Observation of Higgs Boson Production Through Vector Boson Fusion in the WW^* Channel at the ATLAS Detector

by

Sourav Sen

Department of Physics
Duke University

Date: ____________________________
Approved:

____
Ashutosh V. Kotwal, Advisor

____
Ayana T. Arce

____
Phillip S. Barbeau

____
Steffen A. Bass

____
Gleb Finkelstein

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Physics
in the Graduate School of
Duke University

2020

ABSTRACT

Observation of Higgs Boson Production Through Vector Boson Fusion in the WW^* Channel at the ATLAS Detector

by

Sourav Sen

Department of Physics
Duke University

Date: ______________________________
Approved:

____
Ashutosh V. Kotwal, Advisor

____
Ayana T. Arce

____
Phillip S. Barbeau

____
Steffen A. Bass

____
Gleb Finkelstein

An abstract of a dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Physics
in the Graduate School of
Duke University

2020

Abstract

This thesis presents an observation of vector-boson-fusion (VBF) production of Higgs boson in the $H \to WW^* \to \ell\nu\ell\nu$ decay channel. For this analysis, $\sqrt{s} = 13$ TeV proton-proton collision data collected by the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of 139 fb^{-1}, is used. Events with one electron (positron) and one anti-muon (muon), and missing transverse momentum, along with at least two forward jets in the final state, are selected for the study. A major part of the analysis concerns the estimation of backgrounds with misidentified leptons, which generally originate from the production of a W boson and an associated object mistakenly identified as a lepton and are estimated with data-driven techniques. Signal sensitivity has been significantly improved by employing a deep neural network (DNN) in separating signal from background events. The significance of the data assuming the background-only hypothesis to be true has been observed to be 7.0 σ (6.2 σ expected), constituting the first observation of the VBF Higgs boson production in the $H \to WW^* \to \ell\nu\ell\nu$ decay channel. The product of the total VBF cross-section times the $H \to WW^*$ branching fraction is measured to be $0.85 \pm 0.10(\text{stat.})^{+0.17}_{-0.13}(\text{syst.})$ pb, compatible with the Standard Model prediction of 0.81 ± 0.02 pb.

For my parents and my sister

Contents

List of Tables

List of Figures

Acknowledgements

I am extremely grateful to my advisor, Prof. Ashutosh Kotwal, for his valuable guidance and for providing me the opportunities to contribute to many interesting projects across different experiments. I could learn about various aspects of experimental particle physics through them. I am very grateful to all my committee members - Professor Behringer, who passed away in 2018, and Professors Arce, Barbeau, Bass, Finkelstein for their guidance and critiques. I want to thank Prof. Barbeau for all his helpful suggestions and for telling me about the bullet journals, which immensely helped me during my research.

I thank Sergei Chekanov for all the help during my work in the future circular collider project. I am grateful to Louis-Guillaume Gagnon for the initial help and discussions about setting up the CTIDE neural networks. I extend my thanks to the CTIDE group's conveners - Bill Murray, Ben Nachmann, Pierfrancesco, and others for being very helpful and supportive throughout all my projects in the group.

I want to thank Koos, Claudia, David, Konstantin, Benny, and all the HWW analysis team members in ATLAS for fruitful discussions, all the help with the analysis framework, and useful suggestions during the analysis. I would also like to thank Nedaa, Loic, Kate, Alan, and others in the BSM 4 tops analysis team for their help throughout my research.

It was a pleasure sharing office no. 288 at Duke with my fellow graduate students in ATLAS: Doug, Matt, and Mike, who were great resources during the research in ATLAS. We had many interesting discussions on topics, even outside physics. I would also like to thank all my friends at Duke for all the fun times. I learned a lot through helpful discussions with my friends Elham and Harish about ATLAS, particle physics, and other fun topics. I am lucky to have friends like Rahul, Abhranil, and

Swapil, who lifted my spirits even during rough times in my Ph.D.

My parents and sister have always believed in me and have always been there for me, no matter what. I owe an immense debt to my family, as without their constant support and belief in me, I could not have finished this Ph.D.

Chapter 1

Introduction

The Standard Model of particle physics is a theoretical framework developed over the previous century, which provides the current best understanding of the fundamental interactions (except gravitation interaction) and properties of the elementary particles in nature. All the predictions of the SM have been experimentally verified to a very high precision, including the discovery of the particle consistent with the SM Higgs boson in 2012 at the Large Hadron Collider which completed the theory. Despite its remarkable success, the SM is considered to be far from being a complete theory to explain nature. It is not yet able to incorporate the gravitational interaction, as mentioned earlier, as well as provide explanations for the underlying nature of the electroweak symmetry breaking (EWSB), along with several other observations such as the presence of Dark Matter in the Universe and the matter-antimatter asymmetry. To extend the SM, often referred to as Beyond the Standard Model (BSM) theories, two possible approaches can be taken, viz. 1) precise measurement of the SM parameters where any significant deviation from the SM can provide a good direction to formulate BSM theories, and 2) directly testing the parameters for available BSM theories.

Both these routes have been explored through various projects in this thesis. Chapter 7 deals with reducing the systematic uncertainties due to Parton Density Functions (PDFs) in precise measurement of W mass with $\sqrt{s} = 1.96$ TeV proton-antiproton collision data at the Collider Detector at Fermilab (CDF)-II of Tevatron. In Chapter 9, the measurement of vector boson fusion production mode of SM Higgs boson decaying to two W bosons at the ATLAS detector, using $\sqrt{s} = 13$ TeV proton-proton collision data collected between 2015 and 2018, is presented. To observe BSM resonances in the future colliders, Chapter 6 deals with identification techniques of hadronically decaying τ leptons against the multi-jet background at high granularity Future Circular hadron Collider (FCC-hh), in preparation for the proposed $\sqrt{s} = 100$ TeV Hadron Collider experiment. To probe for deeper principles behind the qualitative form of the Higgs potential, Chapter 8 presents the analysis strategies developed to the search for a composite top quark resonance at the ATLAS detector, using $\sqrt{s} = 13$ TeV proton-proton collision data collected between 2015 and 2018. A brief overview of the theoretical framework behind the projects in this thesis is provided in Chapter 2. In Chapter 3, the various subsystems of the ATLAS detector along with a brief introduction to the other experiments at the LHC are presented. The reconstruction of proton-proton collision events in the ATLAS detector is explained in Chapter 4. Chapter 5 presents a study to understand and mitigate the performance loss in the neural networks for splitting merged tracks in the ATLAS silicon pixel detector due to increased particle luminosity in the future LHC data taking periods as well as radiation damages to the silicon pixel detectors.

Chapter 2

Theoretical Background

This chapter provides a brief overview of the theoretical framework on which the rest of the work in thesis depends. The chapter includes a compact review of the Standard Model of particle physics and a discussion of the $H \to WW^* \to \ell\nu\ell\nu$ channel in Section 2.1.3.2 which forms the basis of the main analysis of this thesis in Chapter 9. Further details on these topics can be found in any standard textbook on the subject such as [21].

2.1 The Standard Model

The Standard Model is a highly successful theory that describes the elementary particles of nature and the forces through which they interact [22, 23, 24, 25]. It is a Yang-Mills theory [26] with the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry group, where C indicates the color charge, L the weak isospin and Y the hypercharge. The theory contains the subatomic half-integer spin particles (fermions), and the integer spin particles (bosons) which mediate their strong and electroweak (EW) interactions (see Figure 2.1). The weak interaction is mediated by the massive bosons $W^{\pm}$ and Z, while the electromagnetic interaction by the photon γ, and the strong interaction by the massless gluon g.

There are two classes of fermions: leptons which only interact via the electroweak interaction, and quarks which can also interact strongly.

There are three generations of each class of fermions, with every generation of leptons and quarks organized in $SU(2)_L$ doublets and singlets for the left-handed and right-handed chirality states, respectively. In the SM, the neutrino is assumed to

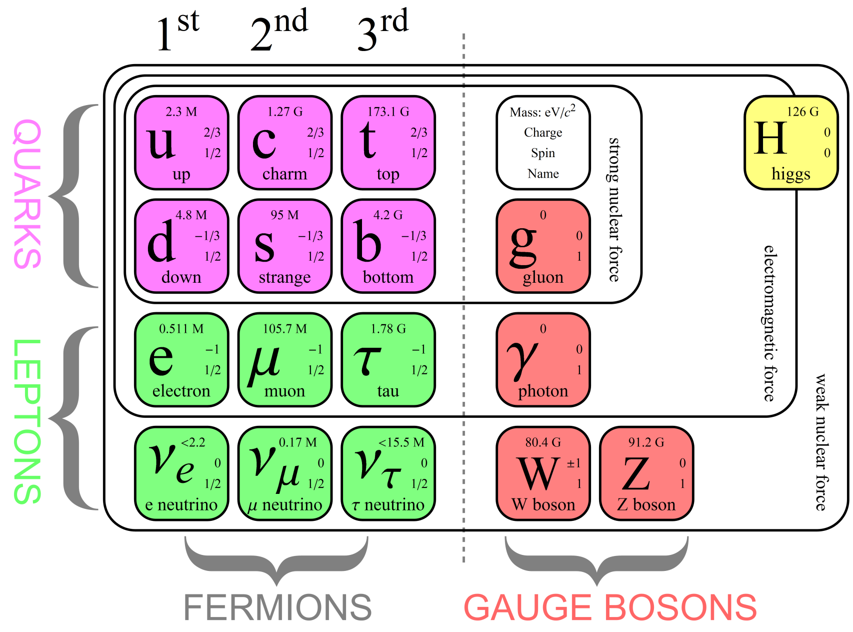


Figure 2.1: The particles of the SM and their properties; three generations of fermion quarks and leptons, the gauge bosons, and Higgs boson [1].

be massless and no right-handed chirality state is present. The quarks can carry any of the three possible color charges - red, green or blue, and are present in $SU(3)_C$ triplets, while leptons, being colorless, are present in $SU(3)_C$ singlets.

The complete SM Lagrangian can be factorized in terms for the electroweak interactions, the Higgs sector, the Yukawa terms for the fermion masses and the strong interactions:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{H} + \mathcal{L}_{Yukawa} + \mathcal{L}_{strong} \tag{2.1}$$

In the following subsections these terms in the SM Lagrangian would be discussed briefly.

2.1.1 The Electroweak Interaction

The electroweak theory describes the electromagnetic and weak interactions between the particles. The fields are separated into left-handed and right-handed chirality states and with $SU(2)_L$ doublets or singlets:

$$Q_i = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L, u_{i,R}, d_{i,R} \tag{2.2}$$

, and

$$L_i = \begin{pmatrix} \nu_i \\ \ell_i \end{pmatrix}_L, \ell_{i,R} \tag{2.3}$$

where the index i runs over the three generations of quarks and leptons.

The electroweak interaction has $SU(2)_L \otimes U(1)_Y$ symmetry

$$\mathcal{L}_{EW} = \sum_{fravors} i(\bar{L}\not{D}L + \bar{Q}\not{D}Q + \bar{\ell}_R\not{D}\ell_R + \bar{u}_R\not{D}u_R + \bar{d}_R\not{D}d_R) - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a \tag{2.4}$$

$\not{D} = D^\mu\gamma_\mu$, where $D = \partial_\mu - \frac{1}{2}ig'YB_\mu + \frac{1}{2}ig\tau^a W^a_\mu$ is the covariant derivative, with γ being the Dirac matrices, τ^a are the Pauli matrices, and $W^a_{\mu\nu}$ and $B_{\mu\nu}$ are the gauge fields defined as:

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc}W^b_\mu W^c_\nu \qquad B^a_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.5}$$

where the fact that the structure constants for $SU(2)$ are represented by the component of the three dimensional Levi-Civita tensor ϵ^{abc} has used and g is the coupling constant for the weak isospin. Here, Y is the hypercharge defined as $\frac{Y}{2} = Q - T_3$, where Q is the electric charge and T_3 is the eigenvalue of the third component of the weak isospin.

2.1.2 Electroweak Spontaneous Symmetry Breaking

In a Yang-Mills theory, mass terms for fermions and gauge boson fields are not included as they would break the gauge invariance, making all the particles massless.

To solve this problem in the SM, the Brout-Englert-Higgs (BEH) mechanism of Electroweak Spontaneous Symmetry Breaking [27, 28, 29, 30, 31, 32] is used. In this mechanism, an $SU(2)_Y$ complex field doublet ϕ field is introduced.

$$\mathcal{L}_H = (D_\mu \phi)^2 - \mu^2 \phi^2 - \lambda \phi^4 \tag{2.6}$$

where the $\mu^2 < 0$ and $\lambda > 0$ such that the ϕ field has a non-vanishing vacuum expectation value (vev) given by $v = \sqrt{\frac{-\mu^2}{\lambda}}$.

The choice of the vacuum state is arbitrary and spontaneously breaks the EW symmetry group into the electromagnetic group $U(1)_Q$ (here Q is the electric charge), giving in this way mass to the weak force carrier bosons and leaving the photon massless, and introducing an extra physical scalar field h. The physical gauge boson fields are obtained by linear combination of the W^1_μ, W^2_μ, and B_μ fields

$$W^\pm_\mu = \frac{(W^1_\mu \mp W^2_\mu)}{\sqrt{2}} \qquad A_\mu = sin\theta_W W^3_\mu + cos\theta_W B_\mu \qquad Z_\mu = cos\theta_W W^3_\mu - sin\theta_W B_\mu \tag{2.7}$$

where the mixing angle θ_W is called Weinberg angle and is defined by the condition $sin\theta_W = g'/\sqrt{g'^2 + g^2}$.

The masses of the physical gauge bosons is given by:

$$M^\pm_W = \frac{1}{2} v g \qquad M_Z = \frac{M_W}{cos\theta_W} \qquad M_A = 0 \tag{2.8}$$

showing that the gauge bosons responsible of the weak interactions gain mass while the photon remains massless. The extra physical field is called the Higgs field and has a corresponding mass given by $M_h = \sqrt{2\lambda} v$.

A mass term for the fermions can also be obtained through the BEH mechanism in a way that preserves gauge invariance adding to the EW Lagrangian the Yukawa terms. Considering a single generation, such terms have the form

$$\mathcal{L}_f = \left(-\lambda_e \bar{E}\phi e_R - \lambda_d \bar{Q}\phi d_R - \lambda_u \epsilon^{ab}\bar{Q}_a \phi_b^\dagger u_R + h.c.\right) \tag{2.9}$$

where ϵ^{ab} is the two dimensional total anti-symmetric tensor with $\epsilon^{12} = 1$ and λ_e, λ_u and λ_d are arbitrary parameters. Due to the spontaneous symmetry breaking the fermions assume masses given by

$$m_e = \frac{1}{\sqrt{2}}\lambda_e v \qquad m_u = \frac{1}{\sqrt{2}}\lambda_u v \qquad m_d = \frac{1}{\sqrt{2}}\lambda_d v \tag{2.10}$$

.

The physical Higgs boson was observed, with mass 125.09 ± 0.21 ± 0.11 GeV by the ATLAS and CMS collaborations in 2012 [33, 34].

2.1.3 Higgs Physics at the LHC

The energy scales of the Large Hadron Collider (LHC), with center-of-mass energy of up to $\sqrt{s} = 14$ TeV, have been able to finally discover the Higgs boson given its heavy mass and small production cross section. In proton-proton collisions at the LHC, the Higgs can be produced through a variety of different processes described Section 2.1.3.1. Once created, it very quickly decays (lifetime of $\sim 10^{-22}$ seconds) before reaching the detector active material one of the modes introduced in Section 2.1.3.2.

2.1.3.1 Higgs Boson Production

The Higgs boson has a higher coupling to massive particles with the coupling strength proportional to the fermion mass and the weak boson mass squared. So, production mechanisms of Higgs boson involving either heavy quarks or massive vector bosons have higher cross-section. The p-p collisions at the LHC mostly involve gluons and light quarks.

As a result, the leading Higgs boson production mechanism at the LHC is gluon-gluon fusion (ggF) (see Figure 2.2).

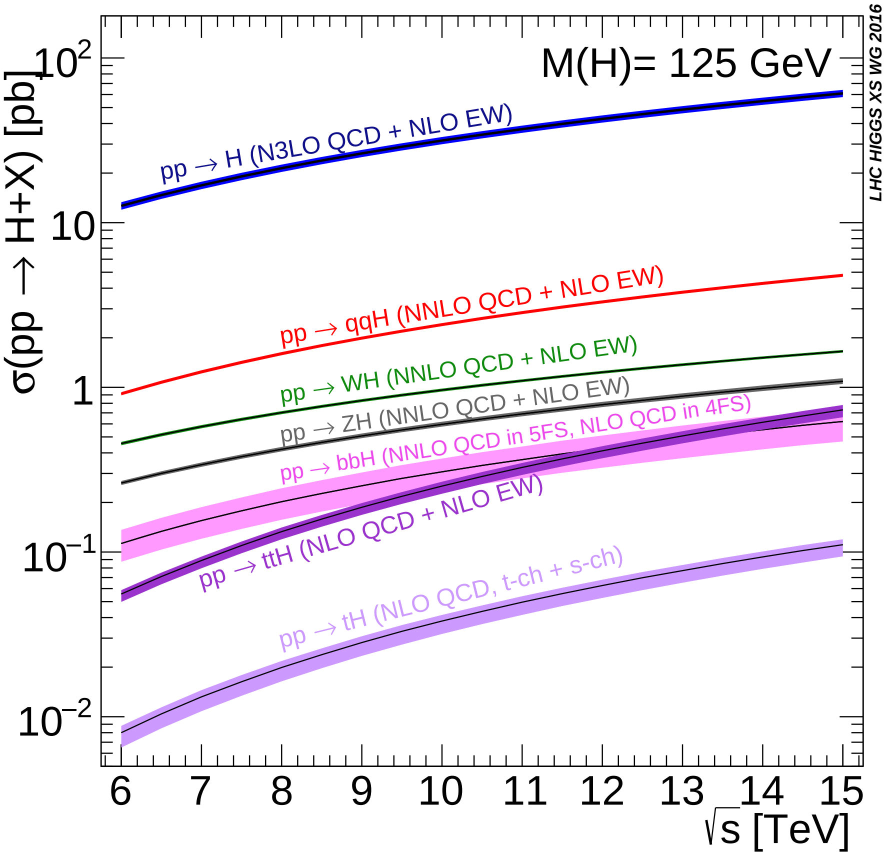


Figure 2.2: Standard Model Higgs production cross sections as determined by theory for different production modes at the LHC. The blue line denotes ggF, while the red line denotes VBF [2].

Since the Higgs boson does not couple to gluons directly, the ggF mechanism involves a quark loop originating from two high energy gluons (see Figure 2.3). Due to higher coupling of Higgs boson to massive fermions, the quark loop is made up of top quarks for the vast majority of events, with only a small contribution from b-quark and lighter quark loops. The ggF production mechanism is useful to probe into the coupling of the Higgs boson to the fermions (Yukawa coupling).

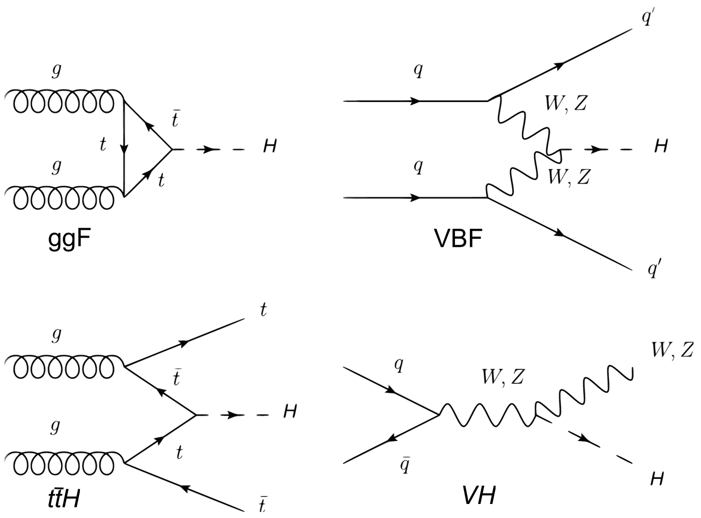


Figure 2.3: Leading Feynman diagrams for different production modes of the Higgs at the LHC in order of largest to smallest cross section. Gluon fusion (ggF) is shown in top-left, vector boson fusion (VBF) is shown in top-right, Higgs strahlung production (WH/ZH) is shown in bottom-left, while associated production with top ($t\bar{t}H$) is shown in bottom-right.

The Higgs boson production mechanism involving weak vector bosons is the vector boson fusion (VBF) which is the second leading Higgs boson production mechanism at the LHC (VBF is ~ 10 times smaller in cross section than the ggF mechanism). In the VBF production mechanism, two high energy quarks radiate weak vector bosons fuse into the Higgs boson as shown in Figure 2.3. The two quarks in the final state of the VBF process produce two forward jets in opposite directions, which serve as an important signature in identifying the VBF Higgs boson production process. The VBF production mechanism gives access to the coupling of the Higgs boson to the weak sector. Since electroweak interactions are theoretically understood better than strong interactions, the VBF cross section have smaller theoretical uncertainties than the ggF cross section. The analysis summarized in the Chapter 9 focuses on the VBF Higgs boson production mechanism.

In addition to ggF and VBF, Higgs boson can also be produced via Higgs Strahlung (see Figure 2.3) and top-Higgs associated production (ttH) (see Figure 2.3) production mechanisms at the LHC, although due to low cross sections their contribution is expected to be significantly less.

2.1.3.2 Higgs Boson Decay

Among the decay modes of the Higgs boson, summarized in Figure 2.4, the $b\bar{b}$ channel has the highest branching ratio ($\sim 58\%$).

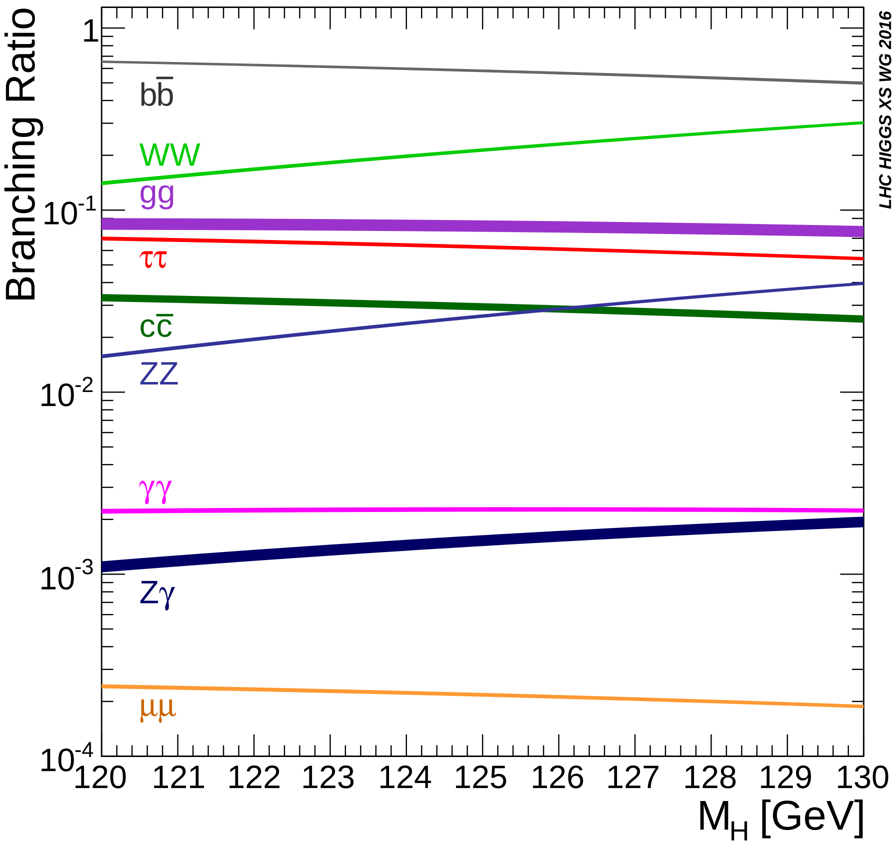


Figure 2.4: Standard Model branching fractions for different Higgs decay modes [2].

However due to difficulty in differentiating the $b\bar{b}$ process from the common backgrounds originating from the strong interaction, the second leading decay mode Higgs boson - $H \to WW^*$ (branching ratio $\sim 21\%$) has been used for the search of the VBF

Higgs boson production in Chapter 9.

Since there are two W bosons in the $H \rightarrow WW^*$ channel, one of the W bosons is off-shell (marked by the asterisk) with a smaller mass than the on-shell W boson, as the combined mass of two on-shell W bosons is larger than the SM Higgs boson mass(m_H = 125.1 GeV). The $H \rightarrow WW^*$ channel is a promising channel to measure Higgs boson properties as it provides access to exclusive coupling of the Higgs boson to the vector bosons. The cleanest signature in the detector is obtained when considering the case where both W bosons decay into an electron or muon and a neutrino. But the branching ratio of the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ is only 1% and, due to the presence of two neutrinos, the mass of the Higgs boson cannot be directly measured in this final state. However, the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ final state provides some distinct kinematic features making it easier to distinguish from its major backgrounds. The Higgs boson being spin 0 particle, decays into two W bosons with

The SM Higgs boson is a spin-0 particle and so the two spin-1 W bosons from its decay must have opposite spin projections, to conserve angular momentum, as indicated by the double arrow sign Figure 2.5.

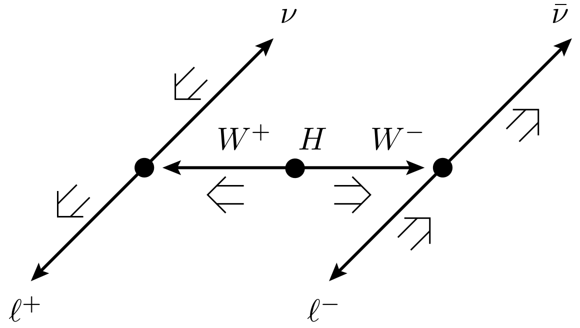


Figure 2.5: Illustration of the $H \rightarrow WW^*$ leptonic decay. The small arrows denote the particle direction of motion, the wide arrows the spin projection on to the direction of motion. Due to angular momentum conservation and the chiral structure of the weak interaction the $\ell^+\ell^-$ pair are preferentially emitted with a small opening angle in the laboratory frame.

Similarly, when a W boson decays into two spin-1/2 leptons, they must have aligned spin projections. Due to the left-handed nature of the weak interaction, neutrinos will have spins opposite to their direction of motion, while anti-neutrinos will have spins parallel to their direction of motion. As a result, the opening angle between the charged leptons is the Higgs boson is expected to be small in the $H \rightarrow WW^*$ rest frame. This small angular gap also leads to a smaller $m_l l \equiv 2E_{\ell 1}E_{\ell 2}(1 - cos\Delta\phi$ (where $\Delta\phi$ is the opening angles between the charged leptons $\ell 1$ and $\ell 2$, and their masses are assumed to be zero for electrons and muons).

2.1.4 The Strong Interaction

The strong interactions are described by the Quantum Chromodynamics (QCD) theory, based on the $SU(3)_C$ gauge symmetry group. Quarks, which have color charges, are represented as triplets while leptons, that do not interact strongly, are represented as color singlets. The $SU(3)$ group has 8 generators T^a (related to the Gell-Mann matrices λ^a in the fundamental representation of the group), representing the 8 gluon fields mediate the strong interaction. The QCD Lagrangian is given by:

$$\mathcal{L}_{strong} = \sum_q \bar{\psi}_{q,i} \left(i\not{D}_{i,j} - m_q \delta_{ij} \right) \psi_j - \frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} \tag{2.11}$$

where $\psi_{q,i}$ is the fermion field q and color i and the covariant derivative operator D^μ and the gluon fields kinetic terms $G^a_{\mu\nu}$ are given by:

$$D^\mu = \partial_\mu + ig_s t^a G^a_\mu G^a_{\mu\nu} = \partial_\mu G^a_\mu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu \tag{2.12}$$

where g_s is the strong running coupling constant and f^{abc} are the $SU(3)$ structure constants given by commutation relations between the generators $[t^a, t^b] = if^{abc}t^c$. The non-vanishing last term for the gluon fields kinetic terms allows self-interactions.

The dynamics of the $SU(3)$ gauge group leads to two important properties of strong interactions: the asymptotic freedom [35] where the strong coupling constant becomes smaller at greater energy scales making perturbative theory applicable at high energies while at low energies analytical calculations are not possible, and the confinement where the quarks are bound into colorless states called hadrons[1]. These important property leads to the fact that when quarks are pulled apart, their interaction becomes so strong that quark-antiquark pairs are created from the vacuum, which leads to the process of hadronization process that leads to spray of hadronic particles from quarks and gluons in the final state, called jets, at the LHC.

---

[1] Hadrons are of two types: mesons which are composed by a quark-antiquark pair and baryons which are composed by three quarks or three anti-quarks.

Chapter 3

LHC and the ATLAS Experiment

This chapter provides a review of the experimental apparatus, i.e. the Large Hadron Collider (LHC) and the ATLAS detector with its subsystems.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [36, 37, 38, 39] is the largest particle accelerator ever constructed. The main LHC ring is a tunnel with circumference of 27 km and is located $\sim$ 100 m below the ground near Geneva, Switzerland at the European Organization for Nuclear Research (CERN). The LHC is a proton-proton collider designed to operate at center-of-mass energy of $\sqrt{s} = 14$ TeV. It delivered $\sqrt{s} = 7$ TeV until 2011 and increased to $\sqrt{s} = 8$ TeV in 2012. In 2015, it has been raised to $\sqrt{s} = 13$ TeV in order to increment in production cross section of Higgs production channels to study their properties as well as heavy particles predicted by beyond the standard model theories, over the increased background rates. In this thesis, all the analyses have been done using $\sqrt{s} = 13$ TeV data collected between 2015 and 2018.

The proton-proton collisions take place at four interaction points (IP) around the ring for the ATLAS, ALICE, CMS, and LHCb experiments (see Figure 3.1). To produce the proton beam, valence electrons are stripped from the hydrogen and accelerated to upto 50 MeV in the Linear Accelerator 2 (LINAC2). The protons are then passed through a series of accelerators before injection into the main LHC ring. The first acceleration in this series is the Proton Synchrotron Booster (PSB) which accelerates the protons to upto 1.4 GeV. The particle beam then enters the Proton Synchrotron (PS) followed by Super Proton Synchrotron (SPS), which accelerates

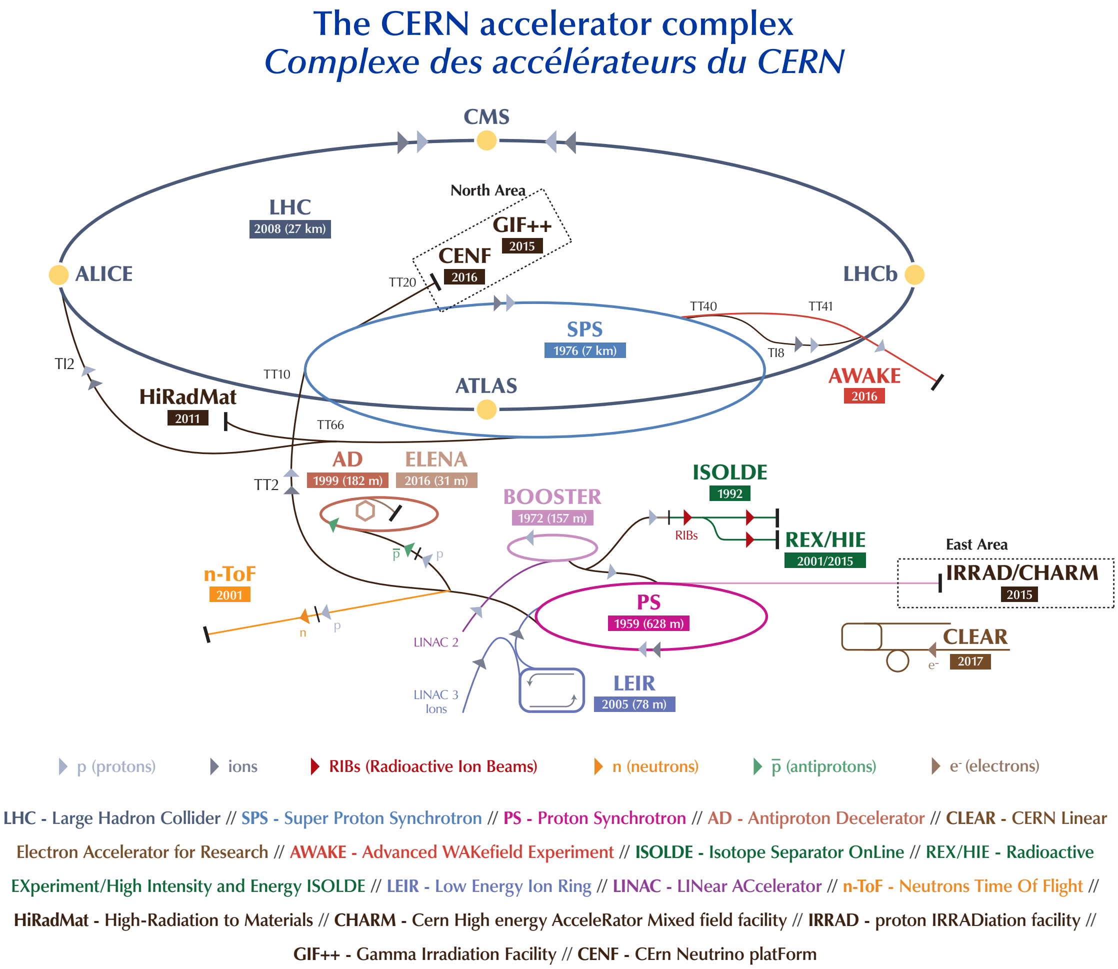


Figure 3.1: The CERN accelerator complex [3].

the protons to upto 450 GeV. The protons in the beams are clumped into bunches, with each bunch contains around 1.1×10^{11} particles, where a bunch is seperated from the next bunch by 25. After the proton beams are injected into the LHC ring, they are accelerated by sixteen superconductive 400 MHz radio frequency (RF) cavities, which are maintained at a temperature of 4.5 K and potential 2 MV to produce of field of 5 MV/m. The proton beams are confined in the LHC ring using 1232 superconducting dipole magnets producing a magnetic field of 8.3 T. The beams are stabilized and focused using 392 quadrupole magnets. ATLAS [5] is a general purpose experiment designed to primarily study the Electroweak Summetry Breaking and the Higgs Mechanism, precisely measure the SM parameters and look for newphysics beyond the SM.

3.1.1 Luminosity

In order to measure the cross sections[1] of scattering processes, one needs to know the luminosity of the collider. The luminosity of a collider is a measure of the number of protons crossing each other per unit time and area. The number of events (N_{proces}) of a scattering process in the data can be expressed in terms of *integrated luminosity* L for data taking period as:

$$N_{process} = \sigma_{process} \cdot \epsilon L \tag{3.1}$$

The ϵ is the experimental efficiency of detecting a process, dependent on the analysis and the detector, and $\sigma_{process}$ is the cross section of the scattering process being studied.

The integrated luminosity of a collider of a period of time can evaluated by inte-

---

[1] Cross section is a measure of probability of a scattering process. It is defined in units of area typically in barns, 1 b = 10^{-28}m^2, measure rare cross sections in particle physics.

grating the *instantaneous luminosity* over that time period:

$$L = \int \mathcal{L} dt \tag{3.2}$$

The instantaneous luminosity depends on a few parameters of the proton beams, expressed as:

$$\mathcal{L} = \frac{N_p^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta_{IP}} F(\theta_c, \sigma_T, \sigma_z) \tag{3.3}$$

where, N_p is the number of protons per bunch and n_b is the number of proton bunches. In Run-2 data-taking conditions, the number of bunches were typically around 2544 in each LHC ring with a maximum possible capacity of 2808 bunches. f_rev is the frequency of revolution of bunch in the LHC ring, γ_r is the Lorentz factor for protons. ϵ_n is the normalized transverse beam emittance. β_{IP} is the beta function at the impact/collision point which represents the transverse size of the beam.

$F(\theta_c, \sigma_T, \sigma_z)$ is geometrical correction factor accounting for the reduction in luminosity due to beam-beam crossing angle, θ_c (σ_T and σ_z represent the transverse and longitudinal bunch lengths, respectively). It is defined (with Gaussian approximation of the beam distribution) as:

$$F = \frac{1}{\sqrt{\left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma_T}\right)^2\right)}} \tag{3.4}$$

3.1.2 Pileup

At high luminosity such as LHC, multiple proton-proton interactions per bunch crossing take place, which is also known as *in-time pileup*. Along with in-time pileup, there are interactions in neighboring bunch-bunch crossings which contribute to detector occupancy of detectors such ATLAS in the LHC, and is known as the *out-of-time*

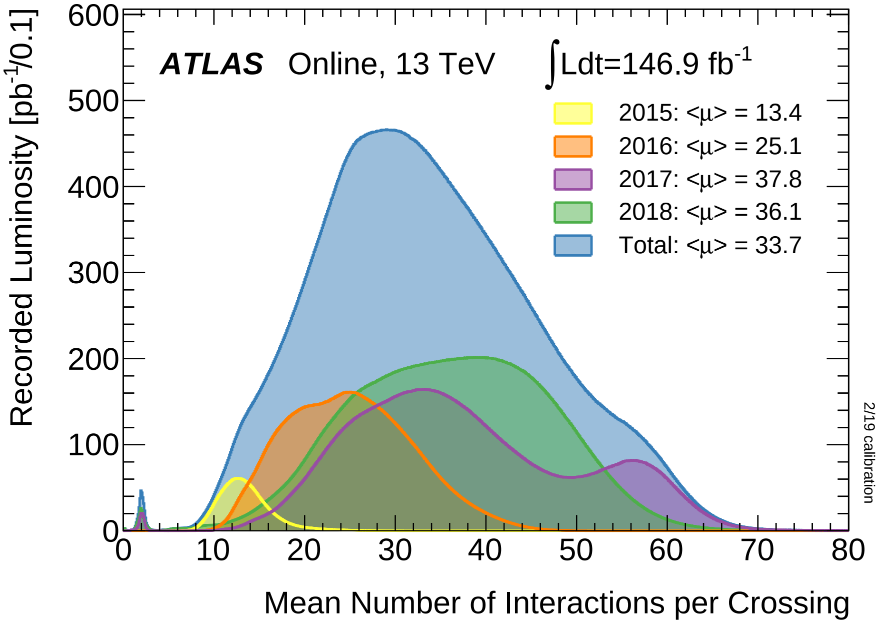


Figure 3.2: The average number of interactions per bunch-bunch crossing $\langle\mu\rangle$ in ATLAS data from 2015 to 2018 [4].

pileup. Pileup is measured using the average number of interactions per bunch crossing $\langle\mu\rangle$, and the number of interaction vertices per event N_{vtx}. For Run 2, the $\langle\mu\rangle$ distribution for ATLAS had a spread between 10 and 70 (as shown in Figure 3.2) due to changing operating conditions of the LHC and the natural decay in the number of interactions per crossing within the lifetime of a beam. Pileup has a significant contribution in misidentified lepton rate which is one of the major experimental backgrounds in the $H \rightarrow WW^*$ analysis results presented in this dissertation.

3.2 The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) experiment [5] is one of the two general purpose particle physics detectors at the LHC, along with the CMS experiment, to test the SM as well as explore the physics beyond. The ATLAS detector is cylindrical

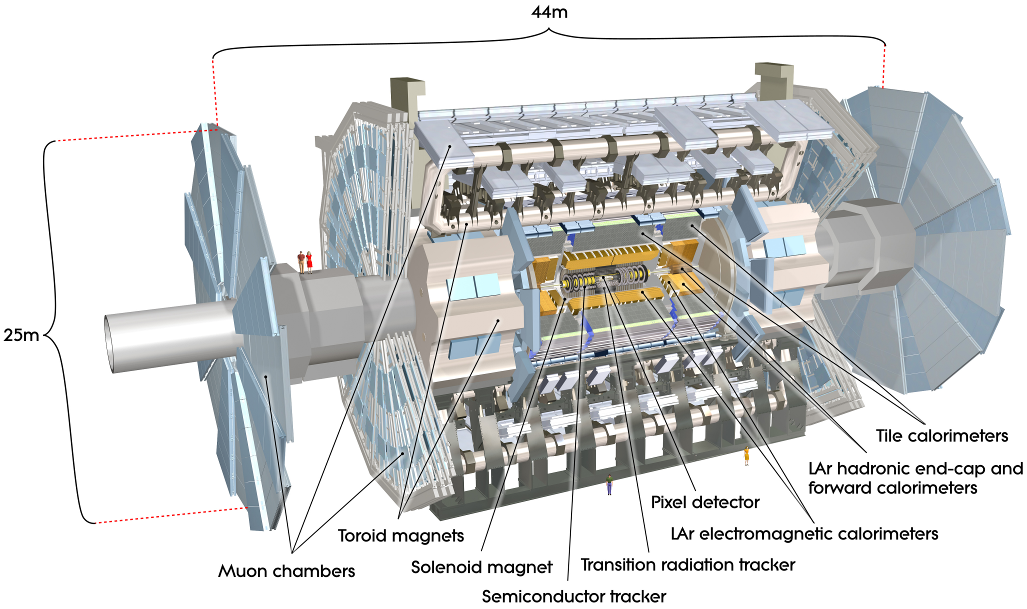


Figure 3.3: Computer generated image of ATLAS, showing the various sub-systems and the size of the whole detector. The figure is taken from [5].

about the beam pipe and is centered around the crossing point of the beams (as shown in Figure 3.3) to have a forward-backward symmetry with respect to the interaction point. The various sub-systems of the ATLAS detectors nearly provide a full solid angle coverage about the interaction point to capture most of the particles from scattering. These sub-systems closest to the beam pipe are the Inner Detector (ID), embedded in a 2T axial magnetic field, followed by the Electromagnetic Calorimeter (ECAL), the Hadronic Calorimeter (HCAL) and the Muon Spectrometers which is torroidal magnets. These sub-systems would be described in the following sections in this chapter.

3.2.1 Coordinate System

The reference frame of the ATLAS detector is a right-handed Cartesian coordinate frame with the origin at the proton-proton interaction point (the center of the detector). The positive x-axis points to the centre of the LHC ring, the positive y-axis points upwards, such that the positive z-axis aligns with the beam pipe (longitudinal direction). Due to the cylindrical symmetry around z-axis, a cylindrical coordinate system is generally preferred to describe the experiment. The azimuthal angle $\phi \in (-\pi, \pi]$ is defined as the angle with respect to the x-axis with $\phi = 0$ corresponding to the positive x-direction. The polar angle $\theta \in [0, \pi]$ is defined with respect to the z-axis with $z = 0$ corresponding to the positive z-direction. The polar direction of the particles is generally expressed in the pseudorapidity $\eta = -ln[tan(\theta/2)]$ ($= \frac{1}{2}ln\left(\frac{E+p_z}{E-p_z}\right)$ for massless particles), instead of θ, since the difference in pseudorapidity is invariant under Lorentz boosts along longitudinal direction. The $\eta = 0$ corresponds to the transverse (x-y) plane, while $\eta \rightarrow \infty$ aligns with the beam-pipe. The angular distances between objects are expressed in Lorentz invariant $\Delta R = \sqrt{(\Delta\phi^2 + \Delta\eta^2)}$.

3.2.2 Inner Detector (ID)

The ATLAS detector is a multipurpose particle physics detector with a forward-backward symmetric cylindrical geometry and nearly 4π coverage in solid angle.[2] The inner detector (ID) provides an $|\eta| < 2.5$ coverage.

The inner detector (see Figure 3.4) consists of three semiconductor subdetectors - IBL [40, 41], pixel detector [42, 43], and SCT [44, 45], to provide high spatial resolution and are radiation hardened. When a charged particle passes through a semiconductor it leaves an ionized track behind with charge carriers (negative elec-

---

[2] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector. The positive x-axis is defined by the direction from the interaction point to the centre of the LHC ring, with the positive y-axis pointing upwards, while the beam direction defines the z-axis. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ by $\eta = -\ln\tan(\theta/2)$. Rapidity is defined as $y = 0.5\ln[(E + p_z)/(E - p_z)]$ where E denotes the energy and p_z is the component of the momentum along the beam direction.

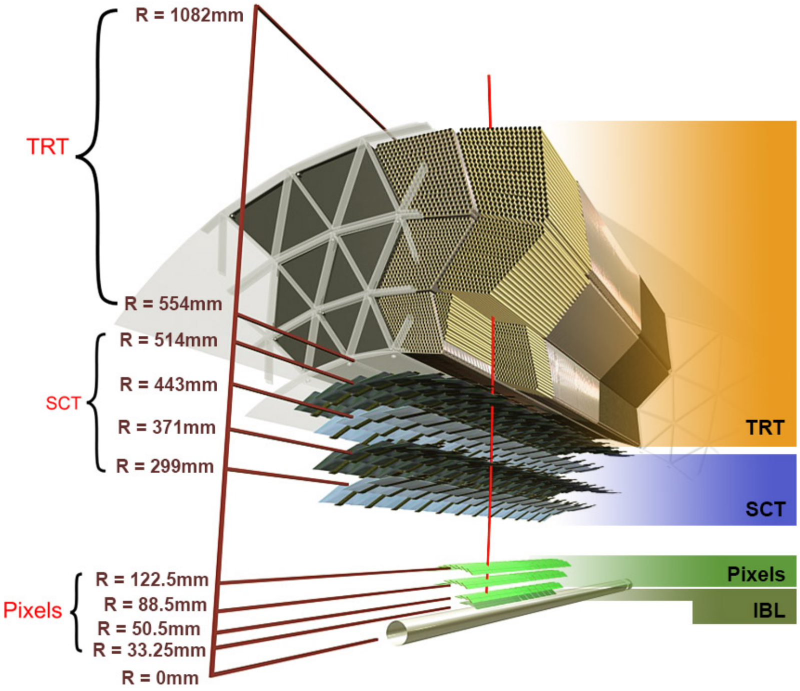


Figure 3.4: The ATLAS Inner Detector [6]

trons or positive holes respectively for n or p-type doped semiconductors). These charge carriers then produce ionization currents under an applied bias voltage which can be picked up by readout electronics. The outlermost subdetector of the inner detector is the Transition Radiation Tracker (TRT) which is a gaseous detector which provides particle identification (PID) information for electrons.

3.2.2.1 The ATLAS Pixel Detector

The pixel detector is made up of three barrel layers and three endcap disks. The pixel modules are formed by 16 front-end (FE) chips, each one with 2880 read-out channels and a n-type pixel sensor with a thickness of 250 μm and area $R\phi \times z = 50$ μm $\times 400$ μm. In total the pixel detector has $\sim$ 80 $\times$ 10^6 read-out channels. The intrinsic hit resolution of each module is 115 μm in the longitudinal direction $(R-\phi)$ and 10 μm in the transverse direction (z).

The insertable B-layer (IBL) [40, 46] can be thought of as a fourth, innermost, pixel layer which has been added in 2014 to improve tracking precision and b-tagging performance, at the higher luminosities of Run 2. The IBL has a mean radius of $33mm$ and a pixel has a size of 50 μm by 250 μm in the transverse and longitudinal directions, respectively, with a thickness of 200 μm and contributes to additional $\sim 6 \times 10^6$ read-out channels. Together the IBL and pixel detector give ATLAS a vertex reconstruction resolution of 11 μm in x and y, and 24 μm in z [40].

The pixel detectors measure the charge collected in each individual pixel using the time over threshold (ToT) [43], which is the time the pulse exceeds a given threshold and is proportional to the deposited energy. The pixel detectors in the IBL store 4-bit ToT while rest of the detectors store 8-bit ToT.

3.2.2.2 Semiconductor Tracker (SCT)

Outside the pixel detectors are the Semiconductor Tracker (SCT) [5] which consists of four double strip layers at radii of 299 mm to 514 mm and nine discs in both the end-caps, giving a coverage of $|\eta| < 2.5$. A typical strip of a barrel SCT sensor has a length of $126mm$ and a pitch of 80 μm. On each layer, the strips are parallel to the beam direction on one side and at a stereo angle of 40 mrad on the other in order to obtain information on the z-coordinate along the strip length and provide a 3D measurement in space. The intrinsic resolution of the SCT modules is of 17 μm in the transverse direction in the barrel and 580 μm in the longitudinal direction.

3.2.3 Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) [47, 48, 49] is the outermost of the ID sub-detectors and is made of 350848 straws filled with a $Xe/CO_2/O_2$ gas mixture (70/27/3%) giving a coverage of $|\eta| < 2.0$. The tubes are 4 mm in diameter and have a 31 μm diameter gold plated tungsten wire as anode. The straws are placed parallel to the beam line and organized in 73 modules of length 144 cm and interleaved with polypropylene fibers in the barrel, while they are arranged radially in the end-caps and organized in 160 layers interleaved with polypropylene foils. The relativistic charged particles traversing the dielectric material embedding the tubes produce transition radiation (TR) photons that are absorbed by the TRT gas mixture and the resulting ionization is collected in the anode wire. The signal on each wire is amplified to get tracking information and particle identification. Since the amount of transition radiation depends of the Lorentz factor $\gamma = E/m$ of the particles, thus for a particular amplitude of collected charge the electron tracks from hadronic tracks can be discriminated by by mass. The TRT provides in average 34 hits, although only a 2D hit information along the transverse plane, in the $|\eta| < 1.7$ region with an intrinsic resolution of 130 μm.

3.2.4 Calorimeters

The calorimeter system is placed outside the Inner Detector and the solenoid. ATLAS has an inner high granularity liquid-argon (LAr) electromagnetic calorimeter (ECAL) [50] and an outer hadronic calorimeter(HCAL) [51], as shown in Figure 3.5. The calorimeters are present in the barrel, end-cap and forward region providing a coverage of $|\eta| < 4.9$, with a transition region gap at $1.37 < |\eta| < 1.52$. Both these calorimeters are non-compensating sampling calorimeters, which consists of alternating layers of dense absorbing material, like lead or steel, and active detector material. The absorbing layers slow down incoming particles (losing some energy of the particles) and promote showering while the active material samples the energy of the resulting showers, through ionization (in ECAL) or scintillation (in HCAL). Through several of such alternating layers in the calorimeters, all the particles, excluding muons and neutrinos, are stopped and their energies are measured. Due to energy lost in the absorbing layers, the detector response is careful calibrated to recover the energy of the particle.

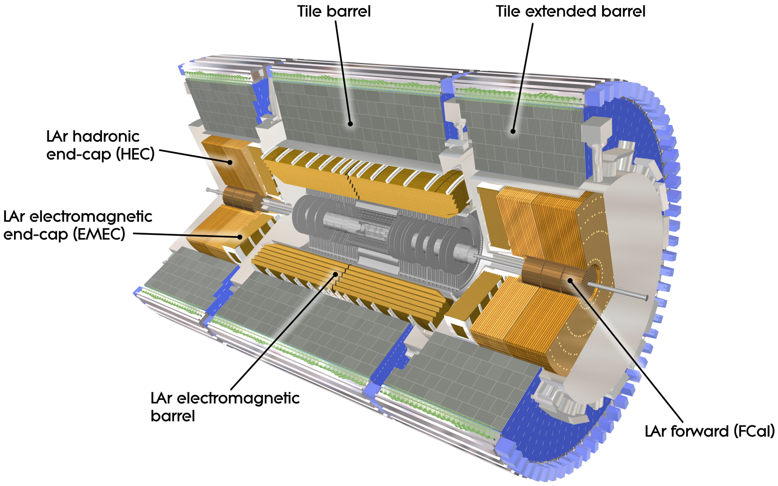


Figure 3.5: The ATLAS calorimeter system [7].

3.2.4.1 Electromagnetic Calorimeter (ECAL)

The Electromagnetic Calorimeter (ECAL) [5, 50] is a lead-liquid Argon (LAr) sampling detector. It has electrodes shaped like an accordion which provides a full ϕ coverage without any crack (and ϕ symmetry). ECAL has three components - one barrel component and two end-cap components. The barrel component has two identical structures separated by a 4 mm gap at $z = 0$. The two end-cap components have two co-axial disks each. Each of the components have three layers (as shown in Figure 3.6).

The first layer has a thickness of 4 radiation lengths[3] (X_0) and granularity of

---

[3] High energy electrons lose their energy by bremsstrahlung, and high-energy photons lose by e^+e^- pair production, as they traverse through the absorbing material in the calorimeters. The radiation length X_0 is the mean distance traversed in the absorbing material by a high-energy

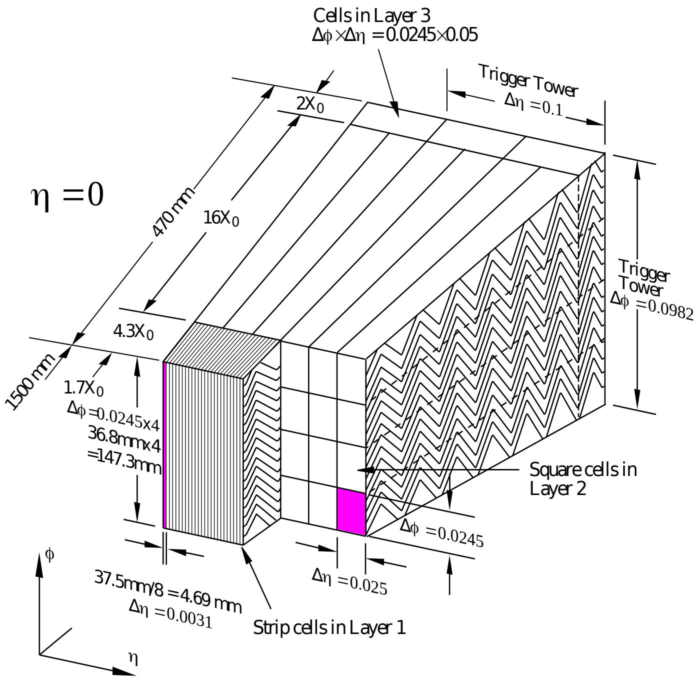


Figure 3.6: ATLAS ECAL barrel geometry [5].

$\Delta\eta \times \Delta\phi = 0.003 \times 0.1$ to precisely measure the energy deposited by electrons and photons used for their identification. The second layer is $16X_0$ thick with granularity $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. This is the thickest layer of the ECAL and is used to absorb the EM shower. The third layer is $2X_0$ thick with granularity $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$. This is the thinnest layer located at the outer edge of the ECAL is used to collect the tail of the EM shower. The Energy resolution of ECAL contains a stochastic term due to EM shower fluctuation and a constant term of 0.17% from local non-uniformities, summed in quadrature $\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E(GeV)}} \bigoplus 0.17\%$.

---

electron to reduce its energy to $\frac{1}{e}$ of its energy by bremsstrahlung, and a high-energy photon to travel $\frac{7}{9}$ of its mean-free path for pair production [52].

3.2.4.2 Hadronic Calorimeter (HCAL)

It is also a sampling calorimeter with three-subsystems to cover different pseudorapidity regions. The *Tile Sampling Calorimeter* uses scintillating plastic tiles with steel absorbers as active material. The UV scintillations in the scintillating tiles are wavelength-shifted to visible region in optical fibers and measured by photomultiplier tubes located on the outer edge of the calorimeter (shown in Figure 3.7).

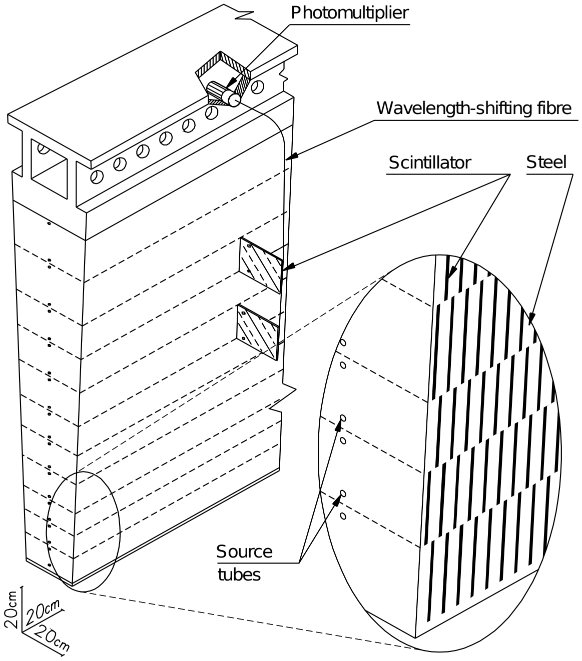


Figure 3.7: ATLAS HCAL tile barrel geometry [5].

It consists of three components - one central barrel (covering $|\eta| < 1.0$) and two extended barrel (covering $0.8 < |\eta| < 1.7$). Each of these components consists of three layers each with thickness 1.5, 4.1, and 1.8 interaction lengths[4] (λ_I) for the central barrel and 1.5,2.6,3.3 λ_I for each of the two extended barrels, respectively. For forward region, *LAr hadronic end-cap calorimeter* (HEC) covers the region $1.5 < |\eta| < 3.2$. It consists of two wheels each with 4 layers in depth. These layers are built using

---

[4] The average nuclear interaction length λ_I is the distance required to reduce the number of particles by a factor of $\frac{1}{e}$ [53].

32 wedge-shaped modules, which use copper plates interleaved with 8.5 mm LAr gaps as active material. The forward most region is covered by the *LAr Forward Calorimeter* (FCAL) which consists of 2 end-caps with 3 modules each. The FCAL uses LAr as active material with copper absorber to absorb EM activity and tungsten as absorber for hadronic activity, providing a thickness of 10 λ_I. Hadronic showers are harder to measure than pure EM showers, as energy of the shower absorbed in nuclear interaction with the absorbing material escapes the detector as neutral long-lived decay products, such as neutrinos, neutrons and kaons. The energy resolution of HCAL [5] for pions is $\frac{\sigma(E)}{E} = \frac{56.4\%}{\sqrt{E(GeV)}} \bigoplus 5.5\%$.

3.2.5 Muon Spectrometer (MS)

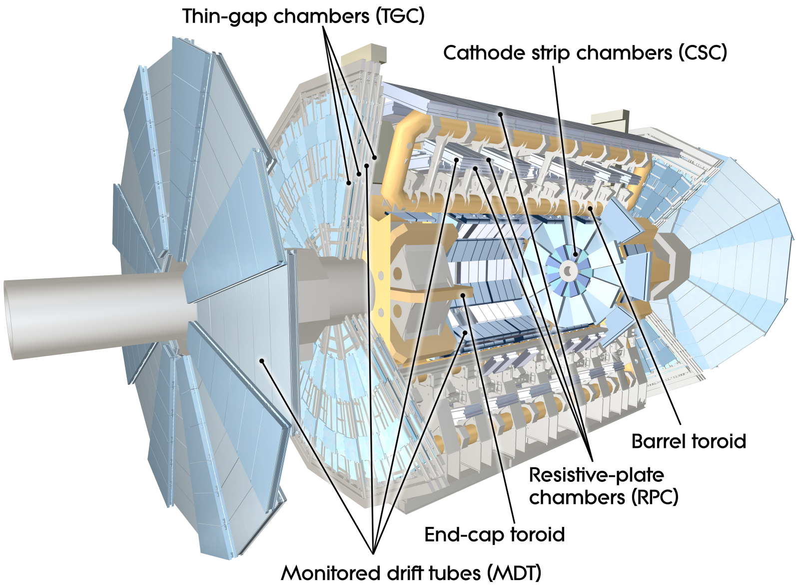


Figure 3.8: The ATLAS Muon Spectrometer (MS) [8]

Muon Spectrometers [54, 55] are the outermost layers of the ATLAS detector. The central part of the MS components (shown in Figure 3.8) is enclosed in three torroidal magent coils, one for the barrel region and one each for the endcaps, providing an azimuthal magnetic field with bending power of 1.0 to 7.5 Tm ($|\eta|$ dependent) within the MS volume.

Muons are able to penetrate most of the ID and calorimeter layers without loosing much of their original energy, due to lack of strong interaction and heavy mass. Muons are $\sim$ 200 times heavier than electrons, and therefore loose their energy using bremsstrahlung ($-\frac{dE}{dx}\propto\frac{E}{m^2}$) 40,000 times slower than electrons. The MS is able identify muon tracks as well as measure their p_{T} (with resolution 10% of 1 TeV muon tracks [5]). The MS consists of three fast triggering layers, stacked in 3 barrel layers and 6 endcap disks. The precise tracking is done by *Monitored Drift Tubes* (MDT) in the barrel and *Cathode Strip Chambers* in the forward region ($2 < |\eta| < 2.7$). The MS also includes the *Thin Gap chambers* (TGC) and the *Restive Plate Chambers* (RPC).

3.3 Trigger and Data Acquisition System (TDAQ)

ATLAS sub-detectors produce readouts of size 1-2 MB per event through $\sim 10^8$ channels. The proton-proton collision rate at the LHC is $\sim$ 40 MHz. Thus to store every event, ATLAS would have to record data at a rate of $40 - 80$ TB/s, which is impractical. Therefore, to conserve readout bandwidth and storage space, only a fraction of events, which are likely to be interesting from physics perspective, are selected in real-time and stored via the ATLAS trigger and data acquisition system (TDAQ). Following the major upgrades after Run-1, the ATLAS TDAQ system consists of two levels, the hardware-based level-1 (L1) [56] and the software-based high-level trigger (HLT) [57].

The L1 system needs to be extremely fast. It uses low-granularity information from the calorimeters and the muon system using custom electronics to rapidly return a L1 acceptance and region-of-interest (ROI) in the $\eta - \phi$ where any interesting activity is detected is passed on to the HLT. Calorimeter trigger towers are fed into sliding window algorithm (discussed in Section 4.2.1) to identify candidate electron, photon, hadronically decaying τ-lepton energy deposits, and jets as well as E_T^{miss}. Trigger muons are identified by simplified linear track finding algorithms with varied levels of coincidence between MS layers to reduce the false positive rate. The L1 system reduces the event rate from 40 MHz down to $\sim$ 100 kHz, reaching a verdict within 2.5 μs.

Chapter 4

Event Reconstruction

In this chapter, various algorithms to efficiently reconstruct the stable particles produced by a proton-proton collision event have been discussed. We start with low level objects such as vertex and track finding, calorimeter clustering and build up to high level physical objects of interest: electrons, muons, jets, b-hadron jets, and missing transverse energy.

4.1 Track & Vertex Reconstruction

Charged particles originating from proton-proton collision leave a series of space-point hits (energy deposits) in the silicon inner detectors and TRT of the ATLAS detector. The charged particles follow a helicoidal trajectory due to the magnetic field along the z direction, and is paramterized in ATLAS as follows:

$$\tau = (d_0, z_0, \phi_0, \theta, q/p) \tag{4.1}$$

Here, d_0 and z_0 are the impact parameters which represent the distance from the point of closest approach to a chosen reference along transverse plane and along the z axis, respectively. The angles ϕ_0 and θ are defined as the azimuthal and the polar angle of the track at its perigee (the point of the track's closest approach in the xy plane), respectively (as depicted in Figure 4.1). Finally, the ratio q/p defines the orientation and the curvature of the helix trajectory.

There are two track-finding strategies employed in ATLAS - *inside-out* and *outside-in.*

The first step in the ATLAS tracking is to group nearby charged silicon pixels,

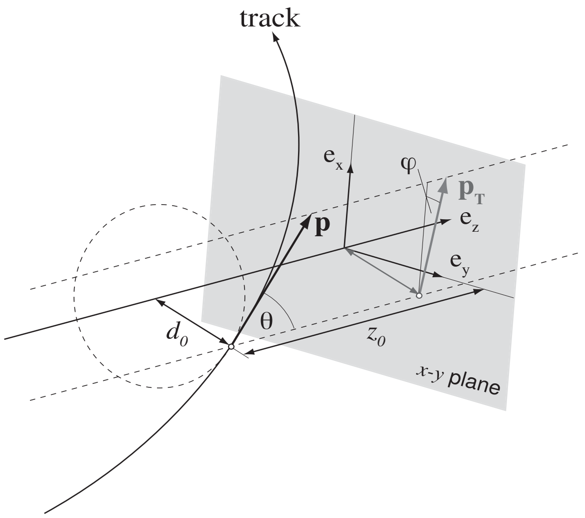


Figure 4.1: ATLAS track paramters [9].

throughout a tracking pixel detector layer, into 7 × 7 pixel clusters using connected component analysis (CCA) [58] algorithm. Then the charge center is linearly interpolated using the charge distribution in a cluster. At least three of these clusters from the silicon detectors are then fitted using pattern recongition algorithms [59] to obtain initial track candidates or seeds (also required to have $p_T > 300$ MeV and $|\eta| < 2.5$). Starting from the seeds, subsequent hits in the outward layers of the silicon tracker which are compatible with the track hypothesis are added to the tracks using combinatorial Kalman Filter [60, 61]. If more than one track candidates pass through a single cluster, the cluster is termed as merged cluster. In the case of merged clusters, the ambiguity is solved by assigning a score to the tracks based on their basic properties and the number of holes [1] and associated hits according to a scoring algorithm [62]. The merged clusters, which are not resolved using the scoring algorithm, are split using a set of artificial neural networks [15] (discussed in Chapter 5), improving the track fit quality.

The inside-out strategy is used to reconstruct the *primary tracks* (originating from

---

[1] Holes are defined as the locations in the detector where a hit is expected by the track fit but not registered.

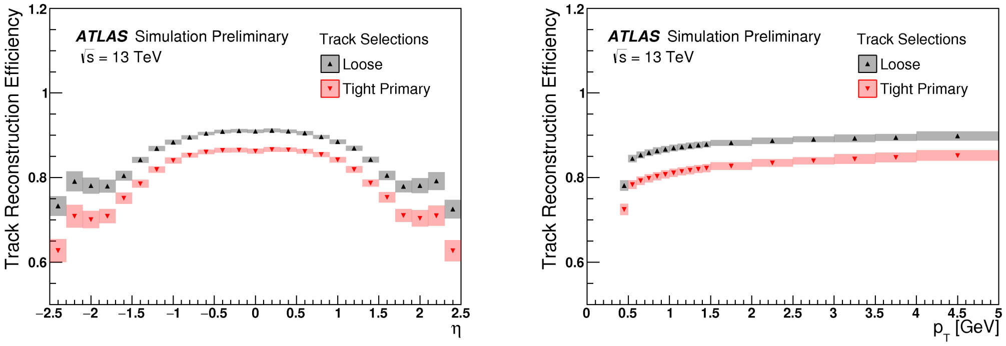


Figure 4.2: Track reconstruction efficiency measured on a PYTHIA [10] Minimum Bias [11] MC simulation [12] as a function of η (left) and p_{T} (right). The error bands indicate the total systematic uncertainty. For (a) $p_{\mathrm{T}} > 400$ MeV and for (b) $|\eta| < 2.5$ requirements are applied. Figures taken from [13].

the proton-proton hard scatter). For *secondary tracks* (decays of long-lived particles) and *conversion tracks* (from particle interaction with the detector), the outside-in algorithm, commonly known as back-tracking, is used which starts from the TRT segments and extrapolates inwards adding silicon layer hits.

The track fitting algorithms can lead to the reconstruction of fake tracks arising from random combinations of hits that cannot be matched with primary or secondary tracks. The rate of fake tracks depend on the quality of tracks. In ATLAS, two sets of quality cuts are defined [12] (see Figure 4.2):

Loose

- $p_T > 400$ MeV
- $|\eta| < 2.5$
- Number of Pixel and SCT clusters on track (‘silicon hits’) $\geqslant 7$
- Number of shared modules $\leqslant 1$
- Number of silicon holes $\leqslant 2$
- Number of pixel holes $\leqslant 1$

Tight-Primary (in addition to the Loose quality cuts)

- Number of silicon hits $\geqslant 9$ (if $|\eta| \leqslant 1.65$)
- Number of silicon hits $\geqslant 11$ (if $|\eta| \geqslant 1.65$)
- At least one hit on one of the two innermost pixel layers
- No pixel holes

When Tight selection is applied, the number of primary tracks increases linearly with pile-up. To estimate the rate of fake tracks, the number of tracks in linearly fitted to the pile-up and any deviation from the fit is considered as due to fake tracks. A vertex is defined as the three-dimensional spatial coordinate from which a track or multiple tracks originate. Vertices associated with primary tracks (originating from p-p scattering in the interaction region) are known as primary vertices, while vertices associated with secondary and conversion tracks (as defined in Section 4.1) are referred as secondary vertices. It is important to have a robust reconstruction of primary vertices in order to suppress in-time pileup. Moreover, the VBF $H \rightarrow WW^*$ analysis rely on identification of b-quarks, which require good resolution to detect displaced vertex from the interaction region.

The reconstruction of primary vertices takes place in two steps, vertex finding and vertex fitting [63]. To begin the process of finding primary vertex position, a tight selection is applied on the ID tracks, and the longitudinal impact parameter z_0 of the selected tracks is evaluated with respect to the center of the beam spot in the z-direction. The longitudinal location of the primary vertex is then estimated to be the global maximum of the longitudinal impact parameter z_0 of all the selected ID tracks. Following the initial estimate of the primary vertex, an iterative method is used to refine its longitudinal position. Tracks close by to the initial estimate of the primary vertex are used as seeds for vertex reconstruction. Each track is assigned a weight, reflecting its compatibility with the primary vertex. The vertex position is recalculated using the tracks taking into account their weights, and their weight are then updated using the new primary vertex position. This procedure is repeated until the vertex position does not change, concluding the vertex finding step.

In the next step of vertex fitting, tracks which are found to be incompatible with the primary vertex by 7σ are removed from the primary vertex. These discarded tracks are used as seeds to search for another primary vertex candidate. The process is repeated until all tracks have been associated to a primary vertex candidate and no additional primary vertex can be found. These primary vertex candidates are also required to fulfill the criteria of having at least two associated tracks and $|z_{PV}| < 200mm$. Out of these primary vertex candidates, the vertex with the largest $\sum p_{T,trk}^2$ is considered as the primary vertex for the event while the rest of the primary vertex candidates are considered as pile-up vertices.

4.2 Calorimeter Clustering

Particle showers deposit their energy in the cells of the EM and hadronic calorimeters. The 4-momentum of particles are recontructed using the direction and total energy of the particle in the ATLAS calorimeter system, in addition to their track and vertex information. To get the total energy deposited by a particle, all the calorimeter cells with energy deposits from that particle have to be clustered. Moreover, the shape of the clusters are also characteristic of the type of the particles - electron, photons, jets, MET. The calorimeter clusters are therefore used in particle identification. To cluster the energy deposits in the ATLAS calorimeters, two algorithms are used, the *sliding window* which useful for clustering deposits from electrons, photons and hadronic τ-leptons, and the *topological clustering* which is used for recontructing jets and MET. These two algorithms have been discussed in details in this section.

4.2.1 Sliding Window Clustering

In the first step of the sliding window method [64], the calorimeter is scanned by a rectangular window of pre-defined dimensions in the η-ϕ plane to find windows with E_T maxima. To do the scan, the calorimeter is divided into a grid with grid elements of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ and the window is moved across each grid element. The dimensions of the sliding window can be defined as $N_\eta \times N_{phi}$, where N_η (N_ϕ) is the number of grid elements along the η (ϕ) direction. The energy of each grid element is defined as the sum of the energies of all cells within the element, across all the calorimeter layers, resulting in a “tower”. At each step of the scan, the E_T inside the window is defined as the total transverse energy of all the towers within the window. The window with a local minima of the E_T (above some threshold energy) is used to find the seeds of the cluster. Position of the cluster seed for each layer within the window is calculated based on the layer type. If two of these seeds fall within two grid units in the η or ϕ directions of each other, the seed with the lower E_T is discarded. All the cells of a layer within $N_\eta \times N_\phi$, centered at the seed of that layer, are included in the cluster. The cells within the window containing the seed is assigned to the cluster.

The dimensions of the window are decided to completely contain the energy deposit of a particle in a layer. So, for electrons and converted photons (photons producing electron-positron pair), which bent under the ID magnetic field in the transverse plane and emit *Bremsstrahlung* photons, the window length in azimuthal (ϕ) direction is wider than for unconverted photons. The dimensions of the window are therefore dependent on type of the particle and the calorimeter layer.

4.2.2 Topological Clustering

Topological Clustering is an iterative algorithm, used to produce variable-size clusters, unlike Sliding Window Cluster. This algorithm produces three-dimensional clusters which is able to reconstruct the shower from a single particle or merged response from several particles, and also part of the shower. The algorithms, as described in [64], works as following:

- Calorimeter cells with $|E_{cell}| > 4\sigma_{noise}$ are selected as the starting seeds of the clusters.

- Around those seed cells, the adjoining cells with $|E_{cell}| > 2\sigma_{noise}$ are added to the cluster.

- The newly added cells become the new seeds for the next iteration and the clustering continues until the $|E_{cell}|$ of neighbouring cells drop below $2\sigma_{noise}$.

- Once the iterations of the above step ends, an additional layer of cells is added to the cluster.

The expected noise of each cell is estimated using test beam data for both LAr and Tile Calorimeter using data collected with triggers.The noise is a cell is defined as the quadrature sum of the electronic noise and the noise due of pile-up,

$$\sigma_{noise} = \sqrt{(\sigma_{noise}^{electronic})^2 + (\sigma_{noise}^{pile-up})^2} \tag{4.2}$$

In Run-2, the effects of pileup dominated over the electronic noise (in Equation 4.2, $\sigma_{noise} \approx \sigma_{noise}^{pile-up}$).

4.3 Physics Object Reconstruction

The particles produced in the pp collision interact differently with the subsystems of the ATLAS detector, based on the particle type. Thus, to reconstruct these particles, their physical properties are exploited in the algorithms which are applied on the low level objects created in the previous sections. In this chapter, the algorithms used to reconstruct the physics objects are discussed.

4.3.1 Electrons & Photons

Electron and photons are reconstructed in the electromagnetic calorimeter using the “sliding window” algorithm (discussed in Section 4.2.1). For electrons, the calorimeter clusters are matched with a track from the inner detector. A track has to match in both position and momentum with the cluster to reconstruct an electron. A multi-variate likelihood method using the shower property variables, track-to-cluster variables, and track quality is used to classify the electron as with either Loose, Medium, or Tight identification working point. A tighter cut gives a higher purity but a lower efficiency.

Photons also induce shower in the electromagnetic calorimeter as electrons, but there are a few variables that can discriminate electron signatures from photon signatures to some extent. One important discriminant variable is that photons are not required to have an associated track in the inner detector. Sometime, when a photon interacts with detector material, it may convert to an electron-positron pair. Such photons are identified by associating an electromagnetic cluster is with a pair of tracks in the inner detector.

4.3.2 Muons

Muons are reconstructed by combining the information from the muon spectrometer and the inner detector, and also from the calorimeter to a lesser extent. Based on the manner in which the information from the different subsystems are combined, there are four types of muons:

- *Combined muons*: a track in the muon spectrometer matched to a track in the inner detector.
- *Extrapolated muons*: a track in the muon spectrometer with a loose requirement on the compatibility with the interaction point.
- *Segment-Tagged muons*: an ID track matched to at least one local track segment in the muon spectrometer.
- *Calorimeter-Tagged muons*: an ID track matched to calorimeter cluster compatible with a minimum ionising particle.

The Extrapolated muons are mainly used for the region $2.5 < |\eta| < 2.7$ as the inner detector is not present. In the case of muons sharing the same ID track, the order of preference to assign the ID track is Combined muons, followed by Segment-Tagged muons and then Calorimeter-Tagged muons. When muons share the same muon system track, the track with better track quality is preferred. Muon identification aims at reconstructing muons at high efficiency while keeping the rate of fakes, coming mainly from kaons and pions decaying in-flight, low. The quality cuts - Loose, Medium, Tight and high-p_{T} selection, are aimed at increasing the reconstruction efficiency of muons while suppressing the misidentified muon background contribution, mainly from kaons and pions decaying in-flight, in varied degree and the type of reconstructed muons. The loose identification criteria is optimized for the $H \to ZZ \to 4\ell$ events and gives a high reconstruction efficiency (above 98% reconstruction efficiency for muons with $p_{\mathrm{T}} > 10$ GeV). The medium identification criteria is only applied extrapolated and combined muons and gives efficiency close to loose criteria. Requirements on the number of hits in the different muon spectrometer are applied based on the type of reconstructed muons. Additionally, for combined muons, a very loose compatibility between the muon spectrometer track and the ID track is required. For combined muons with $p_{\mathrm{T}} > 100$ GeV, the high-p_{T} selection maximizes the momentum resolution. The tight identification criteria can be applied only on combined muons satisfying the medium criteria with additional quality cuts and yields the highest purity with a reconstruction efficiency between 90% and 98%.

4.3.3 Jets

A jet is a spray of tracks in the ID and energy deposits in the calorimeters resulting from the hadronization of color-carrying partons. In ATLAS, these deposits in the calorimeters are first clustered into toplogical clusters Section 4.2.2. These topological clusters are fed to the anti-k_t algorithm [65] which recocntruct the jets. The anti-k_t algorithm combined input objects, topological clusters, sequentially into stable conical jets of raius R specified as a parameter to the algorithm. For this algorithm, a distance measure d_{ij} is calculated for every object i with other objects j. Another distance measure d_{iB} between the object i and the beamline is also calculated. The distance measures are defined as follows:

$$d_{ij} = min(1/p_{T,i}^2, 1/p_{T,j}^2)\frac{\Delta R_{T,j}^2}{R^2} \tag{4.3a}$$

$$d_iB = \frac{1}{p_{T,i}^2} \tag{4.3b}$$

If d_{ij} is the smallest distance measure for object i, it is merged with object j. If d_{iB} is the smallest distance measure for object i, it is declared a jet and not considered for further combinations. This process is repeated until all objects are assigned to jets. The distance measure in Equation 4.3 effectively creates circular jets around high p_T objects. This algorithm is *infrared safe*, which means it is robust against soft radiation and collinear splittings.

In ATLAS, jets are generally of radius $R = 0.4$ as they are optimal for most purposes. For large radius $R = 1.0$ is generally used, were the pileup contributions are limited by trimming. In the process of timming, constituents of the large radius jet are reclustered into subjets of $R = 0.2$ using k_t algorithm [66]. Subjets from $\frac{p_T^{subjet}}{p_T^{jet}} < f_{cut} = 0.05$ or 0.1 are considered to be from pile-up and are removed.

4.3.4 Missing Transverse Momentum

Neutrinos and some waekly interacting hypothized particles in beyond-the-Standard-Model theories can escape the detector without leaving a trace. Such particles can be indirectly detected from the imbalance in momentum in the transverse plane, since if all the particles were directly detected the total tranverse momentum would be zero. This imbalance in the transverse momentum for an event is termed as the *missing transverse momentum* or the *missing transverse energy ($\boldsymbol{E}_T^{miss}$).* The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as follows [67]:

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = (E_{T,x}^{miss}, E_{T,y}^{miss}) E_T^{miss} = \sqrt{(E_{T,x}^{miss})^2 + (E_{T,y}^{miss})^2} \phi_{miss} = arctan(\frac{E_{T,y}^{miss}}{E_{T,x}^{miss}}) \quad (4.4)$$

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is a function of all the observed particles, as expressed in Equation 4.5.

$$E_{T,x(y)}^{miss} = E_{T,x(y)}^{miss,e} + E_{T,x(y)}^{miss,\gamma} + E_{T,x(y)}^{miss,\tau} + E_{T,x(y)}^{miss,jets} + E_{T,x(y)}^{miss,\mu} + E_{T,x(y)}^{miss,soft} \tag{4.5}$$

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is therefore sensitive of double counting due to overlaps between different objects, visible particles escaping due to gaps in the detector, poor reconstruction of objects, inclusion of unassociated pile-up particles. $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ has to be carefully reconstructed and calibrated so that these effects are accounted for and minimized [68]. The "soft term" ($E_{T,x(y)}^{miss,soft}$) is calculated using the ID tracks [69, 70] to include energy from the primary vertex which is not used in any recontructed object, making the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ more pile-up robust.

Chapter 5

CTIDE NN Performance at Reduced Charge Information

The cross-section for very high p_{T}-jets is significantly large at the current energy scale of $\sqrt{s} = 13$ TeV (see Figure 5.1). Due to high Lorentz boost, these high p_{T}-jets are collimated to the extent that they often contain tracks, which are at a smaller separation to each other than the dimensions of the ATLAS silicon pixel detectors (see Figure 5.2). In such cases, the pixel detectors are not able to resolve those tracks, present inside the high p_{T} jets. This leads to merged clusters, degrading the ATLAS tracking efficiency.

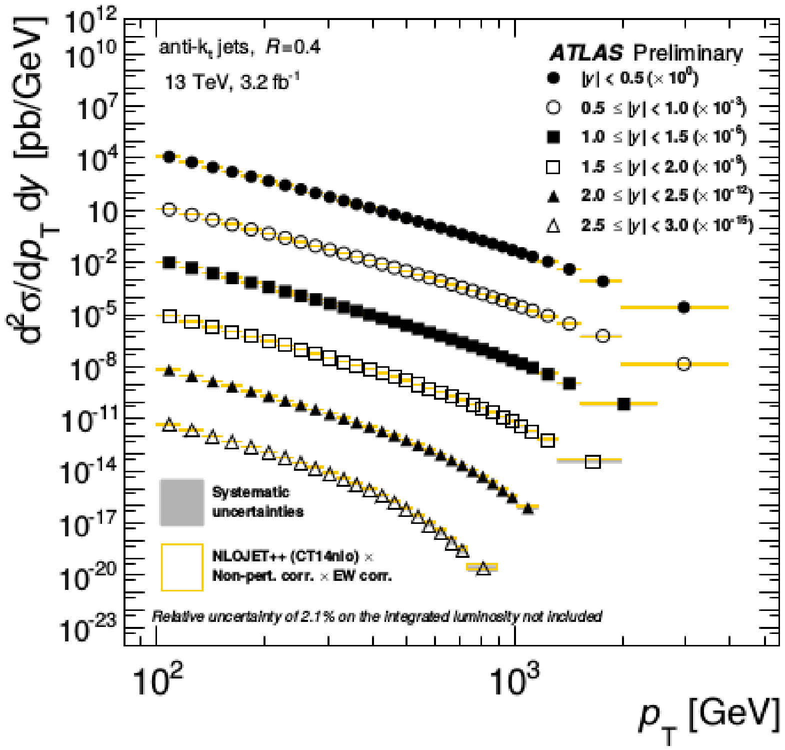


Figure 5.1: Differential inclusive jet cross-section as a function of jet p_{T} (2015 data) [14]. High p_{T} jets have significantly large cross-section at the LHC energy scale.

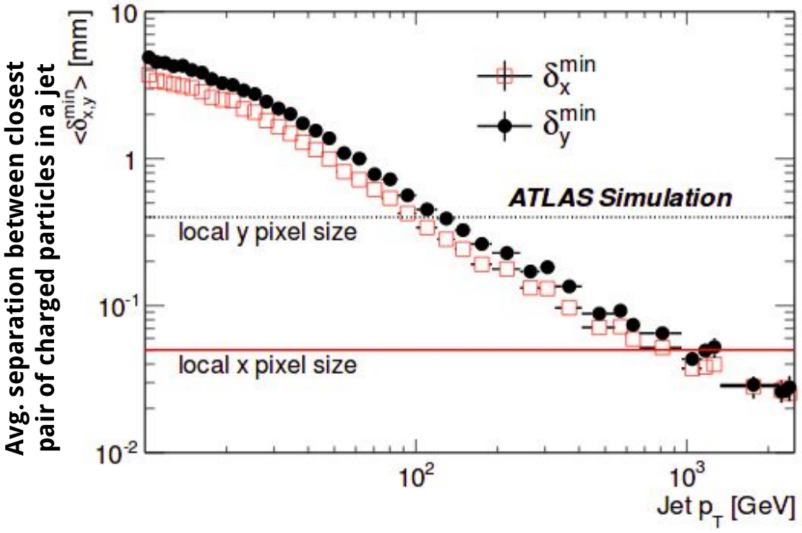


Figure 5.2: Average minimum track separation with respect to jet p_{T} [15]. Higher p_{T} jets can have track separations less than the ATLAS silicon pixel dimensions, leading to merged clusters.

ATLAS employs an ambiguity resolver to resolve these merged clusters. Among other tools, the ATLAS ambiguity resolver also contains a set of artificial neural networks (NNs) which are called the *clustering and tracking in dense environment neural networks* (CTIDE NNs). These networks use all the available low-level information about these merged clusters, such as the charge deposit in every pixel of a cluster, incident angles of tracks on a cluster, etc., to resolve the tracks in these clusters. The overall ATLAS tracking and the CTIDE NNs in the ATLAS ambiguity resolver have been described in Section 5.1 respectively.

At the high luminosity-LHC (HL-LHC), track multiplicity is expected to increase due to high pile-up (see Table 5.1). This would require higher transmission rate of hit information. The Inner Tracker (ITK) is an all-silicon replacement of the current inner detector (see Section 3.2.2) for HL-LHC, and is comprised of pixel and strip detectors. The ITK hardware has bandwidth limitations, hence the storage of charge information i.e. time over threshold (ToT), has to be reduced.

In this chapter, the effects of reducing the charge information on the performance

Table 5.1: Pile-up at the LHC and the HL-LHC (assuming 50ns bunch crossing) [19]. Higher pile-up at the HL-LHC would contribute to more merged clusters.

	$\mathcal{L}_{instantaneous}$	$\frac{avg.\ no.\ of\ interactions}{bunch\ crossing}$ $(<\mu>)$
LHC	$1.5 \times 10^{34} cm^{-2} s^{-1}$	60
HL-LHC	$5 \times 10^{34} cm^{-2} s^{-1}$	140

of ambiguity resolver have been investigated. This work was solely conducted by the author, with some help from a summer student in hyperparameter optimization in Section 5.6 Additionally, the optimum charge information content, which would not deteriorate the tracking performance significantly in the ITK, has been explored.

A straightforward approach would be to study the difference in performance of the CTIDE NNs, for merged clusters from the ITK versus the ATLAS inner detector. However, as the ITK is still in its design phase, neither actual data, nor simulation of the ITK merged clusters is accessible.

Therefore, for this study, the ITK merged clusters were indirectly simulated by reducing the charge information content of merged clusters from the inner detector. To accomplish this reduction in charge, a conversion scheme, developed in the Section 5.4.1, was used. In order to reduce the charge information, the entire range of 8-bit charge values allowed in the inner detector was needed. But, the upper limit of the charge range could not be clearly determined. So, 8-bit ToT was used instead of charge, where the allowed values were just consecutive integers from 0 through 256, in units of 25 ns (1 bunch-crossing)[1]. The conversion scheme was developed and optimized from Section 5.4 through Section 5.6 for ToT. Finally this conversion scheme was also used for charge to study the degradation in performance of the CTIDE NNs as a function of the charge information in merged clusters, as discussed in Section 5.8.

---

[1] IBL which stores in 4-bit also has a time granularity in ToT of 25 ns (1 bunch-crossing), so its ToT range is shorter than 8-bit pixels. Therefore, for 4-bit and 8-bit pixels the allowed charge ranges are about 1.5 and 8.5 times [71] the average charge deposited by a normally incidented minimum ionizing particle (MIP) on these pixels, respectively.

The first step in the ATLAS tracking is to group nearby charged silicon pixels, throughout a tracking pixel detector layer, into 7 × 7 pixel clusters. Then the charge center is linearly interpolated using the charge distribution in a cluster. Any track fit using three such positions (or seeds) and passing the p_{T} and impact parameter cuts is defined as an initial track candidate. A combinatorial Kalman filter [61] uses these initial track candidates and recursively updates their tracking parameters by including more seeds from other detector layers. When more than one track passes through a charged cluster, it is considered to be a merged (or shared) cluster candidate for which the ambiguity resolver is invoked [72]. This step first calculates metric to score the track candidates in the merged cluster, such as number of associated seeds, etc., in order to discard some of the ghost track candidates. But in dense environment, there are genuine merged clusters. In order to resolve these merged clusters, ambiguity resolver has a set of NNs, the CTIDE neural networks. These steps have been schematically depicted in Figure 5.3. The first network is a number neural network, which estimates the track multiplicity of the merged cluster. Then is a position network which estimates the positions (with respect to the center of the cluster) of the tracks in a merged cluster. Finally, there is an error network, which estimates the error in the position of the tracks estimated by the position network. The CTIDE NNs have been depicted schematically in Figure 5.4.

5.1 CTIDE Neural Networks

A general artificial neural network architecture looks like Figure 5.5. ANNs can be used for both classification and regression problems (see Ref. [73]).

In the ATLAS ambiguity resolver (see Figure 5.4), each neural network [74] contains two hidden layers. A stochastic gradient descent method with backpropagation is used to train the weights. A patience-based early-stopping strategy is used in

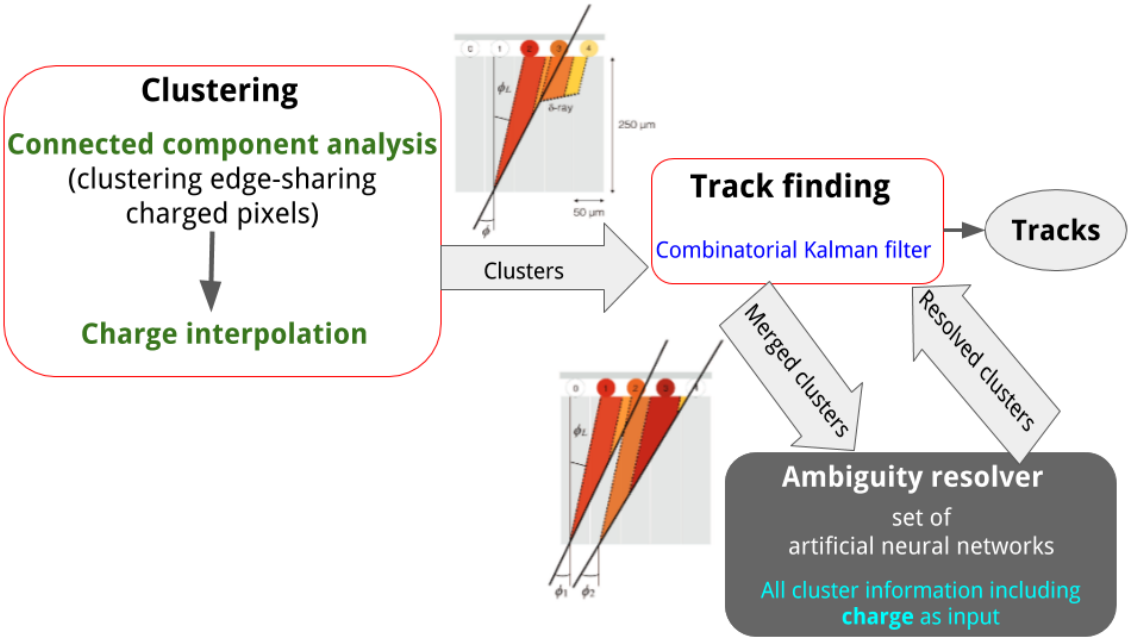


Figure 5.3: Schematic overview of the ATLAS tracking chain.

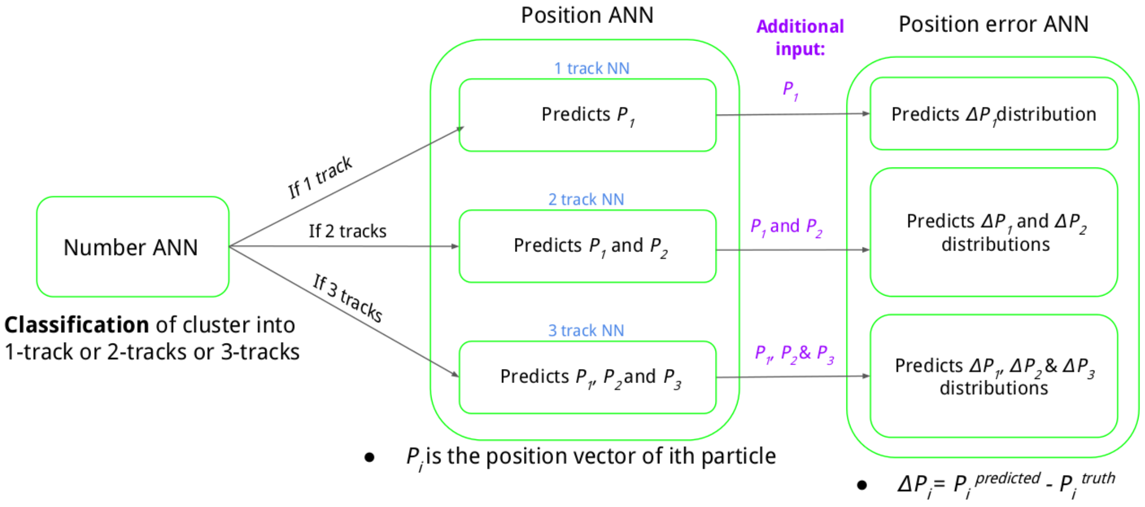


Figure 5.4: Schematic diagram of the CTIDE neural network set in the ambiguity resolver.

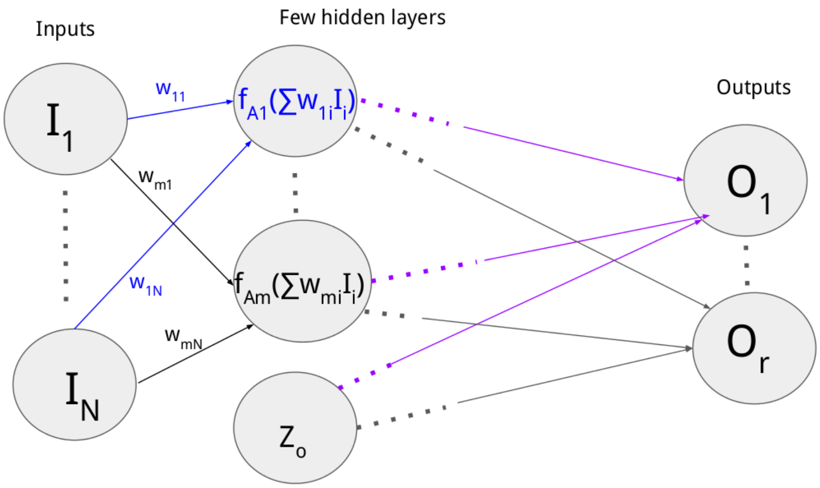


Figure 5.5: General architecture of an artificial neural network. The information in the nodes from the previous layers are added linearly added using a set of weights for a given node (depicted here with the arrows) and acted upon a non-linear activation function. This makes the final output nodes a very non-linear function of the input layers. The error in predicting the true output is propagated back, in every epoch, to the weights in the network until this error is minimized.

the training to avoid overfitting. The number of remaining epochs is increased by 1.75 for every epoch that sees a loss decrease of at least 0.5% for epochs above the 50th one. The required validation loss is computed on 10% of the training set. The hyperparameters used are listed in Table 5.2.

Table 5.2: Hyperparameters used to train the three sets of neural networks. In the Structure row, the numbers in parentheses denote the input and output layer sizes while the numbers in-between represent the hidden layer sizes. The sigmoid function used used here is: $1/(1+e^{-2x})$.

Hyperparameter	Multiplicity network	Position networks	Uncertainty networks
Structure	(60)-25-20-(3)	(60)-40-20-(2/4/6)	(62/64/66)-15-10-(30/50/60)
Hidden activation	Sigmoid	Sigmoid	Sigmoid
Output activation	Sigmoid	Identity	Sigmoid
Learning rate	0.08	0.04	0.3
Weight decay	10^{-7}	10^{-7}	10^{-6}
Momentum	0.4	0.3	0.7
Minibatch size	60	30	50
Loss function	cross-entropy	mean squared error	cross-entropy

The inputs used in these neural networks are:

- A 7×7 matrix of deposited charges in the cluster,
- A 7−dimensional vector of pixel sizes in the longitudinal direction, to identify long pixels,
- The layer number (from 0 to 3) and layer type (barrel or endcap),
- The polar angle (θ) and the azimuthal angle (ϕ) of the track candidates (CTIDE NNs are robust with respect to these angular variables [75], and have a significant improvement in performance by their inclusion [15]).

For error network, the position estimates from the position NN are also inputted. A dijet Monte Carlo sample generated with PYTHIA 8.186 [76] using the A14 set of tuned parameters [77] and the NNPDF2.3LO parton distribution function set [78] is employed to produce the training and validation sets. A filter that keeps only jets with transverse momentum between 1.8 and 2.5 TeV is applied, resulting in a high fraction of multi-particle clusters.

For each neural network training, 1 million clusters each are retained for the training and validation set. For the network used to estimate the particle multiplicity, the dataset is adjusted to contain 22%, 26% and 52% of 1, 2 and $\geqslant$3 particle clusters, respectively. [2] For the position and error neural networks, separate training are carried out for 1, 2 and 3x particle clusters ($>$ 3 particle clusters are not further considered) and each training dataset consists solely of clusters from the corresponding multiplicity class.

5.2 Incident Angle Distributions

Incident angles θ and ϕ are two of the input variables used for the CTIDE networks in resolving merged clusters. These angles are calculated locally for clusters. The

---

[2] This fraction was tuned to help the network to properly learn to recognize 3-particle clusters

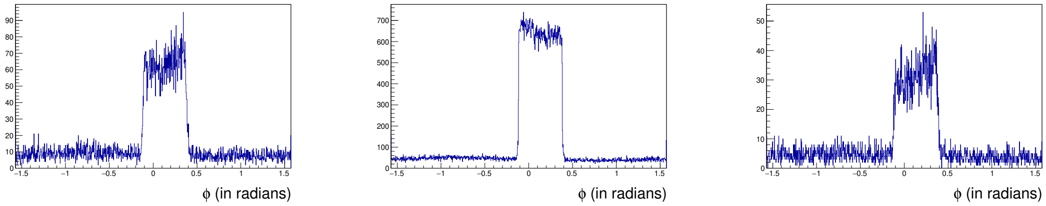
cluster feature might have some dependence of the global η as well. There are 13 η modules in the ATLAS barrel and IBL detectors. Since the global η is not directly inputted in the CTIDE networks, the distributions of θ and ϕ coordinates for 1-, 2- and $\geqslant 3$ clusters at all the barrel layers for each of the η-modules were studied. This helped the dependence of merged cluster profiles on the global η, and if so, the way is it encoded in the local variables such as θ and ϕ over which the CTIDE networks are to be trained.

5.2.1 Azimuthal Angle Distribution

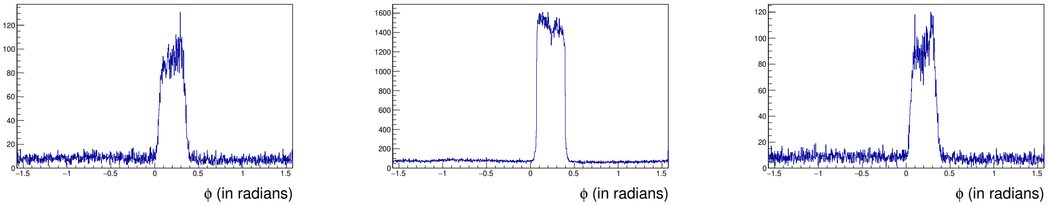
The azimuthal angle ϕ is measured with respect to the normal to the cluster. The ϕ distributions of the merged clusters are uniform only within a range of about [-0.1, 0.4] (asymmetric about 0). The mean positive ϕ is because of a small tilt in the staves along the ϕ direction. This range gets even shorter going radially out, due to the curvature of the charged tracks in the cluster. A flat background outside this range is contributed by delta rays and secondary tracks. Some of the distributions are given in Figure 5.6 (see the distributions for each particle-multiplicity class in [79]):

5.2.2 Polar Angle Distribution

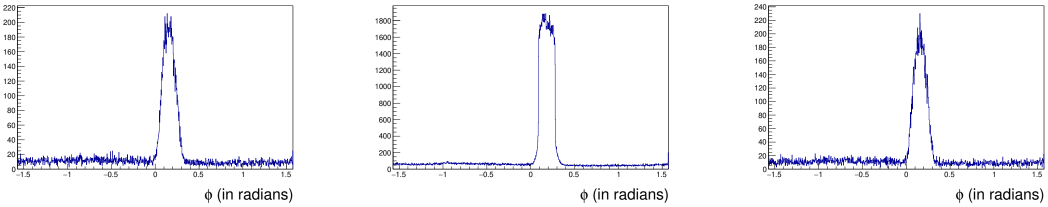
The longitudinal angle θ is measured also measured with respect to the normal to the cluster. The θ distribution, as expected was directly dependent on η, so it was skewed at $-\frac{\pi}{2}$ in the right-most η module (module -6), at 0 in the central η module (module 0) and at $+\frac{\pi}{2}$ in the left-most module (module 6). This ensured that dependence of merged cluster features on global η is well relayed by the local θ coordinate. A few distributions are given in Figure 5.7 (see the distributions for each particle-multiplicity class in [79]):



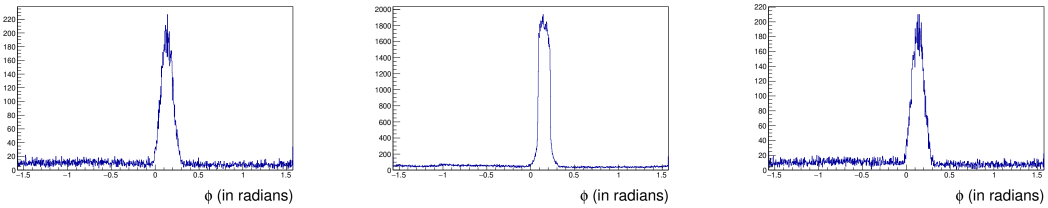
(a) IBL η-module -5 (b) IBL η-module 0 (c) IBL η-module 5



(d) Barrel layer 1 η-module -5 (e) Barrel layer 1 η-module 0 (f) Barrel layer 1 η-module 5



(g) Barrel layer 2 η-module -5 (h) Barrel layer 2 η-module 0 (i) Barrel layer 2 η-module 5

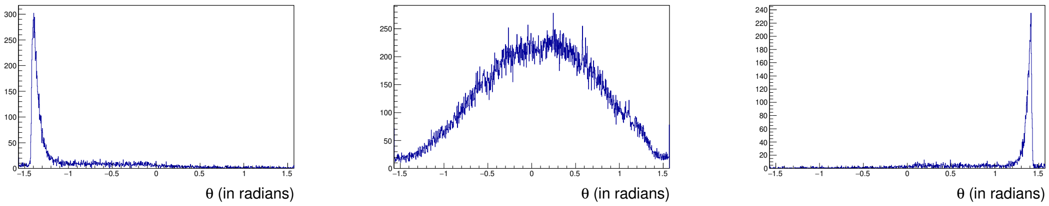


(j) Barrel layer 3 η-module -5 (k) Barrel layer 3 η-module 0 (l) Barrel layer 3 η-module 5

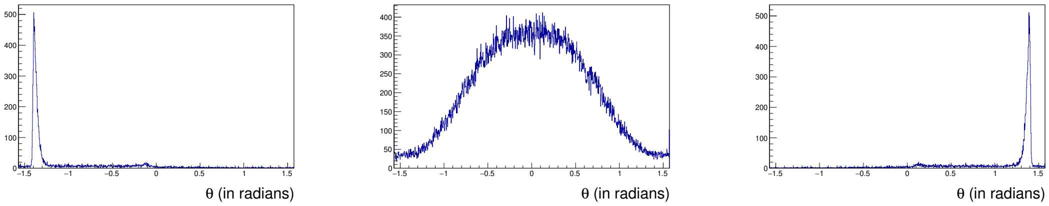
Figure 5.6: ϕ distribution for 2-particle clusters. The leftmost column is for η-module -5, the central column is for η-module 0 and the rightmost column is for η-module +5. The topmost row is for IBL, followed by Barrel layer 1, then Barrel layer 2 and the last row for Barrel layer 3.

5.3 Distribution of Charge in 7×7 Cluster Pixels

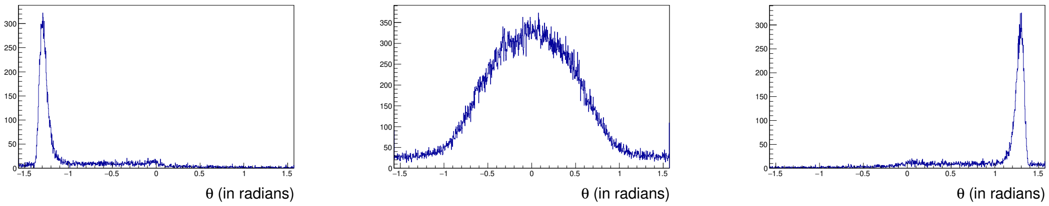
The distribution of charge in the merged cluster pixels depend on the incident angles of the track(s). As discussed in Section 5.2, the incident angles have dependence on



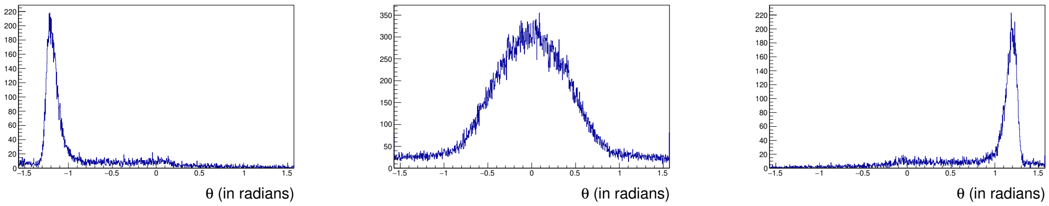
(a) IBL η-module -5 (b) IBL η-module 0 (c) IBL η-module 5



(d) Barrel layer 1 η-module -5 (e) Barrel layer 1 η-module 0 (f) Barrel layer 1 η-module 5



(g) Barrel layer 2 η-module -5 (h) Barrel layer 2 η-module 0 (i) Barrel layer 2 η-module 5



(j) Barrel layer 3 η-module -5 (k) Barrel layer 3 η-module 0 (l) Barrel layer 3 η-module 5

Figure 5.7: θ distribution for 2-particle clusters. The leftmost column is for η-module -5, the central column is at η-module 0 and the rightmost column is at η-module +5. The topmost row is for IBL, followed by Barrel layer 1, then Barrel layer 2 and the last row is Barrel layer 3. The local coordinate θ is proportional to η.

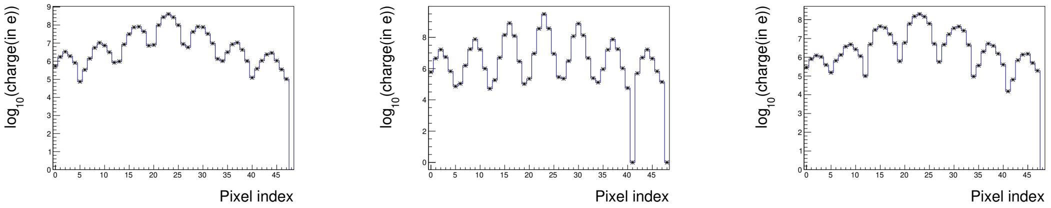
η. Therefore, the η dependence of the charge distribution in the cluster pixels were studied for all the particle multiplicity classes.

Each pixel in the 2-dimensional 7×7 array in a cluster is enumerated with an integer index from [0,48]. Naturally such mapping of 2-D coordinates into 1-D index have non-unique options. Since, actual mapping used to convert the 2-D pixel coordinates into 1-D index is not specified to the CTIDE networks, it can be expected that the networks are somehow capable of learning the scheme being used to get the pixel indices. The charge distribution for 1-D pixel indices were plotted, for each η module at every layer for all the particle multiplicity class. This was to see if the scheme used in mapping the 2-D pixel coordinates to 1-D index is apparent from these aforementioned charge distributions. Some of the distributions are given in Figure 5.8 (see all the distributions in [79]):

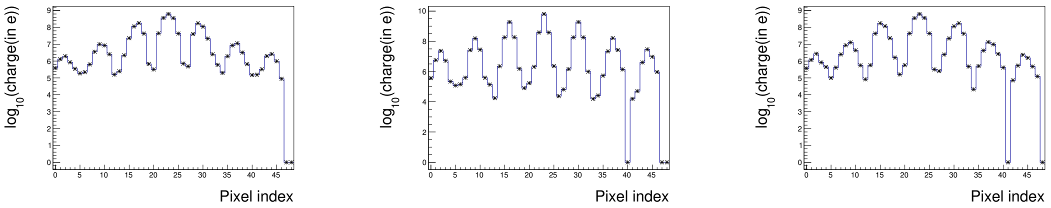
By examining the distributions along pixel indices above, the scheme used to map 2-D coordinates to 1-D pixel could be understood. One can think of the enumeration scheme as lining up the rows of the 7×7 pixels along the azimuthal direction in one long array and then assigning them indices from 0 through 48. Some of the distributions are given in Figure 5.9 (see all the distributions in [79]):

5.4 Clusters with Reduced Charge Information

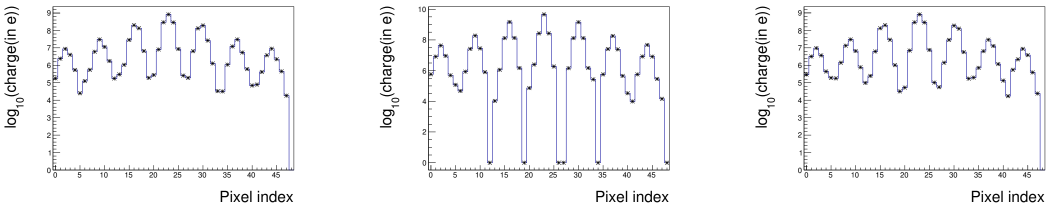
Since the actual charge storage in ITK pixels is still under discussion, a conversion scheme to reduce the charge content of the 8-bit ToT inner detector Si-pixels to any arbitrary low charge content was developed, which could then emulate the ITK merged clusters. Since the barrel and endcap layers in the inner detector store 8-bit charge information, the per-pixel charge distribution was expected to have sharp peaks around 256 distinct charge values. Although lower charge peaks were identified, the charge peaks at higher values could not be well resolved. This could perhaps be attributed to the difference in ToT to charge calibration among the pixels (see Figure 5.10).



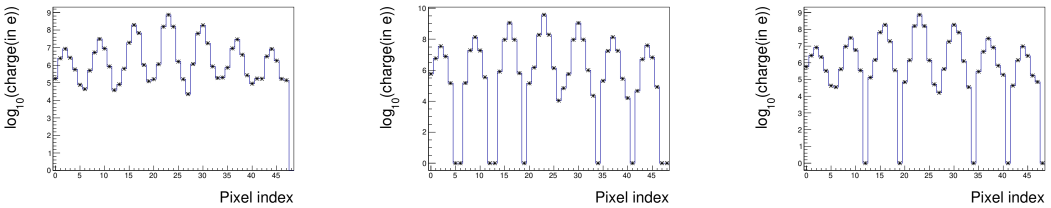
(a) IBL η-module -5 (b) IBL η-module 0 (c) IBL η-module 5



(d) Barrel layer 1 η-module -5 (e) Barrel layer 1 η-module 0 (f) Barrel layer 1 η-module 5

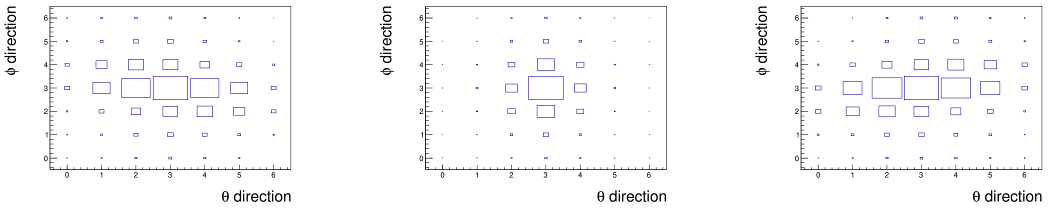


(g) Barrel layer 2 η-module -5 (h) Barrel layer 2 η-module 0 (i) Barrel layer 2 η-module 5

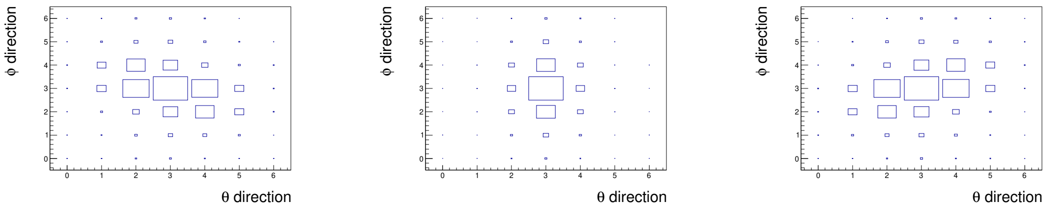


(j) Barrel layer 3 η-module -5 (k) Barrel layer 3 η-module 0 (l) Barrel layer 3 η-module 5

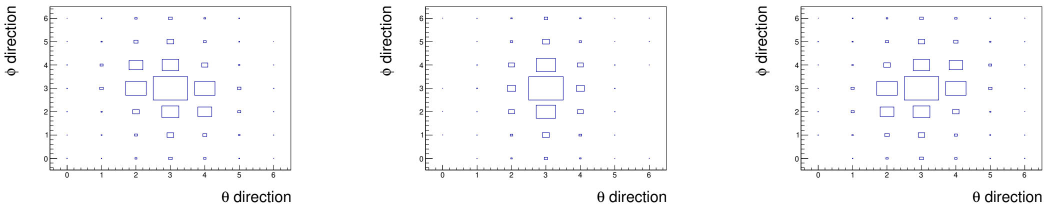
Figure 5.8: Charge distribution over pixel index for 2-particle clusters. The leftmost column is for η-module -5, the central column is at η-module 0 and the rightmost column is at η-module +5. The topmost row is for IBL, followed by Barrel layer 1, then Barrel layer 2 and the last row is Barrel layer 3.



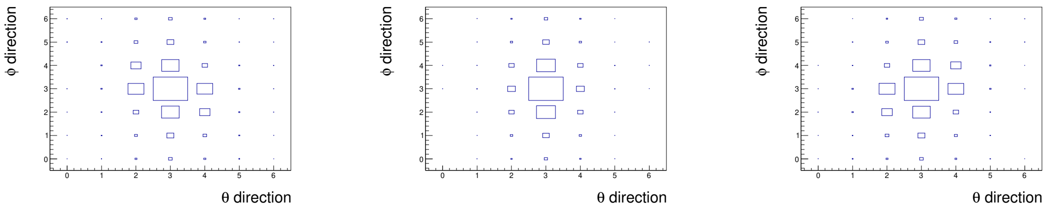
(a) IBL η-module -5 (b) IBL η-module 0 (c) IBL η-module 5



(d) Barrel layer 1 η-module -5 (e) Barrel layer 1 η-module 0 (f) Barrel layer 1 η-module 5

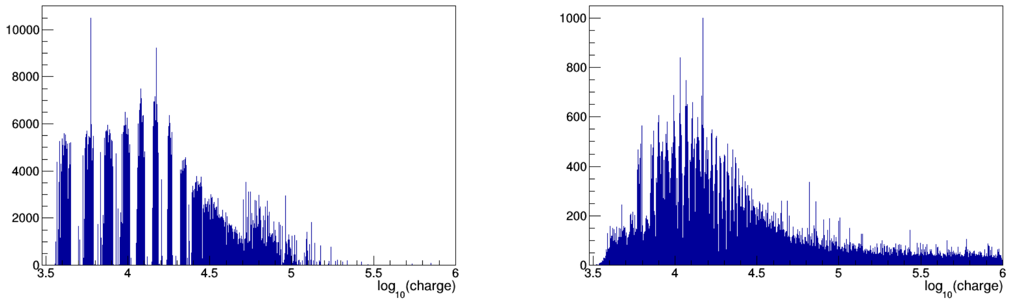


(g) Barrel layer 2 η-module -5 (h) Barrel layer 2 η-module 0 (i) Barrel layer 2 η-module 5



(j) Barrel layer 3 η-module -5 (k) Barrel layer 3 η-module 0 (l) Barrel layer 3 η-module 5

Figure 5.9: Charge distribution over 2-dimensional 7 × 7 pixel array in a cluster (reconstructed using figure 5.8) for 2-particle clusters. The relative size of the boxes represent the total deposited pixel charge (in e) in $\log_{10}$scale. The leftmost column is for η-module -5, the central column is at η-module 0 and the rightmost column is at η-module +5. The topmost row is for IBL, followed by Barrel layer 1, then Barrel layer 2 and the last row is Barrel layer 3.



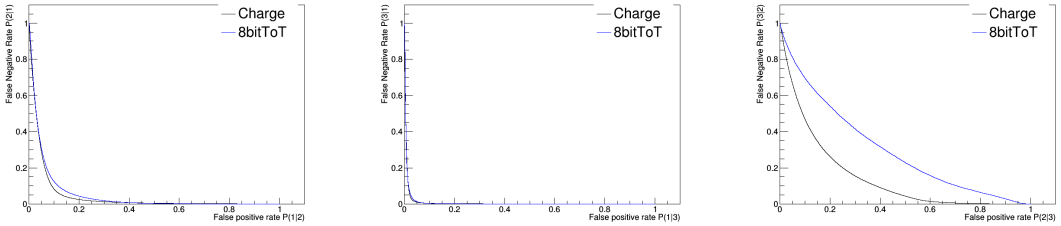
(a) IBL charge

(b) Barrel charge

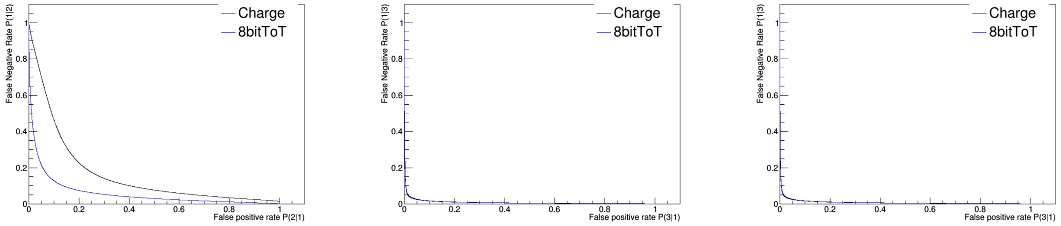
Figure 5.10: Per-pixel charge distribution. Here, (a) shows 4-bit IBL charge distribution and (b) shows 8-bit barrel (all 3 barrel layers included) charge distribution. Merged clusters from all three particle multiplicity classes have been included. The higher charge peaks are not distinctly visible. As the distributions here are inclusive of the entire barrel region, due to the difference in ToT to charge calibration at different parts of the barrel, the charge peaks at higher values merge with each other and therefore unresolvable.

Since, the higher values of the available 8-bit charge range was unclear, the upper limit of the 8-bit charge was also ambiguous. In the case of ToT, however, the range is merely the integers in [0, 256]. Therefore, in order to develop a conversion scheme for charge information and lay the course of this study, a ToT was used in lieu of charge as it was simpler to deal with. The charge inputs were replaced with the ToTs in the CTIDE NN, without altering the architecture of these networks. A quick check on its impact on the performance of the number neural network which estimates the track-multiplicity for the merged clusters, was made by comparing the ROC curves using both types of inputs - charge and ToT (see Figure 5.11).

As expected, the ToT input networks could not out-perform the charge input networks, however the performance difference was generally small. With this caveat, it was decided to proceed with the ToT studies for developing a conversion scheme (see Figure 5.12).

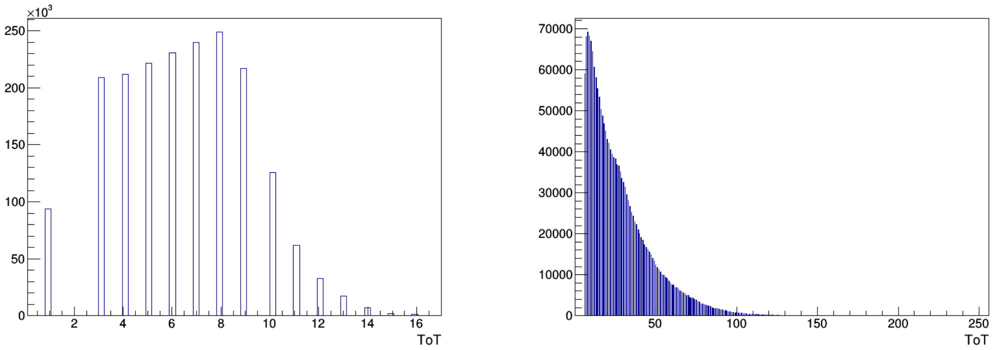


(a) (Barrel) 1vs2 (b) (Barrel) 1vs3 (c) (Barrel) 2vs3



(d) (Barrel) 2vs1 (e) (Barrel) 3vs1 (f) (Barrel) 3vs2

Figure 5.11: The comparison in performance of the number network between pixel charges and pixel ToTs as input. Since the output number network are the three probabilities for the particle multiplicity classes (1-, 2- and ⩾3-particle clusters), there are 6 ROC curves. To compare the performance the ROC curves in the case of 8-bit ToT and charge inputs are overlaid.



(a) IBL ToT (b) Barrel ToT

Figure 5.12: Per-pixel ToT distribution. Here, (a) shows 4-bit IBL per-pixel ToT distribution for 1 particle clusters and (b) shows 8-bit barrel (all 3 barrel layers included) ToT distribution for 2 particle clusters. The (4-bit) IBL ToT covers the entire range through 16, while the (8-bit) Barrel ToT distribution is contained within 120 (less than half the available range).

5.4.1 Conversion Scheme

Among the detector layers of the ATLAS inner detector, only insertable B-layer (IBL) stores 4-bit ToT for charge information. The rest of the layers store 8-bit ToT for charge information. Since, the aim of the conversion scheme was to lower the charge information from existing 8-bit for the same detector, and the IBL was already having a lower charge information, the IBL per-pixel ToT was studied (see Figure 5.12a). It was compared to the per-pixel ToT distribution of the barrel layers (see Figure 5.12b). It was observed that, while for IBL the entire available ToT range was being used, for the barrel layer after a higher ToT value the rest of the available range was unused. It was realized that, since there is anyway a loss of information in conversion of 8-bit ToT to lower n-bit ToT, for maximum information retention there should not be any unused value in the available lower n-bit ToT range. Also, if only the used part of the 8-bit ToT range is used during conversion, it would reduce the resolution loss of information.

Guided by these points, the conversion scheme was developed, where first the 8-bit ToT range was saturated at a given value followed by a mapping of the abridged range (i.e., [0, saturation value]) with 8-bit ToT resolution linearly to lower n-bit ToT range [0, 2^n].

5.4.2 Choice of Saturation ToT

As Figure 5.12b shows, the higher values of the available 8-bit ToT range of the ATLAS inner detector do not contain any information. Therefore, introduction of a saturation ToT (discussed in Section 5.4.1) in the conversion scheme gives a flexibility to use the effective ToT range, which contains all the relevant charge information with 8-bit resolution. If by choosing a large enough saturation ToT, the unused range (or part of it) is included along with the effective range and gets mapped to lower n-bit ToT values, it adds to the resolution loss in the conversion process. The performance of CTIDE networks (and other discriminators) is then sub-optimal for that lower n-bit charge information. On the other hand, a smaller saturation ToT would lose some of the information from the effective 8-bit range. This should would also lead to a sub-optimal performance for the lower n-bit charge content. So, the choice of saturation ToT, which include only includes the effective 8-bit range, would lead to optimal performance for that lower n-bit charge information, and therefore have the least drop in performance from the original 8-bit charge information.

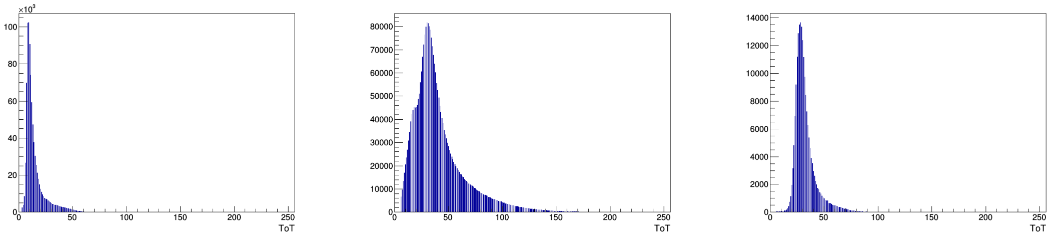
Thus, the performance of CTIDE networks or other discriminators, for a given lower n-bit charge information, can be optimized by tuning the saturation ToT. The general strategy, in this note, has been to choose a saturation which optimizes the performance of a discriminator (except in Section 5.5.4, where the saturation has been tuned to match the performance for modified 4-bit barrel clusters with IBL clusters) for the lower n-bit charge information.

5.5 Cluster ToT as Classifier

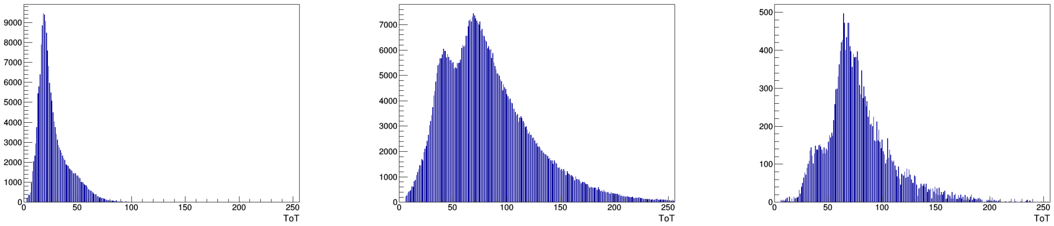
Cluster ToT is defined as the sum of all the ToTs from 7 × 7 pixels in a cluster. It is analogous to the total charge deposited in the cluster and therefore proportional to the track multiplicity. Cluster ToT can be used a classifier between different particle multiplicity classes. It should be noted that cluster ToT does not encode the information about the 2-dimensional spatial moments of charge distribution in the clusters. For that the discriminators produced by the CTIDE networks, by taking into account the entire 2-D spatial profile of the charge distribution, has to be used.

5.5.1 Difference in Barrel Layer Thresholds

When cluster ToT from all the three barrel layers are included, the distribution has more than one peak (see figures 5.13b and 5.13e).



(a) IBL ToT (1 particle) (b) Barrel ToT (1 particle) (c) Endcap ToT (1 particle)



(d) IBL ToT (2 particle) (e) Barrel ToT (2 particle) (f) Endcap ToT (2 particle)

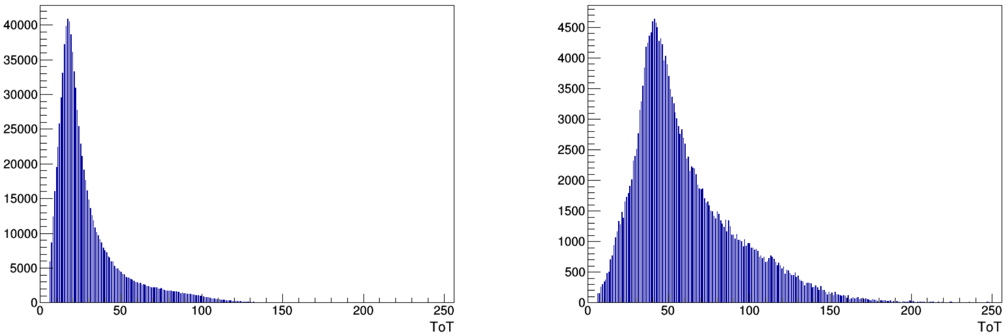
Figure 5.13: Combined cluster ToT distribution for 1-particle and 2-particle clusters. For the barrel region, (b) for 1-particle and (e) for 2-partice clusters have more than one peak.

Table 5.3: Cluster ToT distribution highest peaks for pixel detectors. For barrel, the highest peaks have been listed.

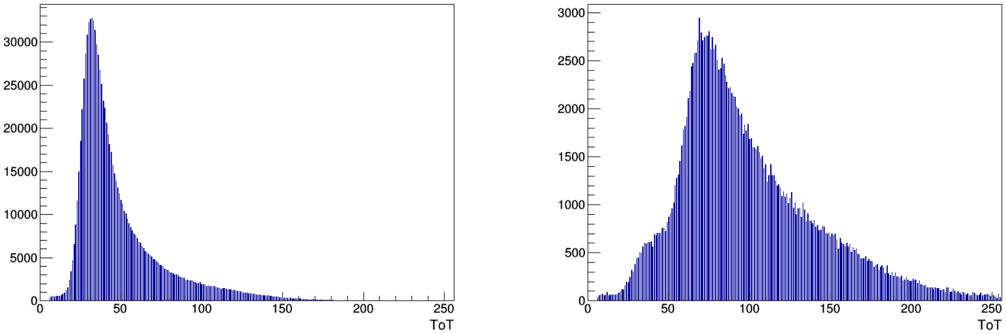
Pixel detector	IBL	Barrel	Endcap
1-particle cluster	7	30	28
2-particle cluster	18	70	65

However, the cluster charge distribution for barrel (all the layers combined) has only one peak (see figure 5.16).

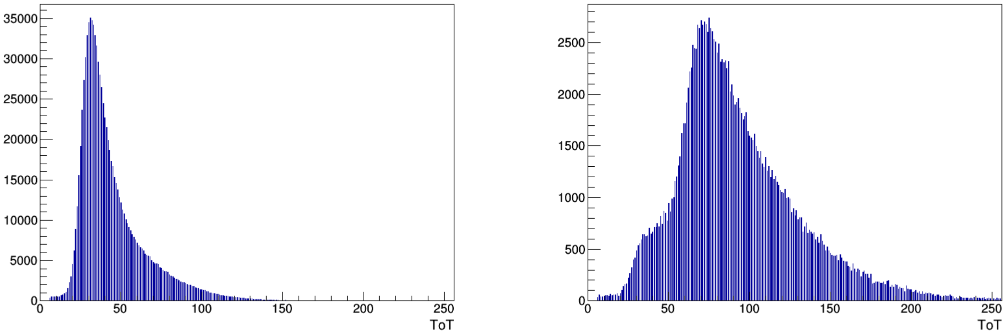
The cluster ToT distribution of each barrel layer has single peak (see Figure 5.14 and Table 5.4). The cluster ToT peaks of barrel L2 and L3 are nearly the same, which make up the higher peaks for combined barrel layer cluster ToT distributions, as listed



(a) Barrel L1 ToT (1 particle) (b) Barrel L1 ToT (2 particle)



(c) Barrel L2 ToT (1 particle) (d) Barrel L2 ToT (2 particle)



(e) Barrel L3 ToT (1 particle) (f) Barrel L3 ToT (2 particle)

Figure 5.14: Cluster ToT distribution peaks for barrel layers.

Table 5.4: cluster ToT distribution peaks for each barrel layer.

Barrel layer	L1	L2	L3
1-particle cluster	16	32	32
2-particle cluster	41	70	74

in table 5.3. This revealed that the barrel pixel layers have different thresholds (L1 threshold > L2 threshold = L3 threshold) and that this difference was taken into consideration while converting ToT to charge. Hence, cluster charge from all the three barrel layers combined, has only one peak.

5.5.2 Si Path Length Correction

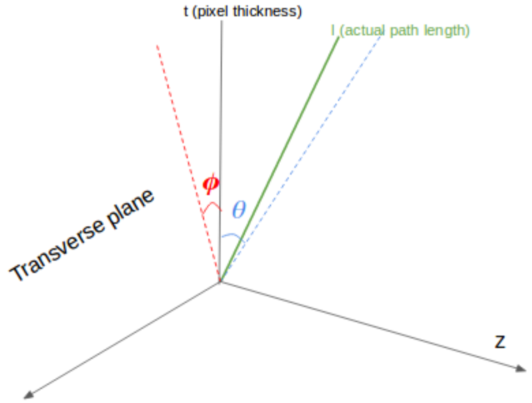


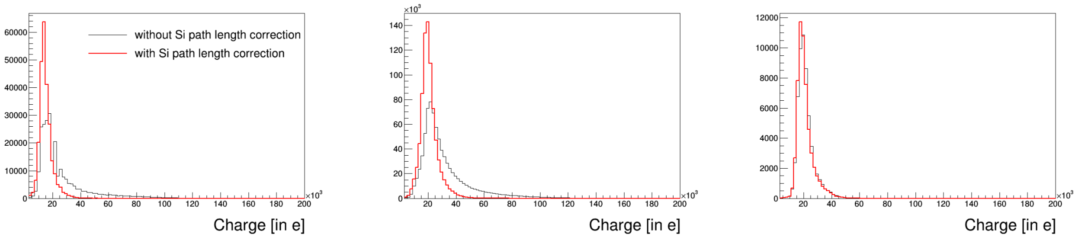
Figure 5.15: The length of the actual path traversed by a track is: $l = t \times \sqrt{1 + tan^2\theta + tan^2\phi}$.

To correct for normal incidence (see Figure 5.15, cluster charges were divided by the factor $\sqrt{1 + tan^2\theta + tan^2\phi}$ as shown in Figure 5.16 and Table 5.5). To study the dependence of cluster charge on ϕ, at a fixed η-module 0, the cluster charge distribution before and after Si-path length correction (see Figure 5.17 and Table 5.6) were studied.

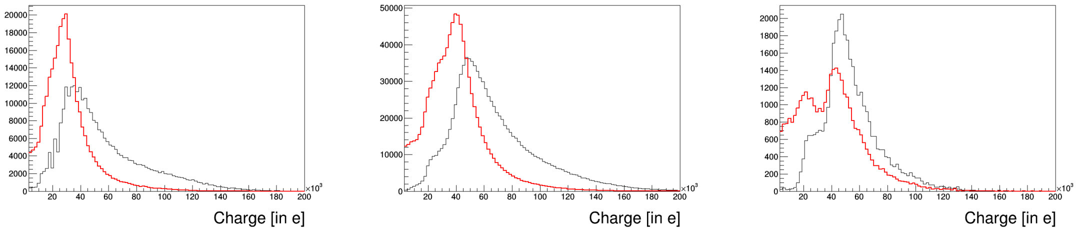
The Si path length correction made the cluster charge distribution comparable to dE/dx (see Figure 5.18).

Table 5.5: corrected cluster charge distribution peaks, before (left columns) and after (right columns) Si path length correction, for pixel detectors.

Pixel detector	IBL (in $\times 10^4$e)		Barrel (in $\times 10^4$e)		Endcap (in $\times 10^4$e)	
Si-path correction	No	Yes	No	Yes	No	Yes
1-particle cluster	1.7	1.4	2.2	2.0	2.0	1.8
2-particle cluster	3.5	2.8	4.8	4.0	4.8	4.4



(a) IBL Charge (1 particle) (b) Barrel Charge (1 particle) (c) Endcap Charge (1 particle)



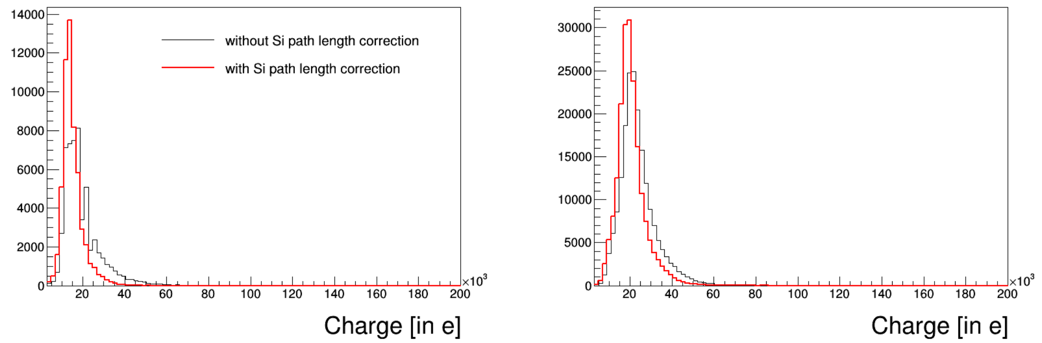
(d) IBL Charge (2 particle) (e) Barrel Charge (2 particle) (f) Endcap Charge (2 particle)

Figure 5.16: Combined cluster charge distribution (with and without Si path length corrected).

Table 5.6: corrected cluster charge distribution peaks, before (left columns) and after (right columns) Si path length correction, for IBL and Barrel at η module 0.

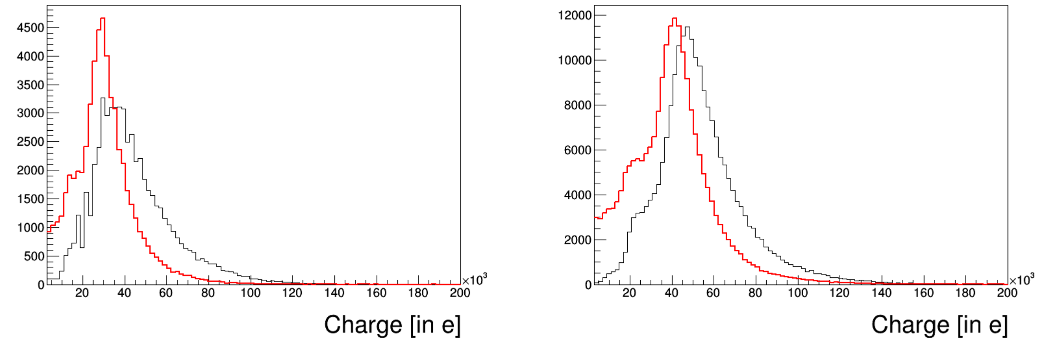
Pixel detector	IBL (in $\times 10^4$e)		Barrel (in $\times 10^4$e)	
Si-path correction	No	Yes	No	Yes
1-particle cluster	1.7	1.7	2.1	2.1
2-particle cluster	3.0	3.0	4.8	4.8

The cluster ToT distribution of 2-particle clusters peak at twice the value as 1-particle clusters (see Table 5.4). Which is why cluster ToT can be used as a discriminator between the three particle multiplicity classes. The discrimination performance of cluster ToT can be given by the difference in the cluster ToT distributions of dif-



(a) IBL (1 particle)

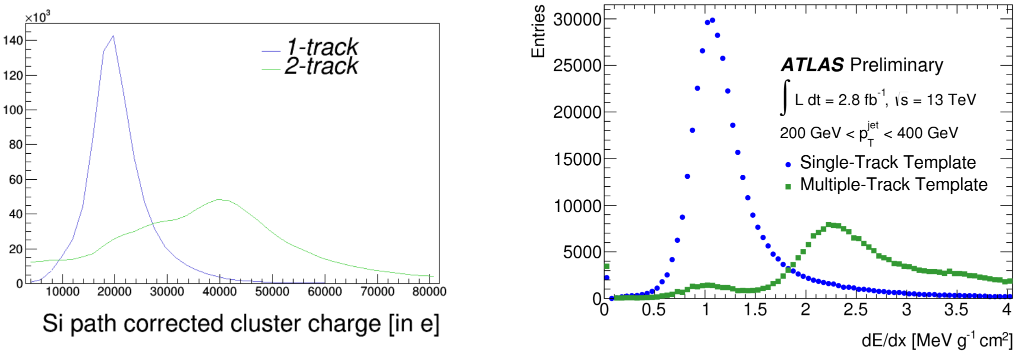
(b) Barrel Charge (1 particle)



(c) IBL Charge (2 particle)

(d) Barrel Charge (2 particle)

Figure 5.17: Cluster charge distribution at η module-0 for pixel detectors, before and after Si path length correction.



(a) 1p and 2p cluster charge in the barrel (Si path length corrected)

(b) 1p and 2p dE/dx distribution

Figure 5.18: Comparison of dE/dx distribution to cluster charge (Si path length corrected) distribution. Since the average θ and ϕ are used instead of individual track θ and ϕ in the correction factor, 2-track path length corrected cluster charge differs from the multi-track dE/dx template.

ferent particle multiplicity classes. Since, no particular shape was assumed to model these cluster ToT distributions, a non-parametric test called the Mann-Whitney U test was used to find the difference between the cluster ToT distributions of different particle multiplicity classes.

5.5.3 Mann-Whitney U Test

Mann-Whitney U test is a non-parametric test which determines how shifted are the respective distributions of two given samples. The steps to calculate the Mann-Whitney U test statistic are as follows:

- **step 1** All the elements of the two samples (containing n_1 and n_2 elements respectively) are placed in a combined set and each element of this set is assigned a global rank.

- **step 2** A rank of an element is defined as one added to the total number of elements less than the given element. Hence, in the case of ties, the elements are assigned an adjusted rank $= R_{tie} + \frac{(n_{tie}-1)}{2}$, where R_{tie} is the rank tied among n_{tie} elements.

- **step 3** The Mann-Whitney U test-statistic is defined as: $U_1 = \left(\frac{1}{n_1 \times n_2}\right)\left[R_1 - \frac{n_1(n_1+1)}{2}\right]$ for sample 1, where R_1 is the sum of the global ranks of all the elements in sample 1, such that $U_1 + U_2 = 1$. The statistic U shows how dispersed are the elements of a given sample in the combined set (defined in step 1).

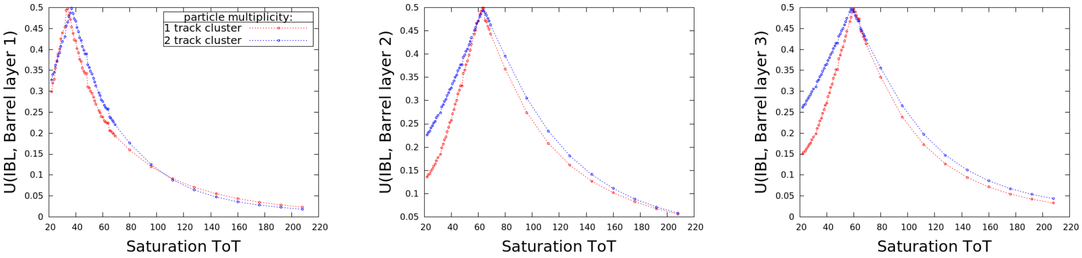
In other words, U gives a measure of the overlap (or separation in case of unimodal distributions) of one sample on (from) the other sample. The test is thus a classifier between the distributions of the two samples, with U being equal to the area under the ROC for classifying these two distributions. U is, therefore, defined to have a maximum value of 0.5 (which is the value between two identical distributions).

5.5.4 Emulating IBL Using Converted Barrel ToT

IBL is a existing real detector which registers 4-bit ToT information. Using conversion scheme, 8-bit barrel ToT clusters can be converted into 4-bit ones. The saturation ToT was tuned such that the cluster ToT of the modified barrel 4-bit clusters match that of the IBL clusters. This way those modified 4-bit barrel clusters could then emulate IBL clusters.

Table 5.7: ToT saturation at U(IBL,Barrel layer) peaks for 1-track and 2-track clusters.

ToT saturation	Barrel L1	Barrel L2	Barrel L3
1-track cluster peak	34(U=0.497)	63(U=0.498)	60(U=0.496)
floor(midpoint)	35	63	59
2-track cluster peak	37(U=0.498)	63(U=0.492)	58(U=0.496)



(a) Barrel layer 1 (b) Barrel layer 2 (c) Barrel layer 3

Figure 5.19: Modified 4-bit Barrel ToT comparison to IBL. Mann-Whitney U versus saturation point for 1-track and 2-track clusters. Leftmost column is layer 1 followed by layer 2 and rightmost column is layer 3.

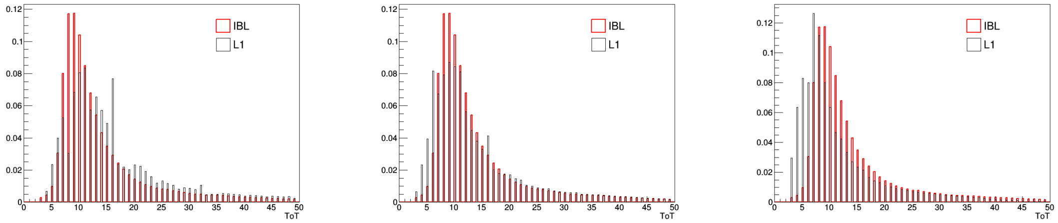
So, a saturation ToT range was scanned and the Mann-Whitney U score between IBL cluster ToT and converted 4-bit barrel layer cluster ToT (see Figure 5.19) for each of those saturation ToTs was computed. The U score was maximized to obtain modified 4-bit barrel ToT closely matching IBL ToT. These saturation ToTs for barrel layers have been summarised in the Table 5.7. ToT saturation of 1-particle and 2-particle clusters at U_{max} were very close, so the midpoints of the saturation at the two peaks could be used for optimal match between barrel layers and the IBL (see Figures 5.20, 5.21 and 5.22).

5.5.5 Optimizing U Score

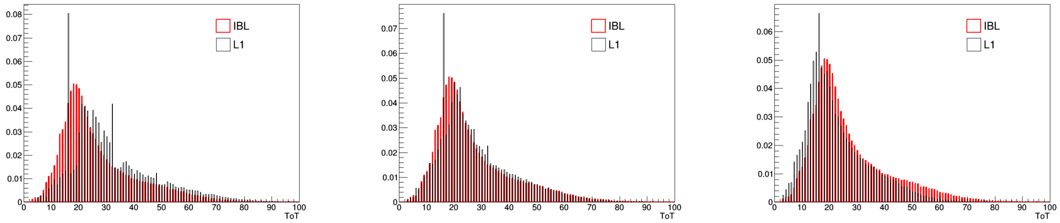
As a proof of concept (discussed in Section 5.4.2), the performance metric U of cluster ToT, for classifying 1-particle from 2-particle merged clusters, was optimized for 3-bit, 4-bit and 5-bit ToT with respect to the saturation ToT (see Figure 5.23).

The cluster ToT distributions for 1-particle and 2-particle clusters, at ToT saturations for optimum U has been compared to performance saturation ToTs with difference in performance by 5% error (see Figures 5.24-5.26 for 3-bit ToT, Figures 5.27-5.29 for 4-bit ToT and Figures 5.30-5.32 for 5-bit ToT).

The ToT saturations for optimum performance in classifying 1-particle and 2-

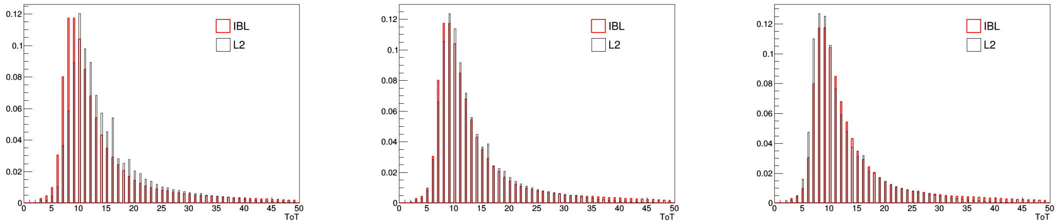


(a) ToT saturation=28 (1-particle)
(b) ToT saturation=35 (1-particle)
(c) ToT saturation=45 (1-particle)

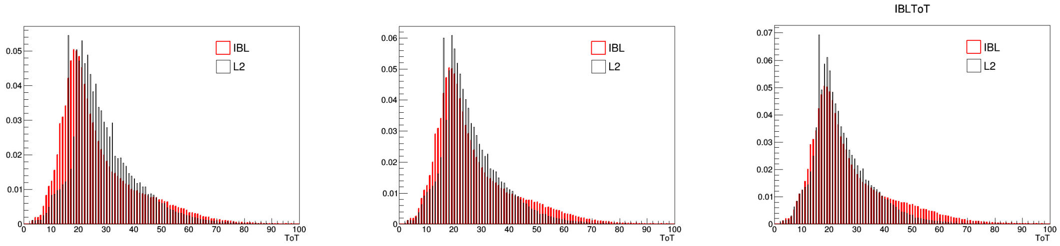


(d) ToT saturation=28 (2-particle)
(e) ToT saturation=35 (2-particle)
(f) ToT saturation=45 (2-particle)

Figure 5.20: Cluster ToT distribution overlay of Barrel layer 1 with IBL.

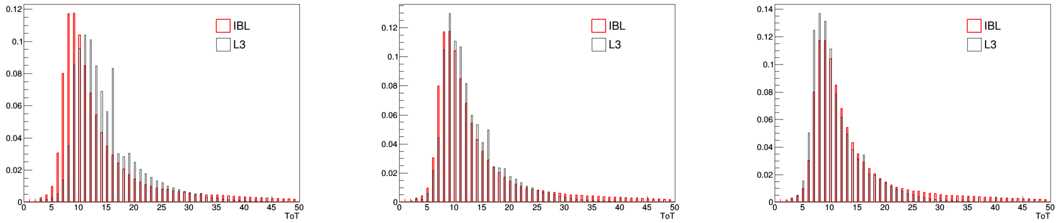


(a) ToT saturation=55 (1-particle)
(b) ToT saturation=63 (1-particle)
(c) ToT saturation=69 (1-particle)



(d) ToT saturation=55 (2-particle)
(e) ToT saturation=63 (2-particle)
(f) ToT saturation=69 (2-particle)

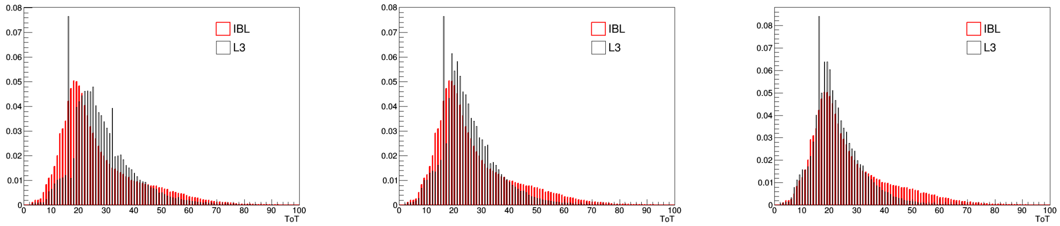
Figure 5.21: Cluster ToT distribution overlay of Barrel layer 2 with IBL.



(a) ToT saturation=50 (1-particle)

(b) ToT saturation=59 (1-particle)

(c) ToT saturation=69 (1-particle)

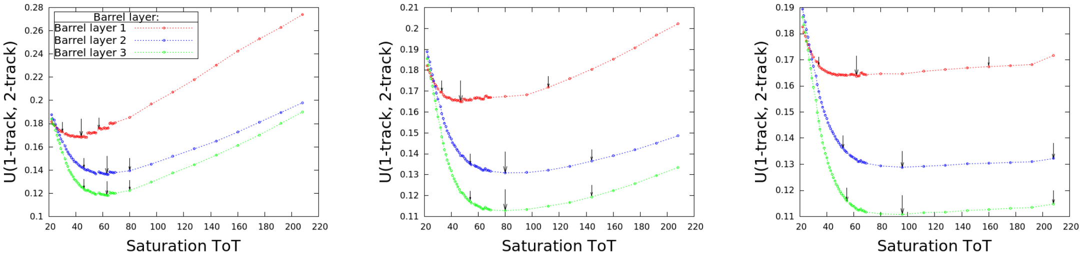


(d) ToT saturation=50 (2-particle)

(e) ToT saturation=59 (2-particle)

(f) ToT saturation=69 (2-particle)

Figure 5.22: Cluster ToT distribution overlay of Barrel layer 3 with IBL.

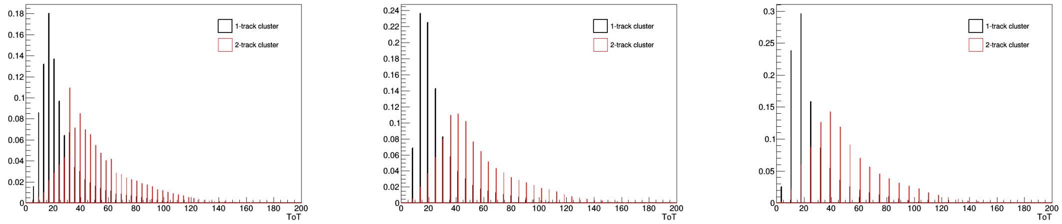


(a) 3-bit ToT

(b) 4-bit ToT

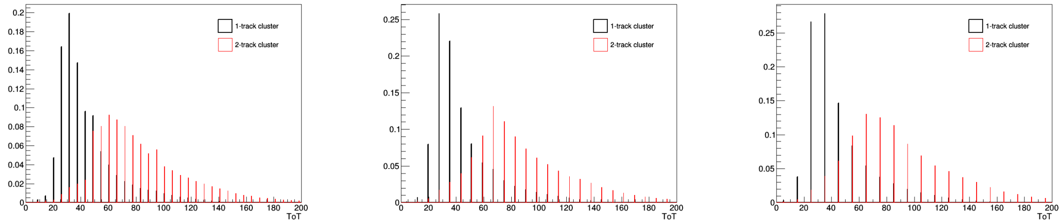
(c) 5-bit ToT

Figure 5.23: Mann-Whitney U(1-track cluster ToT, 2-track cluster ToT) versus saturation point for barrel layers for 3-bit ToT (leftmost), 4-bit ToT (center) and 5-bit ToT (rightmost).



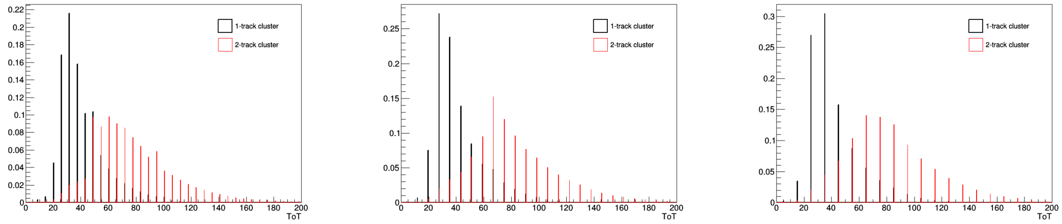
(a) ToT saturation=30 (U=0.173) (b) ToT saturation=44 (U_{min}=0.168) (c) ToT saturation=57 (U=0.176)

Figure 5.24: 1-track and 2-track cluster ToT comparison for 3-bit barrel layer 1.



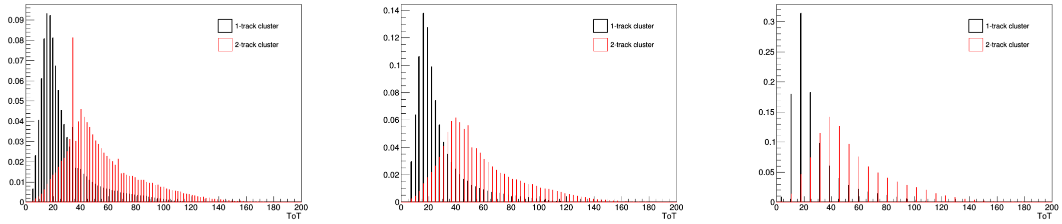
(a) ToT saturation=46 (U=0.141) (b) ToT saturation=63 (U_{min}=0.136) (c) ToT saturation=80 (U=0.139)

Figure 5.25: 1-track and 2-track cluster ToT comparison for 3-bit barrel layer 2.



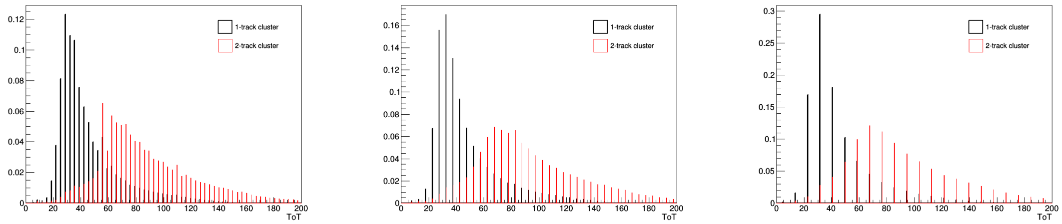
(a) ToT saturation=46 (U=0.124) (b) ToT saturation=63 (U_{min}=0.118) (c) ToT saturation=80 (U=0.122)

Figure 5.26: 1-track and 2-track cluster ToT comparison for 3-bit barrel layer 3.



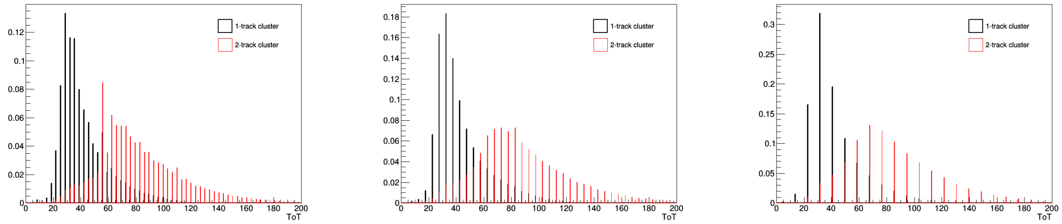
(a) ToT saturation=33 (U=0.169) (b) ToT saturation=47 (U_{min}=0.165) (c) ToT saturation=112 (U=0.172)

Figure 5.27: 1-track and 2-track cluster ToT comparison for 4-bit barrel layer 1.



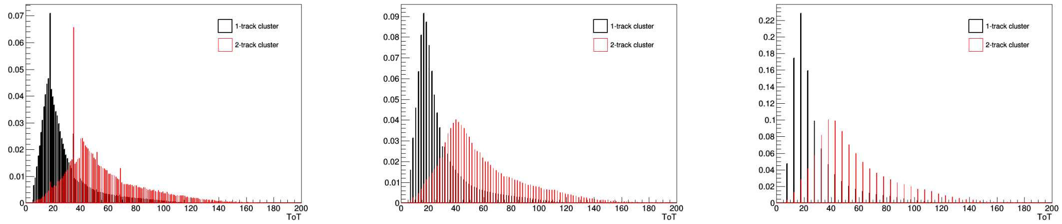
(a) ToT saturation=54 (U=0.135) (b) ToT saturation=80 (U_{min}=0.131) (c) ToT saturation=144 (U=0.136)

Figure 5.28: 1-track and 2-track cluster ToT comparison for 4-bit barrel layer 2.



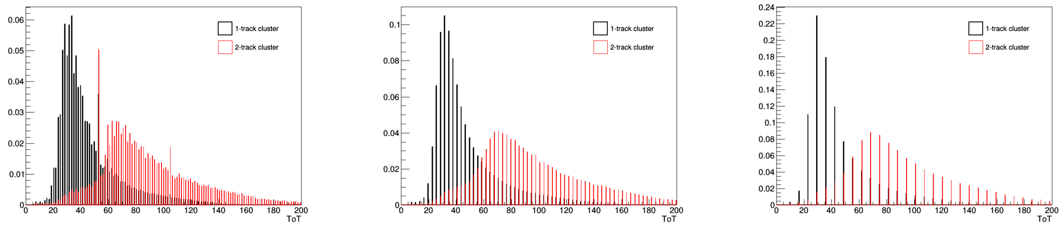
(a) ToT saturation=54 (U=0.117) (b) ToT saturation=80 (U_{min}=0.113) (c) ToT saturation=144 (U=0.119)

Figure 5.29: 1-track and 2-track cluster ToT comparison for 4-bit barrel layer 3.



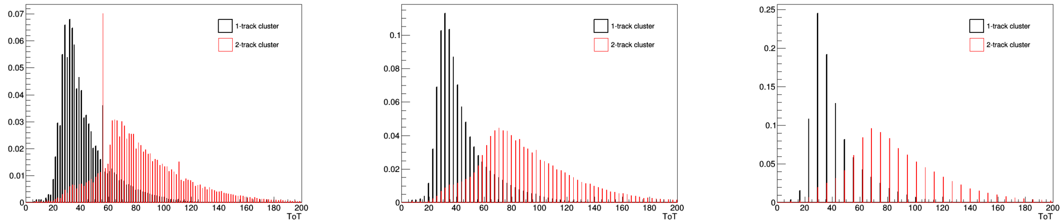
(a) ToT saturation=34 (U=0.168) (b) ToT saturation=62 (U_{min}=0.164) (c) ToT saturation=160 (U=0.167)

Figure 5.30: 1-track and 2-track cluster ToT comparison for 5-bit barrel layer 1.



(a) ToT saturation=52 (U=0.136) (b) ToT saturation=96 (U_{min}=0.129) (c) ToT saturation=208 (U=0.132)

Figure 5.31: 1-track and 2-track cluster ToT comparison for 5-bit barrel layer 2.



(a) ToT saturation=55 (U=0.116) (b) ToT saturation=96 (U_{min}=0.111) (c) ToT saturation=208 (U=0.115)

Figure 5.32: 1-track and 2-track cluster ToT comparison for 5-bit barrel layer 3.

particle clusters has been summarized in Figure 5.33 and Table 5.8.

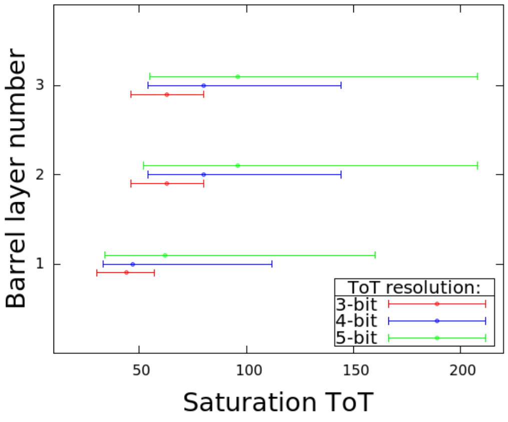


Figure 5.33: ToT saturation for optimum discrimination (with 5% error) between 1-track and 2-track clusters.

Table 5.8: U scores for the optimum saturation ToTs (in Figure 5.33).

	Barrel L1	Barrel L2	Barrel L3
ToT saturation for negative error (3-bit)	30(U=0.173)	46(U=0.141)	46(U=0.124)
ToT saturation for best classification for 3-bit	44(U=0.168)	63(U=0.136)	80(U=0.118)
ToT saturation for positive error (3-bit)	57(U=0.176)	80(U=0.139)	80(U=0.122)
ToT saturation for negative error (4-bit)	33(U=0.169)	54(U=0.135)	54(U=0.117)
ToT saturation for best classification for 4-bit	47(U=0.165)	80(U=0.131)	80(U=0.113)
ToT saturation for positive error (4-bit)	112(U=0.172)	144(U=0.136)	144(U=0.119)
ToT saturation for negative error (5-bit)	34(U=0.168)	52(U=0.136)	55(U=0.116)
ToT saturation for best classification for 5-bit	62(U=0.164)	96(U=0.129)	96(U=0.111)
ToT saturation for positive error (5-bit)	160(U=0.167)	208(U=0.132)	208(U=0.115)

The same strategy to optimize performance was further used for CTIDE NNs.

5.6 CTIDE NN Hyperparameter Optimization

There are three key hyperparameters involved in the training of a NN:

- **Learning Rate**: How fast the network will change its weights at every step.
- **Momentum**: Use a fraction of the previous step in the current step.
- **Regularizer**: Applies penalties for excessively large weights.

Several networks were trained with varying hyperparameters, and their performance evaluated. The trainings used simulated dijet events, whose 8-bit Time-over-Threshold (ToT) measurements were converted to 4 bits by using the conversion scheme in Section 5.4.1. Alongside adjusting hyperparameters and the saturation point, a cross-validation (CV) scheme was introduced. The current NN training algorithm reserves a fixed 10% of its training set as a validation set; training is stopped when the loss against the validation set stops decreasing for 20 epochs. In k-fold CV, the training set is split into k equal parts. Then k networks are trained, where the ith network uses the ith part for its validation set, and the remaining data for its training. To obtain a single output, one simply averages the outputs of the k individual networks. A hyperparameter analysis was also conducted using 4-fold CV.

To analyze the efficiency of the NNs, each NN was evaluated on 100,000 simulated clusters, disjoint from the training set. The Mann-Whitney U statistic was then used to quantify how effective each network is. This analysis results in six U values; where comparison MvsN denotes the classification of M-particle clusters from N-particle clusters using number NN's probability for a M-particle cluster.

The dependence of the U values on the saturation point are shown in Figure 5.34. The six U values and their average are plotted as percent changes from the corresponding values using a NN trained on the original 8-bit data (no 4-bit conversion).

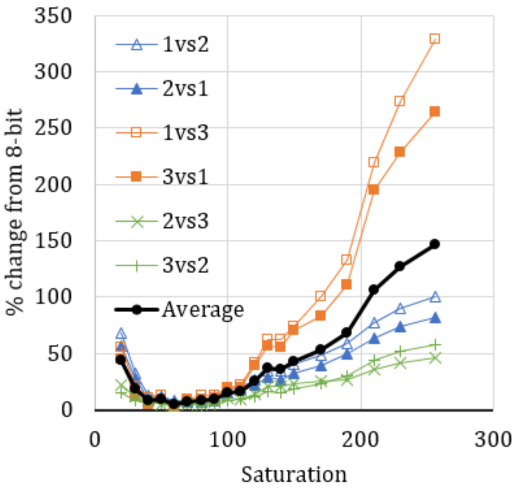


Figure 5.34: U values from varying the saturation point, as percent changes from values using a NN trained on 8-bit data. Hyperparameters were left at current settings, with no cross-validation.

These trials had no CV, and unchanged hyperparameters. A clear minimum appears at a saturation point of $x = 60$, where the 4-bit performance is only around 5% worse than that at full 8-bit resolution.

The effect of introducing CV is shown in Figure 5.35. Because the training algorithm utilizes random numbers when initializing a network's weights, there exists some variance between different networks trained with the same configuration. These correspond to a standard deviation in the U values of around 1%. Figure 5.35 plots the change in the U statistics using k-fold CV from the U values without CV, divided by the uncertainty from randomness (Δ/σ). Negative Δ/σ indicates an improvement in performance. A significant improvement is seen by introducing CV, with k=6 folds producing the best results.

Finally, the dependence of the U statistic on the three hyperparameters is shown in Table 5.9. Here, the changes in the U statistics by altering a single hyperparameter are divided by the uncertainty from randomness (Δ/σ). The values displayed are an average of the individual Δ/σ for the six U comparisons, with $x = 90$ saturation.

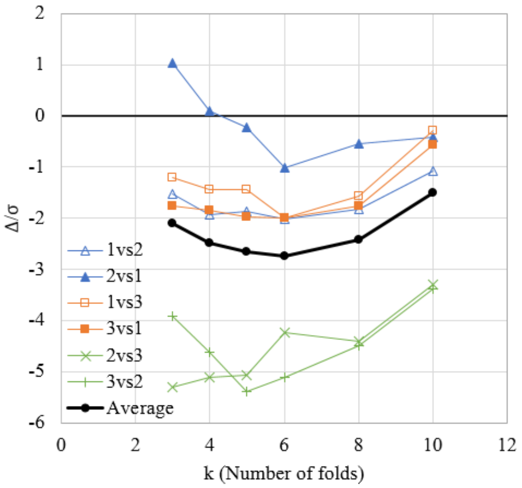


Figure 5.35: $\Delta U/\sigma$ from k-fold cross-validation, compared to no cross-validation, at saturation $x = 90$.

The analysis was conducted for both no CV and 4-fold CV. Notably, the effect of changing a single hyperparameter is often within 1σ, and the only improvement with $\Delta/\sigma < -1$, with negative values indicating improvement, is learning rate $= 0.04$ using 4-fold CV. This suggests the current hyperparameter settings are fairly robust and optimized already.

5.7 Switching Back to Charge from ToT

The reason behind working with ToT to develop a conversion scheme, to reduce the charge information, was that peaks, especially for higher charge values, in the per-pixel charge distribution were not distinctly visible (see Figure 5.10b) and so the entire range of charge could not be determined. The conversion scheme, developed for ToT in Section 5.4, does not however use the entire range, as it is saturated a given value. So, if the same conversion scheme is applied on charge directly, saturating the charge range at a lower value can avoid the issue of unresolvable higher charge peaks. It should be noted that the charge, and not the ToT, is generally used ATLAS-wide in physics and performance studies. Therefore, it would be convenient to use charge

Table 5.9: $\Delta U/\sigma$ from varying a single hyperparameter from its current setting, at saturation $x = 90$. The test was conducted for both no CV and 4-fold CV. Results are an average over the 6 U values. Negative values indicate improvement in NN efficiency, whereas positive values indicate decreased performance.

Hyperparameter Setting		No CV	4-fold CV
Momentum (current = 0.4)	0.2	-0.44	-0.44
	0.6	0.73	0.99
	0.8	1.72	3.07
Learning Rate (current = 0.08)	0.04	-0.61	-1.13
	0.12	0.41	-0.02
	0.16	0.77	1.56
Regularizer (current = 1E-07)	2E-07	-0.18	0.42
	1E-06	-0.23	1.64
	1E-05	1.04	11.67

directly, instead of ToT, to resolve merged clusters at the ambiguity resolver.

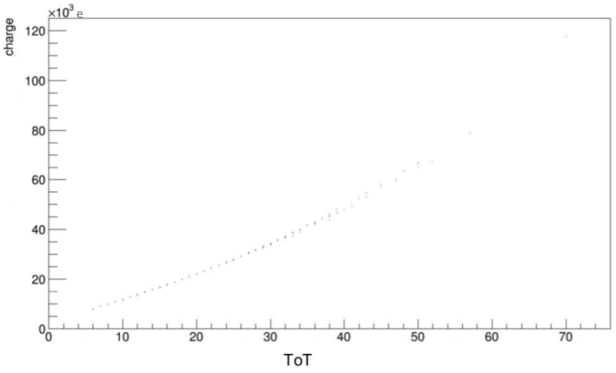


Figure 5.36: Barrel L1 η-module -2, ϕ-module 0.

Figure 5.37: Conversion of truth charge into ToT at a small region in the inner detector barrel region. The actual conversion function is $ToT = A\frac{Charge+E}{Charge+C}$, where parameters A, E and C are calibrated for each pixel based on its η, ϕ and detector layer. This function is approximately linear for lower charge values.

The ToT is approximately linear to the charge, at least for the lower values (see Figure 5.37) in the ATLAS inner detector. The final step in the conversion scheme involves a linear mapping of the saturated 8-bit ToT range to a lower n-bit ToT range. This step would be equivalent when applied on the charge in this linear region. For the rest of the study, charge has been used, instead of ToT, as inputs to study the variation in performance of the CTIDE NNs at reduced charge information, as discussed in Section 5.8.

5.8 Performance Study of Number & Position NNs

This study is intended to provide an estimate of the effect of reduced charge information, among other changes in the HL-LHC tracker upgrade, on the performance of ambiguity resolver at the ITK. Therefore, only the number and position neural networks were studied here. Similar course of action was followed, as with cluster ToT (Section 5.5.5), for the position and number networks. First, metrics to quantify performances of the position and number networks were obtained in Sections 5.8.1 and 5.8.2 respectively. These metrics were used to optimize the performance of these networks by tuning the saturation charge. Finally, optimal performances at different lower bit charges were compared to study the performance as a function of charge information (Section 5.9).

5.8.1 Performance Metric(s) for Number Neural Network

The number neural network helps in classifying merged clusters into three track multiplicity classes, viz. 1-particle, 2-particle and $\geqslant$3-particle classes. It yields three numbers for each cluster: $P(1)$ which is the probability (or a score) of the cluster containing only 1-particle, and similarly $P(2)$ and $P(3)$ for probabilities of the cluster containing 2-particle and $\geqslant$3-particle classes respectively. It can then be decided, whether a merged cluster belongs to a particular particle multiplicity class or any of the other two complementary classes, based on the probability (or score) assigned by the number network for that class to the cluster. For example, if a cluster is assigned $P(2) = 52\%$, one can decide that the cluster is not a 2-particle cluster, but belongs to any of the other two complementary classes, if the minimum cutoff for $P(2)$ is set to be anything greater than 52%. In other words, the score $P(2)$ is a classifier between 2-particle class and any of its two complementary classes and therefore, its performance can be obtained by two receiver operating characteristic (ROC) curves, i.e., 2-particle versus $\geqslant$3-particle class and 2-particle versus 1-particle class ROC curves. Here, the ROC curves have been created by plotting the false negative rate[3] versus false positive rate[4] (see Figure 5.11). Thus, the AUC of this type of ROC curve, for a score $P(n)$, would be equivalent to the Mann-Whitney U score (Section 5.5.3) (or simply U score) between the true positive and the false positive score[5] distribution of $P(n)$.

There are six such ROC curves in total, considering $P(1)$, $P(2)$ and $P(3)$. It should be however noted, that $P(1) + P(2) + P(3) = 1$ for a cluster. So, only four out of six ROC curves are independent. The metrics of performance for the number neural network are the 6 AUCs (or equivalently 6 U's). The lower the U values, the better is the performance of the number network.

5.8.2 Performance Metric(s) for Position Neural Network

The position neural network is a regression network, which estimates the local[6] x (in the local ϕ or transverse direction) and y (in the local θ or longitudinal direction)

---

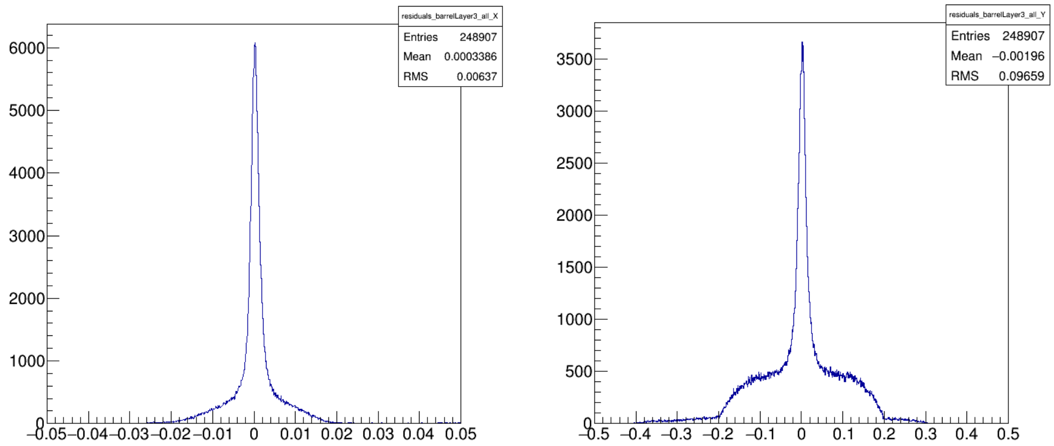
[3] **False negative rate:** probability of classifying a true n-particle multiplicity cluster as m-particle multiplicity cluster based on its $P(n)$ value

[4] **False positive rate:** probability of classifying a true m-particle multiplicity cluster as n-particle multiplicity cluster based on its $P(n)$ value

[5] The **true positive score** is the $P(n)$ score assigned to the true n-particle multiplicity (positive class) clusters, while the **false positive score** is the $P(n)$ score assigned to the true m-particle (complementary or negative class) clusters

[6] Local x and y positions are measured with respect to the center of a cluster.

positions of the particle(s) in a merged cluster. Therefore, each particle multiplicity class has its own position networks i.e., 1-particle position network estimates only one set of local (x,y) position for a cluster, while 2- and ⩾3-particle position network estimates two and three set of local (x,y) positions respectively.



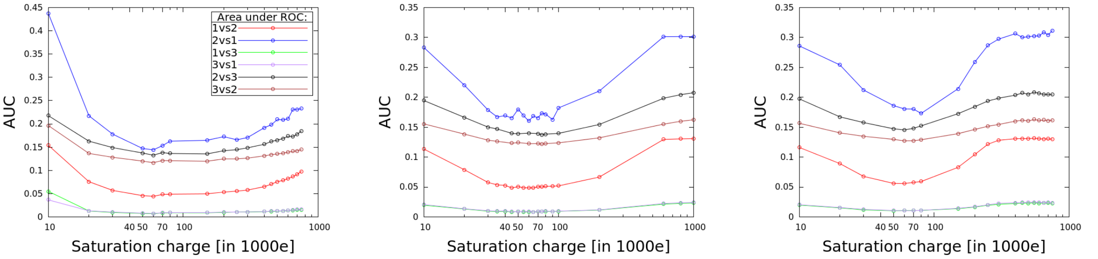
(a) X residual (saturation 100 × 10^3e) (b) Y residual (saturation 90 × 10^3e)

Figure 5.38: Residuals for 1-particle clusters in X and Y directions (in mm) for 8-bit charge resolution.

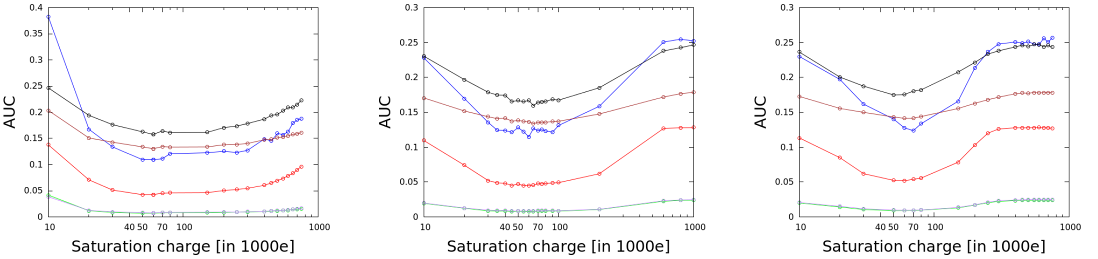
The difference between the estimated local x (y) position and true local x (y) position of a particle, in a merged cluster, is defined as the particle's X (Y) residual (see Figure 5.38). The spread of a residual distribution for a position network gives the quality of its estimation. So, the root mean square (RMS) of both x and y residual distributions have been used as metrics of performance for each of the three position neural networks.

5.9 Results

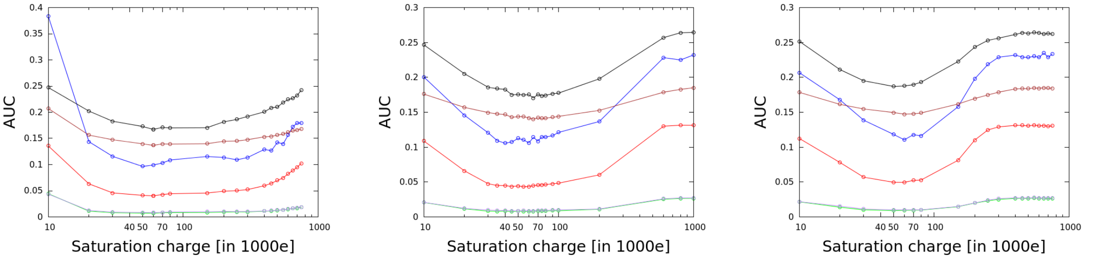
As discussed in Section 5.4.1 about conversion scheme, in 8-bit charge to lower n-bit charge conversion, saturation charge is a free parameter. However, the performance of CTIDE NNs depend on saturation charge, as discussed in Section 5.4.2. Similar to the cluster ToT in Section 5.5.5, the performances of number and position networks, at lower n-bit charge information, was optimized with respect to saturation charge.



(a) Layer 1 (4-bit) (b) Layer 1 (3-bit) (c) Layer 1 (2-bit)



(d) Layer 2 (4-bit) (e) Layer 2 (3-bit) (f) Layer 2 (2-bit)



(g) Layer 3 (4-bit) (h) Layer 3 (3-bit) (i) Layer 3 (2-bit)

Figure 5.39: Number NN AUC versus charge saturation. Leftmost column is for 4-bit charge, center is for 3-bit charge and rightmost column is for 3-bit charge. Topmost row is for layer 1, followed by layer 2 and the last row is layer 3. The legend used for the AUCs in (a) is also shared by the other plots.

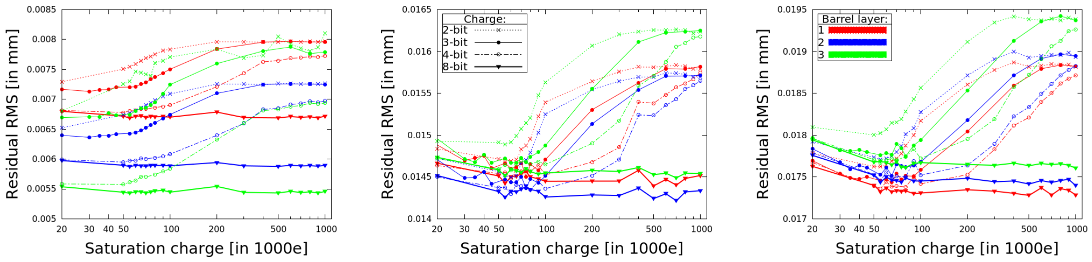
In this study, 8-bit charge information of 1 million barrel merged clusters were converted to 2-bit, 3-bit and 4-bit charge information. After training the number network for each of these three lower charge cases, the 6 AUCs (Section 5.8.1) were minimized

Table 5.10: Saturation for minimum area under the ROC curve for number NN in barrel layers. Minimum step size in saturation charge is 5000e.

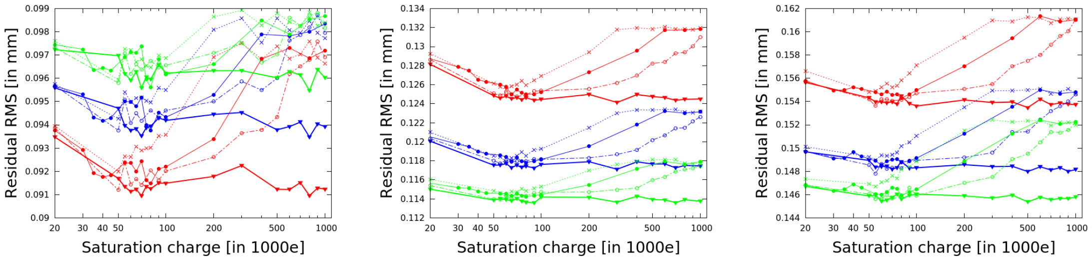
	Saturation charge ($\times 10^3$e)								
	4-bit charge			3-bit charge			2-bit charge		
ROC	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3
1vs2	60	50	60	45	55	55	60	60	60
2vs1	60	50	50	60	60	40	50	70	60
1vs3	60	60	60	45	65	45	60	50	50
3vs1	60	60	60	65	55	60	60	60	50
2vs3	60	60	60	75	65	65	60	50	50
3vs2	60	60	60	75	65	65	60	60	60

with respect to saturation charge. The saturation charge for optimum number network performances are shown in Table 5.10.

After training the position networks, the RMS for X and Y residual (Section 5.8.2) were minimized with respect to saturation charge for each particle multiplicity class.



(a) Position 1 X (b) Position 2 X (c) Position 3 X



(d) Position 1 Y (e) Position 2 Y (f) Position 3 Y

Figure 5.40: Residual RMS versus charge saturation. Rightmost column is for position 1, followed by position 2 and the leftmost column is for position 3 NN. Top row is for X direction and bottom row is for Y direction.

Table 5.11: Saturation for minimum residual RMS (in mm) for position NN in barrel layers. Minimum step size in saturation charge is 5000e.

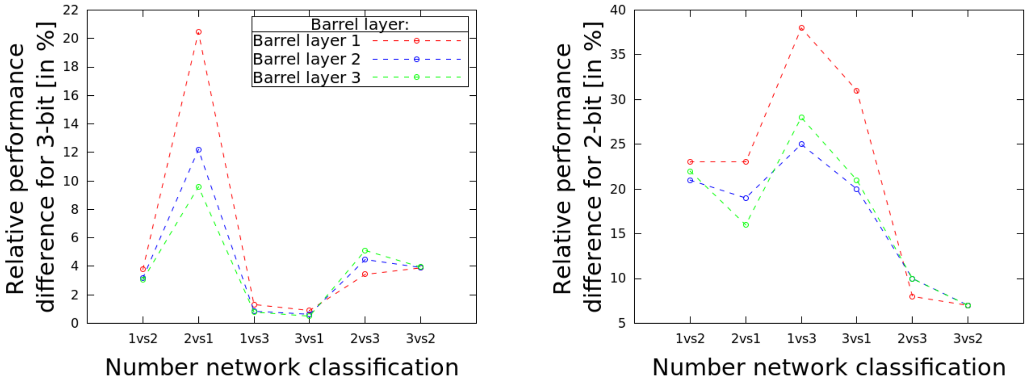
	Saturation charge ($\times 10^3$e)								
	4-bit charge			3-bit charge			2-bit charge		
Direction	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3
X(pos1)	50	50	50	30	30	20	20	20	20
Y(pos1)	50	50	50	80	80	80	50	50	50
X(pos2)	60	60	60	75	75	60	60	60	55
Y(pos2)	70	55	55	65	80	80	60	60	60
X(pos3)	60	60	60	75	75	60	60	60	55
Y(pos3)	70	55	55	65	80	80	60	60	60

The saturation charge for optimum position network performances are shown in Table 5.11. The optimum saturations listed in Tables 5.10 and 5.11 for number and position networks respectively, might not be the actual optimum saturations, as the region around optimum performances have some fluctuation (see the interval of $[5.0 \times 10^4, 7.0 \times 10^4]$ in the plots in Figure 5.39 for number network and Figure 5.40 for position networks).

Nonetheless, the actual optimum charge saturation should generally lie within a small interval of $[5.0 \times 10^4, 7.0 \times 10^4]$, for lower n-bit charge resolution across both number and position networks, since the fluctuation in performance is not very significant (for 4 bit charge, performance variation within $\sim 1\%$ in position networks and around $\sim 10\%$ in number networks). It should be noted here that the optimum saturation charge range lies within the linear range of the charge to ToT dependence (as discussed in Section 5.7). Thus, the conversion scheme, initially developed for ToTs, works just as well for the charge here.

For number network, relative difference in performance (AUC) of lower n-bit charge resolution with respect to 4-bit charge resolution was used to study the performance drop at lower charge information. Optimal performances were used for this comparison. Since there is a significant noise in the range $[4.5 \times 10^4\text{e}, 7.5 \times 10^4\text{e}]$, which seems to contain the actual optimum charge saturation, the relative performance differences were averaged in this range.

The results using relative performance difference have been summarized in the Figure 5.41 and Table 5.12. The drop in 2 vs 1 classification is significant for 3-bit. The relative performance degradation trend for 2-bit is different from 3-bit. This is because the performance of 4-bit charge resolution is lot better than 2-bit, in other words the performance metric AUC for 4-bit is too smaller than that of 2-bit. This makes the relative performance drop of 2-bit charge resolution sensitive to small random fluctuations in 4-bit charge's performance. This becomes obvious when the absolute performance difference with respect to 4-bit is observed (see Figure 5.42), where the trend for 2-bit and 3-bit are the similar.



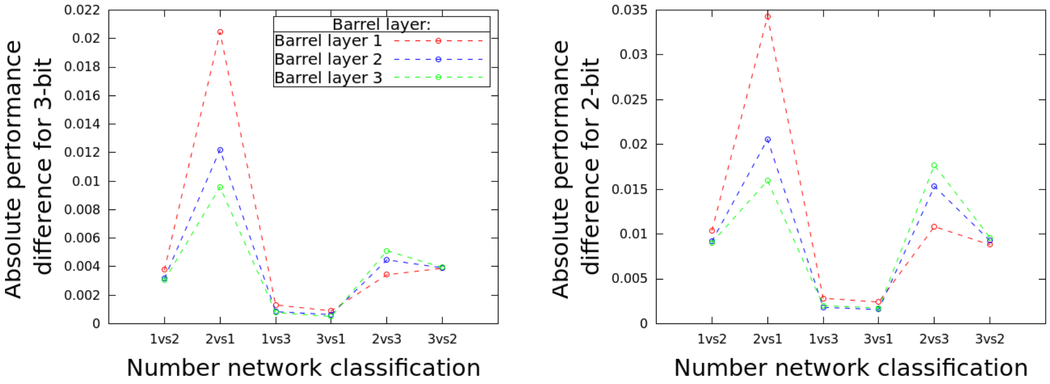
(a) comparison with 3-bit charge (b) comparison with 2-bit charge

Figure 5.41: Relative performance difference in number network: (a) Average relative difference in U for 3-bit with respect to 4-bit resolution charge, (b) Average relative difference in U for 2-bit with respect to 4-bit resolution charge.

Similarly for position networks, the relative difference in the RMS of the residual (performance metric) at the lower n-bit charge resolution with respect to the RMS at 8-bit charge resolution is obtained. This relative performance difference is averaged over a wider saturation charge range of [2.0×10^4e, 8.0×10^4e] as the performance variation is not significantly large in this range. The results have been summarized in the Figure 5.43 and Table 5.13. X position for 1-particle has a significant drop in

Table 5.12: The numerical values of the relative performance of the number network for 2-bit and 3-bit charge resolution with respect to 4-bit charge resolution. The pattern of relative performance drop is not same for 2-bit and 3-bit. This is because 4-bit performance metric is too small compared to that of 2-bit, so small fluctuation in 4-bit performance metric lead to misleading performance drop trend in 2-bit. This becomes clear by evaluating the absolute errors (see Figure 5.42).

	$\frac{U(3bit)-U(4bit)}{U(4bit)}$ (in %)			$\frac{U(2bit)-U(4bit)}{U(4bit)}$ (in %)		
Classifier	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3
1vs2	8	7	7	23	21	22
2vs1	14	11	10	23	19	16
1vs3	18	12	11	38	25	28
3vs1	11	8	6	31	20	21
2vs3	3	3	3	8	10	10
3vs2	3	3	3	7	7	7



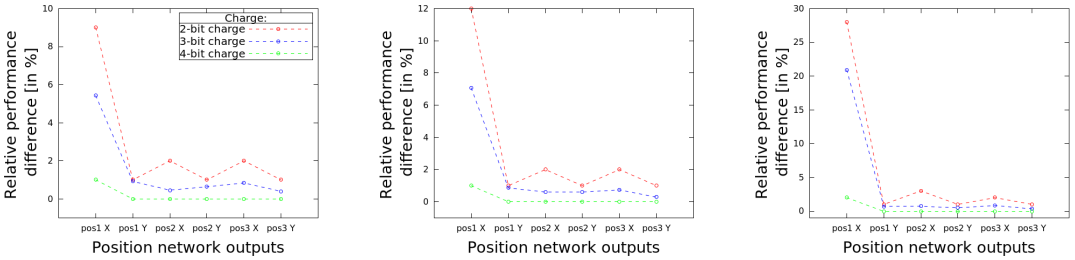
(a) comparison with 3-bit charge (b) comparison with 2-bit charge

Figure 5.42: Absolute performance difference in number network: (a) Average absolute difference in U for 3-bit with respect to 4-bit resolution charge, (b) Average absolute difference in U for 2-bit with respect to 4-bit resolution charge.

performance for 3-bit and 2-bit. The performance drop at 4-bit is not significant.

5.10 Conclusions

The RMS of the position residual has been used as the position NN performance metric. The area under the ROC curve (AUC) has been used as the number NN performance metric. The performance of both number and position neural networks



(a) barrel layer 1 (b) barrel layer 2 (c) barrel layer 3

Figure 5.43: Relative performance difference in position networks: Average relative difference in residual RMS for 2-bit and 3-bit with respect to 8-bit resolution charge respectively in barrel layer 1 (a), layer 2 (b), layer 3 (c).

Table 5.13: The numerical values of the relative performance of the position networks for 2-bit, 3-bit and 4-bit charge resolution with respect to 8-bit charge resolution.

	$\frac{(RMS(2bit)-RMS(8bit))}{RMS(8bit)}$ (in %)			$\frac{(RMS(3bit)-RMS(8bit))}{RMS(8bit)}$ (in %)			$\frac{(RMS(4bit)-RMS(8bit))}{RMS(8bit)}$ (in %)		
position NN	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3
pos1 X	9	12	28	5	7	21	1	1	2
pos1 Y	1	1	1	1	1	1	0	0	0
pos2 X	2	2	3	1	1	1	0	0	0
pos2 Y	1	1	1	1	1	0	0	0	0
pos3 X	2	2	2	1	1	1	0	0	0
pos3 Y	1	1	1	0	0	0	0	0	0

(NNs) generally decrease with lower charge resolution. The drop in performance from 8-bit to 4-bit was smaller than from 4-bit to 3-bit or a similar drop from 3-bit to 2-bit. So, 4-bit charge resolution would have a minimal effect on physics performance compared to 3-bit or 2-bit.

The optimum charge saturation lies generally within a small range of $[5.0 \times 10^4 e, 7.0 \times 10^4 e]$, for lower n-bit charge resolution across all NNs (number and position networks). This charge range does not have significant variation in performance (for 4-bit charge, performance variation within $\sim 1\%$ in position networks and around $\sim 10\%$ in number networks). Since the charge distribution for the 3 barrel layers are similar, so saturating at around $6.0 \times 10^4 e$ includes the relevant charge distribution for CTIDE NN (see Figure 5.44).

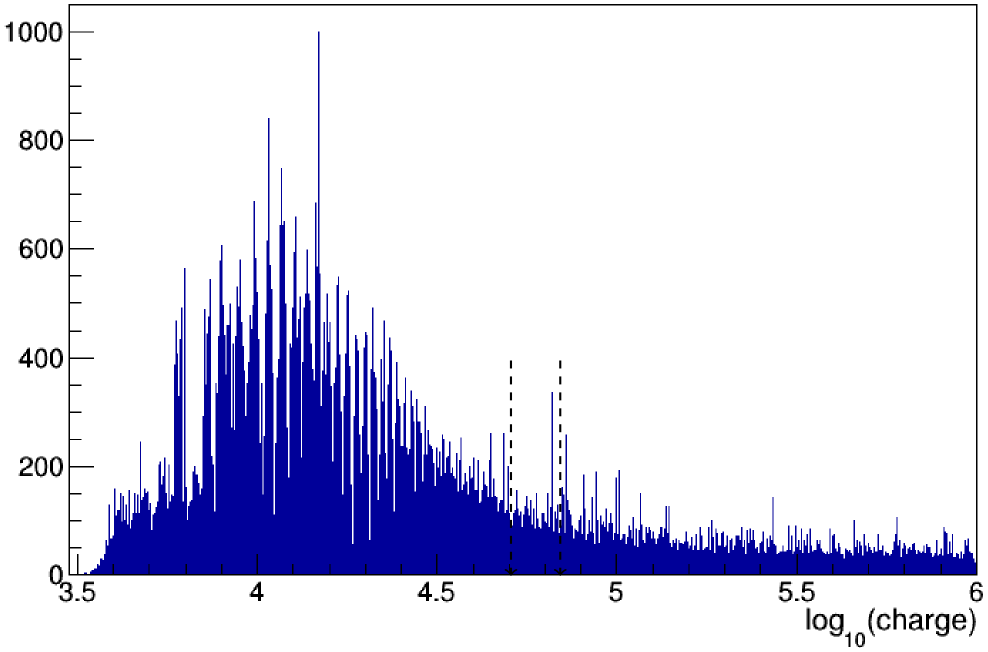


Figure 5.44: Original barrel charge distribution. The dotted arrows indicate the range of 50,000e-70,000e which contained most of the saturation charge points while reducing charge resolution since most of the useful information about the deposited charge in a cluster is contained below this range.

Chapter 6

Superboosted τ-leptons at FCC

Future experiments like FCC-hh (proton-proton collider at very high energy $\sqrt{s} = 100$ TeV [80]) would probe Beyond Standard Model channels involving heavy resonances, many of which have τ leptons in their final states [81, 82]. The detectors should therefore be able to detect the super-boosted final states in order to reconstruct these heavy resonances. Additionally, the τ lepton is the most difficult charged lepton to identify, and it decays both hadronically ($\sim 65\%$) and leptonically ($\sim 35\%$) due to its heavy mass of 1.777 GeV. The hadronically decaying τ leptons (referred here as τ_{had}) have QCD jets as their major background. A high-granularity silicon-tungsten calorimeter with sufficient radiation and interaction lengths to contain all the τ-decay products has been designed. Since the granularity is very high, a fast simulation study would not be able to give the required insights into the detector design. Here we present the detector response to super-boosted τ_{had} jets ($\sim$ 500 GeV) using a full GEANT4-based simulation of the FCC-hh detector. The design of FCC-hh detector was carried out by a team of less than 10 members, where the author contributed to a lesser extent to the analysis framework development for the work and solely performed the benchmarking of the detector design using super-boosted τ-lepton reconstruction, as discussed in this chapter. The work in this chapter presented in the ICHEP conference [83].

6.1 Designing a High-granularity Calorimeter

FCC-hh like detector needs to be a high granularity multipurpose detector, and one such detector already designed for high energy jets was the Silicon Detector (SiD) by

International Linear Collider (ILC). It thus served as a good starting point.

6.1.1 SiD Detector

The SiD detector [84] (see Figure 6.1) is a high-granularity cylindrically symmetric detector for the ILC experiment, which is designed for $\sim$ 500 GeV jets. Among its key features are: its 5 Tesla solenoid and Silicon tracker. Its EM Calorimeter are made of Tungsten absorber with Silicon sensor cells of size 3.5 mm×3.5 mm. Its Hadronic Calorimeter is made of Steel absorber with RPC sensor cells of size 10 mm×10 mm. To contain $\sim$ 500 GeV jets the Hadronic Calorimeter has 40 layers in the barrel, providing about $\sim 5.1X_0$. It is also optimized for Particle Flow Algorithm (PFA).

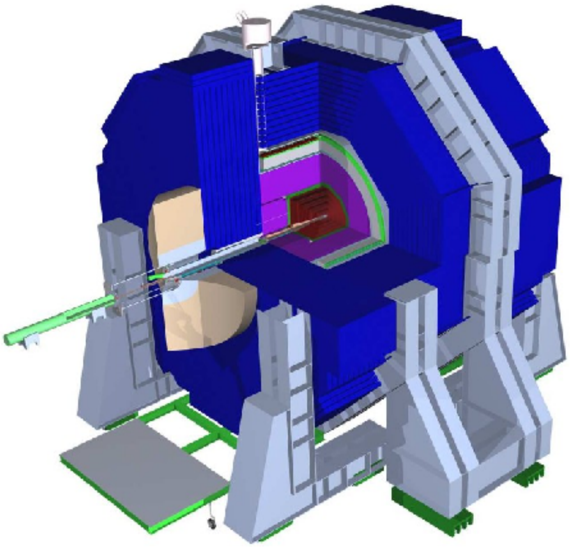


Figure 6.1: The SiD detector of ILC.

6.1.2 Designing a Detector for TeV-scale Boosted Physics

The SiD detector fulfilled almost all the criteria of a FCC-like detector, with its high transverse and longitudinal granularity, magnetic field and optimization to PFA in order to get particle objects. So, we retained all these characteristics of SiD in our detector - Hadronic Calorimeter granularity, detector material and solenoid. But the radiation length of the SiD detector (designed for up to $\sim$ 500 GeV jets) is not enough for the FCC-hh like detector to contain up to several TeV jets (from 100 TeV p-p collisions). So, the size of the Hadronic Calorimeter radius in the barrel was doubled to $\sim$ 10 m (see Figure 6.2), providing $\sim$ 12 λ_I (interaction lengths) to prevent any leakage for up to $\sim 20-30$ TeV jets. These detector geometry in XML file format was used for the GEANT4 simulation of the FCC-hh like detector.

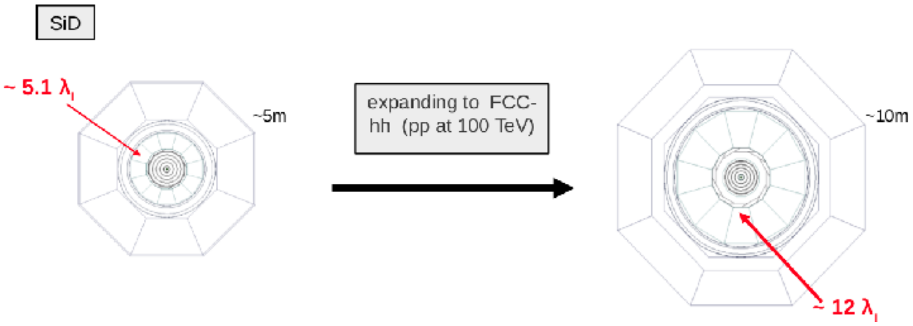


Figure 6.2: Designing a GEANT4 simulation for high granularity calorimeter (10 mm×10 mm) with $\sim 12\lambda_I$ to contain $20-30$ TeV jets.

6.2 Event Simulation and Reconstruction

For this study, Z' (1 TeV) $\rightarrow \tau\tau$ and Z' (1 TeV) $\rightarrow q\bar{q}$ Monte-Carlo events were used to model the signal and background processes respectively. The generator-level (referred to as truth-level) samples were obtained from a public repository for High Energy Physics Monte Carlo simulated events, HepSim [85], where they were produced using PYTHIA 6/MADGRAPH8. These samples were processed through the SiFCC detector simulation. Owing to the di-jet topology of both signal and background events, the Durham Jet Algorithm was used to cluster the final-state particles into two jets. These jets are then passed on to the τ-identification performance study.

6.3 Identification of Hadronically-decaying τ-leptons

τ-leptons being color-neutral decay via the weak interaction. This inhibits quark & gluon radiation in τ_{had} jets, distinguishing them from QCD jets as follows: (i) τ_{had} jets are more collimated and isolated; (ii) low track (charged-particle) multiplicity as τ-leptons predominantly decay via 1-prong ($\tau_{\text{had}} \to \pi^{\pm}\nu_{\tau} + n\pi^0$, branching ratio: $\sim 21\%$) or 3-prong ($\tau_{\text{had}} \to 3\pi^{\pm}\nu_{\tau} + n\pi^0$, branching ratio: $\sim 70\%$) channels; and (iii) presence of secondary vertex in the τ_{had} decay.

Based on these features, we use a set of variables to discriminate τ_{had} jets [86] against QCD background. We set the axis of a jet as the direction of the vector sum of momenta of its constituents. We divide the region containing the jet into two parts viz. the "core region" ($\Delta R \leqslant 0.1$, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ about the jet axis) which contains 90% of the jet energy, and the "isolation region" ($0.1 < \Delta R < 0.4$) which forms the periphery of the jet. The discriminating variables are:

Central energy fraction (f_{cent}) Fraction of transverse energy deposited inside the region $\Delta R < 0.05$ (i.e., half of the core radius) with respect to the total transverse energy deposited in the core region (Figure 6.4).

Number of isolation tracks (N_{track}^{iso}) Number of associated tracks in the isolation region (Figure 6.5).

Leading-track momentum fraction (f_{track}) Fraction of transverse momentum of the hardest associated track inside the core region with respect to the total transverse energy deposited in the core region (Figure 6.6).

Maximum ΔR (ΔR_{max}) Maximum ΔR between an associated track inside the core region and the jet axis (Figure 6.7).

Track radius (R_{track}) p_T-weighted distance of the associated tracks within $\Delta R <$

0.4 (i.e, core and isolation region combined) from the jet axis (Figure 6.8).

Track mass (M_{track}) Invariant mass calculated using all the associated tracks within $\Delta R < 0.4$, assuming $\pi_{\pm}$ mass for each track (Figure 6.9).

6.4 Results

The discrimination variables worked well in capturing the salient features of τ_{had} jets at the truth-level as well as detector-level (with the SiFCC detector design), as shown in Figures 6.4-6.9.

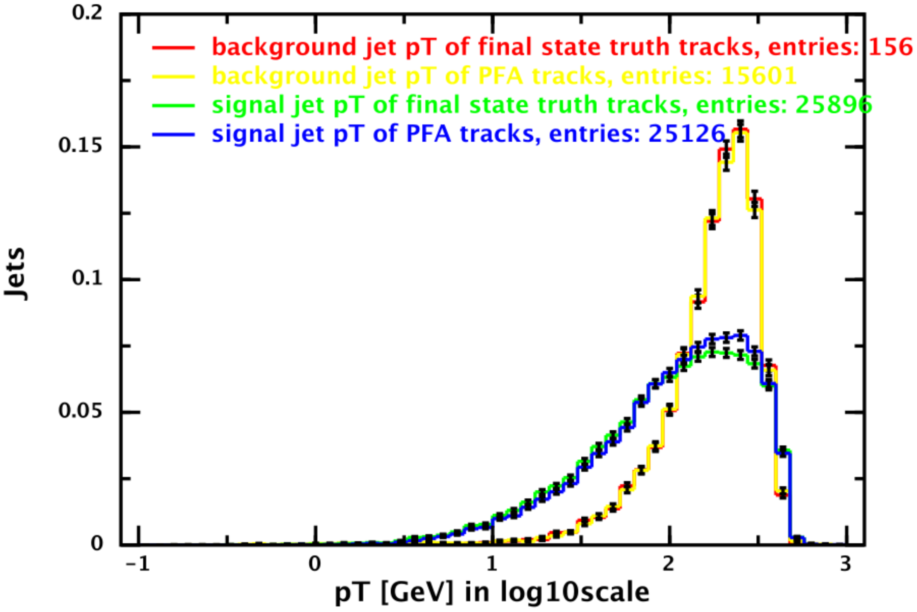
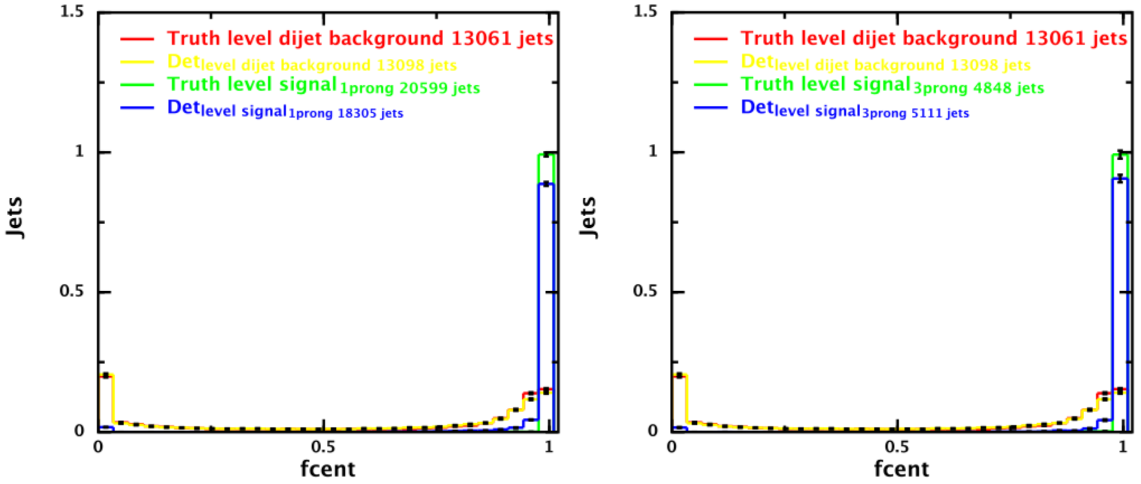


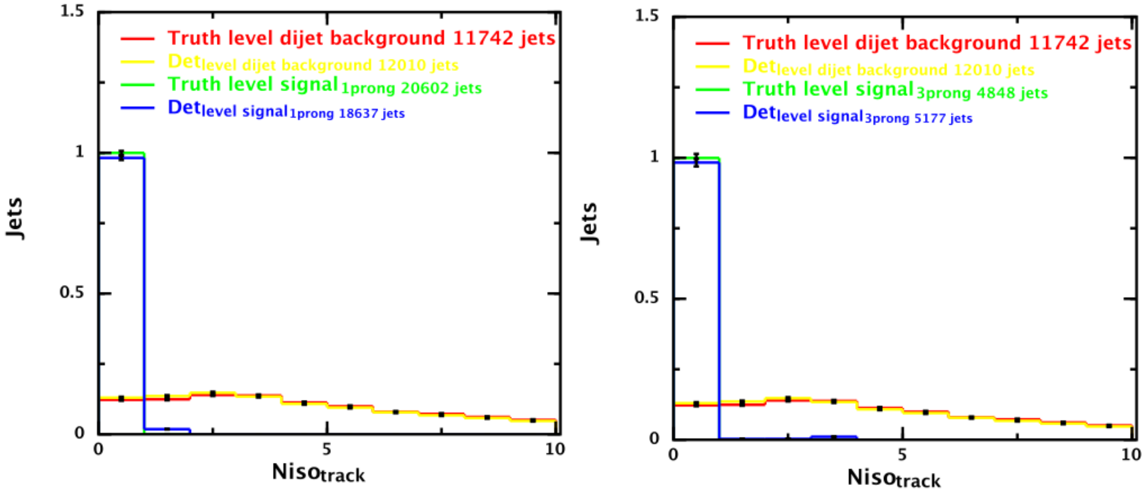
Figure 6.3: p_{T} (GeV) spectrum (logscale).

Due to sufficient granularity of the SiFCC detector design, the super-boosted final state particles could be resolved and reconstructed with high efficiency. This can be inferred from the exceptional agreement between the truth and detector-level distributions (Figures 6.4a, 6.5b) of the discrimination variables for both signal and background. This high reconstruction efficiency also bolsters the performance of the



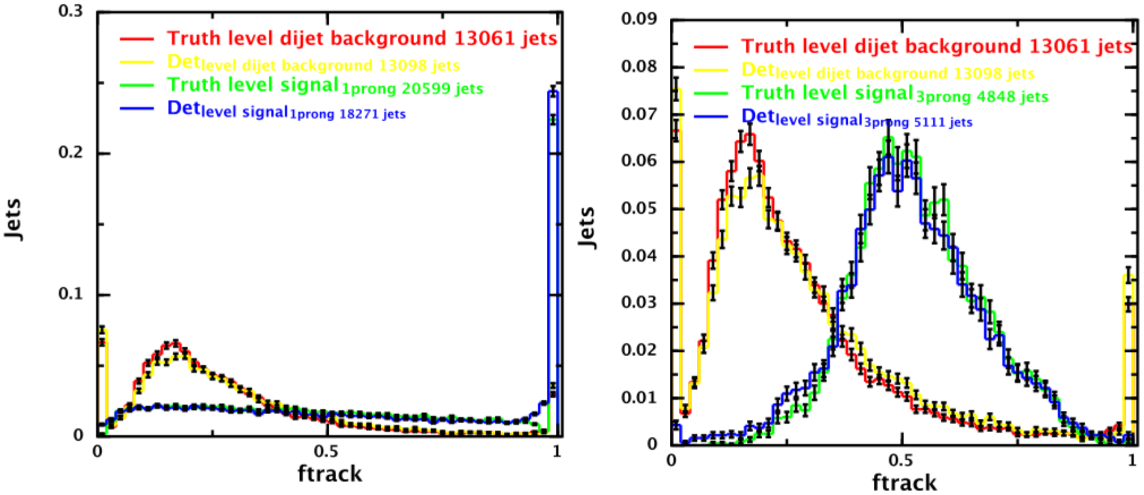
(a) f_{cent} (1 prong) (b) f_{cent} (3 prong)

Figure 6.4: $\tau_{had-vis}$ identification variables: Central energy fraction (f_{cent}).



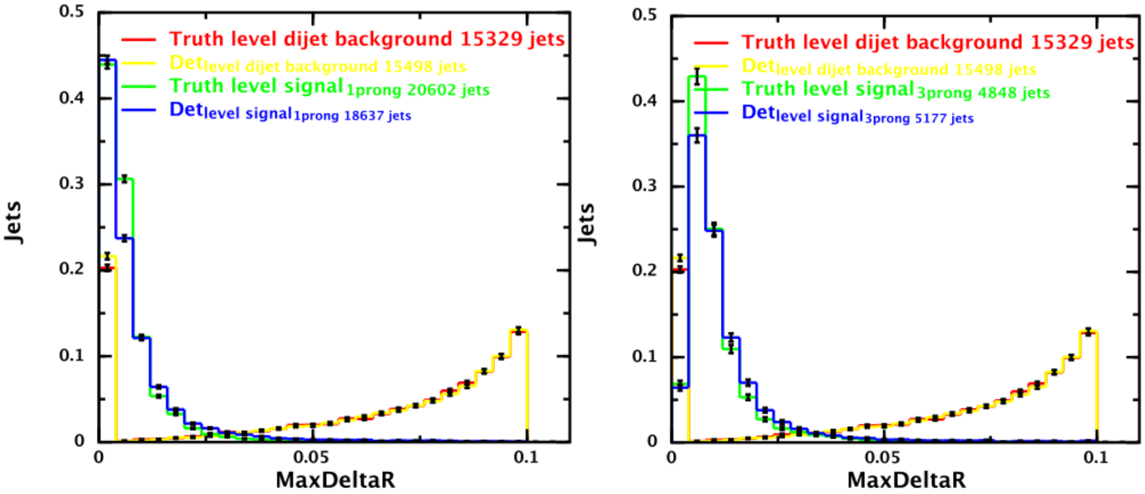
(a) N_{track}^{iso} (1 prong) (b) N_{track}^{iso} (3 prong)

Figure 6.5: $\tau_{had-vis}$ identification variables: Number of associated tracks in the isolation region (N_{track}^{iso}).



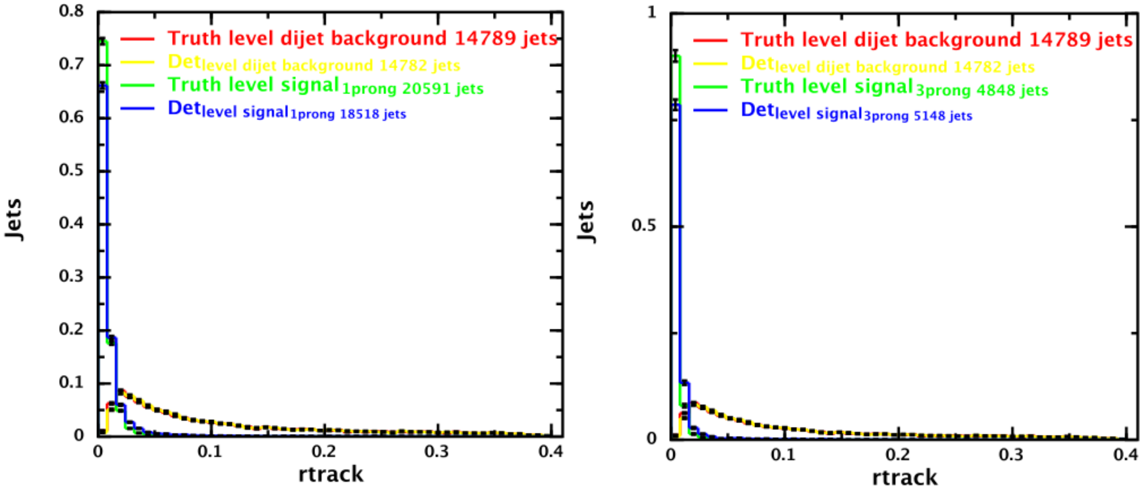
(a) f_{track} (1 prong) (b) f_{track} (3 prong)

Figure 6.6: $\tau_{had-vis}$ identification variables: Leading track momentum fraction (f_{track}).



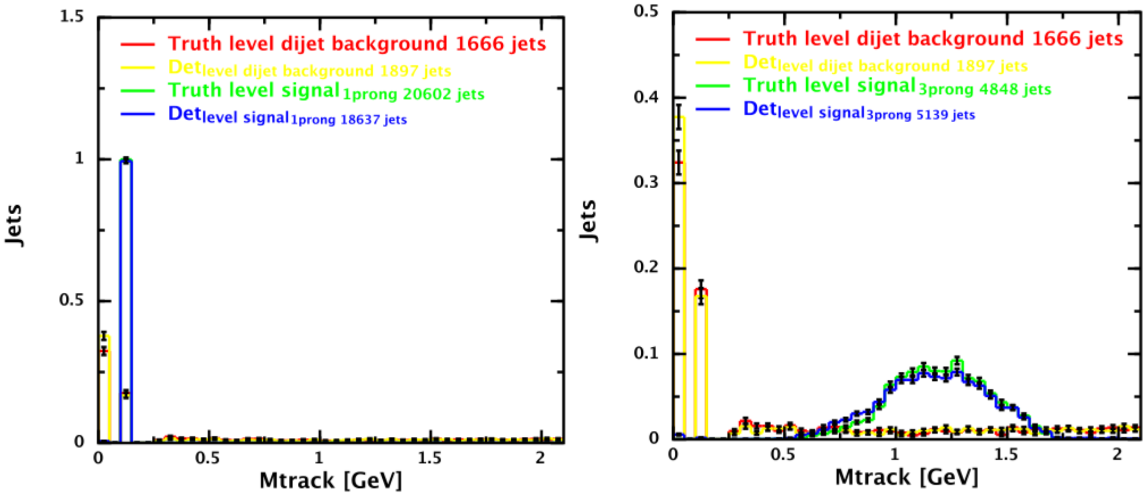
(a) ΔR_{max} (1 prong) (b) ΔR_{max} (3 prong)

Figure 6.7: $\tau_{had-vis}$ identification variables: Maximum ΔR (ΔR_{max}).



(a) R_{track} (1 prong) (b) R_{track} (3 prong)

Figure 6.8: $\tau_{had-vis}$ identification variables: Track radius (R_{track}).



(a) M_{track} (1 prong) (b) M_{track} (3 prong)

Figure 6.9: $\tau_{had-vis}$ identification variables:Track mass (M_{track}).

discrimination variables at the detector level. The signal and background events are clearly separated and even a simple cut-based selection can give a high yield of signal events, with good background rejection. Thus, the discrimination variables are efficient in identification of hadronically-decaying τ-leptons at the FCC-hh energy scale using the current detector design.

6.5 Conclusions and Future Directions

A reasonable agreement between the truth and detector levels can be observed, in the case of both signal and background events. Moreover, these variables are very effective in separating signals from background, showing that these discrimination variables could be useful at the FCC-hh energy scales. Although, the reason behind background multi-jets appearing so hollow, as indicated by the wide empty region in the center of f_{cent} (see Figure 6.4) should be investigated further. A robust classifier between $\tau_{had-vis}$ jets and multi-jets can be obtained at the FCC-hh energy scales by training a Boosted Decision Tree using these discriminating variables as inputs.

Chapter 7

PDF Variations in W Kinematics at CDF

Mass of W is an important consequence of Higgs mechanism and is related to some of the other precisely measured SM parameters (M_Z, α_{EM} and G_F) using: ${M_W}^2 = \frac{\pi\alpha_{EM}}{(\sqrt{2}G_F(1-\frac{M_W^2}{M_Z^2}))(1-\Delta r)}$ [87]. Here, the radiative correction (Δr) has BSM contributions. The largest source of uncertainty, in the current most precise W mass measurement, is due to PDF($\delta M_W^{PDF} \sim 10$ MeV[20]). PDFs [88] heavily influence the longitudinal boosts of the interacting partons, thus affecting the transverse quantities like lepton p_{T} used for measuring M_W. The author was the sole contributor for the PDF uncertainty estimation of the W mass measurement with full CDF-II data.

7.1 The CDF II Detector

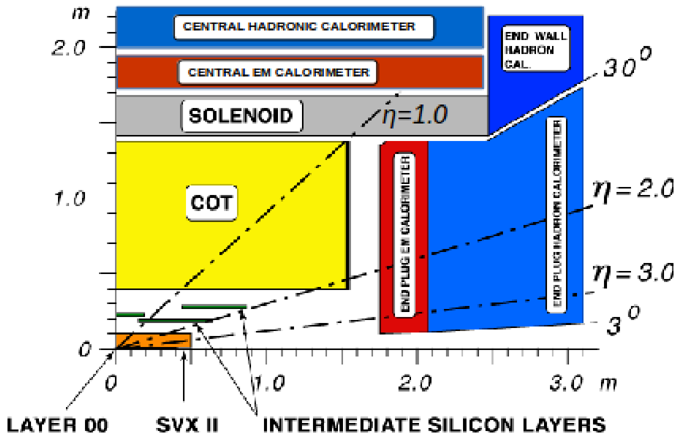


Figure 7.1: A cutaway schematic view of one quadrant of the inner portion of the CDF II.

The **C**ollider **D**etector at **F**ermilab (CDF) II [89, 90, 91, 92] is a cylindrical (symmetric in η & ϕ) detector of Fermilab Tevatron, used to study $p\bar{p}$ collisions occurring at its center. It had a three-level on-line trigger system, which could identify event topologies, such as W and Z production, and record it at a rate of 100 Hz (i.e., one out of every 17000 $p\bar{p}$ crossings). Figure 7.1 shows the cut-away view of a quadrant of the CDF II. After a $p\bar{p}$ collision, its daughter particles would pass through the precision tracking detectors. The first part of which is a Silicon tracking system consisting of an L00 layer (one single sided Si layer over the beam pipe at 1.6 cm radius), then 5 layers of double sided Si wafers from 2.5 cm through 10.6 cm radius called the SVX II, followed by the ISL which is a system 3 double-sided Si layers (1 layer at 23 cm radius in the central region of $|\eta| \leqslant 1$ & 2 layers at 20 cm and 29 cm respectively in region $1 < |\eta| < 2$). The last part of the tracking system is the Central Outer tracking (COT) detector, which is a drift chamber filled with 1 : 1 Argon-ethane gas at 1.4 T field from the solenoid. It stretches within $|z| < 155$ cm from 40 cm < radius < 138 cm, so any charged particle with $p_{\mathrm{T}} \gtrsim 300$ MeV with $|\eta| \lesssim 1$ can pass through the COT. The ionization drift velocity of the COT is 56 μm/ns, used to measure the time of flight. The daughter particle would next encounter the calorimeter system which is disjoint at $\eta = 0$. The Central Electromagnetic Calorimeter (CEM) is a system of alternating 31 layers of scintillators with 30 layers of Pb-Al plates providing $\sim 19X_0$. The Central Electromagnetic Strip Chamber (CES) is placed in the CEM at ($\approx 6X_0$) to get maximum EM shower from a charged particle. This layer is important for our study as it provides precise electron energy measurement. The outermost calorimeter is the Central Hadronic Calorimeter, consisting of a central part ($|\eta| < 0.6$) with 32 layers and wall part ($0.6 < |\eta| < 1.1$) with 15 layers of $4.7X_0$ steel. The Hadronic calorimeter is also very crucial for W mass measurement as it measures the hadronic recoil. To detect the particle within $1.1 < |\eta| < 3.6$ is a barrel plug consisting of both EM and Hadronic Calorimeter. After traversing through a 60 cm thick steel shield, a daughter particle can make it to a system muon detector covering the region until $|\eta| < 1$ (the muon detectors are not shown in Figure 7.1).

7.2 Direct W Mass Measurement

The proton-antiproton ($p\bar{p}$) collisions at Tevatron produced $\sim 80\%$ of W boson via annihilation of valence quark (from p) and valence anti-quark (from $\bar{p}$). Only one-third of them decay leptonically (with one-nine for each lepton flavor) in $W \rightarrow l\nu$. Among the three leptonic channels, only electron and muon channels are chosen for clean signal as the tau leptons can further decay hadronically as well. These events are accompanied by QCD background from the remnants of the $p\bar{p}$ collisions. The charged lepton can be measured in the detector but the neutrinos escape the detector. The missing energy due to the neutrinos is estimated by taking into account these QCD background jets along with the charged lepton in the energy balance equation. Due to incomplete η coverage of the detector, as mentioned above, a substantial longitudinal momentum is lost. For this reason, only the transverse quantities like charged lepton transverse momentum (p^l_{T}) and hadronic recoil momentum (u_T), which is the total transverse QCD background momentum, can be measured, which in principle could be used to obtain the transverse momentum (p^ν_T) of neutrinos (the *missing transverse energy*). But, for a better estimate of the missing transverse energy, only candidate events with $u_T \ll p^l_T$ are selected, since the momentum resolution of leptons is better than jets in an event. This selection makes $p^\nu_T \approx p^l_T + u_{||}$ ($u_{||}$ is the component of u_T parallel to p^l_T). Using the missing transverse energy and the charge lepton's transverse momentum, the transverse mass of W boson, analogous to its mass, can be obtained as $m_T = \sqrt{2p^l_T p^\nu_T (1 - cos\Delta\phi)}$ (where $\Delta\phi$ is the angle between l and ν). The distribution of all these transverse quantities - p^l_T, p^ν_T and m_T, have Jacobian edges which is related to the on-shell W boson mass (M_W).

To understand the reason for a Jacobian edge in these transverse quantities, the

Collin Soper frame (the W rest frame) is useful. As shown in Figure 7.2, assuming the masses of l and ν to be negligible, the transverse momenta of l and ν become $|p_T^l| = |p_T^\nu| = \frac{M_W}{2}$.

Hence, proceeding with the p_T^l distribution: $\frac{d\sigma}{dp_T^l} = \frac{d\sigma}{dcos\theta_{CS}} \frac{dcos\theta_{CS}}{dp_T^l}$.

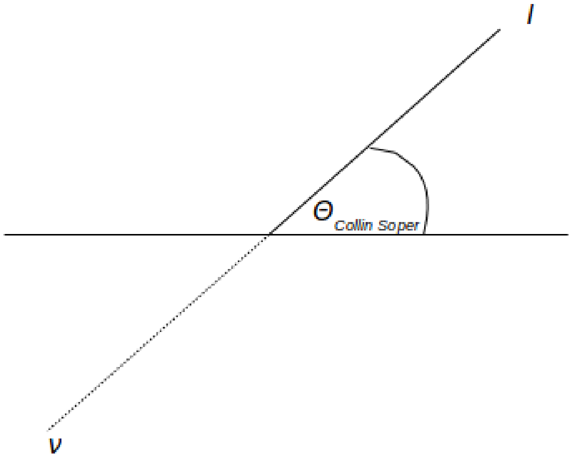


Figure 7.2: Collin Soper frame (W rest frame) where l and ν are back-to-back making $\theta_{CollinSoper}$ (θ_{CS}) with the lab frame W momentum.

Substituting p_T^l, $\frac{d\sigma}{dp_T^l} = \frac{d\sigma}{dcos\theta_{CS}} |\frac{dcos\theta_{CS}}{dsin\theta_{CS}}| \frac{2}{M_W} = \frac{d\sigma}{dcos\theta_{CS}} \frac{p_T^l}{\sqrt{\frac{M_W^2}{4} - p_T^2}}$. Thus, at $p_T^l = \frac{M_W}{2}$ there is a singularity in p_T^l distribution. But, since M_W is smeared by Breit Wigner distribution, the p_T^l or p_T^ν distributions only have a Jacobian edge at $\frac{M_W}{2}$. Consequently, m_T also has a Jacobian edge at M_W, which makes these three transverse quantities useful for M_W measurement. To model the line shapes of p_T^i, p_T^ν and m_T distributions, parametrized Monte Carlo simulation is used. A template with M_W between 80 and 81 GeV is generated for each of the transverse parameters, and the M_W is then extracted using a binned likelihood fit. Finally, the p_T^i, p_T^ν and m_T distributions of both $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$ are combined to measure M_W.

With this method, CDF II detector measured $M_W = 80387 \pm 12_{stat} \pm 14_{syst}$ MeV $= 80387 \pm 19$ MeV (see Table 7.1) [20], using $\sqrt{s} = 1.96$ TeV $p\bar{p}$ at an integrated

Table 7.1: Uncertainties in units of MeV on the final combined result on M_W [20].

Source	**Uncertainty**
PDFs	10
Lepton energy scale and resolution	7
Recoil energy scale and resolution	6
$p_T(W)$ model	5
Photon radiation	4
Backgrounds	3
Lepton tower removal	2
Statistical	12
Total	19

luminosity of $2.2 fb^{-1}$, which is its latest most precise measurement of M_W. The improvements in PDF uncertainty estimation presented in this chapter has been used in the M_W measurement with full Run 2 data set of CDF II.

7.3 PDF Uncertainty

As shown in Table 7.1, the biggest systematic error in M_W measurement happens to be from the Parton Density Functions (PDFs) of the interacting valence partons. For measurement of M_W, only transverse quantities like p_{T}^{l}, p_{T}^{ν} and m_T are used (Section 7.2). PDFs $f_i(x, Q^2)$ gives the probability of finding in the hadron, a parton of flavor i (quarks or gluons) carrying a fraction x of the hadron momentum with Q being the energy scale of the hard interaction [88], hence, only the longitudinal momentum of the W is determined by the momenta of interacting partons. However, due to incomplete rapidity coverage of the detector (Section 7.1), some particles are lost in the forward region through the gaps between the detector and the beampipe. Due to this loss of particles, when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed using the transverse momenta of all the detector particles, it is smaller than the actual $E_{\mathrm{T}}^{\mathrm{miss}}$. This makes the aforementioned transverse quantities dependent on the longitudinal boost of the W though detector cuts and kinematic cuts on the leptons.

Since QCD does not predict the parton content of the proton, the shapes of the PDFs are determined by a fit to data from experimental observables in various processes, using the DGLAP evolution equation [93, 94, 95, 96, 97]. DIS data from fixed-target lepton-nucleon scattering experiments at SLAC, FNAL, CERN and from the electron-proton HERA collider at DESY are mainly used for this purpose. For flavor decomposition of the quark and anti-quark sea or the gluon distribution at large x some additional physical processes which are used in the fits are the single jet inclusive production in nucleon-nucleon interactions for gluon distribution, di-lepton production in the virtual photon Drell-Yan process for sea quark distribution and electroweak Z and W boson production for the up and down quark and anti-quark distributions. Most global fits like CTEQ6.6[98], MSTW08[99], GJR08[100] and NNPDF3.0[101] use data from DIS and proton-nucleon fixed target experiments along with results from the HERA and Tevatron colliders.

The PDFs are first parametrized using these experimental DIS cross-sections at some low $Q^2 = Q_0^2$, which is large enough that the unknown terms of the perturbative equations are assumed to be negligible. Then they are evolved up in higher Q^2 using DGLAP equations. The number of unknown parameters is typically between 10 and 30. The factorization theorems[102], allow to derive predictions for the cross sections. These predictions are then fitted to as much of the experimental data together as possible, to determine the parameters and to provide parton distributions.

The error accrued in these parameters of PDFs are propagated to the observables such as M_W . During the period of study, it was tried to systematically understand the classical method used in such error propagation for most PDFs other than NNPDFs, with the spirit of further investigating any possibility of improvement in reducing the error $\delta M_W^{PDF} = \frac{1}{2}\sqrt{\sum_i (M_W^{i+} + M_W^{i-})^2}$. $M_W^{\pm}$ is obtained using $\pm n\sigma$ in the i^{th}

eigenvector[1]. This method is referred as the *Hessian Method*.

This method can be understood by considering two vector spaces ΔB and ΔO. ΔB is defined here as the error phase space in PDF. Hence, the error in parameters $\vec{b} = \{b_i\}$, defined here as $\delta\vec{b} = \{\delta b_1, \delta b_2, ..., \delta b_N\}$ is a vector in ΔB. As these parameters are independent, the dimensionality of the $\Delta B \geqslant N$. Also, the origin of this phase refers to zero error, i.e. when the $\vec{b}$ minimize the χ^2 of observable O. The δO space is the error phase space in the observable O, which is homomorphic to space δB. The entire objective of the procedure is to find this vector space δO through the homomorphism with δB, so that a volume (error ball) in δB can be translated to a volume (error ball) in δO, thus propagating the error in O. In other words, one has to find $\{E_i\}$, the span of δO, such that $E_i = O(e_i)$, where $\{e_i\}$ is the span of δB.

This is where the Hessian Matrix is introduced. Hessian matrix of $\chi^2(O(\delta\vec{b}))$ is $H^{\chi^2(O)}_{\delta b_i \delta b_j} = \frac{\partial^2 \chi^2(O)}{\partial \delta b_i \partial \delta b_j}$. On Taylor expansion of $\chi^2(O(\vec{b_o} + \delta\vec{b}))$ about its minimum,

$$\chi^2(O(\vec{b_o} + \delta\vec{b})) = \chi^2(O(\vec{b_o})) + \frac{1}{2}\delta\vec{b}^T H \delta\vec{b} + \mathcal{O}(|\delta\vec{b}|^3) \tag{7.1}$$

As, δb is small, the higher order terms in Equation 7.1 can be neglected. The properties of Hessian matrix H can be used to establish homomorphism between δO and δB. H is a N dimensional real symmetric matrix. Hence, a complete set of normalized eigenvectors e_i as the span of δB can be chosen, such that, $E_i = O(e_i)$. Thus now the $n\sigma$ error in any e_i can be propagated, and thus the radius of the spherical envelope covering the entire error volume in δO is the error bar on the observable O.

7.3.1 MC Samples

The MC sample of $W \rightarrow e\nu_e$ events were produced in generator-level at at NLO using POWHEG method and showered using Pythia within the POWHEG-BOX

---

[1] If $M_W^{\pm}$ have same sign, then $\delta M_W^{PDF} = \frac{1}{2}[max(M_W^{i+} - M_W, M_W^{i-} - M_W)]$

Table 7.2: PDF sets used for PDF uncertainties in W mass.

PDF Set	#Error PDFs	#Central PDF
CT10	52	1
CT10(NLO)	52	1
CTEQ6	40	1
MSTW2008(NLO 68% c.l.)	40	1
MSTW2008(NLO 90% c.l.)	40	same as (NLO 68% c.l.)
NNPDF3.0 NLO ($\alpha_s = 0.119$)	100	1

v2 [103, 104, 105] framework. In those events, CT10 central PDF was used for the two incoming partons at $\sqrt{s} = 1.96TeV$. In order to incorporate the kinematics due other PDFs, LHAPDF6[101] framework was used to calculate $f_i(x, Q^2)$ (here i refers to the two incoming partons) using each PDF in Table 7.2, for all the events. Any event when scaled by the weight $\prod_{i=1}^{2} \frac{f_i^{testPDF}(x,Q^2)}{f_i^{CT10}(x,Q^2)}$, would then give the kinematics for the $testPDF$. This scaling applied on the observables like $M_T W$, $p_T e$ and $p_T \nu$ to study their respective variations due to error PDFs of each group, for further investigation.

7.3.2 Analysis of PDF Variations

The shape of the distributions of transverse kinematic variables, denoted here as O (where O can be m_T, p_{T}^{e} and p_{T}^{ν}) of the W boson and its decay products vary with the choice of PDFs. Therefore, any systematic variation in the PDF would lead to a variation in the shape of the O distribution. To understand the shape variation of these distributions due to systematic variation in the PDFs, the difference plot (or distribution) is defined as the difference of the normalized O distribution using an error eigenvector PDF from the normalized O distribution using the central (nominal) PDF of a PDF set. These difference plots essentially give the shape variation of O for an error PDF with respect to its central PDF. The difference plot for m_T distribution for PDF set of CTEQ6, CT10, CT10W, MSTW2008 (NLO 68% CL), MSTW2008

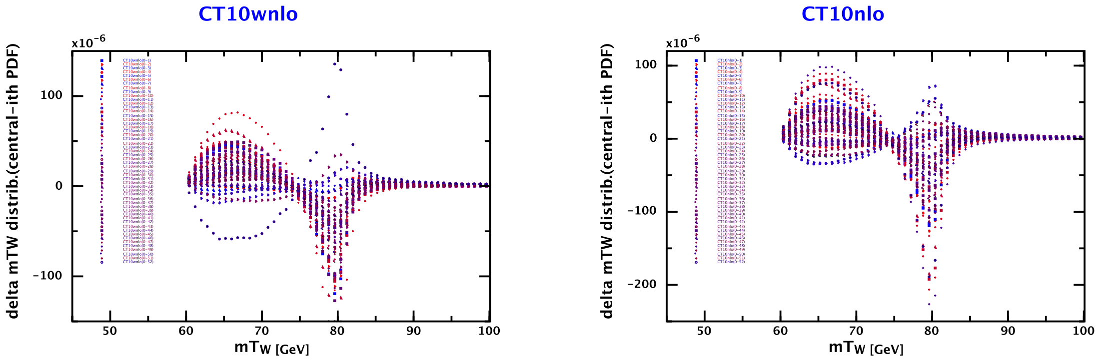
(NLO 90% CL), and NNPDF3.0 NLO($\alpha_s = 0.119$) is shown in Figure 7.3.

These m_T difference plots show that all error PDFs from every PDF set share a common inflection point at around $m_T = 72.8$ GeV. This is an interesting feature of the difference plots (similar feature observed in p_{T}^{e} and p_{T}^{ν} in Appendix A) since the inflection point is close the Jacobian edge of the m_T distribution and is independent of the choice of PDF. For the original m_T distribution, it means that any variation in PDF only leads a change in the fraction of events on the left and right side of the inflection point in the m_T range. Therefore, a ratio $s_{p^{th}\ PDF_j} = f_{left}/f_{right}$, where f_{left} is fraction of events with O less than the infection point and f_{right} is the fraction of events with O greater than the inflection point, characterizes the shape of the O distribution for p^{th} eigenvector (or replica for NNPDF) of PDF_j PDF set. The relative difference of the $s_{p^{th}\ PDF_j}$ (for p^{th} eigenvector) from $s_{central\ PDF_j}$ (for central eigenvector) gives the O shape variation due to p^{th} error eigenvector of the PDF_j PDF set. This ratio is termed as the 'variation parameter' $i(O(p^{th}error\ e{-}vector\ PDF_j))$:
$i(O) := 1000X\left\{\frac{s_{p^{th}\ PDF_j}}{s_{central\ PDF_j}} - 1\right\}$

A 'scale factor' b, which is defined as $b := 1000X\left(\frac{\bar{O}(p^{th}PDF_j)}{\bar{O}(CentralPDF_j)} - 1\right)$, gives an estimate of the relative shift in Jacobian edge of the O distribution due to p^{th} error eigenvector (or replica in case of NNPDF). The scale factor b therefore gives an estimate of the variation in W boson mass M_W from the O distribution. In order to propagate the error due shape variation encoded in $i(O)$, the relation between scale factor(b) and $i(O)$ is studied and the results are discussed in the following section.

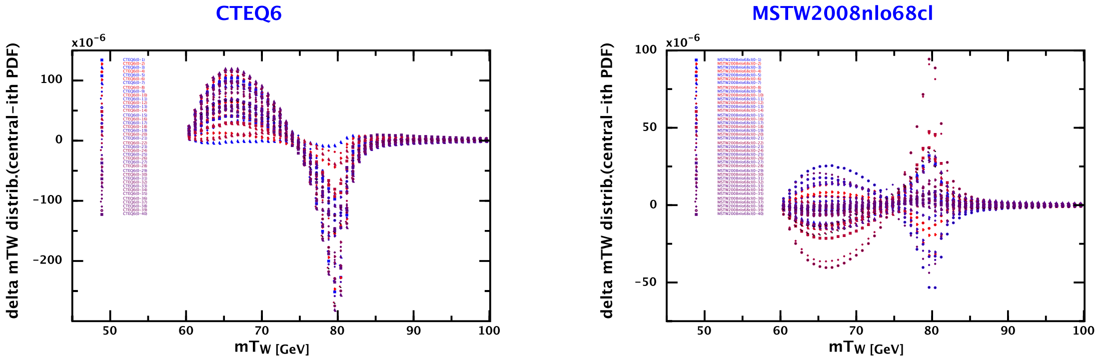
7.4 Results

The statistics of 80 million events were used in the following distributions in Figure 7.3 (including Appendix A. Using binned likelihood template fitting (mentioned in the



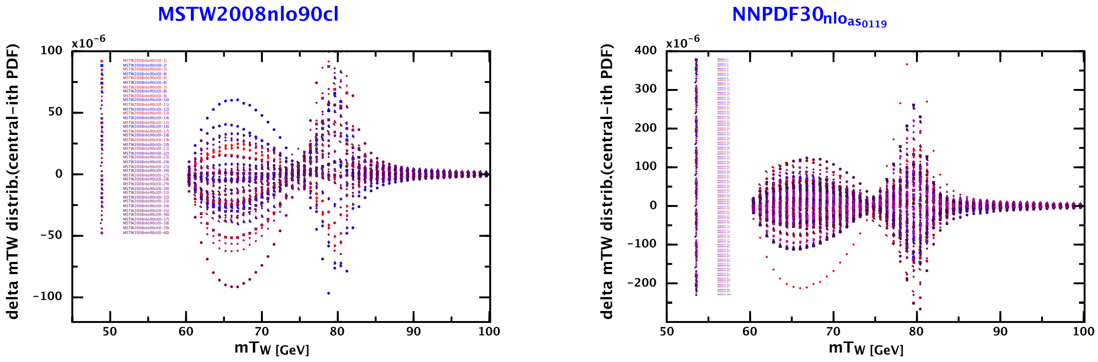
(a) CT10W (m_T)

(b) CT10 (m_T)



(c) CTEQ6 (m_T)

(d) MSTW2008 (NLO 68% CL) (m_T)

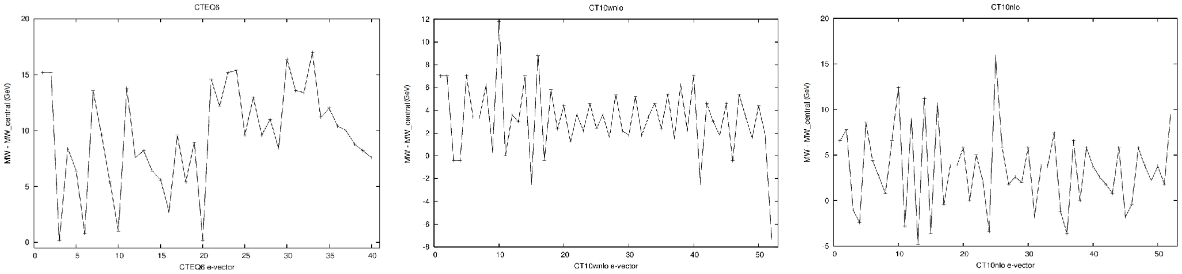


(e) MSTW2008 (NLO 90% CL) (m_T)

(f) NNPDF3.0 (m_T)

Figure 7.3: The transverse W mass (m_T) difference plots for PDF sets of CT10W, CT10, CTEQ6, MSTW2008 (NLO 68% CL), MSTW2008 (NLO 90% CL), and NNPDF3.0 NLO ($\alpha_s = 0.119$). The inflection point at 72.8 GeV is near the Jacobian edge near M_W.

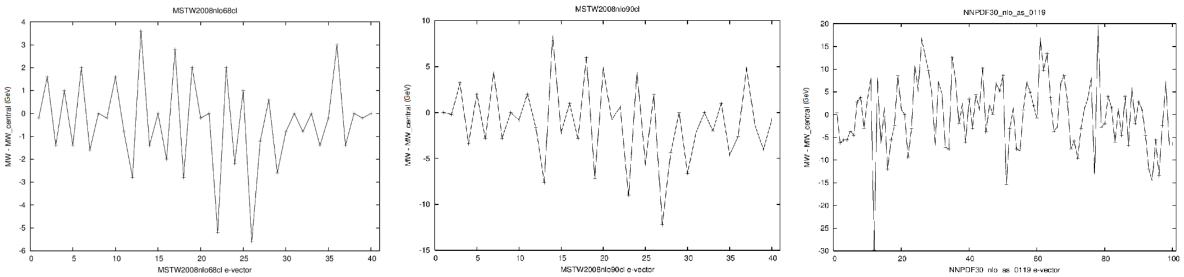
last paragraph of Section 7.2), M_W for each PDF (in Table 7.2) was calculated. Error in M_W for each error eigenvector is calculated for each group to calculate its combined error of the group, as discussed below. The CTEQ PDF groups - CTEQ6 (40 error e-vectors) Figure 7.4a, CT10W (52 error e-vectors) Figure 7.4b & CT10 (52 error e-vectors) Figure 7.4c show positive error in M_W.



(a) CTEQ6 group (b) CT10W group (c) CT10 group

Figure 7.4: Error in M_W due to error eigenvectors for each of the CTEQ groups respectively.

In contrast, the MSTW PDF groups - MSTW2008(NLO 68% c.l.) (40 error e-vectors) Figure 7.5a & MSTW2008(NLO 90% c.l.) (40 error e-vectors) Figure 7.5b along with the NNPDF3.0 NLO ($\alpha_s = 0.119$) (100 error e-vectors) Figure 7.5c give a symmetric deviation in M_W from their respective central PDF values.



(a) MSTW2008(NLO 68% c.l.) group (b) MSTW2008(NLO 90% c.l.) group (c) NNPDF3.0 NLO($\alpha_s = 0.119$) group

Figure 7.5: Error in M_W due to error eigenvectors for each of the MSTW2008(NLO) and NNPDF groups respectively.

The combined error in NNPDF3.0 NLO($\alpha_s = 0.119$) is calculated by taking root mean square of the 100 ensemble PDFs, while for rest of the PDF sets, the combined errors were calculated using Hessian Method Table 7.3:

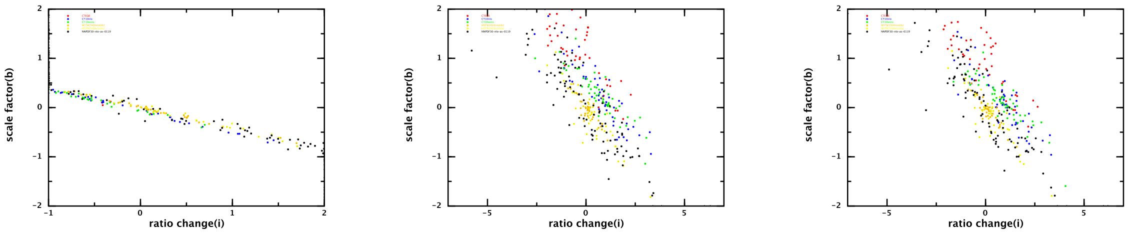
Table 7.3: Combined error in M_W for each PDF set.

PDF set	M_W error	$M_W^{central}$
CTEQ6	13.26	80449.9
CT10	15.23	80453.7
CT10W	11.00	80459.9
MSTW2008 (NLO 68% CL)	6.32	80452.7
MSTW2008 (NLO 90% CL)	15.10	80452.7
NNPDF3.0	7.74	80458.3

From difference plots in Figure 7.3, the inflection point, upper and lower limits of each distribution used:

Table 7.4: Limits used to calculate the variation parameter $i(O)$.

Observable	Lower Limit (GeV)	Inflection Point (GeV)	Upper Limit (GeV)
m_T	60	72.8	100
p_T^e	30	41.5	55
p_T^ν	30	41.5	55



(a) For m_T (b) For p_T^e (c) For p_T^ν

Figure 7.6: Correlation between the scale factor b and the variation parameter $i(O)$(using Table 7.4).

7.5 Conclusions and Future Directions

There is a linear correlation between the scale factor (b) and $i(O)$ parameter as shown in Figure 7.6. Since, the $i(M_T^W)$ parameter encodes the change in shape of the M_T^W distribution due to PDF variations, it is related to the systematic uncertainty in the mass of W (ΔM_W) due to PDF variations. As the computation of b and $i(M_T^W)$ is fairly simple compared to the binned likelihood fit method for obtaining M_W and its systematics, this linear correlation between the b and $i(M_T^W)$ can be exploited to obtain a fair and computationally inexpensive estimate of the PDF systematics in M_W, if the functional relationship between ΔM_W and $i(M_T^W)$ can be obtained.

Chapter 8

Top-philic Vector Resonance Search

With the discovery of the new particle consistent with the SM Higgs boson [33, 34], the SM is complete. However, the underlying nature of electroweak symmetry breaking (EWSB) still remains unknown. Naturalness arguments require that quadratic divergences that arise from radiative corrections to the Higgs boson mass must be canceled by some mechanism beyond the SM in order to avoid fine-tuning. Such mechanisms have been proposed in many Beyond-the-Standard Model (BSM) theories, such as composite Higgs scenarios [106, 107, 108, 109], which predict new truly top-philic vector resonances. This search is done by an analysis team of less than 10 members, where the author contributed mainly in the improvement of the spectator top quark reconstruction by adapting and improving the buckets of tops algorithm which has been discussed in detail from Section 8.2 onward.

8.1 Search for Four Top Quarks Final State

Truly top-philic (exclusively couples to top quarks) vector resonances cannot couple to light quarks to be produced in the s-channel, or via gluon-gluon fusion (ggF) mechanism (due to spin considerations [110, 111]). Therefore, such resonance Z' can only be produced in association with $t\bar{t}$ pair, with the $Z' \rightarrow t\bar{t}$, leading to a four-top quark final state as shown in Figure 8.1

In this analysis, one of the top quarks (preferably a spectator top quark) is required to decay leptonically while the other three top quarks decay hadronically. Using ATLAS $\sqrt{s} = 13TeV$ p-p collision data with an integrated luminosity of 139 fb^{-1}, events with exactly one lepton (e or μ with $p_{\mathrm{T}} > 28$ GeV) several small-R jets and

b-tagged jets are selected. Tau leptons are not explicitly considered in this analysis although some of their decay products can be accepted by the analysis selection criteria.

In this analysis, the invariant mass spectrum of the resonance is inspected for a model-independent search. The top quarks considered to be originating from the decay of the Z' resonance are reconstructed using jet re-clustering [112] with fixed or variable radius (currently being investigated). For reconstruction of the spectator $t\bar{t}$ pair, where the usual top taggers do not work, a combinatorial approach - “Buckets of Tops” algorithm is investigated, which is the main focus of this chapter.

8.2 Modified Buckets of Tops Algorithm

In the above process, due to heavy Z’ resonance, most of the tranverse momentum goes to top quark pairs from the decay of the resonance (referred as “resonance $t\bar{t}$”. As a result, the spectator top quarks are softer with decay products to be at a larger angular separation from each other, such that they are often not contained within a cone or even a hemisphere. The current top taggers such as HEPTOPTAGGER cannot go beyond Cambridge-Aachen jet of size R=1.5 ($p_{\mathrm{T}} \leqslant 200$ GeV) [113].

For this reason, the combinatorial algorithm known as the “Buckets of Tops”

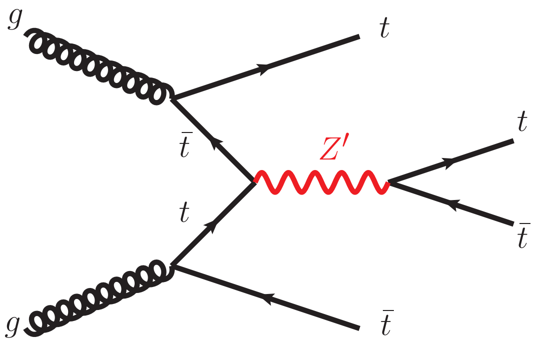


Figure 8.1: Leading Feynman diagrams for the production of a vector resonance that exclusively couples to top quarks.

algorithm appears to be a good candidate for reconstructing the spectator top quarks in the analysis. This algorithm searches through all possible combinations of small-R jets' collection to find ones which best resemble the top quark pairs.

8.2.1 Original Buckets of Tops Algorithm

The Buckets of Tops algorithm [16, 17] does not use any cone to contain the collection of jets (being completely combinatorial), the algorithm works best for moderately boosted top quarks ($p_{\mathrm{T}} \in$ [100 GeV, 400 GeV]) compared to avilable top quark taggers. However, this algorithm has been designed and optimized for identifying and reconstructing hadronically decaying top anti-top quark pairs. The hadronic $t\bar{t}$ version of the Buckets of Tops algorithm, as described below, was able to recontruct top quarks in generator level efficiently (see Figure 8.2):

- **step 1** Complete top quark ('w') bucket search
  - distribute jets into three buckets - B1, B2 and B$_{ISR}$, such that each has 1 b-jet and minimize

$$\Delta^2 = B1wt \cdot \Delta^2_{B1} + \Delta^2_{B2} \tag{8.1}$$

$$\Delta_{B1(2)} = |m_{B1(2)} - m_{top}| \tag{8.2}$$

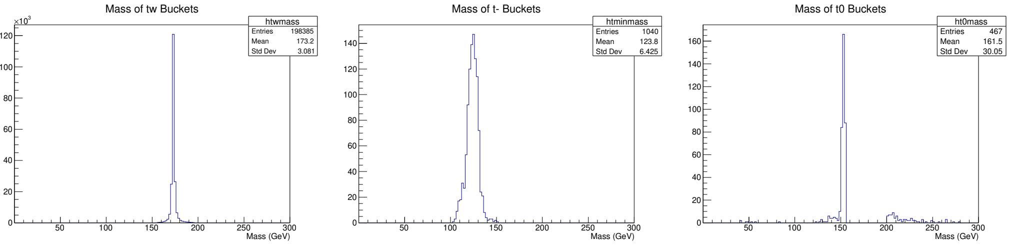
  , where $B1wt = 100$

  - if $|\frac{m_{any\ jet\ pair}}{m_{B_i}} - \frac{m_W}{m_{top}}| < 0.15$ then 'tw' else proto-'t-'
  - if bucket mass not within (155 GeV, 200 GeV) then 't0'
- **step 2** Incomplete top quark ('t-') bucket search
  - exclude all jets in 'tw' buckets

- if event has one proto-‘t-’ bucket, find the b-jet quark-jet pair with minimum Δ_B^{bj}, where:

$$\Delta_B^{bj} = \begin{cases} |m_B - 145| & \text{if } m_B < 155 \text{ GeV} \\ \infty & \text{else} \end{cases}$$

- if event has two proto-‘t-’ buckets, find two b-jet quark-jet pairs with minimum $\Delta_{B1}^{bj} + \Delta_{B2}^{bj}$
- if bucket mass within (75 GeV, 155 GeV) then ‘t-’ else ‘t0’
- any leftover jet is put in extra bucket ‘tX’



(a) tw bucket mass (b) t- bucket mass (c) t0 bucket mass

Figure 8.2: Mass distributions of buckets recontructed by the original hadronic buckets [16, 17] of tops algorithm on generator level $t\bar{t}$ all-hadronic samples (produced using $MadGraph5\_aMC@NLO$ generator).

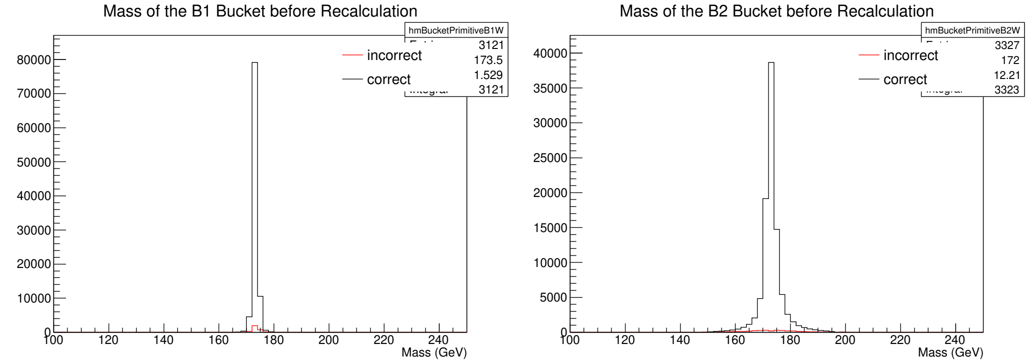
In step 1 of the buckets of tops algorithm, an asymmetry is introduced between buckets B1 and B2 while optimizing the “bucket metric” Δ in Equation 8.1, by setting $B1wt = 100$. This asymmetry puts a tighter constraint on the mass of bucket B1 (to be closer to top quark mass) than B2.

8.2.2 Effect of Detector Smearing

To study the performance of this algorithm in detector level events (detector smearing in jets originating from the decay products of W bosons), detector smearing was simulated on a parton level dataset of $MadGraph5\_aMC@NLO$ $t\bar{t}$ events at $\sqrt{s} = 13$ TeV by adding a gaussian error $\mathcal{N}(0, a \times \sqrt{E})$, where E is the energy of the parton. The value of a determines the degree of detector smearing on a particle. The $p_{\mathrm{T}} > 30$ GeV and the $|\eta| < 5.0$.

8.2.2.1 Choice of $B1wt$

For generator level all-hadronic $t\bar{t}$ sample (with no detector smearing of the momenta of top quark decay products), the reconstructed buckets should have masses very close to the top quark mass (Figure 8.3), and the effect of the aforementioned asymmetry due to $B1wt = 100$ is not apparent.



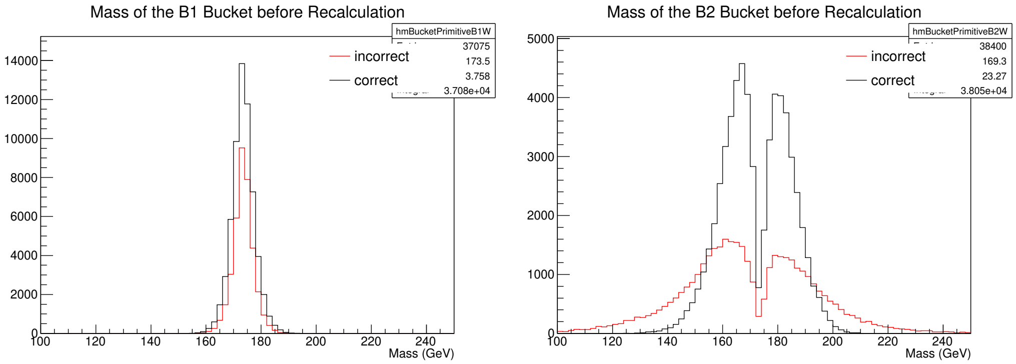
(a) B1 mass

(b) B2 mass

Figure 8.3: Generator-level mass distributions of B1 and B2 buckets in step 1, ie before W tagging and bucket labeling at $B1wt = 100$. Since there is no detector smearing in generator level, the mass of the buckets very close to m_{top}. Due to $B1wt = 100$ (larger constraint on Δ_1, the mass distribution of B1 is narrower than B2.

While, in the case of ATLAS detector-level all-hadronic $t\bar{t}$ events (where hadronic calorimeter smears the momenta of top quark decay products), this asymmetry between B1 and B2 leads to a dip (instead of a peak) at m_{top} for B2 bucket (Figure 8.5).

This m_{top} dip in B2 mass distribution emerges essentially due to detector effects, which can be recreated in generator level as well by introducing a toy smearing on



(a) B1 mass

(b) B2 mass

Figure 8.4: A toy smearing of $\mathcal{N}(E, \sqrt{E})$ is applied on the same top quark decay products used in 8.3. Generator-level (with momenta smeared partons) mass distributions of B1 and B2 buckets in step 1, ie before W tagging and bucket labeling at $B1wt = 100$. Since the recontructed top quark masses may no longer be very close to m_{top}, the same effect of m_{top} dip in B2 mass distribution, as in ATLAS events in Figure 8.5, appears with truth-level event when a toy smearing is introduced. Thus, the m_{top} dip in the B2 mass distribution is essentially due to detector smearing of the top decay products.

final state partons (Figure 8.4).

A mass peak at m_{top} for B2 mass distribution can be recreated, however, when the $B1wt$ is tuned to exactly 1 (Figure 8.6), as $B1wt = 1$ makes B1 and B2 symmetric in Equation 8.1.

In other words, there is an implicit sorting applied between Δ_{B1} and Δ_{B2} (Equation 8.2), when the $B1wt \neq 1$. If an explicit sorting is applied, ie $\Delta_1 < \Delta_2$, B2 mass dip at m_{top} would appear even for $B1wt = 1$ (Figure 8.7).

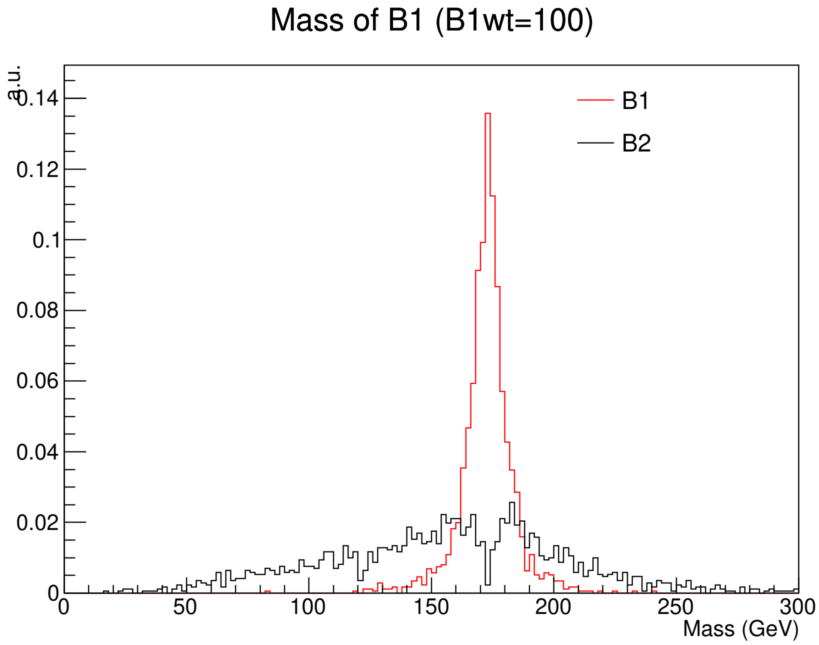


Figure 8.5: Mass distribution of B1 and B2 buckets (in step 1 before W tagging and labeling of buckets) with $t\bar{t}$ (all hadronic)+ISR jets events at $B1wt = 100$. Due to detector smearing, the reconstructed top quark masses need not be closer to m_{top}, so the lower Δ_i bucket (Equation 8.2) is made a B1 bucket.

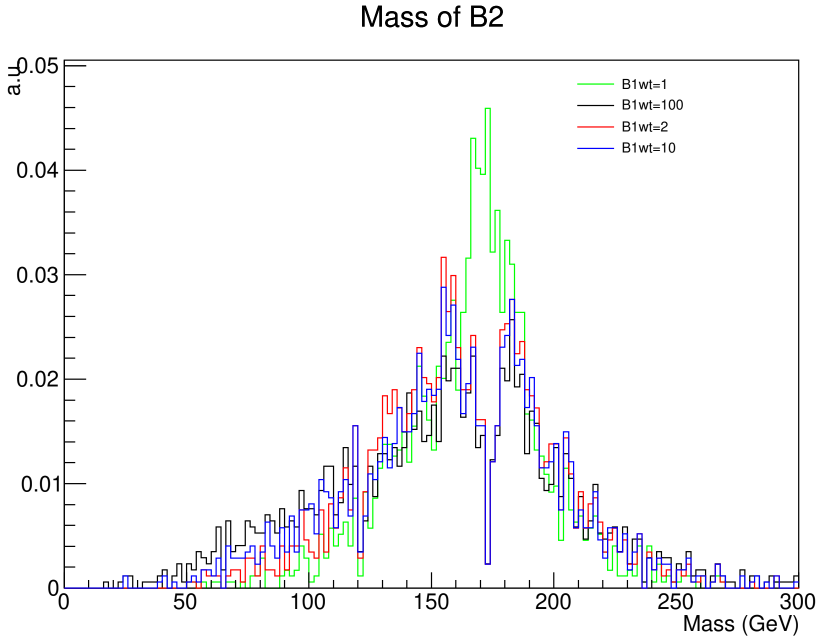


Figure 8.6: Mass distribution of B2 bucket (in step 1 before W tagging and labeling of buckets) with $t\bar{t}$ (all hadronic)+ISR jets events. Only when the $B1wt = 1$, ie when both buckets B1 and B2 become symmetric, the peak at m_{top} appears for B2 mass distribution.

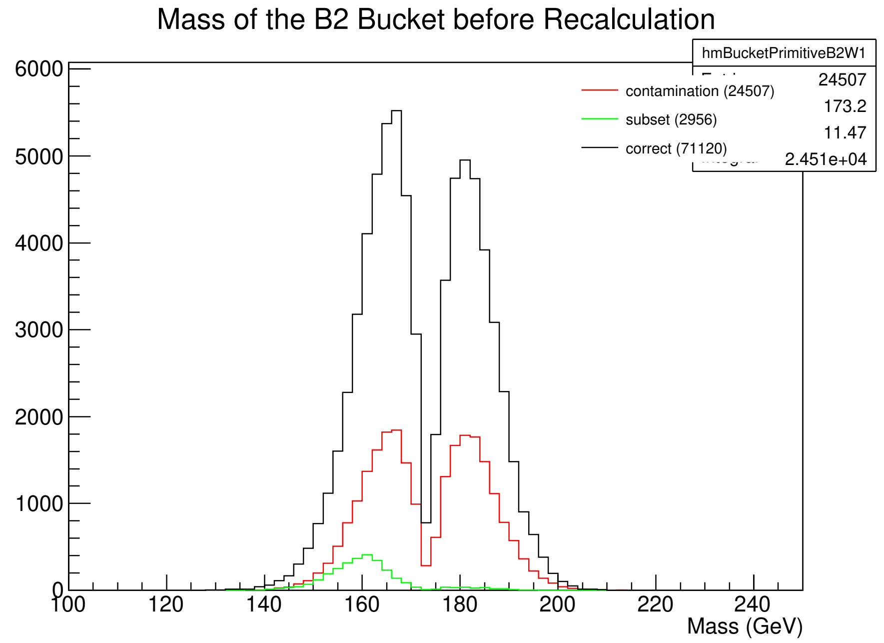


Figure 8.7: Mass distribution of B2 bucket (in step 1 before W tagging and labeling of buckets) with $t\bar{t}$ (all hadronic)+ISR jets events. The dip at B2 mass $= M_{top}$ reappears even at $B1wt = 1$ when explicit sorting is applied between Δ_{B1} and Δ_{B2} such that $\Delta_{B1} > \Delta_{B2}$ (dip at B2 mass$= M_{top}$ absent without explicit sorting in Figure 8.6). This proves that the M_{top} dip in B2 mass distributions for $B1wt > 1$ was due to implicit sorting of Δ_{Bi} when detector smearing is present.

$B1wt$ not only introduces an implicit sorting between Δ_{B1} and Δ_{B2}, but also takes into account the relative smearing of the jet constituents in B1 with respect to B2. If this sorting between Δ_{B1} and Δ_{B2} is explicitly applied, $B1wt$ would just be a hyper-parameter of the algorithm which can be tuned to increase signal efficiency for a given detector smearing.

8.2.2.2 Including the W tagging parameter in the bucket metric

When detector smearing is applied to the parton-level top decay products, the maximum efficiency of the algorithm is significantly smaller (more $\sim 10\%$) than without any detector smearing (as shown in Figure **??**).

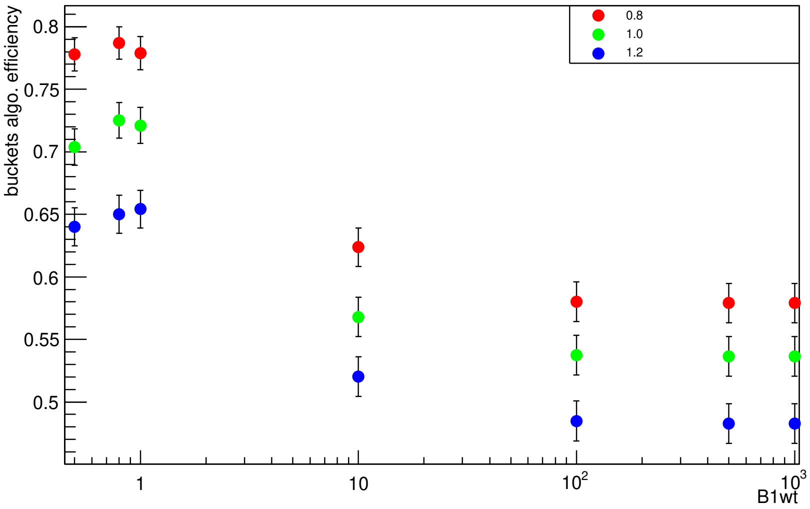


Figure 8.8: Algorithmic efficiency of step 1 of the original algorithm with respect to B1wt after detector smearing applied (smearing four momenta with $\mathcal{N}(0, a \times \sqrt{E})$, where the values of a is depicted in the legend).

To investigate it further, the parton level $t\bar{t}$ pairs which failed to be reconstructed correctly after applying detector smearing were tracked. The $t\bar{t}$ decay products which could be reconstructed correctly using the algorithm were termed as ‘good’, while the pairs which could not be reconstructed correctly were termed as ‘steal’. Hence, the parton level $t\bar{t}$ pairs which would be correctly reconstructed before applying detector smearing but could not be recontrsucted correctly after detector smearing were labelled as ‘good-to-steal’ cases, while the pairs which could still be correctly reconstructed after applying detector smearing were leabelled as ‘good-to-good’ cases.

By plotting the mass of the non-b jets in the respective buckets versus mass of the buckets, as seen in Figure 8.9, it was found that the step 1 of the original algorithm was not putting sufficient constraints on the buckets to ensure the collection of W decay products in the buckets. Rather, the majority of the constraint in the global metric was to make the bucket masses as closes to the top quarks as possible. So, the global metric of the original algorithm can be thought of as the ‘top mass tagging

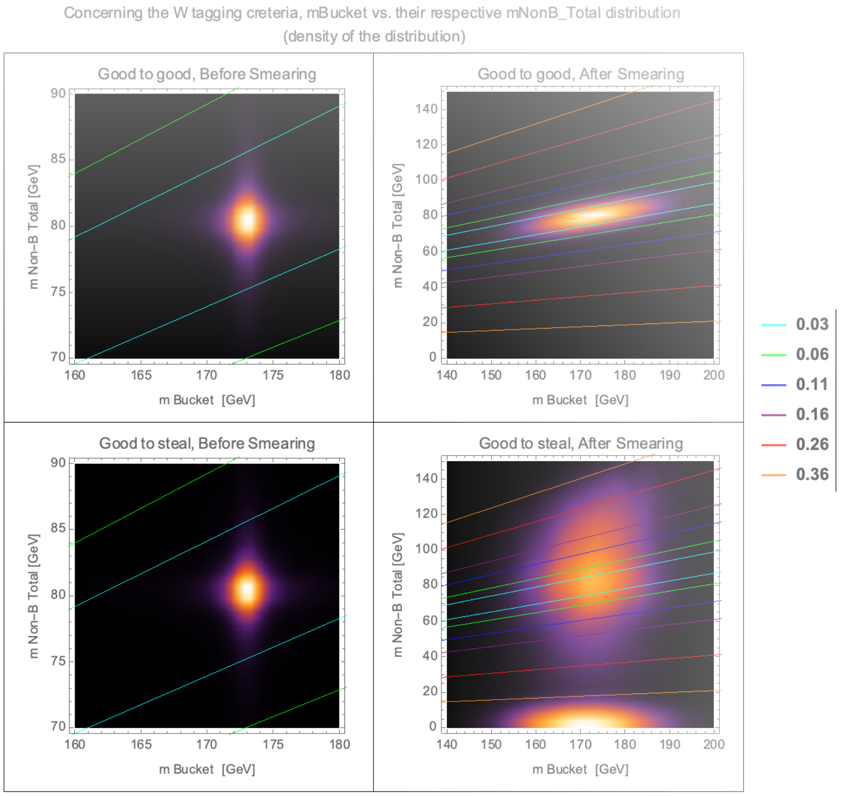


Figure 8.9: Density plot of the mass of the non-b jets in the respective buckets versus mass of the buckets of the classified events. Color lines are the contours of the W parameter.

metric'. This lead to wrong combination of jets mimicking closer masses to the top quark forming final buckets accidentally, when the correct combination might have a larger deviation from top quark mass due to detector smearing.

The solution was to incorporate the W parameter ($|\frac{m_{any\ jet\ pair}}{m_{B_i}} - \frac{m_W}{m_{top}}|$) in the global metric to select combinations of jets as buckets which are more close to physical top quarks. As discussed above, the $B1wt = 1$ gives optimum performance for hadronic $t\bar{t}$ reconstruction, that lifts the asymmetry between B1 and B2. As, a result the respective W parameter and the 'top mass tagging metric' are symmetric to each other. So, a new global metric is defined as a quadrature sum of the standardized W parameters (Δ_W of the buckets algorithm) and the quadrature sum of the top mass tagging metrics (Δ_{top} of the buckets algorithm), and combined together with a pre-factor called Wwt as shown in the modified "Buckets of Tops" algorithm below:

$$\Delta^2 = \Delta^2_{top} + Wwt \times \Delta^2_W$$

and,

$$\Delta^2_{top} = \Delta^2_{top,B1} + \Delta^2_{top,B2}$$

$$\Delta^2_W = \Delta^2_{W,B1} + \Delta^2_{W,B2}$$

where, $\Delta^2_{top,Bi}$ is the top mass tagger metric and $\Delta^2_{W,Bi}$ is the W parameter of the Bi bucket (i being 1 or 2). The value of Wwt is chosen such that it maximizes the algorithmic efficiency.

8.2.3 Modifications for Semi-leptonic $t\bar{t}$ Pair

In the event topology of the signal in the top-philic heavy resonance search, one of the spectator tops decays leptonically while the rest of the three tops decay hadronically. Therefore, the algorithm should be modified to reconstruct semi-leptonic top quark pairs. Since complete decay products of W boson would not be present in the lepton top quark bucket due to missing transverse energy from neutrino, but would only contain the charged lepton (can be jets in the case of τ leptons), no W metric (Δ_W) is defined for leptonic bucket. With the above modification, the "Buckets of Tops" algorithm can be defined for semi-leptonic $t\bar{t}$ pairs as follows:

$$\Delta^2 = \Delta^2_{top} + Wwt \times \Delta^2_W$$

and,

$$\Delta^2_{top} = \Delta^2_{top,had} + \Delta^2_{top,lep}$$

$$\Delta^2_W = \Delta^2_{W,had}$$

where the top bucket metrics are:

- $\Delta_{top} = \left(\frac{m_{bucket} - m_{top}}{(std.\ dev.\ of\ numerator)}\right)$
- $\Delta_{W,had} = \left(\frac{\frac{m_{non-b\ pair}}{m_{bucket\ i}} - \frac{m_W}{m_{top}}}{(std.\ dev.\ of\ numerator)}\right)$

Similar to the all-hadronic case, the Wwt is chosen to maximize algorithm efficiency.

8.3 Results and Discussions

As discussed above, the value of the Wwt are chosen to maximize the algorithm efficiency. To accomplish this, the number of possible values of Wwt are scanned to find the one with maximum efficiency.

8.3.1 All Hadronic $t\bar{t}$ Buckets

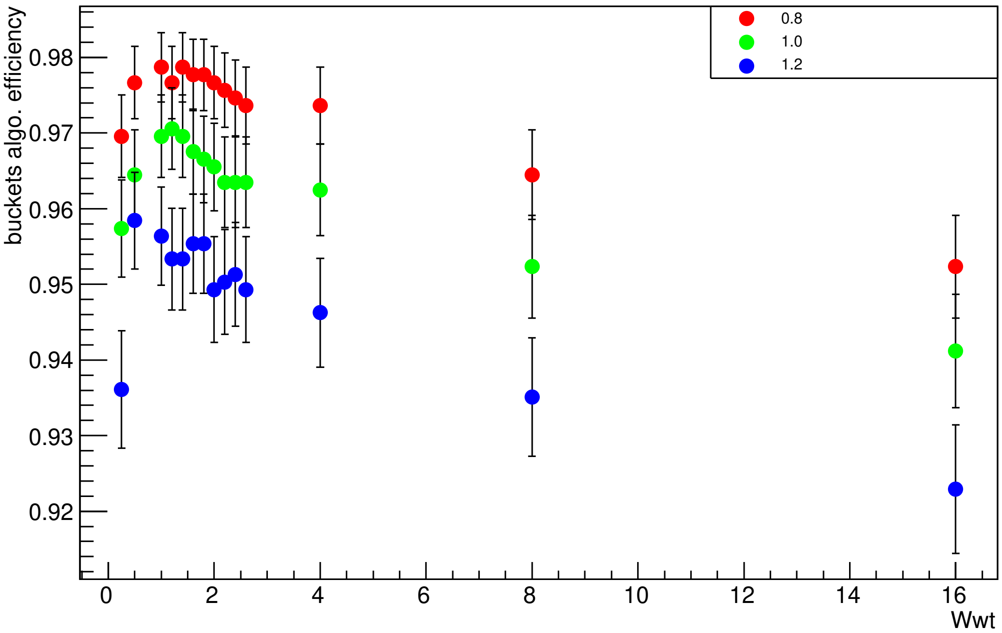


Figure 8.10: Scan of algorithm efficiency of the all hadronic 'Buckets of Tops' algorithm with respect to Wwt, to find optimum value of Wwt. Different detector smearings have been tried (values of a are idicated in the legend for detector smearing of $\mathcal{N}(0, a \times \sqrt{E})$ for quarks).

As seen in Figure 8.10, the maximum efficiency of the all-hadronic Buckets of Tops algorithm was found near $Wwt = 1$ at efficiency greater than 95%.

8.3.2 Semi-leptonic $t\bar{t}$ Buckets

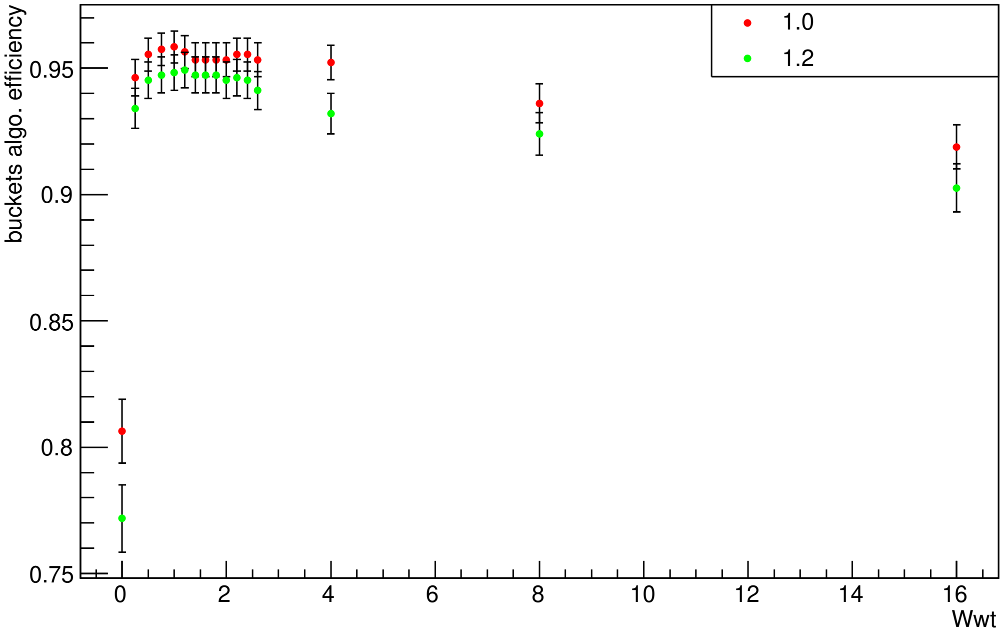


Figure 8.11: Scan of algorithm efficiency of the semi-leptonic 'Buckets of Tops' algorithm with respect to Wwt, to find optimum value of Wwt. Different degrees of detector smearing have been tried (values of a are indicated in the legend for detector smearing of $\mathcal{N}(0, a \times \sqrt{E})$ for quarks). No detector smearing has been applied on the parton which mimics a charged lepton, as lepton momentum resolution is better than jets.

As seen in Figure 8.11, the maximum efficiency of the semileptonic Buckets of Tops algorithm was found near $Wwt = 1$ at $\sim 95\%$.

8.4 Conclusions and Future Directions

Both the versions, all hadronic and semi-leptonic, of the modified 'Buckets of Tops' algorithm in this chapter, work significantly better than the original buckets version with about $\sim 20\%$ improvement in algorithm efficiency. The tuning parameter of the modified algorithm is the Wwt whose value turns out to optimal at 1, for both the versions to get best efficiency. It should be noted, that the maxima in performance at $Wwt = 1$ is universal for any detector smearing, which means that this Wwt can be applied in actual ATLAS events with real detector smearing on jet 4-momenta, without having to re-optimize the Wwt parameter for data and ATLAS simulations again. The stability in performance of the semileptonic and all hadronic versions of the algorithms have to be investigated in more detail, when addition jets are present in the input set of the algorithm, as well as in situation where all the decay products of the top quarks may not be present. Such stability is crucial for its application in the top-philic heavy resonance search, as the resonance top quarks are reconstructed using RC jets, which can miss some of the decay products of the resonance top quarks or can absorb in their cones the jets from the decay products of the spectator top quarks. The comparison of the value of the global metric for signal and background (pure combinatorics) should be studied to understand the discriminative power of the global metric, which can be useful for rejecting background where no real top quark is present. In the case of incomplete decay products for hadronic top buckets, approaches similar to neutrino reweighting [114] can be adopted where the missing decay products can be treated as neutrinos.

Chapter 9

The Vector Boson Fusion (VBF) H $\rightarrow WW^{*} \rightarrow e\nu\mu\nu$ Analysis

This chapter describes the observation of vector-boson-fusion (VBF) production of Higgs bosons in the $H\rightarrow WW^{*}\rightarrow e\nu\mu\nu$ decay channel in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV recorded with the ATLAS detector of the Large Hadron Collider (LHC) [39] in the data collected from 2015 through 2018 (referred to as the full run-2 ATLAS data) corresponding to an integrated luminosity of 139 fb^{-1}. This channel has previously been studied by the CMS [115, 116] and ATLAS [117, 118] collaborations using a dataset with integrated luminosity of approximately 36 fb^{-1} and at lower center-of-mass energies of 8 TeV and 7 TeV of the LHC. In this current version of the analysis, increased data and MC simulated events along with improved object selection, as well as new multi-variate discriminant using a Deep Neural Network (DNN), has significantly improved the analysis sensitivity compared to the previous results and has lead to the observation of the $H\rightarrow WW^{*}\rightarrow \ell\nu\ell\nu$ process for the first time in the ATLAS detector.

This analysis was conducted by a team of approximately 20 people. The main contribution from the author consists of the optimization of lepton selection in Section 9.2.1 and the estimation of the misidentified lepton backgrounds presented in Section 9.6. In the misidentified lepton backgrounds, the author focused mainly on estimation of electroweak backgrounds and its systematic uncertainty on Z+jets fake factor in Section 9.6.3.1. The information in other sections is taken from the ATLAS documents [119, 120, 121]. The analysis presented at the ICHEP-2020 conference [121]. The author has contributed to a lessor extent in the top quark background studies for the analysis done with 2015-16 ATLAS data published in [122] and in the overall analysis software development used for this work.

9.1 Collision Data and Simulation

In this section, the collection of data used and MC generators used for simulating signal and background events for this analysis has been discussed.

9.1.1 Data

The pp collision data used this analysis was collected in the ATLAS detector from 2015 through 2018 at a center-of-mass collision energy $\sqrt{s} = 13\,\mathrm{TeV}$, corresponding to an integrated luminosity of about $139{\pm}2.4\,\mathrm{fb}^{-1}$ and passing data quality requirements (to ensure all detector components were operational during data collection period). Events are triggered using a combination of single-lepton triggers and a dilepton e–μ trigger in order to maximize the total trigger efficiency, summarized in Table 9.1. As the instantaneous luminosity varies with time, the trigger rate is controlled by adjusting the p_T thresholds or quality criteria of the single lepton triggers. To maximize the total trigger efficiency, a logical OR is done between the single lepton triggers and one dilepton trigger. The dilepton (e–μ) trigger requires a lower p_T threshold on the leptons, which allows lowering the requirement of the leading lepton p_{T} and therefore increases the statistics at low p_{T}.

Table 9.1: Summary of triggers used to collect the dataset for VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ analysis. The numbers stand for p_{T} threshold (E_{T} threshold for L1 trigger inside parenthesis) in GeV and the corresponding identification criteria is indicated. Events are selected with an OR between the three types of triggers (single-e/μ and dilepton). Tighter requirements are placed on triggers with lower p_{T} thresholds.

Lepton	**High Level Trigger**
2015	
e	24-medium (L1: 20) OR 60-medium (L1: 22) OR 120-loose (L1: 22)
μ	20 (L1: 15) OR 50 (L1: 20)
$e\mu$	17-loose (L1: 15) (e), 14 (L: 10) (μ)
2016–2018	
e	26-tight (L1: 22) OR 60-medium (L1: 22) OR 140-loose (L1: 22)
μ	26 (L1: 20) OR 50 (L1: 20)
$e\mu$	17-loose (L1: 15) (e), 14 (L1: 10) (μ)

9.1.2 Simulation

The list of Monte Carlo (MC) generators used to produce signal and background processes are listed in Table 9.2. Due to QCD factorization theorem, the process of hard scattering is generally produced using different generators than the ones used for modeling parton showering (PS), hadronization and the underlying event (UE) in most of the processes. For hard scattering process, Powheg-Box v2 [123] with CT10 and NNPDF3.0 parton distributions function (PDF) set [124] is used while with MadGraph 5 (version 2.2.1 and 2.2.2) [125] the NNPDF23LO [78] PDF set is used. For the diboson and signal processes the AZNLO [126] tune is used while for other processes the A14 tune [77]. Sherpa 2.2.1 [127] is used to produce hard scattering at NLO using NNPDF 3.0 NNLO PDF set and showering is produced by its own parton shower generator [128] with a dedicated set of tuned parameters.

For all other hard scatter process generators, the showering is done using Pythia 8.210, Pythia 8.186 [129] or Pythia 6.428 [10] for both signal and background processes.

Table 9.2: MC generators used to model the signal and background processes, and corresponding cross sections (given for $m_H = 125\text{GeV}$ in the case of the Higgs boson production processes). The column "Precision $\sigma_{\text{incl.}}$" gives the precision of the inclusive cross section applied to the sample.

Process	Generator	σ.Br(pb)	Precision $\sigma_{incl.}$
ggF $H \to WW$	POWHEG+ Pythia8	10.4	N^3LO+NNLL
VBF $H \to WW$	POWHEG+ Pythia8	0.808	NNLO
WH $H \to WW$	Powheg +Pythia 8 (MINLO)	0.293	NNLO
ZH $H \to WW$	Powheg +Pythia 8 (MINLO)	0.189	NNLO
inclusive $Z/\gamma^\star \to \ell\ell$ ($40 \geqslant m_{\ell\ell} \geqslant 10\text{GeV}$)	Sherpa 2.2.1	6.80×10^3	NNLO
inclusive $Z/\gamma^\star \to \ell\ell$ ($m_{\ell\ell} \geqslant 40\text{GeV}$)	Sherpa 2.2.1	2.107×10^3	NNLO
$(W \to \ell\nu)\gamma$	SHERPA 2.2.8	453	NLO
$(Z \to \ell\ell)\gamma$	SHERPA 2.2.8	175	NLO
$t\bar{t}$ di-leptonic(e,μ, τ)	POWHEG+Pythia8	76.96	NNLO+NNLL
Wt leptonic	POWHEG+Pythia8	6.99	NLO
$q\bar{q}/g \to WW \to \ell\nu\ell\nu$	SHERPA 2.2.2	12.5	NNLO
$qq \to WWqq \to \ell\nu\ell\nu jj$	SHERPA 2.2.2	0.095	NNLO
$Z^{(*)}Z^{(*)} \to 2\ell 2\nu$	SHERPA 2.1	6.53	NLO
$gg \to WW \to 2\ell 2\nu$	Sherpa 2.2.2	0.47	NLO
$q\bar{q}/g \to \ell\nu\ell\ell$	SHERPA 2.2.2	2.98	NNLO
$q\bar{q}/g, gg \to \ell\ell\ell\ell$	SHERPA 2.2.2	1.269	NNLO

9.1.2.1 Higgs Boson Sample

The four main production modes at the LHC of Higgs boson: gluon fusion (ggF), VBF, and associated WH / ZH productions (The $t\bar{t}H$ production mode is not included due to its negligible contribution) and a decay into a WW^* final state are simulated at $m_H = 125$ GeV. The VBF events are generated with Powheg, interfaced with Pythia 8. The VBF cross-section is normalized to approximate next-to-next-to-leading-order (NNLO) QCD corrections and NLO electroweak corrections [130, 131, 132].

The ggF events are generated using the Powheg NNLOPS program [133, 134], interfaced with Pythia 8. The ggF cross-section is normalized to N^3LO QCD corrections and NLO electroweak corrections [135].

The associated WH / ZH production (collectively referred to as VH) is generated using Powheg and interfaced with Pythia 8. The VH cross-section is normalized to NNLO QCD corrections and NLO electroweak corrections [136].

All the Higgs boson samples use the `PDF4LHC15` [137] PDF set and the AZNLO tune of Pythia 8. The Higgs boson decay branching ratio is calculated with HDECAY [138] and PROPHECY4F [139, 140, 141] for all the Higgs boson samples.

9.1.2.2 Background Sample

The top background is one of the dominant backgrounds which consists of $t\bar{t}$ and single top. The $t\bar{t}$ events are produced using the PowhegBox v2 generator at NLO with the `NNPDF3.0nlo` PDF set and the h_{damp} parameter (a resummation damping factor for matching of Powheg matrix elements to the parton shower which affects the high-p_{T} radiation against which the $t\bar{t}$ system recoils.) set to 1.5 m_{top} and interfaced to Pythia 8 with A14 tuning. An NNLO reweighting is applied to the sample to correct for the known mismodeling of the leading lepton p_{T} due to missing higher-order corrections [142]. The single top quark production in association with W bosons is generated using the PowhegBox v2 generator at NLO in QCD and the `NNPDF3.0nlo` set of PDFs and interfaced to Pythia 8 using the A14 tune and the `NNPDF2.3lo` set of PDFs. The interference and overlap with $t\bar{t}$ production is removed by the diagram removal scheme [143]. The bottom and charm hadron decays are simulated by EvtGen.

Standard model diboson VV (WW, WZ, and ZZ) events are produced by two modes, QCD induced and electroweak induced. The QCD-induced diboson processes are simulated using the Sherpa 2.2.2 generator with matrix elements at NLO accuracy in QCD for up to one additional parton emmision and at LO accuracy for up to three additional parton emissions. Samples for the loop-induced processes $gg \to VV$ are generated using LO-accurate matrix elements for up to one additional parton emission. Electroweak diboson production is generated using Sherpa 2.1.1 in association with two jets ($VVjj$) with LO-accurate matrix elements. The $V\gamma$ events are generated using the Sherpa 2.2.2 generator with matrix elements are at NLO QCD accuracy for up to one additional parton emission and at LO accuracy for up to three additional parton emissions.

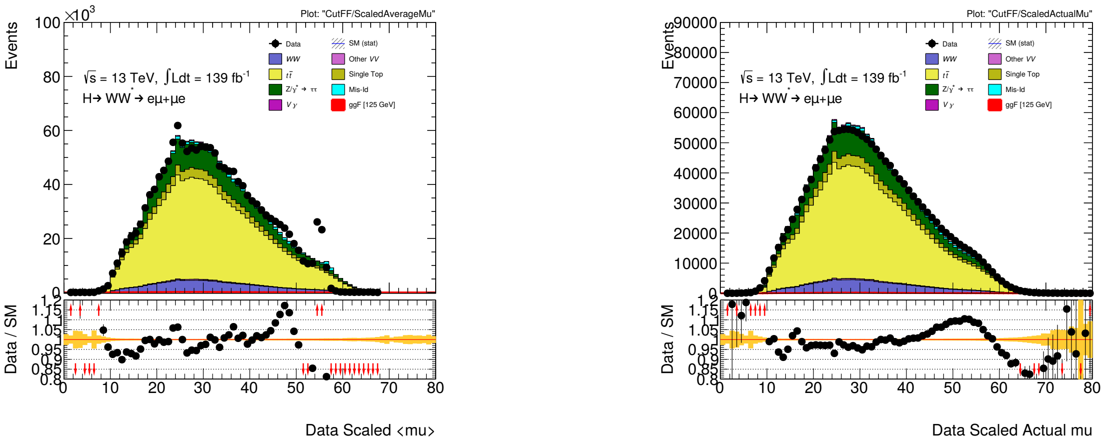
The Drell-Yan background is due to the Z+jets process. The QCD-induced production of Z+jets is simulated using the Sherpa 2.2.1 generator with NLO accuracy for matrix elements for up to two partons, and LO accuracy for up to four parton emmisions. Electroweak production of $\ell\ell jj$ final states are also generated using the Sherpa 2.2.1 using LO matrix element accuracy with up to two additional parton emissions. The cross-section is normalized to an NNLO prediction [144].

The dominant source of mis-identified lepton background is from the W+jets events, where tracks in the jets are mis-identified as leptons. For these events, a data driven estimation is used instead of MC simulation as such misidentification are hard to model. However, for purposes of validation and flavor composition studies MC samples are generated using Powheg interfaced with Pythia 8.

9.1.3 Pileup

For all MC samples, The effect of additional *pp* interactions in the same and neighboring bunch crossings (pile-up) is modeled in the MC samples by overlaying simulated inelastic *pp* events generated with Pythia 8 using the `NNPDF2.3lo` set of PDFs and the A3 tune over the original hard-scattering event.

In order to model pile-up conditions correctly, the μ data scale factor of 1/1.03 has been applied before reweighting the μ distribution in Monte Carlo to match as presented in Figure 9.1.



(a) Average μ, with data scale factor applied

(b) Actual μ, with data scale factor applied

Figure 9.1: Distributions of pile-up proxy observable (μ). Average μ is used to reweight Monte Carlo to match 2015 and 2016 data, while actual μ is used to reweight Monte Carlo to match 2017 and 2018 data. The yellow band includes the MC statistical uncertainties only. No normalization factors are applied.

To apply pile-up reweighting to the MC samples, for 2015 and 2016 data the μ averaged over all bunches in data while for 2017 and 2018 data the actual μ for each bunch in data is used to reduce statistical fluctuations in the MC samples.

9.2 Object Selection

This section discusses the selection of the objects - two charged leptons, two forward jets and missing transverse energy due to two neutrinos in the in the VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ final state, that are used in this analysis.

9.2.1 Leptons

The lepton selection used in this analysis has been optimized to reduce fake lepton background while limiting the corresponding loss in the signal efficiency. The leptons are required to originate from the primary vertex. In run-2, to reconstruct primary vertex (see [145] and [146]), only tracks passing quality cuts along with $p_{\mathrm{T}} > 500$ MeV are considered. In the case when more than one reconstructed primary vertex are reconstructed for an event, the primary vertex with the highest Σp_{T}^2 is considered as the default hard-scatter vertex. To reduce non-prompt lepton backgrounds, selection cuts are made on the impact parameters to discard lepton candidate tracks not associated with the primary vertex. The selection cuts on the impact parameters, and the working points in the final selection criteria are described for the signal electrons and muons in details below. (Looser requirements are used for misidentified leptons, discussed in Section 9.6).

9.2.1.1 Electrons

To define electrons and distinguish their signature from photons and jets, an identification and isolation selection is applied. The available identification working points, which use variables like the number of hits in the pixel and SCT detectors. Since, misidentified lepton backgrounds fall off quickly with increasing lepton p_{T}, lepton selections have been optimized in bins of p_{T}, with likelihood-based identification category *Tight* [147] for $p_{\mathrm{T}} < 25$ GeV. To recover efficiency at high E_T, where the fake background contribution is smaller, the identification category is relaxed to *Medium* for $p_{\mathrm{T}} > 25$ GeV. Further rejection of fake background is accomplished by requiring electrons to be isolated. To assess how well isolation working points perform in the $H \rightarrow WW^*$ analysis, an estimate for the pre-fit significance is calculated to give an indication of the performance of the configuration. The pre-fit significance

$$\sigma = \frac{N_{\text{Sig}}}{\sqrt{N_{\text{Sig}} + N_{\text{Non-W+jets bkg}} + N_{\text{W+jets bkg}} + \sigma^2_{\text{W+jets bkg}} + \sigma^2_{\text{Non-W+jets bkg}}}} \tag{9.1}$$

is used with the number of events N and systematic uncertainty σ to scan different isolation working points. For the systematic uncertainties, percentages of the process yield are used, namely $\sigma_{\mathrm{W+jets\,bkg}} = 0.4 \cdot N_{\mathrm{W+jets\,bkg}}$ and $\sigma_{\mathrm{Non\text{-}W+jets}} = 0.1 \cdot N_{\mathrm{Non\text{-}W+jets\,bkg}}$. The numbers 40% and 10% represent conservative estimates of the total systematic uncertainty affecting each of the processes which are nominally taken from those used in the previous optimization study in the 2015+16 analysis [148]. Based on systematic and statistical uncertainties for VBF and the availability of scale factors for those working points, *FCTight* [147] was chosen. Electrons are required to have a minimum transverse momentum of $p_{\mathrm{T}} > 10$ GeV and a pseudo-rapidity $|\eta| < 2.47$. They need to satisfy $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} < 5$, for rejection against fake electrons arising from pileup vertices. An additional selection of "Author" = 1 is applied to reduce the $W\gamma$ background by nearly half while maintaining a high signal efficiency. All of the electron selection requirements are summarized in Table 9.3.

Table 9.3: Electron selection requirements.

p_{T} range	Electron ID	η range	Author	Isolation	Impact parameter
< 25GeV > 25GeV	LHTight LHMedium	$(\lvert\eta\rvert < 1.37)$ —— $(1.52 < \lvert\eta\rvert < 2.47)$	1	FCTight	$\lvert z_0 \sin\theta\rvert < 0.5$ mm, $\lvert d_0\rvert/\sigma_{d_0} < 5$

9.2.1.2 Muons

"Combined" type muons (discussed in Section 4.3.2) are used in the analysis. The muons are required to pass the *Tight* working point, which maximizes the purity of the muons by requiring that muons are reconstructed from matching tracks in the ID and MS, and by applying stringent quality cuts. The efficiency of the *Tight* working point, evaluated using $t\bar{t}$ MC, is 89.9% (91.8%) for low-p_{T} (high-p_{T}) muons [149]. Additionally, muon are required to have a pseudorapidity $|\eta| < 2.5$. The isolation of a muon candidate is an important way to differentiate prompt muons from those produced in background processes like semileptonic decays. Isolation describes the amount of detector activity in the vicinity of the muon candidate. Many isolation working points are provided, differing in the amount of track- and calorimeter-based activity they accept in a cone around the muon candidate. A selection of recommended isolation working points was studied, and the *FCTight* working point was chosen to maximize sensitivity, in a similar method described for electrons. They need to satisfy $|z_0 \sin\theta| < 1.5$ mm and $|d_0|/\sigma_{d_0} < 15$. All of the muon selection requirements are summarized in Table 9.4.

Table 9.4: Muon selection requirements.

p_{T} range	Muon ID	η range	Isolation	Impact parameter
> 15GeV	Tight	$\|\eta\| < 2.5$	FCTight	$\|z_0 \sin\theta\| < 0.5$ mm, $\|d_0\|/\sigma_{d_0} < 3$

9.2.2 Jets

Jets are reconstructed using The anti-k_t algorithm with a radius parameter of $R = 0.4$ with particle flow objects [150] as input is used to reconstruct jets. For this analysis, jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4.5$ are selected. Furthermore, a multivariate selection called the Jet Vertex Tagger (JVT), which uses calorimeter and tracking information to separate hard-scatter jets from pile-up jets, with JVT Tight WP (JVT > 0.5) is applied on jets with $20 < p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ to reduce contamination from pile-up. Jets with $p_{\mathrm{T}} > 30$ GeV are used for jet multiplicity count, while, the jets with 20 GeV $< p_{\mathrm{T}} <$ 30 GeV jets referred to as the "sub-threshold" jets are used for b-jet counting (as described in Section 9.2.2.1).

9.2.2.1 B-tagged jets

For jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$, the jets containing b-hadrons (b-jets) are identified using a neural network based DL1r b-tagging algorithm at 85% b-jet tagging efficiency working point [169].

9.2.3 Missing Transverse Momentum

The missing transverse momentum $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ due to the two neutrinos in the final state is obtained by the negative vector sum of the p_{T} of all the selected objects (jets and leptons), together with tracks which are not associated with these objects but are consistent with their primary pp vertex [67].

9.2.3.1 Overlap Removal

In order to correctly calculate the $E_{\mathrm{T}}^{\mathrm{miss}}$, any overlap between the objects - leptons and jets should be accounted for properly. The following scheme [152, 153] for overlap removal is used:

- electron-electron: in case of an overlapping second layer cluster (3x5) or a shared track, the lower p_{T} electron is eliminated.
- electron-muon: the electron is removed if a combined muon shares an ID track with an electron, while the muon is removed if a calo-tagged muon shares an ID track with an electron.
- electron-jet: for $\Delta R(jet, e) < 0.2$, the jet is removed after which the electron is removed if $\Delta R(jet, e) < \min(0.4, 0.04 + 10\ \mathrm{GeV} \cdot p_{\mathrm{T}}^{e})$ for any surviving jet.
- muon-jet: If a jet has less than three associated tracks with $p_{\mathrm{T}} > 500$ MeV and $\Delta R(jet, \mu) < 0.2$, or if ratio of p_{T} of the muon and jet is larger than 0.5

($p_{\mathrm{T}}^{\mu}/p_{\mathrm{T}}^{jet} > 0.5$) along with the ratio of the muon p_{T} to the total p_{T} of tracks (associated to the jet) with $p_{\mathrm{T}} > 500$ MeV if larger than 0.7, then the jet is removed and the muon is then removed for any surviving jets, if $\Delta R(jet, \mu) < \min(0.4, 0.04 + 10\ \mathrm{GeV} \cdot p_{\mathrm{T}}^{\mu})$.

A jet is always saved from removal if it is b-tagged.

9.3 Composite Observables

In this section, observables used in the analysis are presented, some of which reject reducible backgrounds, some enhance the Higgs decay topology over the irreducible WW background, and some observables specifically exploiting VBF kinematic signature to enhance the VBF $H \rightarrow WW^{*} \rightarrow \ell\nu\ell\nu$ signal.

9.3.1 Background Rejection

To reduce the reducible backgrounds the following observables are defined:

- $m_{\tau\tau}$ - Invariant mass of the hypothetical $\tau\tau$ system under the assumption of the Collinear Approximation Method [155], which assumes that the τ lepton decay products (charged lepton and neutrinos) are collinear (go along the same direction) if the τ leptons are sufficiently boosted, and that the neutrinos from the τ leptons are the only source of missing energy. The Track based Soft Term (TST) missing transverse energy ($E_{\mathrm{T}}^{miss,TST}$) is used in the $m_{\tau\tau}$ calculation. A cut applied on this variable is effective to suppress not only $Z \rightarrow \tau\tau$, but also $H \rightarrow \tau\tau$.

- N_{b-jet} - Number of jets with $p_{\mathrm{T}} > 20$ GeV identified as b-jets from the b-tagged algorithm (DL1r) at 85% efficiency. This variable is used to suppress and estimate top quark backgrounds.

- p_{T}^{ll} - Transverse momentum of the dilepton system. Large values reflect that the leptons and neutrinos are emitted in opposite hemispheres, corresponding to the signal topology, and is especially effective at rejecting Drell-Yan background.

- $\Delta\phi_{ll,E_T^{miss}}$ - Azimuthal angle between the dilepton system and the TST MET ($E_{\mathrm{T}}^{miss,TST}$). Strongly peaked at back-to-back for signal and most backgrounds. Events in which this angle is close to zero may have mis-measured objects and thus are rejected.

- $\max(m_{\mathrm{T}}^{\ell})$ - The transverse mass using each lepton is defined as:

$$m_{\mathrm{T}}^{\ell} = \sqrt{2p_T^l E_T^{miss,TST}(1 - \cos(\phi^l - \phi^{E_T^{miss,TST}}))} \tag{9.2}$$

  If the process has at least one real W boson, such as the signal, the transverse mass of at least one of the leptons will be a larger value. Therefore, placing a lower bound on the maximum value of m_{T}^{ℓ} helps in rejecting backgrounds with no real W boson, such as $Z/\gamma^* \to \tau\tau$ and QCD.

9.3.2 Topological Variables

The topological variables used to enhance the Higgs signal are defined below:

- $\Delta\phi_{ll}$ - Azimuthal angle between the leading and subleading leptons. $\Delta\phi_{ll}$ between the leptons in the final state of $H\to WW^*\to \ell\nu\ell\nu$ decay is peaked near zero because of the zero spin of the SM Higgs boson and the V-A structure of the weak interactions (discussed in Section 2.1.3.2). The dominant backgrounds from top quark and non-resonant WW processes have relatively flat distributions.

- m_{ll} - Invariant mass of the leading and subleading leptons. Since, $m_{ll} \simeq |p_1||p_2|[1 - cos\Delta\phi_{ll}]$ and $\Delta\phi_{ll} \sim 0$ for signal, m_{ll} also peaks at a smaller value for signal. While the backgrounds from top quark, Z/DY, and non-resonant WW processes have relatively higher values of m_{ll}.

- m_{T} - Since it is not possible to fully reconstruct the mass of the $WW^* \rightarrow l\nu l\nu$ system, its transverse mass is defined as:

  $$m_{\mathrm{T}} = \sqrt{(E_{ll} + E_{\mathrm{T}}^{miss})^2 - |p_{ll} + E_{\mathrm{T}}^{miss}|^2} \tag{9.3}$$

  where $E_{ll} = \sqrt{|p_{ll}|^2 + m_{ll}^2}$. It is is designed to capture the mass of the Higgs and is therefore a powerful discriminant against non-resonant WW background. Similar to Equation 9.2, the $E_{\mathrm{T}}^{miss,TST}$ is used in the definition to improve the resolution of the observable.

9.3.3 VBF Observables

The following observables exploit the kinematic signature of the VBF $H\rightarrow WW^*\rightarrow \ell\nu\ell\nu$ topology:

- $p_{\mathrm{T}}^{\mathrm{tot}}$ - The modulus of the vector sum of transverse momenta all the physics objects in the event, defined as:

  $$p_{\mathrm{T}}^{\mathrm{tot}} = |\boldsymbol{p}_{\mathrm{T}}{}^{l1} + \boldsymbol{p}_{\mathrm{T}}{}^{\mathrm{l2}} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} + \sum \boldsymbol{p}_{\mathrm{T}}{}^{\mathrm{jets}}| \tag{9.4}$$

  Here, all jets are included which pass the final selection criteria as described in Section 9.2.2. Track $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is used in $p_{\mathrm{T}}^{\mathrm{tot}}$, which is the vector sum of all the visible physics objects in the event with an additional soft track term. As a result, after cancellation:

  $$p_{\mathrm{T}}^{\mathrm{tot}} \simeq |\boldsymbol{E}_{\mathrm{T,\,soft}}^{\mathrm{miss,\,trk}}| \tag{9.5}$$

  This quantity is the modulus of the vector sum of soft tracks falling outside of selected jets, as is used to estimate the amount of soft QCD radiation in a given event. Backgrounds with QCD jets have more soft gluon radiation and therefore lie at larger values of $p_{\mathrm{T}}^{\mathrm{tot}}$ with respect to VBF events which have relatively little QCD activity.

- Δy_{jj} - Rapidity distance between the two leading jets in the event, defined as:

$$\Delta y_{jj} = |Y_{jet1} - Y_{jet2}| \tag{9.6}$$

  The signal is characterized by forward jets which have a larger Δy_{jj} value with respect to the backgrounds.

- m_{jj} - Invariant mass of the two leading jets in the event, defined as:

$$\begin{aligned} m_{jj}^2 &= m_{j1}^2 + m_{j2}^2 + 2[E_T^{j1} E_T^{j2} cosh(\Delta y_{jj}) - \boldsymbol{p}_{\mathrm{T}}{}^{j1} \cdot \boldsymbol{p}_{\mathrm{T}}{}^{j2}] \\ &\simeq 2 p_{\mathrm{T}}^{j1} p_{\mathrm{T}}^{j2} [cosh(\Delta y_{jj}) - cos(\Delta\phi_{jj})] \end{aligned} \tag{9.7}$$

  The second equality assumes jet mass smaller than jet p_{T}. Since, m_{jj} is a function of Δy_{jj}, its value increase with larger Δy_{jj} as in the case of signal. Additional discrimination power comes from the other terms in Equation 9.7. For $t\bar{t}$ background, the jets tend to be back-to-back while for VBF signal, jets are distributed quite uniformly in ϕ. In addition, after b-jet veto, $t\bar{t}$ events with ISR jets and a b-jet outside the tracking volume are favored in the signal phase space, resulting in a softer jet p_{T} spectrum for $t\bar{t}$ background. Therefore, signal events tend to have larger m_{jj} value than background events.

- η_{lep} centrality ($\Sigma_l C_l$) - Lepton centrality is an extension of the OLV discussed in Section 9.4.2, and quantifies the position of the leptons with respect to the two leading jets in the η-plane:

$$\mathrm{OLV}_{l_0} = 2 \cdot |\frac{\eta_{l_0} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}}|$$

$$\mathrm{OLV}_{l_1} = 2 \cdot |\frac{\eta_{l_1} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}}|$$

$$\eta_{\mathrm{lep}}\,\text{centrality} = \mathrm{OLV}_{l_0} + \mathrm{OLV}_{l_1} \tag{9.8}$$

where, $\bar{\eta} = (\eta_{j_0} + \eta_{j_1})/2$ is the average η of the two tag jets. OLV_l is defined for each lepton as:

$$\mathrm{OLV}_l \begin{cases} = 0 & : \text{lies in the middle of the rapidity gap between the two tag jets.} \\ < 1 & : \text{lies within the rapidity gap between the two tag jets.} \\ > 1 & : \text{lies outside the rapidity gap between the two tag jets.} \end{cases} \tag{9.9}$$

9.4 Event Selection

In this section, a summary of the event selections for the VBF analysis is provided in Table 9.5. These selection have been designed on the common observable defined above, to exploit the signal event topolgy and reduce backgrounds. After the preselection cuts, a deep neural network (DNN) is trained to separate signal and background which is also discussed here.

9.4.1 Common Preselection

The following set of cuts are applied to select $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ events. Since they effectively select both VBF and ggF mode of Higgs production, they are referred to as common preselection.

- exactly one pair of opposite charge, different flavor ($e + \mu$) leptons
- $p_{\mathrm{T}}^{lead} > 22$ GeV and $p_{\mathrm{T}}^{sublead} > 15$ GeV. Since, one of the W bosons must be off-shell in the Higgs boson decay, one of the lepton is expected to have a lower energy which is the rationale behind asymmetric selections. These cuts have been optimized to reduce misidentified lepton background while maximizing signal acceptance.
- $m_{ll} > 10$ GeV, to suppress low mass meson resonances and DY events

9.4.2 VBF-specific Preselection

After the application of common preselection guided by the $H \to WW^* \to \ell\nu\ell\nu$ final state, the VBF preselection are driven by the distinctive event topology of the VBF signal (with respect to ggF), with the presence of two energetic jets separated by a large gap in rapidity. The VBF preselection requires:

$$N_{\mathrm{jet},p_{\mathrm{T}}>30\ \mathrm{GeV}} \geqslant 2 \tag{9.10}$$

The two leading jets are considered to be originating from the quarks in the final state of the VBF Higgs production, referred to as "VBF induced jets".

In order to further enhance the VBF topology, the following criteria are also applied to suppress the background processes:

- A $Z \to \tau\tau$ veto, which is a cut around the Z mass pole $m_{\tau\tau} < m_Z - 25$GeV is applied to reduce the $Z \to \tau\tau$ background.
- A central-jet veto (CJV), which rejects events with jets above $p_{\mathrm{T}} > 30$ GeV present between the pseudo-rapidity gap of the two leading jets.
- An outside-lepton-veto (OLV), which discards events with the two charged leptons lying outside the rapidity gap between the leading jets.

- a $m_{jj} > 120\,\text{GeV}$ cut is applied to ensure orthogonality with the VH process, and does not have any significant effect on the VBF signal yield.

Table 9.5: Event selection criteria used to define the signal and control regions in the analysis.

<table>
<tr><th></th><th>Signal region</th><th>Z+jets CR</th><th>Top quark CR</th></tr>
<tr><td rowspan="2">Pre-selection</td><td colspan="3">Two isolated, different-flavour leptons ($\ell = e, \mu$) with opposite charge
$p_{\mathrm{T}}^{\mathrm{lead}} > 22\mathrm{GeV}$, $p_{\mathrm{T}}^{\mathrm{sublead}} > 15\,\mathrm{GeV}$
$m_{\ell\ell} > 10\,\mathrm{GeV}$, $N_{\mathrm{jet}} \geqslant 2$</td></tr>
<tr><td>$N_{b\text{-jet},(p_{\mathrm{T}}>20\ \mathrm{GeV})} = 0$</td><td>$N_{b\text{-jet},(p_{\mathrm{T}}>20\ \mathrm{GeV})} = 0$</td><td>$N_{b\text{-jet},(p_{\mathrm{T}}>20\ \mathrm{GeV})} = 1$</td></tr>
<tr><td rowspan="2">Selection</td><td>$m_{\tau\tau} < m_Z - 25\,\mathrm{GeV}$
$m_{jj} > 120\,\mathrm{GeV}$
–</td><td>$|m_{\tau\tau} - m_Z| < 25\,\mathrm{GeV}$
–
$m_{\ell\ell} < 70\,\mathrm{GeV}$</td><td>$m_{\tau\tau} < m_Z - 25\,\mathrm{GeV}$
–
–</td></tr>
<tr><td colspan="3">central jet veto
outside lepton veto</td></tr>
<tr><td colspan="4">A DNN is applied in the SR that uses 15 discriminant variables:
$\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, m_{T}, Δy_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, $\eta_{\ell}^{\mathrm{centrality}}$, $m_{\ell 1 j 1}$, $m_{\ell 1 j 2}$, $m_{\ell 2 j 1}$,$m_{\ell 2 j 2}$,
$p_{\mathrm{T}}^{\mathrm{jet}_1}$, $p_{\mathrm{T}}^{\mathrm{jet}_2}$, $p_{\mathrm{T}}^{\mathrm{jet}_3}$, and $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance</td></tr>
</table>

Signal region is defined after applying all these preselection cuts, summarized in Table 9.5. The distributions of the discriminant variables, described in Section 9.3.3, at the VBF preselection stage are shown in Figure 9.2.

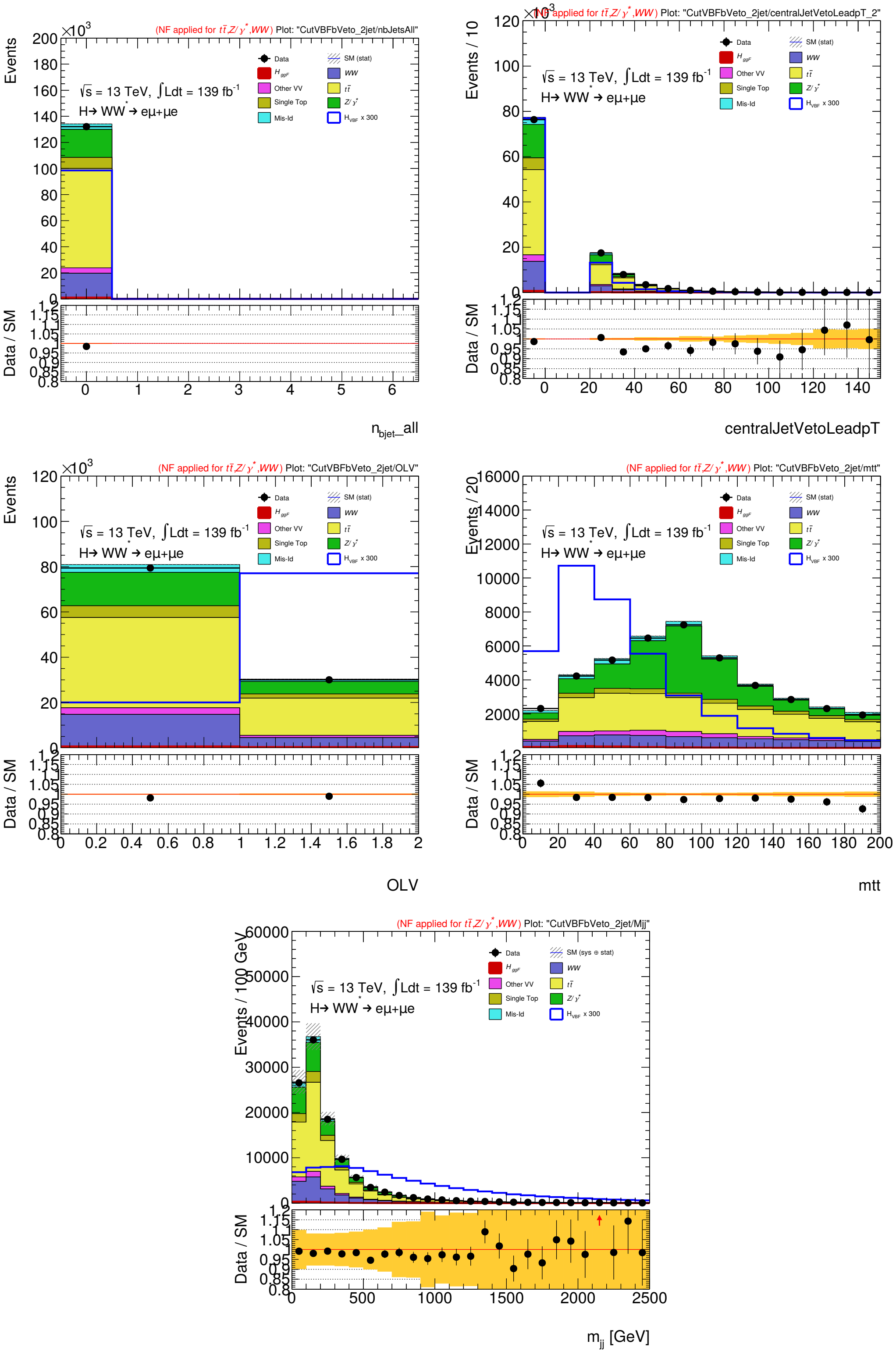


Figure 9.2: Distributions of number of b-tagged jets, CJV, OLV, $m_{\tau\tau}$, and m_{jj} at the VBF pre-selection cut stage. A blinding criteria (Data and MC events with DNN > 0.25 are removed) is applied after the b-veto cut. The yellow band represents the MC statistic only.

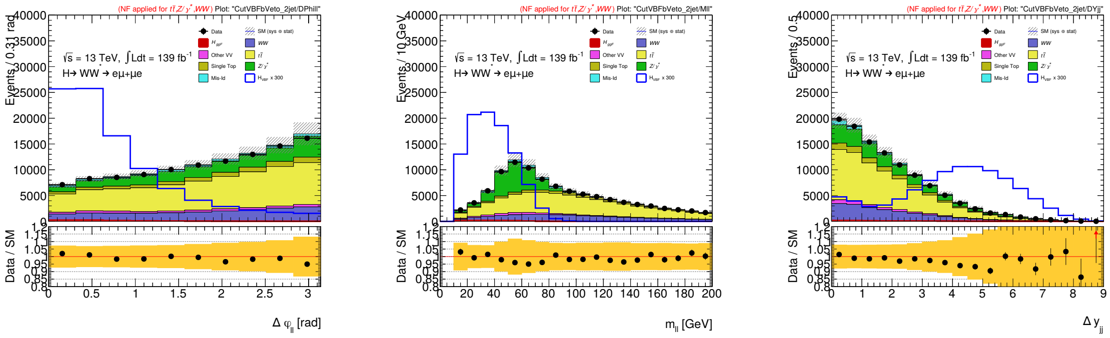
9.4.3 Deep Neural Network (DNN)

Given the low expected VBF signal yield compared to other Higgs production mechanisms, a multivariate analysis based on a deep neural network (DNN), implemented through keras [156] and TensorFlow [157], is employed to extract the most out of the statistically limited dataset.

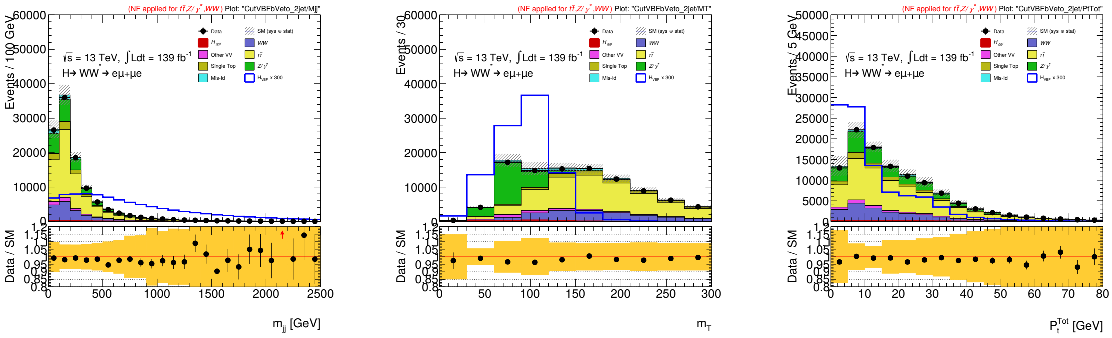
The DNN is trained using a total of 15 input variables:

- VBF topology: Δy_{jj}, m_{jj}, $\eta_{\ell}^{\text{centrality}}$,$m_{\ell 1 j1}$, $m_{\ell 1 j2}$, $m_{\ell 2 j1}$,$m_{\ell 2 j2}$,$p_{\text{T}}^{\text{jet}_1}$,$p_{\text{T}}^{\text{jet}_2}$,$p_{\text{T}}^{\text{jet}_3}$
- H → WW decay: $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, m_{T}
- Top suppression: $p_{\text{T}}^{\text{tot}}$, $\text{E}_{\text{T}}^{\text{miss}}$ Significance

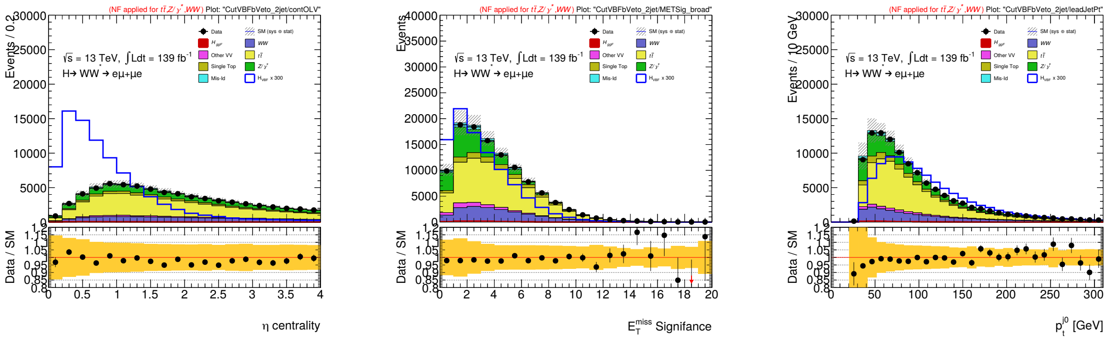
Here, m_{lkjl} refers to the invariant mass of system of the k-th lepton and the l-th jet, with k and l as 1 or 2 referring to the leading object or sub-leading object respectively. The rest of the input variables above have been discussed in Section 9.3, and their distributions in the VBF preselection region are shown in Figures 9.3–9.4.



(a) $\Delta\phi_{\ell\ell}$ (b) $m_{\ell\ell}$ (c) Δy_{jj}

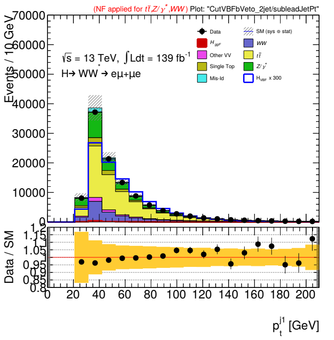


(d) m_{jj} (e) m_{T} (f) $p_{\mathrm{T}}^{\mathrm{tot}}$



(g) $\eta_{\ell}^{\mathrm{centrality}}$ (h) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance (i) $p_{\mathrm{T}}^{\mathrm{jet}_1}$

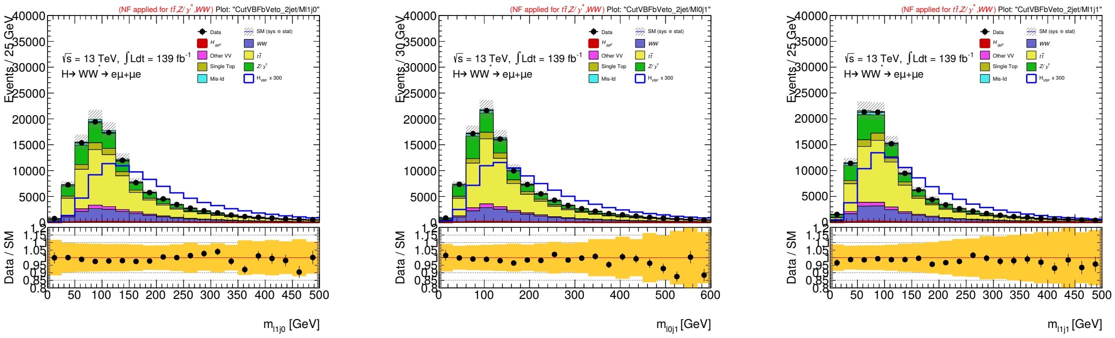
Figure 9.3: Distributions of $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, Δy_{jj}, m_{jj}, m_{T}, $p_{\mathrm{T}}^{\mathrm{tot}}$, $\eta_{\ell}^{\mathrm{centrality}}$, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance, and $p_{\mathrm{T}}^{\mathrm{jet}_1}$ after the VBF preselection. The VBF signal is scaled by a factor of 300. The yellow band represents the MC statistical uncertainties and the main sources of experimental uncertainties (JES/JER/MET/flavor-tagging).



(a) $p_{\mathrm{T}}^{\mathrm{jet}_2}$

(b) $p_{\mathrm{T}}^{\mathrm{jet}_3}$

(c) $m_{\ell 1 j1}$



(d) $m_{\ell 2 j1}$

(e) $m_{\ell 1 j2}$

(f) $m_{\ell 2 j2}$

Figure 9.4: Distributions of $p_{\mathrm{T}}^{\mathrm{jet}_2}$, $p_{\mathrm{T}}^{\mathrm{jet}_3}$, $m_{\ell 1 j1}$, $m_{\ell 2 j1}$, $m_{\ell 1 j2}$, and $m_{\ell 2 j2}$ after the VBF preselection. The VBF signal is scaled by a factor of 300. The yellow band represents the MC statistical uncertainties and the main sources of experimental uncertainties (JES/JER/MET/flavor-tagging).

The DNN is trained on MC events passing the VBF preselection for all major backgrounds, except the W+jets, are included in the training. The reason for excluding W+jets background from training is that it is not well-modeled using simulation (discussed in detail in the Section 9.6). The DNN output reflects the compatibility of an event with having VBF-like kinematics, with higher the output value the more signal-like is the event. The DNN output has been binned to perform binned likelihood fit on its distribution (discussed in Section 9.7.1). The bin boundaries are chosen with an algorithm that aims to split the bins as finely as possible, while having enough statistics in each bin. It starts with splitting the DNN output into very fine bins. Then the algorithm starts combining the adjacent bins together if the bins have less than 10 expected signal and background events each and/or relative statistical uncertainty on the background of more than 20%. After the above requirements are met for the bins, the merging process stops once the bins have at least 20 expected signal events.

The final bin boundaries of the DNN output are: [0.00, 0.25, 0.59, 0.73, 0.83, 0.89, 0.93]. In the bin with the highest DNN output, the expected VBF signal-to-background ratio is approximately 3.5. The DNN output distribution in the final signal region is presented in Figure 9.5.

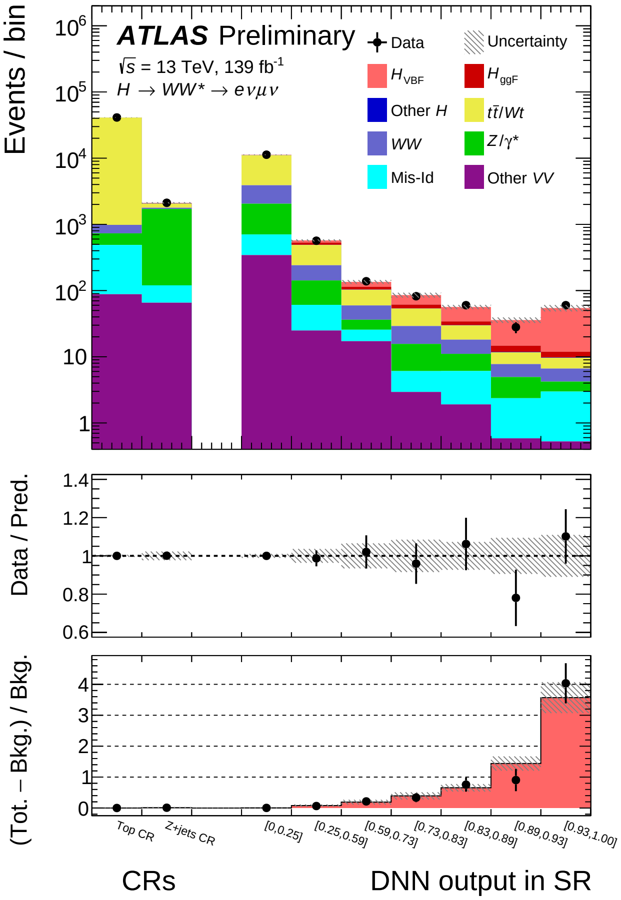


Figure 9.5: Post-fit distribution of the DNN output in the VBF signal region, together with the top quark and Z+jets control regions. The hatched band shows the total uncertainty of the signal and background modeled contributions. The middle panel shows the ratio of the data to the sum of the fitted signal and background. The bottom panel displays the signal to background ratio, where the hatched band indicates the combined statistical and systematic uncertainty for the fitted signal and background. Figure from [121].

9.5 Background Estimation

This section describes estimation of all the major background processes, except the background due to misidentified leptons, that are present in the analysis. The estimation of the background from misidentified leptons, which requires a data-driven approach, is discussed in the Section 9.6.

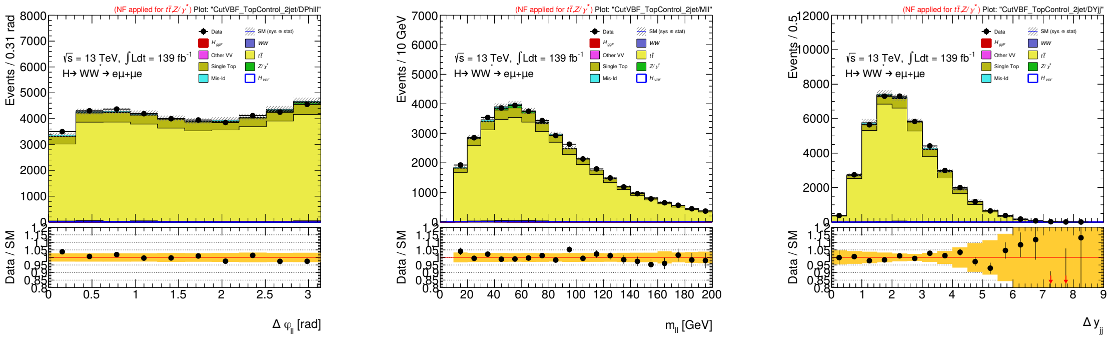
Table 9.6: Cutflow in the VBF signal region, the top CR, the $Z \rightarrow \tau\tau$ CR, and the WW VR. This cutflow includes the W+jets data driven estimate. Only the statistical errors are presented.

$\sqrt{s} = 13TeV$, $\mathcal{L} = 139fb^{-1}$ (Full Run 2)	H_{VBF}	H_{ggF}	WW	Other VV	Top	Zjets	Mis-Id	Total Bkg	Data	Data/MC
Apply fake factor	596.42 ± 0.74	5630.05 ± 7.80	127119.22 ± 114.74	19263.72 ± 206.72	1163914.11 ± 237.39	257414.12 ± 361.98	38973.55 ± 236.41	1612314.76 ± 547.02	1585575	0.98 ± 0.00
2-jet (30,30) fJVT	366.44 ± 0.58	1061.57 ± 3.08	24462.22 ± 30.25	4925.51 ± 98.24	920515.85 ± 203.89	26196.51 ± 90.16	12089.39 ± 162.49	989251.03 ± 294.41	975285	0.99 ± 0.00
b-veto	323.65 ± 0.54	914.35 ± 2.86	21166.54 ± 28.65	3915.68 ± 92.53	64352.98 ± 57.80	21980.21 ± 85.29	4463.54 ± 71.36	116793.30 ± 158.42	109428	0.93 ± 0.00
$Z \rightarrow \tau\tau$ veto	278.04 ± 0.50	798.58 ± 2.68	12310.00 ± 22.65	2086.47 ± 74.83	39234.78 ± 45.07	7801.75 ± 58.49	2627.40 ± 55.13	64858.97 ± 120.87	61311	0.94 ± 0.00
blinding (2-jet)	278.04 ± 0.50	798.58 ± 2.68	12310.00 ± 22.65	2086.47 ± 74.83	39234.78 ± 45.07	7801.75 ± 58.49	2627.40 ± 55.13	64858.97 ± 120.87	61311	0.94 ± 0.00
M_{jj} ¿ 120 GeV	253.95 ± 0.48	488.39 ± 2.07	8284.63 ± 17.71	1397.54 ± 57.08	27103.62 ± 37.10	5134.80 ± 46.09	1554.44 ± 43.08	43963.42 ± 94.52	41466	0.94 ± 0.01
CJV (30GeV)	234.87 ± 0.46	416.06 ± 1.92	6882.96 ± 16.73	1156.26 ± 51.27	21361.54 ± 33.27	4359.13 ± 44.08	1241.61 ± 38.62	35417.55 ± 86.34	33802	0.95 ± 0.01
OLV bool	203.47 ± 0.43	201.90 ± 1.36	2215.65 ± 10.72	402.50 ± 32.93	7731.98 ± 20.12	1696.74 ± 33.49	418.62 ± 23.04	12667.39 ± 57.08	12189	0.95 ± 0.01
VBF SR: DNN ¡= 0.25	36.80 ± 0.18	121.68 ± 1.06	2047.68 ± 10.24	361.50 ± 29.75	7384.46 ± 19.62	1568.09 ± 27.25	361.78 ± 21.78	11845.19 ± 50.92	11256	0.95 ± 0.01
VBF SR: DNN ¿ 0.25	166.66 ± 0.39	80.22 ± 0.85	167.97 ± 3.18	41.01 ± 14.12	347.52 ± 4.45	128.65 ± 19.47	56.83 ± 7.53	822.20 ± 25.80	933	0.94 ± 0.04
VBF SR: DNN ¿ 0.93	39.45 ± 0.19	2.66 ± 0.15	2.63 ± 0.26	0.49 ± 0.13	3.08 ± 0.48	1.51 ± 0.31	2.44 ± 0.78	12.83 ± 1.03	60	1.15 ± 0.15
$Z \rightarrow \tau\tau$ CR: $\lvert m_{\tau\tau} - m_Z \rvert < 25$, bVeto	27.14 ± 0.16	67.22 ± 0.76	1817.53 ± 8.20	713.12 ± 31.62	5684.49 ± 17.28	8860.01 ± 37.18	633.68 ± 26.53	17776.05 ± 58.76	16365	0.92 ± 0.01
$Z \rightarrow \tau\tau$ CR: $M_{ll} < 70$ GeV	23.95 ± 0.15	58.21 ± 0.70	408.53 ± 3.76	429.97 ± 25.77	1163.27 ± 7.65	7890.42 ± 32.44	338.91 ± 18.67	10289.32 ± 46.24	9314	0.90 ± 0.01
$Z \rightarrow \tau\tau$ CR: CJV	22.12 ± 0.14	51.27 ± 0.66	350.69 ± 3.58	373.64 ± 24.51	951.74 ± 6.97	6981.50 ± 31.53	314.02 ± 17.90	9022.85 ± 44.47	8170	0.90 ± 0.01
$Z \rightarrow \tau\tau$ CR: OLV	17.29 ± 0.13	14.78 ± 0.36	83.63 ± 1.80	67.69 ± 12.24	279.27 ± 3.79	1687.34 ± 15.24	54.53 ± 8.87	2187.23 ± 21.88	2114	0.96 ± 0.02
Top CR: 1 b-jets	39.78 ± 0.19	129.19 ± 1.08	3009.78 ± 9.47	907.83 ± 30.17	353256.14 ± 130.15	3656.20 ± 28.80	4644.57 ± 96.60	365603.71 ± 167.64	359321	0.98 ± 0.00
Top CR: $Z \rightarrow \tau\tau$ veto	34.16 ± 0.18	111.77 ± 1.00	1680.17 ± 7.37	491.76 ± 23.03	216323.59 ± 101.55	1293.85 ± 15.80	2798.44 ± 76.70	222699.58 ± 130.50	220476	0.99 ± 0.00
Top CR: CJV	30.80 ± 0.17	95.13 ± 0.93	1375.94 ± 6.93	410.07 ± 21.26	178190.56 ± 92.80	1089.26 ± 15.05	2250.73 ± 70.02	183411.69 ± 119.34	182044	0.99 ± 0.00
Top CR: OLV	22.37 ± 0.14	28.49 ± 0.52	274.00 ± 3.47	90.49 ± 12.66	40411.20 ± 44.32	265.88 ± 8.59	402.20 ± 33.44	41472.27 ± 57.69	41112	0.99 ± 0.01
b-veto	323.65 ± 0.54	914.35 ± 2.86	21166.54 ± 28.65	3915.68 ± 92.53	64352.98 ± 57.80	21980.21 ± 85.29	4463.54 ± 71.36	116793.30 ± 158.42	109428	0.93 ± 0.00
WW VR: M_T¿130 GeV	32.37 ± 0.17	130.64 ± 1.09	16205.26 ± 24.76	1809.25 ± 44.73	50531.02 ± 51.34	864.82 ± 28.64	1947.83 ± 48.70	71488.82 ± 91.88	68076	0.95 ± 0.00
WW VR: M_{T2}¿160 GeV	18.87 ± 0.13	50.21 ± 0.67	6575.65 ± 14.24	706.38 ± 32.95	11645.94 ± 25.06	302.46 ± 12.89	564.35 ± 25.02	19844.98 ± 52.05	18605	0.94 ± 0.01
WW VR: CJV	16.88 ± 0.12	42.85 ± 0.62	5401.88 ± 13.40	582.10 ± 32.52	8527.32 ± 21.85	248.15 ± 12.53	457.98 ± 22.38	15260.28 ± 48.71	14452	0.95 ± 0.01

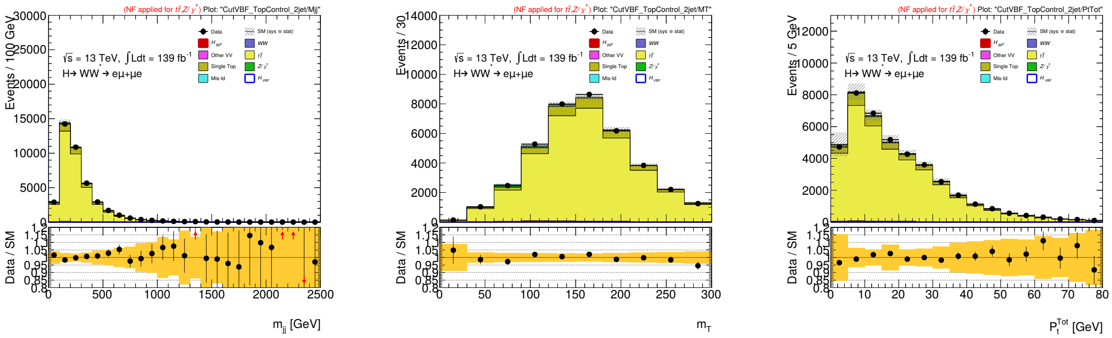
9.5.1 Top Control Region

Although, there is a very good agreement between between theory calculations and ATLAS top quark measurements in general, the top quark background normalization in VBF analysis is constrained using a top-rich control region (CR). Since, the selected jets in the VBF signal topology have a looser η requirement (—$\eta < 4.5$|) compared to the η requirement in the cross section measurements in ATLAS (|$\eta < 2.5$) [158, 159], the existing theoretical models have not been tested in this region by top quark measurements. A common normalization is used for both $t\bar{t}$ and single top (ST) as their kinematic shapes are similar. The top control region (CR) is made orthogonal to the VBF SR by requiring one and only one b-tagged jet ($N_{b-jet} = 1$) instead of b-jet veto, while keeping all the other selection criteria same as SR. The reason for $N_{b-jet} = 1$ instead of an inclusive b-tagged region is to bring the flavor composition of tag jets closer to the one in the b-vetoed SR. The purity of the top control region is of about 97%. The modeling of the DNN input variables are shown in Figure 9.6 and 9.7, and the distribution of the DNN output is shown in Figure 9.8 in the top CR. The yields in top CR are presented in Table 9.6. The top normalization factor simultaneously derived with the $Z \rightarrow \tau\tau$ normalization via simple matrix inversion.

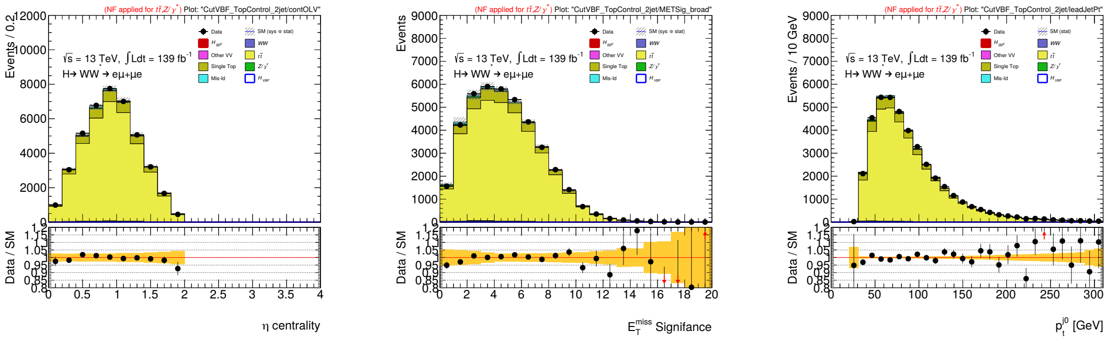
The top NF is measured to be 0.99 ± 0.01 (stat.).



(a) $\Delta\phi_{\ell\ell}$ (b) $m_{\ell\ell}$ (c) Δy_{jj}

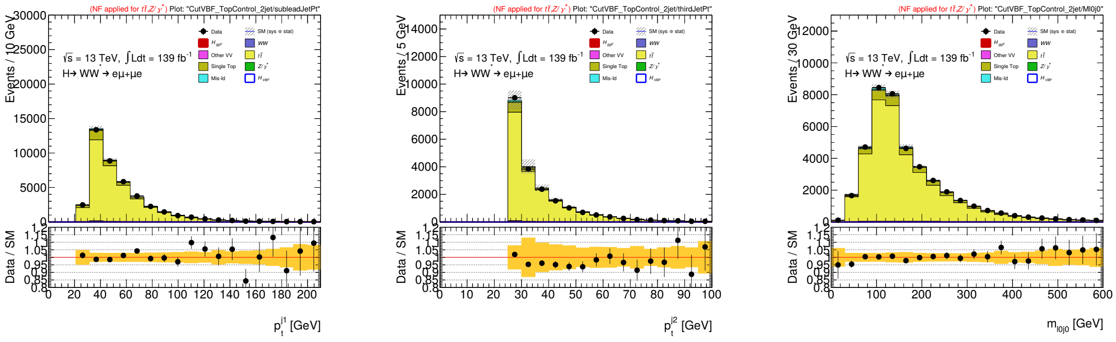


(d) m_{jj} (e) m_{T} (f) $p_{\mathrm{T}}^{\mathrm{tot}}$

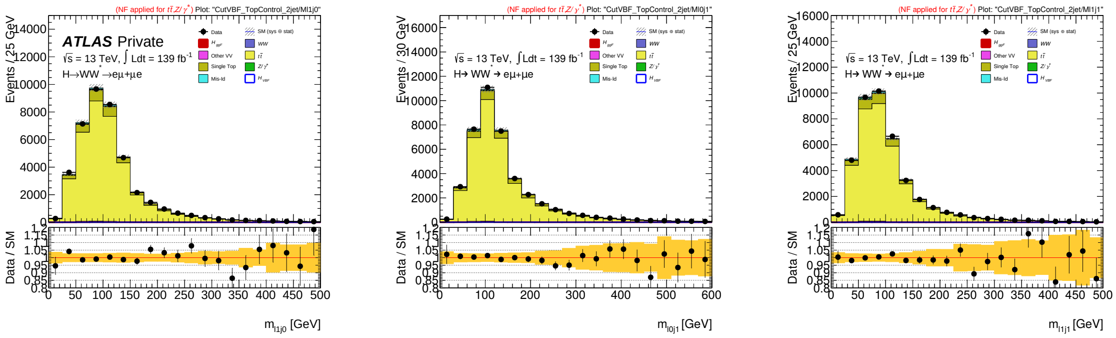


(g) $\eta_{\ell}^{\mathrm{centrality}}$ (h) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance (i) $p_{\mathrm{T}}^{\mathrm{jet}_1}$

Figure 9.6: Distributions of $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, Δy_{jj}, m_{jj}, m_{T}, $p_{\mathrm{T}}^{\mathrm{tot}}$, $\eta_{\ell}^{\mathrm{centrality}}$, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance, and $p_{\mathrm{T}}^{\mathrm{jet}_1}$ in the top CR. The yellow band represents the MC statistical uncertainties.



(a) $p_{\mathrm{T}}^{\mathrm{jet}_2}$ (b) $p_{\mathrm{T}}^{\mathrm{jet}_3}$ (c) $m_{\ell 1j1}$



(d) $m_{\ell 2j1}$ (e) $m_{\ell 1j2}$ (f) $m_{\ell 2j2}$

Figure 9.7: Distributions of $p_{\mathrm{T}}^{\mathrm{jet}_2}$, $p_{\mathrm{T}}^{\mathrm{jet}_3}$, $m_{\ell 1j1}$, $m_{\ell 2j1}$, $m_{\ell 1j2}$, and $m_{\ell 2j2}$ in the top CR. The yellow band represents the MC statistical uncertainties.

9.5.2 $Z \to \tau\tau$ Control Region

The modeling of the DNN input variables are shown in Figure 9.6 and 9.7, and the distribution of the DNN output is shown in Figure 9.8 in the top CR. The yields in top CR are presented in Table 9.6. The top normalization factor simultaneously derived with the $Z \to \tau\tau$ normalization via simple matrix inversion. The top NF is measured to be 0.99 ± 0.01 (stat.). The $Z \to \tau\tau$ control region is made orthogonal to the VBF SR by inverting one $Z \to \tau\tau$ veto ($|m_{\tau\tau} - m_Z| < 25$) that selects a decay product mass around the Z-peak, while keeping all the other selection criteria same as SR. Additionally a cut on $m_{ll} < 70$ GeV is applied to purify the region. The

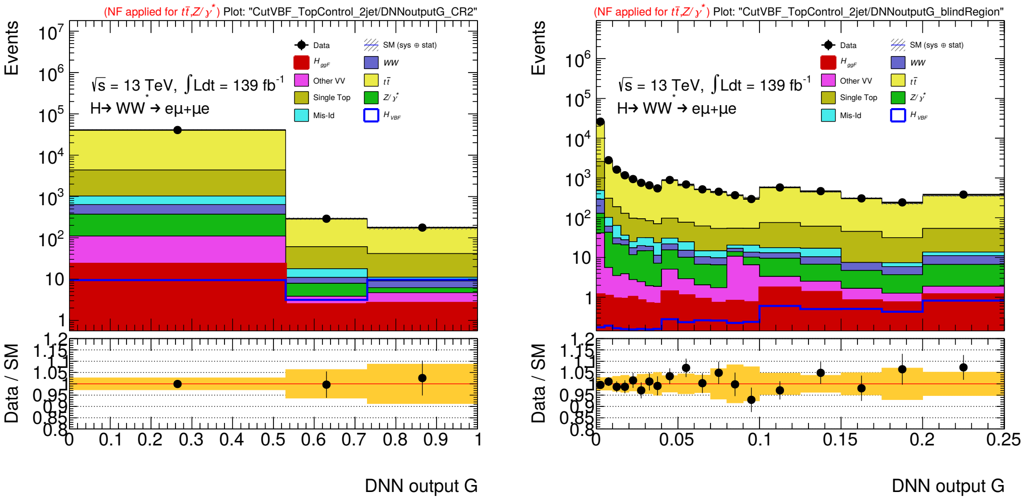
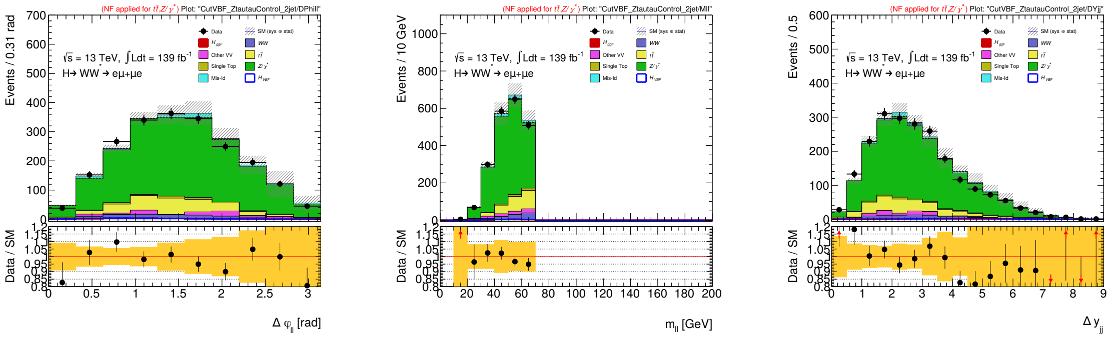
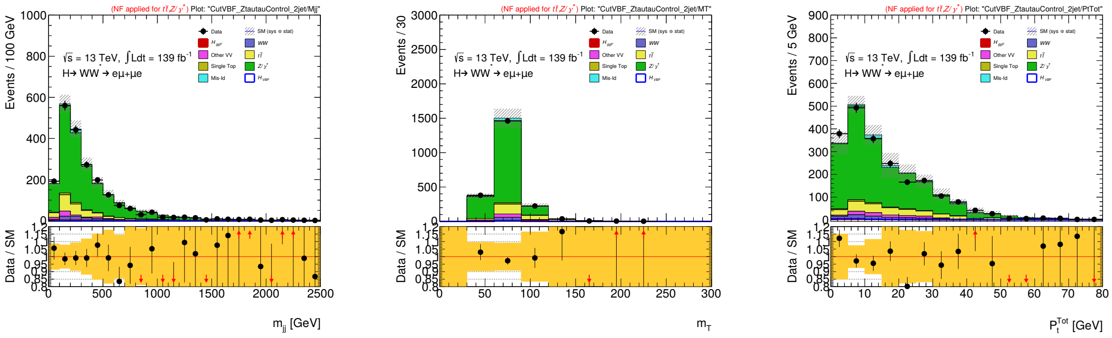


Figure 9.8: The DNN distribution in the top CR with a binning that ensures the signal is not unblinded (left) and the unblinded region with a finer binning (right). The yellow band represents the MC statistical uncertainties and the main sources of experimental uncertainties (JES/JER/MET/flavor-tagging).

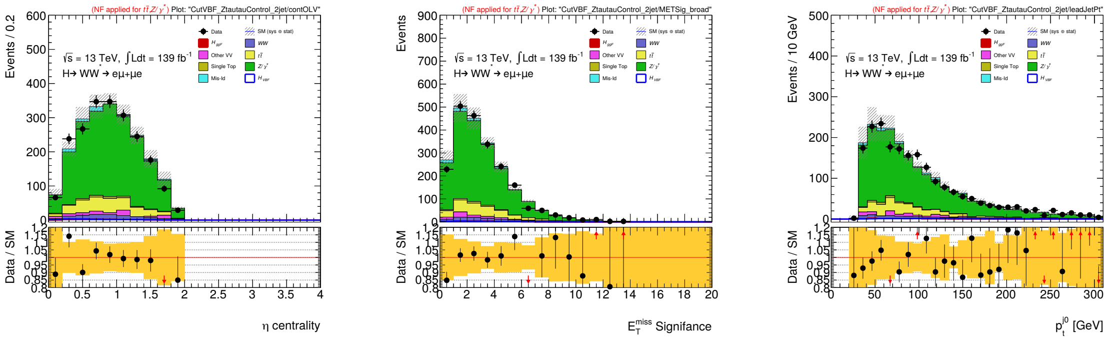
purity of the $Z \rightarrow \tau\tau$ control region is of about 77%. The modeling of the DNN input variables are shown in Figures 9.9 and 9.10, and the distribution of the DNN output is shown in Figure 9.11 in the $Z \rightarrow \tau\tau$ CR. The yields in the $Z \rightarrow \tau\tau$ are presented in Table 9.6. The $Z \rightarrow \tau\tau$ normalization factor simultaneously derived with the top normalization via simple matrix inversion. The $Z \rightarrow \tau\tau$ NF is measured to be 0.98 ± 0.03 (stat.).



(a) $\Delta\phi_{\ell\ell}$ (b) $m_{\ell\ell}$ (c) Δy_{jj}

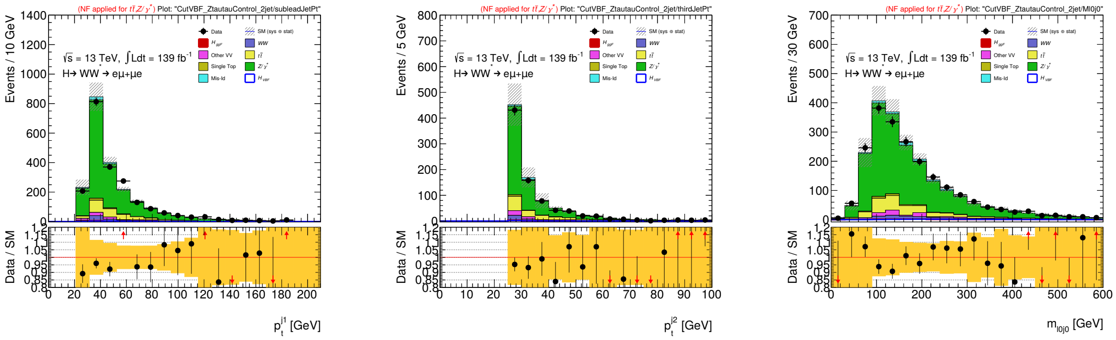


(d) m_{jj} (e) m_{T} (f) $p_{\mathrm{T}}^{\mathrm{tot}}$

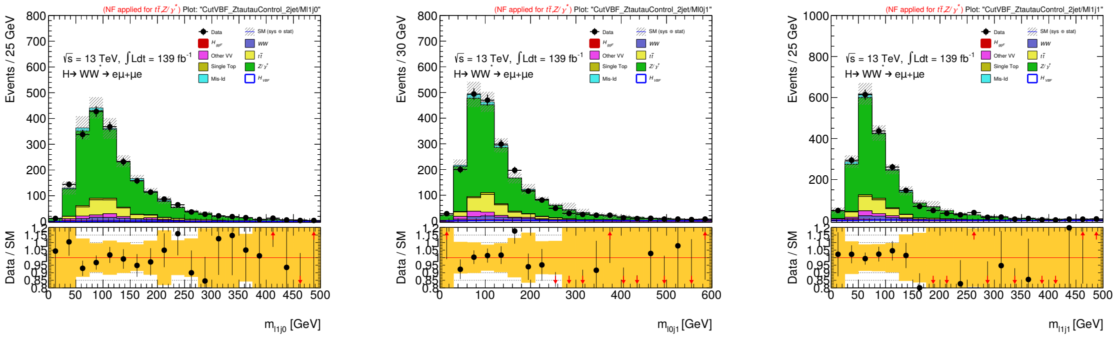


(g) $\eta_{\ell}^{\mathrm{centrality}}$ (h) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance (i) $p_{\mathrm{T}}^{\mathrm{jet}_1}$

Figure 9.9: Distributions of $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, Δy_{jj}, m_{jj}, m_{T}, $p_{\mathrm{T}}^{\mathrm{tot}}$, $\eta_{\ell}^{\mathrm{centrality}}$, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance, and $p_{\mathrm{T}}^{\mathrm{jet}_1}$ in the $Z \rightarrow \tau\tau$ CR. The yellow band represents the MC statistical uncertainties.



(a) $p_{\mathrm{T}}^{\mathrm{jet}_2}$ (b) $p_{\mathrm{T}}^{\mathrm{jet}_3}$ (c) $m_{\ell 1 j1}$



(d) $m_{\ell 2 j1}$ (e) $m_{\ell 1 j2}$ (f) $m_{\ell 2 j2}$

Figure 9.10: Distributions of $p_{\mathrm{T}}^{\mathrm{jet}_2}$, $p_{\mathrm{T}}^{\mathrm{jet}_3}$, $m_{\ell 1 j1}$, $m_{\ell 2 j1}$, $m_{\ell 1 j2}$, and $m_{\ell 2 j2}$ in the $Z \to \tau\tau$ CR. The yellow band represents the MC statistical uncertainties.

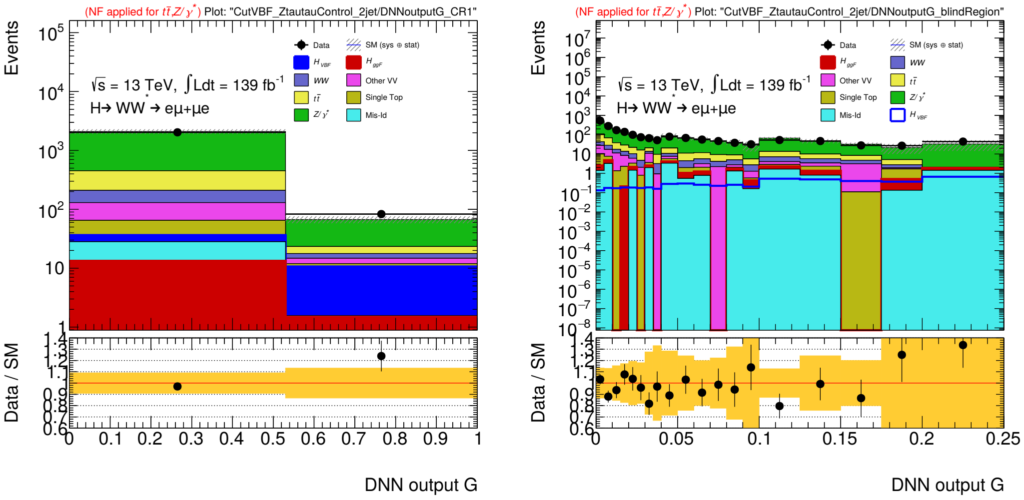


Figure 9.11: The DNN distribution in the $Z \rightarrow \tau\tau$ CR with a coarse binning and the signal stacked in the last bin (left) and the unblinded region with a finer binning (right). The yellow band represents the MC statistical uncertainties and the main sources of experimental uncertainties (JES/JER/MET/flavor-tagging).

9.5.3 WW Validation Region

The WW background process in association with 2 jets can be divided into two distinct types of processes:

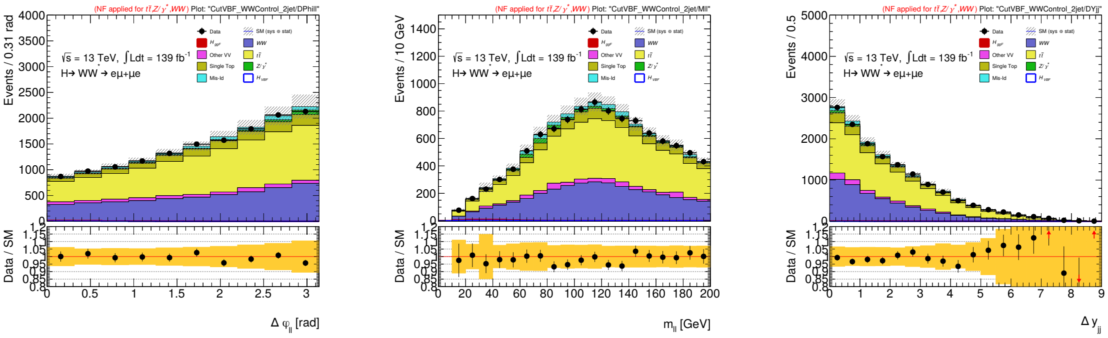
1. EW $WW + 2$ jets: containing only electroweak vertices,

2. QCD $WW + 2$ jets: containing a QCD vertex.

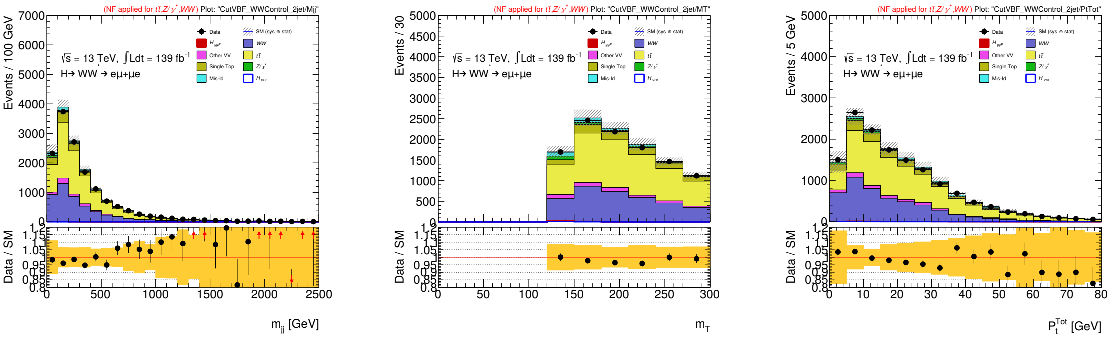
The cross section of the QCD WW+2 jets processes is more than an order of magnitude larger than the EW WW+2 jets one. However, the two contributions have more comparable impact in the VBF phase space region of large m_{jj} and Δy_{jj}. Due to the large $t\bar{t}$ contamination, the purity that can be obtained in the WW events enriched phase space is only between 34% and 46%. Therefore, a validation region being used to check the generators prediction for the QCD WW+2 jets with data while the MC prediction of WW is applied directly in the VBF SR. The following requirements define the $N_{jet} \geqslant 2$ WW validation region:

- VBF preselection;
- $m_{\mathrm{T}} > 130$ GeV;
- $m_{\mathrm{T2}} > 160$ GeV;

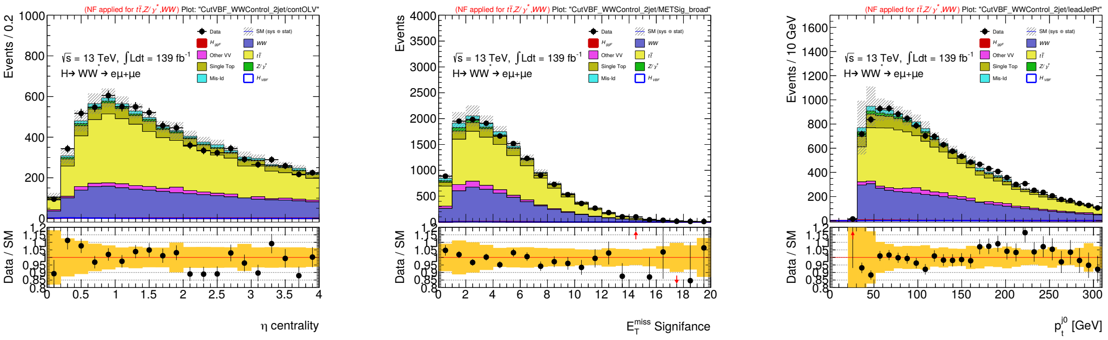
The m_{T2}, which gives a lower bound on the parent particle's mass [160], is defined as $m_{T2} = min_{p_T^1+p_T^2=p_T}(max(m_{\mathrm{T}}^2(p_T^1, p_T^a), m_{\mathrm{T}}^2(p_T^2, p_T^b)))$ The purity of the current validation region is 35%. The modeling of the kinematic distributions in the WW validation region is shown in Figures 9.12 and 9.13. The yields in WW VR are presented in Table 9.6. If the WW validation region is included in a 3x3 matrix inversion procedure, together with the top CR and the $Z \rightarrow \tau\tau$ CR, to extract the normalizations, the WW NF evaluates to 0.89 ± 0.03 (stat.). To reduce discrepancy between data and the MC, this factor in the distributions, as shown in Figures 9.12 and 9.13.



(a) $\Delta\phi_{\ell\ell}$ (b) $m_{\ell\ell}$ (c) Δy_{jj}

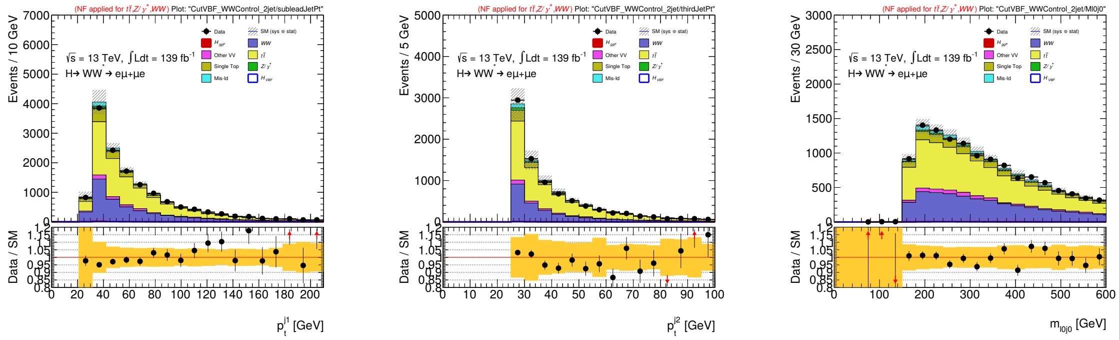


(d) m_{jj} (e) m_{T} (f) $p_{\mathrm{T}}^{\mathrm{tot}}$

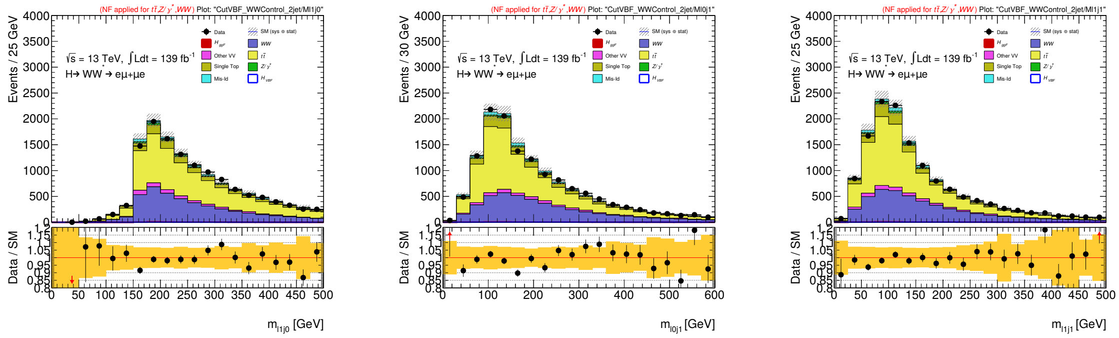


(g) $\eta_\ell^{\text{centrality}}$ (h) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance (i) $p_{\mathrm{T}}^{\mathrm{jet}_1}$

Figure 9.12: Distributions of $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, Δy_{jj}, m_{jj}, m_{T}, $p_{\mathrm{T}}^{\mathrm{tot}}$, $\eta_\ell^{\text{centrality}}$, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ Significance, and $p_{\mathrm{T}}^{\mathrm{jet}_1}$ in the WW VR. The yellow band represents the MC statistical uncertainties.



(a) $p_{\mathrm{T}}^{\mathrm{jet}_2}$ (b) $p_{\mathrm{T}}^{\mathrm{jet}_3}$ (c) $m_{\ell 1 j 1}$



(d) $m_{\ell 2 j 1}$ (e) $m_{\ell 1 j 2}$ (f) $m_{\ell 2 j 2}$

Figure 9.13: Distributions of $p_{\mathrm{T}}^{\mathrm{jet}_2}$, $p_{\mathrm{T}}^{\mathrm{jet}_3}$, $m_{\ell 1 j 1}$, $m_{\ell 2 j 1}$, $m_{\ell 1 j 2}$, and $m_{\ell 2 j 2}$ in the WW VR. The yellow band represents the MC statistical uncertainties.

9.6 Misidentified Lepton Background Estimation

Physics processes with leptons in their final states, as in the $H \to WW^* \to \ell\nu\ell\nu$, have clean signatures because of the significant background rejection due to the excellent lepton identification of the ATLAS detector. With sufficiently tight lepton identification criteria in ATLAS, misidentification rates of events containing only QCD interactions can be heavily reduced to a range within $10^{-5} - 10^{-4}$. However, the remaining jets that are misidentified as leptons populate the non-gaussian tails of the detector response, making them very difficult to model using simulation. Source of misidentified leptons depend upon their flavor, event topology and the lepton identification criteria used. Lepton will refer to electrons or muons, not taus. For electrons, they can originate for instance from charged hadrons, conversion of photons[1], or semi-leptonic heavy-flavor decays. In the latter two cases, although an actual electron is present in the final state from a secondary process, it is still considered misidentified in the sense that it is not produced in isolation as the result of a hard electroweak scattering event (which, in contrast, is referred to as a "prompt" lepton). For muons, nearly all fakes[2] originate from either semi-leptonic heavy-flavor decay or meson decay in flight - both of which contain a real non-prompt muon.

Despite the small misidentification rate, W+jets processes, in which one prompt lepton originates from the W boson decaying leptonically in association with one or more jets where one of them is misidentified as a second lepton, constitutes a non-negligible background due to its large cross section which is more than three orders of magnitude higher than the Higgs boson production, see Figure 9.14.

---

[1] The photons originating in turn from e.g. bremsstrahlung as well as initial or final state radiation.

[2] For this analysis, the terms "fake lepton", "fake" or "misidentified lepton" will refer to cases where a jet is misidentified as a lepton.

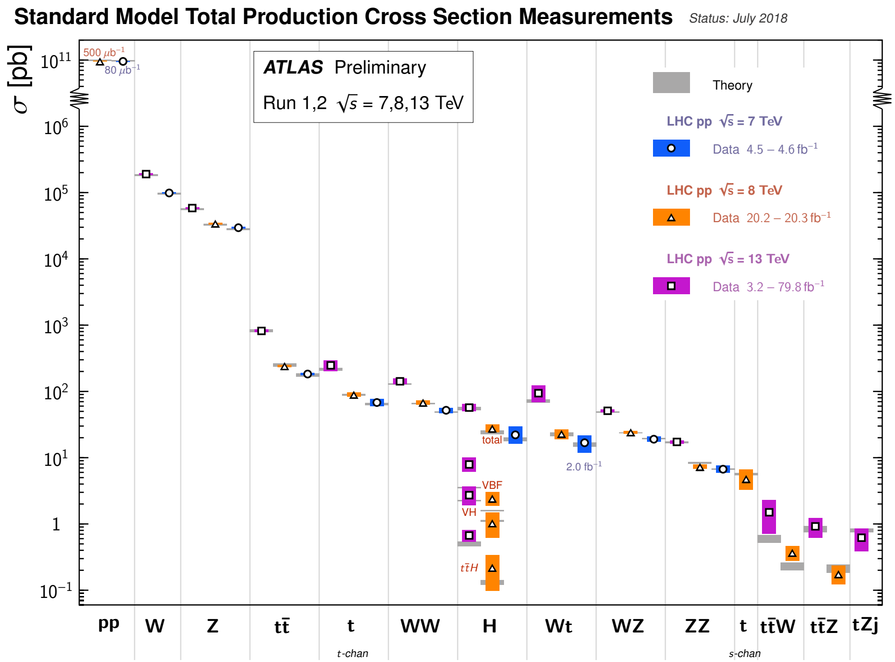


Figure 9.14: Production cross sections of different SM processes as measured by ATLAS, along with theoretical predictions. In purple at the values at $\sqrt{s} = 13$ TeV. Theoretical predictions, calculated at NLO or higher, are shown in grey. Figure taken from [18].

A contribution is also present from pure QCD processes in which two jets are simultaneously misidentified as prompt leptons. However, these backgrounds are significantly suppressed by targeting final states containing high-energy neutrinos, for example by requiring a large missing transverse momentum.

Given the shortcomings of the Monte Carlo modeling of fake leptons, the data-driven fake-factor method is used[3] in this analysis to estimate the W+jets background.

---

[3] Converted photons are not estimated with the fake-factor method, but use $W\gamma^*$ MC.

9.6.1 The Fake Factor Method

The fake factor method for estimating misidentified lepton backgrounds consists broadly of two steps: (1) selecting a *control sample* of observed data enriched in misidentified objects, and (2) applying an *extrapolation factor* to transfer the control sample yields to the signal region. The control sample is designed to be kinematically similar to the signal region by requiring all the kinematic selections same as the signal region, but for one of the two leptons to fail the identification criteria (see section 9.2.1). This lepton is required to satisfy a looser set of selections, and is referred to as the *anti-identified* (anti-id) lepton. The other lepton which satisfies the same identification criteria as the signal region is referred to as the *identified* (id) lepton. The selection criteria for *id* and *anti-id* electrons and muons are listed in Table 9.7.

Table 9.7: Requirements for fully identified and anti-identified leptons.

Electron		Muon	
identified	anti-identified	identified	anti-identified
$p_T > 15$ GeV	$p_T > 15$ GeV	$p_T > 15$ GeV	$p_T > 15$ GeV
$\lvert\eta\rvert < 2.47$,excluding $1.37 < \lvert\eta\rvert < 1.52$	$\lvert\eta\rvert < 2.47$,excluding $1.37 < \lvert\eta\rvert < 1.52$	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
$\lvert z_0 \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \sin\theta\rvert < 0.5$ mm
$\lvert d_0\rvert/\sigma(d_0) < 5$	$\lvert d_0\rvert/\sigma(d_0) < 5$	$\lvert d_0\rvert/\sigma(d_0) < 3$	$\lvert d_0\rvert/\sigma(d_0) < 15$
Pass LHTight if $p_T < 25$ GeV Pass LHMedium if $p_T > 25$ GeV Pass FCTight isolation	Pass LHLoose	Pass Quality Tight Pass FCTight isolation	Pass Quality Medium
AUTHOR = 1	AUTHOR = 1		
	Veto against identified electron		Veto against identified muon

The number of events in the control sample can be expressed as

$$N_{id+\not{id}} = N_{id+\not{id}}^{W+\text{jets}} + N_{id+\not{id}}^{\text{multijet}} + N_{id+\not{id}}^{\text{prompt}} \tag{9.11}$$

where the strike-through denotes anti-id. Due to requirement of at least two forward jets in VBF signal topology, the contribution of multijet background is found to be substantial.

The yield of the misidentified leptons in the control sample is therefore obtained by subtracting the contribution of prompt lepton background, using MC, from the observed data yield. The extrapolation factor, denoted as “fake-factor”, is then applied to the fake lepton yield in the control sample to obtain the estimated fake yield in signal region,

$$N_{id+id}^{W+\text{jets}} = FF \times N_{id+\not{id}}^{W+\text{jets}} = FF \times (N_{id+\not{id}} - N_{id+\not{id}}^{\text{prompt}} - N_{id+\not{id}}^{\text{multijet}}) \tag{9.12}$$

The fake-factor is measured separately for fake eletrons and fake muons, as a function of p_{T} and $|\eta|$, using a Z+jets sample (discussed in Section 9.6.2) for W+jets background, and using a multijet sample (discussed in Section 9.6.4) for multijet background, as

$$FF(p_{\mathrm{T}}, |\eta|) = \frac{N_{id}}{N_{\not{id}}}(p_{\mathrm{T}}, |\eta|) \tag{9.13}$$

9.6.2 Z+jets Fake Factor

In the fake factor method, it is assumed that the fake background efficiency only depends on the properties of the lepton (not on the remainder of the event). This is a reasonable assumption since lepton reconstruction only takes information from a small detector region. Therefore, the fake factor estimated from Z+jets with three lepton selection can be applied on W+jets with two lepton selection. The leading order Feynman diagrams for Z and W boson production in association with a quark or gluon are shown in Figure 9.15.

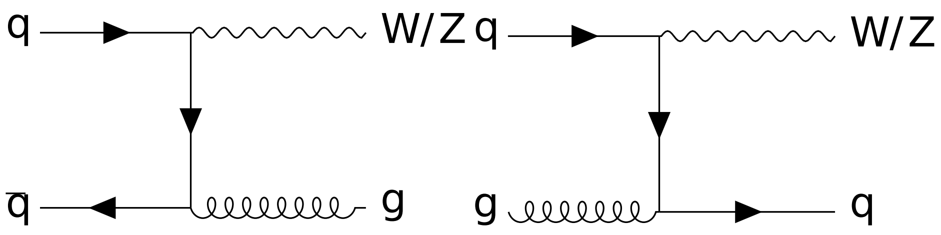


Figure 9.15: Leading order Feynman diagrams for vector boson plus gluon (left) or quark (right) production in pp collisions.

The Z+jets fake factor is expressed as:

$$FF = \frac{N_{id,id,id} - N^{non-Z+jets}_{id,id,id}}{N_{id,id,i\not{d}} - N^{non-Z+jets}_{id,id,i\not{d}}} \tag{9.14}$$

where the superscript i stands for an ID and a for an Anti-ID lepton.

To derive the fake factor in Equation 9.14, an analysis region is defined, which is sufficiently pure in Z+jets. Events with three leptons with $p_{\mathrm{T}} > 15$ GeV are considered. The leptons from Z decay are tagged by requiring the leptons to be of same flavor and opposite charge. These leptons are identified with the same criteria as ID criteria in Table 9.7 with the exception that they only need to pass FCLoose isolation and a “loose-with-b-layer” electron or “medium” muon quality working point. If they are electrons, they need to be in an invariant-mass window of [80, 110] GeV, while muons need to be in a window of [70, 110] GeV. If at least one of the two leptons is matched to an object firing the single-lepton trigger, the leptons are accepted as Z-boson candidates. If multiple lepton combinations satisfy this requirement, the pair with the invariant mass closest to the Z boson is chosen. The remaining lepton is the fake candidate. When applied to a Z+jets MC sample, this algorithm correctly assigns leptons to the Z boson in about 99% of all cases. The reconstructed Z-boson peak can be seen in Figure 9.16.

To reject the leptonic WZ background, $m_T^W = \sqrt{2E_{\mathrm{T}}^{\mathrm{miss}} E_{\mathrm{T}}^{\text{fake cand.}}(1 - \cos\phi)} <$ 50 GeV is applied, where ϕ is the angle between the MET and the fake candidate. The region above a transverse mass m_T^W of 50 GeV, where the fake candidate is required to pass the ID criteria, is used as a control region for the WZ process (with normalization factor of 0.99 ± 0.01 extracted from Figure 9.17). The pseudo-rapidity requirement on the fake candidate is $|\eta| < 2.5$ for muons and $|\eta| < 2.47$ for electrons. Then, for each fake lepton type, electron and muon, two regions are constructed depending on whether the fake lepton passes the ID or Anti-ID cuts.

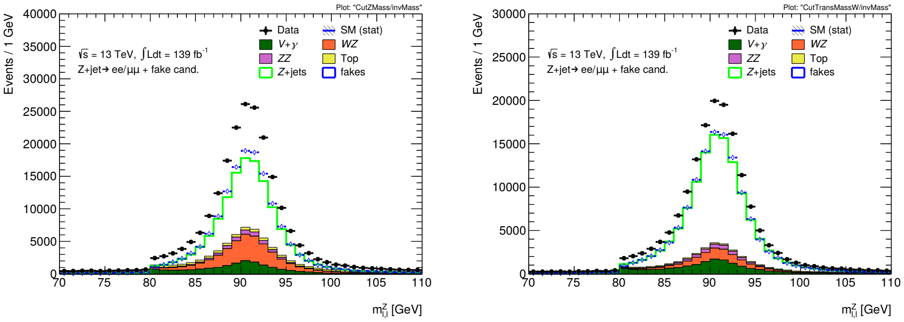


Figure 9.16: Distributions of the invariant mass of the reconstructed Z-boson candidate. Normalization factors are not applied. The stacked histograms are background MC processes not including Z+fake. The measured data is shown in black datapoints. The blue datapoints are the data-driven Z+fake estimate. They are calculated by taking the difference between data and the stacked MC processes.

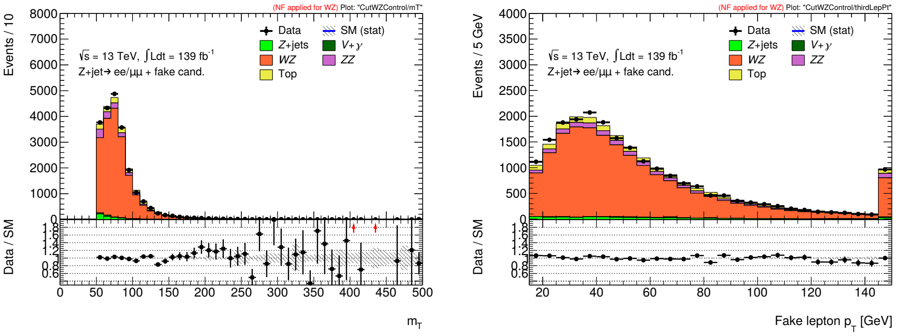


Figure 9.17: Distributions in the WZ CR. A cut of $m_{\mathrm{T}} > 50$ GeV and the requirement that the fake candidate lepton is ID are applied for the WZ CR. A normalization factor of 0.993 is applied for WZ.

The distributions of lepton kinematic variables of the fake candidate leptons are shown in Figure 9.18. The Z+jets prediction agrees with the data-driven estimate of the fake yield quite well in the ID selection. In the AntiID selections of electron fake candidates, MC overestimates the fake yield slightly. The opposite is the case in the muon AntiID selection.

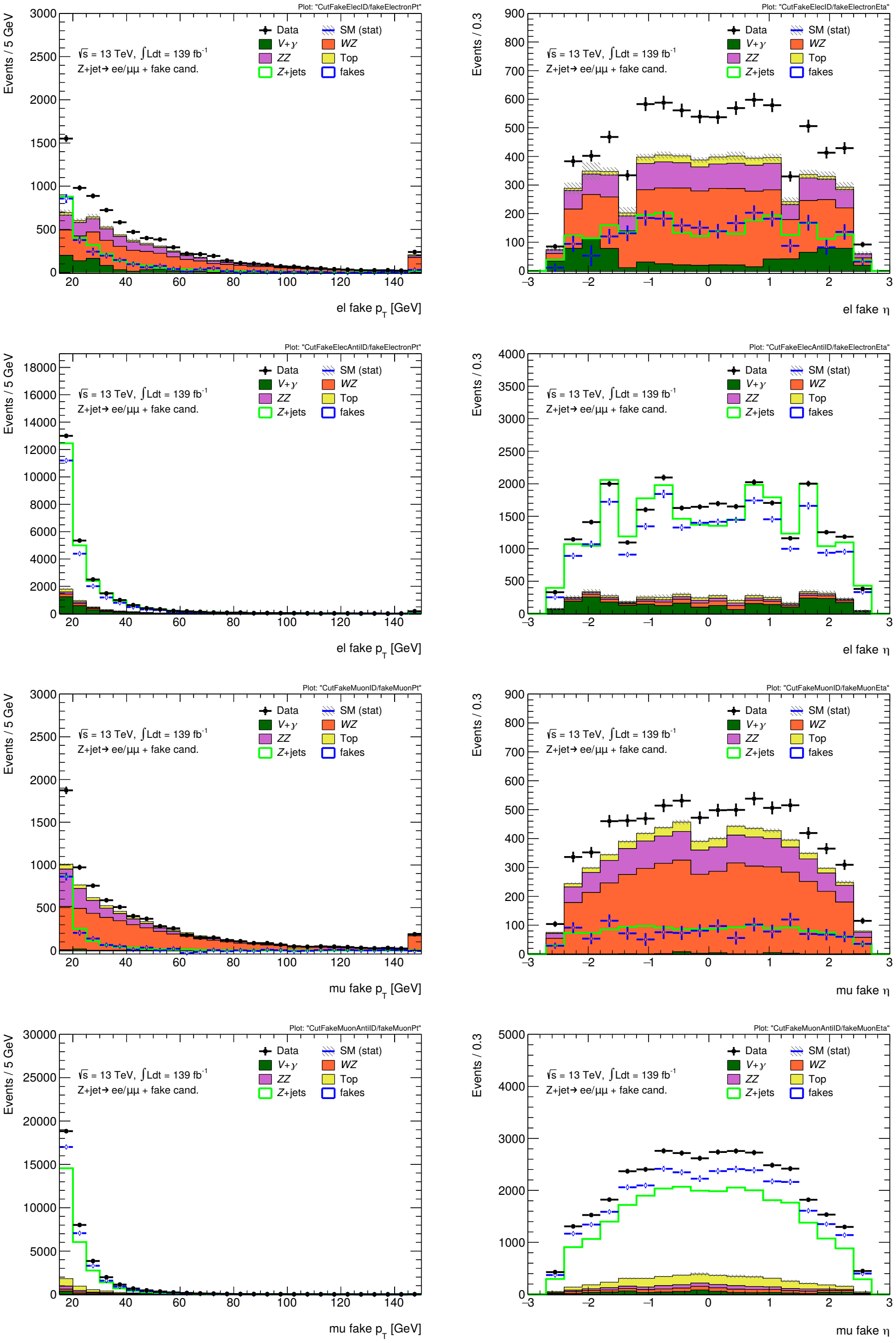


Figure 9.18: Kinematic distributions of the fake lepton candidate. The top four plots are electron fake candidates and the bottom four muon fakes. Between these four plots, the top two show the ID selection and the bottom two the AntiID selection. The transverse momentum is shown on the left and the pseudo-rapidity on the right-hand side.The stacked histograms are background, i.e. not Z+fake, MC processes. The difference between the data yield and the stacked histograms constitutes the estimate for Z+fake. The fake estimate is also shown in blue markers and compared to the not-stacked Z+jets MC prediction.

Table 9.8: Full cutflow of the Z+jets fake factor analysis. The normalization factor applied on WZ yield that is derived in the WZ control region.

$\sqrt{s} = 13TeV$, $\mathcal{L} = 139fb^{-1}$	Z+jets Powheg	$V+\gamma$	WW	WZ	ZZ	Top	VVV	Data	ZFake purity(%)
Channel Selection	246181.21 ± 268.27	105001.82 ± 743.46	1002.92 ± 8.44	49267.44 ± 1727.21	14923.47 ± 114.48	86488.80 ± 62.61	240.09 ± 0.94	605458	57.57 ± 0.34
lll final state	246181.21 ± 268.27	105001.82 ± 743.46	1002.92 ± 8.44	49267.44 ± 1727.21	14923.47 ± 114.48	86488.80 ± 62.61	240.09 ± 0.94	605458	57.57 ± 0.34
All lep $p_T > 15$ GeV	246181.21 ± 268.27	105001.82 ± 743.46	1002.92 ± 8.44	49267.44 ± 1727.21	14923.47 ± 114.48	86488.80 ± 62.61	240.09 ± 0.94	605458	57.57 ± 0.34
Overlap: Vgamma/Vjets	246181.21 ± 268.27	105001.82 ± 743.46	1002.92 ± 8.44	49267.44 ± 1727.21	14923.47 ± 114.48	86488.80 ± 62.61	240.09 ± 0.94	605458	57.57 ± 0.34
Z-tagging	139471.86 ± 198.06	18307.04 ± 317.06	125.51 ± 2.91	32270.51 ± 45.33	5995.85 ± 26.42	10694.83 ± 20.85	86.15 ± 0.44	214484	68.54 ± 0.30
TransMassW	125456.68 ± 187.76	15585.31 ± 296.36	45.58 ± 1.87	10126.93 ± 23.07	4267.69 ± 22.15	4059.09 ± 12.90	21.99 ± 0.22	161198	78.84 ± 0.37
fake lepton ID or AntiID	63284.26 ± 136.45	4592.08 ± 163.39	20.73 ± 1.22	9726.48 ± 22.73	4036.29 ± 21.04	3515.93 ± 11.96	21.14 ± 0.21	86206	74.58 ± 0.47
fake type: electron	34343.60 ± 96.99	3850.64 ± 142.84	12.72 ± 1.01	4375.13 ± 16.19	1849.83 ± 11.37	801.10 ± 5.37	9.42 ± 0.14	41125	73.50 ± 0.71
fake el η cut	27149.91 ± 86.45	3417.59 ± 135.42	9.93 ± 0.93	4131.62 ± 15.89	1742.41 ± 10.48	741.56 ± 5.16	8.93 ± 0.14	34003	70.44 ± 0.78
Normalization factors				**NF = 0.99± 0.01**					
ID cuts	2429.00 ± 26.18	789.22 ± 63.23	1.24 ± 0.31	3367.53 ± 14.08	1272.17 ± 7.08	277.19 ± 2.63	7.40 ± 0.12	7996	28.53 ± 1.42
Normalization factors				**NF = 0.99± 0.01**					
Anti-ID cuts	24720.92 ± 82.39	2628.37 ± 119.75	8.69 ± 0.88	736.64 ± 7.13	470.23 ± 7.72	464.37 ± 4.44	1.53 ± 0.06	26007	83.43 ± 0.93
fake type: muon	28940.67 ± 95.98	741.44 ± 79.33	8.00 ± 0.69	5351.35 ± 15.95	2186.46 ± 17.70	2714.83 ± 10.68	11.72 ± 0.16	45081	75.57 ± 0.62
fake mu η cut	28059.83 ± 94.46	692.08 ± 77.92	7.65 ± 0.69	5240.16 ± 15.77	2142.10 ± 17.54	2637.08 ± 10.51	11.45 ± 0.16	43638	75.41 ± 0.63
Normalization factors				**NF = 0.99± 0.01**					
ID cuts	1429.37 ± 21.21	26.49 ± 9.87	1.01 ± 0.24	4349.23 ± 14.46	1364.69 ± 11.69	387.70 ± 3.28	9.54 ± 0.15	7464	17.76 ± 1.21
Normalization factors				**NF = 0.99± 0.01**					
Anti-ID cuts	26630.46 ± 92.05	665.59 ± 77.29	6.65 ± 0.64	856.10 ± 6.03	777.41 ± 13.08	2249.38 ± 9.98	1.92 ± 0.07	36174	87.40 ± 0.73

Due to small dependence of the electron fake factor dependence on η, the electron fake factor is only measured as a function of p_{T}. The binning in p_{T} is mainly determined by the limited statistics and high background contamination in the ID selections. Since fake leptons are very rare at high p_{T}, the fake factor is only binned at small transverse momenta.

For muons, the bin $p_{\mathrm{T}} > 50$ GeV have very low muon fake yield so the yield is extrapolated from the neighboring p_{T} bin. MC simulation of the Z+jets process is used to determine the ratio between the fake yields in the region above 50 GeV and the region in the next-highest bin between 25 and 50 GeV. This ratio is used as an extrapolation factor to estimate the fake factor in the highest p_{T} bin. The extrapolation factor is extracted in POWHEGand SHERPAindependently and averaged. The difference between the two simulations is used as modeling uncertainty of the extrapolation factor. It is, together with the statistical uncertainty of the extrapolation factor, propagated to the statistical uncertainty of the muon fake factor in the highest p_{T} bin. It can be seen that this uncertainty is still much smaller than the electroweak subtraction uncertainty. The fake factors and their statistical uncertainties are shown in Figure 9.19.

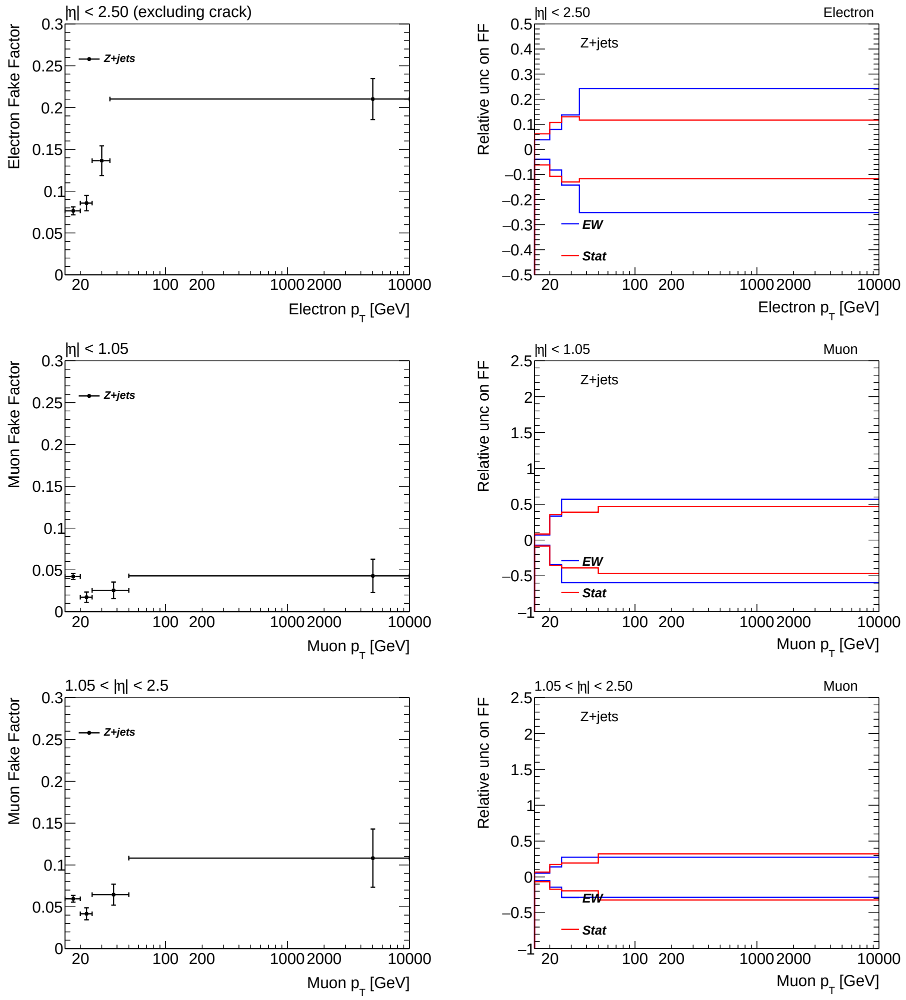


Figure 9.19: Fake factors derived in the three-lepton selection. The top two plots show the fake factor (left) and its relative uncertainties (right) as a function of p_{T} for electrons. The middle two plots show the muon fake factors in the low-η region and the bottom two plots the same in the high-η region. The values are also listed in Table 9.9.

The uncertainties on the Z+jets fake factors have contributions from systematic and statistical sources. The sources of systematic uncertainties are electroweak background subtraction (discussed in Section 9.6.3) and differences of fake factors between Z+jets and W+jets processes (discussed in Section 9.6.5.2). The uncertainties on the Z+jets fake factor are listed in Table 9.9.

Table 9.9: Summary of the fake factors from the Z+jets estimate with uncertainties. All uncertainties are quoted in percent on the nominal value. The column *Total* gives total uncertainty on the Z+jets fake factor by taking the quadrature sum of all individual contributions in the previous columns. In the fit, however, the EW subtraction and sample composition uncertainties are treated as correlated between different bins, while the statistical uncertainty is uncorrelated.

Kinematic Region ($\|\eta\|$ and p_{T} range)	Nominal	Statistical	EW Subtraction	Sample Composition	Total
Electron:					
$0.0 < \|\eta\| < 2.5$					
$15.0 - 20.0$ GeV	0.076	6.2	3.9	7.5	10
$20.0 - 25.0$ GeV	0.086	11	8.1	31	34
$25.0 - 35.0$ GeV	0.14	13	14	7.4	20
$35.0 - \infty$ GeV	0.21	12	25	21	35
Muon:					
$0.0 < \|\eta\| < 1.05$					
$15.0 - 20.0$ GeV	0.042	8.4	7.1	8.1	14
$20.0 - 25.0$ GeV	0.017	35	34	11	50
$25.0 - 50.0$ GeV	0.026	39	58	11	71
$50.0 - \infty$ GeV	0.043	47	58	11	75
$1.05 < \|\eta\| < 2.5$					
$15.0 - 20.0$ GeV	0.060	6.7	5.3	8.1	12
$20.0 - 25.0$ GeV	0.042	17	14	11	25
$25.0 - 50.0$ GeV	0.065	19	28	11	36
$50.0 - \infty$ GeV	0.11	32	28	11	44

9.6.3 Electroweak Subtraction Uncertainty

In order to calculate the fake factor in Equation (9.14), non-Z+jets processes (processes with real leptons) are subtracted from the data yield using MC. Theoretical uncertainties in the normalization of these MC processes can cause a systematic uncertainty in the fake factor, which is referred to as electroweak (EW) subtraction uncertainty. The Anti-ID region in the fake factor derivation has high purity (> 80%) for fakes (the cutflow in Table 9.8), while the ID region is mainly populated with electroweak backgrounds. The MC uncertainties, therefore, affect the ID region in the numerator of the fake factor in Equation (9.14).

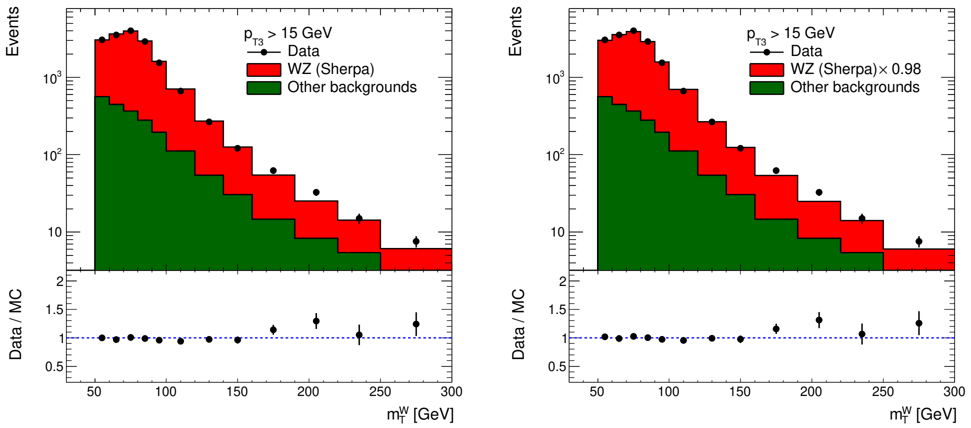
The largest contribution of electroweak backgrounds in the Z+jets ID region comes from WZ and ZZ. For electron fakes, $Z + \gamma$ is also a significant contributor. Theory uncertainties for these processes have been derived and propagated to the fake factor, discussed in the following paragraphs.

9.6.3.1 Normalization of WZ Background

Since WZ background is the largest electroweak background in the Z+jets enriched region, its contribution should be correctly estimated to get a reliable Z+jets fake factor. Therefore, simulations for the WZ process is normalized to data by first defining a WZ enriched control sample and then the WZ MC sample is normalized to data. The WZ CR is defined by requiring the same event selection for the identified leptons as the Z+jets fake factor selection, but inverting the m_T^W requirement, i.e., selecting events with $m_T^W > 50$ GeV with the fully identified lepton. The inversion of m_T^W requirement makes the WZ CR orthogonal to the fake factor estimation sample. In order to estimate the normalization factor, a χ^2 fit is performed to the transverse mass distribution as a function of the WZ scale factor. The χ^2 function is thus defined as:

$$\chi^2 = \sum_k \frac{(x_k - \phi_k(\alpha)^2)}{\sigma_k{}^2} \tag{9.15}$$

In Equation 9.15, the index k runs over bins in the m_T distribution. The $\phi_k(\alpha)$ is defined as $\phi_k(\alpha) = B_k + \alpha{\cdot}S_k$. The B_k is the total number of background events in bin k, S_k is the total number of WZ events in that bin and α is a global normalization factor. The x_k is the number of data events reconstructed in bin k, while σ_k is the sum in quadrature of data and MC statistical uncertainties. The WZ normalization factor (NF) comes out to be 0.98 ± 0.01 (stat.) with $\chi^2/ndof = 17.6/11$ (see Figure 9.20) indicating a good modeling of WZ process using Sherpa 2.2 generator for the Z+jets enriched sample.



(a) prefit (b) postfit (χ^2/dof = 17.6/11)

Figure 9.20: WZ normalization factor calculated using m_T^W distribution.

9.6.3.2 Systematic Uncertainties from Electroweak Backgrounds

Three sources of WZ normalization uncertainties are studied in this section: the p_{T} modeling of the third (fake-able) lepton, the Z+jets modeling in the WZ CR as MC simulation is used for Z+jets here, and the WZ modeling. A good overall agreement in data to MC for the p_{T} modeling of the third lepton can be seen in Figure 9.21. To understand the p_{T} dependence of the mismodeling of third lepton, the p_{T} of fake-able lepton divided into 3 regions: 15 GeV $< p_{\mathrm{T}} <$ 35 GeV, 35 GeV $< p_{\mathrm{T}} <$ 80 GeV, $p_{\mathrm{T}} >$ 80 GeV. The normalization factors (NFs) for each of these three regions is estimated using a χ^2 fit as described in Section 9.6.3.1. The NFs for each of these regions are summarized in Table 9.10 and the third lepton p_{T} modeling after applying the NFs in their respective regions are shown in Figures 9.22, 9.23, and 9.24.

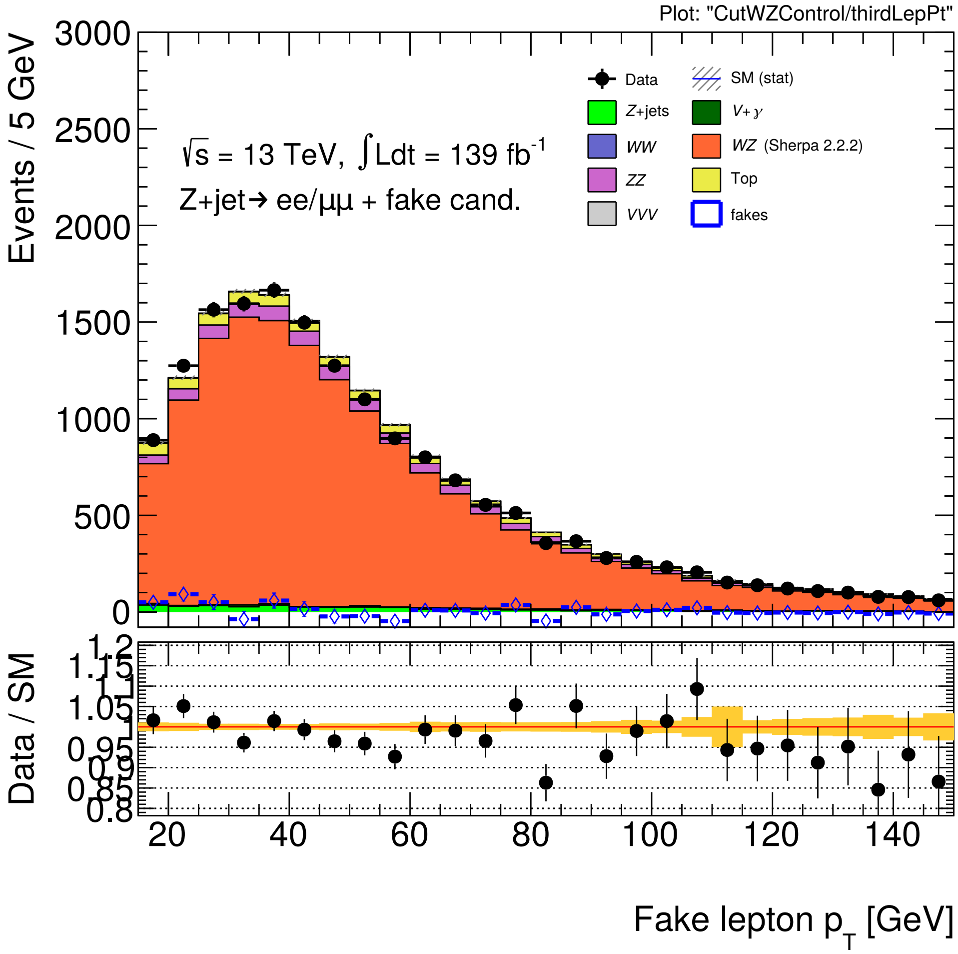
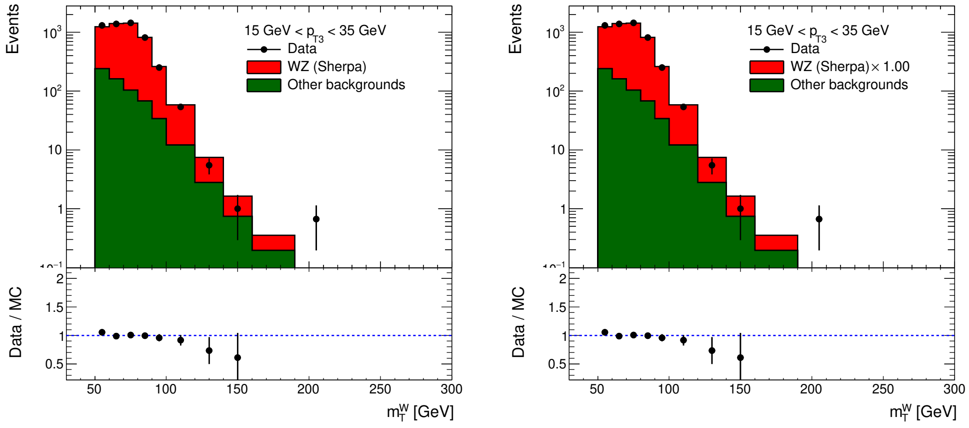


Figure 9.21: Third lepton p_{T} modeling in WZ CR.

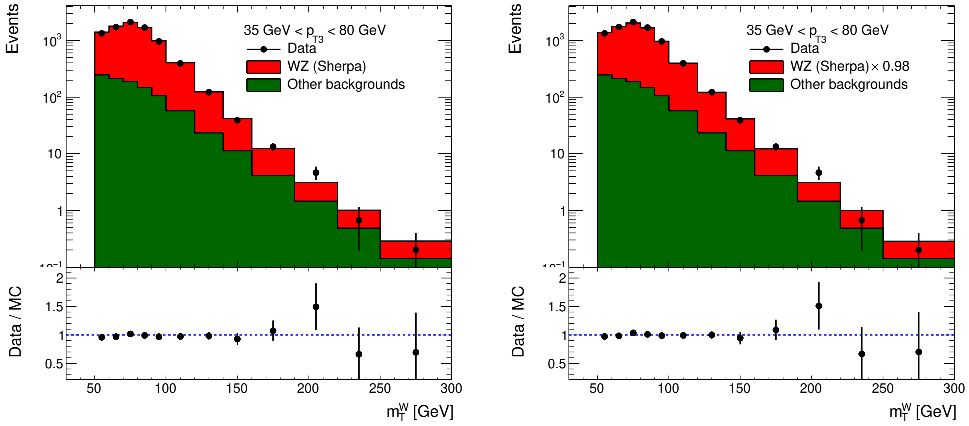
Table 9.10: WZ NF in third lepton p_{T} bins.

WZ CR, $p_{\mathrm{T}3}$ range	WZ NF	$chi^2/ndof$
$p_{\mathrm{T}3} > 15$ GeV	0.97 ± 0.01 (stat.)	19.4/11
15 GeV $p_{\mathrm{T}3} > 35$ GeV	0.99 ± 0.02 (stat.)	7.5/7
35 GeV $p_{\mathrm{T}3} > 80$ GeV	0.98 ± 0.01 (stat.)	12.1/10
$p_{\mathrm{T}3} < 80$ GeV	0.94 ± 0.03 (stat.)	17.9/11



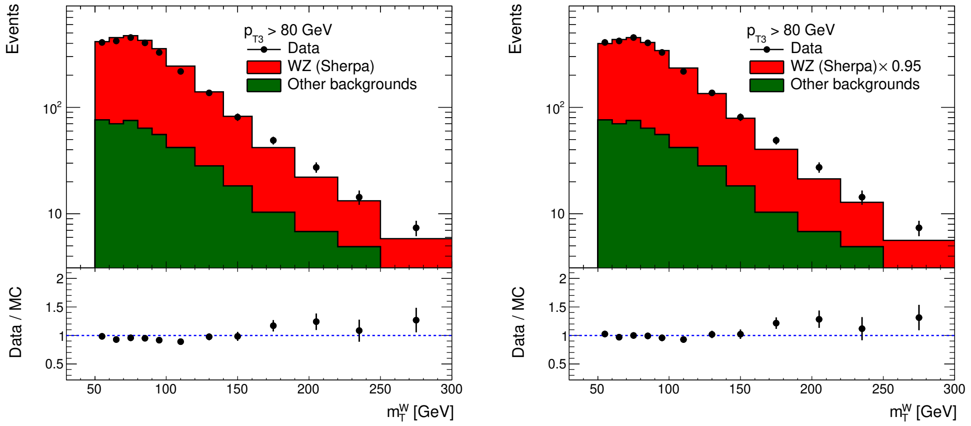
(a) prefit (b) postfit (χ^2/dof = 8.2/7)

Figure 9.22: WZ NF in WZ CR (15 GeV $< p_{T3} <$ 35 GeV).



(a) prefit (b) postfit (χ^2/dof = 6.8/11)

Figure 9.23: WZ NF in WZ CR (35 GeV $< p_{T3} <$ 80 GeV).



(a) prefit (b) postfit (χ^2/dof = 14.5/11)

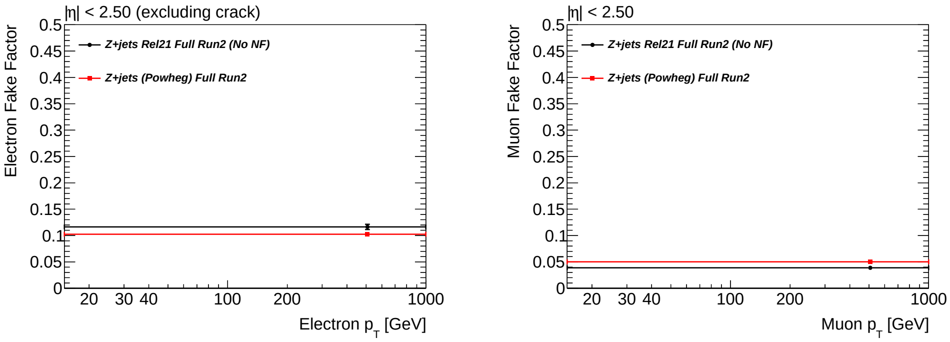
Figure 9.24: WZ NF in WZ CR ($p_{T3} > 80$ GeV).

The variation in NF due to p_{T} dependence is given by:

$$\Delta NF = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(NF_i - NF)^2} = 0.02 \tag{9.16}$$

The relative uncertainty from third lepton p_{T} modeling is then estimated as $\frac{\Delta NF}{NF} =$ 2%, where $NF = 0.98 \pm 0.01$ is taken from Section 9.6.3.1.

Since MC is used for Z+jets fakes while calculating WZ NF, Z+jet mismodeling due MC is a source of systematic uncertainty to the WZ normalization. The nominal generator Sherpa is compared with the alternative generator Powheg for simulating Z+jets fakes by looking at the Z+jets fake factor from both the generators as shown in Figure 9.25. The factor factor with Powheg varies by 30% (up for muons and down for electrons) with respect to Sherpa (nominal).



(a) electron fakefactor (b) muon fakefactor

Figure 9.25: Variation in fake factor (η and p_{T} bins integrated) due to Z+jets generator variation.

So, the systematic uncertainty is estimated by varying the Z+jets MC up and down by a factor of 1.3 and then evaluating the variation in WZ NF as given in Table 9.11.

Table 9.11: Variation in WZ NF by varying Z+jets MC normalization.

Z+jets MC variation	Up (×1.3)	Down (×1/1.3)
$\chi^2/ndof$	16.6/11	18.7/11
WZ NF	0.97 ± 0.01 (stat.)	0.99 ± 0.01 (stat.)

$$\Delta NF = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(NF_i - NF)^2} = 0.01 \tag{9.17}$$

The relative uncertainty due to Z+jets MC modeling is estimated $\frac{\Delta NF}{NF} = 1\%$.

The uncertainty due to the modeling of WZ process is obtained by comparing the transverse mass distribution from the nominal SHERPA WZ sample with the POWHEG +PYTHIA samples as shown in Figure 9.26.

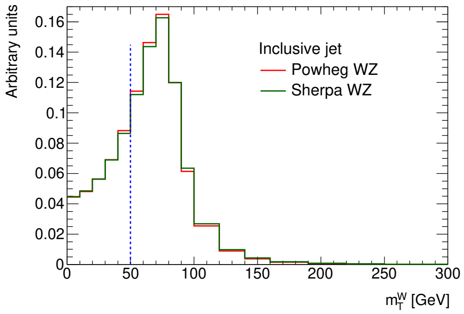


Figure 9.26: Variation in m_T^W distribution using nominal SHERPA and alternate POWHEG WZ samples. It is used to calculate the variation extrapolation factor (between Z+jets enriched SR and WZ CR). WZ generator uncertainty.

The extrapolation factor(R) is then obtained using both MC distributions using:

$$R = \frac{N_{SR}}{N_{CR}} = \begin{cases} \text{SHERPA} : R = 0.4162 \\ \text{POWHEG} : R = 0.4159 \end{cases} \tag{9.18}$$

Here N_{SR} refers to the WZ yield in the Z+jets enriched fake factor region and N_{CR} refers to the WZ yield in th WZ CR. The extrapolation uncertainties from the choice of generators are estimated to be $\Delta R/R = 0.1\%$.

Uncertainties from the renormalization and factorization scales for WZ, ZZ and $Z+\gamma$ are estimated with a 7-point variation with configurations $\{\mu_r, \mu_f\} = \{0.5, 0.5\}$, $\{0.5, 1\}$, $\{1, 0.5\}$, $\{1, 1\}$, $\{1, 2\}$, $\{2, 1\}$, $\{2, 2\}$ and the envelope of all seven variations is quoted as the uncertainty. Uncertainties due to parton distribution function (PDF) for these processes are evaluated by using NNPDF PDFset and taking the standard deviation of the variation in the yields of the processes from the 100 PDFs in the PDFset. The choice of the QCD scale α_s when evaluating the PDFs is also considered as a separate uncertainty. The up and down variations are symmetrized for convenience when they are of similar size. The ZZ theory uncertainties should be largely independent of fake type and are therefore averaged over the fake electron and fake muon cases, while this is not the case for $Z+\gamma$, which behaves differently for both fake flavors. Thus, those uncertainties are handled separately. The uncertainties due to resummation scale (QSF) and matching scale (CKKW) for WZ are estimated using generator-level samples produced with variations in these scales and are compared to the nominal samples. The variation in the extrapolation factor defined as $\Delta\alpha = \frac{N_{\mathrm{SR}}^{\mathrm{var}}/N_{\mathrm{CR}}^{\mathrm{var}}}{N_{\mathrm{SR}}^{\mathrm{nom}}/N_{\mathrm{CR}}^{\mathrm{nom}}}$ is quoted as the resummation and matching scale uncertainties for the WZ events. For ZZ and Z+γ, a conservative estimate for the resummation and matching scale uncertainties are set to be 4% based on the yield difference due to resummation and matching scale variations on the yields of WZ as shown in Table 9.12. This value is added in quadrature to the total uncertainties.

Table 9.12: Theory uncertainties of WZ derived from variation in the yields. These uncertainties are not used for the WZ process. But the uncertainties in the truth analysis give an idea about the impact of QSF and CKKW uncertainties in diboson samples.

WZ Uncertainty	Impact high / low (+/- stat) [%] CutFakeElecID	 CutFakeMuonID
scale (reco)	10.4 / -7.6	11.3 / -9.0
PDF (reco)	1.5	1.4
α_s (reco)	1.6	1.5
CKKW (truth)	0.1 ± 0.7 / -2.9 ± 0.6	-0.8 ± 0.7 / -2.3 ± 0.5
QSF (truth)	2.7 ± 0.7	3.3 ± 0.6
CSSKIN (truth)	0.1 ± 0.7	0.0 ± 0.6

The systematic uncertainties for WZ process is summarized in Table 9.15, and for ZZ and Z+γ in Tables 9.13 and 9.14 respectively.

Table 9.13: Theory uncertainties of ZZ. The columns correspond to different cut stages. Only the ID columns are considered for the systematic uncertainty and the Anti-ID columns (two rightmost columns) are only quoted for completeness. The last row takes the quadrature sum of the rows above and symmetrized uncertainty. For ZZ, fake electron and fake muon contributions are averaged as it does not depend on the flavor of the fakes.

ZZ Uncertainty	Impact high / low (+/- stat) [%]			
	CutFakeElecID	CutFakeMuonID	CutFakeElecAntiID	CutFakeMuonAntiID
scale	10.3 / -7.8	8.3 / -6.8	10.4 / -7.9	11.2 / -8.2
PDF	1.6	1.8	1.6	1.7
α_s	1.3	1.3	1.3	1.3
symmetrized total	8.6		–	–

Table 9.14: Theory uncertainties of Z+γ. The columns correspond to different cut stages. Only the ID columns are considered for the systematic uncertainty. The last row takes the quadrature sum of the rows above and symmetrized uncertainty. Note that the region with fake muons has no practical relevance due to the small yields.

Z+γ Uncertainty	Impact high / low (+/- stat) [%]			
	CutFakeElecID	CutFakeMuonID	CutFakeElecAntiID	CutFakeMuonAntiID
scale	6.7 / -7.8	13.3 / -24.8	13.3 / -9.2	40.7 / -7.3
PDF	3.7	10.0	2.5	2.9
α_s	1.1	0.7	1.2	-0.0
symmetrized total	8.2	22	–	–

Table 9.15: Theory uncertainties on the normalization factor of WZ calculated with $\Delta\alpha = \frac{N_{\mathrm{SR}}^{\mathrm{var}}/N_{\mathrm{CR}}^{\mathrm{var}}}{N_{\mathrm{SR}}^{\mathrm{nom}}/N_{\mathrm{CR}}^{\mathrm{nom}}}$. The columns correspond to different cut stages. Only the ID columns are considered for the systematic uncertainty. The rows show different sources of normalization uncertainty, some of them derived from reconstructed samples (reco) and some derived in a generator level selection (truth). The last row takes the quadrature sum of the rows above and symmetrized uncertainty.

WZ Uncertainty	$\Delta\alpha$ [%] CutFakeElecID	 CutFakeMuonID
scale (reco)	1.3 / -1.3	1.2 / -1.2
PDF (reco)	0.6	0.1
α_s (reco)	0.1	0.0
CKKW (truth)	-0.9 ± 0.7 / 0.0 ± 0.6	-1.8 ± 0.7 / 0.6 ± 0.5
QSF (truth)	-0.2 ± 0.7	0.3 ± 0.6
CSSKIN (truth)	0.2 ± 0.7	0.1 ± 0.6
Fake contamination	1	1
Statistical + p_{T} mismodelling	2	2
symmetrized total	3	3.5

Then the final normalization uncertainties are summarized in Table 9.16. For all the other processes in the electroweak background with negligible contributions, an uncertainties of 10% is assumed conservatively. These uncertainties are propagated to the Z+jets fake factor by varying up and down simultaneously the normalization of all processes.

Table 9.16: Summary of EW subtraction variations. The values for ZZ and $Z+\gamma$ are taken from Tables 9.13 and 9.14 after adding a 4% uncertainty for QSF and CKKW variations. The WZ uncertainties are taken directly from Table 9.15.

Process	Normalization uncertainty [%] CutFakeElecID	 CutFakeMuonID
ZZ	9.5	
$Z+\gamma$	9.1	22
WZ	3	3.5
other	10	

9.6.4 Dijet Fake Factor

Fake factors derived in the dijets-enriched region provides a cross-check to the nominal fake factors derived from Z+jets-enriched region, from a distinctly different physics process of origin. The dijets-enriched samples are also required to estimate the trigger bias in the fake factors (discussed in details in Section 9.6.5.4). The nominal (unbiased, non-triggered) fake factors are derived in a dijet-like, lepton-plus-jet, fake-enriched sample collected with prescaled single-lepton triggers for muons. These prescaled triggers are specifically chosen so as to not introduce trigger bias since they are looser than the anti-id lepton selections. To suppress the contribution of electroweak background and enhance the contribution of dijet events, the following selections are applied before splitting the leptons into id/anti-id category:

- Exactly one fake lepton candidate (electron or muon with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$),
- At least one jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$,
- Angular separation of $\Delta\phi^{\mathrm{lep,jet}} > 2.5$ between the leading p_{T} jet and the fake candidate to enforce lepton+jet topology,
- $E_{\mathrm{T}}^{\mathrm{miss}} < 30\,\mathrm{GeV}$ and the lepton+MET system required to have transverse mass $m_{\mathrm{T}} < 60\,\mathrm{GeV}$, to further suppress W+jets processes

The nominal dijet fake-factor is then derived by first subtracting the prompt lepton contributions from id/anti-id category using MC, and taking the ratio as in Equation 9.13 Figure 9.28 shows the (p_{T}, η) spectra of the electron and muon fake candidate in the id and anti-id regions for the events collected using the prescaled triggers after unprescaling.

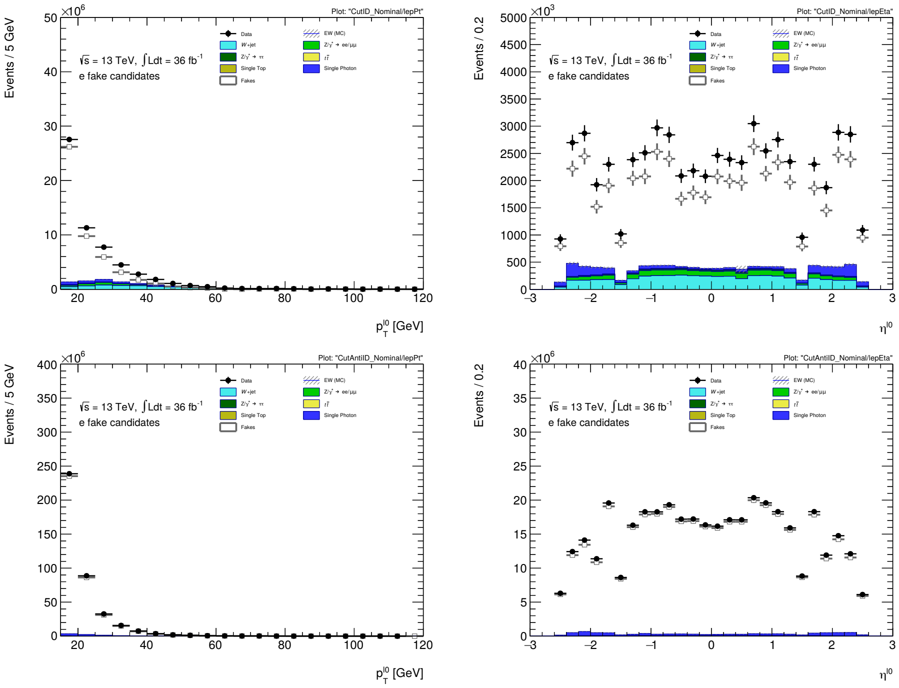


Figure 9.27: p_{T} (left) and η (right) distributions of the fake electron candidate in id (top) and anti-id (bottom) regions observed in 2015+2016 data using unbiased single-lepton triggers.

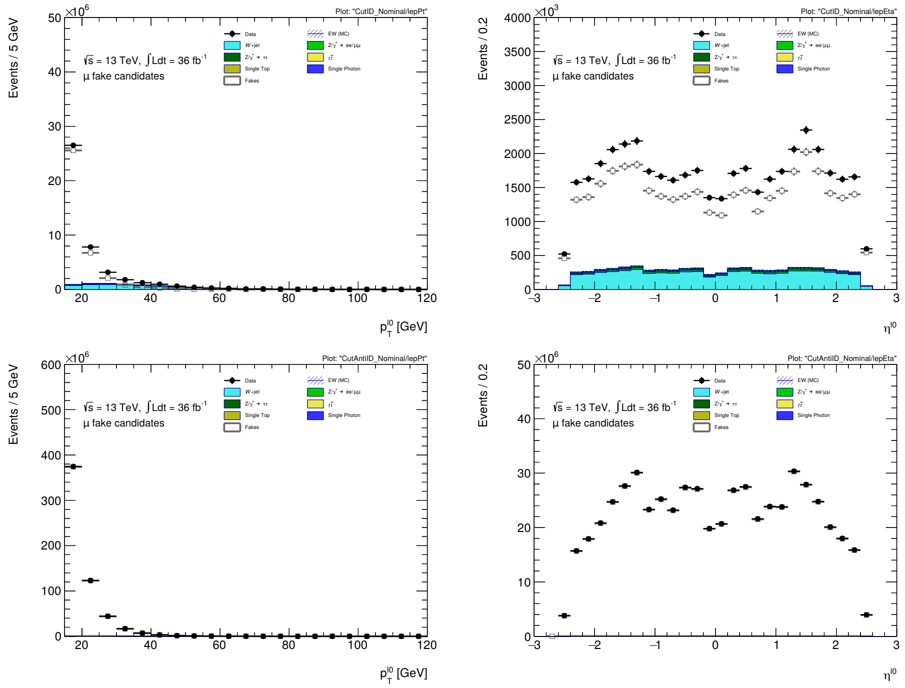


Figure 9.28: p_{T} (left) and η (right) distributions of the fake muon candidate in id (top) and anti-id (bottom) regions observed in 2015+2016 data using unbiased single-lepton triggers.

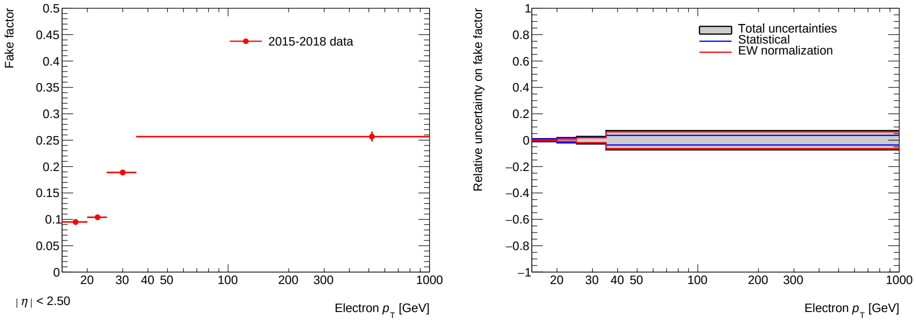


Figure 9.29: Electron fake factors (left) and their relative uncertainties (right).

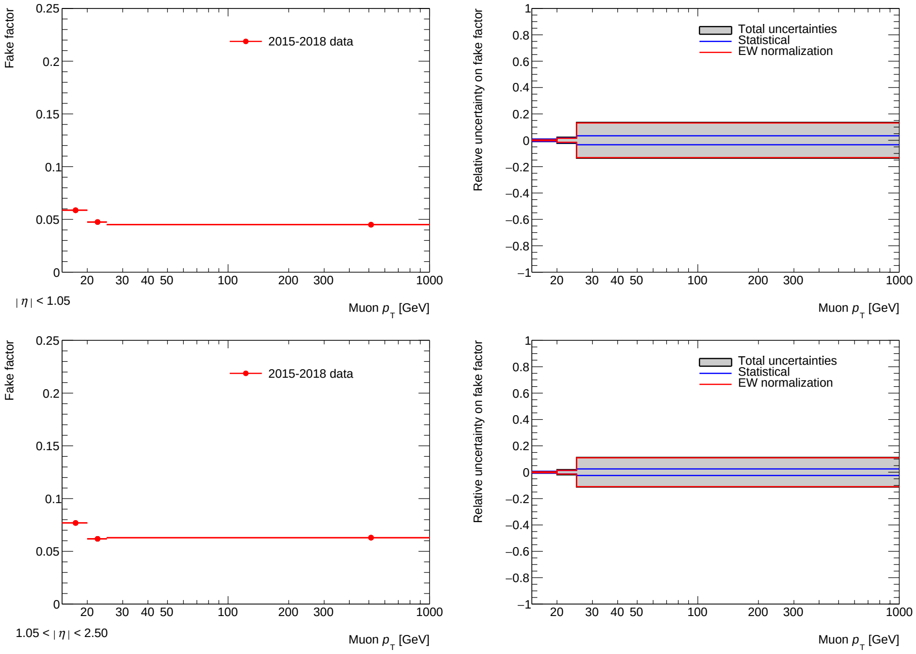


Figure 9.30: Muon fake factors (left) and their relative uncertainties (right) for $|\eta| < 1.05$ (top) and $1.05 \leqslant |\eta| < 2.5$ (bottom).

9.6.5 Fake Factor Corrections

9.6.5.1 Double Fakes Correction

For estimating the W+jets background in the signal region, a single lepton misidentification is generally considered in the fake-factor method. However, it is possible for QCD multijet processes to be selected in the fake yield in the W+jets control sample when both the leptons are misidentified. The is contribution is already included in the fake factor method by design, although it is overestimated, as shown:

$$\begin{aligned}
N^{\text{FF estimate}}_{id+id} &= f^Z_e N_{\mu,\not{e}} + f^Z_\mu N_{e,\not{\mu}} \\
&= f^Z_e \times (N^{\text{data}}_{\mu,\not{e}} - N^{\text{EW MC}}_{\mu,\not{e}}) + f^Z_\mu \times (N^{\text{data}}_{e,\not{\mu}} - N^{\text{EW MC}}_{e,\not{\mu}}) \\
&= f^Z_e \times (N^{W+jets}_{\mu,\not{e}} + N^{\text{QCD}}_{\mu,\not{e}}) + f^Z_\mu \times (N^{W+jets}_{e,\not{\mu}} + N^{\text{QCD}}_{e,\not{\mu}}) \\
&= f^Z_e \times N^{W+jets}_{\mu,\not{e}} + f^Z_e \cdot f^D_\mu N^{\text{QCD}}_{\not{\mu},\not{e}} + f^Z_\mu \times N^{W+jets}_{e,\not{\mu}} + f^Z_\mu \cdot f^D_e N^{\text{QCD}}_{\not{e},\not{\mu}}
\end{aligned} \tag{9.19}$$

where, f_e^Z and f_μ^Z denote the Z+jets fake factors for electrons and muons respectively, f_e^D and f_μ^D denote the dijet fake factors. A strike-through represents an anti-id lepton and the leptons are not p_{T} ordered. In the last line the QCD terms have been expanded using the fake factor method, where if the Z+jets and dijet fake factors are the same, the QCD contribution would be exactly double counted. A correction term $N_{id+id}^{\mathrm{QCDcorr}}$ is therefore added to $N_{id+id}^{\mathrm{FF\ estimate}}$ to account for this overestimation:

$$N_{id+id}^{\mathrm{QCDcorr}} = N_{\not\mu,\not e}^{\mathrm{QCD}} \cdot FF^{QCD} = (N_{\not\mu,\not e}^{\mathrm{data}} - N_{\not\mu,\not e}^{\mathrm{EW\ MC}}) \cdot (f_e^D f_\mu^D - f_e^Z f_\mu^D - f_e^D f_\mu^Z) \tag{9.20}$$

which is derived from a double-fakes control region. In this control region, both leptons in the event are required to pass only anti-ID criteria, and MC processes with two real leptons ($N_{\not\mu,\not e}^{\mathrm{EW\ MC}}$) are subtracted. Fake factors corresponding to both anti-ID leptons (FF^{QCD}) in the selection are then applied.

The impact of this correction on the fakes yields varies is about 25% in the VBF selection. The size of the correction, and the fake yields in the SR before and after the correction, are shown in Table 9.17.

Table 9.17: Corrected yields of fake leptons in the VBF signal regions, after fake factors have been applied. Cutflows are produced at the last cut before the DNN for the VBF signal regions.

$\sqrt{s} = 13TeV$, $\mathcal{L} = 139fb^{-1}$	e Wjets	μ Wjets	e + μ Wjets	Double Fakes Corr.	Double Fakes Relative Contribution (%)	MisID (corrected)
VBF SR (eμ)	81.26 ± 12.54	100.33 ± 3.05	181.59 ± 12.91	-34.25 ± 0.94	23.25 ± 1.75	147.34 ± 12.94
VBF SR (μe)	85.49 ± 11.09	52.45 ± 4.06	137.94 ± 11.81	-30.56 ± 0.94	28.46 ± 2.59	107.38 ± 11.85

9.6.5.2 Correction Factor

The Z+jets fake factor method (described in Section 9.6.2) assumes that the fake factor FF derived in a three-lepton Z+fake selection is applicable to a two-lepton W+fake selection.

However, depending on which process(es) contributes to the topology considered, the origins of the misidentified leptons can have different relative abundance. Furthermore, the charge and flavor configurations of leptons will influence the composition of different origins.

To derive a correction factor, fake factors measured in Z+jets and W+jets MC samples is compared. These fake factors are derived on MC samples with the same selections as applied in data. Furthermore, one (two) leptons in the event are truth-matched to the $W(Z)$ boson, and while the jet is identified as the fake lepton candidate. The ratio of these fake factors give the correction factor (CF),

$$CF = \frac{FF^{W}_{\mathrm{MC}}}{FF^{Z}_{\mathrm{MC}}} \tag{9.21}$$

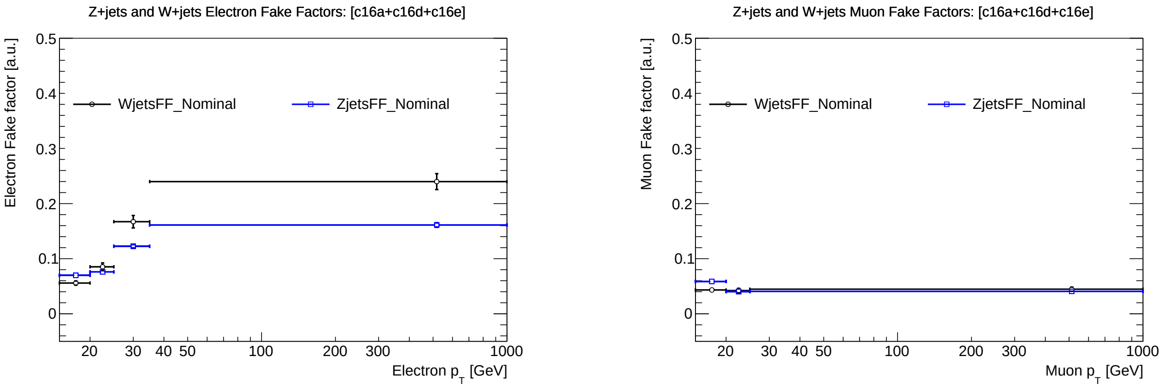
The correction factor derived above is then applied on the three-lepton selection Z+jets fake factor (derived from data) to obtain the fake factor to be applied to the two-lepton selection in W+jets data,

$$FF^{W}_{\mathrm{data}} = CF \times FF^{Z}_{\mathrm{data}} \tag{9.22}$$

Since, correction factor uses MC simulations, the fake factors for W+jets and Z+jets would not be accurate, as simulations do not model fake lepton processes well. However, the requirements on the MC simulations for producing a valid correction factor are quite looser where the relationship between the yields that ID and Anti-ID criteria select be consistent across W+jets and Z+jets samples, which can be understood by rearranging the terms in CF as below,

$$\frac{FF^{W}_{\mathrm{MC}}}{FF^{W}_{\mathrm{data}}} = \frac{N^{W}_{id,MC}}{N^{W}_{id,data}} \bigg/ \frac{N^{W}_{i\not{d},MC}}{N^{W}_{i\not{d},data}} = \frac{N^{Z}_{id,MC}}{N^{Z}_{id,data}} \bigg/ \frac{N^{Z}_{i\not{d},MC}}{N^{Z}_{i\not{d},data}} = \frac{FF^{Z}_{\mathrm{MC}}}{FF^{Z}_{\mathrm{data}}} = const. \tag{9.23}$$

Since, the above requirement cannot be guaranteed for an MC sample, the choice



(a) Electron FF

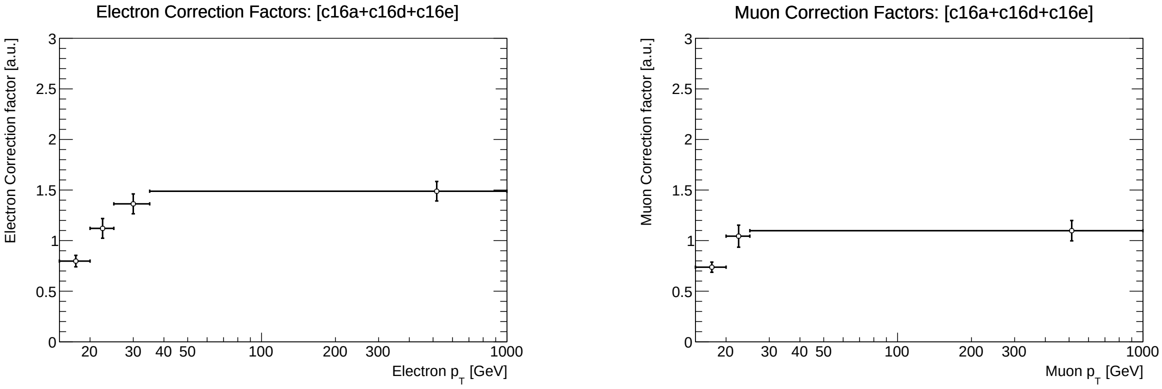
(b) Muon FF

Figure 9.31: Nominal FFs from MC samples that use the POWHEG generator for electrons and muons. Uncertainties are statistical only.

of the generator can introduce a systematic uncertainty in the correction factor. Nominal correction factor was obtained by samples generated by POWHEG due to higher available statistics (see the nominal fake factors in Figure 9.31, and the correction factors produced by taking their ratio in Figure 9.32). To evaluate this systematic uncertainty due to generator, central value of the correction factor obtained using alternate generator MADGRAPH5_aMC@NLO is compared to the central value of the nominal correction factor.

The correction factors are calculated separately for electrons and muons. For the electron correction factors, the same p_T binning are used as the electron fake factors. For the muon correction factors, the two highest p_T bins are combined into one bin above 25 GeV for better statistics and as the correction factors in those two bins were found to be consistent. Due to negligible dependence of the correction factors on η, the correction factors are not binned in η.

The correction factors found with the MADGRAPH5_aMC@NLO and POWHEG samples are given in Tables 9.18 and 9.19, and the nominal correction factors with the statistical and systematic uncertainties are listed in Tables 9.20 and 9.21.



(a) Electron CF

(b) Muon CF

Figure 9.32: Nominal CFs from MC samples that use the Powheg generator for electrons and muons. Uncertainties are statistical only.

Table 9.18: Electron CF in four bins of p_{T} for the two MC generators. Only statistical error is shown.

Generator	Electron CF			
	$15 - 20\,\mathrm{GeV}$	$20 - 25\,\mathrm{GeV}$	$25 - 35\,\mathrm{GeV}$	$35 - 1000\,\mathrm{GeV}$
Powheg	0.80 ± 0.06	1.12 ± 0.10	1.36 ± 0.10	1.49 ± 0.10
MadGraph5_aMC@NLO	0.77 ± 0.12	1.45 ± 0.31	1.35 ± 0.28	1.19 ± 0.16

Table 9.19: Muon CF in three bins of p_{T} for the two MC generators. Only statistical error is shown.

Generator	Muon CF		
	$15 - 20\,\mathrm{GeV}$	$20 - 25\,\mathrm{GeV}$	$25 - 1000\,\mathrm{GeV}$
Powheg	0.74 ± 0.05	1.04 ± 0.11	1.10 ± 0.10
MadGraph5_aMC@NLO	0.76 ± 0.09	1.06 ± 0.22	1.03 ± 0.17

Table 9.20: Electron CF in four bins of p_{T}. The nominal CF is derived using Powheg and the systematic uncertainty is obtained by comparing with MadGraph5_aMC@NLO.

Electron Correction Factor			
$15-20\,\mathrm{GeV}$	$20-25\,\mathrm{GeV}$	$25-35\,\mathrm{GeV}$	$35-1000\,\mathrm{GeV}$
0.80 ± 0.06(stat) $\pm$ 0.03(syst)	1.12 ± 0.10(stat) $\pm$ 0.33(syst)	1.36 ± 0.10(stat) $\pm$ 0.01(syst)	1.49 ± 0.10(stat) $\pm$ 0.30(syst)

Table 9.21: Muon CF in three bins of p_{T}. The nominal CF is derived using Powheg and the systematic uncertainty is obtained by comparing with MadGraph5_aMC@NLO.

Muon Correction Factor		
$15-20\,\mathrm{GeV}$	$20-25\,\mathrm{GeV}$	$25-1000\,\mathrm{GeV}$
0.74 ± 0.05(stat) $\pm$ 0.02(syst)	1.04 ± 0.11(stat) $\pm$ 0.01(syst)	1.10 ± 0.10(stat) $\pm$ 0.12(syst)

9.6.5.3 Flavor Composition Studies

The likelihood of a jet being mis-identified as a lepton also depends on the jet color, which in turn alters the fake factor. The distribution of jet colors is dependent on the sample. Therefore, the difference in fake factor from derived from $W+$ jets and $Z+$ jets events can be attributed to their difference in flavor composition.

To study the difference in fake factor due flavor composition, the MC samples used to find the correction where the reconstructed lepton(s) originating from the leptonically decaying $W(Z)$ is identified by being matched to the truth prompt lepton(s) in W+jets(Z+jets) sample. To identify the flavor of the fake leptons the following scheme is employed. A fake lepton may come from a real lepton produced in decay or in interactions with the detector material. If a real lepton is found within $\Delta R < 0.03$ of the fake lepton, it is classified as "leptonic". Otherwise it is assumed to originate from hadronic activity, denoted "hadronic". To classify whether a fake lepton originates from a quark of a certain flavor, a matching scheme is applied starting from heavier and moving to lighter quarks. A fake lepton is classified as type "bottom" if a bottom-type quark is found within $\Delta R < 0.4$ of the fake lepton. If no bottom-type quark is found, the test is repeated for lighter quarks. If the fake object can not be matched to any of the quark categories, it is classified as "other". For electron fakes, the other-type objects often originate from photons, either from Bremsstrahlung, π_0 decays or final state radiation. The resulting flavor compositions in Powheg samples for ID and Anti-ID electrons and muons are given in Tables 9.22 and 9.23.

Table 9.22: Sample composition of fake electrons in Powheg $Z+$jets and $W+$jets samples.

Samples		Electron Flavor Composition (%)				
		Bottom	Charm	Strange	Light	Other
$W+$jets	ID	1.501 ± 0.379	17.244 ± 1.442	8.162 ± 0.940	66.176 ± 3.338	6.918 ± 0.870
	Anti-ID	1.501 ± 0.132	26.544 ± 0.622	13.358 ± 0.415	57.403 ± 1.010	1.193 ± 0.115
$Z+$jets	ID	21.141 ± 0.693	7.912 ± 0.365	7.483 ± 0.345	54.837 ± 1.139	8.627 ± 0.381
	Anti-ID	13.734 ± 0.165	12.222 ± 0.140	16.013 ± 0.159	55.991 ± 0.346	2.040 ± 0.053

Table 9.23: Sample composition of fake muons in Powheg $Z+$jets and $W+$jets samples.

Samples		Muon Flavor Composition (%)				
		Bottom	Charm	Strange	Light	Other
$W+$jets	ID	7.656 ± 1.243	77.599 ± 5.260	6.400 ± 1.143	5.150 ± 0.966	3.194 ± 0.791
	Anti-ID	4.905 ± 0.212	84.517 ± 1.167	5.066 ± 0.211	4.658 ± 0.207	0.854 ± 0.089
$Z+$jets	ID	58.686 ± 1.717	26.729 ± 0.963	4.329 ± 0.335	4.313 ± 0.333	5.943 ± 0.401
	Anti-ID	58.951 ± 0.389	31.262 ± 0.238	5.447 ± 0.086	3.586 ± 0.068	0.753 ± 0.031

9.6.5.4 Trigger Bias

Due to an inclusive OR-combination of dilepton and single-lepton triggers are used in the analysis (listed in Table 9.1), in the W+jets control region, the trigger decision of each event belongs to one the following three mutually-exclusive categories:

1. Dilepton trigger passed.[4]

---

[4] All events in which both single-lepton triggers fired constitute a strict subset of the first category.

2. Only one single-lepton trigger fired by the ID lepton.
3. Only one single-lepton trigger fired by the Anti-ID lepton.

The dilepton triggers have looser identification and isolation requirements than those of ID and Anti-ID definitions which collect the majority of events in the W+jets control region. A small number of additional events are gained by one of the leptons firing a single-lepton trigger, which has tighter requirements than the Anti-ID definition. If an ID lepton fired the trigger, the requirements being placed at trigger-level are still looser than the ID definition, and so there is no loss in the event yield. If, however, an anti-ID lepton fired the trigger the tighter trigger requirements introduces a bias in the event yield. Table 9.24 shows the relevance of this category. Separate fake factors called the triggered fake factors are derived with the same trigger bias and applied to only the third category listed above. To have the same trigger bias while deriving triggered fake factors, the Anti-ID lepton is required to fire the single-lepton trigger. Since this requirement significantly reduces the statistics in the Z+jets samples, the dijets samples are used to derive the triggered fake factors (FF^{T}):

$$FF^{\mathrm{T}} = \frac{N_{id} - N_{id}^{\text{non-dijet}}}{N_{\not{id}^{T}} - N_{\not{id}^{T}}^{\text{non-dijet}}}, \tag{9.24}$$

The superscript T, on the right-hand side, denotes that the Anti-ID lepton is required to fire the single-lepton trigger.

Table 9.24: Proportion of events in ggF and VBF W+jets control regions for which only the fake lepton candidate fired a single-lepton trigger.

	ggF 0-jet		ggF 1-jet		VBF 2-jet	
	$e\mu$	μe	$e\mu$	μe	$e\mu$	μe
e-fakes	134 / 3321	49 / 10167	80 / 1672	52 / 5729	23 / 578	15 / 1452
μ-fakes	8 / 10426	271 / 2108	10 / 8302	254 / 1954	6 / 3073	131 /925

9.7 Statistical Treatmeant

The aim of the analysis is to measure VBF Higgs boson production cross section times branching ratio for the $H\to WW^*\to \ell\nu\ell\nu$ decay. A likelihood analysis is performed to extract the yield of Higgs boson events observed in data. The likelihood function gives the probability of how likely are a set of model parameters, given an observed dataset. The likelihood function is maximised to obtain the values of the model parameters which best describe the data. The likelihood formalism is introduced in Section 9.7.1 and the corresponding test statistic statistic is defined in Section 9.7.2, followed by a listing of the systematic uncertainties in Section 9.7.3. Finally, The results are presented in Section 9.7.4.

9.7.1 Likelihood Formalism

In a given signal region phase space, if the expected yield of backgrounds is given by B_{SR} and signal is given by S_{SR}, then the PDF of the number of observed events N_{SR} in this region can be given by a Poisson distribution:

$$P(N_{SR}|S_{SR}+B_{SR}) = \frac{(S_{SR}+B_{SR})^{N_{SR}}e^{-(S_{SR}+B_{SR})}}{(N_{SR})!} \tag{9.25}$$

For this analysis, the expected yield of (VBF)$H\to WW^*\to \ell\nu\ell\nu$ signal process is not known a priory. In this case, we introduce a signal strength μ, which scales the predicted yield s_{SR} of the signal process from theory, in this case the SM,

$$S_{SR} = \mu \times s_{SR} \tag{9.26}$$

If $\mu = 0$, it corresponds to no observed signal event, while $\mu = 1$ corresponds to the observed rate of signal process being the same as the expected SM rate. Therefore, in the situation where N_{SR} is measured, the information about μ (which is the *parameter of interest (POI)*) can be gained by recasting the PDF as a likelihood function:

$$\mathcal{L}(\mu) = P(N_{SR}|\mu s_{SR} + B_{SR}) \tag{9.27}$$

By maximizing with respect to μ the value best describing the data is obtained.

Here, it is assumed that the prediction of the background process is perfectly known, which is rarely the case. For estimating the expected rate of the background process(es), sideband (control) region(s) is(are) generally defined such that it is enriched in the background process(es). The expected yield of a background process can be expressed in terms of its background normalization factor β ("background strength") as

$$b = \beta \times B \tag{9.28}$$

with b as the theoretically predicted yield. The likelihood is modified accordingly by multiplying with another Poisson term,

$$\mathcal{L}(\mu) = P(N_{SR}|\mu s_{SR} + \beta b_{SR}) \cdot P(N_{CR}|\mu s_{CR} + \beta b_{CR}) \tag{9.29}$$

Like μ, β is floating in the fit, and can be estimated by maximizing the likelihood function with respect to both μ and β. Here, the background strength β is an example of a *nuisance parameter (NP)*, as the measurement is the experiment is sensitive to but does not it but is not the target for measurement. Nuisance parameters are also used to model the impact of associated systematic uncertainties in the analysis, denoted by θ. The signal and background predictions s and b are expressed as a function of NP θ in the likelihood. A constraint term $C(\theta)$ is included in the likelihood, which serves as the prior distribution (in Bayesian interpretation) or the sampling distribution (Frequentist interpretation), is used to incorporate the associated uncertainty with θ. The inclusion of NPs (θ) broadens the likelihood as a function of μ (compared to the case of fixed NPs), which represents the loss of information on μ due to systematic uncertainties.

The full likelihood function for VBF analysis including all signal and control regions and constraint terms for NPs can be written as:

$$\mathcal{L}(\mu_{VBF}, \vec{\theta}) = \prod_{i}^{SRs\ \&\ CRs} P\left(N_i | \mu_{VBF} \cdot s_i^{VBF}(\vec{\theta}) + \sum_{k}^{backgrounds} \beta_k \cdot b_{i,k}(\vec{\theta})\right) \prod_{\theta \in \vec{\theta}} C(\theta) \tag{9.30}$$

The likelihood function is defined over the regions defined over DNN output (discussed in Section 9.4.3), with i denoting the 7 DNN output bins for SR and combined single bin DNN output for top and $Z \rightarrow \tau\tau$ control regions. The β_k terms denote the normalization for the background processes. The collection of NPs are denoted by $\vec{\theta}$ and are constrained by their respective $C(\theta)$ term. These constraint terms are generally Gaussian, but Poisson terms are used for statistical uncertainties on background predictions. Here, the parameter of interest (POI) is μ_{VBF}.

The likelihood function is maximized with respect to all the parameters to obtain the best-fit value for the signal strength μ_{VBF}, which also yields the best-fit values for the NPs ($\vec{\theta}$). In the fit, a NP may be pulled, meaning that its best-fit value deviates relative to its nominal (pre-fit) value. A NP is over-constrained if the associated uncertainty from the fit is constrained to a smaller value relative to the pre-fit one.

9.7.2 Test Statistic and Interpretation

The search for VBF$H \rightarrow WW^{*} \rightarrow \ell\nu\ell\nu$ process can be framed in terms of hypothesis testing, where the null hypothesis (H_0) is that VBF signal is not present ($\mu_{VBF} = 0$) and any deviation in the observed data is through random chance; while the alternate (H_1) is that VBF signal is present in the observed data. To quantify the agreement between the observed data and either of the hypotheses a test statistic is used. We start with constructing the profile likelihood ratio shown in Equation 9.31. The denominator is the global maximum of the likelihood with respect to all parameters (POI and NPs) simultaneously. The numerator is a profile likelihood with the NPs profiled at some given values of the POI, i.e., the likelihood function is maximized with respect to all the NPs for a fixed value of μ_{VBF} (POI). Thus, the profile likelihood ratio is effectively only parametrized by the POI.

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\vec{\theta}}}(\mu)\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\vec{\theta}}\right)} \tag{9.31}$$

The test statistic is defined using the profile likelihood ratio as,

$$t_{\mu} = -2ln\lambda(\mu) \tag{9.32}$$

A p-value, which is the probability that the background fluctuated to the observed (or a more extreme) value, is defined to evaluate the compatibility between the data and the H_0 hypothesis using the test statistic t_{μ},

$$p_0 = \int_{t_{0,obs}}^{\infty} f(t_0|\hat{\hat{\vec{\theta}}}(\mu = 0))dt_0 \tag{9.33}$$

where, f denotes pdf of t_0, which is the test statistic for $\mu = 0$. In large sample limit, an asymptotic form [162] of the pdf f can be given by a $\chi^2_{dof=1}$ distribution. The statistical significance Z which is the number of standard deviations the measurement is away from $\mu = 0$ can be calculated from the corresponding test statistic t_0 as $Z = \sqrt{t_{0,obs}}$.

The two-sided confidence interval of μ is obtained by scanning the test statistic of μ, where for each point a fit is made with μ fixed. The points where the test statistic is one (two) unit(s) above minimum corresponds to a 68% (95%) confidence level interval.

9.7.3 Systematic Uncertainty

9.7.3.1 Experimental uncertainty

Experimental uncertainties are the systematic uncertainties related to detector and reconstruction effects, and can be classified into two distinct types:

- **four-vector (P4) systematics** - applied as $\pm 1\sigma$ variations to the four-momentum of an object,
- **scale factor (SF) systematics** - applied as $\pm 1\sigma$ variations to the weight of the event or an individual particle.

The complete set of experimental uncertainties that are considered in the analysis is listed in Table 9.25, along with their types. The electron-related systematics include uncertainties on their reconstruction, identification, isolation and trigger efficiencies, as well as on their associated four-momentum scale and resolution. The muon-related systematics are similar, except with an additional uncertainty on the efficiency of track-to-vertex association (TTVA) impact parameter cuts. The lepton uncertainties are derived from studying $J/\psi \rightarrow l^+l^-$, $W^\pm \rightarrow l^\pm\nu$ and $Z \rightarrow l^+l$ decays [163, 164, 165, 147].

The uncertainties on the jet energy scale are derived as a function of p_{T} and η, containing terms that account also for pileup conditions as well as flavor composition of the jet. The uncertainty on the jet energy resolution on the other hand, is modeled with a single parameter. Both are derived based on in-situ studies of dijet, Z+jet, and γ+jet events [166, 167]. A scale factor uncertainty is added to account for the JVT efficiency. Uncertainties related to jet flavor tagging are modeled by parameters that are the result of eigen-vector decomposition [168, 169] as well as a couple of additional parameters for charm quark.

The dominant experimental uncertainties are due to the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement [67]. Uncertainties on the TST missing transverse energy E_T^{miss} are measured in $Z \to \mu\mu$ events using the variable $p_{\mathrm{T}}^{\vec{hard}}$ defined as the p_{T} sum of the hard terms to discriminate between soft term scale and resolution effects, taking into account detector material uncertainties [70].

All components of experimental uncertainties related to shape differences in the DNN output are considered, except for those from $V\gamma$ and Z+jets processes due to large statistical fluctuations as well as for those that are consistent with no shape being observed ($p_0 < 0.05$). Three sources of uncertainty related to the extrapolation factor used in the data-driven fake lepton background estimate are considered: the statistical uncertainty on the extrapolation factor itself, an uncertainty related to the subtraction of processes with two prompt leptons from the Z+jets-enriched sample used to derive the extrapolation factor, and an uncertainty in the sample composition correction factor. While the relative size of these three uncertainties depends on the specific bin of the extrapolation factor, they collectively have only a minimal impact on the overall signal strength.

Other experimental systematics include trigger uncertainties [170, 171], modeling of pile-up, and luminosity measurement [172]. The uncertainty on the integrated luminosity is ±1.7% for the full Run 2 dataset. The luminosity uncertainty is only applied to the Higgs boson signal and to background processes that are normalized to theoretical predictions.

9.7.3.2 Theoretical uncertainty

The theoretical uncertainties concern the normalisation and shape of the MC predictions of signal and background processes. In general, four sources of uncertainty are considered: QCD scale uncertainty, i.e. uncertainty on the factorisation and renormalisation scales, uncertainty in the PDFs, underlying event and parton shower uncertainties (UEPS), uncertainties on the generation of the hard interaction. In some cases, additional process specific theory uncertainties are also included. In other cases, theory uncertainties are excluded if the systematic variation is smaller than the statistical uncertainty of the Monte Carlo sample. An overview of the theory uncertainties included in the analysis is provided in Table 9.26.

Table 9.25: Summary of the experimental sytematic uncertainties considered.

Systematic uncertainty	Short description
	Event
Luminosity	uncertainty on total integrated luminosity
Pileup Reweighting	uncertainty on pileup reweighting
	Electrons
EL_EFF_Trigger_Total_1NPCOR_PLUS_UNCOR	trigger efficiency uncertainty
EL_EFF_Reco_Total_1NPCOR_PLUS_UNCOR	reconstruction efficiency uncertainty
EL_EFF_ID_TotalCorrUncertainty	ID efficiency uncertainty
EL_EFF_ID_CorrUncertaintyNP (0 to 15)	ID efficiency uncertainty splits in 16 components
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP (0 to 17)	ID efficiency uncertainty splits in 18 components
EL_EFF_Iso_Total_1NPCOR_PLUS_UNCOR	isolation efficiency uncertainty
EG_SCALE_ALL	energy scale uncertainty
EG_SCALE_AF2	
EG_RESOLUTION_ALL	energy resolution uncertainty
	Muons
MUON_EFF_TrigStatUncertainty	trigger efficiency uncertainty
MUON_EFF_TrigSystUncertainty	
MUON_EFF_RECO_STAT	reconstruction and ID efficiency uncertainty for muons with p_{T} ¿ 15 GeV
MUON_EFF_RECO_SYS	
MUON_ISO_STAT	isolation efficiency uncertainty
MUON_ISO_SYS	
MUON_TTVA_STAT	track-to-vertex association efficiency uncertainty
MUON_TTVA_SYS	
MUON_ID	momentum resolution uncertainty from inner detector
MUON_MS	momentum resolution uncertainty from muon system
MUON_SCALE	momentum scale uncertainty
MUON_SAGITTA_RHO	charge dependent momentum scale uncertainty
MUON_SAGITTA_RESBIAS	
	Jets
JET_EffectiveNP_Detector (1 to 2)	energy scale uncertainty from the in situ analyses splits into 8 components
JET_EffectiveNP_Mixed (1 to 3)	
JET_EffectiveNP_Mmodeling (1 to 4)	
JET_EffectiveNP_Statistical (1 to 6)	
JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibration (modeling)
JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations (statistics/method)
JET_EtaIntercalibration_NonClosure_highE	energy scale uncertainty on eta-intercalibrations (non-closure)
JET_EtaIntercalibration_NonClosure_negEta	
JET_EtaIntercalibration_NonClosure_posEta	
JET_Pileup_OffsetMu	energy scale uncertainty on pile-up (mu dependent)
JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up (NPV dependent)
JET_Pileup_PtTerm	energy scale uncertainty on pile-up (pt term)
JET_Pileup_RhoTopology	energy scale uncertainty on pile-up (density ρ)
JET_Flavor_Composition	energy scale uncertainty on flavour composition
JET_Flavor_Response	energy scale uncertainty on samples' flavour response
JET_BJES_Response	energy scale uncertainty on b-jets
JET_PunchThrough_MC16	energy scale uncertainty for punch-through jets
JET_SingleParticle_HighPt	energy scale uncertainty from the behaviour of high-p_{T} jets
JET_JER_DataVsMC_MC16	energy resolution uncertainty, each for both MC and pseudo-data
JET_JER_EffectiveNP (1 to 11)	
JET_JER_EffectiveNP_12restTerm	
JET_JvtEfficiency	JVT efficiency uncertainty
FT_EFF_Eigen_B	b-tagging efficiency uncertainties ("BTAG_MEDIUM"): 3 components for b jets, 3 for c jets and 4 for light jets
FT_EFF_Eigen_C	
FT_EFF_Eigen_L	
FT_EFF_Eigen_extrapolation	b-tagging efficiency uncertainty on the extrapolation to high-p_{T} jets
FT_EFF_Eigen_extrapolation_from_charm	b-tagging efficiency uncertainty on tau jets
	MET
MET_SoftTrk_ResoPara	track-based soft term related longitudinal resolution uncertainty
MET_SoftTrk_ResoPerp	track-based soft term related transverse resolution uncertainty
MET_SoftTrk_Scale	track-based soft term related longitudinal scale uncertainty
MET_JetTrk_Scale	track MET scale uncertainty due to tracks in jets

Table 9.26: Overview of the theory uncertainties included in the VBF analyses.

process	uncertainty	Included in fit for VBF
ggF	QCD scale (ST method)	Yes
	matching	Yes
	PS/UE	Yes
	PDF	Yes
VBF	QCD scale (7-point)	Yes
	matching	Yes
	PS/UE	Yes
	PDF	Yes
WW	merging (ckkw)	Yes
	PS (Py8 vs. Hw++/SHERPA vs.MGpy8)	Yes
	PS (QSF, CSSKIN)(*2)	Yes
	PDF	Yes
	QCD scale (7-point)	Yes
	$gg \to WW$ uncertainty	Yes
	EW correction	No
top	matching (Powheg vs MCatNLO)	Yes
	PS/UE (Py8 vs H7)	Yes
	QCD scale (7-point)	Yes
	radiation (ISR/FSR)	Yes
	PDF	Yes
	Top/Wt interference	Yes
Z+jets	alternative generator	Yes
	QCD scale	Yes
	PDF	Yes
	CKKW and QSF	No
$WZ/W\gamma^*$	QCD and merging scales	No
$W\gamma$	QCD scale	No
$W\gamma$	generator/matching/pdf	No

The QCD scale uncertainties are evaluated by independently varying the the renormalization and factorization scales by 2.0 and 0.5 relative to the nominal value, while PDF uncertainties are evaluated using envelope of the 68% confidence level PDF4LHC Hessian PDF eigenvectors added in quadrature. An uncertainty is assigned on the generator and matching scheme by comparing the NLO matching of

POWHEG+HERWIG7 with the NLO matching of the aMC@NLO+HERWIG7. Uncertainties on the modeling of the parton shower are also evaluated by comparing the nominal POWHEG+PYTHIA 8 with POWHEG+HERWIG7. The ggF background includes an uncertainty on the acceptance of the central-jet veto, and is evaluated using the Stewart-Tackmann method [173] with MadGraph5_aMC@NLOat truth level.

For the top background, theory uncertainties are applied to the extrapolation factor between control and signal regions. The QCD scale and PDF uncertainties are evaluated with the same strategy used for the signal processes. For PS/UE uncertainties, POWHEG+PYTHIA 8 is compared with POWHEG+HERWIG7 using fast simulation. For a generator matching uncertainty, POWHEG+PYTHIA 8 is compared with aMC@NLO+PYTHIA 8. An uncertainty on the treatment of the interference between $t\bar{t}$ and Wt is derived by comparing samples with different overlap subtraction schemes [174].

For the WW background, theory uncertainties are applied directly to the nominal SHERPAprediction for normalization and DNN shape in the VBF signal region. The WW theory uncertainties are derived at truth level due to limited Monte Carlo statistics, and the different flavor channels $e\mu$ and μe are combined. The QCD scale and PDF uncertainties are evaluated similarly to the strategy used for the signal processes. For PS/UE uncertainties, a comparison is made between POWHEG+PYTHIA 8 and POWHEG+HERWIG++. Uncertainties on additional electroweak corrections and on the normalization of the $ggWW$ process are applied.

For the $Z \rightarrow \tau\tau$ background, an uncertainty on the generator modeling is evaluated on the extrapolation factor between control and signal regions by comparing the nominal SHERPAprediction to the alternative aMC@NLO+PYTHIA 8.

9.7.4 Results

The signal strength parameter μ_{VBF} is obtained from a simultaneous profile likelihood fit to data in the signal and control regions defined in the previous sections. It is defined as the ratio of the measured signal yield in these regions to that predicted by the Standard Model. The CRs are used to determine the normalization of the corresponding backgrounds and enter the fit as single bins. The systematic uncertainties enter the fit as nuisance parameters in the likelihood function.

Table 9.29 shows the post-fit yields for the VBF SR. Yields are shown separately for the highest DNN output bin, which provides the majority of the signal sensitivity. The number of observed signal events is in agreement with the Standard Model expectation. The observed (expected) VBF signal has a significance of 7.0 (6.2) σ above the background expectation. The signal strength for the VBF production mode in the $H \rightarrow WW^*$ decay channel is measured to be

$$\begin{aligned}\mu_{\text{VBF}} &= 1.04\ ^{+0.24}_{-0.20} \\ &= 1.04\ ^{+0.13}_{-0.12}\ (\text{stat.})\ ^{+0.09}_{-0.08}\ (\text{exp syst.})\ ^{+0.17}_{-0.12}\ (\text{sig. theo.})\ ^{+0.08}_{-0.07}\ (\text{bkg. theo.}).\end{aligned}$$

Furthermore, a consistent value is obtained when allowing the ggF background normalization to float freely in the likelihood fit, with the observed significance reduced slightly to 6.3 σ.

Table 9.27: Summary of uncertainties on the signal strength μ_{VBF} when fitting to data. For each sub-category only the most important sources are listed.

Contribution	$\Delta\mu/\mu$ [%]
Total	21
Total systematics	17.8
Data statistics	12
MC statistics	3.0
Theoretical uncertainties	15.7
VBF	14.2
ggF	5.2
Top	3.3
WW	2.5
Zjets	1.9
Experimental systematic uncertainties	6.2
Missing ET	4.7
Jet energy scale	2.2
Luminosity	1.9
Pileup Reweighting	1.7
Flavor tagging	1.6
Jet energy resolution	1.4
Fake factors	0.9

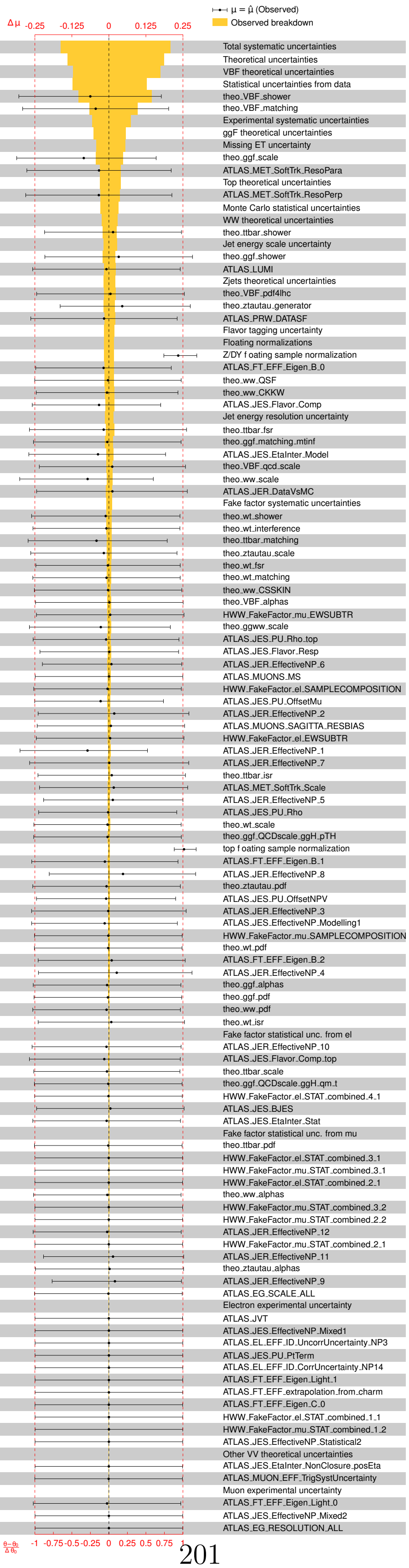


Figure 9.33: Fitted nuisance parameters to data and their impact on the uncertainty on the VBF signal strength, including groups of theory uncertainties.

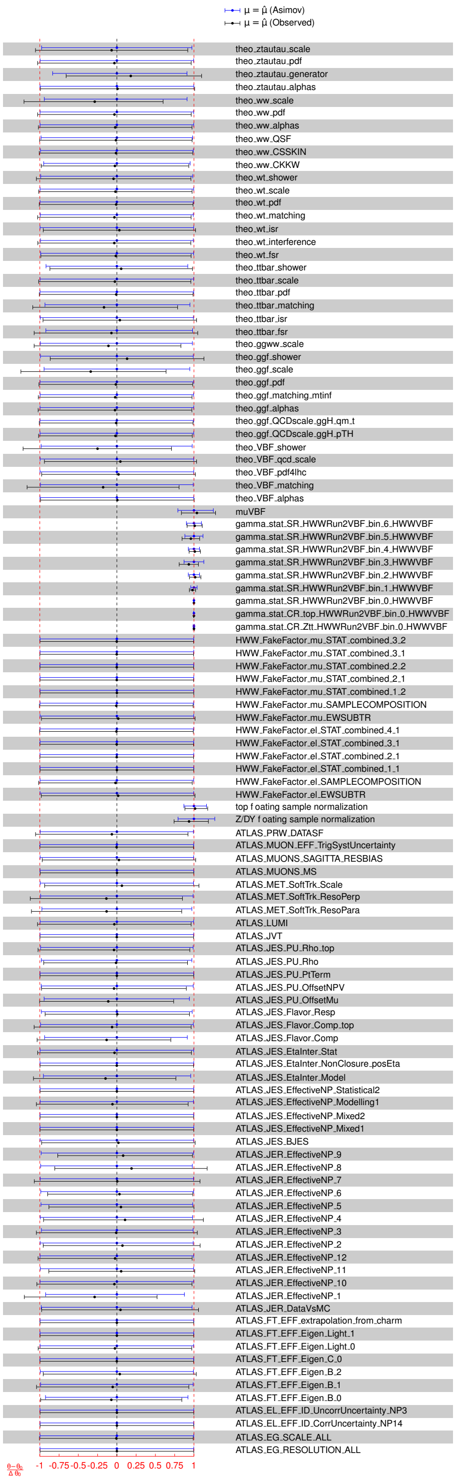


Figure 9.34: Fitted nuisance parameters to the data (black) and to the Asimov data set under the signal plus background hypothesis (blue).

Table 9.28: Post-fit event yields.

	SR	CR $Z \rightarrow \tau\tau$	CR top
VBF	209.15 ± 36.57	17.63 ± 3.6	22.5 ± 4.79
ggF	168.57 ± 61.82	14.31 ± 1.32	27.68 ± 2.61
Other Higgs	28.42 ± 2.22	21.34 ± 1.03	7.78 ± 0.59
$t\bar{t}$	6677.68 ± 516.15	245.16 ± 24.13	36655.73 ± 869.32
Wt	847.28 ± 310.94	29.51 ± 9.65	3412.74 ± 877.01
WW	1996.37 ± 347.35	74.84 ± 13.93	250.41 ± 53.76
VV	392.19 ± 63.76	65.64 ± 21.75	88.15 ± 11.73
Mis-id	415.96 ± 58.46	54.32 ± 7.57	400.35 ± 59.14
Z + jets	1460.15 ± 371.51	1589.5 ± 59.17	243.81 ± 37.71
Predictions	12195.75 ± 121.97	2112.25 ± 47.84	41109.16 ± 226.86
Observed	12189.0 ± 110.4	2114.0 ± 45.98	41112.0 ± 202.76

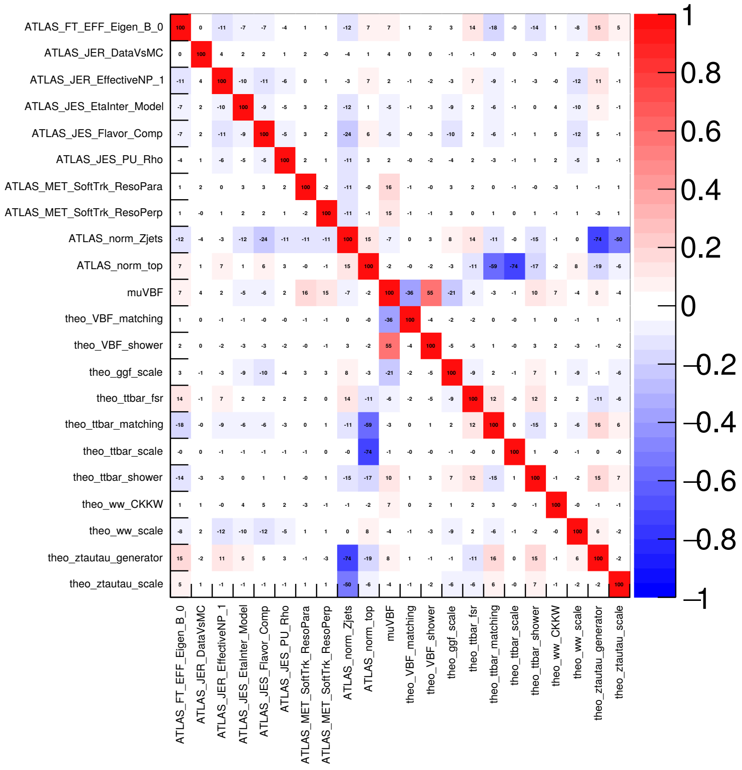


Figure 9.35: Correlation of nuisance parameters under the signal plus background hypothesis in full fit to the Asimov data in the VBF analysis.

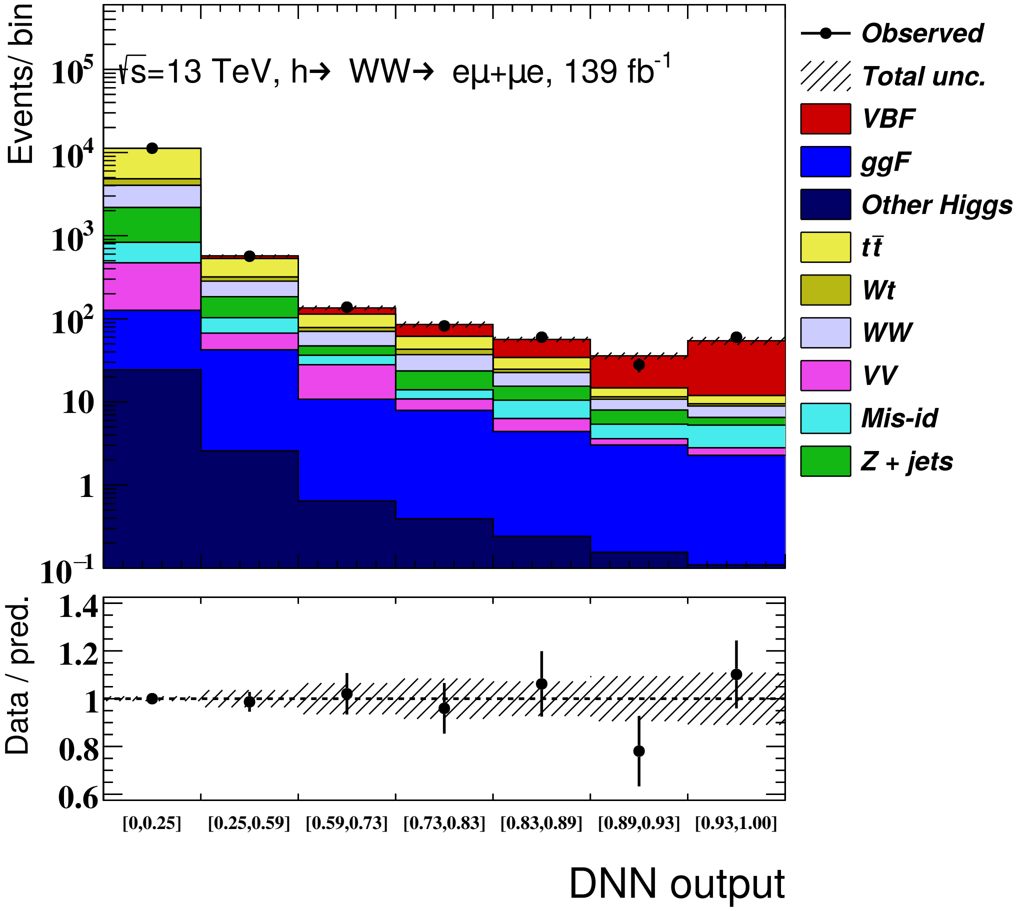


Figure 9.36: Post-fit distribution of the DNN output. The VBF signal is stacked and the ratio of Data / Total MC is shown. A linear fit is performed to the ratio. The grey band represents the MC statistical uncertainties and the main sources of experimental uncertainties (JES/JER/MET/flavor-tagging).

A detailed breakdown of the uncertainties can be found in Table 9.27. The impact of the individual nuisance parameters on the uncertainty of μ_{VBF} and their obtained post-fit values (pulls) are shown in Figure 9.33. A comparison of the pulls in the observed fit and Asimov fit is shown in Figure 9.34. We observe moderate pulls and slight constraints for some of the nuisance parameters.

The correlation matrix of the fitted nuisance parameters to the data can be found in Figure 9.35. Only correlations among parameters with a magnitude greater than 20% are shown. The post-fit distribution for the DNN output is shown in Figure 9.36.

Post-fit event yields for the various control and signal regions on the Observed data set are presented in Table 9.28.

The cross-section times branching fraction, $\sigma_{\mathrm{VBF}} \cdot \mathcal{B}_{H\rightarrow WW*}$, is measured to be

$$\sigma_{\mathrm{VBF}} \cdot \mathcal{B}_{H\rightarrow WW*} = \\ 0.85^{+0.20}_{-0.17}\ \mathrm{pb} = 0.85 \pm 0.10(\mathrm{stat.})^{+0.08}_{-0.07}(\mathrm{exp\ syst.})^{+0.13}_{-0.10}(\mathrm{sig.\ theo.})^{+0.07}_{-0.06}(\mathrm{bkg.\ theo.})\ \mathrm{pb}, \tag{9.34}$$

in comparison to the SM predicted value of 0.81 ± 0.02 pb [135].

Table 9.30 shows the relative impact of the main uncertainties on the measured value of μ_{VBF}. The measurements of the cross-section times branching fraction and the signal strength are dominated by theory uncertainties, of which VBF signal uncertainties make up the largest contribution.

9.8 Conclusions

A search for VBF production of Higgs bosons in the $H\rightarrow WW^*$ decay channel is performed. This result is based on a dataset of proton-proton collisions recorded with the ATLAS detector at the LHC in 2015-2018 at a center-of-mass energy of 13 TeV and corresponds to an integrated luminosity of 139 fb^{-1}. The observed

Table 9.29: Post-fit MC and data yields in the VBF SRs. Yields in the highest DNN output bin are also presented. The quoted uncertainties correspond to the statistical uncertainties, together with the experimental and theory modeling systematics. The sum of all the contributions may differ from the total value due to rounding. Moreover, the total uncertainty differs from the sum in quadrature of the single-process uncertainties due to anti-correlation effects in their systematic sources which dominate over their MC statistical uncertainties.

Process	Total	Highest DNN bin
H_{VBF}	209 ± 37	42.5 ± 6.5
H_{ggF}	169 ± 62	2.2 ± 1.5
Other Higgs	28 ± 2.0	0.1 ± 0.3
$t\bar{t}/Wt$	7520 ± 830	3.0 ± 1.7
Z/γ^*	1460 ± 370	1.2 ± 1.1
WW	2000 ± 350	2.4 ± 1.6
Mis-Id	416 ± 58	2.5 ± 1.6
Other VV	392 ± 64	0.5 ± 0.7
Total	12200 ± 120	54.5 ± 6.0
Observed	12189	60

(expected) signal significance is found to be 7.0 (6.2) σ, while the product of the total VBF cross-section times the $H \rightarrow WW^*$ branching fraction is measured at 0.85 ± 0.10 (stat.) $^{+0.17}_{-0.13}$ (syst.) pb, compatible with the Standard Model prediction of 0.81 ± 0.02 pb. These results provide an observation of the VBF production of Higgs bosons subsequently decaying to a pair of W bosons.

Table 9.30: Breakdown of impacts on the signal strength μ_{VBF}. The uncertainties are estimated by the breakdown method, in which nuisance parameters associated with the uncertainty group in question are first fixed to their best fit value and the uncertainty on the measured signal strength is recomputed [121]. The quadrature difference between the original and recomputed uncertainties present the impact of the uncertainty group. The uncertainties of the main components were calculated by iteratively fixing the respective sets of nuisance parameters and calculating the quadrature difference to the previous step, in reverse order of display.

Source	$\Delta\mu_{\text{VBF}}/\mu_{\text{VBF}}$ [%]
Data statistics	12.5
Total systematics	17.8
Experimental uncertainties	8.8
Missing ET	4.7
MC statistics	3.1
Jet energy scale	2.2
Luminosity	1.9
Mmodeling of pile-up	1.7
b-tagging	1.6
Jet energy resolution	1.4
Misidentified leptons	0.9
VBF signal theory uncertainties	14.4
Background theory uncertainties	7.7
ggF Higgs	5.2
Top-quark	3.3
WW	2.5
Z+jets	1.9
Total	22

Chapter 10

Conclusion

In this thesis studies to test of the Standard Model have been presented. The main analysis focuses on the search of Higgs boson production via vector boson fusion in the $WW^* \rightarrow \ell\nu\ell\nu$ decay channel predicted by the Standard Model. This measurement was done using the full Run-2 data collected by the ATLAS detector at the LHC in 2015-2018 at a center-of-mass energy of 13 TeV. A deep neural network was used as a classifier between signal and background processes which significantly improved the sensitivity of the analysis compared to its previous iterations. With about four times more data and better multivariate classifier than the previous measurement, this analysis lead to the first observation of the VBF production mode in $H\rightarrow WW^*\rightarrow \ell\nu\ell\nu$ channel with an observed significance of 7.0 σ. Moreover, the total VBF cross-section times the $H\rightarrow WW^*$ branching fraction is measured at 0.85 ± 0.10 (stat.) ${}^{+0.17}_{-0.13}$ (syst.) pb, compatible with the Standard Model prediction of 0.81 ± 0.02 pb.

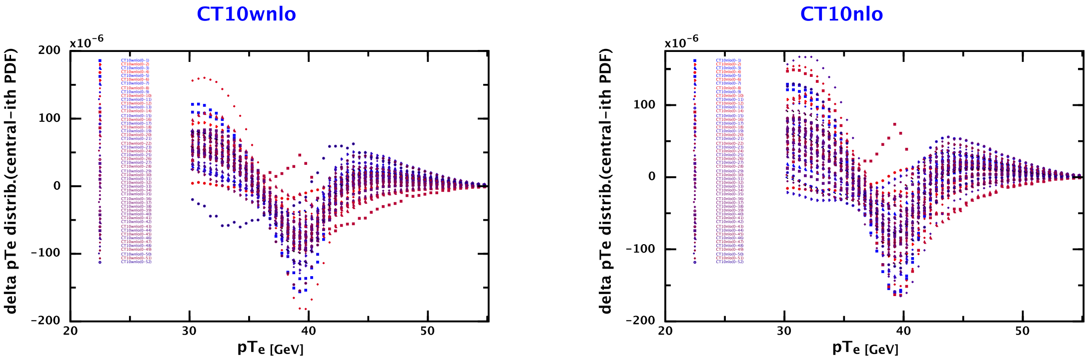
Another important test of the Standard Model was addressed in the precise measurement of the W boson mass using full proton-antiproton collision data set collected by CDF-II detector at the Tevatron at 2 TeV center-of-mass energy. One of the most important sources of systematic uncertainty in this measurement is due to parton density function (PDF). In this project, a method was developed which could be used reduce the computational cost of estimating the systematic uncertainty in W boson mass measurement due to a given PDF. With reduced computational cost to estimate systematics due to PDFs, more latest PDFsets could be tried to reduce the PDF uncertainties in W boson mass measurement.

Search for beyond the Standard Model study was carried out in the search for heavy truly top-philic vector resonance search. For this thesis, the reconstruction of final state, specifically the spectator top quark pairs in the 4-top quark final state was addressed. For this purpose, a combinatorial method ‘Buckets of Tops’ algorithm was modified and significantly improved to reconstruct the spectator semi-leptonic top quark pairs.

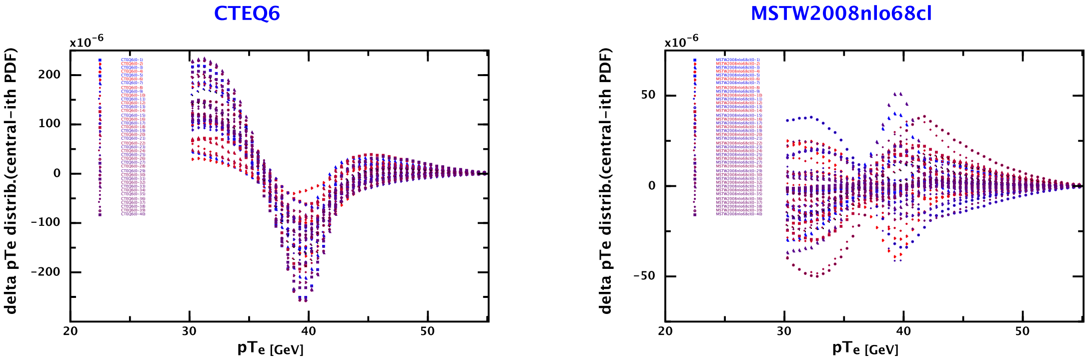
Appendix A

Appendix for Chapter 7: W mass difference plots

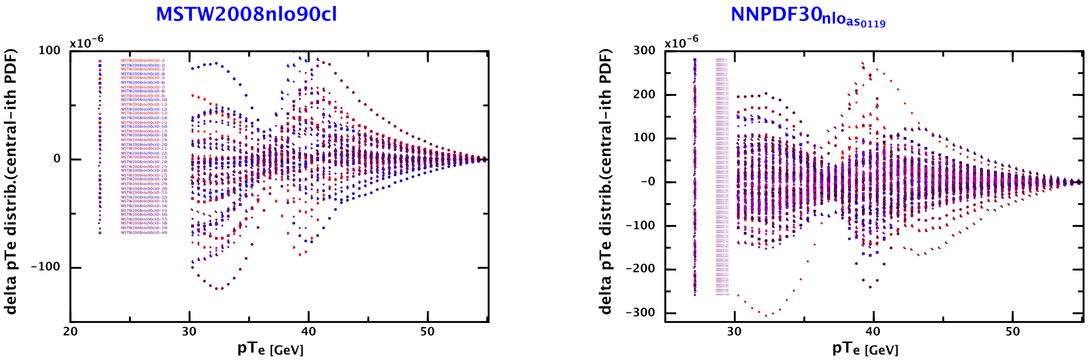
The difference plots are defined as the difference of the normalized O (where O refers to m_T, p_{T}^{e} and p_{T}^{ν} here) can be distribution using an error eigenvector PDF from the normalized O distribution using the central (nominal) PDF of a PDF set, discussed in Section 7.3.2. Here are the difference plots for p_T of electron and neutrino decaying from a W boson, with all the error eigenvectors (replica in the case of NNPDF) for a PDF set overlaid in a single plot. The inflection point is independent of all PDFs and near the Jacobian edge of the O distribution.



(a) CT10W (p_T^e) (b) CT10 (p_T^e)

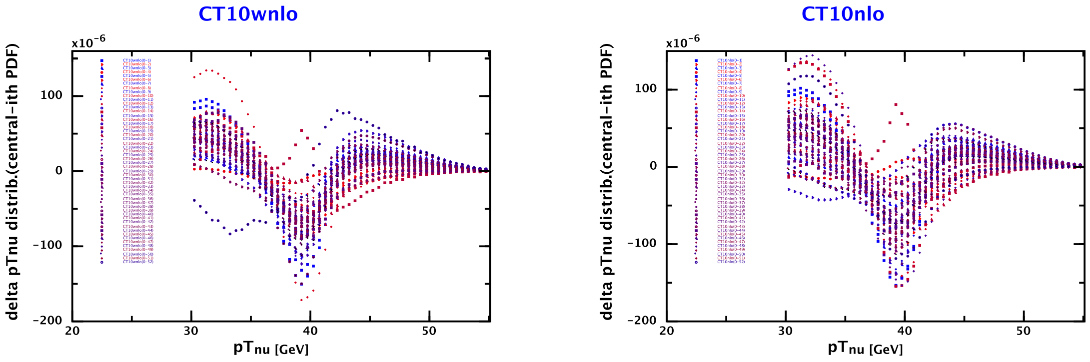


(c) CTEQ6 (p_T^e) (d) MSTW2008 (NLO 68% CL)(p_T^e)



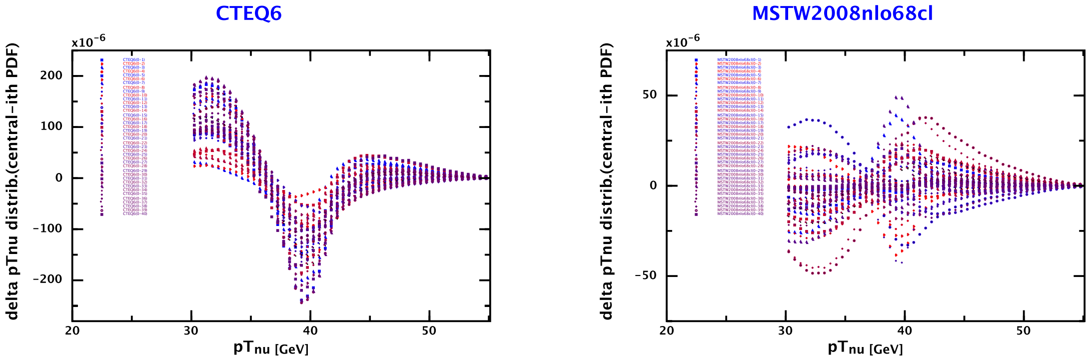
(e) MSTW2008 (NLO 90% CL)(p_T^e) (f) NNPDF3.0 (p_T^e)

Figure A.1: The electron transverse momentum (p_T^e) difference plots for PDF sets of CT10W, CT10, CTEQ6, MSTW2008 (NLO 68% CL), MSTW2008 (NLO 90% CL), and NNPDF3.0 NLO ($\alpha_s = 0.119$). The inflection point is near the Jacobian edge near $M_W/2$.



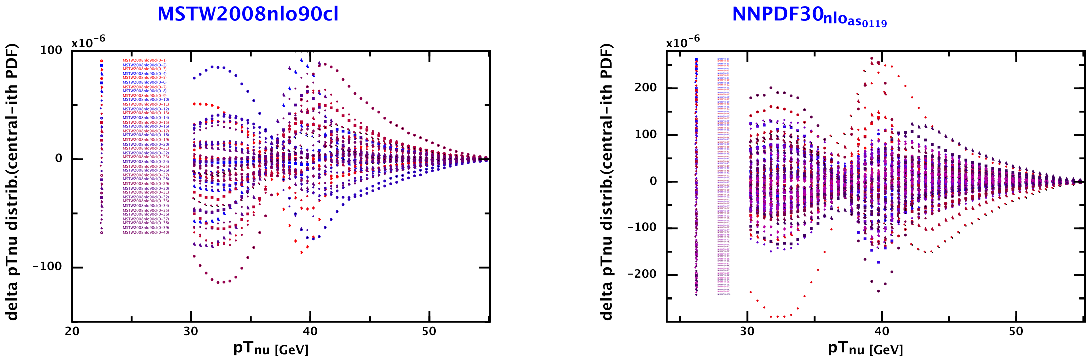
(a) CT10W (p_T^ν)

(b) CT10 (p_T^ν)



(c) CTEQ6 (p_T^ν)

(d) MSTW2008 (NLO 68% CL)(p_T^ν)



(e) MSTW2008 (NLO 90% CL)(p_T^ν)

(f) NNPDF3.0 (p_T^ν)

Figure A.2: The neutrino transverse momentum (p_T^ν) difference plots for PDF sets of CT10W, CT10, CTEQ6, MSTW2008 (NLO 68% CL), MSTW2008 (NLO 90% CL), and NNPDF3.0 NLO ($\alpha_s = 0.119$). The inflection point is near the Jacobian edge near $M_W/2$.

Bibliography

[1] N. Serra, *Standard Model*, `https://www.physik.uzh.ch/groups/serra/StandardModel.html`.

[2] *LHC Higgs Cross Section Working Group. Picture gallery*, `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures`.

[3] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des acclrateurs du CERN - Aot 2018*, `https://cds.cern.ch/record/2636343`, General Photo.

[4] ATLAS Collaboration, *Luminosity Public Results Run 2*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`.

[5] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[6] ATLAS Collaboration Collaboration, K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, Tech. Rep. ATL-PHYS-PROC-2016-104, CERN, Geneva, Aug, 2016. `http://cds.cern.ch/record/2209070`. 15 pages, EPS-HEP 2015 Proceedings.

[7] J. Pequenao, *Computer Generated image of the ATLAS calorimeter*, Mar, 2008.

[8] J. Pequenao, *Computer generated image of the ATLAS Muons subsystem*, Mar, 2008.

[9] A. Salzburger, *The ATLAS Track Extrapolation Package*, Tech. Rep. ATL-SOFT-PUB-2007-005. ATL-COM-SOFT-2007-010, CERN, Geneva, Jun, 2007. `https://cds.cern.ch/record/1038100`.

[10] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`.

[11] A. Sidoti, *Minimum bias trigger scintillators in atlas run ii*, Journal of Instrumentation **9** (2014) C10020.

[12] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, `https://cds.cern.ch/record/1474107`.

[13] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-051, 2015, `https://cds.cern.ch/record/2110140`.

[14] ATLAS Collaboration, *Measurement of track reconstruction inefficiencies in the core of jets via pixel dE/dx with the ATLAS experiment using* $\sqrt{s} = 13$ *TeV pp collision data*, ATL-PHYS-PUB-2016-007, 2016, `https://cds.cern.ch/record/2140460`.

[15] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, `arXiv:1406.7690 [hep-ex]`.

[16] M. Aoki, C. F. Anders, D. E. Ferreira de Lima, S. Suzuki, J. Maeda, K. Terashi, T. Nobe, M. Kolb, D. Sosa, A. Schoening, and O. Sasaki, *A search for top-antitop resonances in all hadronic final state using proton-proton collisions at sqrts=13 TeV*, Tech. Rep. ATL-COM-PHYS-2016-1496, CERN, Geneva, Oct, 2016. `https://cds.cern.ch/record/2226464`.

[17] M. R. Buckley, T. Plehn, and M. Takeuchi, *Buckets of Tops*, JHEP **08** (2013) 086, `arXiv:1302.6238 [hep-ph]`.

[18] ATLAS Collaboration, *Summary plots from the ATLAS Standard Model physics group*, 2018, `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/`.

[19] ATLAS Collaboration, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, Tech. Rep. CERN-LHCC-2012-022. LHCC-I-023, CERN, Geneva, Dec, 2012. `https://cds.cern.ch/record/1502664`. Draft version for comments.

[20] T. Aaltonen, S. Amerio, D. Amidei, A. Anastassov, A. Annovi, J. Antos, G. Apollinari, J. Appel, T. Arisawa, A. Artikov, et al., *Precise measurement of the W-boson mass with the Collider Detector at Fermilab*, Physical Review D **89** (2014) 072003.

[21] D. Griffiths, *Introduction to elementary particles.* John Wiley & Sons, 2008.

[22] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Physical review D **2** (1970) 1285.

[23] S. Weinberg, *A model of leptons*, Physical review letters **19** (1967) 1264.

[24] A. Salam, *Weak and electromagnetic interactions*, pp. , 244–254. World Scientific, 1994.

[25] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear physics **22** (1961) 579–588.

[26] C.-N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Physical review **96** (1954) 191.

[27] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Physical Review Letters **13** (1964) 508.

[28] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.

[29] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Physical Review **145** (1966) 1156.

[30] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Physical Review Letters **13** (1964) 321.

[31] G. S. Guralnik, C. R. Hagen, and T. W. Kibble, *Global conservation laws and massless particles*, Physical Review Letters **13** (1964) 585.

[32] T. W. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Physical Review **155** (1967) 1554.

[33] 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]`.

[34] 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]`.

[35] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Physical Review Letters **30** (1973) 1343.

[36] L. collaboration et al., *LHC Design Report, Vol. 1 The LHC Main Ring*, tech. rep., CERN-2004-003, 2004.

[37] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report. 2. The LHC infrastructure and general services,*.

[38] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report, volume III, The LHC injector chain*, CERN, Geneva (2004).

[39] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[40] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19, 2010, `https://cds.cern.ch/record/1291633`.

[41] ATLAS Collaboration Collaboration, A. La Rosa and A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, Tech. Rep. ATL-INDET-PROC-2016-004. 12, CERN, Geneva, Oct, 2016. `https://cds.cern.ch/record/2221972`.

[42] ATLAS Collaboration Collaboration,, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Tech. Rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, Geneva, Sep, 2017. `https://cds.cern.ch/record/2285585`.

[43] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[44] A. Ahmad, Z. Albrechtskirchinger, P. Allport, J. Alonso, L. Andricek, R. Apsimon, A. Barr, R. Bates, G. Beck, P. Bell, et al., *The silicon microstrip sensors of the ATLAS semiconductor tracker*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **578** (2007) 98–118.

[45] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, `arXiv:1404.7473 [hep-ex]`.

[46] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*, ATLAS-TDR-19-ADD-1, 2012, `https://cds.cern.ch/record/1451888`.

[47] E. Abat, T. Addy, T. Åkesson, J. Alison, F. Anghinolfi, E. Arik, M. Arik, G. Atoian, B. Auerbach, O. Baker, et al., *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, Journal of Instrumentation **3** (2008) P02013.

[48] ATLAS Collaboration, *Particle Identification Performance of the ATLAS Transition Radiation Tracker*, ATLAS-CONF-2011-128, 2011, `https://cds.cern.ch/record/1383793`.

[49] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, `arXiv:1702.06473 [hep-ex]`.

[50] ATLAS Collaboration Collaboration,, *ATLAS liquid-argon calorimeter: Technical Design Report.* Technical Design Report ATLAS. CERN, Geneva, 1996. `https://cds.cern.ch/record/331061`.

[51] ATLAS Collaboration Collaboration,, *ATLAS tile calorimeter: Technical Design Report.* Technical Design Report ATLAS. CERN, Geneva, 1996. `https://cds.cern.ch/record/331062`.

[52] C. Grupen, B. A. Shwartz, H. Spieler, S. Eidelman, and T. Stroh, *Particle detectors*, vol. 11. Cambridge university press Cambridge, 2008.

[53] K. Olive, *Review of Particle Physics*, Chinese Physics C **40** (2016) 100001, `https://doi.org/10.1088%2F1674-1137%2F40%2F10%2F100001`.

[54] ATLAS Collaboration Collaboration,, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. `https://cds.cern.ch/record/331068`.

[55] S. Palestini, *The muon spectrometer of the ATLAS experiment*, Nuclear Physics B-Proceedings Supplements **125** (2003) 337–345.

[56] ATLAS Collaboration Collaboration,, *ATLAS level-1 trigger: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998. `https://cds.cern.ch/record/381429`.

[57] ATLAS Collaboration Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003. `https://cds.cern.ch/record/616089`.

[58] A. Rosenfeld, J. L. Pfaltz, et al., *Sequential operations in digital picture processing.*, J. ACM **13** (1966) 471–494, `http://doi.acm.org/10.1145/321356.321357`.

[59] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. `https://cds.cern.ch/record/1020106`.

[60] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Journal of basic Engineering **82** (1960) 35–45.

[61] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444–450.

[62] D. Wicke, *A new algorithm for solving tracking ambiguities*, tech. rep., 1998.

[63] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, `arXiv:1611.10235 [hep-ex]`.

[64] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, `https://cds.cern.ch/record/1099735`.

[65] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[66] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Physical Review D **48** (1993) 3160.

[67] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, `arXiv:1802.08168 [hep-ex]`.

[68] ATLAS Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015-2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, `https://cds.cern.ch/record/2625233`.

[69] *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at* $\sqrt{s}$= *13 TeV*, Tech. Rep. ATL-PHYS-PUB-2015-027, CERN, Geneva, Jul, 2015. `https://cds.cern.ch/record/2037904`.

[70] *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-PHYS-PUB-2015-023, CERN, Geneva, Jul, 2015. `https://cds.cern.ch/record/2037700`.

[71] Gaudiello, A., *Search for metastable heavy charged particles with large ionisation energy loss in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS experiment*, Nuovo Cim. **C40** (2017) 16.

[72] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, `arXiv:1704.07983 [hep-ex]`.

[73] Goodfellow, I. and Bengio, Y. and Courville, A., *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[74] Gagnon, Louis-Guillaume, *Development and performance of track reconstruction algorithms at the energy frontier with the ATLAS detector*,. `https://cds.cern.ch/record/2242031`.

[75] ATLAS Collaboration, *Robustness of the Artificial Neural Network Clustering Algorithm of the ATLAS experiment*, ATL-PHYS-PUB-2015-052, 2015, `https://cds.cern.ch/record/2116350`.

[76] Sjostrand, T. and Mrenna, S. and Skands, P.Z., *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820 [hep-ph]`.

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

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

[79] S. Sen, R. Xu, and A. Kotwal, *Effects of reduced charge information on CTIDE neural network performances*, Tech. Rep. ATL-COM-INDET-2018-001, CERN, Geneva, Jan, 2018. `https://cds.cern.ch/record/2300393`.

[80] J. Wenninger, M. Benedikt, K. Oide, and F. Zimmermann, *Future circular collider study lepton collider parameters*, CERN EDMS no **1346082** (2014).

[81] A. Kotwal, S. Chekanov, and M. Low, *Double Higgs boson production in the 4 τ channel from resonances in longitudinal vector boson scattering at a 100 TeV collider*, Physical Review D **91** (2015) 114018.

[82] A. V. Kotwal, M. J. Ramsey-Musolf, J. M. No, and P. Winslow, *Singlet-catalyzed electroweak phase transitions in the 100 TeV frontier*, Physical Review D **94** (2016) 035022.

[83] S. Sen, A. Kotwal, S. Chekanov, L. Gray, N. Tran, and S.-s. Yu, *Detectors for Superboosted τ-leptons at Future Circular Colliders*, in *38th International Conference on High Energy Physics*, p. , 788, SISSA Medialab. 2017.

[84] T. Behnke, J. E. Brau, P. N. Burrows, J. Fuster, M. Peskin, M. Stanitzki, Y. Sugimoto, S. Yamada, and H. Yamamoto, *The International Linear Collider Technical Design Report - Volume 4: Detectors*, 2013.

[85] S. Chekanov, *HepSim: a repository with predictions for high-energy physics experiments*, Advances in High Energy Physics **2015** (2015).

[86] G. Aad, B. Abbott, J. Abdallah, S. A. Khalek, O. Abdinov, R. Aben, B. Abi, M. Abolins, O. AbouZeid, H. Abramowicz, et al., *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*, The European Physical Journal C **75** (2015) 303.

[87] A. Sirlin, *Radiative corrections in the SU (2) L× U (1) theory: a simple renormalization framework*, Physical Review D **22** (1980) 971.

[88] R. Placakyte, *Parton Distribution Functions*, arXiv preprint arXiv:1111.5452 (2011).

[89] T. Aaltonen, A. Abulencia, J. Adelman, T. Affolder, T. Akimoto, M. G. Albrow, S. Amerio, D. Amidei, A. Anastassov, K. Anikeev, et al., *First measurement of the W-boson mass in run II of the tevatron*, Physical review letters **99** (2007) 151801.

[90] T. Aaltonen, A. Abulencia, J. Adelman, T. Akimoto, M. G. Albrow, B. A. Gonzalez, S. Amerio, D. Amidei, A. Anastassov, A. Annovi, et al., *First run II measurement of the W boson mass at the Fermilab Tevatron*, Physical Review D **77** (2008) 112001.

[91] A. Abulencia, D. Acosta, J. Adelman, T. Affolder, T. Akimoto, M. Albrow, D. Ambrose, S. Amerio, D. Amidei, A. Anastassov, et al., *Measurements of inclusive W and Z cross sections in collisions at*, Journal of Physics G: Nuclear and Particle Physics **34** (2007) 2457.

[92] D. Acosta, J. Adelman, T. Affolder, T. Akimoto, M. Albrow, D. Ambrose, S. Amerio, D. Amidei, A. Anastassov, K. Anikeev, et al., *Measurement of the J/ψ meson and b-hadron production cross sections in p p^- collisions at s= 1960 G e V*, Physical Review D **71** (2005) 032001.

[93] L. Lipatov, *The parton model and perturbation theory*, Yad. Fiz. **20** (1975) 94–102.

[94] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) 298–318.

[95] Y. L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e+ e- annihilation by perturbation theory in quantum chromodynamics*, Zh. Eksp. Teor. Fiz **73** (1977) 1216.

[96] G. Curci, W. Furmanski, and R. Petronzio, *Evolution of parton densities beyond leading order: the non-singlet case*, Nuclear Physics B **175** (1980) 27–92.

[97] A. Vogt, S. Moch, and J. A. Vermaseren, *The three-loop splitting functions in QCD: the singlet case*, Nuclear Physics B **691** (2004) 129–181.

[98] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, *Implications of CTEQ global analysis for collider observables*, Physical Review D **78** (2008) 013004.

[99] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, The European Physical Journal C **63** (2009) 189–285.

[100] M. Glück, P. Jimenez-Delgado, and E. Reya, *Dynamical parton distributions of the nucleon and very small-x physics*, The European Physical Journal C **53** (2008) 355–366.

[101] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, The European Physical Journal C **75** (2015) 1–20.

[102] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization of hard processes in QCD*, pp. , 1–91. World Scientific, 1989.

[103] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, Journal of High Energy Physics **2004** (2004) 040.

[104] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) 070.

[105] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** (2010) 43.

[106] G. Ferretti and D. Karateev, *Fermionic UV completions of Composite Higgs models*, JHEP **03** (2014) 077, `arXiv:1312.5330 [hep-ph]`.

[107] L. Vecchi, *A dangerous irrelevant UV-completion of the composite Higgs*, JHEP **02** (2017) 094, `arXiv:1506.00623 [hep-ph]`.

[108] K. Agashe, A. Delgado, M. J. May, and R. Sundrum, *RS1, custodial isospin and precision tests*, JHEP **08** (2003) 050, `arXiv:hep-ph/0308036 [hep-ph]`.

[109] K. Agashe, R. Contino, and A. Pomarol, *The Minimal composite Higgs model*, Nucl. Phys. **B719** (2005) 165–187, `arXiv:hep-ph/0412089 [hep-ph]`.

[110] L. Landau, *On the angular momentum of a system of two photons*, Dokl. Akad. Nauk SSSR **60** (1948) 207–209.

[111] C.-N. Yang, *Selection Rules for the Dematerialization of a Particle Into Two Photons*, Phys. Rev. **77** (1950) 242–245.

[112] B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski, and C. Wanotayaroj, *Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, Journal of High Energy Physics **2015** (2015) 75.

[113] T. Plehn, M. Spannowsky, and M. Takeuchi, *How to improve top-quark tagging*, Physical Review D **85** (2012) 034029.

[114] V. M. Abazov, B. Abbott, B. Acharya, M. Adams, T. Adams, J. Agnew, G. Alexeev, G. Alkhazov, A. Alton, A. Askew, et al., *Precise measurement of the top quark mass in dilepton decays using optimized neutrino weighting*, Physics Letters B **752** (2016) 18–26.

[115] CMS Collaboration, *Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV*, `arXiv:1806.05246 [hep-ex]`.

[116] CMS Collaboration, *Observation of a new boson with mass near* 125 *GeV in pp collisions at* $\sqrt{s} = 7$ *and* 8 *TeV*, JHEP **06** (2013) 081, `arXiv:1303.4571 [hep-ex]`.

[117] ATLAS Collaboration, *Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the* $H \to WW^* \to e\nu\mu\nu$ *decay channel in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **789** (2019) 508, `arXiv:1808.09054 [hep-ex]`.

[118] ATLAS Collaboration, *Observation and measurement of Higgs boson decays to* WW^* *with the ATLAS detector*, Phys. Rev. D **92** (2015) 012006, `arXiv:1412.2641 [hep-ex]`.

[119] A. Aggarwal, S. Argyropoulos, M. J. Basso, A. K. Becker, C. Bertella, C. D. Burgard, D. Duda, D. Du, A. Farilla, F. Filthaut, A. Gavrilyuk, N. A. Grieser, R. Gugel, R. L. Hayes, C. Hays, P.-h. J. Hsu, C.-t. Huang, Y.-c. Huang, G. Iakovidis, B. P. Jaeger, C. Kitsaki, K. Koeneke, A. Kotwal, K. Lehmann, H.-c. Lin, K. Lohwasser, Y.-J. Lu, A. Montalbano, T. H. Park, F. Pasquali, E. Ramakoti, C. Schmitt, S. Sen, D. R. Shope, H. A. Smith, J. Strandberg, M. Strauss, I. Tsukerman, S. Wang, B. T. Winter, G. Zemaityte, S. V. Addepalli, G. Barone, L. J. Bergsten, J. Chen, G. Rosin, S. A. Stucci, A. Tricoli, O. Stelzer-Chilton, and B. Stelzer, *Measurements of the Higgs boson production cross section via ggF and VBF in* $H \to WW^* \to \ell\nu\ell\nu$ *with 139* fb^{-1} *of data collected with the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-COM-PHYS-2020-302, CERN, Geneva, May, 2020. `https://cds.cern.ch/record/2716457`.

[120] A. Aggarwal, S. Argyropoulos, M. J. Basso, A. K. Becker, C. Bertella, C. D. Burgard, D. Du, A. Farilla, F. Filthaut, A. Gavrilyuk, N. A. Grieser, R. Gugel, R. L. Hayes, C. Hays, P.-h. J. Hsu, C.-t. Huang, Y.-c. Huang, G. Iakovidis, B. P. Jaeger, C. Kitsaki, K. Koeneke, A. Kotwal, K. Lehmann, H.-c. Lin, K. Lohwasser, Y.-J. Lu, A. Montalbano, T. H. Park, F. Pasquali, E. Ramakoti, C. Schmitt, S. Sen, D. R. Shope, H. A. Smith, J. Strandberg, M. Strauss, I. Tsukerman, S. Wang, B. T. Winter, and G. Zemaityte, *Optimization of the* $H \to WW^*$ *analysis using the full Run 2 dataset*, Tech. Rep. ATL-COM-PHYS-2020-303, CERN, Geneva, May, 2020. `https://cds.cern.ch/record/2716687`.

[121] A. Collaboration, *Observation of vector-boson-fusion production of Higgs bosons in the* $H \to WW^* \to e\nu\mu\nu$ *decay channel in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Tech. Rep. ATLAS-COM-CONF-2020-041, CERN, Geneva, Jul, 2020. `https://cds.cern.ch/record/2724182`.

[122] ATLAS Collaboration, *Measurement of gluon fusion and vector boson fusion Higgs boson production cross-sections in the* $H \to WW^* \to e\nu\mu\nu$ *decay*

channel in pp collisions at at $\sqrt{s} = 13$ TeV with the ATLAS detector , ATLAS-CONF-2018-004, 2018, `https://cds.cern.ch/record/2308392`.

[123] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037, `arXiv:0911.5299 [hep-ph]`.

[124] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, `arXiv:1007.2241 [hep-ph]`.

[125] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, `arXiv:1405.0301 [hep-ph]`.

[126] ATLAS Collaboration, G. Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, `arXiv:1406.3660 [hep-ex]`.

[127] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, `arXiv:0811.4622 [hep-ph]`.

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

[129] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** (2008) 852–867.

[130] M. Ciccolini, A. Denner and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2-jets via weak interactions at the LHC*, Phys. Rev. Lett. **99** (2007), `arXiv:0707.0381 [hep-ph]`.

[131] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys. Rev. D **77** (2008) 013002, `arXiv:0710.4749 [hep-ph]`.

[132] P. Bolzoni et al., *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010), `arXiv:1003.4451 [hep-ph]`.

[133] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, `arXiv:1309.0017 [hep-ph]`.

[134] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, `arXiv:1501.04637 [hep-ph]`.

[135] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, `arXiv:1610.07922 [hep-ph]`.

[136] M. Ciccolini, S. Dittmaier, and M. Kramer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. D **68** (2003) 073003, `arXiv:hep-ph/0306234`.

[137] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016), `arXiv:1510.03865 [hep-ph]`.

[138] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56–74, `arXiv:hep-ph/9704448`.

[139] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber, *Radiative corrections to the semileptonic and hadronic Higgs-boson decays H —¿ W W / Z Z —¿ 4 fermions*, JHEP **02** (2007) 080, `arXiv:hep-ph/0611234`.

[140] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber, *Precise predictions for the Higgs-boson decay H —¿ WW/ZZ —¿ 4 leptons*, Phys. Rev. D **74** (2006) 013004, `arXiv:hep-ph/0604011`.

[141] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber, *Precision calculations for the Higgs decays H —¿ ZZ/WW —¿ 4 leptons*, Nucl. Phys. B Proc. Suppl. **160** (2006) 131–135, `arXiv:hep-ph/0607060`.

[142] M. C. et al, *Top-pair production at the LHC through NNLO QCD and NLO EW*, JHEP **186** (2017).

[143] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, `arXiv:0805.3067 [hep-ph]`.

[144] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. D **69** (2004) 094008, `arXiv:hep-ph/0312266`.

[145] A. Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 protonproton collisions at the LHC*, The European Physical Journal C **77** (2017), `http://dx.doi.org/10.1140/epjc/s10052-017-4887-5`.

[146] ATLAS Collaboration Collaboration,, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, Tech. Rep. ATL-PHYS-PUB-2019-015, CERN, Geneva, Apr, 2019. `https://cds.cern.ch/record/2670380`.

[147] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton coll ision data*, JINST **14** (2019) P12006, `arXiv:1908.00005 [hep-ex]`.

[148] ATLAS Collaboration Collaboration, K. B. et al., *Optimisation note of the ggF+VBF analysis in* $H \rightarrow WW^*$ *using* $36\ fb^{-1}$ *of data collected with the ATLAS detector at* $\sqrt{s} = 13\ TeV$, tech. rep. `https://cds.cern.ch/record/2276101`. ATL-COM-PHYS-2017-1089.

[149] ATLAS Collaboration, *Muon reconstruction performance in early* $\sqrt{s} = 13\ TeV$ *data*, ATL-PHYS-PUB-2015-037, 2015, `https://cds.cern.ch/record/2047831`.

[150] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, `arXiv:1703.10485 [hep-ex]`.

[151] ATLAS Collaboration, *Object-based missing transverse momentum significance in the ATLAS Detector*, ATLAS-CONF-2018-038, 2018, `https://cds.cern.ch/record/2630948`.

[152] Will Buttinger, *Updated Overlap Removal Working Points*, Tech. Rep. (visited on 16/12/2015), 2015. `https://indico.cern.ch/event/457238/`.

[153] tech. rep. `https://indico.cern.ch/event/457238/contribution/17/attachments/1207345/1759805/OR_NotesDec2015_final.pdf`.

[154] tech. rep. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AnalysisRelease`.

[155] T. Plehn, D. L. Rainwater, and D. Zeppenfeld, *A Method for identifying* $H \rightarrow \tau^+\tau^- \rightarrow e^\pm\mu^\mp p_T$ *at the CERN LHC*, Phys. Rev. **D61** (2000) 093005, `arXiv:hep-ph/9911385 [hep-ph]`.

[156] F. Chollet et al., *Keras*, `https://keras.io`, 2015.

[157] *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. `https://www.tensorflow.org/`. Software available from tensorflow.org.

[158] ATLAS Collaboration, *Measurement of the* $t\bar{t}$ *production cross-section in* pp *collisions at* $\sqrt{s} = 8\ TeV$ *using* $e\mu$ *events with* b*-tagged jets*, ATLAS-CONF-2013-097, 2013, `https://cds.cern.ch/record/1600596`.

[159] ATLAS Collaboration, *Measurement of the cross-section for associated production of a top quark and a W boson at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, ATLAS-CONF-2013-100, 2013, `https://cds.cern.ch/record/1600799`.

[160] C. G. Lester and D. J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, Phys. Lett. **B463** (1999) 99–103, `arXiv:hep-ph/9906349 [hep-ph]`.

[161] C. Bertella, C. D. Burgard, L. S. Bruni, R. Castelijn, J. a. Barreiro Guimarães da Costa, K. F. Di Petrillo, D. Du, P. Ferrari, F. Filthaut, A. Gavrilyuk, M. Geisen, R. Gugel, R. Gupta, C. Hays, P.-h. J. Hsu, K. Koeneke, A. Kotwal, J. Llorente Merino, Y.-f. Lo, Y.-J. Lu, L. Ma, J. Oliver, S. Pataraia, S. Sen, P. E. Sidebo, W. Song, C. Schmitt, D. R. Shope, J. Strandberg, M. Strauss, I. Tsukerman, G. Zemaityte, Y. Zhao, M.-j. Tsai, H. Hsu, A. N. Tuna, M. White, P. Jackson, R. Kehoe, B. P. Jaeger, D. Duda, and A. K. Becker, *Measurements of the Higgs boson production cross section via ggF and VBF in* $H \rightarrow WW^{*} \rightarrow l\nu l\nu$ *with 36.1* fb^{-1} *of data collected with the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-COM-PHYS-2017-1094, CERN, Geneva, Jul, 2017. `https://cds.cern.ch/record/2276143`.

[162] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, `arXiv:1007.1727 [physics.data-an]`, [Erratum: Eur. Phys. J.C73,2501(2013)].

[163] ATLAS Collaboration, M. Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, Eur. Phys. J. **C79** (2019) 639, `arXiv:1902.04655 [physics.ins-det]`.

[164] ATLAS Collaboration, M. Aaboud et al., *Electron and photon energy calibration with the ATLAS detector using 20152016 LHC proton-proton collision data*, JINST **14** (2019) P03017, `arXiv:1812.03848 [hep-ex]`.

[165] ATLAS Collaboration, G. Aad et al., *Muon reconstruction performance of the ATLAS detector in protonproton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76** (2016) 292, `arXiv:1603.05598 [hep-ex]`.

[166] *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul, 2015. `http://cds.cern.ch/record/2037613`.

[167] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, `arXiv:1703.09665 [hep-ex]`.

[168] ATLAS Collaboration Collaboration,, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, Tech. Rep. ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun, 2016. `https://cds.cern.ch/record/2160731`.

[169] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisio ns at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, `arXiv:1907.05120 [hep-ex]`.

[170] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, `arXiv:1909.00761 [hep-ex]`.

[171] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, `arXiv:2004.13447 [hep-ex]`.

[172] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, `https://cds.cern.ch/record/2677054`.

[173] I. W. Stewart and F. J. Tackmann, *Theory uncertainties for Higgs mass and other searches using jet bins*, Physical Review D **85** (2012) 034011.

[174] S. Frixione et al., *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, `arXiv:0707.3088 [hep-ph]`.

Biography

Sourav Sen was born in New Delhi, India. He received his integrated BS-MS degree in physics from the Indian Institute of Science Education and Research, Kolkata, India, in 2013. He completed his Ph.D. research in experimental particle physics at Duke University in Durham, North Carolina, in 2020.