}$$

$$N_{\mu\mu}^{fakes}(\mathrm{E}_T^{miss} \geqslant 70\ \mathrm{GeV}) = (N_{\mu\not{\mu}}^{Data} - N_{\mu\not{\mu}}^{\mathrm{Prompt}\ SS}) \times \theta_\mu, \tag{6.6}$$

$$N_{e\mu}^{fakes}(\mathrm{E}_T^{miss} \geqslant 70\,\mathrm{GeV}) = (N_{e\not{\mu}} - N_{e\not{\mu}}^{\mathrm{Prompt}\ SS} - N_{e\not{\mu}}^{\mathrm{QMisID}}) \times \theta_\mu + (N_{\mu\not{e}} - N_{\mu\not{e}}^{\mathrm{Prompt}\ SS} - N_{\mu\not{e}}^{\mathrm{QMisID}}) \times \theta_e. \tag{6.7}$$

At pre-selection level, the estimated fake lepton backgrounds are 65.20±17.32, 26.49±13.43 and 117.07±36.12 for ee, $\mu\mu$ and $e\mu$ channels respectively. Uncertainties taken into account for the fake factors are:

- the statistical uncertainty.
- the uncertainty due to jet composition (the difference in the fractions of heavy flavor jets and light flavor jets between the control region with high E_T^{miss} and the pre-selection region).
- the uncertainty due to the subtraction of the electron's charge mis-identification contribution.

- the uncertainty due to MC uncertainties such as uncertainty of cross section and uncertainty due to detector simulation, etc.
- an additional uncertainty on electron fake factor due to the subtraction of fake muon contribution in the tight+tight control region.

The uncertainty due to jet composition in taken into account because the muon fake factor are expected to be sensitive to the fraction of heavy flavor jets in the sample. To verify the sensitivity to jet composition, two control regions are defined to derive fake factors: one is enriched with heavy flavor jets by requirements on b-jet and the other one is enriched with light flavor jets by requirements on $M_{\ell\ell}$. Table 6.16 and Table 6.17 describe the selections for the two control regions.

Trigger requirement
1 tight electron with p_T > 30 GeV and two loose muons of same electric charge, with p_T > 20 GeV
At least 1 b-jet tagged with MV2c10_70 working point
$E_T^{miss} > 30$ GeV

Table 6.16 Event selections for control region enriched with heavy flavor jets. The selected events are dominated by $t\bar{t}$ events with one extra fake muon (one of the muons of same sign is real and the other is fake). The event selection to study the fake factor of electrons is the same but with electrons and muons interchanged in the table.

Trigger requirement
A pair of e^+e^- passing tight lepton cuts, $M_{e^+e^-}$ between 80 and 100 GeV one additional loose muon with $p_T > 20$ GeV.
No b-jet, MV2c10_70 working point
$E_T^{miss} < 70$ GeV

Table 6.17 Event selections for control region enriched with light flavor jets. The selected events are dominated by Z+jet and WZ. The latter is estimated with MC simulations and subtracted. The event selections to study the fake factor of electrons are the same but with electrons and muons interchanged in the table.

The measurement of lepton fake factors is applied in these two control regions. Take the muon fake factor for instance, there is one real muon and one fake muon passing loose section in the events, with prompt contribution subtracted with MC simulation, following relation should be conserved:

$$\begin{aligned} p(1) &= (1-\varepsilon_r)\varepsilon_f + (1-\varepsilon_f)\varepsilon_r \\ p(2) &= \varepsilon_r\varepsilon_f. \end{aligned}$$

where $p(1)$ and $p(2)$ are the fractions of events with one and two tight muons, ϵ_r and ϵ_f are the relative efficiencies of loose muons passing tight selection for real and fake muons respectively. Therefore, the muon fake factor can be obtained as $\epsilon_f/(1-\epsilon_f)$, similar procedures are performed for electrons. Table 6.18 shows the electron and muon fake factors in heavy-flavor jets enriched and light flavor jets enriched regions. The muon fake factors are very sensitive to heavy flavor jets as expected while electron fake factors are less sensitive.

	Heavy Flavor	Light Flavor
Electron	0.20±0.09	0.31±0.09
Muon	0.15±0.03	0.04±0.05

Table 6.18 Fake factors measured in heavy and light flavor jet enriched regions; uncertainties are statistical only.

To study the fractions of heavy flavor and light flavor components, a template fit method is applied in the pre-selection region to get the fractions of Z+Jets and $t\bar{t}$ contributions. The distribution of jet multiplicity is chosen for the fit. The templates of Z+Jets, $t\bar{t}$ and other components are selected from MC samples. The contribution of other component is fixed while yields of Z+Jets and $t\bar{t}$ are scaled to fit to the data. Figure 6.9 shows the fit to the data. The fraction and its uncertainty of Z+Jets and $t\bar{t}$ are obtained through the fit. By varying the fraction according to its uncertainty together with the fake factors measured in heavy/light flavor enriched regions, the impact of fraction on fake factors is observed. The variations of fake factors caused by the variation of fractions is taken into account as a systematic uncertainty of fake factors which stands for the uncertainty due to jet composition.

Systematic uncertainties of fake factors due to prompt subtraction, charge flip subtraction and the fake muon subtraction are computed by propagating the uncertainties of MC, charge flip rates and muon fake factors to the measurement. Table 6.19 and Table 6.20 summarized the uncertainties of lepton fake factors. The MC systematic uncertainty stands for the uncertainty from prompt subtraction using MC including uncertainties from detector simulation and cross section, this term is mainly from the reconstruction of jets like jet energy resolution.

The measurement of the fake factors in $2\ell^{SS}$ is applied in low E_T^{miss} region while the application is performed in high E_T^{miss} region, therefore, the stability of fake factors among difference E_T^{miss} is a prerequisite of the method. A closure test is performed for the fake factor method in regions with different E_T^{miss} conditions. Fake factors are

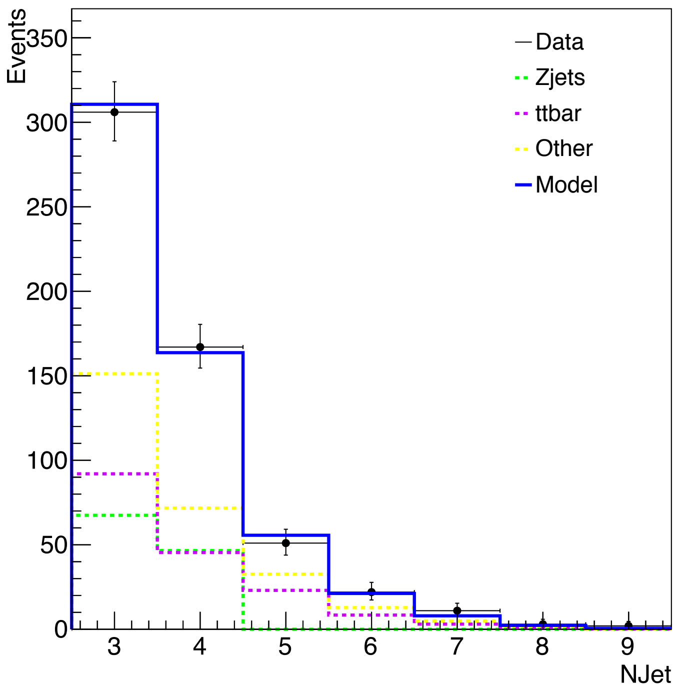


Figure 6.9 Template fit for heavy/light flavor jets study at pre-selection level. Z+Jets contribution is 113.9±40 and $t\bar{t}$ contribution is 172.5±43.7.

Source	Effect in %
Jet flavour composition	14
Pile_Up reweighting	1.5
JVT	7.4
B-jet veto	3.1
MC cross section	32
Lepton ID	3.4
Other MC Systematic Variations	38
Statistic	23
Total	56

Table 6.19 Uncertainties of muon fake factor (in %). Other MC Systematic Variations stand for the uncertainties due to detector simulation that affect the acceptance of signal region selection like uncertainty of Jet energy scales.

Source	Effect in %
QMisID	10
Fake	21
Jet composition	2
Pile_Up reweighting	1.2
JVT	4.7
B-jet veto	1.8
MC cross section	18
Electron ID	2
Muon ID	0,8
Other MC Systematic Variations	11
Statistic	14
Total	35

Table 6.20 Uncertainties of electron fake factor (in %). Note that the first two lines originate from the muon fakes and charge mis-identification contributions (see formula 6.5) and are treated as correlated in the signal extraction procedure. "Fake" uncertainty is due to the uncertainty of muon fake estimation and "QMisID" uncertainty is due to the uncertainty of QMisID estimation. Other MC Systematic Variations stand for the uncertainties due to detector simulation that affect the acceptance of signal region selection like uncertainty of Jet energy scales.

measured in regions similar to the low E_T^{miss} tight+looseNotTight control region except that the E_T^{miss} constraint is different. The result is shown in Figure 6.10, the fake factors are found to be stable among different E_T^{miss} conditions.
With the measured fake factors, the background due to fake leptons in pre-selection region is then estimated together with the high E_T^{miss} tight+looseNotTight control region.

6.2.1.3 Pre-Selection Region to Signal Region

There are major three kinds of backgrounds in the $2\ell^{SS}$ channel: background due to electron's charge mis-identification which is estimated with the electron's charge flip rates described in Section 6.2.1.1, background due to fake leptons which is estimated with the fake factor method described in Section 6.2.1.2, background due to prompt di-leptons (WZ, ZZ and $W^{\pm}W^{\pm}$) which is estimated with MC simulation. Figure 6.11 shows the comparison between the data and the background estimation of $2\ell^{SS}$ channel. The very good agreement between the data and the prediction indicate the background is well controlled. Number of each component of the backgrounds are listed in Table 6.21.

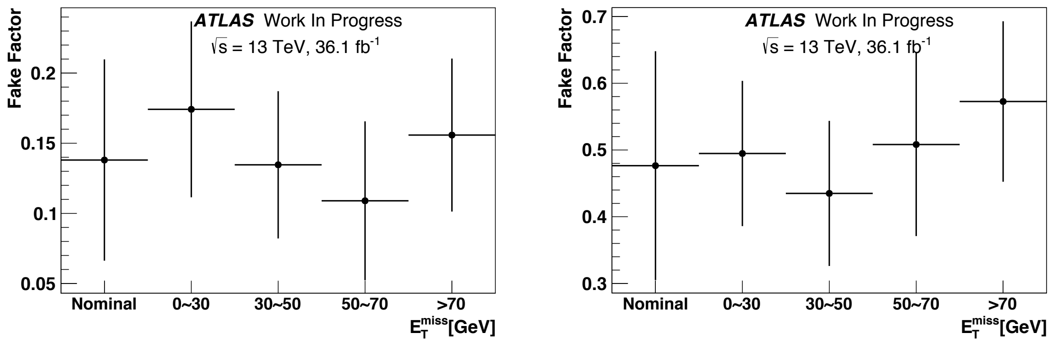


Figure 6.10 Stability of muon (left) and electron (right) fake factors. The fake factors are measured in control regions with different E_T^{miss} requirements, "Nominal" values in the plots indicate fake factors measured in the low E_T^{miss} (< 70 GeV) control region with full uncertainties, other values represent fake factors measures in different E_T^{miss} regions with statistical uncertainty only.

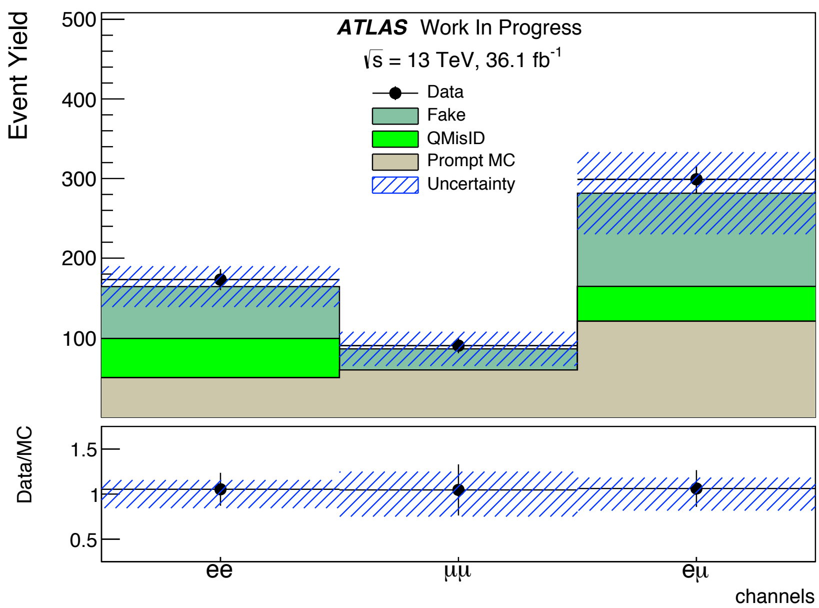


Figure 6.11 The event pre-selection region of the $2\ell^{SS}$ channel. The data (points) is compared to the prediction composed of prompt (estimated with Monte Carlo), charge misidentification and fake lepton contributions. The bottom panel shows the ratio of data to the prediction. The uncertainty includes statistic uncertainty and full systematic uncertainties.

	Prompt	Fake	QMisID	Total	Data
ee	49.90±2.98	65.20±17.32	49.20±13.74	164.58±31.51	173
$\mu\mu$	59.54±3.13	26.49±13.43	-	86.24±14.05	90
$e\mu$	120.84±4.66	117.07±36.12	43.77±12.04	282.15±50.71	299

Table 6.21 Data and background prediction at event pre-selection stage, where the QMisID and fakes contributions were estimated using the data driven methods described above.

On top of the pre-selection region, six variables are defined and optimized to further separate signal from background. These variables are E_T^{miss}, $\Delta R(\ell,\ell)$, $\Delta\phi(\ell\ell, E_T^{miss})$, RMS, $M_{\ell\ell}$ and M_{jets}, definitions of these variables are described in Section 6.1. Distributions of these variables at event pre-selection level are shown in Figure 6.12 - 6.17.

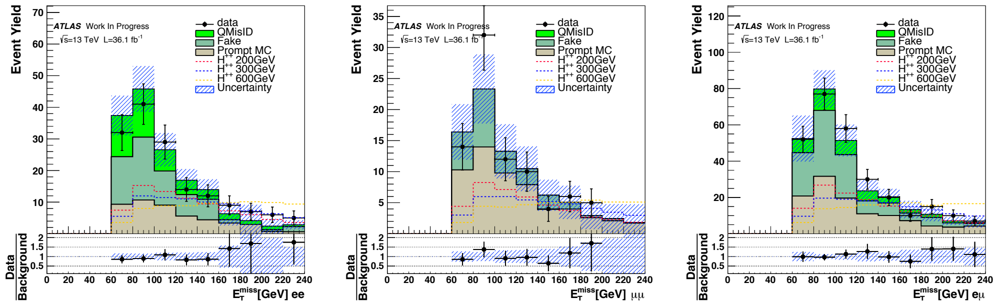


Figure 6.12 E_T^{miss} distribution at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

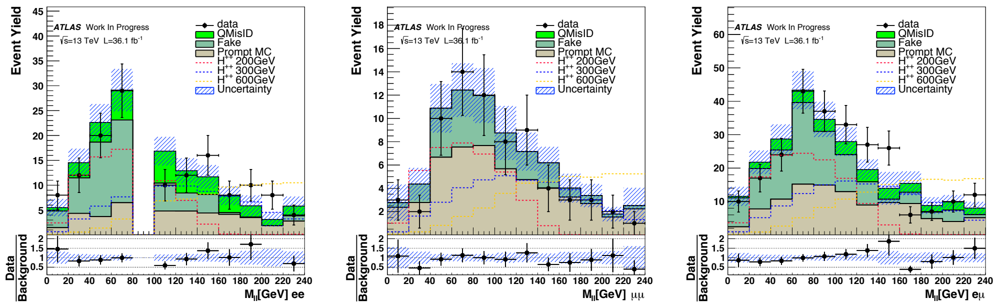


Figure 6.13 $M_{\ell\ell}$ distribution at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

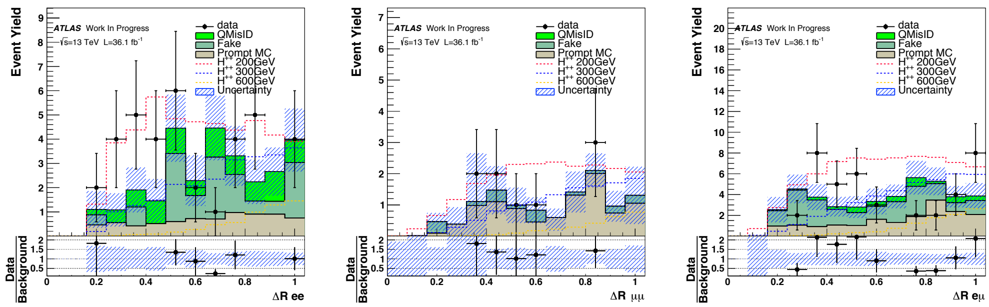


Figure 6.14 Distribution of ΔR of leptons at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

The method of rectangular cuts trained with the Simulated Annealing algorithm implemented in the TMVA tool-kit[64] is used to optimize the definitions of the signal regions.

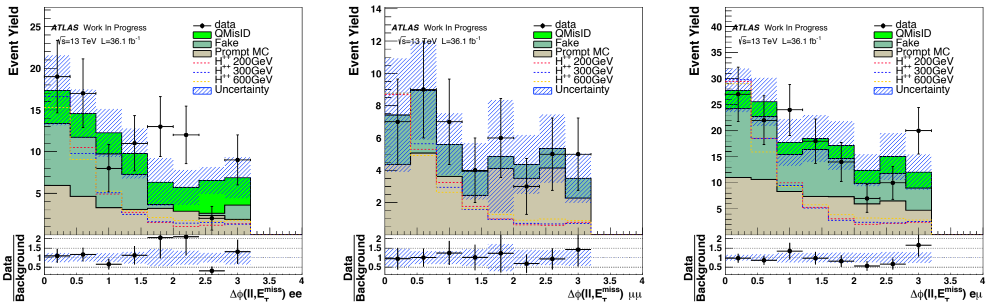


Figure 6.15 Distribution of $\Delta\phi(\ell\ell, E_T^{miss})$ at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

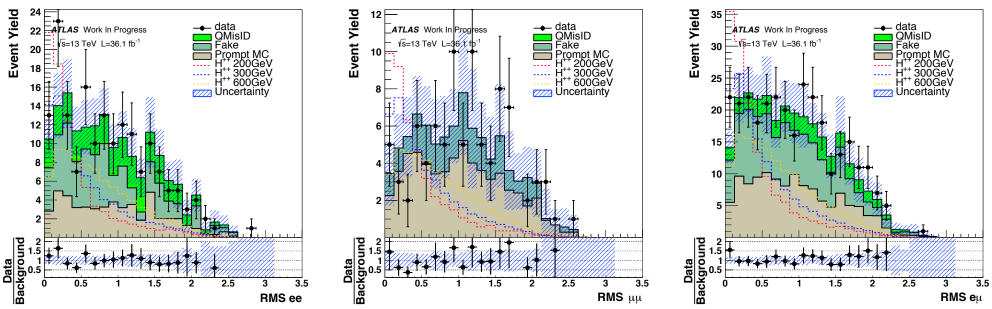


Figure 6.16 RMS distribution at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

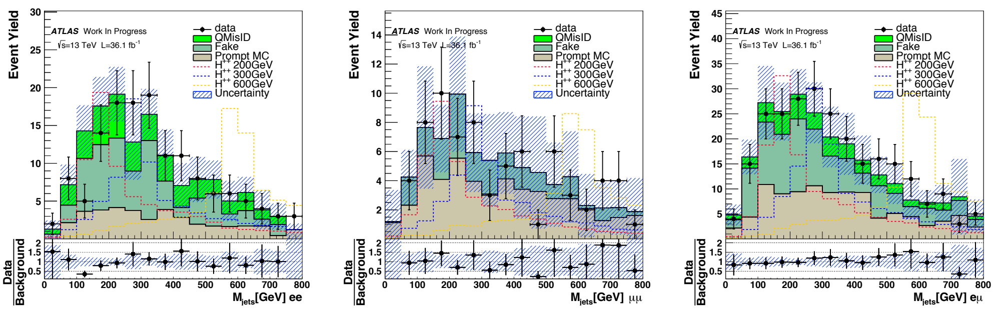


Figure 6.17 Invariant mass of all jets M_{jets} at event pre-selection stage, from left to right are ee, $\mu\mu$ and $e\mu$ channels, signal is rescaled to data for better vision. Several signal masses are shown, all uncertainties included.

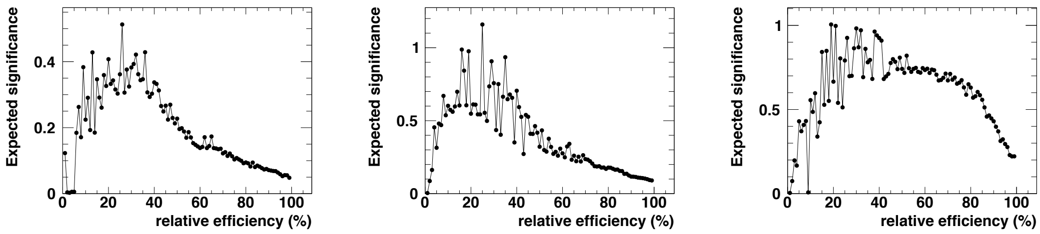


Figure 6.18 Expected significances for different working points of the signal selection for the 200 GeV mass point, Left: ee channel, Middle: $e\mu$ channel Right: $\mu\mu$ channel.

mass	ch.	$M_{jets} >$	$M_{jets} <$	RMS $<$	$\Delta R(\ell, \ell) <$	$\Delta\phi(\ell\ell, E_T^{miss}) <$	$M_{\ell\ell} <$	$M_{\ell\ell} >$	$E_T^{miss} >$
200	ee	140	770	0.3	0.8	1.1	130	25	100
300	ee	180	770	0.4	1.4	2.1	340	105	200
400	ee	280	1200	0.6	2.2	2.4	340	105	200
500-700	ee	440	∞	1.1	2.6	2.6	730	105	250
200	$\mu\mu$	95	310	0.3	1.8	1.3	150	15	100
300	$\mu\mu$	130	640	0.4	1.8	2.4	320	80	200
400	$\mu\mu$	220	1200	0.6	1.8	2.4	350	80	200
500-700	$\mu\mu$	470	∞	1.1	2.2	2.4	440	110	250
200	eμ	95	640	0.2	0.9	1.3	150	35	100
300	eμ	130	640	0.4	1.8	2.4	320	80	200
400	eμ	220	1200	0.5	1.8	2.4	350	80	200
500-700	eμ	470	∞	1.1	2.2	2.4	440	110	250

Table 6.22 Cut values for the definition of the signal regions. All numbers for masses and E_T^{miss} are in unit of GeV.

The optimization is performed independently for each mass point and each of the three channels to achieve maximum signal significance. There are 100 working points chosen during the optimization for signal efficiency from 1% to 100%. Signal significance is computed for each working point using the ttHFitter package[65], Figure 6.18 shows the expected significance as a function of signal efficiency for $M_{H^{\pm\pm}} = 200$ GeV mass point.

The working point with maximum significance is chosen as the baseline of the definition of the signal region, some cut values are refined through re-training on a subset of the cuts while fixing the some cuts and stabler cuts are preferable when the expected significance fluctuate significantly. The final cut values for signal regions are listed in Table 6.22.

6.2.2 3ℓ Channel

Event pre-selections of 3ℓ are described in Table 6.23, the pre-selections are designed in three steps denoted as A, B and C. Events must have exactly three loose leptons with a total charge of ± 1 and pass trigger, trigger match cuts. Z veto and $M_{\ell^+\ell^-}$ requirements are applied to reduce Z+Jets and low mass Drell-Yann backgrounds. B-jet veto is required to suppress the $t\bar{t}$ background.

Step	Selection Criteria
A	Three leptons with $P_T^{0,1,2} > 10, 20, 20$GeV
B	$\lvert M_{\ell^+\ell^-} - M_Z \rvert > 10$ GeV $M_{\ell^+\ell^-} > 15$ GeV $E_T^{miss} > 30$ GeV $N_{\text{jet}} >= 2$
C	$N_{\text{b-jet}} = 0$

Table 6.23 Event pre-selections for the 3ℓ channel.

There are two sources of backgrounds in the 3ℓ channel: background due to fake leptons and background with three prompt leptons. Background with three prompt leptons (WZ, ZZ, etc) is estimated with MC simulation, background due to fake leptons (mainly Z+Jets and $t\bar{t}$ events) is estimated with the data-driven fake factor method. Among the three leptons in a 3ℓ event, the same sign leptons, i.e leptons from the same doubly charged Higgs, are denoted as lepton 1 and 2, the lepton with different charge is denoted as lepton 0. $\Delta R_{01} < \Delta R_{02}$ is used for the ordering of the two same sign leptons, therefore, lepton 0 is assumed to be real and the fake lepton is always among the two same sign leptons. Similar to the fake factor method described for the $2\ell^{SS}$ channel, several regions are designed in the 3ℓ channel:

- Y: Region selected with the pre-selections described in Table 6.23 except that it's selected with low jet multiplicity ($N_{jet} = 1$) and E_T^{miss} requirement is not applied. This is where the fake factors are measured.
- X: Region using cuts described in Table 6.23.
- Z: Z+Jets enriched region. Step A, B and C listed in Table 6.23 are all used. This region is used to study the impact of light flavor jets on fake factors.
- T: $t\bar{t}$ enriched region. Step A and B in Table 6.23 are used. $N_{jets} > 2$ and $N_{b-jet} \geqslant 1$ are also required. This region is to study the impact of heavy flavor jets on fake factors.

These four regions are further divided into a region enriched with fake lepton events and a signal like region. The fake enriched region require at least one of the two same sign leptons to be looseNotTight leptons while the signal like region require the two same leptons to be both tight leptons. Therefore the fake factors are derived in the Y region using the ratio of signal-like events and fake events as:

$$\theta_{e/\mu} = \frac{(Data - N_{prompt})_{\text{xee/x}\mu\mu}}{(Data - N_{prompt})_{\text{x}e\!\!\!/ e\!\!\!/\text{/x}\mu\!\!\!/ \mu\!\!\!/}} \tag{6.8}$$

Region	Data	Prompt	Data-Prompt	MC fakes	
XF $\mathrm{x}e\not{e}$	106	33.7 ± 4.53	72.3 ± 11.2	64.7 ± 6.67	
XF $\mathrm{x}\mu\not{e}$	160	35.9 ± 3.02	124 ± 13	64 ± 8.69	
XF $\mathrm{x}e\not{\mu}$	111	12.3 ± 1.83	98.7 ± 10.7	91 ± 4.58	
XF $\mathrm{x}\mu\not{\mu}$	136	13.3 ± 2.14	123 ± 11.9	98.7 ± 5.53	
Region	Data	Prompt	Data-Prompt	DD fakes	MC fakes
XS $\mathrm{x}ee$	87	55.12 ± 3.4	31.88 ± 9.939	28.25 ± 6.736	42.5 ± 21.4
XS $\mathrm{x}e\mu$	215	135.4 ± 5.6	79.59 ± 15.69	65.64 ± 11.66	32.5 ± 5.87
XS $\mathrm{x}\mu\mu$	90	78.79 ± 4.1	11.21 ± 10.35	21.29 ± 7.108	16.8 ± 5.86

Table 6.24 XF is the fake enriched part of the X region where the fake factors are applied to predict fake contribution in the XS region (signal like part of the X region). DD fakes indicate the fake estimation from the fake factor method while MC fakes are from MC prediction. The Data-Prompt is comparable to the DD fakes and the agreement is good. Errors here are statistical only.

The situation here in the 3ℓ channel is much simpler than that in $2\ell^{SS}$ channel, the only subtraction to be performed is the prompt contamination which is estimated with MC simulation. The measured muon fake factor is 0.17±0.06 and electron fake factor is 0.39±0.07. These factor are then used in the fake enriched part of X region to predict fake contribution in the signal like part of X region as:

$$N_{\mathrm{xe}\mu} = \theta_{\mathrm{e}} \times N_{\mathrm{x}\mu\not{e}} + \theta_{\mu} \times N_{\mathrm{xe}\not{\mu}} \tag{6.9}$$

$$N_{\mathrm{xee}} = \theta_{\mathrm{e}} \times N_{\mathrm{xe}\not{e}} \tag{6.10}$$

$$N_{\mathrm{x}\mu\mu} = \theta_{\mu} \times N_{\mathrm{x}\mu\not{\mu}} \tag{6.11}$$

The signal like part of the X region denoted as XS is a start point towards the final signal region. The data-driven background estimation and optimization for signal region are all performed here with sufficient statistic. Table 6.24 shows the estimation of fake contribution in the X region with data-driven fake factor method and predictions from MC.

The fake factor are measured using xee and $x\mu\mu$ events, then a closure is performed using the $xe\mu$ events. Difference between the fake estimation and "Data-Prompt" is taken into account as a systematic uncertainty.

To investigate the impact of jet composition (heavy/light flavor fractions) on the fake factors, similar to the $2\ell^{SS}$ channel, the fake factors are derived in the Z+Jets enriched region and the $t\bar{t}$ enriched region. Further more, the fake factors are derived in low jet multiplicity region but applied in high jet multiplicity region, therefore, it's necessary to check the difference between these two region, i.e Y and X regions. Figure 6.19 shows the comparison between fake factors derived from these four regions.

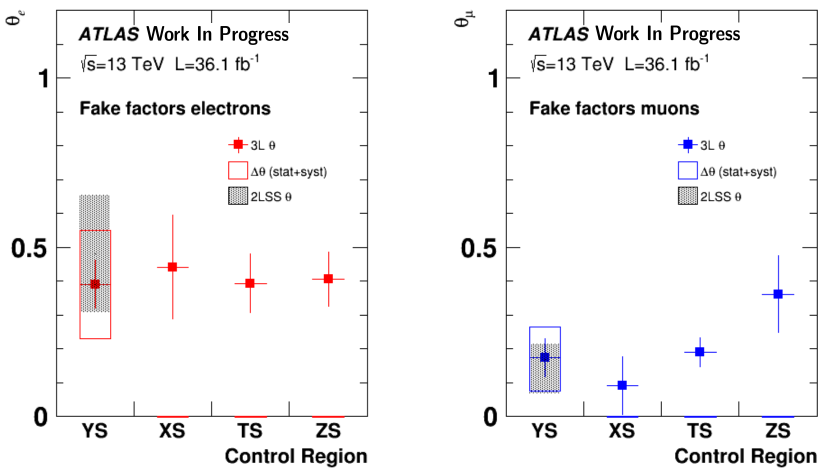


Figure 6.19 Fake factors derived from the four different regions. Total uncertainty is shown for fake factors of YS region but the others are statistical only. The fake factors of the $2\ell^{SS}$ are also shown for comparison.

Good agreement between fake factors derived from different regions is observed. Figure 6.20 shows the background estimation in the XS region, good agreement between the data and the background estimation indicates that the background is well controlled. Apart from the uncertainty due to jet composition, there are several other systematic uncertainties on the fake estimation in 3ℓ. This part will be discussed later since the systematic uncertainty of fake is related to signal region in this channel.

On top of the XS region, five variables are adopted and optimized for the final signal regions: ΔR_{12}, $\Delta R_{\ell,j}$, $p_T^{leadingjet}$, E_T^{miss} and $M_{3\ell}$. Distributions of these five variables in the XS region are shown in Figure 6.21.

The TMVA tool-kit[64] is used to perform the cuts optimization for the signal region. The optimization is performed for each mass point in two regions: 0 SFOS and 1-2 SFOS because the background composition are quite different among these two regions. Similar to the optimization procedure in $2\ell^{SS}$ channel, 100 working points are scanned corresponding to different signal efficiency. Signal significance is computed for each working point to choose the optimal one.

The optimized cuts and their individual efficiencies are described in Table 6.25. Due to statistic constraint, it's impossible to get the fake estimation in signal region by directly applying these cuts. If the cuts are totally uncorrelated, the product of the cut efficiencies can be used to estimate the total efficiency. This idea is then used to extrapolate the fake background estimation from XS region to the signal region by grouping the five

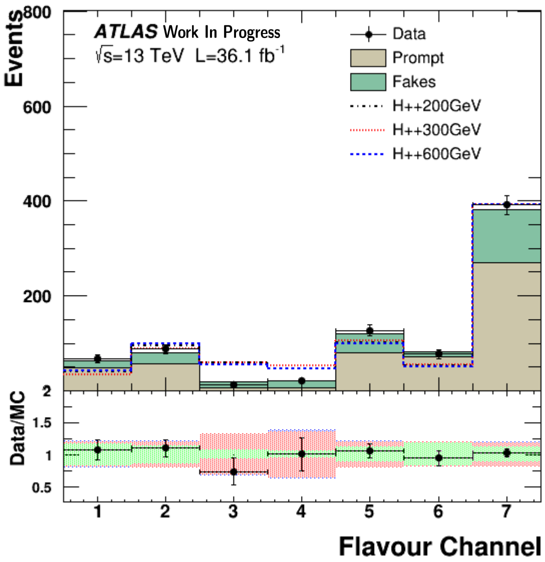


Figure 6.20 Comparison between the data and the background estimation in XS region. Note that the signal is scaled to data integral, for a better visibility. The bottom panel shows the ration between data and the background estimation. The error band includes the uncertainty on fake estimates (red) and the total error obtained by adding in quadrature the Monte Carlo statistics (blue). Bin 1 to bin 6 correspond to the flavour channels eee, $ee\mu$, $e\mu\mu$, μee, $\mu\mu e$, $\mu\mu\mu$. Bin 7 shows the total yields across all the channels.

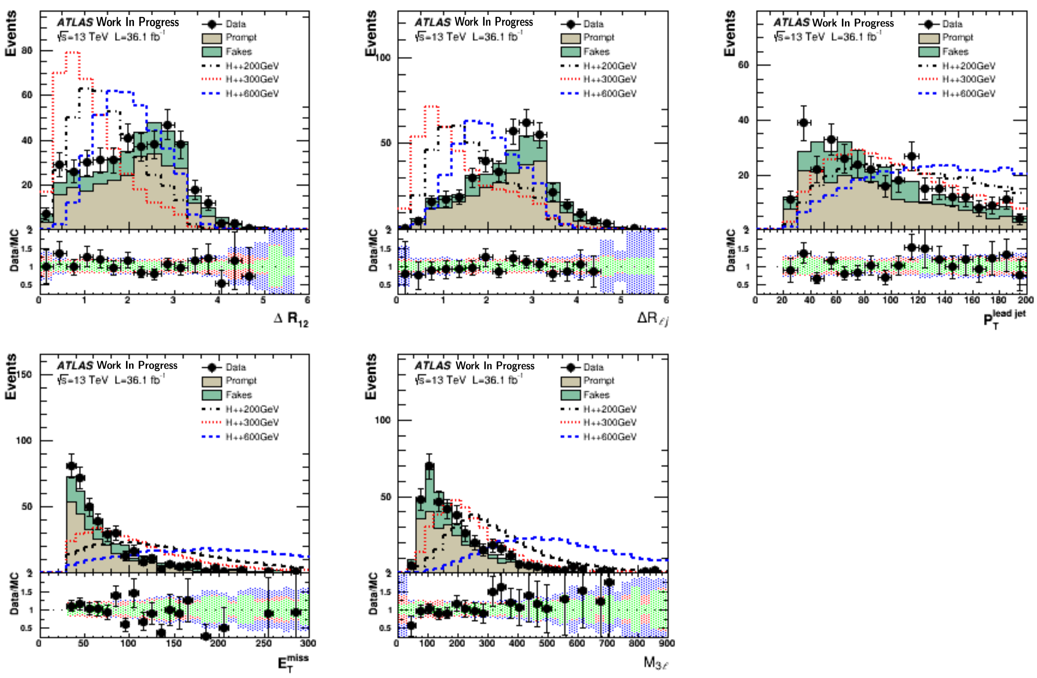


Figure 6.21 Distribution of variables adopted to define signal region for the 3ℓ channel. Errors are full uncertainties.

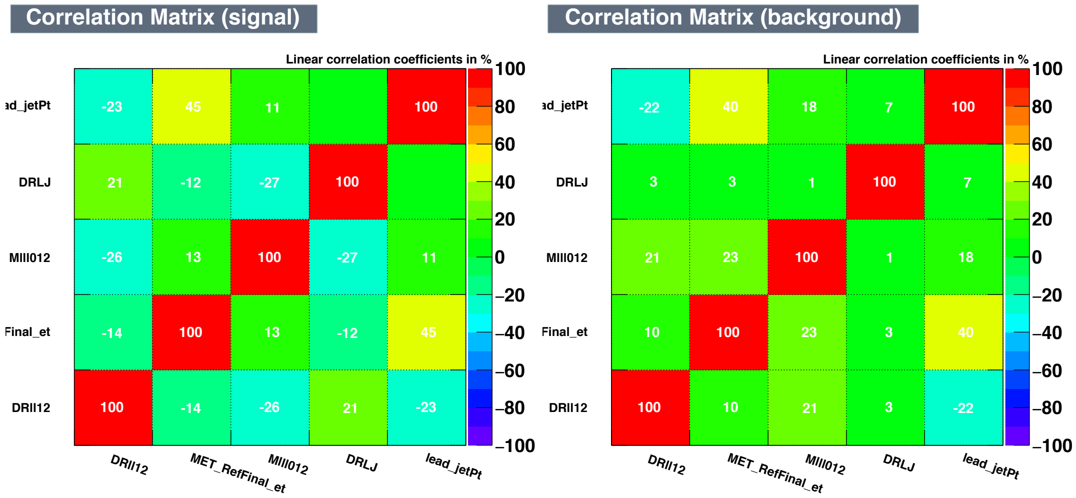


Figure 6.22 The correlation coefficients for the five variables used in the signal region selection for signal (left) and background (right).

variables according to their correlations. The correlation of the five variables is shown in Figure 6.22

	SFOS 0	SFOS 1,2	Data	Prompt	Fakes	H++200GeV
1	$0.15 < \Delta R_{\ell\ell^{ss}} < 1.57$ and $E_T^{miss} > 45$ GeV	$0.00 < \Delta R_{\ell\ell^{ss}} < 1.52$ and $E_T^{miss} > 45$ GeV	0.198 ± 0.046	0.191 ± 0.025	0.16 ± 0.06	0.670 ± 0.004
2	$M_{3\ell} > 160$ GeV and $0.08 < \Delta R_{\ell j} < 1.88$	$M_{3\ell} > 170$ GeV and $0.07 < \Delta R_{\ell j} < 1.31$	0.061 ± 0.050	0.084 ± 0.027	0.038 ± 0.057	0.498 ± 0.005
3	$P_T^{\text{leading jet}} > 80$ GeV	$P_T^{\text{leading jet}} > 55$ GeV	0.751 ± 0.026	0.772 ± 0.014	0.709 ± 0.034	0.821 ± 0.003
4		All cuts	0.008 ± 0.05	0.006 ± 0.019	0.003 ± 0.073	0.330 ± 0.006
5		Factorised efficiency 1234	0.011	0.012 ± 0.000	0.004 ± 0.000	0.274 ± 0.000

Table 6.25 The optimized cut values and their individual efficiencies. The correlated variables are grouped together. The "All cuts" line displays the nominal efficiency when all cuts are applied while the last line "Factorized efficiency" shows the product of the efficiencies of the three groups. Only statistical errors are shown. The systematic uncertainties are not included in this table.

The extrapolation from XS region to the signal region will introduce another systematic uncertainty for the fake lepton background, Table 6.26 summarized all the uncertainties of fake lepton background in the 3ℓ channel. Jet composition and closure in $e\mu$ has been discussed before, prompt uncertainties are from MC such as uncertainty from cross section and detector simulation. ℓ_0 purity uncertainty is from the assumption that ℓ_0 is always real and the fake lepton is always among the two same sign leptons, this term is derived from MC.

6.2.3 4ℓ Channel

The 4ℓ channel imposes that all the four W bosons from $H^{\pm\pm}$ decay to leptons and neutrinos, therefore, the signal yield of this channel is expected to be much lower than

Source	fake factor electron	fake factor muon
Jet composition	25%	35%
Closure $e\mu$	30%	30%
Prompt uncertainties	17.5%	47%
ℓ_0 purity	5%	5%
Extrapolation to the signal region	30%	30%
Total systematics	46%	73%
Total stat+systematics	55%	81%

Table 6.26 Uncertainties of fake lepton background estimation in the 3ℓ channel.

the other two channels as well as background. Background events in this channel can come from fake leptons events or events with four prompt leptons. Background due to fake lepton is expected to be very small and it's estimated with a "Fake Scale Factor" method which will be described later, prompt background (mainly ZZ process) is estimated with MC simulation. Definitions for event pre-selection region are:

- Exactly four loose leptons with $p_T > 10$ GeV, trigger, trigger match, total charge of 0 and event cleaning.
- Z veto with $|M_{\ell^= \ell^-} - M_Z| > 10$ GeV.
- At least 1 SFOS with $M_{\ell^+\ell^-}$ greater than 10 GeV.
- $E_T^{miss} > 30$ GeV and b-jet veto.

Background with prompt four leptons is mostly ZZ process which is estimated with MC simulation. A ZZ enriched control region is selected with four loose leptons, at least one is trigger matched with $p_T > 30$ GeV and two SFOS lepton pairs with $M_{\ell^+\ell^-} > 25$ GeV. The comparison between the data and MC is shown in Figure 6.23, good agreement observed.

Background due to fake leptons in the 4ℓ channel is estimated with the "Fake Scale Factor" method. This is a semi-data-driven method which has been adopted by other analyses[66]. The idea of the method is to scale MC to match to the data and the method is performed using tri-lepton events to provide sufficient statistic. Considering that the fake leptons can originate from heavy flavor or light flavor jets. Two regions are defined for this method: Z+Jets enriched and $t\bar{t}$ enriched:

- Z+Jets enriched region:

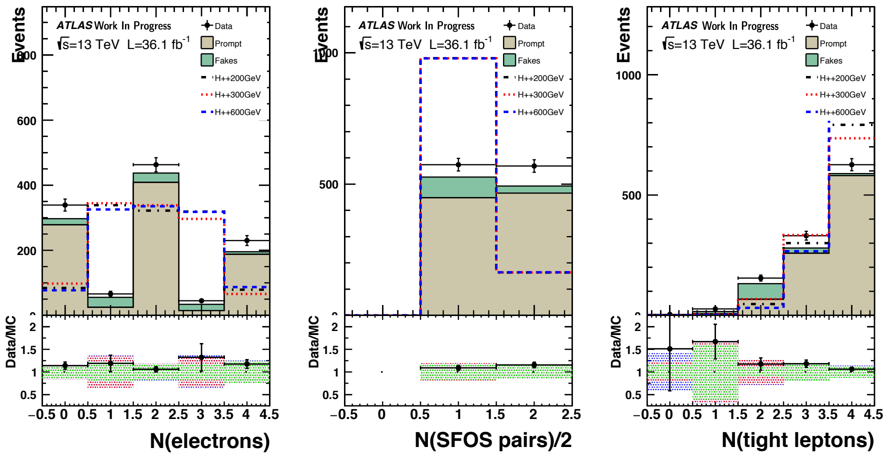


Figure 6.23 Comparison between the data and MC in the ZZ enriched region for number of electrons, number of SFOS pair and number of tight leptons from left to right. Fake is identified using truth information in MC. In the bottom panels, the inner (green) band represents the systematic errors associated to prompt contributions, the next outer band (red) correspond to combining the first with the fake contributions uncertainty, while the outer envelope (blue) includes in addition the MC statistics uncertainties.

  - Exactly three loose leptons with $p_T > 10$ GeV and total charge of ± 1.
  - 1 SFOS lepton pair with $M_{\ell^+\ell^-}$ inside a Z mass window of 10 GeV.
  - 1 or 2 jets with $p_T > 25$ GeV
  - $E_T^{miss} < 50$ GeV and $M_T < 50$ GeV

- $t\bar{t}$ enriched region:
  - Exactly three loose leptons with $p_T > 10$ GeV and total charge of ± 1.
  - No SFOS lepton pair.
  - 1 or 2 jets with $p_T^{1,(2)} > 30(25)$ GeV

Therefore, four scale factors are defined for leptons:

$$\lambda_L^e,\ \lambda_H^e,\ \lambda_L^\mu,\ \lambda_H^\mu$$

where L and H indicate light flavor and heavy flavor. Relations between the MC and data can be written as:

$$e : N_{Data-Prompt} = \lambda_H^e N_{t\bar{t}} + \lambda_L^e N_{Z+Jets}, \tag{6.12}$$

$$\mu : N_{Data-Prompt} = \lambda_H^\mu N_{t\bar{t}} + \lambda_L^\mu N_{Z+Jets}, \tag{6.13}$$

where $N_{Data-Prompt}$ stands for the fake contribution in the data. The scale factor can be derived once the $N_{t\bar{t}}$ and N_{Z+Jets} are measured from the $t\bar{t}$ enriched region and Z+Jets enriched region. Table 6.27 shows the number of events measured in the control regions.

Sample	Data	$t\bar{t}$	Z+jets	others
Z+jets fake electron CR	6299	87.1 ± 3.9	4891.9 ± 322.1	1213.8 ± 29.8
Z+jets fake muon CR	4844	83.0 ± 4.0	4062.5 ± 296.3	920.4 ± 34.9
$t\bar{t}$ fake electron CR	854	712.6 ± 11.8	10.8 ± 4.2	45.8 ± 4.3
$t\bar{t}$ fake muon CR	778	680.9 ± 11.6	5.4 ± 3.1	20.1 ± 6.0

Table 6.27 The tri-lepton control samples used to qualify the fakes contributions to the 4ℓ channel. Only statistical errors are shown for the predictions.

The derived scale factors are:

$$\lambda_H^e = 1.12 \pm 0.05(stat) \pm 0.56(syst) \tag{6.14}$$

$$\lambda_L^e = 1.02 \pm 0.07(stat) \pm 0.51(syst) \tag{6.15}$$

$$\lambda_H^\mu = 1.11 \pm 0.05(stat) \pm 0.55(syst) \tag{6.16}$$

$$\lambda_L^\mu = 0.94 \pm 0.07(stat) \pm 0.47(syst) \tag{6.17}$$

The systematic uncertainty of the scale factor is assigned to be 50%. It is determined to cover the difference between the variation of the control regions. The Z+Jets and $t\bar{t}$ control regions are further divided into four regions according to the number of b-jet:

- A : $N_{\mathrm{b-jets}} = 0$ for Z+Jets and $N_{\mathrm{b-jets}} = 0$ for $t\bar{t}$
- B : $N_{\mathrm{b-jets}} > 0$ for Z+Jets and $N_{\mathrm{b-jets}} = 0$ for $t\bar{t}$
- C : $N_{\mathrm{b-jets}} = 0$ for Z+Jets and $N_{\mathrm{b-jets}} > 0$ for $t\bar{t}$
- D : $N_{\mathrm{b-jets}} > 0$ for Z+Jets and $N_{\mathrm{b-jets}} > 0$ for $t\bar{t}$

The scale factors are recomputed in these four regions, Figure 6.24 shows the comparison.

A check for the stability of this method is performed among low and high p_T region, divide the control regions by lepton's p_T and compare the recomputed scale factors to the nominal values. Non-closure in these tests is taken as a systematic uncertainty of 50%. The scale factor are applied to MC fake events (identified with truth information) to scale the fake prediction from MC. The fake lepton background is the 4ℓ channel at pre-selection level is scaled from 21.0±1.7 to 24.8±2.1 using the scale factors described above.

On top of pre-selection region, five variables are adopted and optimized to go to signal region: E_T^{miss}, $M_{4\ell}$, $p_T^{\ell_0}$, $\Delta R_{\ell^\pm\ell^\pm}^{min}$ and $\Delta R_{\ell^\pm\ell^\pm}^{max}$. Figure 6.25 illustrates the distributions of these five variables in the event pre-selection region.

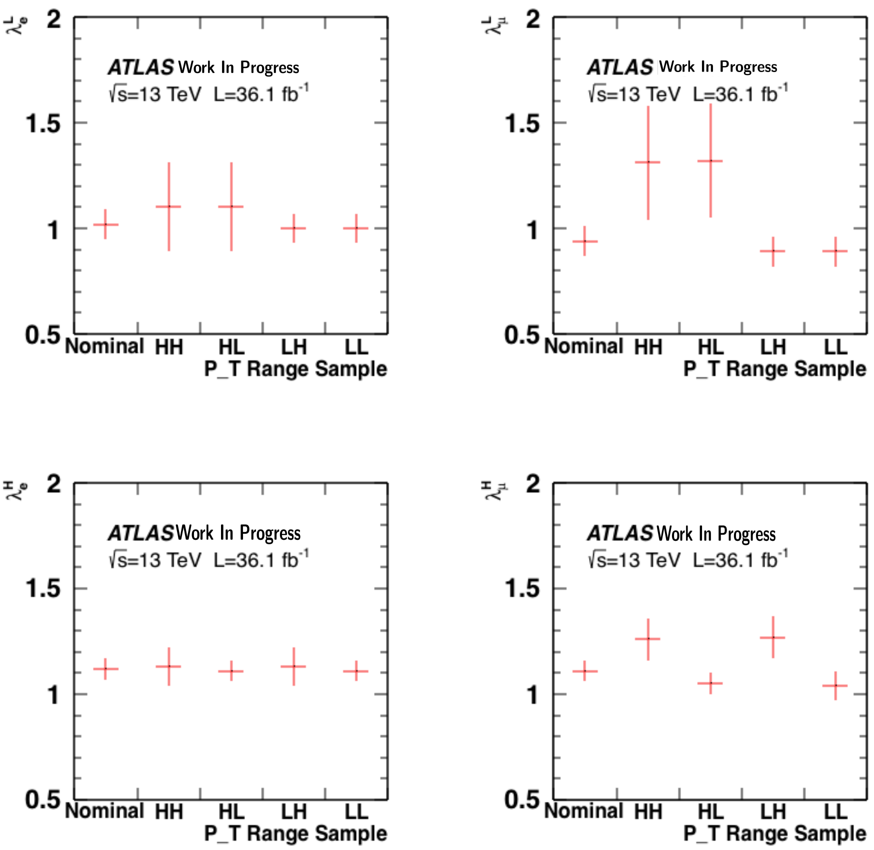


Figure 6.24 Scaling factors deduced from the two tri-leptons CRs, Z+Jets (Light Flavor environment) and $t\bar{t}$ (Heavy Flavor environment) in low and high p_T ranges.

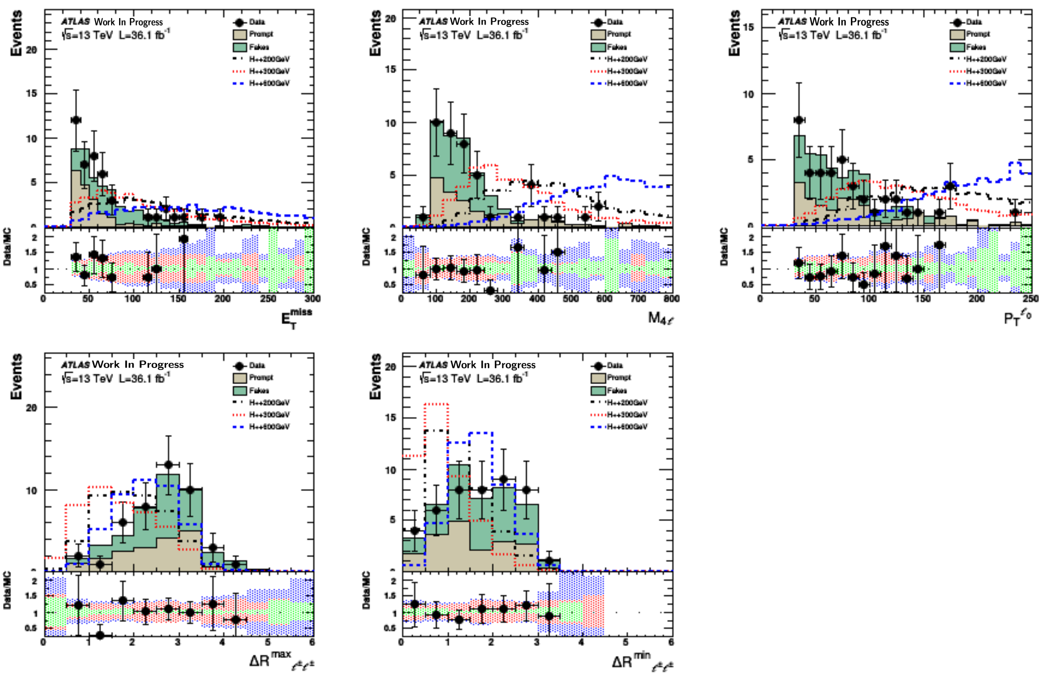


Figure 6.25 Distributions of the five variables used for signal region definition for the 4ℓ channel.

The optimization procedure is similar to the other two channels, it's performed for each mass point and there is no further signal regions split as a function of SFOS. Cut values with optimal signal significance are chosen for each mass point.

6.3 Systematic Uncertainties

Both theoretical and experimental uncertainties are taken into account in this analysis. The signal process and prompt background are estimated with MC while non-prompt background is estimated with data-driven method. Theoretical uncertainty include uncertainties from normalization, acceptance, etc. Experimental uncertainty originates from the finite accuracy of detector simulation and data-driven background estimation. For SM processes, the theoretical uncertainty is treated individually for each MC sample during normalization. For rare processes without delicate measurements, the uncertainty is set to be 50% conservatively. The uncertainties of WZ, VVV , tZ, ZZ, WW, $t\bar{t}W$ and $t\bar{t}Z$ are 7.2%[67], 20%[68], 15%[67], 19.2%[69], 10.1%[70], 53.3%[71] and 33.3%[71] respectively. Phase space of this analysis covers high jet multiplicity region and WZ is the most important prompt background in the signal region as shown in Figure 6.31, therefore, a study comparing the data to background is performed for the WZ uncertainty. Looking into the distribution of jet multiplicity, it's found that data agrees to predicted number with an uncertainty of 10% for $N_{jets} < 4$ and 20% for $N_{jets} > 4$. A 20% uncertainty is assigned to WZ to cover the difference in high jet multiplicity region.

For the signal process, theoretical uncertainty consists of PDF uncertainty, uncertainty due to factorization scale, uncertainty due to parton shower and uncertainty in cross section measurement. The PDF uncertainty is evaluated with the LHAPDF6[72] library for inclusive, $2\ell^{SS}$, 3ℓ and 4ℓ phase spaces, the generated events are re-weighted with:

$$\omega_i = \frac{x_1 f_{1i}(x_1; Q^2)}{x_1 f_{10}(x_1; Q^2)} \frac{x_2 f_{2i}(x_2; Q^2)}{x_2 f_{20}(x_2; Q^2)}, i = 1, 2, \cdots, 40. \tag{6.18}$$

where the functions f_{10} and f_{20} in the denominators are the nominal PDFs; f_{1i} and f_{2i} in the numerators are the eigenvector PDF members. For each eigenvector PDF member, the expected signal yield is estimated by the re-weighted signal samples. The uncertainties are taken as symmetric (average of up-down variations) for simplicity.

Figure 6.26 shows the PDF uncertainties of signal process for phase spaces of the preselections $2\ell^{SS}$, 3ℓ and 4ℓ. The PDF uncertainty of signal process is found to be in the range between 2.5% and 4.5%.

$M_{H^{\pm\pm}}$	200	300	400	500	600	700
μ auto	70.59	14.18	4.11	1.469	0.594	0.2631
$\mu = M_{H^{\pm\pm}}$	73.28	15.21	4.054	1.633	0.6684	0.2866
$\mu = M_{H^{\pm\pm}}/2$	74.3	15.83	4.767	1.748	0.7237	0.3247
$\mu = 2M_{H^{\pm\pm}}$	72.01	14.63	4.268	1.528	0.6208	0.2751
$\Delta\sigma/2\sigma(\mu = M_{H^{\pm\pm}})$	1.56%	3.94%	5.54%	6.74%	7.70%	8.65%

Table 6.28 The variation of the inclusive cross section as a function of the factorization scale (taken to be equal to the normalization scale).

Uncertainty due to factorization scale is obtained through a variation on the scale, a study is presented in Table. 6.28. The variations due to the factorization scale (taken to be equal to the renormalisation scale in this study) ranges from 1.5% to 8.7%. Uncertainty due to parton shower is obtained by comparing the nominal ttH sample with one with the same matrix element calculation but showered using Herwig++, and by comparing the nominal ttV samples with ones with variations in the A14 Pythia 8 tune. An overall uncertainty of 15% (PDF + factorization scale) is assigned for the signal process.

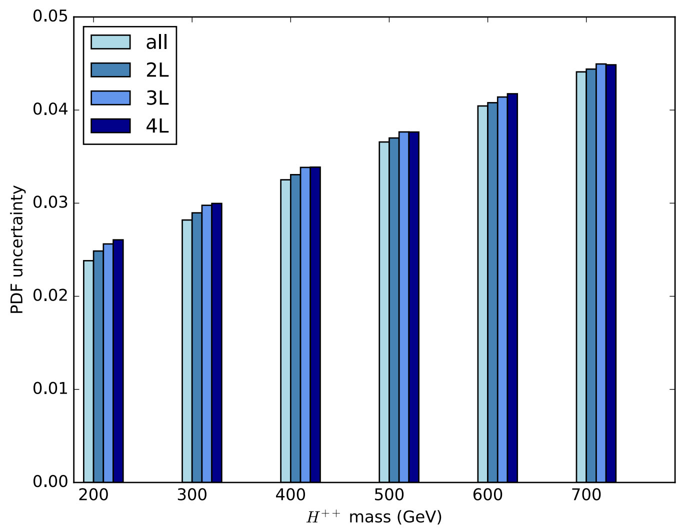


Figure 6.26 The PDF uncertainties of the signal yields for phase-spaces of the pre-selections $2\ell^{SS}$, 3ℓ and 4ℓ analyses. The uncertainties for the inclusive samples are also shown (labelled as "all" in the figure).

Experimental uncertainty is made of two parts: from detector simulation and from data-driven background estimation. Uncertainties from detector simulation are obtained through comparison between the data and MC by the performance groups. Uncertainties due to data-driven background have been discussed concretely in previous sections.

Table 6.29 summarized the experimental uncertainties of $2\ell^{SS}ee$ channel for mass point $M_{H^{\pm\pm}} = 200$ GeV.

	$H^{\pm\pm}$	MisCharge	Prompt	Fake
Pileup	0.02	-	0.13	-
elSF	0.04	-	0.05	-
elReso	0.00	-	0.00	-
elScal	0.01	-	0.00	-
MET	0.00	-	0.00	-
muSF	0.00	-	0.00	-
trigSF	0.00	-	0.00	-
Jet	0.08	-	0.01	-
JVT	0.04	-	0.03	-
MCnorm	0.00	-	0.19	-
QMisID	-	0.22	-	0.16
Fake	-	-	-	0.23
Lumi	0.03	-	0.03	-

Table 6.29 Experimental systematic uncertainties (relative effect) of the $2\ell^{SS}ee$ channel, for the mass point of 200 GeV.

6.4 Statistical Interpretation

Figure 6.27 shows the expected and observed yield in the signal regions for $M_{H^{\pm\pm}} = 200$ GeV and the composition of prompt background is illustrated in Figure 6.31. There are only several events left in the final signal region and the data is consistent with the SM background. Yields for mass points from 300 GeV to 500 GeV are shown in Figure 6.28, Figure 6.29 and Figure 6.30.

The signal significance and exclusion limits are computed with the profile likelihood method using ttHFitter package. The profile likelihood method has been discussed in Section 5.4, parameter of interest is signal strength while the systematic uncertainties are treated as nuisance parameters. The expected and observed signal significances are shown in Figure 6.32, no significant excess is observed, therefore, exclusion limits on the signal strength are derived as illustrated in Figure 6.33. The model can be excluded with 95% CL at $M_{H^{\pm\pm}} < 260$GeV with expected limits combing all channels. The observed limits exclude the model at $M_{H^{\pm\pm}} < 220$GeV.

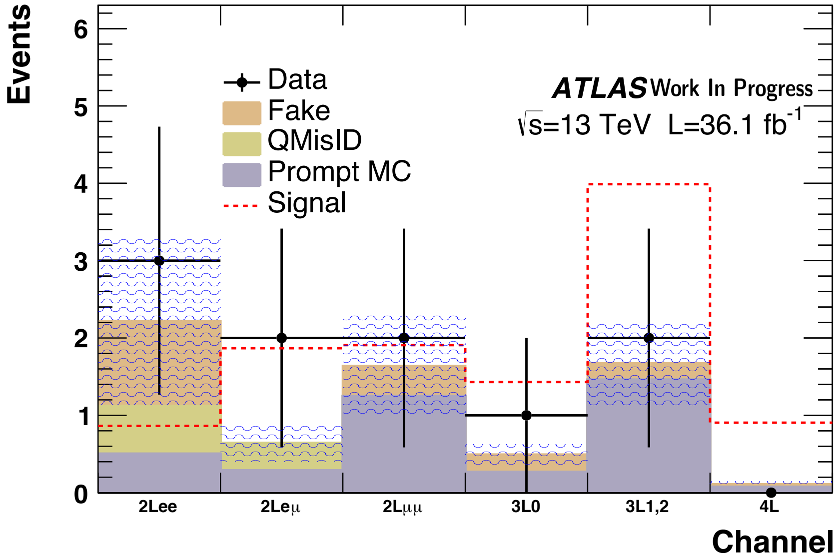


Figure 6.27 Expected and observed yield in the signal regions for all analysis channels at $M_{H^{\pm\pm}}$ = 200 GeV, and used for signal extraction. The error bars represent the full error (statistic and systematic).

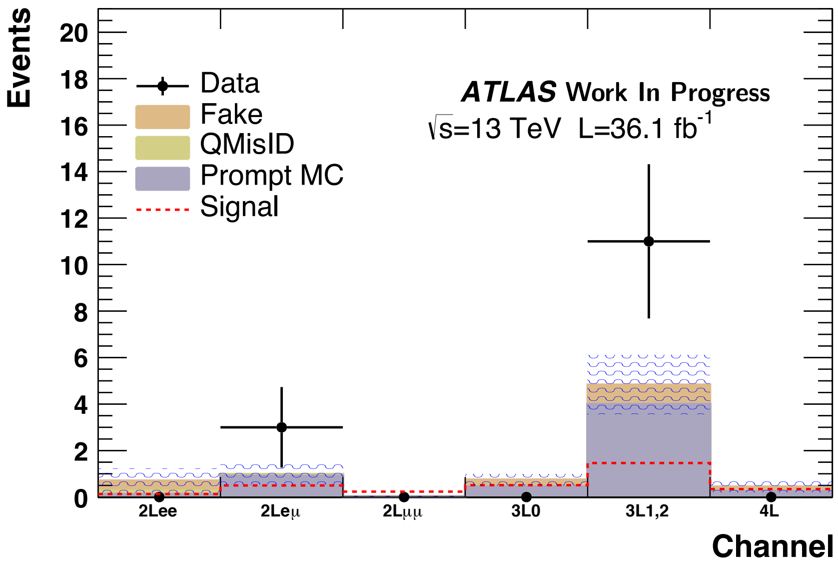


Figure 6.28 Expected and observed yield in the signal regions for all analysis channels at $M_{H^{\pm\pm}}$ = 300 GeV, and used for signal extraction. The error bars represent the full error (statistic and systematic).

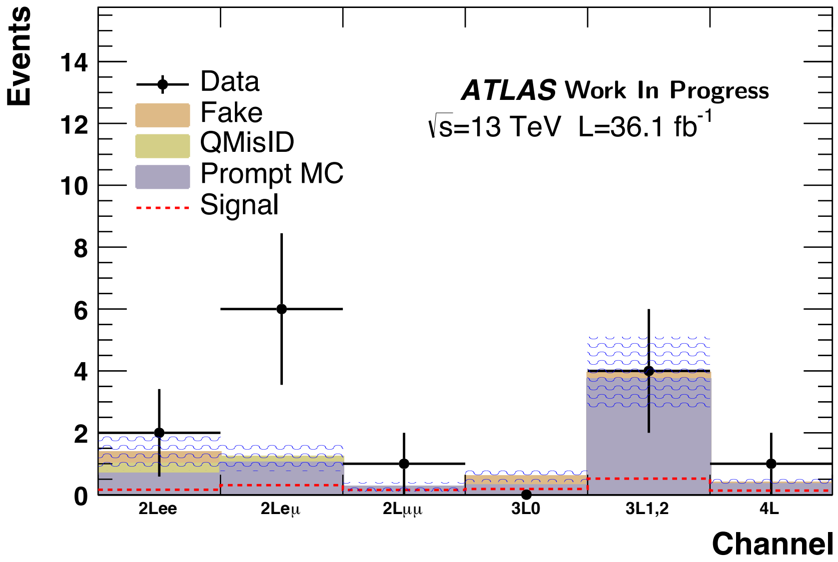


Figure 6.29 Expected and observed yield in the signal regions for all analysis channels at $M_{H^{\pm\pm}}$ = 400 GeV, and used for signal extraction. The error bars represent the full error (statistic and systematic).

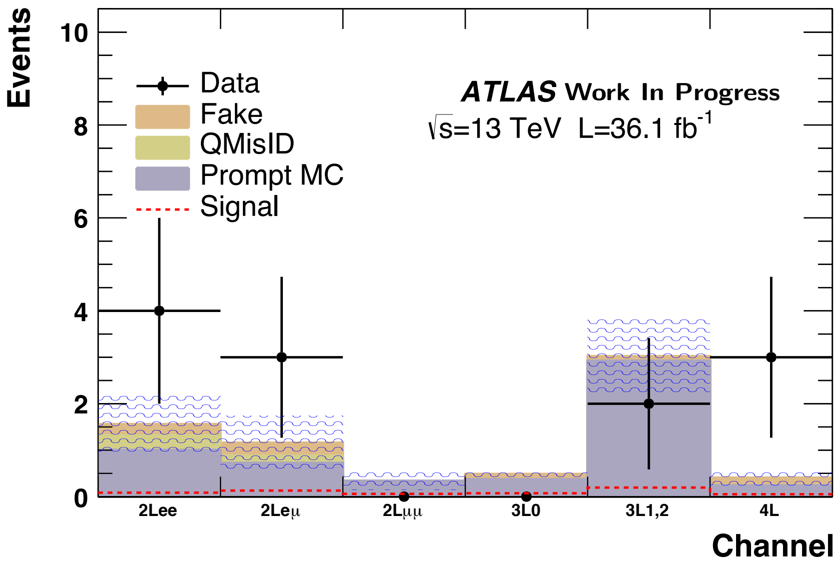


Figure 6.30 Expected and observed yield in the signal regions for all analysis channels at $M_{H^{\pm\pm}}$ = 500 GeV, and used for signal extraction. The error bars represent the full error (statistic and systematic).

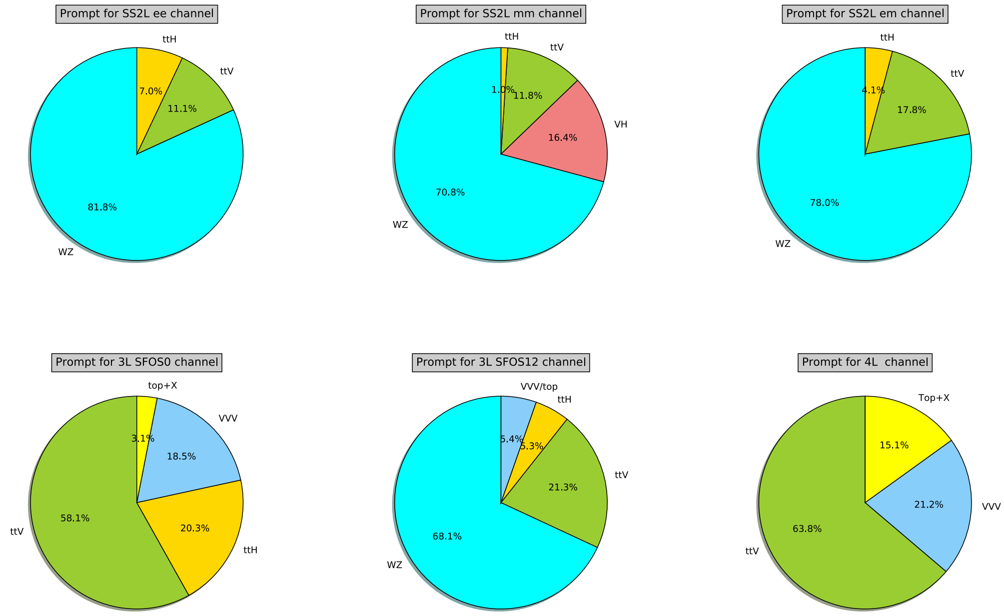


Figure 6.31 Prompt composition in the signal regions for all channels.

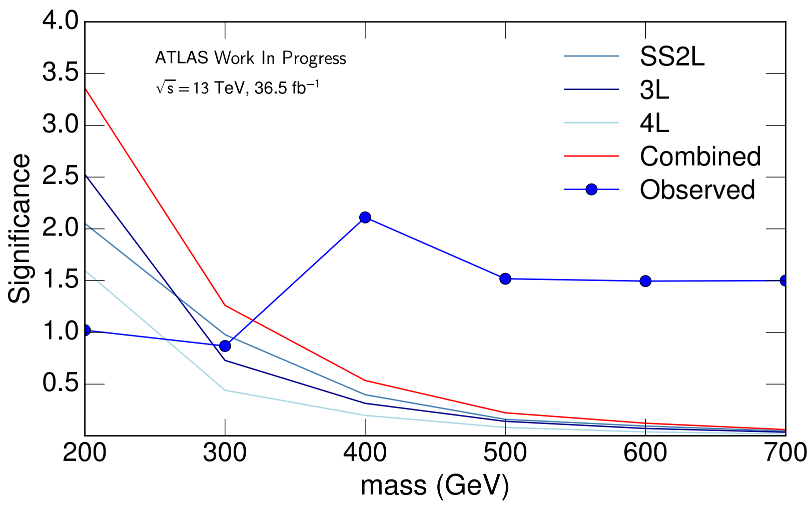


Figure 6.32 Expected and observed significances as a function of $M_{H^{\pm\pm}}$.

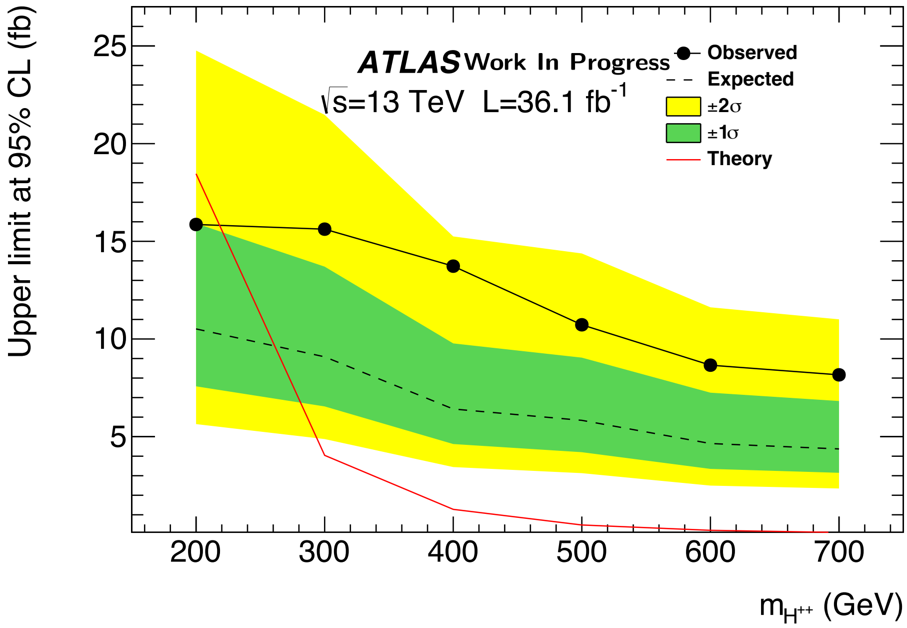


Figure 6.33 Expected and observed upper limits at 95% CL for the combination of $2\ell^{SS}$, 3ℓ and 4ℓ.

6.5 Conclusion

A search for doubly charged Higgs boson in an unexplored phase space is performed using 36.1fb^{-1} of data collected by the ATLAS detector with center-mass-energy of 13 TeV during 2015 and 2016. The search focus on the pair production mode while the doubly charged Higgs decay to W bosons. Various data-driven techniques are adopted during the background estimation. The background is found to be consistent with the data, no significant excess observed, therefore, upper limits on the signal strength are derived. The model is excluded at 95% CL for $M_{H^{\pm\pm}} < 220\text{GeV}$.

Chapter 7 Conclusion

The LHC and ATLAS have be running successfully during the past years. The high quality Run 1 and Run 2 data provide rich physical potential for precise measurements of the SM and searches for new physics beyond the SM. In this document, two physics analyses are presented using Run 1 and Run 2 data.

The search for $W^{\pm}W^{\pm}W^{\mp}$ and study of anomalous quartic gauge couplings utilized the Run 1 data with a center-mass-energy of 8 TeV and an integrated luminosity of 20.3 fb^{-1}. The observation agrees with the SM background prediction and no significant $W^{\pm}W^{\pm}W^{\mp}$ signal could be measured. The observed 95% CL upper limit on the SM $W^{\pm}W^{\pm}W^{\mp}$ cross section is found to be 730 fb with an expected limit of 560 fb in the absence of $W^{\pm}W^{\pm}W^{\mp}$ production. The aQGC limits are also derived on the dimensional-8 operators of the effective field theory using the $WWWW$ vertex.

The search for doubly charged Higgs ($H^{\pm\pm}$) utilized the Run 2 data with a center-mass-energy of 13 TeV and an integrated luminosity of 36.1 fb^{-1}. A simplified scenario is chosen in this analysis, the $H^{\pm\pm}$ are pair produced and they all decay to W bosons. This is a search beyond the SM however the observation is in consistent with the SM background prediction, therefore upper limits are derived and the model is excluded with 95% CL for $M_{H^{\pm\pm}} <$ 220 GeV. Data taking of Run 2 is not finished yet, more data will come in and this analysis will be updated with more data in the future.

The imperfection of the SM imply the existence of new physics beyond the SM. Either precise measurements of the SM productions or the searches for new phenomenas could help us find out the new physics. In this thesis, one measurement of the SM production with Run 1 data and one search for new particle with Run 2 data are performed however no symptom of new physics beyond the SM is found.

Bibliography

[1] *Wikipedia*, https://en.wikipedia.org/wiki/Standard_Model.

[2] K. Arnold, J. Bellm, G. Bozzi, M. Brieg, F. Campanario, et al., *VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.5.0*, arXiv:1107.4038 [hep-ph].

[3] J. Alwall, P. Demin, S. D. Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, *MadGraph/MadEvent v4: The New Web Generation*, JHEP 070907062334.

[4] K. Hagiwara, S. Ishihara, R. Szalapski, and D. Zeppenfeld, *Low-energy effects of new interactions in the electroweak boson sector*, Phys.Rev. **D48** (1993) 2182–2203.

[5] W. Buchmuller and D. Wyler, *Effective Lagrangian Analysis of New Interactions and Flavor Conservation*, Nucl.Phys. **B268** (1986) 621–653.

[6] O. J. P. Eboli, M. C. Gonzalez-Garcia, and J. K. Mizukoshi, $pp \to jje^{\pm}\mu^{\mp}\nu\nu andjje^{\pm}\mu^{\mp}\nu\nu atO(\alpha_{em}^6)andO(\alpha_{em}^4\alpha_s^2)$ *for the study of the quartic electroweak gauge boson vertex at CERN LHC*, Phys. Rev. **D74** (2006) 073005, arXiv:hep-ph/0606118 [hep-ph].

[7] 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].

[8] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[9] *VBFNLO tool to calculate form factor*, https://www.itp.kit.edu/vbfnlo/wiki/doku.php?id=download:formfactor.

[10] P. Fileviez Perez, T. Han, G.-y. Huang, T. Li, and K. Wang, *Neutrino Masses and the CERN LHC: Testing Type II Seesaw*, Phys. Rev. **D78** (2008) 015018, arXiv:0805.3536 [hep-ph].

[11] ATLAS Collaboration, *Search for doubly-charged Higgs boson production in multi-lepton final states with the ATLAS detector using proton-proton collisions at* $\sqrt{s} = 13$ *TeV*,.

[12] A. Belyaev, N. D. Christensen, and A. Pukhov, *CalcHEP 3.4 for collider physics within and beyond the Standard Model*, Comput. Phys. Commun. **184** (2013) 1729–1769, arXiv:1207.6082 [hep-ph].

[13] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[14] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[15] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, http://cds.cern.ch/record/1966419.

[16] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[17] 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].

[18] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[19] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv:1207.1303 [hep-ph].

[20] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv:0803.0883 [hep-ph].

[21] M. H. Seymour and A. Siodmok, *Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data*, JHEP **10** (2013) 113, arXiv:1307.5015 [hep-ph].

[22] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[23] P. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[24] S. Frixione, G. Ridolfi, and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv:0707.3088 [hep-ph].

[25] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[26] 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].

[27] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[28] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. Phys. Res. A **462** (2001) 152.

[29] J.-L. Caron, *The LHC injection complex.. L'ensemble d'injection du LHC.*, May, 1993. https://cds.cern.ch/record/841568. AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[30] https://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc_in_pictures.htm.

[31] ATLAS Collaboration, N. Nikiforou, *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, 2013. arXiv:1306.6756 [physics.ins-det]. https://inspirehep.net/record/1240499/files/arXiv:1306.6756.pdf.

[32] J. Pequenao, *Computer generated image of the ATLAS inner detector*, Mar, 2008. http://cds.cern.ch/record/1095926.

[33] J. Pequenao, *Computer Generated image of the ATLAS calorimeter*, Mar, 2008. https://cds.cern.ch/record/1095927.

[34] J. Pequenao, *Computer generated image of the ATLAS Muons subsystem*, Mar, 2008. http://cds.cern.ch/record/1095929.

[35] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, arXiv:1110.1530[hep-ex].

[36] ATLAS Collaboration, *ATLAS Public Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[37] ATLAS Collaboration, *ATLAS Public Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[38] S. HÃ¶che, *Introduction to parton-shower event generators*, arXiv:1411.4085[hep-ph].

[39] M. Mangano et al., Alpgen*, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:hep-ph/0206293.

[40] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:hep-ph/0204244.

[41] T. Gleisberg, S. Hoeche, F. Krauss, S. S. M. Schoenherr, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, arXiv:0811.4622[hep-ph].

[42] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, https://cds.cern.ch/record/2157687.

[43] A. 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, Geneva, Aug, 2016. http://cds.cern.ch/record/2203514.

[44] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2017-001/index.html.

[45] A. Collaboration, *Muon reconstruction performance of the ATLAS detector in proton?proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv:1603.05598 [hep-ex].

[46] G. S. Matteo Cacciari, Gavin P. Salam, *The* $anti-k_t$ *jet clustering algorithm*, arXiv:0802.1189[hep-ph].

[47] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets*

reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s}$=8 TeV, https://cds.cern.ch/record/2008678.

[48] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV with ATLAS*, arXiv:1108.5602[hep-ex].

[49] ATLAS Collaboration, *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s}$ = 13 TeV* , https://cds.cern.ch/record/2037700.

[50] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008, arXiv:1512.01094 [hep-ex].

[51] ATLAS Collaboration, *Measurement of WZ production in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur.Phys.J. **C72** (2012) 2173, arXiv:1208.1390 [hep-ex].

[52] F. Cascioli, T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, et al., *ZZ production at hadron colliders in NNLO QCD*, Phys.Lett. **B735** (2014) 311–313, arXiv:1405.2219 [hep-ph].

[53] CMS Collaboration Collaboration,, *Measurement of WZ production rate*, Tech. Rep. CMS-PAS-SMP-12-006, CERN, Geneva, 2013. http://cds.cern.ch/record/1564318.

[54] ATLAS Collaboration, G. Aad et al., *Measurements of Wγ and Zγ production in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector at the LHC*, Phys.Rev. **D87** (2013) 112003, arXiv:1302.1283 [hep-ex].

[55] ATLAS Collaboration, G. Aad et al., *Measurement of the $Z\gamma$ and $Z\gamma\gamma$ Production Cross Section in pp Collisions at $\sqrt{s}$ = 8 TeV with the ATLAS Detector at the LHC*, Tech. Rep. ATL-COM-PHYS-2013-1573, CERN, Geneva, Nov, 2013. https://cds.cern.ch/record/1631102.

[56] ATLAS Collaboration, G. Aad et al., *Evidence for Electroweak Production of $W^{\pm}W^{\pm}jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, Phys.Rev.Lett. **113** (2014) 141803, arXiv:1405.6241 [hep-ex].

[57] *Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.*, Tech. Rep. ATLAS-CONF-2015-032, CERN, Geneva, Jul, 2015. https://cds.cern.ch/record/2038143.

[58] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, Eur.Phys.J. **C73** (2013) 2518, arXiv:1302.4393 [hep-ex].

[59] https://rivet.hepforge.org/.

[60] *TGClimPackage*, https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TGClimPackage.

[61] ATLAS Collaboration, M. Aaboud et al., *Search for triboson $W^{\pm}W^{\pm}W^{\mp}$ production in pp*

collisions at $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 141, arXiv:1610.05088 [hep-ex].

[62] C. A. Nelson, *Correlation between decay planes in Higgs-boson decays into a W pair (into a Z pair)*, Phys. Rev. D **37** (1988) 1220–1225, https://link.aps.org/doi/10.1103/PhysRevD.37.1220.

[63] M. Dittmar and H. Dreiner, *How to find a Higgs boson with a mass between 155 and 180 GeV at the CERN LHC*, Phys. Rev. D **55** (1997) 167–172, https://link.aps.org/doi/10.1103/PhysRevD.55.167.

[64] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[65] ATLAS Collaboration, *Search for the associated production of a Higgs Boson with a top quark pair in multilepton final states with the ATLAS detector (ICHEP 2016)*, Tech. Rep. ATL-COM-PHYS-2016-419, CERN, Geneva, May, 2016. https://cds.cern.ch/record/2150024.

[66] ATLAS Collaboration, *Measurement of the tt¯Z and tt¯W production cross sections in multilepton final states using 3.2 fb -1 of pp collisions at ?s = 13 TeV with the ATLAS detector*, Tech. Rep. 1, 1, 2017.

[67] 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].

[68] *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, Tech. Rep. ATL-PHYS-PUB-2016-002, CERN, Geneva, Jan, 2016. https://cds.cern.ch/record/2119986.

[69] 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** (2016) 101801, arXiv:1512.05314 [hep-ex].

[70] 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].

[71] ATLAS Collaboration, *Measurement of the* $t\bar{t}Z$ *and* $t\bar{t}W$ *production cross sections in multilepton final states using 3.2 fb*$^{-1}$ *of pp collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 40, arXiv:1609.01599 [hep-ex].

[72] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, The European Physical Journal C **75** (2015) 132, http://dx.doi.org/10.1140/epjc/s10052-015-3318-8.