

Event reconstruction of  $t\bar{t}$  and  $t\bar{t}H$  using  
the Kinematic Likelihood Fitter in final  
states with two same-sign electric charge  
leptons

Shreya Saha

Master Thesis

Department of Physics

McGill University

Montreal, Quebec

2017-14-08

A thesis submitted to McGill University in partial fulfilment of the requirements of  
Master's Degree

©Shreya Saha 2017  
All rights reserved

## ACKNOWLEDGEMENT

First, I would like to express my deep and sincere thanks to my supervisor Professor Steven Robertson for constantly advising me to push my boundaries and encouraging critical thinking in every area of my research. I would like to thank McGill research assistant, Dr. Tamara Vázquez Schröder, for her immense support in teaching me the core of my analysis from the very basics and explaining every bit of detail with patience and never getting tired of my questions. I would like to extend a special thanks to the professors, research assistants, senior PhDs and fellow MSc students in the McGill- ATLAS group for their insightful discussions and feedback and helping me with all kinds of technical issues I faced during my research work. I would also like to thank the group for sponsoring my stay at CERN which helped me to gain a better insight about my work. I would like to express my gratitude to the ttH-Multilepton group at CERN and to my collaborators, especially, Marco Sessa from the Roma-Tre group for helping me with the basic framework of my analysis code and general discussions. Last, but not the least, I would like to thank my family, friends and office mates who have been a constant support and source of joy during the entirety of two years.

## ABSTRACT

One of the important tests for the validity of the Standard Model of particle physics is the measurement of the top quark Yukawa coupling, which can be directly measured in the production of the top-quark pair in association with a Higgs boson,  $t\bar{t}H$ . In this thesis, a likelihood-based fitting tool, known as KLFitter, is implemented to improve the separation power between the  $t\bar{t}H$  signal and the dominant background contribution arising from the top quark pair production,  $t\bar{t}$ , in the final states containing two same-sign electric charge leptons. NLO Monte Carlo samples of the ATLAS detector at  $\sqrt{s} = 13$  TeV is used to test the fit. Improvement of the separation power of the fitter is studied by modifying parameters in the likelihood definition of the fit. Different cases have been analyzed and a better discriminating variable, after modifications to the original KLFitter, has been achieved regarding the performance of the fit.

## ABRÉGÉ

La mesure du couplage de Yukawa du quark top, en particulier via la production de pair de quarks top en association avec un boson de Higgs, est très pertinente pour tester la validité du Modèle Standard. Cette thèse présente les performances de KLFitter, un outil probabiliste développé pour améliorer le pouvoir de discrimination entre le signal  $t\bar{t}H$  et la contribution de bruit dominante: la production de pair de quarks top dont la désintégration donne deux leptons ayant la même charge électrique. À l'aide d'algorithmes Monte Carlo, des événements sont générés avec une précision au deuxième ordre, leur interaction avec le détecteur est simulée puis les signaux émulsés sont reconstruits et utilisés pour tester KLFitter. L'impact des paramètres du fit de KLFitter sur le pouvoir de discrimination est étudié. Différents cas sont analysés et la variable discriminante optimale est choisie en fonction des performances observées.

## TABLE OF CONTENTS

	ACKNOWLEDGEMENT . . . . .	ii
	ABSTRACT . . . . .	iii
	ABRÉGÉ . . . . .	iv
1	Introduction . . . . .	1
2	Theory . . . . .	3
	2.1 The Standard Model . . . . .	3
	2.2 Top Quark Production and Decay . . . . .	8
	2.3 Higgs Boson Production and Decay . . . . .	10
	2.3.1 Associated Top Quark Pair Production with a Higgs Boson	11
3	Experimental Setup . . . . .	14
	3.1 The Large Hadron Collider . . . . .	14
	3.2 The ATLAS Detector . . . . .	17
	3.2.1 ATLAS Geometry . . . . .	17
	3.2.2 Inner Detector . . . . .	19
	3.2.3 Calorimetry . . . . .	20
	3.2.4 Muon Spectrometer . . . . .	25
	3.2.5 Trigger and Data Acquisition . . . . .	25
4	Monte Carlo Generators and Object Definition . . . . .	27
	4.1 Monte Carlo Generators used . . . . .	27
	4.1.1 Event Generation . . . . .	27
	4.1.2 Detector Simulation . . . . .	29
	4.2 Object Definition and Selection . . . . .	30
	4.2.1 Light Leptons . . . . .	31
	4.2.2 Jets . . . . .	34
	4.2.3 B-jets . . . . .	34

	4.2.4	Missing Transverse Energy . . . . .	34
5		Kinematic Likelihood Fitter . . . . .	36
	5.1	Introduction to Kinematic Fitting . . . . .	36
		5.1.1 Kinematic Fitting . . . . .	36
		5.1.2 The Maximum Likelihood Method . . . . .	36
	5.2	Kinematic reconstruction using KLFitter in the <i>semileptonic</i> decay channel for $t\bar{t}$ production . . . . .	37
		5.2.1 Derivation of Constraints . . . . .	39
		5.2.2 Breit-Wigner Distribution . . . . .	42
		5.2.3 Transfer Functions . . . . .	44
		5.2.4 KLFitter Likelihood . . . . .	45
6		KLFitter Performance Studies . . . . .	48
	6.1	KLFitter <i>lepton + jets</i> likelihood for $t\bar{t}$ and $t\bar{t}H$ event reconstruc- tion in the $2LSS + 0\tau$ final state . . . . .	48
	6.2	Modified KLFitter Likelihood with an Additional Lepton in the $2LSS + 0\tau$ channel . . . . .	50
		6.2.1 Case 1 - Minimizing the angular separation ( $\Delta R$ ) between additional lepton and b-jets . . . . .	51
		6.2.2 Case 2- Merging of the closest b-jet and additional lepton .	52
7		Results . . . . .	54
8		Discussion of Results . . . . .	62
	8.1	Conclusion and Outlook . . . . .	65
		REFERENCES . . . . .	68

## CHAPTER 1

### Introduction

The discovery of the Higgs boson in July 2012 by the ATLAS and CMS experiments at CERN, was a significant contribution in the field of high energy physics [1]. To test whether the observed Higgs boson follows the predictions of the Standard Model (SM) [2, 3, 4, 5, 6], careful study and measurement of its properties is of utmost importance. One of the properties of the Higgs boson [8] currently being studied, is related to its strength of interaction to other elementary particles in the SM, termed as Yukawa coupling [9]. The coupling is directly proportional to the mass of the particles in the SM. Since the top quark is the heaviest elementary particle in nature, the top-Higgs coupling, also referred to as the top-Yukawa coupling, has a measurable value in HEP experiments and is paramount to an understanding of the electroweak symmetry breaking and could provide an essential probe for new physics, if the value deviates from the SM- predicted value of  $0.9956 \pm 0.0043$  [9].

While most production processes at hadronic colliders provide only an indirect measurement of the coupling via loop effects, the associated production of the top quark pair with the Higgs boson,  $t\bar{t}H$ , provides a direct tree-level measurement of the coupling strength [20]. The  $t\bar{t}H$  process is probed through various Higgs boson and top quark decay channels. In this thesis, decay channels characterized by high lepton multiplicity will be studied, namely, the final states producing two same-sign electrically charge leptons. The major challenge for analyzing this particular

decay channel, is the high background contribution associated with the decay of top-antitop quark pair,  $t\bar{t}$ , producing similar final state particles. Powerful discriminating variables are currently being developed to extract the signal from the background events. The primary goal of the analysis is to implement and improve a likelihood fitting tool, known as the Kinematic Likelihood Fitter (KLFitter) [50], to reconstruct the  $t\bar{t}$  event topology in the two same-sign electric charge leptons final state and distinguish between the signal and the background.

The thesis is structured as follows. The theory is outlined and explained in Chapter 1, which introduces the Standard Model and the production and decay of the top quark and the Higgs boson, along with the associated production of the Higgs boson with the top quark pair. Chapter 2 is dedicated to the explanation of the experimental setup for the analysis, focusing on the design and performance of the ATLAS detector. Details of the Monte Carlo generators used and the definition and reconstruction of objects in the detector is elaborated in chapter 4. Chapter 5 provides an extensive explanation about the concept of kinematic fitting and the KLFitter package. Implementation of KLFitter to the related event topologies is provided in chapter 6, along with the modifications made to the fit to improve the discriminating factor between the signal and background processes. The performance of the fit is presented in chapter 7 and the results are discussed in chapter 8 along with the conclusion and future prospects.

Natural units have been used throughout this thesis:  $c = \hbar = k_B = 1$ .

## CHAPTER 2

### Theory

#### 2.1 The Standard Model

The SM is a successful theoretical framework explaining the three fundamental forces (electromagnetic, strong, and weak nuclear force) in nature and their interactions with elementary particles. Gravity is currently not integrated in the theory of the SM.

The SM is a combination of relativistic quantum field theories describing the interactions of half-integer spin particles, fermions, and integer spin particles, bosons, through a local gauge invariance of the following symmetry groups:  $SU(3)_C \times SU(2)_L \times U(1)_Y$ , wherein,  $SU(3)_C$  and  $SU(2)_L \times U(1)_Y$  refer to the strong and electroweak interactions, respectively. The gauge bosons are associated with the specific generators of the symmetry groups, and the conserved charges in the SM, such as the electromagnetic charge, is a result of Noether's theorem, which relates symmetries in nature to conserved quantities [7]. The allowed interactions between fermions and bosons, and between different bosons, are governed by specific properties of these symmetry groups. In Abelian symmetries, such as the  $U(1)_q$  group, self interactions between photons is forbidden, as opposed to the  $SU(3)_C$  symmetry group of quantum chromodynamics (QCD), which is non-Abelian and particles associated with this group (gluons) can carry color charge and interact with each other.

The SM is mathematically formulated in terms of its Lagrangian,  $\mathcal{L}_{SM}$ , which contains kinetic terms describing free fermion and gauge fields,  $\mathcal{L}_{kin}$ , interaction terms describing interactions between fermion fields via gauge fields,  $\mathcal{L}_{int}$ , and terms related to the Higgs field that breaks the electroweak symmetry and provide Yukawa couplings to the fermions,  $\mathcal{L}_{Higgs}$  (refer equations 2.1- 2.5) [11].

$$\mathcal{L}_{SM} = \mathcal{L}_{kin} + \mathcal{L}_{int} + \mathcal{L}_{Higgs} \quad (2.1)$$

$$\mathcal{L}_{kin} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{2}\text{tr}[W_{\mu\nu}W^{\mu\nu}] - \frac{1}{2}\text{tr}[G_{\mu\nu}G^{\mu\nu}] \quad (2.2)$$

Here,  $B_{\mu\nu}$  denotes the gauge field tensor for the U(1) group,  $W_{\mu\nu}$  is associated with the gauge field tensor of the SU(2) group and  $G_{\mu\nu}$  is the gluon field tensor. The trace over the SU(2)<sub>L</sub> and SU(3)<sub>C</sub> indices, respectively, is represented by “tr” above.

The interaction terms in the SM are divided into electroweak interactions governed by the SU(2)<sub>L</sub> × U(1)<sub>Y</sub> symmetry group, and interactions in the strong sector between quarks and gluons within the SU(3)<sub>C</sub> group.

$$\mathcal{L}_{int,EW} = \sum_{\psi} \bar{\psi}\gamma^{\mu}(i\partial_{\mu} - g'\frac{1}{2}Y_W B_{\mu} - g\frac{1}{2}\tau W_{\mu})\psi \quad (2.3)$$

In the above equation, the U(1) gauge field is shown by  $B_{\mu}$ ;  $Y_W$  is the weak hypercharge associated with the U(1) group;  $W_{\mu}$  is the SU(2) gauge field; and Pauli matrices is represented by  $\tau$ .  $\psi$  represents the fermion field and  $\gamma^{\mu}$  is the Dirac gamma matrix.  $g$  and  $g'$  are the SU(2) and U(1) coupling coefficients, respectively.

$$\mathcal{L}_{int,QCD} = i\bar{U}(\partial_{\mu} - ig_s G_{\mu}^a T^a)\gamma^{\mu}U + i\bar{D}(\partial_{\mu} - ig_s G_{\mu}^a T^a)\gamma^{\mu}D \quad (2.4)$$

The interaction terms in the Lagrangian for QCD include the Dirac spinors associated with up and down-type quarks, shown by  $U$  and  $D$ , respectively. Further details about the quark sector is presented in the following sections. In the above equation,  $g_s$  refers to the strong coupling constant,  $G_\mu^a$  is the gluon field tensor and  $T^a$  is the generator of the SU(3) symmetry group.

The Higgs part of the Lagrangian is shown as below, which involves the scalar Higgs field and plays a pivotal role in electroweak symmetry breaking, and further coupling to the fermions via Yukawa coupling.

$$\mathcal{L}_{Higgs} = [(\partial_\mu - igW_\mu^a t^a - ig'Y_W B_\mu)\phi]^2 + \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \quad (2.5)$$

Here,  $\phi$  represents the Higgs field;  $W_\mu^a$ ,  $B_\mu$ ,  $Y_W$ ,  $g$  and  $g'$  follow the same definitions as presented above.  $\lambda$  is the coupling constant and  $\mu$  is associated with the mass of the Higgs boson.

The particle sector of the SM is broadly divided into two major categories, namely, fermions and bosons as seen above. A schematic picture of the particle content is shown in figure 2-1 and is further detailed in the following sections.

**Fermions** - Fermions are half-integer spin particles which are characterized by Fermi-Dirac statistics and obey the Pauli-Exclusion principle. There are a total of 12 distinct fermions in the SM, divided into quarks (6) and leptons (6). Both quarks and leptons are arranged in three weak hypercharge doublets, thus representing three particle generations. Each quark possesses its own unique flavors, with a distinct mass hierarchy increasing from the 1<sup>st</sup> to the 3<sup>rd</sup> generation [5].

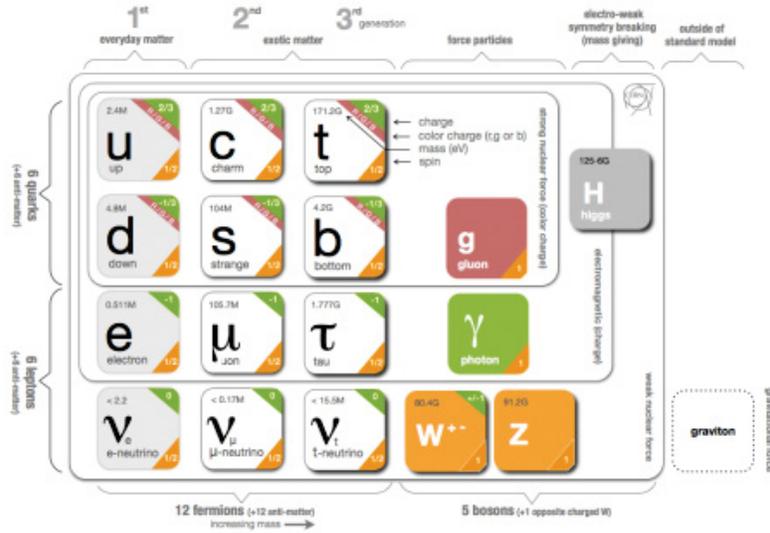


Figure 2–1: Particle content of the Standard Model showing three generations of quarks and leptons, gauge bosons including eight gluons, two W bosons, Z boson and the photon and a scalar Higgs Boson [10].

The quark sector consists of an up-type quark and a down-type quark with charges  $+2/3$  and  $-1/3$ , respectively. Up and down quarks form the  $1^{st}$  generation with charm and strange quarks, and top and bottom quarks making up the  $2^{nd}$  and  $3^{rd}$  generations as seen in figure 2-1. In addition to electrical charges, quarks also possess three different color charges (red, green and blue) and interact via the electromagnetic, weak and strong forces.

Leptons include the negatively-charged electrons, muons and taus along with their corresponding neutral neutrinos. They do not carry color charge and hence do not interact via the strong force.

**Bosons** - Bosons are subject to Bose-Einstein statistics and have an intrinsic integer spin. Gauge or vector bosons include the  $W^+$ ,  $W^-$ ,  $Z$ , photon and gluons which have spin 1 and scalar boson includes the Higgs boson with spin 0. They act as the force mediators in the model, with the exception of the Higgs boson. The  $W^+$ ,  $W^-$  and  $Z$  bosons mediate the short-range weak force and the photon is the force carrier for long-range electromagnetic interaction. Particles interact via the short-range strong interaction through the eight massless and colored gluons. The Higgs boson is an outcome of the spontaneous electroweak symmetry breaking (EWSB) and is essential to understand the mass generation of elementary particles.

**Electroweak Symmetry Breaking** - The mathematical formulation of the SM is described by the following gauge group -  $SU(3)_C \times SU(2)_L \times U(1)_Y$ .  $SU(3)_C$  refers to the gauge symmetry of the strong interaction and is studied under QCD [13]. The subscript  $L$  in  $SU(2)_L$  indicates that the electroweak sector only interacts with the left-handed fermions and  $Y$  in  $U(1)_Y$  refers to the weak hypercharge quantum number carried by the particles [11]. Above the energy scale of around 100 GeV [14], the electromagnetic and weak forces unify giving rise to the electroweak interactions which is explained by the  $SU(2)_L \times U(1)_Y$  symmetry group. The electroweak symmetry is spontaneously broken by the Higgs field, which has a vacuum expectation value ( $v$ ) of 246 GeV. This broken gauge symmetry, also known as the Higgs mechanism, gives mass to the  $W$  and the  $Z$  bosons and predicts the existence of a massive scalar Higgs boson, which was discovered at CERN in 2012 [17] [18]. The  $SU(2)_L \times U(1)_Y$  symmetry is partly broken to a  $U(1)_q$  subgroup, so that three of the four gauge bosons ( $W^+$ ,  $W^-$  and  $Z$  bosons) become massive, while the photon remains

massless due to the residual unbroken  $U(1)_q$  symmetry. Note that the subscript  $q$  in  $U(1)_q$  represents the standard electric charge. The following equations show the mass generation of the weak gauge bosons and the fermions [11]:

$$M_W^2 = \frac{g^2 v^2}{4} \quad (2.6)$$

$$M_Z^2 = \frac{(g'^2 + g^2)v^2}{4} \quad (2.7)$$

In equations 2.6 and 2.7,  $g$  and  $g'$  refer to the gauge coupling of the  $SU(2)_L$  and  $U(1)_Y$  group, respectively.

Further exploitation of the presence of a scalar Higgs field gives rise to the masses of the fermions as shown in equation 2.3 below:

$$m_{fi} = \frac{h_{fi}v}{\sqrt{2}} \quad (2.8)$$

In the above equation,  $f_i$  refers to the generation of quarks and leptons and  $h_{fi}$  refers to the Yukawa coupling, which is defined as the strength of the interaction between the fermions and the Higgs field [11]. It can be seen that the Yukawa coupling has a linear dependence on the mass of the fermion and hence, the top quark has the highest value of the coupling among the fermions.

## 2.2 Top Quark Production and Decay

The top quark is the heaviest known elementary particle, with a mass of  $173.34 \pm 0.27$  GeV [15] and was discovered by the CDF and D0 collaborations at Fermilab in 1995 [16]. It is an up-type quark of the 3<sup>rd</sup> generation with a charge of  $+2/3$ . It decays before it can form any bound states due its extremely short lifetime of  $0.5 \times 10^{-24}$  s, which is somewhat smaller than the average hadronization time of

quarks ( $3 \times 10^{-24}$  s) and hence, the properties of the final state particles are free from effects associated with hadronization of the top quark.

Top quarks at the Large Hadron Collider are produced individually via the weak interaction (figure 2-2) and as top-antitop ( $t\bar{t}$ ) pairs via the strong interaction (figure 2.3). The focus of the analysis is on top-quark pair production.

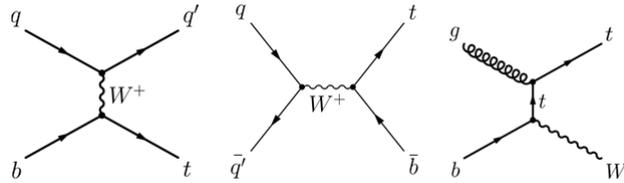


Figure 2-2: Electroweak single top production: t-channel (left), s-channel (center) and  $W$ -associated production ( $tW$ )

Almost 100% of the top quark pairs decay into a pair  $W$  boson and a pair of down-type quarks and the branching ratio of a single top decaying into a  $W$  boson and a bottom quark is around 91% [12].

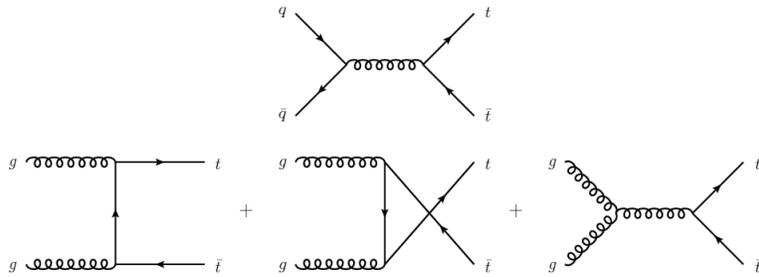


Figure 2-3: Leading order diagrams for top quark pair production at the LHC

The final state signatures in the detector corresponds to the subsequent decay of the  $W$  bosons, giving rise to three different decay topologies of the top quark pair as mentioned below :

- Only jets (Case I - *allhadronic* decay),
- A lepton, its corresponding neutrino, and jets (Case II - *semileptonic* decay),
- Only leptons and neutrinos from  $W$  boson decay and two bottom quark jets from the top quark pair decay (Case III - *dileptonic* decay).

The corresponding values on the right of each decay process is calculated by considering the branching ratio of each decay, where the branching ratio for  $W \rightarrow q\bar{q}$  is 67.41% and 10.86% for the decay process,  $W \rightarrow l\nu_l$  [12].

- I.  $t\bar{t} \rightarrow W^+b W^- \bar{b} \rightarrow q\bar{q}b q\bar{q} \bar{b}$ , (45.7%)
- II.  $t\bar{t} \rightarrow W^+b W^- \bar{b} \rightarrow q\bar{q}bl^-\bar{\nu}_l\bar{b} + l^+\nu_lb q\bar{q} \bar{b}$ , (43.8%)
- III.  $t\bar{t} \rightarrow W^+b W^- \bar{b} \rightarrow l^+\nu_lb l^-\bar{\nu}_l\bar{b}$ , (10.5%)

Leptons originating from the primary decay of the  $W$  boson (here,  $W \rightarrow l\nu_l$ ) are termed as prompt leptons and generally have a higher momentum as compared to non-prompt leptons which mainly come from the *semileptonic* decays of hadrons containing bottom or charm quarks.

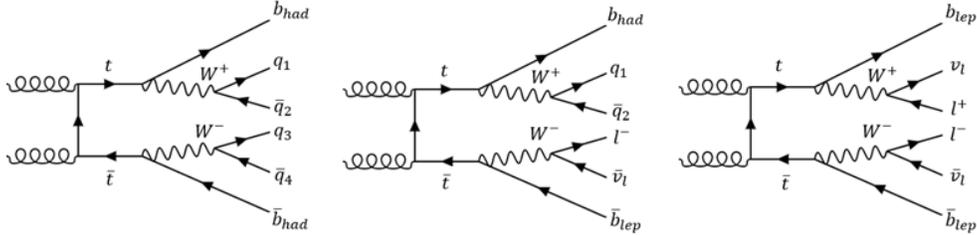


Figure 2–4: Top pair decay channel: From left to right: Case I (*allhadronic* channel), Case II (*semileptonic* channel) and Case III (*dileptonic* channel)

### 2.3 Higgs Boson Production and Decay

The SM- like Higgs boson was discovered by the ATLAS and CMS collaborations at CERN in July 2012 [1]. It is a scalar boson which is measured to have a spin of

0 and has an observed mass of  $125.09 \pm 0.21$  GeV [17]. In the theory of the SM, the Higgs boson is predominantly produced in  $p\bar{p}$  colliders by gluon fusion, followed by, weak-boson fusion, associated production with a gauge boson, and associated production with a top-quark pair. Production mechanisms at the LHC are shown in figure 2-5(a). The Higgs boson mostly decays into a pair of bottom quark jets (b-jets),  $H \rightarrow b\bar{b}$ , followed by decay into a  $W$  boson pair,  $H \rightarrow WW^*$ , as inferred from the branching ratio values at the observed mass of 125 GeV in table 2-5(b).

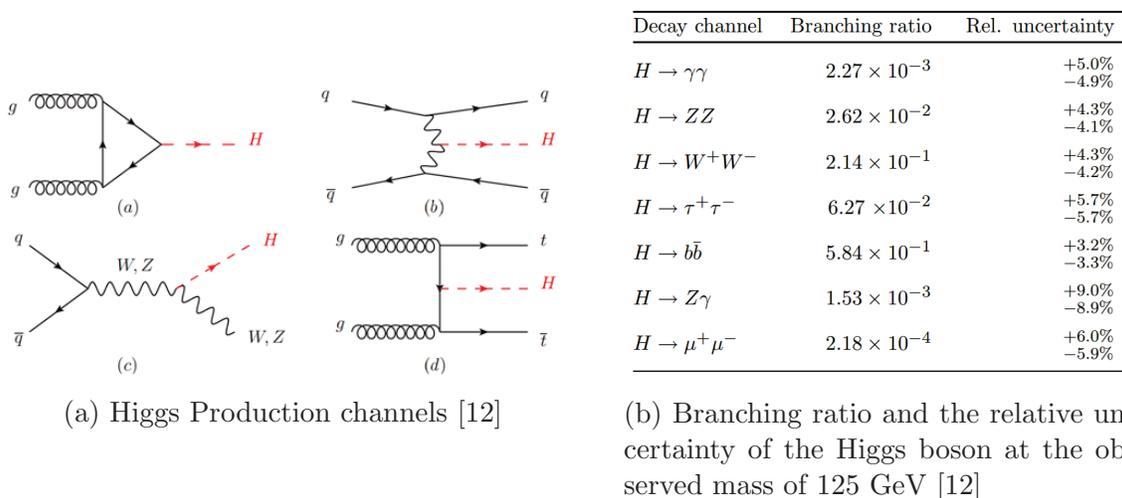


Figure 2-5: Higgs production mechanism and branching ratio

### 2.3.1 Associated Top Quark Pair Production with a Higgs Boson

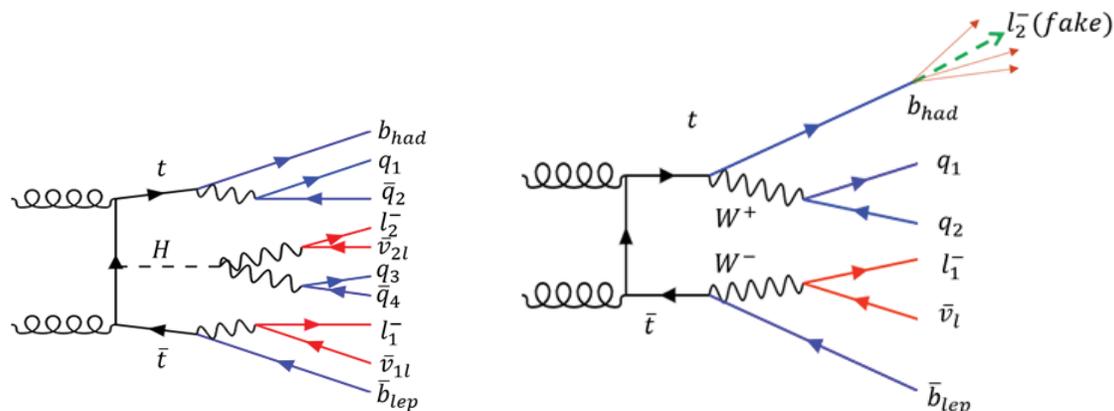
A precise measurement of the top-quark Yukawa coupling from the production rate for the  $t\bar{t}H$  process is essential to validate the electroweak symmetry breaking in the SM. Any substantial deviation from the SM value of the coupling inferred from the measured top mass, could hint towards new physics. The top quark is the only quark which has a value of the coupling in the order of unity owing to its high

mass,  $h_t = 0.9956 \pm 0.0043$  [9]. The Yukawa coupling value can be obtained from the cross section of the production of a Higgs Boson through a top quark loop ( $gg \rightarrow H$ ) and through the top quark loop of the decay  $H \rightarrow \gamma\gamma$ . However, it can also be determined through direct tree-level (leading order) measurement at the lowest order in perturbation theory through the production of a top-quark pair and a Higgs boson ( $gg \rightarrow t\bar{t}H$ ) as seen in figure 2-5 (a).

The analysis presented in this thesis focuses on the final state of  $t\bar{t}H$  involving multiple leptons (“multileptonic”), in particular, final states with two same-sign charged light leptons (electrons or muons) signatures ( $2LSS + 0\tau$ ) as in figure 2-6 (a). In the above notation,  $2L$  stands for two light leptons,  $SS$  stands for “same sign”, which refers to identical electric charges for the leptons, and no  $\tau$  candidates are permitted in this search. Both the final state leptons produced in the  $2LSS$  channel are required to be prompt leptons originating from the decay of one of the  $W$  boson from the top quark pair and the decay of the  $W$  boson produced by the Higgs boson, respectively.

There are several other processes which produce similar final states to those seen in the  $t\bar{t}H$  multileptonic decay channel, which makes this search channel particularly challenging. These processes are collectively referred to as “background” for the  $t\bar{t}H$  signal. The most important of these background processes are  $t\bar{t}$  production with subsequent decays into additional jets and non-prompt leptons, and associated production of  $t\bar{t}$  with a vector boson ( $W$  or  $Z$ ). The dominant background contribution for the  $2LSS + 0\tau$  final state arises from the *lepton + jets* decay mode (Case II in Section 2.2) of  $t\bar{t}$  production, wherein a second non-prompt lepton results from the

decay of the b-jets. Hence, in both  $t\bar{t}H$  and  $t\bar{t}$  processes, two same sign leptons are



(a)  $t\bar{t}H$  semileptonic decay with two same-sign prompt light leptons

(b)  $t\bar{t}H$  semileptonic decay with one prompt and one non-prompt lepton

Figure 2-6:  $2LSS + 0\tau$  final state signatures for  $t\bar{t}$  and  $t\bar{t}H$  decay

being produced in the final state.  $t\bar{t}H$  has two prompt, high momentum leptons, one from the semileptonic decay of one of the top quarks and one from the semileptonic decay of  $H \rightarrow WW^*$ ;  $t\bar{t}$  has one prompt lepton from the  $W$  boson and one non-prompt lepton from the b-jets.

## CHAPTER 3 Experimental Setup

### 3.1 The Large Hadron Collider

The Large Hadron Collider is the world's largest and most powerful particle accelerator, colliding protons or heavy ions and is located at the Swiss-French border at the European Centre for Nuclear Research (CERN). It comprises a 27 km circumference tunnel along with an elaborate accelerator complex as shown in figure 3-1. The LHC has four major ongoing experiments situated 100 m underground, where the two counter-rotating beams intersect at nearly the speed of light: ATLAS and CMS are multipurpose detectors used to test the validity of the SM by performing precision measurements and searches for new physics by studying  $p\bar{p}$  collisions; ALICE is mainly used for studying quark-gluon plasma through heavy ion collisions, e.g lead collisions; LHCb is devoted to the study of the bottom quark to understand the matter-antimatter asymmetry in nature.

Protons isolated from hydrogen gas are accelerated to an energy of 50 MeV in the linear accelerator, LINAC 2. They are brought to an energy of 1.4 GeV in the Proton Synchrotron Booster before being transferred to the Proton Synchrotron, which further accelerates the protons to an energy of 25 GeV. The protons reach an energy of 450 GeV in the Super Proton Synchrotron (SPS) and are then injected into the two beam pipes of the LHC. Superconducting quadrupole magnets at a temperature of 1.9 K are arrayed in the tunnel to focus the beams before the collisions

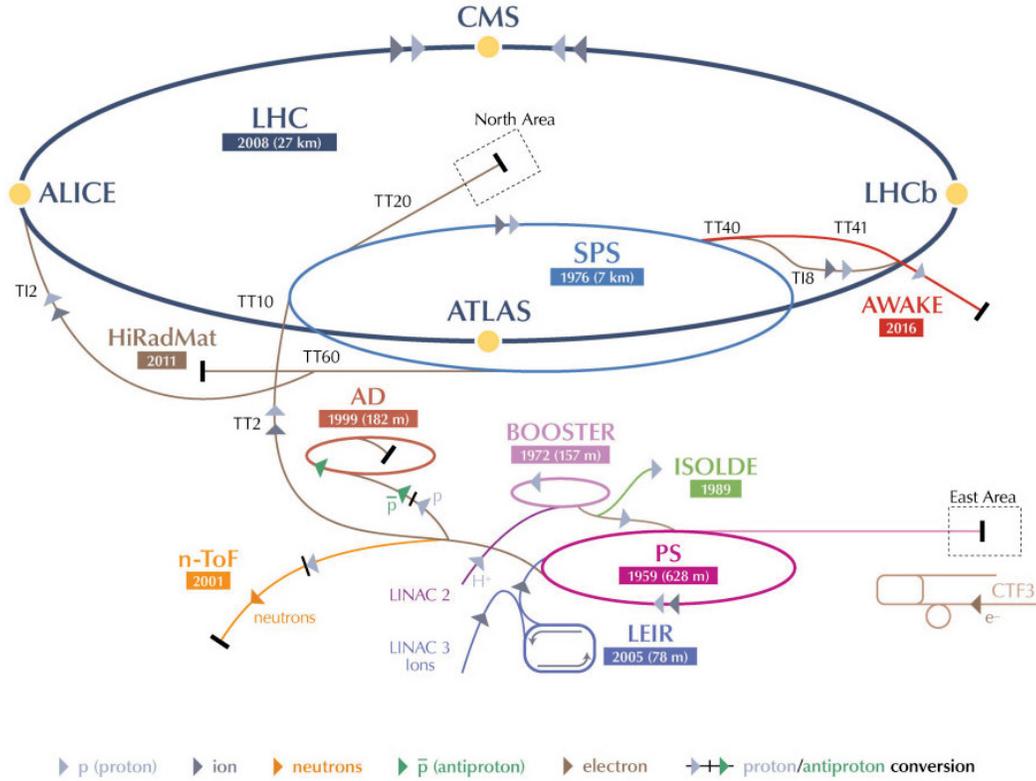


Figure 3–1: CERN Accelerator Complex [21]

at the four interaction points. 1232 superconducting dipole magnets (also at 1.9 K) with a maximum magnetic field strength of 8.33 T are used to keep the proton bunches in a circular path [22].

The designed center of mass energy for proton-proton collisions is  $\sqrt{s} = 14$  TeV, where  $\sqrt{s}$  refers to the invariant Mandelstam variable ( $\sqrt{s} = E_{CoM} = E_{beam_1} + E_{beam_2} = 2 \cdot 7 \text{ TeV} = 14 \text{ TeV}$ ). The roadmap for increasing the energy and luminosity was divided into two sections: Run 1 and Run 2. Run 1 started in 2009 and lasted till

2012 operating at center of mass energy of 7 TeV and later stepped up to 8 TeV. Run 2 commenced in 2015 with center of mass energy,  $\sqrt{s} = 13$  TeV.

Besides the center of mass energy of the particles, instantaneous luminosity,  $L'$ , is an important parameter used to quantify the collisions. It is a measurement of the number of collisions that can be produced per  $\text{cm}^2$  and per second. Considering different number of protons per bunches, and x and y components for the effective geometric cross section of the bunch ( $\sigma$ ), the instantaneous luminosity can be defined by equation 3.1 and has a unit of  $\text{cm}^{-2} \text{s}^{-1}$ .

$$L' = \frac{f N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (3.1)$$

Here,  $N_1$  and  $N_2$  refer to the number of protons in each bunch and  $f$  is the frequency at which the bunches collide (for Run 2,  $f = 40 \times 10^6 \text{ s}^{-1}$ ). Integrated luminosity ( $L$ ) is defined as the integral of the instantaneous luminosity ( $L'$ ) over the time period of operation.

$$L = \int L' dt \quad (3.2)$$

In particle physics, the unit of barn is often used to express the cross-section and is given by, 1 barn (b) =  $10^{-28} \text{ m}^2$ . Integrated luminosity uses the unit of inverse of a barn, which follows from the fact that the total number of events is given by the integral of  $L' \times \sigma$  over time, and hence  $L$  has the unit of inverse cross-section. Run 1 delivered a total integrated luminosity of  $23.3 \text{ fb}^{-1}$ , wherein the ATLAS detector recorded  $21.7 \text{ fb}^{-1}$  [23]. The LHC design instantaneous luminosity is  $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  which has been surpassed in June 2016 and the current luminosity is  $13.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ . The total delivered integrated luminosity to date is  $38.9 \text{ fb}^{-1}$  in

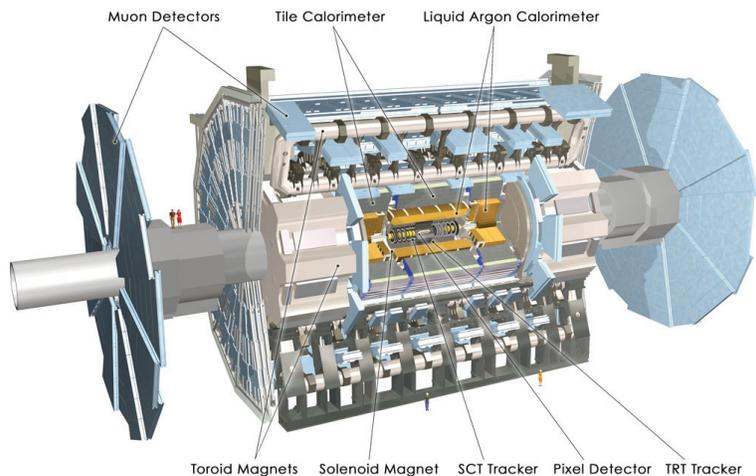


Figure 3–2: The ATLAS Detector

Run 2 collected during 2015 and 2016, and the total luminosity recorded by ATLAS during stable beams at  $\sqrt{s} = 13$  TeV in 2016 is  $36.0 \text{ fb}^{-1}$  before the Extended Year End Technical Stop [24].

## 3.2 The ATLAS Detector

The ATLAS detector is one of the general-purpose detectors at CERN. The detector has a cylindrical geometry, 46 m long, 25 m in diameter and weighs about 7,000 tonnes. It is a multi-layered detector representing a shell-like structure, consisting of an inner detector, electromagnetic and hadronic calorimeters and the muon spectrometer, along with solenoid and toroid magnets.

### 3.2.1 ATLAS Geometry

The detector is forward-backward symmetric about the interaction point of the collision. The coordinate system of the detector is right-handed with the origin at the interaction point. The z-axis is along the beam line and counter-clockwise along the LHC ring as seen from above. The detector is divided into two halves depending

on the positivity of the  $z$ -axis and is referred as the A side for  $z > 0$  and C side for  $z < 0$ . The positive  $x$ -axis is directed towards the inside of the LHC ring and the positive  $y$ -axis is defined as pointing upwards towards the earth's surface.

The precise longitudinal momentum of each of the constituents of the colliding protons (quarks and gluons) is unknown along the  $z$ -axis, whereas the transverse momentum,  $\vec{p}_T$ , is 0 in the initial state. Hence,  $\vec{p}_T$  is a useful quantity to describe the momentum of final state particles and is also used in calculating important variables, such as the missing transverse energy in the detector. Equation 3.3 below gives an expression for the magnitude of  $\vec{p}_T$ .

$$|\vec{p}_T| = \sqrt{\vec{p}_x^2 + \vec{p}_y^2} \quad (3.3)$$

where  $\vec{p}_x$  and  $\vec{p}_y$  are the momenta of the particle in the  $x$  and  $y$ -direction respectively.

Position of particles is well-defined by  $(r, \phi, \theta)$  coordinates and is preferred over the  $(x, y, z)$  coordinates due to the cylindrical symmetry of the detector.  $\phi$  is the azimuthal angle measured around the beam axis and from the positive  $x$ -axis and is in the range of  $[-\pi, \pi]$ .  $\theta$  is the polar angle from the beam axis and ranges from  $[0, \pi]$ . In hadron collider experiments, the concept of rapidity ( $y$ ) is introduced and it tends to the value of pseudorapidity ( $\eta$ ) in the vanishing particle mass limit for large momentum particles ( $m \ll p$ ).

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) \quad (3.4)$$

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (3.5)$$

In the above equations,  $E$  is the energy of the particle,  $p_z$  is the momentum in the  $z$ -direction and  $\theta$  is the polar angle. The difference in rapidity,  $\Delta y$ , is invariant under a Lorentz boost along the beam-axis in the lab frame, unlike the polar angle  $\theta$  in the  $y$ - $z$  plane, which is convenient in the hadron collider environment since this boost is unknown on an event-by-event basis. The  $|\eta|$  coverage of ATLAS detector ranges from 0 to 4.9.  $\eta = 0$  is in the  $y$ -axis direction and is referred as the central part of the detector, and  $|\eta|$  increases towards the forward or backward part of the detector. The angular distance ( $\Delta R$ ) between particles in the  $\eta$ - $\phi$  plane is defined as :

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (3.6)$$

### 3.2.2 Inner Detector

The inner detector (ID) is the innermost portion of the ATLAS detector and is used for pattern recognition and measurements of charged particle momenta. The  $\eta$  coverage of the ID is  $|\eta| < 2.5$  and it has complete  $\phi$  coverage. The ID is surrounded by a superconducting solenoid of 2 T magnetic field for bending charged particle trajectories for momentum determination in the transverse plane. The ID is comprised of four subsequent layers of detectors. The Insertable B-layer (IBL) is the innermost layer and is made up of an array of  $80 \times 336$  array of 26880 cells of silicon pixels [25]. The IBL was installed in 2014 after Run 1 and has been designed to improve flavor jet tagging. It covers the region of  $|\eta| < 3.0$ . The IBL is surrounded by the silicon (Si) pixel detector which is responsible for reliable tracking of the charged particles and robust reconstruction of the primary and secondary vertices and has an  $\eta$  range

of  $|\eta| < 2.5$ . Both the IBL and Si pixel detector generate space-point hits of particles which provides 3-D information of the particles hits in the detector.

The following layer, known as the SemiConductor Tracker, SCT, is composed of silicon microstrips and provides 2-D information about the hits of the particles. The SCT, along with the IBL and Si Pixel detector, produce highly efficient measurements of charged particle momenta and provides precise information for reconstructing the vertices. The outermost layer of the ID is the Transition Radiation Tracker, TRT, and is composed of approximately 300000 thin-walled drift tubes (straws) filled with a gas mixture of Xe, CO<sub>2</sub>, O<sub>2</sub> and with an  $\eta$  range of  $|\eta| < 2$  and  $\vec{p}_T > 0.5$  GeV. It plays a fundamental role in electron identification by detecting the energy of transition radiation photons in the detector which is around 7-10 keV as compared to energy depositions of other particles which are around the range of 2 keV [26].

### 3.2.3 Calorimetry

Calorimetry refers to the measurement of the total deposited energy of the incident particles which aids in particle identification and inferring their properties. The particles create a cascade of associated secondary particles with progressively decreasing energies which is termed as a shower. The longitudinal and transverse profile of the showers have to be known in order to contain them within the active material of the calorimeter. Showers can be broadly classified into two main categories: electromagnetic (EM) showers and hadronic showers. The electromagnetic and hadronic calorimeters (ECAL, HCAL) of the ATLAS detector are designed to absorb the showers completely.

EM showers arise when electrons and photons above the energy of  $\sim 10$  MeV produce secondary photons and electrons/positrons via Bremsstrahlung and pair production, respectively [12]. An important characteristic of the EM showers is its radiation length ( $X_0$ ). Radiation length is defined as the mean distance traversed by the electron before losing  $1/e$  of its energy via Bremsstrahlung. In the case of photons, the incident photon energy would be reduced to  $1/e$  of its value, mainly by pair production, after crossing a mean distance ( $x$ ),  $x = \frac{9}{7}X_0$ , which is related to the probability of the photon interacting with the detector material. The longitudinal spread of the shower ( $d$ ) is determined by equation 3.7 where  $t$  is the thickness of the calorimeter.

$$d = tX_0 \quad (3.7)$$

The estimated thickness to contain 95% of the shower is shown in equation 3.8. Here,  $Z$  is the atomic number of the material,  $E_0$  is the initial energy of the incident particle,  $E_c$  is the critical energy at which energy losses by Bremsstrahlung is equal to the loss in energy by ionisation.  $C_i$  takes the value of  $-0.5$  for electrons and  $+0.5$  for photons.

$$t_{95\%} = t_{max} + 0.08Z + 9.6 ; \quad t_{max} = \ln \frac{E_c}{E_0} + C_i \quad (3.8)$$

The transverse EM shower profile is described by Moliere's radius given by equation 3.9 and 99% of the shower energy is absorbed within  $3.5R_M$ :

$$R_M = \frac{21 \text{ MeV}}{E_c} X_0 \quad (3.9)$$

Hadronization is a process through which quarks form hadrons, which is followed by hadronic showers when mesons or baryons interact in the detector material. The

longitudinal spread of the shower is described in terms of the average interaction length,  $\lambda_I$ , which is generally much larger than the radiation length for EM showers as shown in equation 3.10 [28]:

$$\lambda_I \approx 35\text{g/cm}^2 A^{1/3} \quad (3.10)$$

The lateral spread of the shower is also much wider than EM showers and are governed by momentum transfers within nuclear interactions as shown in figure 3.3. Hence, hadronic calorimeters are of a higher density and larger size to fully contain the showers.

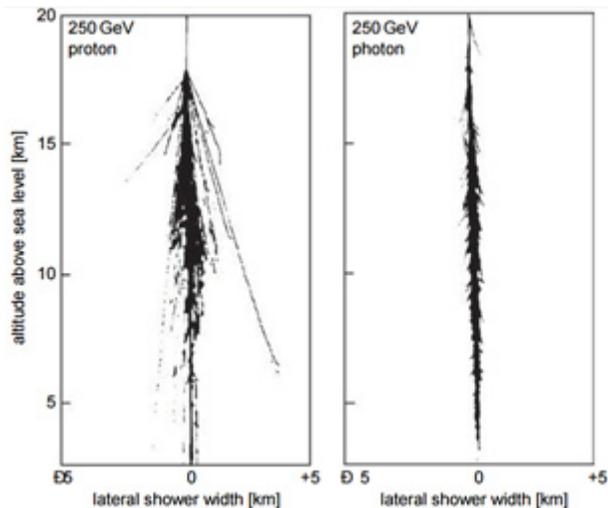


Figure 3–3: Monte Carlo simulations of the distinct development of hadronic and EM showers in the Earth’s atmosphere, induced by 250 GeV protons and photons [28]

The ATLAS calorimeter is comprised of the Electromagnetic Calorimeter (ECAL) followed by the Hadronic Calorimeter (HCAL) with a total coverage of  $|\eta| < 4.9$ .

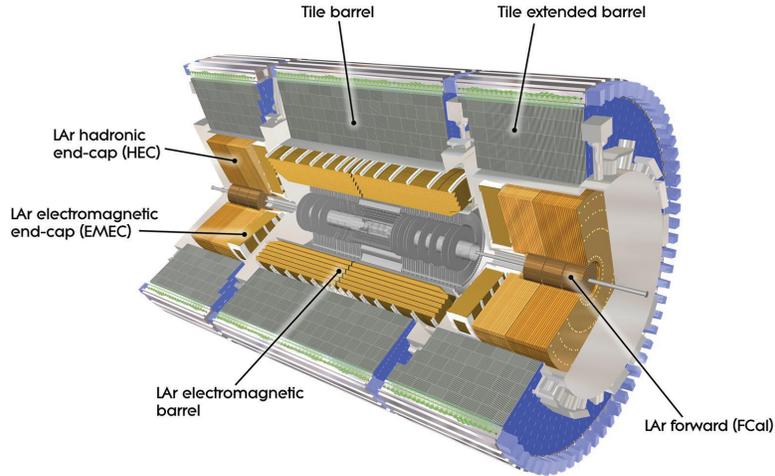


Figure 3–4: Cross-section of the ATLAS Calorimeter

The energy of the shower is efficiently measured by the readout electronics in the calorimeter.

### Electromagnetic Calorimeter

The electromagnetic calorimeter is divided into a three-barrel section ( $|\eta| < 1.475$ ) and two end caps ( $1.375 < |\eta| < 3.2$ ). The barrel is about 6.8 m in length, with an inner radius of 1.15 m and outer radius of 2.25 m and is housed in a cryostat maintained at 88 K. It has an accordion-shaped structure for complete  $\phi$  coverage and uses liquid Argon as the active medium and lead as the absorber material [27]. Liquid argon is chosen due to its high tolerance to radiation and high ionization yield and stability. Kapton electrodes are placed within the gaps of the absorber to measure the output ionization current which is proportional to the incident energy of the particle. The purpose of the ECAL is to accurately measure the energies

deposited by electrons and photons and provide an excellent angular resolution of the detected particles.

### **Hadronic Calorimeter**

The Hadronic calorimeter is designed to measure the energy of hadronic showers which need to be fully absorbed before reaching the Muon Spectrometer. The calorimeter is further divided into three subsystems: the Hadronic Tile Calorimeter, the Hadronic End-cap Calorimeter and the Forward Calorimeter. The barrel region and end-caps have an  $\eta$  coverage of  $|\eta| < 3.2$  and the forward calorimeter covers  $3.1 < |\eta| < 4.9$ . The Hadronic Tile Calorimeter uses steel plates as absorber and plastic scintillators as active material and is further divided into two sections; one central 5.8 m long barrel and two laterally extended 2.6 m long barrel [29]. Each section is segmented into 64 modules and has a total coverage of  $|\eta| < 1.7$ .

The Hadronic End-cap Calorimeter (HEC) consists of copper plates as absorbers and liquid Argon as active material. It is symmetrically placed behind the ECAL end-caps. The HEC covers the  $1.5 < |\eta| < 3.2$  region which overlaps with the Tile calorimeter and the Forward Hadronic calorimeter regions. The Hadronic Forward Calorimeter (FCAL) is placed inside the HEC and covers the high  $|\eta|$  forward regions of the detector within the range of  $3.1 < |\eta| < 4.9$ . FCAL is sub-divided into three distinct longitudinal layers, namely, FCAL1, FCAL2 and FCAL3. All three modules use liquid Argon as the active medium due to high radiation doses in the forward region. FCAL1 is made of copper absorbers and is used for electromagnetic measurements, FCAL2 and FCAL3 consists of tungsten as the absorber and is used to contain and measure hadronic showers.

### 3.2.4 Muon Spectrometer

Muons generally pass through the ID and the calorimeters without losing much energy as their high mass suppresses the Bremmstahlung cross-section and they do not initiate EM showers in the ECAL unlike electrons [28]. Hence, the final layer of the detector, the Muon Spectrometer (MS), is dedicated to tracking and high resolution momentum measurement of muons. Detection of muons in the MS depends on the ionization loss ( $dE/dx$ ) of the particle. The spectrometer consists of four principal sections, namely, Monitored Drift Tube chambers and Cathode Strip Chambers for mapping the trajectory of muons and Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) which act as the muon trigger and provide timing calibration to the event. The Muon Spectrometer is surrounded by eight superconducting toroidal magnets cooled in liquid Helium at 4.5 K which are responsible for bending the trajectories of the muons for momentum measurement. The magnet system consists of two end cap toroids and barrel toroid providing a peak magnetic field of 4.1 T and 3.9 T respectively [30]. The  $|\eta|$  coverage of the MS ranges from  $|\eta| < 2.7$  and the total number of readout channels is  $\sim 1$  million.

### 3.2.5 Trigger and Data Acquisition

Due to very high collision rates and a bunch spacing of only 25 ns, it is nearly impossible to record all the collision events in the LHC. The role of the ATLAS Trigger and Data Acquisition system (TDAQ) is to efficiently record events in real time which have interesting physics signatures and store them for further analysis. ATLAS has a multi-level trigger system consisting of the Level 1 Trigger and the High Level Trigger (HLT).

The Level 1 Trigger is a hardware-based trigger which identifies Region of Interests (ROIs), defined as regions of the detector in  $\eta$  and  $\phi$  which have interesting physics signatures, in the calorimeter and the muon spectrometer. The Level 1 trigger in the calorimeter (L1Calo) mostly identifies electromagnetic clusters (electrons and photons) and jets. The Level 1 muon trigger uses information from the RPCs and TGCs to identify interesting events with high  $\vec{p}_T$  muon signatures. Overall, the L1 triggers reduce the event rate of the collisions from 40 MHz to 100 kHz and has a maximum latency of  $2.5 \mu\text{s}$ . HLT is a software-based trigger and is the combination of the Level 2 Trigger and Event Filter (EF). It reduces the event rate further from 100 kHz to 1 kHz and has a latency of 0.2 s [31]. The Data Acquisition system facilitates the storage of the detector events in huge computing farms for detailed physics analysis.

## CHAPTER 4 Monte Carlo Generators and Object Definition

### 4.1 Monte Carlo Generators used

Monte Carlo (MC) simulation in high energy physics is essential to understand collisions events generated in the LHC in order to test SM predictions and theories beyond the SM, by simulating production of final-state particles, and calibrating the detector for better performance. Simulations of events comprises two steps: event generation and detector simulation. The ATLAS detector uses a C++ language-based control framework known as ATHENA, wherein, the entire offline data processing and physics analysis takes place. ATHENA provides an efficient method to include external MC generators along with their specific interfaces, which are run from within the framework [37]. Additionally, GEANT4 is used for detector simulation to provide precise detector effects and measurements during the collisions [43].

#### 4.1.1 Event Generation

PYTHIA [32], HERWIG [33] and SHERPA [34] are the most widely used event generators within ATLAS. Other dedicated generators, such as MC@NLO [35] and POWHEG [36], are used to cover a broad range of production of events. These generators are fed with initial-state composition and substructure of the colliding hadrons in terms of partons, which are point-like constituents of hadrons, and are matched to gluons and quarks. The momentum distribution of the partons in the colliding protons is termed as parton distribution function (pdf) [38]. The parton

distribution function,  $f_i(x, \mu^2)$  gives the probability of finding a parton of flavour  $i$ , carrying a fraction  $x$  of the proton momentum, with  $\mu$  being the energy scale of the hard interaction.

The following chain of events summarizes a typical event generation process.

**Hard Processes:** Parton collisions belong to the QCD regime are broadly classified into hard and soft processes. Hard processes refer to high momentum transfer during collisions. Some fraction of these collisions produce interesting physics objects, for example, the Higgs Boson or high  $\vec{p}_T$  jets. Event rates and cross sections of hard processes are calculated using perturbative QCD techniques, whereas, soft processes are calculated by non-perturbative QCD. Specific interfaces are used with the above generators to allow for tuning of the MC data to fit the variations of existing physics models and to study specific SM processes such as top physics, and electroweak physics. Adjustable input parameters such as flavors, mass,  $\vec{p}_T$ , spin, production vertex and lifetime of particles are taken into consideration during the production.

**Parton Shower:** Strongly interacting partons often radiate initial (Initial State Radiation- ISR) or final state gluons (Final State Radiation- FSR), which further form quark combinations or radiate additional gluons. These perturbative processes are collectively termed parton showers and are simulated by specialized algorithms [40].

**Hadronization:** As the parton shower continues, the momenta of the daughter particles decrease until the limit during which the perturbative QCD processes are overcome by non-perturbative effects. Due to color confinement, it is not possible

to observe naked quarks (except for the top quark which has a shorter lifetime than the QCD energy scale). Hadronization refers to the process by which quarks bind to form observable hadrons [41]. Hadronization of a single quark in the final state results in a cone of particles, which is known as a jet. In particle detectors, jets are observed rather than individual quarks, whose existence is inferred from the resulting jet.

**Decay:** Hadrons produced during hadronization of quarks are not always stable and further decay into observable particles in the detector. Appropriate models are needed to simulate the decay of short-lived hadrons and lead to the production of final state particles.

#### 4.1.2 Detector Simulation

Event generation is followed by the simulation of the ATLAS detector by taking into consideration the geometry of the detector and the physical processes involving particle interactions and passage of particles through matter. GEANT4 [43] is an analysis toolkit used for simulating the detector setup and the resulting interactions as particles travel through the different layers of material in the detector. The particle trajectory is simulated in steps through the ID, ECAL, HCAL and the muon spectrometer. Additionally, energy losses, directional changes due to scattering, and other processes such as Bremsstrahlung and Cherenkov radiation by the particle are intrinsically modeled within GEANT4. The branching fractions of particles and distribution of secondary particles formed during decays are included in the simulation during the initialization of the process. The detector response and storage of tracks

is user-implemented. Specialized algorithms within ATLAS record and analyse the energy deposits of particles in the detector media.

The MC samples used in the present analysis are produced by the following specific configuration:

- Powheg v2.0 interfaced with Pythia8 is used for generating  $t\bar{t}$  events which have leptons, b-jets and a combination of leptons and jets in their final states (*non-allhadronic*). NLO calculations are implemented in the sample as well.
- Pythia8 interfaced with Madgraph\_aMC@NLO is used to generate the  $t\bar{t}H$  samples along with the detailed NLO calculations [42]. The  $t\bar{t}H$  MC sample is subdivided into three categories depending on the  $t\bar{t}$  decay modes. *Allhadronic* sample refers to the top quark pair decaying only into jets in the final state, *semileptonic* sample includes a combination of lepton and jets and *dileptonic* sample has two leptons being produced from the top quarks.

Table 4.1 shows detailed information about the MC samples used in the analysis.

Table 4-1: Monte Carlo samples

MC Sample	Event Generator	Total number of events
$t\bar{t}$	Powheg v2.0+Pythia6	18818299
$t\bar{t}H$ Allhadronic	Pythia8+Madgraph_aMC@NLO	906454
$t\bar{t}H$ Semileptonic	Pythia8+Madgraph_aMC@NLO	2474771
$t\bar{t}H$ Dileptonic	Pythia8+Madgraph_aMC@NLO	3815609

## 4.2 Object Definition and Selection

In experimental high energy physics, particles formed after the collisions are often not directly observed in the detector. Their presence has to be inferred by observing the tracks or measuring the energy depositions through particle interaction in

various layers of the detector, also termed as detector signatures. These detector signatures are classified into various classes of physics objects, representative of different possible decay topologies, and specific selection criteria (pre-selection) are applied to select the objects as input for the analysis. Along with object pre-selection, the primary vertex in an event is identified by summing over the transverse momentum,  $\sum \vec{p}_T^2$ , of all the associated tracks and selecting the vertex with the highest value of  $\sum \vec{p}_T^2$ , among the other reconstructed vertices.

#### 4.2.1 Light Leptons

Electron and muon candidates are considered to be light leptons and leave distinct detector signatures.

##### Electrons

Electron candidates are expected to leave tracks in the ID and energy deposits in the form of electromagnetic showers in the ECAL (refer section 3.2.3). The  $\eta$  range of the energy cluster in the ECAL is required to satisfy  $|\eta| < 2.47$ . Clusters in the crack region ( $1.37 < \eta < 2.52$ ) between the barrel and the end-caps of the ECAL are excluded. The candidates are reconstructed by associating information from charged particle tracks in the ID originating from a primary vertex, to the ECAL energy deposition. Following the calibration of the electron energy [44], the four-momentum of the electron candidates is computed and the reconstructed electron candidates is required to satisfy  $\vec{p}_T > 10$  GeV in this analysis.

Along with the  $\eta$  and  $\vec{p}_T$  cuts, the transverse and longitudinal impact parameters are two useful quantities used in lepton definition, in order to reduce contributions from the non-prompt leptons and insure that the candidates originated from the

primary vertex. The transverse impact parameter  $d_0$  is defined as the distance of closest approach in the  $R - \phi$  plane of the particle track to the primary vertex. The longitudinal impact parameter  $z_0$  is defined as the value of the z component of the point on the track that determines  $d_0$ . These quantities are graphically explained in figure 4-2. Object pre-selection for light lepton candidates includes a cut on the transverse impact parameter significance, which is defined as the transverse impact parameter divided by the estimated uncertainty on its measurement ( $\sigma_{d_0}$ ).

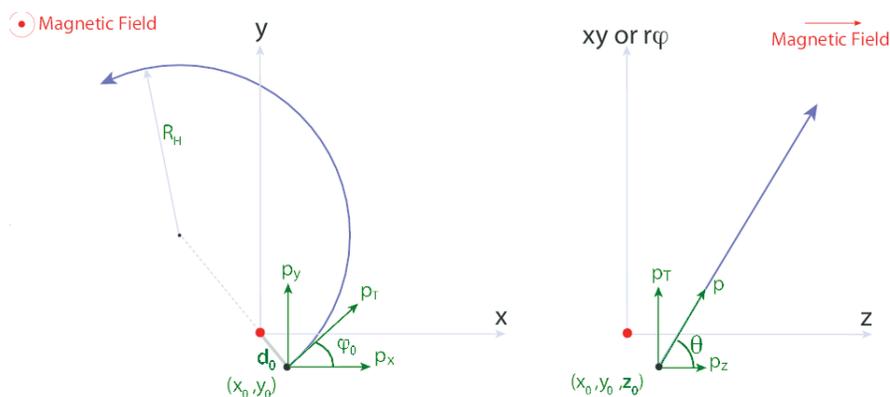


Figure 4–1: From left- Transverse impact parameter ( $d_0$ ) showing the particle track (in blue), the azimuthal angle,  $\phi_0$ , transverse momentum,  $\vec{p}_T$  and radius of the track,  $R_H$ . On the right, longitudinal impact parameter ( $z_0$ ), polar angle,  $\theta$  and  $\vec{p}_T$  is shown. The primary vertex of interaction is shown as the red dot in both the figures [45].

Requirements are imposed on the transverse impact parameter significance,  $|d_0|/\sigma_{d_0}$ , and the longitudinal impact parameter,  $|z_0 \sin \theta_\ell|$ , with respect to the selected primary vertex for electrons, as shown in table 4-1. Candidates are also required to satisfy specific electron definitions, defined as “loose” and “tight” working points. The loose working point does not require isolation cuts for the electron candidates. Isolation of an electron is defined in terms of a  $\vec{p}_T$  cone with a radius

of  $\Delta R$  around the vertex it originated from. The tight definition for an electron candidate requires additional isolation cuts, such that it should be the only object lying within the cone of  $\Delta R < 0.2$ . In the case that the two candidates lie within a cone of  $\Delta R < 0.1$ , the electron with higher  $\vec{p}_T$  is selected. No isolation cut is required at the object preselection level.

### Muons

Muon candidates are either reconstructed by extrapolating tracks in the ID and matching them to the tracks in the MS or by considering full tracks in the MS. Candidates are required to satisfy  $\vec{p}_T > 10$  GeV and  $|\eta| < 2.5$ . In order to reject muons originating from cosmic rays, the transverse impact parameter significance requirement for muon candidates is slightly tighter as compared to electrons, while the longitudinal impact parameter selection is the same. Similar to the electron definition, the isolation cut for muons is not imposed for the object preselection stage.

Table 4–2: Tight and loose light lepton definitions [49].

	Loose		Tight	
	$e$	$\mu$	$e$	$\mu$
Isolation	-	-	Yes	Yes
Identification	Loose	Loose	Tight	Loose
Transverse impact parameter significance $ d_0 /\sigma_{d_0}$	$< 5$	$< 3$	$< 5$	$< 3$
Longitudinal impact parameter $ z_0 \sin \theta_l $	$< 0.5$ mm			

### 4.2.2 Jets

Jets produced from hadronisation of quarks, provide a crucial link between the observed colourless mesons or baryons and the underlying physics at the partonic level. Jets are reconstructed using specific jet reconstruction algorithms that consider the amount and distribution of energy deposited in the calorimeters. Jets are resolved and reconstructed from energy deposits in the HCAL, using anti- $k_t$  jet clustering algorithms [48] with a cone of radius,  $R = 0.4$  around the originating vertex (figure 4-3). Only jets satisfying  $\vec{p}_T > 25$  GeV and  $|\eta| < 2.5$  are used in this analysis.

### 4.2.3 B-jets

Jets containing b-hadrons are identified (“b-tagged”) using specific algorithms which consider ID tracks as input and jet energy deposits in the calorimeters [46]. Due to the long lifetime of b-hadrons, a typical b-hadron topology is characterized by at least one vertex (called a secondary vertex) which is displaced from the point where the hard-scatter collision (primary vertex) takes place as shown in figure 4-2. The b-tagging algorithm used in this analysis has an average efficiency of 70% for identifying b-jets. B-tagged jets are required to satisfy  $\vec{p}_T > 20$  GeV and  $|\eta| < 2.5$ .

### 4.2.4 Missing Transverse Energy

Hard scatter interactions, leading to interesting physics signatures, tend to produce large amounts of transverse energy,  $E_T$ , in the detector, although the  $E_T$  sum should be zero due to conservation of momentum, since the initial state has  $E_T = 0$ . Since the signal modes analyzed in this analysis contain neutrinos, some of this energy is not reconstructed in the detector leading to a non-zero missing transverse energy,  $\vec{E}_T^{miss}$ . The missing energy can only be constrained in the transverse

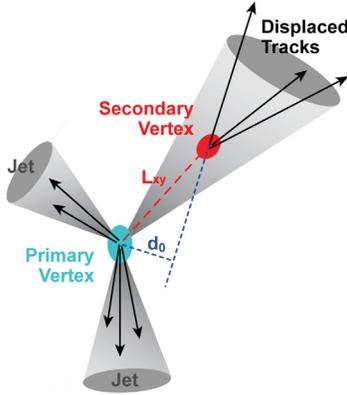


Figure 4-2: B-jet production and reconstruction showing the primary vertex, the displaced secondary vertex associated with a bottom quark and  $\vec{p}_T$  cones for jets and b-jets. [47]

$x - y$  plane, since the longitudinal momentum is non-zero and unknown due to the unknown state of the incident partons. Hence,  $\vec{E}_T^{miss}$  is defined as the negative of the vector sum of the transverse momenta of all reconstructed objects as shown in equation 4.1, where  $\vec{p}_T(i)$  refers to all the final state particles produced after the collision [12].

$$\vec{E}_T^{miss} = - \sum_i \vec{p}_T(i) \quad (4.1)$$

In the ATLAS detector,  $\vec{E}_T^{miss}$  is calculated from the calorimeter energy deposits which are associated with several identified and reconstructed objects, from the reconstructed momenta of muon tracks, and from an estimation of the energy lost in dead materials in the detector. As mentioned above, accurate measurement of  $\vec{E}_T^{miss}$  is crucial for the kinematic reconstruction of the leptonically decaying top quark in this analysis, as it contains a neutrino which is only identified as missing transverse energy in the detector.

## CHAPTER 5

### Kinematic Likelihood Fitter

#### 5.1 Introduction to Kinematic Fitting

##### 5.1.1 Kinematic Fitting

In a typical collision event, the kinematic variables of the final state observable objects, such as energy or momentum of the daughter particles from an object of known mass or particles from a common vertex, are measured by the detector. Kinematic fitting exploits the idea of applying physics constraints (refer section 5.2.1), to identify and match the observed variables of a measured event in the detector with a specific topology. The input measurements which enter the fit are constrained to satisfy certain kinematic properties of the event to correctly associate the measured objects with the final-state particles. Kinematic fitting involves the maximum likelihood method which allows us to choose the best estimator for the free parameters in any physics model and is further detailed in the following section.

##### 5.1.2 The Maximum Likelihood Method

Given a data set consisting of  $N$  measured quantities,  $\mathbf{x} = (x_1, x_2, \dots, x_N)$ , described by a probability distribution function (p.d.f.),  $f(\mathbf{x}, \boldsymbol{\theta})$ , where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$  is a set of  $n$  unknown parameters, a likelihood function can be defined as  $L(\boldsymbol{\theta})$ . If  $x_i$  are statistically independent, the likelihood can be expressed as the product of

individual p.d.f.s as shown in equation 5.1 [12].

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N f(x_i|\boldsymbol{\theta}) \quad (5.1)$$

The maximum likelihood method provides an estimation of the unknown parameters, where  $\hat{\boldsymbol{\theta}}$  are the values of  $\boldsymbol{\theta}$  for which the likelihood is maximized. In most data analyses, the negative log likelihood, which is the negative of the sum of the log-likelihood values of each observation, is minimized. As the log function is a monotone transformation, it returns the same value for  $\hat{\boldsymbol{\theta}}$ . Additionally, the likelihood values can often be very small and a log transform would result in large values which can be handled more easily during calculations. Taking log values of the likelihood converts a product of p.d.f.'s into a summation, which makes it convenient to differentiate and maximize it as shown in equation 5.2.

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \sum_{i=1}^N \ln f(x_i|\boldsymbol{\theta}) = 0 \quad \forall j = 1, 2, \dots, n \quad (5.2)$$

## 5.2 Kinematic reconstruction using KLFitter in the *semileptonic* decay channel for $t\bar{t}$ production

The Kinematic Likelihood Fitter (KLFitter) is a standalone package [50] that aids in kinematic reconstruction by employing a kinematic fit and using the maximum likelihood method as discussed in the preceding section. In this analysis, KLFitter is used to reconstruct *semileptonic*  $t\bar{t}$  events to discriminate them from the  $t\bar{t}H$  signal in the  $2lSS+0\tau$  final state (figure 5-1). The *semileptonic*  $t\bar{t}$  decay channel, also referred to as the *lepton+jets* decay channel, yields a final state containing four jets, a charged light lepton (either electron or muon) and its corresponding neutrino. Two b-jets,

$b_{had}, b_{lep}$ , originate from the top quarks decaying in a hadronic or leptonic final state, respectively, while the other two quark jets,  $q_1, q_2$ , are produced from the hadronic decay of one of the  $W$  bosons. The lepton and neutrino result from the leptonic decay of the other  $W$  boson. The second same sign non-prompt lepton originates from the further decay of the b-jets, which contributes to a high background in  $t\bar{t}H$  final state consisting of two same sign leptons.

On the other hand, the  $2LSS + 0\tau$  final state for  $t\bar{t}H$  comes from the decay of a Higgs boson into two  $W$  bosons and contains two neutrinos with six jets in total, produced by the two top quarks and one of the  $W$  bosons from the Higgs boson. The presence of two neutrinos means that the leptonic top mass cannot be reconstructed, unlike the  $t\bar{t}$  topology, where the energy of the sole neutrino can be calculated by using  $\vec{E}_T^{miss}$  information. Moreover, one of the  $W$  bosons coming from the decay of the Higgs boson, is off-shell, which limits the availability of kinematic constraints that can be used to identify this topology. KLFitter is originally designed to cleanly identify and efficiently reconstruct the  $t\bar{t}$  topology, as compared to reconstructing  $t\bar{t}H$  events, thus providing discrimination between the background and signal processes.

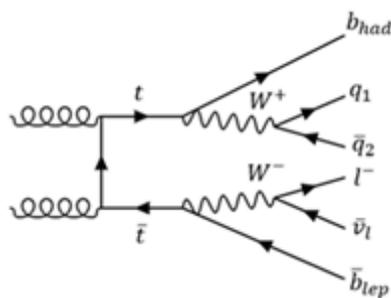


Figure 5–1: *Semileptonic (lepton +jets) decay of a  $t\bar{t}$  event*

Constructing a p.d.f. for kinematic fitting requires a defining set of unknown parameters ( $\theta_i$ ) and measured quantities ( $x_i$ ) as shown in equation 5.1. The measured values of the physics observables are obtained from MC simulation and are listed below:

- The energies  $\tilde{E}_i$  and directions  $\tilde{\Omega}_i = (\tilde{\eta}_i, \tilde{\phi}_i)$  of four or more quark jets.
- The energy  $\tilde{E}_l$  and direction  $\tilde{\Omega}_l = (\tilde{\eta}_l, \tilde{\phi}_l)$  of the lepton.
- The x and y component of the missing transverse energy,  $\tilde{E}_{T_x}^{miss}$ ,  $\tilde{E}_{T_y}^{miss}$ .

The measured physics observables and constraints (section 5.2.1) are taken as input to the KLfitter to fit the unknown parameters by using the maximum likelihood method.

### 5.2.1 Derivation of Constraints

Physics constraints on the final state objects are derived from the energy and direction measurements of the decay objects in the detector. As seen in figure 5-1, the two top quarks decay into two b-jets and  $W$  bosons, which further decay into two quark jets and a lepton and its corresponding neutrino respectively. The invariant masses of the  $W$  bosons and top quarks are calculated from the input four-momentum vectors ( $p$ ) of the final state particles. In the set of equations below,  $m$  refers to the invariant mass of the particle and  $\vec{p}$  is the 3-vector of the particle.

$$p = (E, \vec{p}_x, \vec{p}_y, \vec{p}_z) \quad (5.3)$$

$$p^2 = E^2 - |\vec{p}|^2 = m^2 \quad (5.4)$$

The derivation of each constraint used in the KLfitter likelihood for the *semileptonic*  $t\bar{t}$  decay channel is detailed as below.

### Hadronically Decaying $W$ boson

The invariant mass of the two light quark jets,  $q_1, q_2$ , is calculated from the 4-vectors, to reconstruct the mass of the hadronically decaying  $W$  boson. Here,  $\cos \theta_{q_1 \bar{q}_2}$  is the angle between the momentum vectors of the two quarks,  $\vec{p}_{q_1}$  and  $\vec{p}_{\bar{q}_2}$ . The jet masses,  $m_{q_1}, m_{\bar{q}_2}$  are explicitly set to zero in the jet 4-vector during the calculations. The invariant mass of the two jet combination,  $m_{q_1 \bar{q}_2}$ , is given by:

$$\begin{aligned} m_{q_1 \bar{q}_2}^2 &= (p_{q_1} + p_{\bar{q}_2})^2 \\ &= m_{q_1}^2 + m_{\bar{q}_2}^2 + 2E_{q_1} E_{\bar{q}_2} - 2|\vec{p}_{q_1}| |\vec{p}_{\bar{q}_2}| \cos \theta_{q_1 \bar{q}_2} \\ &\approx 2E_{q_1} E_{\bar{q}_2} (1 - \cos \theta_{q_1 \bar{q}_2}) \end{aligned} \tag{5.5}$$

### Leptonically Decaying $W$ boson

The 4-vectors of the lepton and neutrino are considered as input for constructing the invariant mass of the leptonically decaying  $W$  boson. The measurement of final states containing a neutrino only provides the x and y components of  $\vec{E}_T^{miss}$  (refer section 5.2), which is lacking information about the z-component to form the 4-momentum vector of the neutrino. The z component is constrained by the mass of the leptonic  $W$  boson by solving the relevant quadratic equation providing two possible solutions to the z-component in the 4-vector. Lepton mass is neglected and

the neutrino is treated as massless:

$$\begin{aligned}
m_W^2 &= (p_l + p_\nu)^2 \\
&= 2E_l|\vec{p}_\nu| - 2\vec{p}_l \cdot \vec{p}_\nu \\
m_W^2 + 2\vec{p}_{T_l} \cdot \vec{p}_{T_\nu} &= 2E_l|\vec{p}_\nu| - 2p_{zl}p_{z\nu} \\
\alpha^2 + 4p_{zl}p_{z\nu}\alpha + 4p_{zl}^2p_{z\nu}^2 &= 4E_l^2(\vec{p}_{T_\nu}^2 + \vec{p}_{z\nu}^2) \tag{5.6} \\
\underbrace{(p_{zl}^2 + E_l^2)}_a p_{z\nu}^2 + \underbrace{p_{zl}\alpha}_{b} p_{z\nu} + \underbrace{\frac{\alpha^2}{4} - E_l^2\vec{p}_{T_\nu}^2}_c &= 0 \\
\Rightarrow p_{z\nu}^{1,2} &= \frac{-b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}
\end{aligned}$$

In the above equations,  $\alpha = m_W^2 + 2\vec{p}_{T_l} \cdot \vec{p}_{T_\nu}$ . The more likely value of the two neutrino  $p_z$  solutions is used as a fit parameter in the likelihood. The z-component of the neutrino momentum has to be within a range of  $\pm 1000$  GeV. The invariant mass reconstruction of the  $W$  boson is calculated as per the following set of equations:

$$\begin{aligned}
m_{l\nu}^2 &= (p_l + p_\nu)^2 \\
&= m_l^2 + m_\nu^2 + 2E_lE_\nu - 2|\vec{p}_l||\vec{p}_\nu| \cos \theta_{l\nu} \tag{5.7} \\
&\approx 2E_l\sqrt{p_{x\nu}^2 + p_{y\nu}^2 + p_{z\nu}^2}(1 - \cos \theta_{l\nu})
\end{aligned}$$

### Hadronically Decaying Top quark

The reconstruction of the hadronically decaying top quark takes the energies and directions of the b-jet ( $b_{had}$ ) and two light quark jets as input. The masses of the two light quarks is neglected. The invariant mass of the 3-jet combination,

$m_{q_1\bar{q}_2b}$ , is given by:

$$\begin{aligned}
m_{q_1\bar{q}_2b}^2 &= (p_{q_1}^2 + p_{\bar{q}_2}^2 + p_b^2) \\
&\approx m_b^2 + 2E_{q_1}E_{q_2}(1 - \cos\theta_{q_1\bar{q}_2}) \\
&\quad + 2E_{q_1}E_b(1 - \frac{|\vec{p}_b|}{E_b}\cos\theta_{q_1b}) + 2E_{\bar{q}_2}E_b(1 - \frac{|\vec{p}_b|}{E_b}\cos\theta_{\bar{q}_2b})
\end{aligned} \tag{5.8}$$

### Leptonically Decaying Top quark

The top quark which decays via a leptonically decaying  $W$  is reconstructed by using the input 4-vectors of the b-jet ( $b_{lep}$ ), lepton and the corresponding neutrino. The mass of the lepton and neutrino are neglected. The invariant mass of the top quark candidate,  $m_{l\nu b'}$ , is given by

$$\begin{aligned}
m_{l\nu b'}^2 &= (p_l^2 + p_\nu^2 + p_{b'}^2) \\
&\approx m_{b'}^2 + 2E_lE_\nu(1 - \cos\theta_{l\nu}) \\
&\quad + 2E_lE_{b'}(1 - \frac{|\vec{p}_{b'}|}{E_{b'}}\cos\theta_{lb'}) + 2E_\nu E_{b'}(1 - \frac{|\vec{p}_{b'}|}{E_{b'}}\cos\theta_{\nu b'})
\end{aligned} \tag{5.9}$$

### 5.2.2 Breit-Wigner Distribution

The decay lineshape of the  $W$  boson and the top quark is parametrized using a Breit-Wigner distribution ( $BW$ ). The Breit-Wigner distribution describes a resonance, i.e. an unstable particle in quantum field theory. The peak value of the curve is at the pole mass of the particle at which the particle is produced at its ‘‘on-shell mass’’. The width of the curve is related to the mean lifetime of the resonance using the following equation, where  $t$  is the lifetime of the particle and  $\Gamma$  is the decay width.

$$t = 1/\Gamma \tag{5.10}$$

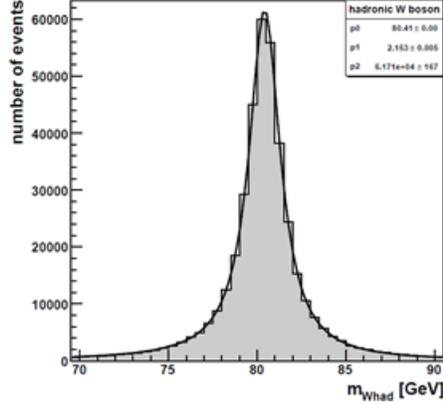


Figure 5–2: Breit-Wigner fitting of hadronic W boson

A particle with pole mass  $M$  and decay width,  $\Gamma$  will produce a distribution of observed masses,  $m$ , given by  $BW(m|M, \Gamma)$  [12]:

$$BW\{m|M, \Gamma\} = \frac{2\Gamma M^2}{\pi(m^2 - M^2)^2 + M^2\Gamma^2} \quad (5.11)$$

The first constraint is derived from the decay width of the hadronically and leptonically decaying  $W$  bosons, where the invariant mass of the two jets ( $m_{q_1\bar{q}_2}$ ) and the invariant mass of the lepton and its neutrino ( $m_{l\nu_l}$ ) are required to be distributed in the  $BW$  curve around the pole mass of the two individual  $W$  bosons, each of mass 80.4 GeV and within a decay width of 2.1 GeV [12]. The second constraint comes from the decay of the two top quarks, where the invariant mass of the three jets (2 quark jets and one b-jet from the hadronically decaying top,  $m_{q_1\bar{q}_2b}$ ) and the invariant mass of the lepton and its neutrino and one b-jet from the leptonically decaying top ( $m_{l\nu_l b'}$ ) should be compatible with the  $BW$  distribution of the top quark pole mass (172.5 GeV) with a decay width of 1.5 GeV [12]. In this analysis, the top mass is fixed at 172.5 GeV. An example of the hadronically decaying  $W$  boson fitted with a

$BW$  function is shown in figure 5.2. The first fitting parameter ( $p_0$ ) corresponds to the pole mass, the second ( $p_1$ ) to the width and the third ( $p_2$ ) to the amplitude.

### 5.2.3 Transfer Functions

The KLfitter likelihood includes the transfer functions ( $TFs$ ), which map the measured response (energy, direction or momentum) of the final state jets and leptons in the detector, to the true values of the associated single particle (quarks and leptons in this case).  $TFs$  are essential in the likelihood as they take the detector resolution effects into account. They are represented as  $W(\tilde{E}|E)$ , where  $\tilde{E}$  is the measured energy and  $E$  is the true value of the energy. The general form of  $TFs$  for particle energy is shown in equation 5.12.

$$W(\tilde{E}, E) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)} \left( e^{-\frac{(\Delta E - p_1)^2}{2p_2^2}} + p_3 e^{-\frac{(\Delta E - p_4)^2}{2p_5^2}} \right) \quad (5.12)$$

Here,  $\Delta E = \frac{E - \tilde{E}}{E}$  and  $p_i$  are parametrized functions of the particle's true energy obtained from MC truth information, as opposed to the particle energy measured in the detector. Moreover, the  $TFs$  are parametrized for different  $\eta$  regions to consider the detector resolution efficiency.

TFTool is a dedicated package which uses  $t\bar{t}$  sample generated by MC@NLO generator to produce the transfer functions for various final state objects [51]. The  $TFs$  currently used in the analysis are derived from 8 TeV MC simulation. The  $\vec{p}_T$  requirement for the jets and the charged lepton is set to  $\vec{p}_T > 7$  GeV. The model objects are matched to the reconstructed objects in the detector and the matching is dependent on the fact that both the model object and reconstructed object lie within an angular separation,  $\Delta R < 0.3$  [51]. The  $TFs$  for the energy of particles are

fitted by double Gaussian functions consisting of a narrow and a wide component, to account for the asymmetric tails in energies, except for  $\vec{E}_T^{miss}$ , which is described by a single Gaussian function. The global  $TF$  fit is defined as the sum of the two individual Gaussians and is shown in figure 5-3.

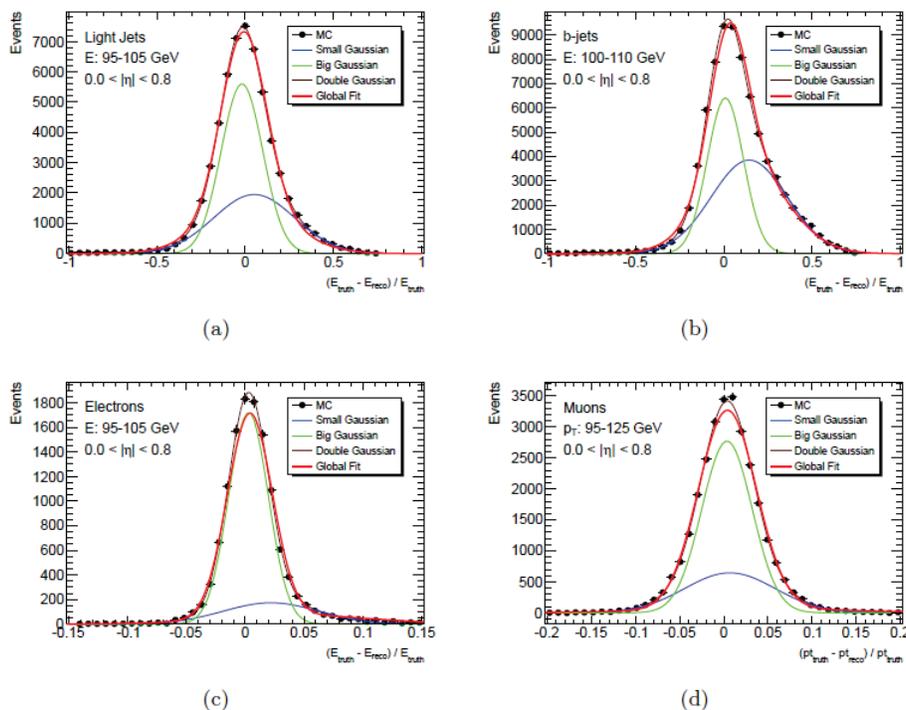


Figure 5-3: Transfer functions for (a) light jets, (b) b-jets, (c) electrons and (d) muons for truth energies / transverse momenta of around 100 GeV and  $0 < \eta < 0.8$  [51].

#### 5.2.4 KLFitter Likelihood

The likelihood function of the KLFitter comprises a product of p.d.f.s of two distinct parts: parametrization of the physics constraints on the fit, in terms of the Breit-Wigner distribution ( $BW$ ) of the reconstructed objects,  $L_{con}$  and measurement resolutions of the kinematic quantities of the decay products in terms of transfer

functions,  $L_{TF}$ . The general form of the likelihood is shown below.

$$L = L_{con} \cdot L_{TF} \quad (5.13)$$

The following fit parameters are used to calculate the likelihood:

- The energies  $E_i$  of the quark jets;
- the energy  $E_l$  of the lepton;
- the momentum of the neutrino  $p'_{x,y,z}$ ;
- top pole mass, which is fixed to 172.5 GeV in this case.

Equation 5.14 below defines the likelihood function explicitly for the *semileptonic*  $t\bar{t}$  decay topology:

$$\begin{aligned}
L = & BW\{m(q_1q_2)|m_W, \Gamma_W\} \cdot BW\{m(l\nu)|m_W, \Gamma_W\} \cdot BW\{m(q_1q_2b_{had})|m_{top}, \Gamma_{top}\} \cdot \\
& BW\{m(l\nu b_{lep})|m_{top}, \Gamma_{top}\} \cdot W(\tilde{E}_{jet1}|E_{bhad}) \cdot W(\tilde{E}_{jet2}|E_{blep}) \cdot \\
& W(\tilde{E}_{jet3}|E_{q1}) \cdot W(\tilde{E}_{jet4}|E_{q2}) \cdot W(\tilde{E}_{T_x}^{miss}|p_{x,\nu}) \cdot W(\tilde{E}_{T_y}^{miss}|p_{y,\nu}) \cdot \\
& \begin{cases} W(\tilde{E}_1|E_l) & \text{single electron channel} \\ W(\tilde{p}_{T,1}|\vec{p}_{T,l}) & \text{single muon channel} \end{cases}
\end{aligned} \quad (5.14)$$

The first four terms in equation 5.14 are the *BW* parametrization of the constraints and the  $W(\tilde{E}|E)$  terms represent the transfer functions. As input to the likelihood, at least 4 jets have to be selected with a maximum limit of 8 jets. Out of these selected jets, some of the jets can be flagged as b-jets by using b-tagging algorithms, thus distinguishing them from the light quark jets. There are a total of  $4! = 24$  permutations for the jets i.e., 24 possible ways in which the 4 observed

jets can be mapped to the 4 model particle labels  $(q_1, q_2, b_{had}, b_{lep})$ . The two (light) quark jets,  $q_1$  and  $q_2$  are interchangeable in the likelihood, since both combinations yield the same  $W$  candidate. This reduces the number of jet permutations to 12 and is computed for each candidate event.

The KLFitter package also provides user- defined options, wherein, the choice of using b-tagging information can be controlled by the user. The most relevant options used in this analysis are listed below:

- **kVetoNoFit**: Permutations in which a b-tagged jet is in the position of a model light quark  $(q_1, q_2)$  are removed before the fitting procedure. The number of b-tags required is equal to or greater than 1.
- **kVetoNoFitBoth**: Both permutations with a light jet associated to a bottom quark or with a b-tagged jet associated to a light quark are removed before the fitting procedure. Two or more b-tagged jets are required to be present in order to use this option within KLFitter.

The negative log of the likelihood is calculated for all the different permutations of the input quark jets and the specific permutation producing the minimum value of the negative log likelihood is selected as output by the fitter.

## CHAPTER 6 KLFitter Performance Studies

### 6.1 KLFitter *lepton + jets* likelihood for $t\bar{t}$ and $t\bar{t}H$ event reconstruction in the $2LSS + 0\tau$ final state

The goal of the analysis is to implement the KLFitter package to  $t\bar{t}$  and  $t\bar{t}H$  topologies to discriminate between the signal and background contributions in the  $2LSS + 0\tau$  final state. Since the KLFitter *lepton + jets* likelihood is applied to the *semileptonic*  $t\bar{t}$  events as mentioned in section 5.2, the  $t\bar{t}$  MC sample would be identified and reconstructed efficiently and would produce a higher likelihood value on average as an output for the fit, as compared to the other topologies. Hence, a worse likelihood value for KLFitter is expected from the reconstruction of the  $t\bar{t}H$  topology as compared to  $t\bar{t}$  events. Thus, the reconstruction efficiency of the fit, which would be reflected in the likelihood values, could potentially be used as a discriminant to distinguish between the signal and background.

As a preliminary test, the KLFitter tool is tested with the *lepton+jets* likelihood as defined in equation 5.14 by applying it to the *non-allhadronic*  $t\bar{t}$  and  $t\bar{t}H$  *hadronic, semileptonic and dileptonic* MC samples. Following the testing of fit,  $2LSS + 0\tau$  selection cuts are applied to the  $t\bar{t}$  and  $t\bar{t}H$  events before they enter KLFitter *lepton + jets* likelihood. The final state objects reconstructed in the detector are required to pass the set of requirements listed in table 6-1. The two light leptons are ranked according to a descending order of  $\vec{p}_T$  and are required to have the same electric

charge. They are additionally required to pass at least one of the triggers depending on the run year, as specific trigger requirements differ between runs. Tight definition for leptons is used in the  $2LSS + 0\tau$  selection. The loose and tight lepton definition requirements are listed in table 4-1, in which the tight definition is applied to the leptons only after it satisfies the loose lepton definition cuts.

The events satisfying  $2LSS + 0\tau$  selection criteria are required to pass additional KLFitter selection cuts to enter the fit. At least four jets and less than a maximum of 8 jets are taken as input in a descending  $\vec{p}_T$  order for the KLFitter. The use of b-tagging information is not required by the fit in the *lepton + jets* KLFitter likelihood, but is required in the modified version of the fit. Since the *lepton + jets* likelihood takes only one lepton as input to the fit, the lepton with the highest  $\vec{p}_T$  is selected out of the two same-sign leptons, to calculate the likelihood. Permutations are performed for the selected jets and the permutation minimizing the negative log likelihood value is selected as the output. A tabular representation of  $\vec{p}_T$  and  $\eta$  cuts for jets taken as input by the fit is presented in table 6-2.

In addition to the output from the combined lepton channel, three combinations of lepton final states are studied in all the presented cases in the analysis: dielectron ( $ee$ ), dimuon ( $\mu\mu$ ), and mixed lepton channel ( $e\mu/\mu e$ ). Histograms and the discriminating power of the fit related to this section have been presented in the following chapter. Additional modifications are made to the original *lepton+jets*  $t\bar{t}$  likelihood, in order to study the potential for improvement of the  $t\bar{t}H$   $2LSS + 0\tau$  analysis. These are described in the following sections.

Table 6–1: Cutflow implementation for the selection of jets and leptons in the  $2LSS + 0\tau$  final state decay for  $t\bar{t}H$  signal region [52].

Selection/Cuts	Implementation
Trigger (2015)	HLT_mu20_iloose_L1MU15, HLT_mu50, HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium, HLT_e120_lhloose, HLT_2e12_lhloose_L12EM10VH, HLT_e17_lhloose_mu14, HLT_mu18_mu8noL1
Trigger (2016)	HLT_mu26_ivarmedium, HLT_mu50, HLT_e26_lhtight_nod0_ivarloose, HLT_e60_lhmedium_nod0, HLT_e140_lhloose_nod0, HLT_2e17_lhvloose_nod0, HLT_e17_lhloose_nod0_mu14, HLT_mu22_mu8noL1
No. of light leptons	2
Same Sign selection	Sum of lepton charges $\pm 2$
Lepton $\vec{p}_T$	$> 10$ GeV for both selected leptons
Lepton Definition	Tight
No. of Jets	$\geq 4$
No. of B-jets	$\geq 1$
No. of $\tau_{had}$	0

Table 6–2: KLFitter selection criteria for  $lepton+jets$  and modified KLFitter likelihood

KLFitter selection cuts	Lepton+jets likelihood	Modified likelihood
Number of jets (nJets)	$\geq 4, < 8$	$\geq 4, < 8$
Jet $\vec{p}_T$ cut	$> 25$ GeV	$> 25$ GeV
Jet $\eta$ cut	$< 2.5$	$< 2.5$
B-tagging	-	kVetoNoFit: $\geq 1$ b-tags, kVetoNoFitBoth: $\geq 2$ btags
Lepton $\vec{p}_T$ cut	$> 10$ GeV	$> 10$ GeV for both leptons

## 6.2 Modified KLFitter Likelihood with an Additional Lepton in the $2LSS + 0\tau$ channel

The KLFitter likelihood function is further extended to include the  $2LSS + 0\tau$  event topology within the  $lepton+jets$  likelihood, such that the  $t\bar{t}H$  signal can be discriminated efficiently against the  $t\bar{t}$  background. The KLFitter  $lepton+jets$

likelihood definition is modified to include an additional non-prompt lepton ( $l_2$ ) from either of the b-jets in the  $t\bar{t}$  decay. Equation 6.1 shows the modified likelihood function where  $L$  refers to the *lepton+jets* likelihood defined in equation 5.14. Here,  $W(\tilde{E}_2|E_{l_2})$  refers to the transfer function of the additional second lepton which has been added to the likelihood. The original high  $\vec{p}_T$  lepton in equation 5.14 is used to reconstruct the  $W$  and top quark masses and is referred to as the “leading lepton”. The second lepton, having a lower  $\vec{p}_T$  compared to the leading lepton, is referred to as the “sub-leading lepton” in the analysis. The modified likelihood is given by:

$$L_{mod} = L \cdot \begin{cases} W(\tilde{E}_2|E_{l_2}) & ee \text{ channel} \\ W(\tilde{\vec{p}}_{T,2}|\vec{p}_{T,l_2}) & \mu\mu \text{ channel} \\ W(\tilde{\vec{p}}_{T,2}|\vec{p}_{T,l_2}) & e\mu/\mu e \text{ channel} \end{cases} \quad (6.1)$$

Further modifications are made within the above likelihood to increase the discriminating power of KLFitter and two cases have been presented in the following sections.

### 6.2.1 Case 1 - Minimizing the angular separation ( $\Delta R$ ) between additional lepton and b-jets

Assuming that the sub-leading lepton is a non-prompt lepton,  $\Delta R$  is calculated between the sub-leading lepton and each of the b-jets. The smaller value of  $\Delta R$ , indicating the closest b-jet to the sub-leading lepton, is entered as a new term in the modified likelihood,  $L_{mod}$ , and is represented in equation 6.2. A weighting factor of 10 has been applied to the  $\Delta R$  component in the likelihood, such that it scales accordingly to the other likelihood components, since this particular term is expected

to act as a discriminating factor between  $t\bar{t}$  and  $t\bar{t}H$  decay processes.

$$L_{\Delta R} = L_{mod} \cdot [\min \Delta R(l_2, b - jet)]^{10} \quad (6.2)$$

The cut implementation for the  $2LSS + 0\tau$  final state and the selection criteria for the two leptons and jets is identical to that which is described in section 6.2, except for the inclusion of the b-tagging information in the fit as shown in table 6-2. Two options for using the b-tagging information in the fit, kVetoNoFit and kVetoNoFitBoth, are tested. Other kinematic variables, such as the reconstructed top mass pair and the  $\vec{p}_T$  distribution of the top quark pair is studied to further analyze the discriminating power of the fit.

### 6.2.2 Case 2- Merging of the closest b-jet and additional lepton

The non-prompt lepton originating from the b-jets in the  $t\bar{t}$  decay may carry away some energy from the b-jets. The angular separation,  $\Delta R$ , is calculated between the sub-leading lepton and the b-jets ( $b_{had}, b_{lep}$ ), similar to Case 1. The smaller value of  $\Delta R$  indicates the closest b-jet to the lepton and indicates that the  $l_2$  most likely originated from that particular b-jet. Hence, the closest b-jet and  $l_2$  are merged by adding their respective Lorentz vectors. The new b-jet momenta is entered as input to the likelihood and the  $2^{nd}$  lepton is omitted in the likelihood, which leads to a *lepton + jets* likelihood as shown in equation 5.14. The object selection criteria and KLFFitter user options are identical to section 6.2.1, along with the application of kVetoNoFit and kVetoNoFitBoth. Other kinematic variables such as the reconstructed top mass pair and the  $\vec{p}_T$  distribution of the reconstructed top quark pair by the fit are also tested. The results of these studies are presented in the following

chapter.

## CHAPTER 7

### Results

The plots presented in this chapter are generated with version 23 nTuples for the MC samples within the AnalysisTop 2.4.27 release and are based on the selection procedures described in sections 6.1, 6.2.1 and 6.2.2, respectively.

The plots in figures 7-1, 7-2 and 7-3 show the negative log likelihood plotted on the x-axis and the number of events with scaled normalization on the y-axis for  $t\bar{t}$  and  $t\bar{t}H$  MC samples in the  $2LSS + 0\sigma$  final state. Four outputs corresponding to *non-allhadronic*  $t\bar{t}$ ,  $t\bar{t}H$ , *semileptonic*  $t\bar{t}H$  and *dileptonic*  $t\bar{t}H$  samples are plotted. Here,  $t\bar{t}H$  represents a merged output including *semileptonic*, *dileptonic* and *all-hadronic*  $t\bar{t}H$  outputs. The histograms show the log likelihood for (a) the combined lepton channel, (b)  $ee$  channel, (c)  $\mu\mu$  channel, and (d) the mixed lepton channel, respectively, in figures 7-1, 7-2 and 7-3. Figure 7-4 shows the reconstructed invariant mass and  $\vec{p}_T$  of the  $t\bar{t}$  system in units of GeV plotted on the x-axis for case 1, along with the arbitrary normalization of the events on the y-axis.

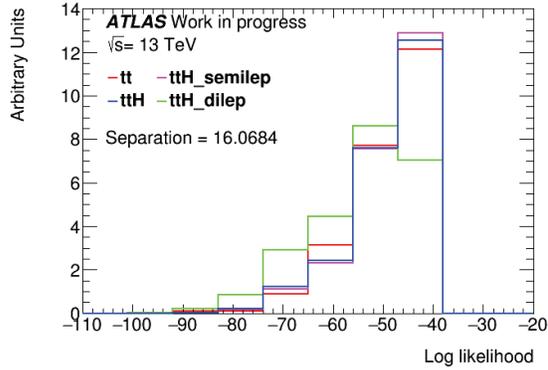
The normalization of the histograms is calculated by dividing the  $t\bar{t}$  histogram by the total number of events in its MC sample for each bin. The area of the  $t\bar{t}$  histogram is calculated with the new values of the entries in the bins and is multiplied by a factor of 1000, 100 and 10 in the cases of the log likelihood, invariant mass and  $\vec{p}_T$  distribution of the  $t\bar{t}$  system, respectively, for convenience. Each of the  $t\bar{t}H$  histograms in the plots are normalized by having an area equal to the calculated

area above. This method would result in each event in the MC sample having an identical weight in terms of the area under the histograms, when calculating the figure of merit. Two b-tagging options, kVetoNoFit and kVetoNoFitBoth, are used in all the cases, except for the plots in figure 7-1 which uses the KL Fitter *lepton + jets* likelihood definition without b-tagging information required by the fit. Along with signal background discrimination, a comparison study is done for the efficiency of the b-tagging options used within the fit.

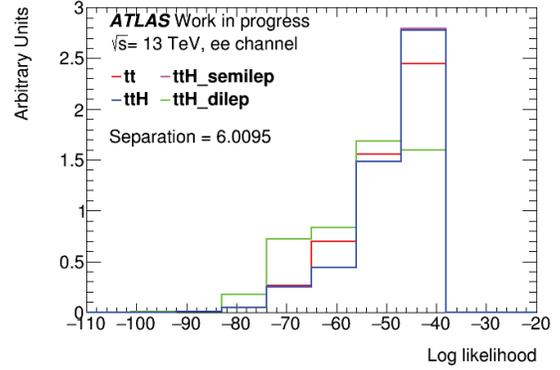
The figure of merit used to analyze the performance of the fit in all the cases is referred as the separation value or discriminating power, which is calculated between the  $t\bar{t}$  sample (in red) and  $t\bar{t}H$  (in blue) by subtracting the area under the two histograms to calculate the total overlapping area. The overlapping area integral is computed as the sum of the bin contents multiplied by the bin width in the given x-axis range shown in the plots in figures 7-1 to 7-4. The values for uncertainty in the separation value arising from the statistics of the MC samples, is computed using error propagation of the two individual histogram uncertainties from the bin errors assuming that all the bins are uncorrelated. The errors are further multiplied by the bin width shown on the x-axis. The bin width varies for the different variables; it corresponds to a value of 9 units on the x-axis for all the log likelihood plots in figures 7-1, 7-2 and 7-3, a value of 45 GeV for the reconstructed invariant mass of the  $t\bar{t}$  system in figures 7-4 (a) and 7-4 (b) and a value of 80 GeV for the  $\vec{p}_T$  distribution plots in figures 7-4 (c) and 7-4 (d). The separation values for the three distinct variables are studied within the specific distribution (log likelihood distribution, invariant mass and  $\vec{p}_T$  distributions) and are not compared between distributions, as they have

variable bin width. The discriminating values are mentioned in all the plots and are also listed along with their relative uncertainties in tables from 7-1 to 7-7 for the respective cases and lepton channels considered in the analysis.

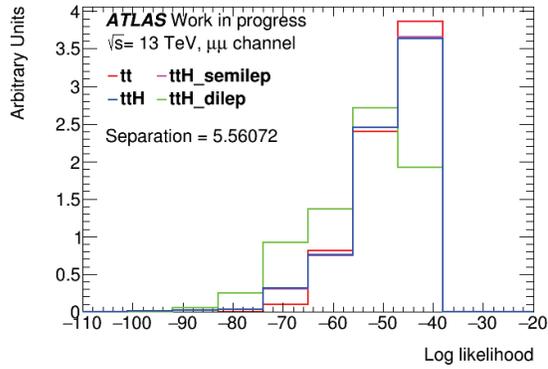
Note that the lepton channel notation in the tables below correspond to the flavour of the two same-sign leptons that satisfy the  $2LSS + 0\tau$  selection and subsequently enter the fit. In tables 7-1 and for case 2 in tables from 7-2 to 7-7, only the leading lepton is included in the likelihood, hence, the lepton notation also corresponds to the flavour of the leading lepton. Specifically,  $ee$  would refer to a leading electron and  $\mu\mu$  would indicate a leading muon and the mixed lepton channel could either comprise a leading electron or a muon depending on the lepton  $\vec{p}_T$ . In the lepton notation for case 1, both the lepton flavours listed under the lepton channel are included in the likelihood.



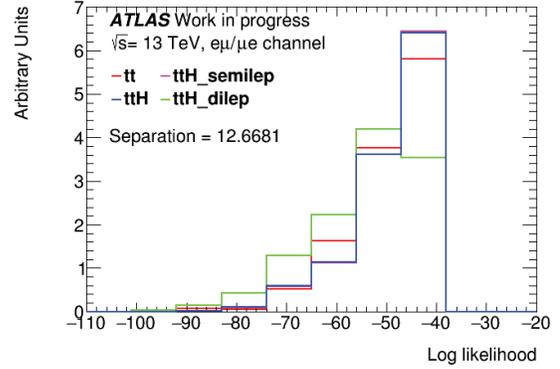
(a) Log likelihood for combined lepton channel



(b) Log likelihood for  $ee$  channel



(c) Log likelihood for  $\mu\mu$  channel

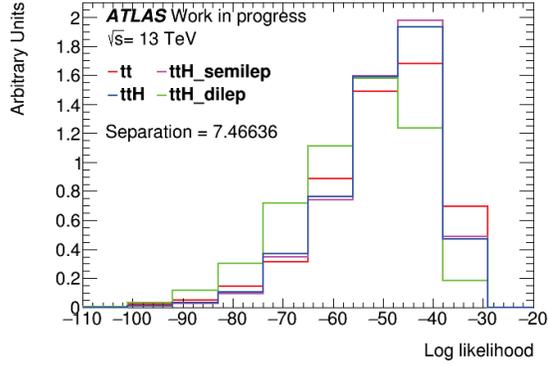


(d) Log likelihood for mixed lepton channel

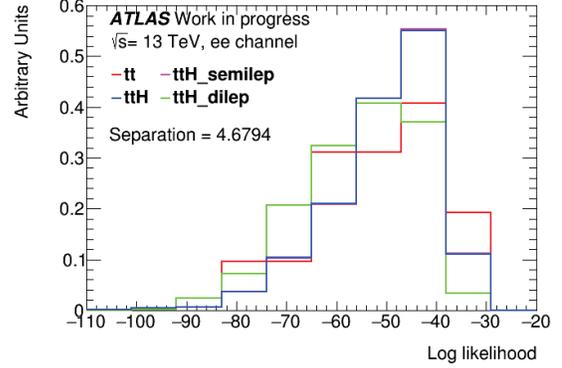
Figure 7–1: KLFitter log likelihood plots for  $2LSS + 0\tau$  selection applied to  $lepton + jets$  likelihood, highest  $\vec{p}_T$  lepton is selected as input to the fitter and b-tagging information is not required by the fit (section 6.1).

Table 7–1: Discriminating power of KLFitter for log likelihood values of  $t\bar{t}$  and  $t\bar{t}H$  event topologies in the  $2LSS + 0\tau$  final state for  $lepton+jets$  KLFitter likelihood.

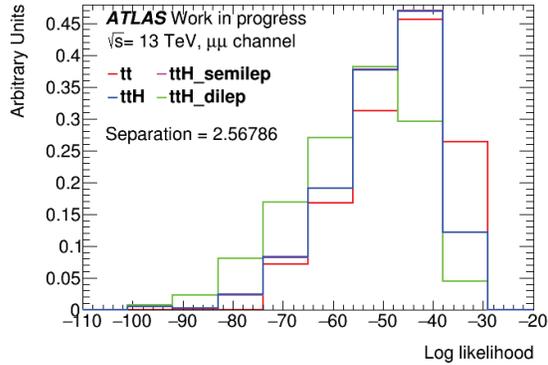
Lepton channel	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Separation value	$6.00 \pm 3.78$	$5.56 \pm 2.92$	$12.67 \pm 7.08$	$16.07 \pm 8.17$



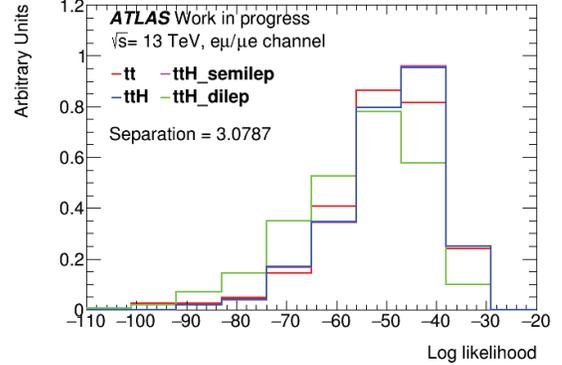
(a) Log likelihood for combined lepton channel



(b) Log likelihood for  $ee$  channel



(c) Log likelihood for  $\mu\mu$  channel

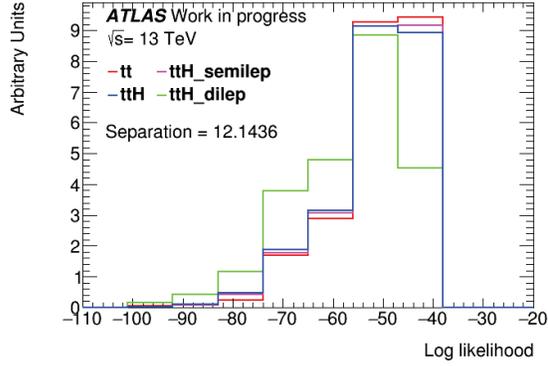


(d) Log likelihood for mixed lepton channel

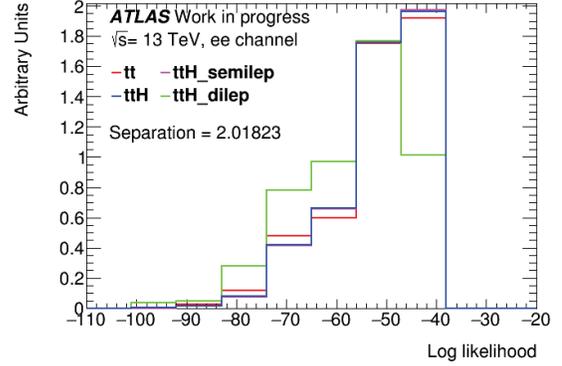
Figure 7-2: KLFitter log likelihood plots for Case 1 - Minimizing  $\Delta R$  between  $2^{nd}$  lepton and b-jets,  $n_{\text{Jets}} \geq 4, < 8$ ,  $k_{\text{VetoNoFitBoth}}$  is used as the KLFitter b-tagging option in the above plots (section 6.2.1).

Table 7-2: Discriminating power of KLFitter for log likelihood values of  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using  $k_{\text{VetoNoFit}}$ .

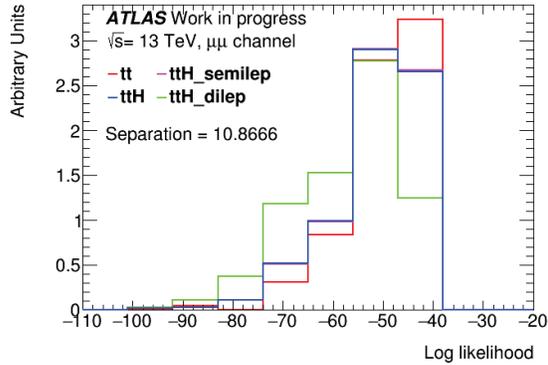
Cases/B-Tagging	kVetoNoFit			
	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$8.02 \pm 4.07$	$9.72 \pm 4.38$	$12.70 \pm 6.15$	$16.34 \pm 8.96$
Case 2	$2.02 \pm 0.99$	$10.87 \pm 5.90$	$6.66 \pm 3.40$	$12.14 \pm 6.00$



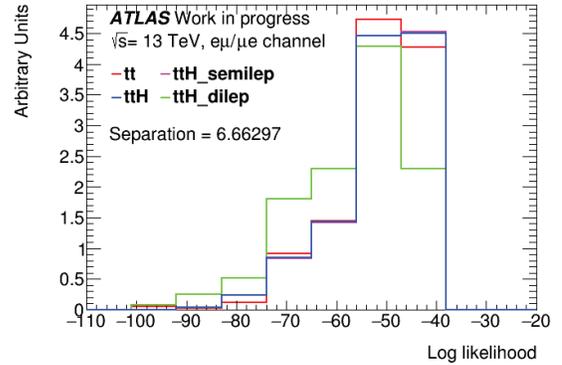
(a) Log likelihood for combined lepton channel



(b) Log likelihood for  $ee$  channel



(c) Log likelihood for  $\mu\mu$  channel

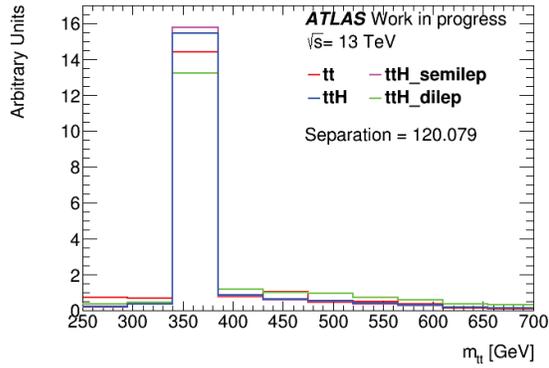


(d) Log likelihood for mixed lepton channel

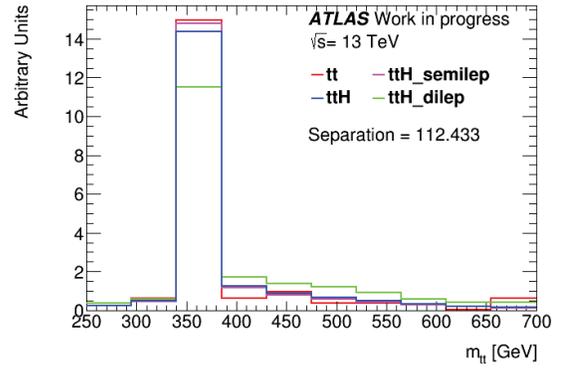
Figure 7-3: KLFitter log likelihood plots for Case 2 - Merging  $2^{nd}$  lepton and b-jets,  $n_{\text{Jets}} \geq 4, < 8$ , kVetoNoFit is used as the KLFitter b-tagging option in the above plots (section 6.2.2).

Table 7-3: Discriminating power of KLFitter for log likelihood values of  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using kVetoNoFitBoth.

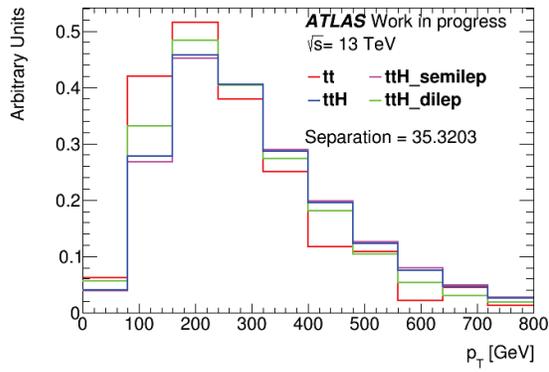
Cases/B-Tagging	kVetoNoFitBoth			
	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$4.68 \pm 2.05$	$2.56 \pm 1.44$	$3.08 \pm 1.53$	$7.46 \pm 3.43$
Case 2	$4.62 \pm 2.21$	$3.88 \pm 2.17$	$2.70 \pm 1.33$	$6.97 \pm 3.21$



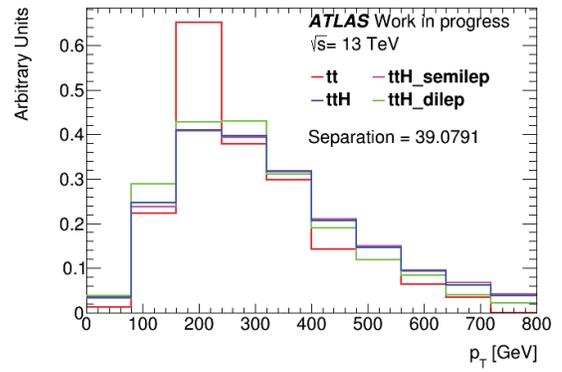
(a) Reconstructed  $t\bar{t}$  mass for combined lepton channel for kVetoNoFit



(b) Reconstructed  $t\bar{t}$  mass for combined lepton channel for kVetoNoFitBoth



(c) Reconstructed  $\vec{p}_T$  distribution of  $t\bar{t}$  system using kVetoNoFit for combined lepton channel



(d) Reconstructed  $\vec{p}_T$  distribution of  $t\bar{t}$  system using kVetoNoFitBoth for combined lepton channel

Figure 7-4: From top: Combined lepton channel plots for reconstructed top pair mass and  $\vec{p}_T$  distribution of the  $t\bar{t}$  system for Case 1-  $n_{\text{Jets}} \geq 4, < 8$ , both the KLFitter b-tagging options are shown here.

Table 7-4: Discriminating power of KLFitter for reconstructed  $t\bar{t}$  mass in  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using kVetoNoFit.

Cases/B-Tagging	kVetoNoFit			
Lepton channel	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$9.00 \pm 4.49$	$16.12 \pm 7.91$	$21.73 \pm 11.61$	$120.01 \pm 57.71$
Case 2	$4.40 \pm 1.64$	$6.28 \pm 2.22$	$6.36 \pm 2.38$	$52.98 \pm 20.49$

Table 7–5: Discriminating power of KLFitter for reconstructed  $t\bar{t}$  mass in  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using kVetoNoFitBoth.

Cases/B-Tagging	kVetoNoFitBoth			
Lepton channel	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$8.33 \pm 3.23$	$11.63 \pm 5.34$	$13.66 \pm 5.53$	$112.43 \pm 47.31$
Case 2	$9.74 \pm 4.32$	$14.41 \pm 6.69$	$18.02 \pm 9.33$	$143.08 \pm 71.35$

Table 7–6: Discriminating power of KLFitter for reconstructed  $t\bar{t} \vec{p}_T$  distribution in  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using kVetoNoFit.

Cases/B-Tagging	kVetoNoFit			
Lepton channel	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$2.49 \pm 1.04$	$4.53 \pm 2.13$	$7.41 \pm 2.97$	$35.32 \pm 14.99$
Case 2	$2.67 \pm 1.12$	$4.41 \pm 1.96$	$6.43 \pm 2.56$	$32.08 \pm 13.65$

Table 7–7: Discriminating power of KLFitter for reconstructed  $t\bar{t} \vec{p}_T$  distribution in  $t\bar{t}$  and  $t\bar{t}H$  event topologies for modified KLFitter in the  $2LSS + 0\tau$  final state for Case 1 and Case 2 using kVetoNoFitBoth.

Cases/B-Tagging	kVetoNoFitBoth			
Lepton channel	$ee$	$\mu\mu$	$e\mu/\mu e$	Total
Case 1	$3.17 \pm 1.67$	$5.70 \pm 3.01$	$6.15 \pm 3.17$	$39.07 \pm 20.79$
Case 2	$2.51 \pm 1.18$	$5.58 \pm 2.39$	$5.92 \pm 2.58$	$40.40 \pm 17.60$

## CHAPTER 8 Discussion of Results

A comparison study of the separation power and b-tagging options between the  $t\bar{t}$  and the  $t\bar{t}H$  topologies is performed for the modified likelihood definition cases outlined in chapter 6 in the  $2LSS + 0\tau$  final state.

A common trend observed in all the plots (figures 7-1, 7-2, 7-3, 7-4) is that, the output for *semileptonic*  $t\bar{t}H$  traces the  $t\bar{t}H$  output. It is known that the dominant contribution for the  $2LSS$  final state arises from the *semileptonic* decay of  $t\bar{t}$  and the overall  $t\bar{t}H$  decay in the  $2LSS$  final state can be decomposed into  $t\bar{t}H = (t\bar{t} \rightarrow \textit{semileptonic}) + H (\rightarrow WW \rightarrow l\nu q\bar{q})$ . Hence, we see the  $t\bar{t}H$  and the *semileptonic*  $t\bar{t}H$  histograms overlapping in all the plots.

The  $t\bar{t}$  sample is expected to produce a higher log likelihood value in the KL-Fitter *lepton + jets* likelihood definition for the reconstruction of its events in the *semileptonic* decay channel as compared to the  $t\bar{t}H$  events. All the four MC samples in the log likelihood plot in figure 7-1 show a peak towards the right of the x-axis with a less negative log likelihood value, which corresponds to events in the samples being increasingly  $t\bar{t}$ - like or satisfying the  $t\bar{t}$  hypothesis. Hence, the discrimination between the two topologies for the original KL-Fitter likelihood definition is not very large as seen in the values presented in table 7-1.

The characteristics of the sub-leading lepton is further exploited to modify the KL-Fitter likelihood definition as outlined in chapter 6. Comparing the log likelihood

separation values in tables 7-1 and 7-2, 7-3 for the two different KLFitter likelihood definitions, it can be observed that there is a marginal improvement, considering the uncertainties, in the discriminating power of the fit with the modified KLFitter, especially in the  $ee$  and  $\mu\mu$  channel for kVetoNoFit option in case 1. Case 1 with the b-tagging option of kVetoNoFit provides the highest separation value in this scenario as the angular separation between either of the b-jets and the sub-leading lepton is one of the important differences between the two topologies.

A distinct feature observed in the separation values for the btagging option of kVetoNoFitBoth in both case 1 and case 2, is the presence of a strong statistical correlation between the values, despite the uncertainties. This feature arises due to the fact that most of the input events entering the fit in the sample is also entering the fit for the other samples. This would mean that events which tend to satisfy the  $t\bar{t}$  hypothesis in one likelihood definition would also satisfy the hypothesis in the other likelihood definition, which is dependent on how efficiently the KLFitter reconstructs the topology. The values in table 7-3 are not influenced much by the value of the errors, meaning that the errors do not reflect the significance of the separation value.

The distribution of the reconstructed mass of the top quark pair for all four MC samples is shown in plots 7-4 (a) and 7-4 (b) for Case 1 for the combined lepton channel. The peak lies around 370 GeV, which is approximately the total invariant mass of the two top quarks, showing that the assignment of the jets to the final state quarks in the KLFitter likelihood definition is quite efficient. Case 2 with the kVetoNoFitBoth option performs the best, giving a discriminating value of 143.08

$\pm 71.35$ . This outcome may be intuitive as the merging of the sub-leading lepton to the closest b-jet would lead to a better reconstruction of the leptonic top quark mass, as the lepton is assumed to carry away a substantial amount of energy from the b-jet. As seen in the above case, statistical correlation is observed in the case of kVetoNoFitBoth.

Plots 7-4 (c) and 7-4 (d) show the reconstructed  $\vec{p}_T$  distribution of the two top quarks for case 1 with kVetoNoFit and kVetoNoFitBoth. The  $t\bar{t}$  events peaking around 200 GeV indicate that, in a large number of events, the  $\vec{p}_T$  distribution of the two top quarks in the  $t\bar{t}$  topology lie around 200 GeV, unlike the top quarks in  $t\bar{t}H$ . Comparing the two b-tagging options, kVetoNoFitBoth provides an enhanced separation value as compared to kVetoNoFit for case 1 and case 2. Case 2 with kVetoNoFitBoth option performs better than the other cases. Overall, kVetoNoFitBoth also shows a better statistical correlation between the discrimination values for different likelihoods in all three variables above. This is because, by definition, kVetoNoFitBoth enhances the probability for picking the correct jet permutation among the selected input jets corresponding to the final state quarks, which are compatible with the *semileptonic*  $t\bar{t}$  decay model before they enter the fit.

Ultimately, these variables would enter a multivariate analysis in the future, in which cuts would be applied to the log likelihood, invariant mass and  $\vec{p}_T$ , with the outcome that the combined background rejection power of the KLfitter is improved relative to the individual separation values of the three variables.

Another aspect regarding the application of the KLfitter to the two topologies is related to the improved kinematic fitting which results in better reconstructed

objects. Important features used in the fit include the permutation of the jets to correctly assign them to the final state quarks and considering the kinematic properties of the top quarks, leptons and neutrinos. An example of enhanced reconstruction capability of the tool is seen in the invariant mass peak of the  $t\bar{t}$  pair, in which the tool is efficient in selecting the correct jets and leptons. KLFitter is a multifunctional tool which can be subsequently exploited in reconstructing various event topologies. As an example, the signal  $t\bar{t}H$  topology can be reconstructed to improve the background rejection by selecting the closest daughter particles produced in the decay. Better understanding of the reconstruction of objects would also enable fine-tuning of the subsequent selection cuts imposed on these objects, such as, isolation or  $\vec{p}_T$  cuts on jets and leptons, associated with different parent particles.

## 8.1 Conclusion and Outlook

As seen in the above discussion, background  $t\bar{t}$  decay events mimic the kinematics of the  $t\bar{t}H$  signal topology in the  $2LSS + 0\tau$  final state which is demonstrated in the log likelihood plots using the *lepton + jets* likelihood definition for KLFitter. Crucial differences in the decay properties of the two topologies in the  $2LSS + 0\tau$  channel are exploited and efforts were made to differentiate between the kinematics of the two processes by modifying the KLFitter likelihood definition and building two specific cases to be analyzed. Overall, the modified KLFitter likelihood gives a better discriminating value as compared to the *lepton + jets* likelihood. Within the modified likelihood definition, Case 1 with the b-tagging option of kVetoNoFit performs better than Case 2 in terms of the KLFitter log likelihood used as a discriminant. Additionally, the reconstructed mass and  $\vec{p}_T$  distribution of the  $t\bar{t}$  system

also provide a separation power to some extent; Case 2 (kVetoNoFitBoth) performed better in both the variables as compared to the other cases.

Apart from the uncertainties originating from the statistics of the MC samples, contributions from potential systematic uncertainties is currently not taken into account, when calculating the values for the discriminating power of the fit. These uncertainties would essentially arise from the individual terms in the likelihood definitions such as the  $BWs$  and  $TFs$ , which depend on the MC and data agreement. Since validating the MC samples against the data is beyond the scope of this thesis, systematic uncertainties are not considered in the separation value calculations. The future goal of the analysis would be to include the kinematic reconstruction discriminant and its related systematic uncertainties from KLFitter as an input variable in the global event multivariate analysis currently being used to separate the high  $t\bar{t}$  background from the  $t\bar{t}H$  signal.

A further extension of the project would be using information from the MC truth for the events to check how often the right candidate is taken into consideration by the fit. This would help in an improved understanding of the reconstruction of the topologies, for instance, it would tell us how often the correct lepton is being assigned as the non-prompt lepton arising from the b-jets in the  $t\bar{t}$  topology. However, several technical challenges were encountered in studying the MC truth information, such as the relevant truth information not being available in the ntuple version used in the analysis and ntuple production being centrally scheduled and managed by a large group, leading to the complexity in its production and rerunning. As a result, having access to the MC truth information was not possible on the timescale of this analysis

and hence, this section is beyond the scope of the thesis and is a part of future analysis.

## REFERENCES

- [1] G. Aad *et al.* [ATLAS Collaboration], *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. B 716, (2012).
- [2] Sheldon L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. 22, (1961).
- [3] A. Salam and J.C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. 13, (1964).
- [4] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys. Rev. 127, (1962).
- [5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. 19, 1264, (1967).
- [6] D. Denegri, *The discovery of the W and Z*, Phys.Rept. 403-404, (2004).
- [7] S. F. Novaes, *Standard model: An Introduction*, Particles and fields, Proceedings, 10th Jorge Andre Swieca Summer School, Brazil, (1999).
- [8] T. W. B. Kibble, *Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism*, Scholarpedia, 4, No. 1, ID.6441, (2009).
- [9] M. Cristinziani and M. Mulders, *Top-quark physics at the Large Hadron Collider*, arxiv:1606.00327, (2016).
- [10] <https://cds.cern.ch/record/1473657/files/>.
- [11] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory*, Avalon Publishing, (1995).
- [12] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C, 40, 100001, (2016).
- [13] J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. 30, Issue 26, (1973).

- [14] B. C. Allanach, *et al.*, *SUSY parameter analysis at TeV and Planck scales*, Nucl. Phys. Proc. Suppl. 135, (2004).
- [15] [ATLAS, CDF, CMS, D0 collaborations], *First combination of Tevatron and LHC measurements of the top-quark mass*, arxiv:1403.4427, (2014).
- [16] S. Abachi *et al.* [D0 Collaboration], *Observation of the top quark*, Phys. Rev. Lett. 74, (1995).
- [17] G. Aad *et al.* [ATLAS and CMS Collaborations], *Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. 114, (2015).
- [18] S. Chatrchyan *et al.* [CMS collaboration], *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B716, (2012).
- [19] A. Denner *et al.*, *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur.Phys.J. C71, 1753, (2011).
- [20] G. Aad *et al.* [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. B749, (2015).
- [21] C. De Melis, <https://cds.cern.ch/record/2197559>, As of 2016-07-08.
- [22] A. Hoecker, *Physics at the LHC Run-2 and Beyond*, European School of High-Energy Physics, arxiv:1611.07864, (2016).
- [23] [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#2012\\_data](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#2012_data), As of 2015-07-20.
- [24] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>, As of 2017-04-22.
- [25] A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, JINST 11, (2016).
- [26] V. Mitsou [ATLAS TRT collaboration], *The ATLAS Transition Radiation Tracker*, arxiv:hep-ex/0311058, (2003).
- [27] C. W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, Reviews of Modern Physics 75, (2003).

- [28] C. Grupen and B. Shwartz, *Particle Detectors, 2nd Edition*, Cambridge Monographs on Particle Physics, (2008).
- [29] P. Francavilla [ATLAS Collaboration], *The ATLAS Tile Hadronic Calorimeter performance at the LHC*, Journal of Physics: Conference Series vol. 404, (2012).
- [30] ATLAS muon spectrometer: Technical design report [ATLAS collaboration], CERN-LHCC-97-22, (1997).
- [31] A. Hamilton [ATLAS Collaboration], *The ATLAS Trigger System Commissioning and Performance*, arXiv:1010.0017, (2010).
- [32] T. Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. 191, (2015).
- [33] M. Bahr *et al.*, *Herwig++ Physics and Manual*, Eur. Phys. J. C58, (2008).
- [34] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP. 02, (2009).
- [35] S. Frixione *et al.*, *The MCaNLO 4.0 Event Generator*, arxiv:1010.0819 (2010).
- [36] C. Oleari, *The POWHEG-BOX*, Nucl. Phys. Proc. Suppl. 205-206, (2010).
- [37] C. Ay *et al.*, *Monte Carlo generators in ATLAS software*, Journal of Physics: Conference Series, 219, (2010).
- [38] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C63, (2009).
- [39] T. Sjöstrand, *Status and developments of event generators*, Proceedings, 4th Large Hadron Collider Physics Conference LHCP2016, (2016).
- [40] S. Höche, *Introduction to parton-shower event generators*, arxiv:1411.4085, (2014).
- [41] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. A15S1, (1999).
- [42] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP, (2002).
- [43] S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. A506, (2003).

- [44] G. Aad *et al.* [ATLAS collaboration], *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J., C74(10):3071, (2014).
- [45] <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/InDetTrackingDC14>, As of 2014-09-18.
- [46] G. Aad *et al.* [ATLAS collaboration], *Performance of b-Jet Identification in the ATLAS Experiment*, JINST 11, (2015).
- [47] [www-d0.fnal.gov/Run2Physics/top/singletop\\_observation/singletop\\_observation\\_updated.html](http://www-d0.fnal.gov/Run2Physics/top/singletop_observation/singletop_observation_updated.html), As of 2009-05-22.
- [48] M. Cacciari, G. P. Salam, G. Soyez, *Anti-k(t) jet clustering algorithm*, JHEP 04, (2008).
- [49] K. Liu *et al.* [ATLAS Collaboration] , *Search for the Associated Production of a Higgs Boson and a Top Quark Pair in multilepton final states in pp Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*, ATL-COM-PHYS-2017-101, (2017).
- [50] J. Erdmann *et al.*, *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nucl. Instrum. Meth. A748, (2014).
- [51] B. Lemmer, *Measurement of Spin Correlations in  $t\bar{t}$  Events from pp Collisions at  $\sqrt{s} = 7$  TeV in the Lepton + Jets Final State with the ATLAS Detector*, Gottingen U., II. Phys. Inst., (2014).
- [52] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TTHtoLeptonsCutflow3011>, As of 2017-04-26.