Searches for associated Higgs Boson production with top quark pair and Higgs pair production in multi lepton final states with the ATLAS detector

Merve Nazlim Agaras

UNIVERSITE CLERMONT AUVERGNE

ECOLE DOCTORALE DES SCIENCES FONDAMENTALES

THESE

présentée pour obtenir le grade de

DOCTEUR D'UNIVERSITE en physique

Spécialité : *Physique des Particules*

Par **Merve Nazlim Agaras**

Master

Searches for associated Higgs Boson production with top quark pair and Higgs pair production in multi lepton final states with the ATLAS detector

Soutenue publiquement le devant le jury : 9 Septembre 2020

Président: M. Stéphane MONTEIL	- Professeur, Université Clermont Auvergne
Rapporteur et examinateur: Mme. Lydia FAYARD	- Chercheuse, IJClab Orsay
Rapporteur et examinateur: M. Henri BACHACOU	- Chercheur, CEA Saclay
Examinateur: Mme. Susan GASCON-SHOTKIN	- Professeur, Université Claude Bernard, Lyon
Examinateur: M. Stéphane MONTEIL	- Professeur, Université Clermont Auvergne
Examinateur: M. Stefan GUINDON	- Chercheur, CERN
Directeur de thèse: M. Djamel BOUMEDIENE	- Chercheur, LPC Clermont-Ferrand

Acknowledgements

I would firstly like to thank my supervisor, Djamel Boumediene, for his patience and continuous guidance, for always being available to help. I would also like to express my gratitude to the members of the LPC ATLAS team for their hospitality, support and help during these three years, especially David Calvet for his special care and support in both physics and non-physics related subjects.

I would like to extend my sincere thanks to $t\bar{t}H$-multilepton analysis group, we worked hard together through challenging subjects to achieve the most interesting results presented in this dissertation, and I wish them luck for the continuing analysis in the future. Special thanks to analysis contacts, Tamara Vazquez Schroeder and Ximo Poveda Torres who always supported me, gave me chance to prove my skills, and shared their deep understanding of ATLAS and particle physics with me.

Many thanks to TileCal group, it has been a pleasure to be part of the group and I wish to thank everyone I have known and worked with there.

To the reviewers of this document, Lydia Fayard and Henri Bachacou, as well as the other jury members, Susan Gascon-Shotkin, Stephane Monteil and Stefan Guindon, thank you for taking the time to read my work and listen to my defense.

Finally, I would like to thank to my family for their support, encouragement and constant love, it would not have been possible without you.

Contents

Introduction

Towards the end of 19^{th} century, the known theory of what the universe is made was more than 100 different chemical elements. Thanks to Dalton's theory, it was also known that every element is an atom, which was the fundamental and indistructable little thing. The turn of the 20^{th} century, the discovery of the first elementary particle, the electron, led to the revision of the atom structure, which yielded to the more complex structure of the atom, a nucleus and electrons around. Later in the 1930s, the nucleus of the atom was found to be made of smaller particles, neutrons and protons. Subsequently around 1960s, it was found out that even these particles inside the nucleusis is made of much smaller particles, called *quarks*. Now it is known that all of matter is made of three different elemantary particles, (u, d, e, ν) and nature provided few additional copies of these particles, which are more massive and unstable, and also the force carrying particles, which are mediators of three fundamental forces. Modern particle physics assumes that all particles have associated fields, and they are actually the excitaions of these fields. The theory which is closest to a complete description of the universe at the fundamental level is called the *Standard Model (SM)*. The last missing piece of the theory is another field that acquires the property of mass of the particles, the Higgs field, and the excitation of this field, Higgs boson. After the discovery of this particle, the theory is completed. The SM is an incredible achivement, it can explain most of the physics phenomena around. However, despite all its success the absence of a description of the last fundamental force, gravity, as well as various observations such as the existence of dark matter tend to prove that the theory needs more extensions. Therefore, many models attempt to complete the Standard Model and they are grouped under the name of New Physics or Physics Beyond the Standard Model.

One of the most capable tool today for testing the predictions of the SM is the LHC, a proton-proton collider that produces collisions with an energy at a center of mass of 13 TeV. Its the most important observation is the discovery of the Higgs boson. The various predicted extensions are intensively studied at the LHC, as well as the properties of the Higgs boson such as mass, spin, parity and couplings of the Higgs boson to itself and to the other fundamental particles. Its interaction with the heaviest SM elementary particle, the top quark, is particularly interesting. Finding a disagreement could be actually a hint of the New Physics. This dissertation mainly presents a probing of the Higgs boson coupling to the top quark by searching $t\bar{t}H$ associated production in the multilepton final states by using the data provided by ATLAS, which is one of the main experiment at the LHC.

The Higgs-self coupling is also predicted in the SM and its value defines the shape of the Higgs potential. The natural way to measure the Higgs-self coupling is to study the Higgs

boson pair production process. Given the small predicted cross section for this process, the aim of the study is to set limits on the predictions of the SM. Modifications to the couplings from Beyond Standard Model physics could also significantly alter the cross section, and can make the process incompatible with the predictions. The Higgs Boson pair production through the multilepton final states is a new group in the ATLAS experiment, and some studies are presented in this thesis.

In addition to these two physics analyses, this thesis discusses the hadronic Tile Calorimeter (TileCal) of the ATLAS experiment. The TileCal provides crucial input to the estimation of jet kinematics and to the missing transverse energy reconstruction. Therefore, the monitoring of the TileCal channels and spotting possible problems during the data taking is crucial.

This dissertation is structured as follows. Chapter 1 introduces the theoretical background along with the motivation behind the searches described in the thesis. Chapter 2 describes the LHC and the ATLAS experiment. This is followed by the detailed overview of the TileCal detector and the monitoring of the channels with laser and high voltage system in Chapter 3. In Chapter 4, the process of simulating the collected experimental data from ATLAS experiment and different programs are briefly discussed. Chapter 5 outlines the particle reconstruction and identification algorithms used to define the objects in the physics analyses. Chapter 6 introduces the statistical formulation needed to interpret the results along with the handling and evaluating the effects of the uncertainties. Chapter 7 describes the $t\bar{t}H$ multilepton analysis in detail, and Chapter 8 discusses the studies for Higgs boson pair production in the 2ℓSS final states. Finally, Chapter 9 summarises the work presented in the thesis and provides the concluding remarks.

Personal contributions

This thesis is based on the data delivered by LHC and collected by the ATLAS experiment. The whole operation of the experiment is achieved by a large number of scientists from different international institutes. In October 2017 I joined the High Energy Physics (HEP) group at Laboratoire de Physique de Clermont (LPC) and shortly after became a member of the ATLAS collaboration. As a member of the ATLAS collaboration, I have contributed to several studies within the common shared activities enabling the detector operation and analysis of physics. My main contribution has been to the *searches for the $t\bar{t}H$ production in multilepton final states* during Run2 of the LHC, which is given in Chapter 7 and I have also started working on the new course of the related study, *$t\bar{t}W$ cross-section measurements with Full Run2 data*. The other physics analysis that I was involved in is the *search for Higgs boson pair production in the $2\ell SS$ channel* and the work is described in Chapter 8. I have joined the TileCal group of ATLAS experiment and participated to the performance, maintenance and monitoring activities of the detector during the course of PhD period and the related work is summarised in Chapter 3. The corresponding contributions will be indicated explicitly at the start of each chapter.

CHAPTER 1

The Standard Model of particle physics

This chapter presents an outline of the SM as well as the theoretical and experimental motivation of the physics analyses subject to the thesis. Particular emphasis is devoted to describe the Higgs mechanism and its relation with the top quark. Section 1.1 introduces the SM particles and their interactions, which is followed by the mathematical principals of the model in Section 6.1 and the introduction to electroweak symmetry breaking in Section 1.2.5. The Higgs boson and its properties are described in Section 1.3 including its relation with top quark. Finally the limitations of the SM are described in Section 1.4.

The theoretical information of this chapter is taken from the Ref. [1–3], and the mathematical formalisims are adapt within each other.

1.1 Elementary particles and interactions

The SM provides an elegant theoretical framework that characterize the fundamental nature of matter and its interactions. The known elementary particles (Table 1.1) are grouped according to their spins: half-integer particles *fermions* and integer spin *bosons*. The spin property of particles indicates whether the particle obeys Fermi-Dirac statistics (fermions) or Bose-Einstein statistics (bosons). The fermions, which are responsible for making up the matter are then divided into two major groups, quarks and leptons, and are grouped into three families (or generations) which are typically copies of each other with different range of masses. For each charged lepton (electron (e^-), muon (μ^-) and tau (τ^-)), there is a corresponding neutral partner (electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ)). Quarks, on the other hand, are not present as free particles; they can only exist in groups, which are then called *hadrons*. Antiparticles are defined as the objects with same properties as of particles except having opposite charge.

SM particles are represented by *quantum fields* using the Lagrangian densities. Therefore, the dynamics of the system can be derived by equations of motions where the final Lagrangian is the sum of free and interaction Lagrangians. SM is a *Quantum Field Theory* (QFT) which is renormalisable, unified and invariant under both spacetime (space, rotations and

	Leptons			Quarks		
Generation	Particle	Q/e	Mass [GeV]	Particle	Q/e	Mass [GeV]
First	ν_e	0	$< 2 \times 10^{-9}$	u	+2/3	2.2×10^{-3}
	e^-	-1	0.511×10^{-3}	d	-1/3	4.7×10^{-3}
Second	ν_μ	0	$< 0.19 \times 10^{-3}$	c	+2/3	1.27
	μ	-1	0.106	s	-1/3	95×10^{-3}
Third	ν_τ	0	$< 18.2 \times 10^{-3}$	t	2/3	173
	τ	-1	1.777	b	-1/3	4.18

Table 1.1: Summary of the SM fermions and their properties. The charge is given per fraction of the electron charge (e) [4].

Lorentz boosts) and three local gauge transformations: $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. These three symmetry groups correspond to the three fundamental interactions described by the SM. Noether's theorem [5] indicates that if the system is invariant under certain transformations (symmetric), there is an associated conservation law and therefore a converved quantity. Thus, the corresponding conserved quantities are *colour*, *weak isospin* and *weak hypercharge* for the given local symmetries in SM. The only interaction that is not included in this framework so far is the gravity, since currently available theories are not able to include the general relativity into QFT description.

$SU(3)_C$ non-abelian symmetry characterises the *strong force*, which binds quarks together into the composite particles and nucleons into atomic nucleus. Particles with colour charge can interact through this force and the colour charges are labelled as red, green or blue for quarks, and anti components for anti-particles. The quarks are combined into neutral combinations of colours, meaning that only these combinations can be observed individually in nature. Eight massless vector boson fields mediate the interaction with coupling g_s, which are called *gluons*. The QFT, describing the strong force, is called *Quantum Chromodynamics* (QCD).

$SU(2)_L$ represents the *weak force*, carried by massive W^+, W^- and Z^0 bosons, determining the short range of the interactions between all (right) left-handed (anti-)chiral fermions. The weak isospin (T) is the charge of the weak force. Weak hypercharge (Y) is the generator of $U(1)_Y$ symmetry group. Together the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ describes the *electroweak force* that unifies the weak and electromagnetic interactions known as the Glashow-Weinberg-Salam [6–8] theory. The relation between the two conserved quantities T and Y defined as $Q = T_3 + Y/2$, where Q is the electric charge. The Table 1.2 summaries the gauge bosons involved in the theory.

The gauge invariance nature of SM requires that the fermions, as well as the W and Z bosons to be massless. A spontaneous symmetry breaking mechanism is introduced to justify the masses of the particles [9–11]. So-called *Brout-Englert-Higgs* mechanism predicts the massive spin-0 *Higgs boson*, H, and generates masses of the massive particles.

Force	boson	Mass [GeV]
Electromagnetic	γ	0
Weak	$W^{\pm}$	80.38
	Z	91.19
Strong	g	0

Table 1.2: Summary of the gauge bosons properties[4].

1.2 Formalism of the Standard Model

The mathematical structure of the SM is based on the QFT, which relies on the Lagrange formalism and the group theory. The dynamics of the system are defined by a mathematical object called *Lagrangian density* ($\mathcal{L}_{\mathrm{SM}}$), where the interactions and the free motions of particles are encoded. This object is invariant under transformations of the gauge symmetry group $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ and the Poincaré group. In a consequence of the variations in the Lagrangian, sets of Feynmann rules are constructed which allows to obtain variables such as scattering cross sections and decay rates of particles. In the following sections, a few basic mathematical concepts are introduced briefly in order to derive the physical laws from the theory.

1.2.1 Basic concepts: Group symmetry and Lagrangian description

Lie groups The transformations of physical systems are carried out by continuous groups, where the variables can take infinite number of values and are often invariant under the transformations. In Lie groups, the elements E are differentiable functions of their real parameters. Any element in a Lie group can be written in the form;

$$E(\theta_1, \theta_2, ...\theta_n) = exp\left(\sum_{i=1}^{n} i\theta_i F_i\right), \tag{1.1}$$

where n indicates the number of F_i parameters and are called the generators of the Lie group. The group of unitary matrices can be described as follows:

$$U = exp^{i\theta_i F_i}, \tag{1.2}$$

where θ_i are the numbers parametrizing the group elements. Then the *Lie algebra* can be described by its commutation relations from the f^{ijk} structure constants as follows:

$$\left[F_i, F_j\right] = i f^{ijk} F_k, \tag{1.3}$$

A Lie group is abelian if $f^{ijk} = 0 \,\forall\, i, j, k$ and it is non-abelian for other cases.

The Lie groups which are then most relevant with the context of SM are;

- the unitary groups $U(N)$ can be identified through the sets of $N \times N$ unitary matrices,
- the *special* unitary groups $SU(N)$ can be described by the sets of unitary $N \times N$ matrices with determinant $D = 1$.

Lagrangian density and Noether's theorem All fundamental physics interactions can be encoded in a mathematical expression called the *action* (S) which is an integral over the phase space of the Lagrangian density. Lagrangian must be scalar in every relevant phase space and invariant under transformations, since the action is invariant. Requiring $\mathcal{L}$ invariant under Lorentz transformations guarantees that all predictions of the theory are Lorentz invariant. From the principle of least action ($\delta S = 0$), the equations of motion can be written as;

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi \right)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \,, \tag{1.4}$$

where the ∂_μ stands for $\frac{\partial}{\partial x^\mu}$ and $\phi(x^\mu)$ are the fields. If the system is symmetric, the infinitesimal change in the field $\delta\phi_i$ yields the following equation;

$$\mathcal{L}(\phi + \delta\phi) = \mathcal{L}(\phi) \,. \tag{1.5}$$

Since the Lagrangian depends on the fields and their derivatives, the variation on Lagrangian is given as

$$\delta\mathcal{L}(\phi) = \mathcal{L}(\phi + \delta\phi) - \mathcal{L}(\phi) = \frac{\partial \mathcal{L}}{\partial \phi_j} \partial\phi_j + \frac{\partial \mathcal{L}}{\delta(\partial_\mu \phi_j)} \delta(\partial_\mu \phi_j) \,. \tag{1.6}$$

Using the equation of motion formula given in Equation 1.4 the following equality can be found;

$$\delta_\mu \left[\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi_j)} \delta\phi_j \right] = 0 \,. \tag{1.7}$$

Whenever the Lagrangian or the action is invariant under a set of continuous transformations, there is a associated conserved charge. Thus, the conserved current satisfies the $\partial_\mu J^\mu = 0$, then the current is derived as,

$$J^\mu = \frac{\mathcal{L}}{\delta(\partial_\mu \phi_j)} \delta\phi_j \,. \tag{1.8}$$

Therefore integrating over spacetime gives the conserved charge Q,

$$Q = \int d^3x J_0(x)\,. \tag{1.9}$$

1.2.2 The Quantum Electrodynamics

The QFT that describes the electromagnetic force is called *the Quantum Electrodynamics* (QED). The basic forms of matter in the SM, the quarks and the leptons are spin $1/2$ particles. When considering a free fermion field ψ, their motion can be expressed by the Dirac Lagrangian as follows:

$$\mathcal{L} = \overline{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi\,, \tag{1.10}$$

where m indicates the mass of the fermion and γ^μ are the 4×4 Dirac matrices represented as follows:

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix},$$

where σ^i are the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.11}$$

and $\bar{\psi} = \psi^\dagger\gamma^0$ is the hermitian transformation of the field ψ. Since the free Lagrangian is already invariant under Lorentz transformations and spacetime, this yields to an energy and momentum conservation through Noether theorem. The QED must be invariant under the local gauge transformations of the $U(1)$ Lie group, resulting in electric charge conservation. Using the Equation 1.4, one can obtain through 1.10,

$$\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0\,, \tag{1.12}$$

which is the Dirac equation of motion for a free fermionic field ψ of mass m. Equation 1.12 is invariant under local gauge transformation therefore,

$$\psi(x) \rightarrow \psi' = U(\alpha)\,\psi(x)\,, \tag{1.13}$$

where the transformation element is,

$$U(\alpha) = e^{i\alpha} \tag{1.14}$$

form the abelian Lie group $U(1)$ and α is a fixed phase transformation. The infinitesimal change yields $\psi \rightarrow (1 + i\alpha)\,\psi$, through the Equation 1.12, the following equality can be found;

$$\partial_\mu j^\mu = 0\,, \tag{1.15}$$

where,

$$j^{\mu} = \frac{ie}{2}\left(\frac{\partial \mathcal{L}}{\partial\left(\partial_{\mu}\psi\right)}\psi - \overline{\psi}\frac{\partial \mathcal{L}}{\partial\left(\partial_{\mu}\overline{\psi}\right)}\right). \tag{1.16}$$

Here the quantity $e = 2\alpha$. Employing Equation 1.10, this yields to

$$j^{\mu} = -e\overline{\psi}\gamma^{\mu}\psi, \tag{1.17}$$

where the quantity j^{μ} is defined as the electromagnetic charge current density. Hence from the Noether theorem electromagnetic charge is defined as;

$$e = \int d^3x j^0, \tag{1.18}$$

which is the conserved charge associated with the $U(1)$ gauge symmetry.

In the Equation 1.13, the phase transformation can have spacetime dependence and therefore, the Lagrangian is not anymore invariant under these transformations;

$$\mathcal{L} \rightarrow \mathcal{L}' = \mathcal{L} + \overline{\psi}\gamma^{\mu}\psi(\partial^{\mu}\alpha)\,. \tag{1.19}$$

In order to guarantee a local gauge invariance of the Lagrangian, it is imperative to establish the gauge field $A_{\mu}(x)$, which transforms as follows:

$$A_{\mu}(x) \rightarrow A'_{\mu}(x) = A_{\mu}(x) + \frac{1}{e}\partial_{\mu}\alpha\,(x)\ . \tag{1.20}$$

The derivative term ∂_{μ} is substituted by *covariant derivative* D_{μ}:

$$\mathcal{D}_{\mu} = \partial_{\mu} - ieA_{\mu}(x)\,. \tag{1.21}$$

Therefore the Lagrangian that is invariant under local $U(1)$ transformations is obtained as;

$$\mathcal{L} = i\overline{\psi}\gamma^{\mu}\mathcal{D}_{\mu}\psi - m\overline{\psi}\psi\,, \tag{1.22}$$

$$= i\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\overline{\psi}\psi + e\overline{\psi}\gamma^{\mu}\psi A_{\mu}\,, \tag{1.23}$$

$$= i\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\overline{\psi}\psi - j^{\mu}A_{\mu}\,. \tag{1.24}$$

where $J^{\mu} = \bar{\psi}\gamma^{\mu}\psi$ is a term associated to the electromagnetic current, and can be interpreted as an interaction term between the electron and the *photon* field A_{μ}. The resulting Lagrangian

is therefore written as;

$$\begin{aligned}\mathcal{L}_{U(1)} &= \bar{\psi}(i\gamma^{\mu}D_{\mu} - m)\psi\,, \\ &= \mathcal{L}_{free} - J^{\mu}A_{\mu}\,.\end{aligned}$$

The A_{μ} field is interpreted as the mediator of the electromagnetic interaction, which is the *photon*. This local gauge transformation yields to the existence of gauge mediating boson for an interaction between charged leptons with an electromagnetic charge e. The kinetic term is also needed to be added to the Lagrangian for its propagation. Without violating the group symmetry, the solution is to introduce the field strength tensor $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ which can be obtained by computing the transformation of the commutation operator $[D_{\mu}, D_{\nu}]$ with respect to the $U(1)$ local gauge transformations. Combining these informations yield to the Lagrangian of QED as follows:

$$\begin{aligned}\mathcal{L}_{QED} &= \bar{\psi}(i\gamma^{\mu}D_{\mu} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \\ &= \mathcal{L}_{free} - \underbrace{qJ^{\mu}A_{\mu}}_{\text{interaction term}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{kinetic term}}\,.\end{aligned}$$

The mass term of the photon is defined as

$$\frac{1}{2}m_{\gamma}^{2}A_{\mu}A^{\mu}. \tag{1.25}$$

However, the gauge invariance is violated and can be easily seen by using Equation 1.2.2;

$$A_{\mu}A^{\mu} \rightarrow \left(A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha\right)\left(A^{\mu} + \frac{1}{e}\partial^{\mu}\alpha\right) \neq A_{\mu}A^{\mu}\,. \tag{1.26}$$

Since the mass term is violating the local symmetry, the photon must be massless.

Another representation of the QED Lagrangian is the *Weyl* or *chiral* representation of the fields. The 4-component Dirac spinors (Equation 1.2.2) can be represented by 2-component spinor;

$$\psi = \begin{pmatrix}\psi_L \\ \psi_R\end{pmatrix}$$

where L and R represent the left-handed and right-handed spinors components. The Lagrangian can be defined using these spinors and the terms with the γ_{μ} matrices are attaching the components with the same chirality while the mass term is connecting the left-handed to the

right-handed parts:

$$\mathcal{L}_{QED} = \underbrace{\bar{\psi_L} i\gamma^\mu \partial_\mu \psi_L + \bar{\psi_R} i\gamma^\mu \partial_\mu \psi_R}_{\text{kinetic term}} - \underbrace{qA_\mu(\bar{\psi_L}\gamma^\mu \psi_L + \bar{\psi_R}\gamma^\mu \psi_R)}_{\text{interaction term}} - \underbrace{m(\bar{\psi_R}\psi_L + \bar{\psi_L}\psi_R)}_{\text{mass term}} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

In summary, imposng the local symmetry on a free Lagrangian yield to the new vector field, expressed as the photon field, where its mass term is not allowed.

1.2.3 The Quantum Chromodynamics

The Quantum Chromodynamics (QCD) is the QFT of the strong interactions between quarks and gluons, which are colour carrying particles. It is a $SU(3)$ Yang-Mills theory [12] which is a non-abelian gauge group. There are 8 gauge fields (*gluons*) corresponding to the eight generators of group representation and since they carry colour charge, theory allows the gluon to self interact. The quark fields (ψ) are represented by 3-dimensional vectors called *triplet*, and are required to transform as;

$$\psi \rightarrow \psi' = e^{i\alpha^a(x)t^a}\psi = U(x)_{ij}\psi \, .$$

This local symmetry is non-commuting which makes it non-abelian. The generator t^a of the gauge group (represented by eight 3×3 Gell-Mann matrices) obey the Lie algebra commutation relation;

$$[t^a, t^b] = i f_{abc} t^c \, , \tag{1.27}$$

where f_{abc} are the structure constants of $SU(3)$ and the Gell-Mann matrices are defined as;

$$t^1 = \frac{1}{2}\begin{pmatrix} & 1 & \\ 1 & & \\ & & \end{pmatrix}, \quad t^2 = \frac{1}{2}\begin{pmatrix} & -i & \\ i & & \\ & & \end{pmatrix}, \quad t^3 = \frac{1}{2}\begin{pmatrix} 1 & & \\ & -1 & \\ & & \end{pmatrix},$$

$$t^4 = \frac{1}{2}\begin{pmatrix} & & 1 \\ & & \\ 1 & & \end{pmatrix}, \quad t^5 = \frac{1}{2}\begin{pmatrix} & & -i \\ & & \\ i & & \end{pmatrix}, \quad t^6 = \frac{1}{2}\begin{pmatrix} & & \\ & & 1 \\ & 1 & \end{pmatrix},$$

$$t^7 = \frac{1}{2}\begin{pmatrix} & & \\ & & -i \\ & i & \end{pmatrix}, \quad t^8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix}.$$

As in QED, to ensure the gauge invariance, the covariant derivate is defined as $D_\mu = \partial_\mu - igt^a G^a_\mu$ where G^a_μ represent the gluon fields and g is the interaction coupling constant. Gluon fields as well are required to be transformed to contain the field and its derivative as follows:

$$G^a_\mu \rightarrow G'^a_\mu = U(x)(G^a_\mu t^a + \frac{i}{g}\partial_\mu)U(x)^\dagger \, . \tag{1.28}$$

The QCD Lagrangian can be written while respecting the non-abelian structure of the group;

$$\mathcal{L}_{\text{QCD}} = \overline{\psi}_f(i\gamma^\mu D_\mu - m)\psi_f - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a \, , \tag{1.29}$$

where $F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g f_{abc} G^b_\mu G^c_\nu$ is the strength tensor defined by

$$[D_\mu, D_\nu] = -igF^a_{\mu\nu}t^a \, . \tag{1.30}$$

In comparison to the field tensor of QED, there is an extra term coming from the non-abelian nature of the QCD and lead to triple and quadrate gluon self-interactions, which then allow the properties of confinement and asymptotic freedom. This quantity has the infinitesimal transformation,

$$F^a_{\mu\nu} \rightarrow F^a_{\mu\nu} - f_{abc}\alpha^b F^c_{\mu\nu}, \tag{1.31}$$

where the corresponding classical equation of motion is derived as;

$$\partial_\mu F^a_{\mu\nu} + g f_{abc} G^{b\mu} F^c_{\mu\nu} = -g j^a_\nu, \tag{1.32}$$

where $j^a_\nu = \overline{\psi}\gamma_\nu t^a \psi$ is the global symmetry current of the fermion field. There are several kinds of hadrons that are observed, which can be categorised by their quark content. Hadrons are divided into *mesons* consisting of two quarks, and *baryons* made of three quarks. Free Lagrangian without the interaction terms can be obtained by setting the coupling constant to zero. Therefore Lagrangian contains eight gluon fields and three fermion fields corresponding to the dimension of the gauge group $SU(3)$. Thus, the QCD Lagrangian density can be written in terms of free and interaction parts of it;

$$\begin{aligned}\mathcal{L}_{QCD} = &\underbrace{\overline{\psi}_f(i\gamma^\mu\partial_\mu - m)\psi_f}_{\text{free fermion term}} - \underbrace{\frac{1}{2}(\partial_\mu G^a_\nu \partial^\mu G^\nu_a - \partial_\nu G^a_\mu \partial^\nu G^\mu_a)}_{\text{free gluon term}} \\ &\underbrace{+ g G^a_\mu \overline{\psi}_f \gamma^\mu t^a \psi_f - g f_{abc}(\partial_\mu G^a_\nu) G^\mu_b A^\nu_c - \frac{1}{4} g^2 f_{abc} G^b_\mu G^c_\nu f_{ade} G^\mu_d G^\nu_e}_{\text{interaction term}}.\end{aligned} \tag{1.33}$$

Interaction term contains three different interactions: quark-gluon vertex with strength proportional to g, gluon-gluon interactions where the second term is between three gluons with strength proportional to g and the third term is between four gluons with strength proportional to g^2.

Additional Fadeev-Popov ghost term is included to be consistent with the path integral formalism which is used to quantize the theory. The ghost term is given as

$$\mathcal{L}_{ghost} = \partial_\mu \overline{c}^a D^\mu c^a, \tag{1.34}$$

where c^a is a ghost field, serves to cancel it out. Ghosts show up as virtual particles in internal closed loops and can only couple to gluons.

1.2.4 The Electroweak Unification

The Electroweak theory (EWK) is the unified definition of electrodynamic and weak interactions. The unification is succeeded under the gauge symmetry group of $SU(2)_L \times U(1)_Y$, where L represents the *left-handed weak isospin* and Y stands for *hypercharge*. EWK is a chiral theory,

meaning that nature treats left-handed fermions differently from right-handed fermions. The chiral representation of the fermions is defined by considering the left-handed fields as a doublet (weak isospin $T = 1/2$) and the right-handed fields as a singlet (weak isospin $T = 0$):

$$\begin{aligned} L_L &= \begin{pmatrix} \nu_L^e \\ e_L \end{pmatrix} \begin{pmatrix} \nu_L^\mu \\ \mu_L \end{pmatrix} \begin{pmatrix} \nu_L^\tau \\ \tau_L \end{pmatrix}, & L_R &= e_R, \mu_R, \tau_R \\ Q_L &= \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, & U_R &= u_R, c_R, t_R \\ & & D_R &= d_R, s_R, b_R \end{aligned} \tag{1.35}$$

The theory based on the $SU(2)$ local gauge invariance is predicted which transforms the fields as:

$$\begin{aligned} \psi_L &\rightarrow \psi'_L = e^{i\beta(x).\sigma}\psi_L\,, \\ \psi_R &\rightarrow \psi'_R = \psi_R\,, \end{aligned}$$

where σ are the Pauli matrices given in the Equation 1.2.2. This transformation yields to the generation of a massless vector boson triplet $W_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$. Therefore the same convention is followed as QED and the covariant derivative functioning on left-handed doublets and right-handed singlets are derived as follows:

$$\begin{aligned} \partial_\mu\psi_L \rightarrow D_\mu\psi'_L = (\partial_\mu - i\,g\,\frac{\sigma_i}{2}W_\mu^i)e^{i\theta(x).\sigma}\psi_L\,, \\ \partial_\mu\psi_R \rightarrow \partial_\mu\psi'_R = \partial_\mu\psi_R\,, \end{aligned} \tag{1.36}$$

The non-abelian $SU(2)$ Yang-Mills theory is introduced and therefore the fields are represented as,

$$W_\mu^i \rightarrow W_\mu^{i'} = W_\mu^i - \frac{1}{g'}\cdot\partial_\mu\theta^i(x) + f^{abc}W_\mu^b\theta^c(x)\,,$$

where f^{abc} is defined from the generators of the subgroup algebra as it is in the QCD case (Equation 1.2.3). The Lagrangian kinetic term for leptons can be given as

$$\begin{aligned} \mathcal{L}_{\text{kin.}} &= i\bar{L_L}\gamma^\mu D_\mu L_L + i\bar{L_R}\gamma^\mu\partial_\mu L_R\,, \\ &= \underbrace{\frac{g}{\sqrt{2}}(J^{+,\mu}W_\mu^+ + J^{-,\mu}W_\mu^-)}_{\text{charged current}} + \underbrace{g(\bar{\nu_L}\gamma^\mu\nu_L - \bar{l_L}\gamma^\mu l_L)W_\mu^3}_{\text{neutral current}}\,, \end{aligned}$$

where the charged currents $J^{\pm,\mu}$ are defined as currents. The quark model can be also represented by replacing the lepton fields with quark fields. The similarities between weak and electromagnetic interactions yield to unification of these forces. Therefore, in 1961 the Glashow-Salam-Weinberg model of electroweak interactions is built by combining SU(2) Yang-

Mills Lagrangian and U(1) Lagrangian while ensuring that weak bosons couple to left-handed fermions only, thus the electroweak Lagrangian ($\mathcal{L}_{\mathrm{EW}}$) is formed as follows:

$$\mathcal{L}_{\mathrm{EW}} = \sum_{j=1}^{3} \bar{\psi}_j [i\gamma^\mu D_\mu]\psi_j - \frac{1}{4} W_i^{\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \,. \tag{1.37}$$

Local gauge transformations of $SU(2)_L \times U(1)_Y$ is required with following laws:

$$\begin{aligned}
\psi_L &\rightarrow \psi'_L = e^{i(\beta(x).\sigma + \alpha(x)\cdot Y)} \psi_L \,, \\
\psi_R &\rightarrow \psi'_R = e^{i\alpha(x).Y} \psi_R \,, \\
\partial_\mu \psi_L &\rightarrow D_\mu \psi'_L = (\partial_\mu - i\, g\, \frac{\sigma_i}{2} W^i_\mu - i\, g'\, \frac{Y}{2} B_\mu) e^{i(\theta(x).\sigma + \alpha(x)\cdot Y)} \psi_L \,, \\
\partial_\mu \psi_R &\rightarrow D_\mu \psi'_R = (\partial_\mu - i\, g'\, \frac{Y}{2} B_\mu) \psi_R \,.
\end{aligned}$$

Therefore the Lagrangian of the electroweak interaction can be derived as:

$$\begin{aligned}
\mathcal{L}_{\mathrm{int}} = \frac{g}{\sqrt{2}} (J^{+,\mu} W^+_\mu + J^{-,\mu} W^-_\mu) + g \underbrace{(\bar{\psi}_L \gamma^\mu \frac{\sigma_3}{2} \psi_L)}_{\text{weak isospin current}} W^3_\mu + g' \underbrace{(-\bar{\psi}_L \gamma^\mu \frac{1}{2} \psi_L - \bar{\psi}_R \gamma^\mu \psi_R)}_{\text{weak hypercharge current}} B_\mu \\
= \frac{g}{\sqrt{2}} (J^{+,\mu} W^+_\mu + J^{-,\mu} W^-_\mu) + g J^{3,\mu} W^3_\mu + g' J^{Y,\mu} B_\mu \,,
\end{aligned} \tag{1.38}$$

where the field strength tensors are:

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g_W \epsilon_{ijk} W^j_\mu W^k_\nu \,, \tag{1.39}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \,. \tag{1.40}$$

In the Equation 1.37 the last two terms represent the gauge boson self interactions and ψ_i are the fermion fields for the first-generation quarks or leptons. W^i and B_μ represent the $SU(2)$ and $U(1)$ boson fields, respectively and the g' and g are coupling strength factors for the W^i and B_μ fields. In nature the SU(2) fields (W^i_μ) and the U(1) fields (B_μ) are heavily mixed together to the form of the observed electroweak gauge bosons

$$A_\mu = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W \;\; \text{(Photon)} \,, \tag{1.41}$$

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W \;\; (Z \text{ boson}) \,, \tag{1.42}$$

$$W^\pm_\mu = \frac{1}{\sqrt{2}} (W^1_\mu \pm W^2_\mu) \;\; (W^\pm \text{ boson}) \,. \tag{1.43}$$

Particle	T	T^3	Q	Y
$\nu_e,\ \nu_\mu,\ \nu_\tau$	$\frac{1}{2}$	$\frac{1}{2}$	0	-1
$e_L^-,\ \mu_L^-,\ \tau_L^-$	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1
$e_R^-,\ \mu_R^-,\ \tau_R^-$	0	0	-1	-2
$u_L,\ c_L,\ t_L$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$
$d_L,\ s_L,\ b_L$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{3}$
$u_R,\ c_R,\ t_R$	0	0	$\frac{2}{3}$	$\frac{4}{3}$
$d_R,\ s_R,\ b_R$	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$

Table 1.3: The fermion charges for weak isospin T, its third component T^3, electric charge Q, and hypercharge Y.

where θ_W is the *weak mixing angle* (also called the Weinberg angle) which relates the g and g' couplings as:

$$\sin\theta_W = \frac{g}{\sqrt{g^2+g'^2}},\ \cos\theta_W = \frac{g'}{\sqrt{g^2+g'^2}}. \tag{1.44}$$

In order to fix the degrees of freedom, one can choose $g\sin\theta_W = g'\cos\theta_W$, and restoring the electromagnetic part of the interaction together with the weak interactions, the Lagrangian can be written as;

$$\begin{aligned}\mathcal{L}_{\text{int}} &= \frac{g}{\sqrt{2}}(J^{+,\mu}W_\mu^+ + J^{-,\mu}W_\mu^-) + gJ^{3,\mu}W_\mu^3 + g'J^{Y,\mu}B_\mu \\ &= \frac{g}{\sqrt{2}}(J_{CC}^{+,\mu}W_\mu^+ + J_{CC}^{-,\mu}W_\mu^-) + \frac{g}{\cos\theta_W}J_{NC}^\mu Z_\mu \\ &\quad + g\cdot\sin\theta_W\, J_{EM}^\mu A_\mu\end{aligned} \tag{1.45}$$

by including the weak charged current (J_{CC}^μ) , the weak neutral current (J_{NC}^μ) and the electromagnetic current (J_{EM}^μ) for the interactions of fermions with the W boson, with the Z boson and with the photon, respectively.

However, the Lagrangian derived until now does not contain the mass terms for fermion fields as well as the gauge fields. This situation is in contrast with the experimental findings, therefore the solution will be given by introducing a new scalar field for vector boson and by Yukawa interactions for fermions.

1.2.5 Symmetry Breaking: The Higgs Mechanism

To the framework described so far, adding mass terms for the gauge bosons, or the fermions, would break the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ local gauge symmetry of the EW theory since the mass terms involve mixing $SU(2)$ doublets with $SU(2)$ singlets. Therefore, the *spontaneous symmetry breaking* mechanism is introduced into the model by Brout-Englert-Higgs. The motivation is

typically arising from the fact that all the local symmetries of a Lagrangian are not necessarily also symmetries of the vacuum state of the system when it is the degenerated. *Spontaneous* implies that the preference on the vacuum state is arbitrary. This is accomplished by defining a spin-0 scalar complex *Higgs field.*

In order to explain the Higgs mechanism, one can start with a simple case to grasp the basic concept of the structure. Therefore, for the simple abelian $U(1)$ scalar field ϕ, the Lagrangian is defined as

$$\mathcal{L}_\phi = \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi), \tag{1.46}$$

where the form of the potential energy $V(\phi)$ is chosen to be the most general renormalizable potential allowed by the $U(1)$ gauge invariance

$$V(\phi) = [\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2]. \tag{1.47}$$

The Lagrangian given in Equation 1.46 is invariant under local $U(1)$ gauge transformations. Therefore, it requires the introduction of a massless vector field A_μ and the Lagrangian must be written in terms of the covariant derivative

$$\begin{aligned} \phi &\rightarrow \phi' = e^{-i\cdot\beta(x)}\phi, \\ A_\mu &\rightarrow A'_\mu = A_\mu - \frac{1}{g}\partial_\mu \beta(x), \\ \partial_\mu &\rightarrow D_\mu = \partial_\mu - i\, g\, A_\mu, \end{aligned} \tag{1.48}$$

Thus the full Lagrangian can be written as;

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{1.49}$$

The parameters μ and λ are initially free parameters of the potential term. The sign of μ^2 will give the shape of the potential. If $\mu^2 > 0$, the lowest energy state rises at $\phi = 0$, where the vacuum is empty. In this case the theory is simply QED with massless gauge bosons and a charged scalar field ϕ with mass μ without breaking the symmetry. On the other hand, in the case of $\mu^2 < 0$, the potential can be written as;

$$V(\phi) = -|\mu^2||\phi|^2 + \lambda(|\phi|^2)^2, \tag{1.50}$$

which has the Mexican hat shape shown in Figure 1.1. Therefore the vacuum is then unstable and a minimum occurs at;

$$\langle \phi \rangle = \sqrt{-\frac{\mu^2}{2\lambda}} \equiv \frac{\upsilon}{\sqrt{2}}, \tag{1.51}$$

where $\langle \phi \rangle$ represents the non-zero vacuum expectation value (vev) of the scalar field ϕ which breaks the $U(1)$ symmetry. To study the physics around the ground state, the expansion of the

field is applied as follows:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \upsilon + h(x) \end{pmatrix}, \tag{1.52}$$

with real $h(x)$ value. Therefore the Lagrangian can be written as;

$$\mathcal{L}_\phi = (\partial_\mu + igA_\mu)\frac{1}{2}(\upsilon + h(x))(\partial_\mu - igA_\mu)(\upsilon + h(x)) - \mu^2(\frac{1}{2}(\upsilon + h(x)))^2 + \frac{\lambda}{4}(\upsilon + h(x)))^4,$$

$$= \underbrace{\frac{1}{2}(\partial_\mu h)^2 - \lambda \cdot \upsilon^2 h^2}_{\text{massive scalar h}} + \underbrace{\frac{1}{2}g^2\upsilon^2 A_\mu^2}_{\text{mass term of the gauge boson}} + \underbrace{\frac{1}{2}g^2h^2A_\mu^2 + g^2\upsilon A_\mu^2 h}_{\text{coupling between h and the gauge boson}} \underbrace{- \lambda\upsilon h^3 - \frac{1}{4}\lambda h^4}_{\text{scalar self coupling}}$$

$$+\mathcal{O}(\text{higher order term in h}) - \frac{1}{4}F_{\mu\nu}\,. \tag{1.53}$$

Here the symmety is *hidden* i.e. it is spontaneously broken through the choice of the ground state. The mass of the gauge boson appeares with $m_A = g\upsilon$. Furthermore, $\beta(x)$ field is eliminated from the theory as it is transformed into the longitudinal component of massive gauge boson in a phase transition, therefore the gauge boson gained its extra degree of freedom by *absorbing* the Goldstone boson [13], which is a massless scalar field.

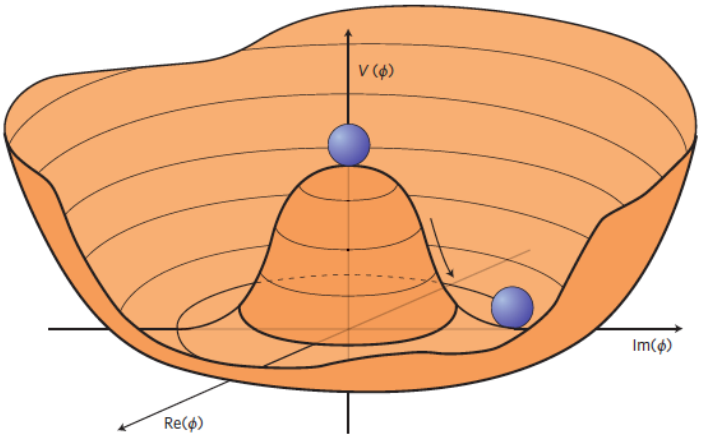


Figure 1.1: The Higgs potential shape as a function of Φ.

This model can be applied for the system with non-abelian gauge symmetry as well, such as the Weinberg-Salam Model with $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge theory. In this model a complex scalar field in the form of a $SU(2)$ doublet with $Y = 1$ is introduced as follows:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.54}$$

The associated Lagrangian of this field can be written as;

$$\mathcal{L}_{higgs} = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) - V(\Phi), \tag{1.55}$$

where $D_{\mu} = \partial_{\mu} + i\frac{g_2}{2}\sigma \cdot W_{\mu}^{a} + i\frac{g_1}{2}B_{\mu}Y$ and the potential term $V(\Phi)$ is given by

$$V(\Phi) = \mu^2|\Phi^{\dagger}\Phi| + \lambda\left(|\Phi^{\dagger}\Phi|\right)^2 . \tag{1.56}$$

As it was in the abelian case, one can study the physical consequences of *electroweak symmetry breaking* by perturbating the ground state by substituting the Equation 1.52. Finally the Lagrangian gains the extra terms of

$$\phi^{\dagger}\left(i\frac{g_1}{2}B_{\mu}Y + i\frac{g_2}{2}\sigma \cdot W_{\mu}\right)^{\dagger}\left(i\frac{g_1}{2}B_{\mu}Y + i\frac{g_2}{2}\sigma \cdot W_{\mu}\right)\phi\,. \tag{1.57}$$

The vacuum must be invariant under $U(1)$ gauge transformations therefore hypercharge must be $Y = 1$. The Equation given in 1.57 becomes,

$$\frac{1}{8}\frac{g_2^2}{8}(\upsilon + h(x))^2|W_{\mu}^1 + iW_{\mu}^2|^2 + \frac{1}{8}(\upsilon + h(x))^2|g_2W_{\mu}^3 + g_1B_{\mu}|^2\,. \tag{1.58}$$

The spontaneous symmetry breaking is generating the mass terms for W and Z bosons, which are defined as

$$W^{\pm} = \frac{1}{2}(W^1 \mp iW^2) \tag{1.59}$$

and

$$Z = \frac{g_2W^3 - g_1B}{\sqrt{g_2^2 + g_1^2}}\,. \tag{1.60}$$

The mass terms of the gauge boson are derived as

$$m_W = \frac{g_2\upsilon}{2}, \tag{1.61}$$

and

$$m_Z = \sqrt{g_2^2 + g_1^2}\frac{\upsilon}{2}\,. \tag{1.62}$$

Consequently the masses of the gauge bosons are obtained from the interaction with a scalar field whose ground state is broken under the symmetry $\mathrm{SU(2)}_L \otimes \mathrm{U(1)}_Y$ group. The Higgs boson mass is not predicted by the SM since λ and μ are free parameters. Furthermore the λ parameter gives rise to the trilinear and quartic self-coupling of the Higgs. It is still to be measured and is one of the particular interest for the physicists allowing to compute the full Higgs potential and thus understanding possible instabilities.

The Higgs mechanism explained previously, describes the mechanism of the gauge boson

masses, however introducing directly the fermion masses to the theory would still break the symmetries. Therefore, the new gauge invariant interaction term between the fermion field and the Higgs field, known as *Yukawa term*, is added supposing that the chiral fermions can gain their masses through the interaction with Higgs field after the spontaneous symmetry breaking. Since the Higgs field is defined as $SU(2)$ doublet, an $SU(2)$ invariant interaction term of fermions with Higgs field can be written. Therefore, the following interaction term can be added to the Lagrangian given in Equation 1.58 in order to conserve the symmetries of the total Lagrangian,

$$\mathcal{L}_{int} = -y_l \left[\overline{\psi}_L \Phi \psi_R + \overline{\psi}_R \Phi^\dagger \psi_L \right] . \tag{1.63}$$

The fermion fields can be represented by $SU(2)$ doublets or singlets, depending on their chirality as it is shown in Equation 1.2.4. Using this representation, the most general gauge invariant and renormalisable interaction term of the leptons can be written as;

$$\mathcal{L}_{leptons} = -\frac{y_l}{\sqrt{2}} \left(\overline{L}_L (\upsilon + h(x)) L_R + \overline{L}_R (\upsilon + h(x)) L_L \right) \Rightarrow -\frac{y_l \upsilon}{\sqrt{2}} \overline{\ell}\ell - \frac{y_l}{\sqrt{2}} \overline{\ell} h(x) \ell \, . \tag{1.64}$$

From this equation, the mass term for leptons can be extracted $m_l = {}^{y_l \upsilon}/\sqrt{2}$ where y_l represents a $3x3$ complex matrix, known as *Yukawa coupling* of the leptons. These matrices can be diagonalised by unitary transformations between the weak and mass eigenstates. Typically the gauge eigenstates defined as the weak isospin doublets, which are transformed by interaction through the W boson and the mass eigenstates, which are created by their interaction with the higgs boson. For leptons, these eigenstates are matching and they can be observed alone.

In the quark case, since the Higgs interactions yield the quark weak eigenstates to mix together, the resulting mass eigenstates are not the same as the weak eigenstates. Thus, the quark mass eigenstates do not take part as pure states in weak interactions. This indicates to apply diagonalization procedure which yields eventually to the observed mass eigenstate by mixing among the weak eigenstates of the quark fields. This unitary transformation connecting the two states of mass and weak basis is represented by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [14] as follows:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} . \tag{1.65}$$

1.3 The Higgs Boson

Searches of the Higgs boson have started right after the discovery of W and Z bosons by the UA1 and UA2 Collaborations at CERN. One of the first activities were carried out in 1990s at the Large Electron Positron (LEP) Collider, at the Tevatron and at the Stanford Linear Collider. As a result of the direct search of $e^+e^- \rightarrow ZH \rightarrow b\bar{b}$ up to $\sqrt{s} = 209$ GeV at LEP, the lower bound on Higgs boson was measured to be 114.4 GeV at 95% confidence level (CL)

[15]. On the other hand, the mass range 162 – 166 GeV at 95% CL [16] was excluded by Tevatron. Furthermore, the upper limit on the mass of Higgs had been improved to 152 GeV from precision EW measurements [17].

In 2009, the first protons started circulating in LHC and the experiments started recording data. LHC has reached centre of mass energy of 7 TeV in 2010 and data taking continued in 2011. An additional 5 fb^{-1} at 8 TeV was collected in early 2012. The combination of the 2011 and 2012 datasets collected by ATLAS and CMS were enough to claim the discovery of the Higgs boson. On July 4 2012, both collaborations made a joint announcement on the discovery of the Higgs boson, which is considered as one of the most important discoveries for particle physics [18, 19]. The discovery channels $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ determined the Higgs boson mass $m_H = 125.09 \pm 0.21(stat.) \pm 0.11(syst.)$GeV [20, 21].

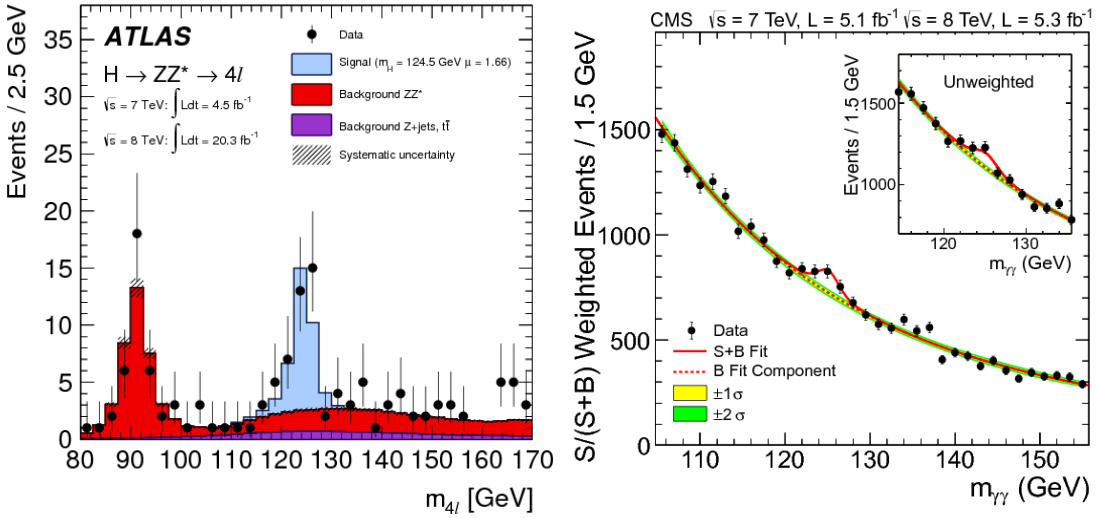
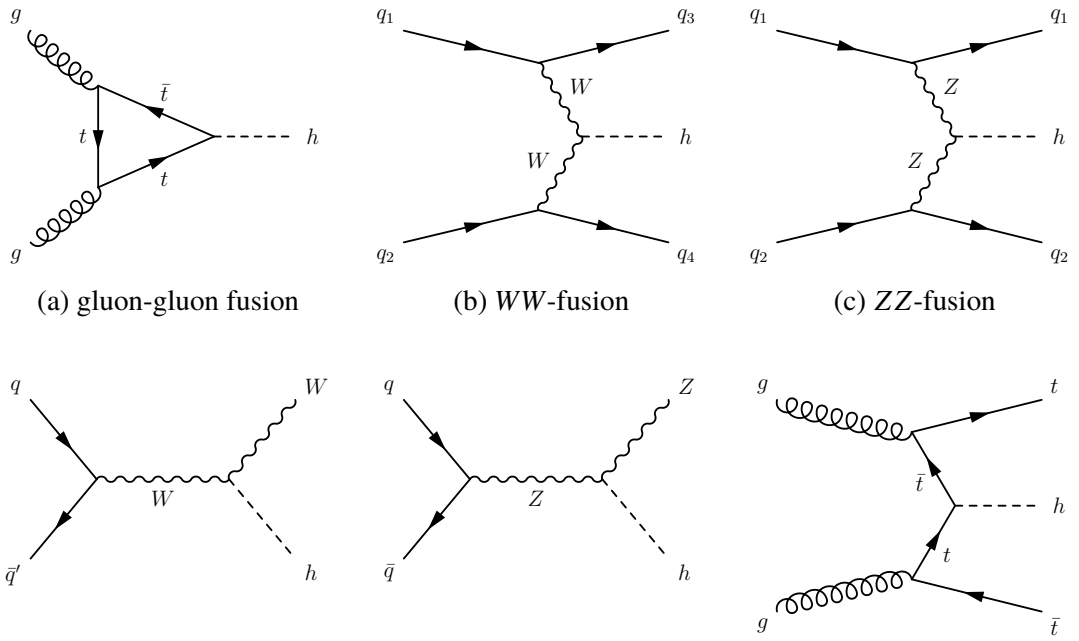


Figure 1.2: Higgs boson discovery by ATLAS experiment (***upper***) showing the distribution of the four-lepton reconstructed mass and by CMS experiment (***lower***) distribution of the diphoton invariant mass [18, 20].

1.3.1 Production and decay properties

The Higgs boson production at the LHC is governed by four modes. At LHC, the initial state consists of quarks and gluons thus the largest cross section of Higgs production mechanism is the gluon-initiated top loop (ggF). Another production mechanism of the Higgs is associated production with a pair of top quarks ($t\bar{t}H$). Alternatively, Higgs can be produced by its couplings to the W and Z bosons. Here the most important process is vector boson fusion (VBF) and Higgs propagator is involved to a t-channel W or Z exchange between two quarks. Besides VBF, Higgs can be produced via radiation from an s-channel W or Z boson, which is then called associated production of the gauge bosons. The leading order Feynman diagrams are shown in Figure 1.3.



(a) gluon-gluon fusion (b) WW-fusion (c) ZZ-fusion

(d) Wh production (e) Zh production (f) top quark fusion

Figure 1.3: Leading order Feynman diagrams of Higgs boson production processes.

The decays of the Higgs are governed by its coupling to the other SM particles. Although the largest coupling is to the top quark, the decay of the Higgs to $t\bar{t}$ is kinematically forbidden due to the large top mass. The most dominant decay of the Higgs is to b quarks (57.7%). Decays to weak vector bosons (W^+W^- and ZZ) have also large branching ratios although at least one of them needs to be off-shell. The SM branching ratios for the dominant decay modes are shown in Figure 1.4.

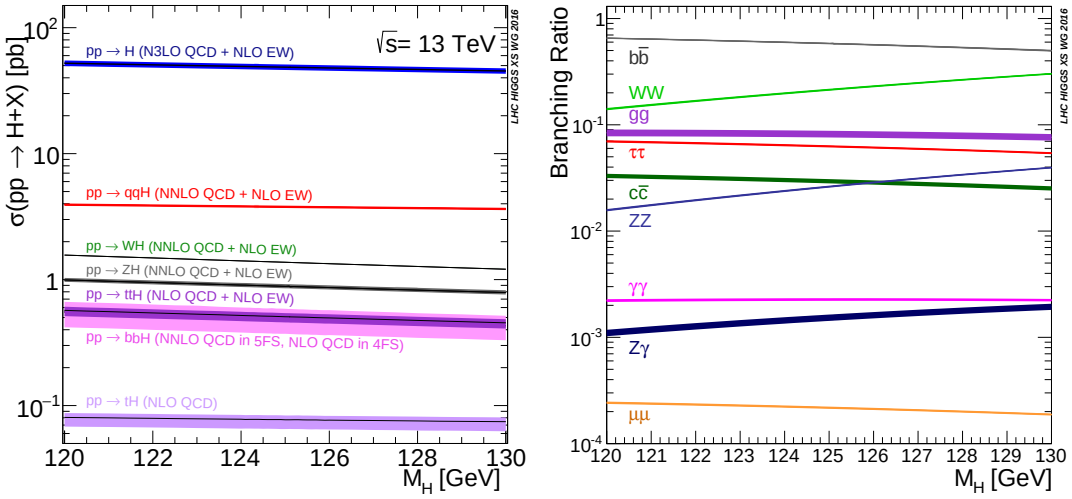


Figure 1.4: Cross sections for dominant production modes of Higgs boson at $\sqrt{s} = 13$ TeV (***left***) and Higgs boson branching fraction ratios to SM particles as a function of Higgs mass around 125 GeV (***right***) [22].

1.3.2 Higgs Boson pair-production

After the EWSB and the Higgs field acquires the vev, the Higgs potential can be obtained as follows:

$$V(\phi) \rightarrow V(\phi)_{\text{EWSB}} = -\lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 + \text{const.} \tag{1.66}$$

The first term of the above equation is the Higgs mass term, and the remainings are the trilinear and quadri-linear Higgs-self couplings,

$$\underbrace{m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2}}_{\text{Higgs boson mass}} \qquad \underbrace{\lambda_{hhh} \propto \frac{m_h^2}{v} \qquad \lambda_{hhhh} \propto \frac{m_h^2}{v^2}}_{\text{Trilinear and quadri-linear Higgs self-couplings}} \quad . \tag{1.67}$$

A measurement of these couplings would therefore give a hint about the actual structure of the potential, whose shape can have theoretical consequences. The quartic Higgs coupling, λ_{hhhh}, can not be measured at LHC since the cross section of triple Higgs production is small [23] [24], while the trilinear coupling can be probed directly through the Higgs pair production. At LHC, as for the single-Higgs production, the Higgs pair production (diHiggs) is dominantly in the loop induced gluon-fusion mechanism, ggF, through the destructive interference of two LO diagrams. Figure 1.5 shows the main diagrams that are involved in diHiggs production: top quark loops (box) and the triple Higgs production (triangle) [25]. The top quark loop diagram contribution to the overall cross section of the production is proportional to y_t^2, and

the contribution of the triangle production is the product of two couplings, λ_{hhh} and y_t. Cross section of diHiggs production is reduced due to the existence of two heavy particles in the final state, and the destructive interference between these two diagrams (the value of maximal destructive interference is $\kappa_\lambda = {}^{\lambda}/_{\lambda_{SM}} = 2.4$). Therefore resulting in even smaller cross section such that an observation is particularly challenging at LHC. Overall cross section is calculated as 31.05 fb, which is $O(10^3)$ times smaller than that of the dominant production modes of single higgs boson [26]. Further details about the calculation at different orders with available QCD corrections and top quark mass dependence can be found in the References [27–34].

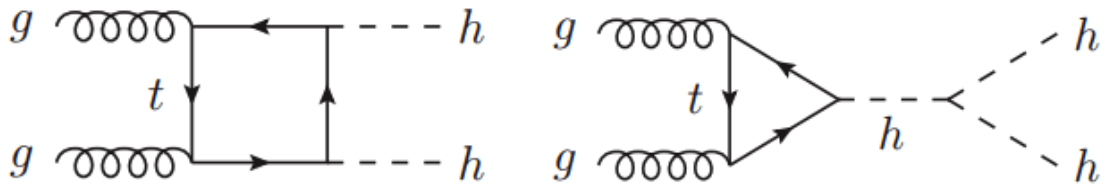


Figure 1.5: Contributing diagrams to ggF Higgs boson pair production via top quark loop (***left***) and via triple Higgs production (***right***).

Due to its lower cross section, the SM diHiggs production is not expected to be observed in the Run2, however it is possible to define limits on the measurement to constrain the BSM physics theories. The effective cross section of diHiggs production and the decay kinematics would be affected by κ_λ variations, considering the SM top quark coupling y_t. Currently at LHC, non-resonant searches are assuming that the total the cross section is affected by BSM constributions, which means only the Higgs boson self-coupling is affected by the BSM contribution while the rest of the coupling constants remain unaffected. The latest CMS and ATLAS constraints at 95% confidence level by combining various channels are $-11.8 \leq \kappa_\lambda \leq 18.8$ [35] and $-5 \leq \kappa_\lambda \leq 12.1$ [36], respectively (Figure 1.6). In addition, the Higgs self-coupling is sensitive to the single Higgs production through EW one-loop corrections [37]. The combined observed (expected) 95% CL interval constraint from single Higgs and diHiggs production on κ_λ is $-2.3 \leq \kappa_\lambda \leq 10.3$ $(-5.1 \leq \kappa_\lambda \leq 11.2)$ [38] at $\sqrt{s} = 13$ TeV using a dataset corresponding to an integrated luminosity of 36.1 fb^{-1}.

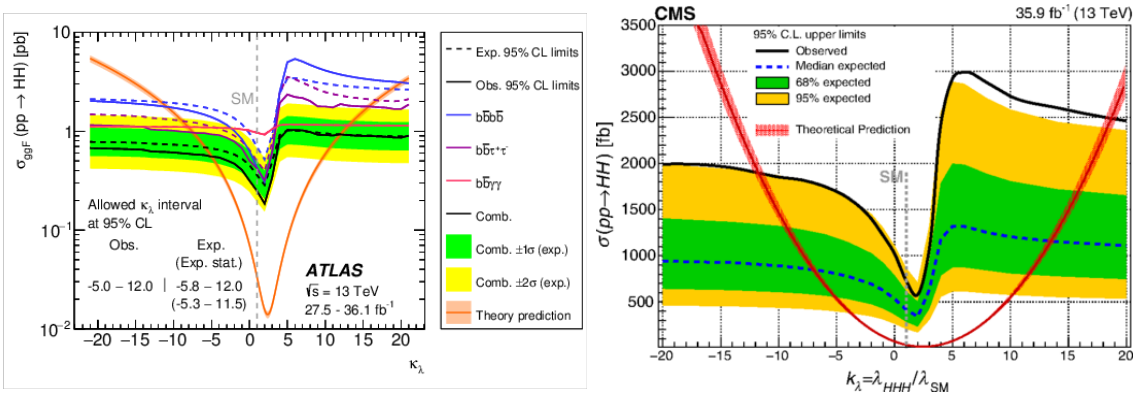


Figure 1.6: Expected and observed upper limits (95% CL) on the diHiggs production cross section as a function of κ_λ for ATLAS (***left***) and CMS (***right***) experiments at LHC.

The ATLAS experiment exploits several final states of diHiggs production. The analysis groups are categorised according to the Higgs boson decay modes in which they are shown in Figure 1.7 as a function of branching ratios. Among them, diHiggs multilepton group focuses on the diHiggs decay modes to WW^*WW^*, ZZ^*ZZ^*, $\tau\tau\tau\tau$, $WW^*\tau\tau$, $ZZ^*\tau\tau$, $ZZ^*b\bar{b}$ as well as the $\gamma\gamma + \ell$, where the signature of final states can be $2\ell/3\ell/4\ell + X$. The overall cross section for this channel is ≈ 2.8 fb.

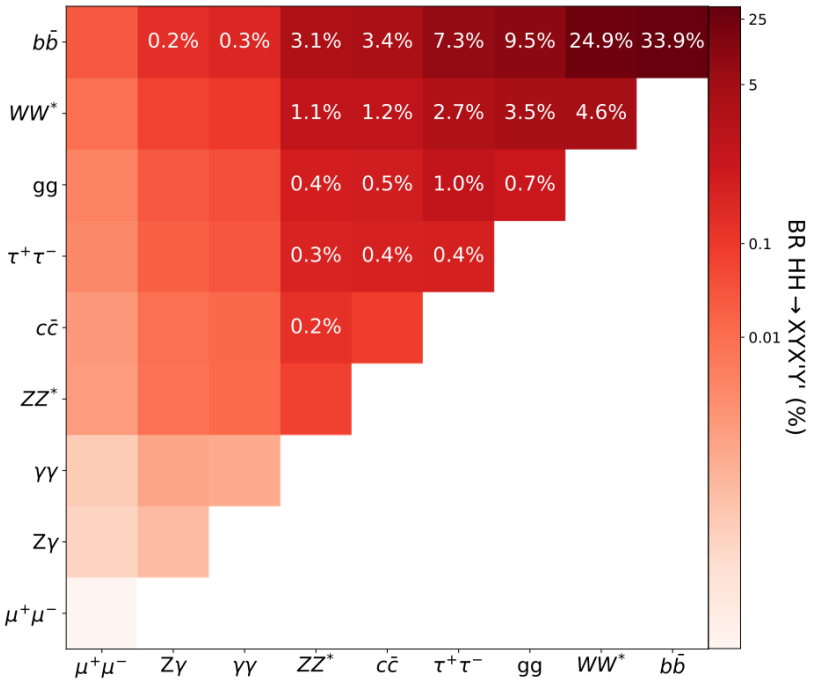


Figure 1.7: Branching ratios of the decay of an diHiggs pair to the selected group of final states. The axises arranged in increasing probability order [26].

1.3.3 Higgs Boson and top quark

The SM expectations lead the theory to remain consistent at the energy scales up to and beyond the Plank scale. The energy scale where the observations contradict theoretical expectations to be able to probe new physics. Due to the renormalised nature of the QFT, the hint might be around the renormalisation evolution of the coupling constants. In the theories which are asymptotically free, the theory can be valid arbitrarily in high energies. However, the theories which have opposite behaviour, the couplings become strong at high energies and diverges at the Landau pole. This is the motivation to search for a new physics to avoid the divergence. It is already known that around the Plank scale, the theory must be modified to incorporate the quantum gravity.

Some of the SM predictions are suitable to search for the new physics. For instance, the Higgs field itself has unique properties that may allow to probe the new physics such that its vacum expectation breaks the EWK symmetry and provides mass to W and Z bosons, define the group representation of left-right handed fermions. The top quark contribution to the effective potential is essential, due to having the largest Yukawa coupling constant. The variations of top Yukawa coupling can lead to the new minima on the Higgs potential at high energies, which could directly affect the vacuum expectation value and the metastability of the Universe. The dependency of the effective potential on Top coupling is shown in Figure 1.8.

The running of the Higgs self-coupling $\lambda(\mu)$ gets positive (negative) contribution from

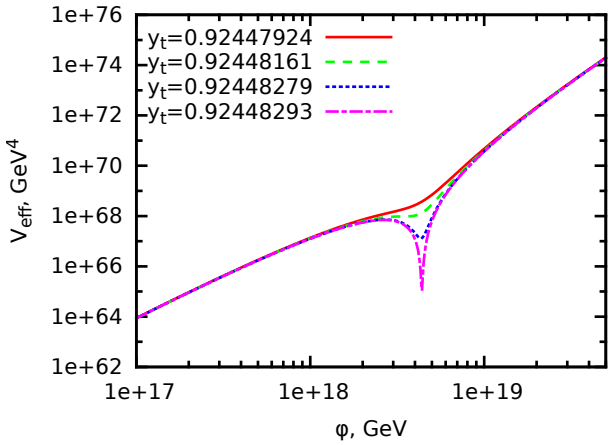


Figure 1.8: Dependence of the effective potential for the Higgs field with different top quark coupling constants [39].

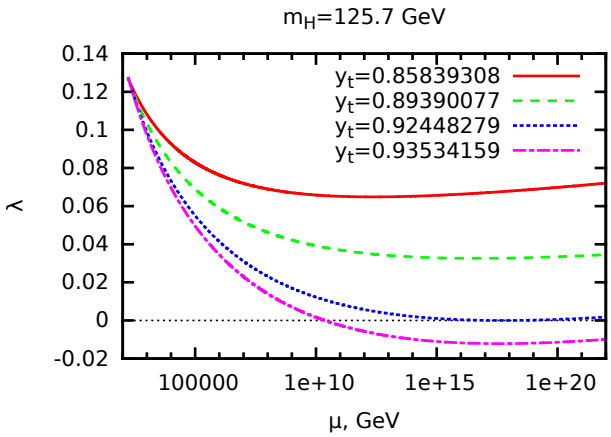


Figure 1.9: Running of the renormalised Higgs self-coupling parameter depending on the top quark coupling strengths [39].

bosonic(fermionic) interactions and the contribution is proportional to the mass of the particles. At large energy scales, depending on the masses of bosons and fermions, the divergence would point some new physics. Therefore, if fermion couplings are large enough to contribute dominantly than non-fermionic positive terms in the running of $\lambda(\mu)$, negative values of $\lambda(\mu)$ can be observed. Due to the dependence of Yukawa couplings on the particle mass and the fact that the top quark mass is far greater than the mass of other fermions, the major contribution to Equation 1.68 will be from the Top Yukawa coupling. The dependence of $\lambda(\mu)$ to the renormalisation scale for different values of y_t can be seen in Figure 1.9.

$$\frac{\mathrm{d}\lambda(\mu)}{\mathrm{d}\ln\mu} \propto -y_f^4 \tag{1.68}$$

The other possibility is that the couplings might remain finite around the Plank scale, which mainly depends on the Higgs boson and the top quark masses. Currently, the best experimental results show that this is an intermediarie state. Therefore the SM can actually remain valid all the way to the Plank scale. The Figure 1.10 shows the SM vacuum stability depending on Higgs and top quark masses.

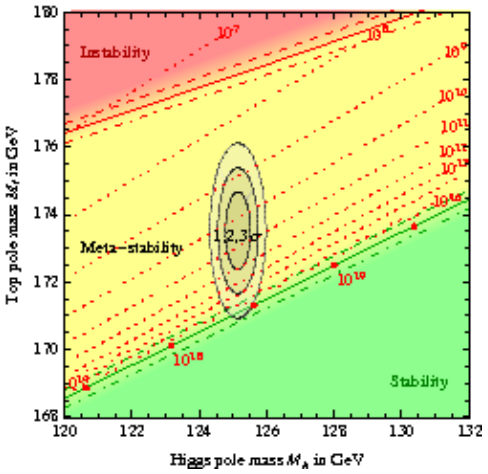


Figure 1.10: SM phase diagram as a function of Higgs and top pole masses. The area is separated into regions of stability, meta-stability, instability of the SM vacuum, and non perturbative region of the Higgs quartic coupling [40].

1.4 Limitations and open questions of the SM

Although the SM is able to describe successfully the experimental observations, there are still several known phenomena that are not addressed in this theory. A few items are described below.

Gravity The best existing theory for describing the gravitational force is *General Relativity*. However, formulating this theory in the quantized and renormalizable model frame is not achieved.

Dark matter and dark energy The existence of Dark Matter (DM) is proved by several types of measurements such as gravitational lensing, galaxy rotation curves, microwave background and so on. DM is assumed to interact weakly with any known matter, however the nature of DM is still unknown and does not have a fundamental description. The knowledge of dark energy is even poorer. It is known that the universe is expanding with increasing velocity from astrophysical observations. In order to explain this phenomenon, the dark energy is introduced and can interact only gravitationally. One of the most common theory indicates that the entire

spacetime is filled by low homogeneous constant energy, which contributes to 68% of the total energy in universe.

Matter-antimatter asymmetry In the SM, the only parameter allowing for CP violation is the CP-violating phase in the CKM matrix, which is too small to explain the asymmetry in the Universe. Thus, additional CP violating effects are necessary to account for the observed asymmetry.

Neutrino mass The evidence of neutrino oscillations [41] provides a clear indication for non-zero mass neutrinos, which indicates that the physical eigenstates are not the flavour eigenstates rather the mass eigenstates. Whereas in the SM, the neutrinos are left-handed particles, thus the non-zero term in the Lagrangian leads to the presence of the right-handed term. This observation contradicts with SM.

The hierarchy problem The observed mass of the Higgs boson is approximately 125 GeV, which indicates that the EWK symmetry breaks at the scale of $O(100\ \text{GeV})$. This is sensitive to high energy cut-off scales (Λ), which is around the Planck mass $M_P \approx 10^{18} - 10^{19}$ GeV. The higher-order corrections to the Higgs mass are contributing to this behaviour since the loop corrections to m_H^2 are proportional to Λ^2, whereas the fermionic corrections to the Higgs masss are proportional to the Λ as follows:

$$\Delta m_h^2 \, |y_f|^2 \left[\Lambda^2 + O\left(m_f^2 \ln\left(\frac{\Lambda}{m_f} \right) \right) \right] \tag{1.69}$$

where y_f is the Yukawa coupling, m_f is the associated fermion mass. Similar terms appear also for the gauge bosons, which yields the correction to be $O(10^{38}\ \text{GeV}^2)$ in order to reach the observed Higgs mass. The correction coming from the top quark itself contributes highest as shown in Figure 1.11.

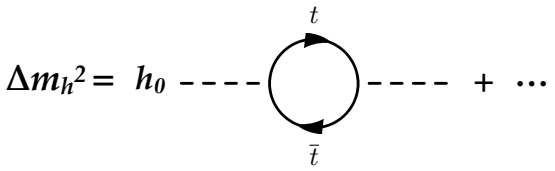


Figure 1.11: Top-quark loop contribution to the higher-order computation of the Higgs mass.

1.5 Search for Higgs Boson production in association with a Top-Antitop quark pair

Following the discovery of the Higgs boson, its properties are still being studied at the LHC with high precisions to understand better the model and to probe possible hints to BSM approaches. One of the most essential properties of the Higgs boson is the couplings to the fermions especially the heaviest fermion, the top quark, as the importance is discussed in Section 1.3.3.

The top Yukawa coupling can be probed indirectly through the top quark loops coupling to Higgs boson, Higgs boson production with the gluon-initiated top loop (ggF). Since the calculation of the loop processes require some assumptions, the potential contribution from new physics would vary the top Yukawa coupling constant. Nevertheless in the case of the Higgs boson production through association with $t\bar{t}$, the top Yukawa coupling constant can be measured directly since the process is tree-level at first order perturbation theory. The cross section (σ) of the production is proportional to squared of the matrix element of the process, which is proportional to the products of the couplings for each vertex in the Feynmann diagram. As a consequence, the production cross section can be interpreted as the $|y_t|^2$ while not being sensitive to the sign of the coupling. Furthermore, any deviation between the indirect and direct measurements could be a crucial proof for new physics.

The $t\bar{t}H$ process has a complex and rich diversity of possible final states, which are classified in the particular physics analysis by the Higgs decay mode. Among these, the $t\bar{t}H(H \to b\bar{b})$ process is the dominant one with large number of jets and b-jets[1] with one or more leptons coming from W boson. As a consequence of having large combinatorics in the final state and difficulties in modelling of the backgrounds, this channel is particularly challenging. On the contrary, $t\bar{t}H(H \to \gamma\gamma)$ offers a pure final state with small background and small branching ratio. Finally, the multilepton analysis which is targetting the $t\bar{t}H(H \to WW^*, ZZ^*, \tau\tau)$ stays in between the other analyses in terms of branching ratio and the complexity of the background modellings with one or more leptons (e, μ) in the final states. Tree-level Feynman diagrams contributing to the analysis is given in Figure 1.12.

As it is pointed out in Secion 1.3.1, $t\bar{t}H$ production has relatively low cross section ($\sigma(t\bar{t}H) = 500^{=35}_{-50}$ fb at NLO) , $O(100)$ with respect to ggF. Hence the measurements of the production cross section was challenging during the data taking periods of 2009–2013. At a centre of mass energy $\sqrt{s} = 7, 8$ TeV with analysis sensitive to $H \to WW^*$, $\tau\tau$, $b\bar{b}$, and $\gamma\gamma$ decays [42–46], the ATLAS and CMS collaborations have searched the $t\bar{t}H$ production in pp collisions and the combined result from the two collaborations yielded to a best fit ratio of observed cross section, $\mu_{t\bar{t}H} = {}^{\sigma}/_{\sigma_{SM}}$ of $2.3^{+0.7}_{-0.6}$. With an increased luminosity and centre of mass energy, evidence for $t\bar{t}H$ production was reported by ATLAS using 36.1 fb^{-1} of data collected during 2015 and 2016 [47], with an observed (expected) significance of 4.2σ (3.8σ). The best-fit result $\mu_{t\bar{t}H}$ of the $t\bar{t}H$ production to the SM expectation was found to be 1.2 ± 0.2 (stat) $^{+0.3}_{-0.2}$ (syst) and the production cross section at $\sqrt{s} = 13$ TeV is 790^{+230}_{-210} fb, both

[1] The detailed explanation of jets and b-jets are given in Section 5.3 and Section 5.3.3, respectively.

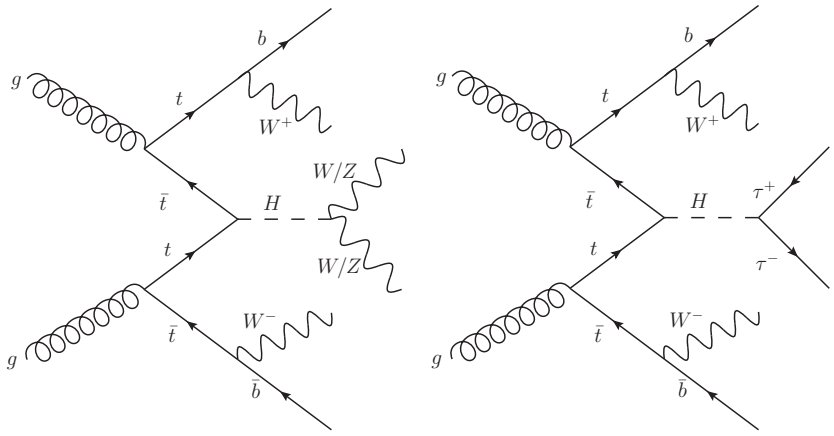


Figure 1.12: Illustration of tree-level Feynman diagrams for the production of $t\bar{t}H$ process. Higgs boson decays to WW^*/ZZ^* (***left***) or $\tau\tau$ (***right***) are shown.

consistent with the SM expectation.

Finally the observation of $t\bar{t}H$ production was announced by ATLAS using the data corresponding to an integrated luminosity of 79.8 fb^{-1} for $\gamma\gamma$ and ZZ^*, 36 fb^{-1} for multilepton and $b\bar{b}$ channels with 5.8σ observed significance, compared to 4.9σ expected significance [48].

The ATLAS $t\bar{t}H$-multilepton analysis presented in Chapter 7 using data corresponding to an integrated of 79.8 fb^{-1}. It is targetting to estimate the cross section of the process more precisely comparing to the previous studies (Ref. [42, 47]) by adding more data and improving the background modellings.

CHAPTER 2

The Large Hadron Collider and the ATLAS experiment

The CERN facility, the European Organization for Nuclear Research, is one of the largest particle physics laboratory located in Franco-Swiss border close by Geneva. Its purpose is to supply essential instruments for high energy physics searches such as particle accelerators, detectors or alternative infrastructures.

The Large Hadron Collider LHC [49], is a superconducting circular accelerator and collider at CERN. The operations and properties of LHC are given in Section 2.1. Detailed explanation of ATLAS experiment, which is the subject of the thesis is given is Section 2.2.

2.1 Large Hadron Collider

The LHC is a multi purpose particle collider, such that it uses protons and ions (lead and Xenon) to study the nature of the elementary particles. LHC ring is situated between 45 m and 17 m below the surface with a circumference of 26.7 km. The main function of LHC is accelerating and colliding two high energy proton beams (pp collisions), currently at the centre of mass energy of $\sqrt{s} = 13$ TeV. LHC is composed of multi-step accelerator systems in order to reach the required enegy of proton beams. The four biggest experiments at the LHC are ALICE [50], ATLAS [51], CMS [52] and LHCb [53]. CMS and ATLAS are the general purpose detectors, while LHCb is studying mainly the heavy-flavour physics and ALICE is specialised in heavy-ion physics. The special focus will be given to the ATLAS detector in the following chapters since the work exploits the data of this detector. Figure 2.1 displays the LHC accelerator complex together with experiments located in.

Protons begin their journey after they are separated from Hydrogen atom by electrical field and accelerated gradually by different accelerator mechanisms. The proton beams composed of spatially separated *bunches* of $\approx 1.15 \times 10^{11}$ protons and they are fed to the detector complex by 25 ns (40 MHz) bunch-crossings. Physics processes produced during the particle collisions are determined by a parameter called *instantaneous luminosity*, which is based on the properties of

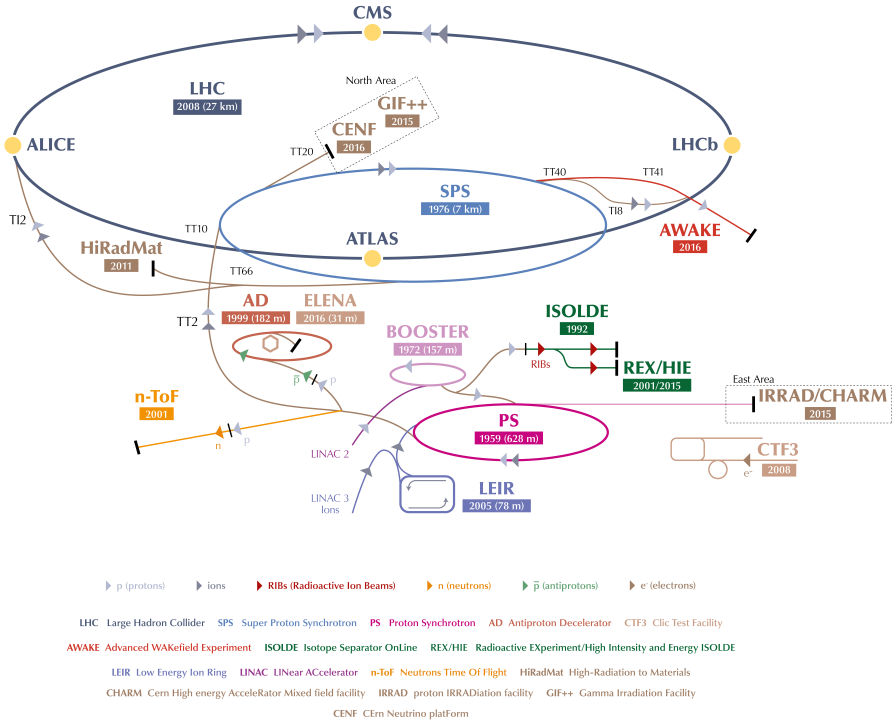


Figure 2.1: Sketch of the CERN accelerator complex including LHC and the experiments [54].

the bunch-crossing and their sizes. It is defined for gaussian beam distributions as:

$$L = \frac{N_1 N_2 n_b f_{\text{rev}}}{4\pi\sigma_x\sigma_x} F\,, \tag{2.1}$$

where N_1 and N_2 refer to the number of particles per bunch, n_b is the number of bunches per beam, $f_r = 11.245$ kHz is the revolution frequency, $\sigma_{x(y)} \approx 17\mu m$ represents the beam widths and since the particle beams do not collide head-to-head, the parameter F is added as geometric correction factor. The peak instantaneous luminosity of the LHC has reached approximetely $2.1 \times 10^{34} cm^{-2} s^{-1}$ in 2018. The number of collision events produced by LHC based on instantaneous luminosity delivered in a t time period is

$$N_{events} = \sigma_{event} \int L dt = \sigma_{event} \mathcal{L}\,, \tag{2.2}$$

where σ_{event} is the process cross section and $\mathcal{L}$ refers to *integrated luminosity*, which is widely used to define the collected or delivered data in a certain time of period. Luminosity measurements are done by several detectors in ATLAS, however a special detector called

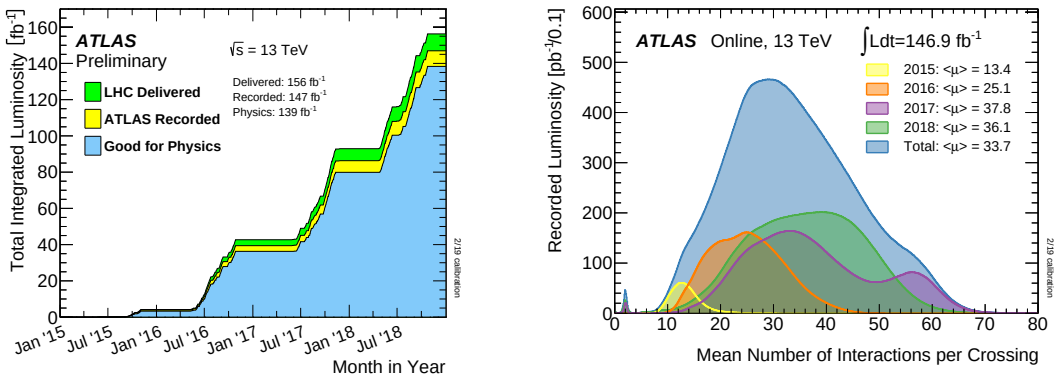


Figure 2.2: Integrated luminosity delivered by LHC and collected by ATLAS (***left***) and the pile-up profiles of the data recorded by ATLAS (***right***) during the years between 2015 and 2018 of Run2 [58].

LUCID [55] is dedicated to online monitoring of luminosity and the uncertainty calculations. The total integrated luminosity delivered by LHC and collected by ATLAS experiment for 2015-2018 data taking periods are given in on the left side of the Figure 2.2. ATLAS has recorded $\approx 36\, fb^{-1}$ in 2015-2016, $\approx 43\, fb^{-1}$ in 2017 and $\approx 60\, fb^{-1}$ in 2018 [56].

Due to the high instantaneous luminosity delivered by the LHC, each bunch crossing can result in more than one hard-scatter interaction and many secondary interactions. This kind of interactions are called *pile-up* interactions and the mean number of pp interactions per bunch crossing is usually denoted by $< \mu >$. These events are likely to interfere with hard process, forming as background that is hard to model but has to be considered in the physics analyses [57]. The pile-up distributions of the data reported by ATLAS experiment is illustrated in right side of the Figure 2.2 as a function of data taking years of LHC.

2.2 The ATLAS detector

ATLAS (A Toroidal LHC ApparatuS) is one of the general purpose detectors at LHC, located at the *point one* interaction site. It is a hermetic, cylindrically shaped detector, almost 4π coverage in solid angle around the interaction point, which permits to have great knowledge of pp collisions. A wide range of high energy physics analyses are carried on, including the precision measurements of SM parameters and searches for BSM.

The ATLAS detector is composed of multiple sequential layers and each serves for different purposes. These layers of subdetectors are organised cylindrically around the beam pipe (*barrel*) and/or as discs at the end of two sides of barrel regions (*endcap*). Additionally, ATLAS exploits a hyrbid system of superconducting magnets [59] to measure the particle momentum by bending their trajectories. The solenoidal magnet [60] in barrel region provides a 2 T field in the direction of z to bend the particles in the direction of ϕ, while toroidal magnets [61, 62] located at muon spectrometer level supplying magnetic field between 0.5 T and 4 T allowing a

precise momentum and spatial measurements of muons. Figure 2.3 displays the illustrative view of ATLAS detector with all the subdetectors and magnet systems.

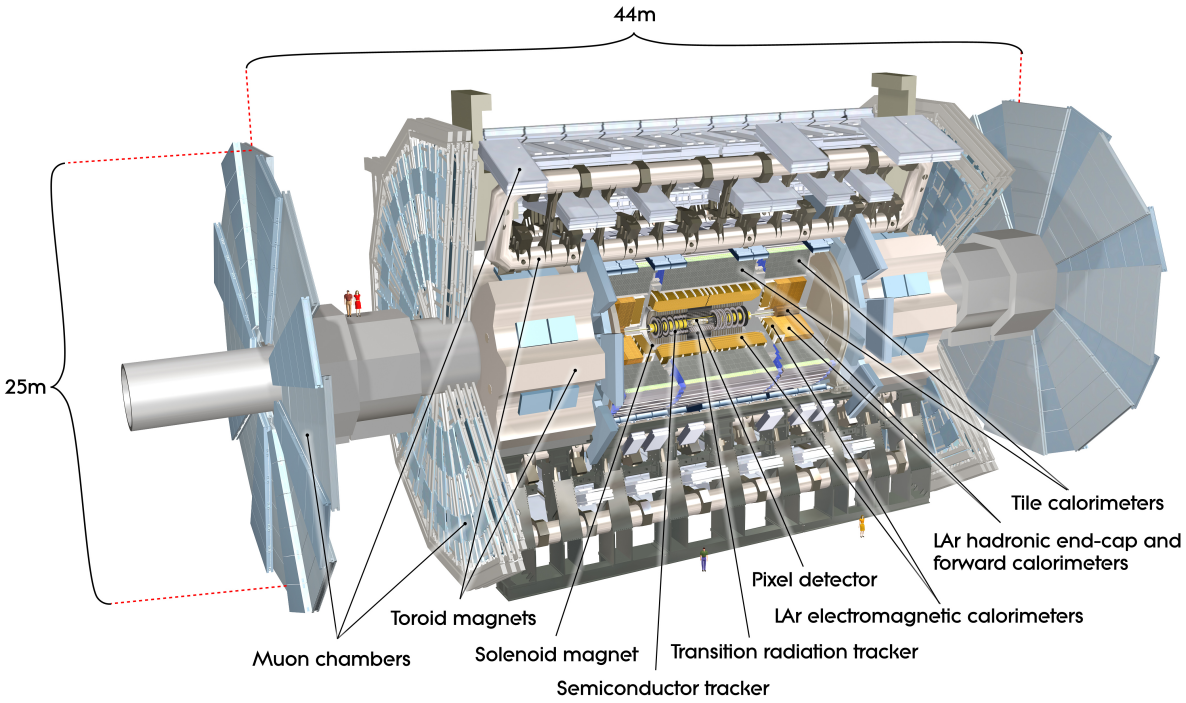


Figure 2.3: The cut-away view of complete ATLAS detector [63].

2.2.1 Coordinate system

ATLAS detector adapted a right-handed coordinate system, where its origin is at the *interaction point (IP)*. The direction of beampipe represents the z-axis, the x-axis points radially towards to the centre of LHC ring and the y-axis towards to the surface. Azimuthal angle (ϕ) is defined as the x-y plane and it is diagonal to z direction, where the polar angle (θ) represents the y-z plane. Figure 2.4 demonstrates the schematic view of the coordinate system of ATLAS detector.

Due to the relativistic nature of the particle interactions at LHC, the quantity *rapidity* is benefited from relativistic nature of physics. It is described as follows:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) , \tag{2.3}$$

where the momentum vector of the particle can be written as $\boldsymbol{p} = (p_T \cos(\phi), p_T \sin(\phi), p_z)$, and the $p_T = |\boldsymbol{p}| \sin(\theta)$ is the *transverse momentum*. In the relativistic regime of massive particles, where their rest mass is negligible with respect to their energies, rapiditiy is equivalent to the

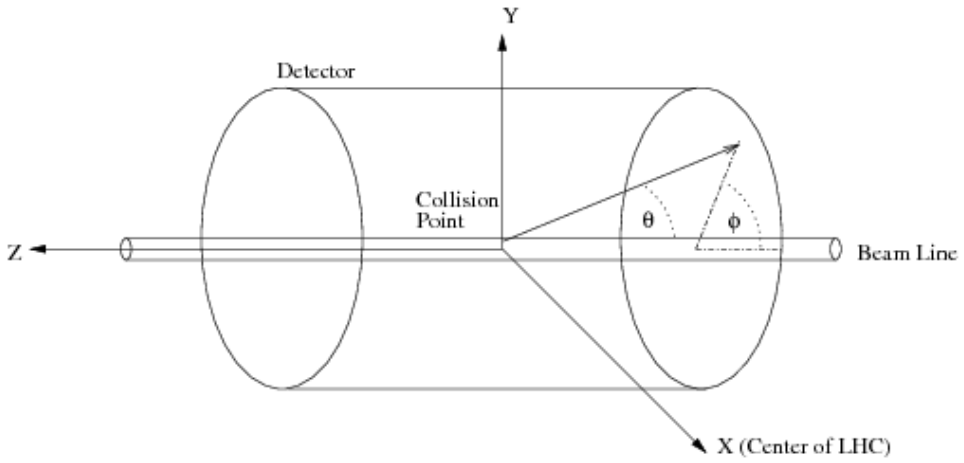


Figure 2.4: Illusturation of ATLAS coordinate system [64].

quantity called *pseudorapidity*:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) . \tag{2.4}$$

Therefore, the boosted measurements along the z direction are consistent within each other and Lorentz invariant in both lab and centre-of-mass frame. The angular distance between two objects in (η, ϕ) plane is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

2.2.2 The inner detector

Tracks of the charged particles are reconstructed in the Inner detector (ID) [65, 66] with a use of magnetic field. Tracks permit to reconstruct the main interaction point, *the primary vertex*. It covers the pseudorapidity range of $|\eta| \leq 2.5$ and formed as concentric cylinders centered around the beam pipe in the barrel, and disks in both end-caps. The ID is composed of three different complementary parts: the Pixel Detector, the Semiconducter Tracker (SCT) and the Transition Radiation Tracker (TRT).

Pixel detectors are located in the innermost part of the ID. Detector technology is based on the electron-hole pairs that are created in the semiconductor material during the passage of the charged particles. It is made up from three different layers (in addition insertable B-Layer (IBL)) and each layer contains several number of modules each consist of $\approx$ 46000 (27000) silicon pixels with a size of $50\mu m \times 400$ (250) μm. Therefore, it provides a precise measurement of space-points of particles with $\approx$ 92 million channels.

The SCT consists of four barrel layers and two end-caps with nine paralel discs, in total $\approx$ 4000 modules. Each silicon strip has a dimentionality of $80\,\mu m \times 12\,cm$. Layers at barrel region are assembled from double sided strips, each one tilted at 40 mrad provides better precision of the $r - \phi$ plane, in addition to z position measurements. In total 6.3 million readout

channels are used.

The TRT is situated at the most external part of the ID detector. It is using single wire drift tubes with *straw tubes* technology. When charged particles traverse the gas filled straws, surronding gas molecules are ionised. Later the resulting electrons are accelerated resulting in secondary particles. This process continues until the particle cascade reaches the wire, where the number of particles are proportional to the energy of the incoming particle. TRT measures position only in the $r - \phi$ plane with less spatial resolution with respect to other parts of ID detector, however it provides complementary information for electron identification due to good electron-pion separation power.

2.2.3 The calorimeter system

ATLAS calorimeter system [67] consist of several types of sampling detectors, where the interaction of the particles happen in absorbing material and the measurements of energies are evaluated in active medium. They are located outside the ID and solenoid, covering pseudorapidity of $|\eta| < 4.9$. The calorimeters of ATLAS are divided into two parts, electromagnetic calorimeters to measure the electron and photon energies with fine granularity, and hadronic calorimeters contributing to the measurement of the jet energy and missing transverse momentum. Both calorimeters share the similar mechanisms to evaluate the particle energies. While particles passing through the absorbing material, *showers* of primary and secondary particles are created and they enter to a detection medium, where the signals are produced via ionization or scintillation. In both cases, several different active layers are summed up to measure the total signal and the energy deposit by the particles.

Electromagnetic calorimeter (EM) [68] is made of lead and liquid argon (LAr) detectors and splitted into barrel and end-cap regions. Due to the radiation hardness in the region, active material is choosen to be lead absorbers. The barrel region covers $|\eta| < 1.475$ where the region of $1.375 < |\eta| < 3.2$ is complemented by end-cap regions. The *crack* region at $1.375 < |\eta| < 1.52$ is used to cool and instrument the detector readout system, where the energy resolution is degraded. EM has ≈ 170000 readout channels.

Hadronic calorimeter is situated outside of the EM. It consists of three parts: the Tile Calorimeter, the Hadronic End-cap Calorimeter (HEC) and the Forward Calorimeter (FCal). Tile calorimeter, which uses steel absorber and scintillating tiles, is defined thoroughly in Chapter 3. HEC is located right after the EM calorimeter and it uses copper plates for absorbing material, while it shares the LAr active material. FCal is covering the forward region of $3.1 < |\eta| < 4.9$. Due to its location, the particle flow and energies are relatively higher and thus the material is choosen to be more dense. It is made of three parts, one is dedicated to electromagnetic measurements and uses copper as a absorbing material and shares the active material of LAr, the other two are designed to measure the hadronic showers, and tungsten is used for absorbing material for its high density.

2.2.4 The muon spectrometer

The Muon Spectrometer (MS) [69] shapes the outmost of ATLAS detector due to muons' high penetration capability. It covers the range of $|\eta| < 2.7$ and is composed of four different parts. Monitored drift tubes (MDT) and cathode strip chambers (CSC) provide accurate measurements of the muon track coordinates, while resistive plate chambers (RPC) and thin gap chambers (TGC) supply complementary identification information with fast signals. Each part adapt different technologies, while only TGC make use of copper multi-wires and the rest is using gas mixtures for active material.

2.2.5 The trigger and data acquisition system

As mentioned in Section 2.1, proton bunches are crossing in 25 ns interval at LHC. Therefore, at the peak luminosity, the interaction rate reaches $\approx$ 1 GHz, and the required disk space would be 1 Petabyte per second. Since it is not possible to keep this amount of data, ATLAS performs a two stage trigger selection to reduce the bandwidth by storing only interesting features, resulting to the event rate of $\approx$ 1 KHz [70, 71].

Hardware based Level-1 (L1) trigger is the first step of the trigger system. It collects the events with high transverse momentum objects and defining the *Regions-of-Interests* (RoIs) from calorimeters and muon system, where these objects are populated. At this stage, the event rate is reduced to 100 KHz, within $\approx 2.5\mu s$. Afterwards, these events are fed to the next trigger selection step.

High-level-trigger (HLT) is a software based system, composed of fast and sophisticated reconstruction and identification algorithms to run over either RoIs or full regions from inner detector, calorimeters and muons sytems. The reconstruction and identification algorithms are similar to the ones described in Chapter 5, with less precision since it needs to be faster. This level reduces the event rate by two orders of magnetide $\approx$ 1 kHZ with a processing time of 0.2 s. The reconstructed objects in HLT are provided to trigger performance groups to define trigger menus with some threshold selections at the momentum and energy of the reconstructed objects to keep the required event rate [72–75].

The data acquisition (DAQ) system is responsible from monitoring the data to be recorded on the storage disks. If an event triggers the L1 system, it is moved from read-out electronics of each subdetector to the read-out drivers (ROD). Furhtermore, if this event passes the HLT as well, the data of the event is merged and recorded to disk.

CHAPTER 3

The Tile calorimeter

I contributed to TileCal activities in several ways. To be qualified as an ATLAS author, I updated and maintained the Laser calibration algorithm to produce regular monitoring of TileCal channels using calibration data, in particular Laser and HV systems, and prepared an overview of the problematic channels list and documented in an internal note, as described in Section 3.3.3. I was Data Quality Leader (DQL) of TileCal in total during a period of 2 months. Some of the responsibilities were overseeing all operational data quality activities and reporting the results to the TileCal group and the broader ATLAS community, maintaining and updating the channel status list in the Tile conditions database. Futhermore, I was involved in the maintenance activities of TileCal during a 2 months period, and some of the responsibilities were to perform detector hardware repairs in the ATLAS cavern, to study causes and understanding of hardware failures. I analysed the 2018 HV data to monitor the HV related problems in TileCal (Section 3.4). The algorithm is developed by David Calvet.

The TileCal is the hadronic calorimeter systems of the ATLAS detector and supplies an input to the jet energy measurements and to the missing transverse momentum reconstruction, as well as the L1 trigger system. TileCal is a sampling calorimeter composed of scintillating tiles as active material and steel plates as absorber. It is divided into three cylinders: Long-Barrel (LB) located in the ATLAS barrel, covers the region $|\eta| < 1.0$, while the others are in the endcaps, called Extended-Barrel (EB), covers the region $0.8 < |\eta| < 1.7$. Tile cells and electronics are organised into 4 partitions segmented radially, LBA and LBC for the A-side and C-side of the barrel region, and separate EBA and EBC partitions in the extended barrel region. Each has 2280 mm inner and 4230 mm outer radius and the LB part is 5640 mm in length along the z-axis, while EBA and EBC have a length of about 2910 mm. Each of the TileCal cylinders is divided into 64 independent azimuthally oriented modules. These modules lie in the $r - \phi$ plane with 18.2 mm distance, separated by the steel plates and are grouped in 11 tile rows along the z-axis. The scintillation light is collected at the exposed edges of each tile from the two opposite sides by wavelength-shifting (WLS) fibres, arranged in groups defining the readout cells, and connected to a pair of photomultiplier tubes (PMTs) within a module to increase the uniformity of the response and the reliability of the light collection. There are 45 cells in

each LB module and 14 cells in each EB module, read by around 10,000 PMTs (or named as *channel*) in total. The scintillators are grouped in *cells* in three categories: the A-cells being the closest to the beam axis, followed by the BC and D-cells with depth 1.5, 4.1, and 1.8 interaction lenghts at $|\eta| = 0$, respectively. Additional cells are located to cover the gap (E1-E2) and crack regions (E3-E4) at $1.0 < |\eta| < 1.6$, while at higher $|\eta|$ regions, the Minimum Bias Trigger Scintillators (MBTS) are utilised to monitor minimum-bias event rates. The E cells are partially closer to the beam axis with respect other cells and are exposed to high radiation level. The $|\eta|$ and radial structure of the TileCal cells is shown in Figure 3.1.

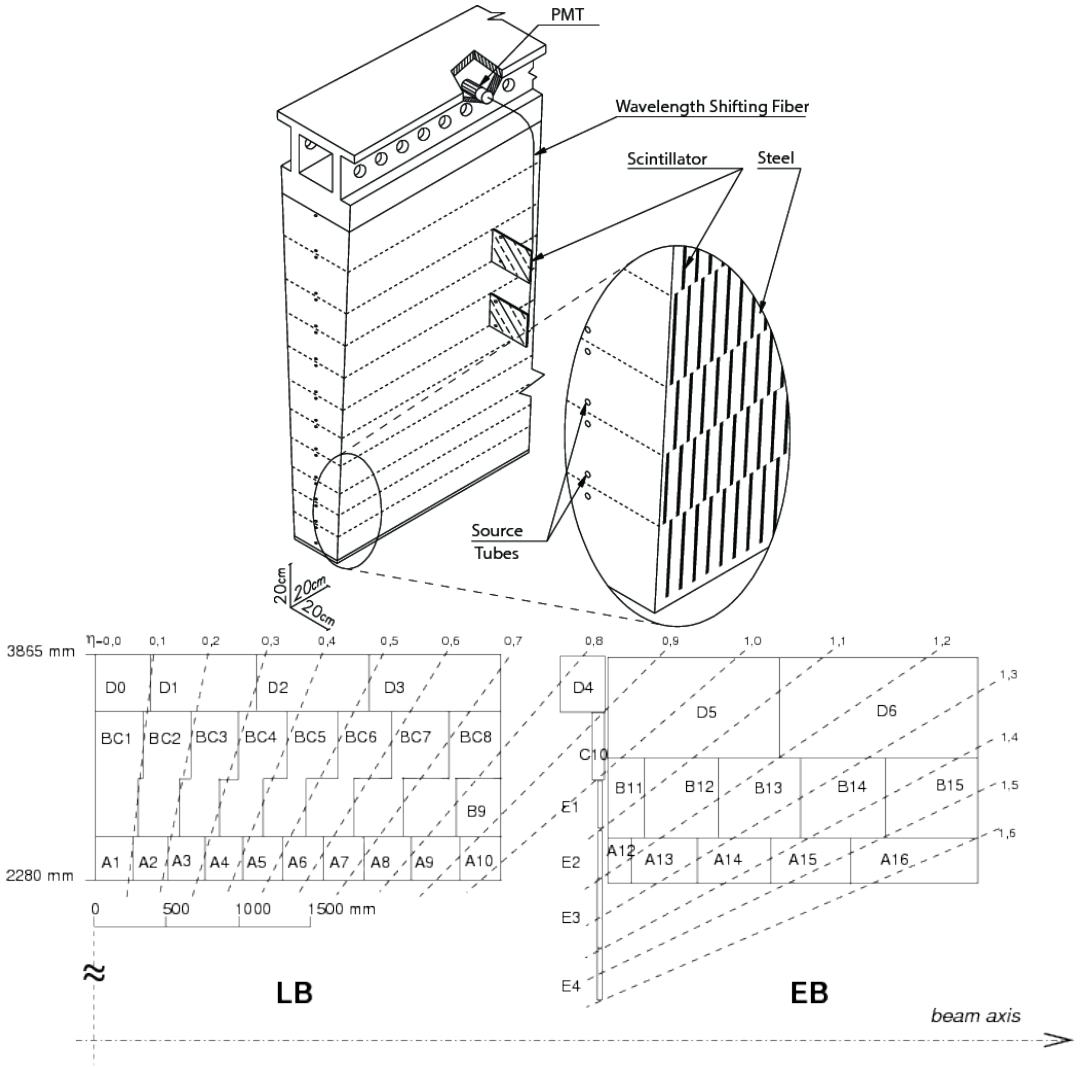


Figure 3.1: Tile Calorimeter module structure showing the mechanical assembly and the optical readout (***upper***). TileCal cell and scintillator design, in total 256 of such modules exist in the detector (***lower***).

3.1 Readout system and signal reconstruction

Light collected by each PMT is a current pulse whose amplitude is proportional to the energy deposited by traversing particles in the associated cell [76]. The output pulse is shaped and amplified by the electronics located on the PMTs. The width of the signal is 50 ns at half-maximum. Later the shaped signal is amplified into two analogue pulses: high and low gain, which are then digitized by two 10-bit ADCs. The high gain signal is taken into account, while in some cases low gain can be also used if one of the the time series for each pulse (samples) saturated in ADC. The resulting pulse shape is then represented by seven samples, which are accepted by first level trigger to be sent to the back-end electronics (RODs).

Amplitude of the signal pulse (Figure 3.2), timing and pedestal for each channel are reconstructed by the Optimal Filtering (OF) technique [77]. Each sample in ADC count referred as S_i is linearly combined to evaluate the pulse amplitude A, the time t and the pedestal P, using the reference pulse shape obtained by test beam studies as follows:

$$A = \sum_{i=1}^{7} a_i \cdot S_i,\ t = \frac{1}{A}\sum_{i=1}^{7} b_i \cdot S_i,\ P = \sum_{i=1}^{7} c_i \cdot S_i\,, \tag{3.1}$$

where the a_i, b_i and c_i are constants determined from reference pulse shape and its optimisation performed in order to minimize the effect of noise. These coefficients are also a function of the phase of the pulse. This allows to calibrate it in a way that the reconstructed peak is at the center of the sample. The value of t indicates the time difference between the peak of expected pulse and the real time of the reconstructed peak, mainly originating from the variations in particle path and read-out electronics uncertainties.

Following the signal reconstruction, series of corrections and calibrations are applied [78]. So called *channel time calibrations* are used to equalise the time for each channel in one drawer by using laser signals in order to avoid time delay effect. Additional corrections on *electronic* and *pile-up* noise are evaluated and applied to correct such affects.

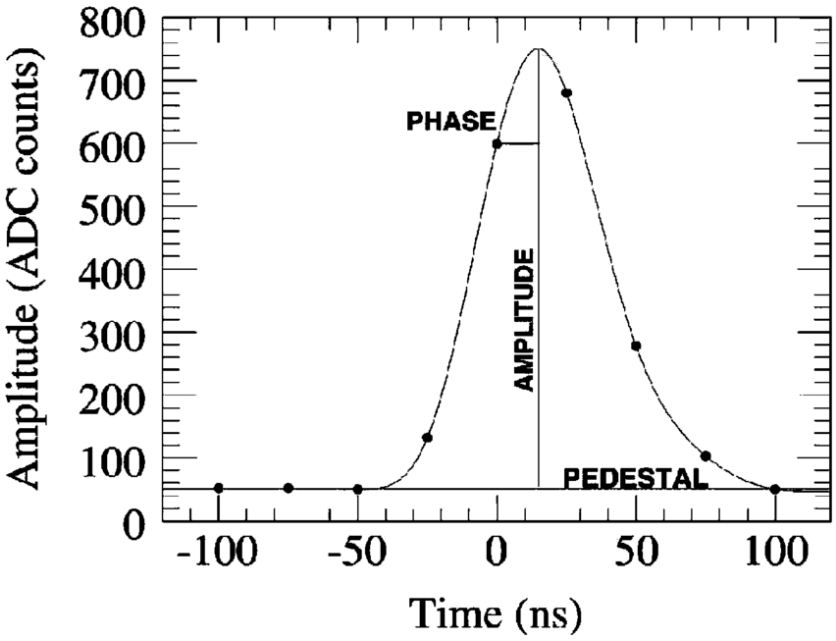


Figure 3.2: Sketch of TileCal pulse shape in ADC counts [79].

3.2 Calibration systems

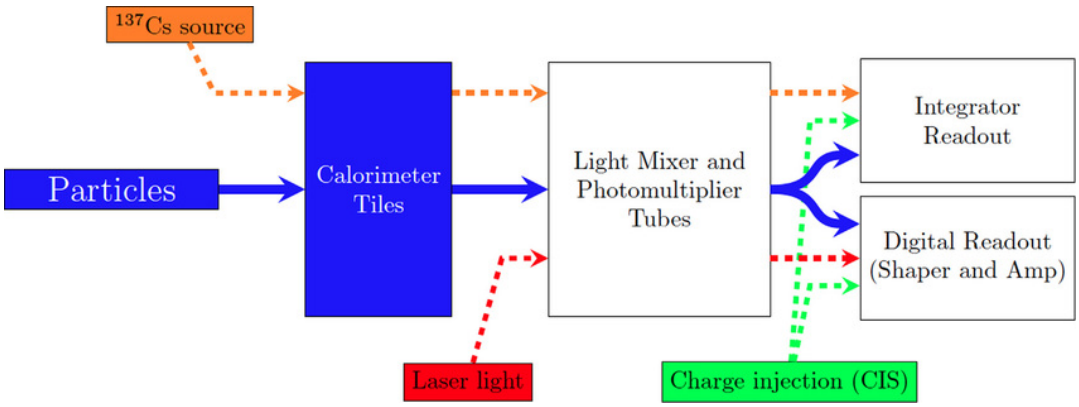


Figure 3.3: Schematic representation of the readout signal path from different TileCal calibration systems.

Several calibration systems are used to monitor the stability of the different stages of the signal processing elements of TileCal. These systems provide per channel calibration constants. Figure 3.3 shows the different calibration systems along with the signal path. The paths of different monitoring and calibration systems are relatively overlapping with each other, implying that some parts of the detector can be monitored by more than one system. This allows for validation of potential problems from different monitoring systems. The signal amplitude of

each channel in ADC counts is extracted using OF as described in Section 3.1 and is converted into an energy by the following formula:

$$E[GeV] = A[ADC].C_{ADC\rightarrow pC}.C_{pC\rightarrow GeV}.C_{Cesium}.C_{Laser}.., \tag{3.2}$$

where $C_{ADC\rightarrow pC}$ is the fixed factor to data taking, from ADC counts to charges in pC during dedicated test beam activities, $C_{pC\rightarrow GeV}$ is the relation between the charges and the energy deposit measured during dedicated test beam campaigns. C_{Cesium} and C_{Laser} are the calibration constants used to maintain the stability of the energy scale, which are then explained briefly in the following paragraphs.

Cesium calibration system

The TileCal uses three radioactive ^{137}Cs sources, which are displaced using a hydraulic system to scan all TileCal cells in order to sustain the global electromagnetic scale response and to monitor the electrical and optical response of each PMT [80]. Each Cs source travels through the calorimeter using a chain of stainless steel tubes embeded into each tile scintillator and delivers 0.665 MeV photons through β decay at rate of 10^6 Hz. The generated scintillating light is converted to the electrical signal and finally the signal is collected through a integrator read-out chain. Therefore the response is a measurement of the integrated current in each PMT. As it is shown in Figure 3.3, the cesium system is used for monitoring the quality of the optics (scintillators and fibres), the PMTs and the integrator readout. Any deviation of TileCal response to Cesium signals from the reference signal can be interpreted as a cell light collection or PMT gain variations and translated into calibration constants, C_{Cesium}. The precision of the system is at the order of 0.3%. Cesium calibration scans were taken in one to three months intervals up to 2015. During 2016 the frequency was reduced, and scans were taken only at the beginning and end of the proton-proton collisions period. The laser system is used between two Cesium scans to maintain the energy scale constant.

Laser calibration system

The Laser calibration system [81] is used to monitor and measure the gain stability of each PMT by sending a controlled amount of laser light to the photocathode and comparing with a reference light. The response of each channel with respect to its nominal value (at the time of the latest Cesium calibration) is translated into a calibration constant: C_{Laser}. The laser calibration runs are usually taken twice a week. During the LHC Long Shutdown from 2012 to 2015, a new Laser II system was developed with new electronics and optical components and light monitoring, which provides an improved resolution. The precision on gain variation is better than 0.5%. The detailed explanation of the laser calibration system and the calculation of the gain variations are given in Section 3.3.1.

Charge injection system

Charge injection system [82] is used to calibrate and monitor the readout electronics. A signal of a known charge is injected to the electronics to evalate the ADC response over its full charge (approximately 0–800 pC in low gain and 0–12 pC in high gain). The conversion factor from ADC counts to pC, $C_{ADC \to pC}$, is calculated with a linear fit to the peak amplitude versus injected charge and complemented by nonlinear correction factors coming from signal processing. The charge injection calibration runs are taken two times a week when there are no collisions. In case the calibration constant of the channel differs by more than 1%, the calibration value in the system is updated. The precision of the calibration system is $\approx 0.7\%$ and the variation of the constants are approximately around 0.05% for individual channels.

Minimum bias integration

LHC proton-proton collisions are dominated by soft parton interactions, which is called *Minimum Bias (MB)* events [83]. The integrated PMT currents over a time window of about 10 μs are continuously recorded during the collisions, allowing to monitor and measure the response of the detector to MB events. The response of the TileCal channels to signals created by MB interactions scales with instantaneous luminosity. Therefore, this provides an additional way to measure the luminosity delivered to ATLAS due to the dependency of the currents to the instantaneous luminosity. The response stability is used to produce calibration constants providing an independent cross-check of the Cesium calibration since both systems measure the signal coming from scintillators and the variation in PMT response over time. In addition, the MB system is used to calibrate E-cells and MBTS cells, which are not calibrated by the Cesium system. The variations observed by the MB and Cesium systems are sensitive to PMT gain drift and scintillator irradiation (ageing). Thus, a difference between MB, or Cesium, and the laser system measurements can be interpreted as a loss of efficiency of the scintillators through irradiation. An extra calibration from the integrator system can be applied to some channels on top of the laser calibration.

3.3 The Laser system

The purpose of the laser calibration system is to monitor the PMTs and the read-out electronics in TileCal. An upgraded version of the laser calibration system (Laser II) of the ATLAS TileCal has been installed in USA15 during Long Shutdown 1. The main concern of the Laser II project was to design a calibration system without Laser I limitations while maintaining a high performance level in terms of precision and stability. The major improvements and the detailed studies are summarized in the Ref. [84].

The components of Laser II system are the optics box, the diodes and the internal calibration system composed by an alpha source and a LED (Figure 3.4). The path of light of laser starts in the optics box and continues to the laser expander, then to the filter wheel and finally to the beam expander and mixer. Following that, the light can be collected in three different positions: before the laser expander, after the filter wheel and after beam expander, allowing to monitor

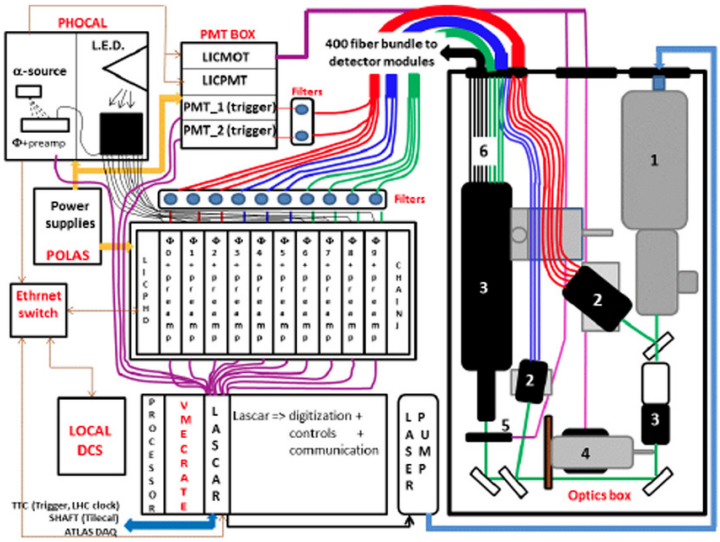


Figure 3.4: Scheme of the Laser II system. Numbers given in parentheses indicate the components of the system. (1) Laser, (2) mixer, (3) expander, (4) filter, (5) shutter, and (6) fibres.

each step of the light. The signal from optics box are then read out by ten photodiodes, which are calibrated using LED pulses. The two PMTs in the PMT box are utilized to trigger the local DAQ with laser pulses. The material for the active mechanical devices, such as the shutter and the filter wheel, are located inside the optics box. Light mixers expand the laser beam allowing to distribute the light over the fibres leading to the PMTs

3.3.1 Determination of the PMT gain variation with laser system

The method to obtain the laser calibration constants C_{Laser} is based on the analysis of certain laser calibration runs. The procedure consists of calculation of constants with the gain variations.

Laser calibration runs are taken with two different setups:

- Low Gain run (LG) with 10000 pulses and the filter attenuation factor equal to 3.
- High Gain run (HG) with 100000 pulses and the filter attenuation factor equal to 330.

For each run and for each TileCal channel, the response of the laser pulse is reconstructed after pedestal subtraction and combined with the diode response. The normalised response $R_{i,p}$ of each channel i to a pulse p is defined as:

$$R_{i,p} = \frac{E^{pmt}_{i,p}}{D1_p}, \tag{3.3}$$

where $D1_p$ is the signal of the photodiode measured after the signal is reflected by the mirror and $E^{pmt}_{i,p}$ is the signal of the TileCal channel i. The mean value of the Equation 3.3 over all pulses is used in the measurement of the constants, $R_i =< R_{i,p} >$.

The calibration constants are computed with respect to a reference laser run taken after the cesium runs. The relative gain variation of the i^{th} channel is given as follows:

$$\Delta_i = \frac{R_i - R_i^{ref}}{R_i^{ref}}, \tag{3.4}$$

where R_i^{ref} is the i^{th} channel mean response for the reference laser run.

However, the inhomogenities after the light splitter causes some variations in the light intensity carried by the fibre. Therefore, Δ_i given in Equation 3.4 is corrected by the term $\Delta^{fibre}_{f(i)}$ (*f(i)* indicates the number of the fibres from the laser, which are connected to the channel *i*). The term $\Delta^{fibre}_{f(i)}$ is evaluated by using an iterative method. Firstly, the distribution of Δ_i (raw relative gain variation of the channel) of all the PMTs fed from the same fibre is evaluated. In this distribution, only the PMTs associated to a set of cells, which are considered as the most stable cells in terms of the PMT gain, are used. These are the D cells for the long barrel (LB) modules and the B13, B14, B15, D5 and D6 cells for the extended barrel (EB) modules and assuming that they have a similar drift[1]. The median and the RMS of the Δ_i distribution is evaluated. For the next iteration, to reject the outlier drifting channels, the channels with gain variation larger than twice the mean RMS are excluded from the calculation. After five iterations, the mean gain variation yields to the correction $\Delta^{fibre}_{f(i)}$. Therefore, the correction factor C^i_{laser} for channel *i* is defined as:

$$C^i_{laser} = \frac{1}{1 + \Delta_i - \Delta^{fibre}_{f(i)}}, \tag{3.5}$$

which is the laser calibration constant that goes in to the Equation 3.2 and the gain variation of each channel is,

$$\left(\frac{\Delta_G}{G}\right)_i = \frac{R_i - R_i^{ref}}{R_i^{ref}} - \Delta^{fibre}_{f(i)}. \tag{3.6}$$

3.3.2 Automated laser monitoring of TileCal channels

The channels having pathological problems have to be promptly identified during the data taking periods of the LHC. Therefore, an automated monitoring of the laser system was setup in order to monitor and to diagnose the issues of the channels. This is achieved by analysing the latest laser runs taken each day. After each processing, several representative illustrations are produced by the software, as well as the list of channels with their possible source of problems. All the TileCal channels including the masked channels after data quality (DQ) checks are

[1] In Run 1, it was assumed that they have a similar drift and fixed to 0, while for Run 2 the value of the drift is floating as no cell is perfectly stable.

analysed and flagged according to the algorithm described in the following paragraph.

The algorithm of the automated laser monitoring is based on the behavior of the gain variation (Equation 3.7) of each channel and its comparison with other type of measurements such as High Voltage (HV) (Section 3.4). In order to study the variations of the behavior of channels in time, the algorithm exploits the laser runs (HG and LG) taken in the 15 preceeding days. Three categories of channels are defined depending on the source of the problem:

- **Normal channels:** Channels that have no deviation or a deviation compatible with the mean deviation of similar cells. These channels can be calibrated safely and do not require a special attention.
- **Suspicious channels:** Channels with a deviation slightly higher (5-10%) than the mean deviation of similar cells or a deviation compatible with an HV variation. These channels can be calibrated safely but some follow up can be needed.
- **Channels to be checked:** Channels with large deviations (>10%) that cannot be explained by the behavior of similar type of cells or by the HV system, and channels having a non-linear behaviour during the 15 preceeding days (jumps or fast drifts). These channels should not be calibrated unless the origin of the effect is understood. In most of the cases, especially in case of a fast drift, the channels need to be masked.

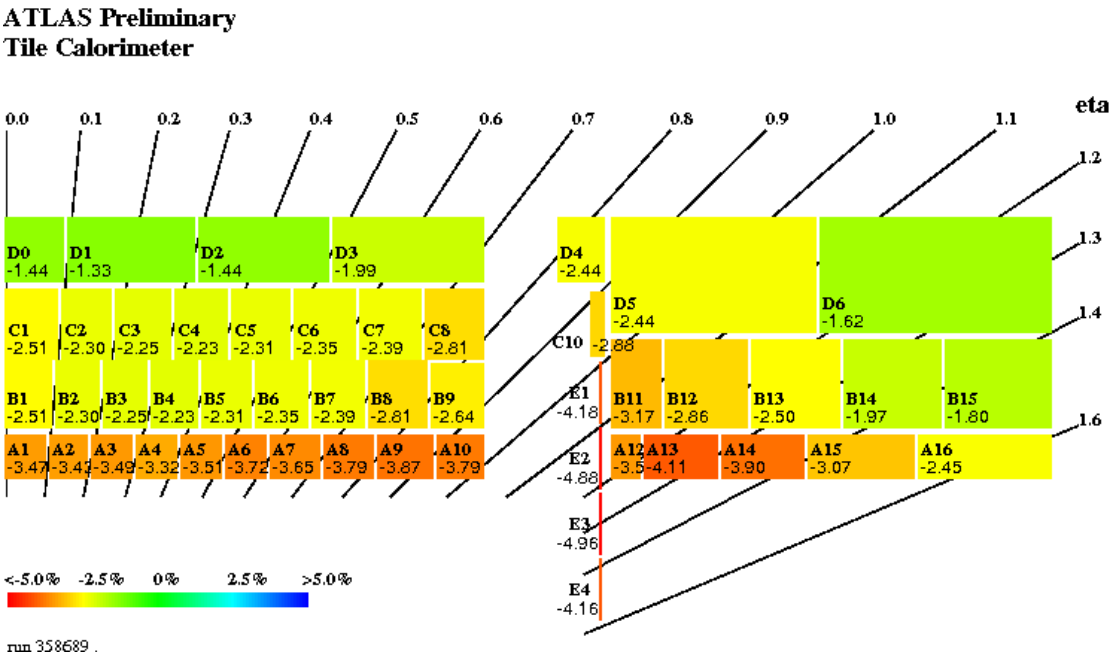


Figure 3.5: The mean gain variation of the cells (in %) in the ATLAS TileCal PMTs as a function of η and radius for the laser run number 358689 taken in 22 August, 2018.

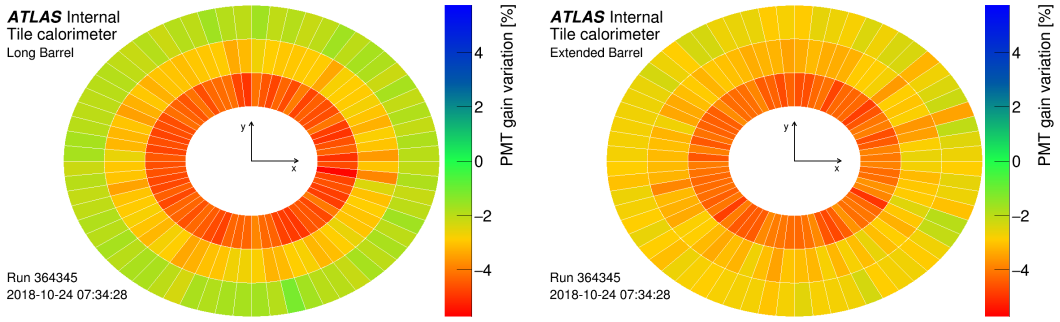


Figure 3.6: The mean gain variation (in %) in the ATLAS TileCal PMTs, as a function of ϕ bins and radius long barrel (***left***) and extended barrel (***right***) for the laser run number 358689 taken in 22 August, 2018.

Figures 3.5 and 3.6 show the illustrations made by the laser monitoring system, which are produced in the same day with the calibration run. Figure 3.5 is a map of the TileCal cells showing the mean gain variation (in %) as a function of η and radius. Figure 3.6 shows the mean gain variation (in %) of the channels of the TileCal as a function of the polar angle (ϕ) and the layer. One ϕ bin corresponds to one module per partition, averaged for two sides.

Furthermore, the history plots (gain variation as a function of time evaluated by HV and laser system) are produced for the suspicious channels and the channels to be checked for 15 preceding days (Figure 3.7). These plots show the comparison of the gain variations calculated from both the analysed LG run (blue points) and the HG run (green points). In order to understand whether the gain variation is stable or not in HV, the gain variation predicted by the HV system is included. The HV data is extracted from DCS Data Viewer (DDV) during the laser data taking time interval and gain variation between the time t and a reference time t_r is calculated from the formula given below:

$$\left(\frac{\Delta_G}{G}\right)_i = \frac{HV_{out}^{\beta}(t)}{HV_{out}^{\beta}(t_r)} \,. \tag{3.7}$$

where the parameter β is specific to each PMT and its value is ≈ 7.

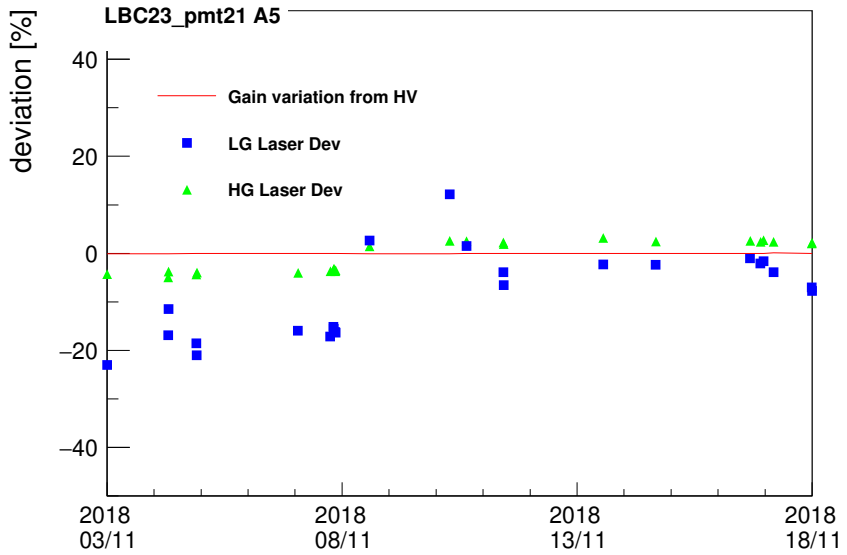


Figure 3.7: An example of history plot produced by automated laser monitoring (LBC23 pmt21).

3.3.3 Full year monitoring of TileCal channels

As described in the Section 3.3.2, automated monitoring of the TileCal channels is useful to diagnose the channels having issues in a short interval of time. However, in order to spot the channels with pathological behaviours during all the data taking period, the full year analysis is performed for each 2017 and 2018 years. The study is based on the automated monitoring algorithm. The laser runs (HG and LG) are chosen with approximately 10 days intervals for the 2017 and 2018 data taking periods individually. For each chosen date, both HG and LG run are

analysed and all the channels are flagged according to their problems. The categorisation of the channels are done in the following way:

- **Bad channels:** If the gain deviation of the channel is large (>10%) or having a bad behaviour such as fluctuations during the 15 preceding days,
- **HV unstable:** If the gain deviation of the channel is compatible with HV and greater than 10%,
- **No laser data:** If the amplitude of the channel signal is small (fibre problem or low signal),
- **Bad laser data:** If the data during the run is corrupted or having problematic reference.

Then a channel is reported if it shows a problem in any gain. An illustration shows the overall number of channels in each category is produced for each year, as given in Figure 3.8.

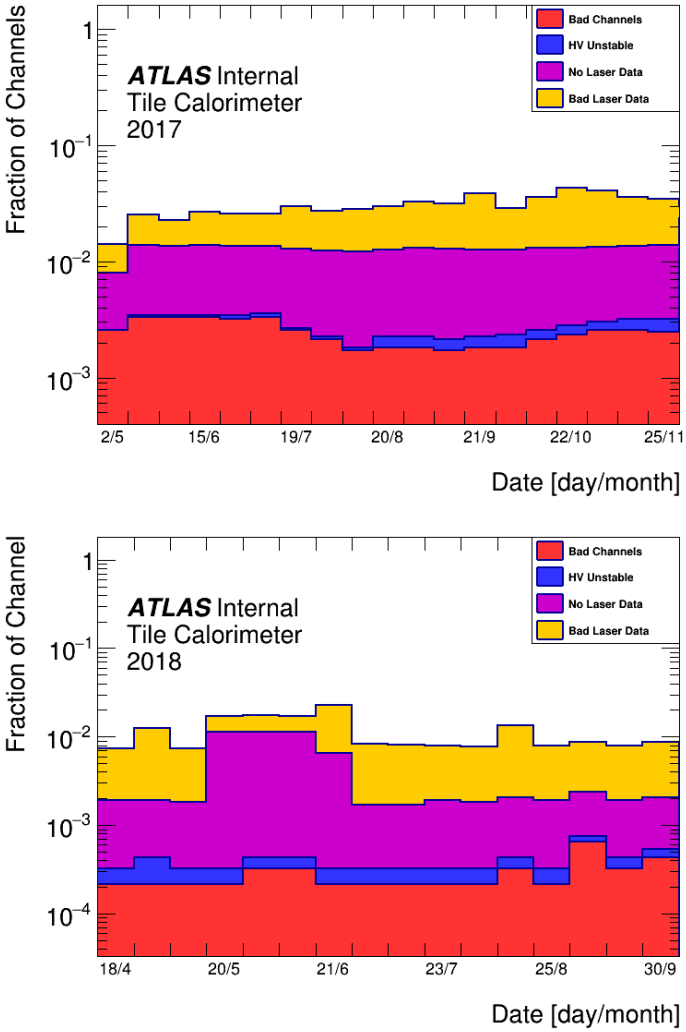


Figure 3.8: The overall fraction of the problematic channels in 2017 (***upper***) and in 2018 (***lower***).

The number of problematic channels in 2017 is greater than in 2018. The reason of this behavior can be explained by the various problems that occured in the corresponding years such as wrong settings of HV or the fact that the response recovers slowly when the beam is off. History plots are produced for the *bad channels* and *HV unstable* categories which show the gain variation of this particular channel during the all period. Figure 3.9 illustrates an example of history plot, which demonstrates succesfully the problem in HG for EBA15 PMT09. One possible reason for this behaviour is that the 3-in-1 card situated on each PMT is broken, which is the interface with the electronics read-out. The result of this study summarised in the internal note and the problematic channels are provided to the maintenance group of TileCal.

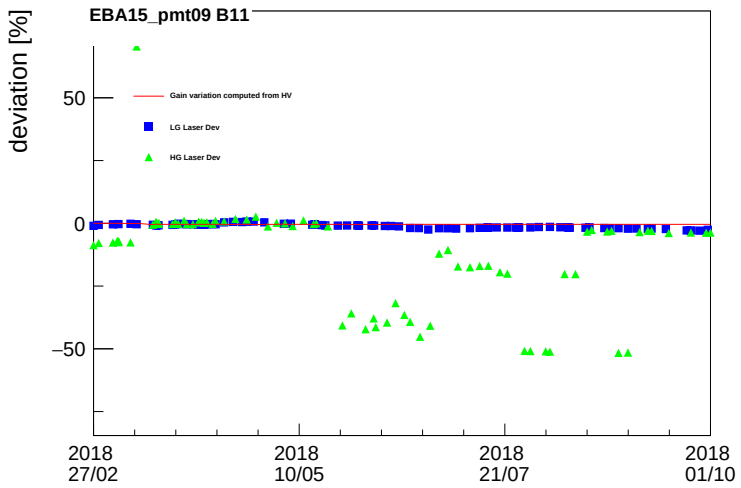


Figure 3.9: An example of history plot produced for all data taking period in 2018. The green (blue) data point corresponds to one HG (LG) run and the red line shows the gain variation evaluated from HV system.

3.4 High Voltage distribution system

The High Voltage distribution system [85] is designed to control and monitor the high voltages (HV) applied to the PMTs of the TileCal drawers. Gains of these PMTs are fixed to a specific value through the stable applied HV (within 0.5 V) to maintain reliable energy measurements. The behaviour of HV system is being studied every year to monitor the values and spot possible HV problems in PMTs. In this section, the analysis of 2018 year data taking period is given. Previous year studies as well as the 2018 study is documented in Ref. [86–88].

In total there are 256 HV power supplies one for each super-drawer[2] located in USA15. In one super-drawer, there is one HVmicro board and two HVopto boards together with two HVbus boards. The micro-controller is located on each HVmicro board that controls the 48 output HV channels. The HV distribution system for a given drawer is given in Figure 3.10 and these boards are provided with +5 V, +15 V and -15 V through an external power supply. For each channel, there is a regulation loop and a monitoring part as it is shown in Figure 3.11.

The HV_{in}value is applied for each super-drawer in order to supply only one output to each PMT and can take two values: -830 V and -950 V. Later each super-drawer can adjust individually the applied high voltages for each PMT, which is done by HVmicro and HVopto cards. Finally the regulation loop allows to deliver the high voltage to the PMT lying in the range:

$$HV_{in} - 360\,V < HV^{i}_{out} < HV_{in} - 1\,V\,. \tag{3.8}$$

The value of 360 V is indicative and could fluctuate by 20 V around this value.

[2] A super-drawer corresponds to the pair of two drawers and contains 48 PMT holes in total.

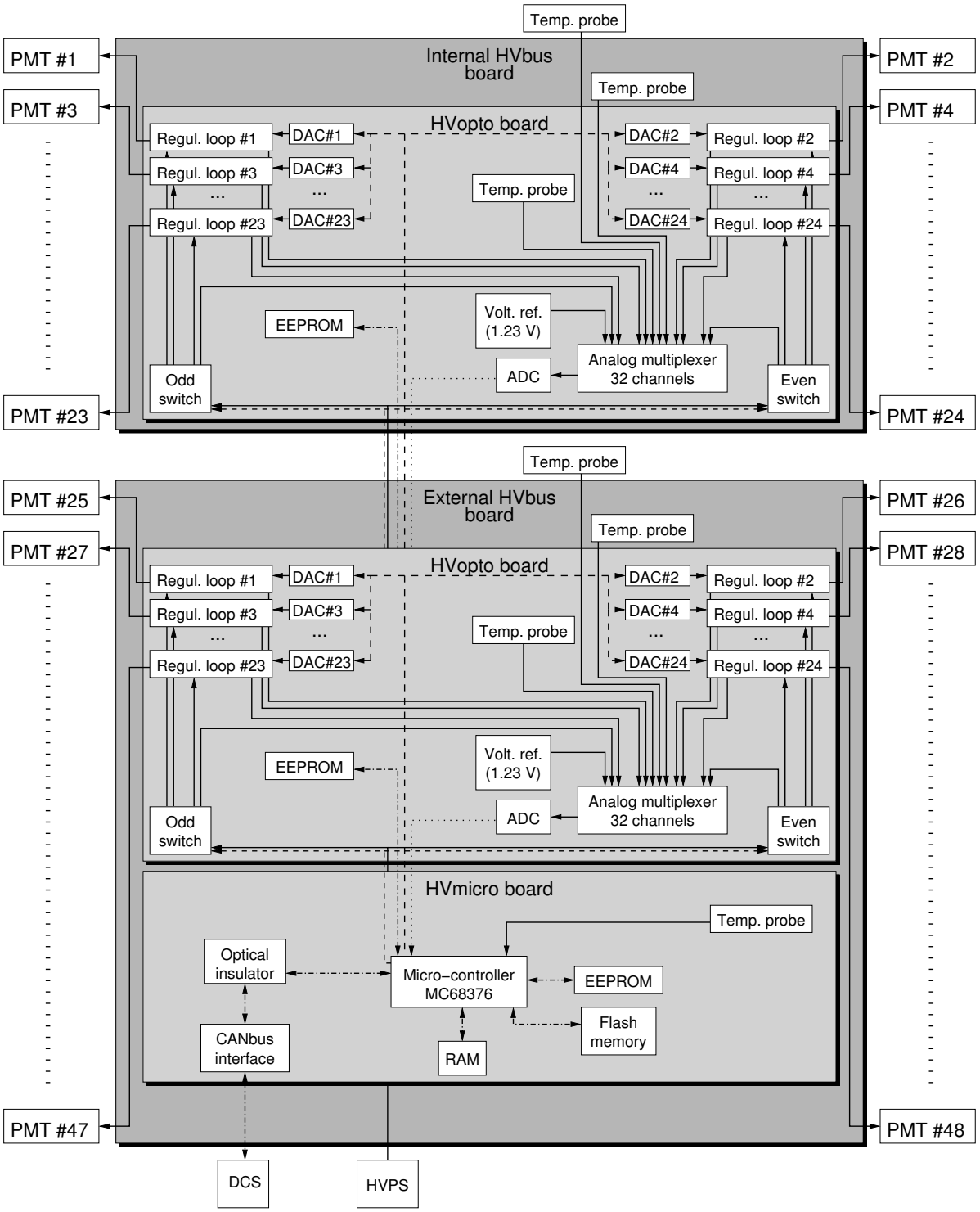


Figure 3.10: Illusturation of the HV distribution system.

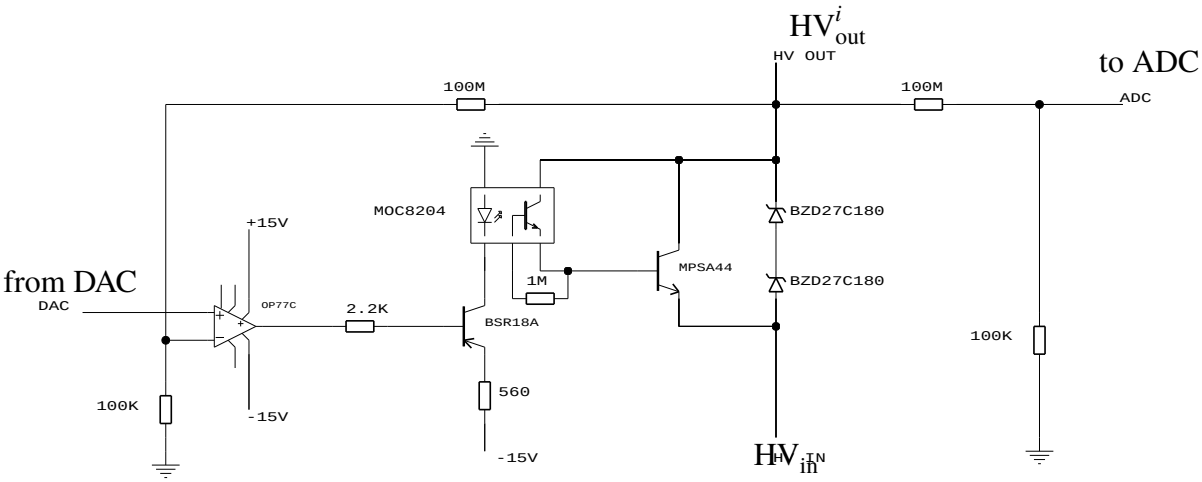


Figure 3.11: Scheme of the HVopto regulation loop: the HV_{in}, coming from the high voltage power supply (HVPS), is modified by the opto-coupler under the control of the DAC value, in order to have HV^i_{out} value (applied to the PMT) equal to the required (HV^i_{req}) one. This output value is monitored by the ADC on the right hand side.

3.4.1 Algorithm

This section describes the algorithm used to analyse the HV behaviour, based on the data extracted from DCS Data Viewer (DDV) [89]. The main goal of this analysis is to study the stability of the HV distribution system, in particular the high voltage that is provided to each PMT (HV^i_{out}), by studying the measured high voltage in each regulation loop (HV^i_{meas}) and to compare it with the required voltage (HV^i_{req}). In order to have stable analysis, the external conditions must be eliminated such as change in the setting of HV of MBTS channels. Therefore, the periods of stable conditions (PSC) are defined according to different parameters explained in the following paragraphs. First step is to categorise the super-drawers depending on the power supply status of low-voltage (LV),for the control and regulation, and HV.

- on: when both LV and HV supplies are on;
- no-HV: when the LV supply is on but not the HV;
- reduced-HV: when the HV supply is on but not the LV[3];
- off: when both LV and HV supplies are off.

If the super-drawer stayed more than 5% of the time in one of the faulty states described above (off, no-HV or reduced-HV), it is flagged with the corresponding status. Following this, the PSC is defined where the super-drawers are in *on* stage.

[3] Emergency mode, in this mode the regulation is no longer active and the voltage supplied to the PMTs is of the order of HV_{in} + 360 V, depending on the setting of the HV power supply.

Secondly, in order to better control the PMTs, and switch off only the related part when it is needed, the HV_{in} value is measured in each quarter of the super-drawer (two times from same HVopto). Therefore the algorithm evaluates the states of HV_{in} only for the defined PSC for each quarter as follows since the HV power supply can deliver either -830V or -950V:

- HV0: if the measured HV_{in} is above -5 V;
- HV830: if the measured value is in the range $[-835, -825]$ V;
- HV950: if the measured value is in the range $[-955, -945]$ V;
- HVERR: if the measured HV_{in} does not satisfy the previous conditions.

Similarly in the case of the power supplies, if the quarter-drawer stays in same state for more than 5% of the time in one of the two faulty states (HV0 or HVERR), the quarter is flagged with the corresponding status. Furthermore, the PSC are split if the second HV_{in} of the same HVopto is different.

As a final step, the study of the required high voltage value for each PMTs (HV^i_{req}) is performed. Almost all the channels have the same value of HV^i_{req} in all the periods however the MBTS PMTs are modified several times during the year and therefore the PSC are defined for the unchanged periods.

3.4.2 High Voltage stability

Following the final definition of PCS per channel, the study of the HV behaviour can be performed. Access to HV^i_{out} value is done by the HV^i_{meas} value, which is stored in DDV. For each channel, the $\Delta HV^i = HV^i_{meas} - HV^i_{req}$ is calculated for each HV^i_{meas} value. According to that value, channels are categorised as follows:

- if $|\Delta HV^i| \leq 0.5\ V$ for all the measurements, the channel is stable for the corresponding PSC;
- if $\sigma^i_{\Delta HV} > 0.5\ V$, the channel is unstable with dispersion flag for the corresponding PSC;
- if $\mu^i_{\Delta HV} > 0.5\ V$, the channel is unstable with outliers flag for the corresponding PSC;
- if the channel is stable but with $|\mu^i_{\Delta HV}| > 5\ V$, the channel has a large offset for the corresponding PSC.

Finally, since some channels may have more than one PSC, the channel is considered unstable if there is at least one PSC which is unstable.

3.4.3 Results

The analysis of HV behaviour of 9846 channels in TileCal is performed for the period of 11 April - 18 December 2018 with the algorithm explained above. In total, 25 channels are flagged as unstable, 2 channels with an offset and 6 channels with bad setting. The summary of the behavior of channels is given in Figure 3.12 and Table 3.1. Distributions of $\mu^i_{\Delta\mathrm{HV}}$ and $\sigma^i_{\Delta\mathrm{HV}}$ (the standard deviation of the distribution $\mu^i_{\Delta\mathrm{HV}}$) are given in Figure 3.13.

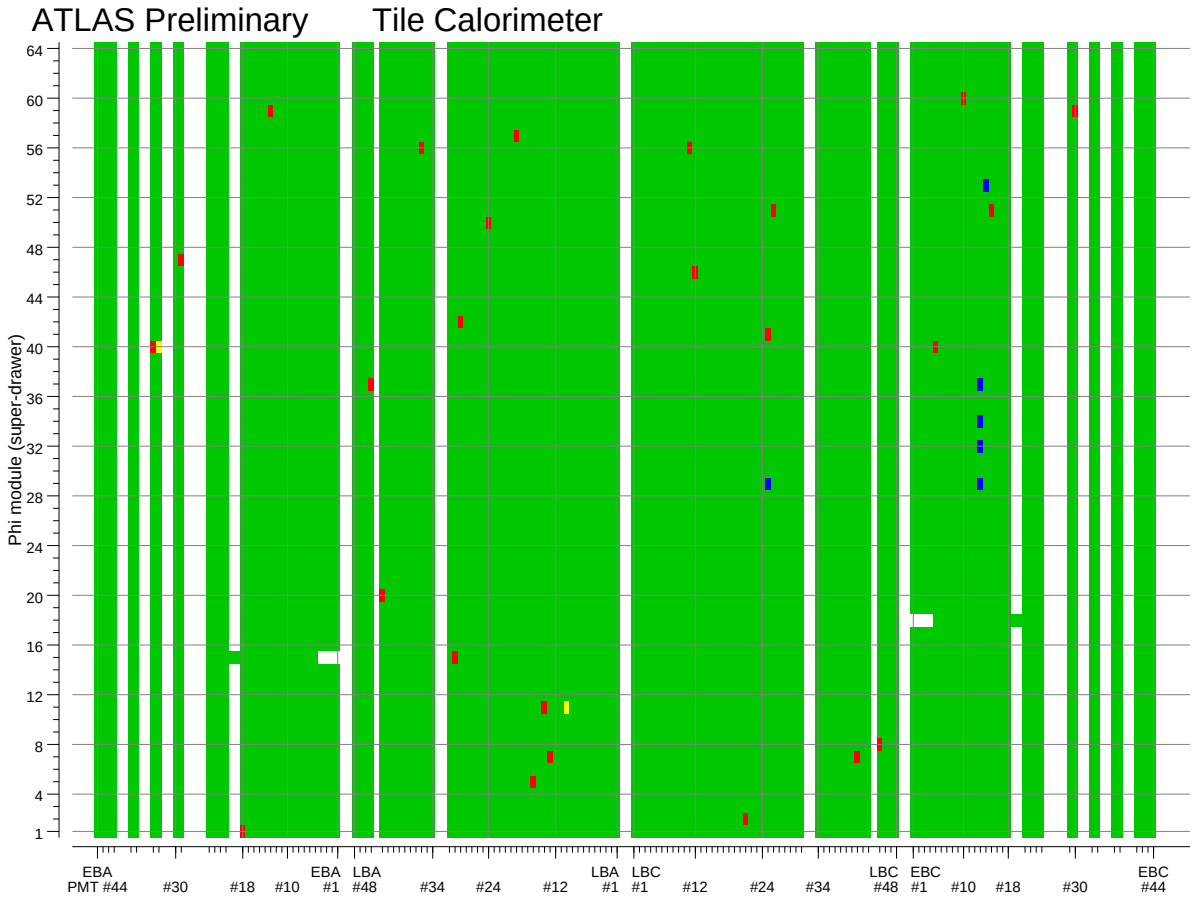


Figure 3.12: Summary of the behaviour of the TileCal channels in 2018. Green channels are stable, yellow channels large offset, red channels are unstable. Black channels are off and blue channels have a bad setting.

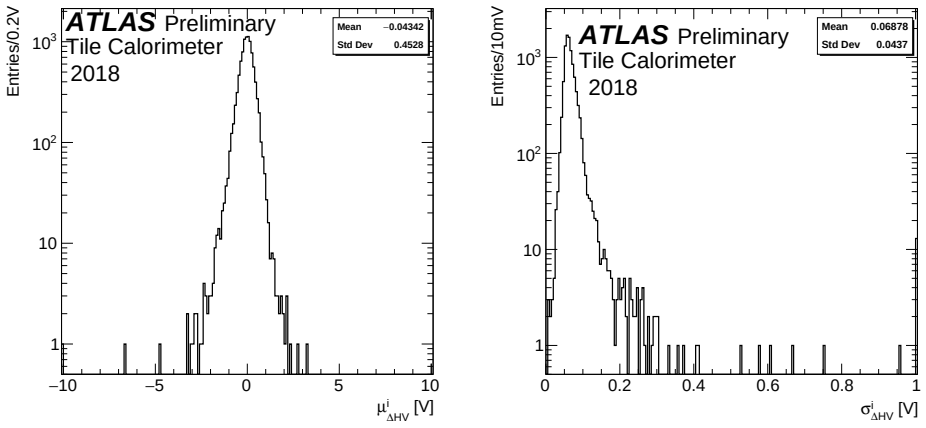


Figure 3.13: Distribution of $\mu^i_{\Delta HV}$ (***left***) and of $\sigma^i_{\Delta HV}$ (***right***) for all stable channels in 2018. The left-most and right-most bins are the under- and over-flow bins.

Table 3.1: List of faulty channels in 2018.

Module	PMT	Analysis	Comment
EBA40	33	Offset	$\mu^i_{\Delta\mathrm{HV}} = -10.2$ V
LBA11	10	Offset	$\mu^i_{\Delta\mathrm{HV}} = -6.7$ V
EBA01	18	Unstable	due to sudden jumps
EBA40	34	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.30$ V for $\mathrm{HV}^i_{\mathrm{req}} - \mathrm{HV}_{\mathrm{in}} = 294$ V
EBA47	29	Unstable	due to sudden jumps
EBA59	13	Unstable	outliers for $\mathrm{HV}_{\mathrm{in}} = 950$ V, OK for $\mathrm{HV}_{\mathrm{in}} = 830$ V
LBA05	16	Unstable	due to sudden jumps
LBA07	13	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.33$ V
LBA11	14	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.36$ V
LBA15	30	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.41$ V
LBA20	43	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 3.67$ V
LBA37	45	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.25$ V
LBA42	29	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.29$ V
LBA50	24	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.96$ V
LBA56	36	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.19$ V
LBA57	19	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.58$ V
LBC02	21	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.67$ V
LBC07	41	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.75$ V
LBC08	45	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.14$ V
LBC41	25	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 2.13$ V
LBC46	12	Unstable	due to sudden jumps
LBC51	26	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.58$ V
LBC56	11	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.28$ V
EBC40	5	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 1.25$ V for $\mathrm{HV}^i_{\mathrm{req}} - \mathrm{HV}_{\mathrm{in}} = 330$ V
EBC51	15	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.61$ V
EBC59	30	Unstable	outliers for $\mathrm{HV}_{\mathrm{in}} = -830$ V, $\sigma^i_{\Delta\mathrm{HV}} = 0.53$ V for $\mathrm{HV}_{\mathrm{in}} = -950$ V
EBC60	10	Unstable	$\sigma^i_{\Delta\mathrm{HV}} = 0.42$ V
LBC29	25	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.7$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.6$ V
EBC29	13	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.4$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.0$ V
EBC32	13	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.4$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.0$ V
EBC34	13	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.4$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.0$ V
EBC37	13	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.0$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.0$ V
EBC53	14	Bad setting	$\mathrm{HV}_{\mathrm{in}} = -829.4$ V, $\mathrm{HV}^i_{\mathrm{req}} = -830.0$ V

CHAPTER 4

Event simulation and physics modellings

In order to make sensible deductions from the observed data recorded by the detectors, an accurate simulation, based on the theoretical and detector modellings, is required. The MONTE CARLO (MC) method, based on the random number generation, is used to generate large amount of random particle interaction experiments at LHC. The simulation process consists of estimating the cross section of the given process by calculating the amplitudes of Feynman diagrams with extra radiation procedures (fragmentation and parton shower), and simulating the detector response to the emerged particles.

Colliding protons are interacting predominantly with strong interactions at LHC experiments as a consequence of having color charge. Therefore, the precise measurement of the underlying physics processes driven by QCD is necessary and special emphasis will be given in the consecutive sections to describe the simulations of pp collisions. Physics processes driven primarily by EWK are better understood and hence the choices of the parameters have less impact on the physics analyses.

Main pp interactions occur at the very high energy scales, since the incoming partons have energy levels around TeV, which is $O(10^3)$ bigger than their rest mass. The running of the coupling constant (α_s) yields to different behaviour of QCD at low and high energy scales. The strength of the coupling is decreasing with increased energies, thus the terms of higher orders become less important (asymptotic freedom or confinement). First step of event generation is called *hard process*, where the matrix element (ME) calculation is performed (Section 4.1), and can be calculated with sufficient precision perturbatively by including fixed order α_s or mixed QCD-EWK calculations at higher orders, typically the next-to-leading (NLO) or next-to-next-to-leading order (NNLO) with modern MC event generators. The soft hadronic processes such as hadronisation and the evaluation of the underlying events at larger distance scales (lower energies) are non-perturbative and their computation process is based on the quantitative QCD-like algorithms. This is denoted with *parton shower (PS)* and *hadronisation* (Section 4.2). This factorised steps are illustrated in Figure 4.1, and described over the successive sections, together with discussion of detector simulation in Section 4.5. Finally, MC generators are compared in Section 4.4.1 that are more relevant to this work.

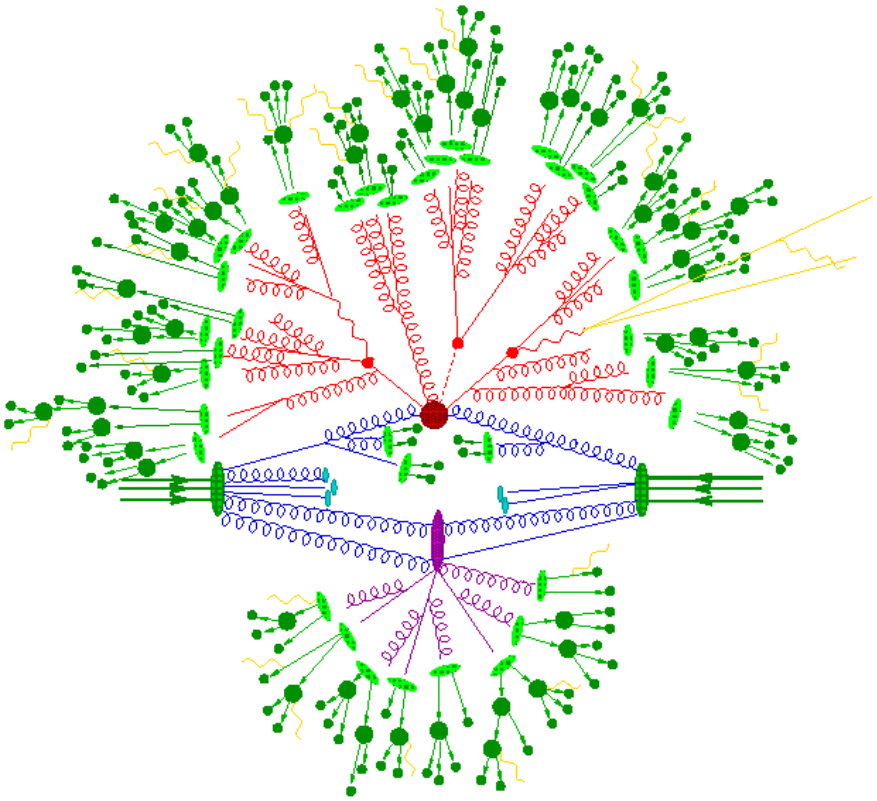


Figure 4.1: Outline of a *pp* collision at LHC, showing the hard scattering with dark red, particle decays and final state radiation with ligth red, initial state radtion with blue, underlying events with purple and photon radiation with yellow [90].

4.1 Hard process

The ME generation is the primary step of the physics event simulation. It is also called as *the hard process* or *primary interaction*. In order to compute the cross section of the interaction and kinematic observables of the partons, perturbative calculations are performed at this stage. The arbitrary scales, depending on the choice of scale Q^2 are defined as the reference scales enabling the perturbative QCD computations to proceed on which physical observables should not depend. Therefore, the *renormalisation scale* μ_R represents the momentum scale, where the particles couple to each other. In theory, the choice of scale is arbitrary, however the scale is typically set to the hard scattering energy scale Q^2 to evade large computations might result in breking the perturbation calculations.

Cross section calculation is evaluated by considering all Feynman diagrams up to a given fixed order of α_s in the hard process and the more complicated diagrams are ignored as the approximation of α_s values. Usually, the events are generated at NLO precision, LO precision can be used for some complex processes and some rare procceses at NNLO precision. The

NLO calculation is determined by taking into account all LO diagrams with the addition of loop corrections of gluons and quarks, such as emission of a gluon or gluon splitting into a quark-antiquark pair. Finally, in order to complete the cross section calculation, one needs to intagrate over the full phase space by defining the differential cross section convolution. While evaluating the integration, the four momentum of all final state particles are obtained, which are then used as an input to the next process of MC production.

Another feature of the ME calculation is the parton distribution function (PDF), which describes the momentum distribution of the quarks and gluons in the protons and can be interpreted as a probability of a parton to be inside a proton with a given fraction of the total momentum. Since the PDFs illustrate the non-perturbative conditions of the partons initiating the hard-scatter process, they are determined experimentally generally through *deep inelastic scattering* [91, 92] experiments at specific scales and then extrapolated to the energy scale relevant to the calculated physics process. In Figure 4.2 the momentum density as a function of partons is shown at different energy scales with NLO precision. The scale dependence of PDF is governed by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation, which evaluates the PDFs from one scale to another, and this equation depends on the *factorisation scale* μ_F, which defines the cut-off scale. Below this scale, the process is absorbed by the PDFs and above this, phenomena contribute to the hard-scatter calculation.

Therefore, the cross section can be obtained for the incoming partons *i,j* as follows

$$\begin{aligned}\sigma_{tot} &= \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 f_i(x_1) f_j(x_2) \hat{\sigma}_{ij}(x_1 x_2 S) \rightarrow \sigma_{tot}(\mu_F, \mu_R) \\ &= \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \hat{\sigma}_{ij}(x_1 x_2 S, \mu_R),\end{aligned} \tag{4.1}$$

where the $x_{i,j}$ are the momentum fractions of partons and $f_i(x_i, \mu_F)$ are the PDFs. The energy of the scattering process is $s = x_1 x_2 S$, where LHC proton energy $\sqrt{s} = 13$ TeV.

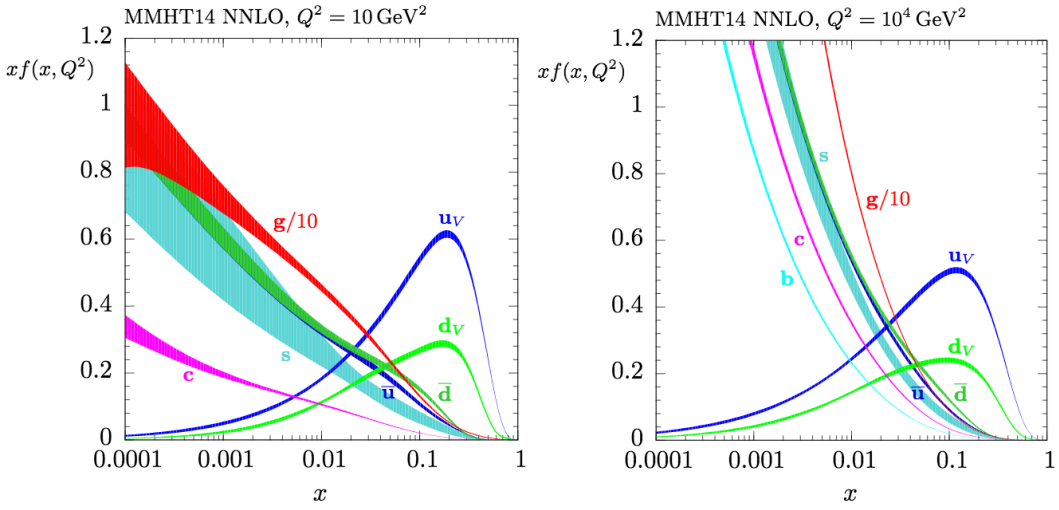


Figure 4.2: PDFs computed at energy scales $Q^2 = 10\,\mathrm{GeV}^2$ (***left***) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (***right***). The uncertainty bands demonstrate the 68% confidence-level [93].

Finally, the computed hard process must be matched to the parton shower. At this point, choosing the right scale choice is essential since the decision will be propagated to the next levels of the simulation and this freedom will lead to systematics uncertainties in the measurements.

4.2 Parton shower and hadronisation

Even though the protons transfer a significant amont of their momentum to the initial partons and consecutively partons transfer their momentum to the final state partons, highly energetic coloured particles can radiate gluons. Therefore, these soft and collinear emissions initiate strong interactions and create initial (final) state radiations. The final-state radiation (FSR) of PS simulation performed by a forward estimation of a form factor describes the probability that a parton yields to smaller scales, from energy scale of Q^2 until $Q^2 \approx 1\,GeV^2$. Adding these partons is performed iteratively by scaling the probabilities and by splitting into additional partons as a function of the energies. This procedure continues until all momentum transfers are below the cut-off scale in which the hadronization process begins. In case of the initial-state radiation (ISR) simulation, a backward estimation is operated, at which the initiated partons radiate.

At this point, the partons coming from fragmentation of PDFs are getting energy to be included in the hard-scatter. The overlap of identical final states may occur between evalution of PS and ME. This overlap needs to be removed when matching the ME and PS. One of the most common approaches are the CKKW method [94–96], which performs an event reweighting to veto the showers, and the MLM method [97] relies on the event rejection.

Following the evolution of partons in PS, they enter the hadronisation scale where they are hadronised into colour neutral particles due to colour confinement of QCD. The main models used at LHC simulations are the Lund string model [98] and the cluster model [99]. The decay of unstable particles which are produced as a result of hadronisation are also considered at this point, supported by external tools such as EvtGen [100], which is particulary used in heavy hadron decays, b- and c-quarks.

4.3 Underlying event

The underlying event (UE) describes the additional interactions in an event, which is not directly associated to the initiated hard process. Usually these scattering of remaining partons are relatively soft and can interfere with the hard process. These additional occurrences in the hard process is also known as *multiple parton interactions* (MPI). These processes yield to long distance QCD color effects and as a result can lead to higher multiplicity of jets coming from the additional radiation of quarks and gluons. This effect can be simulated with tunable models provided by some of the event generators.

4.4 Monte Carlo event generators

There are several MC event generators dedicated to collider physics at LHC. Although the procedures are similar among different generators, the numerical implementation can differ, and since the models are based on some arbitrary parameter choices which can not be evaluated from perturbation theory or first principles, their validity degree vary. Typically, this implies that the modelling of the observables must be evaluated and compared by different generators, and finally the resulting difference must be added as a systematic uncertainty in the physics analysis. In the following sections, the relevant generators with this study and their differences are presented.

4.4.1 Matrix Element generators

Matrix element MC event generators supply computations of the hard process corresponding to the given pp process. Usually the PS calculation is performed by a separate generator, which is then interfaced to the ME generator.

MadGraph5_aMC@NLO [101] begins from automatically generated code. It computes the amplitude of a requested process during the operation. It permits the hard process calculations at LO or NLO accuracy. The PS algorithm must be provided by generators such as Pythia.

POWHEG-BOX [102] provides NLO calculations using the POWHEG formalism [103]. The differences with MADGRAPH5_aMC@NLO is in the numerical calculation and the PS matching and merging algorithm.

4.4.2 General purpose event generators

General purpose event generators are performing ME calculations together with the PS, hadronisation and UE modellings.

PYTHIA [104, 105] can simulate all the event processes, starting from the hard process at LO, performing parton shower, hadronisation and underlying event. PYTHIA is often incorporated to a ME generator and used to compute the PS or UE. In PYTHIA the hadronisation method relies on the Lund string model.

Herwig++[106, 107] is focusing on the accurate simulation of QCD PS at LO and uses cluster model for hadronisation process.

SHERPA [90, 108] permits to obtain both LO and NLO accuracy and includes a comprehensive PS generator relies on the Catani-Seymour dipole formalism [109]. The hadronisation method is based on the cluster model.

4.5 Simulation of the detector response and pile-up

The simulation of events described above, from the matrix element to the final state objects, results into four-vectors of final state particles after the hadronisation. Until now, so called the *truth level* or *particle level* generation of the events are discussed, and this is useful for studying physics processes without detector effect folded in. However, in order to make real predictions with experimental data, the interactions of the particles with detector and the response of the detector have to be included.

To acheive this, ATLAS uses the GEANT4 software package [110]. This includes the simulation of the particles passing through the various parts of the detector and signals that are created by the interaction of particles with detector material. Different custom algorithms are employed for each ATLAS subdetector to convert the energy depositions into signal while modelling the detailed response of each subdetector to incident particles. The simulated signals are digitised by the electronic readout systems. The reconstruction algorithms convert these signals into meaningful high-level physics objects, where the detailed information given in Chapter 5.

To reproduce the real detector response, the pile-up effect must be taken into account. In MC simulation, the pile-up is modelled by generating a large sample of minimum bias events using PYTHIA event generator with the dedicated PDF set and the shower tune [111]. Minimum bias events are inelastic collisions without any specific selections, hence having high occupancy of

soft particles. Following the fully simulated detector response and the object reconstruction, obtained pile-up profile is corrected by reweighting the event.

CHAPTER 5

Object reconstruction

Following the *pp* collisions, stable particles travel across to the ATLAS detector, resulting an interaction with the layers of the detectors, which are eventually shaped in electrical signals. In order to convert these signals into meaningful and well-defined representations of the physics objects, several processes must take place. As described in Section 2.2.5 some low level reconstruction procedures are applied at the online trigger level to make an acceptance decision. However, whole event reconstruction is done offline with more sophisticated detector models and with higher accuracy. This chapter summarises the algorithms applied to the low level detector signals in order to obtain final reconstructed objects, which are then used in physics analyses. Section 5.1 describes the process of reconstruction of charged particle tracks in the ID and their association to the primary vertex (PV). The lepton reconstruction is discussed in Section 5.2 focusing mainly on electrons and muons. Section 5.3 gives an overview of reconstruction of jet objects and the identification of heavy flavour jets. The reconstruction of missing transverse energy is discussed in Section 5.4, which relies on kinematic properties of other reconstructed particles such as leptons and jets.

A illustrative view of the fundamental particle signatures in the ATLAS detector is given in Figure 5.1.

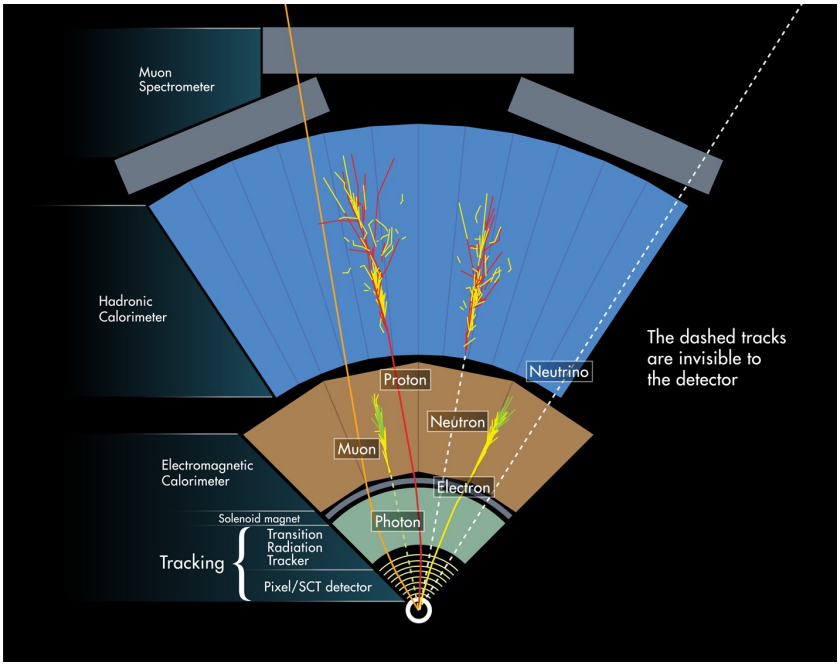


Figure 5.1: Illustrative representation of elementary particles signatures interacting with the ATLAS detector, allowing particle reconstruction and identification to be performed [112].

Electrical charged particles leave signatures in the ID, therefore this information is used in the process of reconstruction. The muons are able to pass through the whole detector, and their identification process uses the information from both calorimeters and the MS, while the electrons result a electromagnetic shower in EM calorimeter. Photons have similar signature as electrons, however they do not interact with ID since they are neutral particles. Protons and neutrons yield to a hadronic shower in hadronic calorimeter, with and without having associated tracks in ID, respectively.

5.1 Charged track and vertex

The reconstruction of electrically charged particle tracks and primary vertices are based on the information given by the ID, which is composed of pixel, SCT and TRT subdetectors. The track reconstruction algorithm utilises hit coordinates provided by the subdetectors to feed the fitting algorithm based on provided Kalman filters [113–115].

The procedure starts by clustering the pixel and SCT signals above a certain threshold. Then the three dimensional spacepoints of the particles are evaluated from the clusters. The track algorithm groups the sets of spacepoints, which are then used as seeds of the track candidates. A track candidate is built by applying Kalman filtering, which produces the track parameters by iterating through each detector layer of SCT and pixel. Kalman filter approach is designed to

reduce effects of multiple scattering and measurement errors. For low momentum particles, where the effects of multiple scattering are larger, the filter produces tracks which are bended proportional to the particle scatter. For particles with large momentum, for which the effect of measurement errors is larger, the Kalman filter produces tracks consistent with a straight line. At this stage, the hits can be associated to several tracks, therefore a score for each track candidate is calculated by taking into account the assigned number of tracks, the number of holes, the χ^2 of the fitting track, and the track momentum. An ambiguity-solver algorithm is applied to have further requirements on the track candidates, which decreases the scores of the candidates. Further quality requirements can be found in Ref. [116].

The primary vertex reconstruction algorithm is fed by reconstructed track candidates. The reconstruction algorithm is called *adaptive vertex fitting* [117, 118] and is applied in two stages: vertex finding and vertex fitting. The tracks that satisfy certain quality requirements are taken into account for the estimation of primary vertices. One of the commonly used parameter to associate the track candidate to primary vertex is *the transverse impact parameter* (d_0) described as the closest point of a track to the primary vertex in the transverse plane, and the second parameter is *the longitudinal impact parameter* (z_0), which is described as the distance between the point where the d_0 is measured and the primary vertex, along the beam axis. The algorithm starts with the vertex seeds obtained from the z-position of a reconstructed track, then the global maximum of the longitudinal impact parameter z_0 is obtained for all selected ID tracks. In the second step, the vertex position is evaluated by an iterative χ^2 fit. Tracks which are close by the z_0 maximum are used to seed vertex reconstruction. In the iterative fit, a weight is assigned to each track candidate. The larger weights are assigned to tracks with smaller χ^2 values. The vertex position is recalculated using the weighted track candidates and the weights are recomputed until the vertex location no longer changes.

The efficiency of the track reconstruction according to p_T and η is given in Figure 5.2. The efficiency is described as the ratio of particles reconstructed in the ID with respect to a reconstructed track in simulation and its up to 90 %.

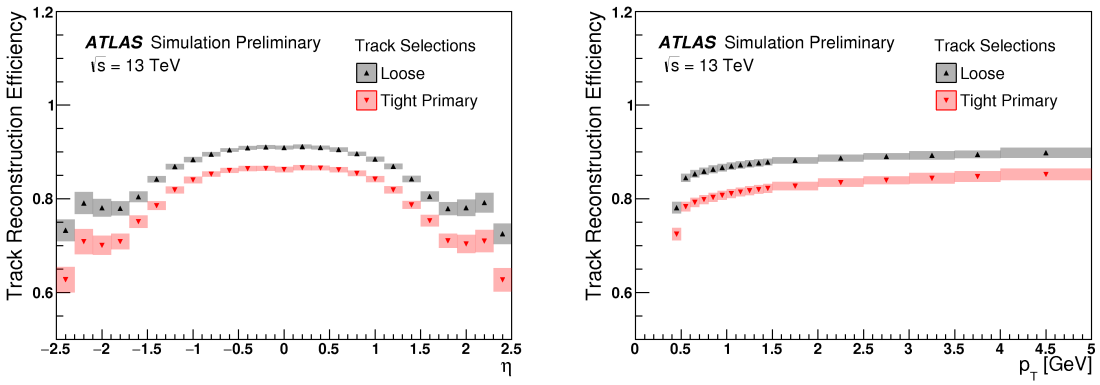


Figure 5.2: ID track reconstruction efficiency as a function of p_T and η [119].

Due to the high pile-up profile at LHC, several primary vertices may be associated with each pp bunch crossing. Therefore, once all vertices are reconstructed, the additional requirement is applied in order to define the *primary hard-scatter vertex*: at least two tracks are associated to it and $|z_{PV}| < 200$ mm. For events which have multiple primary vertices, the vertex with the highest sum of p_T squared is selected as the primary vertex, and the other vertices are considered as pile-up vertices.

5.2 Leptons

Charged leptons leave distinguishable tracks in the ID and energy deposit information in the calorimeters. As a consequence, these informations are used initially in the process of reconstructing the objects. The electron reconstruction algorithm is described in Section 5.2.1, which harmonizes the track information with the information provided by the EM ,while adding the knowledge provided by TRT through transition radiation. The muon reconstruction, given in Section 5.2.2, combines the track knowledge from the ID and and the MS independently.

5.2.1 Electrons

Identification of electron signatures is based on two components (Figure 5.3): energy deposits found in EM calorimeter and close matched associated particle tracks in ID [120, 121]. Additional identification and isolation criteria are applied in order to reduce the wrong reconstruction of electrons. Reconstruction, identification and isolation procedures are explained briefly in the following sections.

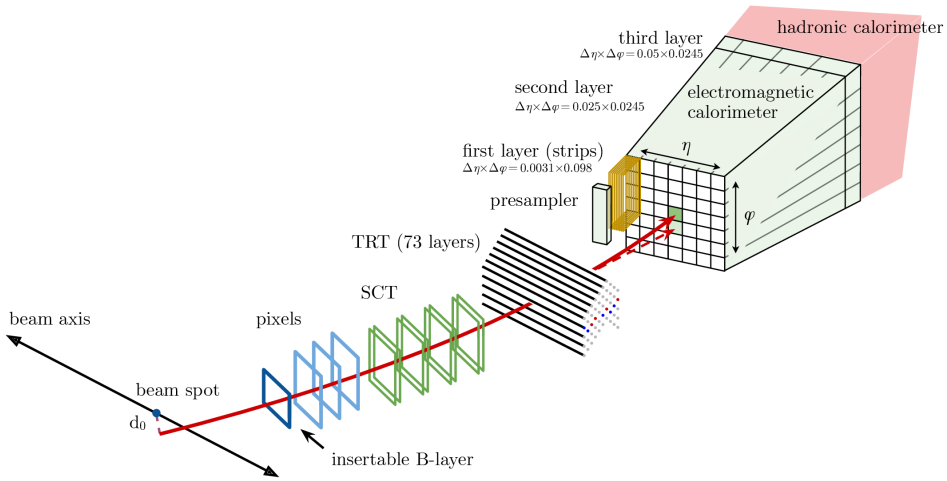


Figure 5.3: A schematic illusturation of the path of an electron through the detector. The electron path is shown with red solid line and the dashed red line shows the photon produced by Bremsstrahlung in the material of the detector [122].

Electron Reconstruction

Electrons typically loose non-negligible amount of their energies while passing through detector medium through bremsstrahlung. The radiated photons can go into a photon conversion, which yields electron-positron pairs. These electrons, positrons and photons are usually emitted closely each other and can share the same electromagnetic cluster matched to multiple tracks from same primary vertex. This type of interactions can take place inside the ID medium or in the beam pipe.

Clustering of the energy deposits begins with the clusters having energies above 2.3 GeV [122]. Clustering algorithm adapts *scanning sliding window* [123] mechanism, which scans over the group of cells whose granularity is determined by the middle layer of the EM calorimeter (3×5 towers with size of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ each) and search for the localised energy depositions. Electron candidates are seeded around the localised clusters, which are then used to find the charged tracks associated to the clusters. The ΔR value of the extrapolated track and the center of the cluster at the intermediate layer of EM calorimeter is used to match the clusters with tracks. The tracks are reconstructed using a Gaussian Sum Filter (GSF) method [124] to account for the bremsstrahlung effects. After the track candidate and the cluster seed are coupled, it is considered as an electron candidate and its momentum is described by the associated charged tracks. Further requirements are applied on the significance variables to match with primary vertex ($d_0/\sigma_{d_0} < 5$ and $|z_0 \sin\theta| < 0.5$ mm). Scale factors are calculated by comparing the simulated electron and data efficiency. For electrons that satisfy the requirement of $p_T \geq 15$ GeV, the efficiency to reconstruct an electron with a good quality track (at least one signal in pixel and at least seven signals at silicon) varies from 97% to 99% [122] depending on the p_T of the electron.

Electron identification

Electron identification (ID) is based on likelihood (LH) algorithm defined by comparing number of variables such as cluster- and track-based information, and calorimeter shower shapes. Those variables are choosen to improve discrimination power on whether a reconstructed electron is so-called signal-like (prompt) where the electrons originate from the primary hard-scatter vertex or background-like (non-prompt electron eg. photon conversions, misidentification of jets or pions, and non-prompt electrons from semileptonic decays of heavy flavour hadrons[1]). The likelihoods of signal and background are based on the probability density functions (PDFs) calculated by using sets of variables as given in Equation 5.1,

$$L_{S(B)}(\mathbf{x}) = \prod_{j=1}^{n} P_{S(B),j}(x_j)\,, \tag{5.1}$$

[1] This type of leptons are produced during the hadronisation process of heavy hadrons such as $b \to X_{c,u}\ell\nu$.

where **x** is the discriminating variables and the $P_{S(B),i}(x_j)$ are the PDFs for variable i at value x_i for the signal and background. Simulated processes are used to define both signal and background PDFs [125]. The final discriminant (d_L) is defined as the relative probability for the object to be signal-like as follows;

$$d_L = \frac{L_S}{L_S + L_B}\,, \tag{5.2}$$

where d_L values are limited by 0 and 1, which makes it inconvenient to define various working points (WPs). For that reason, the final discriminant is transformed by inverse sigmoid function and Figure 5.4 shows the electron LH disriminant after transformation. Depending on the

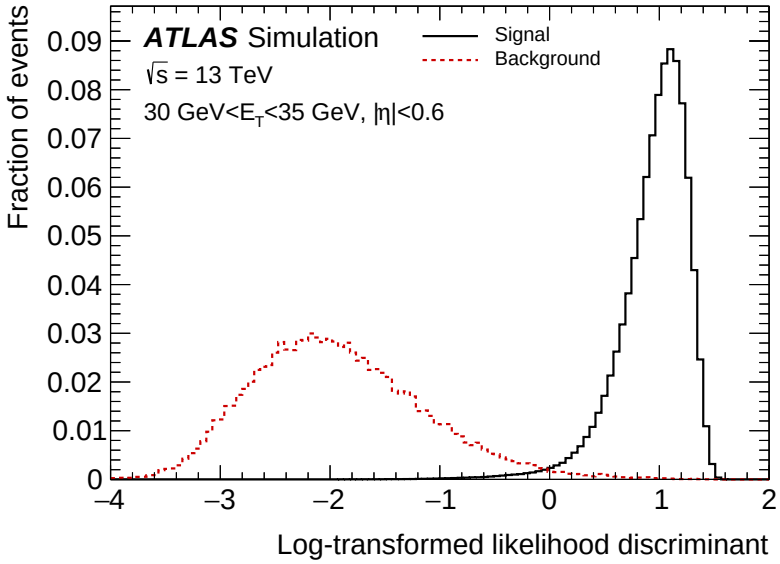


Figure 5.4: The transformed electron identification discriminant for electron candidates [120].

discrimination power, four fixed WPs are introduced for electrons: VeryLoose, Loose, Medium and Tight, grouped by increasing order on the final LH discriminant. The tag-and-probe method is used to obtain the efficiency to select a prompt lepton in samples of $Z \to ee$ and $J/\psi \to ee$. For the Loose, Medium, and Tight WPs, the efficiencies are 93%, 88%, 80% [125] and Figure 5.5 shows the reconstruction and identification efficiencies of electron candidates.

Electron isolation

In addition to the identification criteria on the electron candidates, complementary requirement related to the isolation of electrons can be applied in order to reduce the proportion of electrons coming from non-prompt processes. The criteria is based on the variables related to the

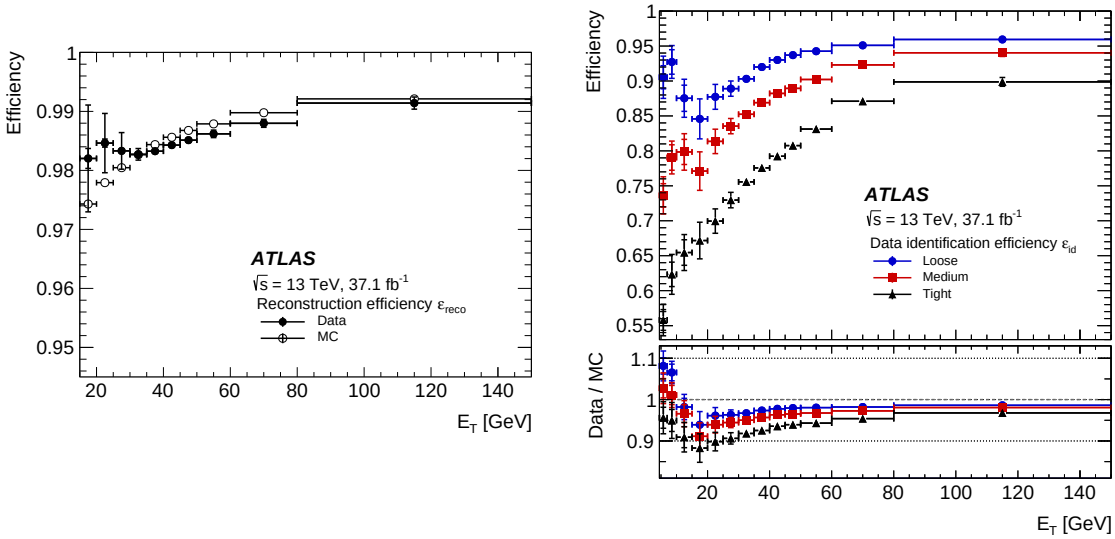


Figure 5.5: The efficiencies of electron reconstruction in simulation and in data as a function of electron E_T (***left***), and identification efficiencies only in data depending on electron E_T for different LH identification WPs Loose (blue), Medium (red), and Tight (black) (***right***). The ratio plots located at lower panel displays the efficiencies obtained in data over in MC.

energy deposits in the surrounding regions of reconstructed electron clusters, which allows for determination of whether a reconstructed electron is isolated. Currently, two types of isolation variables are being used in ATLAS physics analyses [120]:

- Calorimeter-based isolation variable $E_{\mathrm{T}}^{cone\Delta R}$ is the summation of electron candidate transverse energies of positive energy topological clusters around a cone size of $\Delta R = 0.2$ with an energy deposit.
- Track-based isolation variable $p_T^{cone\Delta R}$ described as the sum of the p_T of selected charged tracks within a given cone size R. Different cone sizes can be defined depending on the energy of the cluster $\Delta R = \max(10\ \mathrm{GeV}/p_T,\ \Delta R_{max})$ where $\Delta R_{max} = 0.2 - 0.4$.

The choice of the isolation criteria is specific to the physics analysis requirements. Therefore, several isolation working points are establised using the variables described above defined in bins of E_T and η [125]. Some of the relevant WPs are:

- Fixed, imposing certain fixed requirements on the values of calorimeter and/or the track isolation variables.
- Gradient, targeting a fixed isolation efficiency depending on the p_T and constant in η.

5.2.2 Muons

Due to their high level of penetration inside the detector medium, the identification of the signature of the muon candidates uses additional information from MS on top of the ID and the calorimeter knowledge [126]. The calorimeter information can be used when the muons are out of the MS acceptance. The details related to the muon reconstruction, identification and isolation are given briefly in the following sections.

Muon reconstruction

Reconstruction of muon uses the track information of muon candidates in the ID and in the MS, independently. Afterwards these track informations are merged to obtain the complete muon candidate. The reconstruction of tracks in the ID is described in Section 5.1. The reconstruction algorithm starts with forming the track segments from the individual layers of MS by identifiying the hits. Track segments are combined starting from the middle, and later the inner and outer layers. Aferwards these combined track segments are fit together using a global χ^2 fit to form the candidate. In order to represent the muon track candidate, it is required to have at least two matched track segments. The χ^2 fit is repeated to remove the outer signals several times until all associated signals are below a certain threshold. The final track candidate is built using the combination of tracks from ID, with several different algorithms depending on the interest of the analysis. The four standard combination algorithms are described according to the subdetectors utilised in the reconstruction phase:

- **Combined Muon (CB)** This kind of muon candidates are merged with a global refit algorithm exploiting the muon track candidate signals in the ID and the MS. During this procedure, signals from the MS may be added or removed to improve the fit quality. Muon candidates are reconstructed firstly in the MS and extrapolated inwards to the ID. This is the most common type of muons used in the physics analyses.

- **Segment-tagged Muon (ST)** An ID seed track is considered as a muon candidate, if the CSC or MDT chambers contain at least one associated track segment. This type of muons are useful when a muon candidate interacts only with one layer in the MS chambers due to their low p_T or going into not instrumented regions of the MS.

- **Calorimeter-tagged Muon (CT)** If there is a paired energy deposit in the calorimeter consistent with a minimum ionising particle, a track in the ID is considered as a muon candidate. This algorithm provides relatively low purity of all muon types since it is not using the information from MS, however it is useful to recover the acceptance regions of the MS, specially for the region of $|\eta| < 0.1$.

- **Extrapolated Muon (ME)** ME muons are reconstructed using uniquely the MS track candidates and required to be arising from the primary vertex. This type of muons are typically used to extend the acceptance of muon reconstruction into the region where the ID is not able to cover ($2.5 < |\eta| < 2.7$).

Muon Identification

Muon identification process indicates applying further quality benchmark on the reconstructed muon candidates to retrain contribution from non-prompt background sources, such as hadron decays, while preserving the highest possible rates for the acceptance of prompt muons. The muon identification algorithm make use of variables, which are related to the ID and MS muon track candidates. Therefore, depending on the reconstruction efficiency and the rate of non-prompt muon contribution, four muon identification WPs are developed for physics analyses in ATLAS:

- Loose is developed in order to maximise the reconstruction efficiency and uses all types of reconstructed muons explained above excluding $|\eta| > 0.1$ for CT and ST muons.
- Medium is choosen to be the default criteria for muons in ATLAS physics analyses, which minimises the systematic uncertainties related to the reconstruction algorithm. Only CB and ME types of reconstructed muons are used, and a loose criteria on the hits and regions are required to suppress the misidentification of hadrons as muons. Additionaly, a variable cut related to the change in trajectory is used to control whether it is a decay in flight.
- Tight aimed to maximise the purity of muons with some efficiency lost. Only CB muons having signals at least in the two layers of the MS with satisfying the Medium selection criteria are considered.
- Hight p_T is designed to improve the momentum resolution for the muon candidates with $p_T > 100$ GeV. This type of muons is mainly developed for high-mass resonances searches. CB muons passing the Medium selection are used with additional requirements on the number of signals per layer in the MS.

The reconstruction efficiency for the Medium muons, which are the most relevant ones to the study, as a function of the p_T of the muon candidate shown in Figure 5.6.

Muon isolation

The isolation requirement applied on top of the reconstructed muons, similar to electrons, it uses the track-based and calorimeter-based isolation variables. Muons from the decays of heavy bosons, such as $W \rightarrow \mu\nu_\mu$ and $Z \rightarrow \mu\mu$, are mostly present isolated from other particles, unlike muons from the semileptonic decays originating from jets. The muon isolation efficiency, which is used as a baseline in this work is shown in Figure 5.7.

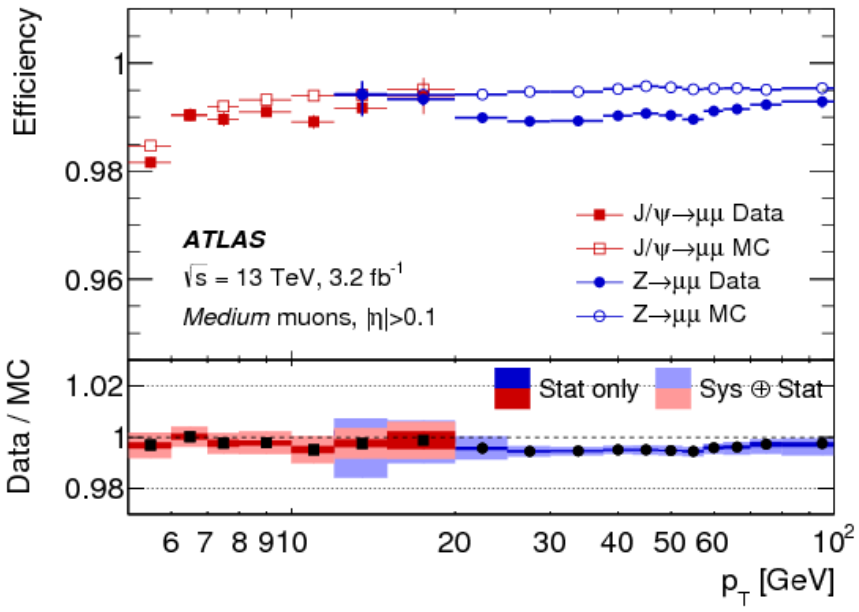


Figure 5.6: The MEDIUM ID muon reconstruction efficiencies as a function of the p_T in the $0.1 < |\eta| < 2.5$ region based on $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events [126].

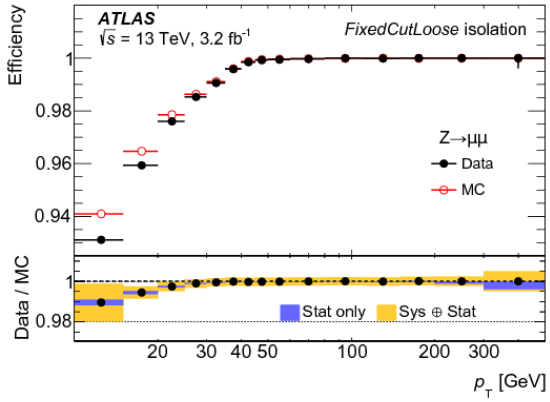


Figure 5.7: Muon isolation efficiencies as a function of the muon p_T for the FIXEDCUTLOOSE working point based on $Z \to \mu\mu$ events [126].

5.2.3 Taus

Unlike the other charged leptons in the SM, tau leptons have a relatively large mass, 1.777 GeV and remarkably short lifetime $O(10^{-13})$ seconds. Therefore, they decay before arriving the detector material, and can be detected via their decay products. Taus are likely to decay hadronically, 65% of the time, into one or three charged hadrons and additional neutral hadrons, or leptonically into lighter leptons [127]. Reconstruction of leptonically decaying taus are

experimentally challenging since it is hard to distinguish decay products from the prompt light leptons. Tau reconstruction and identification studies focus on hadronically decaying taus, so called *hadronic taus*, despite the large background contributions from many hadronic processes at the LHC. The signature of hadronic taus in the ATLAS detector are the electromagnetic and hadronic showers in the calorimeters, associated to tracks in the ID. Jets which represent the tau candidates are formed from topo-clusters with anti-k_t algorithm explained in detail in Section 5.3.1. Futhermore, the identification of the hadronic taus are based on machine learning methods such as boosted decision tree [127] or neural network [128].

5.3 Jets

As a consequence of the confinement nature of QCD, color charged gluons and quarks produced in pp collisions can not be detected as free isolated particles in the detector. Instead, after a set of hadronisation and fragmentation processes, due to the color field which binds them, color-neutral particles such as pions, kaons, protons are produced and finally the special object referred as *jets* are constructed as collimated sprays of these hadrons. The experimental signature of the jets are clusters of energy deposits in the calorimeter cells and the associated charged particle tracks in the ID. Unlike the other objects described above, jets are not formed into cone-shaped regions. Therefore, the common sequential or cone algorithms are underestimating the presence of the soft and collinear emissions. The standard reconstruction method for jets in ATLAS is introduced in Section 5.3.1. The clustering algorithm of jets suffers from several effects, typically detector resolution, granularity and inefficiencies. Therefore, series of calibrations are applied in order to represent them accurately in the physics analyses, as discussed in Section 5.3.2. Finally, the identification of jets initiated by heavy flavoured quarks is described in Section 5.3.3.

5.3.1 Jet reconstruction

The procedure of jet reconstruction begins with the clustering of the calorimeter cells, defined by the readout channels in the Tile and LAr calorimeters. Using the *topological clustering* algorithm [123, 129], three dimensional shape and location information is assigned by grouping neighbouring cells with significant signal-to-expected-noise ratio (S/N) for energy measured in a calorimeter cell relative to expected noise. The signal referred as the absolute calorimeter cell energy, and the noise is described as summation in quadrature of electronic signal RMS and the contribution originating from pile-up.

In order to begin the clustering, signal-to-noise ratio of seed cell is required to be $S/N > 4$. The neighbouring cells are described as three dimensional cells nearby the seed cell within the same calorimeter layer or partially overlapped cells in the (η, ϕ) plane in different calorimeter layer that pass the significance threshold.

Later the algorithm merges iteratively these neighbouring cells into topo-clusters until the last set of neighbouring cells included does not pass the required significance threshold. After

the collection of topo-clusters, the finding of jet algorithm begins. There are several ways to identify jets in ATLAS depending on the purpose of the analysis. The default method is called anti-k_t jet clustering algorithm [130], inspired by k_t [131] and Cambridge/Aachen [132, 133] inclusive jet finding algorithms that are sequential recombination algorithms served as general class algorithms.The anti-k_t algorithm is generalizing the k_t and Cambridge/Aachen algorithms and additionally providing infrared and collinear (IRC) safe properties with geometrically circular jets in the y (rapidity) $-$ ϕ plane, as seen in Figure 5.8. The IRC safety indicates the property which is defined to avoid the change in the clustered jet coming from the additional soft collinear particles arising from other processes. Thus, jets are favorable in opposition to QCD divergent regime. Therefore, it can be used in all the phase space of partons without needing cut-off or divergences scales to remove the effects.

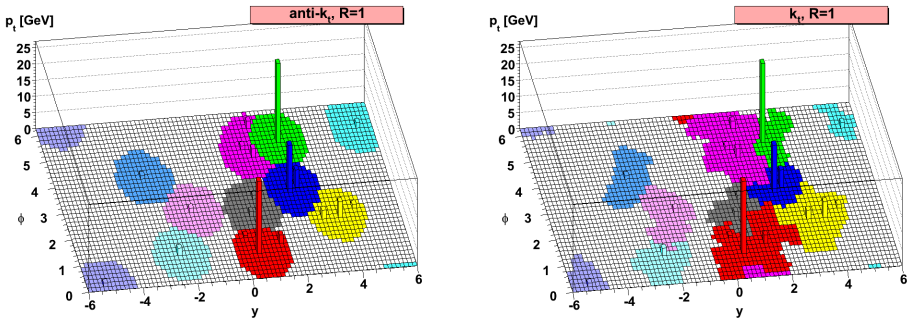


Figure 5.8: An illustration of parton level event with jet components clustered by the anti-k_t algorithm (***left***) and k_t algorithm (***right***). The colored circles referred as the areas populated by jet components, in the case of anti-k_t algorithm hard jet components are approximately circular, while the softer ones have more complex shape in k_t algorithm [130, 134].

The algorithm uses two variables: $d_{i,j}$ represents the distances between two jet topo-clusters i and j, and $d_{i,B}$ defines the distance between the jet topo-clusters i and the beam-line B, which is introduced to improve the rejection of clusters originating from soft proton interactions. The definitions of $d_{i,B}$ and $d_{i,j}$ are given in Equation 5.4 and 5.3.

$$d_{ij} = \min\left(k_{t,i}^{2p}, k_{t,j}^{2p}\right)\frac{\Delta_{ij}}{R^2}, \tag{5.3}$$

$$d_{iB} = k_{t,i}^{2p}, \tag{5.4}$$

where $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ denotes the angular distance between the topo-clusters, where y_i is the rapidity of the constituent i, R is the pre-defined radius parameter of the jet and k_t, i are the transverse momentum. The R parameter is defined as $R = 0.4$ for small$-R$ jets and $R = 1.0$ for large$-R$ jets. The jets ordering of inputs for reconstruction is governed by the

parameter p. The k_t algorithm has value of $p = 1$, giving priority to cluster the objects with low p_T. The Cambridge/Aachen algorithm has value of $p = 0$, which yields to energy-independent clustering. Finally the anti-k_t algorithm corresponds to $p = -1$, in favor of clustering the objects with higher p_T, whose distance between the clusters are the smallest, can be seen by Equations 5.3 and 5.4. In the anti-k_t algorithm, the choice of p ensures that the smallest distances are associated with the hardest particles and since the clustering algorithm begins from the hardest particle, the jet direction is governed by the hardest topo-cluster.

Among the set of input topo-clusters, the jet components satisfying the smallest d_{ij} are recombined to form a single cluster in the list. If $d_{ij} > d_{iB}$, then the component i is treated as the complete jet candidate and removed from the list. This process repeats, by recomputing the d_{ij} and d_{iB} with new object, until no particle remains.

5.3.2 Jet calibration and pile-up suppression

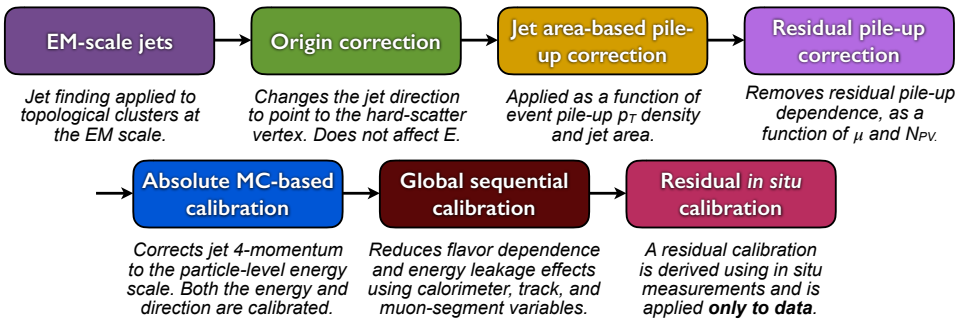


Figure 5.9: Flow diagram illusturating the jet calibration steps in sequence [135].

The reconstruction of topo-clusters described so far does not naturally represent the whole energy deposited by a particle. The electromagnetic showers are narrower than hadronic showers, therefore topo-clusters may contain the hadronic showers partially. In order to restore the jet energy scale series of corrections are applied to EM-scaled jets. The full set of jet energy scale (JES) calibration procedure for small-R jets adopted in Run2 for anti-k_t is demonstrated in Figure 5.9, at which the corrections are applied to the jet four-momentum, except the jet origin correction. The direction of the reconstructed EM-scale jet is corrected to be originated from the primary vertex by comparing the jet four-momentum calculated according to the center of the detector and the four-momentum from hard-scatter primary vertex, while keeping the jet energy constant. This correction is applied in order to improve the pseudorapidity resolution.

Subsequently, the pile-up contribution is calculated on EM-scale jets by benefiting from two variables: the median p_T density in η-ϕ plane (ρ), the number of primary vertices (N_{PV}), and the average number of interactions ($< \mu >$). The resulting calibration procedures are then called *area-based* and *residual* corrections. The active area (A) of the jet is obtained by the ghost particles associated to jet, which is described in Ref [134] in detail. The p_T density of

each jet (p_T/A), is then calculated and the average p_T density (ρ) is determined as the median of all p_T densities in the event. Due to the higher number of jets in the forward regions, this correction is only applied in the region $|\eta| < 2$.

The next correction to pile-up contribution aims to reduce the residual dependence of the reconstructed jet p_T on higher occupancy regions or higher occupancy jets by comparing to the truth jets in simulated events. Residual dependence corrections are derived from the number of reconstructed primary vertices N_{PV}, and the average number of interactions $< \mu >$. These corrections are evaluated by fitting linearly the jet p_T as a function of the pseudorapidity of the detector.

After estimating all the parameters, the final pile-up correction on p_T of the reconstructed jet is given in Equation 5.5 as follows;

$$p_T^{\mathrm{corr}} = p_T^{\mathrm{reco}} - \underbrace{\rho \times A}_{\text{Area-based}} - \underbrace{\alpha \times (N_{\mathrm{PV}} - 1) - \beta \times \mu}_{\text{Residual}}, \tag{5.5}$$

where p_T^{reco} and p_T^{corr} are p_T of jet before and after the pile-up correction, respectively. A is referred as the jet area, and the α and β expressions are obtained from the linear fits, and defined as $\alpha = \partial p_T / \partial N_{\mathrm{PV}}$ and $\beta = \partial p_T / \partial \mu$, respectively.

The next step of jet calibration process is the absolute calibration of jet energy scale and η. This phase of corrections are required to be applied to account for the effect of having different calorimeter technologies, the mismodelings in the active material of the detector or changes in the calorimeter granularity. Purely MC-based corrections are derived by comparing the energy of the truth particle level jets from Pythia sample to the EM-scale reconstructed jets on top of the origin and pile-up corrections. The average energy response, described as the mean of a Gaussian fit to E_{reco}/E_{truth}, is computed as a function of E_{truth} and η_{det}, where η_{det} is the pseudorapidity according to the center of the detector. The second correction is also evaluated as the difference between the η_{reco} and the η_{truth} as a function of E_{truth} and η_{det} similarly. The inverse of the results are taken as a correction to the reconstructed EM-scaled calorimeter jets in simulation. Figure 5.10 demonstrates the average energy response and the derived η correction for different energy of matched truth jets, as a function of η_{det}. Following this correction step, the jets are referred as EM+JES scaled jets.

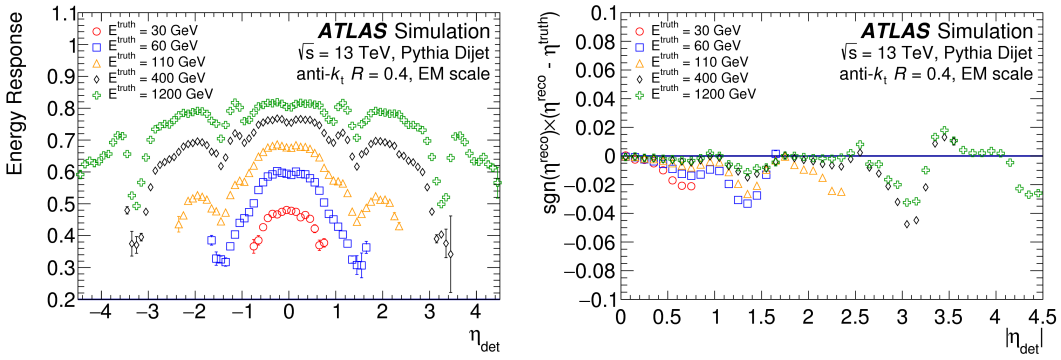


Figure 5.10: Average energy response of the jets after the corrections on the origin and pile-up, as a function of η_{det} (***left***) and the difference η_{truth} and η_{reco} (***right***), as a function of η_{det}. The colors represent the different energies for the truth jets that are matched to EM-scaled jets. The largest correction is in the low p_T region. The largest bias is seen for $|\eta_{\text{det}}|$ 1.4 (3.1), corresponding to the barrel-endcap transition regions.

Another correction process is applied to the reconstructed jets is *the Global Sequential Calibration, GSC*. This correction intended to catch the remaining dependencies such as difference in the jet shower shapes and in the jet flavour composition or jet energy distribution, mostly between quark- and gluon-initiated jets. Typically the quark-initiated jets are consist of fewer number of components, but with higher p_T, wherease the gluon-initiated ones usually populate more components and softer in p_T, therefore are not penetrating further in the detector medium. Five different observables are defined in order to improve the corrections through the GSC, and each is corrected independently as again the inverse of the jet response binned in p_T and η_{det}. These variables are mainly sensitive to the shower profile of the jets and different layers of calorimeters while using the track information associated to the reconstructed jet. Among these parameters, the number of muon track segments ghost-associated with jet is particularly important since it reduces the effect of the leak into MS, called *punch-through jets*.

The final step of calibration is called *in-situ calibrations*, which takes into account the difference of reconstructed jet response between data and MC. The correction consists of two stages. One is correcting the forward jet response ($0.8 < |\eta| < 4.5$) according to well described central jets (using the dijet events which are back-to-back in ϕ), which is called *η-intercalibration*. The other correction is applied to the central region jets ($|\eta| < 0.8$), which calibrates the EM-JES scaled jets with respect to a well defined reference object. The final correction on the response of jets depending on the p_T of the object, derived by using in-situ methods is shown in Figure 5.11.

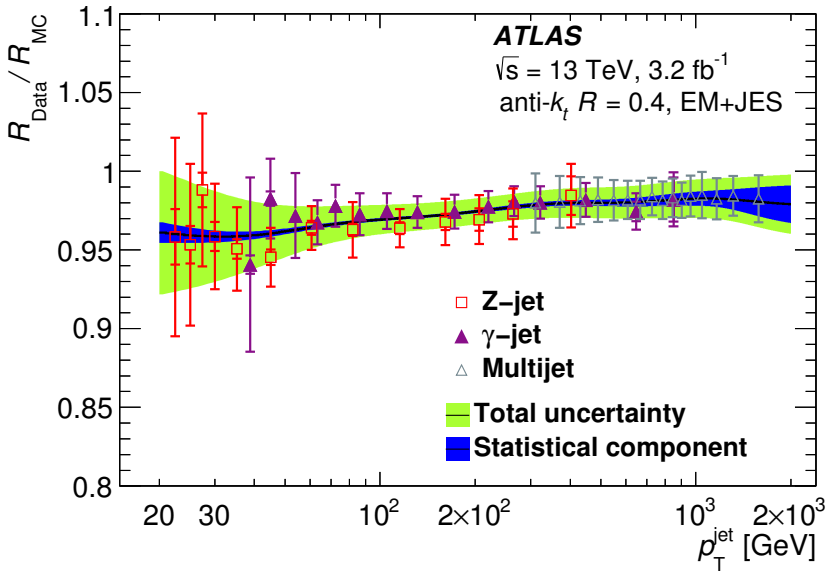


Figure 5.11: The ratio of response of EM+JES jet as a function of jet p_T, in data over MC simulation. The different type of markers demonstrate different reference objects used in in-situ corrections. The final correction is showed with black line, its statistical and total uncertainties are given by dark blue line and light green line, respectively [135].

The final combination of all uncertainties on the JES response is computed and provided as 80 systematic uncertainties. Figure 5.12 illustrates the total uncertainty on the JES response in 2015 conditions and it is seen that for $20 < p_T < 50\,\text{GeV}$ range, the uncertainty is 4%, wherease for the higher p_T regimes it is around 1%.

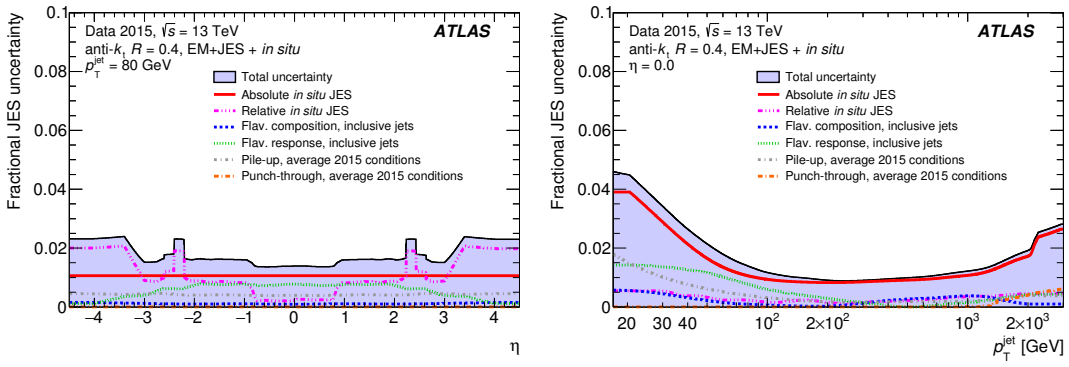


Figure 5.12: The combined JES uncertainties on the calibrated jets according to p_T of the jet at $\eta = 0$ (***left***). The combined JES uncertainties on the calibrated jets as a function of η of the jet with $p_T = 80\,\text{GeV}$ (***right***). The colors represent the different sources of uncertainties, as well as the total and relative uncertainties [135].

Following the reconstruction and the calibration of the jets, the pile-up jets need to be identified since the pile-up interactions may result in the final states, which are not associated to the hard-scatter. A multivariate algorithm is developed in order to differentiate the pile-up jets from jets originating from the hard-scatter, called *Jet Vertex Tagger (JVT)*. This algorithm utilises the associated track-based information to the reconstructed jet to define how likely the corresponding jet is coming from hard-scatter [136]. The variables called *corrJVF (Jet Vertex Fraction)* and R_{p_T} are defined. The parameter R_{p_T} is defined as the ratio between scalar sum of the track p_T in jet coming from the PV, $\sum_k p_T^{trk,k}(PV_0)$, to p_T of the jets. The corrJVF parameter is defined as the summation of the p_T of all tracks associated to the jet divided by jet p_T, regardless from the vertex they originate from. Furthermore, the pile-up tracks are weighted with respect to the total number of pile-up tracks in the event n_{trk}^{PU} in order to remove the dependency of the number of pile-up vertices. The value of R_{p_T} is expected close to be higher for jets from hard-scatter and close to 0 for the pile-up jets.

These variables are combined into a two-dimensional likelihood that relies on the k-nearest neighbours (k-NN) algorithm. The relative probability of the jet originated from hard-scatter is given at each point in the two-dimensional plane. The distribution of JVT for pile-up and hard-scatter jets are given in Figure 5.13. The higher JVT value corresponds to the higher probability of being a jet originated from hard-scatter.

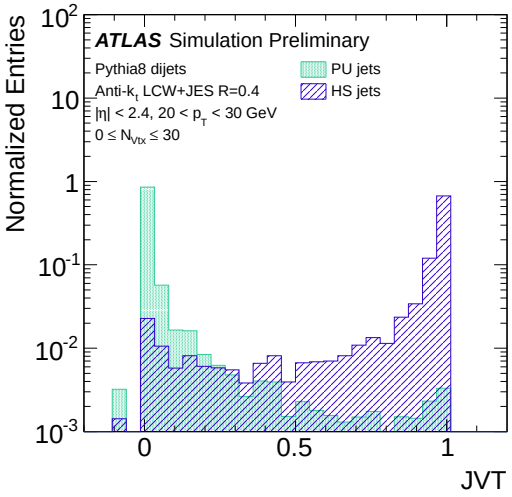


Figure 5.13: A distribution of JVT score for the pile-up (green) and hard-scatter (purple) jets. Negative values represent the jets with no associated tracks.

As last step, many of the analyses make use of an algorithm to remove the *bad* jets. This kind of reconstructed jets are not originating from a pp collision but rather from different sources such as noise fluctuations occured in the calorimeter, cosmic muons or beam protons which are scattered with gas in the beam pipe. Several variables are defined to differentiate bad jets

according to their sources. For instance, in order to reject the jets produced due to fluctuations or noise in the calorimeters, the quality variable exploiting the cell information is used. In total, seven variables are defined exploiting tracking information and enegy depositions [137].

5.3.3 Jet heavy flavour tagging

Identifying the jets containing heavy-flavoured hadrons, such as the jets originating from the fragmentation and hadronization of a b-quark is fairly achievable due to relatively long lifetime and distinguishable experimental signature from other hadrons. The identification of this type of jets are called *heavy flavour tagging* and the resulting jets are then called *b-jets*. Therefore, all other jets are called *light jets* or simply *jets*. b-quarks are the lightest member of the third generation of elementary particles and hence may only decay into first or second generations (c,u) weakly and they are suppressed. Other kind of decays are not allowed (FCNC). Therefore, the average life time of b-hadron ($\tau \approx 1.6\ ps$) yield to travel $c\tau \approx 450\,\mu m$ away from the primary hard-scatter vertex before it decays inside the detector [138]. The decay product forms a secondary vertex that can be reconstructed using the ID which is displaced from the primary vertex as illustrated in Figure 5.14.

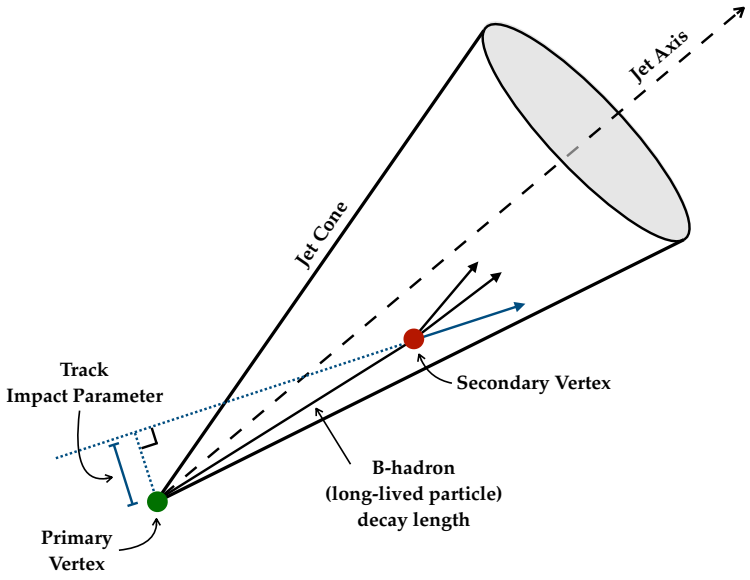


Figure 5.14: An illusturation of expected signature of b-jet in the ID.

In the process of discriminating the b-jets in ATLAS experiment, three basic algorithms are employed:

- **IP2D and IP3D algorithm:** These algorithms exploit the transverse and longitudinal impact parameter significance of the tracks associated to jets. Generally, the tracks

associated to the secondary vertex have a positive signed impact parameter, whereas the others are symmetric. The log-likelihood ratios are constructed in order to differentiate the b-jets.

- **Secondary Vertex Finding Algorithm (SV1):** This algorithm aims to reconstruct a single inclusive displaced secondary vertex within the jet by using all possible two track vertices except the tracks coming from long-lived particles, such as K_s and Λ, photon conversions, or vertices due to detector material interactions. The inclusive secondary vertex and associated tracks are then used to construct discriminating observables sensitive to the differences between b- and light-jets.

- **Multi-vertex Finding Algorithm (JetFitter, JF):** JetFitter aims to reconstruct the full decay chain of b-hadron using a Kalman filter [113].

Some of the variables [139] used by the tagging algorithms described above are shown in Figure 5.15.

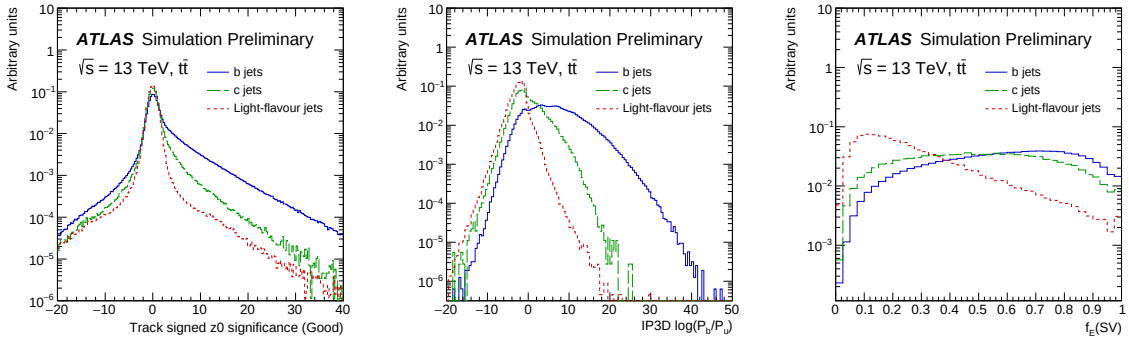


Figure 5.15: Some of the variables used by ATLAS b-tagging algorithms for b-jets (solid blue line), c-jets (dashed green line) and light-flavour (dotted red line). Longitudinal impact parameter significance (z_0) for tracks associated to jets (***left***). The log-likelihood ratio for the parameter of IP3D (***middle***). The energy fraction of the tracks in the displaced vertex with respect to the energy of all the tracks in the jet, used by the SV1 algorithm (***right***).

The outputs of the algorithms decribed above are combined with jet kinematics (η and p_T) using multivariate technique called *Boosted Decision Tree (BDT)* and an *Deep Neural Network (DNN)*, in order to maximise the performance. The resulting algorithms are called MV2 and DL1, respectively. The MV2 is trained depending on their charm composition in training and the best found to be MV2c10 with 7% c-fraction. The DL1 algorithm exploits also the same variables as in MV2 but with a different architecture of multivariate technique which results in a better performance [140]. Furthermore, both algorithms can be used at different WPs depending on the efficiencies of b- and c-tagged jets. Figure 5.16 shows the light-flavour and c-jets rejections with respect to b-tag efficiencies for different algorithms.

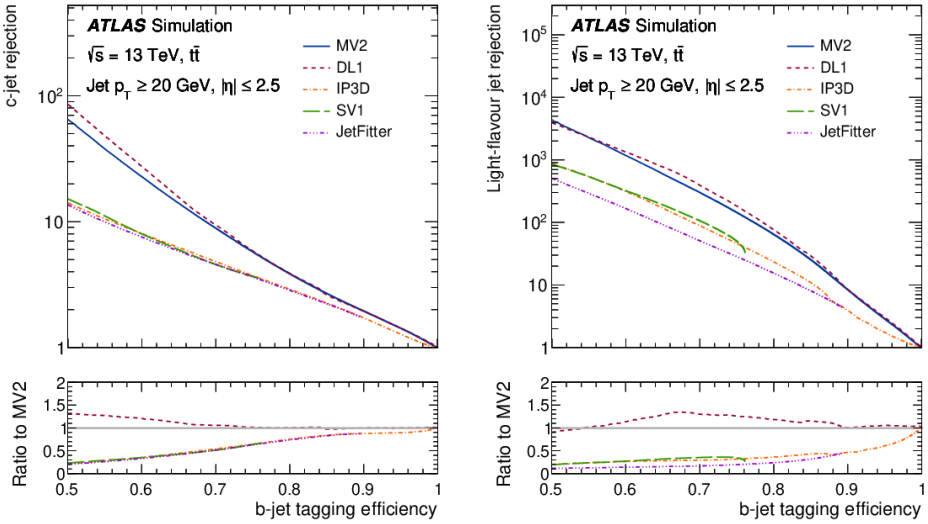


Figure 5.16: The light-flavour (***right***) and c-jet rejections (***left***) with respect to b-jet tagging efficiencies for the MV2, DL1, IP3D, SV1 and JetFitter algorithms evaluated on the $t\bar{t}$ events.

5.4 Missing transverse momentum

Momentum conservation implies that the total momentum before and after collisions must be identical. At LHC, the pp interactions are presumed to be head-to-head, meaning that the initial momenta of the particles has only parallel component to the beam-axis and the outgoing transverse components should cancel out. Thus, the sum of the transverse momentum of all objects, which originated from the primary hard-scatter vertex must be zero. Missing transverse momentum ($\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$) can be defined as the negative sum of the all transverse momenta of the objects associated to the corresponding event [141] with additional soft radiation, and its magnitude denoted $E_{\mathrm{T}}^{\mathrm{miss}}$. The non-zero value of $E_{\mathrm{T}}^{\mathrm{miss}}$ may be originated by weakly interacting SM particles such as neutrinos, as well as BSM particles which can escape the detector covarage. Therefore the $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated using the following formula:

$$E_{x\,(y)}^{\mathrm{miss}} = -\sum_{i\in\{\text{hard objects}\}} p_{x\,(y)}, i - \sum_{j\in\{\text{soft signal}\}} p_{x\,(y)}, j\,, \tag{5.6}$$

where the hard objects represent the stable final state objects such as electons and muons, and the soft signal is sum over the tracks only originating from the primary vertex which are not part of another reconstructed object in the x- (y-) direction.

CHAPTER 6

Statistics

Interpretation of the compatibility between the theory predictions and experimental observations is evaluated through statistical concepts. This section gives a description of the statistical procedures used in the physics analyses at LHC in the context of a *frequentist* approach. Within this approach, the probability is associated with the relative recurrence of an outcome from repeatable experiments. The likelihood formalism that is widely used to quantify the probabilistic behaviour of the experiments is described in Section 6.1, and one of the most efficient methods to estimate the parameters of the model, the *maximum-likelihood* method is given in Section 6.2. The formalism used in physics searches to make decisions and to draw conclusions based on *test statistic* is described in Section 6.3. Two important concepts of statistical measurements known as assessment of the *confidence intervals* and the *upper limits* are described in Section 6.4. The natural generalisation of including parameters to the measurements *profile likelihood* method is described in Section 6.5. Estimation of the expected median significance is performed on *asimov* dataset, as described in Section 6.6. Finally, some assets of the systematic uncertainties is given in Section 6.7.

The theoretical information of this chapter is taken from the Ref. [142–145], and the mathematical formalisms are adapted within each other.

6.1 Formalism

When considering frequentist probabilities, statistical claims are based on the outcome of an experiment, while using constructed ensemble of experiments from simulations or mathematical functions. Replication of the experiment will lead to different values of observables, denoted with x, and this ensemble will result to *probability density function (PDF)* of a variable, which can be normalized to unity (Equation 6.1). The family of PDFs are written by paramaterizing the probability density ($f(x|\theta)$) as a function of model parameters, denoted by θ, and can

typically symbolize parameters of a physical theory or other parameters of the system.

$$\int f(x|\theta)d\boldsymbol{x} = 1\,. \tag{6.1}$$

The probability of obtaining a result x given value of a model parameter θ can be given by *likelihood*, $\mathcal{L}(x|\theta)$. The likelihood function describes all the information relevant to the measurement of θ and its value indicates the plausibility of a particular value of θ.

If the individual experiments are considered independent and following the same underlying distribution, then the probability density of a dataset is simply the product of densities for each experiment. Given a finite number of experiments (N) from the distribution $x = \{x_1, x_2, ...x_N\}$, the likelihood can be constructed as

$$\mathcal{L}(x|\theta) = \prod_{i=1}^{N} f(x_i|\theta)\,, \tag{6.2}$$

where the individual experiment x_i follows the probability density of $f(x|\theta)$, and θ is a set of unknown parameters to be estimated. When repeating an experiment with same distribution, the observed rate of an experiment may fluctuate around the predicted value denoted by ν, according to a Poisson distribution. Therefore, determination of *absolute rates* can be obtained by *extended likelihood* as given following:

$$\mathcal{L}(x|\nu,\theta) = \mathrm{Pois}(N|\nu)\prod_{i=1}^{N} f(x_i|\theta)\,. \tag{6.3}$$

Here, the product of the f is the standard likelihood. It is also quite common to work with logarithm of the likelihood since the terms can lead to negative values. Therefore, logaritmic likelihood function can be written as

$$-\ln(\mathcal{L}(x|\nu,\theta)) = \underbrace{\nu - N\ln\nu}_{\text{extended term}} - \sum_{i=1}^{N}\ln f(x_i|\theta) + \underbrace{\ln n!}_{const.}\,. \tag{6.4}$$

While estimating the parameters from large data sample, where the estimation method is described in Section 6.2, the likelihood function can be binned in order to be computationally efficient. Each individual experiment can be grouped in bins of a variable x, and this may change the shape and rate of the likelihood function. This procedure is accepted as long as the information on the parameter from the variation of the probability function $f(x|\theta)$ within bin is not significant compared to the variation derived over all the bins. Auxiliary measurements or control regions can be also added to constrain or reduce the effect of systematic uncertainties. The construction of the likelihood function is often the most crucial step in the physics analysis. In ATLAS, the framework commonly used to create the likelihoods is called *HistFactory* [146].

6.2 Parameter estimation and maximum likelihood

An estimator is a function of the data and its value is used to obtain the estimation of a given parameter. There are various abstract properties of the estimators such as variance, bias, efficiency, robustness, and choice of the estimator is usually done by taking into account all its properties. The most widely used estimator in high energy physics is the *maximum likelihood estimator (MLE)*. The MLE of the parameter θ is defined as the value of θ for which the likelihood function $\mathcal{L}(x|\theta)$ has its global maximum $\hat{\theta}$. Equivalently, this value maximizes the $\ln \mathcal{L}(x|\theta)$, minimizes $-\ln \mathcal{L}(x|\theta)$ and this procedure is then called *fitting*. Typically in physics analysis, several parameters enter to this procedure, and the estimation becomes more complicated. Therefore the MLEs of these parameters are used in the global measurement of the experiments. Some are referred as *unconditional MLE*, referred as parameter is *floating*. While some of the parameters can be also constrained, which is then called *conditional MLE*, denoted by $\hat{\hat{\theta}}(\theta_0)$. Since the value of a likelihood function is arbitrary, it holds the meaning on its own, however it is more common to bring meaning to its value by comparing it at two different points in parameter space. In order to achieve this, the test statictics is defined by using the log-likelihood ratio. In the case of a dependency of many parameters, the *profile likelihood ratio* is used. These concepts will be explained in the following sections.

There are several important aspects of MLE, which makes it beneficial for the needs of physics analysis at LHC. Assuming that the number of experiments N goes to infinity, the MLE is consistent, which means the parameter converges to its true value. In this asymptotic limit, the MLE is unbiased and reaches the minimum-variance bound, which makes it the most efficient estimator. The common tool used in ATLAS to evaluate the MLE is Minuit [147].

6.3 Test statistic and hypotheses testing

In order to draw a conclusion from the observed data, test statistic (q) is defined by comparing one or multiple hypothesis, often called *hypothesis testing*. The conclusion on a given hypothesis is done based on the observed value of the test statistics and where the value is placed compared to the pre-defined *critical region*. In the physics searches, two hypotheses are considered. The first is the *null hypothesis*, denoted by H_0, which is subject to the hypothesis to test and considered to be true. The null hypothesis is commonly referred to as the *background-only hypothesis*. The second hypothesis is the *alternate hypothesis*, denoted by H_1, and corresponds to a complementary hypothesis, which serves as the comparison of H_0 to optimise the hypothesis tests. This is represented by the *signal-plus-background hypothesis* in physics and sometimes refers to the presence of both SM and the addition of the BSM physics process sought after. In high energy physics, the claim of discovery is a statement that data is incompatible with the background-only hypothesis.

The determination of critical region needs to be done. If the measured value of test statistics lies within this region, the hypothesis H_0 is rejected, and if it lies in *acceptance region* then it is failed to reject the H_0 hypothesis. In this context, the threshold to reject the H_0 hypothesis α, or

also known as the size of the test (or significance level or the rate of a Type-I error) is defined as follows:

$$\int_{q_c}^{+\infty} f(q|H_0)\,\mathrm{d}q = \alpha\,, \tag{6.5}$$

where the critical region corresponds to the domain with $q \geq q_c$ and $f(q|H_0)$ is the probability distribution for the test statistic under assumption of the background-only hypothesis, which will be necessary to determine. In the case of existence of H_1 hypothesis, the quantity β, also known as Type-II error rate is defined as follows:

$$\int_{-\infty}^{q_c} f(q|H_1)\,\mathrm{d}q = \beta\,, \tag{6.6}$$

where $f(q|H_1)$ is the probability distribution for the test statistic under the signal-plus-background hypothesis. The quantity $1-\beta$ is referred to as the *power of the test*. The performance of the test can be evaluated by its ability to discriminate the null hypothesis from alternative hypothesis; the more H_0 and H_1 are separated, the smaller α will be.

The choice of the size of a test α determines the size of the acceptance region, where the null hypothesis being rejected. For a chosen size α, an ideal test should maximize the β value to not have the test biased ($1-\beta \leq \alpha$). The Neyman-Pearson lemma [148] indicates that the most powerful acceptance region with high statistics for a significance level α can be defined as

$$q_x = \frac{\mathcal{L}(x|H_1)}{\mathcal{L}(x|H_0)} > \alpha\,, \tag{6.7}$$

which corresponds to the most powerful test statistics called *likelihood ratio*. Even though this test is powerful, its usability is often limited due to the presence of systematic uncertainties in the complex models. After the test statistic is defined and the expected distribution is obtained, the concept of p-value or its corresponding Gaussian significance can be used to evaluate the statistical significance of an observation. The p-value of the experiment is described as the probability for observing the test statistic values equal to or greater than the value of q_{obs} as follows:

$$p = \int_{q_{\mathrm{obs}}}^{+\infty} f(q|H_0)\,\mathrm{d}q\,, \tag{6.8}$$

where $f(q|H_0)$ is the PDF of q under the presumption of null hypothesis and q_{obs} is the observed value of the test statistic in data. If the pre-defined significance value α is greater than observed p-value, then the hypothesis is rejected and the small p-value is an indication of the

inconsistency with the hypothesis. The value of α, which serves as the threshold to reject the hypothesis, is conventional. In high energy physics, a discovery is claimed for $\alpha = 2.87 \times 10^{-7}$, which corresponds to significantly small Type-I error. It is also conventional to represent the p-value with significance Z, and the relation between them is given as

$$Z = \Phi^{-1}(1 - p), \tag{6.9}$$

where Φ^{-1} is the cumulative distribution function of the standard Gaussian. Therefore, in order to claim the discovery under the null hypothesis, the corresponding significance value calculated as $Z = 5$, or equally can be expressed as 5σ.

6.4 Confidence intervals and upper limits

The physics models are parametrized by theoretical parameters, and usually the aim is to determine which region in the parameter space is allowed or excluded by the dataset. In the frequentist approach, this region is called *confidence intervals*, and instead of a single numerical estimate, two limits are provided together with a level of confidence about the expected value of the parameter lying between these limits. Confidence level is the primary characteristic of an interval construction, and the procedure of building it is called *Neyman Construction* [149]. This method is based on reversing a series of hypothesis tests. Particularly, for each point in the parameter space, the hypothesis testing is performed. If one wants to obtain 95% confidence interval, then a series of hypothesis test must be performed with a size of 5%. The confidence interval is then constructed by taking the set of parameter points where the null hypothesis is accepted.

Figure 6.1 illustrates the Neyman construction for a parameter θ. Here initially the point estimator is chosen ($\hat{\theta}$), later the PDF distribution of $\hat{\theta}$ for several values of θ is estimated. In this case, for each point in the parameter space, the acceptance interval of $\hat{\theta}$ values with fixed integrated probability is selected. Afterwards the interval boundaries are connected, and the observed value of x_0 is evaluated. The confidence belt is used to derive the corresponding interval $[\theta_1, \theta_2]$ for θ variable. The green region in Figure 6.1 is forming the confidence belt.

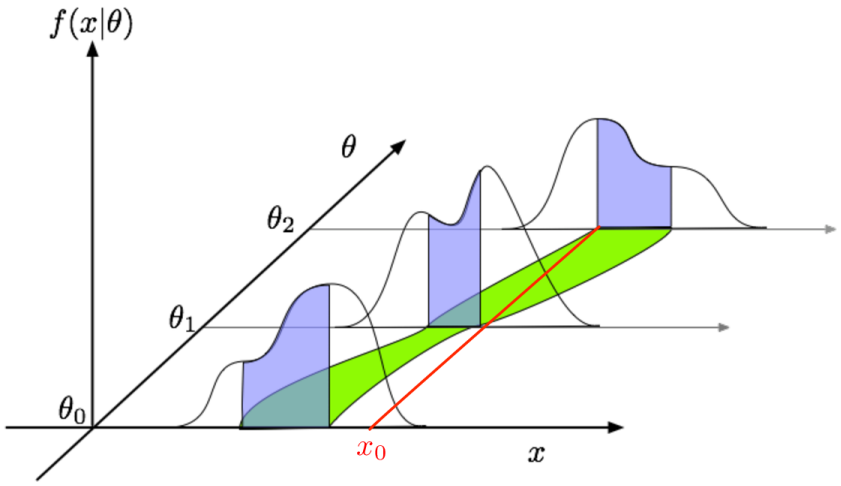


Figure 6.1: A schematic visualization of the Neyman Construction [142].

6.4.1 The CL_s construction

The CL_s construction [150] is the technique for setting limits to determine the exclusion intervals under certain confidence level. In this method, the probability to exclude the signal hypothesis is aimed to be reduced, and the exact confidence in the signal hypothesis is determined, while the background contribution is efffectively higher than the signal contribution. The CL_s metric is given by,

$$CL_s(\mu) = \frac{p_\mu}{1 - p_0}, \tag{6.10}$$

where p_0 and p_μ assess the compatibilities between the data and the background-only and signal hypothesis, respectively (similar to Equation 6.8). The experiments with relatively small sensitivity, the downward fluctuations in observed data can yield to larger values of p_0, and therefore small upper limits on μ. However, resulting CL_s values will remove this effect and avoid exclusion on the signal. A given signal hypothesis is considered as excluded at 95% CL if $CL_s \leq 0.05$, and the *upper limit* is the largest signal hypothesis value satisfying $CL_s \geq 0.05$.

6.5 Profile likelihood

The test statistic application for one dimensional parameter hypothesis is straight forward. However, if the hypothesis has more than one parameters, the method described in the previous sections become more complex and computationally expensive. Therefore in such cases, the *profile likelihood ratio* is used, which approximates the proportion of the parameter space by using their corresponding maximum likelihood estimators. Within the full parameter space, *parameter of interest POI*, is typically a free parameter, and *nuisance parameters NPs*, are

the parameters whose values can be constrained with aN auxiliary knowledge, usually from data while unconstrained nuisance parameters also can be used. The parameter of interest is often the overall signal strength denoted by $\mu = {}^{\sigma}/_{\sigma_{SM}}$ where σ_{SM} is the SM production cross section. The nuisance parameters are the systematic uncertainties that permit variations of the expectations of signal and background, subject to Gaussian or log-normal constraints of the experimental setup. In the physics analyses, the test statistic is based on counting experiments, performed by taking into consideration the number of events from the predicted signal and background. The events can be also grouped in bins of observables, and the counted event numbers in each bin are then used in the likelihood function. For each event and a variable, the measured values are used to construct a histogram $n = (n_1, n_2...., n_R)$ of the variable. Therefore the expected number of events is given by

$$N_{n_i}^{exp}(\mu_{\text{sig}}\,\boldsymbol{\nu}, \boldsymbol{\theta}) = \mu_{\text{sig}} \cdot s_i + \boldsymbol{\nu} \cdot b_i\,, \tag{6.11}$$

where μ_{sig} and $\boldsymbol{\nu}$ are the signal strength and the normalisation factors for the backgrounds, respectively. The mean number of entries in the bin i for signal and background are given as

$$s_i = s_{\text{tot}} \int_{i\text{th bin}} f_s(x|\boldsymbol{\theta}_s)dx\,, \tag{6.12}$$

$$b_i = b_{\text{tot}} \int_{i\text{th bin}} f_b(x|\boldsymbol{\theta}_b)dx\,, \tag{6.13}$$

where $f_s(x|\boldsymbol{\theta}_s)$ and $f_b(x|\boldsymbol{\theta}_b)$ defines the PDFs of the variable x and the total number of signal and background events are given with parameters s_{tot}, b_{tot}, respectively. The set of the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_{\text{tot}})$ represent the NPs. In addition to the measured histogram n, it is common to make further complementary measurements that constrain the nuisance parameters. For instance, adding a control region where the background events expected to contribute more, and from them constructing a histogram of some chosen kinematic variables where their shapes are dissimilar for background and signal processes. Usually a systematic uncertainty is parametrised by the its expected nominal value. Only the values of $\theta_i = \pm 1$ are known, which correspond to their variation of $\pm 1\sigma$ around the nominal value to simplify the calculations.

The effects of the NPs associated with each of the systematic uncertainties in the analysis are implemented as *constraint terms* in the likelihood function and typically implemented as Gaussians centered at zero and with a width of unit, appearing in Equation 6.14. The observed data yield of the analysis is expected to obey Poisson statistics. Therefore, the likelihood

function $L(\mu_{\text{sig}}, \boldsymbol{\nu}, \boldsymbol{\theta})$ is formulated as a product of Poisson probabilities for all bins:

$$L(\mu_{\text{sig}}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \prod_{i=1}^{\text{R}} \frac{\left[N_{n_i}^{\text{exp}}(\mu_{\text{sig}}, \boldsymbol{\nu}, \boldsymbol{\theta})\right]^{N_{n_i}^{obs}}}{N_{n_i}^{obs}!} \cdot \exp\left[N_{n_i}^{\text{exp}}(\mu_{\text{sig}}, \boldsymbol{\nu}, \boldsymbol{\theta})\right] \cdot \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta_j^2}{2}\right), \tag{6.14}$$

where R is the total number of bins, $N_{n_i}^{obs}$ is the observed number of data events in i^{th} bin of the histogram n and $N_{n_i}^{\text{exp}}(\mu_{\text{sig}}, \boldsymbol{\lambda}, \boldsymbol{\theta})$ is given in Equation 6.11. The ML estimates $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are determined in the fit to data by maximising the likelihood, or equivalently minimising the $-\ln L$. The values of the parameters before (after) the minimisation procedure are indicated as the *pre-fit* (*post-fit*) values. The pre-fit values of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ parameters are set to *1* and *0*, respectively. The *profile likelihood ratio* is then evaluated in order to be used in the test statistics,

$$\lambda(\mu_{\text{sig}}) \equiv \frac{L(\mu_{\text{sig}}, \hat{\hat{\boldsymbol{\nu}}}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}_{\text{sig}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})}. \tag{6.15}$$

Here $(\hat{\hat{\boldsymbol{\nu}}}, \hat{\hat{\boldsymbol{\theta}}})$ values in the numerator indicate their MLE values for a given value of the parameter μ_{sig}, called conditional maximum-likelihood of $\boldsymbol{\nu}, \boldsymbol{\theta}$ parameters, thus they are the function of μ_{sig}. The denominator is the unconditional maximum likelihood, where all set of parameters are equal to their MLE values. The method of choosing the particular values of the NPs for a given value of the parameter μ_{sig} is often called *profiling*. From the definition of profile likelihood given in Equation 6.15, test statistic can be defined as follows

$$q_{\mu_{\text{sig}}} = -2 \ln \lambda(\mu_{\text{sig}}), \tag{6.16}$$

with $\lambda(\mu_{\text{sig}})$ close to unity indicates good agreement of data and the expected value of μ_{sig}. In high energy physics, the contribution of the signal process is assumed to be non-negative, thus estimator of μ_{sig} must be non-negative. The use of an effective estimator, which can take negative values is more convenient because this will allow to model the distribution of test statictics and avoid the complications of a boundary at $\mu_{\text{sig}} = 0$. The alternative likelihood ratio then is given by

$$\tilde{\lambda}(\mu_{\text{sig}}) = \begin{cases} \frac{L(\mu_{\text{sig}}, \hat{\hat{\boldsymbol{\nu}}}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}_{\text{sig}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0 \\ \frac{L(\mu_{\text{sig}}, \hat{\hat{\boldsymbol{\nu}}}, \hat{\hat{\boldsymbol{\theta}}})}{L(0, \hat{\boldsymbol{\nu}}(0), \hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0\,. \end{cases} \tag{6.17}$$

This indicates that if the observed data yields are $\hat{\mu} < 0$, then the best level of agreement between data and any value of μ is at $\mu = 0$. The test statistics in Equation 6.16 can be redefined

based on the likelihood ratio given in Equation 6.18

$$\tilde{q}_{\mu_{\text{sig}}} = -2\ln\tilde{\lambda}(\mu_{\text{sig}}) = \begin{cases} \frac{-2\ln L(\mu_{\text{sig}},\hat{\hat{\nu}},\hat{\hat{\theta}})}{L(\hat{\mu}_{\text{sig}},\hat{\nu},\hat{\theta})}, & \hat{\mu} \geq 0 \\ \frac{-2\ln L(\mu_{\text{sig}},\hat{\hat{\nu}},\hat{\hat{\theta}})}{L(0,\hat{\nu}(0),\hat{\hat{\theta}}(0))} & \hat{\mu} < 0. \end{cases} \tag{6.18}$$

An important special case of test statistic $\tilde{q}_{\mu_{\text{sig}}}$ is used to test $\mu = 0$ where $\mu \geq 0$ is assumed. Using the definition given in 6.18, the test statistic for background-only hypothesis q_0 ($\mu = 0$) is given as follows:

$$q_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0\,. \end{cases} \tag{6.19}$$

where the profile likelihood ratio for $\mu = 0$ is $\lambda(0)$ as in Equation 6.15. This way, the lack of agreement between data and background-only hypothesis is considered only when $\hat{\mu} > 0$, while the $\hat{\mu}$ values lower than zero may also represent the evidence against background-only hypothesis, however also this kind of discrepancy does not necessarily refer that the data contain signal events, rather can indicate fluctuations. In order to quantify the disagreement between data with observed value q_0 and hypothesis of $\hat{\mu} = 0$, *p-value* is obtained as

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0)\,\mathrm{d}q_0\,, \tag{6.20}$$

where $f(q_0|0)$ is the PDF of the test statistic q_0 under the background-only hypothesis. In the case of the signal-plus-background hypothesis, the test statistics can be written as

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu\,. \end{cases} \tag{6.21}$$

As in the of background-only testing, the *p-value* of the test can be obtained as follows:

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu)\,\mathrm{d}q_\mu\,. \tag{6.22}$$

6.6 Approximate distributions and expected significance

The test statistic given in Equation 6.16 can be interpreted as an outcome of the experiment, with a real number. When the experiment repeated many times, the value of the test statistic will differ, and therefore the test statistic will have a distribution. This distribution allows then to obtain the expected p-value, the median significance and the limits. One way to compute these parameters is generating *pseudo-experiments* with MC simulations, based on the test statistic distribution, and using the properties of the generated distribution to obtain them. However, this method is not very efficient due to being time consuming and since it randomizes the

measurements and auxiliary measurements, it is usually not ideal. Alternatively, it is possible to use an approximation method for profile likelihood ratio, which will allow to evaluate the required distributions and the parameters. The Wald [151] approximation denotes that in the case of a large data sample and a single parameter of interest, the statistical test has form of

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + O(1/\sqrt{N}), \tag{6.23}$$

where the $\hat{\mu}$ can be described as a Gaussian distribution[1] with the μ' mean value, which can be zero for background-only or non-zero for an upper limit, σ is the standart deviation, and N serve as the data set size. Here the σ of the $\hat{\mu}$ distribution can be acquired from the covariance matrix of the estimators in all parameter space, $V_{ij} = cov[\hat{\theta}_i, \hat{\theta}_j] = E[\partial d^2/\partial\theta_i\partial\theta_j]$ where $\hat{\theta}$ values can represent both signal strength and nuisance parameters.

Similarly for a sufficiently large data sample, the distribution of the test statistic, given in Equation 6.24, obeys a distribution of *non-central chi-square* with one degree of freedom

$$f(q_\mu;\Lambda) = \frac{1}{2\sqrt{q_\mu}}\frac{1}{\sqrt{2\pi}}\left[exp\left(-\frac{1}{2}(\sqrt{q_\mu}+\sqrt{\Lambda})^2\right) + exp\left(-\frac{1}{2}(\sqrt{q_\mu}-\sqrt{\Lambda})^2\right)\right], \tag{6.24}$$

where the parameter of Λ refers to non-centrality term of

$$\Lambda = \frac{(\mu-\mu')^2}{\sigma^2}. \tag{6.25}$$

The results given above can be generalised to more than one parameter. Derived formulae above requires the parameter σ and $\hat{\mu}$ of the distribution to be evaluated. The *Asimov dataset* can be used as a result of the approximations above. Asimov dataset is defined in order to evaluate the ML estimators of all parameters to obtain their expected values. This is done through setting the derivatives of $\ln\mathcal{L}$ (Equation 6.14) with respect to parameters to zero,

$$\frac{\partial\ln\mathcal{L}}{\partial\theta_j} = \sum_{i=1}^{N}\left(\frac{n_i}{\nu_i}-\right)\frac{\partial\nu_i}{\partial\theta_j} + \sum_{i=1}^{M}\left(\frac{m_i}{u_i}-\right)\frac{\partial u_i}{\partial\theta_j} = 0\,. \tag{6.26}$$

where $\nu_i = E[\nu_i] = \mu' s_i + b_i$. This condition applies if the Asimov data, n_i and m_i are equal to their expected values:

$$n_i = E[n_i] = \nu_i = \mu' s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}), m_i = E[m_i] = u_i(\boldsymbol{\theta}). \tag{6.27}$$

Afterwards the Asimov likelihood and corresponding profile likelihood ratio are obtained as

[1] In the case of large data samples drawn randomly, the distribution of means of the observables would follow a normal distribution even if the measured quantities are not normally distributed. This is a consequence of the *central limit theorem*.

follows since it satisfies the $\hat{\mu} = \mu'$ condition,

$$-2\ln\lambda(\mu) \approx \frac{(\mu-\mu')^2}{\sigma^2} = \Lambda, \tag{6.28}$$

where

$$\sigma^2 = \frac{(\mu-\mu')^2}{q_\mu}. \tag{6.29}$$

Here μ corresponds to the hypothesis being tested. In case of background-only hypothesis testing, $\mu' = \hat{\mu} = 0$ the expected limit $\mu_{\text{up}}^{\text{med}}$ corresponds to the median of $f(q_\mu|0)$ and it can be obtained by solving the following equation

$$\text{med}[Z_0|0] = \sqrt{q_0}. \tag{6.30}$$

6.7 Systematic uncertainties and interpretations

The Gaussian constraint terms in Equation 6.14, centered on 0 and with σ of 1, are comparable by transforming the $\boldsymbol{\theta}$ parameters as follows:

$$\boldsymbol{\theta}' = \frac{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\sigma}}}, \tag{6.31}$$

where $\boldsymbol{\sigma}$ corresponds to the initial widths of the $\pm1\sigma$ variations associated to the systematic variation represented by $\boldsymbol{\theta}$. The normalisation of the NPs to their values given in Equation 6.31, is done in order to simplify the comparison between pre- and post-fit values and uncertainties. It is calculated in each bin and for each nuisance parameter. The effect of systematic uncertainty in each bin is obtained for ±1 values of NP, and then bin-by-bin linear interpolation of distribution performed to extrapolate for any value of $\boldsymbol{\theta}$. Likelihood fit described above can change the background prediction if the best-fit $\boldsymbol{\theta}$ is different from its nominal value, or can reduce the uncertainty on background. The likelihood procedure can adapt these effects in the following ways:

- Changes in the post-fit NP central values are referred to as *pulling*. The pull can be described as the difference between the pre-fit and the post-fit central values of the NP normalized to its pre-fit uncertainty. The central values may be pulled during the fit procedure to data to maximise the overall agreement of the background prediction. However, if the pull exists without the constrain in the error, this situation might be worth to check.

- Decrease of the uncertainty value of the parameter are referred to as *constraint*. The constraint is described as the post-fit over pre-fit uncertainty ratio of a corresponding NP, which serves to constrain the value of the NP to its measured value. Such cases

demonstrate that the initial prior on the uncertainty is much larger than the prediction, thus the uncertainty may be reduced to be more compatible with the data.

- The correlations between NPs and signal strength are obtained from covariance matrix, and built by the fit. Even if the sources are independent before the fit, correlations may occur with improved knowledge of the parameters. Correlations can reduce the number of degrees of freedom in the fit, and can lead to misinterpretation of the results. Additionaly (anti-)correlations may reduce total post-fit uncertainty.

This prospective of the profiling likelihood mechanism to reduce the overall impact of systematic uncertainties on the results can be interpreted as a general benefit of the profile likelihood method. The investigation of the impact of the NP on the signal strength error is crucial, thus the *ranking method* may be benefited. In this method for each NP, four individual fits are performed, where the NP is fixed to its pre-fit $\pm\sigma$ value and its post-fit $\pm\sigma$ values (all the other NPs are allowed to float from their initial values to take into account correlations), and compute the best-fit signal strength value. The difference in the signal strength $\Delta\mu^{\pm} = \mu_{\pm 1\sigma} - \hat{\mu}$ for each NP is ranked by largest impact in signal strength, and plotted. Due to the correlation effects, the positive and negative variations may not be identical resulting in asymmetrical representation.

Further investigation may be performed by the *breakdown method*. The NPs are generally merged into categories in order to evaluate the impact of each group. The impact is calculated by performing a fit where the NPs in the group are fixed to their best-fit values, and then subtracting the resulting uncertainty from the nominal uncertainty in quadrature. The calculation can be performed in two ways: parallelized or sequential computation. In parallelized method, the calculation is done independently in each group, whereas in sequential method, the fit is evaluated by fixing NPs from the other groups one after the other, and compute the difference in quadrature. In the first method, the other NPs are allowed to free-float to recover some of the impacts of the NPs in the group, therefore this might lead some misinterpretations. By construction, the sums of all uncertainties must give the total impact of the uncertainties on the signal strength, however the order of the fits may cause the individual values to change.

CHAPTER 7

Search for the Higgs boson production with a top quark pair in multilepton channels

The search for the Higgs boson production in association with a top-antitop quark pair at a center-of-mass energy of $\sqrt{s} = 13$ TeV using a dataset corresponding to an integrated luminosity of 79.8 fb^{-1} [152] is presented in this chapter through the multilepton final states, which corresponds to decay of the Higgs boson into a pair of WW^*, ZZ^* or $\tau^+\tau^-$. The main background contributions arise from $t\bar{t}W$, $t\bar{t}Z/\gamma^*$, and diboson (VV) productions, as well as from $t\bar{t}$ production with additional light leptons (electron, muon) from heavy-flavour hadron decays, misidentified jets or photon conversions (together referred to as *non-prompt leptons*), and other processes where the electron charge is incorrectly assigned or where jets are misidentified as τ_{had} candidates.

Section 7.1 presents the dataset and MC simulation samples used for this analysis. It is followed by Section 7.2, where the event and object preselection is defined in order to categorise the channels orthogonally, as well as the tight lepton selection used by the channels. In Section 7.3, the definition of control and signal regions are given. Section 7.4 discusses the background estimations with special focus on light lepton estimations. Section 7.5 gives the various sources of systematic uncertainties included in the analysis and the procedures to estimate them. The analysis results has been presented two times to the physics community. Therefore, the final expected and observed combined fit results are given in Section 7.6, while the first results are given in Appendix C. Finally, several numbers of cross-checks are performed to validate the fit results and some of them are given in Section 7.7.

My contribution in this group begins with the development of the analysis framework and migrating it to a new ATLAS release. Shortly after I became the coordinator of the analysis framework, which is used by the whole multilepton group. Following closely the recommendations from different physics groups, implementing various Athena[1] based

[1] Athena is the name of the ATLAS software framework that manages almost all ATLAS production workflows

computation tools and managing the small framework group for the sample production were some of my main responsibilities. My major contribution was to develop the combined fit model of the signal strength and the statistical interpretation of the results. In this role, I worked on the definition of the fit model described in Section 7.6 and its implementation in the statistical framework called TRExFitter, which is available only for the collaboration. I did extensive work on the preparetion of the results given in Section 7.6 and Appendix C, as well performed the various fit cross-check studies together with the other main analyzer Zhi Zheng. I also performed the studies documented in Appendix D. The most crucial cross check studies are shown in Section 7.7. In addition to the work on the analysis framework and on the results, I had several minor contributions to the group such as the estimation of breakdown of the uncertainties, investigation of the impact of the different binning setups on the results, derivation of PDF uncertainties for $t\bar{t}W$ background and many more. The implementation of Matrix Element Method to improve the 2ℓSS channel purity, described in Section 7.8, is my own work.

7.1 Data and simulation

This section gives an overview of the collision dataset and MC samples used in this analysis. All set of samples are being prepared with xAOD format [154] and later to the more specialized format derived xAOD (*DxAOD*) using *HIGG8D1* derivation framework. Derivation framework provides a particular reduction for the events with leptons in the final states. The selection is applied in three steps: smart slimming (remove unnecessary variables), thinning (remove entire unnecessary object from events), augmenting (include special variables) and skimming (removing whole events) as follows:

- Events passing the selection of at least two light leptons (the highest transvers momentum lepton to be $p_T > 15$ GeV and the second one is with $p_T > 5$ GeV, within $|\eta| < 2.6$)
- Events passing the selection of at least one light lepton (with $p_T > 15$ GeV and $|\eta| < 2.6$), and two hadronic taus (with JetBDTSigLoose, $p_T > 15$ GeV and one or three associated tracks).

7.1.1 Collision dataset

The analysis uses dataset corresponding to an integrated luminosity of 79.8 fb^{-1} of proton-proton collision data recorded by the ATLAS detector at $\sqrt{s} = 13$ TeV during 2015 and 2017 data taking periods, with a peak instantaneous luminosity of $\mathcal{L} = 2.1 \times 10^{34}\ cm^{-1}s^{-1}$ and the Inner B-Layer (IBL) on. The dataset has been collected with a bunch crossing of 25 ns and verifying data quality criteria defined by the *good run list* (GRL) procedure [56].

such as event generation, simulation, reconstruction and derivation production [153].

7.1.2 Monte Carlo samples

$t\bar{t}H$ multilepton final state phase space is populated by many different physics processes. Therefore, besides the signal process, there are many background processes that need to be included. A summary of the MC samples used to simulate those processes, as well as the additional samples used to estimate the systematic uncertainties (in parentheses), are given in Table 7.1 [152]. The masses of the top quark and SM Higgs boson were set to 172.5 GeV and 125 GeV for all MC samples. The ATLAS detector response is simulated by using the GEANT4 package, and the pile-up reweighing procedure is applied to reflect the mean number of additional interactions observed in collected data.

$t\bar{t}H$ Signal

The simulation of nominal signal sample $t\bar{t}H$ is done using NLO generator POWHEG-BOX v2, with the NNPDF3.0 NLO PDF set computed by MADGRAPH5_AMC@NLO, interfaced to $t\bar{t}$ Pythia8 for parton showering using A14 tune [155] and for hadronisation of the particles. It includes off-shell Z and γ^* contributions with $m(\ell\ell) > 5$ GeV. Model parameter h_{damp} is set to $1.5 \times (2m_t + m_H)/2 = 352.5$ GeV, which accounts for missing contributions in ME to PS matching, effectively regulates the high-p_T radiation beyond the LO Feynman diagram. The renormalisation and factorisation scale choice μ_R and μ_F are set to geometric mean of the transverse energies of the top quark, antitop quark and the Higgs boson. The modelling uncertainties are derived using the nominal sample with the uncertainties on the A14 tune. The branching ratios are calculated using HDECAY [156]. The overall $t\bar{t}H$ cross section is calculated as 507^{+35}_{-50} fb by accounting for the NLO contributions in QCD with the leading NLO EWK corrections (such as $O(\alpha_s^2\alpha^2)$) [157–162].

$t\bar{t}W$ background

The nominal $t\bar{t}W$ is generated using SHERPA 2.2.1 [163] with the NNPDF3.0 NLO PDF set. The ME is estimated including up to one additional parton at NLO and up to 2 additional partons at LO with COMIX [164] and OPENLOOPS [165]. The merging of the PS and ME is done with the SHERPA parton shower [109] using the MEPS@NLO prescription [166] at a merging scale of 30 GeV. The overall cross section for $t\bar{t}W$ is 601 ± 76 fb at NLO in QCD with the leading NLO electroweak corrections ($O(\alpha_s^2\alpha^2)$) [22, 167, 168]. The choice of renormalisation and factorisation scales is $\mu_R = \mu_F = H_{\text{T}}/2$, where H_{T} is represents the scalar sum of the transverse masses $\sqrt{p_{\text{T}}^2 + m^2}$ of all final state particles and obtained from same sample as nominal. In order to account for QCD and EWK corrections at NLO, additional scaling factor is applied to the inclusive cross section value. The related work given in Ref. [169] has shown that NLO QCD corrections to $t\bar{t}W$+1jet production can be non-negligible due to the qg initiated production contributions and therefore the scale factor of 1.11 is estimated using dedicated samples generated with SHERPA 2.2.5 with the MEPS@NLO prescription. The corrections for accounting the NLO EWK effect are calculated based on the study [170], which has showed that

sub-leading NLO EWK corrections are as well non-negligible because of the large $O(\alpha_s\alpha^3)$ (NLO_3) term driven by the $t\bar{t}W$+1-jet diagrams with a Higgs boson exchanged in the t-channel and becoming increasingly important in LHC phenomenology. Therefore, the corresponding scale factor of 1.09 is calculated as follows:

$$\begin{aligned} \delta(NLO_2) &= -0.042\,, \\ NLO_2 = LO_1 \times \delta(NLO_2) &= -0.042 \times 363 = -15.2460\ fb\,, \\ k^{\text{Missing EWK}} = (\text{LO} + \text{QCD})/(LO_1 + NLO_1 + NLO_2) &= 577/(544 - 15.2460) = 1.09124\,. \end{aligned} \tag{7.1}$$

Here the $(N)LO_i$ are representing the corrections on different perturbative orders, LO_1 ($\alpha_s^2\alpha$), NLO_1 ($\alpha_s^3\alpha$), NLO_2 ($\alpha_s^2\alpha^2$) and LO ($\alpha_s^2\alpha + \alpha_s\alpha + \alpha^2$). The study [170] quotes 1.06 of scaling factor for EWK corrections, on top of the NLO QCD corrections. Due to the fact that in the analysis, the normalisation is done with respect to YR4 [22], not to the NLO QCD, the recalculation is done by removing YR4 EWK correction before applying the full one. Therefore, after applying these two scaling factors, the inclusive cross section used to normalise the $t\bar{t}W$ sample is obtained 727 ± 92 fb[2]. As the $t\bar{t}W$ normalisation will be estimated from data (see Section 7.4), cross section uncertainty is not included to the fit model. QCD radiation is assessed by comparing the nominal $t\bar{t}W$ prediction with an alternative sample, which is generated at NLO (thus at lower order than for the nominal sample) with the MadGraph5_aMC@NLO 2.2.1 generator using the same scale choice and PDF set as for the nominal sample, and interfaced to Pythia 8.2 in combination with the A14 tune.

Other backgrounds

The production of MC samples for $t\bar{t}Z/\gamma^*$, VV , and $t\bar{t}$ follow Ref. [171–173]. The $t\bar{t}$ sample is generated with Powhegv2.0 and merged with Pythia8 for the parton showering using A14 tune and hadronisation, where the overall $t\bar{t}$ cross section is 832 pb. For the $t\bar{t}Z/\gamma^*$ sample, the inclusive $pp \to t\bar{t}\ell^+\ell^- + X$ ME is computed at NLO in QCD and EWK couplings, including off-shell Z and γ^* contributions with 1 GeV $< m(l^+l^-) <$ 5 GeV and $m(l^+l^-) \geq$ 5 GeV, the cross section is calculated as 162 ± 21 fb.

Dedicated samples are produced for $t \to Wb\gamma^*(\to l^+l^-)$ radiative decays in $t\bar{t} \to W^+bW^-\bar{b}l^+l^-$ process is using the ME at LO and with $m(l^+l^-) > 1$ GeV, and is called *rare top decay*. In this sample the photon can be radiated from the top quark, the W boson, or the b-quark. The $t\bar{t}Z/\gamma^*$ and $t\bar{t} \to W^+bW^-\bar{b}l^+l^-$ samples are combined and together form the $t\bar{t}Z/\gamma^*$ (*high mass*) sample. Non-trivial overlaps were found between the $t\bar{t}$, rare top and the low invariant mass $t\bar{t}\ell\ell$ samples. Therefore, the study has been performed to avoid both double-counting of contributions and uncovered regions of phase space when combining them. The contribution from photon conversions such as ($\gamma^* \to l^+l^-$) with $m(l^+l^-) <$ 1 GeV are modelled by QED multiphoton radiation via the PS in an inclusive $t\bar{t}$ sample and is referred to

[2] The theoretical uncertainties are not rederived based on the additional corrections but are scaled proportionally to the scaling factors applied.

Process	Generator	ME order	Parton shower	PDF	Tune
$t\bar{t}H$	POWHEG-BOX [174, 175]	NLO	PYTHIA 8 [176]	NNPDF3.0 NLO [177]/ NNPDF2.3 LO [92]	A14
	(POWHEG-BOX)	(NLO)	(HERWIG7)	(NNPDF3.0 NLO/ MMHT2014 LO [93])	(H7-UE-MMHT)
$tHqb$	MG5_aMC [101]	LO	PYTHIA 8	CT10 [178]	A14
tHW	MG5_aMC	NLO	HERWIG++	CT10/ CTEQ6L1 [179, 180]	UE-EE-5
$t\bar{t}W$	SHERPA 2.2.1	MEPS@NLO	SHERPA	NNPDF3.0 NNLO	SHERPA default
	(MG5_aMC)	(NLO)	(PYTHIA 8)	(NNPDF3.0 NLO/ NNPDF2.3 LO)	(A14)
$t\bar{t}(Z/\gamma^*)$	MG5_aMC	NLO	PYTHIA 8	NNPDF3.0 NLO/ NNPDF2.3 LO	A14
	(SHERPA 2.2.0)	(LO multileg)	(SHERPA)	(NNPDF3.0 NLO)	(SHERPA default)
$t\bar{t} \to W^+bW^-\bar{b}l^+l^-$	MG5_aMC	LO	PYTHIA 8	NNPDF3.0 LO	A14
tZ	MG5_aMC	LO	PYTHIA 6 [181]	CTEQ6L1	Perugia2012
tWZ	MG5_aMC	NLO	PYTHIA 8	NNPDF2.3 LO	A14
$t\bar{t}t$, $t\bar{t}t\bar{t}$	MG5_aMC	LO	PYTHIA 8	NNPDF2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_aMC	LO	PYTHIA 8	NNPDF2.3 LO	A14
$t\bar{t}$	POWHEG-BOX	NLO	PYTHIA 8	NNPDF3.0 NLO/ NNPDF2.3 LO	A14
Single top (t-, Wt-, s-channel)	POWHEG-BOX [182–184]	NLO	PYTHIA 8	NNPDF3.0 NLO/ NNPDF2.3 LO	A14
VV, $qqVV$, VVV	SHERPA 2.2.2	MEPS@NLO	SHERPA	NNPDF3.0 NNLO	SHERPA default
$Z \to l^+l^-$	SHERPA 2.2.1	MEPS@NLO	SHERPA	NNPDF3.0 NLO	SHERPA default

Table 7.1: The MC generator configurations used for signal and background processes. In this table the names of the generators, the order of the strong coupling constant of the perturbative calculation, PS generators, PDF evaluation methods and the underlying-event tune of the PS generator is given for corresponding physics process, respectively. All samples include leading-logarithm photon emission, either modelled by the parton shower generator or by PHOTOS [185].

as $t\bar{t}\gamma^*$ (*low mass*).

Diboson (VV) process, where V represents the electroweak boson production of W or Z/γ^*, are generated with SHERPA v2.2.2 using NNPDF30NNLO PDF set. Processes with four leptons, three leptons and one neutrino, or two leptons and two neutrinos are produced with up to one extra parton at NLO and up to 3 extra partons included at LO. The $W^\pm W^\pm jj$ process (with non-resonant contributions) were generated at LO with up to one extra parton, separately for QCD-induced $\left(O(\alpha^6_{\mathrm{EM}})\right)$ and VBS-induced $\left(O(\alpha^6_{\mathrm{EM}})\right)$ production. Additional samples for VBS-induced $qq \to 3\ell\nu jj$ and $qq \to 4\ell$ and loop-induced $gg \to WZ^{(*)}/ZZ^{(*)}$ processes are also included with the same configuration. The backgrounds, which has less contribution to the $t\bar{t}H$ multilepton phase space, (tZ, $t\bar{t}t\bar{t}$, $ttWW$, WtZ, VVV, $tt\bar{t}$, $tHjb$ and WtH) are normalised using their NLO theoretical cross sections and assigned a 50% normalisation uncertainty.

7.2 Object and event preselection

All analysis channels share a common object and event preselection in order to ensure the orthogonality between the channels and further selections are applied to each channel individually to improve the sensitivity of $t\bar{t}H$ signal.

7.2.1 Object preselection

Physics objects are reconstructed and identified based on the ATLAS standard procedures using various methods described in Chapter 5.

Electron candidates, which are reconstructed from energy clusters along with associated tracks in the ID, are accepted if they satisfy the condition of $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$[3]. The identification WP is choosen to be LooseLH and no isolation is required. To improve the efficiency of leptons originating from hard scatter, further requirements are applied on track-primary vertex association variables $|d_0|/\sigma_{d_0}$ and $|\Delta z_0 \sin\theta_\ell|$.

Muon candidates are reconstructed by combining ID tracks with track segments in the MS. The preselection cut is $p_T > 10$ GeV and $|\eta| < 2.5$, while the transverse impact parameter requirement is tighter than electrons. Passing loose identification is required without any isolation selection.

The reconstruction of jets are based on the topological clusters formed from energy deposits in the calorimeters, exploiting the anti-k_t algorithm with a radius parameter $R = 0.4$. The candidates passing the *jet cleaning algorithm* (Section 5.3.2) are considered. Additionaly, a JVT cut is applied to the jets with $p_T < 60$ GeV and $|\eta| < 2.4$ to improve the acceptance of jets originating from primary vertex through removing the contribution from pile-up. Jets containing b-hadrons are identified through MV2c10 algorithm, as described in Section 5.3.3. The WP used in the analysis corresponds to an average efficiency of 70% with $p_T > 20$ GeV and $|\eta| < 2.5$. Hadronically decaying τ leptons (τ_{had}) are reconstructed from clusters in the calorimeters and associated ID tracks. One or three associated tracks with a total charge of ± 1 is required. The selections of the candidates are $p_T > 25$ GeV and $|\eta| < 2.5$ and being originated form the primary vertex. A BDT is trained to ameliorate the discrimination between the jet backgrounds and τ_{had} candidates, and the Medium WP is choosed due to its higher efficiency in the analysis.

The missing transverse momentum p_T^{miss} (magnitude of $E_{\text{T}}^{\text{miss}}$) cut is not required in the preselection while its selection is used later for certain control or signal regions.

The two of the reconstructed objects, which are considered very close to each other geometrically, can overlap and yield duplicated objects or mis-reconstructed object such as a lepton is reconstructed as jet or a lepton inside the b-jet coming from a hadron decay. Thus, a set of criteria is applied depending on the angular distances between the objects and one of the them is preserved. Table 7.2 shows the criteria applied to each object in the analysis in order to avoid this phenomena.

[3] Candidates in the transition region $1.37 < |\eta_{\text{cluster}}| < 1.52$ are rejected.

Keep	Remove	Cone size (ΔR)
electron	electron (low p_T)	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	min(0.4, 0.04 + 10[GeV]/p_T (muon))
electron	tau	0.2
muon	tau	0.2
tau	jet	0.3

Table 7.2: Summary of the overlap removal procedure between electrons, muons, hadronically decaying taus, and jets.

7.2.2 Event preselection and categorisation

Depending on the number and flavour of the lepton candidates after applying the pre-selection defined in Section 7.2.1, the analysis is categorised into six channels. The most sensitive channels are then further split into signal and control region categories, where the purity of $t\bar{t}H$ signal events and different background process events are more populated, respectively. The orthogonality between channels is preserved in a way that each event only contributes to one channel. The six channels are:

- 2ℓSS: two light leptons with same-sign charge and no τ_{had} candidates;
- 3ℓ: three light leptons and no τ_{had} candidates;
- 4ℓ: four light leptons;
- 1ℓ+$2\tau_{\text{had}}$: one light lepton and two opposite-charge τ_{had} candidates;
- 2ℓSS+$1\tau_{\text{had}}$: two light leptons with same-sign charge and one τ_{had} candidates;
- 3ℓ+$1\tau_{\text{had}}$: three light leptons and one τ_{had} candidates.

Following the primary vertex selection[4], the events are required to pass certain un-prescaled triggers as summarised in Table 7.3 and 7.4 with their corresponding p_T thresholds. Channels having at least two light leptons are required to pass dilepton triggers, where the channels with one light lepton required to pass single lepton triggers.

[4] The primary vertex with the largest $\sum p_T^2$ of associated tracks in an event [186].

	Dilepton triggers (2015)
$\mu\mu$ (asymm.)	HLT_mu18_mu8noL1
ee (symm.)	HLT_2e12_lhloose_L12EM10VH
$e\mu$, μe (~symm.)	HLT_e17_lhloose_mu14
	Dilepton triggers (2016)
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
ee (symm.)	HLT_2e17_lhvloose_nod0
$e\mu$, μe (~symm.)	HLT_e17_lhloose_nod0_mu14
	Dilepton triggers (2017)
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
ee (symm.)	HLT_2e24_lhvloose_nod0
$e\mu$, μe (~symm.)	HLT_e17_lhloose_nod0_mu14

Table 7.3: List of lowest p_T-threshold, un-prescaled dilepton triggers used for 2015-2017 data taking.

	Single lepton triggers (2015)
μ	HLT_mu20_iloose_L1MU15, HLT_mu50
e	HLT_e24_lhmedium_L1EM20VH, HLT_e60_lhmedium, HLT_e120_lhloose
	Single lepton triggers (2016, 2017)
μ	HLT_mu26_ivarmedium, HLT_mu50
e	HLT_e26_lhtight_nod0_ivarloose, HLT_e60_lhmedium_nod0, HLT_e140_lhloose_nod0

Table 7.4: List of lowest p_T-threshold, un-prescaled single lepton triggers used for selecting 1ℓ+$2\tau_{\text{had}}$ events in 2015-2017 data taking.

For the di-electron triggers the p_T thresholds on the two electrons are 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017, while for the di-muon triggers the p_T thresholds on the leading highest p_T (sub-leading, second highest p_T) muon are 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016 and 2017. For the electron+muon triggers, the p_T thresholds on the electron (muon) are 17 GeV (14 GeV) for all datasets. Events in the one lepton channel are required to pass a single-electron (single-muon) trigger with $p_T > 24$ (20) GeV in 2015, while for 2016 and 2017, the lepton p_T threshold are increased to 26 GeV. Additionaly, selected electrons or muons are required to match with the corresponding leptons reconstructed by the trigger, with $\Delta R < 0.15$, and their p_T must pass the trigger p_T threshold by 1 GeV or 2 GeV.

7.2.3 Tight lepton definition

Tight requirements on the leptons are defined on top of the general object pre-selection in order to suppress reducible background contributions. Detailed description of different background sources and their estimation methods are given in Section 7.4.1.

A BDT discriminant based on the isolation and b-tagging variables, PROMPTLEPTONVETO (PLV), is trained in order to reduce the non-prompt light leptons coming from heavy-flavour decays, which are produced with a displacement from the interaction point [187]. Two different WPs are defined for electrons and muons as given in Table 7.5.

Physical processes that mistakenly yield to two prompt same-charge electrons are further suppressed by a BDT discriminant based on calorimeter and tracking quantities [122]. A cut on the BDT output is defined for electrons to reject this type of backgrounds. The 95% efficiency for electrons with correct charge assignment is obtained with a rejection factor of ~17 for electrons with misassigned charge.

Another type of background is due to photon conversions, where an electron pair originates from the radiated photon and one is reconstructed as prompt electron. Conversions are classified in two categories: *material* and *internal* conversions. Material (internal) conversions have a reconstructed displaced vertex with radius $r > 20$ mm ($r < 20$ mm) with respect to the beamline, and include the track associated to the electron. Therefore, two variables ($m_{trk-trk,CV}$ and $m_{trk-trk,PV}$) are defined in order to reject the conversion electrons, along with the AMBIGUITY BIT variable, which is used to surpress material conversion electrons. Related studies and the definition of the variables are summarised in Appendix B.

The full definition of tight light leptons are given in Table 7.6 and different selections of leptons in the analysis are summarised in Table 7.5.

	electron	muon
ID	TightLH && ambiguityType == 0	Medium
Isolation	FixedCutLoose && PLV < -0.7	FixedCutLoose && PLV < -0.5
conv. suppression	$\Delta R_{ll} > 0.5$ && $\lvert\eta_e\rvert < 2.0$	
	!ExtCo && !IntCo	
QmisID MVA	> 0.7	
impact parameter	$\lvert d0\rvert/\sigma(d0) < 5$	$\lvert d0\rvert/\sigma(d0) < 3$
	$z_0 \sin\theta < 0.5$ mm	$z_0 \sin\theta < 0.5$ mm

Table 7.5: Summary of light lepton selection criteria defining the tight WPs.

	e					μ			
	L	L†	L*	T	T*	L	L†	L*	T
Isolation	No	Yes				No	Yes		
Non-prompt lepton BDT	No	Yes				No	Yes		
Identification	Loose	Tight				Loose			
Charge mis-assignment veto	No	Yes				N/A			
Tranverse impact parameter significance $\lvert d_0\rvert/\sigma_{d_0}$	< 5					< 3			
Longitudinal impact parameter $\lvert z_0 \sin\theta\rvert$	< 0.5 mm								

Table 7.6: Loose (L), loose and isolated (L†), less loose (L*), tight (T) and very tight (T*) light lepton definitions. Selections for tighter light leptons are applied in addition to the loose selections.

7.3 Signal and control region definitions

Following the selection and categorisation of events into channels explained in Section 7.2.2, each channel further applied selections to define the *Signal Regions* (SRs), where the $t\bar{t}H$ signal purity is higher. Furthermore, in order to increase the confidence on the background estimations, some channels make use of *Control Regions* (CRs). The CRs must be orthogonal to the SRs, such that no events that satisfy the requirements of the SRs can enter to the CRs. Some of the channels exploit multivariate techniques (2ℓSS, 3ℓ and 1ℓ+$2\tau_{\text{had}}$) for both SRs and CRs definitions, while some use cut optimisation methods (2ℓSS+$1\tau_{\text{had}}$ and 3ℓ+$1\tau_{\text{had}}$) with event reconstruction variables (4ℓ). A comprehensive summary for each channel is given in the following paragraphs.

2ℓSS channel

In 2ℓSS channel, five different event categories are defined to control the different source of backgrounds. There are three regions called *low jet multiplicity (LJ)*, which provide high statistics CRs enriched in non-prompt leptons from semileptonic b-decays. Events of this CRs are required to contain either two or three reconstructed jets. Events in these three categories are classified according to the sub-leading lepton flavour and those with a sub-leading electron are split into one or two b-jets (2ℓLJ(e1), 2ℓLJ(e2), and 2ℓLJ(μ)). The remaining two control regions are dedicated to photon conversions, either an internal conversion (2ℓIntC) or a material conversion (2ℓMatC) by requiring at least one of the electron to be conversion electron.

The signal region is optimised by using two different BDTs (Appendix A), one is trained to separate $t\bar{t}H$ from $t\bar{t}$ (non-prompt background) and the other to separate $t\bar{t}H$ from $t\bar{t}V$ (V=W or Z). The regions, which are enriched in $t\bar{t}H$, $t\bar{t}W$ and $t\bar{t}$ contributions, are defined by a two dimensional cut on BDT outputs. Additionaly, $t\bar{t}$ region is divided according to flavour of sub-leading leptons. In order to exploit the charge asymmetry in the $t\bar{t}W$ process, the regions

are further split into categories according to the lepton charge, and therefore six different background categories are defined (2ℓttW+, 2ℓttW-, 2ℓtt(e)+, 2ℓtt(e)-, 2ℓtt(μ)+, and 2ℓtt(μ)-). The signal region as well is split into two categories depending on the charge of the leptons (2ℓttH+, 2ℓttH-). These categories including signal and control regions are referred as *high jet multiplicity (HJ)* regions. In many cases, categories are split according to the flavour of the sub-leading lepton because this lepton is more likely to be a non-prompt lepton and therefore splitting provides a better control on different non-prompt background flavours. This yield to 13 categories in the 2ℓSS channel.

3ℓ channel

The 3ℓ channel exploit a five-dimensional multinomial BDT targetting five different event classifications: $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$ and diboson. The outputs of the BDTs are used to define five categories (3ℓttH, 3ℓttW, 3ℓttZ, 3ℓtt, and 3ℓVV). Since some of the non-prompt lepton processes are as well contributing to the 3ℓ channel, two additional regions are defined to control the internal and material conversion backgrounds: (3ℓIntC and 3ℓMatC) by loosining the lepton definitions as given in Table 7.6 and requiring at least one lepton to be conversion lepton. Finally seven categories are defined in the 3ℓ channel.

4ℓ channel

In 4ℓ channel, two categories are determined according to the presence (4ℓZenr) or absence (4ℓZdep) of a same-flavour, opposite-charge lepton pair. The purity of the Z-enriched region is enhanced by a BDT discriminant, which is trained to discriminate signal from the main backgrounds, $t\bar{t}Z$ and ZZ production, which exploits kinematic reconstruction of the events using multivariative method.

τ_{had} channels

In the 1ℓ+2τ_{had} channel, three categories are defined based on a BDT trained to discriminate the $t\bar{t}H$ signal from the $t\bar{t}$ background. In the 2ℓSS+1τ_{had} and 3ℓ+1τ_{had} channels, a cut and count method is used, along with the non-prompt leptons and fake τ_{had} estimations. The selection requirements for all CRs and SRs are summarised in Table 7.7.

Table 7.7: Selection criterias applied to each channel and the corresponding number of regions. Same-charge (opposite-charge) lepton pairs are referred to as SS (OS).

Channel	Selection criteria
Common	$N_{\text{jets}} \geq 2$ and $N_{b-\text{jets}} \geq 1$
2ℓSS	Two same-charge (SS) very tight (T*) leptons, $p_T > 20$ GeV No τ_{had} candidates $m(\ell^+\ell^-) > 12$ GeV for all SF pairs **13 categories**: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}$, mat. conv, int. conv., $N_{\text{jets}} \geq 4$; $N_{b-\text{jets}} = 1, 2$ for $t\bar{t}H$ SRs split by lepton flavour, charge, jet and b-jet multiplicity
3ℓ	Three loose (L) leptons with $p_T > 10$ GeV; sum of light-lepton charges = ± 1 Two SS very tight (T*) leptons, $p_T > 15$ GeV One OS (w.r.t the SS pair) loose-isolated (L*) lepton, $p_T > 10$ GeV No τ_{had} candidates $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\text{ GeV}\rvert > 10$ GeV for all SFOS pairs $\lvert m(3\ell) - 91.2\text{ GeV}\rvert > 10$ GeV **7 categories**: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, VV , $t\bar{t}$, mat. conv, int. conv
4ℓ	Four loose-isolated (L*) leptons; sum of light lepton charges = 0 $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\text{ GeV}\rvert > 10$ GeV for all SFOS pairs $m(4\ell) < 115$ GeV or $m(4\ell) > 130$ GeV **2 categories**: Zenr (Z-enriched;1 or 2 SFOS pairs) or Zdep (Z-depleted; 0 SFOS pairs)
1ℓ+$2\tau_{\text{had}}$	One tight (T) lepton, $p_T > 27$ GeV Two OS τ_{had} candidates At least one tight τ_{had} candidate $N_{\text{jets}} \geq 3$
2ℓSS+$1\tau_{\text{had}}$	2ℓSS selection, except: One medium τ_{had} candidate $N_{\text{jets}} \geq 4$
3ℓ+$1\tau_{\text{had}}$	3ℓ selection, except: One medium τ_{had} candidate, of opposite charge to the total charge of the light leptons Two SS tight (T) leptons

7.4 Background estimations

The procedure of estimating background compositions depends on the type and the source of background processes. Background contributions in $t\bar{t}H$ signal region can be categorised as irreducible and reducible backgrounds. Irreducible backgrounds contain only prompt leptons from SM processes. Reducible backgrounds have at least one non-prompt light lepton originating from various different sources: lepton originating from heavy-flavour hadron decay or radiated photon conversion, lepton with mis-assigned charge (*QMisID*), or jets misidentified as τ_{had} .

7.4.1 Irreducible backgrounds

Irreducible backgrounds originate from a wide range of physics processes and its importance is varying by channel, depending as well the cross section of the processes. The main irreducible backgrounds are arising from $t\bar{t}W$ and $t\bar{t}Z/\gamma^*$ production, followed by VV production. Minor SM processes also contribute to the $t\bar{t}H$ signal regions such as tZ, tW, WtZ, $t\bar{t}WW$, VVV, $t\bar{t}t$, and $t\bar{t}t\bar{t}$ productions.

$t\bar{t}W$ background

SM $t\bar{t}W$ process is one of the dominant background in $t\bar{t}H$ phase space, particularly in LJ and HJ regions in the 2ℓSS and 3ℓ channels. Although the analysis makes use of well modelled $t\bar{t}W$ MC simulations, the improvement on the knowledge of the normalisation of this process is necessary. Therefore, number of CRs, sensitive to $t\bar{t}W$ phase space, are introduced into the analysis to study and constrain this background. Due to an observed disagreement between data and simulation in 2ℓSS and 3ℓ channels in $t\bar{t}W$ phase space, further investigation is performed, and resulting studies are required to include additional uncertainties on the modelling of the b-jet multiplicity and charge asymmetry in W-boson, and define two corresponding uncorrelated normalisation factors for LJ and HJ categories of the 2ℓSS and 3ℓ channels. Detailed explanation of these studies are provided in Section 7.5.3. The obtained normalisation factors are: $\hat{\lambda}^{2\ell\text{LJ}}_{t\bar{t}W} = 1.56^{+0.30}_{-0.28}$, $\hat{\lambda}^{2\ell\text{HJ}}_{t\bar{t}W} = 1.26^{+0.19}_{-0.18}$, and $\hat{\lambda}^{3\ell}_{t\bar{t}W} = 1.68^{+0.30}_{-0.28}$. The pre-fit fractions of $t\bar{t}W$ process per analysis regions is given in Table 7.8.

Other backgrounds

As indicated in Section 7.3, total yields in the control regions of 3ℓttZ and 3ℓVV are utilised in the likelihood fit to ameliorate the estimation of the background contribution from the $t\bar{t}Z/\gamma^*$ and VV processes.

Another prompt process is mainly coming from conversion of a virtual (non-propagating) photon at the PS level in $t\bar{t}$ process. The normalisation of this background is estimated with $m(e^+e^-) < 1$ GeV in two dedicated control regions (2ℓIntC and 3ℓIntC), within the same background estimation method described in Section 7.4.2. Total yield of each region is used in

Region	**Fraction (%)**	**Region**	**Fraction (%)**	**Region**	**Fraction (%)**
2ℓttH+	33	2ℓttW+	50	2ℓtt(e)-	21
2ℓttH-	22	2ℓttW-	41	2ℓtt(μ)+	31
3ℓttH	12	2ℓMatC	5.9	2ℓtt(μ)-	22
4ℓZenr	0	2ℓIntC	0.4	3ℓMatC	10
4ℓZdep	0	2ℓLJ(e1)	19	3ℓIntC	0.4
2ℓSS+$1\tau_{\text{had}}$	18	2ℓLJ(e2)	40	3ℓtt	13
3ℓ+$1\tau_{\text{had}}$	3.5	2ℓLJ(μ)	23	3ℓttZ	9
1ℓ+$2\tau_{\text{had}}$	0.3	2ℓtt(e)+	30	3ℓVV	10
3ℓttW	37				

Table 7.8: $t\bar{t}W$ pre-fit fractions per region.

the likelihood fit to obtain the normalisation factor $\hat{\lambda}_e^{\text{IntC}} = 0.83 \pm 0.32$, which is dominated by statistical uncertainty. Overall good agreement between data and simulation is observed, as it is shown in Figure 7.1.

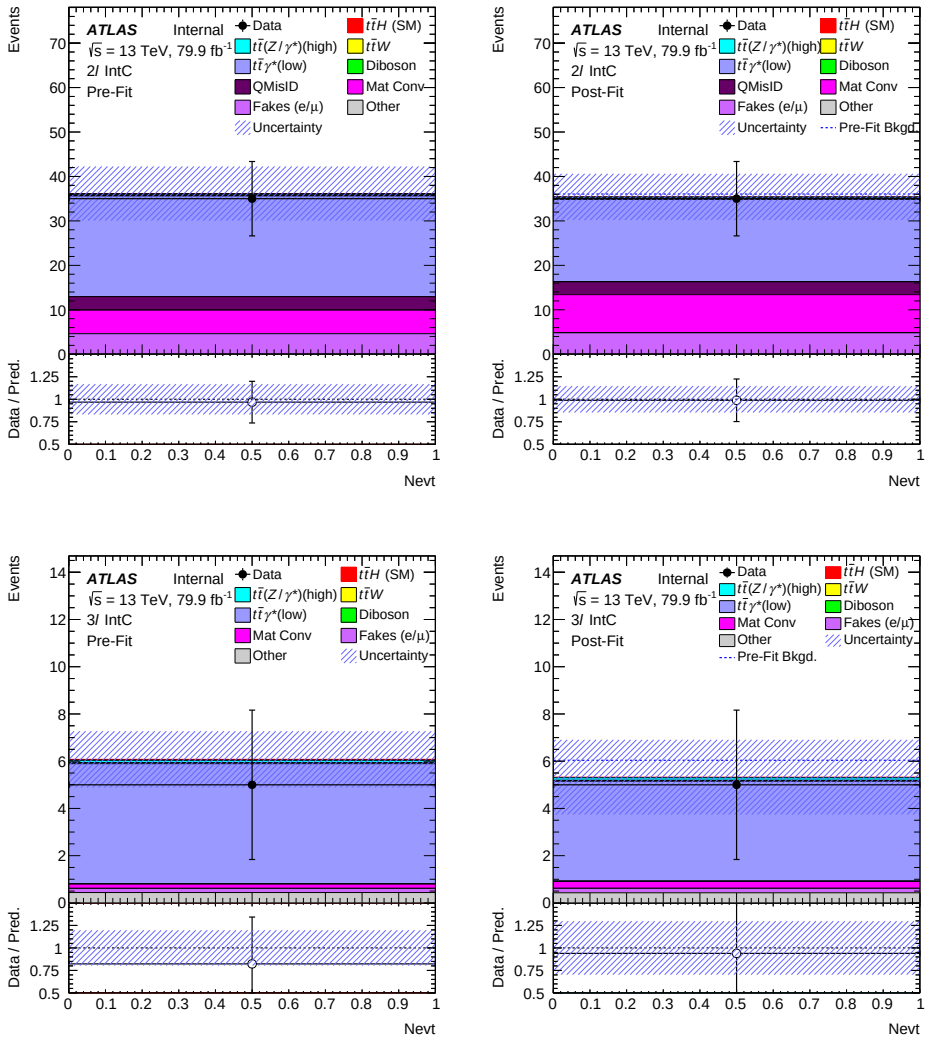


Figure 7.1: The event count of internal conversion CRs for the 2ℓIntC region (***upper***) and 3ℓIntC region (***lower***) for the pre-fit (***left***) and the post-fit (***right***).

7.4.2 Reducible backgrounds

Despite the very tight definitions of leptons, non-prompt backgrounds can contribute to the SRs. The normalisations of these backgrounds must be adjusted through the real data. The types and sources of non-prompt leptons typically have different experimental signatures than those leptons that actually originate from the pp hard-scatter. There are several background estimation techniques used by ATLAS. The reducible backgrounds QMisID and fake τ_{had} are modelled by data-driven methods, while the other background contributions are based on MC simulation shapes and its normalisation taken from data fit in CRs.

Charge misassignment

An important background process, which can contribute to the $t\bar{t}H$ signal phase space in two lepton regions, is the charge mis-assignment of the one of the two leptons. This incident can have two origins:

- One electron having encountered a hard bremsstrahlung ($e^{\pm} \rightarrow e^{\pm}\gamma^{*} \rightarrow e^{\pm}e^{+}e^{-}$). In this case, QMisID occurs when the EM cluster is associated to the track of the opposite sign electron in the trident. Since the probability of this process depends on the traversed detector material, dependence of the QMisID rate on $|\eta|$ is expected.
- Mismeasurement of the electron track-curvature. This effect is more important in the high p_T range (smaller curvature), therefore dependence of the rate on p_T is also expected.

The mis-assignment of the muon charge is negligible since the combination of ID and MS is used in the reconstruction level, which yields to an improved accuracy on the muon identification. The QMisID background rates are estimated by using a data-driven approach that relies on measuring the probability that an electron can have its charge wrongly assigned (ϵ). The rates are calculated in three dimensional bins $|\eta|$, p_{T}, and by each region in 2ℓSS channel such as internal/material conversion. To determine the expected number of events from this process, a data sample is used with same selection for the analysis, except that the electrons required to be opposite sign.

Considering an event containing two electrons with opposite charges ($e^{+}e^{-}$) in the final state with the rate of QMisID (ϵ), this event can be reconstructed in three different ways:

- $e^{+}e^{-}$ can be reconstructed with their correct signs with a probabilty of $(1-\epsilon_i)(1-\epsilon_j)$,
- $e^{+}e^{-}$ can be reconstructed with their wrong signs with a probabilty of $\epsilon_i\epsilon_j$,
- $e^{\pm}e^{\pm}$ can be reconstructed with one of the two electrons have mis-reconstruction of the charge with a probability of $2\epsilon_i(1-\epsilon_j)$.

where ϵ_i and ϵ_j are the QMisID rates for each of the two electrons. Therefore the fraction of events that are reconstructed as same-sign electrons is $\epsilon_i(1-\epsilon_j) + \epsilon_j(1-\epsilon_i) = \epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j$. For $e^{\pm}\mu^{\mp}$ events, the respective fraction is equal to the QMisID rate ϵ_i of the electron. A weight

can be obtained for each event according the values of the QMisID rates of the electrons in the event, and the expected number of mis-assigned same-sign events, N_{SS}, can be computed from the observed number of opposite-sign events N_{OS} as follows:

$$\bar{N}_{\mathrm{SS}} = \underbrace{\frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1 - (\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j)}}_{\text{weight}} N_{\mathrm{OS}} \quad \text{and} \quad \bar{N}_{\mathrm{SS}} = \frac{\epsilon_i}{1 - \epsilon_i} N_{\mathrm{OS}}\,, \tag{7.2}$$

for the ee and $e\mu$ channels, respectively. This weight is associated with the probability that this opposite-sign dilepton event could have been reconstructed as a same-sign dilepton pair. The QMisID rates are derived from the data, based on the fraction of $Z \to e^+e^-$ decays that are reconstructed as a same-sign electron pair, and events in the m_{ee} region around the reconstructed Z-boson mass peak are used. For N^{ij} electron pairs falling in the bin combination i, j (where each of i, j uniquely represents a 3D bin) the expected number of same-sign events is:

$$\bar{N}^{ij}_{\mathrm{SS}}(\epsilon_i, \epsilon_j) = N^{ij} \cdot (\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j)\,. \tag{7.3}$$

Assuming that all of the observed same-sign events N^{ij}_{SS} in the m_Z window are products of electron QMisID, they follow a Poisson distribution around the expectation value:

$$f(N^{ij}_{\mathrm{SS}}|\bar{N}_{\mathrm{SS}}(\epsilon_i, \epsilon_j)) = \frac{[\bar{N}^{ij}_{\mathrm{SS}}]^{N^{ij}_{\mathrm{SS}}} e^{-\bar{N}^{ij}_{\mathrm{SS}}}}{N^{ij}_{\mathrm{SS}}!}\,, \tag{7.4}$$

which is integrated into a likelihood as follows

$$L(\vec{\epsilon}|N_{\mathrm{SS}}) = \prod_{i,j} f(N^{ij}_{\mathrm{SS}}|\bar{N}_{\mathrm{SS}}(\epsilon_i, \epsilon_j))\,. \tag{7.5}$$

This expression can be maximised (minimisation of $-2\ln L$) to obtain the rates that best describe the data. The contributions from other processes such as non-prompt backgrounds, described in Section 7.4.2, is substracted from sidebands of the m_Z window for each bin combination i, j, separately for same-sign ($N^{ij}_{\mathrm{SS,BG}}$) and opposite-sign ($N^{ij}_{\mathrm{OS,BG}}$) events. The widths of upper and lower sidebands are defined with width equal to the m_Z window so that the background can be obtained from the average yield. This estimate is then used to correct the expected number of events in Equation 7.3:

$$\bar{N}^{ij}_{\mathrm{SS}} = N^{ij}_{\mathrm{SS,BG}} + (N^{ij} - N^{ij}_{\mathrm{SS,BG}} - N^{ij}_{\mathrm{OS,BG}}) \cdot (\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j)\,. \tag{7.6}$$

The QMisID rates are evaluated separately for the three different types of electron definitions (internal conversion, material conversion and tight electrons) and parameterised as a function of p_T and η. Further studies are done in order to investigate the dependence of the rates on positron and electron, however due to the lack of statistics, it is decided to maintain the inclusive

approach. In Figure 7.2, the measured QMisID rates are given.

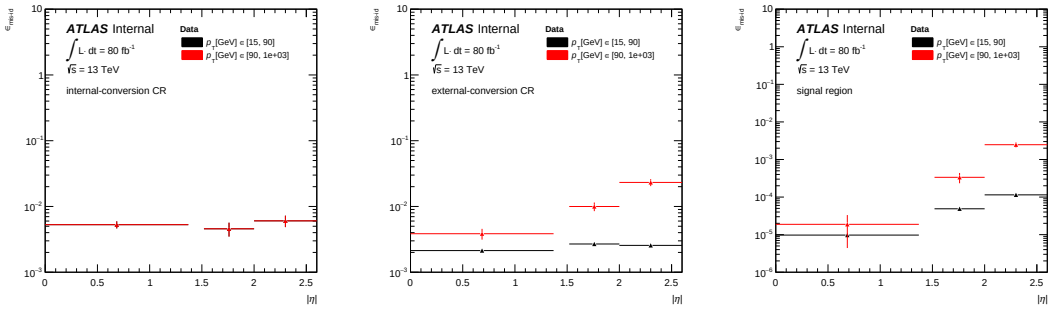


Figure 7.2: Electron QMisID rates derived from the data with the likelihood method. The rates are presented as a function of $|\eta|$, parameterised in p_T for internal-conversion (***left***), external-conversion (***middle***) and prompt candidates (***right***).

The electron charge mis-assignment measurement is validated by a closure test in simulation using same-charge electron pairs, with the observed difference between measured and predicted rates taken as the systematic uncertainty.

Non-prompt light leptons

Non-prompt background contribution may originate from heavy-flavour hadron decays (b or c hadrons), material conversions, or mis-reconstruction of other particles depending on the lepton definition and may differ across event categories.

The estimation of this background is defined between 2ℓSS and 3ℓ together, and 4ℓ. The main contribution to the non-prompt-lepton background is $t\bar{t}$ production, followed by much smaller contributions from V+jets and single-top processes and its estimation is based on simulation with data corrections.

Template method in 2ℓSS and 3ℓ channels

Non-prompt lepton backgrounds in 2ℓSS and 3ℓ channels are a mixture of leptons from semileptonic heavy-flavour decays (HF), light-flavour (LF) hadron decays and photon conversions (MatC). Therefore, a template method is developped to estimate each background contribution. This method relies on the normalisation of the different non-prompt contribution templates obtained from data, where their shapes are given by MC simulations. Based on the truth information of events, main contributions listed below are determined and normalisation factors (NF) which are free floating are attibuted to each of them:

- $\lambda_e^{Mat.CO}$: NF applied to events with one electron originating from photon material conversion.
- λ_e^{HF}: NF applied to events with one non-prompt electron from b- or c hadron decay, or light hadron decays.

- λ_{μ}^{HF}: NF applied to events with one non-prompt muon from b- or c hadron decay, or light hadron decays.

Contribution from internal conversions ($\gamma* \rightarrow \ell\ell$) is also predicted by MC simulation within this method and a NF is assigned, as described in Section 7.4.1. Several event categories described in Section 7.3 are used to determine and control the non-prompt lepton background. These CRs are designed to be enriched in specific background processes. CRs of 2ℓMatC and 3ℓMatC are dedicated for material conversions, and further details about the regions can be found in Appendix B.

In 2ℓSS LJ regions, where the non-prompt backgrounds exist with high statistics, different kinematic distributions are exploited. The shape of these regions provide a further control and constraint of different sources of backgrounds and as well with $t\bar{t}W$ background. The separation power of the following distributions are found to be most efficient for the simultaneous fit:

- ΔR_{ll} in $\mu e + ee$ region with exactly 1 b-jet requirement, provides separation against internal conversions, which are characterised by lower values of $\Delta R(\ell, \ell)$ since the virtual photon is usually radiated by one of the reconstructed leptons.

- $H_{T,lep}$ in $e\mu + \mu\mu$ region labelled as 2ℓLJ(μ), used to estimate NF_{μ}^{HF} and decrease its correlation with $t\bar{t}W$ NFs.

- $H_{T,lep}$ in $\mu e + ee$ region with ≥ 2 b-jets labelled as 2ℓLJ(e2), used to provide improved discrimination between conversions and $t\bar{t}W$ NFs.

Different b-jet requirements across the regions provide better discriminant between non-prompt background originating from conversions and heavy-flavour sources, where the HF is mostly populated in $2b$ region. The 3ℓtt control region exploits the output of the BDT corresponding to the $t\bar{t}$ category, which provides additional constraints on non-prompt background estimations. The likelihood fit is performed to estimate these NFs simulteniously with internal conversions, $t\bar{t}H$ signal and $t\bar{t}W$ normalisation factors. The measured normalisation factors are: $\hat{\lambda}_{e}^{\text{MatC}} = 1.61 \pm 0.48$, $\hat{\lambda}_{e}^{\text{had}} = 1.12 \pm 0.38$, and $\hat{\lambda}_{\mu}^{\text{had}} = 1.20 \pm 0.18$. Several systematic uncertainties are considered and are described in Section 7.5, although the impact on the $t\bar{t}H$ signal is negligible. The final pre- and post-fit distributions of non-prompt background control regions are shown in Figures 7.3- 7.10 and generally good agreement with the pre-fit background prediction is observed.

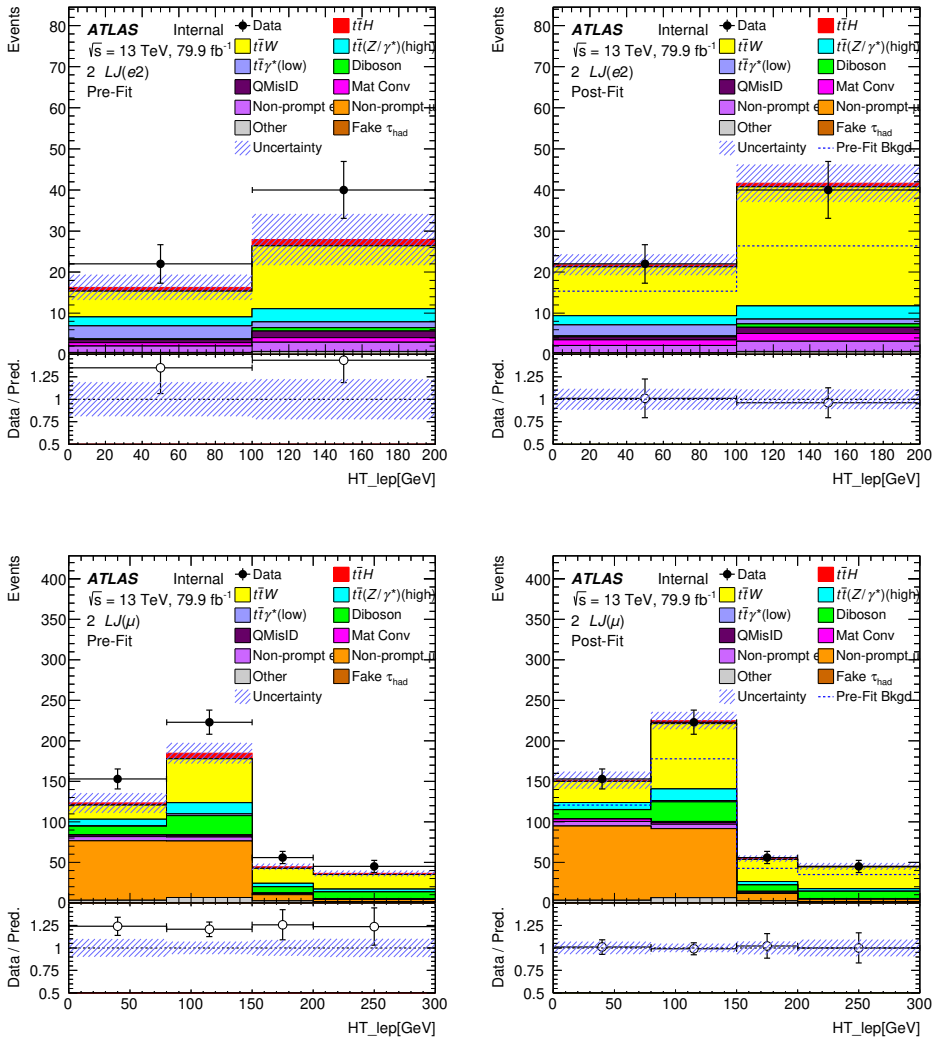


Figure 7.3: Distributions of $H_{\mathrm{T,\,lep}}$ in $\mu e + ee$ (***upper***) and in $e\mu + \mu\mu$ (***lower***) region in 2ℓSS channel, for pre-fit (***left***) and post-fit (***right***). Post-fit plots obtained from final fit.

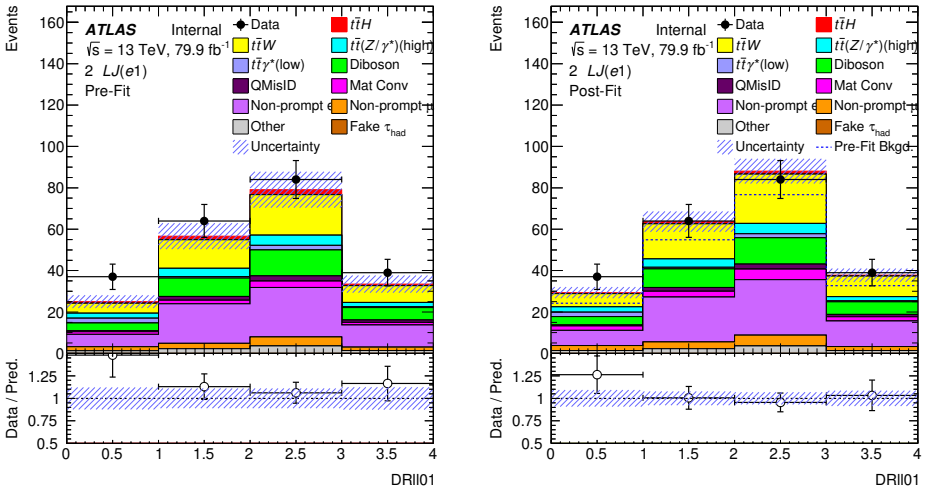


Figure 7.4: Distribution of ΔR_{ll} in $\mu e + ee$ region in 2ℓSS channel, for pre-fit (***left***) and post-fit (***right***). Post-fit plot obtained from final fit.

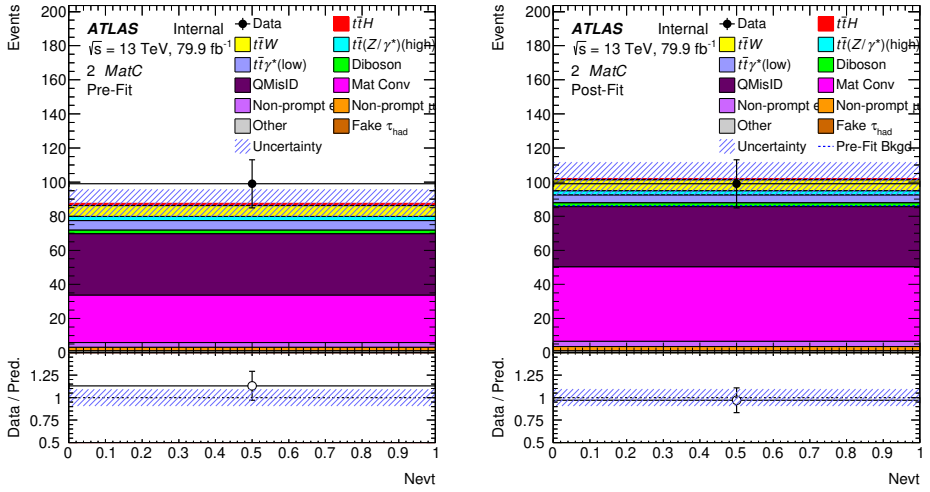


Figure 7.5: The event count of external conversion CRs in 2ℓSS channel for pre-fit (***left***) and post-fit (***right***). Post-fit plot obtained from final fit.

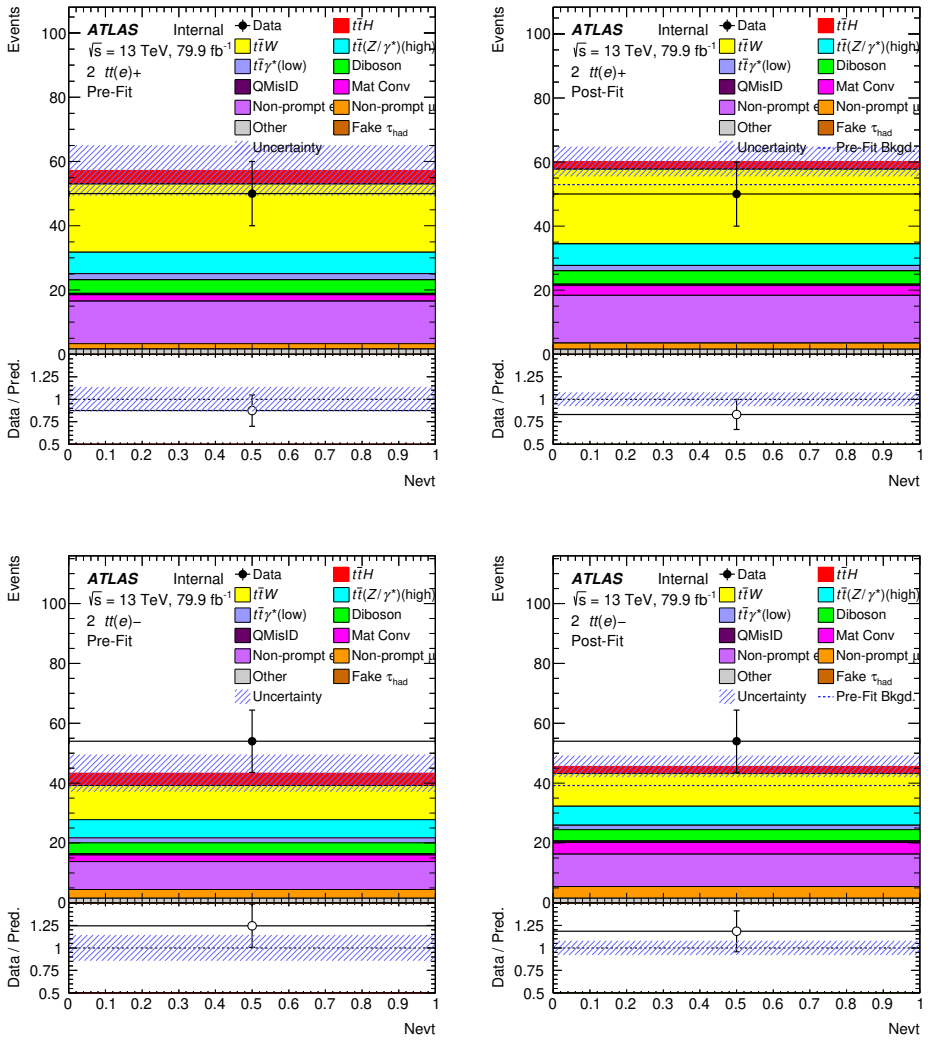


Figure 7.6: The event count of $t\bar{t}$ CRs for $\mu e + ee, plus - plus$ (***upper***) and $\mu e + ee, minus - minus$ (***lower***) in 2ℓSS channel for pre-fit (***left***) and post-fit (***right***). Post-fit plots obtained from final fit.

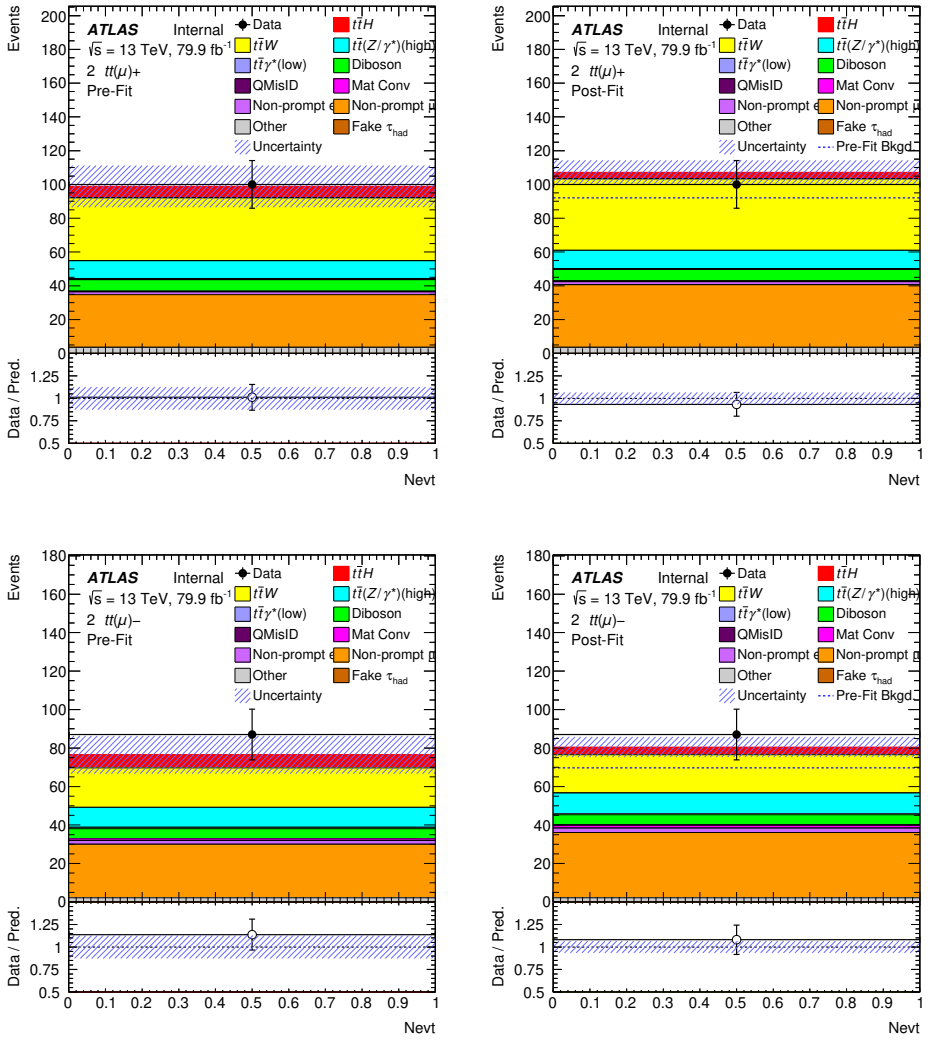


Figure 7.7: The event count of $t\bar{t}$ CRs for $e\mu + \mu\mu, plus - plus$ (***upper***) and $e\mu + \mu\mu, minus - minus$ (***lower***) regions in 2ℓSS channel for pre-fit (***left***) and post-fit (***right***). Post-fit plots obtained from final fit.

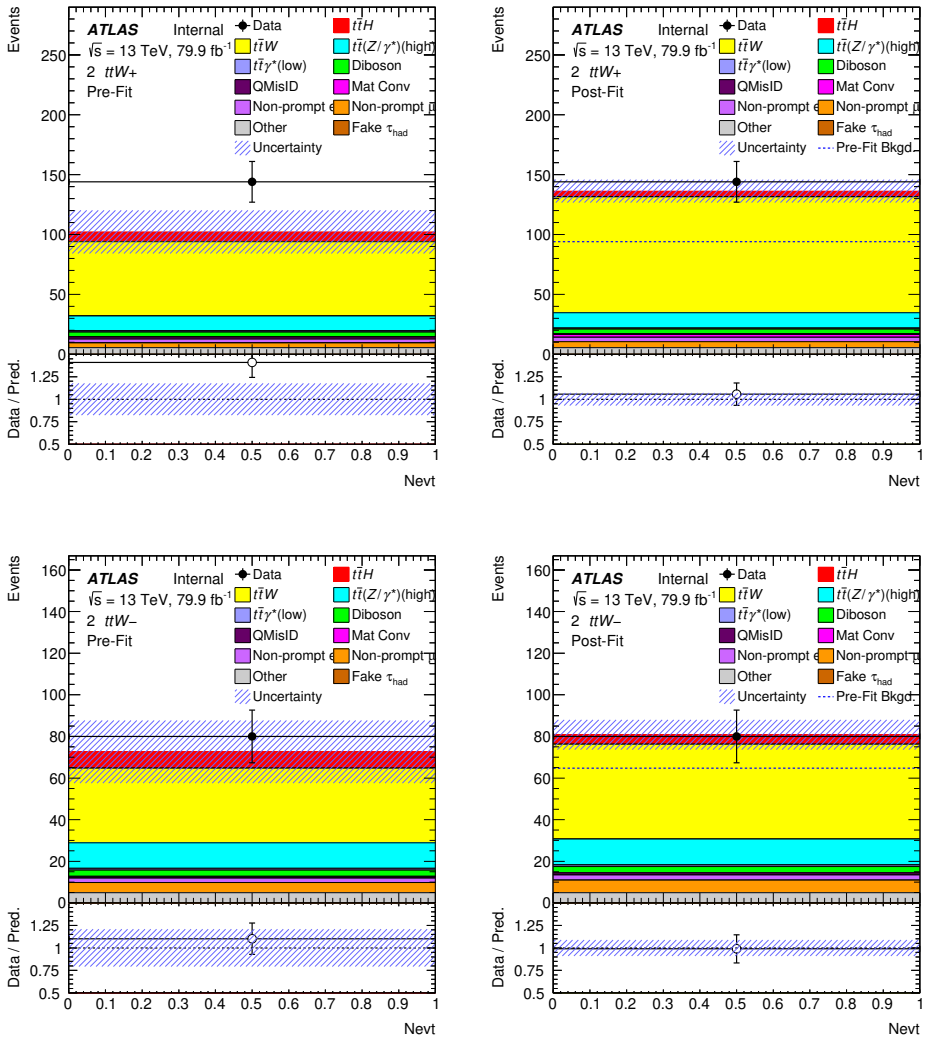


Figure 7.8: The event count of $t\bar{t}W$ CRs for $plus - plus$ (***upper***) and $minus - minus$ (***lower***) regions in 2ℓSS channel for pre-fit (***left***) and post-fit (***right***). Post-fit plots obtained from final fit.

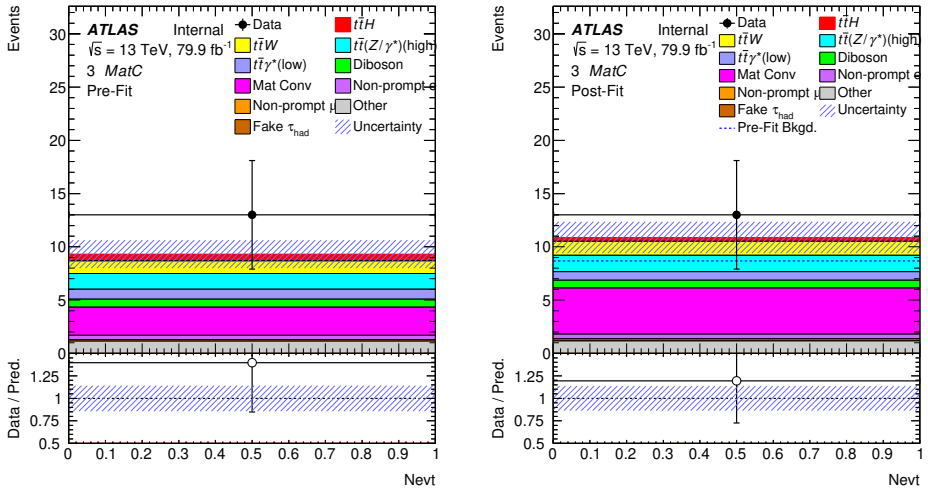


Figure 7.9: The event count of material conversion CRs in 3ℓ channel for pre-fit (***left***) and post-fit (***right***). Post-fit plot obtained from final fit.

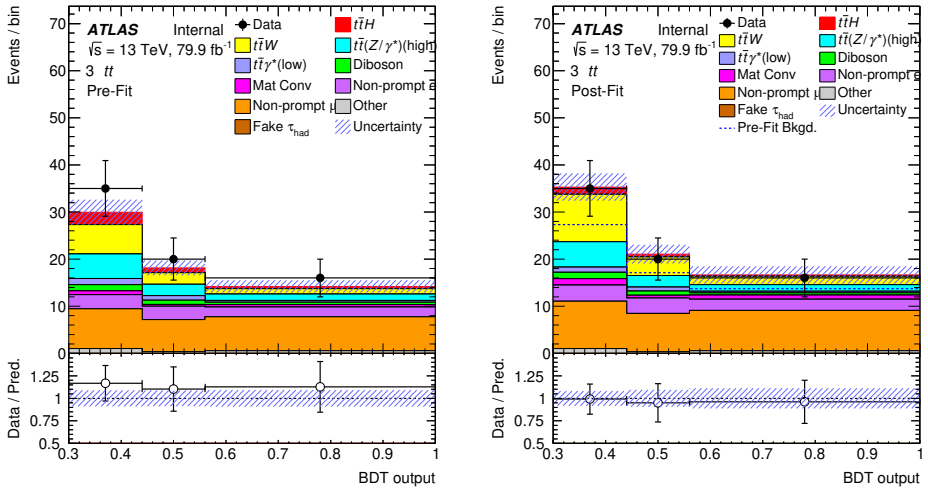


Figure 7.10: The event count of $t\bar{t}$ CRs in 3ℓ channel (***left***) and post-fit (***right***). Post-fit plot obtained from final fit.

Due the limited number of non-prompt leptons in 4ℓ phase space, a semi data-driven method is adapted. The fake factors for events, where one or more lepton candidates are originating either from heavy-flavour b- or c-decay, or photon conversions and light flavour decays are assigned. The simultaneous fit is performed in 6 bins of $E_{\mathrm{T}}^{\mathrm{miss}}$ in 3ℓ CRs with leptons having loose selections. The associcated normalisation factors are: $\lambda_e^{\ell} = 1.37 \pm 1.41$, $\lambda_e^{h} = 0.89 \pm 0.41$ and $\lambda_{\mu} = 1.07 \pm 0.43$. Additional uncertainties to cover the dependency on the lepton WPs are added.

Fake τ_{had} candidates

The τ_{had} -jet fake estimate is performed in the 2ℓOS+$1\tau_{\mathrm{had}}$ and utilised in the 3ℓ+$1\tau_{\mathrm{had}}$ and 2ℓSS+$1\tau_{\mathrm{had}}$ analyses. The 1ℓ+$2\tau_{\mathrm{had}}$ uses a fully data driven estimate of the fakes. The estimation of the fake tau background is based on a data-driven method called *scale-factor* method. According to this method, scale factors are derived for each event by comparing the rate of fake-tau events in data to the rate of fake tau events in MC from the dedicated control regions, and used to correct the normalization of the MC yield. The scale factors are derived as a function of p_T for 1- and 3-prong taus separately using a formula:

$$Fake\ \tau_{had}\ SF\ =\ \frac{DD_{Fakes}}{MC_{Fakes}}\ =\ \frac{Data\ -\ MC_{Real\tau}\ -MC_{FakeTauFromLepton}}{MC_{FakeTauFromJet}}. \qquad (7.7)$$

For 2ℓSS+$1\tau_{\mathrm{had}}$ and 3ℓ+$1\tau_{\mathrm{had}}$ channels, the main contribution is arising from $t\bar{t}$ and $t\bar{t}V$, while for 1ℓ+$2\tau_{\mathrm{had}}$ is mainly from $t\bar{t}$ physics processes. For the one-prong (three-prong) τ_{had} candidates, the normalisation factors lies in the range of 1.05 ± 0.06 (1.25 ± 0.42) for $p_{\mathrm{T}}^{ensuremath\tau_{\mathrm{had}}}$ in the

range of 25–45(25 – 50) GeV, to 0.64 ± 0.12 (0.52 ± 0.71) for $p_{\mathrm{T}}^{\tau_{\mathrm{had}}} \geq 70(75)$ GeV.

7.5 Systematic uncertainties

There are a number of systematic uncertainty sources that affect the final results of $t\bar{t}H$ measurements. These uncertainties categorised into experimental uncertainties (Section 7.5.1), and (signal and background) modelling uncertainties (Secion 7.5.2). These uncertainties can affect the shape and the normalisation of the distribution or a process, or as well can affect the both.

Systematics uncertainties, also called nuisance parameters, are defined with a Gaussian constraint as described in Section 6.5. The polynomial functions are used in interpolation of the $\pm 1\sigma$ effect of systematic variation, while the extrapolation step varies for the normalisation and shape uncertainties. An exponential extrapolation is used for normalisation components, in order to avoid that total yield become negative, while a linear interpolation is applied for the shape components. Therefore, the normalisation uncertainties behave like linearly extrapolated, with a log-normal constraint.

Some NPs are described by a varying the nominal setting in single direction. In such cases, the variation is described as one sided ($+1\sigma$) effect, and is symmetrized to acquire the equivalent variation for the -1σ effect of the NP. When both NP variations are given, the variations corresponding to the $\pm 1\sigma$ effects are mostly symmetrized, and the average of the affect is applied, or directly the real affect of $\pm 1\sigma$ is applied individually.

Additionaly, a smoothing technique is applied to some uncertainties in order to eliminate the fluctuations caused by limited statistical components, which may lead to bigger constraints artificially in the fit.

Finally, if the effect of a NP behaving on a sample for a given region is negligible, this effect is removed from the likelihood function. This threshold is 1% for shape uncertainties, and 0.5% for the normalisation uncertainties.

Statistical uncertainties are predicted by the nominal model coming from the number of MC events. The nominal model estimation is done as a suplementary measurement, with a statistical uncertainty in each bin, which corresponds to one nuisance parameter. These uncertainties are controlled by Gaussian constraints.

7.5.1 Experimental uncertainties

Experimental uncertainties are associated to the lepton reconstruction, identification and isolation, trigger efficiency, jet calibration and b-tagging efficiencies. The ATLAS performance groups provide the experimental systematics variations often evaluated by data-MC scale factors, and these are applied either as a rescaling the energy or momentum of the object or as an overall event re-weighting in the analysis.

Conditions of data taking

The uncertainty on the integrated luminosity for 2015-2017 is 2.1%. It is derived from the luminosity scale calibration following a method detailed in Reference [188] and is applied to each MC processes. To account for the variation on re-weighting of the pile-up distributions in MC, with respect to actual pile-up distributions observed in data, additional uncertainty is considered.

Lepton uncertainties

For muons, the uncertainties related to ID track smearing, MS track smearing, charge-independent scale momentum and two uncertainties for the charge-dependent scale momentum (so-called saghitta effects) are considered in addition to reconstruction, identification, isolation uncertainties. For electrons, uncertainties covering energy scale and resolution, as well as the correction factor uncertainties related to efficiency of trigger, reconstruction, identification and isolation are applied.

Jet uncertainties

The systematic uncertainties on reconstructed jet objects are associated to the jet energy resolution (JER), jet energy scale (JES), and JVT. The JVT uncertainty is obtained using the dedicated tool and varying up and down the JVT cut [136]. To determine the JES uncertainties, data and simulation are used [189], and the globally reduced uncertainty set is provided by the dedicated performance group. As a results, several NPs are defined related to statistical, detector, modelling effects, jet flavour compositions, pile-up corrections, and η-dependence effects, each with up/down variations symmetrised. JER uncertainties as well are measured separetely for data and MC using *in-situ* techniques, described in Reference [190]. Unlike the JVT uncertainties, JES and JER uncertainties are applied on the p_T of jets with data-MC obtained scale factor.

Flavour tagging uncertainties

The recommended uncertainties from the heavy flavour tagging group [191] are used with several components related to statistical, experimental, and modelling uncertainties arising from the flavour tagging calibration procedures. The NPs are obtained after diagonalisation of the error matrix across all pairs of kinematic bins, which are used to derive heavy-flavour efficiency corrections. The obtained eigenvectors correspond to independent variations include the proper correlations in all different kinematic regions.

Missing transverse energy

The systematic variations of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is modelled by three independent NPs, which are related to the track-based soft term such as energies are not associated to any reconstructed object. The systematic uncertainty on these objects are estimated by comparing events without $E_{\mathrm{T}}^{\mathrm{miss}}$, and the difference is considered to be an estimation of the corresponding uncertainty. This soft vectoral component is decomposed into a parallel and perpendicular component.

7.5.2 Signal and background modelling uncertainties

This section summarizes systematic uncertainties related to the signal and background modellings. Uncertainties on the modelling of specific processes are typically assessed by comparing the nominal MC simulation with the certain theoretical or phenomenological parameters varied versions of it. Therefore the impact of the underlying assumptions made in the MC simulation generation are measured on the final results. Uncertainties having impact on the modelling of the acceptance and event kinematics contain variations in the QCD factorisation and renormalisation scales, choice of hadronisation and parton shower model, the modelling of ISR/FSR, and PDF uncertainties.

Uncertainties related to non-prompt background modelling are derived from data-MC comparisons in the dedicated regions, as explained in the following paragraphs.

$t\bar{t}H$ signal

The uncertainties on QCD scale choice are obtained by varying the factorisation and renormalisation scales independently by factor of 2.0 and 0.5 according to central value, and resulting an overall uncertainty of $^{+5.8\%}_{-9.2\%}$. The uncertainty on $t\bar{t}H$ cross section due to the PDF choice is (including α_s strong coupling uncertainties) ±3.6%. Uncertainties related to the choice of hadronisation and parton shower model are estimated by comparing the nominal sample to an alternative sample generated with Powheg-BOX interfaced to Herwig7 [192], while keeping Powheg for the hard-scatter generation in both cases. The uncertainty associated to the modelling of ISR is obtained by considering the Var3c A14 tune variation [155], which corresponds to a variation of α_s in the A14 tune. The uncertainty on the choice of PDF used in the hard-scatter generation is evaluated by using the PDF4LHC15 prescription [193], comprised by 30 eigenvector shifts derived from fits to multiple NNLO PDF sets. Finally, the uncertainties associated with the predicted Higgs-boson branching ratios [22] are also considered.

$t\bar{t}W$ background

Systematic uncertainties due to missing higher-order QCD terms are estimated by varying the factorisation and renormalisation scales in the nominal sample simultaneously by a factor of 2.0 and 0.5 with respect to the central value. To assess the impact of ME generator, the comparison is done between the nominal $t\bar{t}W$ prediction with an alternative sample MadGraph5_aMC@NLO 2.2.1 generator while keeping the same scale choice and PDF set as for the nominal sample, and interfaced to Pythia 8.2 in combination with the A14 tune. Finally, the uncertainty due to the choice of PDF set is obtained using the PDF4LHC15 prescription, as in the case of $t\bar{t}H$ signal sample. Additional uncertainties assigned to $t\bar{t}W$ dominated regions are discussed later in Section 7.5.3. They are motivated by the observation of tension between data and MC predictions in these regions. An example of the effect of one PDF error set variation on the $t\bar{t}W$ yields is shown in Figure 7.11.

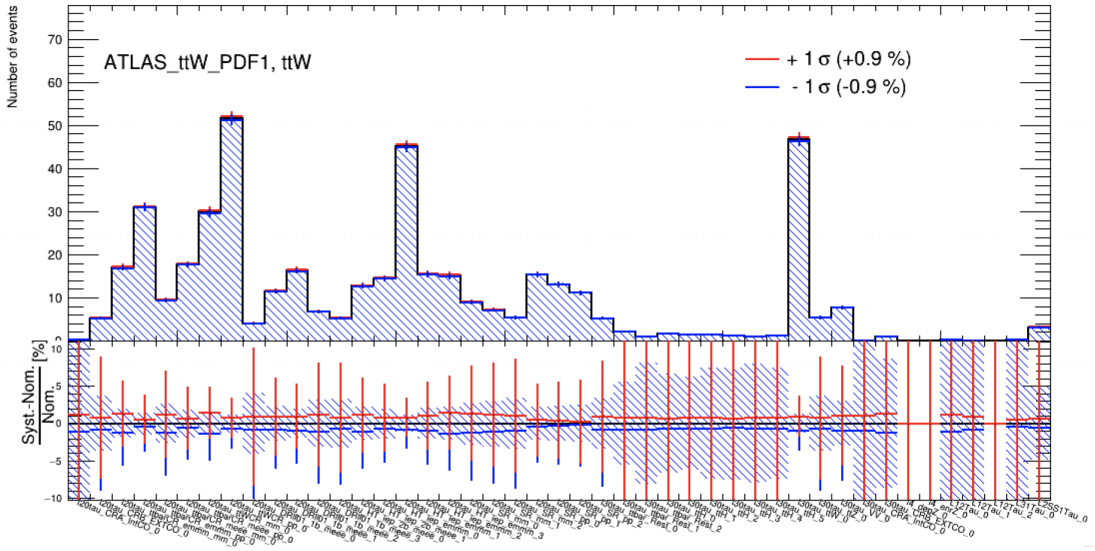


Figure 7.11: Effect of one PDF4LHC error set variation on the $t\bar{t}W$ yields in each region.

Other backgrounds

Uncertainties affecting the modelling of the acceptance and event kinematics for the $t\bar{t}\ell\ell$ sample include the same QCD scale and tune variations as considered for the $t\bar{t}H$ and $t\bar{t}W$ samples, PDF variations using the PDF4LHC15 prescription and the comparison to an alternative LO multileg sample. In the case of the inclusive $t\bar{t}$ sample, the uncertainty on the modelling of additional QCD radiation is evaluated with two alternative samples generated with settings that increase or decrease the amount of radiation [173]. The uncertainties assigned to diboson backgrounds are 50% and treated as uncorrelated among different subprocesses (WZ+light-jets, WZ+$\geq 1c$, WZ+$\geq 1b$, and ZZ+jets). An uncertainty of 50% on the normalisation is considered for rare background contributions (tZ, $t\bar{t}t\bar{t}$, $ttWW$, WtZ, VVV, $tt\bar{t}$, $tHjb$ and WtH).

Charge misassignment

The uncertainty sources on this background estimation are grouped into four categories:

- The statistical uncertainty from the likelihood method.
- The difference between the measured rates from the likelihood method and from the truth-matching with simulated $Z \to e^+e^-$ events.
- The variation of QMisID rates with definition of invariant mass of the dilepton system around the m_Z window and its sidebands, which are used to measure the contamination from non-prompt leptons.
- The difference between the QMisID rates of electrons and positrons, estimated by truth-matching with simulated $Z \to e^+e^-$ events.

The total uncertainty is defined as the quadratic sum of the whole contributions discussed above. (Figure 7.12).

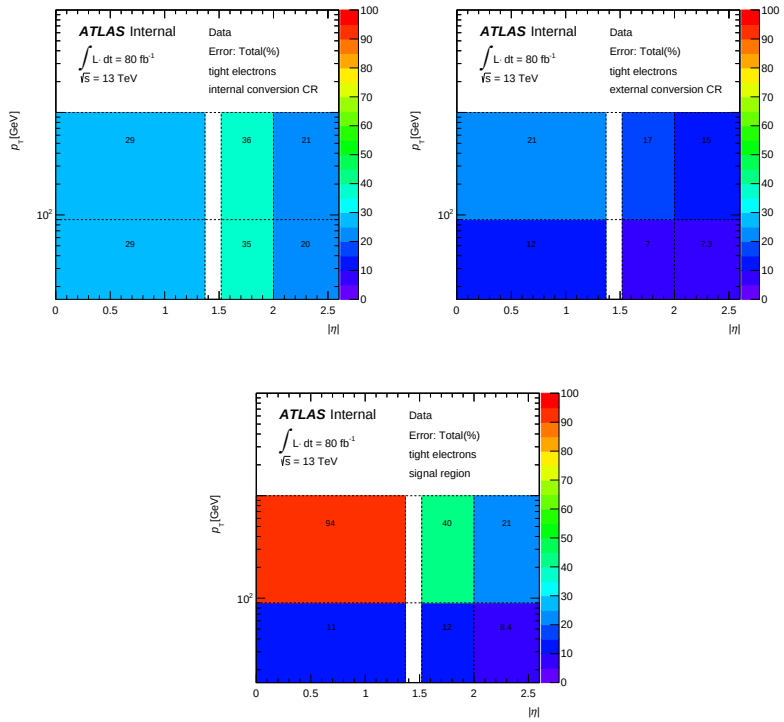


Figure 7.12: Total relative systematic uncertainty (in %) on the QMisID rate in bins of $|\eta|$ and p_T for internal-conversion (***upper***), external-conversion (***middle***) and prompt electron (***lower***) candidates.

Non-prompt backgrounds

Since the overall normalisation of this background is a free parameter in the final fit, the uncertainty on the non-prompt background estimations originates from the shape of the template distributions used in the template fit method. Depending on the source of the non-prompt background, several shape uncertainties are assigned. Since the light and heavy-flavour electron backgrounds are originally treated as one single background contribution and both scaled by the same normalisation factor, a 100% uncertaintiy is assigned to light-flavour electron brackground samples. This value is justified by studies investigating the distributions of input variables for the PLV discriminant with loose lepton selection, which has a significant light-flavour component.

Shape uncertainties on the non-prompt templates are derived by relaxing the tight lepton selection to enrich the contributions and later by comparing the data with the MC simulation, after the subtraction of remaining backgrounds. Two shape uncertainties are derived, which are targeting different sources of non-prompt backgrounds, namely material conversions

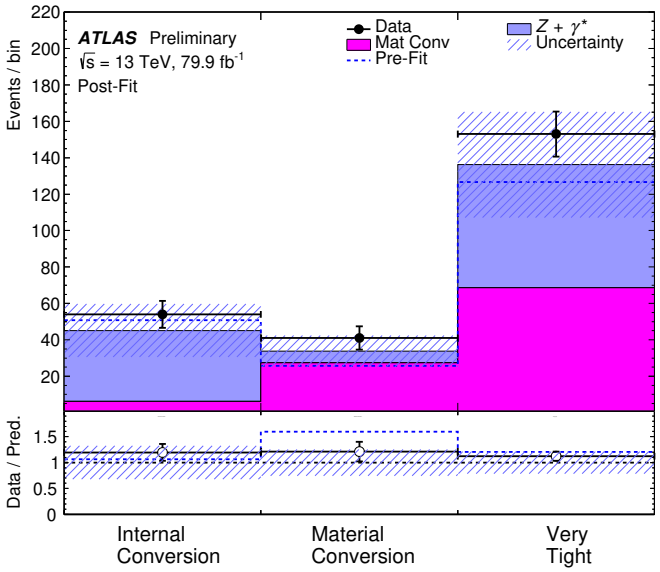


Figure 7.13: Comparison between data and prediction for the distribution of $m_{\mu\mu e}$ in the 3ℓ validation region enriched in $Z \to \mu^+\mu^-\gamma^*(\to e^+e^-)$ candidate events, after correcting the simulation by the corresponding normalisation factors, coming from the likelihood fit to data in all categories.

and heavy-flavour components by reversing the ambiguity bin and loosing the cut on PLV, respectively.

A shape uncertainty of up to 15% (6%) is assigned to the HF non-prompt electron (muon) background component from a comparison between data and simulation. The resulting differences between data and scaled simulation, after removal of the overall normalisation effect is considered as an additional uncertainty correlated across all distributions and event categories. Furthermore, the validation of the internal and material conversion background modelling is done through comparison of data and the scaled simulation in a validation region enhanced with $Z \to \mu^+\mu^-\gamma^*(\to e^+e^-)$ candidate events. The difference arising from the observed and predicted yields is assigned as an additional uncertainty (25%) to the fit model, corresponding the extrapolation of the internal/material conversion control regions to the other event categories, and can be seen in Figure 7.13.

Fake τ_{had} backgrounds

Uncertainties from the τ_{had} background estimations are coming from statistical uncertainties on the control regions, the substruction of real τ_{had} candidates in the control regions and the difference between the normalisation factors derived from control regions and the validation region, which is enriched in Z+jets. The total systematic uncertainty is obtained as a function of p_T and found to be around 13% and 60% for one and three prong candidates, respectively.

7.5.3 Updated $t\bar{t}W$ uncertainty model

Changes in the modelling of $t\bar{t}W$ phase space are mainly motivated by the observation of the tension arising from the difference between data-MC in $t\bar{t}W$ enriched regions (Appendix C). In order to obtain a more robust extraction of the $t\bar{t}H$ signal strength, additional degrees of freedom, which can absorb the observed mismodellings, are included to the fit setup. This reduces the overall sensitivity to the $t\bar{t}W$ normalisation and resulting a minimal impact on the sensitivity on the $t\bar{t}H$ measurement. Three major sources of tension between data and MC are observed and corresponding additional degrees of freedom have been introduced into the fit model for 2ℓSS and 3ℓ channels:

- Low-/high-jet multiplicity regions (Figure 7.14).
- Charge asymmetry (Figure 7.16).
- b-jet multiplicity (Figure 7.15).

Hence, in order to reduce the dependence of $t\bar{t}H$ signal measurement on $t\bar{t}W$ prediction, three independent NFs for $t\bar{t}W$ background are introduced to fit the model ($\hat{\lambda}^{2\ell\mathrm{LJ}}_{t\bar{t}W}$, $\hat{\lambda}^{2\ell\mathrm{HJ}}_{t\bar{t}W}$ and $\hat{\lambda}^{3\ell}_{t\bar{t}W}$). Figure 7.14 demonstrates the pre- and post-fit distributions of number of jets in 2ℓSS and 3ℓ regions, respectively. Clearly, in order to have better data-MC closure, the normalisation of $t\bar{t}W$ should be higher in the low jet multiplicity shown in the right upper of the Figure 7.14. Overall jet multiplicity agreement is improved compared to 1 NF $t\bar{t}W$ fit model scenario, as shown in Appendix D. The main motivation behind this mainly originated from $3\ell_{t\bar{t}W}$ NF, which is slightly higher in both CR-only and 3 NF fit model, which is absorbed by $t\bar{t}H$ signal in the case of 1 NF model (see Section 7.7.4), and to show the consistency of the $t\bar{t}W$ measurement across different phase spaces.

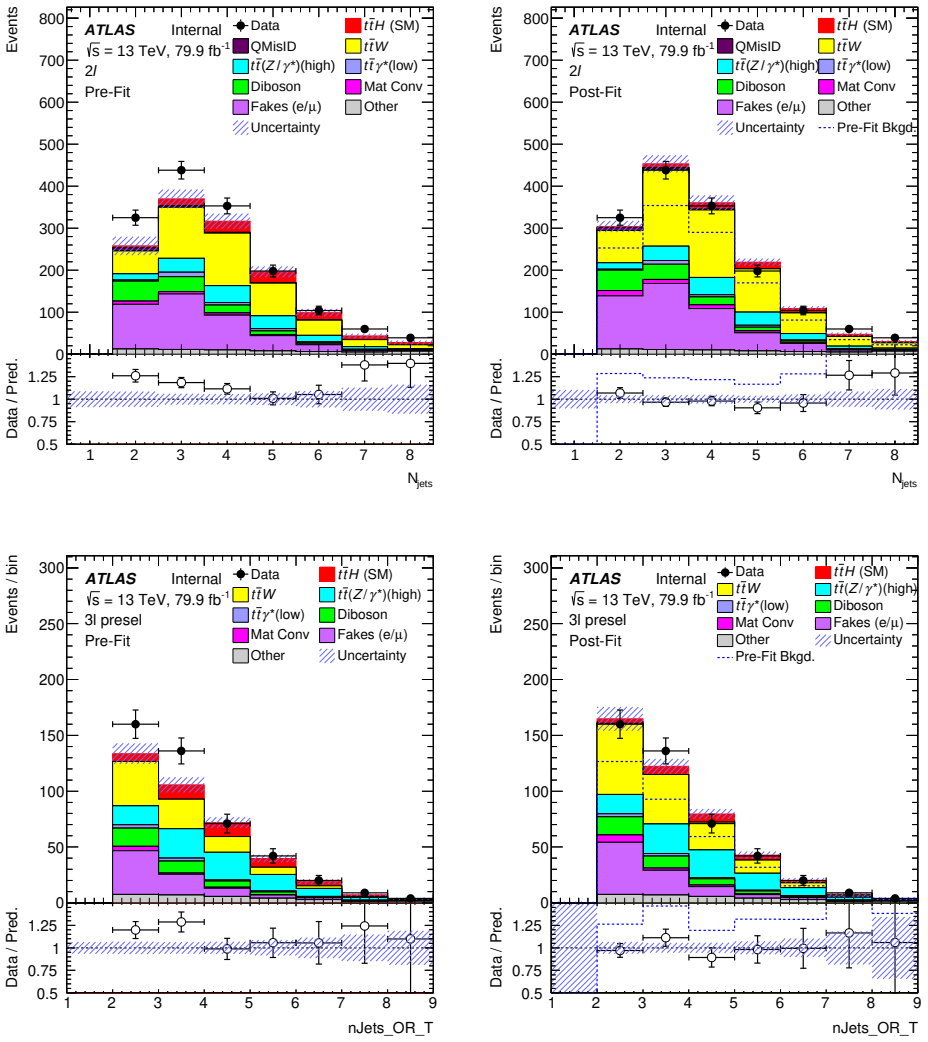


Figure 7.14: Pre-fit (***left***) and post-fit (***right***) distributions of number of jets in 2ℓSS channel (***upper***) and 3ℓ channel (***lower***). The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel. Fakes (e/μ) represent the total heavy-flavour non-prompt background for electrons and muons.

Additional extrapolation uncertainties associated to the modelling of the b-jet multiplicity and W-boson charge asymmetry in the $t\bar{t}W$ background are derived from the observed discrepancies in the shape of these distributions between data and pre-fit background predictions in 2ℓSS and 3ℓ channels. The uncertainties are calculated by taking the data/MC differences between the relevant bins in the 2ℓSS and 3ℓ pre-MVA regions. This difference is taken as the size of the new nuisance parameter ($|+1\sigma| + |-1\sigma| = [\text{data} - \text{MC}^{\text{non-}t\bar{t}W}]/\text{MC}^{t\bar{t}W}$ difference) and the variations are normalised to their overall effect, so that the extrapolation uncertainties do not affect the overall normalisation of the $t\bar{t}W$ sample and 80% prior is assigned to $t\bar{t}W$+ and $t\bar{t}W$-2b region. The resulting variations for the b-jet multiplicity distribution is ±25% (∓35%) for events with exactly one (at least two) b-jets, and for the modelling of the total charge distribution is ±20% (∓35%) for events with positive (negative) total charge. These additional uncertainties are considered as uncorrelated between the 2ℓSS and 3ℓ channels and Figure 7.17 shows the variations in the bins of b-jet multiplicity and charge distributions for 2ℓSS and 3ℓ channels. Figure 7.15 and 7.16 demonstrate the pre- and post-fit distributions of b-jet multiplicity and total charge with these additional uncertainties.

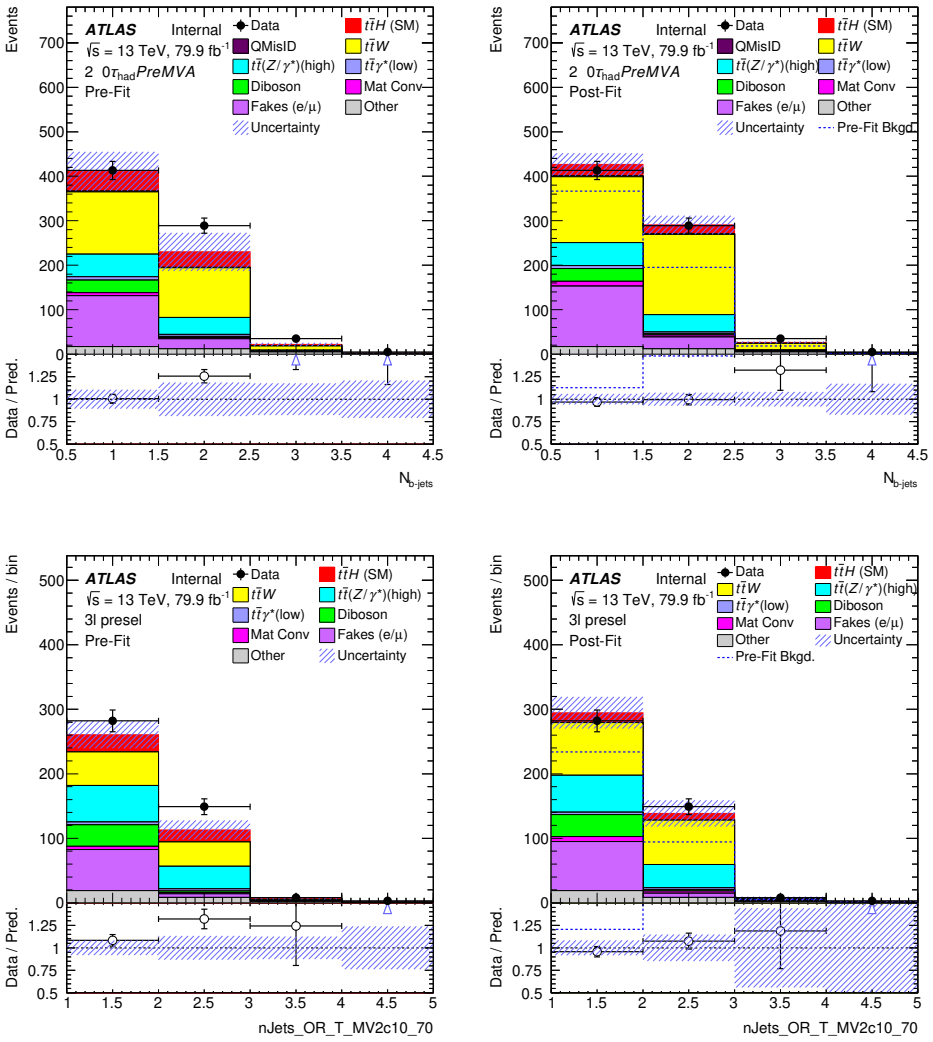


Figure 7.15: Pre-fit (***left***) and post-fit (***right***) distribution of number of b-jets in 2ℓSS channel (***upper***) and 3ℓ channel (***lower***). The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel. Fakes (e/μ) represent the total heavy-flavour non-prompt background for electrons and muons.

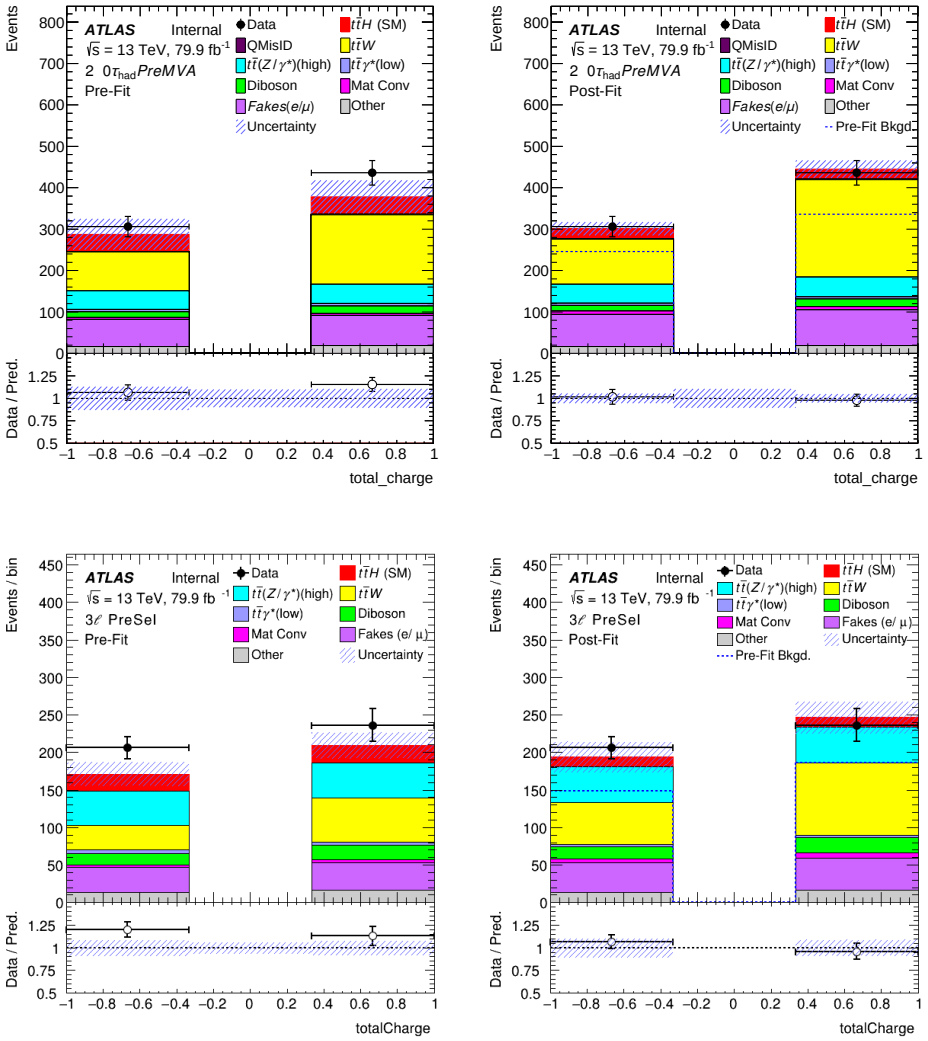


Figure 7.16: Pre-fit (***left***) and post-fit (***right***) distribution of total charge in 2ℓSS channel (***upper***) and 3ℓ channel (***lower***). The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel. Fakes (e/μ) represent the total heavy-flavour non-prompt background for electrons and muons.

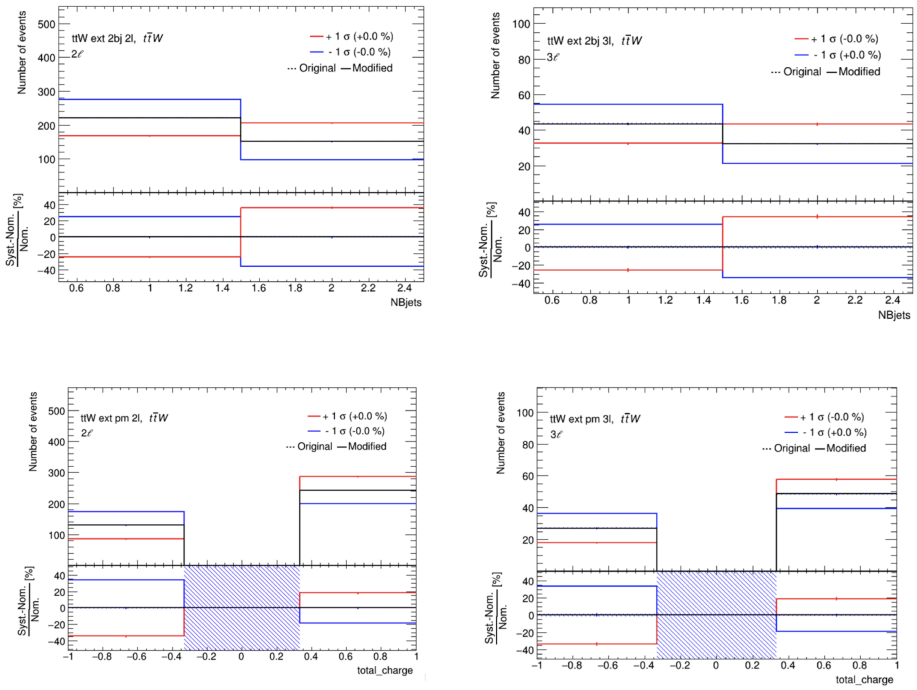


Figure 7.17: Variation of the number of events depending on extrapolation uncertainties in bins of b-jet multiplicity (***upper***) and total charge (***lower***) distributions for 2ℓSS (***left***) and 3ℓ (***right***) channels.

Figure 7.18 shows the overlayed bins of lepton charge and b-jet multiplicity for 2ℓSS and 3ℓ channels, and the agreement is well improved after the application of the additional uncertainties. The tensions observed in the *positive* charge and *2b-jet* pre-fit regions are solved in the post-fit regions. The detailed initial studies related to assigning these uncertainties are given in Appendix D.

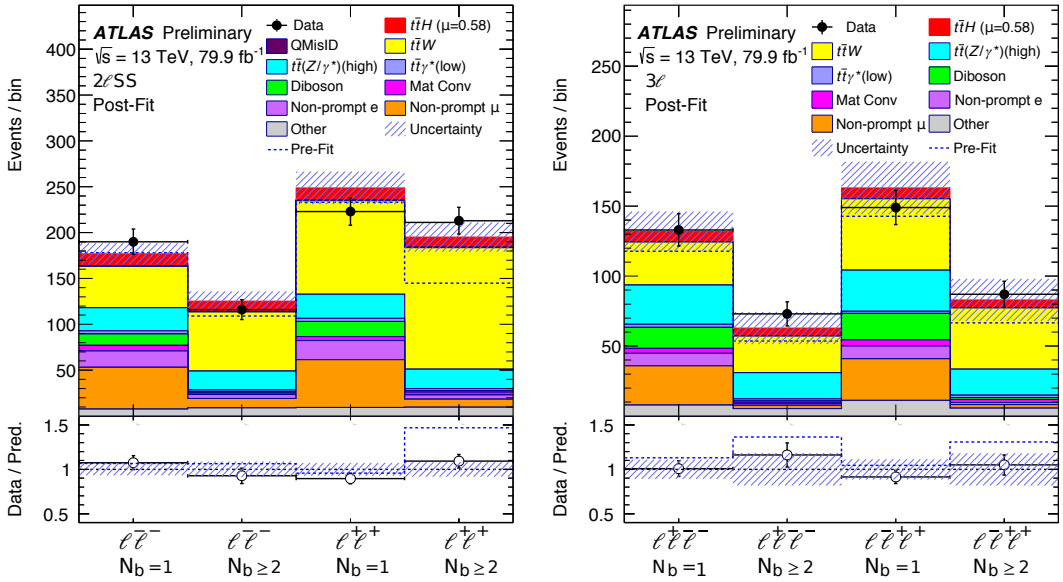


Figure 7.18: Distributions of the total charge and b-jet multiplicity split in four separate categories in 2ℓSS and 3ℓ channels. The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel.

7.5.4 Summary of systematic uncertainty sources

All systematic uncertainties entering to the final likelihood fit is listed in Table 7.9, grouped by their sources. The type of each source indicates whether the NP affects only normalization (type N) or both shape and normalization (not indicated) of the samples on which it acts. The amount of these components per source is listed in the last column.

Table 7.9: Sources of systematic uncertainties considered in the analysis. “N" means that the uncertainty is taken as normalisation-only for all processes and channels affected. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column. The PDF uncertainties are correlated between the $t\bar{t}H$ signal and the $t\bar{t}W$ background.

Systematic uncertainty	Components
Luminosity (N)	1
Pile-up modelling	1
Physics objects	
Electron	8
Muon	11
Tau	7
Jet energy scale and resolution	28
Jet vertex fraction	1
Jet flavour tagging	17
$E_{\mathrm{T}}^{\mathrm{miss}}$	3
Total (Experimental)	77
Data-driven background estimates	
Non-prompt light-lepton estimates (3ℓ, 3ℓ+1τ_{had})	1
Fake τ_{had} estimates	6
Electron charge misassignment	2
Total (Data-driven reducible background)	9
Template fit uncertainties	
Material conversions	1
Internal conversions	1
HF non-prompt leptons	18
LF non-prompt leptons	2
Total (Template fit)	22

Systematic uncertainty	Components
$\boldsymbol{t\bar{t}H}$ **modelling**	
Renormalisation and factorisation scales	3
Parton shower and hadronisation model	1
Higgs boson branching ratio	4
Shower tune	1
PDF	32
$\boldsymbol{t\bar{t}W}$ **modelling**	
Radiation	1
Generator	1
PDF	32
Extrapolation	4
$\boldsymbol{t\bar{t}Z/\gamma^*}$ **(high mass) modelling**	
Cross section (N)	2
Generator	1
Renormalisation and factorisation scales	3
Shower tune	1
$\boldsymbol{t\bar{t}}$ **modelling**	
Radiation	1
$\boldsymbol{WZ}$ **modelling**	
HF composition (N)	3
Shower tune	1
Other background modelling	
Cross section (N)	22
Total (Signal and background modelling)	120
Total (Overall)	218

7.6 Statistical model and results

The results of the statistical analysis for the $t\bar{t}H$ cross section measurement in multilepton final states is presented in this section. A maximum profile likelihood fit is performed, including all bins in the 25 event categories defined in the analysis (Section 7.3) and with the observed data in all regions, used as a constrain. In total seventeen regions from 2ℓSS and 3ℓ channels are used as CRs to determine or constraint different source of backgrounds (material conversions, internal conversions, heavy-flavour non-prompt backgrounds and irreducible backgrounds such as $t\bar{t}W$, $t\bar{t}Z$ and VV). The total event yield is used in thirteen of the control regions, and different kinematic variables are exploited for the rest of the four CRs. The remaining eight regions are defined as SRs to improve the $t\bar{t}H$ signal purity. In the $t\bar{t}H$ SRs of the $2\ell SS$, 3ℓ, and $1\ell + 2\tau_{\mathrm{had}}$ channels, a BDT discriminant is exploited while the total event yield is used in the remaining four signal regions, and the binning of all distributions is optimized for sensitivity.

The likelihood function $\mathcal{L}(\mu, \boldsymbol{\lambda}, \boldsymbol{\theta})$ of the signal strength parameter μ is defined as the yield for the $t\bar{t}H$ signal events in several regions, which is normalised to the SM predictions, where the

parameter $\boldsymbol{\lambda}$ is the normalisation factors (for $t\bar{t}W$ background, internal and material conversions and non-prompt background for electron and muon), and parameter $\boldsymbol{\theta}$ is defined as a set of NPs encoding systematic uncertainties in the signal and background expectations. The statistical uncertainty on the prediction represents the statistical uncertainty of the MC simulation events. This uncertainty is included in the form of additional nuisance parameter for each bin of the distributions, by using the Beeston-Barlow technique [194].

Profile likelihood ratio is used to define the test statistic q_0 as follows:

$$q_0 = -2\ln\Big(\mathcal{L}(0, \hat{\hat{\boldsymbol{\lambda}}}_0, \hat{\hat{\boldsymbol{\theta}}}_0)/\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\lambda}}_{\hat{\mu}}, \hat{\boldsymbol{\theta}}_{\hat{\mu}})\Big), \tag{7.8}$$

where $\hat{\mu}$, $\hat{\boldsymbol{\lambda}}_{\hat{\mu}}$, and $\hat{\boldsymbol{\theta}}_{\hat{\mu}}$ are the values that maximises the likelihood function, and $\hat{\boldsymbol{\lambda}}_0$ and $\hat{\boldsymbol{\theta}}_0$ are the parameter values that maximise the likelihood function, while fixing μ to zero. The test statistic is assessed with the RooFit package [195] and is used to describe the agreement of the observed data with the background-only hypothesis. The expected sensitivity of the analysis is obtained by fitting the model to an Asimov dataset, as the procedure of creating the dataset is described in Section 6.6.

Figure 7.19 shows the distributions of signal regions for the BDT discriminants, where the pre-fit predictions are indicated with dashed-lines and the uncertainty bands include all sources of systematic uncertainties described in Section 7.5.2. Figure 7.20 shows the final discriminant distribution of $t\bar{t}H$ signal yield (S), fitted background yield (B), as well as the data in all regions combined into bins of $\log_{10}(S/B)$. It is constructed by assessing the value of $\log_{10}(S/B)$ for every bin in the analysis, and merging the bins with similar values. Top panel shows the total fitted background, best fit $t\bar{t}H$ signal strength in red, and SM prediction in orange, as well as the distribution of data. The lower panel displays the difference between data and post-fit background-only prediction, divided by its uncertainty (dashed lines) in red for for best-fit value of signal strength and orange for SM prediction. The significance of the observed (expected) signal strength of $t\bar{t}H$ under the background-only hypothesis ($\mu = 0$) is 1.8 (3.1) standard deviations. The best-fit value of μ is:

$$\hat{\mu} = 0.58^{+0.26}_{-0.25}\,(\text{stat.})^{+0.19}_{-0.15}\,(\text{exp.})^{+0.13}_{-0.11}\,(\text{bkg. th.})^{+0.08}_{-0.07}\,(\text{sig. th.}) = 0.58^{+0.36}_{-0.33}\,. \tag{7.9}$$

The measured cross section for $t\bar{t}H$ production is evaluated in inclusive phase space by extrapolating the SM kinematics as follows:

$$\hat{\sigma}(t\bar{t}H) = 294^{+132}_{-127}\,(\text{stat.})^{+94}_{-74}\,(\text{exp.})^{+73}_{-56}\,(\text{bkg. th.})^{+41}_{-39}\,(\text{sig. th.})\ \text{fb} = 294^{+182}_{-162}\ \text{fb}\,. \tag{7.10}$$

The predicted SM cross section is $\sigma(t\bar{t}H) = 507^{+35}_{-50}$ fb computed at NLO in QCD and EWK couplings [22]. The measured cross section is compatible at 1.5 σ level with the SM prediction within uncertainties.

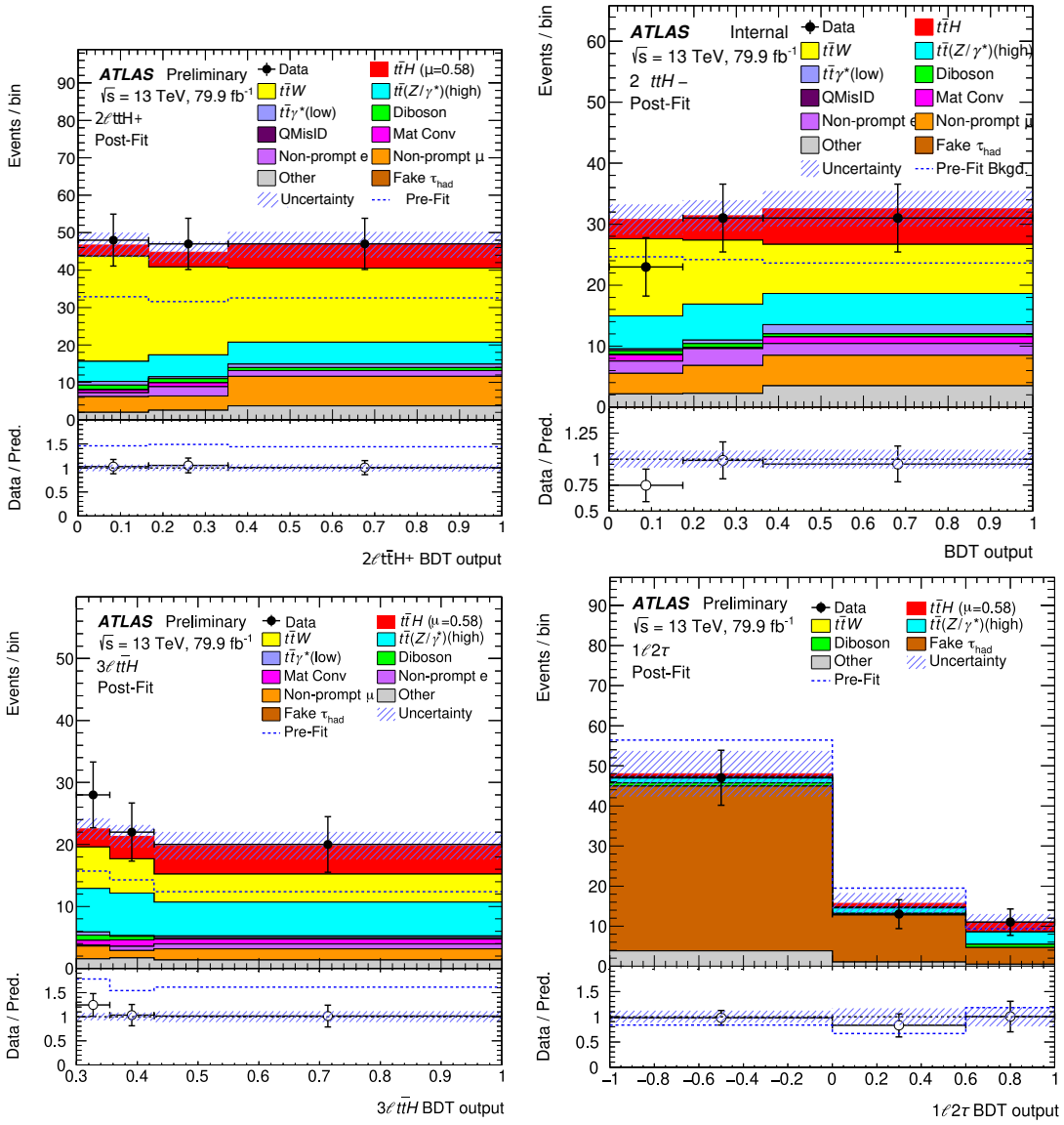


Figure 7.19: Signal region distributions for the BDT discriminants used in likelihood fit for 2ℓttH+ (***upper-left***), 2ℓttH- (***upper-right***), 3ℓttH (***lower-left***) and 1ℓ+2τ_{had} (***lower-right***) channels.

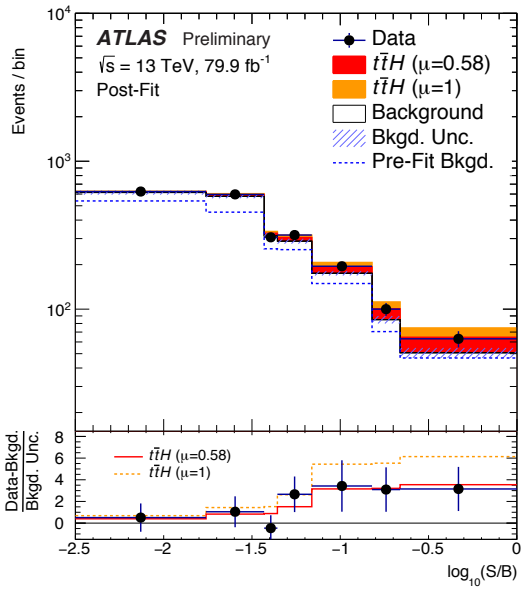


Figure 7.20: Total event yields in a bins of $\log_{10}(S/B)$ for data, estimated background, the $t\bar{t}H$ signal yield for SM prediction and best fit value.

Simultaneous fit is performed with several normalisation factors dedicated to reducible and irreducible backgrounds. The summary of these NFs are given in Table 7.10.

Backgrounds	Normalisation Factors
$\hat{\lambda}^{2\ell\mathrm{LJ}}_{t\bar{t}W}$	$1.56^{+0.30}_{-0.28}$
$\hat{\lambda}^{2\ell\mathrm{HJ}}_{t\bar{t}W}$	$1.26^{+0.19}_{-0.18}$
$\hat{\lambda}^{3\ell}_{t\bar{t}W}$	$1.68^{+0.30}_{-0.28}$
$\hat{\lambda}^{\mathrm{IntC}}_{e}$	0.83 ± 0.32
$\hat{\lambda}^{\mathrm{MatC}}_{e}$	1.61 ± 0.48
$\hat{\lambda}^{\mathrm{had}}_{e}$	1.12 ± 0.38
$\hat{\lambda}^{\mathrm{had}}_{\mu}$	1.20 ± 0.18

Table 7.10: Normalisation factors for several backgrounds, evaluted from the simultaneous signal strength fit to data.

The normalisation factors for $t\bar{t}W$ background are compatible through the three regions around 0.3 σ ~ 1.3 σ level and systematically above unity, which indicates the higher data yields than the $t\bar{t}W$ theoretical cross section (including the higher order corrections decribed in Section 7.1.2). Since the introduced uncertainties on $t\bar{t}W$ phase space are not affecting the total yields but the shapes, the normalisation factors represent the scale factor for events in $t\bar{t}W$ phase space.

Figure 7.21 shows the regions used in the likelihood fit. In all categories, the observed yields

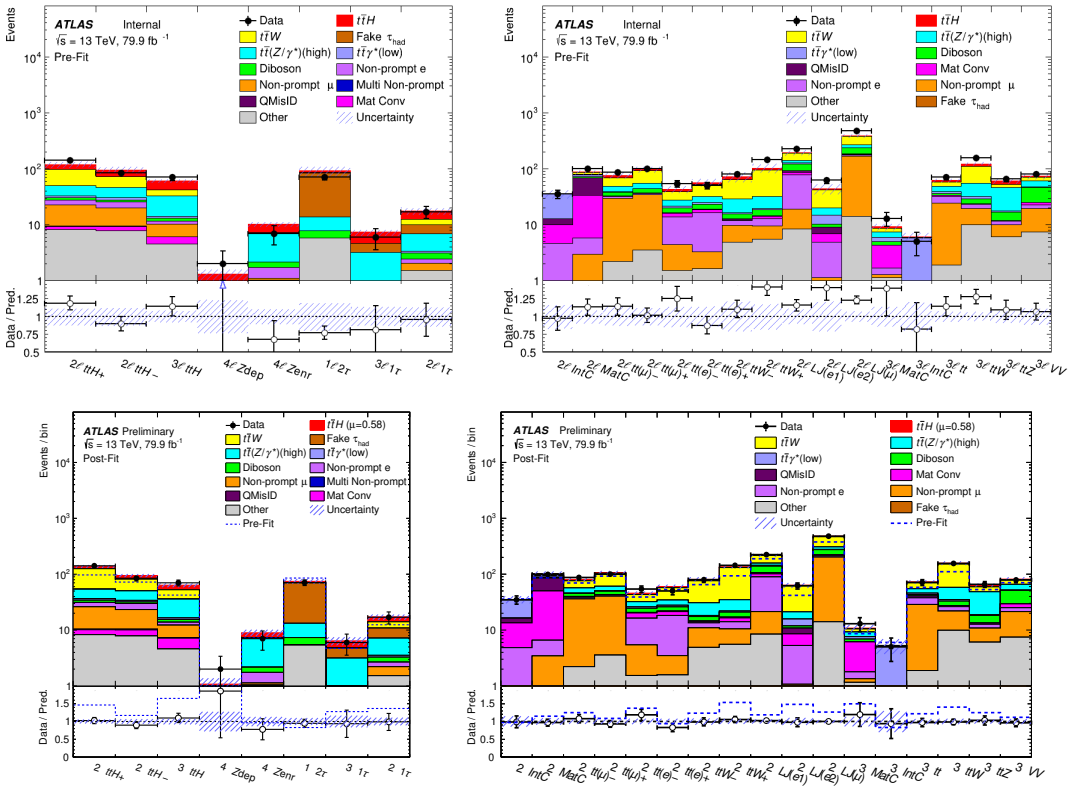


Figure 7.21: Comparison of the data and prediction for the event yields in the eight $t\bar{t}H$ signal regions (***left***) and 17 control-region categories (***right***) for pre-fit (***upper***) and post-fit (***lower***).

agree with the fitted prediction within uncertainties. The total uncertainties are decreased in the post-fit distributions compared to the pre-fit distributions due the correlations between NPs and their constraints.

The impact of a corresponding NP on the fitted signal strength $\mu_{t\bar{t}H}$ is computed by re-evaluating the likelihood fit when the NP fixed to the specific values. The pre-fit impact is calculated by varying the related NP by $\Delta\theta = \pm 1\sigma$, while the post-fit impact can be evaluated by variying $\Delta\hat{\theta}$. The most impacting 15 NPs in the fit is ranked in the Figure 7.22. The upper axis shows the $\Delta\mu$ obtained from the NPs. The empty (filled) rectangles with cyan (blue) colours display the pre-fit (psot-fit) impact on $t\bar{t}H$ signal strength. The pulls are determined as the difference between the best-fit value ($\hat{\theta}$) and the nominal value of related NP (θ_0) divided by its pre-fit uncertainty, and are shown by the black points. The scale of the pull is shown on the lower axis. The vertical black lines illustrate the post-fit NP uncertainty ($\Delta\hat{\theta}$). Free-floating normalisation factors for $t\bar{t}W$ are also shown with red vertical lines for which the pre-fit impact on signal strength is not defined. Some of the NPs are pulled and/or constrained and more detailed version of the plot is given in appendix Figure E.1. Among them, the NPs associated with the b-jet multiplicity in 2ℓSS channel is pulled by +0.75 and the uncertainty is reduced by factor of 3. While the NP associated with the total charge extrapolation is pulled by +33

and the uncertainty is reduced by factor of 2. The correlations between the NPs and the signal strength, with a lower threshold of 20% for the combined fit, are shown in Figure 7.23.

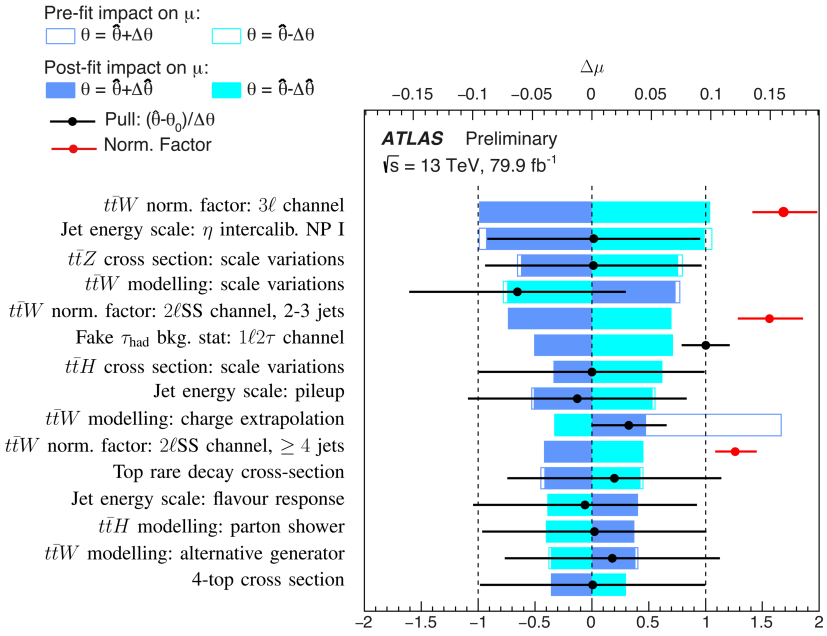


Figure 7.22: The most impacting 15 NPs entered to the fit according to their effect on the signal strength $\mu_{t\bar{t}H}$. The upper axis shows the $\Delta\mu$ obtained from the NPs. The empty (filled) rectangles with cyan (blue) colours display the pre-fit (psot-fit) impact on $t\bar{t}H$ signal strength. The pulls are shown with the black points and its scale is shown on the lower axis. The vertical black lines illustrate the post-fit NP uncertainty. Free-floating normalisation factors for $t\bar{t}W$ are also shown with red vertical lines for which the pre-fit impact on signal strength is not defined.

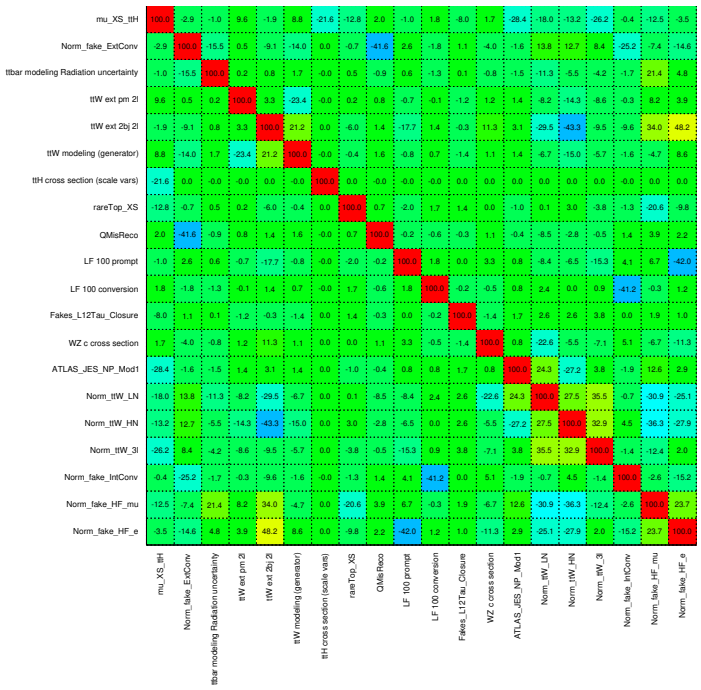


Figure 7.23: Correlation matrix between nuisance parameters and the signal strength, with a lower threshold of 20% for the combined fit.

In Table 7.11, the uncertainties contributing to the measured signal strength are shown. The breakdown of the contributions is evaluated by fixing all NPs within the corresponding group to their best-fit values and repeating the fit. Finally the impact of the total systematic uncertainty in the group is assessed by substracting the uncertainty in quadrature from the total nominal uncertainty, while the statistical uncertainty is obtained by fixing all the NPs except the free-floating background normalisation factors. The leading contribution to the measured uncertainty emerge from the jet energy scale and resolution due to the presence of jets in both control and signal regions, and due to the assumptions made in the reconstruction of jet procedure. This is followed by the uncertainties related to $t\bar{t}W$ prediction and the normalisation of the $t\bar{t}Z/\gamma^*$ background.

Table 7.11: The breakdown of the grouped systematics uncertainties on signal strength. The simulation sample size uncertainty includes effects from statistical uncertainties in nominal simulation samples. Statistical uncertainties from data-driven background estimates are included within the experimental uncertainties. Due to rounding effects and the correlations between the different sources of uncertainties, the total systematic uncertainty is different from the sum in quadrature of the individual sources.

Uncertainty source	$\Delta\hat{\mu}$	
Jet energy scale and resolution	+0.13	−0.13
$t\bar{t}Z/\gamma^*$ (high mass) modelling	+0.09	−0.09
$t\bar{t}W$ modelling (radiation, generator, PDF)	+0.08	−0.08
Fake τ_{had} background estimate	+0.07	−0.07
$t\bar{t}W$ modelling (extrapolation)	+0.05	−0.05
$t\bar{t}H$ cross section	+0.05	−0.05
Simulation sample size	+0.05	−0.05
$t\bar{t}H$ modelling	+0.04	−0.04
Other background modelling	+0.04	−0.04
Jet flavour tagging and τ_{had} identification	+0.04	−0.04
Other experimental uncertainties	+0.03	−0.03
Luminosity	+0.03	−0.03
Diboson modelling	+0.01	−0.01
$t\bar{t}\gamma^*$ (low mass) modelling	+0.01	−0.01
Charge misassignment	+0.01	−0.01
Template fit (non-prompt leptons)	+0.01	−0.01
Total systematic uncertainty	+0.25	−0.22
Intrinsic statistical uncertainty	+0.23	−0.22
$t\bar{t}W$ normalisation factors	+0.10	−0.10
Non-prompt leptons normalisation factors (HF, material conversions)	+0.05	−0.05
Total statistical uncertainty	+0.26	−0.25
Total uncertainty	+0.36	−0.33

7.7 Fit validation studies

Studies of the compatibility of the results for both $t\bar{t}H$ signal strength and background normalisation factors is performed. Some of the studies are discussed in this chapter. The $t\bar{t}H$ signal strength found to be robust across the cross-checks, as well as the normalisation factors.

7.7.1 Multi-μ fit model

Different signal strength parameters are introduced for each channel in the fit model. The results of this so-called *multi-*μ fit, together with the nominal fit results are given in Figure 7.24. The purpose of this study is to test the compatibility of the measurement across the fitted channels.

The compatibility calculation is done for the two signal strengths in the multi-μ fit through evaluating a χ^2 test with five degree of freedom (indicating the difference between signal strengths in the multi-μ fit and the nominal fit model) which is calculated by using the negative logarithm of the likelihood fit for the nominal and multi-μ at the corresponding best-fit points. The results between the channels with respect to nominal fit are as large or larger than the 0.003 sigma level. Observed signal strengths of individual channels have some differences among

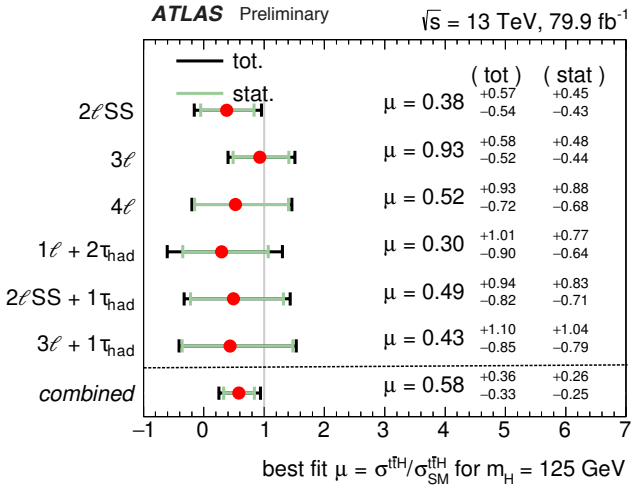


Figure 7.24: The observed best-fit values of the $t\bar{t}H$ signal strength and their uncertainties by analysis channel, together with the nominal fit results. Each channel μ values are evaluated from one simultaneous fit to the each parameter floating independently.

them, therefore in addition to multi-μ fit, a 2-μ fit is performed to evaluate the overall impact of two groups: non-tau and tau channels. Table 7.12 shows the results of this fit setup, and Figure 7.25 shows the impact of the NPs for each group to the fitted μ signal strength.

Group.	Observed $t\bar{t}H$ significance [σ]: stat.+syst. (stat.)	Expected $t\bar{t}H$ significance [σ]: stat.+syst. (stat.)
Combined Tau	0.80 (0.8)	1.79
Combined Non-Tau	1.77 (2.52)	2.67

Table 7.12: Observed and expected $t\bar{t}H$ significance of the combination of non-tau channels (2ℓSS, 3ℓ, and 4ℓ) and tau channels (1ℓ+$2\tau_{\text{had}}$, 2ℓSS+$1\tau_{\text{had}}$, and 3ℓ+$1\tau_{\text{had}}$).

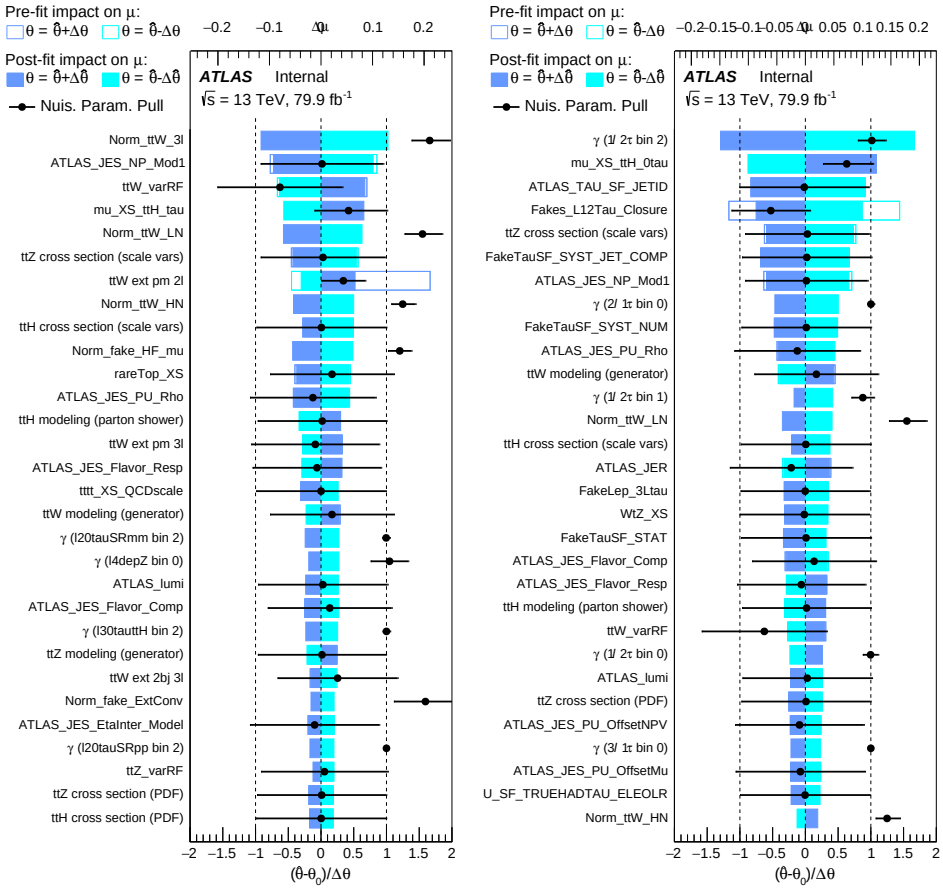


Figure 7.25: Impact of NPs on fitted signal strength uncertainty and best-fit values together with their uncertainties for combined non-tau (***left***) and tau (***right***) channels. Gammas refer to MC statistical uncertainty of the NP.

7.7.2 Expected sensitivity results

The expected sensitivity studies are performed to verify the results from real data. The expected sensitivity of the analysis is obtained by fitting the model to an Asimov dataset, when the procedure of building an Asimov dataset is described in Section 6.6. The extracted normalisation factors, the values of the nuisance parameters, and their correlations from CR-only fit are used to build the Asimov dataset (so called *Hybrid Asimov*). This Hybrid-Asimov is an estimate to present expected sensitivity, and provides the total uncertainty from the available statistics and systematics. The results of the expected signal strength measurement $\mu_{exp.} = 1^{+0.39}_{-0.35}$ corresponding to 3.06 σ. The normalisation factors from CR-only fit model, and correlation of the NPs and normalisation factors are given in Figure 7.26.

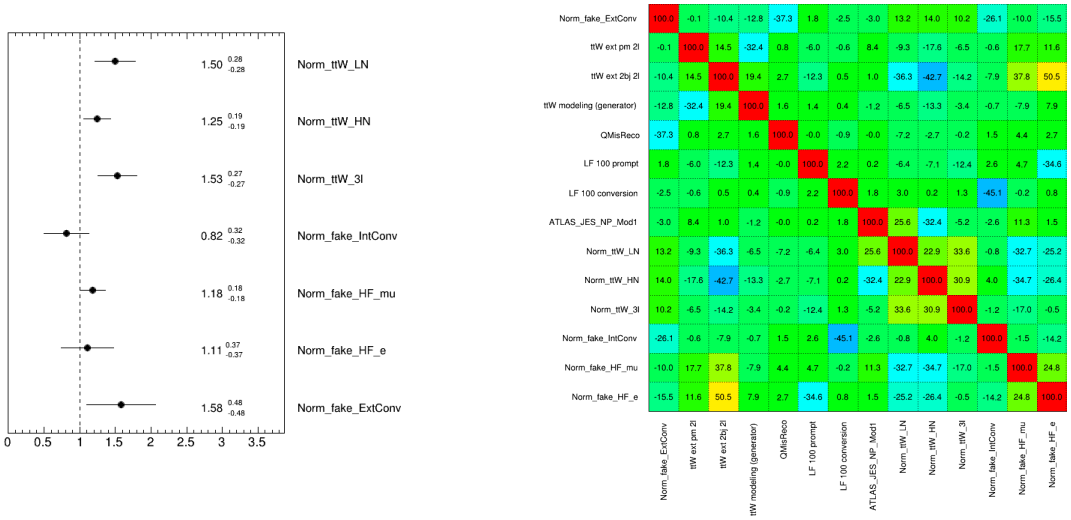


Figure 7.26: Normalisation factors (***left***) and correlation matrix (***right***) for the CR-Only fit results with updated $t\bar{t}W$ uncertainty model.

7.7.3 Single $t\bar{t}W$ normalisation factor

The fit setup has changed to adapt one $t\bar{t}W$ normalisation factor and the results are evaluated to test the compatibility of the NFs. Figures 7.27-7.28 show the comparison of the normalisation factors between the single $t\bar{t}W$ and nominal model. There are few differences observed between two fit setups. The scale variations for $t\bar{t}W$ sample is slightly pulled for the 3 normalisation factor fit model, and the reason could be explained as follows: Due to the lower $\hat{\lambda}^{2\ell\mathrm{LJ}}_{t\bar{t}W}$ value, there is a tension in 2ℓSS signal regions, which is not there in the 1 normalisation factor fit model and the scale variation NP has a significant effect on the BDT shape in the 2ℓSS signal region, which is compensating that effect in the 1 NF fit model. The scale variation uncertainty for each region entering to likelihood fit is summarised in the Figure 7.29. The compatibility of this fit model with the nominal model is 28%, which corresponds to 0.59 standard deviations.

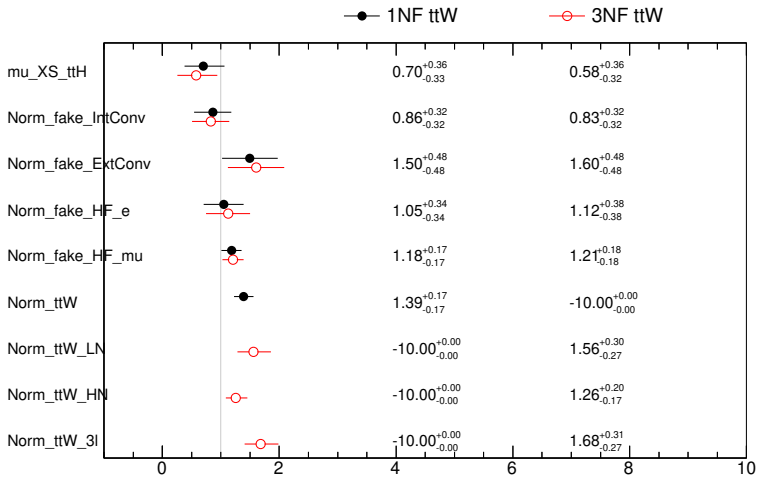


Figure 7.27: Comparison of normalisation factors and signal strength for 1 normalisation factor $t\bar{t}W$ model between nominal model. The value -10.0 is default placeholder value from the plotting framework.

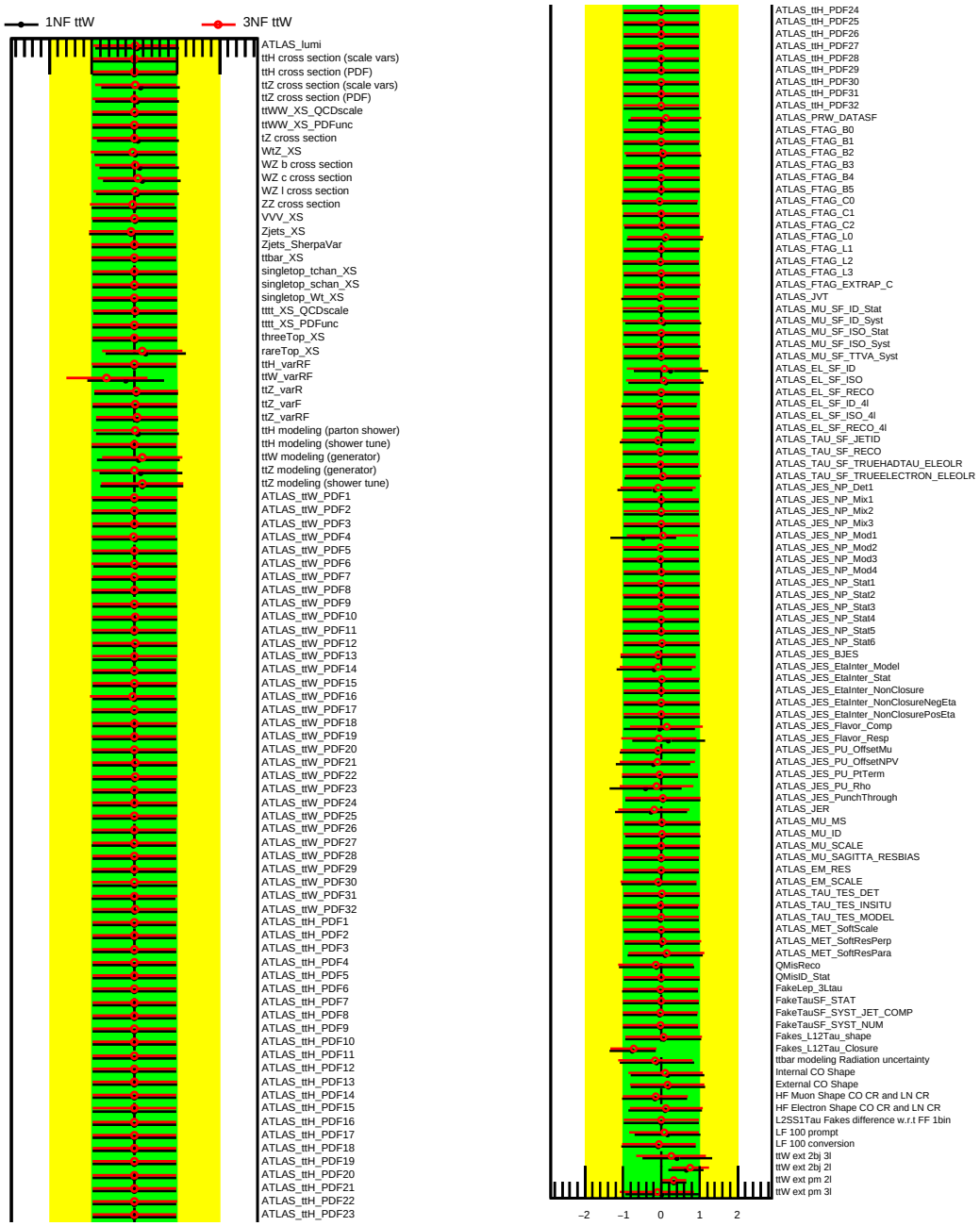


Figure 7.28: Comparison of NP pulls for 1 normalisation factor $t\bar{t}W$ model between nominal model. x-axis is arranged in 1 σ bins. The post-fit uncertainty is given with black or red lines and the pulls are given with dots.

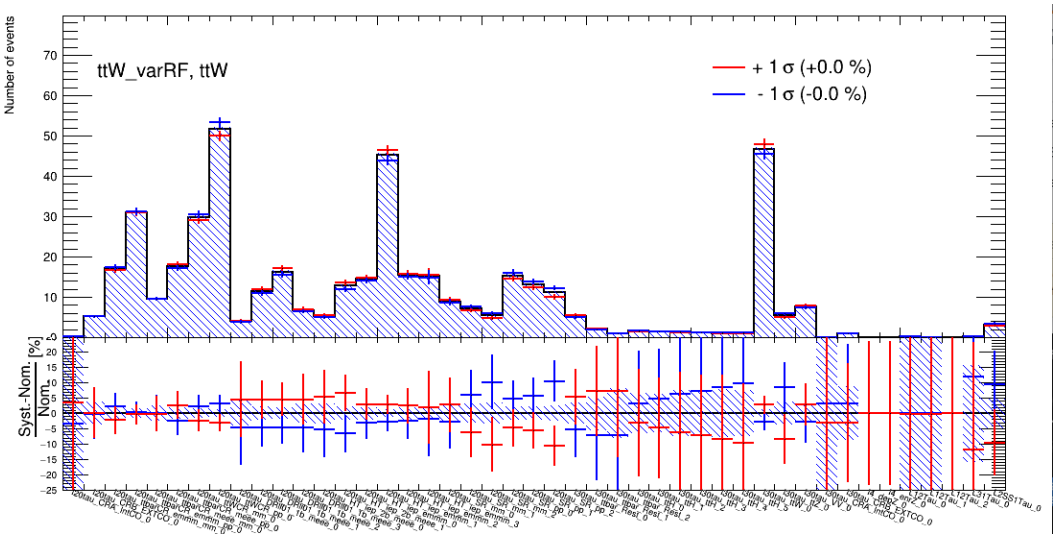


Figure 7.29: Renormalisation and factorisation scale shape uncertainty for each fit region.

7.7.4 Control region fit comparisons

In order to validate the consistency of normalisation factors and the pulls of the NPs, the comparison is performed between the nominal fit, control region fit and multi-μ fit, as shown in Figures 7.30-7.31. This study shows the consistency of the larger $t\bar{t}W$ normalisations compared to the theoretical predictions along with the different fit setups, and the good estimate of the non-prompt lepton backgrounds. It is seen that CR only and nominal fit models leads to higher $t\bar{t}W$ normalisation factor for 3ℓ channel compared to the 1 $t\bar{t}W$ NF model, which yields to lower $\mu_{t\bar{t}H}$. The explanation of this behaviour is that in the 1 $t\bar{t}W$ model, the 3ℓ $t\bar{t}W$ effect is absorbed by the $t\bar{t}H$ signal strength.

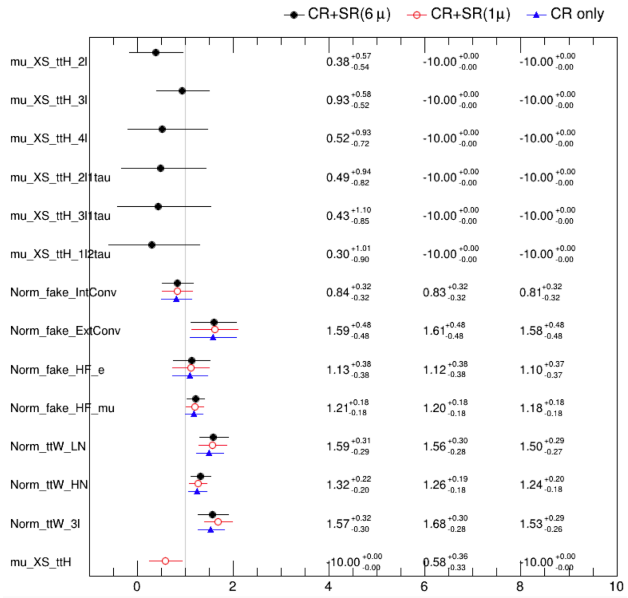


Figure 7.30: Comparison of normalisation factors and signal strength for multi-*mu* (black), nominal (blue) and contol region only (red) fit models. The value -10.0 is default placeholder value from the plotting framework.

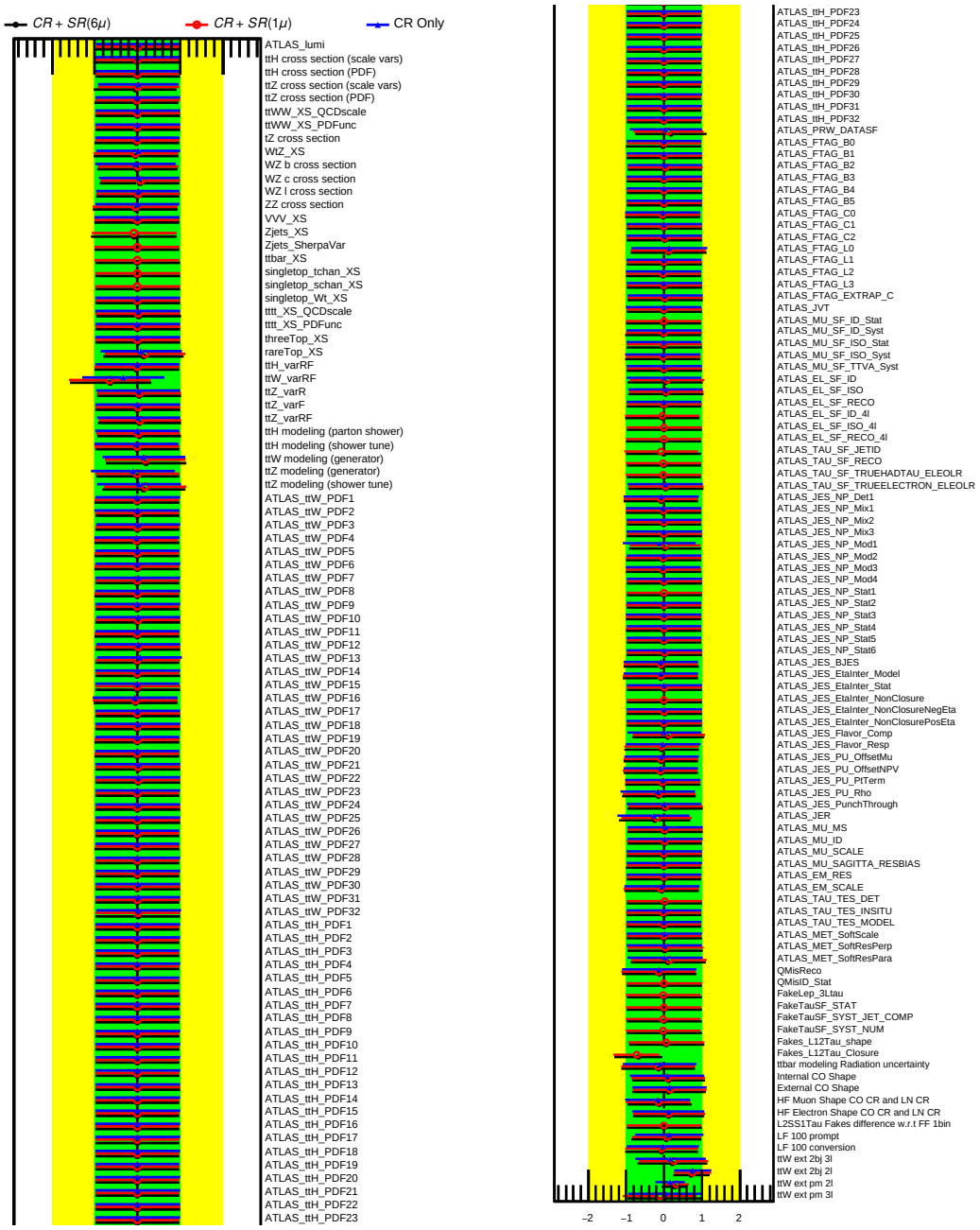


Figure 7.31: Comparison of NP pulls for multi-*mu* (black), nominal (blue) and contol region only (red) fit models.

7.8 Possible improvements

As discussed in Section 6.3, the way to prove a theory is to construct an optimal test statistics based on the likelihood ratios of hypotheses. There are several ways to build the likelihoods, and Matrix Element Method (MEM) is one way to do it by using the matrix element calculations of the corresponding hypotheses. In this method, the likelihoods of each event for given two hypotheses is evaluated and this results to an event based weight, which shows how compatible an event with the given hypotheses.

Likelihood is calculated with approximation of probability density of the reconstructed objects under the given hypothesis, which is proportional to the differential cross section of the process. In this study, the signal hypothesis (L_S) is $t\bar{t}H$ SR with four number of jets, two b-jets and two tight leptons. The background hypothesis (L_B) is $t\bar{t}$ process with two jet, two b-jet and one tight lepton. The choice of the regions is done to simplfy the calculation of the weights for this preliminary study. Therefore, the likelihoods are defined as follows:

$$L_i = \sum \int \frac{f_1\left(x_1, Q^2\right) f_2\left(x_2, Q^2\right)}{|\vec{q}_1||\vec{q}_2|} \left|\mathcal{M}_i\left(\mathbf{Y}\right)\right|^2 T\left(\mathbf{X};\mathbf{Y}\right) \mathrm{d}\mathbf{\Phi}_n\left(\mathbf{Y}\right) \ . \tag{7.11}$$

It involves the parton distribution functions f_1, f_2, matrix element (M_i) for the phase space configuration of Y at parton level (LO) and T is the transfer function which describes the probabilty function for a given reconstructed object to be originating from a parton level configuration, typically can be calculated from truth matching of the particles and describes parton shower, hadronization, and detector and reconstruction effects. The combination of signal and background likelihoods, which are evaluated on both data and MC, results in a powerful discrimination variable:

$$\mathrm{MEM}_{D1} = log_{10}\left(L_S\right) - log_{10}\left(L_B\right) \ . \tag{7.12}$$

The full phase space integration is a difficult numerical problem and computationally expensive due to the presence of peaks in the integrand (propagators in the matrix element, transfer functions on angles and energies). Solving it in a reasonable amount of time requires the use of both adaptive MC integration algorithms and smart phase-space parametrisations. Additionally the presence of permutations in the final state yields even longer duration for the final likelihood calculations. Therefore, the software package called *MoMEMta* [196] is used for calculation of the likelihoods. Within this tool, the transfer functions, change of variables for the phase-space parametrisation are done, and a plugin is used to extract the ME from MadGraph. Finite number of block definitioins are exploited in order to align the phase-space along the peaks while preserving the momentum conservation laws.

The discriminant given in Equation 7.8 is shown in Figure 7.32. The discrimination is observed between the two processes and this output can be used either as a final discriminant or an input for the multivariative techniques used in this channel. Here the TFs was not included. Even though the application is not optimal, this preliminary result shows that the possible

improvement in the discrimination of $t\bar{t}H$ versus $t\bar{t}$ background can be obtained by MEM method.

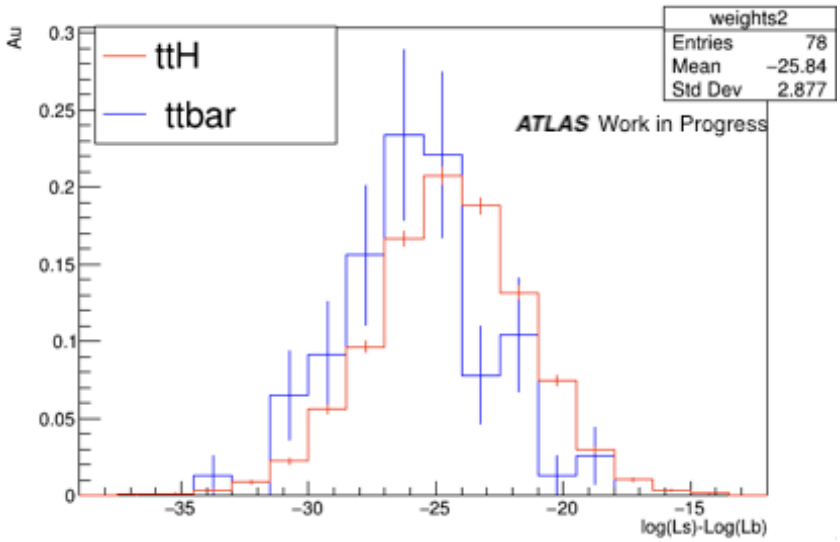


Figure 7.32: Distribution of the $t\bar{t}H$ signal and $t\bar{t}$ processes in 2ℓSS signal region a function of the logarithm of the signal minus background likelihoods. Both processes are normalized to the unity.

CHAPTER 8

Search for Higgs boson pair production in the 2ℓSS channel

This chapter discusses the search for the Higgs pair production (diHiggs production), mainly focusing one of the sensitive category in the multilepton analysis group, 2ℓSS channel. At the time of writing the thesis, the studies have been on-going towards a full combination of diHiggs production using 2015-2018 data, corresponding to an integrated luminosity of the 139 fb^{-1} at a center-of-mass energy of $\sqrt{s}$ = 13 TeV.

All the studies shown in this chapter were performed by me, and my main contribution to diHiggs multilepton group is reconstructing the analysis framework for the needs of whole analysis channels and being responsible for the production of samples, as well as being internal note editor of the analysis group.

Multilepton analysis is using the same set of data and MC samples as in $t\bar{t}H$ studies, described in Section 7.1. The nominal sample used to model diHiggs signal is generated using MadGraph5_aMC@NLO [101] for ME and Herwig++[106, 107] for PS with the CT10 [178] PDF set. The object and event preselections are as well similar, however in this study, the trigger matching of the leptons are not required in order to higher the statistics in the optimisation studies (Section 7.2).

8.1 2ℓSS channel studies

2ℓSS final state particularly is interesting since the its contribution to the overall significance is one of the highest due to relatively higher S/B ratio among other channels in diHiggs multilepton analysis group. In 2ℓSS channel, the most common decay process is the Higgs bosons decay into two W bosons, one being off-shell, and followingly the two W bosons arising from different Higgs boson decays semileptonically. Therefore the ideal signal region topology of 2ℓSS channel is two same-sign leptons, four reconstructed jets with $E_{\mathrm{T}}^{\mathrm{miss}}$ which is a result of the double neutrino system arising from the leptonically decaying W bosons. Figure 8.1 illusturates the 2ℓSS statistical significance ($s/\sqrt{b}$) as a function of number of jets ($2 \leq N_{\mathrm{jets}} \leq 5$)

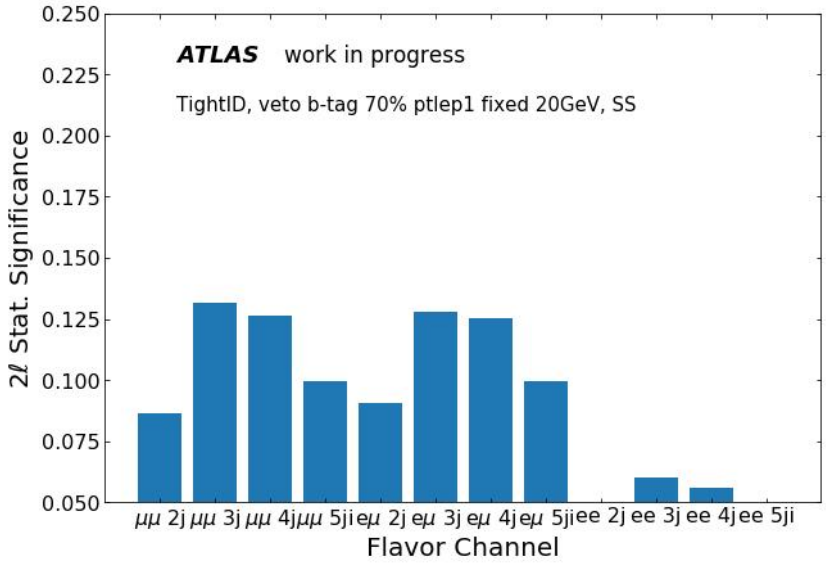


Figure 8.1: Statistical significance of the 2ℓSS channel split in flavour compoisition in bins of number of jets.

in ee, $e\mu$ and $\mu\mu$ channels. It can be seen that the highest sensitivity is coming from $N_{\text{jets}} = 3, 4$ regions, while the ee channel having the least contribution to the significance of 2ℓSS channel.

Giving the b-jet veto in the signal regions, the non-prompt lepton contribution from $t\bar{t}$ process is highly reduced, however it is still non-negligible due to its high cross section.

The SM processes that are contributing with prompt leptons, such as VV, $t\bar{t}V$, $V\gamma$ and tV ($V = W\,Z$) are described using MC simulations. Sensitivity estimates described in this section are based on MC simulation of non-prompt leptons. Therefore the processes such as $t\bar{t}$ and $Vjets$ which are contributing mainly through non-prompt lepton, described by MC without their estimations since it was not finalised at the time of writing the thesis.

In order to find the best estimation method for non-prompt lepton background, the origin of the processes needs to be known. A study has been performed to evaluate the non-prompt lepton background composition in signal region of 2ℓSS channel, using the truth information in MC simulation. The sources of the non-prompt lepton backgrounds are similar to the $t\bar{t}H$ case, where the heavy-flavour decay is expected to be the largest, subsequently the photon conversions (internal and material) and others such as π^0 decays.

The composition matrices are produced for 2ℓSS ee, $e\mu$ and $\mu\mu$ final states using the tight lepton selection given in Section 7.2.3. Figure 8.2 shows the composition of non-prompt background sources for each background process in ee, $e\mu$ and $\mu\mu$ channels, respectively. Sum of each row is 100% in the right set of plots, sum of each column is 100% in the middle set of plots and the left set of plots are showing the fractions of background contributions in whole phase space. Each set of plot is pointing out different aspects of the background sources.

It can be seen from the figures that the VV, VVV and $t\bar{t}W$ SM processes are contributing mainly with prompt leptons, whereas the $t\bar{t}$ process mainly with non-prompt heavy-flavour

decays. In the elecron channels, the sources of backgrounds are distributed in the different processes, such as conversions arising both from $t\bar{t}$, $V\gamma$ and $Vjet$ processes. Another important background source, electron with mis-assigned charge, is mainly rising from the $t\bar{t}$ and $Zjets$ physics processes. In the case of $\mu\mu$ channel, non-prompt lepton backgrounds are coming from $t\bar{t}$ with heavy-flavour decays. Looking at these results, the estimation method must be choosen by taking into account each sources to define CRs where there is high purity of the corresponding dominant background process.

The preliminary object working point optimisation studies are performed in 2ℓSS channel. Therefore, the scanning of the p_T, PLV working point and QMisID BDT working point requirements are performed for both leptons and the comparison of the statistical significance is performed. Figure 8.3 illustrates that the selection of PLV working point is improving the significance of the channel since the tool is targetting to reduce the non-prompt leptons from heavy-flavour decays. Furthermore, the effect of the QMisID is evaluated and found out that the choice of working point does not affect the significance of the channel drastically since the ee does not contribute largely to the 2ℓSS channel (Figure 8.4).

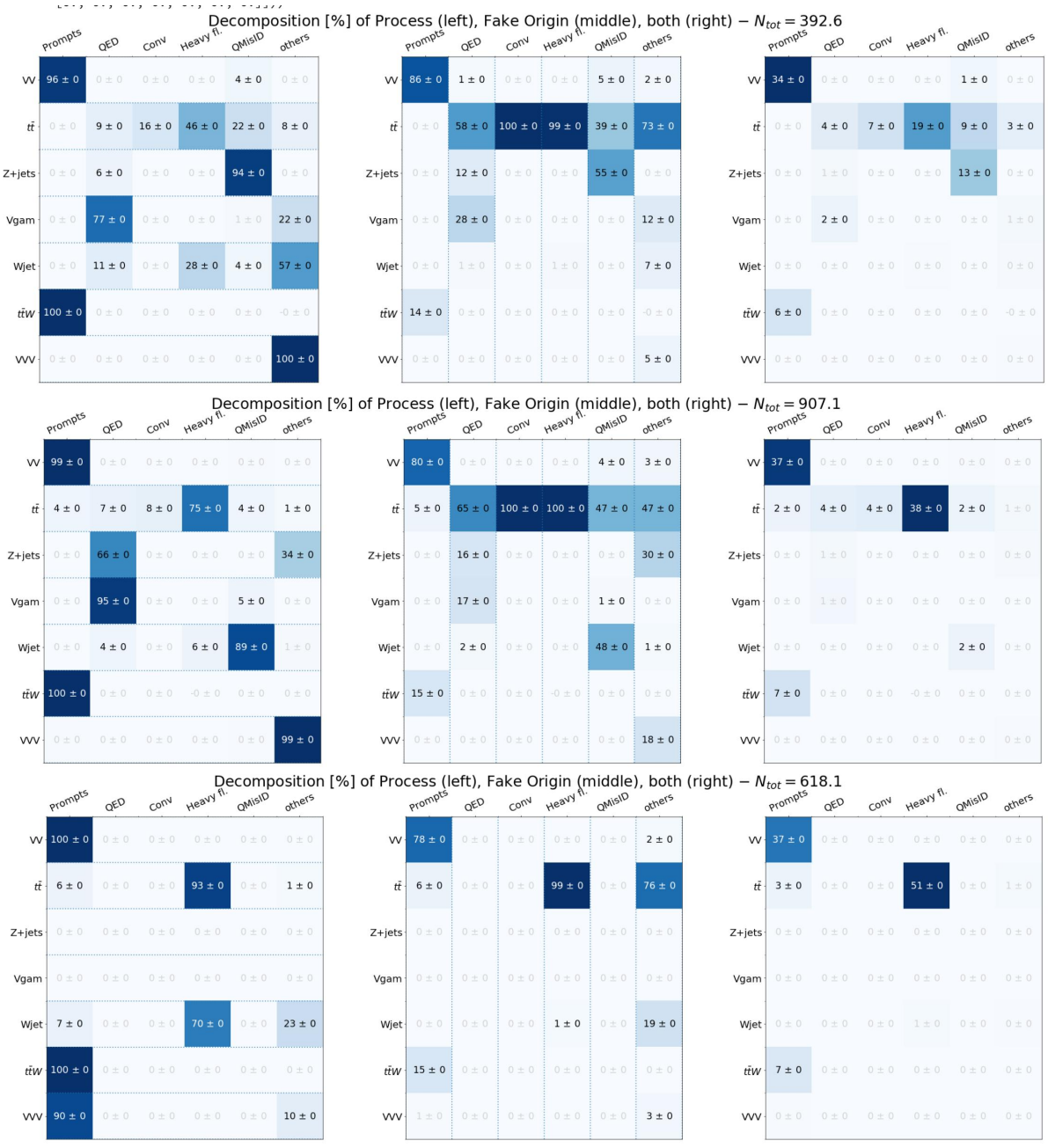


Figure 8.2: Background composition matrices in 2ℓSS ee (***upper***), $e\mu$ (***middle***), $\mu\mu$ (***lower***) channels.

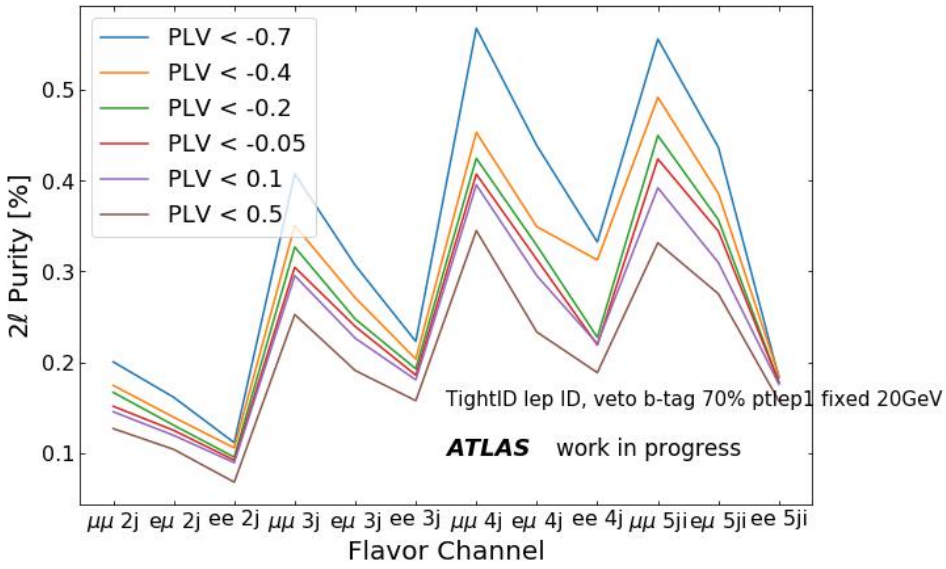


Figure 8.3: Purity of the diHiggs signal in 2ℓSS as a function of PLV working point in bins of number of jets.

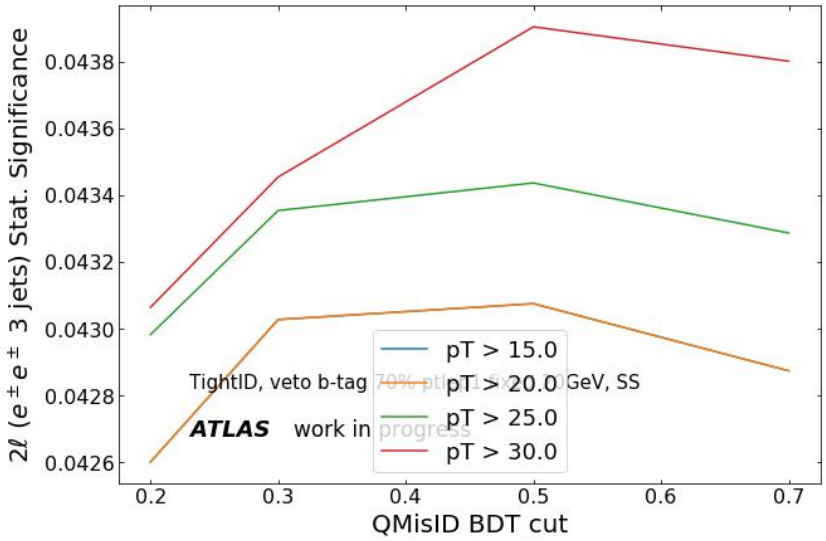


Figure 8.4: The impact of QMisID BDT selection on 2ℓSS (ee) channel as a function of p_T.

8.2 Template fit method in 2ℓSS channel

Template fit method is applied to 2ℓSS channel to estimate the non-prompt light lepton background, as described in Section 7.4.2. In total five CRs are used to extract the normalisation factors; one each for internal and external conversions and three low jet multiplicity regions (2-3 jets) where the kinematic variables such as ΔR_{ll} and $H_{T,lep}$ are used. The extracted normalisation factors from the data corresponding to integrated luminosity of 139 fb^{-1} are given in Table 8.1. The fitted CRs are given in Figure 8.6 and 8.5 and several kinematic distributions are given in Appendx F.

Backgrounds	Normalisation Factors
$\hat{\lambda}_e^{\text{IntC}}$	$1.08^{0.32}_{0.27}$
$\hat{\lambda}_e^{\text{MatC}}$	$2.01^{0.47}_{0.42}$
$\hat{\lambda}_e^{\text{had}}$	$0.71^{0.21}_{0.20}$
$\hat{\lambda}_\mu^{\text{had}}$	$0.83^{0.21}_{0.20}$

Table 8.1: Normalisation factors for backgrounds evaluted from the CR fit to data.

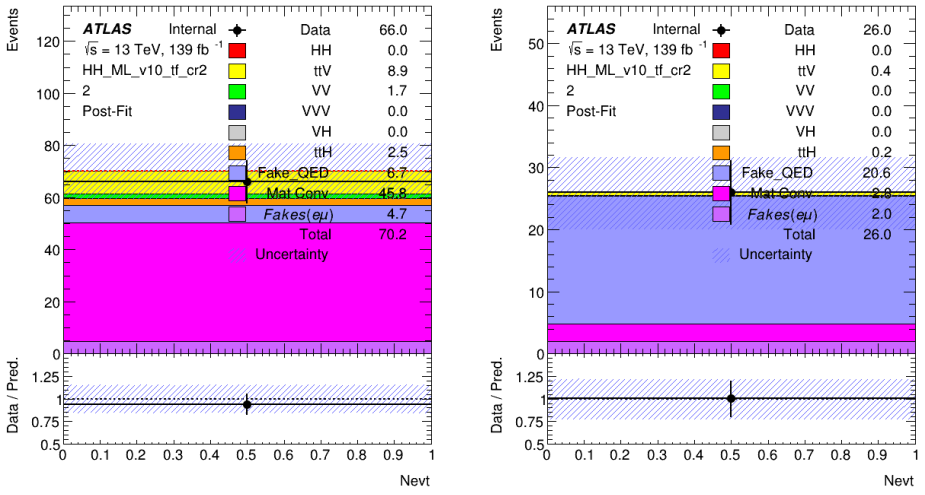


Figure 8.5: The event count of external (***left***) and internal (***right***) conversion CRs in 2ℓSS channel for post-fit.

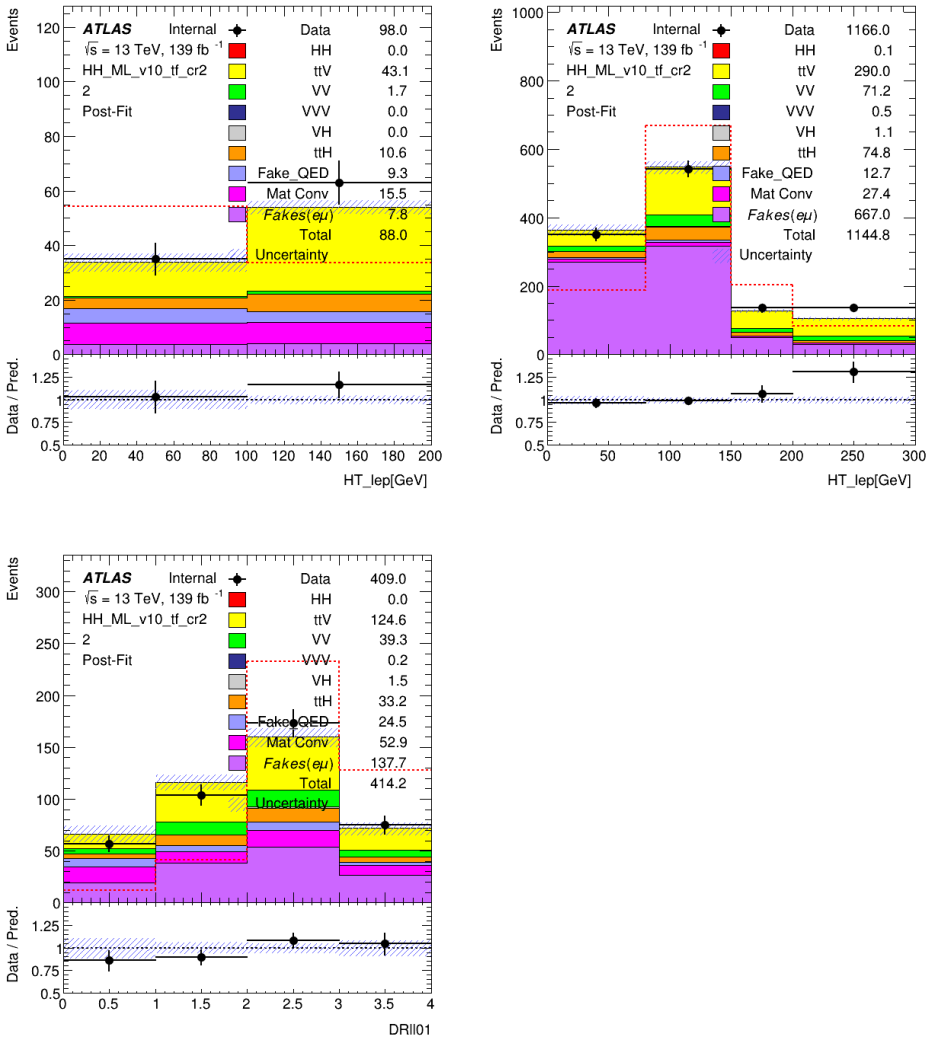


Figure 8.6: Distributions of $H_{\mathrm{T,\,lep}}$ in $\mu e + ee$ (***upper,left***), in $e\mu + \mu\mu$ (***upper,right***) and ΔR_{ll} in $\mu e + ee$ (***lower,left***) region in 2ℓSS channel for post-fit.

8.3 Results

As with the selections presented above, the expected SM upper-limit on cross section at 95% CL is calculated by using *Asimov* dataset, corresponding to a total integrated luminosity of 139 fb^{-1} and the results are presented in Table 8.3. Since the non-prompt contribution is modelled by MC simulations, the overall 20% uncertainty is assigned to the background processes, and the normalisation factors from CR fit are included.

There are several different approaches applied to the process of samples between the ones used for the optimisation studies given in the previous section. In order to preserve the orthogonality through all the analysis channels in diHiggs group, the standard overlap removal precedure given in Table 8.2 is adapted. The improved jet reconstruction algorithm called *Particle Flow (PFlow)* [197] is used in this part of the analysis. As for the flavour tagging algorithm, DL1r is used with 70% working point. Lepton working point definition is adapted from $t\bar{t}H$ analysis given in Section 7.2.3, except the additional selections on the external and internal conversions.

Keep	**Remove**	**Cone size (ΔR)**
electron	electron (low p_T)	0.1
muon	electron	Shared ID track
electron	jet	0.2
jet	muon	0.4
electron	tau	0.2
muon	tau	0.2
tau	jet	0.2
muon	jet	NumTrack< 3 and (ghost-associated or ΔR < 0.2)

Table 8.2: Summary of the overlap removal procedure between electrons, muons, hadronically decaying taus, and jets.

The fitted kinematic distributions are split in flavour of leptons to control the different source of backgrounds for each flavour. The choice of distributions to be used in the fit is done based on the comparison of shapes between signal and background, and it is found that the distributions of $\Delta\phi_{\ell\ell}$ and $\Delta R_{\ell_0,jet}$ have the most distinctive shape in signal with respect to other backgrounds. The fitted distributions are shown in Figure 8.7, and the pie charts illustrate the fraction of background composition is given in Figure 8.8.

Table 8.3: Expected 95% CL upper-limits on the SM, non-resonant diHiggs production cross section ($\sigma(pp \to hh)$). The $\pm1\sigma$ and $\pm2\sigma$ excursions from the median expected value shows the statistical effects and the uncertainty on the $t\bar{t}$ process and luminosity.

	-2σ	-1σ	**Median Expected**	$+1\sigma$	$+2\sigma$
95% CL $\sigma(pp \to hh)\,{}^{\sigma}/_{\sigma_{SM}}$	34.02	45.67	63.38	105.26	137.47

At the time of writing the thesis, the samples including the systematic uncertainties were not ready, therefore the results include only the statistical uncertainties, uncertainty on the luminosity (2%) and uncertainty on $t\bar{t}$ sample (20%). Combination with other multilepton channels, as well as with other diHiggs channels were also not finalised.

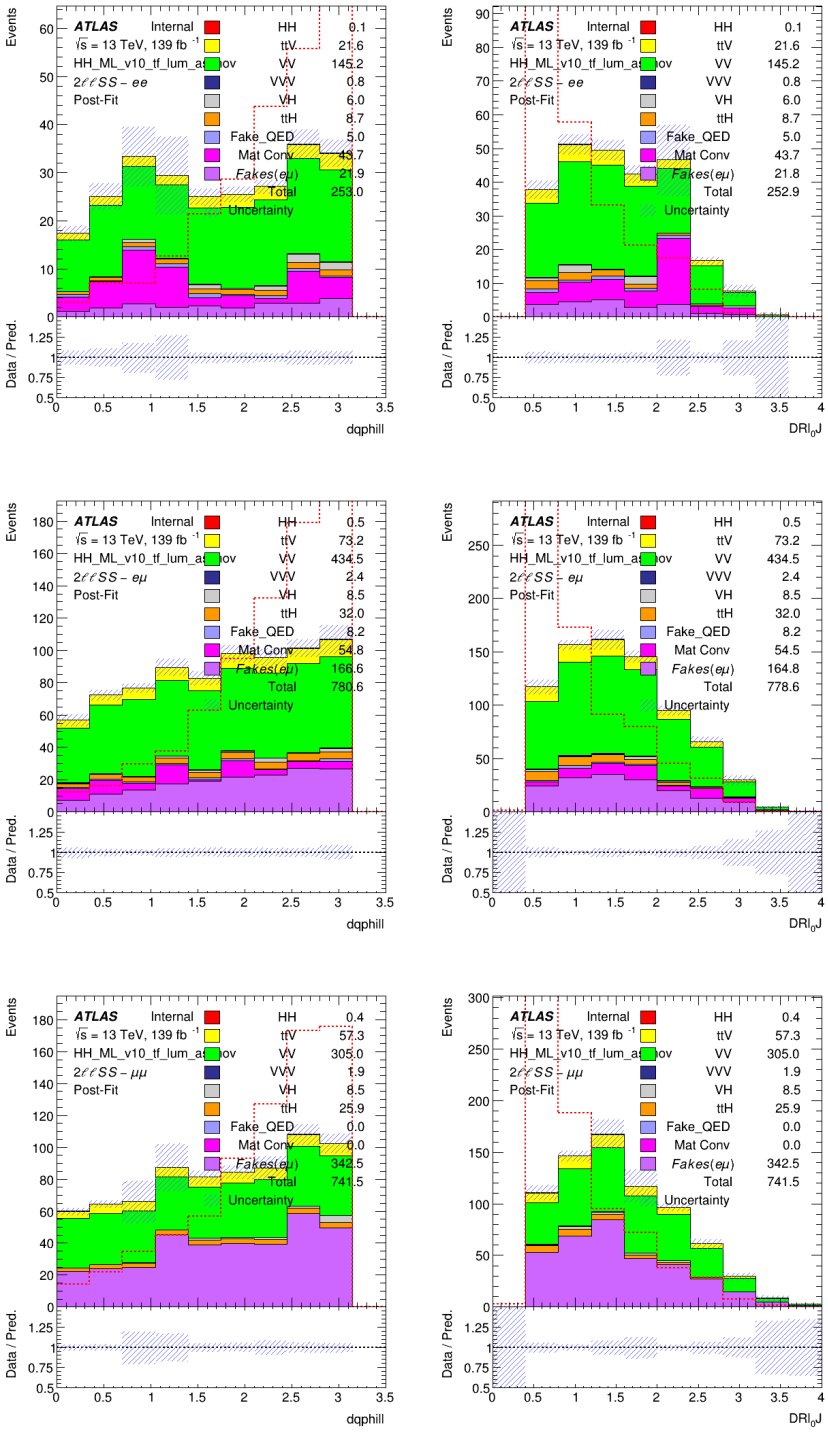


Figure 8.7: The distributions of $\Delta\phi$ between two leptons (***right***) and minimum angular distance between the first lepton and the jet (***left***) for ee (***upper***), $e\mu$ (***middle***) and $\mu\mu$ (***lower***) channels. Signal without normalisation is shown with red line.

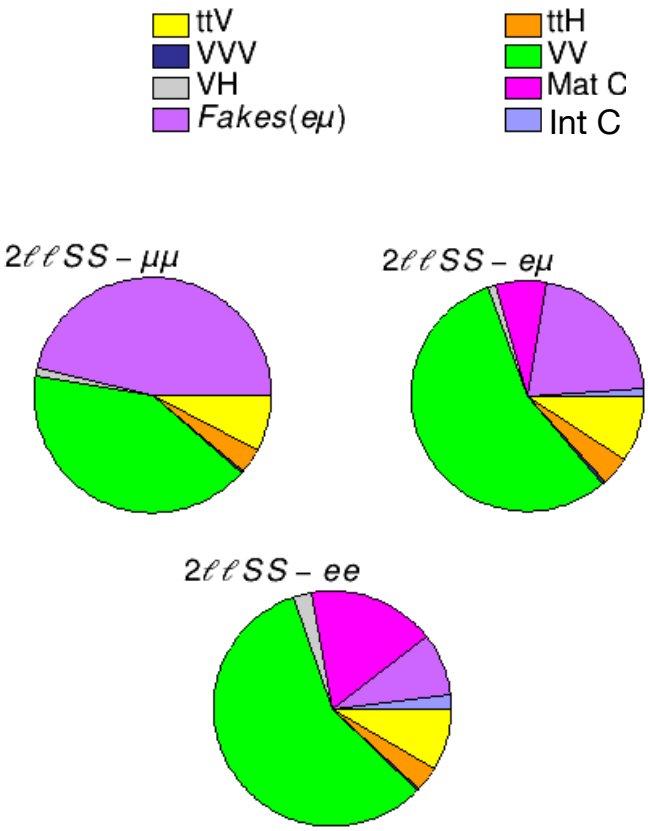


Figure 8.8: Fraction of background compositions in 2ℓSS channel.

CHAPTER 9

Conclusions

After the Higgs boson discovery in July 2012 by two big collaborations of LHC, CMS and ATLAS, its property measurements became the main goal of LHC physics programs. Besides the measurements of its fundamental properties such as its spin and parity, the study of the nature of the coupling to other particles has a special role since any deviation from SM prediction can give hints to the nature of new physics. During Run2, the observation of the top quark, bottom quark and tau lepton Yukawa interactions are achived, which provided a significant validation of the SM. Whereas the first and second generation fermion coupling strenghts continue to be a big challenge for the forthcoming of the particle physics programs at LHC. Beyond the Higgs boson coupling to any particle, the observation of Higgs self-coupling will permit an important test of the SM and more specifically of the Higgs potential. Due to its very low cross section, the measurement of the diHiggs production suffers from the low statictics with the current operation properties of LHC, and it will be one of the main measurement with in the future of LHC physics.

In this disertation, the production of a Higgs boson in association with a top quark pair in multilepton channels at $\sqrt{s}$ = 13 TeV centre of mass using 79.8 fb^{-1} data collected by ATLAS is presented. Likelihood fit to signal strength and to seven normalisation factors for several background processes is performed in 25 event categories. The best fit observed result of the cross section is $\hat{\sigma}(t\bar{t}H) = 294^{+182}_{-162}$, in agreement with the SM $t\bar{t}H$ prediction of 507^{+35}_{-30}. This result is directly sensitive to the top quark Yukawa coupling since the $t\bar{t}H$ production rate is proportional to y_t^2 at NLO, and therefore the top Yukawa coupling is measured with a precision of 10%. The measured normalisation factors for $t\bar{t}W$ in the phase space selected by the analysis are between 30%-70% above the higher order predicted values by the SM. In addition, some modelling issues are observed in the $t\bar{t}W$ dominated phase space. Therefore, a better understanding of the $t\bar{t}W$ process is imparative.

The operation of the High Luminosity Large Hadron Collider (HL-LHC) [198] planned to start in 2026 and expected to deliver total integrated luminosity of 3000 fb^{-1} at a $\sqrt{s}$ = 14 TeV center of mass energy. With the current obtained normalisation factors, the analysis result can be naively extrapolated to the HL-LHC conditions, and overall around 48% reduction on the

total uncertainty is expected on the $t\bar{t}H$ cross section measurement.

This thesis also presented search for Higgs boson pair production in the 2ℓSS final state at $\sqrt{s} = 13$ TeV centre of mass using 139 fb^{-1} data. The expected limit on SM Higgs-self coupling at 95% CL is $90 \times \sigma^{SM}_{ggF}(pp \to HH)$. The combination with other diHiggs production channels will be performed using full Run2 statistics. Furthermore, higher order corrections (one-loop corrections to the $t\bar{t}H$ vertex and from one-loop box and pentagon diagrams) to $t\bar{t}H$ measurements are sensitive to the Higgs trilinear coupling and would be exploited to extract a competitive limit with Run2 data. In HL-LHC, the projections show that the combined expected significance of the ATLAS analysis for diHiggs production will be $\approx 4\ \sigma$, and such high luminosities will permit milestone observation of diHiggs production.

In the prospect of next-generation collider projects, electron-positron (e^+e^-) Higgs factories will provide high level of precision measurements for the Higgs physics. Currently proposed four e^+e^- Higgs factories, CEPC, ILC, CLIC, and FCC-ee, can access the Higgs-self coupling through the single Higgs measurements, at which the precision on the single Higgs measurements will reach or exceed the 1% level. On the other hand, diHiggs production can also be used to measure the Higgs self-coupling ($e^+e^- \to ZHH$ and $e^+e^- \to \nu\hat{\nu}HH$), which requires much higher energies than the nominal threshold energy (250 GeV and 341 GeV) to have significant event sample. Therefore, ILC and CLIC proposals provide promising extrapolation results for diHiggs production measurements among the future experiments. Furthermore, for the future hadron colliders, such as FCC-hh, the direct measurement of diHiggs production will be possible within the level of 5% precision.

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*, Includes exercises, Westview, 1995, URL: https://cds.cern.ch/record/257493 (cit. on p. 3).

[2] F. Halzen and A. D. Martin, *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*, 1984, ISBN: 978-0-471-88741-6 (cit. on p. 3).

[3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on p. 3).

[4] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001, URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001 (cit. on pp. 4, 5).

[5] E. Noether, *Invariante Variationsprobleme*, ger, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235, URL: http://eudml.org/doc/59024 (cit. on p. 4).

[6] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579 (cit. on p. 4).

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on p. 4).

[8] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367 (cit. on p. 4).

[9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 4).

[10] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132, ISSN: 0031-9163, URL: http://www.sciencedirect.com/science/article/pii/0031916364911369 (cit. on p. 4).

[11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 4).

[12] C. N. Yang and R. L. Mills,
Conservation of Isotopic Spin and Isotopic Gauge Invariance,
Phys. Rev. **96** (1 1954) 191,
URL: https://link.aps.org/doi/10.1103/PhysRev.96.191 (cit. on p. 11).

[13] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,
Phys. Rev. **127** (3 1962) 965,
URL: https://link.aps.org/doi/10.1103/PhysRev.127.965 (cit. on p. 17).

[14] M. Kobayashi and T. Maskawa,
CP-Violation in the Renormalizable Theory of Weak Interaction,
Progress of Theoretical Physics **49** (1973) 652, ISSN: 0033-068X, eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf,
URL: https://doi.org/10.1143/PTP.49.652 (cit. on p. 19).

[15] R. Barate et al., *Search for the standard model Higgs boson at LEP*,
Phys. Lett. **B565** (2003) 61, arXiv: hep-ex/0306033 [hep-ex] (cit. on p. 20).

[16] T. Aaltonen et al., *Combination of Tevatron Searches for the Standard Model Higgs Boson in the W^+W^- Decay Mode*, Phys. Rev. Lett. **104** (6 2010) 061802,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.061802
(cit. on p. 20).

[17] L. E. W. Group,
Precision Electroweak Measurements and Constraints on the Standard Model, (2010),
arXiv: 1012.2367 [hep-ex] (cit. on p. 20).

[18] S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv: 1207.7235 [hep-ex]
(cit. on p. 20).

[19] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1,
arXiv: 1207.7214 [hep-ex] (cit. on p. 20).

[20] G. Aad et al., *Combined Measurement of the Higgs Boson Mass inppCollisions ats=7and 8 TeV with the ATLAS and CMS Experiments*,
Physical Review Letters **114** (2015), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.114.191803 (cit. on p. 20).

[21] G. Aad et al., *Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data*,
Phys. Rev. **D90** (2014) 052004, arXiv: 1406.3827 [hep-ex] (cit. on p. 20).

[22] D. de Florian et al.,
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,
(2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 22, 100, 101, 127, 139).

[23] T. Binoth et al., *Multi-Higgs boson production in the standard model and beyond*, Phys. Rev. D **74** (11 2006) 113008, URL: https://link.aps.org/doi/10.1103/PhysRevD.74.113008 (cit. on p. 22).

[24] B. Fuks, J. H. Kim and S. J. Lee, *Probing Higgs boson self-interactions in proton-proton collisions at a center-of-mass energy of 100 TeV*, Phys. Rev. D **93** (3 2016) 035026, URL: https://link.aps.org/doi/10.1103/PhysRevD.93.035026 (cit. on p. 22).

[25] D. A. Dicus, C. Kao and W. W. Repko, *Interference effects and the use of Higgs boson pair production to study the Higgs trilinear self-coupling*, Phys. Rev. D **92** (9 2015) 093003, URL: https://link.aps.org/doi/10.1103/PhysRevD.92.093003 (cit. on p. 22).

[26] J. Alison et al., "Higgs Boson Pair Production at Colliders: Status and Perspectives", *Double Higgs Production at Colliders Batavia, IL, USA, September 4, 2018-9, 2019*, ed. by B. Di Micco et al., 2019, arXiv: 1910.00012 [hep-ph], URL: https://lss.fnal.gov/archive/2019/conf/fermilab-conf-19-468-e-t.pdf (cit. on pp. 23, 25).

[27] S. Borowka et al., *Full top quark mass dependence in Higgs boson pair production at NLO*, JHEP **10** (2016) 107, arXiv: 1608.04798 [hep-ph] (cit. on p. 23).

[28] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme*, Eur. Phys. J. **C79** (2019) 459, arXiv: 1811.05692 [hep-ph] (cit. on p. 23).

[29] D. de Florian and J. Mazzitelli, *Two-loop virtual corrections to Higgs pair production*, Phys. Lett. **B724** (2013) 306, arXiv: 1305.5206 [hep-ph] (cit. on p. 23).

[30] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **111** (2013) 201801, arXiv: 1309.6594 [hep-ph] (cit. on p. 23).

[31] J. Grigo, K. Melnikov and M. Steinhauser, *Virtual corrections to Higgs boson pair production in the large top quark mass limit*, Nucl. Phys. **B888** (2014) 17, arXiv: 1408.2422 [hep-ph] (cit. on p. 23).

[32] D. de Florian et al., *Differential Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD*, JHEP **09** (2016) 151, arXiv: 1606.09519 [hep-ph] (cit. on p. 23).

[33] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05** (2018) 059, arXiv: 1803.02463 [hep-ph] (cit. on p. 23).

[34] G. Heinrich et al., *Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects*, JHEP **06** (2019) 066, arXiv: 1903.08137 [hep-ph] (cit. on p. 23).

[35] A. M. Sirunyan et al., *Combination of Searches for Higgs Boson Pair Production in Proton-Proton Collisions at* $\sqrt{s}$ = 13 TeV, Phys. Rev. Lett. **122** (12 2019) 121803, URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.121803 (cit. on p. 23).

[36] G. Aad et al., *Combination of searches for Higgs boson pairs in pp collisions at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, Phys. Lett. **B800** (2020) 135103, arXiv: 1906.02025 [hep-ex] (cit. on p. 23).

[37] *Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements*, tech. rep. ATL-PHYS-PUB-2019-009, CERN, 2019, URL: http://cds.cern.ch/record/2667570 (cit. on p. 23).

[38] *Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment*, tech. rep. ATLAS-CONF-2019-049, CERN, 2019, URL: https://cds.cern.ch/record/2693958 (cit. on p. 23).

[39] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, Journal of Experimental and Theoretical Physics **120** (2015) 335, ISSN: 1090-6509 (cit. on p. 26).

[40] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089. 49 p, Comments: 43 pages, 7 figures. v3: updated with latest Higgs and top mass, corrected NNLO result for g2 and gY, URL: https://cds.cern.ch/record/1561905 (cit. on p. 27).

[41] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (8 1998) 1562, URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562 (cit. on p. 28).

[42] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, Phys. Lett. B **749** (2015) 519, arXiv: 1506.05988 [hep-ex] (cit. on pp. 29, 30).

[43] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into* $b\bar{b}$ *in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 349, arXiv: 1503.05066 [hep-ex] (cit. on p. 29).

[44] ATLAS Collaboration, *Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. B **740** (2015) 222, arXiv: 1409.3122 [hep-ex] (cit. on p. 29).

[45] CMS Collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, JHEP **09** (2014) 087, arXiv: 1408.1682 [hep-ex] (cit. on p. 29).

[46] CMS Collaboration, *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*, Eur. Phys. J. C **75** (2015) 251, arXiv: 1502.02485 [hep-ex] (cit. on p. 29).

[47] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (2018) 072003, arXiv: 1712.08891 [hep-ex] (cit. on pp. 29, 30).

[48] M. Aaboud et al., *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173, arXiv: 1806.00425 [hep-ex] (cit. on p. 30).

[49] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001 (cit. on p. 31).

[50] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 31).

[51] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 31).

[52] S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 31).

[53] J. Alves A.Augusto et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 31).

[54] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, (2016), General Photo, URL: https://cds.cern.ch/record/2197559 (cit. on p. 32).

[55] M. Bruschi, *The new ATLAS/LUCID detector*, tech. rep. ATL-FWD-PROC-2015-002, CERN, 2015, URL: https://cds.cern.ch/record/2025000 (cit. on p. 33).

[56] ATLAS collaboration, *Good Run Lists For Analysis Run 2*, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/GoodRunListsForAnalysisRun2 (cit. on pp. 33, 99).

[57] Z. M. and, *Simulation of Pile-up in the ATLAS Experiment*, Journal of Physics: Conference Series **513** (2014) 022024, URL: https://doi.org/10.1088%2F1742-6596%2F513%2F2%2F022024 (cit. on p. 33).

[58] ATLAS collaboration, *TWiki*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, Accessed: 24.07.2018 (cit. on p. 33).

[59] *ATLAS magnet system: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/338080 (cit. on p. 33).

[60] *ATLAS central solenoid: Technical Design Report*, Technical Design Report ATLAS, Electronic version not available, CERN, 1997, URL: https://cds.cern.ch/record/331067 (cit. on p. 33).

[61] J. P. Badiou et al., *ATLAS barrel toroid: Technical Design Report*, Technical Design Report ATLAS, Electronic version not available, CERN, 1997, URL: https://cds.cern.ch/record/331065 (cit. on p. 33).

[62] *ATLAS end-cap toroids: Technical Design Report*, Technical Design Report ATLAS, Electronic version not available, CERN, 1997, URL: https://cds.cern.ch/record/331066 (cit. on p. 33).

[63] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 34).

[64] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at $\sqrt{s}$ = 7 TeV. Review of single vector boson production in pp collisions at $\sqrt{s}$ = 7 TeV*, Eur. Phys. J. C **74** (2014) 60 p, Comments: 60 pages, 64 figures, For Eur. Phys. J. C, URL: https://cds.cern.ch/record/1699952 (cit. on p. 35).

[65] *ATLAS inner detector: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/331063 (cit. on p. 35).

[66] S. Haywood et al., *ATLAS inner detector: Technical Design Report, 2*, Technical Design Report ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/331064 (cit. on p. 35).

[67] *ATLAS calorimeter performance: Technical Design Report*, Technical Design Report ATLAS, CERN, 1996, URL: https://cds.cern.ch/record/331059 (cit. on p. 36).

[68] *ATLAS liquid-argon calorimeter: Technical Design Report*, Technical Design Report ATLAS, CERN, 1996, URL: https://cds.cern.ch/record/331061 (cit. on p. 36).

[69] *ATLAS muon spectrometer: Technical Design Report*, Technical Design Report ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/331068 (cit. on p. 37).

[70] *ATLAS level-1 trigger: Technical Design Report*, Technical Design Report ATLAS, CERN, 1998, URL: https://cds.cern.ch/record/381429 (cit. on p. 37).

[71] P. Jenni et al.,
ATLAS high-level trigger, data-acquisition and controls: Technical Design Report,
Technical Design Report ATLAS, CERN, 2003,
URL: https://cds.cern.ch/record/616089 (cit. on p. 37).

[72] M. Aaboud et al., *Performance of the ATLAS Trigger System in 2015*,
Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 37).

[73] *Trigger Menu in 2016*, tech. rep. ATL-DAQ-PUB-2017-001, CERN, 2017,
URL: https://cds.cern.ch/record/2242069 (cit. on p. 37).

[74] *Trigger Menu in 2017*, tech. rep. ATL-DAQ-PUB-2018-002, CERN, 2018,
URL: https://cds.cern.ch/record/2625986 (cit. on p. 37).

[75] *Trigger menu in 2018*, tech. rep. ATL-DAQ-PUB-2019-001, CERN, 2019,
URL: https://cds.cern.ch/record/2693402 (cit. on p. 37).

[76] P. Klimek,
"Signal reconstruction performance with the ATLAS Hadronic Tile Calorimeter", 2012,
URL: http://cdsweb.cern.ch/record/1473499/files/ATL-TILECAL-PROC-2012-007.pdf (cit. on p. 40).

[77] W. E. Cleland and E. G. Stern, *Signal processing considerations for liquid ionization calorimeters in a high rate environment*, Nucl. Instrum. Meth. **A338** (1994) 467
(cit. on p. 40).

[78] M. Aaboud et al., *Operation and performance of the ATLAS Tile Calorimeter in Run 1*,
Eur. Phys. J. **C78** (2018) 987, arXiv: 1806.02129 [hep-ex] (cit. on p. 40).

[79] G. Aad et al., *Readiness of the ATLAS Tile Calorimeter for LHC collisions*,
Eur. Phys. J. **C70** (2010) 1193, arXiv: 1007.5423 [physics.ins-det]
(cit. on p. 41).

[80] E. Starchenko et al., *Cesium monitoring system for ATLAS Tile Hadron Calorimeter*,
Nucl. Instrum. Meth. **A494** (2002) 381 (cit. on p. 42).

[81] J. Abdallah et al.,
The Laser calibration of the Atlas Tile Calorimeter during the LHC run 1,
JINST **11** (2016) T10005, arXiv: 1608.02791 [physics.ins-det] (cit. on p. 42).

[82] K. Anderson et al.,
Design of the front-end analog electronics for the ATLAS tile calorimeter,
Nucl. Instrum. Meth. **A551** (2005) 469 (cit. on p. 43).

[83] 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] (cit. on p. 43).

[84] A. Blanco et al,
Upgrade of the Laser calibration system of the ATLAS Tile Calorimeter,
ATL-TILECAL-INT-2016-004 (2016) (cit. on p. 43).

[85] R. Chadelas et al.,
High voltage distributor system for the Tile hadron calorimeter of the ATLAS detector,
tech. rep. ATL-TILECAL-2000-003. PCCF-RI-2000-04,
Clermont-Ferrand 2. Lab. Phys. Corpusc. Cosmol., 2000,
URL: https://cds.cern.ch/record/436230 (cit. on p. 51).

[86] M. N. Agaras, D. Calvet and M. Marjanovic, *New analysis of the performance of the TileCal High Voltage distribution system during run 2*,
tech. rep. ATL-TILECAL-INT-2020-001, CERN, 2020,
URL: https://cds.cern.ch/record/2706096 (cit. on p. 51).

[87] D. Calvet and M. Marjanovic,
Stability of the TileCal High Voltage distribution system in 2015 and 2016,
tech. rep. ATL-TILECAL-INT-2017-004, CERN, 2017,
URL: https://cds.cern.ch/record/2293471 (cit. on p. 51).

[88] S. Calvet et al.,
Stability of the High Voltage system in the Tile Calorimeter of the ATLAS detector.,
tech. rep. ATL-TILECAL-INT-2012-010, CERN, 2012,
URL: https://cds.cern.ch/record/1500523 (cit. on p. 51).

[89] *DCS Data Viewer*, https://atlas-ddv.cern.ch/DDV.html (cit. on p. 53).

[90] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007,
arXiv: 0811.4622 [hep-ph] (cit. on pp. 60, 64).

[91] R. Placakyte, "Parton Distribution Functions",
Proceedings, 31st International Conference on Physics in collisions (PIC 2011): Vancouver, Canada, August 28-September 1, 2011, 2011,
arXiv: 1111.5452 [hep-ph] (cit. on p. 61).

[92] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244,
arXiv: 1207.1303 [hep-ph] (cit. on pp. 61, 102).

[93] L. A. Harland-Lang et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*,
Eur. Phys. J. **C75** (2015) 204, arXiv: 1412.3989 [hep-ph] (cit. on pp. 62, 102).

[94] S. Catani et al., *QCD matrix elements + parton showers*, JHEP **11** (2001) 063,
arXiv: hep-ph/0109231 [hep-ph] (cit. on p. 62).

[95] L. Lonnblad,
Correcting the color dipole cascade model with fixed order matrix elements,
JHEP **05** (2002) 046, arXiv: hep-ph/0112284 [hep-ph] (cit. on p. 62).

[96] N. Lavesson and L. Lonnblad, *W+jets matrix elements and the dipole cascade*,
JHEP **07** (2005) 054, arXiv: hep-ph/0503293 [hep-ph] (cit. on p. 62).

[97] S. Mrenna and P. Richardson,
Matching Matrix Elements and Parton Showers with HERWIG and PYTHIA,
Journal of High Energy Physics **2004** (2004) 040,
URL: https://doi.org/10.1088%2F1126-6708%2F2004%2F05%2F040
(cit. on p. 62).

[98] B. Andersson et al., *Parton Fragmentation and String Dynamics*,
Phys. Rept. **97** (1983) 31 (cit. on p. 63).

[99] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*,
Nucl. Phys. **B238** (1984) 492 (cit. on p. 63).

[100] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. Phys. Res. A **462** (2001) 152 (cit. on p. 63).

[101] 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] (cit. on pp. 63, 102, 156).

[102] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043,
arXiv: 1002.2581 [hep-ph] (cit. on p. 64).

[103] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070,
arXiv: 0709.2092 [hep-ph] (cit. on p. 64).

[104] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 64).

[105] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*,
Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 64).

[106] G. Corcella et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010,
arXiv: hep-ph/0011363 [hep-ph] (cit. on pp. 64, 156).

[107] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76** (2016) 196,
arXiv: 1512.01178 [hep-ph] (cit. on pp. 64, 156).

[108] E. Bothmann et al., *Event Generation with SHERPA 2.2*, (2019),
arXiv: 1905.09127 [hep-ph] (cit. on p. 64).

[109] S. Schumann and F. Krauss,
A Parton shower algorithm based on Catani-Seymour dipole factorisation,
JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on pp. 64, 100).

[110] S. Agostinelli et al., *GEANT4: A Simulation toolkit*,
Nucl. Instrum. Meth. **A506** (2003) 250 (cit. on p. 64).

[111] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 64).

[112] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector", 2013, URL: https://cds.cern.ch/record/1505342 (cit. on p. 67).

[113] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/0168900287908874 (cit. on pp. 67, 84).

[114] R. Frühwirth et al., *Recent results on adaptive track and multitrack fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **502** (2003) 702, Proceedings of the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900203005497 (cit. on p. 67).

[115] L. Bugge and J. Myrheim, *Tracking and track fitting*, Nuclear Instruments and Methods **179** (1981) 365, ISSN: 0029-554X, URL: http://www.sciencedirect.com/science/article/pii/0029554X8190063X (cit. on p. 67).

[116] M. Aaboud et al., *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 68).

[117] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. **C77** (2017) 332, arXiv: 1611.10235 [physics.ins-det] (cit. on p. 68).

[118] G. Piacquadio, K. Prokofiev and A. Wildauer, *Primary vertex reconstruction in the ATLAS experiment at LHC*, Journal of Physics: Conference Series **119** (2008) 032033, URL: https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032033 (cit. on p. 68).

[119] *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s}$ = 13 TeV*, tech. rep. ATL-PHYS-PUB-2015-018, CERN, 2015, URL: http://cds.cern.ch/record/2037683 (cit. on p. 68).

[120] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 639. 40 p, URL: https://cds.cern.ch/record/2657964 (cit. on pp. 69, 71, 72).

[121] The ATLAS collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, (2016) (cit. on p. 69).

[122] M. Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV*, Submitted to: Eur. Phys. J. (2019), arXiv: 1902.04655 [physics.ins-det] (cit. on pp. 69, 70, 106).

[123] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, (2008) (cit. on pp. 70, 76).

[124] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, (2012), URL: https://cds.cern.ch/record/1449796 (cit. on p. 70).

[125] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, (2019), arXiv: 1908.00005 [hep-ex] (cit. on pp. 71, 72).

[126] G. Aad et al., *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] (cit. on pp. 73, 75).

[127] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$ = 13 TeV*, tech. rep. ATLAS-CONF-2017-029, CERN, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on pp. 75, 76).

[128] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, tech. rep. ATL-PHYS-PUB-2019-033, CERN, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on p. 76).

[129] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. **C77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 76).

[130] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 77).

[131] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Physical Review D **48** (1993) 3160, ISSN: 0556-2821, URL: http://dx.doi.org/10.1103/PhysRevD.48.3160 (cit. on p. 77).

[132] Y. Dokshitzer et al., *Better jet clustering algorithms*, Journal of High Energy Physics **1997** (1997) 001, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/1997/08/001 (cit. on p. 77).

[133] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998 270, arXiv: hep-ph/9907280 [hep-ph] (cit. on p. 77).

[134] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: 0802.1188 [hep-ph] (cit. on pp. 77, 78).

[135] M. Aaboud et al., *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on pp. 78, 81).

[136] T. A. collaboration, *Tagging and suppression of pileup jets*, (2014) (cit. on pp. 82, 126).

[137] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*, tech. rep. ATLAS-CONF-2015-029, CERN, 2015, URL: https://cds.cern.ch/record/2037702 (cit. on p. 83).

[138] *Performance ofb-jet identification in the ATLAS experiment*, Journal of Instrumentation **11** (2016) P04008, URL: https://doi.org/10.1088%2F1748-0221%2F11%2F04%2Fp04008 (cit. on p. 83).

[139] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016, URL: http://cds.cern.ch/record/2160731 (cit. on p. 84).

[140] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, tech. rep. ATL-PHYS-PUB-2017-013, CERN, 2017, URL: http://cds.cern.ch/record/2273281 (cit. on p. 84).

[141] M. Aaboud et al., *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. **C78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on p. 85).

[142] K. Cranmer, *Practical Statistics for the LHC*, (2015) 267, Comments: presented at the 2011 European School of High-Energy Physics, Cheile Gradistei, Romania, 7-20 September 2011 I expect to release updated versions of this document in the future, URL: https://cds.cern.ch/record/2004587 (cit. on pp. 86, 91).

[143] G. Cowan, *Use of the profile likelihood function in searches for new physics*, (2011) 109, URL: https://cds.cern.ch/record/2203244 (cit. on p. 86).

[144] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, [Erratum: Eur. Phys. J.C73,2501(2013)], arXiv: 1007.1727 [physics.data-an] (cit. on p. 86).

[145] O. Behnke et al.,
Data analysis in high energy physics: a practical guide to statistical methods,
Wiley-VCH, 2013, URL: https://cds.cern.ch/record/1517556 (cit. on p. 86).

[146] K. Cranmer et al.,
HistFactory: A tool for creating statistical models for use with RooFit and RooStats,
tech. rep. CERN-OPEN-2012-016, New York U., 2012,
URL: https://cds.cern.ch/record/1456844 (cit. on p. 87).

[147] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) 343 (cit. on p. 88).

[148] E. S. P. J. Neyman,
IX. On the problem of the most efficient tests of statistical hypotheses,
Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **231** (1933) 289, ISSN: 0264-3952, URL:
http://rsta.royalsocietypublishing.org/content/231/694-706/289
(cit. on p. 89).

[149] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Phil. Trans. Roy. Soc. Lond. **A236** (1937) 333 (cit. on p. 90).

[150] A. L. Read, *Presentation of search results: The CL(s) technique*,
J. Phys. **G28** (2002) 2693, [,11(2002)] (cit. on p. 91).

[151] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large*, American Mathematical Society, 1943,
URL: https://books.google.fr/books?id=igL7xgYxsJEC (cit. on p. 95).

[152] *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*,
tech. rep. ATLAS-CONF-2019-045, CERN, 2019,
URL: http://cds.cern.ch/record/2693930 (cit. on pp. 98, 100).

[153] ATLAS collaboration, *ATLAS Athena Guide*,
https://atlassoftwaredocs.web.cern.ch/athena/ (cit. on p. 99).

[154] I. Bird et al., *Update of the Computing Models of the WLCG and the LHC Experiments*, tech. rep., 2014 (cit. on p. 99).

[155] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021, 2014,
URL: https://cds.cern.ch/record/1966419 (cit. on pp. 100, 127).

[156] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*,
Comput. Phys. Commun. **108** (1998) 56, arXiv: hep-ph/9704448 [hep-ph]
(cit. on p. 100).

[157] W. Beenakker et al., *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805, arXiv: hep-ph/0107081 [hep-ph] (cit. on p. 100).

[158] W. Beenakker et al., *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. B **653** (2003) 151, arXiv: hep-ph/0211352 [hep-ph] (cit. on p. 100).

[159] S. Dawson et al., *Associated top quark Higgs boson production at the LHC*, Phys. Rev. D **67** (2003) 071503, arXiv: hep-ph/0211438 [hep-ph] (cit. on p. 100).

[160] S. Dawson et al., *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. D **68** (2003) 034022, arXiv: hep-ph/0305087 [hep-ph] (cit. on p. 100).

[161] Y. Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. B **738** (2014) 1, arXiv: 1407.1110 [hep-ph] (cit. on p. 100).

[162] S. Frixione et al., *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, JHEP **09** (2014) 065, arXiv: 1407.0823 [hep-ph] (cit. on p. 100).

[163] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 100).

[164] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 100).

[165] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 100).

[166] S. Hoeche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 100).

[167] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{\pm}$ production and decay at NLO*, JHEP **07** (2012) 052, arXiv: 1204.5678 [hep-ph] (cit. on p. 100).

[168] S. Frixione et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184, arXiv: 1504.03446 [hep-ph] (cit. on p. 100).

[169] 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] (cit. on p. 100).

[170] R. Frederix, D. Pagani and M. Zaro, *Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions*, JHEP **02** (2018) 031, arXiv: 1711.02116 [hep-ph] (cit. on pp. 100, 101).

[171] ATLAS Collaboration,
Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses,
(2015), ATL-PHYS-PUB-2016-005,
URL: https://cds.cern.ch/record/2120826 (cit. on p. 101).

[172] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*,
(2015), ATL-PHYS-PUB-2016-002,
URL: https://cds.cern.ch/record/2119986 (cit. on p. 101).

[173] ATLAS Collaboration,
Simulation of top quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV,
(2015), ATL-PHYS-PUB-2016-004,
URL: https://cds.cern.ch/record/2120417 (cit. on pp. 101, 128).

[174] 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] (cit. on p. 102).

[175] H. B. Hartanto et al.,
Higgs boson production in association with top quarks in the POWHEG BOX,
Phys. Rev. D **91** (2015) 094003, arXiv: 1501.04498 [hep-ph] (cit. on p. 102).

[176] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*,
Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph]
(cit. on p. 102).

[177] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040,
arXiv: 1410.8849 [hep-ph] (cit. on p. 102).

[178] H.-L. Lai et al., *New parton distributions for collider physics*,
Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph] (cit. on pp. 102, 156).

[179] J. Pumplin et al.,
New Generation of Parton Distributions with Uncertainties from Global QCD Analysis,
JHEP **07** (2002) 012, arXiv: hep-ph/0201195 [hep-ph] (cit. on p. 102).

[180] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*,
Phys. Rev. D **78** (2008) 013004, arXiv: 0802.0007 [hep-ph] (cit. on p. 102).

[181] T. Sjöstrand et al., *High-energy-physics event generation with Pythia 6.1*,
Comput. Phys. Commun. **135** (2001) 238, arXiv: hep-ph/0010017 [hep-ph]
(cit. on p. 102).

[182] 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]
(cit. on p. 102).

[183] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: $s-$ and $t-$channel contributions*, JHEP **09** (2009) 111, arXiv: 0907.4076 [hep-ph]
(cit. on p. 102).

[184] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: 1207.5391 [hep-ph] (cit. on p. 102).

[185] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv: hep-ph/0506026 [hep-ph] (cit. on p. 102).

[186] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: https://cds.cern.ch/record/2037717 (cit. on p. 104).

[187] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (7 2018) 072003, URL: https://link.aps.org/doi/10.1103/PhysRevD.97.072003 (cit. on p. 106).

[188] M. Aaboud et al., *Luminosity determination in pp collisions at* $\sqrt{s}$ = *8 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 126).

[189] *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s}$ = 13 *TeV*, (2015), URL: https://cds.cern.ch/record/2037613 (cit. on p. 126).

[190] G. Aad et al., *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s}$ = 7 *TeV*, Eur. Phys. J. **C73** (2013) 2304, arXiv: 1112.6426 [hep-ex] (cit. on p. 126).

[191] G. Aad et al., *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. **C79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 126).

[192] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on p. 127).

[193] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 127).

[194] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219 (cit. on p. 139).

[195] W. Verkerke and D. Kirkby, *RooFit Users Manual*, URL: http://roofit.sourceforge.net (cit. on p. 139).

[196] S. Brochet et al., *MoMEMta, a modular toolkit for the Matrix Element Method at the LHC*, The European Physical Journal C **79** (2019) 126, URL: https://doi.org/10.1140/epjc/s10052-019-6635-5 (cit. on p. 154).

[197] M. Aaboud et al.,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 163).

[198] A. G. et al.,
High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1,
CERN Yellow Reports: Monographs, CERN, 2017,
URL: https://cds.cern.ch/record/2284929 (cit. on p. 167).

Appendix

APPENDIX A

2ℓSS BDT

2ℓSS channel exploits two BDTs in order to discriminate the different processes. One training is done between $t\bar{t}H$ and $t\bar{t}$, where the non-prompt backgrounds originating from. Second traning is to discriminate $t\bar{t}H$ against $t\bar{t}V$ ($V = W, Z$). Both $t\bar{t}H$-vs-$t\bar{t}V$ and $t\bar{t}H$-vs-$t\bar{t}$ trainings use MC simulations scaled by the normalisation factors. Nine input variables are used:

- Number of jets with $p_T > 25$ GeV, N_{jets};
- Number of b-jets with $p_T > 25$ GeV, $N_{b\text{-jets}}$ – MV2c10 algorithm is used with 70% b-tag efficiency;
- leptonic flavour, ee, $e\mu$ μe and $\mu\mu$ where the first lepton corresponds to the one with the highest p_T, $lFlav = 2\times\text{N}_{\mu_0}+\text{N}_{\mu_1}$;
- Distance between leading lepton and its closest jet, $\Delta R(\ell_0, \text{jet})$;
- Distance between sub-leading lepton and its closest jet, $\Delta R(\ell_1, \text{jet})$;
- Maximum between lepton $|\eta_{\ell_0}|$ and $|\eta_{\ell_1}|$, Max($|\eta_\ell|$);
- Sub-leading lepton p_T, $p_T(\ell_1)$;
- $E_{\text{T}}^{\text{miss}}$
- Distance between the two same-sign leptons, $\Delta R(\ell_0, \ell_1)$;

The catogarisation of some CRs and SRs in 2ℓSS is done by combination of the two BDTs' output. Figure A.1 shows the $t\bar{t}$, $t\bar{t}W$, and $t\bar{t}H$ events distributed in the 2-dimensional BDTG plane, separately for ee+eμ and eμ+$\mu\mu$. Pink coloured lines define the cuts on the corresponding regions. A CR enriched in $t\bar{t}$ (split in the two lepton flavour channels), a CR enriched in $t\bar{t}W$, and SRs, where the remaining BDTG shape of $t\bar{t}H$ vs $t\bar{t}V$ is used in the fit (3 bins).

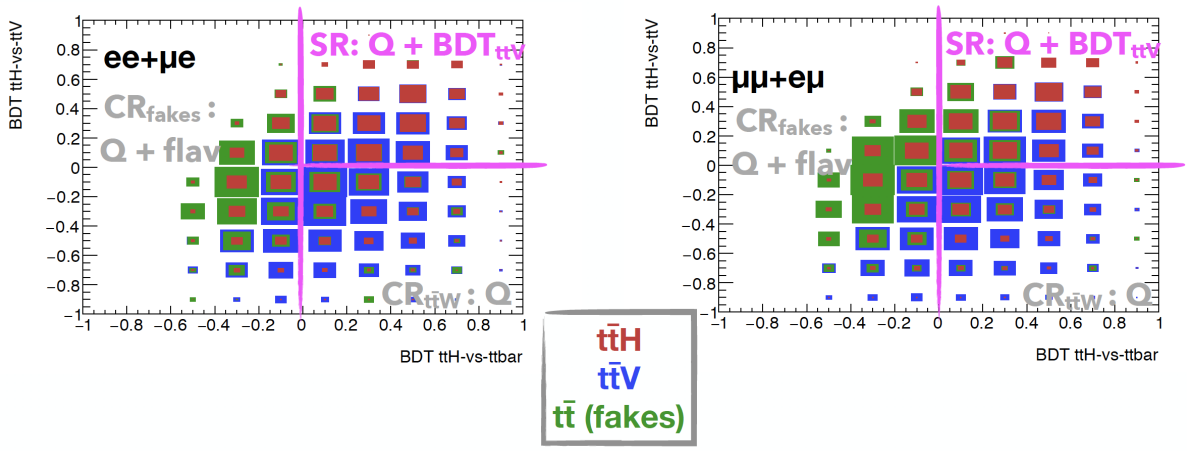


Figure A.1: Fit categories in the 2ℓSS region defined in the 2D space formed by the BDT ($t\bar{t}H$ vs $t\bar{t}V$) in the y-axis and the BDT ($t\bar{t}H$ vs $t\bar{t}$) in the x-axis. The rectangular cuts are used to define each region.

APPENDIX B

Conversion studies

It is found that the conversion electrons can originate from two different processes. Photons with virtual mass larger than twice 511 keV converting to a pair of electrons at the interaction point, is called *internal conversions*, and the converted photons resulting a pair of electrons in the detector material. These two processes have different kinematic properties and therefore needs to be evaluated differently in the background estimation techniques. The truth level studies showed that the conversion radius of internal conversions are observed $< 1mm$ while for the material conversions its $> 1mm$ mainly after 25mm where the beam-pipe is at 25-30mm in ATLAS (Figure B.1).

Therefore following variables are found to be efficient to distinguish those two sources of background electrons: the conversion radius, the invariant mass of the track associated to the electron and its closest track calculated at conversion vertex ($m_{trk-trk,CV}$), and the same invariant mass calculated at the primary vertex ($m_{trk-trk,PV}$). Figure B.2 shows the distribution of these variables for electrons originating from internal conversions and external conversions, as well as for prompt and isolated electrons. The final selection on the conversion electrons are following:

- External conversion candidate: a conversion vertex is found with radius $r > 20$mm, and the mass of the vertex is 0<m(trk-trk)$atCV$ <100MeV
- Internal conversion candidate: not an External conversion candidate and 0<m(trk-trk)$atPV$ <100MeV

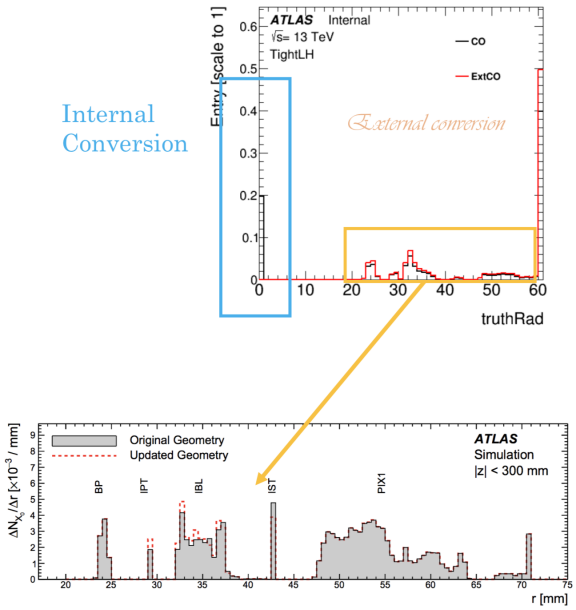


Figure B.1: Position of beam pipe and inner track detectors of ATLAS, 0.7mm is much smaller than any materials.

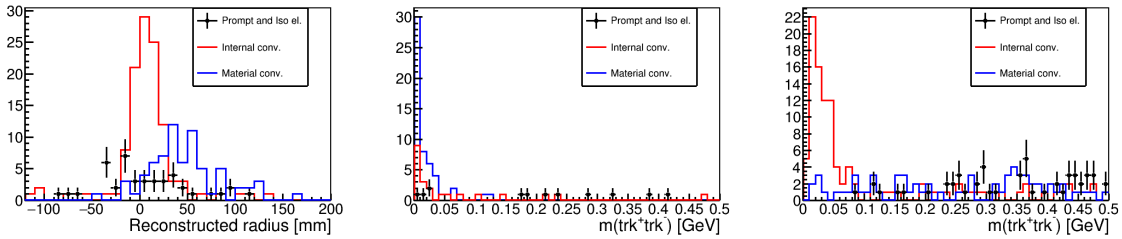


Figure B.2: Reconstructed radius for prompt, internal conversions, and material conversions (***left***); mass of the two associated tracks at the conversion vertex for candidates with radius $r > 20$mm (***middle***); mass of the two associated tracks at the primary vertex when no conversion vertex with $r > 20$mm has been found (***right***).

APPENDIX C

Unblinding results from May 2019

This section contains the results from the fit model that was approved for unblinding in March 2019. After this unblinding, some minor bug-fixes that were identified in the unblinding cross-checks and fixed. Detailed results of these cross-checks are not given and each step is performed by myself. These steps are:

- A small fraction of the data was not processes correctly, this is now added.
- The 3ℓ channel was approved with a 3-bin BDT, the 6-bin BDT was used in the original unblinding, this has been fixed
- Some minor backgrounds (W+jets and $V\gamma$) were omitted in the original unblinding results, these have now been added.

As a result of these improvements, observed $t\bar{t}H$ significance of the individual and combined non-tau/tau channels is shown in Table C.1. The 1.2 factor coming from EWK/QCD corrections were not included for $t\bar{t}W$ NF. Figure C.1 shows the ranking of the nuisance parameters in terms

Channels.	Observed $t\bar{t}H$ significance [σ]: stat.+syst. (stat.)	Expected $t\bar{t}H$ significance [σ]: stat.+syst. (stat.)
2ℓSS	0.44 (0.56)	1.89(2.59)
3ℓ	2.45 (2.94)	2.43(2.92)
4ℓ	0.72 (0.79)	1.22 (1.29)
Combined Non-Tau	2.12 (2.54)	2.89
1ℓ+$2\tau_{\text{had}}$	0.60 (0.60)	1.17 (1.35)
3ℓ+$1\tau_{\text{had}}$	0.48 (0.49)	1.06(1.12)
2ℓSS+$1\tau_{\text{had}}$	0.36 (0.05)	0.98(1.35)
Combined Tau	0.88 (0.67)	1.80
Combined Fit	2.13 (2.45)	3.16

Table C.1: Observed and expected $t\bar{t}H$ significance of the individual non-tau channels (2ℓSS, 3ℓ, and 4ℓ) and its combination, the individual tau channels (1ℓ+$2\tau_{\text{had}}$, 2ℓSS+$1\tau_{\text{had}}$, and 3ℓ+$1\tau_{\text{had}}$) and its combination and the full combination.

of its impact on the fitted signal strength and normalisation factors in the full fit.

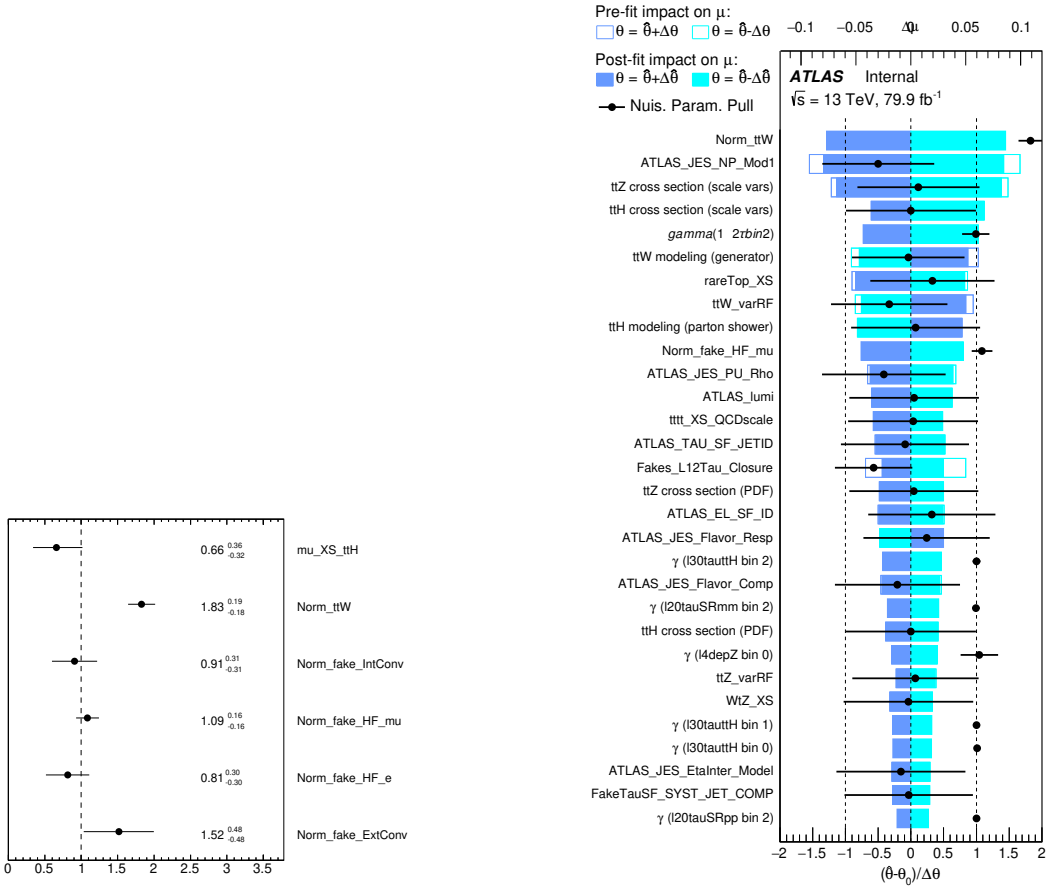


Figure C.1: The ranking of the nuisance parameters in terms of its impact on the fitted signal strength and normalisation factors in the full fit.

APPENDIX D

$t\bar{t}W$ studies

The pre- and post-fit plots for number of jets in 2ℓSS chanell in given in Figure D.1. The post-fit results are obtained from the fit described in Appendix C. The tension between fitted $t\bar{t}W$ NFs across the different jet multiplicity regions is improved in 3 $t\bar{t}W$ NF fit setup.

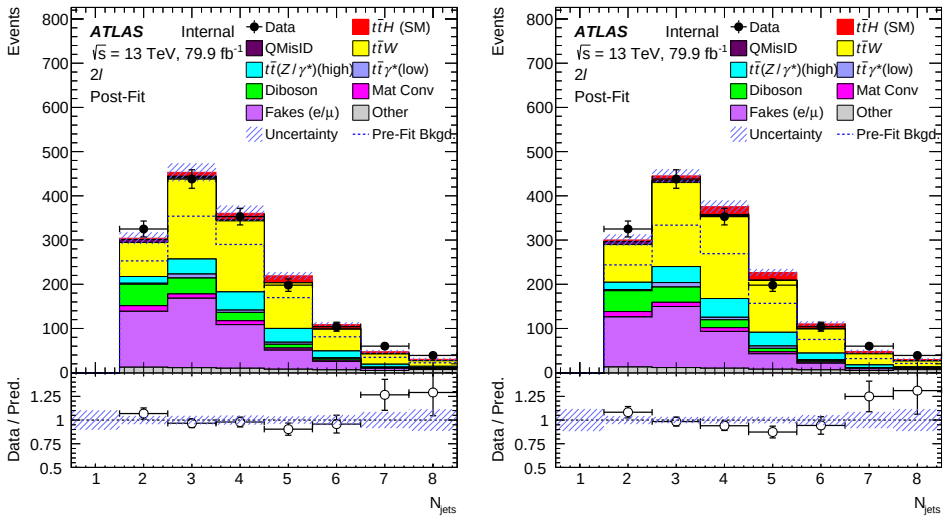


Figure D.1: Distributions of the jet multiplicity after the fit for 3 $t\bar{t}W$ NF (***left***) and 1 $t\bar{t}W$ NFs (***left***) in the 2ℓSS regions.

The comparison between the two fit setups, with and without extrapolation uncertainties for b-jet multiplicity and lepton charge, is shown in Figure D.2. The data-MC agreement is significantly improved with additional uncertainties.

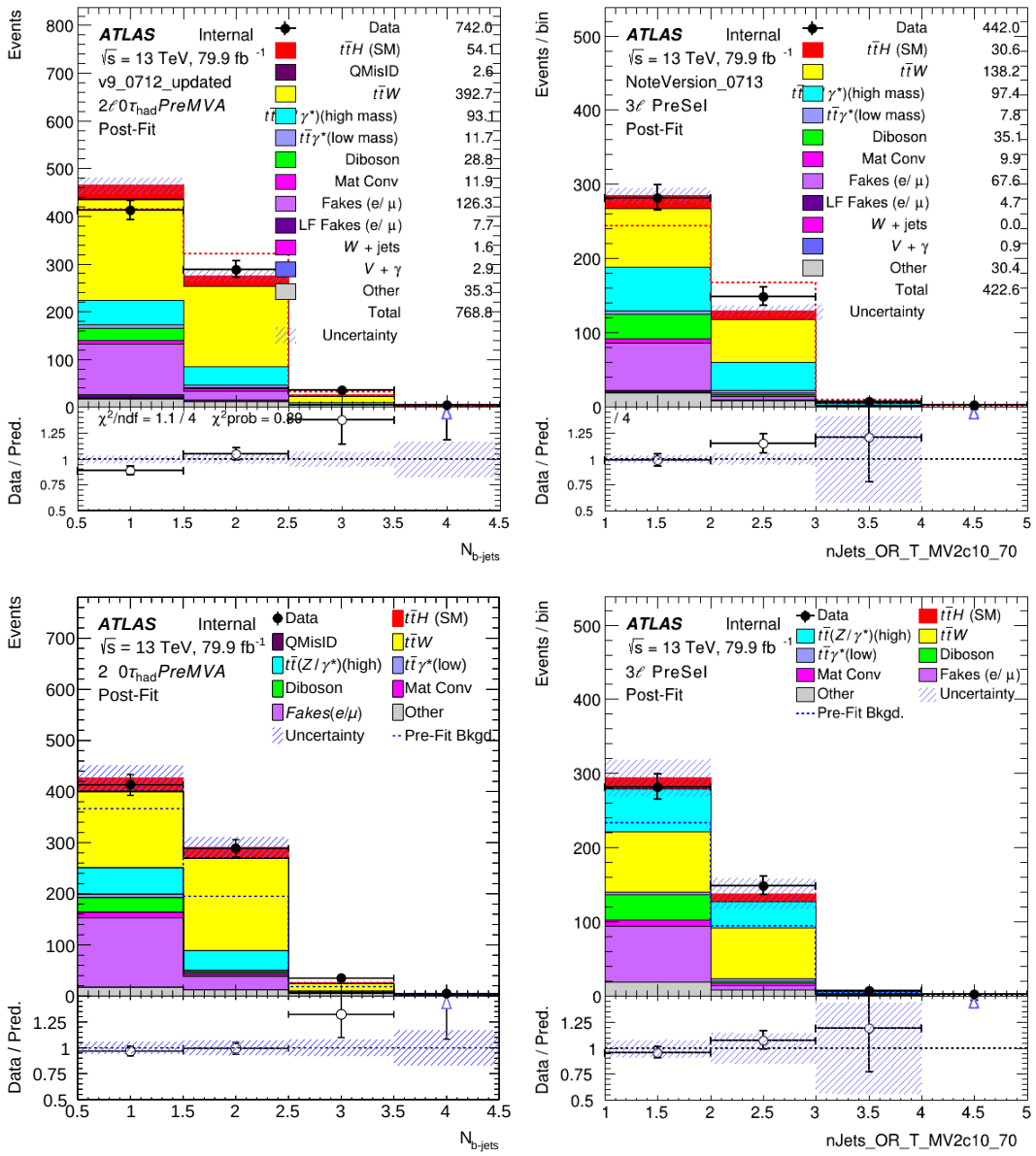
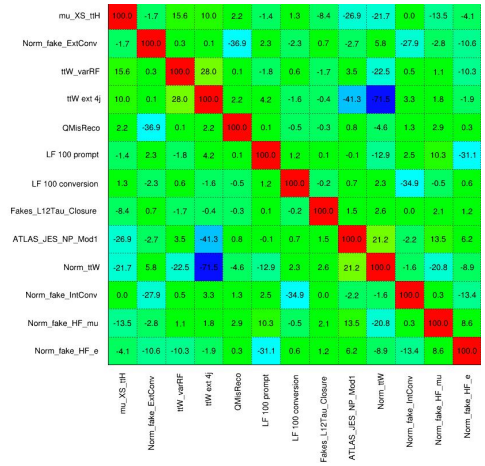


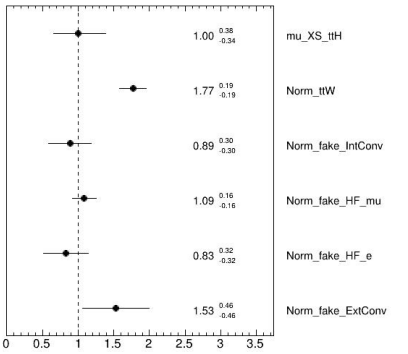
Figure D.2: The post-fit b-jet multiplicity is shown for the old fit mode (***top***) and new fit model (***bottom***) for the 2ℓSS (***left***) and 3ℓ (***right***) channels.

D.1 $t\bar{t}W$ extrapolation studies

Before deciding the final fit setup on $t\bar{t}W$, several studies were performed. During the time of these studies, the $t\bar{t}W$ was not scaled with 1.2 for higher order corrections. The nominal fit results without any extrapolation uncertainty on $t\bar{t}W$ background are shown in Figure D.3 and the expected significance is 3.16 σ. In order to resolve the observed mismodellings summarised



(a) Correlations of nuisance parameters, normalisation factors and signal strength.



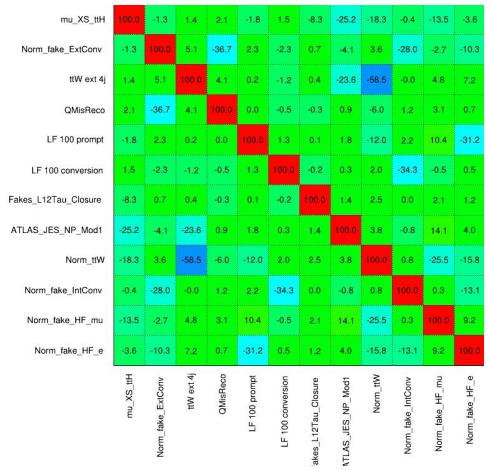
(b) Results of the Asimov fit of the signal strength and CR-only fit of the normalisation factors.

(c) NP pull and constraint on $t\bar{t}W$ related NPs.

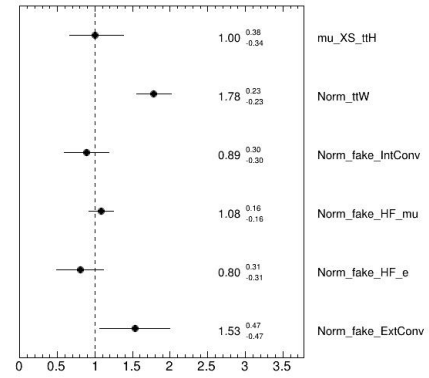
Figure D.3: The CR-only and asimov fit results without the extrapolation uncertainties.

in Section 7.5.3, the uncertainties were assigned to each of them and CR-only fit was evaluated to test the behaviour of the new NPs. Firstly, the prior of 50% extrapolation uncertainty were assigned to high number of jet region ($N_{\text{jets}} \geq 4$). The result of the fitted NFs in CR-only fit, the Asimov signal strength and their correlations with NPs are shown in Figure D.4. The expected significance were found to be 3.16 σ.

In the second step, the extrapolation uncertainty on the $N_{b-\text{jets}} \geq 2$ region on $t\bar{t}W$ with 50% prior was assigned. The result of the fitted NFs in CR-only fit, the Asimov signal strength and their correlations with NPs are shown in Figure D.5. The expected significance were found to be 3.17 σ. In the third step, the extrapolation uncertainty on the $plus - plus$ region on $t\bar{t}W$ with 50% prior was assigned. The result of the fitted NFs in CR-only fit, the Asimov signal strength and their correlations with NPs are shown in Figure D.6. The expected significance were found to be 3.10 σ. The uncertainties obtained so far were implemented together to the fit model and the result of this is given in Figure D.7. Overall significance on $t\bar{t}H$ remained unchanged with 3.15 σ. There are few conclusions can be drawn: $b-$jet extrapolation results in 4 $\sigma \rightarrow$ 1.3 σ reduction on the normalisation of $t\bar{t}W$ but there is large pull observed, the highest correlation is found to be between the normalisation of $t\bar{t}W$ and the N_{jets} extrapolation uncertaintiy. Another study was performed by assigning three normalisation factors for high number of jets, low number of jets for 2ℓSS and 3ℓ channels. It is seen that overall the higher $t\bar{t}W$ normalisation factor is preserved and therefore it was decided to proceed with this option to show also the consistent $t\bar{t}W$ normalisation factor through different channels. In addition, the uncertainties for $b-$jet and total charge is decorrelated between the 2ℓSS and 3ℓ channel since their behaviour are slighlty different, 3ℓ channel has less tention than 2ℓSS channel.



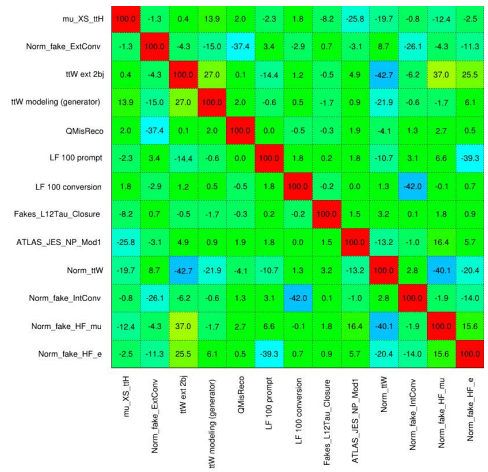
(a) Correlations of nuisance parameters, normalisation factors and signal strength.



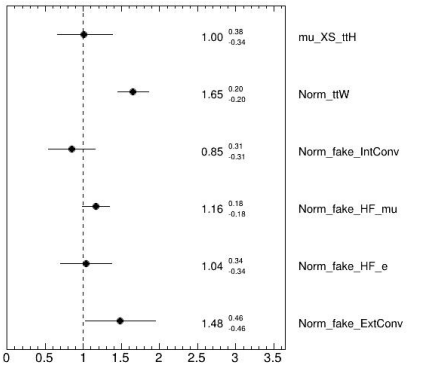
(b) Results of the asimov fit of the signal strength and CR-only fit of the normalisation factors.

(c) Associated NP pull and constraint.

Figure D.4: The CR-only and asimov fit results with extrapolation uncertainty on the jet multiplicity.



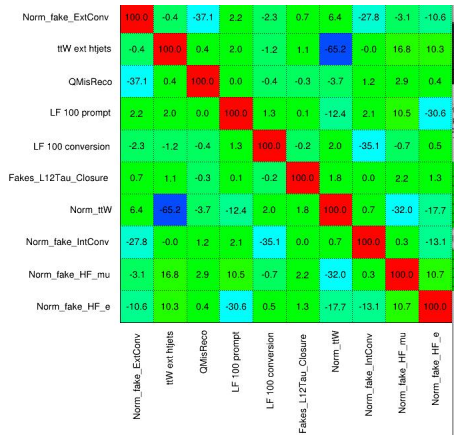
(a) Correlations of nuisance parameters, normalisation factors and signal strength.



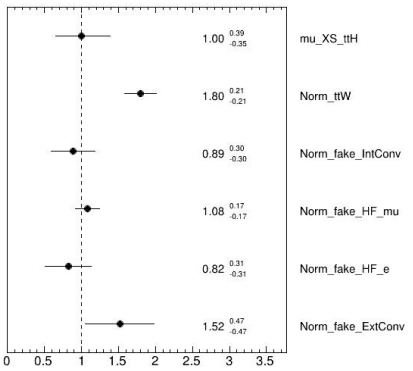
(b) Results of the Asimov fit of the signal strength and CR-only fit of the normalisation factors.

(c) Associated NP pull and constraint.

Figure D.5: The CR-only and asimov fit results with extrapolation uncertainty on the b-jet multiplicity.



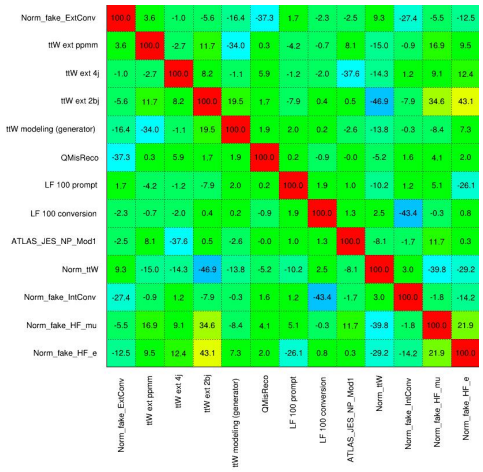
(a) Correlations of nuisance parameters, normalisation factors and signal strength.



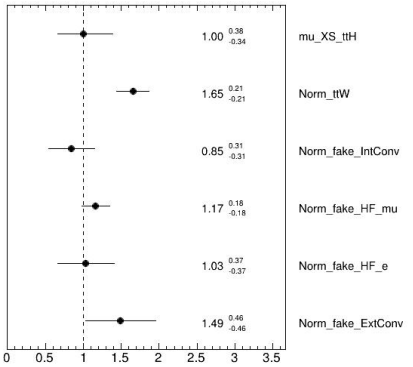
(b) Results of the Asimov fit of the signal strength and CR-only fit of the normalisation factors.

(c) Associated NP pull and constraint.

Figure D.6: The CR-only and asimov fit results with extrapolation uncertainty on the total charge of the leptons.



(a) Correlations of nuisance parameters, normalisation factors and signal strength.



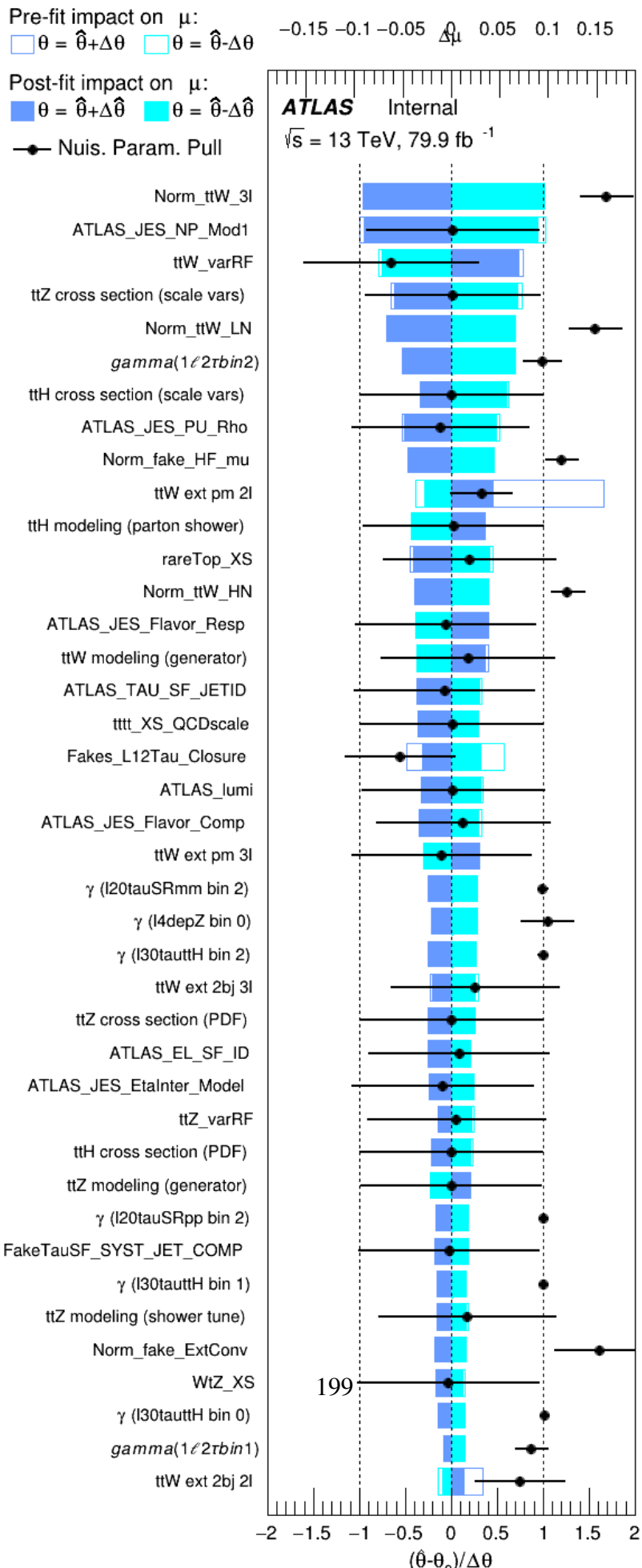
(b) Results of the Asimov fit of the signal strength and CR-only fit of the normalisation factors.

(c) Associated NP pulls and constraints.

Figure D.7: The CR-only and asimov fit results with all the extrapolation uncertainties explained above.

APPENDIX E

Full ranking of the nuisance parameters


Pre-fit impact on μ:
θ = θ̂+Δθ
θ = θ̂-Δθ
Post-fit impact on μ:
θ = θ̂+Δθ̂
θ = θ̂-Δθ̂
Nuis. Param. Pull
−0.15 −0.1 −0.05 0 0.05 0.1 0.15
Δμ
ATLAS Internal
√s = 13 TeV, 79.9 fb -1
Norm_ttW_3l
ATLAS_JES_NP_Mod1
ttW_varRF
ttZ cross section (scale vars)
Norm_ttW_LN
gamma(1ℓ2τbin2)
ttH cross section (scale vars)
ATLAS_JES_PU_Rho
Norm_fake_HF_mu
ttW ext pm 2l
ttH modeling (parton shower)
rareTop_XS
Norm_ttW_HN
ATLAS_JES_Flavor_Resp
ttW modeling (generator)
ATLAS_TAU_SF_JETID
tttt_XS_QCDscale
Fakes_L12Tau_Closure
ATLAS_lumi
ATLAS_JES_Flavor_Comp
ttW ext pm 3l
γ (l20tauSRmm bin 2)
γ (l4depZ bin 0)
γ (l30tauttH bin 2)
ttW ext 2bj 3l
ttZ cross section (PDF)
ATLAS_EL_SF_ID
ATLAS_JES_EtaInter_Model
ttZ_varRF
ttH cross section (PDF)
ttZ modeling (generator)
γ (l20tauSRpp bin 2)
FakeTauSF_SYST_JET_COMP
γ (l30tauttH bin 1)
ttZ modeling (shower tune)
Norm_fake_ExtConv
WtZ_XS
γ (l30tauttH bin 0)
gamma(1ℓ2τbin1)
ttW ext 2bj 2l
−2 −1.5 −1 −0.5 0 0.5 1 1.5 2
(θ̂-θ0)/Δθ

APPENDIX F

diHiggs 2ℓSS kinematic distributions

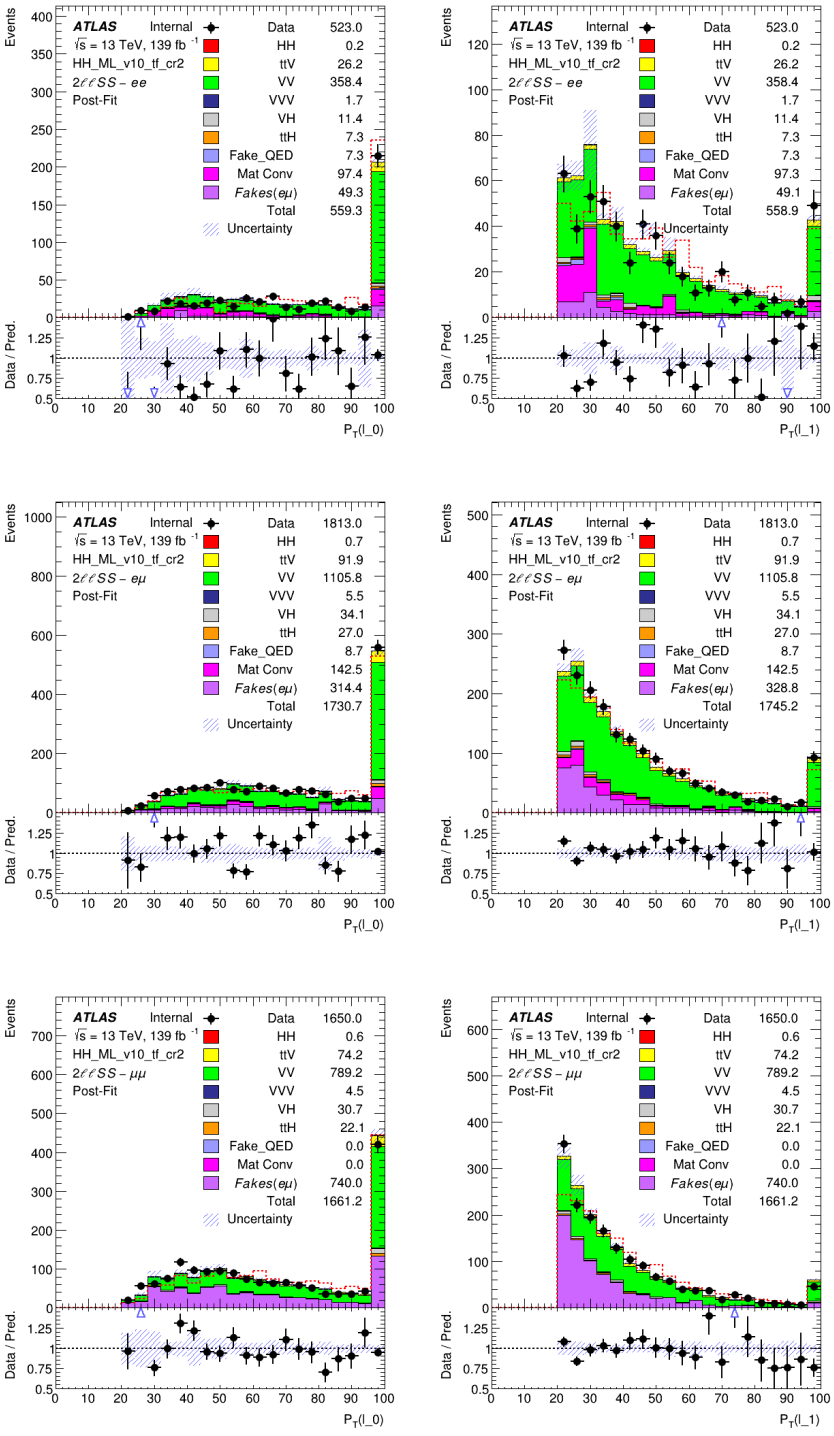


Figure F.1: The distributions of the p_T of the first lepton (***right***) and the p_T of the second lepton (***left***) for ee (***upper***), $e\mu$ (***middle***) and $\mu\mu$ (***lower***) channels. Signal without normalisation is shown with red line.

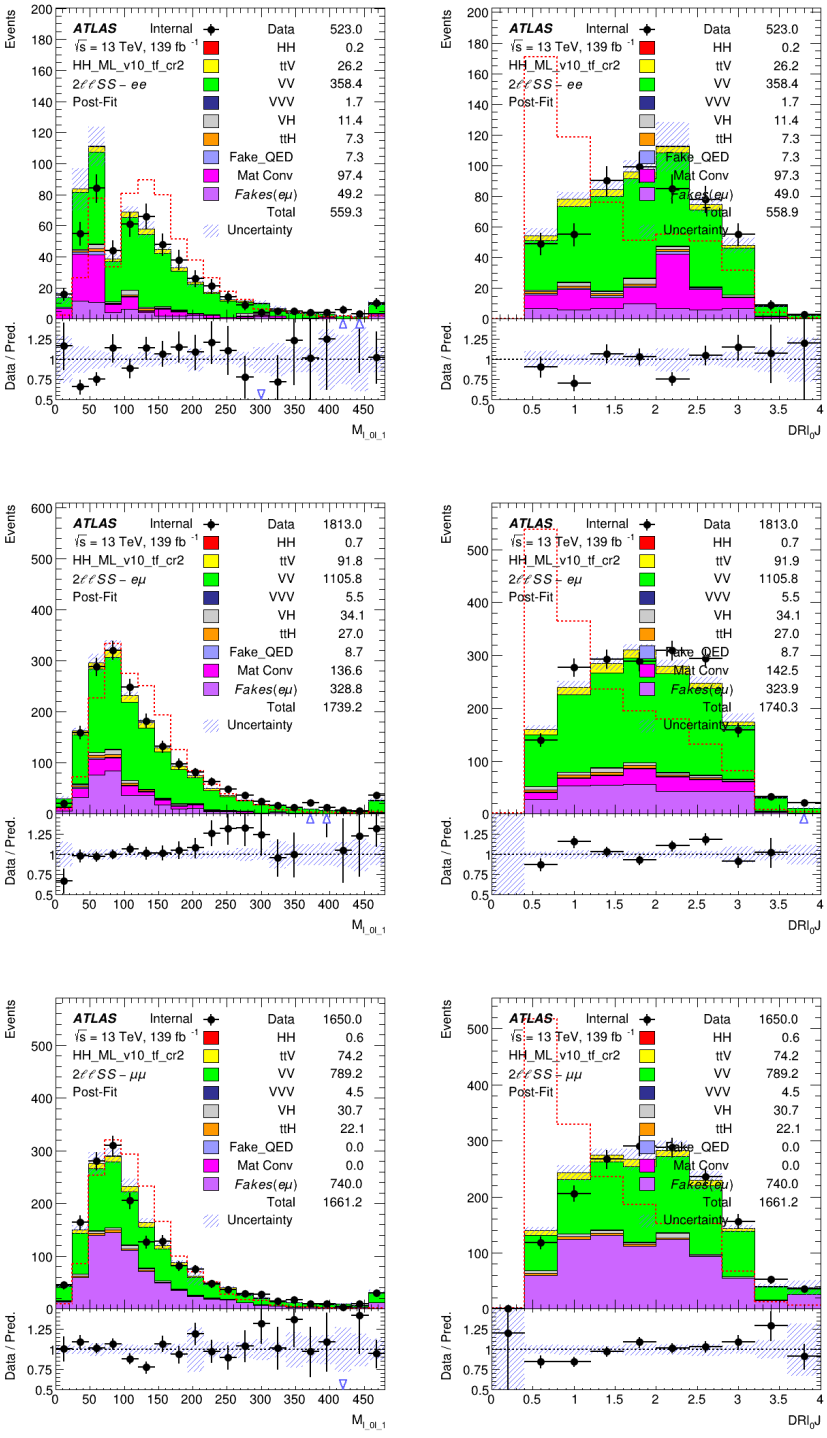


Figure F.2: The distributions of the invariant mass of two leptons (***right***) and the min angular distance between the jet and first lepton (***left***) for ee (***upper***), $e\mu$ (***middle***) and $\mu\mu$ (***lower***) channels. Signal without normalisation is shown with red line.

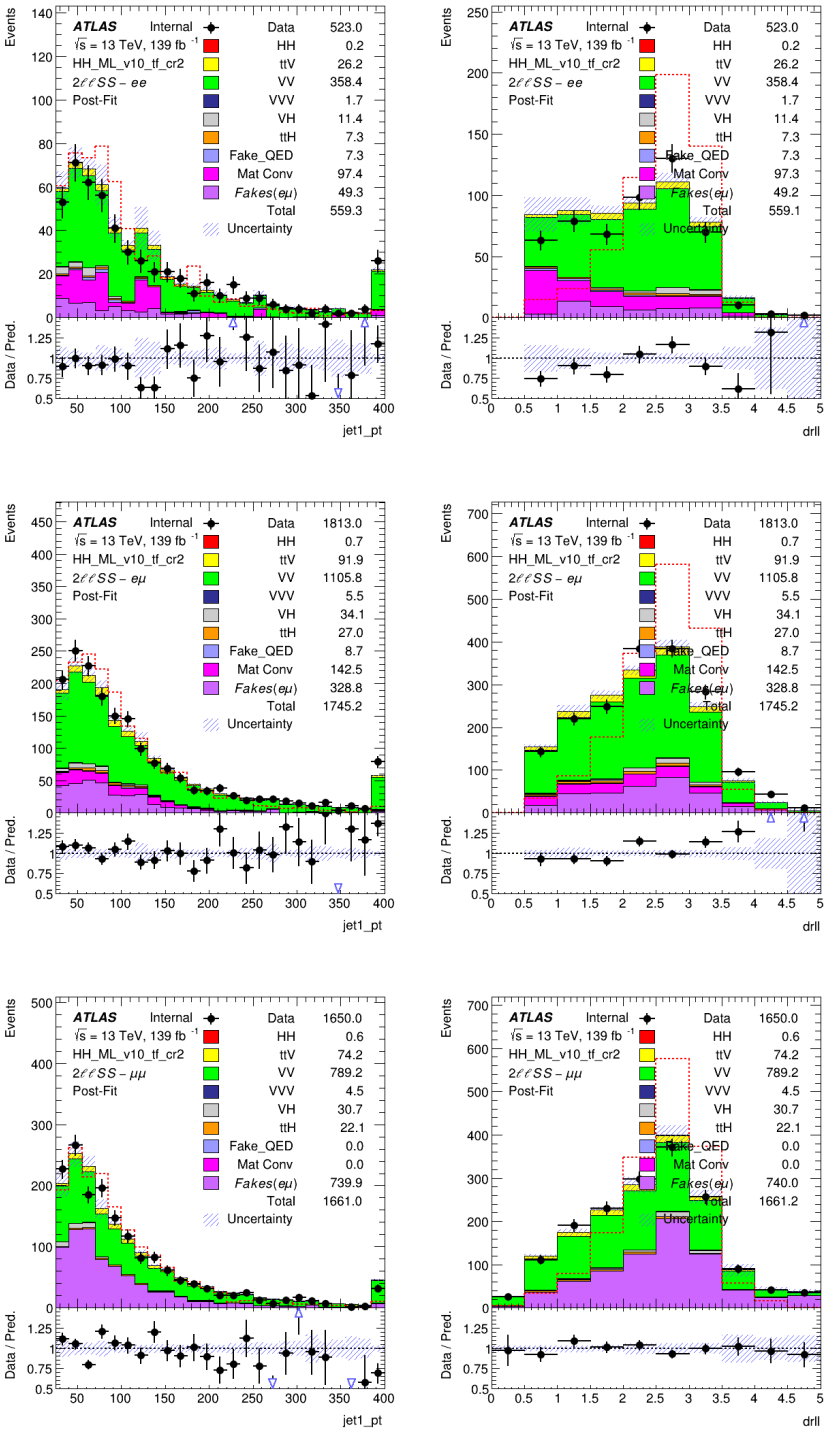


Figure F.3: The distributions of the leadind jet p_T (***right***) and the $\Delta R_{\ell,\ell}$ (***left***) for ee (***upper***), $e\mu$ (***middle***) and $\mu\mu$ (***lower***) channels. Signal without normalisation is shown with red line.

List of Figures

List of Tables