ABSTRACT

SEARCH FOR HIGGS PAIR PRODUCTION IN THE $bb\tau\tau$ FINAL STATE WITH THE ATLAS DETECTOR AT THE LARGE HADRON COLLIDER

Puja Saha, Ph.D.
Department of Physics
Northern Illinois University, 2020
Dr. Dhiman Chakraborty, Director

Abstract: This is a search for a Beyond Standard Model (BSM) heavy scalar in the ATLAS detector at 13 TeV center of mass energy using Run 2 data set recorded in 2018 with an integrated luminosity 58.5 fb^{-1}. The heavy scalar (X) decays to two Standard Model (SM) like light Higgs (H) bosons that decay into the $bb\tau\tau$ final state. Particularly, the semi-leptonic final state of two τ's is considered in this search where one τ decays in leptonic final state and other to a hadronic final state. No evidence of a new scalar found, so upper limits are set on the cross-section times branching ratio of the $X \to HH \to bb\tau\tau$ process with 95% confidence interval on heavy scalar (X) masses between 251 and 1000 GeV.

CERN-THESIS-2020-030
24/03/2020

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

MAY 2020

SEARCH FOR HIGGS PAIR PRODUCTION IN THE $bb\tau\tau$ FINAL STATE WITH THE ATLAS DETECTOR AT THE LARGE HADRON COLLIDER

BY

PUJA SAHA
© 2020 Puja Saha

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

Dissertation Director:
Dr. Dhiman Chakraborty

ACKNOWLEDGEMENTS

First, I would like to thank to Dr. Dhiman Chakraborty for giving me the opportunity to pursue PhD at Northern Illinois University and in the ATLAS experiment at the LHC. His guidance on choosing my thesis analysis and working on the ATLAS Tile calorimeter was a great help and prepared me well for my career in particle physics.

I am also very thankful to Dr. Serban Protopopescu for his support towards getting US-ATLAS fellowship at Brookhaven National Laboratory and his supervision on my thesis research. I also thank to Dr. George Redlinger and Dr. Ketevi Assamagan at BNL.

I really appreciate the help I received from my senior graduate student Blake Oliver Burghgrave, who helped me to start my work in the ATLAS. I learned a lot of things from him, specially python and other computational skill. He spent significant time on guiding me during the Tile work which I took over from him in 2016.

Next I will thank to all the members of the NIU ATLAS group - Dr. Jahred Adelman, Dr. Elizabeth Brost, Dr. Pawel Klimek, Dr. Iouri Smirnov, Tyler Burch, Brianna Stamas, Elliot Parrish, Rafael Oreamuno.

I am very grateful to my supervisor in Tile Calormeter Sanya Solodkov to supervise my Tile Data Quality work. Also I want to mention Tibor Zenis, Oleg Solovyanov for their guidance and support. My Tile testbeam work was supervised by Caludio Santoni and got huge help from my colleague Tamar Zakareishvili. I thank both of them.

I got great support and guidance from the ATLAS analysis group. Specially I want to convey my gratitude to Prof. Allison MacCarn who spent her valuable time to answer all my questions and solving problems I faced during my analysis. I also want to thank Agni Bethani, Tatjana Lenz our analysis contacts. I appreciate the help I got from Katharine

Leney, Danilo Ferreira de Lima, Alessandra Betti, Peter Bokan, Carl Guilliam, Tulin Varol, Rodrigo Gombao, Marisilvia Donadelli, Katy Grimm.

I personally want to thank my friend and family members for their motivation and encouragement to pursue PhD. Thanks to my parents Mr. Paritosh Saha and Mrs. Ashima Saha for what they have done for me to see me at this stage of my life. A small thank is not enough to express my love and respect for them. I say thank to my maternal uncle Dr. Soumyajit Saha, all my aunts, uncles, grandparents for their constant support in my education and successes. The person who was always beside me and gave me mental support in difficult time during my PhD is my husband Dr. Arnab Kumar Maity.

I thank my friends Nandini Midya, Rumita Guha, Taniya Modal, Prasanta Char, Suvankar Roy Chowdhury, Utsha Paul Chowdhury, Zinnia Podder, Shabalima Saha Biswas. My friend Late Debasish Chakraborty who was one my best friends, supported me all the time and loved me, will always remain in my heart and thoughts.

The person whose teaching motivated me to pursue physics as my career is Satrajit Roy. I also got great support from my college and university professors, Dr. Bidisha Chottopadhay, Dr. Lachima Gupta, Dr. Anuradha Basu, Dr. Sudeshna Dasgupta, Dr. Jonaki Chowdhury, Dr. Anirban Kundu, Dr. Abhijit Bhattacharjee.

TABLE OF CONTENTS

Page

LIST OF TABLES viii

LIST OF FIGURES. x

LIST OF APPENDICES xiv

Chapter

1 THEORY OF PARTICLE PHYSICS 1

1.1 Introduction 1

1.2 Fundamental forces of nature 1

1.2.1 Gravity 1

1.2.2 Electromagnetic Interaction 2

1.2.3 Strong Interaction. 2

1.2.4 Weak Interaction 2

1.3 Particles in the Standard Model 3

1.3.1 Fermions 3

1.3.2 Bosons. 4

1.4 Spontaneous Symmetry Breaking and the Higgs Mechanism 6

1.5 The Two-Higgs-Doublet Model. 10

1.6 Cross-section and Branching Ratio 13

1.7 Motivation 14

1.8 Previous result 16

2 LARGE HADRON COLLIDER. 19

Chapter Page

3 ATLAS DETECTOR AT THE LHC 24

3.1 Design Consideration of a General Purpose Detector 24

3.2 ATLAS Detector 25

3.3 ATLAS Magnet System 25

3.3.1 Solenoid 27

3.3.2 Toroid 27

3.4 Inner Detector 28

3.4.1 Pixel Detector 29

3.4.2 SCT 29

3.4.3 TRT 30

3.5 Calorimeter 30

3.5.1 Liquid Argon Calorimeter (LArg) 30

3.5.2 Tile Calorimeter (TileCal) 31

3.6 Muon System 33

3.6.1 RPC 34

3.6.2 TGC 34

3.6.3 MDT 34

3.6.4 CSC 35

3.7 The ATLAS Trigger System 35

4 OBJECT RECONSTRUCTION 37

4.1 Object Reconstruction 37

4.1.1 Electrons 37

4.1.2 Jets 39

4.1.3 Taus 39

Chapter Page

4.1.4 Muons . . . 42

4.1.5 b-jets . . . 44

4.1.6 $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 46

5 DATA AND MONTE CARLO SAMPLE . . . 48

6 OBJECT AND EVENT SELECTION . . . 52

6.1 Object Selection . . . 52

6.1.1 Electrons . . . 52

6.1.2 Muons . . . 53

6.1.3 Jets . . . 53

6.1.4 Taus . . . 53

6.1.5 Fake Taus . . . 54

6.1.6 Missing Transverse Energy . . . 54

6.1.7 Overlap Removal . . . 55

6.2 Event Selection . . . 55

6.2.1 Trigger . . . 56

6.2.2 Offline Event Selection . . . 56

7 BACKGROUND ESTIMATION . . . 58

8 MULTIVARIATE ANALYSIS AND B-TAGGING OPTIMIZATION . . . 65

8.1 Boosted Decision Tree . . . 65

8.2 b-tagging Working Point Optimisation . . . 70

9 SYSTEMATIC UNCERTAINTIES . . . 80

10 RESULT . . . 89

10.0.1 2 b-tag the Signal Region . . . 90

10.0.2 1 and 2 b-tag as the Signal Region . . . 90

Chapter Page

10.0.3 2 b-tag with 85% b-tagging efficiency as the Signal Region 94

10.0.4 Summary and Conclusion 96

REFERENCES 99

APPENDICES 110

LIST OF TABLES

Table Page

1.1 Mass, charge of leptons [1]. . . . 3

1.2 Mass, charge of quarks [1]. . . . 4

1.3 2HDM 12

1.4 Higgs BR table 15

2.1 LHC design 22

6.1 Object Selection 57

6.2 Ebject Selection 57

7.1 Yield in 2btag, 2018/MC16e. . . . 62

7.2 Yield in 1btag, 2018/MC16e. . . . 63

8.1 Optimised BDT hyper-parameter settings.. . . . 69

8.2 Ranking of input variables in 1 b-tag. The variables are listed according to their importance. . . . 70

8.3 Ranking of input variables in 2 b-tag. The variables are listed according to their importance. $b2_{Pt}$ was not useful for high mass BDT training. . . . 76

9.1 $t\bar{t}$ modeling systematics 83

9.2 Systematics 87

10.1 PostfitYield 2018 2btag 300GeV 91

10.2 PostfitYield 2018 2btag 600GeV 91

Table Page

10.3 PostfitYield 2018 2btag 900 92

10.4 LimitcombineSyst2btag 92

10.5 LimitcombineSyst12btag 96

10.6 Limit17Systs2btag. 97

10.7 Limit17Systs2btag85WP 97

A.1 Tile modules and their electronics in June 2016 testbeam setup. 115

C.1 Coupling constants considered in MC generation of VBF hh production. . . . 132

LIST OF FIGURES

Figure Page

1.1 Standard Model 5

1.2 The SM Higgs boson production cross sections as a function of the center of mass energy, for pp collisions [6].. 13

1.3 The branching ratios for the main decays of the SM Higgs boson near m_H = 125 GeV [7]. 14

1.4 Higgs pair production (a) triangular, (b) box diagrams. 16

1.5 Branching ratio of di-τ decay [10]. 17

1.6 Limit on 2HDM and RSG model with 36 fb^{-1} data combining semi-leptonic and fully hadronic final state [11] 18

2.1 CERN 21

3.1 ATLAS 26

3.2 ATLAS Magnet 26

3.3 ATLAS Magnet 27

3.4 ATLAS Inner Detector 28

3.5 ATLAS LArg Calorimeter 31

3.6 ATLAS Tile Calorimeter 32

3.7 ATLAS Muon System 33

4.1 ATLAS online luminosity in 2018 [38]. 38

4.2 Electron identification efficiency in $Z \to ee$ events as a function of E_{T} and η [39]. 39

Figure Page

4.3 Tau reconstruction efficiency as a function of truth p_{T} [42].. 40

4.4 Inverse of the efficiency for mis-tagging QCD jets as a function of the identification efficiency for $\tau_{\mathrm{had-vis}}$ candidates. The "loose", "medium", "tight" working points are shown on the curves [42]. The working points do not stay exactly on the line because they implement variable cuts to achieve a reduced p_{T}-dependency of the efficiency. 42

4.5 Identification efficiency of 1-prong and 3-prong $\tau_{\mathrm{had-vis}}$ candidate in $Z \to \tau\tau$ sample [42]. 43

4.6 Reconstruction efficiency of medium, loose and tight muons in $Z \to \mu\mu$ sample [43]. On the left plot the loose selection (squares) in the region $|\eta| < 0.1$ where the Loose and Medium selections differ significantly. 44

4.7 Light and c-jet rejection as a function of b-jet efficiency [44]. 46

5.1 Monte Carlo simulation in High energy physics [52]. 49

7.1 Kinematic variable distributions in 2018 and MC16e in 2b-tag region 59

7.2 Kinematic variable distributions in 2018 and MC16e in 2b-tag region 59

7.3 Fake factor 2btag 61

7.4 Fake factors in 1b-tag 64

7.5 Combine fake factors in 1b-tag 64

8.1 BDT 66

8.2 2018/MC16e in 2 b-tag 71

8.3 2018/MC16e in 2 b-tag 72

8.4 2018/MC16e in 1 b-tag 73

8.5 2018/MC16e in 1 b-tag 74

8.6 BDT over training plots. 75

8.7 Correlation matrix: (a) signal $m_{\mathrm{X}} = 300$ GeV and (b) $t\bar{t}$ processes. 77

8.8 ROC curve 77

Figure Page

8.9 btagging in 1b-tag 78

8.10 btagging in 2b-tag 79

9.1 ISR 84

9.2 FSR. 85

9.3 Fake factor in both regions. 86

9.4 Fake factor due to $t\bar{t}$ generator sample 87

9.5 Fake factor due to $t\bar{t}$ cross-section uncertainty 88

10.1 Postfit in 2018 data/MC16e of 300 GeV and 600GeV mass 2 b-tag region. . . 93

10.2 2018 Limit 94

10.3 Postfit in combine data of 300 GeV and 600GeV mass 1 b-tag region 95

A.1 TileCal signal flow in the current system (Legacy SD) [84]. 112

A.2 Signal flow in Phase 2 upgrade electronics [84]. 112

A.3 June 2016 testbeam setup [85]. 114

A.4 Tile Module setup during 2016 testbeam period [85]. 114

A.5 Normalized pulse shape of Physics data [86]. 117

A.6 The charge in ADC and time distribution in PMT3 with different reconstruction methods, FIT (red) and OF (blue).. 120

A.7 Total energy of 4th Tile row using OF method. 121

A.8 Response of PMTs 3 (left) and PMT4 (right) in BC1 cell after the E_{tot} cut.. 122

A.9 Energy Vs time response of PMT45 and PMT46 in the same cell. 122

A.10 $\frac{dE}{dZ}$ using OF method for 4th Tile row.. 123

B.1 Luminosity-weighted distribution of the mean number of interactions per crossing,μ, for the full Run 2 pp collision data set at 13 TeV [88]. 126

Figure Page

B.2 Schematic illustrating the nominal Run 2 operations work flow for the data quality assessment of ATLAS data [88].. 127

B.3 Evolution of channel and cell Masking in TileCalorimeter [94]. 129

C.1 Leading order diagram of VBF hh production 131

C.2 $\Delta\eta$ of two VBF jets in semi-leptonic final state 133

C.3 Invariant mass distribution of two VBF jets [97]. 133

C.4 Invariant mass distribution of two VBF jets. 135

C.5 $\Delta\eta$ of two VBF jets in semi-leptonic final state with two different parton shower [97]. 136

C.6 Invariant mass distribution of two VBF jets with two different parton shower. 136

LIST OF APPENDICES

Appendix Page

A RESPONSE OF TILE DEMONSTRATOR MODULE TO MUON BEAM 110

A.1 Introduction 111

A.2 Tile Phase 2 upgrade 111

A.3 Testbeam setup 113

A.4 Reconstruction Method 115

A.4.1 Fit Reconstruction Method 116

A.4.2 Optimal Filtering Reconstruction Method 118

A.5 Response of individual PMTs in BC Layer to -90° muon beam 119

A.6 Summary 123

B TILE DATA QUALITY MONITORING 124

B.1 ATLAS Data Quality 125

B.2 Contribution 127

C GENERATION AND VALIDATION OF HIGSS PAIR MONTE CARLO SAMPLES IN VECTOR BOSON FUSION PRODUCTION 130

C.1 Introduction 131

C.2 Non-resonant Production of VBF Higgs Pair 131

C.2.1 MC Generation of $bb\tau\tau$ Final State 132

C.3 Resonant Production of VBF Higgs pair 134

C.4 VBF jet modelling with PYTHIA8 134

C.5 Conclusion 135

CHAPTER 1

THEORY OF PARTICLE PHYSICS

1.1 Introduction

This chapter gives a brief description of particles physics: fundamental forces of nature and particles in the Standard Model (SM). The deficiencies in SM theory further motivate to look beyond SM physics. The last section explains the two-Higgs-doublet (2HDM) that motivates our search of heavy scalar that decays to two SM-like Higgs having $b\bar{b}\tau^+\tau^-$ in the final state with the ATLAS detector.

1.2 Fundamental forces of nature

In this section we will discuss fundamental forces of nature and about their mediator:

1.2.1 Gravity

Four fundamental forces are known to exist in nature. Gravity, is the force that keeps planets in their orbits around the Sun. It is the weakest force and has infinite range. The carrier of this force is gravitons. Gravity is not incorporated in the Standard Model.

1.2.2 Electromagnetic Interaction

Electromagnetic force is mediated by the photon(γ) and interacts with only charged particles, keeping the charged conserved. The photon is massless so electromagnetic force is a long range force.

1.2.3 Strong Interaction

The strong interaction is mediated by two force carriers. On a larger scale (about 1 to 3 fm), it is carried by mesons, that binds protons and neutrons (nucleons) together to form the nucleus of an atom. On the smaller scale (less than about 0.8 fm) this force is mediated by gluons. This holds quarks together inside protons, neutrons, and other hadron particles. In the later context, it is often known as the color force. Quark changes color by exchanging color quantum number. The quarks have three different color quantum numbers names as Red (R), Green (G), Blue (B) and anti-quark have $\bar{R}$, $\bar{G}$, $\bar{B}$. These are not real colors but indicates quantum number.

1.2.4 Weak Interaction

The weak interaction is mediated by the $W^{\pm}$ and the Z vector bosons. $W^{\pm}$ interacts with only left-handed fermions and right-handed anti-fermions.

The weak interaction has been consolidated with electromagnetism as a single interaction called the "electroweak interaction" which will be discussed in detail in later section to

Table 1.1: Mass, charge of leptons [1].

Lepton	Mass (MeV)	Charge (e)
e	0.511	-1
μ	105.65	-1
τ	1776.86	-1

understand how vector bosons acquire mass through the electroweak symmetry breaking and the Higgs mechanism.

1.3 Particles in the Standard Model

The Standard Model consists of seventeen particles are shown in Figure 1.1 . Different categories of particles are discussed in the following sections.

1.3.1 Fermions

Quarks and leptons follow Fermi-Dirac statistics and have half integral ($\frac{1}{2}$,$\frac{2}{3}$....) spin, so in general are known as fermions. Up (u), down (d), charm (c), strange (s), top (t) and bottom (b) are known as quarks. Electron (e), muon (μ), tau (τ) and corresponding neutrinos (ν_e,ν_μ, ν_τ) are leptons. Each of the fermions has it's anti-particle ($\bar{e}$,$\bar{\mu}$,$\bar{\tau}$,$\bar{\nu_e}$,$\bar{\nu_\mu}$,$\bar{\nu_\tau}$,$\bar{u}$,$\bar{d}$,$\bar{c}$,$\bar{s}$,$\bar{t}$,$\bar{b}$). The mass, charge of the fermions and their anti-particles are given in the Table 1.1 and Table 1.2. Anti-particles have same mass as their particle but opposite in charge. Each generation of lepton its own lepton number equal to 1 which is zero for other generation. The three quarks bound states called baryons, such as protons and neutrons. Baryons are fermions.

Table 1.2: Mass, charge of quarks [1].

Quarks	Mass (MeV)	Charge (e)
u	2.3	+2/3
c	1275	+2/3
t	173210	+2/3
d	4.8	-1/3
s	95	-1/3
b	4180	-1/3

1.3.2 Bosons

As described in the previous section each of the forces in nature is carried by bosons, which have integral spins, and therefore obey Bose-Einstein statistics. Gluons (g), photons (γ), W/Z bosons ($W^{\pm}$, Z^0), these are vector bosons. Only one scalar boson exists that is known as Higgs boson (H).

The Electroweak part of the SM is represented by $SU_L(2) \times U_Y(1)$ symmetry. $SU_L(2)$ has three guage bosons ${W_\mu}^a$, where a=0,1,2, couple to weak-isospin I. The generators in $SU_L(2)$ group are the Pauli Spin matrices. $U_Y(1)$ subgroup is known as weak hypercharge, has a vector boson B_μ. The linear combinations of ${W_\mu}^1$ and ${W_\mu}^2$ give the two $W^{\pm}$ bosons. The ${W_\mu}^3$ must be neutral vector boson. The the photon and the Z boson are formed by the mixing between the B_μ and ${W_\mu}^3$ fields. The weak hypercharge Y is a conserved charge just like electric charge ($Q = I_3 + \frac{Y}{2}$, I_3 is third component of weak-isospin) but it is different for left-handed and right-handed fermions. Both members of an $SU_L(2)$ doublet must have the same weak hypercharge in order to satisfy the $SU_L(2)$ gauge invariance.

A quark and anti-quark bound state can form mesons, mesons are bosons. π^0, $\pi^{\pm}$ are knows as pions, they are the lightest mesons.

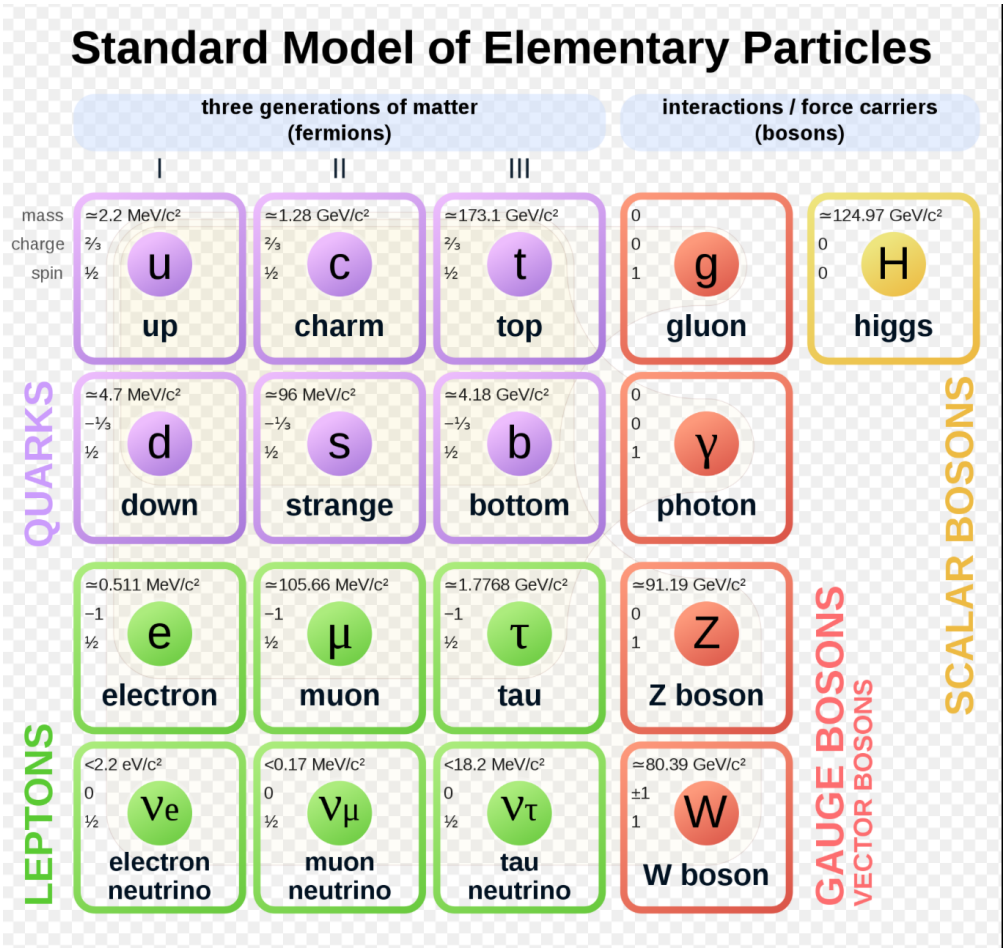


Figure 1.1: Particles in Standard Model [2].

1.4 Spontaneous Symmetry Breaking and the Higgs Mechanism

The situation in which the laws of physics are invariant under some symmetry transformations, but the vacuum state is not, is called spontaneous symmetry breaking.

In this section we will discuss the spontaneous symmetry breaking with a simple example. The following is the Lagrangian of a scalar field ϕ:

$$L(\phi) = (\partial^\mu \phi)^*(\partial_\mu \phi) - \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 \tag{1.1}$$

With potential energy:

$$V(\phi, \phi^*) = \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 \tag{1.2}$$

This Lagrangian is invariant under U(1) symmetry:

$$\phi(x) \to \phi'(x) = e^{i\alpha}\phi(x) \tag{1.3}$$

Here, α is a real number, x is the position. When $\mu^2 > 0$ and $\lambda > 0$, then minimum of the potential is at $\phi = 0$. The ground state of the system is where the expectation value of scalar field $\phi(x)$ vanishes.

If $\lambda < 0$, the expectation value of $\phi(x)$ will grow without bound. If $\mu^2 < 0$ and $\lambda > 0$, the potential looks like a "Mexican hat" shape with a local maximum $\phi = 0$ and a degenerate set of minima with :

$$|\phi_{min}|^2 = \frac{v^2}{2} \tag{1.4}$$

where, $v = \sqrt{\frac{-\mu^2}{\lambda}}$ and the Vaccum Expectation Value (VEV) of the ϕ is $< 0|\phi(x)|0 >= \frac{v}{\sqrt{2}}$. Now the initial $U(1)$symmetry does not hold anymore, VEV is not invariant under $U(1)$ transformation.

$$< 0|\phi'(x)|0 >= e^{i\alpha} < 0|\phi(x)|0 >= e^{i\alpha}\frac{v}{\sqrt{2}} \neq \frac{v}{\sqrt{2}} \tag{1.5}$$

If we define the scalar field in different way:

$$\phi(x) = \frac{1}{\sqrt{2}}[v + h(x)]e^{\frac{iG(x)}{v}} \tag{1.6}$$

where, $h(x)$ is the Higgs boson and $G(x)$ is the Goldstone boson. The Lagrangian density for the scalar and vector degrees of freedom is:

$$L(\phi) = (D^\mu\phi)^*(D_\mu\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.7}$$

$$D^\mu\phi = (\partial_\mu + igA_\mu)\phi \tag{1.8}$$

and D^μ is the gauge covariate derivative given by Equation 1.8 and this Lagrangian is symmetric under $U(1)$ local symmetry.

After the $U(1)$ transformation:

$$D^\mu\phi = \frac{1}{\sqrt{2}}[\partial_\mu h + ig(A_\mu + gv\partial_\mu G)(v + h)]e^{iG/v} \tag{1.9}$$

It is convenient to define a new vector:

$$V_\mu = A_\mu + \frac{1}{gv}\partial_\mu G \tag{1.10}$$

After the symmetry breaking, Equation 1.7 becomes:

$$L = \frac{1}{2}[\partial^{\mu} h \partial_{\mu} h + g^2(v+h)^2 V^{\mu} V_{\mu}] - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} - \lambda(vh + h^2/2)^2 \tag{1.11}$$

These gives the mass to $h(x)$ and vector boson V_{μ}:

$$m_h{}^2 = 2\lambda v^2 \tag{1.12}$$

$$m_V{}^2 = g^2 v^2 \tag{1.13}$$

After the symmetry breaking, the mass term of G disappears and it can be seen why the disappearance of the field G causes the appearance of the massive boson now. A massless spin-1 vector boson (like the photon) has only two possible polarization states, transverse to its direction of motion but a massive spin-1 vector boson has three possible polarization states; the two transverse, and one longitudinal (parallel) to its direction of motion. The additional polarization state degree of freedom had to come from somewhere, so one real scalar degree of freedom had to disappear. One common description of this is that the vector boson becomes massive by eating (not in real sense) the would-be Nambu-Goldstone boson G, which becomes its longitudinal polarization degree of freedom. This is called the Higgs mechanism [3]. The original field $\phi(x)$ is called a Higgs field, and the surviving real scalar degree of freedom $h(x)$ is called by the generic term Higgs boson.

When $SU_L(2) \times U_Y(1)$ symmetry is spontaneously broken, it gives mass to the vector bosons $W^{\pm}$, Z.

The Lagrangian containing the complex scalar filed can be written as:

$$L(\phi) = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) - \mu^2 \phi^{\dagger}\phi - \lambda(\phi^{\dagger}\phi), \tag{1.14}$$

ϕ is a SU(2) doublet. It is known as Higgs field, each member of this doublet is a complex scalar field:

$$\phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}$$

The covariant derivative is given by:

$$D^\mu = \partial_\mu + igT_a{}^{(n)}W_\mu{}^a + ig^{'}YB_\mu, \tag{1.15}$$

where $T_a^{(n)}$ denote the generators of $SU(2)$ in the n-dimensional representation and Y is the identity matrix times the hypercharge.

Under guage transformation ϕ transforms like:

SU(2): $\phi(x) \to \phi'(x) = e^{\frac{i\alpha^a(x)\sigma^a}{2}}\phi(x)$

U(1): $\phi(x) \to \phi'(x) = e^{\frac{i\alpha(x)}{2}}\phi(x)$

when $\mu^2 < 0$, the SU(2) symmetry is spontaneously broken, the VEV $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is no longer zero. This non-zero VEV must be assigned to the neutral member of the doublet.

$$< 0|\phi|0 >= \begin{bmatrix} 0 \\ v/\sqrt{2} \end{bmatrix}$$

The Higgs field ϕ has two complex or four real degrees of freedom. We can now write the Higgs field like:

$$\phi(x) = e^{\frac{iG^a(x)\sigma^a}{2}} \begin{bmatrix} 0 \\ (v + h(x))/\sqrt{2} \end{bmatrix}$$

After the symmetry breaking , $SU_L(2) \times U_Y(1) \to U_Q(1)$. The Nambu-Goldstone boson can be removed in the $U_Q(1)$ symmetry.

$$\phi(x) = \begin{bmatrix} 0 \\ (v + h(x))/\sqrt{2} \end{bmatrix}$$

The kinematics part of the Lagrangian Equation 1.14 will give:

$$L_{\phi,\text{kinematic}} = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) = \partial_{\mu}h\partial^{\mu}h + \frac{(v+h)^2}{4}[g^2 {W_{\mu}}^{+} W^{-\mu} + \frac{g^2 + g^{'2}}{2} Z_{\mu}Z^{\mu}] \quad (1.16)$$

The second and third terms give the masses of the $W^{\pm}$ and Z bosons, respectively. This is how the massive ($W^{\pm}$, Z) and massless (γ) vector bosons appear in the SM. The last missing piece of the electroweak SM Lagrangian is the Yukawa Lagrangian:

$$L_{Yuk} = {\Gamma^{u}}_{mn}\bar{q}_{m,L}\bar{\phi}u_{n,R} + {\Gamma^{d}}_{mn}\bar{q}_{m,L}\phi d_{n,R} + {\Gamma^{e}}_{mn}\bar{l}_{m,L}\phi e_{n,R} + {\Gamma^{\nu}}_{mn}\bar{q}_{m,L}\bar{\phi}\nu_{n,R} + h.c, \quad (1.17)$$

$\bar{q}_{m,L}$ and $\bar{l}_{m,L}$ are the left handed doublets, where $u_{n,R}$, $d_{n,R}$, $e_{n,R}$, $\nu_{n,R}$ are called the right handed singlet. m and n are the fermion family indices. The matrices Γ_{mn} is called the Yukawa couplings between Higgs doublet ϕ and fermions. The Yukawa Lagrangian is gauge invariant since the combination of $\bar{L}\phi R$ is a $SU(2)_L$ singlet.

1.5 The Two-Higgs-Doublet Model

It is clear that the SM cannot be the complete picture. Among other shortcomings, it doesn't have a suitable dark matter candidate, nor a satisfactory explanation for the observed level of baryon-antibaryon asymmetry. Among a number of extensions of the SM that offer

a means to overcome these problems is supersymmetry, which requires an extension of the minimal Higgs sector postulated in the SM. The simplest such extension which incorporates two, instead of one, complex doublet scalar fields, is known as two Higgs doublet model (2HDM). The motivation behind the 2HDM is that by adding another scalar doublet to the SM we observe some interesting properties, such as the appearance of additional Higgs particles, sources of CP-violation [4, 5].

The Lagrangian in 2HDM having two scalar doublets, ϕ_1 and ϕ_2 is given by:

$$L(\phi) = (D^\mu \phi_1)^\dagger (D_\mu \phi_1) + (D^\mu \phi_2)^\dagger (D_\mu \phi_2) - V(\phi_1 \phi_2) \tag{1.18}$$

If we expand the scalars around their minima, then it becomes:

$$\phi_a = \begin{bmatrix} {\phi_a}^{\pm} \\ \frac{v_a + \rho_a + i\eta_a}{\sqrt{2}} \end{bmatrix}$$

where a=1,2. After expanding the the scalar doublets in the 2HDM potential we get bi-linear terms of ${\phi_a}^{\pm}$, ρ_a and η_a, these terms give mass to the propagators. The 2HDM potential exhibits two CP-conserving neutral minima $< \phi_1 >$ and $< \phi_2 >$ of the form:

$$< \phi_1 > = \begin{bmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{bmatrix}$$

$$< \phi_2 > = \begin{bmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{bmatrix},$$

v_1 and v_2 are the VEV's of the scalar doublets. To get the physical fields mass-squared matrices are diagonalized. The fields ${\phi_a}^{\pm}$, ρ_a and η_a in the guage basis are transformed into physical fields. The massless particles G^0 and $G^{\pm}$ are the three Goldstone bosons of the

Table 1.3: Natural flavour conservation models.'i' is a generation index.

Models	Type I	Type II	Lepton-specific	Flipped
${u_R}^i$	ϕ_2	ϕ_2	ϕ_2	ϕ_2
${d_R}^i$	ϕ_2	ϕ_1	ϕ_2	ϕ_1
${e_R}^i$	ϕ_2	ϕ_2	ϕ_1	ϕ_2

2HDM and are "eaten" up to give the $W^{\pm}$, Z^0 gauge bosons. The remaining five physical particles H^0(heavy) and h^0(lighter) are the CP-even, A^0 is a CP-odd and the $H^{\pm}$ are the charged Higgs:

$$h^0 = \rho_1 \sin\alpha - \rho_2 \cos\alpha, \tag{1.19}$$

$$H^0 = -\rho_1 \cos\alpha - \rho_2 \sin\alpha, \tag{1.20}$$

$$H^{\pm} = \phi_1^{\pm} \sin\beta - \phi_2^{\pm} \cos\beta, \tag{1.21}$$

$$A^0 = \eta_1 \sin\beta - \eta_2 \cos\beta, \tag{1.22}$$

where α is a mixing angle between the CP-even scalars (h^0, H^0) and β is a rotational angle that diagonalizes the mass-squared matrices charged Higgs and the psuedoscalars. In SM, we don't see FCNC (floavour changing neutral current) at tree level because diagonalizing the mass matrix diagonalizes the Yukawa interactions. In 2HDM introduces the possibility of having FCNC at tree-level. The interaction between fermions and scalar boson in 2HDM is determined by the Yukawa Lagrangian like SM interaction. The transformation from a general to a flavour-conserving 2HDM can be achieved by imposing a discrete or continuous symmetries on the Higgs doublets ϕ_1 and ϕ_2. These discrete symmetries suppress FCNC at the tree-level, these constraint the coupling of the Higgs doublets with the fermions.

To avoid confusion from now onward we use X for the 2HDM neutral CP-even heavy scalar and SM like Higgs with H.

1.6 Cross-section and Branching Ratio

Cross-section and branching ratios are the two important definitions in the particle physics.

When two particles interact, their mutual cross section is the area transverse to their relative motion within which they must meet in order to scatter from each other. If the particles are hard inelastic spheres that interact only upon contact, their scattering cross section is related to their geometric size. Figure 1.2 [6] shows the SM Higgs cross-section at different center-of-mass energy ($\sqrt{s}$) in different production modes.

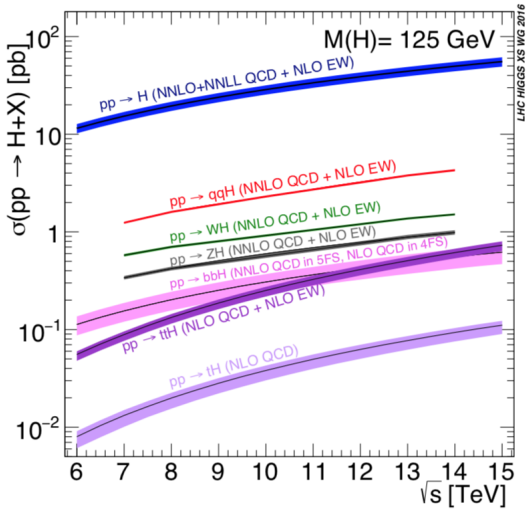


Figure 1.2: The SM Higgs boson production cross sections as a function of the center of mass energy, for *pp* collisions [6].

Branching ratio (BR) is the fraction of particles decay to an individual decay mode with respect to the total number of particles which decay. Figure 1.3 [7] shows the branching ratio of SM Higgs and Table 1.4 shows the value of SM Higgs BR.

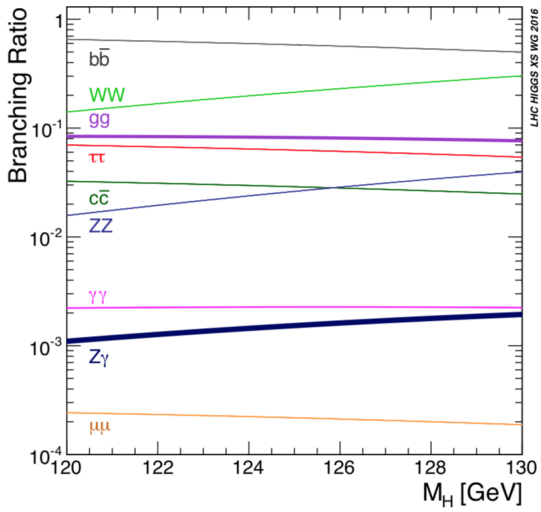


Figure 1.3: The branching ratios for the main decays of the SM Higgs boson near $m_H = 125$ GeV [7].

1.7 Motivation

The discovery of the SM-like Higgs boson (H) by the ATLAS [8] and CMS [9] experiments in 2012 was one of the most significant milestones in the history of particle physics, but the mystery of Electroweak Symmetry Breaking (EWSB) does not end with it. A full understanding of the Higgs potential is still incomplete without the measurement of Higgs self-couplings (λ). The direct measurement of λ requires Higgs pair production, in the SM the cross-section for Higgs boson pair-production (Figure 1.4) is small, 33.41 fb at $\sqrt{s}$=13 TeV. Even with the full projected luminosity (3000 fb^{-1}) of the High-Luminosity LHC (HL-LHC) this process will be hard to detect. A deviation in the strength of λ from its SM value that can significantly enhance Higgs pair production opens door to physics Beyond the SM (BSM). BSM theories can predict heavy resonances such as a heavy spin-0 scalar, X, that could decay to a pair of SM Higgs bosons of mass m_H=125 GeV, in two Higgs doublet models (2HDM) [4].

Table 1.4: BR and their relative uncertianty [7].

Higgs decay channels	BR	Uncertainty
$H \to \gamma\gamma$	2.27×10^{-3}	$+5.0\%$
		-4.9%
$H \to ZZ$	2.62×10^{-2}	$+4.3\%$
		-4.1%
$H \to W^+W^-$	2.14×10^{-1}	$+4.3\%$
		-4.2%
$H \to \tau^+\tau^-$	6.27×10^{-2}	$+5.7\%$
		-5.7%
$H \to bb$	5.84×10^{-1}	$+3.2\%$
		-3.3%
$H \to Z\gamma$	1.53×10^{-3}	$+9.0\%$
		-8.9%
$H \to \mu^+\mu^-$	2.18×10^{-4}	$+6.0\%$
		-5.9%

This analysis is a search for resonant heavy scalar (X) in the ATLAS detector at $\sqrt{s}$ of 13 TeV using data recorded in 2018. Here BSM heavy scalar X decays to two SM-like Higgs, H. The two Higgs bosons then further decay to two b-quarks and two τs. These are two decay modes with significant BR of Higgs boson.. Tau can decay to leptons (electrons(e), muons(μ)) and hadrons (π^0,π^-,π^+). This analysis is focused on semi-leptonic final state of two τ's, where one τ decays leptonically and other in hadronic final state. The lepton in the final state helps to trigger the events.

The $\tau_e\tau_{\text{had}}$ (branching ratio = 23%), $\tau_\mu\tau_{\text{had}}$ (branching ratio = 23%) (Figure 1.5) decay channels are considered in this analysis, where the subscript on the τ indicates the nature of its visible decay products (e=electron, μ=muon, had=hadrons).

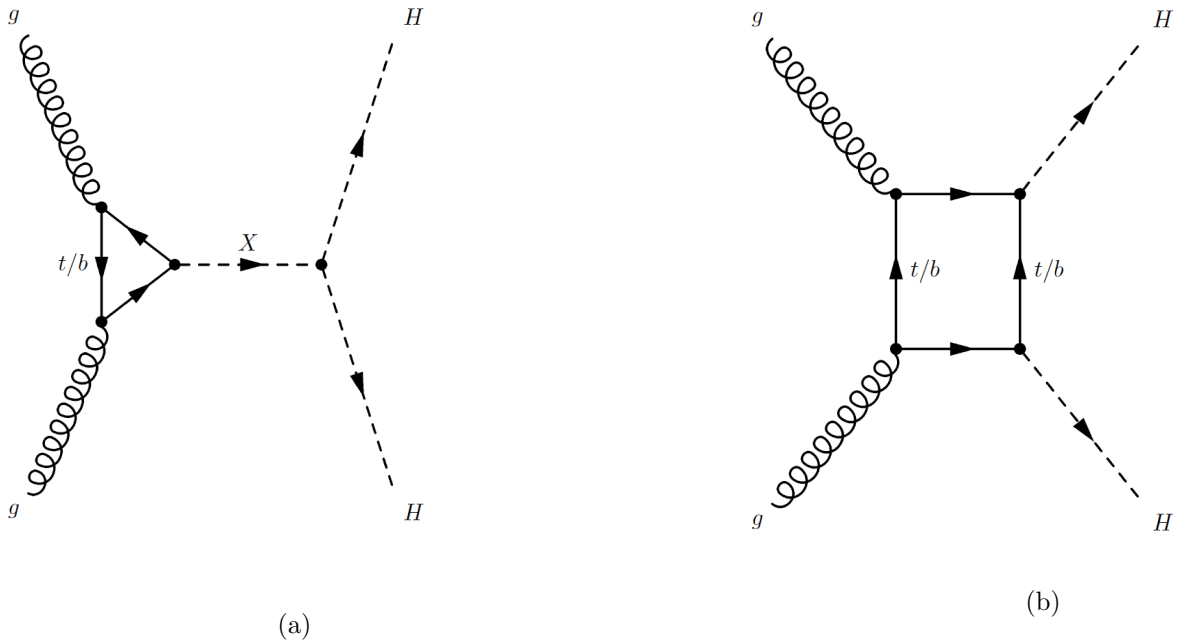


Figure 1.4: Higgs pair production (a) triangular, (b) box diagrams.

1.8 Previous result

The current observed (expected) upper limit on the cross-section of Higgs boson production (in 2HDM) times branching ratio (BR) of $bb\tau\tau$ final state with 36 fb$-$1 data is reported in [11] and shown in Figure 1.6. The mass range 305 GeV $< m_X <$ 402 GeV is excluded at 95% CL for $\tan\beta = 2$, where $\tan\beta$ is the ratio of the vacuum expectation values of the scalar doublets.This result is produced combining semi-leptonic final state and fully hadronic final state of two τs.

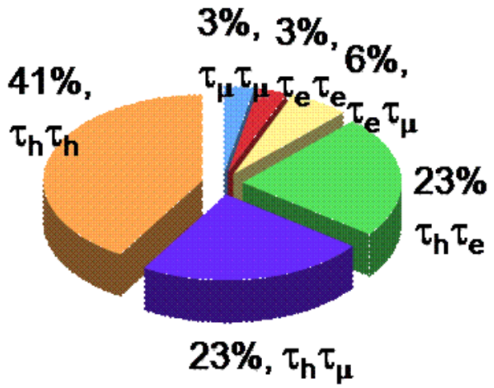


Figure 1.5: Branching ratio of di-τ decay [10].

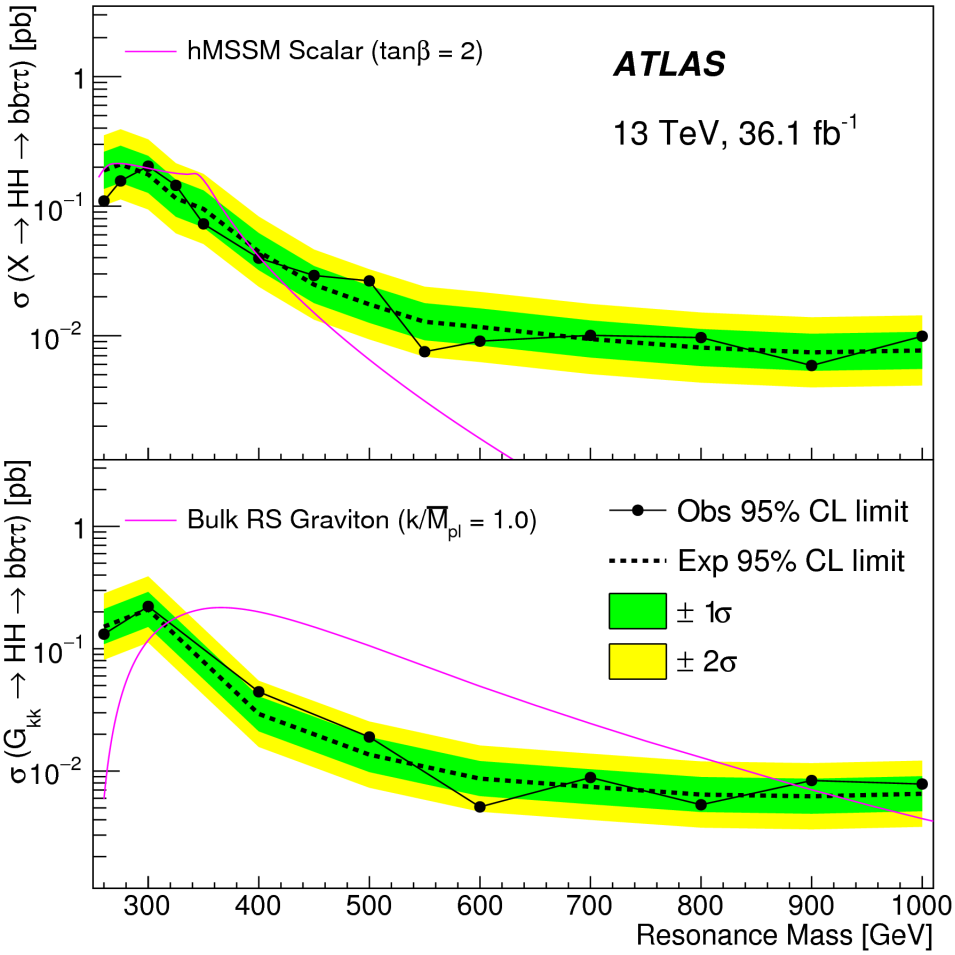


Figure 1.6: Limit on 2HDM and RSG model with 36 fb^{-1} data combining semi-leptonic and fully hadronic final state [11] .

CHAPTER 2

LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) [12], the world's largest particle accelerator, first started operation from 10th September, 2008. It sits in a 27-km long circular tunnel 100 m underground, astride the border between Switzerland and France near Geneva.

The LHC, shown schematically in Figure 2.1, is a proton-proton (pp) collider. The p bunches are produced from the ionization of hydrogen atoms and steered through a series of accelerators. First the protons are accelerated up to 50 MeV in the LINear particle ACcelerator (LINAC). This is followed by passage through the Proton Synchroton Booster (PBS) where they are accelerated to 1.4 GeV. Proton energy is further increased to 26 GeV in the Proton Synchrotron (PS). Finally, the Super Proton Synchrotron (SPS) accelerates the protons to 450 GeV, the minimum energy needed by LHC to maintain stable beam. Protons are injected into two parallel beam pipes, accelerated to 6.5 TeV each in opposite directions, and made to collide at up to four interaction points. To accelerate particles accelerators are fitted with metallic chambers containing an electromagnetic field known as radiofrequency (RF) cavities. Inside this field charged particles receive electrical pulse which accelerates them. The volume inside the beam tubes are kept at ultrahigh vacuum, where protons are guided by superconducting magnets operating at -271.3°C. The superconducting magnets are cooled with liquid helium. Magnets of in a variety of arrangements direct the beams around the accelerator and also "squeeze" them near the interaction point to maximize the number of collisions. The proton beams collide at four different points where the detectors are situated.

An important feature of the LHC is the instantaneous luminosity ($\mathcal{L}_{ins}$). It means how many particles we are able to squeeze through a given space in a given time. This depends on the beam parameters :

$$\mathcal{L}_{ins} = f.\frac{{N_b}^2 n_b \gamma_r}{\mathcal{E}_n \beta^* A}, \tag{2.1}$$

where f is the frequency of bunch crossings, N_b the number of particles per bunch, n_b the number of bunches per beam, γ_r the relativistic gamma factor, $\mathcal{E}_n$ the normalized transverse beam emittance, β^* the optical beta function at the collision point and A is the effective cross-sectional area of the beams, with the angles of the beam collision taken into account. The rate of collisions is defined as:

$$\frac{dN}{dt} = \mathcal{L}_{ins}\sigma, \tag{2.2}$$

where, $\frac{dN}{dt}$ is the event production rate in the LHC collisions and σ is the cross-section of the process in question.

In Run 1 LHC was run at 7 TeV and 8 TeV center-of-mass (CM) energy in 2010-2011 and 2012 respectively. The LHC experiments began taking 13 TeV *pp* collision data with a nominal bunch-spacing of 50 ns in July 2015. A month later before the LHC progressed to 25 ns bunch-spacing.

The integrated luminosity is a measure of the recorded data in the certain period of time. Accumulating large amounts of data is essential to the LHC's success to discover rare SM processes or hypothetical new particles. The LHC running conditions in Run 2 is summarized in Table 2.1 [14].

There are four main experiments ongoing in the LHC : ATLAS [15], CMS [16], LHCb [17] and ALICE [18]. There are also two other smaller experiments TOTEM [19] and LHCf [20] located along the LHC ring.

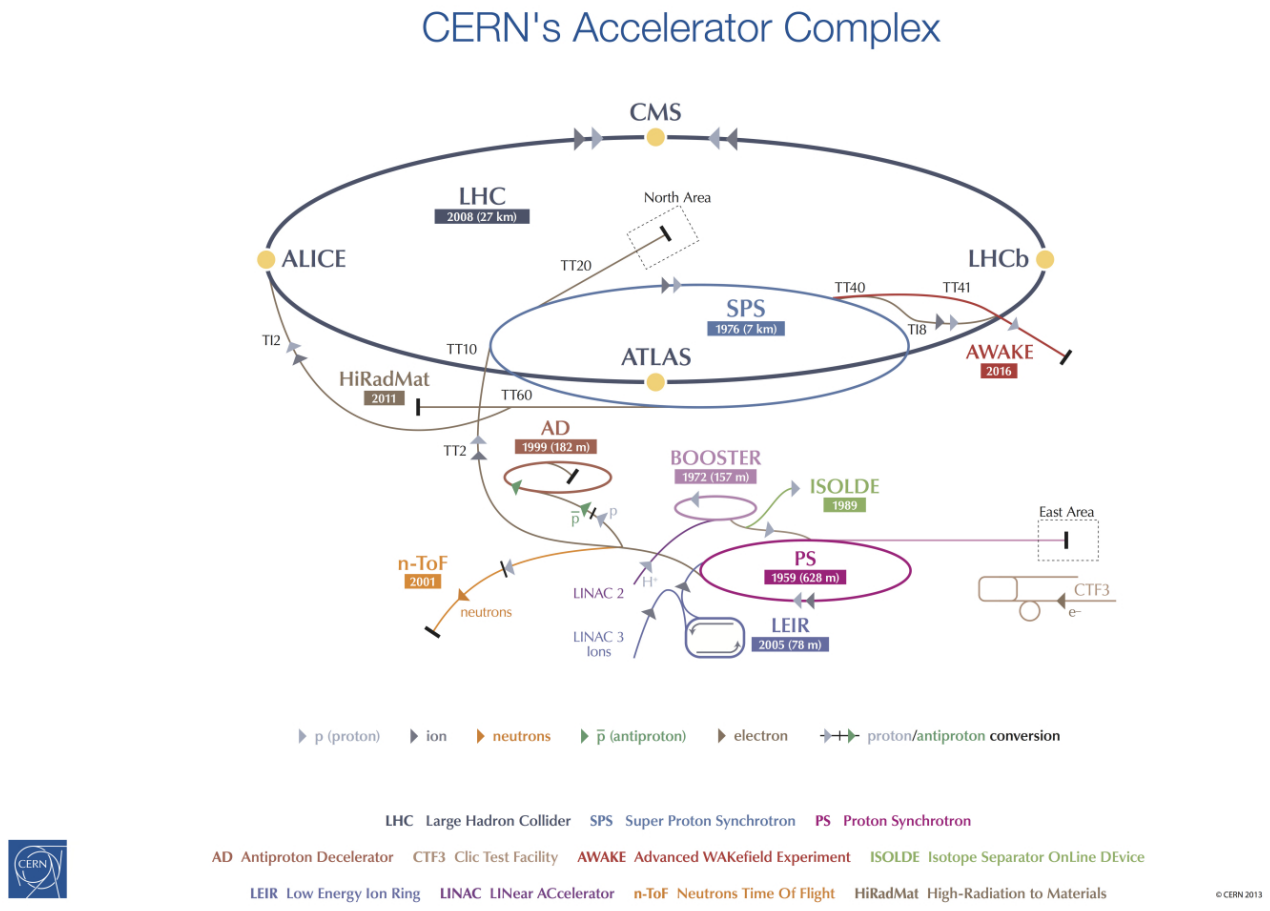


Figure 2.1: The Accelerator complex at CERN [13].

Table 2.1: Machine parameters in Run 2.

Parameters	Design
Particle used	Protons or heavy ions (Pb)
Circumference (m)	26659
Magnetic field at 7 TeV (Tesla)	8.33
Bunch crossing frequency at each interaction point (MHz)	31.6
Power consumption (MW)	∼180
Beam energy (TeV)	7.0
Proton/bunch (average at start of collisions), N_b ($\times 10^{11}$ p)	1.15
Maximum number of bunches, n_b	2808
Maximum stored energy per beam (MJ)	362
Bunch spacing (ns)	25
Transverse normalised emittance, $\mathcal{E}_n$ (μm)	3.75
Smallest allowed magnet aperture, σ	8.4
Amplitude functionβ^*(m)	0.55
Maximum peak luminosity ($10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	1.0

A Toroidal LHC ApparatuS (ATLAS) and the Compact Muon Solenoid (CMS) are two multipurpose detectors with the same broad physics charter but differ in choices of technology and, in some cases. These detectors are designed to scrutinize the SM, and look for physics beyond the SM, at the energy frontier. In this analysis uses the data recorded by the ATLAS detector, a detailed description of which will be presented in the next chapter. Many subsystems in the ATLAS detector have sensitivity window larger than 25 ns, this is the interval between pp bunch crossings. The average number of interactions per bunch crossing ($< \mu >$) increased from 13.4 (2015) to 36.1 (2018). So the physics objects are affected by the pile-up (pp collisions in addition to collisions of interest) in some ways. The in-time pile-up is the additional pp collisions occurring in the same bunch crossing as the collision of interest. The out-of-time pile-up is the additional collisions occurring in bunch crossing before and after the collision of interest [21].

LHCb is an experiment designed to study heavy flavour physics. It focuses mainly on new physics in CP violation and rare decays of bottom and the charm quarks.

ALICE is a detector dedicated to study physics of strongly interacting matter at extreme temperature and energy density in collisions of heavy (Pb) ions, where a new phase of matter and the quark-gluon plasma play a dominant role.

CHAPTER 3

ATLAS DETECTOR AT THE LHC

3.1 Design Consideration of a General Purpose Detector

The main motivation to build a general purpose detector is to detect particles of different types (charged, neutral, hadrons, electromagnetic) through their footprint left at different component of the detectors. At LHC, ATLAS and CMS are two general purpose detectors, they have similar detector components (calorimeter, tracker, magnet) but the material of detectors are of different types with different detection efficiency.

A general purpose detector has a tracker which measures the momentum of the charged particles. A tracker can not produce visible track of the particles but produces electrical signal generated by the particle moves through it. This particle tracker is usually placed inside a strong magnetic field. The path of high momentum particle is relatively straight and remains minimally affected by the magnetic field but low momentum particle tracks are bent inside the magnetic field, that helps measuring the particle momentum. The direction of bending also gives the charge of the particle.

The energy of a particle is measured by the calorimeter. Calorimeter generally has an"active" and a "passive" medium. The "active" medium absorbs the secondaries and other daughter objects created by the particles interactions with the "passive" medium. Particles which interact through electromagnetic interaction such as electrons and photons deposit energy in electromagnatic calorimeter and hadrons (containing quarks) interact via strong force and deposit energy in the hadronic calorimeter.

Weakly interacting neutrinos do not interact with the detector material at all so they can not be detected. Muons due to Lorentz boost can travel long distance before they decay so muon spectrometer is the outer most layer of a detector.

This analysis uses the data recorded by the ATLAS detector, and its components are described in detail below.

3.2 ATLAS Detector

ATLAS (Figure 3.1) [15] is one of two general purpose detectors at the Large Hadron Collider (LHC), sitting in a cavern 100 meters below in Geneva, Switzerland. This 7000 tonne detector detects particle produced in proton-proton (pp) collision in the LHC ring. It has a dimension of a cylinder, 46m long, 25m in diameter. It is a multi-purpose detector that measures the momentum and energy of the particles produced in pp collisions. The magnet system allows to bend the trajectory of the charged particles needed for precision measurements of momenta of charged particles. The co-ordinate system of the ATLAS detector is given by Figure 3.2.

In the sections below the main components of the ATLAS detector are discussed:

3.3 ATLAS Magnet System

The ATLAS magnetic system (Figure 3.3) is built in a central solenoid magnet and three toroidal magnets in barrel and endcap region. Solenoid magnet is located between the inner detector and electromagnetic calorimeter.

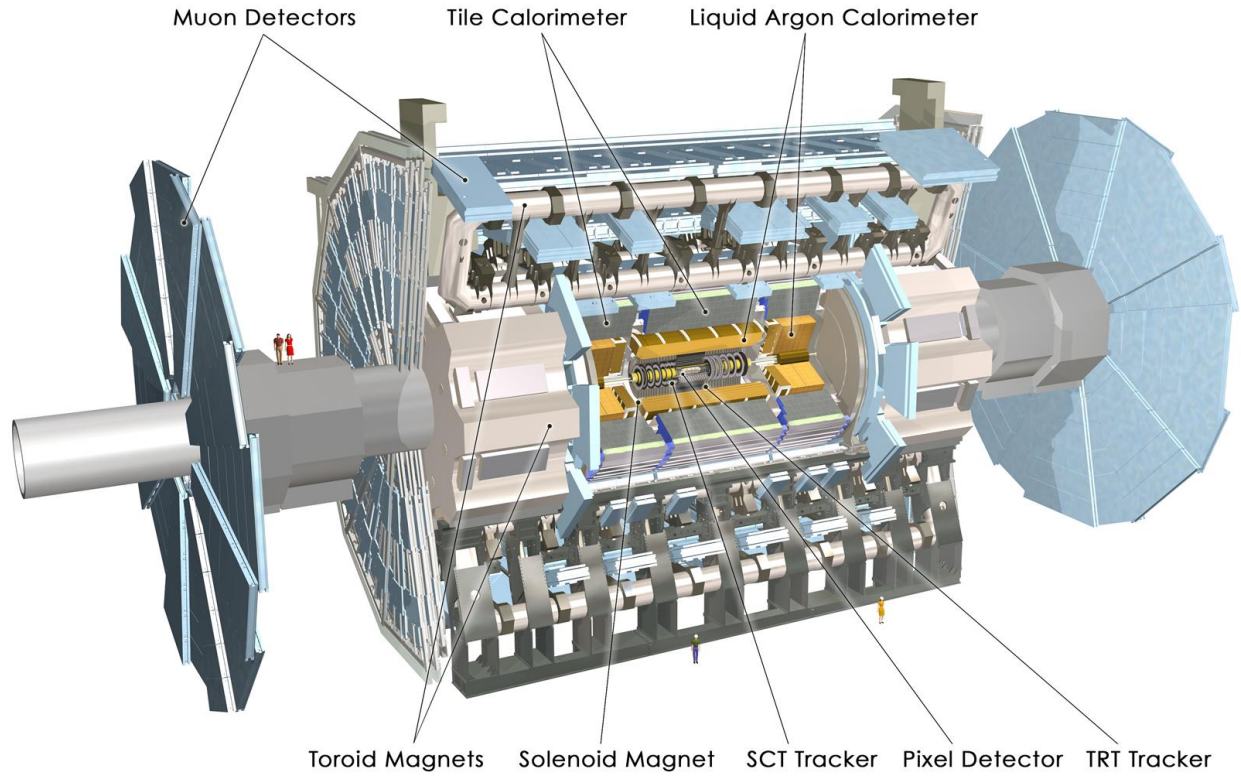


Figure 3.1: The ATLAS detector [26].

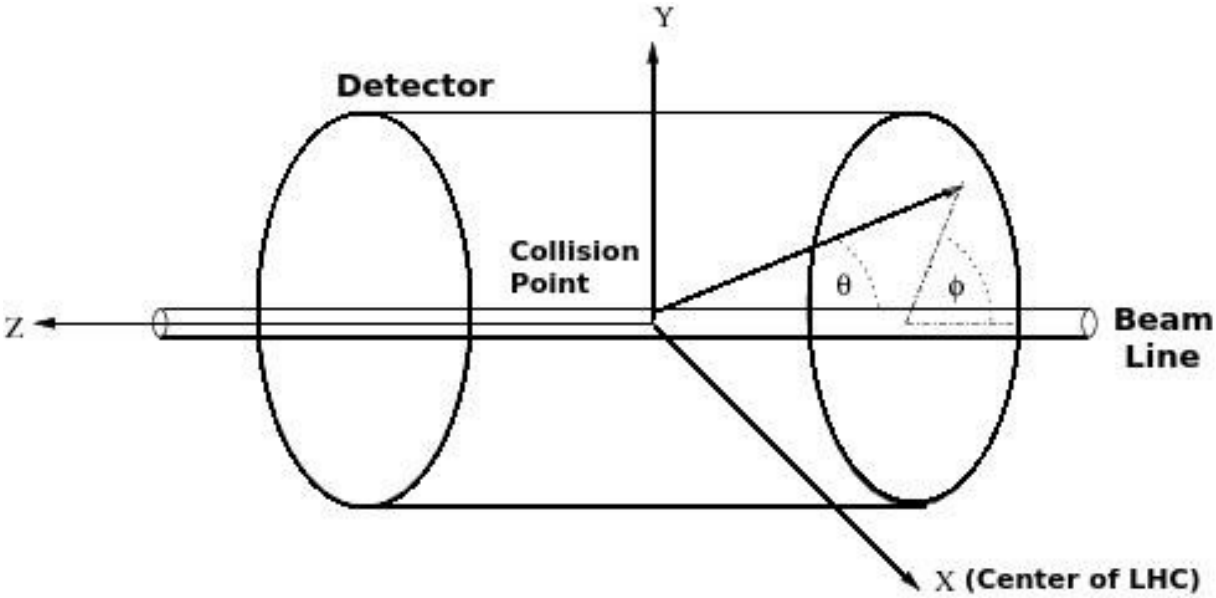


Figure 3.2: Co-ordinate system of the ATLAS detector [22].

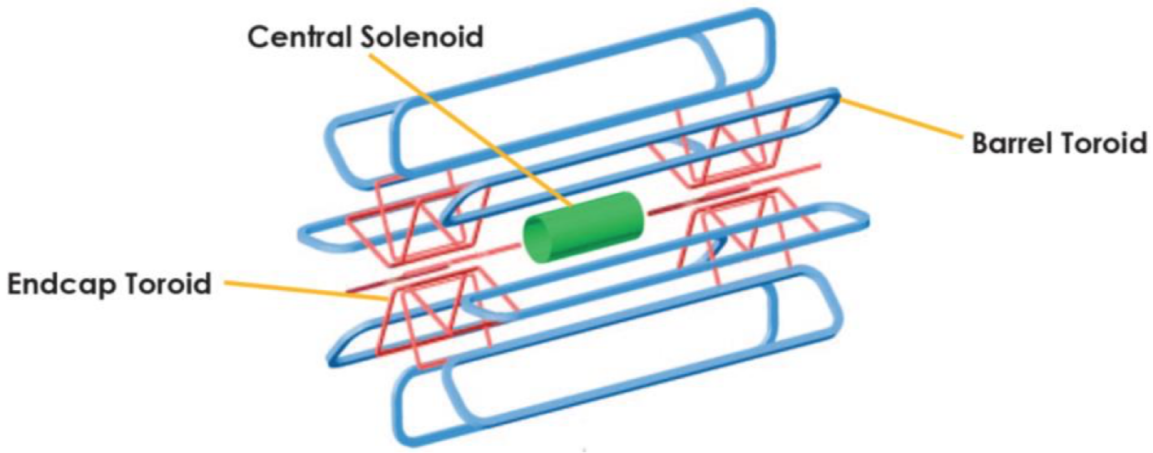


Figure 3.3: Schematic diagram of the ATLAS magnet system [23].

3.3.1 Solenoid

The ATLAS solenoid magnet [24] is a superconducting magnet, 5.3m long, 2.4m in diameter and 4.5 cm thick. It is aligned along the beam axis needs 7.73 kiloampere nominal current to provide 2T of magnetic field with a stored energy of 38 megajoules. The flux is returned by the steel of the ATLAS hadronic calorimeter and its girder structure. The solenoid is charged and discharged in about 30 minutes. In the case of a quench, the stored energy is absorbed by the enthalpy of the cold mass which raises the cold mass temperature to a safe value of 120 K maximum. Re-cooling to 4.5 K is achieved within one day.

3.3.2 Toroid

One barrel and two end-cap toroid magnets [25] produce 0.5T and 1T magnetic fields in the muon system in barrel and end-cap region, respectively. The toroidal magnet is cylindrical in shape and located behind the calorimeters. The barrel toroid has eight coils

encased in individual racetrack-shaped, stainless-steel vacuum vessels. The overall size of the barrel toroid system is 25.3 m in length, with inner and outer diameters of 9.4 m and 20.1 m, respectively.

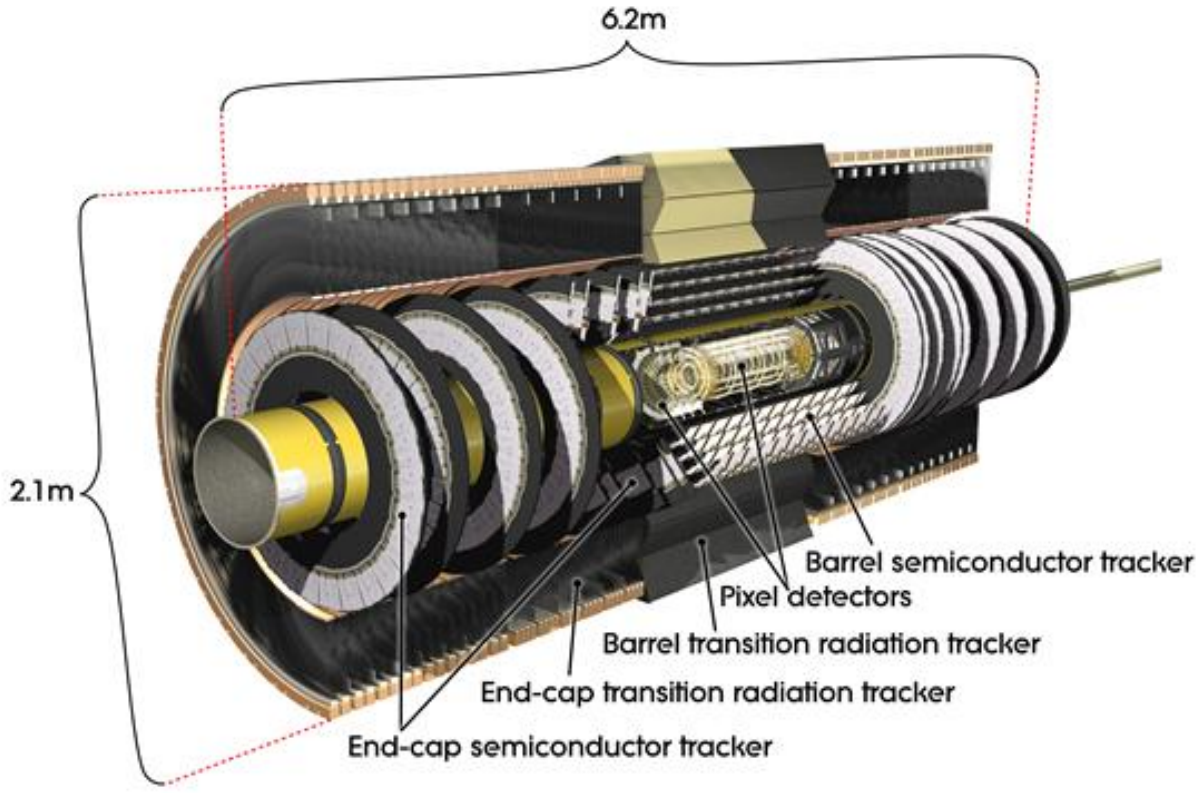


Figure 3.4: The ATLAS Inner detector system [28].

3.4 Inner Detector

The Inner Detector (ID) (Figure 3.4) [27] is the first component of the ATLAS to see the daughter particles produced in pp collisions. The ID is mainly built with three parts inside the superconducting solenoidal magnetic field: Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT).

3.4.1 Pixel Detector

The pixel detector [29] is the innermost element of the Inner Detector just outside the LHC beam pipe, it contains approximately 80 million channels and provides pattern recognition capability to reconstruct track of the charged particles at the full luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. It also helps to identify the secondary vertex needed for the b-jet reconstruction. It has pseudorapidity ($\eta = -\ln[\tan\frac{\theta}{2}]$) range $|\eta| < 2.5$ and transverse impact parameter resolution of better than 15 microns. It has very good b-tagging capabilities both in the high-level trigger and in the offline reconstruction. The ATLAS pixel detector has 1,744 modules with 46,080 readout channels per modules. It has three barrel layers and three disks in each end-cap layer.

3.4.2 SCT

SCT provides space-point information that is useful for vertex detection, pattern recognition, momentum and impact parameter measurements. This silicon micro strip tracker is placed behind the pixel detector and consists of 4,088 two sided modules and over 6 million implanted read out strips. 60 m^2 of silicon is distributed over four cylindrical barrel and 18 planar end cap discs. The read out strips on the silicon measures the position of charged particle to an accuracy of 17 μm per layer in the transverse direction of the beam pipe.

3.4.3 TRT

TRT [30] is a straw tube based tracker behind SCT. Each straw is 4 mm in diameter and 150 cm in length, it has a gold plated tungsten wire at the central region of the tube. The barrel region has 50000 straws and the end cap has 320000 radial straws. The straws are filled with a gas mixture Xe, CO_2, O_2. When a charged particle passes through the straw it ionizes the gas and generates free electrons which move towards the wire where they get amplified and read out. The space between two straws is filled with polymer fiber in barrel and with foils in end cap. The transition radiation created by charged particle when crossing the material boundary has intensity dependence on $\gamma = \frac{E}{m}$ and is strongest for particles of low mass.

3.5 Calorimeter

Calorimeter measures the energy of a particle by absorbing or stopping the particle and forcing them to deposit energy within the detector. Calorimeter also contain the particle shower.

3.5.1 Liquid Argon Calorimeter (LArg)

ATLAS LArg is an electromagnetic (EM) calorimeter [31] measures the energy of electrons and photons as they interact with matter. In LArg calorimeter, lead/stainless and liquid argon layer are inter spaced, short radiation length of lead provides information about shower development and secondary electron produces ionization in liquid argonne layer. An

inductive signal from the ionization electrons drifting in the electric field across the gas-gap is registered by copper electrodes.

LArg (Figure 3.5) has two electromagnetic wheels in the middle, at each end it has two concentric electromagnetic wheels, two hadronic wheels and three forward calorimeter wheels. The energy resolution for the LArg hadronic and forward regions are $50\%/\sqrt{E}$ in $|\eta| < 3$ and $100\%/\sqrt{E}$ in $3 < |\eta| < 4.9$. The energy resolutions for the EM shower is $10\%/\sqrt{E}$. The EM barrel is 6.4m long, 53 cm thick and it has 110,000 read out channels.

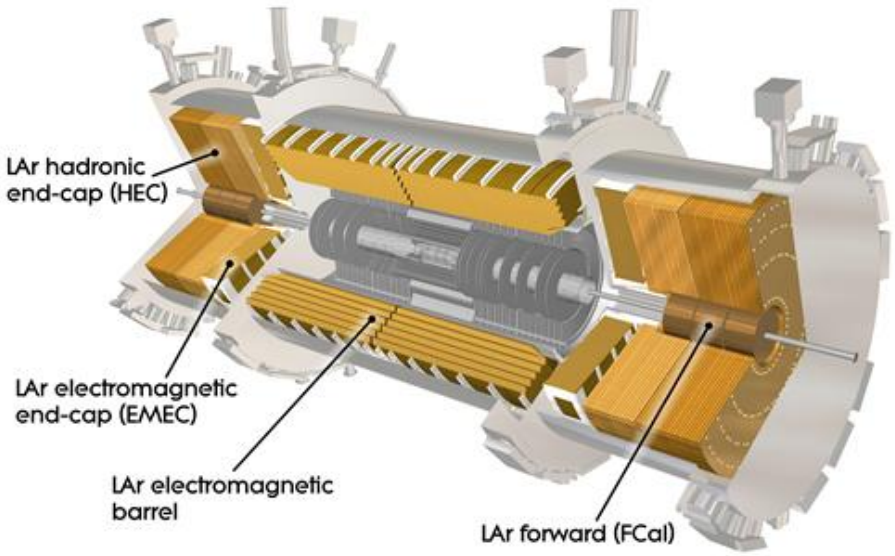


Figure 3.5: Schematic diagram of the ATLAS LArg calorimeter system [32].

3.5.2 Tile Calorimeter (TileCal)

The ATLAS Tile calorimeter (Figure 3.6) [33] is a steel-scintillator sampling hadronic calorimeter covering the central part of the detector ($|\eta| < 1.7$) and situated behind LArg.

TileCal is a key detector for the measurement of hadrons, jets, taus, missing transverse energy. Numerous particles produced in collisions in the LHC produce light in scintillator tiles, generating analog pulses in photomultiplier tubes. The amplitude of the analog pulse is proportional to the energy deposited in the active medium. The TileCal is divided in four sections along Z (beamline) : the long barrel (LB) and extended barrel (EB) on either side (A and C) of the interaction point. Each of the TileCals four partitions is comprised of 64 modules in ϕ, each with three radial layers (A, BC, D). In total, there are 5198 scintillators cells each read out with two PMTs. Additionally, the gap and crack regions are instrumented with special C10 cell and E cells, which have one PMT. Each module has 48 channels, some of which are not read out.

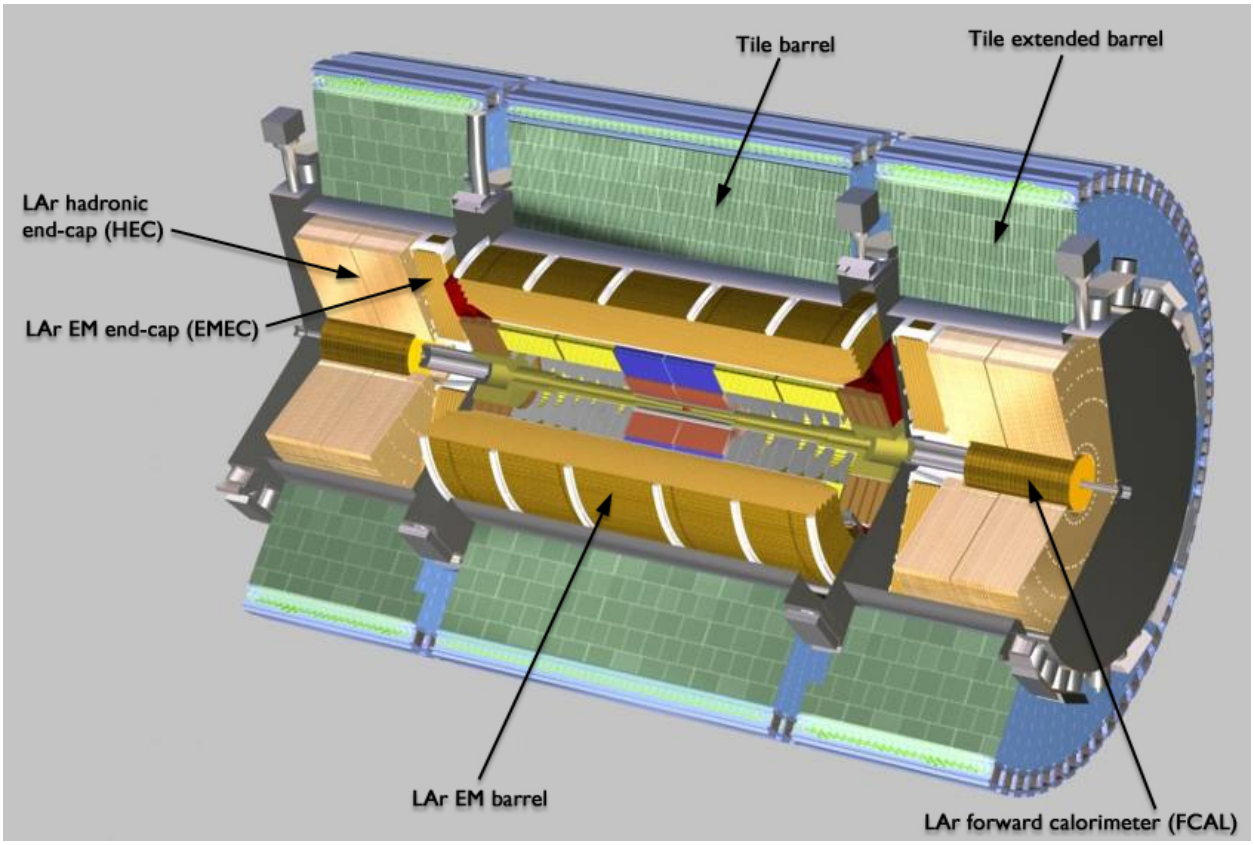


Figure 3.6: Schematic diagram of the ATLAS Tile calorimeter system [35].

3.6 Muon System

The Muon Spectrometer (MS) (Figure 3.7) [34] is the outer most part of the ATLAS detector, measures energy and trajectory of muon of momentum at 1 TeV to an accuracy of 10%. Muon system consists of Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC). The magnetic field in the muon spectrometer is provided by three air-core toroidal magnets, each consist of eight coils. This magnetic fields cause the track to bend in η, the sagitta of this bending curve is the main variable to calculate the momentum of the muon.

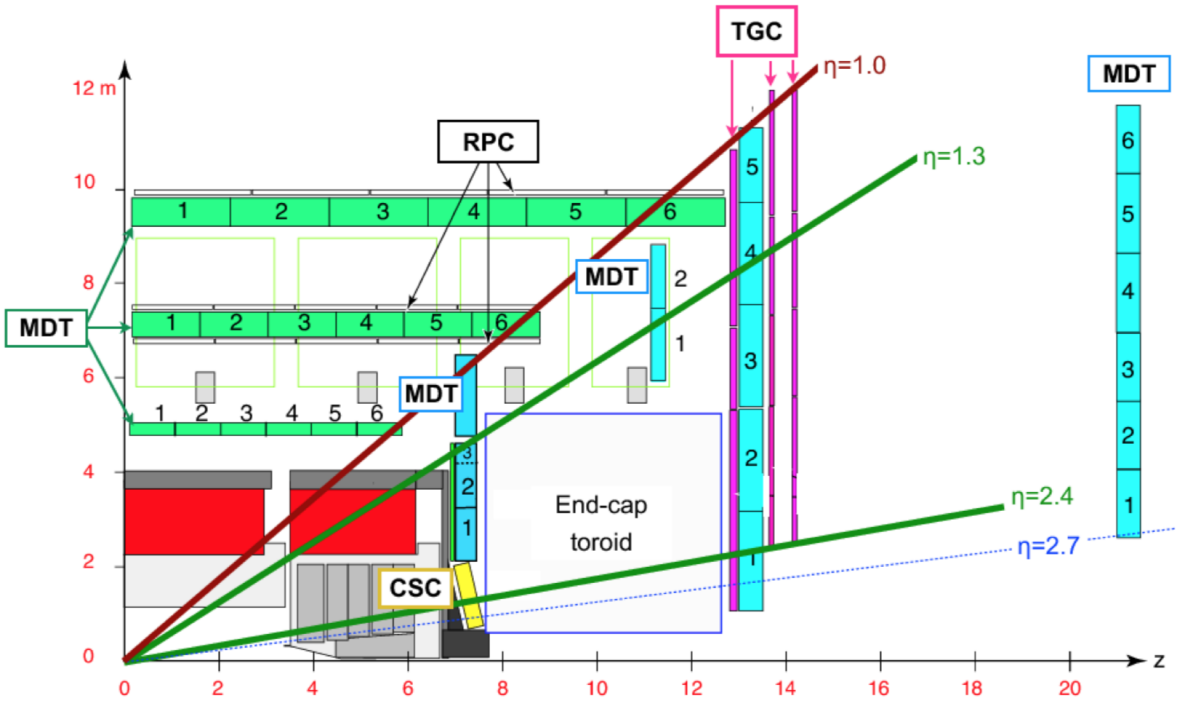


Figure 3.7: Schematic diagram of the ATLAS Muon system [36].

3.6.1 RPC

The RPC is a three doublet layers within $|\eta| < 1.05$ mainly used to trigger on events with muons. The RPC consist of three chambers that are in the middle part of MS provide the trigger for low p_{T} muons and outer chamber station triggers the high p_{T} muons. These two layers are 2 mm thick, bakelite laminated and kept apart from each other by 2 mm insulating material. The space between these two layers are filled with gas mixture $C_2H_2F_4$, C_4H_{10} and SF6. RPC works in avalanche mode, the discharge electron drifts the gas and the signal induced in the copper strip further read out via front-end-electronic system.

3.6.2 TGC

The TGC is located in the middle tracking station, has one triplet layer followed by two doublet layers within $1.0 < |\eta| < 2.4$. The cathode strip in the gas chamber is coated with graphite and the gas used is 55%-45% mixture of carbon dioxide and n-pentane. TGC operates in high gain mode and provides large saturated signal with respect to the mechanical noise. TGC also provides narrow width signal due to the narrow spacing between adjacent wires. Thin Gap Chamber is mainly for triggering second co-ordinate detection at the end region, where Resistive Plate Chamber does the same job in the central region.

3.6.3 MDT

The muon trajectories in the muon spectrometer are measured with three layers of MDT covers up to $|\eta| < 2.7$. The MDT chambers are built of two quadruple layers of cylindrical

drift tubes in the innermost layer and of two triple layers of drift tubes in the remaining two layers. The tubes have an outer diameter of 30 mm and a 400 μm thick tube walls made of aluminum. The anode wire of 50 μm diameter is set to a voltage of 3080 V with respect to the tube walls leading to a gas gain of 20,000 and an average spatial single-tube resolution better than 100 μm. The chamber is filled with Ar-CO_2 (93%-7%) and kept at a constant pressure. When muon pass through the gas, it is ionizes and generates free electrons which move towards the anode.

3.6.4 CSC

Precision tracking at the inner-most station in the high pseudorapidity region, $2.0 < |\eta| < 2.7$, is performed by 16 four-layered CSC on each end-cap. These are multi-wire proportional chambers, wires are oriented in the radial direction, and multiple cathode strips oriented at a right angle with respect to the anodes. Each chamber is filled with an Argon-based gas mixture. Muons passing through the CSC ionize the gas and the electrons are attracted to the anodes. Interaction of electrons with the gas causes an electron avalanche that is read out at the anode. The positively charge ions, also produced by the avalanche, are attracted to the cathode strips. The orientation of the anode wires and cathode strips allows for the position of the muon to be reconstructed in the two dimensional plane of the end-cap.

3.7 The ATLAS Trigger System

The ATLAS Trigger system [37] is composed of two sequential steps: a hardware-based Level-1 Trigger System and a CPU-based High Level Trigger (HLT) system with its Level-2. The Level-1 trigger uses information from the calorimeter and the muon spectrometer and

provides Regions-of-Interest (RoIs). Event rates from the LHC bunch crossings decreases from 40 MHz to 100 kHz after the Level-1 trigger selection, which takes 2.5 μs. The RoIs passing Level-1 go to HLT where the rate is further reduced from 100kHz to 1kHz in 200ms.

Level-1 calorimeter trigger system uses coarse segmentation in the electromagnetic and hadronic calorimeters. This information is needed to flag electron (e), muon (μ), photon (γ), tau (τ) and jet candidates and to estimate the missing transverse momentum ($E_{\mathrm{T}}{}^{\mathrm{miss}}$). In Run 2 there is a new introduction in Level-1 trigger, L1Topo, which allows topological selection based on angular separation, invariant mass and sum of transverse momenta of Level-1 trigger objects.

CHAPTER 4

OBJECT RECONSTRUCTION

The online event selection and filtering is done by the ATLAS trigger system. The LHC instantaneous luminosity in 2018 at 13 TeV center of mass energy is given by Figure 4.1 and bunch-crossing frequency is 40 MHz. For every bunch-crossing around 23 inelastic proton-proton collisions are expected, which means $O(10^9)$ interactions per second. Considering the data storage capability and event size to the trigger output rate of 200 Hz, the ATLAS trigger system can only accept one bunch crossing out of 200,000, on average. In attaining such rejection, it is vitally important to maintain high efficiency for interesting events.

4.1 Object Reconstruction

The algorithms and software that are used in the HLT event selection are also extensively used in the offline object reconstructions. The reconstruction of different physics objects in the ATLAS detector are discussed in detail in next few sections.

4.1.1 Electrons

Electrons in the central region of the ATLAS detector are reconstructed from isolated energy deposits in the electromagnetic calorimeter and that are subsequently matched with isolated tracks found in the inner detector [39]. Electromagnetic cluster in the barrel of size corresponding to 3×7 cells in the middle layer is used (equivalent to an area of size

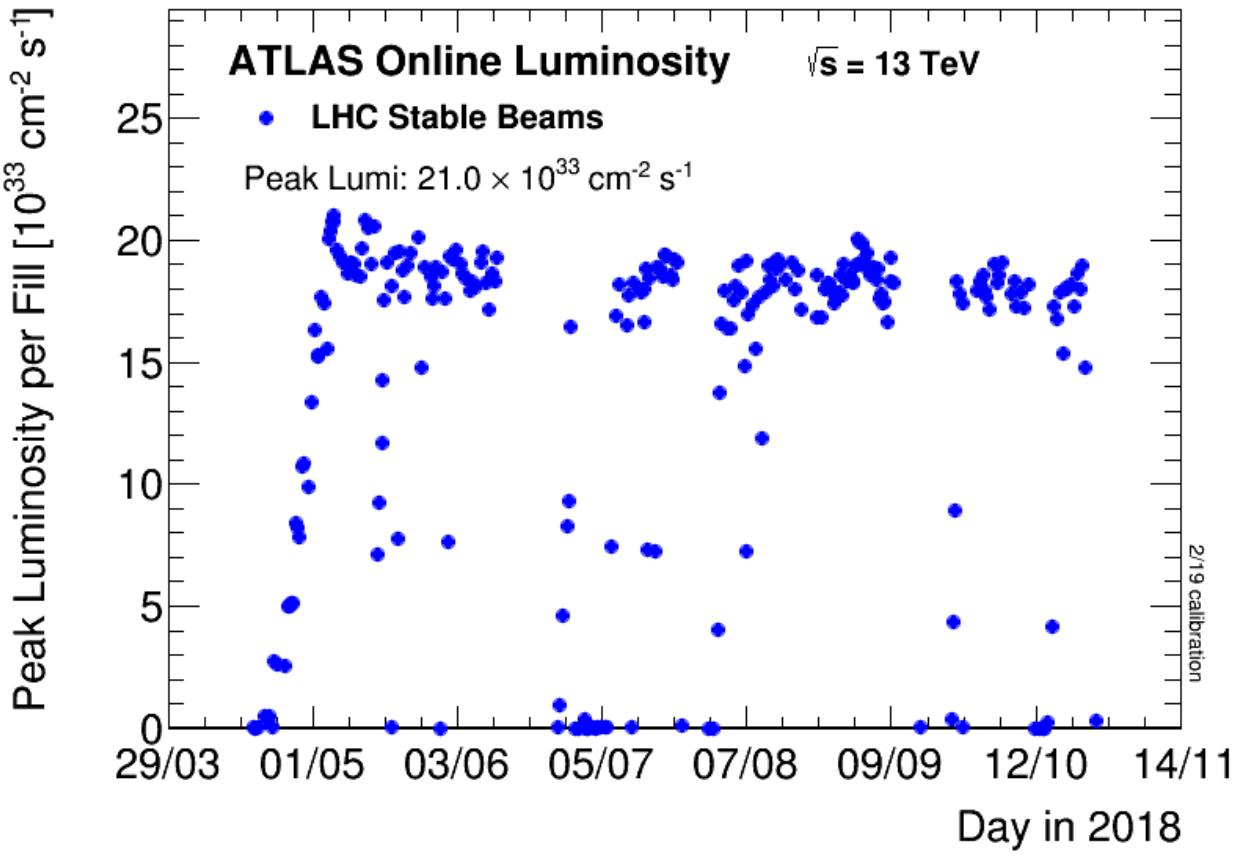


Figure 4.1: ATLAS online luminosity in 2018 [38].

$\delta\eta \times \delta\phi = 0.075 \times 0.175$), while a cluster of 5×5 cells in the middle layer is used in the endcaps (equivalent to an area of $\delta\eta \times \delta\phi = 0.125 \times 0.125$). Tracks reconstructed in the inner detector should matched to the electromagnetic clusters, are selected, with additional minimum number of hits in the pixel and overall silicon layers. To define loose, medium, and tight working points particular cuts are applied on the discriminating variables [39] that depend on the $|\eta|$ and transverse momentum p_{T}. The electron identification efficiency in the central region (Figure 4.2) at the loose, medium, and tight working points are 95%, 90% and 85%, respectively, on average. For this analysis "loose" working point is used. For this analysis the"loose" working point the requirement on the number of hits in the innermost layer of tracker are removed.

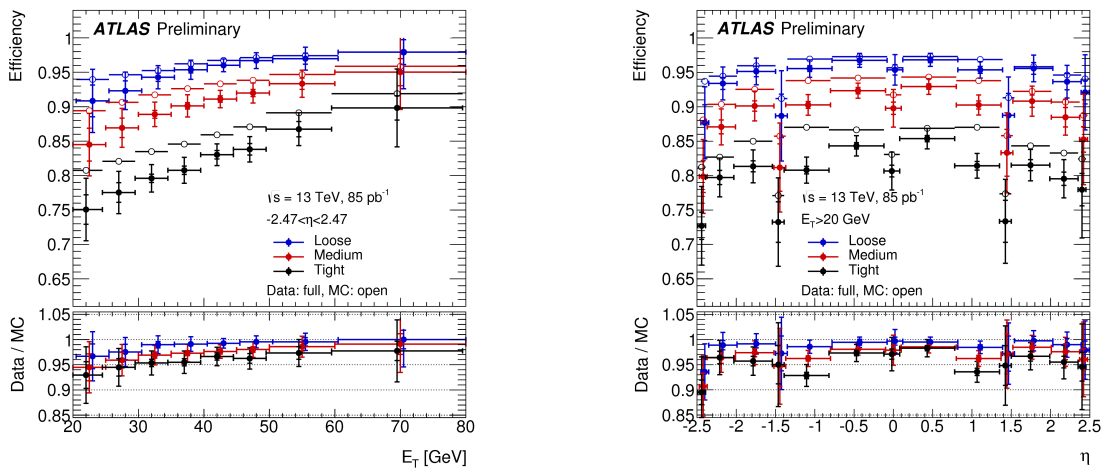


Figure 4.2: Electron identification efficiency in $Z \to ee$ events as a function of E_{T} and η [39].

4.1.2 Jets

The fine granularity in the ATLAS calorimeter reconstructs jets with high precision. The calorimeter cells are grouped to three dimensional clusters if topologically connected; these cells are called Topoclusters. The first step to build a Topocluster is to identify cells with energy deposits $E > 4\sigma$, where σ is the noise defined as a sum in quadrature of electronic and pile-up noise. The second step is the iteration within the neighboring cells with $E > 2\sigma$ to the seeds. In the third step, an extra layer of cells with $E > 0$ on the perimeter of the clustered cells are added. The standard jet finding algorithm that is used in the ATLAS experiment is the anti-kt clustering algorithm with distance parameters R = 0.4 [41]

4.1.3 Taus

Tau (τ) are heavier than electron and muon in the lepton group and have a small lifetime so they can only be detected from their decay products. A τ decays to either a leptonic final

state or hadronically. Hadronic final state of a tau decay could be in association with one or more π^0. A leptonic final state contains e/μ and neutrinos. As discussed earlier a neutrino (invisible) does not interact with detector material and so remains undetected. leptons from τ decay are not isolated leptons. In this section the reconstruction, energy calibration and identification of hadronically decaying tau (τ_{had}) in the ATLAS calorimeter is discussed [42].

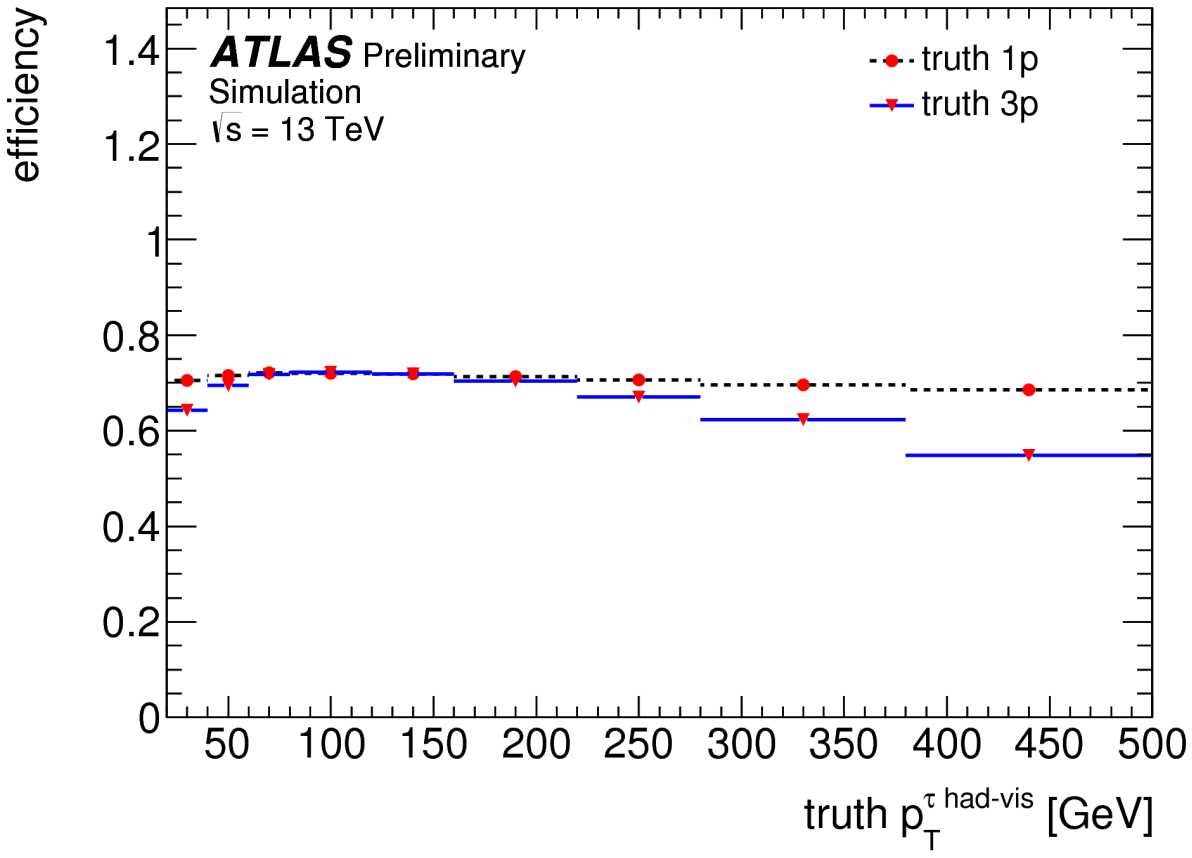


Figure 4.3: Tau reconstruction efficiency as a function of truth p_{T} [42].

The visible part (hadrons) of a hadronically decaying tau ($\tau_{\text{had-vis}}$) is reconstructed using Anti-Kt jet algorithm with distance parameter R=0.4 and calorimeter cluster, calibrated with local hadronic calibration. The momentum of $\tau_{\text{had-vis}}$ is set to the total energy deposited in $\Delta R < 0.2$ of the TopoCluster. The hadronic component of a TopoCluster is calculated from cells in the last layer of the electromagnetic calorimeter and the entire depth of the

hadronic calorimeter. The electromagnetic component is calculated from the upstream layer of the EM calorimeter.

The τ vertex algorithm sums the transverse momentum (p_{T}) of the tracks within ΔR= 0.2 of the direction of the seed jet. The primary vertex to which the largest fraction of the p_{T} sum is matched, is chosen as the production vertex of the tau.

Tracks within $\Delta R = 0.2$ around $\tau_{\mathrm{had-vis}}$ are associated with the τ candidate. Only tracks with at least two hits in pixel detector and at least seven hits in total pixel and SCT are accepted. The distance of closest approach in transverse plane must satisfy $d_0 < 1.0$ mm and longitudinally $|\Delta z_0 \sin\theta| < 1.5$ mm.

Figure 4.4 shows the τ reconstruction efficiency, that is the fraction of 1-prong and 3-prong hadronic tau decays that are reconstructed as 1-track and 3-track $\tau_{\mathrm{had-vis}}$ candidates respectively. The reconstruction efficiency for 1-prong tau is almost constant in the tau p_{T} range with a slow drop at high momentum. The reconstruction efficiency of 3-prong tau decays drops noticeably at both low and high momenta. At low momenta, tau decay products often fail to meet the p_{T} threshold and at high momenta the decay products are so collimated that the occurrence of incorrect vertex assignment becomes more frequent, resulting in the drop in efficiency.

The τ reconstruction method described above has very low rejection rate for jet background. Jet rejection is done in the tau identification step where a Boosted Decision Tree (BDT) is used to discriminate between real taus and jets. The discriminating variables used for separate BDT training for 1-prong and 3-prongs decays are discussed in Ref [42]. The $Z/\gamma^* \rightarrow \tau\tau$ sample is used as signal and di-jet events are used as background. The output of BDT training has a p_T and μ (average interaction per bunch crossing) dependence which was further flatten by using a cumulative distribution function. This gives a constant tau identification (Figure 4.5) for each working point ("loose", "medium", "tight") across the whole p_{T} and μ range.

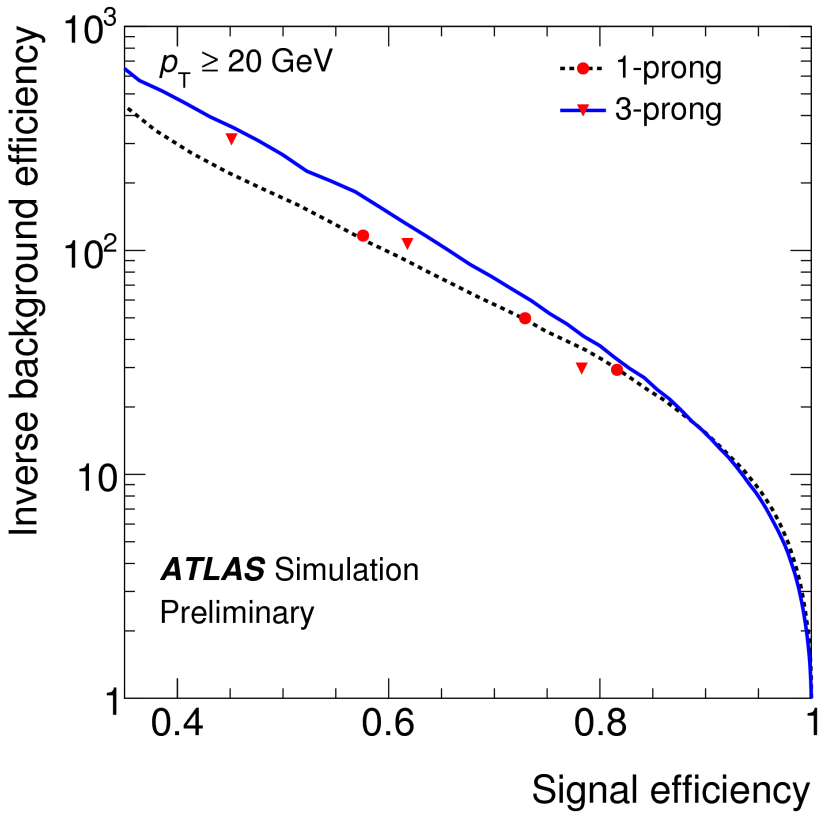


Figure 4.4: Inverse of the efficiency for mis-tagging QCD jets as a function of the identification efficiency for $\tau_{\text{had-vis}}$ candidates. The "loose", "medium", "tight" working points are shown on the curves [42]. The working points do not stay exactly on the line because they implement variable cuts to achieve a reduced p_{T}-dependency of the efficiency.

4.1.4 Muons

Muon reconstruction [43] inside the MS is done by fitting a straight line to the hit patterns inside each muon chamber. The RPC or TGC hits measure the coordinate orthogonal to the bending plane. Segments in the CSC detectors are built using a separate combinatorial search in the η and ϕ detector planes. Four different types of muons are defined based on the combined ID-MS muon reconstruction with different algorithms.

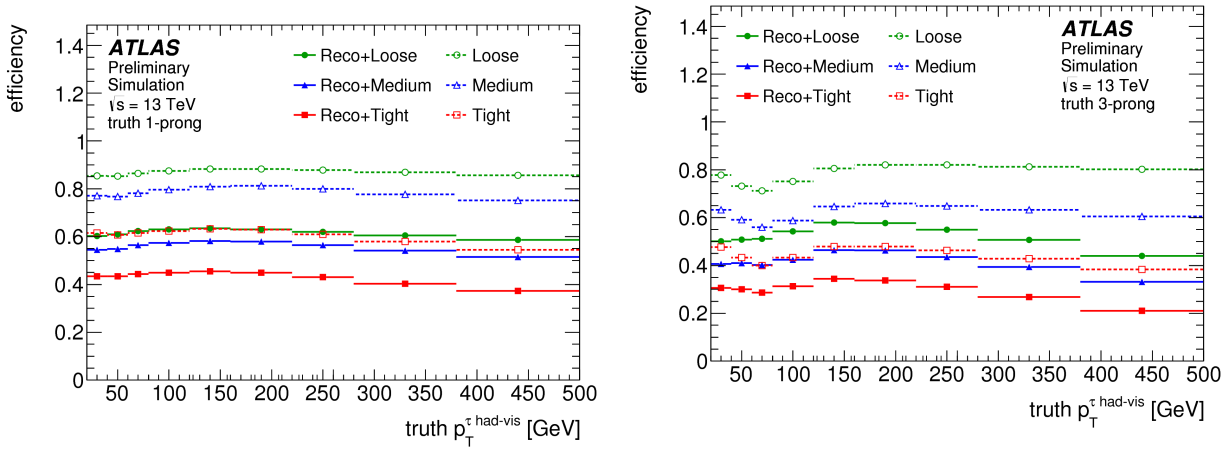


Figure 4.5: Identification efficiency of 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidate in $Z \to \tau\tau$ sample [42].

- Combined (CB) muon: tracks segments are reconstructed independently in the ID and MS, and a combined track is formed by fitting hits from both the ID and MS sub-detectors after matching segments are found.

- Segment-tagged (ST) muon: a track in the ID is considered as muon when an extrapolation to the MS is associated with at least one local track segment in the MDT or CSC chambers.

- Calorimeter-tagged (CT) muon: a track in the ID is classified as a muon if it is matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle. The identification of CT muon is optimized in the region $|\eta| < 0.1$ and momentum range of 15 GeV $< p_{\text{T}} <$ 100 GeV.

- Extrapolated (ME) muons: the muon trajectory is reconstructed based on MS track and a loose requirement on it originating from the interaction point (IP). ME muons are mainly used to extend the acceptance of muon reconstruction into the region 2.5 $< |\eta| <$ 2.7, which is not covered by the ID.

Figure 4.6 shows the reconstruction efficiency of loose, medium and tight muons.

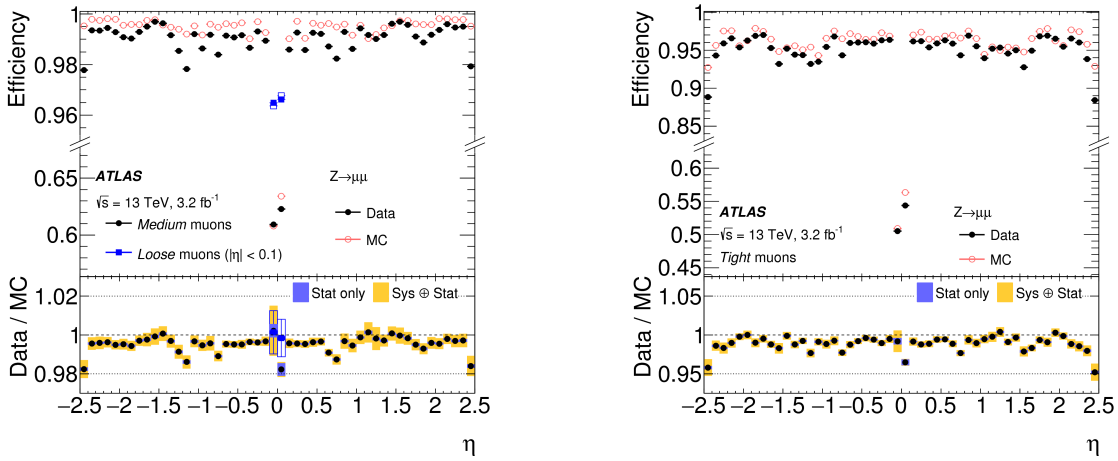


Figure 4.6: Reconstruction efficiency of medium, loose and tight muons in $Z \to \mu\mu$ sample [43]. On the left plot the loose selection (squares) in the region $|\eta| < 0.1$ where the Loose and Medium selections differ significantly.

4.1.5 *b*-jets

Identification of jets coming from b-hadron decay is important in many analyses. The top quark decays most of the time exclusively into a b quark and a W boson, the SM Higgs boson decays most of the time to $b\bar{b}$. The relatively large decay length of a high-p_{T} B hadron means that its decay (secondary) vertex can be resolved from its production (primary) vertex. Three different algorithms are used to identify b-jets in the ATLAS detector: 1. Impact Parameter Based Algorithm (IP), 2. Secondary Vertex Reconstruction Algorithm (SV), 3. A decay chain multi-vertex reconstruction algorithm (JetFitter). The output of these three algorithms are then used as input to a multivariate discriminant to classify jets into different flavors. The transverse impact parameter is defined as the distance of closest approach in the r-ϕ plane of the track to the primary vertex while the longitudinal impact parameter ($z_0 \sin\theta$) is defined as the distance of the track to the primary vertex in the longitudinal

plane at the point of closest approach in r-ϕ . Due to the long lifetime of b-hadrons, tracks generated from b-hadron decay products tend to have large impact parameters enabling their contribution to be separated from the contribution of tracks from the primary vertex. It is possible to assign a sign to the impact parameter by measuring distance relatively to the jet direction, if the primary vertex is in front of the secondary vertex or behind. Secondary vertices behind the primary are normally due to background and are used to help to calibrate the light quark tag rate.

The secondary vertex finding algorithm aims to find an inclusive displaced secondary vertex within a jet. All track pairs within a jet are tested for a two-track vertex hypothesis. Any two track vertices are rejected if they were likely to originate from the decay of a long-lived particle, photon conversions or hadronic interactions with the detector material. Several refinements in the selections of tracks and two-track vertices have been made in order to improve the performance of the algorithm, especially against pileup contamination and for high-p_{T} jets [44].

The input variables that are selected to train the multivariate classifier (MV2) where jets from $t\bar{t}$ events with b-jets being considered as signal, and c and light-flavour jets being considered as background are listed in [44].

The performance of the optimized MV2c00, MV2c10 and MV2c20 b-tagging algorithms is shown in Figure 4.7 for the light and c-jet rejection as a function of the b-jet efficiency in comparison to the 2015 MV2c20 configuration. The name of the algorithm indicates the c-jet fraction in the training the background sample, for example MV2C20 is composed of 20% (80%) c- (light-flavour) jets. It is noted that the current MV2c10 (2016 configuration) discriminant provides a similar light-flavour jet rejection (improvement of approximately 4% at 77% b-jet efficiency) to the 2015 MV2c20 configuration, but a significantly better c-jet rejection of 40% at the 77% b-jet efficiency working point. The 2016 MV2c20 setup provides even better charm rejection, but at the expense of a reduced light-flavour jet rejection, which

may be detrimental to some analyses. The MV2c10 tagger algorithm has been chosen as the standard b-tagging discriminant for this analyses.

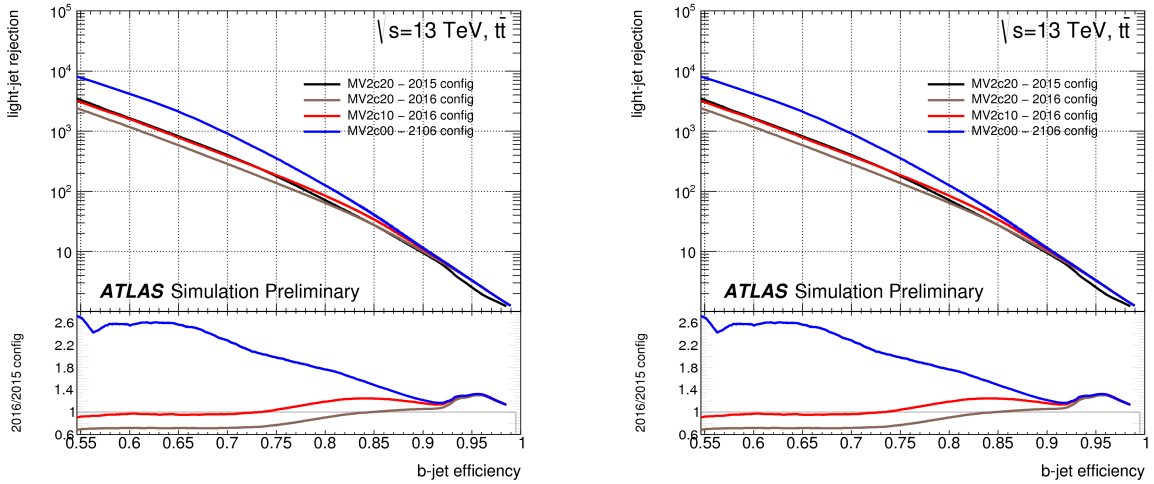


Figure 4.7: Light and c-jet rejection as a function of b-jet efficiency [44].

4.1.6 $E_{\mathrm{T}}^{\mathrm{miss}}$

The conservation of momentum implies that the sum of the total transverse momentum of the particles produced in pp collision will be zero because before collision the sum of transverse momentum of the colliding particle was zero (p-p beams comes along z direction). If the sum of the total transverse momentum is not zero in a any collision that indicates "missing transverse momentum ", or $E_{\mathrm{T}}^{\mathrm{miss}}$ [45], and may be indicative of weakly-interacting, stable particles in the final state. Within the Standard Model, the only such particles are neutrinos but there are other prospective candidates in theories beyond the Standard Model, which makes $E_{\mathrm{T}}^{\mathrm{miss}}$ an important quantity in searches for new physics. Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ can also result from Standard Model particles which escape the acceptance of the detector, are badly reconstructed, or fail to be reconstructed. The reconstructed TST $E_{\mathrm{T}}^{\mathrm{miss}}$ has contribution from "hard term" and track-based "soft term". The hard term is the reconstructed and

fully calibrated physics objects. The track-based soft term contains reconstructed tracks associated with the hard-scatter vertex that are not associated to the hard term. The main difficulty to reconstruct ${E_{\mathrm{T}}}^{\mathrm{miss}}$ is in-time and out-of-time pile up.

The following equation shows the component of ${E_{\mathrm{T}}}^{\mathrm{miss}}$:

$$E_{\mathrm{(x)y}}^{\mathrm{miss}} = E_{\mathrm{(x)y}}^{\mathrm{miss,e}} + E_{\mathrm{(x)y}}^{\mathrm{miss},\gamma} + E_{\mathrm{(x)y}}^{\mathrm{miss},\tau} + E_{\mathrm{(x)y}}^{\mathrm{miss,jets}} + E_{\mathrm{(x)y}}^{\mathrm{miss},\mu} + E_{\mathrm{(x)y}}^{\mathrm{miss,soft}} \tag{4.1}$$

Each terms is the negative vectorial sum of the momentum of the respective calibrated physics object. The soft term is reconstructed from detector signal objects not associated with any object.

Specific selections on the objects and their kinematics relevant to the $HH \to bb\tau\tau$ analysis will be described in Chapter 6 in detail.

CHAPTER 5

DATA AND MONTE CARLO SAMPLE

Monte Carlo event generators comprise an extremely important tool in particle physics. All collider physics experiments use MCEG-simulated events to understand the data recorded by their detectors. Some commonly used Monte Carlo event generators programs are MG5_aMC@NLO [46], Herwig [47], POWHEG [48], Pythia [49], Sherpa [50]. They can be used to design and tune the detectors and analysis strategies.

The simulation of a proton-proton (pp) collision event at the LHC consist of the following steps:

1. The hard scattering process.
2. parton showering.
3. hadronisation.
4. underlying events.
5. decays of unstable particles.

Most of the particle produced in pp collisions are "boring" in the sense that they involve transfer of only a very small fraction of the momentum of the colliding particles and add nothing to our knowledge. Events with high physics importance are simulated by calculating the probability distribution of a particular "hard scatter" which has highest momentum transfer in the event from the perturbation theory. MCNLO, Powheg are the well known event generators mainly used to simulate hard scatter process.

In the hard scattering what happened to the incoming and outgoing particles is given by the "parton shower". Partons involves in hard scatters are mainly quarks and gluons. Gluons, being colored, radiate more gluons, resulting in an extended shower. The parton shower starts from the hard scatter process and goes downward to lower and lower momentum scale until perturbation theory breaks down. Programs commonly used for modeling parton shower include Pythia, Herwig and Sherpa.

"Hadronisation" takes account of the confinement of a system of partons into hadrons, which are detected by the detector. When two protons collide with each other there is high probability that there will be other interactions apart from the hard scattering. These events are known as "underlying events". The last stage of event generation is the decay of the unstable hadrons produced in the event. The Monte Carlo simulation in high energy physics is given by the Figure 5.1 [51] .

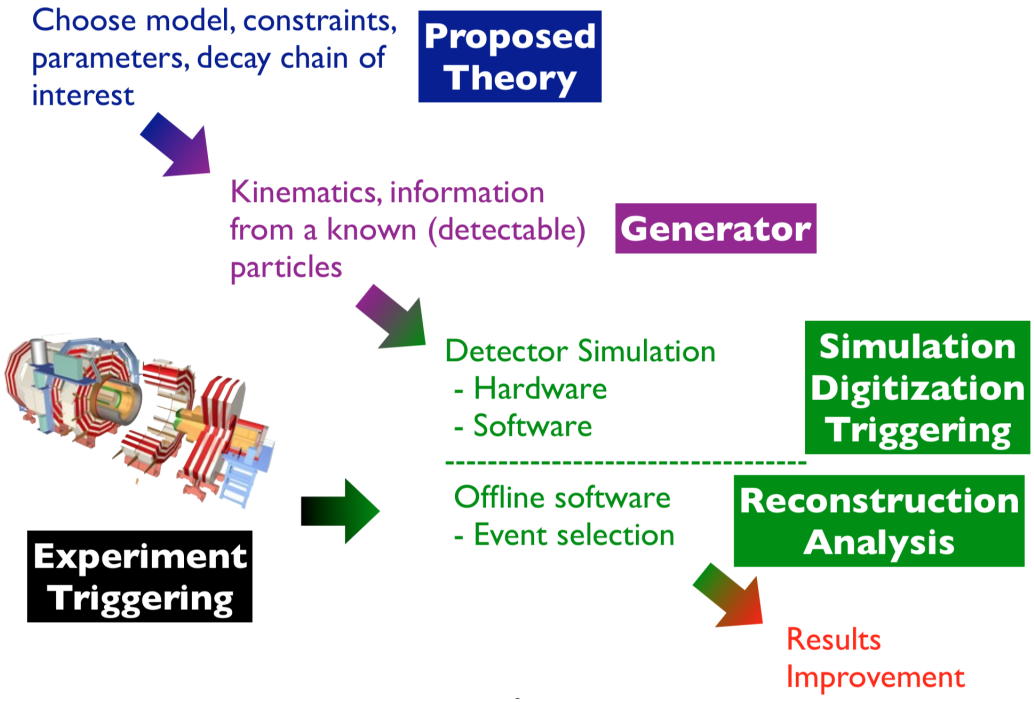


Figure 5.1: Monte Carlo simulation in High energy physics [52].

This section gives a brief overview on the generation of signal and background samples used in this analysis. The 2HDM next -to-leading order (NLO) heavy scalar (X) of fifteen different masses are generated in MG5_aMC@NLO and showered with Herwig7 where it decays to SM-like Higgs boson (H) pair, one of these light scalars decays to two b-jets and other one decays to two τs. Then a final state filter is applied to select semi-leptonic final state of two τs, one of the two τs decay leptonically (e or μ) and other to a hadronic final state. The light scalar mass, m_H is set at 125.0 GeV and the decay width of the heavy scalar, X, is fixed at 4 MeV.

For the generation of $t\bar{t}$ and single top quarks in the Wt and s-channel the Powheg-Box v2 generator is used. The single top quark production in t-channel is generated using the Powheg-Box v1 generator, which uses a "4-flavor scheme" for the NLO calculation of matrix elements. The parton shower and the underlying event are simulated using Pythia 8 and the corresponding Perugia 2012 tune (P2012) [53]. The top quark mass is set to 172.5 GeV. The EvtGen v1.2.0 program [54] is used to model bottom and charm hadron decays. the simulated $t\bar{t}$ background is normalized to the $t\bar{t}$ production cross-section at (next-to-next-to leading orde) NNLO + next-to-next-to leading log (NNLL) [55]. For single top processes, the generator NLO cross-sections are used.

Events containing W or Z bosons with associated jets are simulated using the Sherpa 2.2.1 generator. Matrix elements are calculated for up to two partons at NLO and four partons at LO using the Comix [56] and OpenLoops [57] matrix element generators and merged with the Sherpa parton shower [58] using the ME+PS@NLO prescription [59].

Diboson processes with one of the bosons decaying hadronically and the other leptonically are simulated using the Sherpa2.1.1 generator. They are calculated for up to one (ZZ) or zero (WW, WZ) additional partons at NLO and up to three additional partons at LO.

The $qqZH(Z \to \tau\tau,\ H \to bb)$ and $qqZH(Z \to bb,\ H \to \tau\tau)$ processes are generated with Pythia 8.186 [60]. The gluon-fusion initiated $ZH(Z \to \tau\tau,\ H \to bb)$ process is generated with Powheg-Box v2.

SM-like Higgs production in association with a top pair is generated with MG5_aMC@NLO and Pythia 8 is used to simulate the parton shower. Vector Boson Fusion (VBF) production of di-Higgs and WH have negligible contribution to this final state.

All signal and background events are passed through the full GEANT4-based [61, 62] simulation of the ATLAS detector response and are reconstructed with the same program as that used for data.

This analysis is based on the full 2018 data set, which corresponds to an integrated luminosity of 58.5 fb^{-1}.

CHAPTER 6

OBJECT AND EVENT SELECTION

6.1 Object Selection

Details of reconstruction and identification method of objects - es, γs, μs, τs, jets (both heavy and light flavors), and $E_{\mathrm{T}}^{\mathrm{miss}}$ - were discussed in Chapter 4. More specific selection criteria applied in this analysis are discussed below.

6.1.1 Electrons

The electrons are identified using a likelihood technique [39] and required to pass the "loose" [39] working point. The "loose" electron identification efficiency is 95%. The electron candidates are required to have $p_{\mathrm{T}} > 7$ GeV and $|\eta| < 2.47$. The p_{T} of an electron is calculated from the four-vector of an electron candidate based on the energy deposited in the calorimeter and the direction of track in the inner detector. Electron candidates in the gap-and-crack region ($1.37 < |\eta| < 1.52$) are removed. Electrons are required to be isolated and selected using the "loose" [39] isolation working point of efficiency 99% (flat as function of electron p_{T}). Isolated electrons do not have any nearby tracks or energy deposits in the calorimeter a in a ΔR cone, whose size depends on p_{T}, around the electron. The isolation requirement is inverted in the control region (CR) to estimate the QCD background (mainly gluon jets).

6.1.2 Muons

Muon candidates are selected with MS and ID information. Muon candidates must have $p_{\mathrm{T}} > 7$ GeV and $|\eta| < 2.5$ are required to pass "loose" [43] isolation requirement of efficiency 99%. The isolation requirement is inverted like it is done for electron in the CR to estimate the QCD background.

6.1.3 Jets

As previously mentioned in Chapter 4, jets are reconstructed with the anti-k_t algorithm [41] with a radius parameter R=0.4 applied to topological clusters of energy deposited in calorimeter cells. The jet energy is determined from reconstructed cluster energies at the electromagnetic scale with a correction factor derived from simulation and data to account for the response of the calorimeter to hadrons. All jets are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.4$, and are cleaned for detector effects, beam backgrounds and cosmic rays. Reconstruction of b-jets is discussed in Chapter 4. A Jet Vertex Tagger (JVT) "medium" requirement is applied on the jets having $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ to remove contribution from pile up jets. JVT is a 2D likelihood discriminant based on the R_{pT} and corrJVF described in [65].

6.1.4 Taus

Jets reconstructed with the anti-k_t [63] algorithm are further calibrated using Local hadronic Calibration (LC) [64] as inputs. These jets are used to seed the search for the visible part of hadronic τ (τ_{had}) decays, if they have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$. To have

better rejection against jet misidentified (fake) as τ_{had}, τ_{had} is chosen using a Boosted Decision Tree (BDT) discriminant, as explained in Chapter 8 below. The τ_{had}s passing the "medium" identification working point are chosen. This working point has 55% efficiency for 3-prong and 65% efficiency for 1-prong τ_{had}s. The BDT output is flattened using a cumulative distribution function (interpolation between bins) so that the working point will be fixed across the whole τ_{had} p_{T} range and independent of the event's pile-up condition. Selected τ_{had} candidates are required to have $p_{\text{T}} > 20$ GeV, $|\eta| < 2.5$, one or three tracks and unit charge for 3-prong candidates, where the sum of the charges of the three tracks is required to be ± 1.

6.1.5 Fake Taus

To construct a control region (CR) region having fake τ_{had}, an 'anti-τ' selection ID is defined. 'anti-τ' are required to pass "veryLoose" identification and fail "medium" [42] identification. The minimum BDT score requirement ensures that the jets still have some τ_{had}-like properties and the composition of quark and gluon initiated jets are closer to that of the signal region.

6.1.6 Missing Transverse Energy

The reconstruction of missing transverse energy, $E_{\text{T}}^{\text{miss}}$ in the ATLAS detector is described in Chapter 4. In the case of events containing anti-τs used for fake-τ_{had} background estimation, the $E_{\text{T}}^{\text{miss}}$ is computed considering the anti-τs as τ_{had}-like objects, in order to ensure that they are calibrated at the τ_{had} energy scale rather than the jet energy scale.

6.1.7 Overlap Removal

If two objects are reconstructed within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, one of them is removed following the procedure stated below:

- e-jet within ΔR=0.2. Priority is given to the e and jet is removed.
- jet-e in 0.2< ΔR <0.4. Priority is given to the jet.
- μ-jet within ΔR=0.4 and fewer than three tracks with $p_{\mathrm{T}} > 500$ MeV in the jet. Priority is given to the μ.
- jet-μ within $\Delta R = 0.4$ and three or more tracks with $p_{\mathrm{T}} > 500$ MeV in the jet. Priority is given to the jet.
- e-μ within ΔR=0.2 and sharing an inner detector track. Priority is given to the μ.
- μ-τ within ΔR=0.2. Priority is given to the μ.
- e-τ within ΔR=0.2. Priority is given to the e.
- τ-jet within ΔR=0.2. In case of the b-tagged jets and anti-τ, we prioritize the physics object as BDT "medium" τ_{had} > b-tagged jet > anti-τ > Light jet. For example, if a "medium" τ_{had} overlaps with a b-jet we will remove b-jet and if an anti-τ overlaps with b-jet then the b-jet is kept and the anti-τ is removed.

6.2 Event Selection

The trigger used for 2018 data periods, the final event selection for the signal region (SR) and different CRs are explained in this section.

6.2.1 Trigger

In this analysis we use single lepton triggers (SLTs) because in total they have larger signal acceptance than the lepton tau triggers (LTTs). LTTs are considered as a separate channel that is not included in this analysis. SLTs include Single Electron Triggers (SETs) and Single Muon Triggers (SMTs). Events are first tested to see if they pass any of the single lepton triggers (SLTs). Several different SLTs are used to maximize acceptance. SET events are required to have at least one electron with $p_{\mathrm{T}} > 24$ GeV that satisfied the "medium" identification criteria and "loose" isolation requirements (in later data-taking periods the p_{T} threshold is raised to 26 GeV and identification changed to "tight") or at least one electron with $p_{\mathrm{T}} > 60$ GeV that satisfies the "medium" identification criteria, with no requirements on its isolation, or at least one electron with $p_{\mathrm{T}} >$ 120-140 GeV (depending on the data-taking period) that satisfies the "loose" identification criteria. In the muon channel, events are required to have a muon with $p_{\mathrm{T}} >$ 24-26 GeV (depending on the data-taking period) and to satisfy a "loose" isolation criteria or at least one muon with $p_{\mathrm{T}} > 50$ GeV with no isolation requirement.

6.2.2 Offline Event Selection

The events that pass the trigger requirements need to have either exactly one electron passing the "tight" identification criteria or one muon passing the "medium" identification criteria. The p_{T} of all objects are required to be greater by 1 GeV, at least, than their corresponding trigger thresholds. At least two jets with $p_{\mathrm{T}} > 45$ and 20 GeV are also required. The lepton and the τ_{had} candidate charges are required to have opposite signs. The invariant mass of τ_{had} and e/μ including neutrinos (MMC_{TauLep}) greater than 60 GeV.

Table 6.1: Summary of object selection.

Object	Selection
e	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.47$, remove es $1.37 < \lvert\eta\rvert < 1.52$
μ	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.5$
τ_{had}	$p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.4$, "medium" τ_{had}
anti-τ	"VeryLooseTau" but not "medium" τ_{had}
jet	sub-leading jet of $p_{\mathrm{T}} > 20$ GeV, leading jet of $p_{\mathrm{T}} > 45$ GeV, $\lvert\eta\rvert < 2.4$
b-jet	MV2C10 algorithm, 70% WP

Table 6.2: Summary of Event selection.

Event Selections
Single Lepton Trigger
$MMC_{TauLep} > 60$ GeV
$\mu > e > \tau_{\mathrm{had}} > b$-jet > anti-$\tau$ > light jet
Exactly one τ_{had} and one e/μ, opposite sign
Events should have at least two jets with the p_{T} requirements.
2 b-tag: events with 2 b-tagged jets.
1 b-tag: events with 1 b-tag jet and the second jet is the leading non b-tagged jet.
0 b-tag: events with no b-tagged jets.

Table 6.1 and Table 6.2 summarize the object and event selections.

CHAPTER 7

BACKGROUND ESTIMATION

The backgrounds that share the same final state (one e/μ, exactly one τ_{had}, 1 or 2 b jets) as our signal are the irreducible backgrounds. In the 2 (1) b-tag category $t\bar{t}$ with true τ_{had} contributes around 70% (50%) of the total background. The second dominant background in 2 and 1 b-tag is fake tau (25-30%), where jets are mis-identified as τ_{had}. The next irreducible background which shares the same final state is $Z \to \tau\tau$+jets, especially heavy flavour (HF) jets that dominates over light flavour (LF). Both are estimated from simulation. Contribution from true τ_{had} in the semi-leptonic channel is estimated from simulation using the truth information of reconstructed τ_{had}s. τ_{had} matched with truth electrons or muons are also estimated from simulation. Only jet faking τ_{had} are estimated using a data driven method. Other backgrounds like SM Higgs ($t\bar{t}H$, VH), diboson and W+HF which have negligible contribution in this final state are estimated from simulation.

A disagreement between data and MC was observed as a function of τ_{had} p_{T} (Figure 7.1). This difference is more prominent among 3-prong τ_{had} (Figure 7.2b).

The fake τ_{had} background are coming mainly from $t\bar{t}$ in 2 b-tag and from W+jets in 0 b-tag, while the two sources contribute evenly to 1 b-tag. The fakes are estimated in different control regions dominated by a specific background. The regions are defined as follows:

- $t\bar{t}$ control region (CR): ${m_{\text{T}}}^{LepMET} > 40$GeV, 2 b-tag,
- W+jets CR: ${m_{\text{T}}}^{LepMET} > 40$GeV, 0 b-tag,

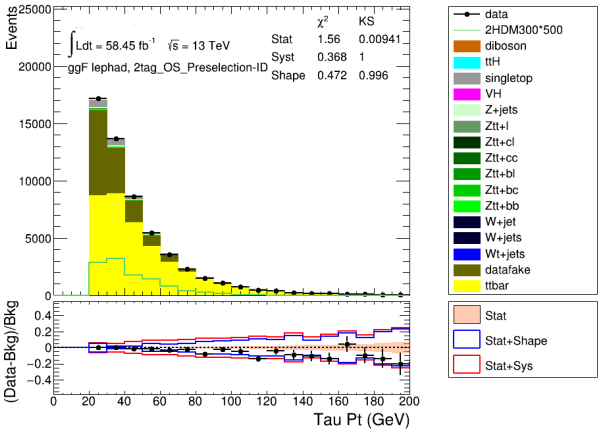
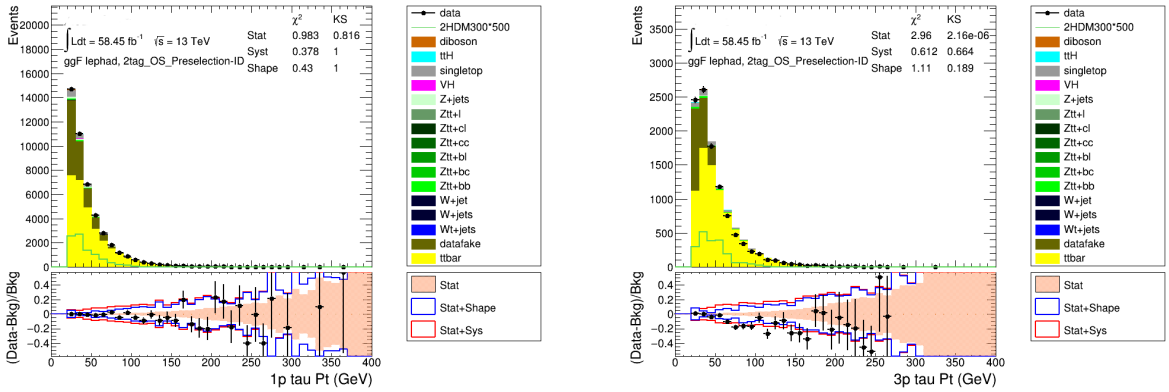


Figure 7.1: Kinematic variable distributions with 2018 data and MC16e in 2 b-tag region.



(a) (b)

Figure 7.2: Kinematic variable distributions (a) 1-prong τ_{had} p_{T}, (b) 3-prong τ_{had} p_{T} with 2018 data and MC16e in 2 b-tag region.

- Multi-jet CR: inverted lepton isolation (an e/μ candidate that fails the respective 'loose' isolation requirement, but passes all other requirements for a 'tight' electron or a 'medium' muon), 0 or 1 b-tags.

Each fake-τ_{had} enriched sample is defined with all the selection in signal region and specific control region but requiring 'anti-tau' defined in section 6.3.5 instead of a the τ_{had}. Fake-τ_{had} background is estimated from data using a Fake Factor (FF) method, defined as:

$$\mathrm{FF} = \frac{N(\tau)}{N(\text{anti-}\tau)} \tag{7.1}$$

where, $N(\tau)$ ($N(\text{anti-}\tau)$) is estimated by subtracting true hadronic tau (anti-tau) contribution of $t\bar{t}$, W+jets and other minor background processes from data. FFs are calculated separately for each process $t\bar{t}/W$+jets, multi-jet, for 1- and 3-prong τ_{had} and parameterized in terms of τ_{had} p_{T}. Since the fake contribution from multijet background in the 2 b-tag region is negligible, only FF($t\bar{t}$) is applied in the 2 b-tag region. A combined FF is applied in 1 and 0 b-tag regions to include fakes coming from the multijet background. This combined fake factor is defined as:

$$\mathrm{FF(combine)} = \mathrm{FF(QCD)} \times r_{\mathrm{QCD}} + FF(W + \mathrm{jets}/t\bar{t}) \times (1 - r_{\mathrm{QCD}}) \tag{7.2}$$

Where r_{QCD} (Equation: (6.9)) is defined as a fraction of fakes in the anti-τ region and is measured as a function of τ_{had} p_{T} :

$$r_{\mathrm{QCD}} = \frac{N(\text{multi-jet, data})}{N(\text{data}) - N(\text{true}\,\tau_{\mathrm{had}}, \mathrm{MC})} \tag{7.3}$$

$$N(\text{multi-jet, data}) = N(\text{data}) - N(\text{true}\,\tau_{\mathrm{had}}, \mathrm{MC}) - N(\text{fake}\,\tau_{\mathrm{had}}, \mathrm{MC}) \tag{7.4}$$

The N(multi-jet, data) is calculated by subtracting all background contributions, both fake and true taus, apart from multi-jet, from the data in the anti-τ region. The r_{QCD} factor is computed separately in each tag region, since the amount of QCD contribution changes with the number of b-tag jets required.

Figure 7.3 shows the fake factor in 2 b-tag region. In 1 b-tag FF($t\bar{t}$) (Figure 7.4a) and FF(QCD) (Figure 7.4b) are calculated separately and then combined with correct r_{QCD} (Figure 7.5b) following the Equation:(6.7). In 1 b-tag FF($t\bar{t}$) includes fakes from $t\bar{t}$ and

W+jets. The negative fake factors are set at zero assuming no fake contribution in that p_{T} range.

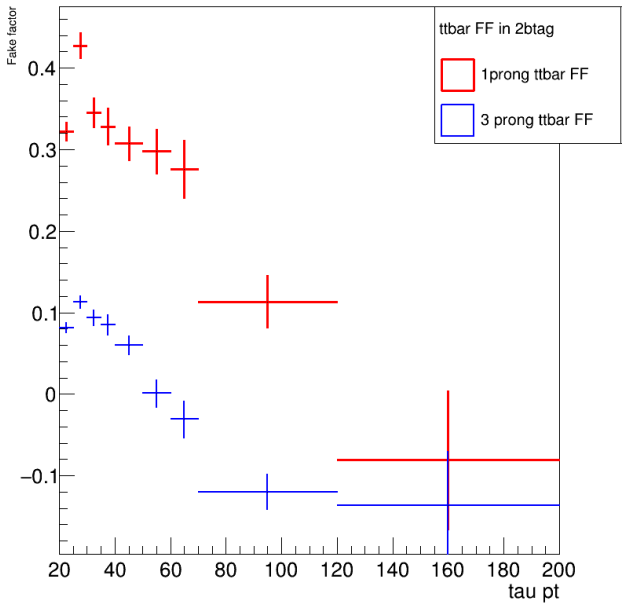


Figure 7.3: Fake factor in 2 b-tag category.

The estimated background in 1 and 2 b-tag regions are shown in Table 7.1 - 7.2, where the sample names mean:

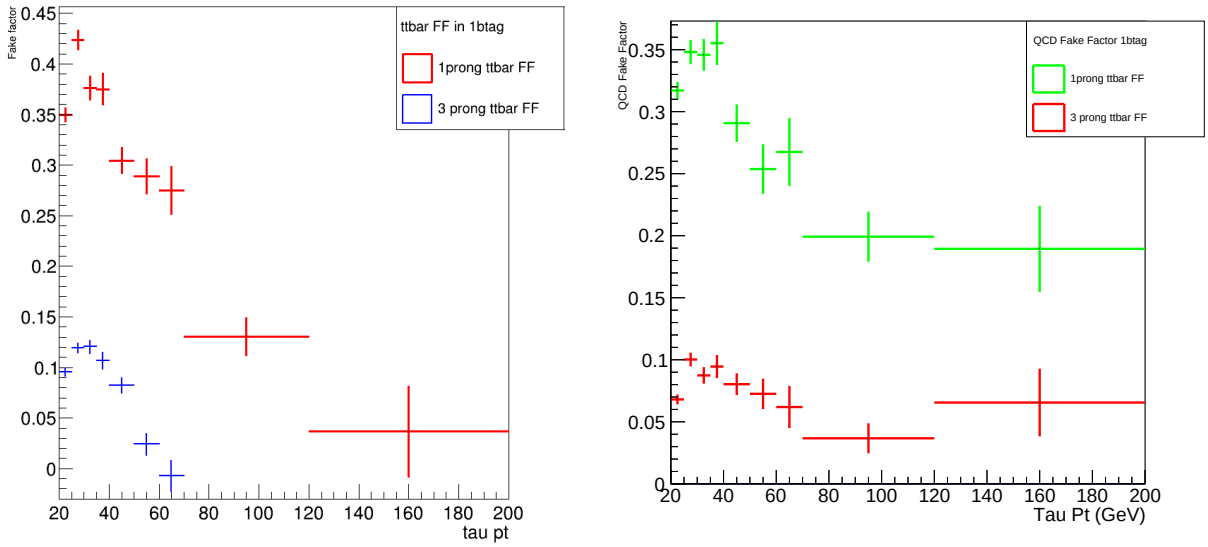
- '$Zttbb/bc/bl/cc/cl/l/$': $Z \rightarrow \tau\tau$ in association with jets of different flavours ('b','c','l').
- '$Zbb/bc/bl/cc/cl/l$': $Z \rightarrow ee/\mu\mu$ in association with jets of different flavours.
- '$Wtbb/bc/bl/cc/cl/l$': $W \rightarrow \tau\nu$ in association with jets of different flavours.
- '$Wbb/bc/bl/cc/cl/l$': $W \rightarrow e\nu/\mu\nu$ in association with jets of different flavours.
- 'stop$s/t/Wt$': single top sample with different production channels 's', 't', 'Wt'.
- Fakes ("datafakes" in Figure: 8.2- 8.5): jet faking taus are estimated using the data-driven method explained in section 6.5.

Table 7.1: Yield table for 2018 data and Monte Carlo (MC16e) in the 2 b-tag region before fitting the MC to data.

Sample	Yield	Stat uncertainty	contribution to total background %
data	55894	± 236	
$t\bar{t}$	38227.55	± 82.67	68.4277
stopWt	1827.52	± 22.13	3.2713
stop$t$	310.91	± 9.13	0.5565
stop$s$	22.48	± 0.82	0.0402
Fakes	15010.21	± 79.35	25.8792
ttH	68.79	± 0.62	0.1217
$ggZllH$	0.99	± 0.01	0.0017
$qqZllH$	23.45	± 0.72	0.0415
$Zttl$	1.59	± 0.02	0.0028
$Zttcl$	5.55	± 0.10	0.0099
$Zttcc$	18.34	± 0.59	0.0328
$Zttbl$	5.03	± 0.21	0.0090
$Zttbb$	448.12	± 11.62	0.8021
$Zttbc$	10.95	± 0.54	0.0196
Wtl	0.01	± 0.001	0.0000
$Wtbb$	1.19	± 0.34	0.0021
$Wtcl$	0.07	± 0.02	0.0001
$Wtcc$	0.15	± 0.03	0.0003
$Wtbl$	0.14	± 0.02	0.0003
$Wtbc$	0.49	± 0.10	0.0009
WW	5.29	± 0.52	0.0095
WZ	17.84	± 0.94	0.0319
ZZ	26.19	± 0.99	0.0469
Wl	0.10	± 0.004	0.0002
Wcl	0.67	± 0.03	0.0012
Wbb	18.45	± 1.78	0.0330
Wcc	1.79	± 0.16	0.0032
Wbl	1.53	± 0.09	0.0027
Wbc	4.71	± 0.34	0.0084
Zl	1.09	± 0.02	0.0020
Zcc	16.60	± 0.70	0.0297
Zbl	13.40	± 0.50	0.0240
Zbb	389.05	± 9.20	0.6964
Zcl	4.46	± 0.13	0.0080
Zbc	26.77	± 1.13	0.0479
Total Bkgd	56505.73	± 0.21	
signal(300 GeV, 1pb)	21.99	± 0.49	
signal(600 GeV, 1pb)	86.06	± 1.01	
signal(900 GeV, 1pb)	130.71	±1.32	
data/MC	0.9892		

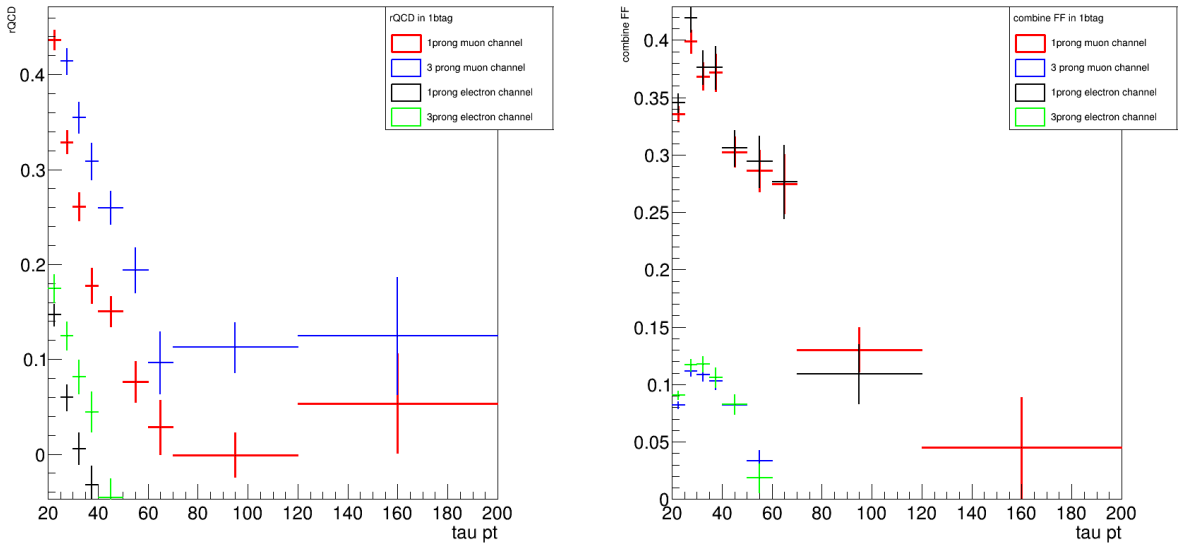
Table 7.2: Yield table for 2018 data and Monte Carlo (MC16e) in 1 b-tag region before fitting MC to the data.

Sample	Yield	Stat uncertainty	contribution in total background %
data	110536	± 332	
$t\bar{t}$	56186.75	± 102.10	50.2013
stopWt	6745.10	± 42.79	6.0274
stop$t$	1070.74	± 17.12	0.9567
stop$s$	56.91	± 1.34	0.0508
Fakes	37360.47	± 127.57	33.3806
ttH	76.20	± 0.66	0.0680
diboson	3.77	± 0.03	0.0034
$Zttl$	455.22	± 4.39	0.4067
$Zttcl$	931.73	± 14.91	0.8325
$Zttcc$	221.42	± 8.77	0.1978
$Zttbl$	2811.38	± 34.52	2.5119
$Zttbb$	606.09	± 16.43	0.5415
$Zttbc$	224.05	± 8.06	0.2002
Wtl	3.58	± 0.25	0.0032
$Wtbb$	1.32	± 0.60	0.0012
$Wtcl$	8.81	± 2.25	0.0079
$Wtcc$	1.62	± 0.41	0.0014
$Wtbl$	22.23	± 1.75	0.0199
$Wtbc$	2.87	± 0.57	0.0026
WW	143.07	± 3.24	0.1278
WZ	163.52	± 2.98	0.1461
ZZ	111.40	± 2.52	0.0995
Wl	34.39	± 1.33	0.0307
Wcl	110.32	± 5.07	0.0986
Wbb	22.85	± 2.99	0.0204
Wcc	17.53	± 2.19	0.0157
Wbl	237.41	± 8.65	0.2121
Wbc	25.53	± 2.07	0.0228
Zl	370.21	± 4.79	0.3308
Zcc	177.49	± 9.88	0.1586
Zbl	2328.88	± 34.12	2.0808
Zbb	518.37	± 17.34	0.4632
Zcl	793.40	± 16.55	0.7089
Zbc	157.28	± 8.62	0.1405
Total Bkg	111992.8047	± 0.1616%	
signal(300 GeV, 1pb)	36.89	± 0.66	
signal(600 GeV, 1pb)	116.02	± 1.19	
signal(900 GeV, 1pb)	158.96	± 1.47	
data/MC	0.9870		



(a) (b)

Figure 7.4: (a) $t\bar{t}/W$+jets fake factor, (b) QCD fake factor in 1 b-tag.



(a) (b)

Figure 7.5: (a) r_{QCD} and (b) combine fake factor in the 1 b-tag.

CHAPTER 8

MULTIVARIATE ANALYSIS AND B-TAGGING OPTIMIZATION

The first section of this chapter describes multivariate technique used to achieve better sensitivity between resonant signal and the dominant background $t\bar{t}$ Boosted Decision Tree (BDT) [66] is used. The Toolkit for Multivariate Analysis (TMVA) [66] provides a ROOT integrated environment for processing, parallel evaluation and application of multivariate classifications. Later section discusses an optimization study of b-tagging working point.

8.1 Boosted Decision Tree

A decision tree is a classifier with a binary branching structure that grows with repeated classification based on one characteristic input variable at a time. This procedure is shown in Figure 8.1. The tree starts from a root node which selects the variables which have the best ability to discriminate between signal and background. Each node can use the same variable and other variables might not be used at all. The leaf nodes at the bottom end of the tree are labeled 'S' for signal and 'B' for background depending on the majority of events that end up in the respective nodes. The boosting of decision tree extends this idea from one tree to several trees which make a forest. The trees are generated with the same training samples by reweighting events and finally combined into single classifier that is an average of the individual decision trees.

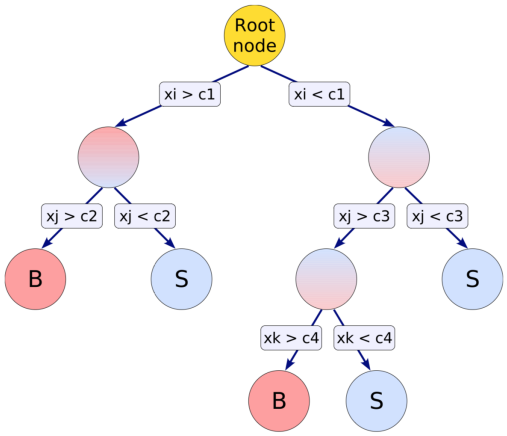


Figure 8.1: Schematic view of a decision tree [66].

Events that are misclassified during the training of a decision tree are given higher event weight in training of the next tree. The first tree starts with the original event weights provided in the training events, subsequent are trained with weights modified by earlier misclassifications and multiplied by a boost weight α. The boost weight is derived from the misclassification rate, err, from the previous tree:

$$\alpha = \frac{1 - \text{err}}{\text{err}} \tag{8.1}$$

$$\text{err} = 1 - \max(\text{p}, 1 - \text{p}) \tag{8.2}$$

Misclassification rate (Equation:(8.2)) is defined by the purity (p) of the sample. The purity of a node is given by the ratio of signal events to all events in that node. For a completely mixed sample purity is set at 0.5 and fall off to zero when sample consists of fully background. The error rate is err ≤ 0.5 because the same training events are used to classify the output nodes of the previous tree and these are also used to calculate the error rate.

The weight of the full training sample is renormalized so that the total sum of weights remains constant. To classify an event as either signal-like or background-like, the following equation is used:

$$y_{\text{Boost}}(x) = \frac{1}{N_{\text{collection}}} . \sum_{i} \ln(\alpha_i) . h_i(x) \tag{8.3}$$

Where sum is over all classifiers (i) in collection, $h(x)$ is the result of individual classifiers and x is the tuple of input variables. $h(x) = 1$ and -1 for signal and background respectively. A small (large) value of $y_{\text{Boost}}(x)$ indicates background-like (signal-like) event.

The most popular boosting algorithm that is used for this analysis is Adaptive Boost (AdaBoost) [66]. Another boosting technique that was considered but was not used finally is Gradient Boost (GradBoost). GradBoost does not give better separation between signal and background in the signal region than AdaBoost.

The list of input variables used for the training sample are listed below.

- MMC_{TauLep}: invariant mass of tau and lepton system, includes the missing transverse momentum of the neutrinos. This variable is calculated with Missing Mass Calculator (MMC) explained in Ref [67].

- m_{HH}: invariant mass of the two postulated SM-like Higgs boson candidates. It is reconstructed from di-tau and di-b-jet masses. Scale factors m_H/MMC_{TauLep} and m_H/m_{bb} (where m_H is the light scalar boson mass of 125 GeV) are applied to the four-momenta of the di-tau and di-b-jet systems, respectively, in order to improve the mass resolution. For signal this variable will have a peak at the m_X but not for the background, so it has a good discriminating power to select signal.

- $DeltaR_{TauLep}$: ΔR between tau and lepton.

- m_{bb}: invariant mass of di-b system.

- $DeltaR_{bb}$: ΔR between two b-jets.

- ${m_{\mathrm{T}}}^{LepMET}$: the transverse mass between the lepton and ${E_{\mathrm{T}}}^{\mathrm{miss}}$, defined as:

$$ {m_{\mathrm{T}}}^{LepMET} = \sqrt{2{p_{\mathrm{T}}}^{l}{E_{\mathrm{T}}}^{\mathrm{miss}}(1-\cos\Delta\phi)}, \tag{8.4} $$

  where ${p_{\mathrm{T}}}^{l}$ is the transverse momentum of the lepton and $\Delta\phi$ is the angle between the lepton and ${E_{\mathrm{T}}}^{\mathrm{miss}}$.

- MET: the missing transverse momentum of the event, ${E_{\mathrm{T}}}^{\mathrm{miss}}$.

- $METcentrality$: it quantifies the position in $\Delta\phi$ of the ${E_{\mathrm{T}}}^{\mathrm{miss}}$ with respect to the visible decay products of the two taus:

$$ METcentrality = \frac{A+B}{\sqrt{A^2+B^2}} \tag{8.5} $$

  Where A and B are given by:

$$ A = \frac{\sin(\phi_{{E_{\mathrm{T}}}^{\mathrm{miss}}} - \phi_{\tau_{\mathrm{l}}})}{\sin(\phi_{\tau_{\mathrm{had}}} - \phi_{\tau_{\mathrm{l}}})}, B = \frac{\sin(\phi_{\tau_{\mathrm{had}}} - \phi_{{E_{\mathrm{T}}}^{\mathrm{miss}}})}{\sin(\phi_{\tau_{\mathrm{had}}} - \phi_{\tau_{\mathrm{l}}})}, \tag{8.6} $$

  The $METcentrality$ is equal to:

  1. $\sqrt{2}$ when the ${E_{\mathrm{T}}}^{\mathrm{miss}}$ lies exactly between the two taus; or
  2. 1 if the ${E_{\mathrm{T}}}^{\mathrm{miss}}$ is perfectly aligned with either of the taus; or
  3. < 1 if the ${E_{\mathrm{T}}}^{\mathrm{miss}}$ lies outside of the angular region defined by the two taus.

  Signal events tend to have larger values of the METcentrality as in these cases the two taus are produced from the decay of a light scalar boson and the reconstructed ${E_{\mathrm{T}}}^{\mathrm{miss}}$ generally falls in between the two visible τ decay products.

Table 8.1: Optimised BDT hyper-parameter settings.

Parameters	Value
BoostType	AdaBoost
AdaBoostBeta	0.15
NTrees	200
MaxDepth	4
MinNodeSize	5
NCuts	100

- $DeltaPhi_{HH}$: the $\Delta\phi$ between the two reconstructed 125 GeV bosons, where the di-tau direction is taken from the MMC fit.
- $b2_{Pt}$: p_{T} of the sub-leading b-jet.
- $DeltaPt_{TauLep}$: difference in p_{T} of the lepton and the visible hadronic tau.
- $visM_{TauLep}$: invariant mass of lepton and visible decay product of τ_{had}.

The input variable distributions of signal (m_X = 300 GeV) and different backgrounds in 2 b-tag region are shown in Figure 8.2 - Figure 8.3. The same distributions in 1 b-tag can be found in Figure 8.4 - Figure 8.5. BDT training in 1 b-tag and 2 b-tag regions are done separately for every signal mass and the main background $t\bar{t}$.The training and application are done using a k-fold (k=5) method to ensure that they are statistically independent. For this, the signal and background events are divided in five sets of equal size. The BDT is trained on four of the sets, and applied to the fifth. This process is repeated for each of the five possible choice of the fifth set in order to minimize statistical fluctuations.

The optimized hyper-parameters [66] setting in AdaBoost are listed in Table 8.1. Varying the hyper-parameters did not improve the sensitivity of our channel, so these settings are taken from the previous analysis [11] and used in both tagging regions.

The input variable ranking for the 1 and 2 b-tag regions are shown in Table 8.2 and Table 8.3 respectively for the 300, 600, 900 GeV mass points. The correlation between each

Table 8.2: Ranking of input variables in 1 b-tag. The variables are listed according to their importance.

Ranking	m_X=900 GeV	m_X=600 GeV	m_X=300 GeV
1	m_{HH}	$DeltaR_{TauLep}$	$visM_{TauLep}$
2	$DeltaR_{TauLep}$	m_{HH}	MMC_{TauLep}
3	m_{bb}	MMC_{TauLep}	MET
4	$DeltaR_{bb}$	m_{bb}	${m_{\mathrm{T}}}^{LepMET}$
5	MMC_{TauLep}	${m_{\mathrm{T}}}^{LepMET}$	$\mathrm{DeltaR_{TauLep}}$
6	${m_{\mathrm{T}}}^{LepMET}$	$b2_{Pt}$	m_{bb}
7	$visM_{TauLep}$	$DeltaR_{bb}$	m_{HH}
8	$DeltaPt_{TauLep}$	$visM_{TauLep}$	$METcentrality$
9	$b2_{Pt}$	$METcentrality$	$b2_{Pt}$
10	$METcentrality$	$DeltaPt_{TauLep}$	$DeltaPt_{TauLep}$
11	MET	MET	$DeltaR_{bb}$

pair of input variables in signal and background samples are shown in Figure 8.7. The number of input variables is optimized using area under the ROC (Figure 8.8) [66] curve as a figure of merit.

The agreement between training and testing samples in 1 and 2 b-tag regions are shown in Figure 8.6.

8.2 b-tagging Working Point Optimisation

The MV2C10 algorithm described in Chapter 4 is used to tag b- jets. Different working points are defined in terms of the algorithm's parameter settings that result in different b-tagging efficiencies: 60%, 70%, 77%, 85%. Separate BDT training in 85% and 70% WP were performed and significance in exactly 1 and exactly 2 b-tagging regions were investigated. For signal, which has a very small cross-section, the following significance (ξ) formula was used:

$$\xi = \sqrt{2((N_{\mathrm{S}} + N_{\mathrm{B}})ln(1 + \frac{N_{\mathrm{S}}}{N_{\mathrm{B}}}) - N_{\mathrm{S}})} \tag{8.7}$$

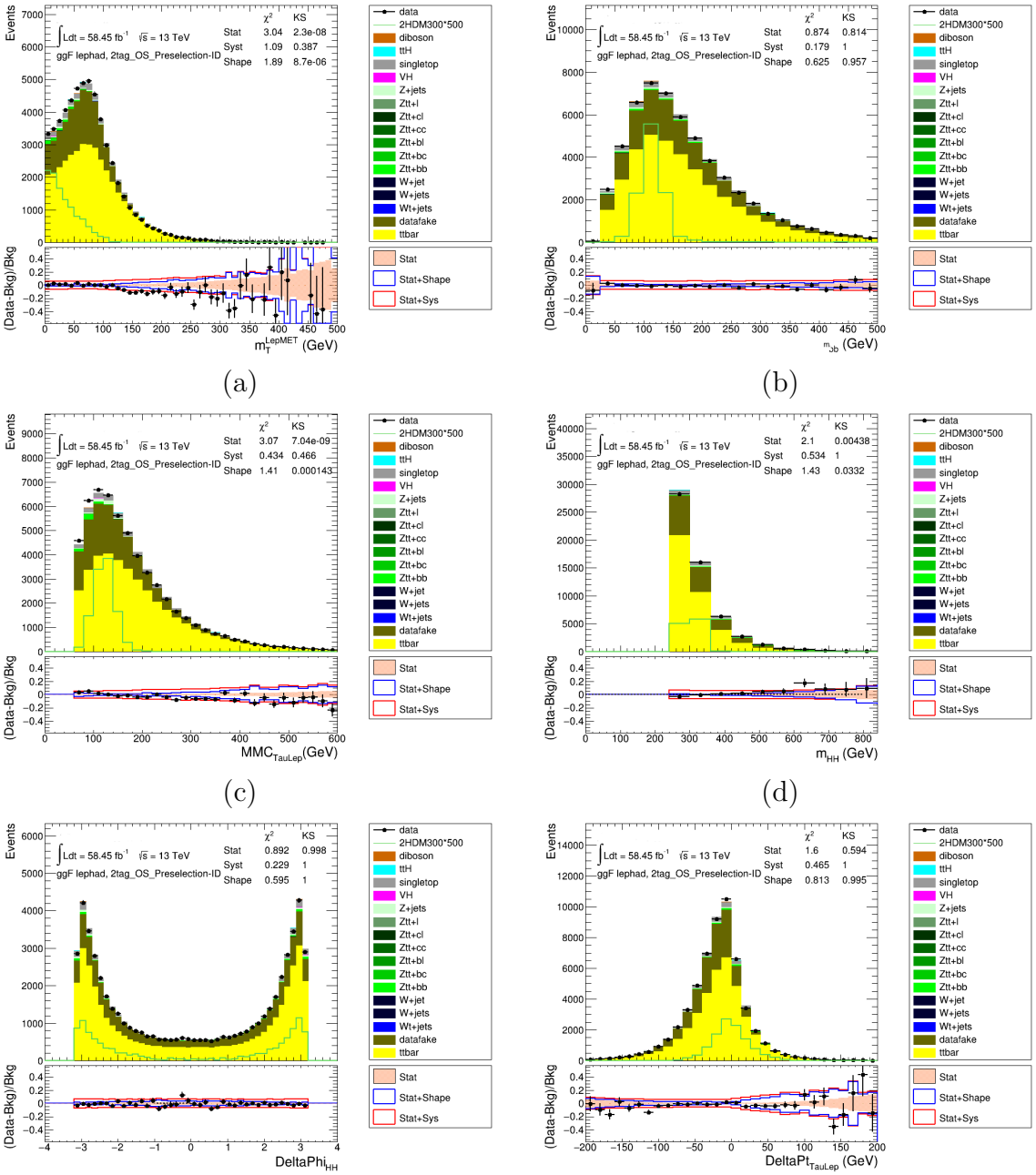


Figure 8.2: Examples of distributions of BDT input variables: (a) $m_{\mathrm{T}}{}^{LepMET}$, (b) m_{bb}, (c) MMC_{TauLep}, (d) m_{HH}, (e) $DeltaPhi_{HH}$, (f) $DeltaPt_{TauLep}$ in 2018 data and MC16 in the 2 b-tag region. "Stat+Shape" indicates the statistical and systematic uncertainties from the shape difference. "Stat+Syst" indicates the statistical and systematic uncertainties from the overall normalization difference.

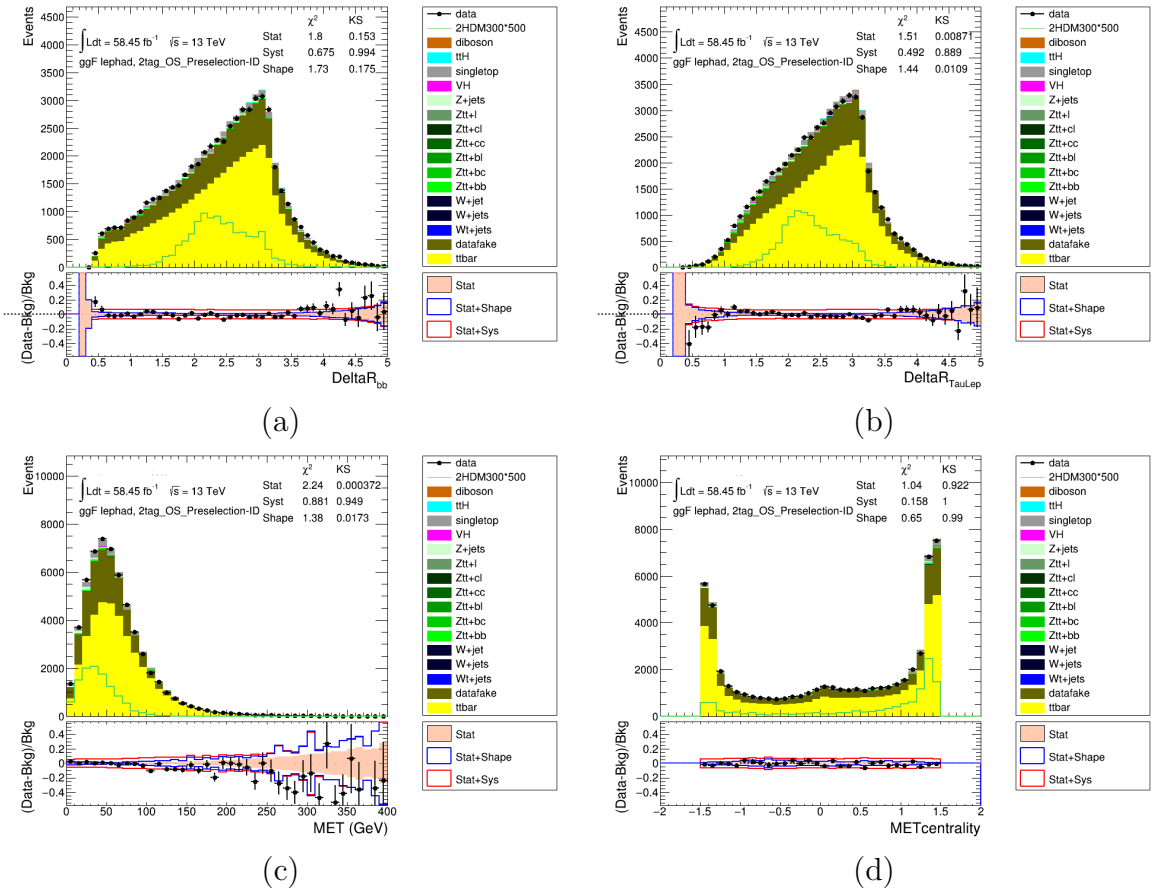


Figure 8.3: Examples of distributions of BDT input variables: (a) $DeltaR_{bb}$, (b) $DeltaR_{\mathrm{TauLep}}$, (c) MET, (d) $METCentrality$ in 2018 data and MC16 in the 2 b-tag region. "Stat+Shape" indicates the statistical and systematic uncertainties from the shape difference. "Stat+Syst" indicates the statistical and systematic uncertainties from the overall normalization difference.

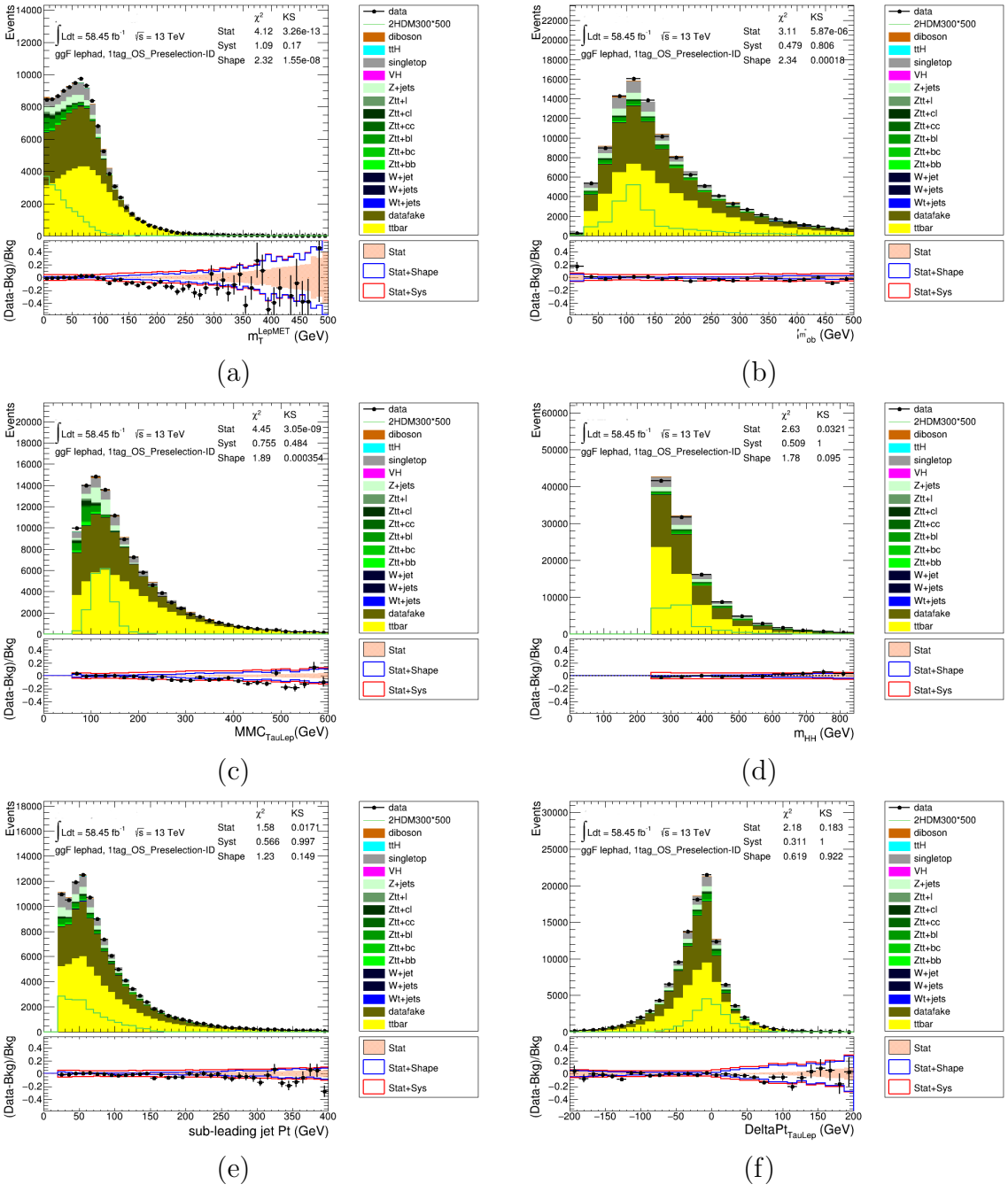


Figure 8.4: Examples of distributions of BDT input variables: (a) $m_{\mathrm{T}}{}^{LepMET}$, (b) m_{bb}, (c) MMC_{TauLep}, (d) m_{HH}, (e) $DeltaPhi_{HH}$, (f) $DeltaPt_{TauLep}$ in 2018 data and MC16 in the 1 b-tag region. "Stat+Shape" indicates the statistical and systematic uncertainties from the shape difference. "Stat+Syst" indicates the statistical and systematic uncertainties from the overall normalization difference.

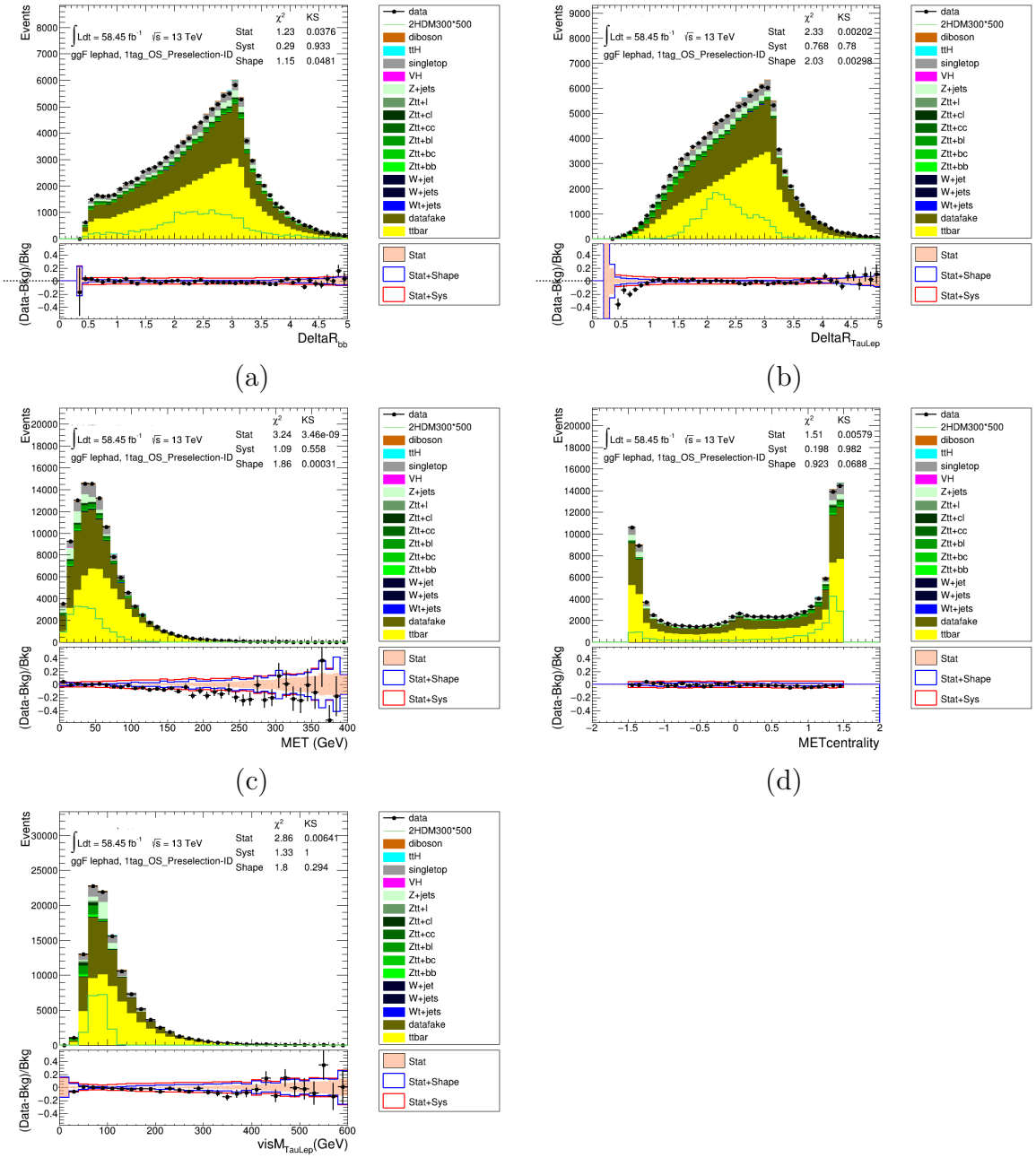


Figure 8.5: Examples of distributions of BDT input variables: (a) $DeltaR_{bb}$, (b) $DeltaR_{TauLep}$, (c) MET, (d) $METCentrality$, (f) $visM_{TauLep}$ in 2018 data and MC16e in the 1 b-tag region."Stat+Shape" indicates the statistical and systematic uncertainties from the shape difference. "Stat+Syst" indicates the statistical and systematic uncertainties from the overall normalization difference.

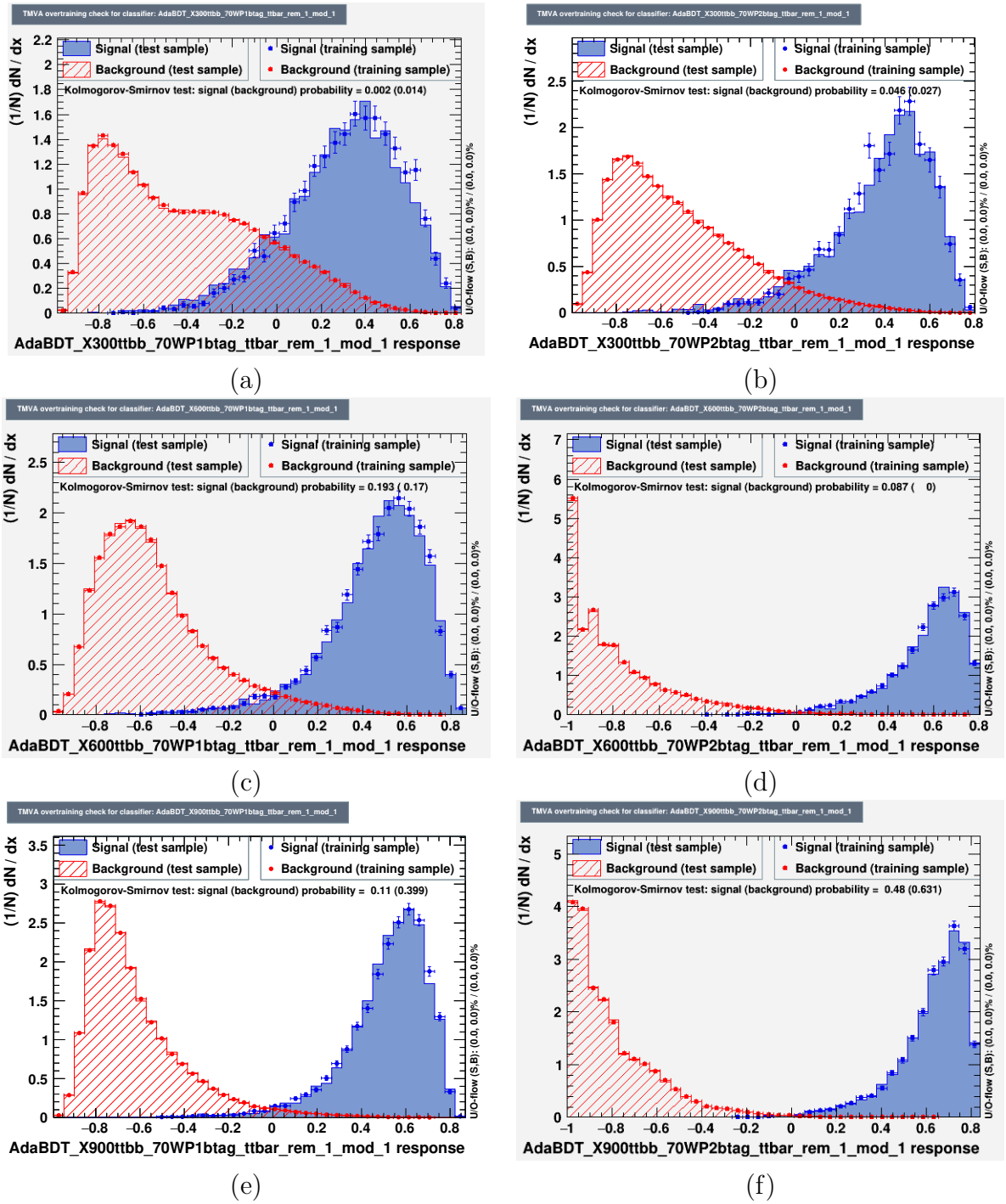
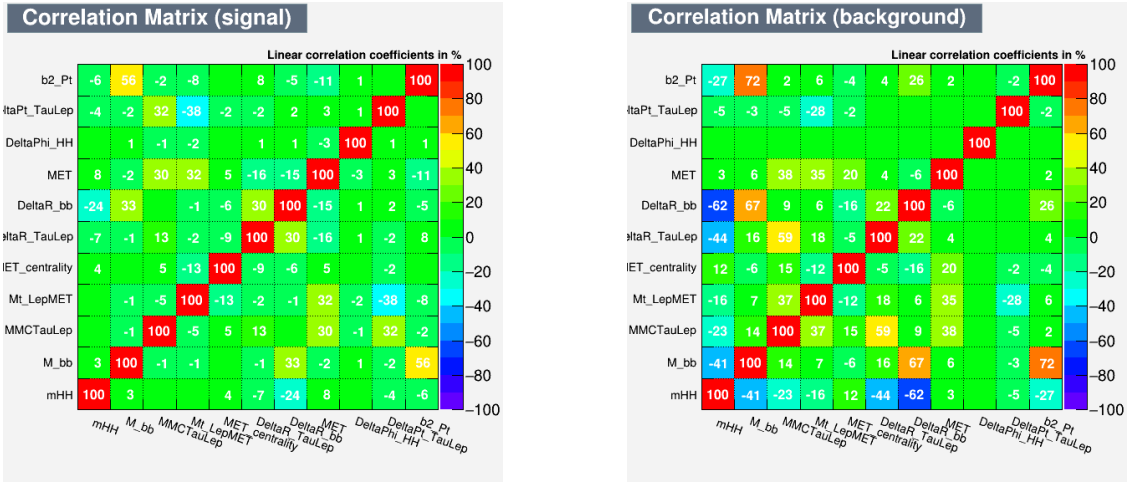


Figure 8.6: Distribution of BDT output for signal (solid) and background (hatched) MC training and test samples for 1 b-tag (left) and 2 b-tag (right) regions. (a), (b) $m_X = 300$ GeV, (c), (d) $m_X = 600$ GeV, (e), (f) $m_X = 900$ GeV.

Table 8.3: Ranking of input variables in 2 b-tag. The variables are listed according to their importance. $b2_{Pt}$ was not useful for high mass BDT training.

Ranking	m_X=900 GeV	m_X=600 GeV	m_X=300 GeV
1	m_{HH}	m_{HH}	MMC_{TauLep}
2	$DeltaPt_{TauLep}$	$DeltaPt_{TauLep}$	${m_{\mathrm{T}}}^{LepMET}$
3	m_{bb}	m_{bb}	m_{HH}
4	$DeltaR_{bb}$	MMC_{TauLep}	m_{bb}
5	$\mathrm{MMC_{TauLep}}$	$\mathrm{DeltaR_{bb}}$	$\mathrm{DeltaPt_{TauLep}}$
6	${m_{\mathrm{T}}}^{LepMET}$	${m_{\mathrm{T}}}^{LepMET}$	$DeltaR_{bb}$
7	MET	$DeltaPhi_{HH}$	$METcentrality$
8	$METcentrality$	$METcentrality$	MET
9	$DeltaPhi_{HH}$	MET	$DeltaPhi_{HH}$
10			b2_Pt

The signal at $m_X = 280$ GeV (cross-section of 1 pb) mass point as signal and $t\bar{t}$, Z+jets as background. Figures 8.9 and Figure 8.10 show the signal significance of two different b-tagging efficiency as a function of the BDT score cut in 1 and 2 b-tag categories. The significance in each bin was calculated with the number of weighted signal (N_{S}) and background (N_{B}) events with BDT scores greater than the lower edge of the bin. In the 2 b-tag region the 85% WP has better sensitivity (Figure 8.10) than the 70%, however the 70% WP showed higher sensitivity (Figure 8.9a) in the 1 b-tag region. If we use 85% efficiency for 2 b-tag and 70% for 1 b-tag, then we are double counting the events that qualify in both categories. Therefore, further checks are made in the 1 b-tag category by removing events that also belong to the 85% b-tagging WP in the 2 b-tag category. After removing the overlapped events, it is observed that there is no significant difference between in the signal significance between using the 70% WP and the 85% WP response in the 1 b-tag category (Figure 8.9b) was observed. So the same 70% b-tagging efficiency WP is adopted for both tagging regions to avoid double counting.



(a) (b)

Figure 8.7: Correlation matrix: (a) signal $m_{\mathrm{X}} = 300$ GeV and (b) $t\bar{t}$ processes.

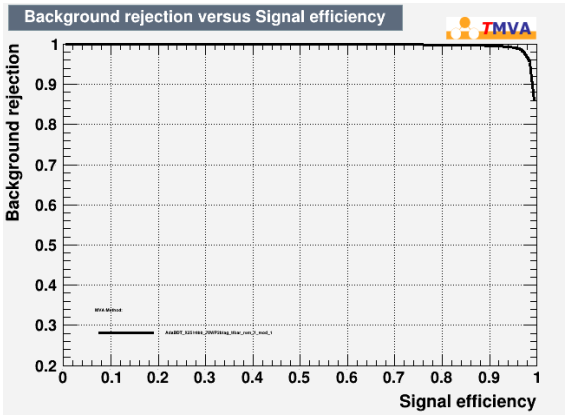
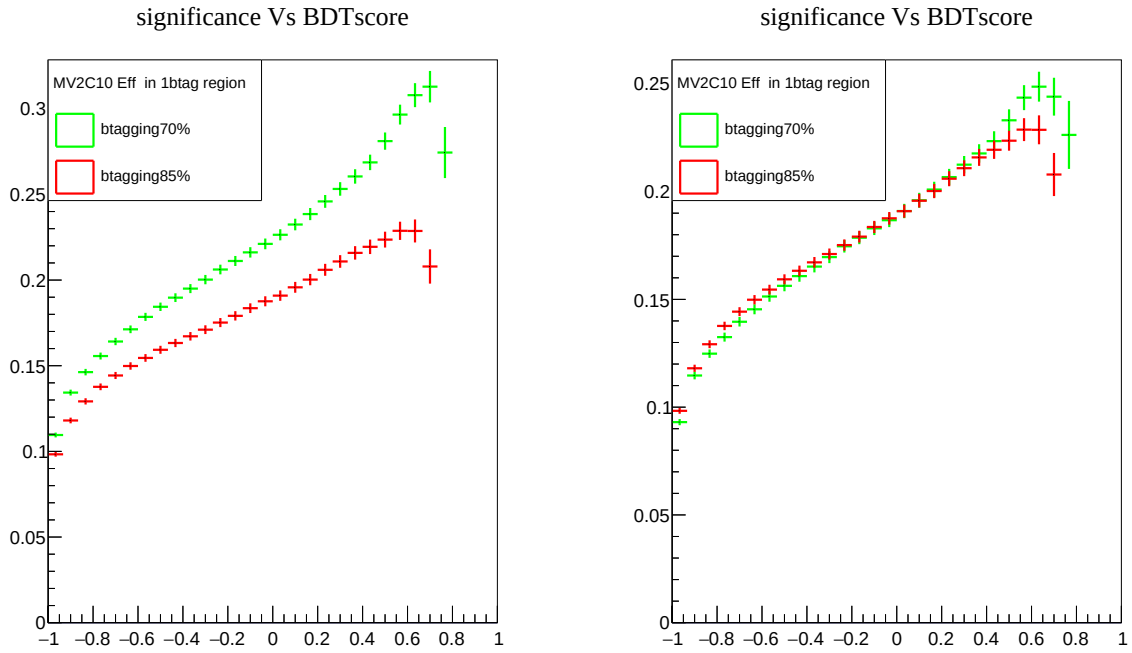


Figure 8.8: ROC Curve for m_X= 300 GeV.



(a) (b)

Figure 8.9: *b*-tagging WP optimisation (a) before, (b) after removal of overlap events in 1 *b*-tag category.

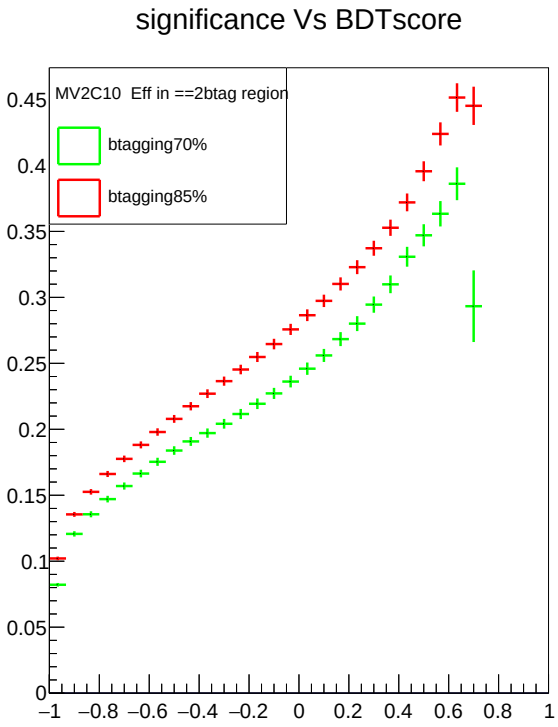


Figure 8.10: *b*-tagging WP optimisation in 2 *b*-tag category

CHAPTER 9

SYSTEMATIC UNCERTAINTIES

The uncertainty on the integral luminosity of the data was 2.1% [68]. This was applied on signal and all the backgrounds, except fakes, which were estimated by data-driven methods. In order to account for potential acceptance differences between control and signal regions, the normalization of the $t\bar{t}$ background containing true τ_{had}, determined predominantly from the SR in the final fit, is allowed to vary within a range determined by the acceptance variations associated with the $t\bar{t}$ modeling uncertainties.

The electron energy scale and resolution [69] in MC were corrected to get better agreement with data. Electron trigger, identification and reconstruction efficiencies, and uncertainties associated with the isolation requirements were also considered.

Scaling and smearing corrections were applied on the p_{T} of simulated muons [70] to have better matching between MC and data. These uncertainties were used as shape (of the BDT output distribution) dependent systematic uncertainties in the final fit. Systematics related to muons and electrons are small compared to τ_{had} and jets.

Energy scale and identification efficiency of τ_{had} [71] were corrected to match MC with data. The τ_{had} identification and reconstruction systematics had the highest variation among all variations for evaluating systematic uncertainties. The τ_{had} reconstruction uncertainty depends on pile up that effects the τ_{had} vertex. The probability to incorrectly identify the τ_{had} candidate production vertex depends on the amount of interactions reconstructed in the event. The reconstruction also depends on inner detector material, effects track counting. The largest contribution effecting the identification efficiency of $p_{\text{T}} < 100$ GeV is

the calorimeter calibration and performance, which causes variation in shapes of calorimeter based and tracking-based BDT input variables.

The jet energy scale (JES) [72, 73] uncertainty had several of primary sources, such as dijet η-inter-calibration, where the forward jet scale was calibrated with respect to central jets. This depends on p_{T} and η and is relevant for forward jets. The uncertainty for central jet was derived from "Z+jet" balancing where JES scale was calibrated with respect to a well-understood object ($Z \to ee$, $Z \to \mu\mu$). In this case the low-p_{T} (below 100 GeV) jets are affected. The medium p_{T} (100 GeV to 1 TeV) jets are affected when scale was calibrated with respect to a well understood object (photon). Lastly, the high-p_{T} (1 TeV to 2 TeV) jets were affected by the energy scale which was calibrated against the recoil system comprised of previously calibrated lower-p_{T} objects. The jet energy resolution (JER) and JES uncertainties both comes from pile up, change in topoclustering algorithm, change in noise threshold, change in bunch spacing and sampling reconstruction affects the mapping between the electronic signals and the energy measurements.

Correction factors to account for flavour-tagging efficiency differences between simulation and data are measured separately for b, c, and light-flavour jets. The calibrations and uncertainties on the mis-identification of c and light-flavour jets were derived from samples of W+c-jet events [74] and multi-jet events in Run2 data. The uncertainty on c-jet calibration was inflated by the difference measured in W+c-jet events and semi-leptonic $t\bar{t}$ decay events. Correction factors for b-jets using $t\bar{t}$ events was derived using Run 2 data [44]. All three correction factors had many sources of uncertainty and were decomposed into uncorrelated components, which were then treated independently resulting in 3 components for c-jets, 4 for b-jets, and 5 for light-flavour jets.

For resonant signals the renormalization and factorization scale is varied by a factor of 2 (up), or 0.5 (down). This variation is added with parton shower uncertainty quadratically having a 12% total effect. The uncertainty on $Z \to \tau\tau$+heavy flavour jets (bb, bc, cc)

was set to 30% [76, 11]. A 20% normalization uncertainty was applied on $Z \to e$ sample related to the mis-identification of electrons as taus, is derived by comparing data and simulation in a $Z \to ee$ CR. In this channel a 30% normalization uncertainty was assigned on the associated production of W+HF. This was estimated in the VH analysis [75]. The uncertainty for $Z \to \tau\tau$+light jets and diboson production were 5% and 6%, respectively, calculated combining PDF, α_s and scale variation uncertainties in quadrature [76].

A normalization uncertainty of 28% is applied to the ZH background based on ATLAS measurement [77]. Similarly, a 30% uncertainty was applied to the ttH background [78] .

The main backgrounds in the semi-leptonic final state are $t\bar{t}$ and fakes. The true τ contribution was estimated from simulation and comes mainly from $t\bar{t}$ in the 2b-tag region. $t\bar{t}$ nominal sample was generated in Powheg (hdamp (parameter to limit hard radiation) = $1.5m(t)$) and showered by Pythia8. To get the systematics variation from different generator and parton shower, MC@NLO and Herwig7 samples were considered respectively. Shape and normalization variations from initial state radiation (ISR) and final state radiation (FSR) samples were also included in the $t\bar{t}$ modeling uncertainty. The reweighting method for ISR and FSR samples will be discussed in detail below.

The ISR and FSR samples were reweighted following the ATLAS Particle Modeling Group (PMG) recommendation [79]. ISR sample had hdamp = $3m(t)$, at the same time dividing by 2 the renormalization (μ_{R}) and factorization (μ_{f}) scales and varying the showering (Var3cUp). The shape variation (normalized to unity) compared to the nominal $t\bar{t}$ sample is shown in Figure 9.1 (left).

The ISR down sample was obtained using nominal sample but doubling the renormalization and factorization scales and varying the showering (Var3cDown). The shape variation (normalized to one) compare to the nominal $t\bar{t}$ sample is shown in Figure 9.1 (right).

Table 9.1: Normalisation variation of $t\bar{t}$ generator systematics compared to nominal sample (Powheg+Pythia8) in the 2 b-tag region.

$t\bar{t}$ samples	Normalization variation (%)
Powheg, Herwig7	-11.5
MCNLO, Pythia8	-7.2
ISR Up	-2.7
ISR Down	0.6
FSR Up	-5.2
FSR Down	3.4

The nominal $t\bar{t}$ samples are reweighted to get the FSR up and down variation samples. The shape variation (normalized to one) of FSR up and FSR down samples compare to the nominal $t\bar{t}$ sample is shown in Figure 9.2.

Table 9.1 gives the normalisation variation of the different $t\bar{t}$ generator samples compared to the nominal sample. For the low mass BDT training MCNLO and ISR have the dominant shape variations among all the other $t\bar{t}$ modeling systematics after pruning. For the high mass BDT training all of them have shape variation compare to nominal.

The fake τ_{had}s were estimated from the data-driven method explained in Section 6.5. The uncertainty on estimated fake τ_{had}s were calculated from four different sources. First, the fake factor on Figure 7.3 were varied up to it's statistical uncertainty on both side. Second, the fake factor was calculated in the $t\bar{t}$ CR defined in Section 6.5 and then used in the signal region to estimate fake taus assuming that FF in both regions are same. The difference between FF in both regions is another source of uncertainty in the fake tau estimation. But the difference in FF (Figure 9.3) of both regions are found to be covered by the statistical error so it is not used in the the final fit.

Third, the data-driven fake τ_{had} estimation depends on the true τ_{had} estimated from simulation. The dominant true τ_{had} contribution is from $t\bar{t}$. So the fake factor uncertainty also arises from the different $t\bar{t}$ generator samples. The variation in fake factor due to

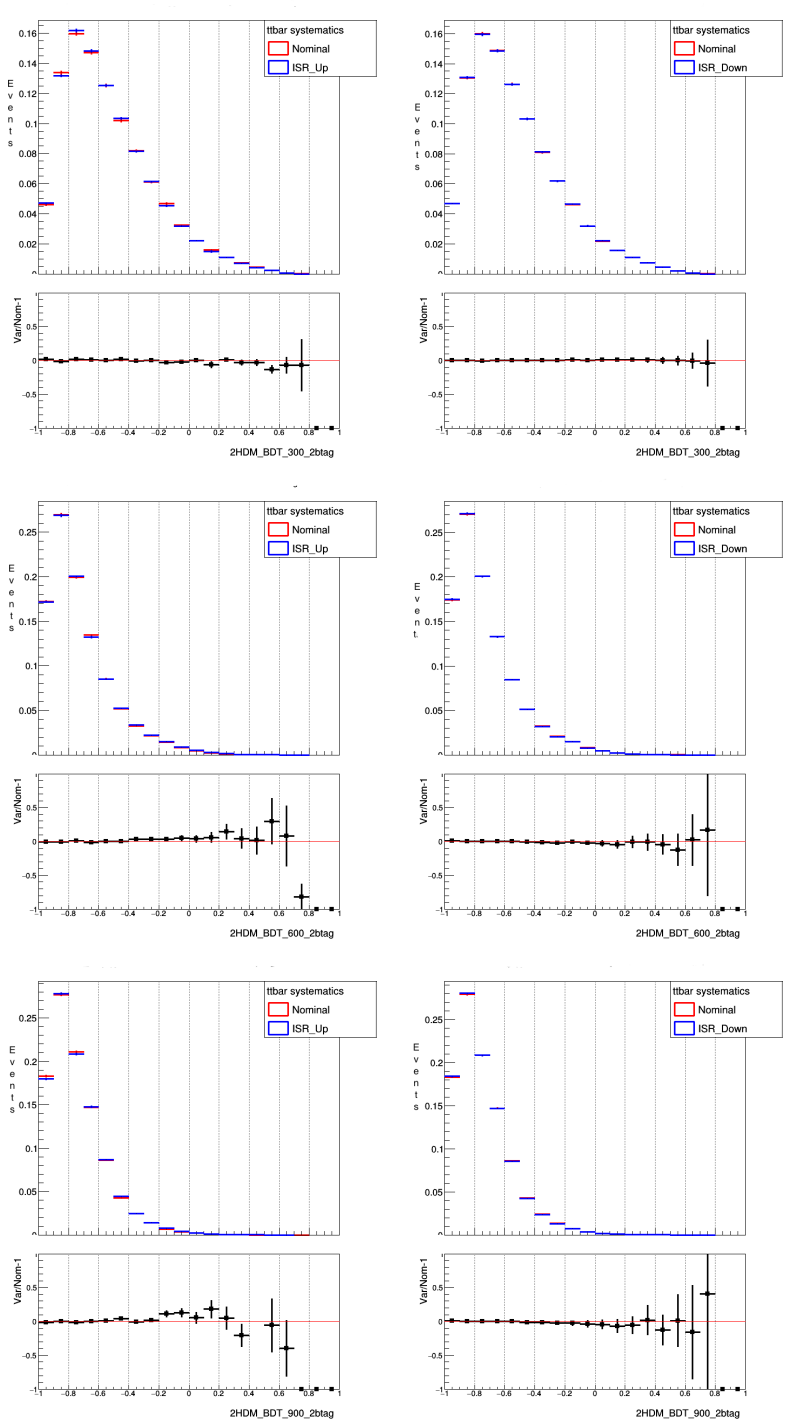


Figure 9.1: Comparison between ISR up (blue) (left) and ISR down (blue) (right) variation with nominal (red) as a function of BDT output of different signal masses and $t\bar{t}$ training.

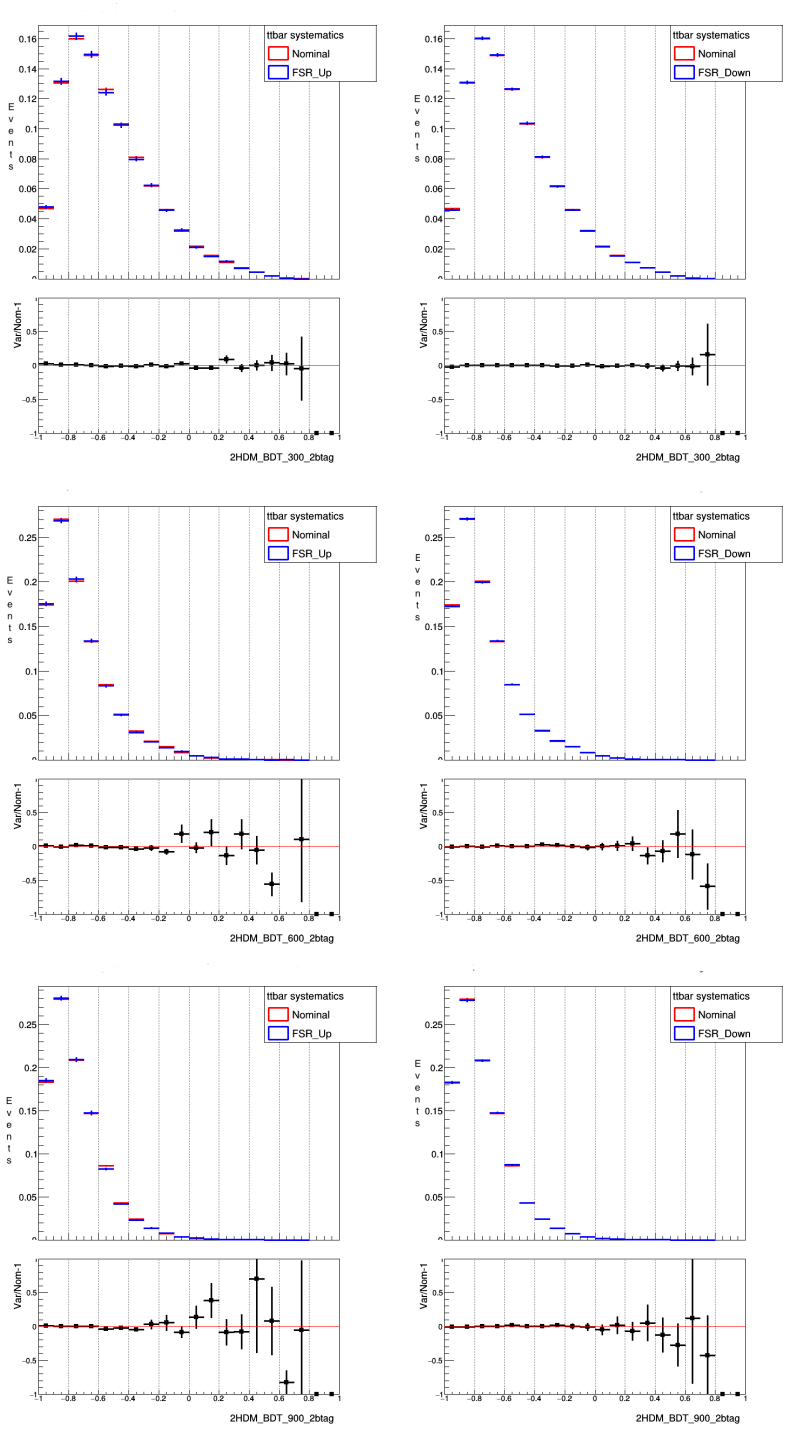
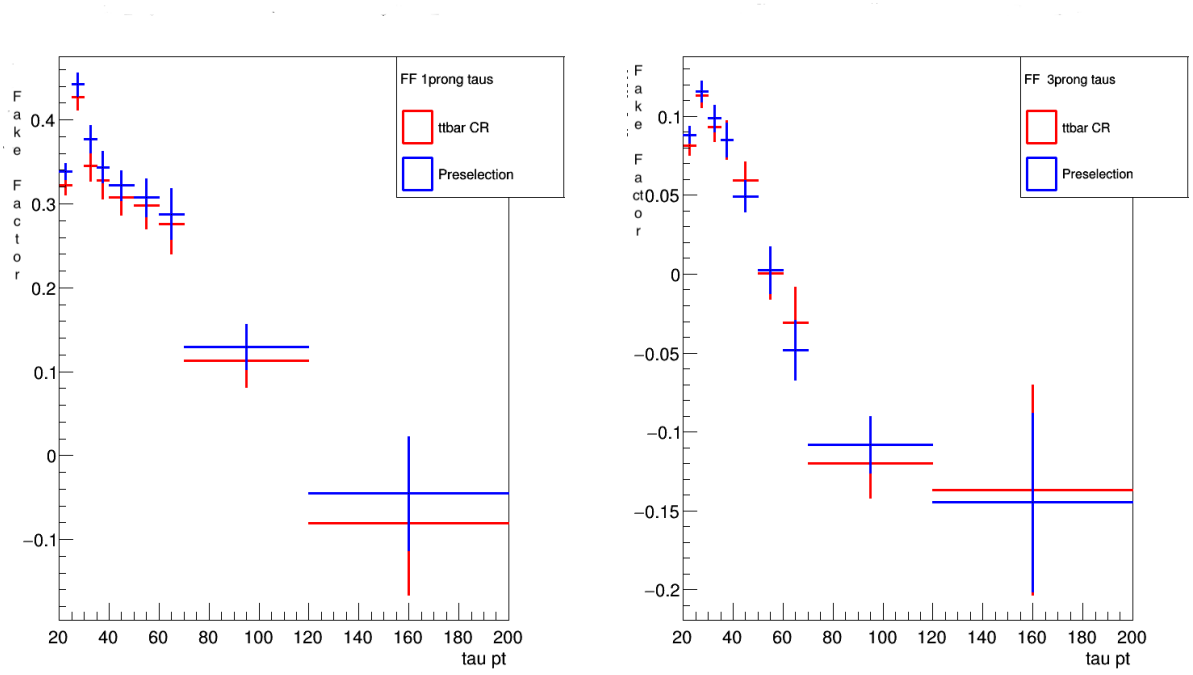


Figure 9.2: Comparison between FSR up (blue) (left) and FSR down (blue) (right) variation with nominal (red) as a function of BDT output of different signal masses and $t\bar{t}$ training.

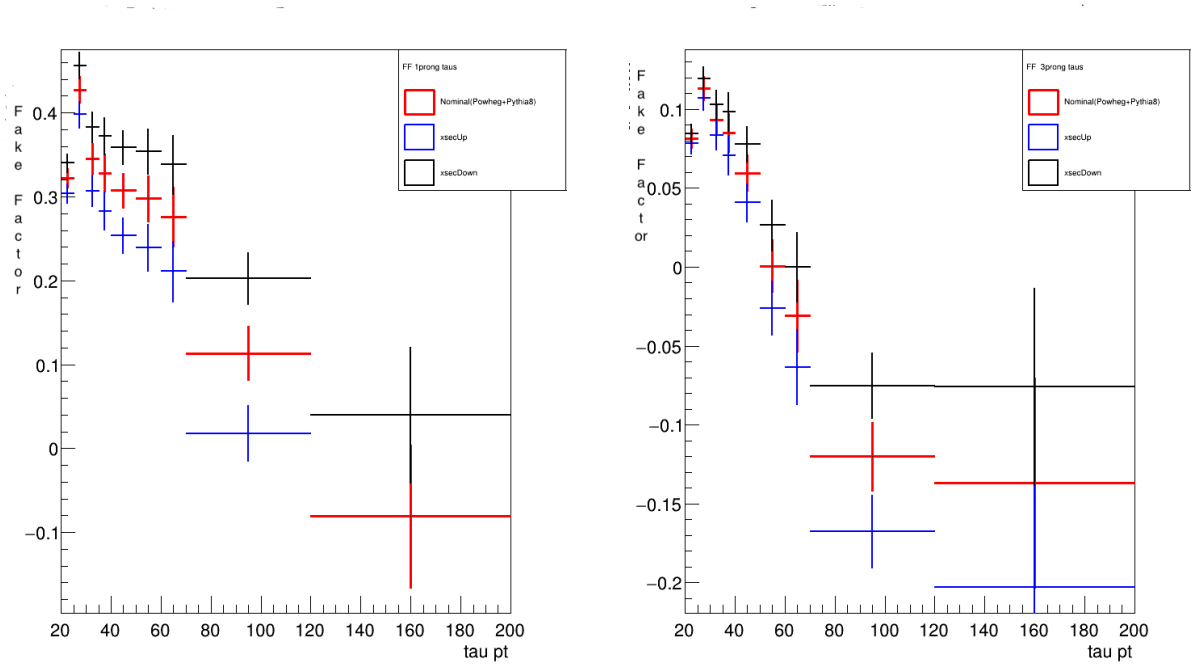
different $t\bar{t}$ generator samples is on Figure 9.4. The highest up side variation on FF is due to Herwig7 $t\bar{t}$ sample and highest down side variation is due to FSR down $t\bar{t}$ sample. For the final fit only the Herwig7 and FSR down $t\bar{t}$ sample is used to get systematic uncertainty on fake tau estimation due to different $t\bar{t}$ generator samples. The last source of fake factor uncertainty is the cross-section of $t\bar{t}$. A 6% theoretical uncertainty on the $t\bar{t}$ cross-section was included in estimating the variation of the fake factor shown in Figure 9.5.

Table 9.2 shows the effect of main systematic uncertainties on the event yield of signal (300 GeV) and main background $t\bar{t}$ in the 2 b-tag region.



(a) (b)

Figure 9.3: Fake factor in $t\bar{t}$ CR region (red) and in pre-selection (signal region) (blue) (a) 1-prong τ_{had}, (b) 3-prong τ_{had}.

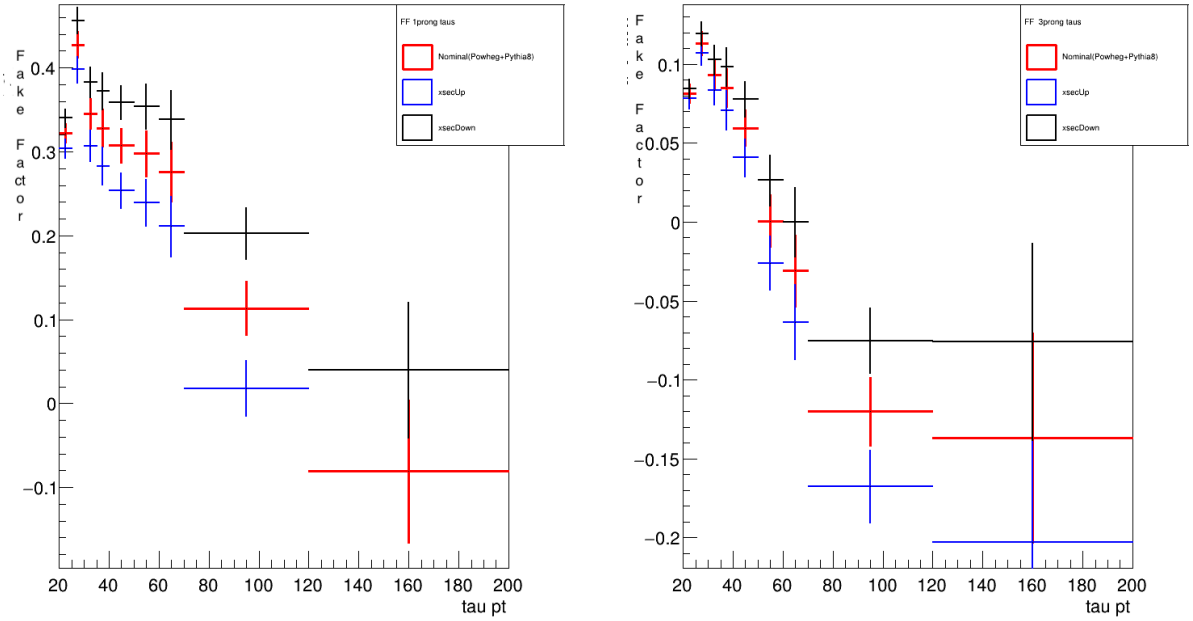


(a) (b)

Figure 9.4: Variation in fake factor from different $t\bar{t}$ generator sample (a) 1-prong τ, (b) 3-prong τ.

Table 9.2: Effect of important systematic uncertainties on the expected event yield of signal m_X= 300GeV and main background $t\bar{t}$ in the 2 b-tag region.

Systematics	$t\bar{t}$ (%)	signal m_X= 300GeV(%)
τ_{had} identification uncertainty	±6.03	± 6.42
τ_{had} reconstruction efficiency	±4.47	± 4.91
τ_{had} identification (1prong and $p_{\text{T}} > 40$ GeV)	±1.12	± 1.11
τ_{had} energy scale	±0.42	± 0.18
b-tagging efficiency	± 2.22	± 3.41
b-jet energy scale	± 0.78	± 0.68
JES uncertainty from pile up effect	± 0.71	± 0.54
JER difference between data and MC	± 0.11	± 0.31



(a) (b)

Figure 9.5: Fake factor variation from $t\bar{t}$ cross-section uncertainty (a) 1-prong τ_{had}, (b) 3-prong τ_{had}.

CHAPTER 10

RESULT

In the absence of signal in our search for a 2HDM heavy scalar (X) we set an exclusion limit with 95% CLs on the cross-section times branching ratio ($\sigma\times$BR) for each heavy scalar mass point and its decay to two SM-like Higgs boson candidates, which further decay to two b-quarks and two τs.

The BDT discriminant was used in the statistical analysis of the data, for which binned likelihood method was used, with uncertainties implemented as nuisance parameters, to determine the level of exclusion of the signal (in terms of 95% confidence level) [80, 81].

To set the exclusion limit on the $\sigma\times$ BR , BDT score distribution in different regions are used. The result section is organized as follows:

- First, 2 b-tag with 70% b-tagging WP is used as the signal region. Only the 2018 data is used here.

- Second, both 1 b-tag and 2 b-tag regions with 70% b-tagging WP are used as the signal regions. No improvement in the limit results compare to 2 b-tag only SR when only CP and detector uncertainties are taken into account on 2017 and 2018 data. This result does not include $t\bar{t}$ and fakes systematics.

- Finally, 85% b-tagging efficiency is used to get better sensitivity in 2 b-tag region but it did not show improvement for all heavy scalar masses compare to the 70% WP. This study is done using only 2017 data and CP, detector systematics.

10.0.1 2 b-tag the Signal Region

The BDT output score distribution in the 2 b-tag with 70% b-tagging WP is used as the final discriminant. The expected upper limit at 95% CL is calculated from the fit of MC to Asimov data (simulation of data) and real ATLAS data (Figure 10.1) is used to calculate the observed limit. Table 10.1 - Table 10.3 give the post-fit yields (after fitting MC to data) of the backgrounds and signal.

A spin-0 heavy scalar boson (X) as predicted by 2HDM models is used here in the narrow width approximation as benchmark model. The heavy resonance decays into two SM-like Higgs bosons of $m_H = 125$ GeV. Resonance masses between 251 GeV and 1 TeV are considered, with the lower mass range being constrained as twice the mass of the SM Higgs boson. The binning for the output BDT score distributions is optimised to ensure that the background uncertainty is kept to below 20% of the signal fraction. If there is no signal in that bin the uncertainty is required to be below 1%. In addition a minimum number of events in each bin is required 10.

Figure 10.2 shows the observed and expected limit with 2018 data of luminosity 58.5 fb^{-1}. The one (green) and two sigma (yellow) bands include statistical and all systematics uncertainties discussed in Chapter 9.

10.0.2 1 and 2 b-tag as the Signal Region

Theoretically, by using 70% b-tagging WP, we select 49% of the signals in the 2 b-tag category. By including the 1 b-tag category, we will have 42% of signal that could improve the sensitivity depending on the background rate in 1 b-tag. To investigate if 1 b-tag adds anything to the signal sensitivity a simultaneous fit with Asimov data in the 1 and 2 b-tag

Table 10.1: Postfit yield from BDT output when training is done between $t\bar{t}$ and signal of mass 300 GeV, using 2018 data/MC16e in 2b-tag region, signal cross-section is set at 1 pb.

Sample	Postfit Yield (syst+stat uncertainty)
Fakes	18926.11 ± 1426.67
ZttHF	441.73 ± 132.65
$t\bar{t}$	33747.40 ± 1422.74
stop	2172.21± 176.43
ZttLF	12.10± 4.67
ZHF	401.52 ± 125.19
ZLF	18.98 ± 7.78
W+jets	27.61 ± 8.61
Wt+jets	2.06 ± 0.16
diboson	43.78 ± 4.15
VH	24.29 ± 7.27
ttH	68.69 ± 21.12
Total Background	55886.52 ± 399.44
Data	55894
signal m_X=300 GeV	21.98 ± 3.47

Table 10.2: Postfit yield from BDT output when training is done between $t\bar{t}$ and signal of mass 600 GeV, using 2018 data/MC16e in 2 b-tag region, signal cross-section is set at 1 pb.

Sample	Postfit Yield (syst+stat uncertainty)
Fakes	16547.81 ± 2447.53
ZttHF	496.54± 93.87
$t\bar{t}$	36046.47 ± 2917.86
stop	2149.49± 161.48
ZttLF	11.79 ± 4.48
ZHF	451.61 ± 89.65
ZLF	18.42± 7.54
W+jets	26.90± 8.32
Wt+jets	2.06 ± 0.16
diboson	43.51 ± 3.81
VH	25.26 ± 7.02
ttH	69.16 ± 20.98
Total background	55889 ± 705
Data	55894
Signal m_X=600 GeV	85.61 ± 12.85

Table 10.3: Postfit yield from BDT output when training is done between $t\bar{t}$ and signal of mass 900 GeV, using 2018 data/MC16e in 2 b-tag region, signal cross-section is set at 1 pb.

Sample	Postfit Yield (syst+stat uncertainty)
Fakes	15942.60 ± 2640.54
ZttHF	524.63± 96.20
$t\bar{t}$	36552.64 ± 2941.86
stop	2193.30 ± 160.70
ZttLF	12.89± 4.74
ZHF	470.90 ± 90.40
ZLF	19.84 ± 7.91
W+jets	27.70 ± 8.42
Wt+jets	2.03 ± 0.18
diboson	44.20 ± 3.75
VH	25.54 ± 7.11
ttH	70.11 ± 21.22
Total background	55886 ± 559
Data	55894
Signal m_X= 900 GeV	132.53 ± 19.67

Table 10.4: Limit including statistical and systematic uncertainties from combine (2017+ 2018) in the 2 b-tag category.

m_X in GeV	-2σ	-1σ	expected limit in fb	+1 σ	+2σ
251	38.55	51.75	71.82	99.96	134.00
260	74.87	100.52	139.50	194.15	260.27
280	96.10	129.02	179.05	249.19	334.06
300	97.03	130.27	180.79	251.61	337.30
325	85.13	114.28	158.61	220.73	295.91
350	65.86	88.42	122.71	170.78	228.94
400	32.81	44.05	61.13	85.08	114.06
450	20.20	27.12	37.63	52.38	70.21
500	12.62	16.94	23.51	32.71	43.86
550	7.40	9.94	13.79	19.20	25.73
600	6.07	8.15	11.31	15.74	21.10
700	4.60	6.17	8.57	11.92	15.99
800	3.27	4.39	6.09	8.48	11.37
900	2.38	3.19	4.43	6.17	8.27
1000	2.09	2.81	3.89	5.42	7.26

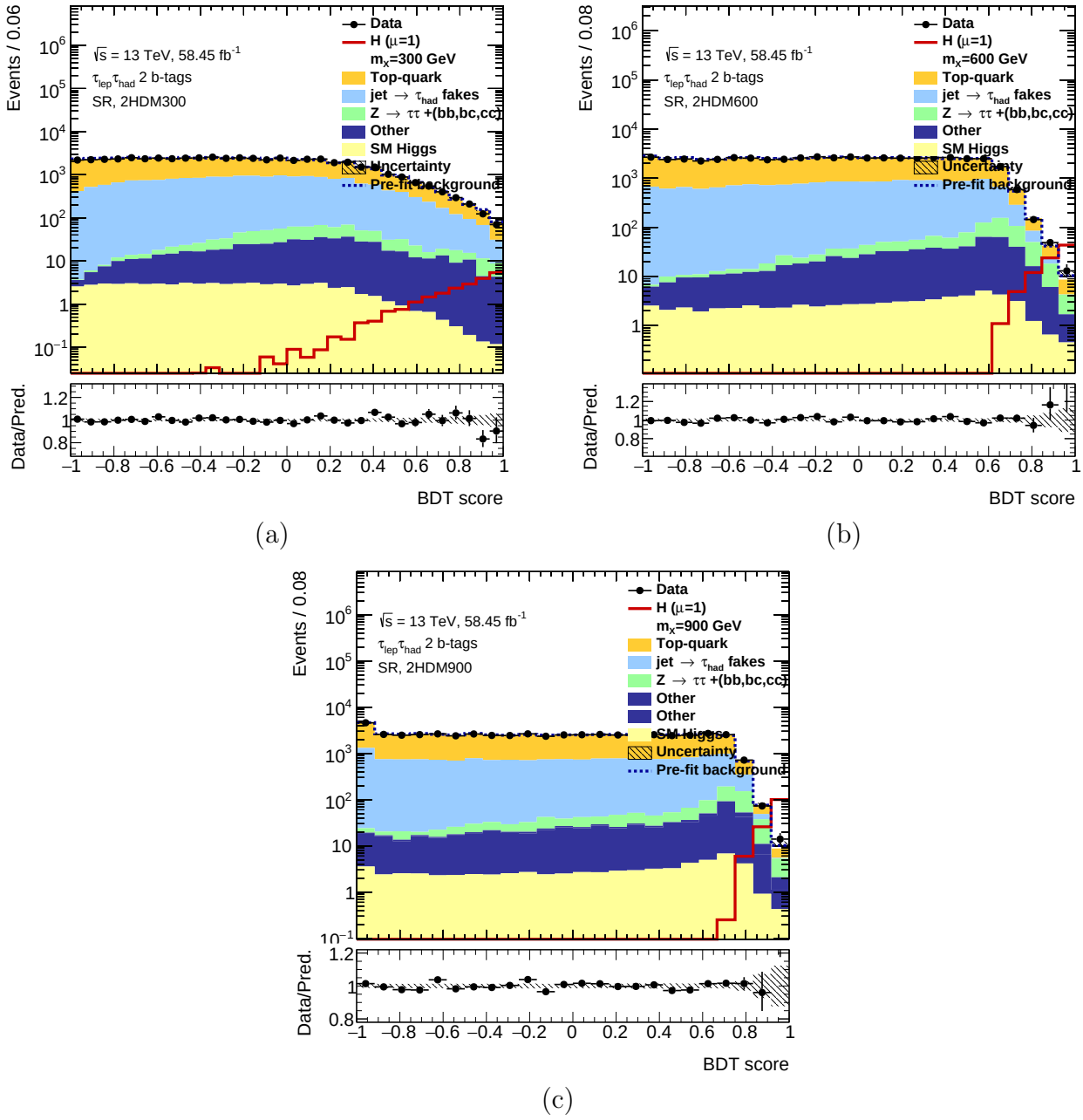


Figure 10.1: Postfit (MC fitted to data) distributions of the BDT output of (a) m_X=300 GeV and $t\bar{t}$, (b) m_X=600 GeV and $t\bar{t}$, (c) m_X=900 GeV and $t\bar{t}$ using 2018 data and MC16e in the 2 b-tag region. The signal normalization corresponds to a cross section of 1 pb.

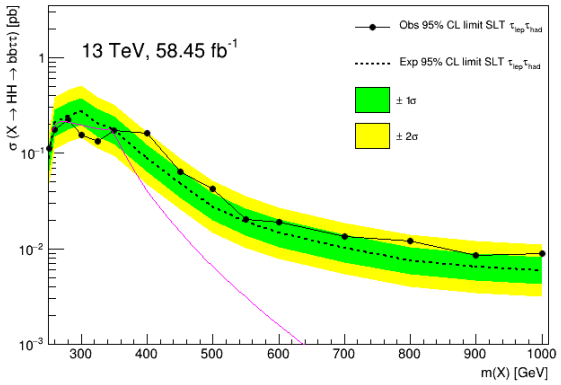


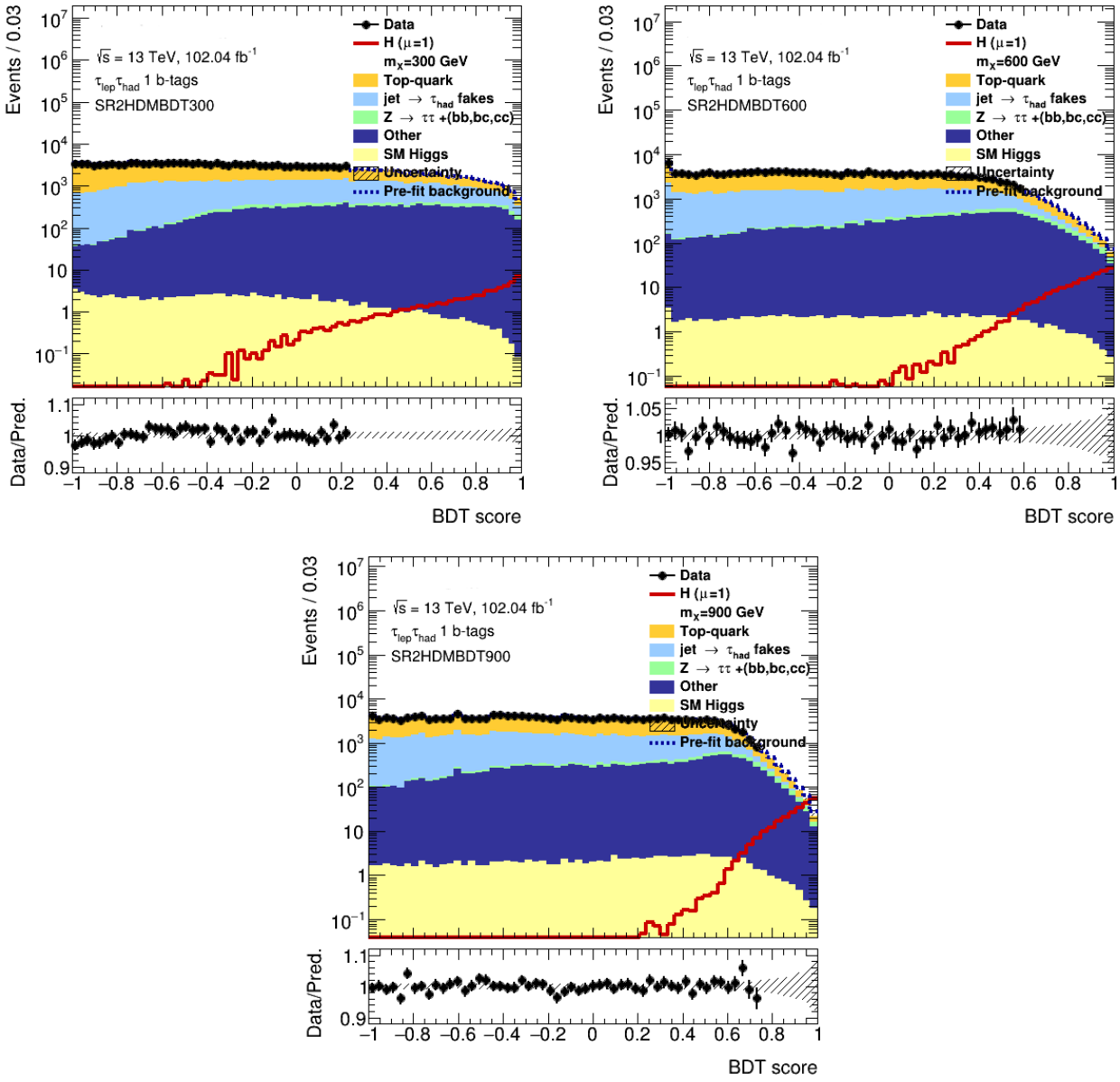
Figure 10.2: Upper limit on the signal cross section as a function of the mass of the heavy scalar X all systematics and statistical uncertainties using the 2018 data and MC16e. The pink line shows the Type 2 2HDM cross-section×branching ratio at $\tan\beta$=2.

regions was performed. The postfit plot in 1 b-tag region for three heavy scalar masses are shown in Figure 10.3. In these plots the data is blinded in the higher BDT score where 90% signal appears.

The expected upper limit with statistical and systematic uncertainties (Table 10.5) gets weaker than 2 b-tag only fit (Table 10.4) for some of the heavy scalar masses. Only systematics from combine performance group and detector systematics are included in both b-tagging regions for this study. Due to lack of improvement this choice of signal region is not used in the analysis.

10.0.3 2 b-tag with 85% b-tagging efficiency as the Signal Region

As shown in Chapter 8, that 85% b-tagging WP for 2 b-tag category is the best optimal choice after the 1 and 2 b-tag combined fit at the 70% WP. That study was only done for the lowest mass of the heavy scalar, X, namely m_X = 280 GeV. This section shows the gain in sensitivity when the fit is done using the 2 b-tag BDT score with 85% b-tagging WP. Table 10.7 shows that the sensitivity increases in low mass points but we lose sensitivity



(c)

Figure 10.3: Postfit (MC fitted to data) distributions of the BDT output of (a) m_X=300 GeV and $t\bar{t}$, (b) m_X=600 GeV and $t\bar{t}$, (c) m_X=900 GeV and $t\bar{t}$ using 2018 data and MC16e in the 1 b-tag region. The signal normalization corresponds to a cross section of 1 pb.

Table 10.5: Limit with systematics and statistical uncertainty from simultaneous fit of 1 and 2 b-tag BDT score distributions using 2017 and 2018 data. The change is calculated with respect to Table 10.4.

m_X in GeV	-2σ	-1σ	expected limit in fb	+1 *sigma*	+2 *sigma*	change(%)
251	38.87	52.18	72.42	100.78	135.11	0.83
260	85.23	114.42	158.80	221.00	296.27	13.83
280	112.39	150.89	209.40	291.43	390.68	16.95
300	111.76	150.04	208.23	289.79	388.48	15.17
325	100.04	134.30	186.39	259.40	347.74	17.51
350	75.96	101.98	141.52	196.96	264.04	15.32
400	41.00	55.04	76.39	106.31	142.52	24.96
450	23.03	30.91	42.90	59.71	80.04	14.04
500	12.48	16.76	23.25	32.36	43.38	-1.11
550	7.50	10.07	13.98	19.46	26.09	1.37
600	5.50	7.38	10.25	14.26	19.12	-9.37
700	4.22	5.66	7.86	10.94	14.66	-8.28
800	2.88	3.87	5.37	7.47	10.02	-11.82
900	2.16	2.90	4.03	5.60	7.51	-9.02
1000	1.95	2.62	3.63	5.05	6.77	-6.68

for high mass points compare to Table 10.6. This study was only done with 2017 data and MC16d. Due to the lack of a clear improvement in the 85% WP compare to 70%, this study was not further carried out.

10.0.4 Summary and Conclusion

We have carried out a search of 2HDM heavy scalar particle (X) in the LHC by the ATLAS detector using the 2018 data corresponding to a total luminosity of 58.5 fb^{-1}. The MC is fitted to data in the BDT output distribution in the 2 b-tag region of 70% b-tagging WP. Having found no evidence of signal in data, we set an upper limit on the cross-section times branching ratio of the heavy scalar production followed by its decay to two SM-like Higgs bosons that subsequently decay to the $b\bar{b}\tau^+\tau^-$ final state and no more than one of

Table 10.6: Limit including statistical and systematic uncertainties in 2btag with 70% b-tagging efficiency from 2017/MC16d.

m_X in GeV	-2σ	-1σ	expected limit in fb	+1 σ	+2σ
251	69.29	93.02	129.10	179.67	240.86
260	117.17	157.30	218.31	303.82	407.29
280	166.11	223.01	309.49	430.72	577.42
300	184.73	248.00	344.18	479.00	642.14
325	140.37	188.45	261.53	363.97	487.93
350	111.98	150.33	208.63	290.35	389.23
400	53.48	71.80	99.64	138.68	185.91
450	31.67	42.51	59.00	82.11	110.08
500	20.73	27.83	38.62	53.75	72.05
550	13.53	18.16	25.20	35.08	47.02
600	10.50	14.09	19.56	27.22	36.49
700	7.45	10.00	13.88	19.31	25.89
800	5.80	7.78	10.80	15.03	20.15
900	4.60	6.17	8.57	11.92	15.98
1000	4.02	5.40	7.50	10.44	13.99

Table 10.7: Limit including statistical and systematic uncertainties in 2btag with 85% b-tagging efficiency from 2017/MC16d. Change is calculated comparing Table 10.6.

m_X in GeV	-2σ	-1 σ	expected limit in fb	+1 σ	+2 σ	change(%)
251	62.38	83.75	116.23	161.75	216.84	-9.9
260	116.86	156.89	217.73	303.02	406.22	-0.2
280	141.48	189.94	263.60	366.86	491.80	-14.8
300	153.94	206.67	286.82	399.17	535.12	-16.6
325	125.59	168.61	234.00	325.66	436.57	-10.5
350	97.51	130.90	181.67	252.83	338.94	-12.9
400	47.85	64.24	89.15	124.08	166.33	-10.5
450	31.09	41.74	57.92	80.61	108.06	-1.8
500	22.56	30.29	42.03	58.50	78.42	8.8
550	15.47	20.77	28.82	40.11	53.77	14.3
600	12.57	16.88	23.43	32.60	43.70	19.7
700	7.93	10.65	14.78	20.57	27.57	6.4
800	5.93	7.96	11.04	15.37	20.60	2.2
900	4.27	5.73	7.95	11.06	14.83	-7.2
1000	3.77	5.06	7.03	9.78	13.11	-6.2

the τs decaying to a charged lepton. The observed upper limit on σ.BR of the heavy scalar mass of 251 GeV $< m_X <$ 1000 GeV is set between 110.73 fb and 8.97 fb.

REFERENCES

[1] M.Tanabashi. "Particle Data Group." *Phys.Rev.D 98, 030001 (2018) and 2019 update* http://pdg.lbl.gov/2019/tables/rpp2019-sum-leptons.pdf

[2] Standard Model, from Wikipedia. https://en.wikipedia.org/wiki/Standard_Model.

[3] Higgs, P. (1964). "Broken Symmetries and the Masses of Gauge Bosons". Physical Review Letters." *1964PhRvL..13..508H.* doi:10.1103/PhysRevLett.13.508.

[4] G. Branco et al. "Theory and phenomenology of two-Higgs-doublet models." *Phys. Rept. 516 (2012) 1.* arXiv: 1106.0034 [hep-ph].

[5] G. Branco et al. "Two-Higgs doublet model and the LHC." http://inspirehep.net/record/1720213.

[6] "LHC Higgs Cross Section Working Group." https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[7] S. Heinemeyer et al. "LHC Higgs Cross Section Working Group." arXiv:1307.1347 [hep-ph] (2013).

[8] 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) 1, arXiv: 1207.7214 [hep-ex].* https://arxiv.org/abs/1207.7214.

[9] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." *Phys. Lett. B 716 (2012) 30, arXiv: 1207.7235 [hep-ex].* https://arxiv.org/abs/1207.7235.

[10] CDF experiment. "Search for a low mass Standard Model Higgs boson in the tau tau decay channel." https://www-cdf.fnal.gov/physics/new/hdg/Results_files/results/smtautau_jul10/analysis.htm

[11] ALTAS Collaboration. "Search for Resonant and Nonresonant Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ Decay Channel in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector." *Physical Review Letters 121, 191801 (2018).* https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.121.191801.

[12] Bruning, Oliver Sim, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. 2004. LHC Design Report. CERN Yellow Reports: Monographs. Geneva: CERN. https://cds.cern.ch/record/ 782076.

[13] Mobs, Esma. "The CERN accelerator complex." https://cds.cern.ch/record/2197559.

[14] R. Bruce, G. Arduini, H. Bartosik, R. de Maria, M. Giovannozzi, G. Iadarola, J.M. Jowett, K. Li, M. Lamont, A. Lechner, E. Metral, D. Mirarchi, T. Pieloni, S. Redaelli, G. Rumolo, B. Salvant, R. Tomas, J. Wenninger. "LHC RUN 2: Results and Challenges." http://accelconf.web.cern.ch/AccelConf/hb2016/papers/moam5p50.pdf.

[15] ATLAS Collaboration, G. Ada *etal*. 2008. "The ATLAS Experiment at the CERN Large Hadron Collider." *JINST 3: S08003.* https://doi.org/10.1088/1748-0221/3/08/S08003.

[16] CMS Collaboration, S.Chatrchyan *etal*. 2008. "The CMS experiment at the CERN Large Hadron Collider." *JINST 3 S08004.* https://cds.cern.ch/record/1129810.

[17] LHCb Collaboration, A.Augusto *etal*. 2008. "The LHCb Detector at the CERN Large Hadron Collider." *JINST 3 S08005.* https://cds.cern.ch/record/1129809/files/jinst8_08_s08005.pdf.

[18] ALICE Collaboration, K.Aamodt *etal*. 2008. "The ALICE experiment at the CERN Large Hadron Collider." *JINST 3 S08002.* http://cds.cern.ch/record/1129812.

[19] TOTEM Collaboration, G.Anelli *etal*. 2008. "The TOTEM experiment at the CERN Large Hadron Collider." *JINST 3 S08007.* https://cds.cern.ch/record/1129807.

[20] LHCf Collaboration, O.Adrian *etal*. 2008. "The LHCf experiment at the CERN Large Hadron Collider." *JINST 3 S08006.* https://cds.cern.ch/record/1129808.

[21] Zachary Marshall for the ATLAS Collaboration. "Simulation of Pile-up in the ATLAS Experiment." https://iopscience.iop.org/article/10.1088/1742-6596/513/2/022024/pdf.

[22] Tsiskaridze S. (2017) The ATLAS Experiment at the Large Hadron Collider. "Search for Flavor-Changing Neutral Current Top Quark Decays $t \to Hq$, with $H \to b\bar{b}$, in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector." https://doi.org/10.1007/978-3-319-63414-2_2.

[23] Tiago Barreiro Goncalves. "3D Printing the ATLAS Barrel Toroid." https://cds.cern.ch/record/2210547/files/Report.pdf.

[24] ATLAS Collaboration. 1997. "ATLAS central solenoid : Technical Design Report." https://cds.cern.ch/record/331067.

[25] ATLAS Collaboration. "ATLAS barrel toroid: Technical Design Report." https://cds.cern.ch/record/331065.

[26] Pequenao, Joao. 2008. "Computer generated image of the whole ATLAS detector. https://cds.cern.ch/record/1095924.

[27] ATLAS Collaboration. 1997. "ATLAS inner detector: Technical Design Report 1." https://cds.cern.ch/record/331063.

[28] Pequenao, Joao. 2008. "Computer generated image of the ATLAS inner detector." https://cds.cern.ch/images/CERN-GE-0803014-01.

[29] ATLAS Collaboration. 1998. "ATLAS pixel detector : Technical Design Report." *Journal of Instrumentation, Volume 3, February 2008.* http://cds.cern.ch/record/381263.

[30] The ATLAS TRT Collaboration. "The ATLAS TRT Barrel Detector." https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02014/meta.

[31] Henric Wilkens and the ATLAS LArg Collaboration. 2009. "The ATLAS Liquid Argon calorimeter: An overview." *J. Phys.: Conf. Ser. 160 012043.* https://iopscience.iop.org/article/10.1088/1742-6596/160/1/012043/pdf.

[32] Pequenao, Joao. 2008. "Computer generated image of the ATLAS Liquid Argonne. http://cdsweb.cern.ch/record/1095928.

[33] ATLAS Collaboration. "Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter." https://cds.cern.ch/record/2285583.

[34] ATLAS Collaboration. 1997. "ATLAS muon spectrometer: Technical Design Report." https://cds.cern.ch/record/331068.

[35] Pequenao, Joao. 2008. "Computer generated image of the ATLAS calorimeter. https://cds.cern.ch/record/1095927.

[36] ATLAS Collaboration. "Performance of the ATLAS muon trigger in pp collisions at ps = 8 TeV". *Eur. Phys. J. C75 (2015), p. 120. doi: 10.1140/epjc/s10052-015-3325-9.* arXiv: 1408.3179 [hep-ex].

[37] Patrick Czodrowski, ATLAS Collaboration. "ATLAS Trigger System : Ready for RunII." https://pos.sissa.it/234/273/pdf.

[38] ATLAS Data Quality. https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/DataSummary/2018/filldata/figures/peakLumiByFill.png

[39] ATLAS Collaboration. " Electron identification measurements in ATLAS using $\sqrt{s}$ =13 TeV data with 50 ns bunch spacing." https://cds.cern.ch/record/2048202/files/ATL-PHYS-PUB-2015-041.pdf.

[40] ATLAS Collaboration. "Photon identification in 2015 ATLAS data." http://cdsweb.cern.ch/record/2203125/files/ATL-PHYS-PUB-2016-014.pdf.

[41] Peter Berta, On behalf of the ATLAS Collaboration. "ATLAS jet and missing-ET reconstruction, calibration, and performance." *Nuclear and Particle Physics Proceedings 273275 (2016) 11211126.* https://doi.org/10.1016/j.nuclphysbps.2015.09.176.

[42] ATLAS Collaboration. " Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC." http://cdsweb.cern.ch/record/2064383/files/ATL-PHYS-PUB-2015-045.pdf.

[43] ATLAS Collaboration. "Muon reconstruction performance of theATLAS detector in protonproton collision data at $\sqrt{(s)}$=13 TeV." *Eur. Phys. J. C76 (2016) 292.* http://dx.doi.org/10.1140/epjc/s10052-016-4120-y.

[44] ATLAS Collaboration. "Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run." https://cds.cern.ch/record/2160731/files/ATL-PHYS-PUB-2016-012.pdf.

[45] ATLAS Collaboration. "Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at $\sqrt{(S)}$= 13 TeV ." http://cdsweb.cern.ch/record/2037904.

[46] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer. "MadGraph 5: going beyond." *JHEP 06 (2011) 128.* arXiv: 1106.0522 [hep-ph].

[47] J. Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note." *Eur. Phys. J. C76 (2016) 196.* arXiv: 1512.01178 [hep-ph].

[48] S. Alioli, P. Nason, C. Oleari and E. Re. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX." *JHEP 06 (2010) 043.* arXiv: 1002.2581 [hep-ph].

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

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

[51] Michael H. Seymour and Marilyn Marx. "Monte Carlo Even Generation." https://arxiv.org/pdf/1304.6677.pdf.

[52] N. Srimanobhas. "Introduction to Monte Carlo for Particle Physics Study." https://indico.cern.ch/event/92209/contributions/2114409/attachments/1098701/1567290/CST2010-MC.pdf.

[53] P. Skands. "Tuning Monte Carlo generators: The Perugia tunes." *Phys. Rev. D 82 (2010) 074018.* arXiv: 1005.3457 [hep-ph].

[54] D. J. Lange. "The EvtGen particle decay simulation package." *Nucl. Instrum. Meth. A 462 (2001) 152.*

[55] M. Czakon, M. L. Mangano, A. Mitov and J. Rojo, "Constraints on the gluon PDF from top quark pair production at hadron colliders." JHEP 07 (2013) 167. arXiv: 1303.7215 [hep-ph].

[56] T. Gleisberg and S. Hche. "Comix, a new matrix element generator." *JHEP 12 (2008) 039.* arXiv: 0808.3674 [hep-ph].

[57] F. Cascioli, P. Maierhofer and S. Pozzorini. "Scattering Amplitudes with Open Loops." *Phys. Rev. Lett. 108 (2012) 111601.* arXiv: 1111.5206 [hep-ph].

[58] S. Schumann and F. Krauss. "A Parton shower algorithm based on Catani-Seymour dipole factorisation." *JHEP 03 (2008) 038.* arXiv: 0709.1027 [hep-ph].

[59] S. Hche, F. Krauss, M. Schnherr and F. Siegert. "QCD matrix elements + parton showers: The NLO case." *JHEP 04 (2013) 027.* arXiv: 1207.5030 [hep-ph].

[60] ATLAS Collaboration. "ATLAS Pythia 8 tunes to 7 TeV data". *ATL-PHYS-PUB-2014-021, 2014.* https://cds.cern.ch/record/1966419.

[61] S. Agostinelli et al. "GEANT4 - a simulation toolkit." *Nucl. Instrum. Meth. A 506 (2003).*

[62] ATLAS Collaboration. "The ATLAS Simulation Infrastructure." Eur. Phys. J. C 70 (2010) 823. arXiv: 1005.4568 [physics.ins-det].

[63] M. Cacciari, G. P. Salam and G. Soyez. "The anti-kt jet clustering algorithm." *JHEP 04 (2008) 063.* arXiv: 0802.1189 [hep-ph].

[64] ATLAS Collaboration. "Jet energy scale and its systematic uncertainty in proton-proton collisions at s = 7 TeV in ATLAS 2010 data, tech. rep." https://arxiv.org/pdf/1112.6426.pdf.

[65] ATLAS Collaboration. "Tagging and suppression of pileup jets with the ATLAS detector." *ATLAS-CONF-2014-018, 2014.* https://cds.cern.ch/record/1700870.

[66] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss. "TMVA 4,Toolkit for Multivariate Data Analysis with ROOT, Users Guide." https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf.

[67] A. Elagin, P. Murat, A. Pranko and A. Safonov. "A new mass reconstruction technique for resonances decaying to $\tau\tau$." *Nucl. Instrum. Meth. A 654 (2011) 481.* arXiv: 1012.4686 [hep-ex].

[68] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{8}$= 8 TeV using the ATLAS detector at the LHC." *Eur. Phys. J. C 76, 653 (2016).*

[69] ATLAS Collaboration. "Electron identification measurements in ATLAS using $\sqrt{s}$=13 TeV data with 50 ns bunch spacing." https://cds.cern.ch/record/2048202.

[70] ATLAS Collaboration. "Muon reconstruction performance in early 13 TeV data." https://cds.cern.ch/record/2047831.

[71] ATLAS Collaboration. "Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$=13 TeV." https://cds.cern.ch/record/2261772.

[72] ATLAS Collaboration. "et Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s}$=13 TeV." https://cds.cern.ch/record/2037613.

[73] ATLAS Collboration. "In *situ* calibration of large-radius jet energy and mass in 13 TeV protonproton collisions with the ATLAS detector." https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/JETM-2018-02/.

[74] ATLAS Collaboration. " Calibration of the b-tagging efficiency for c jets with the ATLAS detector using events with a W boson produced in association with a single C quark." https://cds.cern.ch/record/1640162/files/ATLAS-CONF-2013-109.pdf.

[75] S. Argyropoulos et al. "Search for resonances decaying to a W or Z boson and a Higgs boson in the bb, lbb and llbb final states with s = 13 TeV ATLAS data in the context

of models with 2 Higgs doublets or additional heavy vector triplets." *tech. rep. ATL-COM-PHYS-2016-479, CERN, 2016.* https://cds.cern.ch/record/2151842.

[76] ATLAS Collaboration. "Measurements of the production cross section of a Z boson in association with jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector." https://arxiv.org/abs/1702.05725.

[77] ATLAS Collaboration. "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector." *J. High Energy Phys. 12 (2017) 024.*

[78] ATLAS Collaboration. "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector." Phys. Rev. D 97, 072003 (2018).

[79] ATLAS Collaboration. "Systematic affecting Top MC for 13 TeV analysis in release 21." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopMCSystematicsR21.

[80] Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells. "Asymptotic formulae for likelihood-based tests of new physics." https://arxiv.org/pdf/1007.1727.pdf.

[81] Kyle Cranmer. "Practical Statistics for the LHC." https://arxiv.org/abs/1503.07622.

[82] "Signal Significance in Particle Physics." https://arxiv.org/pdf/hep-ex/0208005.pdf.

[83] Max Baak, Geert-Jan Besjes, Jeanette Lorenz, Sophio Pataraia. "Introduction to statistics, statistical tools and the HistFitter framework." https://indico.cern.ch/event/524296/contributions/2147359/attachments/1288747/1918288/HistFitter_Stats_Introduction.pdf.

[84] A. Gomes, on behalf of the ATLAS Tile Calorimeter System. "The new Front-End Electronics for the ATLAS Tile Calorimeter Phase 2 Upgrade." https://cds.cern.ch/record/2066732/files/ATL-TILECAL-PROC-2015-017.pdf.

[85] https://twiki.cern.ch/twiki/bin/viewauth/Atlas/TileCalTestbeamAnalysis2016

[86] ATLAS Collaboration. "Readiness of the ATLAS Tile Calorimeter for LHC collisions." https://arxiv.org/pdf/1007.5423.pdf.

[87] J. M. de Seixas, on behalf of the ATLAS Collaboration. "The TileCal Energy Reconstruction for LHC Run2 and Future Perspectives." https://arxiv.org/pdf/1510.01690.pdf.

[88] ATLAS Collaboration. "ATLAS Data Quality Operations and Performance for 2015-2018 data-taking." https://arxiv.org/abs/1911.04632.

[89] R. Trentadue et al. " LCG Persistency Framework (CORAL, COOL, POOL): Status and Outlook in 2012." *J. Phys.: Conf. Ser. 396 (2012) 052067 (cit. on p. 5).*

[90] A. Barriuso Poy et al. "The detector control system of the ATLAS experiment." JINST 3 (2008) 05006 (cit. on p. 6).

[91] P. Adragna et al. " GNAM: A low-level monitoring program for the ATLAS experiment." *IEEE Trans. Nucl. Sci. 53 (2006) 1317 (cit. on p. 6).*

[92] D. Salvatore et al.. "The GNAM system in the ATLAS online monitoring framework." Nucl. Phys. Proc. Suppl. 172 (2007) 317 (cit. on p. 6).

[93] A. Corso-Radu et al. " Data Quality Monitoring Framework for the ATLAS experiment: Performance achieved with colliding beams at the LHC." *J. Phys.: Conf. Ser. 331 (2011) 022027 (cit. on p. 6).*

[94] "Approved plots Tile Detector Status." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileDetectorStatus.

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

[96] F. Bishara, R. Contino and J. Rojo. "Higgs pair production in vector-boson fusion at the LHC and beyond." *Eur. Phys. J. C77 (2017) 481.* arXiv: 1611.03860 [hep-ph].

[97] ATLAS Collaboration. "Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector." https://cds.cern.ch/record/2660207

APPENDIX A

RESPONSE OF TILE DEMONSTRATOR MODULE TO MUON BEAM

A.1 Introduction

The TileCal is a hadronic calorimeter in the central region of the ATLAS detector in the Large Hadron Collider (LHC) experiment. The existing TileCal inside ATLAS is designed to operate at 13 TeV center-of-mass ($\sqrt{s}$) energy and integrated luminosity 10^{34} of LHC in RUN 2. However, a series of upgrades are planned for TileCal to make it operational in 5-7 times larger than the current luminosity and high $\sqrt{s}$ after the long shutdown planned in 2019.

This result shows the response of the Tile Demonstrator module which consists of the Tile Phase 2 upgrade electronics to the muon beam entering at +/-90 degree.

A.2 Tile Phase 2 upgrade

The electronics Figure A.1 of the existing TileCal are aging and so to operate the TileCal smoothly in the high luminosity (HL-LHC) environment new electronics system with low maintenance cost is needed. The current analog trigger system will be replaced by a full digital Level-1 trigger. "Demonstrator" is a functional prototype of the new system associated with a TileCal module, this will be integrated into ATLAS for evaluation of the LHC Phase 2 new electronics architecture [84].

New electronics modification is also done in the read out system. Each drawer is split into two mini drawers. This is designed such a way that handling will be easier and also to get easy access to the mini-drawers during maintenance.

The new read out electronic is the Demonstrator is assembled with:

1. 12 Front End (FE) board: 1 out of 3 different options. It provides the conditions for PMT signals, amplifies and then digitizes the signal. It is also equipped with calibration

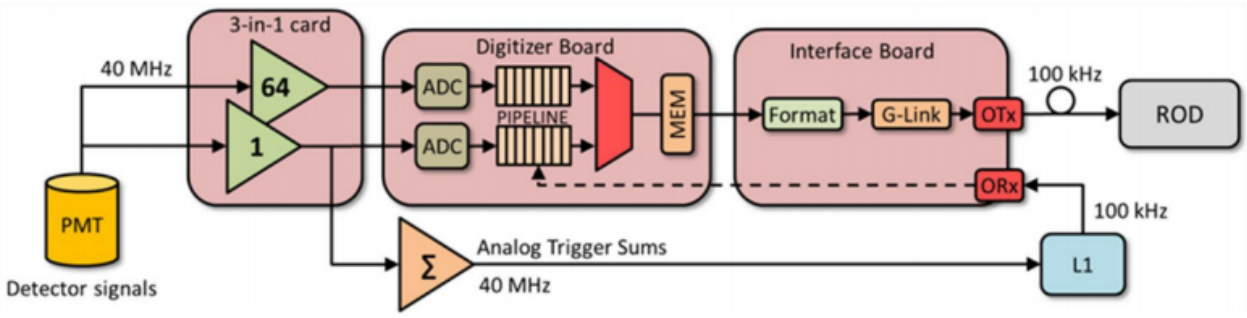


Figure A.1: TileCal signal flow in the current system (Legacy SD) [84].

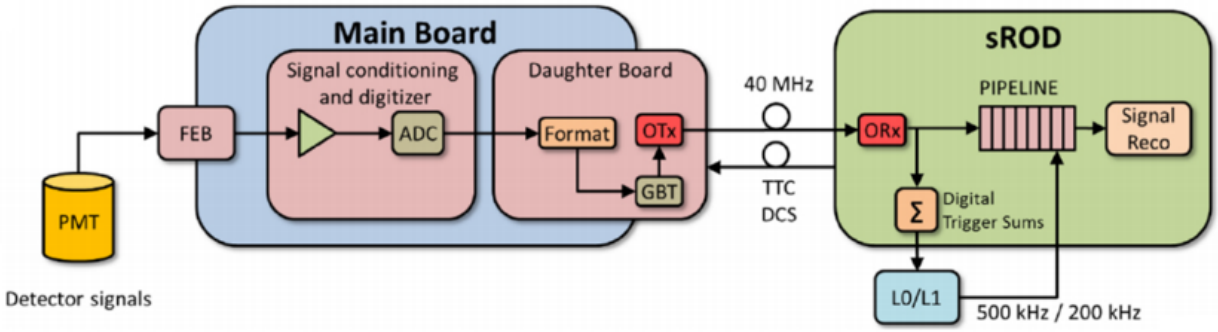


Figure A.2: Signal flow in Phase 2 upgrade electronics [84].

functionalities.

2. 1 Main Board (MB): it provides digitization of signals depending on FE system and control.

3. 1 Daughter Board (DB): this is used for data processing and interface with the back-end electronics via optical links at up to 40 GHz rate with redundancy.

4. 1 High Voltage (HV) regulation board : 1 out of 2 options viz. remote HV and HVOpto.

5. 1 adder base board plus 3 adder base cards for the simulation of the analog trigger signal to send to L1 and L0 trigger.

6. New Low Voltage Power Supply (LVPS) which provides redundancy and Point of Load Regulators.

7. Another important change on the drawer architecture is the replacement of a single interface board by four Daughter Boards, improving the system robustness and decoupling the drawer electronics into independent units.

Signal flow in Phase 2 system shown in Figure A.2. These signals are fully digitized and sent off-detector via high-bandwidth optical links to the Tile Preprocessor (TilePPr). The TilePPr sends trigger tower signals to trigger, receives and transmits slow controls to the drawer, and establishes the system timing to synchronize the read out stream with the Data Acquisition System (DAQ).

A.3 Testbeam setup

Figure A.3 gives an idea how different component were setup in testbeam session. Figure A.4 shows the setup of the Tile modules. The hybrid prototype - the Demonstrator has been tested together with one long barrel module (LBC module - M0 module) of the TileCal

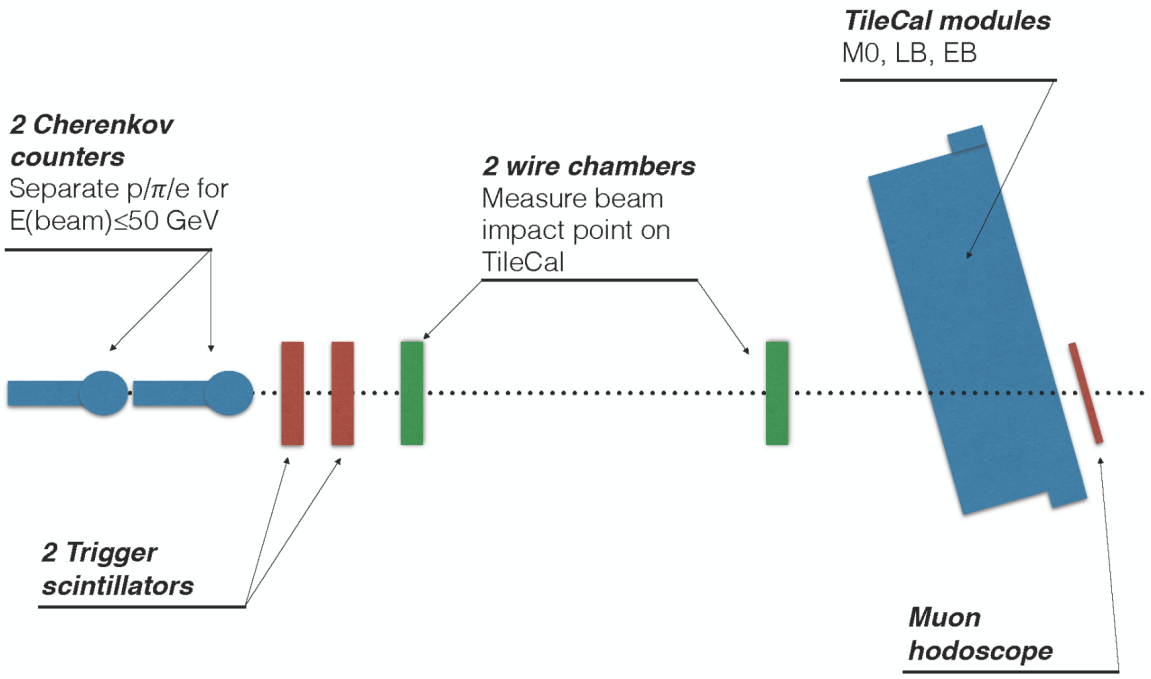


Figure A.3: June 2016 testbeam setup [85].

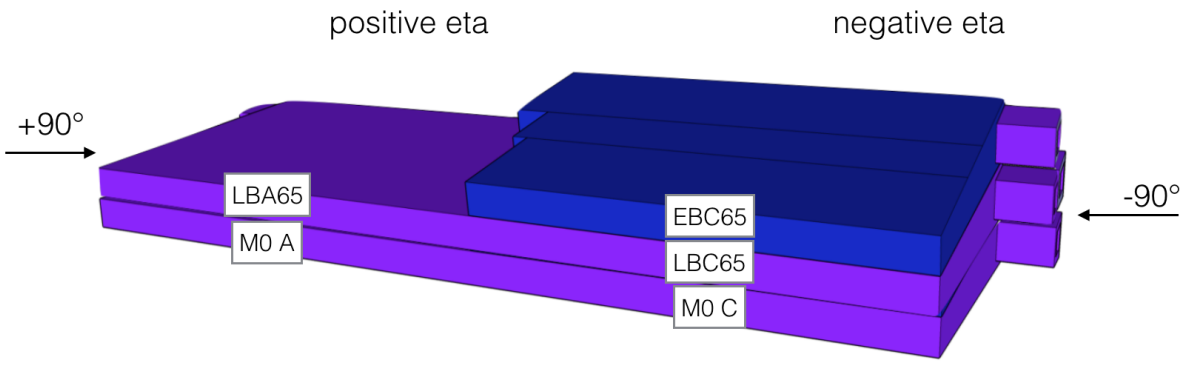


Figure A.4: Tile Module setup during 2016 testbeam period [85].

Table A.1: Tile modules and their electronics in June 2016 testbeam setup.

Module	Electronics	Module name in ntuple
M0A	Fatalic, QIE	A01
M0C	Legacy SD	C01
LBA65	Legacy SD	A02
LBC65	Demonstrator	C02
EBC65	Legacy SD	E03

with beams of different particles (π, e and μ) from the SPS accelerator of CERN at different incident angles (20°, 90°, -90°) and energies (50,150,165,180 GeV).

Half of LBC65 has been equipped with so-called Hybrid Demonstrator which provides all the upgrade functionalities compared to previous electronics (Legacy SD). This analysis is based on the Demonstrator response which has upgraded 3-in-1 FE baseline option. The 3-in-1 cards are built with discrete commercial off-the-shelf components. The upgrade improve the digitization of the signal resolution using 12 bits ADC instead of 10 bit. This new 3-in-1 card also reduce the electronics noise. This card is compatible with the analog trigger and can be installed into TileCal before the HL-LHC upgrade.

M0 A is instrumented with two front-end electronics known as QIE and FATALIC, these are under inspection. LBA65, EBC65, M0C are equipped with Legacy SD electronics. Different components' name in the data and the type of electronics inside it are listed in Table A.1.

A.4 Reconstruction Method

Data in June 2016 testbeam is reconstructed with two different method, Optimal Filtering (OF) and Fit (FIT) method. The first one we use for *pp* collision data taking. This section will give you a brief overview of both methods.

A.4.1 Fit Reconstruction Method

The pulse shape of the front-end electronics is proportional to the energy deposited by the particle in the Tile calorimeter cell. This pulse is digitized by 7 samples which are spaced by 25 ns as shown in Figure A.5.

From the energy deposition in the scintillator tile we get the 7 samples $S_{\rm i}$, where i = 1 to 7. The value of the sample can be expressed by,

$$S_{\rm i} = Ag(t_{\rm i} - t) + ped \tag{A.1}$$

Here A is the amplitude of the signal, $t_{\rm i}$ is the time where the sample i was obtained, the ped is the pedestal value and t is the time of the peak of the pulse.

In Fit method [86], we use a known pulse shape which reduces the background due to the electronic noise. The timing of the energy deposition can also be determined by this method and it depends on the fit parameter explained in Eq.(3), amplitude, time and pedestal. To get the best fit result the following expression needs to be minimized:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{S_{\rm i} - [Ag(t_{\rm i}) - Atg^{'}(t_{\rm i}) + ped]}{\sigma_{\rm i}} \right)^2 \tag{A.2}$$

$\sigma_{\rm i}$ is the error (standard deviation) of the sample i, This is around 1.5 ADC in high gain and 0.6 ADC for low gain channels.

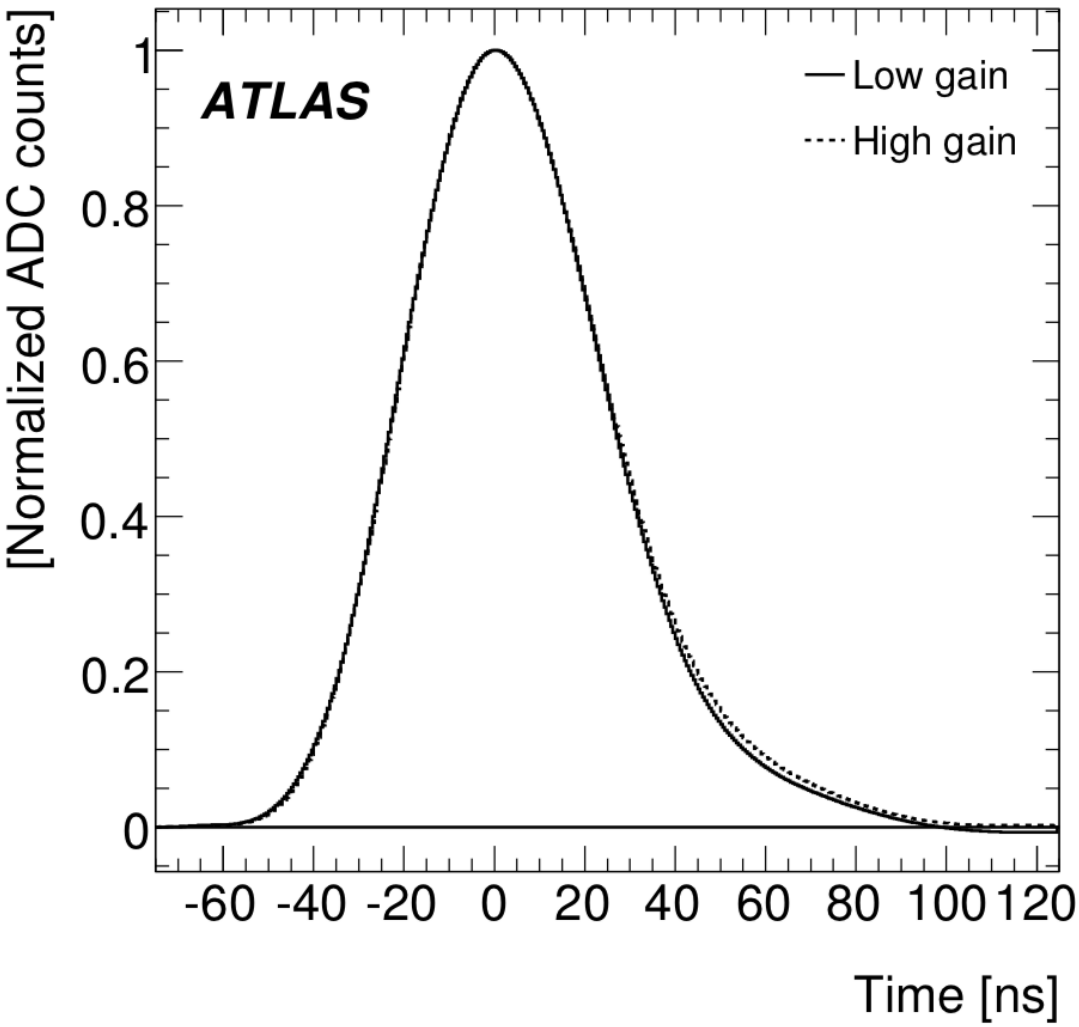


Figure A.5: Normalized pulse shape of Physics data [86].

A.4.2 Optimal Filtering Reconstruction Method

The other reconstruction method that is being used for the real collision data is the Optimal Filtering method (OF) [87]. Instead of fitting a known pulse shape we calculate amplitude of a given pulse using the linear combination of the sample S_{i}.

There are two versions of the OF method, first one is OF1 which gives two output amplitude A and time t. OF2 gives three outputs A, t and pedestal ped. The expression for OF1 are:

$$A = \sum_{i=1}^{n} a_{\mathrm{i}}(S_{\mathrm{i}} - ped) \tag{A.3}$$

$$At = \sum_{i=1}^{n} b_{\mathrm{i}}(S_{\mathrm{i}} - ped) \tag{A.4}$$

The expression of the OF2 are the following:

$$A = \sum_{i=1}^{n} a_{\mathrm{i}}(S_{\mathrm{i}} - ped) \tag{A.5}$$

$$At = \sum_{i=1}^{n} b_{\mathrm{i}}(S_{\mathrm{i}} - ped) \tag{A.6}$$

$$ped = \sum_{i=1}^{n} c_{\mathrm{i}} S_{\mathrm{i}} \tag{A.7}$$

In the OF1 the pedestal is calculated from the average of the first (S_1) and last sample (S_{n}). OF2 which is used for this testbeam analysis gives better pedestal value. In both type of OF the quality of the reconstruction is measured by the quality factor Q:

$$Q = 50 \sum_{i=1}^{n} |\frac{S_{\rm i} - (Ag_{\rm i} + ped)}{S_{\rm i}}| \tag{A.8}$$

In OF the timing should be very close to zero, the quality factor decreases when the timing is far from zero.

For the testbeam data, the OF filtering is used with iteration and without iteration. When the first and maximum sample value ($S_{\rm max} - S_1$) is below certain threshold then it does not iterate over different OF weight and gives timing to zero otherwise, it iterates to reconstruct the amplitude and timing of the signal.

For the iterative OF, the phase of the sample is the time of the first sample,

$$\tau_0 = time(S_{\rm max}) \tag{A.9}$$

This is done to start the iteration, then this phase is used to calculated the phase for the next iteration, like following:

$$\tau_{\rm i} = \tau_{\rm i-1} - int(t_{\rm i}) \tag{A.10}$$

The iteration stops if the t is small enough or it exceeds maximum number of iteration. The final value of the signal is set to $t = t_{\rm i} - \tau_{\rm i}$. If the time $|t| > [(n-1)/2] * 25$ ns, i.e 75 ns for 7 samples the events are then forced to be at $\pm$75 ns.

A.5 Response of individual PMTs in BC Layer to -90° muon beam

The dominant energy loss process of muon inside the detector is ionization and the energy loss is essentially proportional to the muon track path length. The high-energy muons avail-

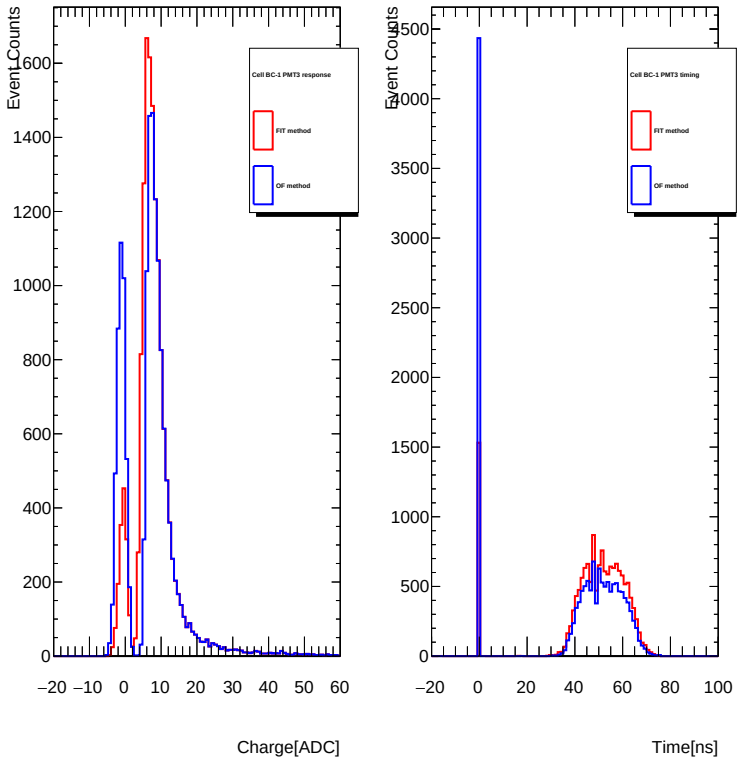


Figure A.6: The charge in ADC and time distribution in PMT3 with different reconstruction methods, FIT (red) and OF (blue).

able at the beam-line traverse the entire TileCal modules for any angle of incidence. Muon response has been studied determining the ratio between the energy deposited in a calorimeter cell (dE) and the track path-length in the cell (dX). The timing and energy response of PMT3 in BC Layer to -90° muon beam and reconstructed with two different methods (OF and FIT) is shown in Figure A.6. The peak near zero (in both plots) corresponds to events reconstructed with non-iterative FIT and OF filtering method so the reconstructed timing for those events are zero. The noise threshold for FIT method is set at 3.2 ADC counts and OF is set at 5 ADC counts. The peak near zero is smaller for FIT method in compare to OF, that is due to the different noise threshold used in these two reconstruction processes. The final $\frac{dE}{dZ}$ calculation is done in OF and the noise threshold is set at 3ADC counts.

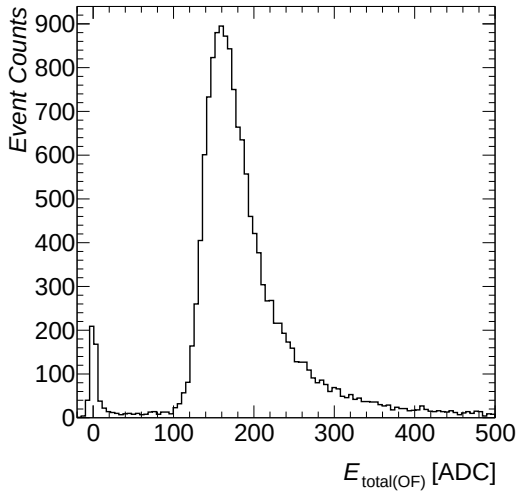


Figure A.7: Total energy of 4th Tile row using OF method.

In Figure A.7 shows the total energy distribution of PMTs in the 4th TileCal row. The cut $40 < E_{\text{tot}} < 400$ ADC counts applied to select real muon events and reject events from fake trigger, electronic noise. The individual PMT response after $40 < E_{\text{tot}} < 400$ cut is shown Figure A.8. The OF filtering reconstruction method is used and the noise threshold is set at 3 ADC counts, with increasing noise threshold the events near zero increases, this events are real signal events but outside the time window and the estimated energy is not correctly reconstructed because the reconstructed time is very far from the average of the time window.

Each PMT responses in 4th Tile row was checked separately to measure the energy loss (dE) in the BC cells. PMT46 response has been found different with respect to other PMTs in the same row, so PMT46 response is replaced with PMT45 times the ratio of the two PMTs in the neighbourhood cell of the same size. Figure A.9 shows the energy Vs time distribution of good PMT45 and bad PMT46. PMT46 most of the events are outside time window of [20,75] ns which is beyond the maximum possible time, so the reconstructed energy is underestimated. PMT46 is excluded from the analysis. However, PMT45 also has few events which are outside the given time window [20,75] ns can not be rejected since these

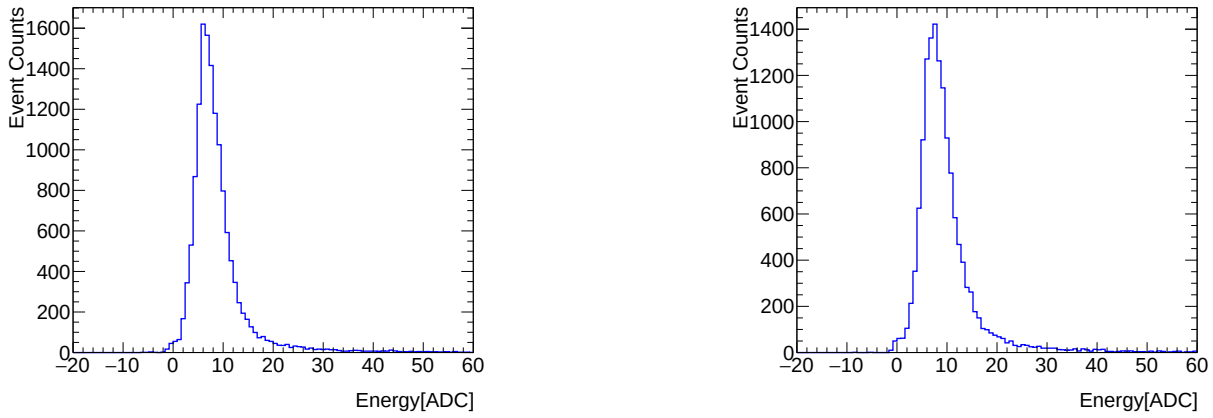


Figure A.8: Response of PMTs 3 (left) and PMT4 (right) in BC1 cell after the E_{tot} cut.

are real signal events at low energy. The reconstruction method was improved later to have these events inside the window and also the time window should be same for all the good PMTs.

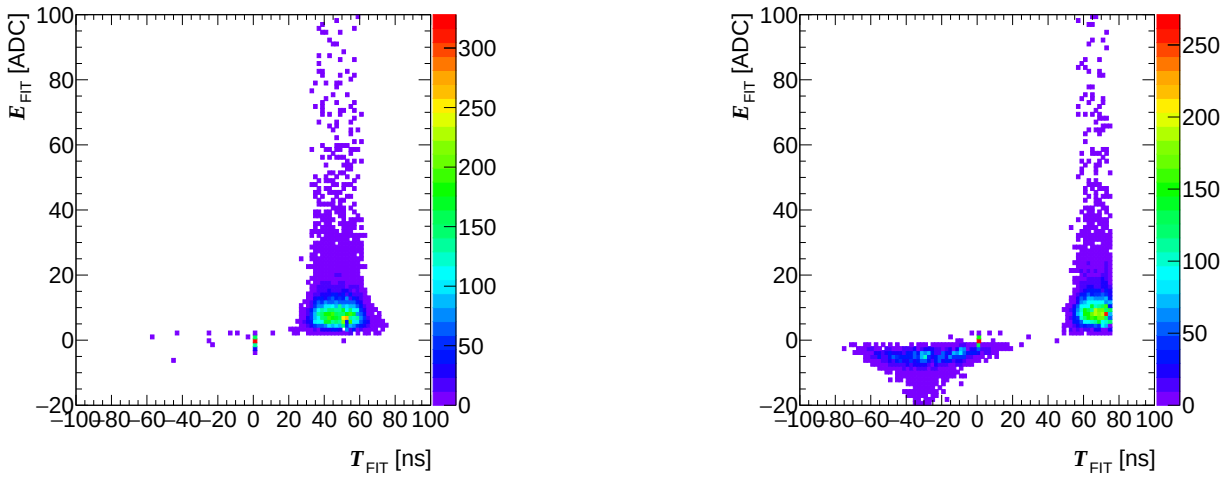


Figure A.9: Energy Vs time response of PMT45 and PMT46 in the same cell.

The truncated (95%) mean of the energy distribution shown in Figure A.8 is used to calculate the energy loss dE per unit length (dZ, length of BC cells in 4th Tile Row) of the BC cells, $\frac{dE}{dZ}$, in the 4th Tile row Figure A.10. The increase in noise threshold did not make significant change in the $\frac{dE}{dZ}$ values of each cell.

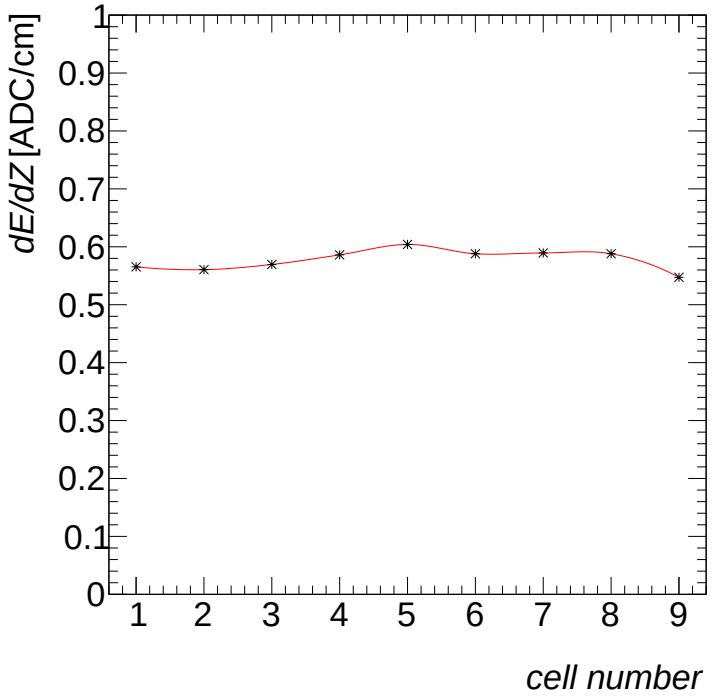


Figure A.10: $\frac{dE}{dZ}$ using OF method for 4th Tile row.

A.6 Summary

In this chapter we have discussed the PMT response of the Demonstrator module using the 2016 testbeam data. Demonstrator is equipped with new phase 2 upgrade electronics which will be used for HL-LHC. The Demonstrator response is studied with 150 GeV muon beam hitting Tile rows from - 90°. Two different reconstruction method is used, OF and FIT. Comparison between different noise threshold was also shown in Figure A.7. A unbiased cut on total energy of 4th Tile row is applied to remove events coming from fake trigger. To make the $\frac{dE}{dZ}$ distribution flat the truncated mean (95%) on the energy distribution is considered.

APPENDIX B

TILE DATA QUALITY MONITORING

This appendix gives brief overview of the ATLAS Data Quality Monitoring (DQM) followed by my contribution in Tile DQM. The next section describes the ATLAS data taking process and monitoring tools [88]. The last section is my contribution in Tile DQM as a part of my service task in the ATLAS.

B.1 ATLAS Data Quality

An uninterrupted period when beam are circulating in the LHC machine is known as *fill*. Typically ATLAS records data continuously during a single LHC *fill*. Data-taking for physics begins as soon as possible after the LHC declares "stable beams", then the high-voltage is ramped up in the ATLAS tracking detectors and muon system. Once the preamplifiers of the pixel system are turned in ATLAS "ready for physics" is declared. During each stable beam ATLAS data is recorded against a 6-digit run number. Each run is further divided into so-called luminosity blocks (LBs). A luminosity block is a period of time (60 sec) during which instantaneous luminosity, detector and trigger configuration and data quality conditions are considered constant. The LB serves as a standard unit of granularity for data quality book-keeping and also helps to reject inferior-quality of data with finer granularity.

The detector conditions such as trigger, configuration and calibration constants are stored in the ATLAS Oracle database hosting a COOL technology schema [89]. This database stores the detector condition in specific Interval of Validity (IOV). This IOV could be an interval in time (in ns), run number and LBs, between which the stored conditions are valid for data.

Detector Control System (DCS) [90] shows the operational status of detector hardware components like temperatures, high voltage values and operational capacity. The low-level detector information is provided by GNAM [91, 92]. GNAM is a light-weight tool that monitors detector status at various stages of the data-flow by producing monitoring histograms.

Trigger monitoring tool monitors events which are passed through L1 trigger and which fails to meet the L1 trigger requirement to ensures that interesting events are not unintentionally discarded. In addition, the high-level trigger system generates a set of histograms that monitor aspects of its operation. For monitoring of higher-level information, including the quality of reconstructed physics objects, full event reconstruction is run on a small subset of trigger-accepted events and histograms produced with code similar to that used for monitoring during the offline processing.

The algorithms that run in the Data Quality Monitoring Framework (DQMF) [93] must be lightweight such that they comply with data processing time restrictions and must be written and tested offline, independently from online data-taking. This means that the DQMF is compatible with the online (during real-time data-taking at ATLAS) and offline (during offline data reconstruction) environments, and the use in the two environments differs only in the methods used to deliver histograms to the framework.

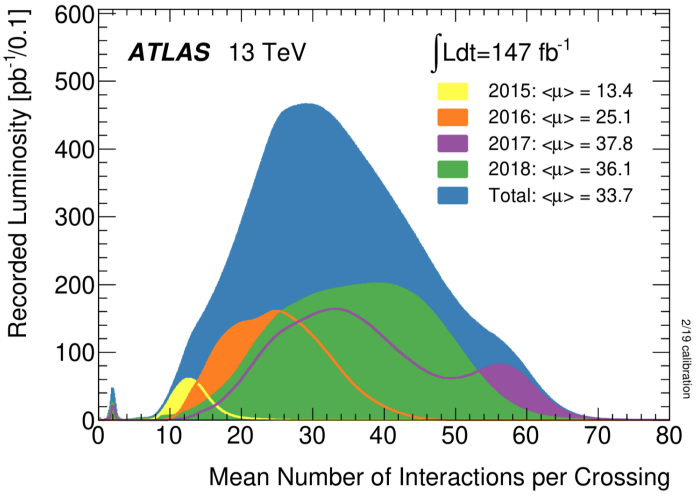


Figure B.1: Luminosity-weighted distribution of the mean number of interactions per crossing,μ, for the full Run 2 pp collision data set at 13 TeV [88].

The online and the offline data monitoring system is shown in B.2. The data recorded by ATLAS detector goes thorough the elaborate online monitoring at P1 (ATLAS control room)

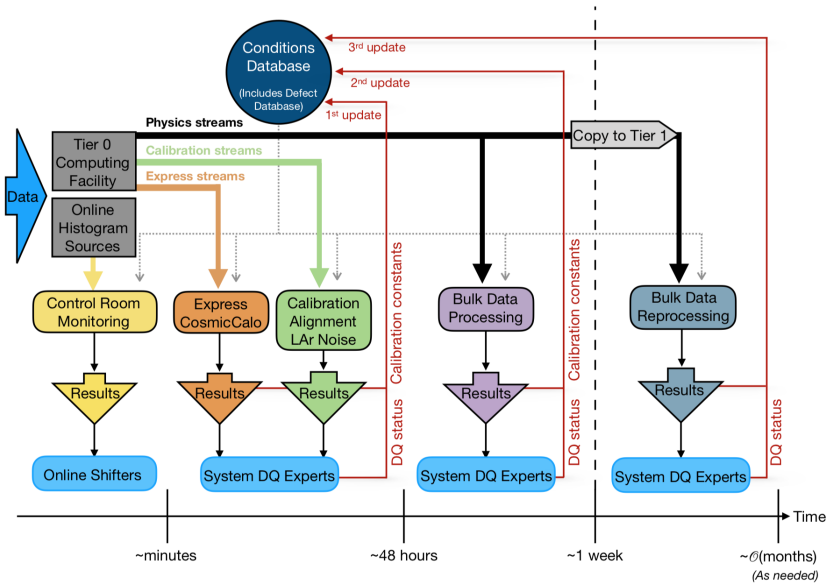


Figure B.2: Schematic illustrating the nominal Run 2 operations work flow for the data quality assessment of ATLAS data [88].

by all the sub-systems. Once a physics run is over the data is processed and the"express" stream processing starts for the first offline monitoring. In this stream 10% of the total data are processed. All the detectors are given 48 hours to check the"express" stream and update the database if needed. This 48 hours time span is called "calibration Loop" (CL).

B.2 Contribution

I worked in the ATLAS TileCal as Data Quality (DQ) coordinator for two years starting from 2016 April to 2018 June. Before that I started taking over my responsibility of a DQ coordinator from September 2015 from my senior Blake Oliver Burghgrave.

As a DQ coordinator my main responsibility was to monitor the offline DQ plots and assess quality of data recorded by the TileCal. I was mainly responsible for offline checks and reprocessing of data. During the CL DQ Validator does preliminary checks of the

"express_express" stream, produce the DQ report for DQ Leader. DQ Leader checks the report and make the necessary change in the Tile database. These changes are mainly regarding calibration constants and masking any channels which appear cold or hot with respect to the neighborhood channels. Once the CL is over the database is locked, no more changes are allowed, after that the full processing of the recorded data starts. It takes few days to get the full processed data. There are several physics streams produced with different DQ plots depending on sub-system requirements. From Tile side the important stream for DQ monitoring is "physics_main". As a DQ coordinator I was responsible to do the final checks of the "physics_main" stream and mark the luminosity blocks containing bad data in the ATLAS DQ data base. Keeping record of any minor problem that could be fixed during the reprocessing of data at the end of every year was another important duty of DQ coordinator. After all the checks are done, bad data are removed from Good Run List (GRL) by setting intolerable defects for the affected LBs and then physics run is ready to use for the analysis.

During my service task as DQ coordinator I also supervised DQ Validator and Leader. I prepared report of the physics data quality on behalf of TileCal and presented on the ATLAS DQ meeting. At the end of the year ATLAS DQ team reprocesses all the data of that year and previous years with new software releases. During that time I prepared reprocessing conditions for Tile masked channels and after reprocessing I did the DQ assessment again.

In between two physics runs, each detector records calibration runs to calculate and update calibration constants to ensure good quality physics data and identify detector issues. The DQ of calibration runs are assessed by the DQ validator every week. I developed a python script that will show the TileCal channels and modules status in calibration runs and helped implementing that in web interface to make it user friendly during my ATLAS authorship project .

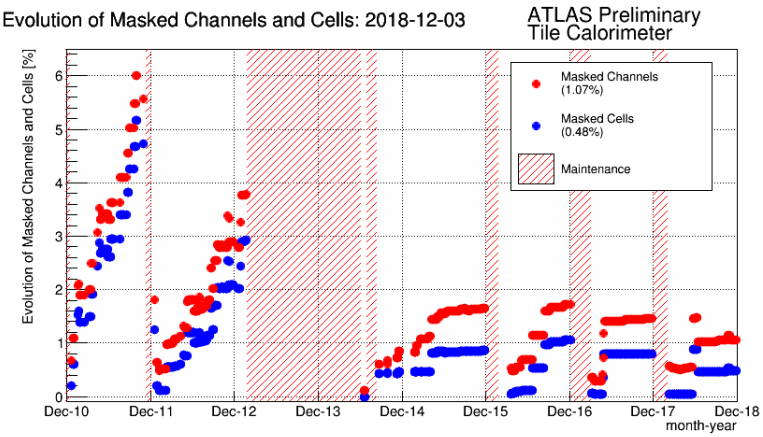


Figure B.3: Evolution of channel and cell Masking in TileCalorimeter [94].

During my shift, in 2016 and 2017 ATLAS recorded 76 fb^{-1} data and Tile efficiency during this period 99.8%. The plot on Figure B.3 shows the evolution of Tile masked channels and cells starting from December 2010 (RunI) upto December 2018 (end of RunII). Performance of TileCal improves significantly with time.

A paper based on the full Run-II DQ is in preparation, will be submitted to JINST. https://cds.cern.ch/record/2674873/

APPENDIX C

GENERATION AND VALIDATION OF HIGSS PAIR MONTE CARLO SAMPLES IN VECTOR BOSON FUSION PRODUCTION

C.1 Introduction

Vector boson fusion (VBF) has the second highest cross-section in Higgs pair production after the gluon-gluon fusion (ggF) at 13 TeV. The Figure C.1 shows the leading order (LO) VBF production of Higgs pair, the left diagram is sensitive to c_{2V}, middle diagram to ${c_V}^2$ and right diagram to $c_V k_\lambda$ couplings. The Standard Model (SM) values of c_{2V} and c_V are both 1.

C.2 Non-resonant Production of VBF Higgs Pair

The Monte Carlo (MC) signal sample for VBF $HH \rightarrow b\bar{b}\tau^+\tau^-$ are generated with MadGraph_aMC@NLO2.6.0 The NNPDF 2.3 LO PDF set [95] is used in the matrix element, which is interfaced to Herwig 7.0.4 using the H7-UE-MMHT tune for underlying events and the H7-MMHT2014LO tune for parton shower and hadronisation. The set of coupling constants considered in the MC generation are listed in the Table C.1. The plots from semi-leptonic and fully hadronic final states will be discussed here. I contributed in semi-leptonic final state of two τs mainly.

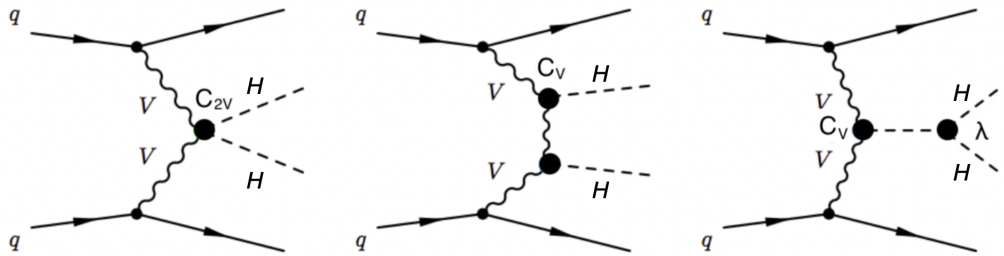


Figure C.1: Leading order diagram of VBF hh production

Table C.1: Coupling constants considered in MC generation of VBF hh production.

k_λ	c_{2V}	c_V
1	0	0.5
1	1	0.5
0	0	1
1	0	1
1	0	1
1	0.5	1
1	1	1
1	1.5	1
1	2	1
1	4	1
0	1	1
2	1	1
10	1	1
1	1	1.5

C.2.1 MC Generation of $bb\tau\tau$ Final State

MadGraph processing card is used to generate VBF HH sample at 13 TeV center of mass energy and Herwig7 is used for HH decay in $bb\tau\tau$ (only $H \to bb$ and $H \to \tau\tau$ are allowed) and then filtering final state where one τ decays leptonically and other τ hadronically. Also another filter was added to remove particles having low transverse momentum. $p_{\mathrm{T}} > 13$ GeV and $p_{\mathrm{T}} > 15$ GeV are applied on the lepton (e/μ) and visible decay product of τ respectively. The decay particles of final state τs are required to be within $|\eta| < 3.0$.

Generated processes are $pp \to HHqq$, which are dominantly from the VBF HH production. However they also contain contributions from Vhh hadronic (where V decays hadronically) and Higgsstrahlung processes. These additional contributions can be seen at low values of the distribution of the η separation of the VBF jets on Figure C.2. These VBF jets are selected by looping over jet container in xAOD, and selecting events which have two

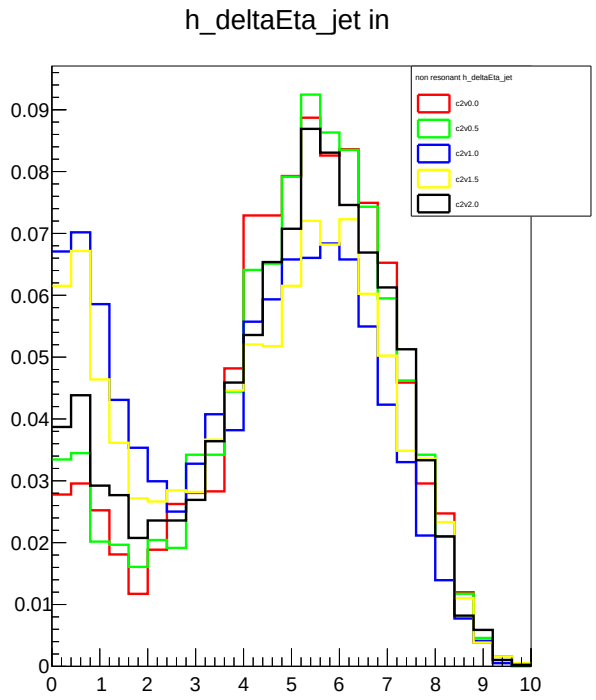


Figure C.2: $\Delta\eta$ of two VBF jets in semi-leptonic final state

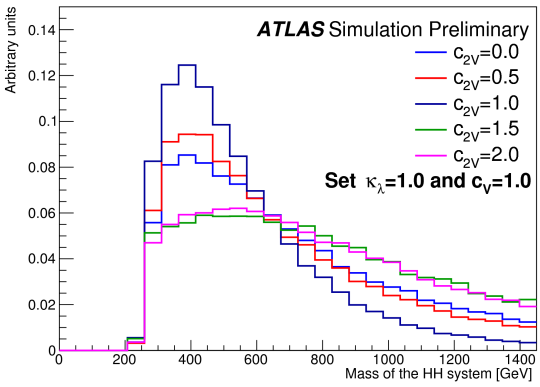


Figure C.3: Invariant mass distribution of two VBF jets [97].

jets not overlapped with any of the other particles in that event. VBF jets are supposed to be very far from the decay products of two SM Higgs, H. No other cut is applied on the kinematics variables of the final state objects for the validation of MC sample. A threshold on m_{jj} can suppress Vhh and Higgsstrahlung processes.

In fully hadronic final state only a different filter is applied in Herwig7 to select two hadronic τs in the final state. Two p_{T} filters are applied, $p_{\mathrm{T}} > 13$ GeV and $p_{\mathrm{T}} > 15$ GeV are applied on the lepton (e/μ) and visible decay product of τ respectively. Then $p_{\mathrm{T}} > 35$ GeV filter was applied on the leading τ p_{T}. The decay particles of final state τs are required to be within $|\eta| < 3.0$. During the validation (not in generation) of signal samples, further cuts are applied:

- At least 2 hadronic taus with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ using visible decay products,

- Leading hadronic tau with $p_{\mathrm{T}} > 40$ GeV and $|\eta| < 2.5$ using visible decay products,
- At least 2 b-jets ($\Delta R < 0.4$ with respect to b-quarks), using the anti-kt algorithm with a distance R=0.4,
- $\Delta R > 5.0$ for the 2 light jets that have the highest m_{jj} and $m_{jj} > 300$GeV as well as with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$.

C.3 Resonant Production of VBF Higgs pair

The MC sample for resonant Higgs pair in VBF process is generated using Powheg-Box-V2 and Pythia 8.235. The 2HDM heavy scalar (X) is generated using Powheg and then the decay of X decay to two SM-like H and $bb\tau\tau$ final state is done with Pythia 8.235 with "dipoleRecoil = on" . The same selection like non-resonant VBF HH production is applied to select semi-leptonic final state and to remove low p_{T} objects.

Figure C.4 shows the invariant mass distribution of two SM H produced from X decay. In total 15 different set of mass points of X are generated. The long tail on the left side of each peak appears due to the final state b-quark (after gluon radiation) from H decay. This peak can be removed by considering immediate b-quark from the H decay. The same selection mentioned in Section C.2 is applied to select VBF jets, no other cuts are applied on the kinematics variables of the final state objects.

C.4 VBF jet modelling with PYTHIA8

Pythia 8.235 is connected to the VBF matrix element generator for parton-shower studies related to VBF jet modeling. The option "dipoleRecoil" is turned on to have better VBF jet

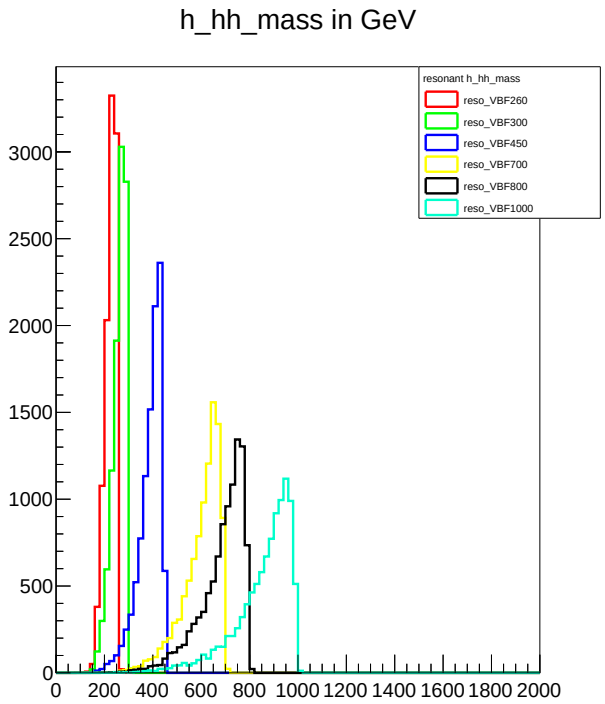


Figure C.4: Invariant mass distribution of two VBF jets.

modeling as recommended by the LHCHXSWG [16]. The way Pythia8 deals with the recoil of the radiation in VBF processes leads to the tendency of generating too much hard and central radiation. The option "dipoleRecoil = on" can mitigate this effect. Figure C.5 shows relevant distributions with this option on and off, also compared to Herwig 7.0.4 which is the nominal parton-shower generator. With "dipoleRecoil = on", Pythia 8.235 agrees better with Herwig 7.0.4.

C.5 Conclusion

In this chapter I discussed the generation of di-Higgs sample in VBF production mode and $bb\tau\tau$ final state both in SM and BSM scenarios. Then truth level plots are prepared to

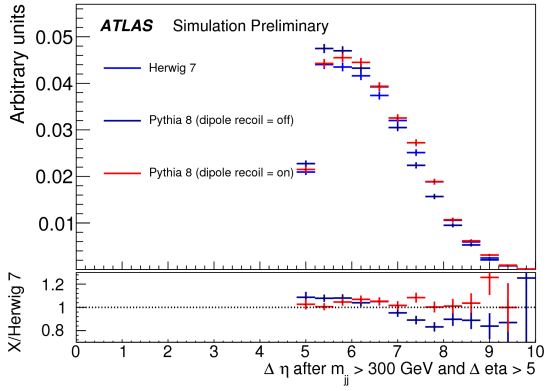


Figure C.5: $\Delta\eta$ of two VBF jets in semi-leptonic final state with two different parton shower [97].

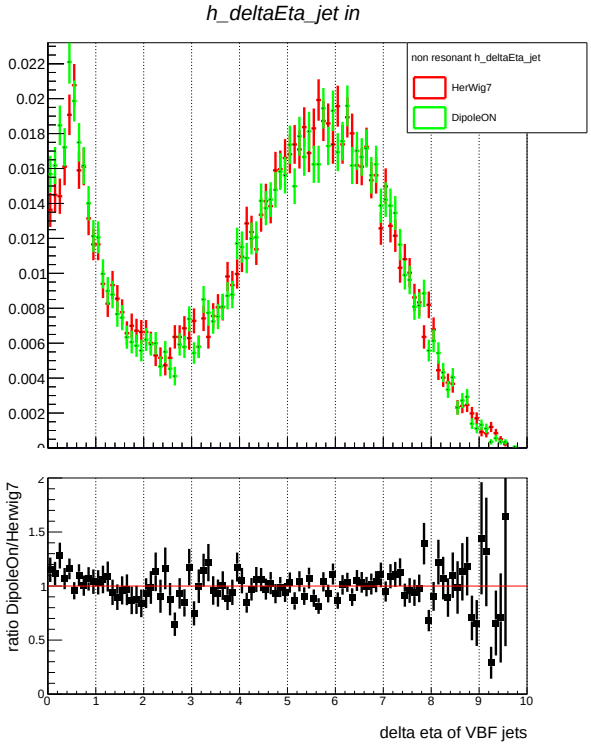


Figure C.6: Invariant mass distribution of two VBF jets with two different parton shower.

validate the sample in semi-leptonic channel and fully hadronic channel. Finally VBF jet modeling was compared in Pythia8 with DipoleON and Herwig7.