UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

OBSERVATION OF A SINGLE TOP QUARK WITH AN ASSOCIATED CHARM QUARK WITH THE ATLAS DETECTOR

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

VEENA BALAKRISHNAN
Norman, Oklahoma
2025

OBSERVATION OF A SINGLE TOP QUARK WITH AN ASSOCIATED CHARM QUARK WITH THE ATLAS DETECTOR

A DISSERTATION APPROVED FOR THE

HOMER L. DODGE DEPARTMENT OF PHYSICS AND ASTRONOMY

BY THE COMMITTEE CONSISTING OF

Dr. Braden K Abbott, Chair

Dr. Michael Strauss

Dr. Howard Baer

Dr. Arne Schwettmann

Dr. Michele Galizia

Acknowledgments

I would like to express my sincere gratitude to my PhD advisor, Dr. Brad Abbott, for his support and encouragement throughout my PhD process. He is an outstanding mentor, always ready to help whenever I was stuck or needed guidance. He was kind in providing me with the flexibility in work that increased my productivity and improved my personal well-being. During a challenging time in my academic journey, his support was a constant source of motivation to move forward. Thank you for being the best advisor one could possibly ask for, and I am forever grateful.

Next, I would like to thank my doctoral committee members, Dr Strauss, for all the engaging conversations you were kind enough to include me in, Dr Baer, for your patience and help in guiding me through the theory section during my general exam, Dr Schwettmann, for generously stepping in as a committee member when one member had to leave and finally Dr Galizia for readily being part of my committee when I reached out. Thank you all for taking the time to review my work.

I would like to thank my collaborators in the top+charm analysis. Dr Gutierrez and Marija, thank you for your valuable advice and guidance throughout the analysis. Lubos, thank you for your help with the analysis, the technical tips, and your statistical knowledge, I learned so much from you. And thank you, Taher, for your hard work on the ML section, which helped make this thesis complete.

I would also like to thank Dr Stupak, Horst, and Daniel for their support and help in various aspects of my PhD journey. Daniel, thank you for always being patient and for helping me with my coding questions without any hesitation. I also thank my authorship technical supervisor Dr Chris Meyer, for his incredible help and guidance through the project. I would also like to thank the Department of Energy (DOE) for supporting this work through their grant.

Finally I would like to thank all the wonderful people in my life, who encouraged and supported me throughout this long journey. Naming just a few, Daida, for always being one call away during tough times and being a steady source of comfort for me. Mayuri and Umang, we started this journey together and are finishing it together, thank you for pushing me forward when I got behind in between. Mayuri, I will always cherish our hallway chitchats, work discussions and travel memories. Hari and Eliza , for making the first two years of my PhD life truly memorable, and for being my family away from home. Jugal, for the great conversations and words for encouragement to keep moving forward. I am grateful to have such supportive in-laws who have been genuinely invested in my journey. Thank you all for your encouragement. To My younger sister, for brightening my dull days with your humor and words of wisdom. To my husband Rohit, my comfort zone, thank you for calming me whenever I felt overwhelmed and being my constant source of support and encouragement, I love you and could not have done this without you. And finally, to my parents, for encouraging and supporting me in every possible way to move thousands of miles away from home to chase my dream. Thank you for believing in me.

Table of Contents

List of Figures

List of Tables

Abstract

The analysis presented in this thesis is expected to provide the first observation of single top quark production with an associated charm quark. The data used for the analysis are recorded by the ATLAS detector at the LHC using the proton-proton collisions at a center-of-mass energy of 13 TeV with a full integrated luminosity of $140.1 fb^{-1}$. The recorded data is from the 2015-2018 data-taking period. The analysis is performed by defining a signal region with events consisting of a lepton (e, μ), zero light jets, one b-tagged jet, one c-tagged jet, and $E_T^{\text{miss}} > 25$ GeV in the final state and two control regions dominated by $t\bar{t}$ and $W^{\pm}$+jets samples. A Boosted Decision Tree (BDT) is trained, and the BDT output is used in the signal region to improve the measurement's sensitivity. A binned likelihood fit is performed simultaneously on the signal region and control regions using the BDT distribution in the signal region and E_T^{miss} distribution in the control regions. The data remain blinded in the signal region, and therefore only expected results are deduced from the fit. The expected uncertainty on the signal strength, μ_{tc}, is measured to be 6%, and the expected significance is more than 17, which is well above the 5σ discovery threshold, showing strong sensitivity towards the observation of this process.

Chapter 1

Introduction

The Big Bang theory is the existing scientific explanation of how the universe began 13.8 billion years ago from an initial stage of extremely high density and temperature. As the universe expanded, it cooled sufficiently to form the subatomic particles. High-energy particle accelerators, such as the LHC, simulate conditions of the early universe, helping scientists to understand fundamental particles and their interactions in detail. These fundamental particles are described by the Standard Model of particle physics, governed by Quantum Field Theory.

This thesis presents the first expected observation of the single top quark with an associated charm quark as predicted by the Standard Model at a center-of-mass energy of 13 TeV by the ATLAS detector at LHC. The rest of this chapter discusses the Standard Model, Quantum Field Theory, and the single-top + charm production mechanism.

Chapter 2 will introduce the Large Hadron Collider and the ATLAS detector. Chapter 3 describes the reconstruction of physics objects from the data collected from the ATLAS detector. The data and the Monte Carlo samples used in the analysis are presented in Chapter 4. Chapter 5 gives an overview of the selection criteria of the events and analysis regions, as well as the machine learning technique used to increase the sensitivity of the signal. Chapter 6 discusses fake background calculation and validation. Systematic uncertainties are described in Chapter 7. Lastly, statistical analysis and results are discussed in detail in Chapter 8. A binned-likelihood fit procedure is used to obtain the expected signal strength uncertainty and significance. The data remain blinded in the signal region, so only expected results are deduced from the fit.

1.1 The Standard Model

The Standard Model (SM) [1] of particle physics describes the elementary particles that constitute our universe together with the theory of interactions between them. They interact through the four fundamental forces of nature: strong, electromagnetic, weak, and gravitational. These subatomic particles are classified as fermions and bosons based on their intrinsic spin. Fermions include quarks and leptons with half-integer spin, and they obey the Pauli's exclusion principle[1]. Leptons do not respond to strong interactions, and they are affected only by electromagnetic, weak, and gravitational interactions. There are six known leptons: electron, muon, tau, electron neutrino, muon neutrino and tau neutrino, and each has its own antiparticle. The antiparticles have the same mass and spin but have opposite charge and parity[2]. There are six types of quarks: up, down, strange, charm, top, and bottom. Bosons have integer spin and are the force carriers that mediate the interaction between particles. The Higgs boson [2], a massive scalar boson discovered in 2012, explains how fundamental particles acquire mass (Figure 1.1). All the fundamental particles are shown in Figure 1.1

The fundamental forces of nature can be explained by Quantum Field Theory (QFT),

[1] Pauli's exclusion principle states that no two identical fermions can occupy the same quantum state simultaneously.

[2] Symmetry of a system under spatial inversion.

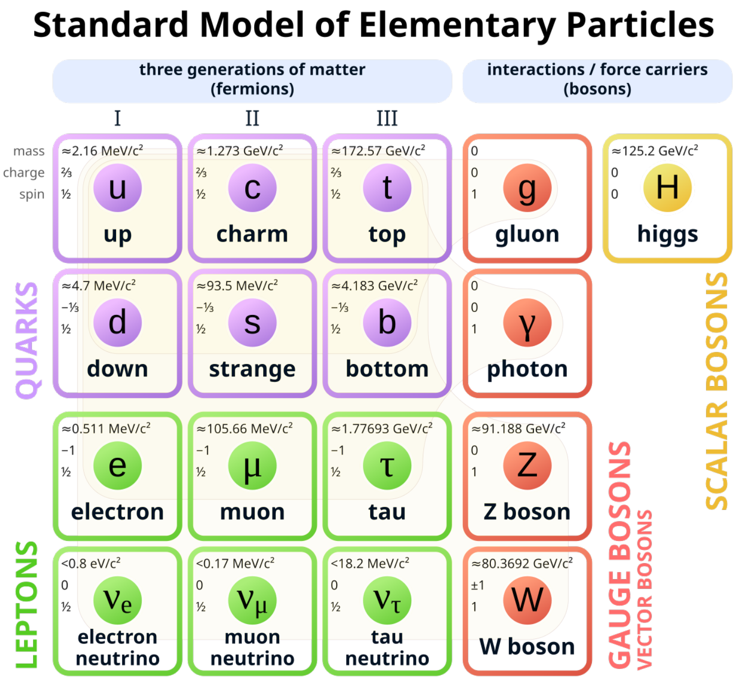


Figure 1.1: Subatomic particles in the SM

even for gravity at tree level. Gravitational interactions are governed by Einstein's General Theory of Relativity. There is no renormalizable quantum theory of gravity to help explain its effect at the subatomic level. Nevertheless, at tree level, it has a minimal effect at these scales. The particle interactions and their relative strengths are summarized in Table 1.1, and it can be seen that gravitational interactions are insignificant at the particle scale. Electromagnetic forces mediated by the exchange of photons are described by Quantum Electrodynamics (QED). Weak forces are mediated via intermediate vector bosons, $W^{\pm}$ and Z, and they affect both quarks and leptons. According to the Glashow-Weinberg-Salam (GWS) theory, weak and electromagnetic interactions can be considered as different manifestations of a single Electroweak Force. Quantum Chromodynamics (QCD) is the theory associated with strong forces, and these are mediated by gluons. The construction of the Standard Model is governed by the principles of symmetry. It is a local gauge quantum field theory based on the internal symmetries of the unitary product group $SU(3) \otimes SU(2) \otimes U(1)$. A group is a mathematical concept that is often represented by matrices. It is a set of transformations that leaves the laws of physics unchanged. According to Noether's theorem, every symmetry leads to a conservation law (See Table 1.2).

Interaction	**Relative Strength**
Strong	1
Electromagnetic	$\sim 10^{-2}$
Weak	$\sim 10^{-13}$
Gravitational	$\sim 10^{-39}$

Table 1.1: Particle interactions and their relative strengths

Symmetry	**Conservation law**
Translation in time	Energy
Translation in space	Momentum
Rotation	Angular momentum
Gauge transformation	Charge

Table 1.2: Symmetries and associated conservation laws.

Quantum Field Theory

QFT [3] combines quantum mechanics with the special theory of relativity. The special theory of relativity comes into play when we deal with velocities near the speed of light. In this fundamental theoretical framework, particles are considered as excitations of their corresponding quantum fields. For example, photons are the quanta of the electromagnetic field. A scalar field, defined at every point of space and time, is denoted as $\phi(\vec{x}, t)$. How the field evolves in time can be understood using the Lagrangian $\mathcal{L}$. The Euler-Lagrangian equation of motion for the scalar field ϕ is given as:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0 \tag{1.1}$$

where $\mu = 0, 1, 2, 3$. The Lagrangian is invariant under the Poincaré group. The Poincaré group is a set of symmetries that includes translations and the Lorentz group (rotations and boosts). The Klein-Gordon Equation, which is the relativistic wave equation for spin 0 particles, can be derived from Eq 1.1 and it is given as:

$$\left(\Box + m^2\right) \phi = 0 \tag{1.2}$$

where m is the mass of the particle and $\Box$ is the d'Alembertian operator, which represents the Laplacian in Minkowski space. For the spin $\frac{1}{2}$ particles, the relativistic wave equation is of the form:

$$\left(i\hbar\gamma^\mu \partial_\mu - mc\right) \psi = 0 \tag{1.3}$$

where ψ is a four-element column matrix, known as the Dirac spinor field, and γ^μ represents the Dirac gamma matrices.

Quantum Electrodynamics (QED)

QED [4] is a QFT which describes the electromagnetic interaction using a U(1) symmetry group. A U(1) group is a collection of 1×1 unitary matrices. Under the local U(1) symmetry, the spinor field ψ get transformed as:

$$\psi(x) \to e^{i\alpha(x)}\, \psi(x) \tag{1.4}$$

where $\alpha(x)$ can vary from point to point in space time. The QED Lagrangian is given as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}(F_{\mu\nu})^2 - Q\bar{\psi}\gamma^\mu \psi A_\mu \tag{1.5}$$

where A_μ is the electromagnetic vector potential, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, and Q is the charge. A_μ transforms as $A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu \alpha(x)$. It is a gauge field introduced to preserve the gauge invariance, and its excitation produces photons. The QED Lagrangian could be written in a simplified form as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}(F_{\mu\nu})^2 \tag{1.6}$$

where $D_\mu \equiv \partial_\mu + iQA_\mu(x)$ is the gauge covariant derivative.

Electroweak Theory

The unified theory of weak and electromagnetic interaction, also known as the GWS theory [5, 6], is based on the $SU(2)_L \times U(1)_Y$ local gauge symmetry, where L refers to left-handed fields and Y is the weak hypercharge [3] quantum number. The left-handed fermion fields transform as doublets under $SU(2)$ and participate in the weak interaction, while right-handed fermion fields transform as $SU(2)$ singlets. In particle physics, handedness describes the orientation of the spin relative to its direction of motion.

[3]It is related to the electric charge Q and the third component of weak isospin, I_3 via the equation, $Q = I_3 + \frac{1}{2}Y$

Consider a single family of quarks from the three-fold family structure of fermions:

$$\psi_1(x) = \begin{pmatrix} u \\ \\ d \end{pmatrix}_L, \qquad \psi_2(x) = u_R, \qquad \psi_3(x) = d_R \tag{1.7}$$

The above notation is also applicable to leptons, given by:

$$\psi_1(x) = \begin{pmatrix} \nu_e \\ \\ e^- \end{pmatrix}_L, \qquad \psi_2(x) = \nu_{eR}, \qquad \psi_3(x) = e^-_R \tag{1.8}$$

The free fermion Lagrangian is given by:

$$\mathcal{L}_0 \ = \ \sum_{j=1}^{3} i\,\bar{\psi}_j(x)\,\gamma^\mu\,\partial_\mu\,\psi_j(x) \tag{1.9}$$

where $\psi_j(x)$ are left-handed(L) or right-handed(R) fermion fields depending on j. $\mathcal{L}_0$ is invariant under global transformations. In order for the Lagrangian to be also invariant under $SU(2)_L \times U(1)_Y$ gauge transformation, four gauge fields are introduced, W^1, W^2, W^3, and B. Thus the Electroweak Lagrangian is given by:

$$\mathcal{L}_{\mathcal{EWK}} \ = \ \underbrace{\sum_{j=1}^{3} i\,\bar{\psi}_j(x)\,\gamma^\mu\,D_\mu\,\psi_j(x)}_{\mathcal{L}_f} \ - \ \underbrace{\frac{1}{4}B_{\mu\nu}B^{\mu\nu} \ + \ \frac{1}{4}W^i_{\mu\nu}W^{i\mu\nu}}_{\mathcal{L}_{\rm kin}} \tag{1.10}$$

where $\mathcal{L}_f$ is the fermion term and $\mathcal{L}_{\rm kin}$ is the gauge field kinetic term. The covariant derivative D_μ is given by $D_\mu = \partial_\mu + ig\,\frac{\sigma^i}{2}W^i_\mu + ig'\,\frac{Y}{2}B_\mu$. Here, in the second term $ig\frac{\sigma^i}{2}W^i_\mu$, W^i_μ describes the three gauge boson fields associated with $SU(2)$, σ^i are the Pauli matrices, $\frac{\sigma^i}{2}$, the weak isospin operators (T^i) and g, the coupling constant of $SU(2)$. Similarly, in the third term, g' is the coupling constant for $U(1)$, Y is the weak hypercharge quantum number, and B_μ is the gauge boson of $U(1)$. The Lagrangian of the electroweak theory (equation 1.10) only contains massless fields. The gauge symmetry does not allow writing explicit mass terms for the gauge bosons. For fermions, adding a mass term would link left and right-handed fields, which is not possible since they have different transformation properties and hence would break the gauge symmetry [7].

In order to generate massive fermions and bosons, a complex scalar doublet known as the Higgs field is introduced. The $SU(2)_L$ complex scalar field with a charged and a neutral component is given by:

$$\Phi(x) = \begin{pmatrix} \phi^{(+)}(x) \\ \phi^{(0)}(x) \end{pmatrix} \tag{1.11}$$

The Higgs Lagrangian, also known as the gauged scalar Lagrangian, is invariant under $SU(2)_L \times U(1)_Y$ transformations and is expressed as:

$$\mathcal{L}_S \;=\; (D_\mu\phi)^\dagger(D^\mu\phi) \;-\; \mu^2\,\phi^\dagger\phi \;-\; h\,(\phi^\dagger\phi)^2, \qquad (h > 0,\; \mu^2 < 0) \tag{1.12}$$

with a potential term, $V(\phi) = \mu^2\,\phi^\dagger\phi \,+ h\,(\phi^\dagger\phi)^2$. It's sometimes referred to as the Mexican hat potential shown in Figure 1.2, because of its shape. For $\mu^2 < 0$, the Higgs field ϕ develops a vacuum expectation value, $\frac{v}{\sqrt{2}} = \frac{|\mu|}{h}$, where $v = 246.66$ GeV, breaking the electroweak gauge symmetry [9]. Since electric charge is conserved, only the neutral component, ϕ^0 of the scalar doublet can acquire the vacuum expectation value. There are numerous degenerate states

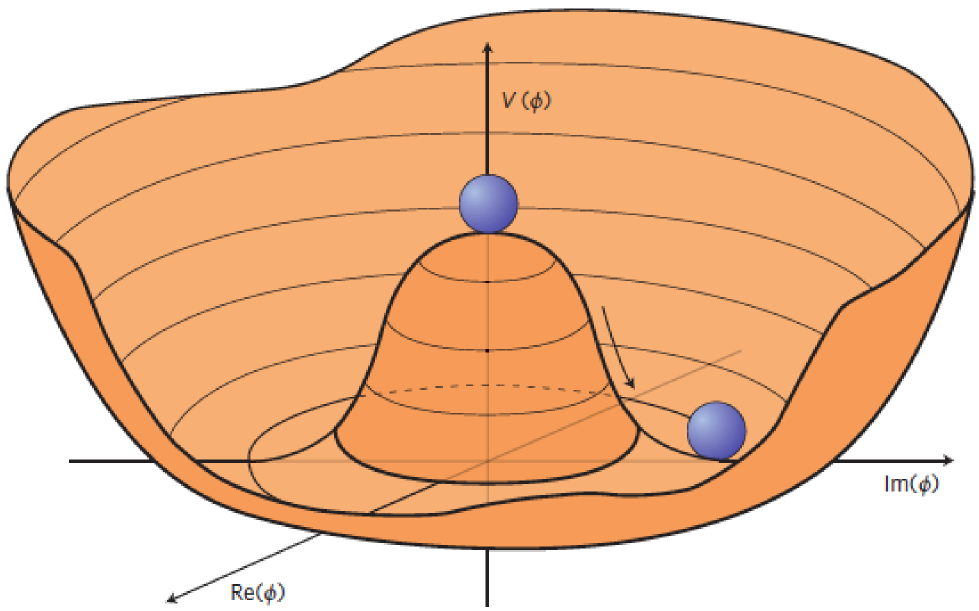


Figure 1.2: An illustration of the Higgs potential, where $Re(\phi)$ and $Im(\phi)$ are the real and imaginary part of the field [8].

with minimum energy, satisfying:

$$\langle 0|\phi^{(0)}|0\rangle = \sqrt{\frac{-\mu^2}{2h}} \equiv \frac{v}{\sqrt{2}} \tag{1.13}$$

Once a particular ground state is chosen, the $SU(2)_L \times U(1)_Y$ gets spontaneously broken to the electromagnetic subgroup $U(1)_{QED}$, after which ϕ becomes:

$$\phi(x) = \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.14}$$

with the real field $H(x)$, which appears as the Higgs Boson. Substituting the equation 1.14 into the kinematic part of the equation 1.12 results in a quadratic term for the $SU(2)$ gauge bosons due to the vacuum expectation value of the neutral scalar given by:

$$(D_\mu\phi)^\dagger(D^\mu\phi) \;\rightarrow\; \frac{1}{2}\,\partial_\mu H\,\partial^\mu H \;+\; (v+H)^2\left(\frac{g^2}{4}\,W_\mu^\dagger W^\mu \;+\; \frac{g^2}{8\cos^2\theta_W}\,Z_\mu Z^\mu\right) \tag{1.15}$$

These bosons are the $W^\pm$ and the Z, and they acquire mass (equation 1.16). θ_W is the weak mixing angle, also known as the Weinberg angle:

$$M_Z\cos\theta_W \;=\; M_W \;=\; \tfrac{1}{2}\,g\,v \tag{1.16}$$

The gauge boson associated with $U(1)$, γ, is massless since $U(1)_{QED}$ is an unbroken symmetry. Therefore, the mechanism of spontaneous symmetry breaking generating mass for gauge bosons is known as the Higgs mechanism.

As mentioned in the previous discussions, a fermionic mass term, $L_m = -m\bar{\psi}\psi = -m(\bar{\psi_L}\psi_R + \bar{\psi_R}\psi_L)$ is not allowed since it breaks the gauge symmetry. But with the introduction of the additional scalar doublet, fermion masses can be generated using the spontaneous symmetry-breaking mechanism. Through the Yukawa interaction, the coupling between the Higgs field and fermion fields can be described. The Yukawa Lagrangian is given by:

$$\mathcal{L}_Y \;=\; y_f\left[\bar{\psi}_R\,(\phi^\dagger\psi_L) \;+\; (\bar{\psi}_L\phi)\,\psi_R\right] \tag{1.17}$$

where $y_f = \sqrt{2}\frac{m_f}{v}$ is the Yukawa coupling constant and m_f is the mass of the fermion.

Quantum Chromodynamics (QCD)

The theory of strong interaction is called Quantum Chromodynamics [10], which describes the interaction of quarks mediated by gluons. QCD is based on the $SU(3)$ gauge symmetry group. Quarks carry a quantum number called color or color charge and can be red, green, or blue. The QCD Lagrangian is given by:

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} A^C_\mu - m_q \delta_{ab} \right) \psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu}, \tag{1.18}$$

where $\psi_{q,a}$ are the quark field spinors for a quark flavor q and mass m_q. a is the color index that runs from 1 to 3. The A^C_μ corresponds to the gluon fields, and C runs from 1 to 8, depicting that there are eight gluon fields. The t^C_{ab} are the eight 3×3 matrices that are the generators of the $SU(3)$ color group, g_s is the QCD coupling constant, and $F^A_{\mu\nu}$ is the field tensor given by:

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - g_s f^{ABC} A^B_\mu A^C_\nu\,, \tag{1.19}$$

where f^{ABC} are called the structure constants of $SU(3)$ and the term $g_s f^{ABC} A^B_\mu A^C_\nu$ depicts gluon self interactions. Thus, quarks and gluons are not found in isolation, and hadrons are color-neutral combinations of quarks and anti-quarks held together by gluons.

1.2 Parton Distribution Function

Protons are made up of two up quarks and one down quark, bound together by gluons, and are surrounded by a 'sea' of quark-antiquark pairs. These quarks, antiquarks, and gluons within hadrons are called partons. The momentum distribution functions of the partons within the proton are called Parton Distribution Functions (PDFs). They describe the probability density of finding a parton with a given momentum fraction x at a squared energy scale Q^2.

Asymptotic freedom [11] is a central feature in QCD, which describes the interaction between partons. According to this property, the interaction between partons becomes asymptotically weaker as the energy scale increases and the length scale decreases. QCD provides quantitative predictions on the rate of change of parton distributions with Q^2. Perturbative calculations can be applied at very high energies, where the coupling constant, $\alpha_s(Q^2) << 1$. These calculations can be made at different levels of approximation relative to different powers of $\alpha_s(Q^2)$, referred to as Leading-Order (LO), Next-to-Leading-Order (NLO), and Next-to-Next–Leading-Order (NNLO). Even though these calculations provide Q^2 dependence, the x dependence of the parton distribution has to be determined from the experimental data. QCD factorization theorems are used to relate parton distributions to the observable cross sections given by:

$$\sigma(x, Q^2) \;\approx\; \sum_a C_a \otimes f_{a/A}(x, Q^2) \;+\; \text{remainder} \tag{1.20}$$

where C_a is the part where perturbative calculation can be done, $f_{a/A}$ is the parton distribution of a in a hadron of type A and it is summed over all type of partons, a. The first term is called the leading twist contribution, and the remainder is called the higher twist correction. An example of the parton distribution function is shown in Figure 1.3 for $Q = 2$ GeV and $Q = 100$ GeV [12]. The example shown is based on the four-flavor scheme. In the four-flavor scheme, only u, c, s, and d are considered within the PDFs, and b is treated as massive, while in the five-flavor scheme, the bottom quark is included in the PDFs. At low energy scale ($Q = 2$ GeV) and large momentum fraction ($x > 0.1$), the up quarks dominate compared to d, and at low x values, there are fewer strange quarks than up and down quarks, and the charm density is negligible. As the energy scale increases, as seen in the $Q = 100$ GeV PDF, sea quarks and gluons appear in abundance, which would help in the production of t and c, critical for the analysis presented.

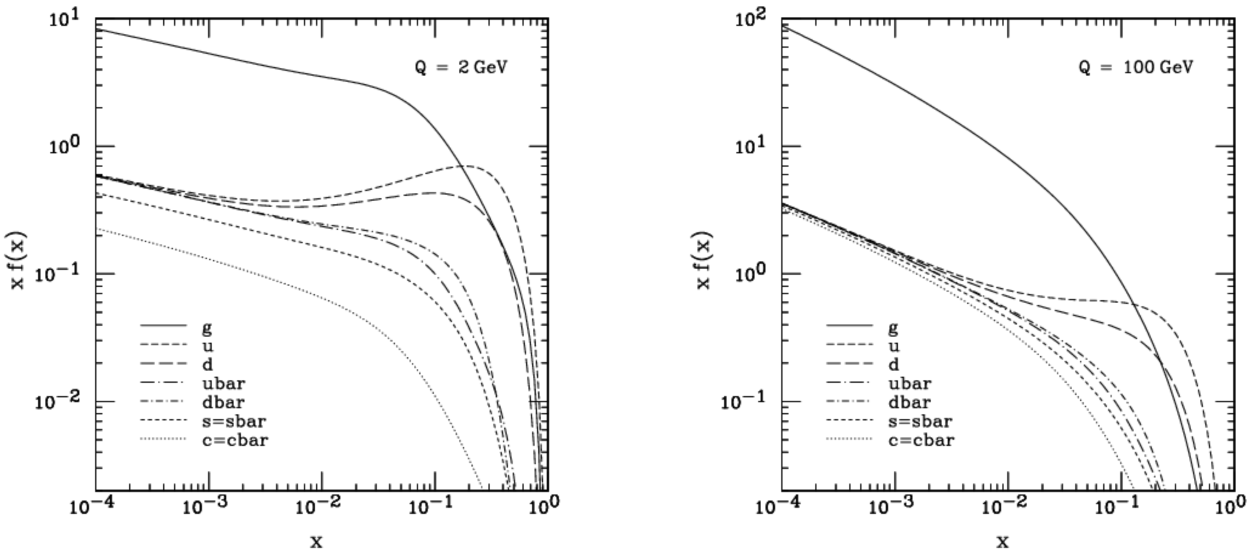


Figure 1.3: Overview of the CTEQ6M parton distribution at $Q = 2$ GeV and $Q = 100$ GeV

1.3 Production of single top with charm

The production of a single top quark in association with a charm quark, $pp \to t\bar{c}$, is a standard model t-channel process where the weak interaction is mediated by the exchange of a virtual W boson. The Feynman diagram for the single top + charm process is shown in Figure 1.4.

This interaction enables quarks to couple with one another, and the Cabibbo-Kobayashi-Maskawa (CKM) matrix determines the strength of these couplings. The CKM matrix is

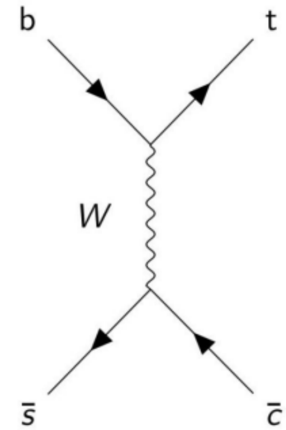


Figure 1.4: Feynman diagram depicting single top + charm process.

3×3 unitary matrix given by:

$$V_{\text{CKM}} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \tag{1.21}$$

The strength of the coupling between the quarks involved is described by the matrix elements V_{cs} and V_{cd}. The value of $V_{cs} = 0.973$ gives $|V_{cs}|^2 = 0.947$, resulting in a significant contribution to the process, whereas the value of $V_{dc} = 0.225$ and $|V_{dc}|^2 = 0.051$. These values imply a roughly 5% asymmetry between these contributions. Since the d quark is a valence quark, its contribution is modified by the structure of the hadronic initial state, leading to an approximate 20% higher production rate for $\bar{t}$ compared to the standard t quark production.

Top quarks are primarily produced in pairs via the strong interaction. In contrast, single top production occurs through electroweak interactions, and while the cross-section for single top production is smaller than that of top pair production, it is associated with a much larger relative background. This makes single top production more challenging to identify. However, with larger data samples and increased luminosity, more single top states could be probed.

The analysis presented in the thesis reports the first observation of single-top production with an associated charm quark as predicted by the SM. Since the process occurs via electroweak interactions, a precision measurement provides a test of the electroweak sector of the SM and of the CKM matrix elements. Achieving sensitivity to this measurement can also lead to the search for beyond the SM (BSM) interactions in this channel.

Chapter 2

Experimental Apparatus

The ATLAS detector (A Toroidal LHC Apparatus) is a general-purpose detector at the Large Hadron Collider (LHC) at CERN, the European Organization for Nuclear Research, located near Geneva, Switzerland. The ATLAS experiment is one of the largest scientific collaborations ever attempted, and it involves over 6000 scientists, engineers, and students from 259 institutions across 45 countries (2025) [13]. It allows scientists to probe fundamental particles produced in the high-energy collisions at the LHC and test the predictions of the Standard Model and search for physics beyond the Standard Model.

2.1 The Large Hadron Collider

The LHC [14], built to collide protons or heavy ions at very high energies, is the world's largest particle accelerator. It is a powerful two-ring superconducting hadron accelerator and collider situated inside a 26.7 km tunnel buried about 175 meters underground. Since the collisions are done between two beams of the same type of particle, two separate rings with beams rotating in opposite directions are used. Out of the possible eight interaction points, four are active and occupied by ATLAS at Point 1, CMS (Compact Muon Solenoid) at Point 5, ALICE (A Large Ion Collider Experiment) at Point 2, and LHCb (LHC-beauty) at Point 8 [15].

Run 1 began in 2009 and was concluded in 2012 with a center of mass energy of 8 TeV.

The Run 2 data recording period was from 2015 to 2018, with an increase in center of mass energy to 13 TeV. The ongoing data-taking is Run 3, which started in 2022 with a center-of-mass energy of 13.6 TeV and is expected to run until 2026. The analysis results presented in this thesis are based on Run 2.

The schematic diagram of the CERN accelerator complex, of which the LHC is a part, is

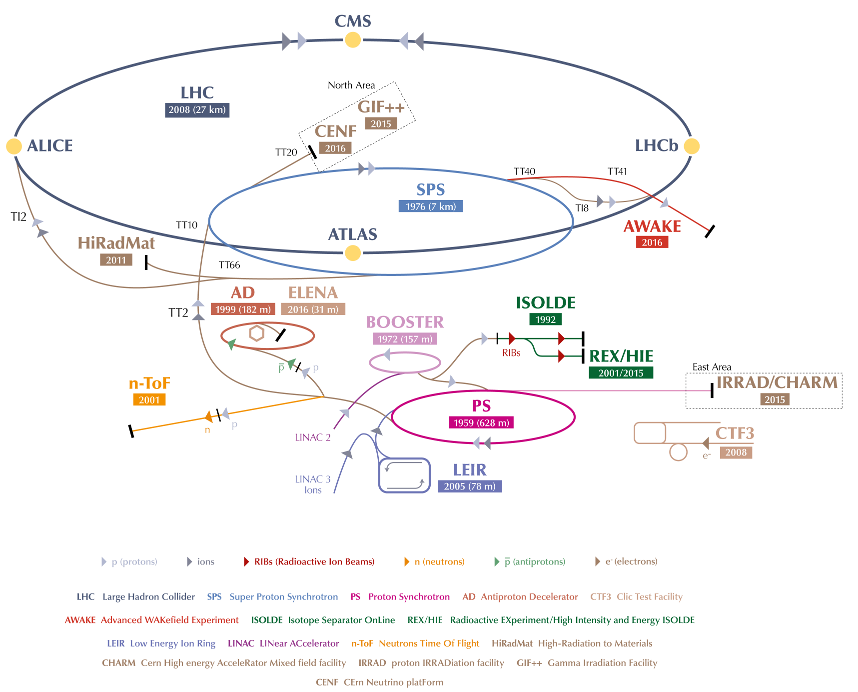


Figure 2.1: Schematic diagram of the CERN Accelerator Complex [16]

illustrated in Figure 2.1. The protons in the LHC originate from a hydrogen gas source. The hydrogen is ionized to H^{-1} ions after which it passes through a LINAC (Linear Accelerator) where it is accelerated to 160 MeV. Then the electrons are stripped to obtain H^{+1} and enter the Proton Synchrotron Booster (PSB). Before moving to the main ring it goes through the Super Proton Synchrotron (SPS) where the protons are accelerated to 450 GeV. In the LHC, it reaches the maximum energy.

The proton beams move in opposite directions through the side-by-side beam pipes in the LHC tunnel and collide at the four interaction points, where they are studied. The LHC uses 1232 dipole magnets to keep the beams moving in circular paths, and 392 quadrupole magnets help focus the beams, increasing the chance of collisions. There are as many as 1600 superconducting magnets in total, and in order to maintain superconductivity, they are cooled to 1.9 Kelvin using liquid helium. At full power, each proton beam is accelerated to the total collision energy.

The proton beams are not a continuous stream but rather 'proton bunches' and are grouped into 2808 bunches per beam. These bunches collide every 25 nanoseconds, resulting in a collision rate of 40 MHz. The number of collisions occurring per unit time is given by:

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{2.1}$$

where $\mathcal{L}$ is the Instantaneous Luminosity measured in number of events per unit area per unit time and σ is the cross section, which is the probability of a given interaction. The total number of events N over a time interval is obtained by:

$$N = \int L(t) \cdot \sigma \, dt = \sigma \cdot \int L(t) \, dt = \sigma \cdot \mathcal{L}_{\text{int}} \tag{2.2}$$

where $\mathcal{L}_{\text{int}}$ is the integrated luminosity. Figure 2.2 shows the collision data collected by the ATLAS detector during the LHC Run 2. The total integrated luminosity recorded by ATLAS suitable for physics analysis is 139 fb^{-1}. In the ATLAS detector, on average, each bunch crossing contains approximately 34 pp collisions. Among these, only one is typically for physics analysis, and the rest of the interactions contribute to what is known as pileup [17].

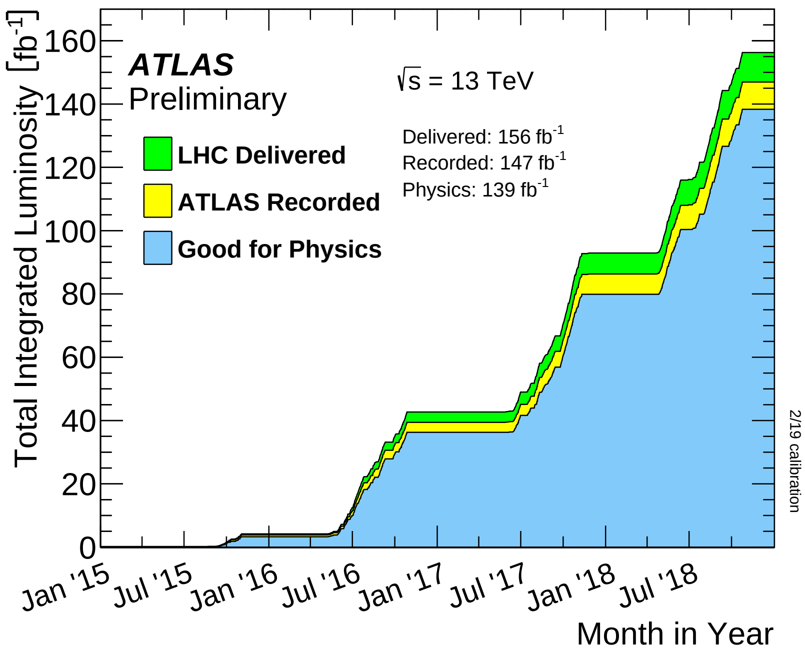


Figure 2.2: Integrated luminosity delivered by the LHC and recorded by the ATLAS detector during Run2 (2015–2018) [18]

2.2 The ATLAS Detector

The ATLAS detector [19] at the LHC is designed to detect decay products from the energetic proton-proton and heavy ion collisions. It is a massive cylindrical-shaped detector, 25 meters tall, 44 meters long, and weighs approximately 7000 tonnes and is mostly symmetric in both directions along the beam axis with the interaction point at the center. The detector is built with several layers, and each layer identifies particles based on their properties, such as whether they are charged or neutral, or whether they are hadrons or leptons in origin, and how they interact with matter. The main layers or sub-detectors in ATLAS are :

- The Inner Detector: It tracks charged particles and determines their charge and momentum, and is used to reconstruct the interaction vertex. The sub-components are the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

- The Calorimeter System: The purpose of this sub-detector is to measure the energy of the particles. There are two types of calorimeter systems, the Electromagnetic Calorimeter (ECAL), which measures the energy of electrons and photons, and the hadronic calorimeter (HCAL), which measures the energy of hadrons.

- The Muon Spectrometer: This sub-detector detects muons and, along with the ID, provides a secondary measurement, thereby improving the resolution.

The schematic diagram of the ATLAS detector with the sub-detectors is shown in Figure 2.3.

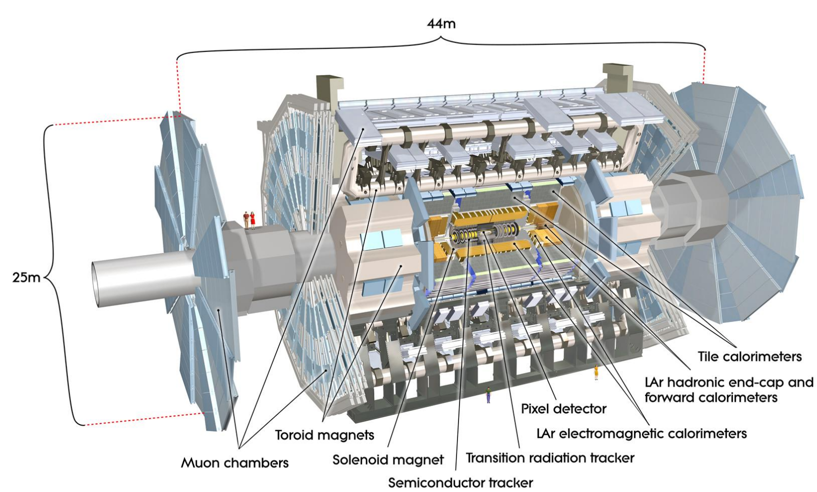


Figure 2.3: The ATLAS detector with its sub detectors [20]

2.2.1 Detector Coordinate Framework

The ATLAS detector uses a standard coordinate system (as shown in Figure 2.4) to describe its geometry and the properties of the particles from the collision [21]. The interaction point is taken as the origin. The z-axis is along the beam line and the $x-y$ plane is perpendicular to the beam direction, the positive x-axis points from the origin towards the center of the LHC and the positive y-axis points upwards. The positive z side is called side A, and the negative z side is called side C. The azimuthal angle (ϕ) is measured from the positive x-axis

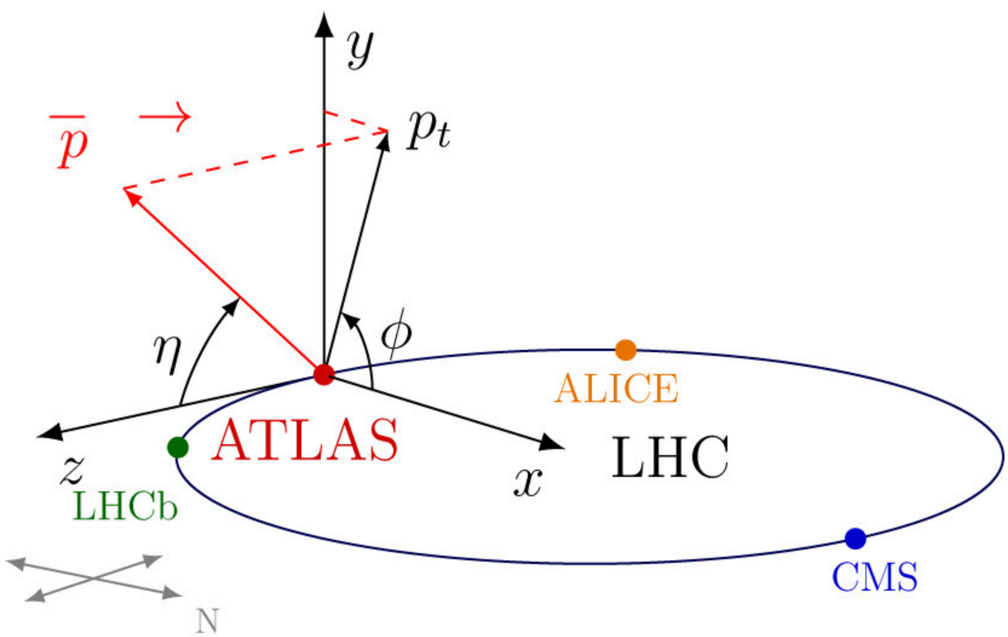


Figure 2.4: ATLAS detector coordinate system [22]

around the z-axis. The polar angle (θ) is the angle from the positive z-axis. It is used to calculate pseudorapidity, which is used to describe how far a particle is emitted from the beam axis:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.3}$$

It is called 'pseudo' for massless particles. For massive particles, rapidity is used:

$$y = \frac{1}{2} \ln \left(\frac{E - p_z}{E + p_z} \right) \tag{2.4}$$

where E is the energy and p_z is the momentum along the z axis. Pseudorapidity is used instead of rapidity for particles produced at very high energies, where $E >> m$. In this limit, η becomes approximately equal to y. We use η or y instead of θ since particles are relativistic and differences in rapidity are invariant under Lorentz transformations, unlike θ. The angular separation between two particles is given by:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.5}$$

The quantities defined in the x-y plane are :

- Transverse momentum (p_T): This is the momentum of the particle perpendicular to the beamline,

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.6}$$

- Transverse energy (E_T): This is the energy component of particles transverse to the beam direction. It is equal to p_T for massless particles.

- Missing transverse energy (E_T^{miss}): The total p_T of the colliding particles is zero since the collision is along the beam axis. According to the conservation of momentum, the total p_T before collision is equal to the total p_T after collision. If some particle (such as a neutrino) escapes the detector without detection, we measure an imbalance in the momentum calculation. This imbalance is referred to as E_T^{miss}.

2.2.2 Inner Detector

The Inner Detector (ID) [23] consists of mainly three parts; the pixel detector, the semi-conductor tracker (SCT), and the transition radiation tracker (TRT). A diagram of the ID with all the sub-components is shown in Figure 2.5. The ID surrounds the beam pipe and is enclosed by a thin superconducting solenoid with a length of 5.3 m and a diameter of 2.4 m. The solenoid provides a magnetic field of 2T pointing along the beam axis. The Lorentz force[4] experienced by the charged particles traversing through the ID, causes their path to curve and allows a measurement of their momentum, with curvature being proportional to the momentum. It measures particles within a range of $|\eta| < 2.5$. The transverse momentum

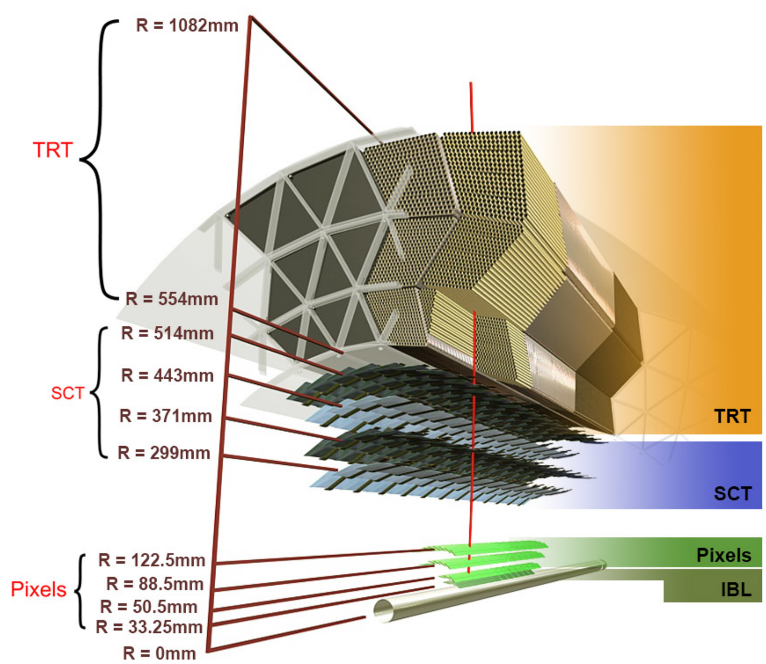


Figure 2.5: Inner Detector [24]

[4] $F = qvB = \frac{mv^2}{R} \Rightarrow p_T = qBR$

resolution of the ID is given by :

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% \times p_T \text{ [GeV]} \oplus 1\% \tag{2.7}$$

The ID also helps to reconstruct primary vertices, where the initial collision between particles occurs, and secondary vertices, where the decay and interaction of any unstable particles occur. The vertex resolution can be studied using the longitudinal impact parameter ($Z_0 \sin\theta$) and the transverse impact parameter[5] (d_0). The ID provides a d_0 resolution of 10 μm for high p_T particles in the central η region [25]. The rest of the section describes the technology of the sub-components of the ID that helps in the high precision measurement.

Pixel Detector

The Pixel Detector [26] is a high-granularity silicon detector and is located at the innermost part of the ID. Initially, it consisted of 3 pixel layers, but with an increase in luminosity and hence pileup, a fourth inner layer was added at approximately 3.3 cm from the beam line. This layer is called Insertable B-Layer (IBL) [27] and is designed to increase the measurement precision. The IBL consists of 280 silicon pixel modules, and the rest of the 1744 silicon pixel modules are arranged in three barrel layers and two end caps with 3 disks each. Each of these modules contain 47232 tiny pixels with a pixel pitch[6] of 50 μ wide ($R\phi$ plane) and 400 μ long (z direction) for the three layers and 50 μ wide ($R\phi$ plane) and 250 μ long (z direction) for the IBL. A reverse bias voltage when applied across these silicon layers creates a charge depletion region, and the charged particles, when passing through it, ionize the

[5] Impact parameter is the shortest distance between a track and the primary vertex

[6] Pitch is the center to center distance between two pixels or strips in a detector.

silicon and create an electron-hole pair. The charges drift towards the readout channel and are collected as a signal. The detector has about 80.4 million readout channels.

Semiconductor tracker (SCT)

The semiconductor tracker, also known as the silicon strip tracker, consists of 4088 silicon strip sensor modules [28] arranged in four barrel layers and two end caps with 9 disks each. The strips in the barrel are parallel to the beam axis and have a constant pitch of 80 μm. The module has two sensors on one side, arranged side by side and wire-bonded together to form 12 cm-long strips. While the other side has another set of identical sensors glued to the back of the first pair, tilted at an angle of 40 mrad relative to each other. This helps to reconstruct the position of the particles. The SCT has approximately 6.3 million readout channels.

Transition Radiation Tracker (TRT)

The transition radiation tracker [29] is the outermost layer of the inner detector, which utilizes straw drift tube technology for particle detection. It comprises approximately 300000 straw tubes that are 4 mm in diameter. The tubes are made of a material called Kapton, which is efficient in withstanding heat and radiation. The tube encloses a thin wire made of tungsten. The straws are arranged parallel to the beam line in the barrel region and perpendicular to the beam line in the end cap region. The gas mixture used in the TRT is that of Xenon and Argon. When charged particles pass through the TRT straws, they ionize the gas mixture, creating ionization clusters. The straw wall has a negative voltage applied to it, which creates an electric field that pulls the electrons towards the thin wire, producing a signal.

2.2.3 Calorimeters

The ATLAS calorimetry system [30], consisting of electromagnetic and hadronic calorimeters, is used to measure energies of particles and has full ϕ symmetry and coverage around the beam axis. They provide energy measurement for particles within $|\eta| < 4.9$. These are sampling calorimeters, and they contain layers of active and absorber materials. The absorbers are dense materials, and showers of secondary particles are created when high-energy particles pass through them. These particles then deposit their energy in the sensitive active layer, which is detected in the form of light or charge. Figure 2.6 shows the schematic diagram of the ATLAS calorimeter detector system.

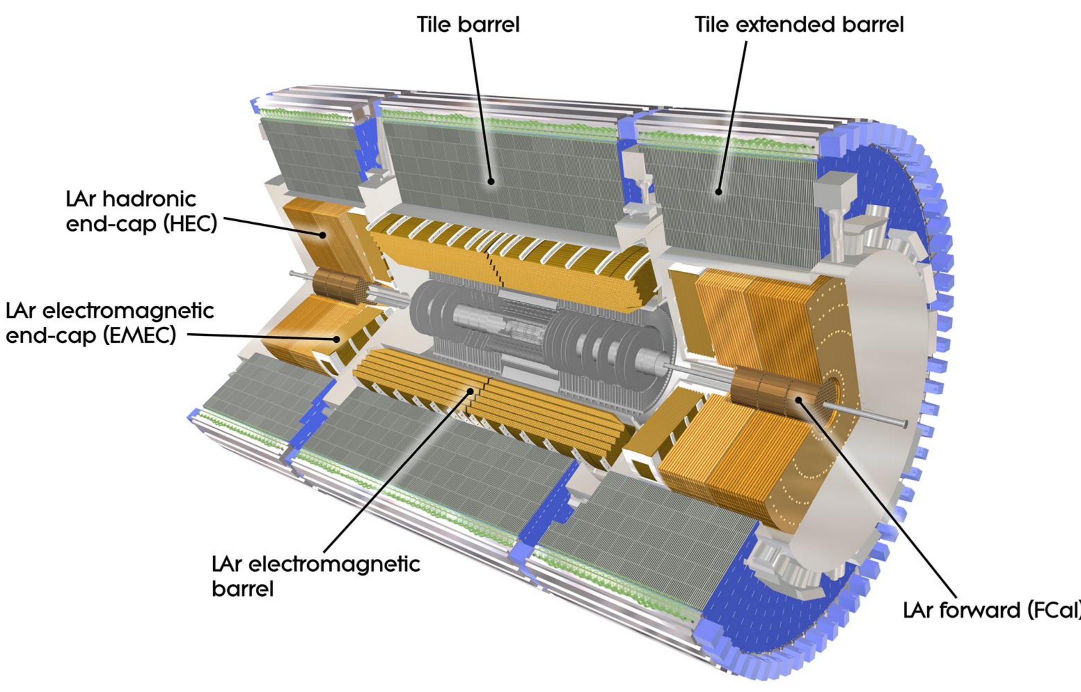


Figure 2.6: ATLAS Calorimeter Detector [31]

The Electromagnetic calorimeter (ECAL) helps in the measurement of the energy of electrons and photons. They are located at the barrel section ($|\eta| < 1.475$) and two end cap sections ($1.375 < |\eta| < 3.2$) and are made of lead, acting as the absorber layer, and liquid argon, acting as the active layer. Electrons and photons interact with the absorber layer via bremsstrahlung radiation[7] and pair production[8] respectively. These electromagnetic showers can be described using a parameter known as the radiation length (X_0). In the case of electrons, radiation length represents the rate at which electrons lose energy by bremsstrahlung. The average distance x that an electron needs to travel in a material to reduce its energy to $\frac{1}{e}$ of its original energy E_0 can be depicted as [32]:

$$\langle E(x) \rangle = E_0 e^{-x/X_0} \tag{2.8}$$

For photons, the radiation length is described in terms of Intensity. If a photon beam is traveling through a material with initial intensity I_0, it is reduced to $\frac{1}{e}$ of its original intensity after traveling a distance of $x = \frac{9}{7} X_0$:

$$\langle I(x) \rangle = I_0 e^{-\frac{7}{9}\frac{x}{X_0}} \tag{2.9}$$

The LAr (liquid argon) is ionized by the showering particles. The charges drift under the high voltage and the current is recorded in the readout channels.

The Hadronic calorimeter (HCAL) is used for hadron energy measurement. It is placed outside the ECAL and has sections at the barrel region, also known as the tile calorimeter,

[7] $e^- \to e^- + \gamma$
[8] $\gamma \to e^- + e^+$

to cover $0.8 < |\eta| < 1.7$, end cap region for $|\eta|$ between 1.7 and 3.2, and forward region or FCal for $|\eta| > 3.2$. FCal probes energetic and forward-directed processes. Particles, when passing through HCAL, interact with the absorber material mostly via strong interactions. The secondary hadrons are described in terms of nuclear interaction length or λ. The barrel region of the HCAL has scintillating tiles as the active medium, and it emits light when traversed by particles. The absorber material is made of iron.

The energy resolution of the calorimetry system can be expressed in a general way as:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.10}$$

where the first term a is called the stochastic term and is related to shower fluctuations. The second term b is called the noise term, and c is the constant. $\oplus$ indicates quadratic sum. For ECAL, the energy resolution was measured to be:

$$\frac{\sigma_E}{E} \sim \frac{10\%}{\sqrt{E}} \oplus 0.7\% \tag{2.11}$$

while for HCAL, it was estimated to be:

$$\frac{\sigma_E}{E} \sim \frac{50\%}{\sqrt{E}} \oplus 5\% \tag{2.12}$$

2.2.4 Muon Spectrometer

The Muon Spectrometer [33] is the outermost and the last layer that surrounds the rest of the detector system. It has three regions, a barrel region covering $|\eta| < 1.2$ and two endcap regions for $1 < |\eta| < 2.7$ as shown in Figure 2.7. As the name suggests, the main goal of the muon spectrometer is to track muons. The magnetic field generated by the large superconducting air-core toroid magnets bends the trajectory of the muons while traversing, which helps in the calculation of momentum. The muon system consists of precision tracking chambers as well as trigger chambers. In the case of tracking chambers, Monitored Drift Tubes (MDTs) are used for most of the η range to measure the position of the muons accurately. For forward regions, $|\eta| < 1.2$, Cathode Strip Chambers (CSCs) are used. CSCs have high granularity and provide high momentum resolution. The trigger chambers consist of Resistive Plate Chambers (RPCs) in the barrel region and Thin Gap Chambers (TGCs) in the end-cap region. They specifically help in selecting events containing muons through bunch cross identification and applying a transverse momentum threshold.

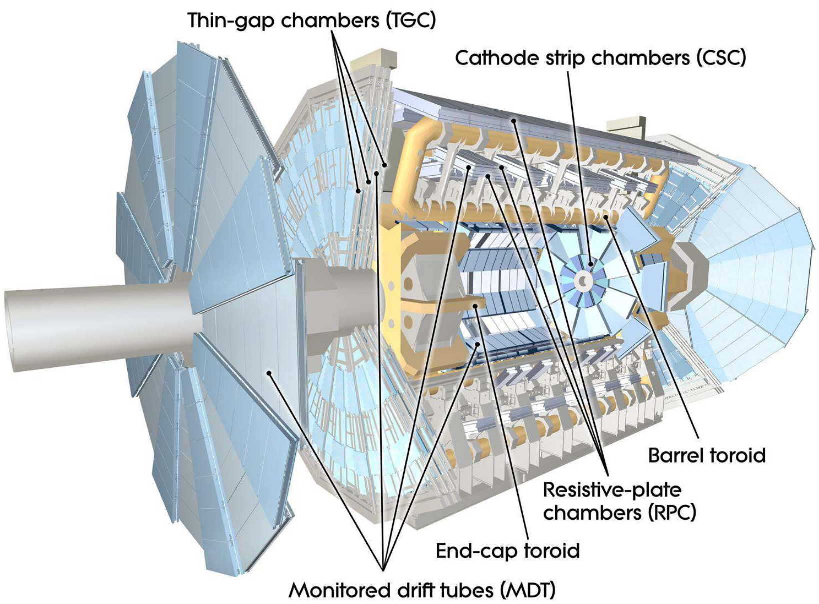


Figure 2.7: ATLAS Muon System [34]

2.2.5 Trigger System

The ATLAS Trigger System [35] is an important part of the ATLAS detector designed to select only interesting events from the huge number of proton-proton collisions. Collisions are happening at the rate of 40 MHz in the LHC. The raw data rate could be approximately 1 petabyte per second, which is very difficult to store and analyze. The trigger system filters out uninteresting events and records interesting data for offline analysis. It consists of a hardware-based first-level trigger (L1) and a software-based high-level trigger (HLT).

The L1 trigger selects events based on the event level qualities, such as total energy in the calorimeter, number of particles above a certain threshold, as well as the kinematic relationship between the particles. The output rate of the L1 trigger is around 100 kHz from an input rate of 40 MHz with a decision time of 2.5 μs. The software-based HLT reduces the rate further to 1.2 kHz. It uses more precise CPU-intensive algorithms and makes event selection decisions within a few hundred milliseconds.

Chapter 3

Physics Object Reconstruction

This chapter provides an overview of the process involved in reconstructing physics objects from the raw data collected by the ATLAS detector following the proton-proton collisions in the LHC. The requirements and an outline of the selection criteria for the physics objects used in the analysis presented in this thesis are also provided. The physics objects considered are electrons ($e^{\pm}$), muons ($\mu^{\pm}$), jets, and missing transverse energy (E_T^{miss}). Tau leptons ($\tau^{\pm}$) are also taken into account; however, they decay either leptonically[9] or hadronically[10] before reaching the detector due to their short lifetime. Leptonic τ decays are considered in this analysis, while hadronic τ decays are not explicitly used.

3.1 Tracks and Vertex Reconstruction

Tracks are the charged particle trajectories. The ID of the ATLAS detector is involved in track reconstruction, as mentioned in Section 2.2.2, through several algorithmic processes. The one explained here is the Inside-Out Track Seeding and Finding. The clusters in the Pixel and SCT, as well as the drift circles in the TRT, formed from the particle hits, can be transformed into 3D space points. The set of three space points is used to form a seed. A seed can be Pixel-only, SCT-only, or Pixel + SCT. The seeds that pass the initial transverse momentum and impact parameter resolution cuts are kept and matched to a fourth

[9] $\tau \rightarrow l\nu_l\nu_\tau, l = e, \mu$

[10] $\tau \rightarrow hadrons$

space point to confirm their trajectory. A combinatorial Kalman filter [36] is then used to complete a possible trajectory. Additionally, the Bremsstrahlung recovery method is used if the seed fails the usual track finding and is located within a region of interest (ROI) in the calorimeter. The track recovery process is repeated by allowing a change in trajectory due to Bremsstrahlung radiation from electrons. Figure 3.1 gives a pictorial description of the above steps mentioned. An ambiguity solver algorithm is used to select track candidates based on track scores, and later a global χ^2 method is used to obtain a high-precision track candidate. A neural network clustering algorithm is used to separate pixel clusters assigned to multiple track candidates. Finally, the track is extended to the TRT to check if it matches the drift circle [37].

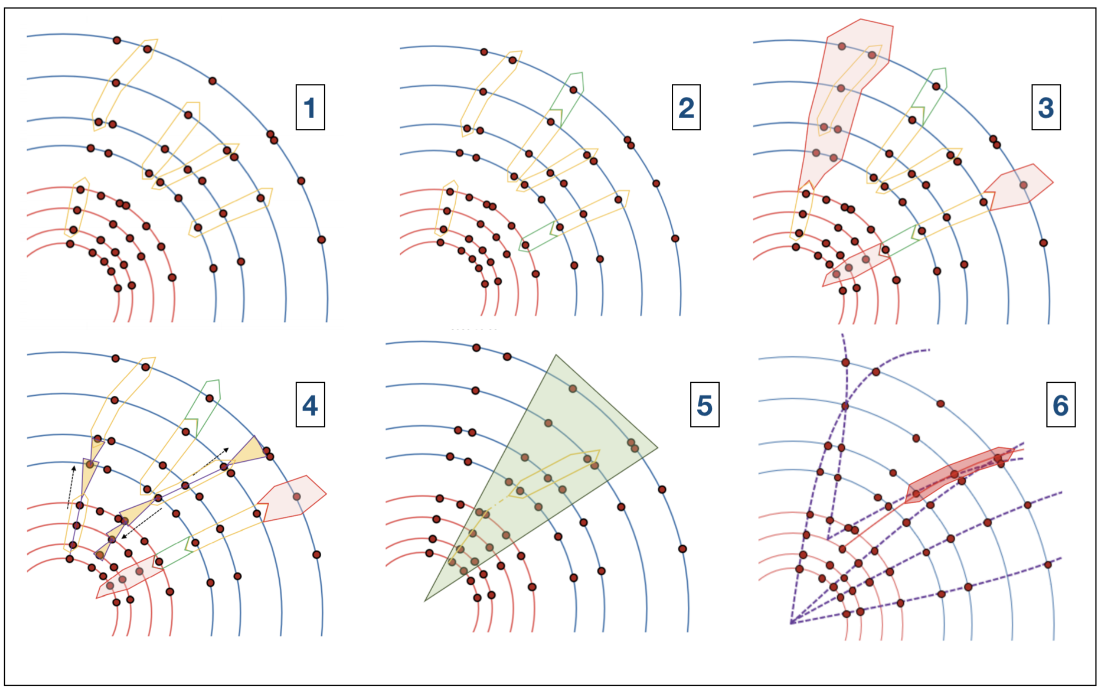


Figure 3.1: Inside Out method, 1) Seed formation, 2) Seed confirmation, 3) Search for clusters compatible with the seed, 4) combinatorial Kalman filter, 5) Brem recovery diagram 6) Completed track finding diagram [38]

The reconstructed tracks are used to determine the primary and secondary vertices. The vertex reconstruction uses an iterative algorithm known as the Adaptive Multi-Vertex Reconstruction, and it occurs in two steps: vertex finding and vertex fitting. An initial guess of a vertex position, known as the seed, is determined by the distribution of track impact parameters along the beam axis, using the density of tracks. The seed information, along with the reconstructed tracks, is used to fit the best vertex position. During each iteration, the less compatible tracks are given a lower weight, and the vertex position is recalculated. These less compatible tracks are then used to determine another vertex. This process is repeated until all the tracks are associated with a vertex or no vertex can be found for the remaining set of tracks. The vertex with the highest squared sum of track p_T is taken as the primary vertex and others as the pileup vertices [39]. After the primary vertex reconstruction is completed, transverse impact parameter (d_0), the distance between the particle's track and the primary vertex in the plane perpendicular to the beam axis, and longitudinal impact parameter (Z_0), the distance of a particle's track from the primary vertex along the beam axis, can be calculated. The secondary vertex reconstruction can be done using d_0. It is assigned a positive sign if the track crosses the path of the decaying hadron or τ lepton downstream of the primary vertex and is considered a candidate for a secondary track [40].

3.2 Topological Clusters

Particles deposit energy across multiple cells in the calorimeter system. These energy deposits or signals in the cells are collected together to form topological clusters. The cluster formation depends on the cell signal significance ς_{cell}^{EM}, which is the ratio of the cell signal to the average noise $\sigma^{EM}_{noise,cell}$ (includes electronic noise, pileup contributions, etc) in the cell:

$$\varsigma^{EM}_{cell} = \frac{E^{EM}_{cell}}{\sigma^{EM}_{noise,cell}} \tag{3.1}$$

A topo-cluster algorithm is used for the formation of clusters. It starts by selecting those cells with $\varsigma^{EM}_{cell} > S$, where S is the primary seed threshold, with a default value of S = 4, and $\varsigma^{EM}_{cell} > N$, where N is the growth control, with default value of N = 2. The direct neighboring cells should also satisfy $\varsigma^{EM}_{cell} > P$, where P is the principal cell filter, with default value of P = 0. S, N, and P are the three parameters in this algorithm that control the seeding, growth, and boundary features of topo-clusters. This process is repeated until all the adjacent cells that pass the P threshold are added to the cluster [41].

3.3 Electrons

For the electron reconstruction, information from the tracks in the inner detector, as well as the energy deposited in the electromagnetic calorimeter, is used. The electromagnetic topological clusters are matched to the tracks reconstructed in the inner detector. The procedure for track reconstruction and topo-cluster formation is explained in sections 3.1 and 3.2. The track candidates matched to the EM clusters are re-fitted with a Gaussian sum filter (GSF) [42] algorithm to account for energy loss due to bremsstrahlung.

The reconstructed electrons are either prompt or non-prompt. Prompt leptons are produced directly from the main interaction, while non-prompt leptons originate due to semileptonic decay of hadrons, photon conversion, as well as misidentification. Three working points are defined based on a likelihood (LH) discriminator to reduce the non-prompt electrons. The electron LH is based on the probability distribution functions (pdfs) P of prompt electrons, LS, and non-prompt electrons, LB:

$$L_S(x) = \prod_{i=1}^{n} P_{S,i}(x_i), \quad L_B(x) = \prod_{i=1}^{n} P_{B,i}(x_i) \tag{3.2}$$

where x_i is the value that is based on shower width, TRT information, track conditions, track cluster matching, etc. For each electron candidate, a discriminant d_L is defined, given by:

$$d_L = \frac{L_S}{L_S + L_B} \tag{3.3}$$

Then an inverse sigmoid function is applied to obtain:

$$d_L^{'} = -\tau^{-1} ln(d_L^{-1} - 1) \tag{3.4}$$

where τ is fixed to 15. Thus, the threshold values on $d_L^{'}$ are used to define the electron identification working points, which are LooseLH, MediumLH, and TightLH with increasing order of background rejection power.

Additionally, electron isolation requirements are also applied to remove the non-prompt electrons. The ATLAS recommended isolation working points are : HighPtCaloOnly, TightTrackOnly_VarRad, TightTrackOnly_FixedRad, Tight_VarRad, and Loose_VarRad. Two types of isolation variables are considered, calorimeter-based and track-based. The calorimeter isolation variable, $E_{T,raw}^{isol}$, is the sum of the transverse energy of the topological clusters with positive energy. The sum is calculated within a cone around the electron cluster center and later corrected for leakage, pileup etc to obtain the corrected calorimeter isolation. A track-based isolation variable, p_T^{coneXX} is the sum of the transverse momenta of all tracks within a cone size XX, where the typical value of XX is 20 or 30 and the tracks that belong to the electron candidate itself are excluded [43, 44].

The identification and isolation working points, as well as the p_T and η requirements for the top+charm analysis, are summarized in the Table 3.1. Due to the poor electron identification efficiency, electrons in the LAr crack region, $1.37 < |\eta| < 1.52$, are vetoed. Additional selections are also done on d_0 and Z_0. The d_0 significance, $\frac{d_0}{\sigma_{d_0}}$ is the ratio of the transverse impact parameter and its uncertainty. $|\Delta Z_0 sin(\theta)|$ measures the allowed maximum difference between the track's longitudinal impact parameter and the primary vertex, scaled by the sine of the angle θ.

	Electron	
	Loose	Tight
p_T	10 GeV	
$\lvert\eta\rvert$	< 2.5 & $\notin (1.37, 1.52)$	
$\lvert\Delta Z_0 \sin\theta\rvert$	< 0.5 mm	
$\lvert d_0\rvert/\sigma_{d_0}$	< 5	
Identification	LooseLH	TightLH
Isolation	None	Tight_VarRad

Table 3.1: Selection criteria for electrons.

3.4 Muons

Muon reconstruction [45, 46] requires information from the ID and the Muon Spectrometer, as well as the calorimeters. Depending on where the information is processed, muon candidates are divided into four categories:

- Combined (CB) Muon: The muon track is reconstructed in the ID and the muon spectrometer independently, and with the hits from both the sub-detectors, a combined track is formed with a global refit. Reconstruction can be done by either using the outside-in method, where reconstruction occurs first in the Muon Spectrometer and then extrapolated to match the ID, or the inside-out method, where ID tracks are extrapolated outward and matched to the tracks in the Muon Spectrometer. CB Muons are considered to have the most accurate measurement.

- Segment-tagged (ST) Muon: The muon candidate is identified as ST if the ID track extrapolated matches to at least a track segment in the MDT or CSC chambers. They have low p_T and reduced acceptance in the Muon Spectrometer.

- Calorimeter-tagged (CT) Muon: If the track in ID can be matched to an energy deposited in the calorimeter, resembling a minimum ionizing particle, the muon reconstructed can be classified as CT. These muons have the lowest purity of all.

- Extrapolated (ME) Muon: Muon reconstruction is only based on muon spectrometer tracks with a loose requirement of originating from the interaction point, and it cannot be matched to an ID track. Thus, muon reconstruction extends to the region outside of the ID, $2.5 < |\eta| < 2.7$.

After the reconstruction, the prompt muons are selected based on certain identification requirements. The selection working points for muon identification are Loose, Medium, Tight, and High p_T. The Loose identification criteria are used to increase the reconstruction efficiency as well as provide good quality muons. Medium selection is the default selection for ATLAS analysis, and it reduces the systematic uncertainties associated with muon reconstruction. With Tight selection, good-quality muons are obtained with lower efficiency. High p_T selection is considered to maximize momentum resolution above 100 GeV. Isolation criteria are also considered to reject non-prompt muons from heavy-flavor hadron decays. They can be track-based, calorimeter-based, or a combination of both. Muon selection criteria for the top + charm analysis are summarized in the Table 3.2.

	Muon	
	Loose	Tight
p_T	10 GeV	
$\|\eta\|$	< 2.5	
$\|\Delta Z_0 \sin\theta\|$	< 0.5 mm	
$\|d_0\|/\sigma_{d_0}$	< 3	
Identification	Medium	Medium
Isolation	None	PflowTight_VarRad

Table 3.2: Selection criteria for muons.

3.5 Jets

Jets, showers of collimated particles, are produced by the hadronization of quarks and gluons in the high-energy proton-proton collisions. The reconstruction of jets uses an approach called Particle flow [47, 48], which utilizes information from both the calorimeter and the ID. The Particle Flow (PFlow) algorithm removes the energy deposited by charged particles matched to the tracks in the ID. The energy deposited by the neutral objects is retained. The combination of the track and the topological cluster created by the PFlow algorithm is called a Particle Flow Object (PFO), and the jet reconstruction is performed on these PFOs. This is achieved using the anti-kt algorithm [49], with a radius parameter of R = 0.4 (AntiKt4EMPFlowJets). Various jet selection requirements are considered for the top + charm analysis, such as jets should have a transverse momentum p_T greater than 20 GeV and a pseudo-rapidity of $|\eta| < 2.5$. To reduce jets from in-time pileup[11], a discriminant known as the jet-vertex-tagger (JVT) [50] is created using a two-dimensional likelihood method. The JVT must be greater than 0.59 for jets with $p_T > 60$ GeV and $|\eta| < 2.4$.

3.5.1 Jet Flavor Tagging

Jet Flavor Tagging [51] helps in the identification of jets originating from heavy-flavor quarks. The tagging algorithm classifies the jets into those containing b-hadrons (b-jets), c-hadrons (c-jets), hadronic τ-lepton decays (τ-jets), and light-jets, which are not included in any of the above-mentioned. The latest flavor tagging algorithm recommended by the ATLAS collaboration is GN2, a graph neural network and a type of deep learning algorithm. GN2

[11] In-time pileup arises from additional pp interactions in the current bunch-crossing.

uses a deep learning model to learn the track and jet information by training on MC data with jet flavor known. It also groups tracks by origin and predicts the underlying physics process of a track origin.

The b-tagging algorithm rejects c-jets, τ-jets, and light jets with a b-jet efficiency of 70% in the case of the top + charm analysis. A jet is considered b-tagged if the D_b score is greater than a given value. D_b is a single discriminant given by:

$$D_b = \log\left(\frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_u}\right) \tag{3.5}$$

where p_b is the algorithm's jet flavor prediction output probabilities of a jet being a b-jet, p_c being a c-jet, p_τ being a τ-jet, and p_u being a light-jet. f_c is set to 0.2 and f_τ is set to 0.01. The values of f_c and f_τ are determined through an optimization procedure which would maximize the rejection of c-jets, τ-jets) and light-flavor jets.

Similarly for a c-tagging algorithm, it rejects b-jets, τ-jets, and light jets with a c-jet efficiency of 30%. The discriminant D_c is given by:

$$D_c = \log\left(\frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_u}\right) \tag{3.6}$$

where the free parameter values, f_b and f_τ that control the relative rejection between b, τ and light jets are 0.3 and 0.05 respectively.

3.6 Missing Transverse Momentum

The missing transverse momentum, E_T^{miss}, represents the imbalance in the transverse momentum of the event. The reconstruction of the E_T^{miss} [52] involves the calculation of the hard term, p_T^{hard}, as well as the soft term, p_T^{soft}. The hard term includes p_T of the reconstructed electron, muons, τ leptons, and jets, while the soft term consists of the p_T charged particle tracks not associated with any reconstructed objects but with the primary collision vertex. Therefore the E_T^{miss} vector is given by:

$$\vec{p}_T^{\text{miss}} = \underbrace{-\left(\sum_{\text{selected electrons}} \vec{p}_T^{e} + \sum_{\text{selected } \mu} \vec{p}_T^{\mu} + \sum_{\text{accepted jets}} \vec{p}_T^{\text{jet}} \right)}_{\text{hard term}} \underbrace{- \sum_{\text{unused tracks}} \vec{p}_T^{\text{track}}}_{\text{soft term}} \qquad (3.7)$$

3.7 Transverse Mass of the W boson

The transverse mass of the W boson, MWT, is given by:

$$MWT = \sqrt{2\, p_T^{\ell}\, E_T^{\text{miss}}\, (1 - \cos \Delta\phi_{\ell,\nu})}, \qquad (3.8)$$

where p_T^{ℓ} is the lepton p_T and $\Delta\phi_{\ell,\nu}$ is the azimuthal angle between the lepton and the missing transverse momentum.

3.8 Overlap Removal

After the physics object reconstruction, the overlap removal method is used to eliminate any duplication caused by the same physical object being reconstructed twice. The procedure given below is followed:

- An electron is removed if it shares a track with a muon and $\Delta R < 0.01$.
- A jet is removed if the distance between the jet and an electron is $\Delta R < 0.2$. If multiple jets satisfy this condition, only the closest one is removed.
- An electron is removed if the distance between a jet and a baseline electron is $\Delta R < 0.4$.
- If the distance between a jet and a muon candidate is $\Delta R < 0.4$, the muon is removed if the jet has more than two associated tracks; otherwise, the jet is removed.

Chapter 4

Data and Monte Carlo Samples

4.1 Data

The data collected from 2015 to 2018 (Run 2) by the ATLAS detector in pp collisions at a center-of-mass energy of 13 TeV is used in this analysis. The dataset has a full integrated luminosity of $140.1 \pm 1.2\,\mathrm{fb}^{-1}$ (an uncertainty of 0.83%) after standard data quality selections [53]. The integrated luminosity corresponding to each year is given in Table 4.1. The ATLAS

Year	**Int. lumi. (fb^{-1})**
2015	3.24
2016	33.40
2017	44.63
2018	58.78

Table 4.1: Integrated luminosity collected in each data-taking year.

Trigger System records data events for offline analysis, as mentioned in subsection 2.2.5. They use information from the Inner Detector, Calorimeter, and the Muon System. The electron trigger uses a Neural Network-based algorithm called Ringer [54] to measure a pattern of energies that matches the electron track in the ID. The electrons are identified using LooseLH, MediumLH, or TightLH working points based on a likelihood discriminator. The muon trigger uses the muon stand-alone (SA) algorithm [55], which constructs a track using hits within a region of interest (ROI) identified by the L1 trigger. The tracks are then extrapolated back to the interaction point and matched to the ID tracks. The working points based on the likelihood discriminant for muons are Tight, Medium, and Loose. For the analysis, events from the main physics stream are considered if they are accepted by at least one of the single electron or single muon triggers outlined in Table 4.2. For an electron to pass the HLT_e24_lhmedium_L1EM20VH trigger, p_T should be greater than 24 GeV with a medium identification seeded by an L1 EM cluster with $p_T > 20$ GeV.

Year	**Single Electron Triggers**	**Single Muon Triggers**
2015	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 HLT_mu40
2016–2018	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0	HLT_mu26_ivarmedium HLT_mu50

Table 4.2: List of single-lepton triggers used for different years.

4.2 Monte Carlo Samples

The Monte Carlo (MC) samples are generated from computer simulations that imitate LHC collisions. These simulations take into account the complex physics behind the collisions as well as the detector properties. The Run 2 MC samples are divided into three distinct subsets:

- MC16a generated with pileup conditions of the 2015 and 2016 data.
- MC16d generated with pileup conditions of the 2017 data.
- MC16e generated with pileup conditions of the 2018 data.

These samples are normalized to the integrated luminosity of the corresponding data-taking periods as listed in the Table 4.1.

The MC simulations [56] have a key role to play in various aspects of physics analysis, such as event selection, where they are used to select interesting events. They also help in background estimation, enabling the modeling of background processes in detail. Additionally, they are also used for efficiency assessment to correct biases or limitations in real data, as well as to evaluate systematic uncertainty, validate data, and perform calibration. Various MC scale factors are applied to the simulated events to match data and MC samples, accounting for differences in trigger, reconstruction, and identification efficiencies, as well as pileup, between data and MC. Simulated events are also normalized by their production cross sections.

The MC sample generation utilizes various event generators, such as Pythia [57], Herwig [58], Sherpa [59], MadGraph [60], and POWHEG [61], to create theoretical particle collision events that contain mathematical descriptions of particle properties, such as energy and momentum. The most common tools for detector simulation are the Geant4 [62] software package (for full simulation) and the AtlFast3 [63] software package (for fast simulation). These help in simulating the interaction of particles with the ATLAS detector. Descriptions of the various MC samples used in the single top with an associated charm analysis are provided in the following sections.

4.2.1 Signal Sample

The signal samples are MC samples that represent the physics process of interest. The top + charm signal samples are generated based on the t-channel single top JO[12] (Job Option file) using the SM CKM matrix. The single-top t-channel samples are generated using

[12]Python script used to configure and run an analysis job

the POWHEG BOX v2 event generator at NLO with the NNPDF3.0NLO_NF4 [64] PDF set implementing the four-flavor scheme. Parton showers, hadronization, and the underlying event are modeled using Pythia 8.230 with the A14 tuned parameters and the NNPDF2.3LO PDF set [65]. Subsequently, the LHE (Les Houches Event) [66] filters are applied to separate the events with charm quarks, referred to as the $t + c$ signal, from the events with u, d, or s quarks, which constitute the $t + q$ background.

4.2.2 Background Samples

The background samples are MC simulations that mimic the signal process. They are of two types: reducible and irreducible backgrounds. Irreducible backgrounds have the same final state as the signal and are well-modeled by the MC, while reducible backgrounds do not have identical final states to that of the signal but can still pass the final state selection due to detector effects or misidentification. They are often estimated using data-driven methods. Data-driven methods are discussed in detail in Chapter 6.

V+jets and VV Backgrounds

V+jets (V = $W^{\pm}$, Z), samples are generated using the event generator, SHERPA 2.2.11 [67]. In this process, matrix elements are calculated with NLO precision for up to 2 jets and LO precision for up to 5 jets in the five-flavor scheme. The b-quark and c-quark are considered massless at the matrix-element level and massive in the parton shower. Comix [68] is used to compute these matrix elements, while OPENLOOPS [69] provides virtual QCD corrections for matrix elements at NLO accuracy. The PDF set used for this Sherpa version is the

Hessian NNPDF3.0NNLO [64]. For the parton shower, the default Catani–Seymour dipole scheme [70] is utilized, and for hadronization, the cluster hadronization model is used. A colour-exact variant of the MC@NLO algorithm matches the NLO matrix elements of a given jet multiplicity with the parton shower. All the different jet multiplicities are then combined into an inclusive sample using an improved CKKW [71] matching procedure.

The diboson (VV) samples are produced using SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The matrix elements are calculated with NLO accuracy for events with up to one additional jet and with LO accuracy for events with two or three additional jets. They are matched and merged with the SHERPA parton shower.

Top Backgrounds

The top backgrounds considered in this analysis are $t\bar{t}$, tW, t-channel single-top (tq), s-channel single-top, and raretop processes. The top backgrounds, $t\bar{t}$, tW, single-top s-channel samples are generated with the POWHEG BOX v2 event generator at NLO with the PDF set NNPDF3.0NLO using the five-flavor scheme. The tq sample generation is described in Section 4.2.1. These are then interfaced with PYTHIA 8.230 using the A14 parameter tune to model parton shower, hadronization, and underlying events. In the case of $t\bar{t}$, the h_{damp} parameter controls the matrix element to parton showering matching, and it is set to 1.5 times the top mass, m_t, where m_t = 172.5 GeV. The top decays are modeled by MAD-SPIN [72] in the case of tq, while they are handled directly by POWHEG BOX in the case of others. The interference between the $t\bar{t}$ and tW sample production is reduced using the diagram-removal scheme [73].

The contribution of raretop processes is minimal in this analysis. They include the production of top in association with electroweak bosons($t\bar{t}W$, $t\bar{t}W$, tWZ), the higgs boson($t\bar{t}H$), and in 3 top or 4 top final states. These are generated using MadGraph5_aMC@NLO and interfaced with Pythia.

All the MC samples, along with their DSIDs and event generators used in this analysis, are summarized in Table 4.3.

Sample	DSID	Generator
tc	604227,604229	Powheg +Pythia8
tq	604228,604230	Powheg +Pythia8
$t\bar{t}$	410470	Powheg +Pythia8
tW	601352,601355	Powheg +Pythia8
single-top s-channel	410644,410645	Powheg +Pythia8
$W + jets$	700338-700349,700843	Sherpa 2.2.11
$Z + jets$	700320-700325,700335-700337,700792-700794	Sherpa 2.2.11
Diboson	700587-700590,700600-700605	Sherpa 2.2.11
rareTop	410276-410278	MadGraph +Pythia8

Table 4.3: DSIDs and Event Generators of MC Samples.

Chapter 5

Event Selection and Analysis Regions

The selections applied to the physics objects are used to select events of interest and to minimize background. The selected events are typically divided into several analysis regions. They are signal regions (SR), control regions (CR), and validation regions (VR). The signal region is chosen to maximize the signal-to-background ratio. They are used for hypothesis testing, which will be discussed in detail in Chapter 8. The control regions are enhanced with specific dominant backgrounds and are used to constrain systematic uncertainties. The validation regions are used to validate the background modeling.

5.1 Signal Region

The single top + charm process is given by:

$$pp \to tc, \tag{5.1}$$

with,

$$t \to W^{\pm}b \to l^{\pm}\nu b, \tag{5.2}$$

where $l = e, \mu$.

The analysis presented considers $W^{\pm}$ decaying leptonically, rather than hadronically, due to a large QCD background in the hadronic decay. With an isolated lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ from the neutrino in the final state, leptonic decay provides a clean signature. Thus, the single top + charm SR contains events with:

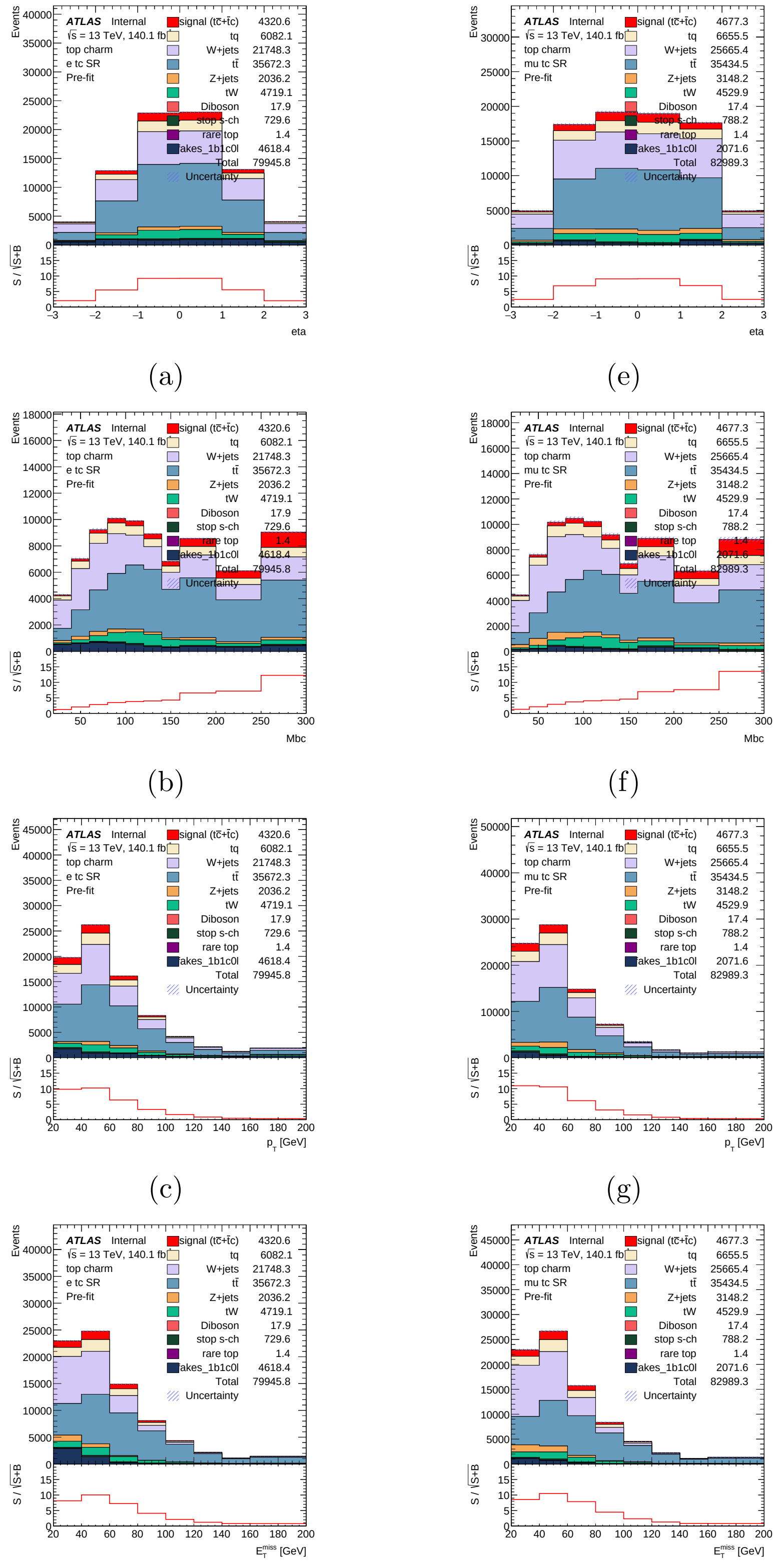
- exactly one lepton (e, μ) with $p_T > 27$ GeV,
- exactly one b-tagged jet and one c-tagged jet with $p_T > 20$ GeV,
- zero light jets,
- $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV due to the neutrino in the final state, and
- $MWT > 40$ GeV to suppress the multijet background.

Distributions of a few kinematic variables in the SR are shown in Figure 5.1.

5.2 Control Regions

5.2.1 $t\bar{t}$ CR

The $t\bar{t}$ background is one of the dominant backgrounds in the single top + charm analysis. Since both the top quarks decay to $W^{\pm}b$, there can be 2 b-tagged jets in the final state. Thus the $t\bar{t}$ CR contains events with exactly one lepton (e, μ) with $p_T > 27$ GeV, two b-tagged jets and two light jets with $p_T > 20$ GeV, zero c-tagged jets, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV and $MWT > 40$ GeV. The $t\bar{t}$ CR helps in constraining the $t\bar{t}$ normalization in a simultaneous fit with the SR. The lepton p_T and η distributions in the $t\bar{t}$ CR are shown in Figure 5.2.



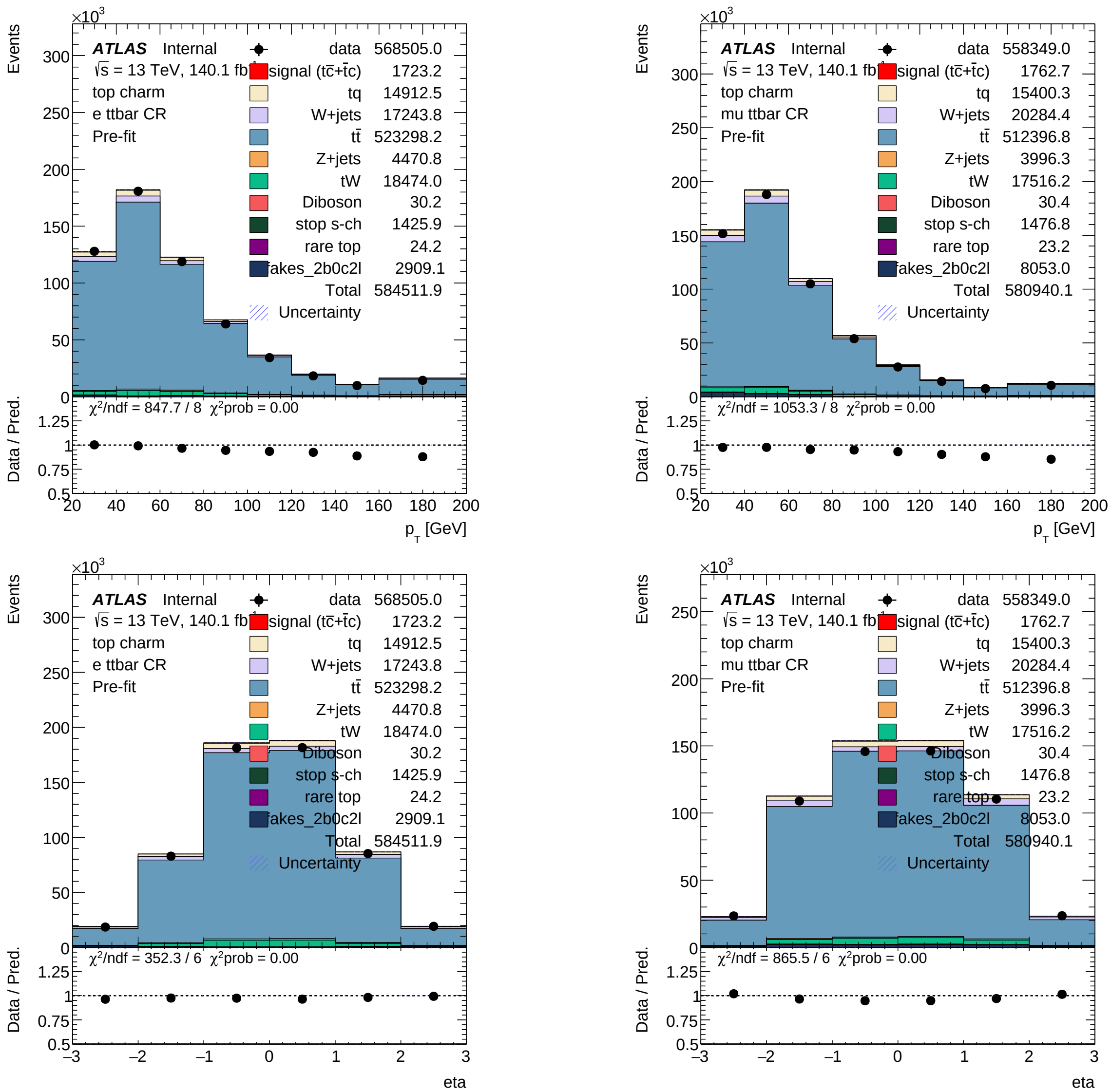
(a) (e)

(b) (f)

(c) (g)

(d) (h)

Figure 5.1: (a)(e) η, (b)(f) M_{bc}, (c)(g) lepton p_T and (d)(h) E^{miss} distributions in SR for electron (left) and muon channel (right). Uncertainties are statistical only.



(a) (b)

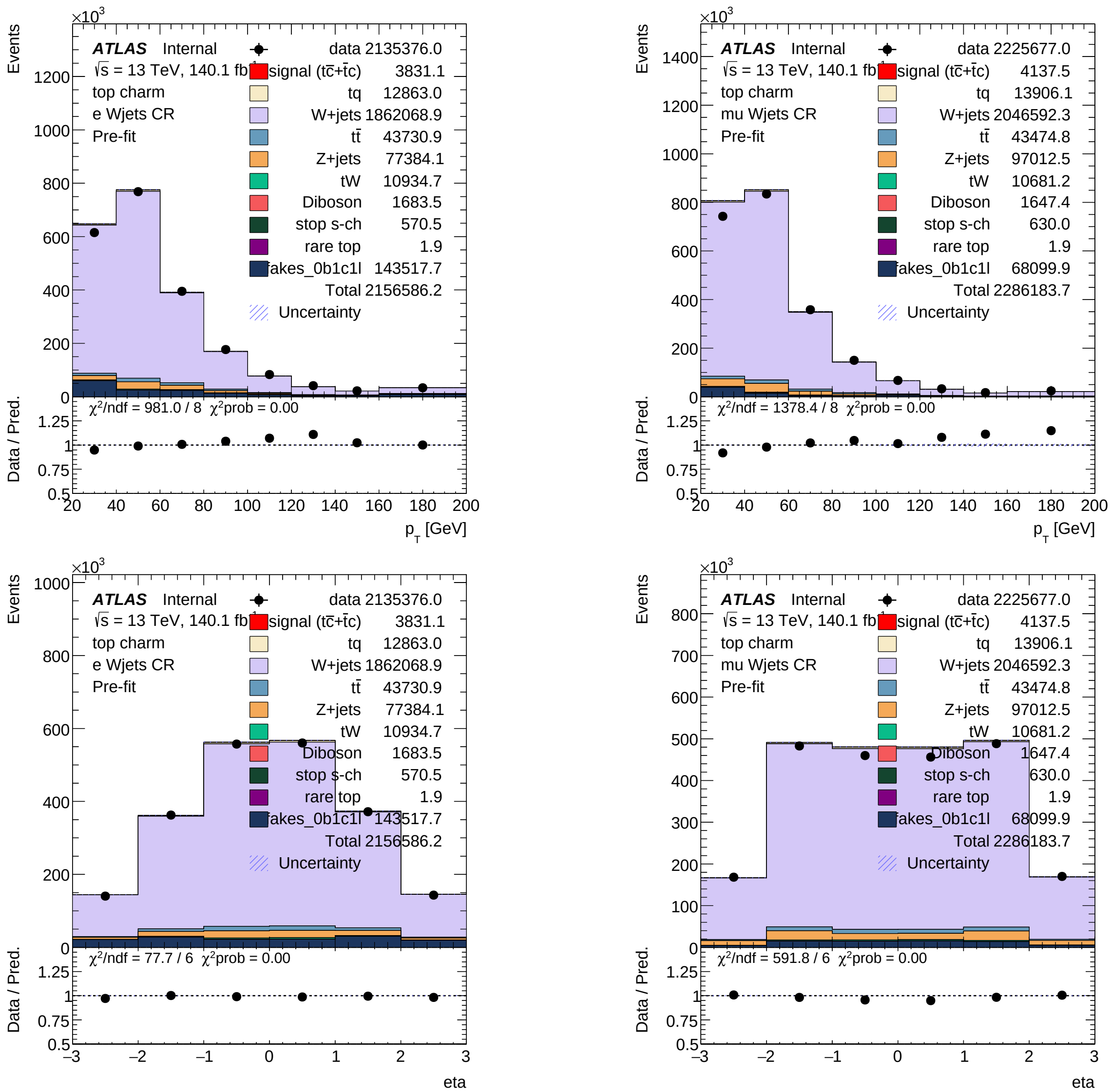
Figure 5.2: Distribution of lepton p_T and η in $t\bar{t}$ CR for (a) electron and (b) muon channel. Uncertainties are statistical only.

5.2.2 $W^{\pm}$+jets CR

The largest background in the single top + charm analysis is $W^{\pm}$+jets. The $W^{\pm}$+jets events contain light jets and c-tagged jets in the final state. Thus, $W^{\pm}$+jets CR contains events with exactly one lepton (e, μ) with $p_T > 27$ GeV, one c-tagged jet, and one light jet with $p_T > 20$ GeV, zero b-tagged jets, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, and $MWT > 40$ GeV. Similar to $t\bar{t}$, the W+jets CR helps in constraining the W+jets normalization in a simultaneous fit with the SR. The lepton p_T and η distributions in the W+jets CR are shown in Figure 5.3. A 15-20% disagreement between data and MC is observed in the high p_T region and is especially visible in the electron channel. This indicates a mismodelling of the $W^{\pm}$+jets background. The treatment of this mismodelling is described in Chapter 6.

5.3 Machine Learning

A machine learning technique called a Boosted Decision Tree (BDT) [74] is used in the analysis to increase the sensitivity of the signal sample. A decision tree algorithm consists of nodes, branches, and leaves. A node represents a condition that evaluates the input features, and the outcome of the evaluation is split into branches. Each branch provides one outcome based on the condition from the previous node. A node at the end of the branch, which is not split further, is called a leaf. Figure 5.4 shows a pictorial representation of a decision tree. The algorithm begins at the root node with the entire training dataset, which contains both signal and background events. At each node, it checks the best cut value for each kinematic variable or feature and the best overall cut that separates signal and background



(a) (b)

Figure 5.3: Distribution of lepton p_T and η in W+jets CR for (a) electron and (b) muon channel. Uncertainties are statistical only

events, then divides the data into two branches. The splitting stops when there is no further reduction in impurity, and the node becomes a leaf. The impurity is described using the Gini index [75] given by:

$$\text{Gini} = (S + B)P(1 - P) = \frac{SB}{S + B}, \tag{5.3}$$

where $P = S/(S + B)$ is the signal purity and S and B are signal and background events, respectively. A BDT is an ensemble of these decision trees that combine into a strong classifier for better signal background separation.

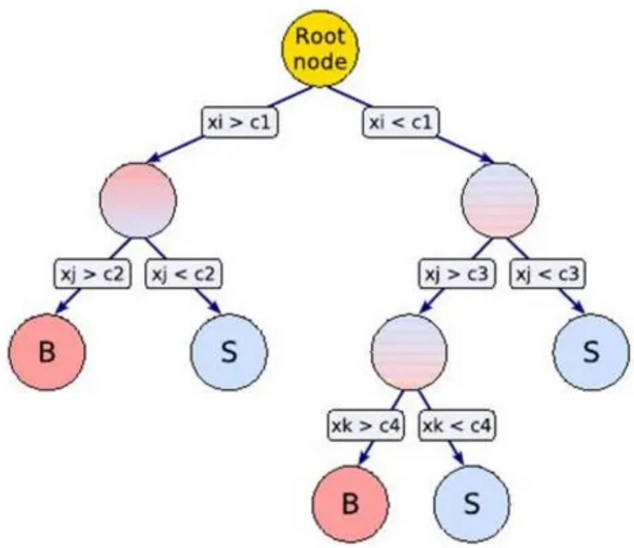


Figure 5.4: Decision Tree Diagram

5.3.1 XGBoost

XGBoost [76], or Extreme Gradient Boosting, is a Python package used to train the BDT. XGBoost provides several hyperparameters. These are settings that determine an algorithm's learning process. The hyperparameters used in the BDT for this analysis are :

- Learning rate: Shrinks the feature weights after each boosting step to make the learning process smoother and avoid overfitting.

- Maximum depth: Indicates the maximum number of splits per tree. Increasing this value can make the model more complex and more likely to overfit.

- L1 and L2 regularization: L1 regularization term (reg_alpha) and L2 regularization term (reg_lambda) are added to leaf weights to ensure sparsity and smoothness and prevent overfitting.

- Subsample: It is the fraction of training events.

- Min child weight: It is the minimum sum of weights needed in a leaf node.

- Gamma: Minimum loss reduction needed to make a split on a leaf node of the tree.

The range of values used for the hyperparameters is given in the Table 5.1.

Hyperparameters	Range of Values
learning_rate	[0.01, 0.02]
max_depth	[3, 6]
reg_alpha (L1)	[1, 5]
reg_lambda (L2)	[0.5, 2]
subsample	[0.7, 0.9]
min_child_weight	[3, 10]
gamma	[0.1, 0.5]

Table 5.1: Hyperparameter values used for the BDT training.

The BDT is trained using the events in the SR. The training includes events with negative weights, but the absolute values of the events are taken and rescaled by the total sum of weights to the total sum of absolute weights. An additional scale factor is also applied to make the number of signal and background events equal.

A 5-fold cross-validation technique is used, where the dataset used to train the BDT is split into five equal parts or folds. Four folds are used for training and one for testing. This is repeated 5 times, by changing the fold used for testing each time. 85% of the dataset is used as the training set, and 15% as the testing set. The train set is again split into a 70% train set and a 15% validation set.

The BDT is evaluated using the area under the ROC (Receiver Operating Characteristic) curve, known as the AUC (Area Under Curve). The ROC curve is obtained using the True Positive Rate (TPR) and the False Positive Rate (FPR). By integrating the ROC curve, the AUC value can be obtained. It is the probability that the BDT score of a randomly selected signal event has a higher value than that of a randomly selected background event. A higher AUC value indicates better separation. Typical AUC values and their descriptions are shown in Table 5.2. The ROC curves of the training and testing set used in the BDT training are shown in Figure 5.5. An AUC score of 0.84 indicates good/excellent discriminatory ability. Similar curves for the training and test sets indicate that there is no overfitting.

The variables used in the BDT training and their impact on the AUC value are shown in Figure 5.6. The relationship between the features and the AUC value is described through permutation feature importance (PFI) [77]. PFI measures a feature's contribution to model accuracy by how much the AUC drops when the feature's values are permuted. The variables used in the training are shown in Table 5.3. The BDT output distribution is given in Figure 5.7.

AUC Range	Interpretation
1.0	Perfect – the model correctly classifies every instance.
0.9 – 1.0	Outstanding / Excellent discriminatory ability.
0.8 – 0.9	Good / Excellent discriminatory ability.
0.7 – 0.8	Acceptable / Fair discriminatory ability.
0.5 – 0.7	Poor discriminatory ability; limited predictive power.
0.5	No discriminatory ability; performance equals random guessing.
< 0.5	Worse than random guessing; model predictions are systematically incorrect.

Table 5.2: Interpretation of AUC values.

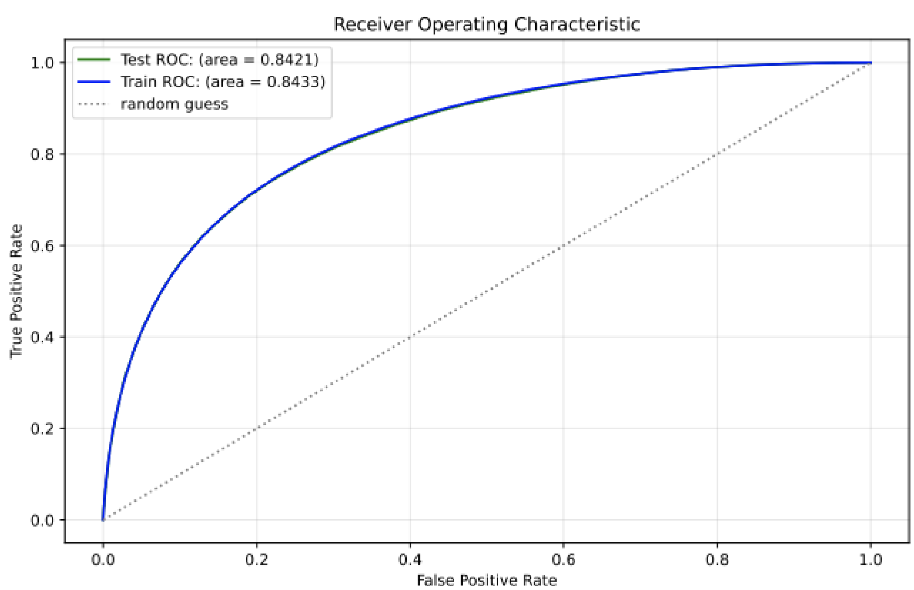


Figure 5.5: The ROC curves of the training and testing dataset

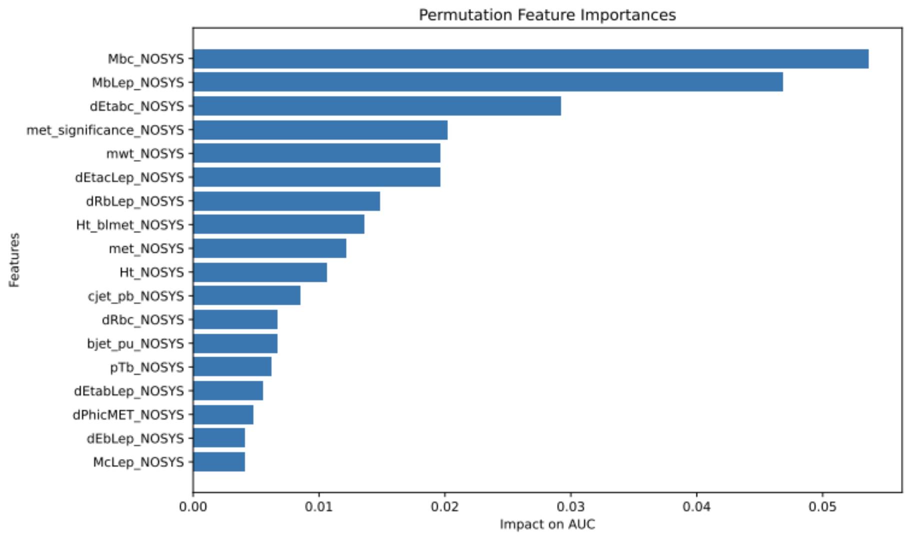
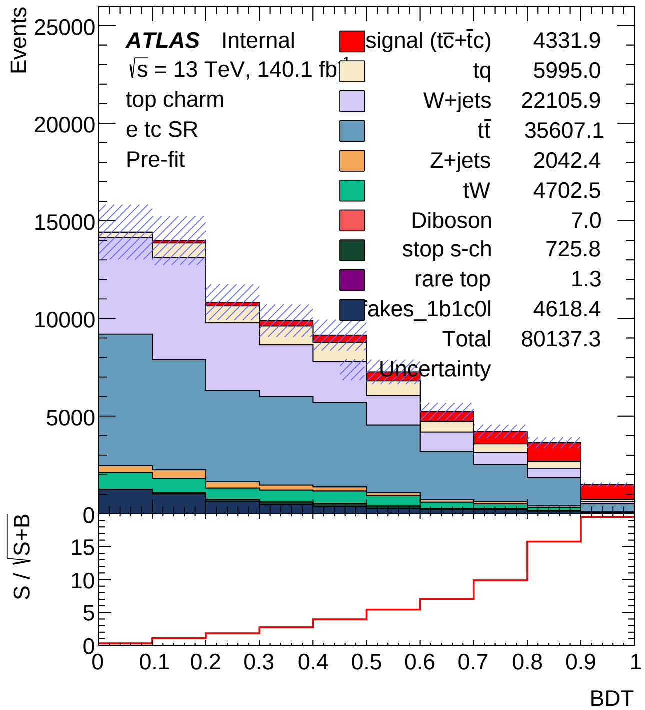


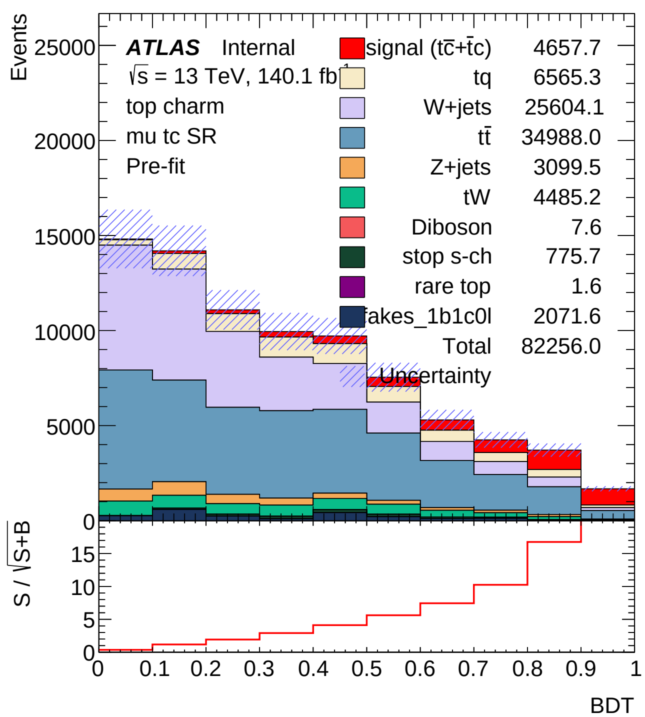
Figure 5.6: The variables used in BDT and its permutation feature importance

Variable	**Description**
M_{bc}	The invariant mass of the bjet and cjet in the event.
$M_{b\ell}$	The Invariant mass of the bjet and lepton.
$\Delta\eta_{bc}$	Difference in pseudorapidity between the bjet and cjet.
met significance	Ratio of E_T^{miss} to its uncertainty.
MWT	Transverse mass of the W boson.
$\Delta\eta_{c\ell}$	Difference in pseudorapidity between the cjet and lepton.
$\Delta R_{b\ell}$	Angular separation between the b-jet and lepton.
$H_T^{b\ell+E_T^{\text{miss}}}$	Scalar sum of p_T of the bjet, lepton, and E_T^{miss}.
E_T^{miss}	Missing transverse energy.
H_T	Scalar sum of the transverse momenta of jets and the lepton.
cjet pb	Flavor tagging output probability
ΔR_{bc}	Angular separation between the bjet and cjet.
bjet pu	Flavor tagging output probability.
bjet p_T	Transverse momentum of the bjet
$\Delta\eta_{b\ell}$	Difference in pseudorapidity between the bjet and lepton.
$\Delta\phi(c, E_T^{\text{miss}})$	Azimuthal angle difference between the cjet and the missing transverse energy.
$\Delta E(b, \ell)$	Energy difference between the bjet and the lepton.
$M_{c\ell}$	The Invariant mass of the cjet and lepton.

Table 5.3: Input variables used in the BDT training and their description.



(a)

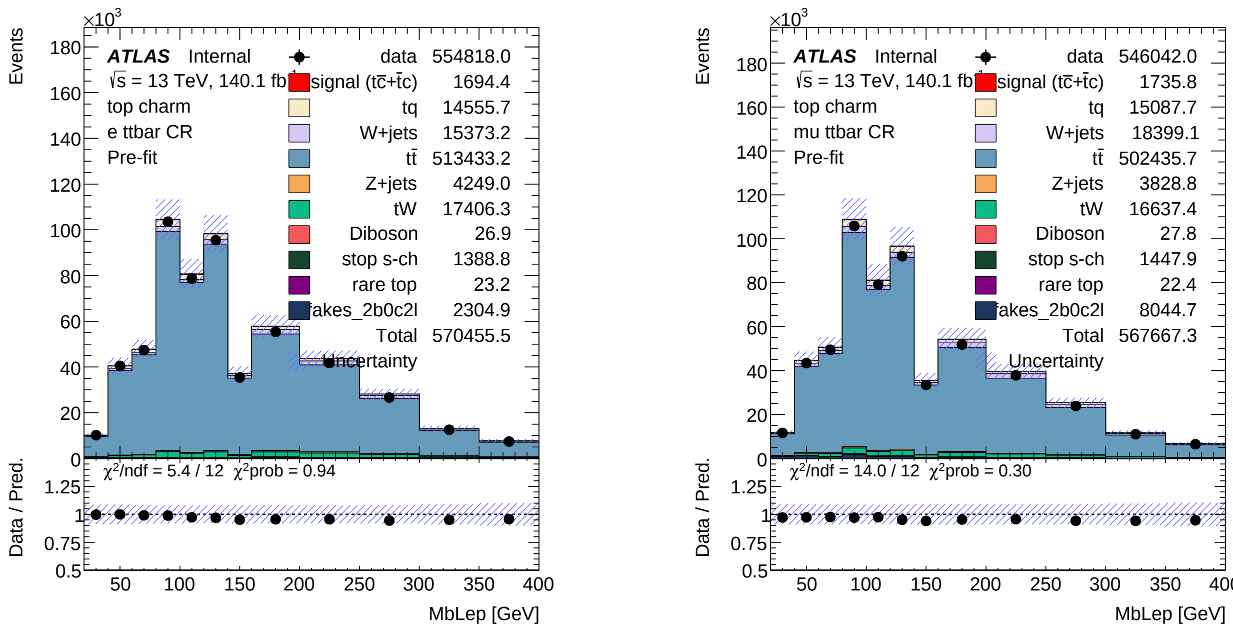


(b)

Figure 5.7: BDT output distribution in the signal region for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.

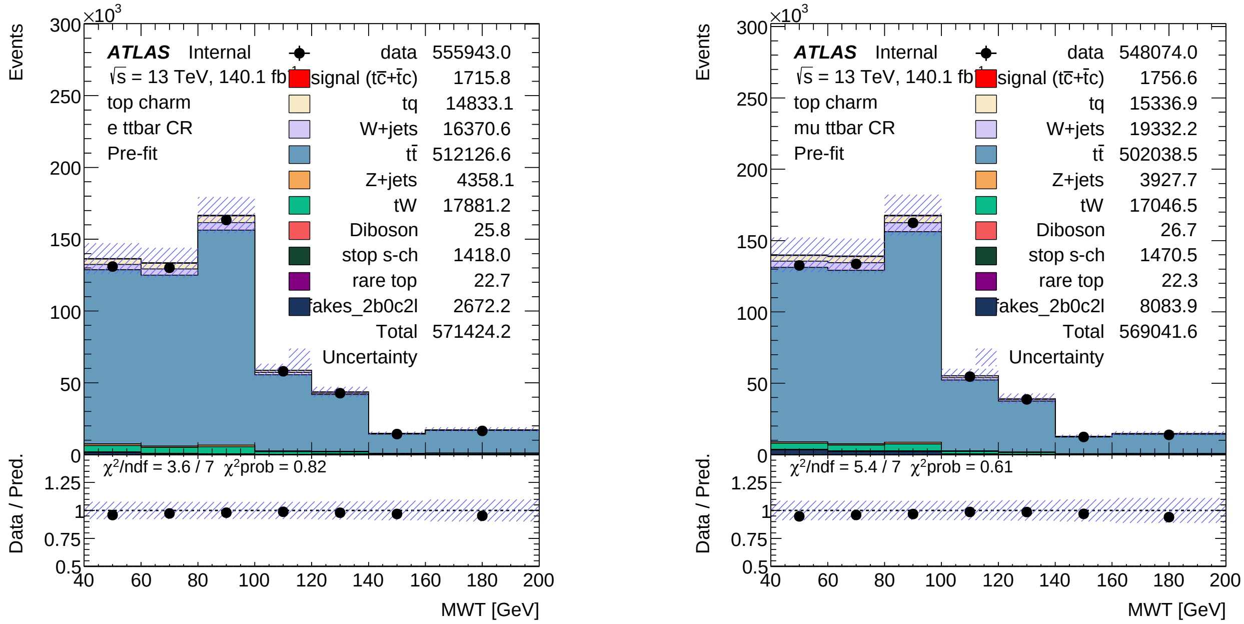
5.3.2 BDT validation

The BDT is validated by comparing the data-to-mc ratios of the variables used in the BDT training in the $t\bar{t}$ and W+jets CRs to ensure they are well modeled. The plots are shown in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11 for the $t\bar{t}$ CR and Figure 5.12, Figure 5.13 Figure 5.14 and Figure 5.14 for the W+jets CR. Reasonable agreement within systematics is obtained for both the CRs.



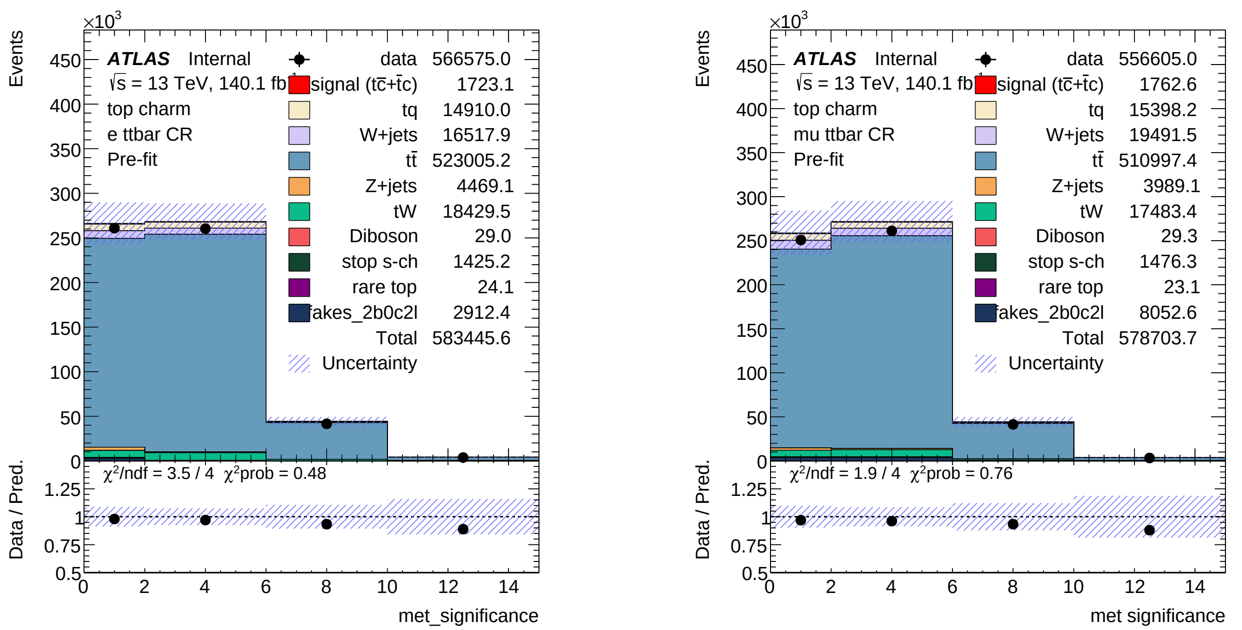
(a) (b)

Figure 5.8: Distribution of $M_{b\ell}$ in the $t\bar{t}$ CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



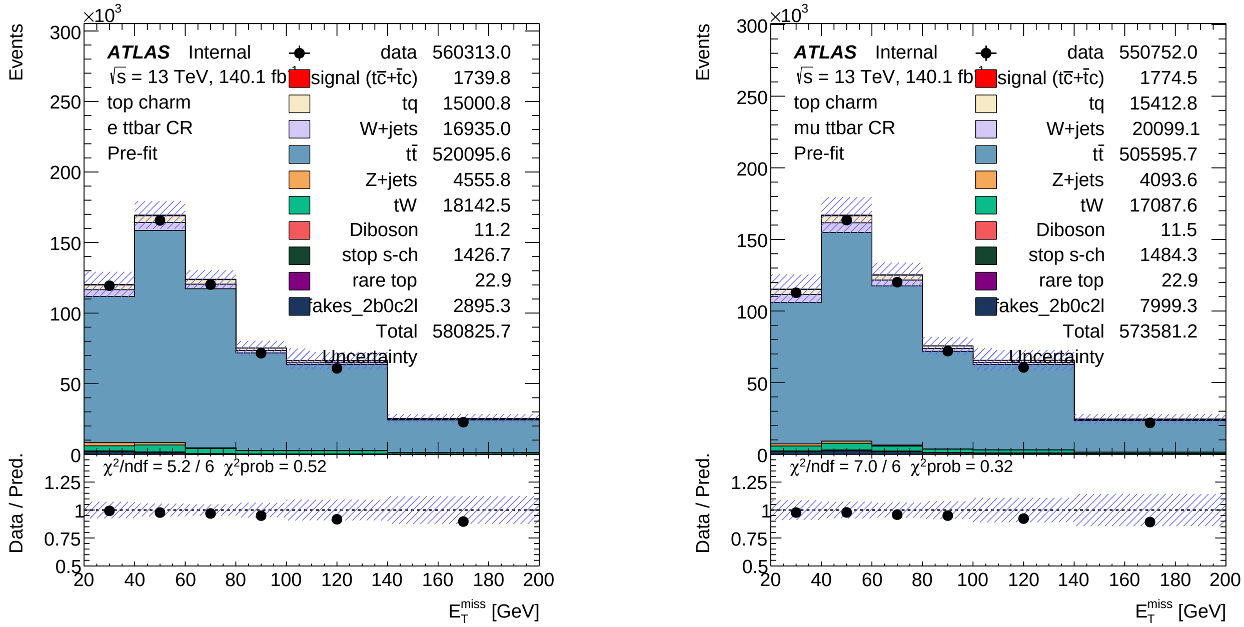
(a) (b)

Figure 5.9: Distribution of MWT in the $t\bar{t}$ CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



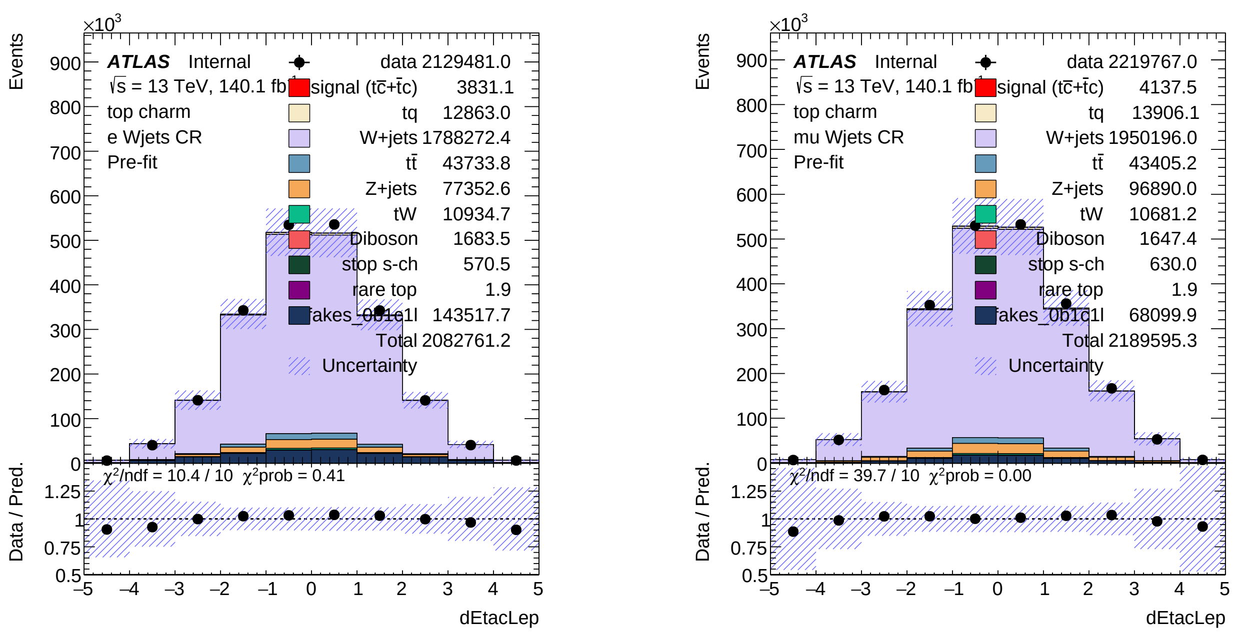
(a) (b)

Figure 5.10: Distribution of met significance in the $t\bar{t}$ CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



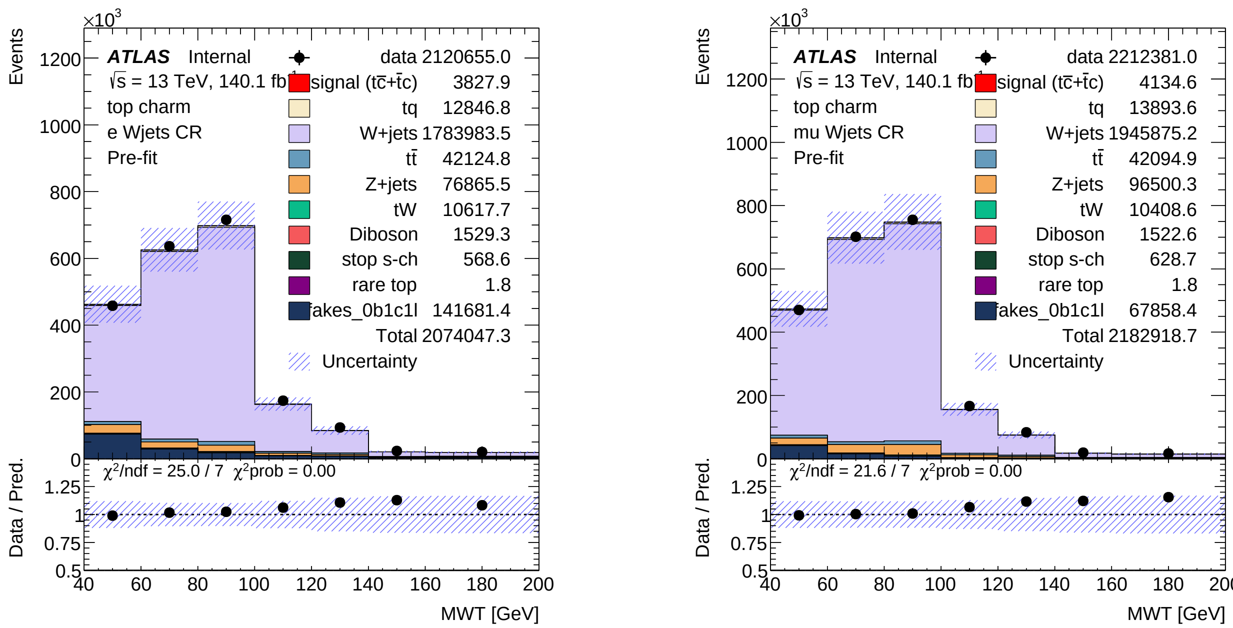
(a) (b)

Figure 5.11: Distribution of E_T^{miss} in the $t\bar{t}$ CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



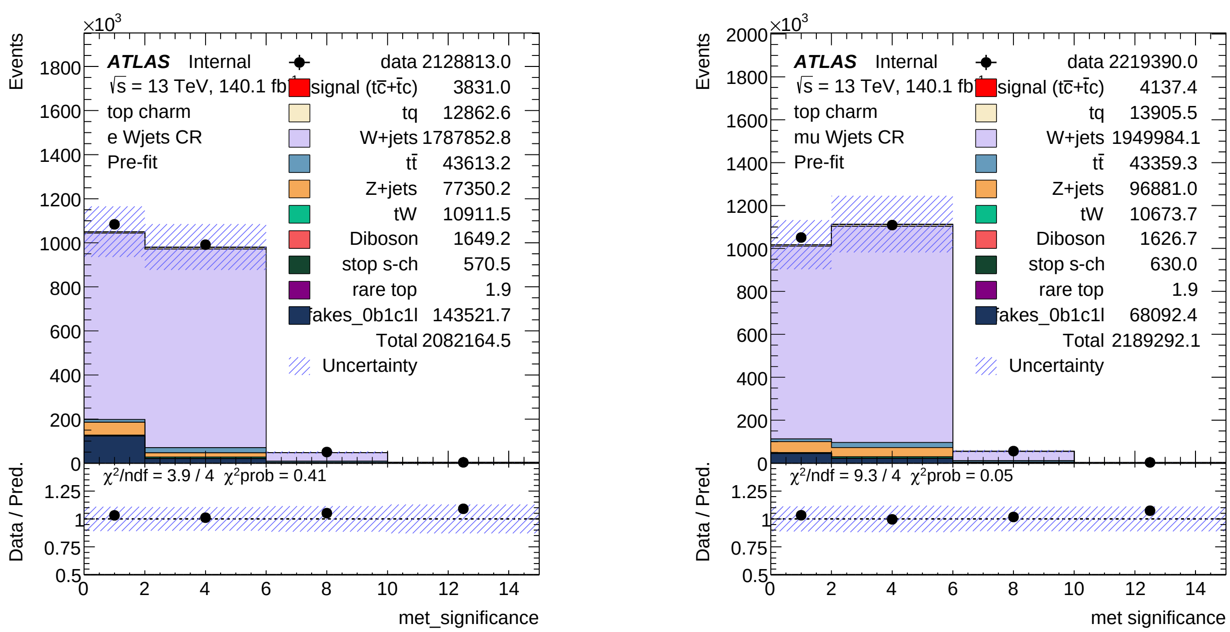
(a) (b)

Figure 5.12: Distribution of $\Delta\eta_{c\ell}$ in the $W^{\pm}$+jets CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



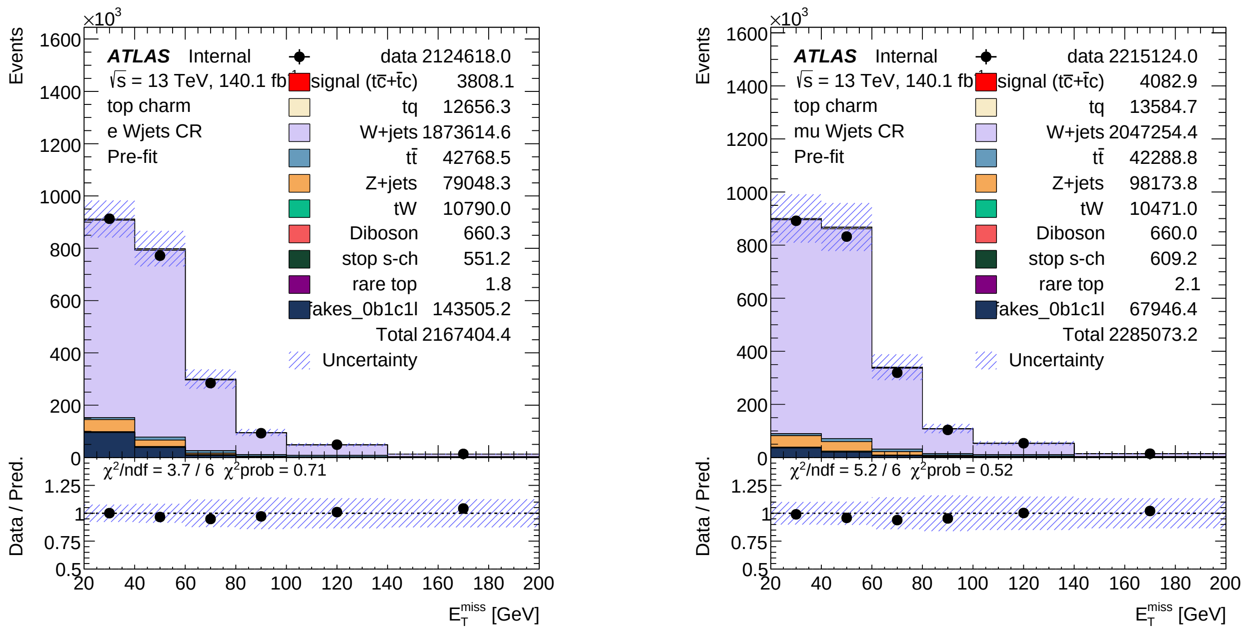
(a) (b)

Figure 5.13: Distribution of MWT in the $W^{\pm}$+jets CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



(a) (b)

Figure 5.14: Distribution of met significance in the $W^{\pm}$+jets CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.



(a) (b)

Figure 5.15: Distribution of E_T^{miss} in the $W^{\pm}$+jets CR for (a) electron channel, (b) muon channel. Uncertainty includes both statistical and systematic.

Chapter 6

Fake Background Estimation and Validation

A substantial source of background is introduced in the analysis due to misidentification. Therefore, non-prompt leptons, also known as fake leptons, are expected to be present in the analysis. Accurately predicting fake contributions from simulation is challenging. These fake rates depend on the details of the physics simulation. In non-perturbative regions where the physics becomes more complex and less predictable, the simulation's reliability decreases. Rates are also influenced by material composition and the response of the detector. Hence, a data-driven technique called the Matrix Method [78] is used to estimate the fake/non-prompt leptons. TopCPToolKit [79], along with FastFrames [80], is used to generate ntuples containing information to determine the fake backgrounds.

6.1 Matrix Method

In this technique, the fake and the real efficiencies are used to calculate the fake yield. The real/prompt leptons and fake/non-prompt leptons are selected using the loose criteria outlined in Tables 3.1 and 3.2. The probability of real leptons passing the tight selection is called real efficiency, and the probability of fake leptons passing the tight selection is called fake efficiency. The real efficiencies are calculated using MC in the region given by:

- exactly one lepton (e, μ) with $p_T > 27$ GeV,
- $E_{\mathrm{T}}^{\mathrm{miss}} > 25$,

while the fake efficiencies are calculated using data in the fake-enriched regions, which are essentially SR and CRs, discussed in Chapter 5. They are selected with the requirement of $E_{\mathrm{T}}^{\mathrm{miss}} < 25$ and no MWT cut. The Truth Classification Package [81] is incorporated in the FastFrame code to identify whether a lepton is prompt or non-prompt.

The number of events containing loose and tight leptons in data can be written using the equations:

$$
\begin{aligned}
N_{\mathrm{data}}^{\mathrm{loose}} &= N_{\mathrm{real}}^{\mathrm{loose}} + N_{\mathrm{fake}}^{\mathrm{loose}} \\
N_{\mathrm{data}}^{\mathrm{tight}} &= N_{\mathrm{real}}^{\mathrm{tight}} + N_{\mathrm{fake}}^{\mathrm{tight}}
\end{aligned} \tag{6.1}
$$

The real and fake efficiencies are denoted as ε_{r} and ε_{f} respectively. The prompt lepton contribution is subtracted from the tight and loose data samples to obtain the fake efficiency:

$$\varepsilon_{\mathrm{f}} = \frac{N_{\mathrm{fake}}^{\mathrm{tight}}}{N_{\mathrm{fake}}^{\mathrm{loose}}} = \frac{N_{\mathrm{data}}^{\mathrm{tight}} - N_{\mathrm{prompt,\ MC}}^{\mathrm{tight}}}{N_{\mathrm{data}}^{\mathrm{loose}} - N_{\mathrm{prompt,\ MC}}^{\mathrm{loose}}} \tag{6.2}$$

$$\varepsilon_{\mathrm{r}} = \frac{N_{\mathrm{real}}^{\mathrm{tight}}}{N_{\mathrm{real}}^{\mathrm{loose}}} \tag{6.3}$$

The known values, $N_{\mathrm{data}}^{\mathrm{tight}}$ and $N_{\mathrm{data}}^{\mathrm{loose}}$ can be related to the unknown values, $N_{\mathrm{real}}^{\mathrm{loose}}$ and $N_{\mathrm{fake}}^{\mathrm{loose}}$ using a matrix M_{ϵ} and hence the name, matrix method:

$$\begin{pmatrix} N_{\mathrm{data}}^{\mathrm{tight}} \\ \\ N_{\mathrm{data}}^{\mathrm{loose}} \end{pmatrix} = \begin{pmatrix} \varepsilon_{\mathrm{r}} & \varepsilon_{\mathrm{f}} \\ \\ 1-\varepsilon_{\mathrm{r}} & 1-\varepsilon_{\mathrm{f}} \end{pmatrix} \begin{pmatrix} N_{\mathrm{real}}^{\mathrm{loose}} \\ \\ N_{\mathrm{fake}}^{\mathrm{loose}} \end{pmatrix} \tag{6.4}$$

$$N_{\mathrm{data}}^{\mathrm{tight,loose}} = \mathrm{M}_{\epsilon} N_{\mathrm{real,fake}}^{\mathrm{loose}} \tag{6.5}$$

From equation 6.1, $N_{\text{real}}^{\text{loose}}$ can be solved:

$$N_{\text{real}}^{\text{loose}} = N_{\text{data}}^{\text{loose}} - N_{\text{fake}}^{\text{loose}} \tag{6.6}$$

By the inversion of the matrix M_{ϵ}, $N_{\text{fake}}^{\text{loose}}$ can be determined:

$$N_{\text{fake}}^{\text{loose}} = \frac{1}{\varepsilon_r - \varepsilon_f}\left[(\varepsilon_r - 1)N_{\text{data}}^{\text{tight}} + \varepsilon_r N_{\text{data}}^{\text{loose}}\right] \tag{6.7}$$

Now using equation 6.7 and the fact that:

$$N_{\text{fake}}^{\text{tight}} = \varepsilon_{\text{f}} \cdot N_{\text{fake}}^{\text{loose}} \tag{6.8}$$

$N_{\text{fake}}^{\text{tight}}$ can be obtained by:

$$N_{\text{fake}}^{\text{tight}} = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f}\left[(\varepsilon_r - 1)N_{\text{data}}^{\text{tight}} + \varepsilon_r N_{\text{data}}^{\text{loose}}\right] \tag{6.9}$$

An event weight $w_{fake,i}$ is defined using these efficiencies and applied to each event in a loose data sample. Therefore:

$$w_{fake,i} = \begin{cases} \dfrac{\varepsilon_{fi}}{\varepsilon_{ri} - \varepsilon_{fi}}(\varepsilon_{ri} - 1), & \text{if lepton candidate } i \text{ is tight } (N_{\text{data}}^{\text{tight}} = 1 \text{ and } N_{\text{data}}^{\text{loose}} = 0) \\ \\ \dfrac{\varepsilon_{fi}}{\varepsilon_{ri} - \varepsilon_{fi}}\,\varepsilon_{ri}, & \text{if lepton candidate } i \text{ is loose } (N_{\text{data}}^{\text{tight}} = 0 \text{ and } N_{\text{data}}^{\text{loose}} = 1) \end{cases} \tag{6.10}$$

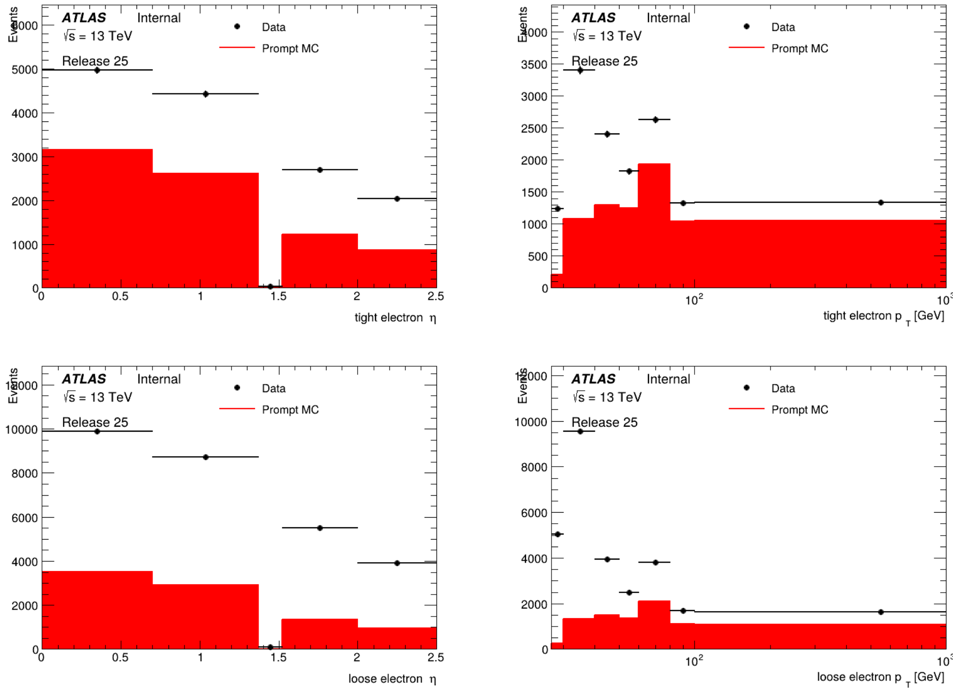
where ε_{fi} and ε_{ri} are the values of fake and real efficiencies corresponding to the lepton i. The results of the loose sample are extrapolated to the tight sample. The total fake yield in the tight sample is estimated by:

$$N_{\text{fake}}^{\text{tight}} = \sum_{\text{events}} w_{fake,i} \tag{6.11}$$

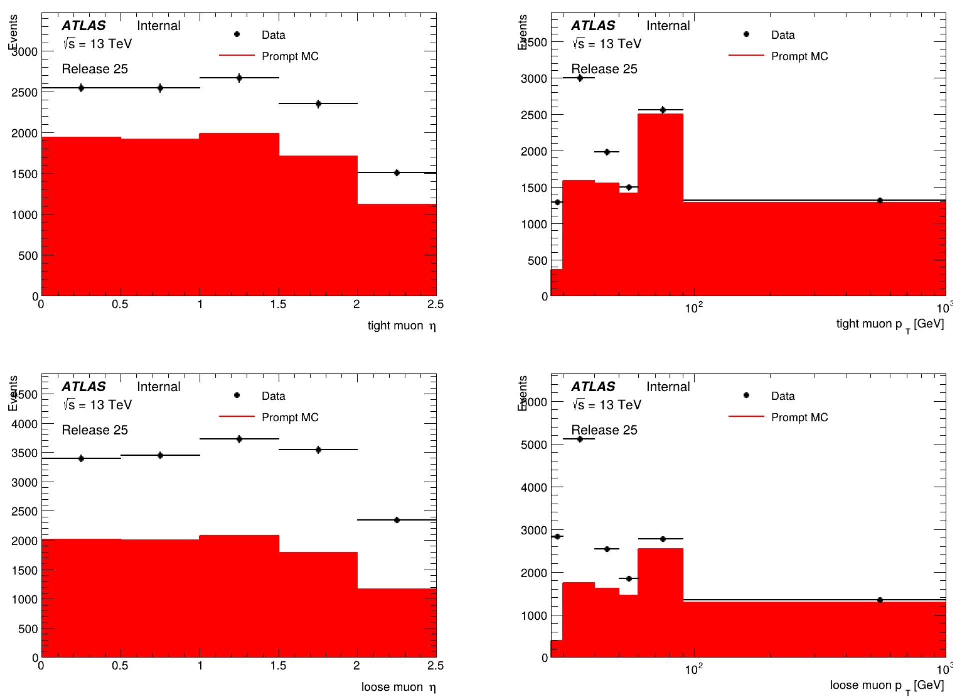
In this method, each event has its corresponding weight $w_{fake,i}$ and hence histograms can be filled and binned in any variable of interest. Negative events are expected since ε_{ri} is always less than 1, and therefore, when a lepton candidate passes the tight criteria, $w_{fake,i}$ would be negative, and as a result, summing over all the events, the total fake yield could be negative in bins of lower statistics. Since the estimated fake yields depend on the real and fake efficiency values, any change in these values due to binning or limited statistics can affect the fake yield. The statistical uncertainty of $N_{\text{fake}}^{\text{tight}}$ is given by:

$$\sigma_{N_{\text{fake}}^{\text{tight}}} = \sqrt{\sum_{\text{events}} w_{fakes,i}^2} \tag{6.12}$$

The efficiencies are measured in bins of p_T and $|\eta|$. They are estimated using the number of tight and loose leptons as described in the equation 6.2. Figure 6.2 shows the number of leptons in p_T and $|\eta|$ bins for the fake enriched SR for the 2018 data-taking year. Similar measurements are performed for all regions and all data-taking years. The efficiencies are calculated separately for each year due to varying pileup conditions. Figure 6.2 shows 2D real efficiency plots in p_T and $|\eta|$ bins. They are typically in the range 0.8-0.95. Figures 6.3, 6.4, and 6.5 show the fake efficiencies for SR and CRs for electron and muon channels for the three data taking periods. These 2D efficiency maps are incorporated with TopCPToolkit, where the fake weights are calculated and applied to the loose data sample.

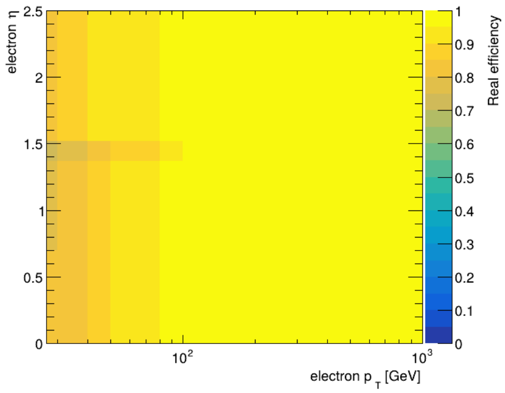


(a)

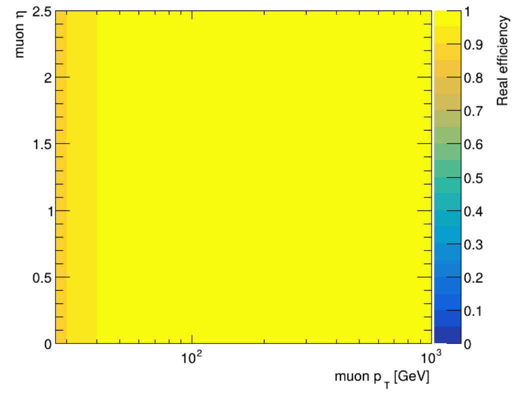


(b)

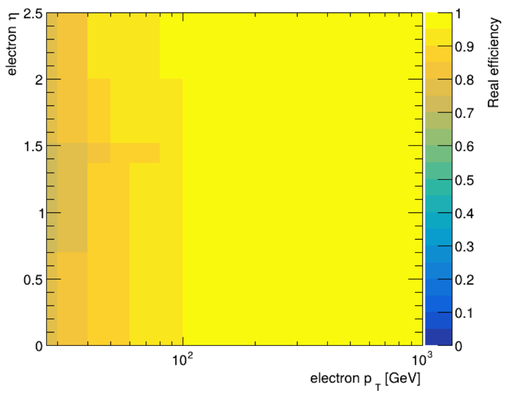
Figure 6.1: The number of loose and tight electrons (a) and muons (b) in the SR in the 2018 data-taking year.



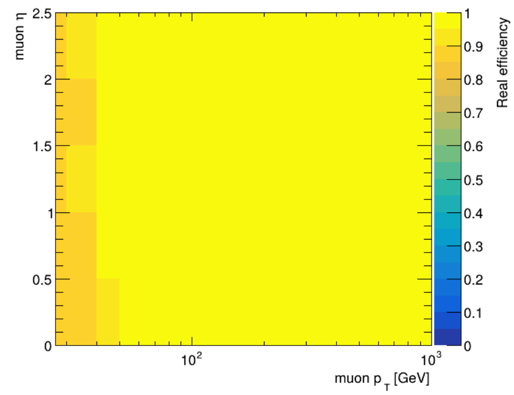
(a)



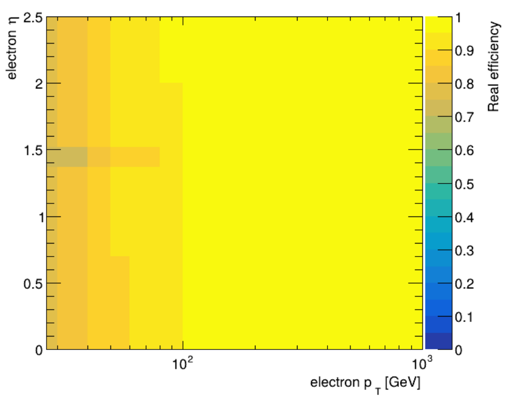
(d)



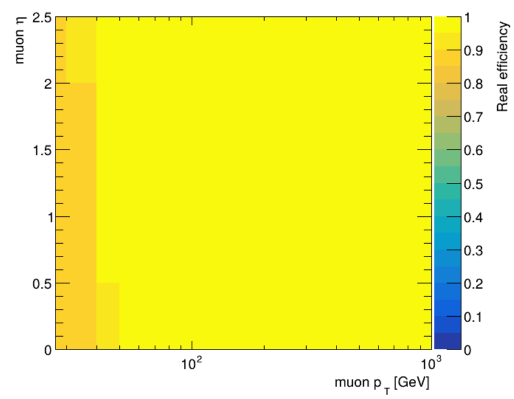
(b)



(e)

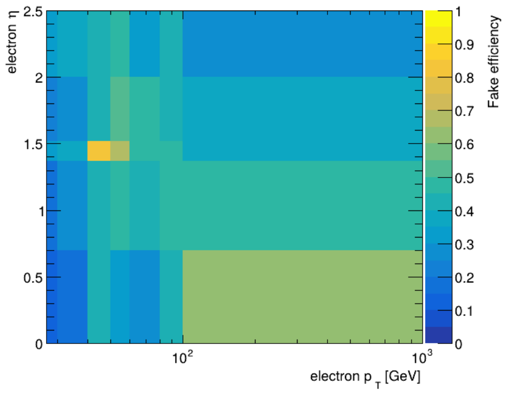


(c)

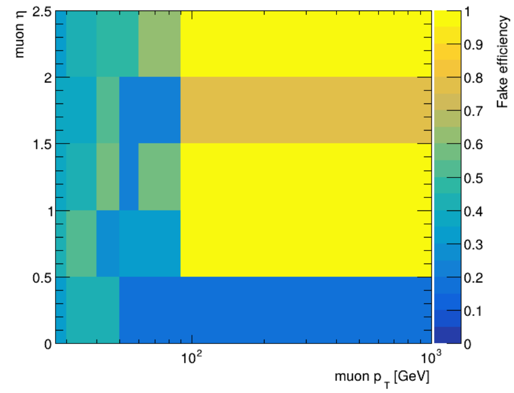


(f)

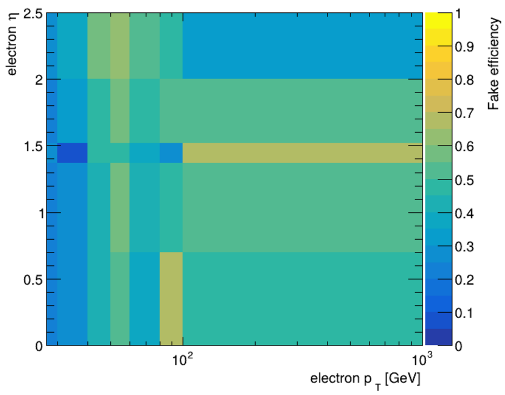
Figure 6.2: Real efficiency in the 2015 and 2016 data taking period for electron (a) and muon channel (d); 2017 for electron (b) and muon channel (e); 2018 for electron (c) and muon channel (f).



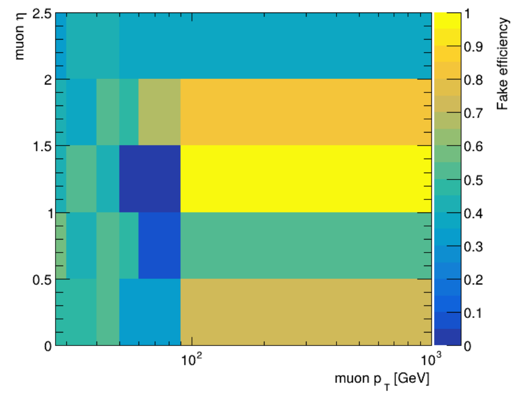
(a)



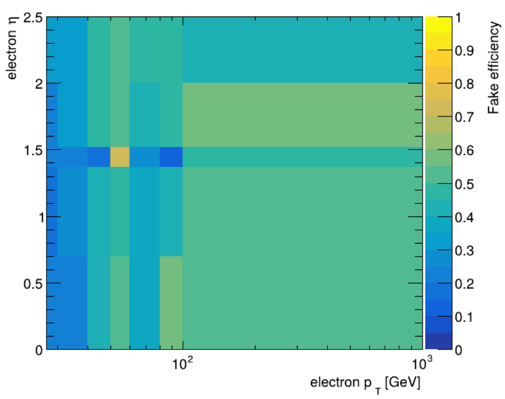
(d)



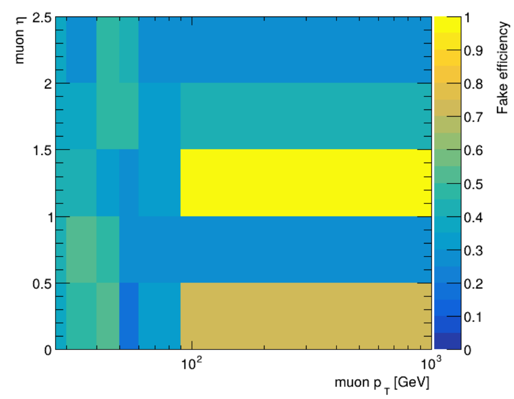
(b)



(e)

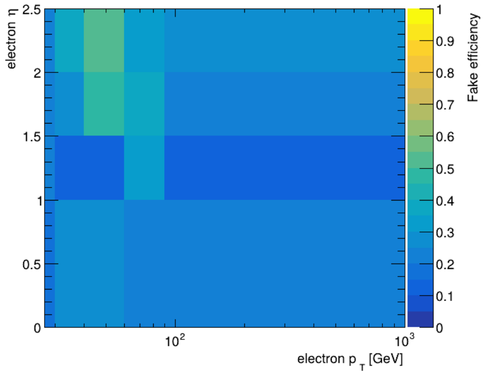


(c)

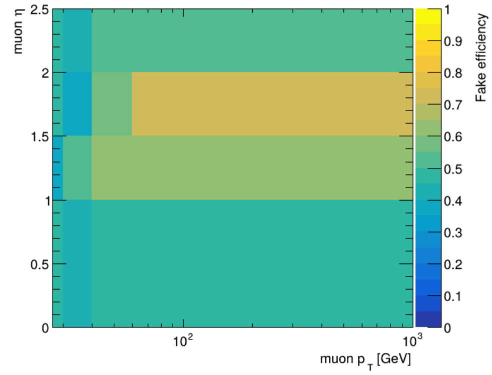


(f)

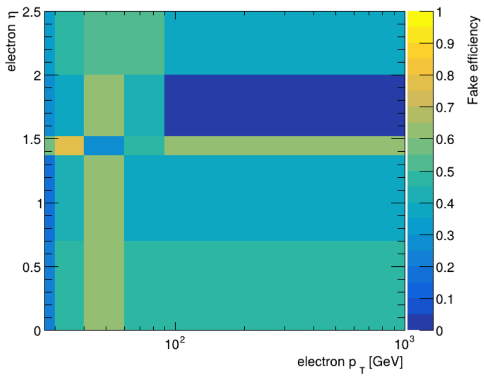
Figure 6.3: Signal region fake efficiency plots in the 2015 and 2016 data taking period for electron (a) and muon channel (d); 2017 for electron (b) and muon channel (e); 2018 for electron (c) and muon channel (f).



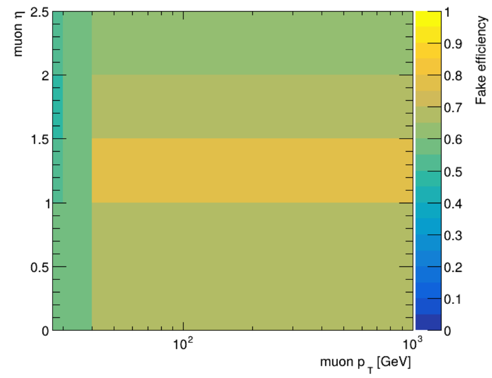
(a)



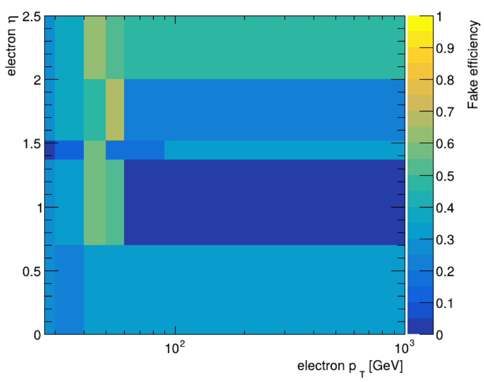
(d)



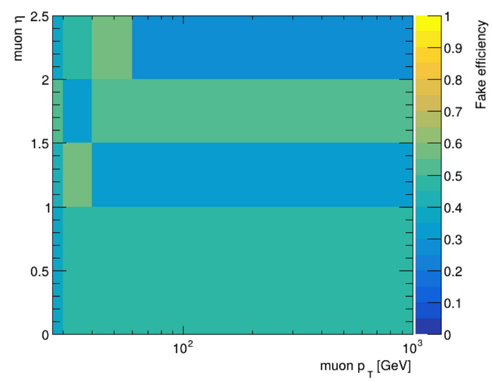
(b)



(e)



(c)



(f)

Figure 6.4: $t\bar{t}$ control region fake efficiency plots in the 2015 and 2016 data taking period for electron (a) and muon channel (d); 2017 for electron (b) and muon channel (e); 2018 for electron (c) and muon channel (f).

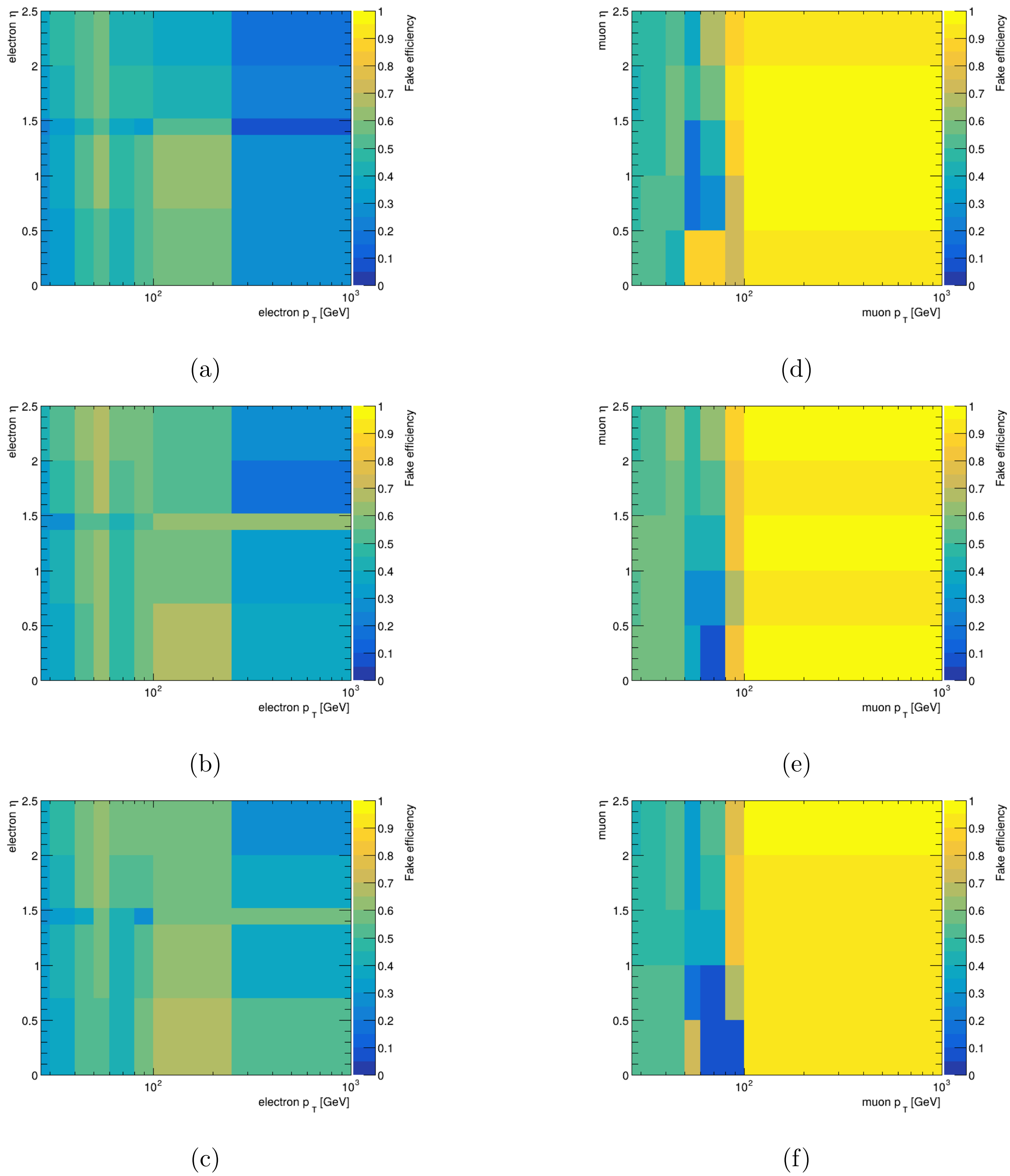
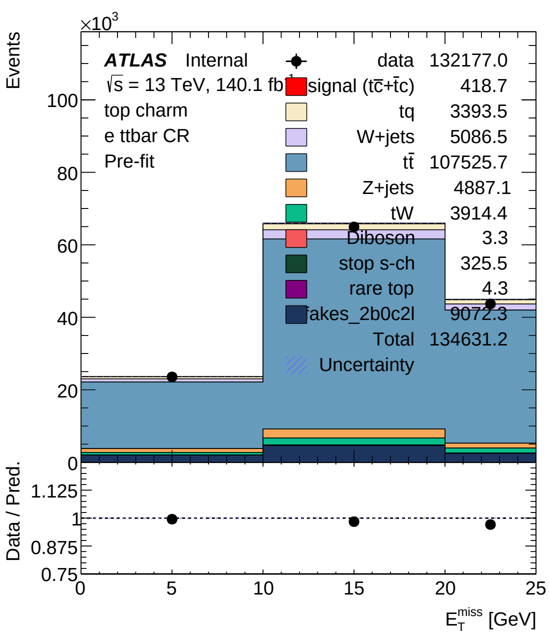


Figure 6.5: W+jets control region fake efficiency plots in the 2015 and 2016 data taking period for electron (a) and muon channel (d); 2017 for electron (b) and muon channel (e); 2018 for electron (c) and muon channel (f).

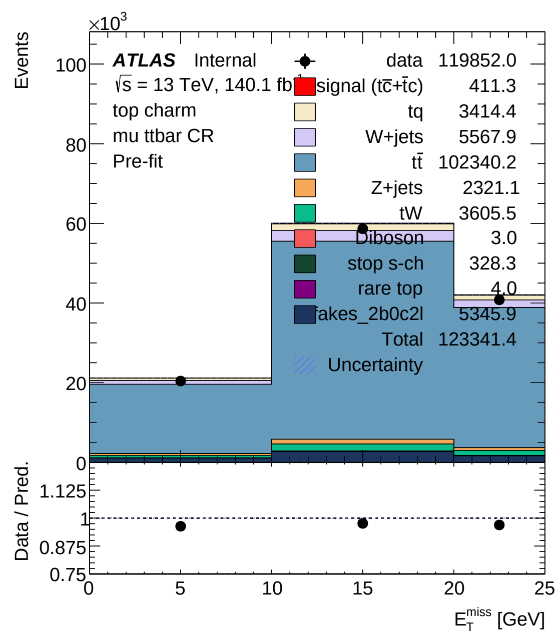
6.2 Fake Validation

The fake yield estimate is computed from the loose data sample, and the assumption is made that it extrapolates correctly to the tight sample in the $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ region. To validate this assumption, a closure test is performed by examining the data-to-MC ratio agreement using the kinematic variable, $E_{\mathrm{T}}^{\mathrm{miss}}$ for the tight sample, in the $t\bar{t}$ and W+jets control regions for $E_{\mathrm{T}}^{\mathrm{miss}} < 25$ and no MWT cut for electron and muon channels.

Figures 6.6 and 6.7 shows the validation plots for $t\bar{t}$ and $W^{\pm}$+jets control regions. From these plots, it can be seen that the data and MC agree within 5% for both the $t\bar{t}$ and $W^{\pm}$+jets control regions.

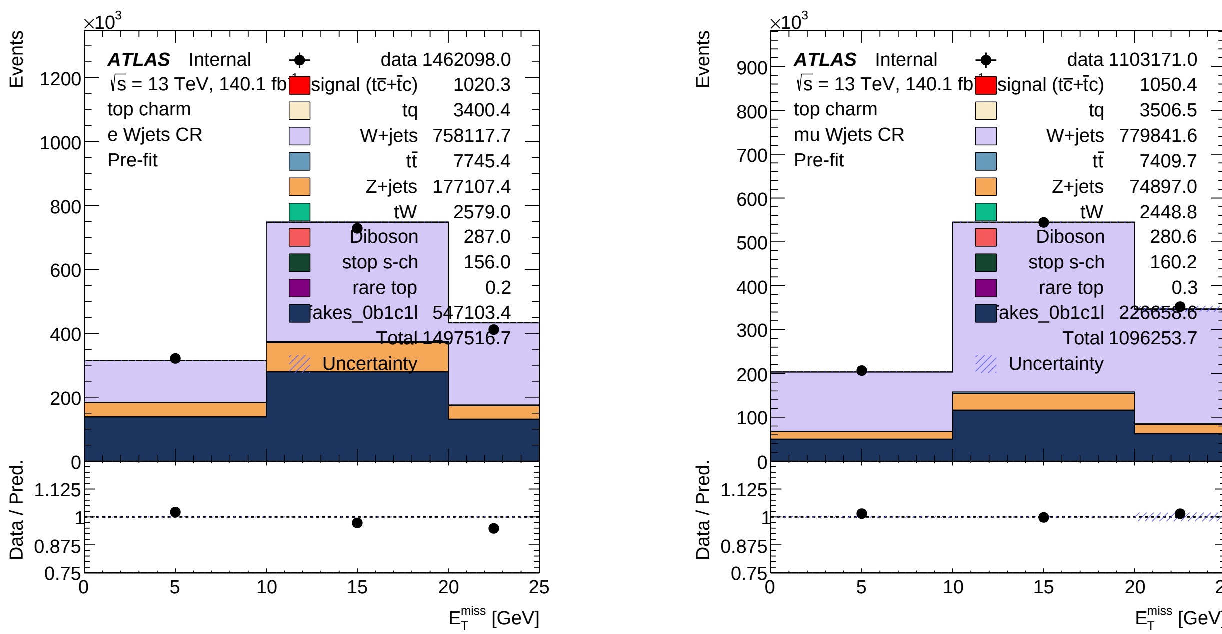


(a)



(b)

Figure 6.6: Fake validation plots in $t\bar{t}$ control region for (a) electron channel, (b) muon channel. Uncertainty is statistical only.



(a)

(b)

Figure 6.7: Fake validation plots in W+jets control region for (a) electron channel, (b) muon channel. Uncertainty is statistical only.

Chapter 7

Systematic Uncertainties

Systematic uncertainties refer to the inaccuracies that are not directly related to the statistics of the sample. They occur due to theoretical modeling of MC, data-taking conditions, detector effects, etc. Various sources of systematic uncertainty are taken into account in this analysis, which can affect both the normalization and the shape of the kinematic distributions. The effect of any given systematic uncertainty is applied consistently across all signal and control regions. Up-and-down variations are generated for each systematic and stored as histograms. They are treated as independent nuisance parameters, and measurements are done using statistical tools like TRExFitter [82].

7.1 Experimental Systematic Uncertainties

The systematic uncertainties arising from the reconstruction and identification of physics objects due to detector performance, calibration, and data-taking conditions are referred to as experimental systematic uncertainties. These include calibration uncertainties, such as jet energy scale and efficiency uncertainties arising from inaccurate measurements during the reconstruction and identification of physics objects. The luminosity and trigger uncertainties, as well as uncertainties related to pileup modeling and flavor tagging, are also taken into account in this analysis. These are discussed in detail in the following sub-sections.

7.1.1 Luminosity

The full Run 2 data sample, with an integrated luminosity of 140 ± 1.2 fb^{-1} collected from 2015 to 2018, has an uncertainty of 0.83%. The uncertainty is measured using the van der Meer beam (vdM) [83] separation scans and measurements from the luminosity-sensitive detector, LUCID-2 [84].

7.1.2 Pileup Reweighing

Additional interactions in proton-proton collisions, known as pileup, are modeled by overlaying minimum bias events (low-energy collisions) onto the hard scattering events. These simulated events are then reweighted to match the distribution of the number of proton-proton interactions per bunch crossing observed in the data. The uncertainty due to pileup reweighing is estimated by changing the amount of pileup in simulated events to generate alternative event weights [85], which are used as systematics in the analysis.

7.1.3 Lepton systematic uncertainties

The systematic uncertainties associated with the electron and muon triggers, identification, isolation, and reconstruction efficiency, as well as the combined muon (CB) momentum scale (MS) and resolution uncertainty, and the track-to-vertex-association (TTVA) impact-parameter selection uncertainty for muons, fall under lepton systematic uncertainties.

To correct for the small differences between the lepton trigger, identification, isolation, and reconstruction efficiency in the real data compared to the simulation, event-by-event scale factors are applied to simulated events, affecting both yield and shape. These scale factors are calculated using the tag-and-probe method. In this method, a well-known process, such as $Z \rightarrow l^+l^-$, is considered, where one lepton is considered a tag that passes the strict selection criteria and the other a probe. The mass of the tag and probe pair should fall inside a specified window around the Z-mass peak. By subtracting the background, efficiencies can be measured in both data and MC. The ratio of these efficiencies provides the scale factor. The uncertainties associated with this are propagated to the variables used in the analysis.

The uncertainty due to the energy or momentum scale is calculated by varying the correction factors applied to the simulated events by their uncertainties. Similarly, uncertainty due to energy or momentum resolution is evaluated by varying the amount of smearing applied to the simulated events according to their uncertainties [86].

7.1.4 Jet systematic uncertainties

The main systematic uncertainties associated with jets are the jet energy scale (JES) and the jet energy resolution (JER) uncertainties. The measurement of the jet energy scale and resolution is provided in the reference [87].

The full JES uncertainty consists of 125 individual terms or nuisance parameters derived from data-driven calibrations, pile-up effects, flavor dependence, etc. These parameters cover the effects of MC modeling issues, analysis selection cuts, event topology, and physics object uncertainties.

A smearing procedure is performed to ensure that the resolution of the jet energy scale in MC matches that in data. This procedure is done in the regions of jet p_T of MC, where the resolution in data is larger than in MC. The MC sample is smeared until its average resolution matches that of the data. In regions with lower data resolution, no smearing is performed. The JER uncertainties are propagated in the analysis by smearing jets depending on a Gaussian function with width σ_{smear}.

7.1.5 Flavor tagging uncertainties

Flavor tagging is explained in detail in section 3.5.1. As mentioned in the section 3.5.1, the latest flavor tagging algorithm used by ATLAS is GN2. The calibration of the b-tagging efficiency of GN2 is performed using a sample enriched in $t\bar{t}$ di-leptonic decay events. Jets are divided into 6 PCBT bins (pseudo-continuous bins of the GN2 discriminant) with b-tagging probability. The b-tagging probabilities are obtained via a maximum-likelihood approach, along with the determination of the jet flavor composition using different control regions. The scale factor is then defined in each PCBT bin and jet p_T bin. Systematic uncertainties are determined by the variation in MC that results in a shift in scale factors [88].

7.1.6 Missing transverse energy systematic uncertainties

The systematic uncertainties associated with E_T^{miss} depend on the composition of the hard terms and on the magnitude of the corresponding soft term (refer to section 3.6). The uncertainties due to hard terms are affected by the systematic uncertainty of the reconstructed physics object, which propagates to E_T^{miss}. In the case of the soft term, it is estimated by comparing data to MC for a well-known process such as $Z \rightarrow l^+l^-$, where E_T^{miss} is expected to be negligible. This expectation does not hold due to experimental inefficiencies. By examining the contribution of the soft term to the total E_T^{miss} in the simulation compared to the data, the difference contributes to the systematic uncertainties [89].

7.2 Theoretical Systematic Uncertainties

Theoretical systematic uncertainties arise from the theoretical modeling of the physics process, such as assumptions or approximations made during event generation and cross-section calculation. Common sources of these theoretical uncertainties are:

- PDF variations: Different PDF sets, such as NNPDF and PDF4LHC [90] variations, are chosen, which leads to a change in event yield and is taken as the uncertainty.
- Renormalization (μ_R) and Factorization (μ_F) scales: These are usually varied by factors of 0.5 and 2 with respect to the nominal to estimate the uncertainty. Different combinations of μ_R and μ_F are considered, say, $\mu_R, \mu_F = (0.5, 0.5), (0.5, 1.0), (1.0, 0.5), (1.0, 2.0), (2.0, 1.0), (2.0, 2.0)$ and the envelope of these variations results in the scale uncertainty.
- Initial state radiation (ISR): α_S is varied for ISR in the A14 tune using weights Var3cUp and Var3cDown to estimate the uncertainty.

Alternative samples of $t\bar{t}$ and $W^{\pm}$+jets are also used as systematic uncertainty to provide shape variation to the respective samples. For $t\bar{t}$, four alternative samples are considered by varying the Powheg hard-emission damping scale, h_{damp}, and the shower veto parameter (pThard), and using the recoil-to-top scheme and Herwig7 [91] for the parton shower. For W+jets, alternate samples are generated using MadGraph. Furthermore, an additional shape variation is applied to the W+jets sample to account for the mis-modeling observed as discussed in Chapter 5.

7.3 Fake Systematic Uncertainties

This section provides an overview of the systematic uncertainties associated with fakes. One of the main sources of uncertainty is the E_T^{miss} bias, introduced by the E_T^{miss} cut. Since there are underlying jets around fake leptons, they can affect both E_T^{miss} and p_T. Hence, fake efficiency depends on E_T^{miss}. Therefore, fake efficiency is calculated in the E_T^{miss} region between 15 and 25 using the matrix method mentioned in Chapter 6 and the sample obtained is used as a systematic variation to nominal fake yield.

The fake efficiency can also depend on the d_0 cut, as it can alter the composition of fakes. Hence, the efficiencies are calculated with different d_0 cuts: 2 and 4 for muons (the default value is 3), and 4 and 6 for electrons (the default value is 5). The fake samples obtained are thus used as systematic uncertainty to the nominal.

Estimation of fakes involves fake efficiency calculation where prompt lepton contributions are subtracted from data as described in Chapter 6. MC modeling of prompt leptons is not perfect, so there can be inaccuracies in the fake yield estimation. Hence, a flat systematic uncertainty can be applied to cover the possible inaccuracy.

7.4 Treatment of Systematic Uncertainties

Symmetrization: Systematic uncertainties are treated as nuisance parameters during the fit procedure. There is a Gaussian constraint/prior term added to each of the systematics given by:

$$\pi(\theta_i) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\theta_i^2}, \tag{7.1}$$

where θ is the nuisance parameter. The nominal is set to be at the mean of the Gaussian distribution, and the systematic uncertainties are implemented as $\pm 1\sigma$ variations. Since the uncertainties are not always symmetric, symmetrization is implemented. Symmetrization methods used in this analysis are ONESIDED and TWOSIDED symmetrization. ONESIDED symmetrization is used when only one variation is provided; say the up variation and the down variation will be added as a mirrored version. In TWOSIDED symmetrization, (up-down)/2 variation is calculated bin-by-bin and is used as up variation and then mirrored to down variation.

Smoothing: Smoothing is applied to the systematic uncertainties to reduce statistical fluctuations that can cause variations in shape that are not physical. For the analysis, a smoothing option of 40 [92] is set to treat the systematic uncertainty as correlated with the nominal.

Pruning: Pruning is used to remove systematic uncertainties which has a negligible effect on the result. The threshold for pruning for this analysis is 0.5%.

Chapter 8

Statistical Analysis and Results

The cross-sectional measurement of the single top + charm can be evaluated using the signal strength μ given by:

$$\mu = \frac{\sigma_{measured}}{\sigma_{theory}}, \tag{8.1}$$

where, σ_{theory} is the SM or theory cross section and $\sigma_{measured}$ is the measured cross section. A simultaneously binned likelihood fit is performed on the BDT distribution in the SR and the E_T^{miss} distribution in the CRs to determine signal strength. The data in the signal region remain blinded for the analysis until it is fully finalized and validated, as per the ATLAS recommendation to avoid potential bias. At the same time, data in the CRs are already unblinded for the analysis and help in background modeling and validation. The fitting procedure is done using the TRExFitter framework.

8.1 The Likelihood function

Consider a set of variables in a data sample, $x_1, x_2, x_3, ...x_n$. They are distributed according to a probability distribution function (pdf), $f(x, \theta)$, where θ describes external or nuisance parameters represented by systematic uncertainties. It is assumed that each measurement is independent and hence the probability of the dataset, $x_1, x_2, x_3....x_n$ is the product of individual probabilities and this product is called the likelihood function $\mathcal{L}(\theta)$ [93], given by:

$$\mathcal{L}(\theta) = P(x_1, x_2, \dots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{8.2}$$

The likelihood [94] is defined as the probability of observing n events of a dataset, given the expected number of signal and background events. For a histogram with N bins, the expected number of events in the i^{th} bin can be written as:

$$E[n_i] = \mu s_i + b_i, \tag{8.3}$$

where s_i is the number of signal events and b_i is the number of background events in the i^{th} bin, with:

$$s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x; \theta_s)\, dx, \tag{8.4}$$

$$b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x; \theta_b)\, dx, \tag{8.5}$$

where s_{tot} and b_{tot} are the total expected signal and background events, $f_s(x; \theta_s)$ and $f_b(x; \theta_b)$ are the pdfs of the variable x for signal and background events and θ_s and θ_b represent the nuisance parameters. The parameter of interest (POI) μ determines the strength of the signal, and $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ corresponds to the signal+background hypothesis.

Additional measurements in the CR with mainly background events can also be made to constrain the nuisance parameters. For this, consider a histogram with M bins with a set of values $m = (m_1, m_2, m_3, ...m_n)$ corresponding to the number of events in each of the M bins. The expectation value of m_i is given by:

$$E[m_i] = u_i(\theta), \tag{8.6}$$

where u_i is a calculable quantity and it depends on θ. The likelihood function is thus the product of Poisson probabilities for all the bins and is given by:

$$\mathcal{L}(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}. \tag{8.7}$$

The profile likelihood ratio is defined to test a hypothesized value of μ, given by:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \tag{8.8}$$

The numerator describes the likelihood when μ is fixed and the nuisance parameter θ is varied to maximize $\mathcal{L}$. The denominator describes the maximum likelihood when μ and θ can vary freely. From the equation 8.8, it can be seen that $0 \le \lambda \le 1$ with $\lambda \approx 1$ shows good agreement between data and the hypothesized value of μ. As the basis of a statistical test, a test statistic t_μ can be defined:

$$t_\mu = -2 \ln \lambda(\mu) \tag{8.9}$$

Higher values of t_μ indicate disagreement between data and μ. To evaluate the level of disagreement, the p value can be computed as follows:

$$p_\mu = \int_{t_{\mu,\mathrm{obs}}}^{\infty} f(t_\mu \mid \mu)\, dt_\mu, \tag{8.10}$$

where $t_{\mu,\text{obs}}$ is the test statistic value observed from data, and $f(t_\mu \mid \mu)$ is the pdf of t_μ at μ. The μ values are rejected if the p value is below a certain threshold α (usually 0.05). This can happen in two ways: if the estimated signal strength $\hat{\mu}$ is found to be greater or less than the hypothesized value μ.

An alternative test statistic $\tilde{t}_\mu$ is defined for $\mu \geq 0$. A special case of this alternative test statistic is used to test $\mu = 0$, where it is assumed that $\mu \geq 0$. Rejecting $\mu = 0$, a background-only hypothesis leads to the discovery of a new signal. In this case, $\tilde{t}_\mu = q_0$ and

$$q_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0, \end{cases} \tag{8.11}$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$. The p value is thus given by:

$$p_0 = \int_{t_{q_0,\text{obs}}}^{\infty} f(q_0 \mid 0)\, dq_0, \tag{8.12}$$

where $f(q_0 \mid 0)$ is the pdf of q_0 under background only hypothesis. The significance can be derived from the p value, and it is given by:

$$Z = \Phi^{-1}(1 - p_0), \tag{8.13}$$

where Φ is the cumulative distribution of the standard (zero mean, unit variance) Gaussian. The significance threshold for discovery is 5σ with p_0 value 2.87×10^{-7} and evidence is 3σ with p_0 value 1.35×10^{-3}.

8.2 Expected Results

A binned likelihood fit is performed simultaneously on the SR and CRs. Since the data in the CR is blinded, a pseudo-data set called Asimov data is used in the SR. Any data set for which the parameter estimators provide the true values is called an Asimov data set [95]. The expected results are obtained using mixed data and MC fit. In this technique, a background-only fit is performed in the CR using real data. This initial fit gives a set of fitted values for all nuisance parameters (or at least for all the nuisance parameters that affect the background prediction in the control regions). The next step involves using these fitted nuisance-parameter values to construct modified pseudo-data in the SR, where the real data is not used. Finally, the fit is performed on this mixed data set, where real data are used in the CRs and modified pseudo-data in the SR.

The fit includes free-floating parameters, known as normalization factors (NFs), for the dominant backgrounds, $t\bar{t}$ and $W^{\pm}$+jets, as well as nuisance parameters associated with the systematic uncertainties. The post-fit values of these parameters are extrapolated to all analysis regions, including the validation regions. The validation regions are not used in the fitting procedure, but they are used to validate the results. The per-bin statistical uncertainties known as the γ parameters are also included. They are uncorrelated between regions and bins.

A pull plot is used to assess the nuisance parameter θ after the fitting procedure. A pull is defined as $\frac{\hat{\theta}-\theta_0}{\Delta\theta}$, where $\hat{\theta}$ is the post-fit value of the uncertainty, θ_0 is the nominal or pre-fit value, and $\Delta\theta$ is the pre-fit uncertainty. A pull of ± 1 means a 1σ shift from nominal. A pull near zero indicates that the data is consistent with the prediction, and it is considered significant if it exceeds 1σ. A significant pull might point to mis-modeling. The uncertainty on the nuisance parameters also provides important information about how they are treated in a fit. A nuisance parameter is constrained if post-fit uncertainty is $<< 1$. The constraint arises because the data contain sufficient information to reduce the post-fit uncertainty. For example, if the prior on the nuisance parameter is 15% during the fit, and the data indicate that the uncertainty is 3% rather than 15%, the post-fit uncertainty decreases, and a constrained pull plot is obtained. The pull plots of the various nuisance parameters used in the fitting procedure are shown in Figures 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, and 8.9.

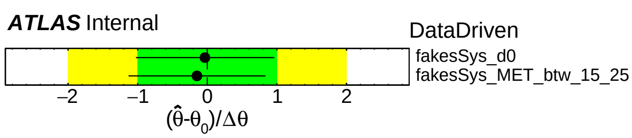


Figure 8.1: Datadriven Nuisance Parameters

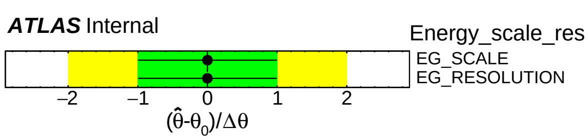


Figure 8.2: Energy Scale and Resolution Nuisance Parameters

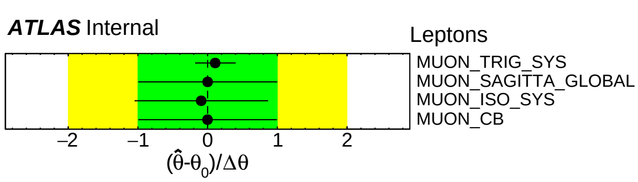


Figure 8.3: Leptons Nuisance Parameters

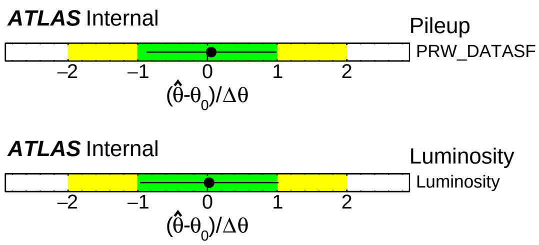


Figure 8.4: Pileup and Luminosity Nuisance Parameters

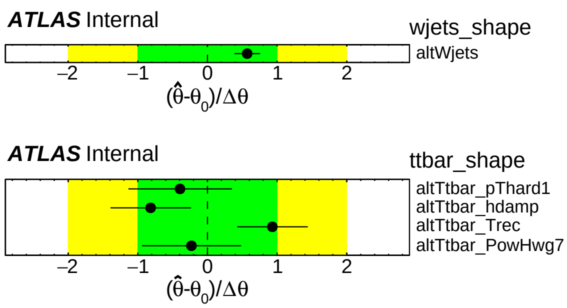


Figure 8.5: $t\bar{t}$ and $W^{\pm}$+jets shape variation Nuisance Parameters

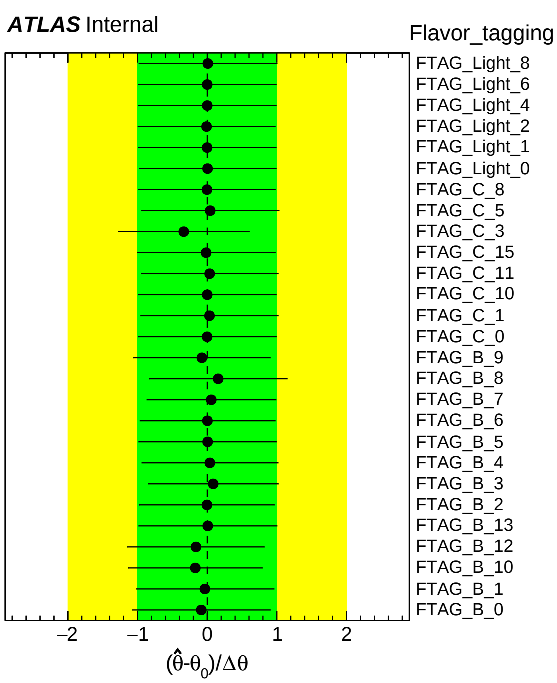


Figure 8.6: Flavor Tagging Nuisance Parameters

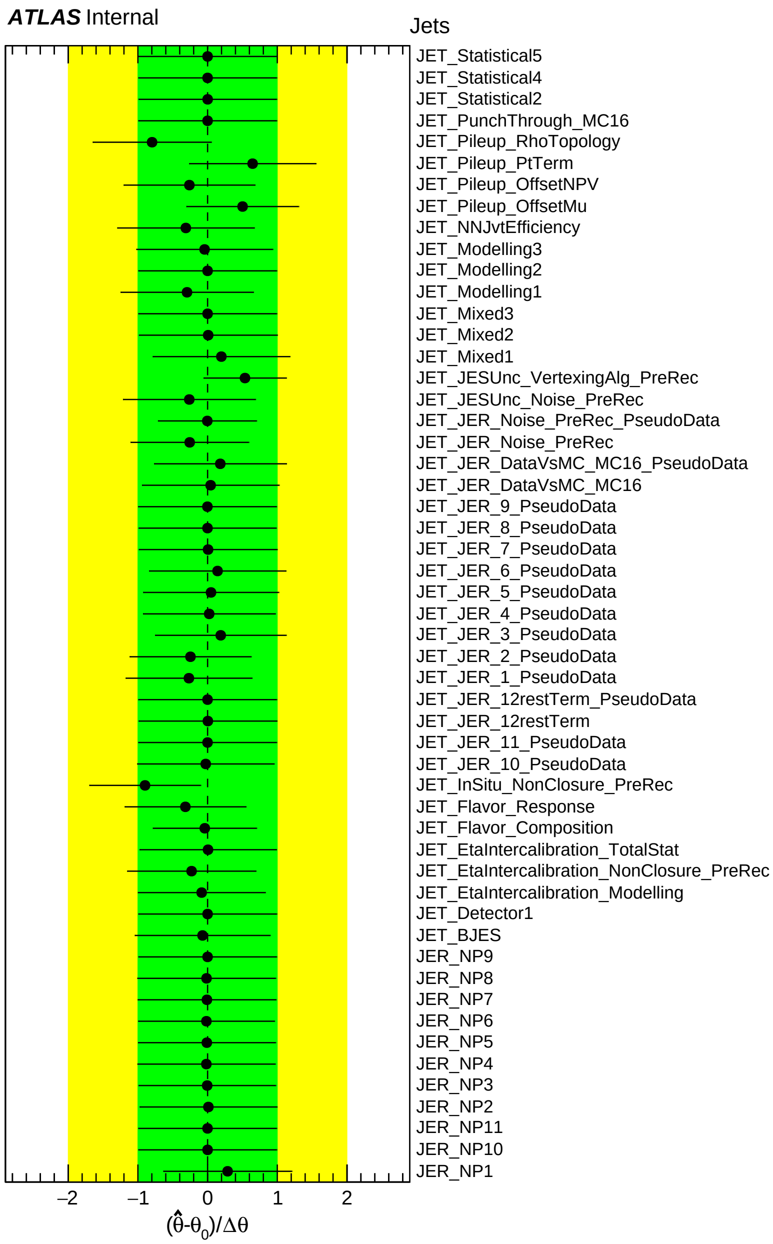


Figure 8.7: Jets Nuisance Parameters

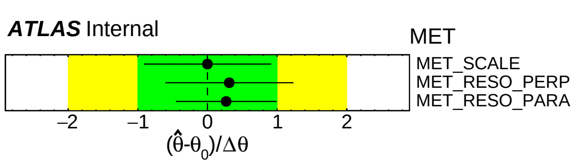


Figure 8.8: E_T^{miss} Nuisance Parameters

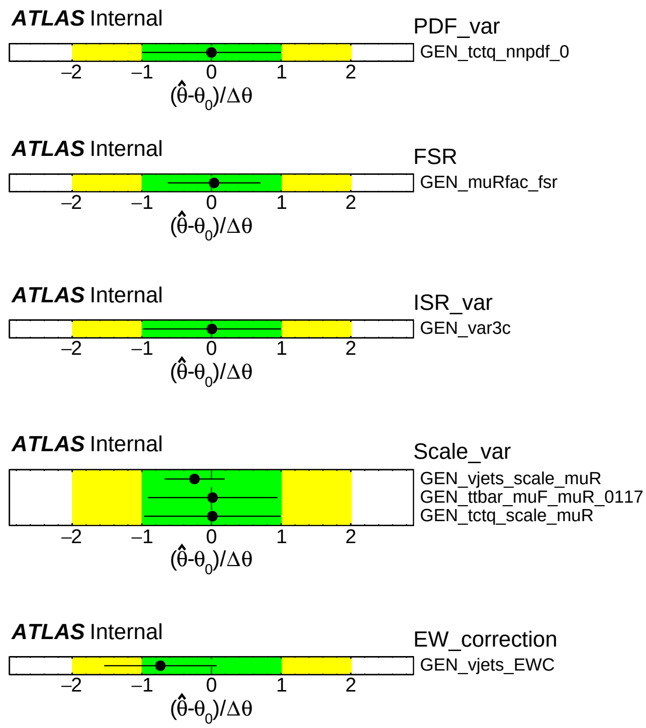


Figure 8.9: Theoretical modeling Nuisance Parameters

From the pull plots, it can be seen that there are no significant pulls and all are within 1σ. Meanwhile, a few nuisance parameters are constrained, MUON_TRIG_SYS, the systematic uncertainty associated with muon trigger efficiency, as shown in Figure 8.3, altWjets, the systematic uncertainty associated with the $W^{\pm}$+jets shape as shown in Figure 8.5, and Gen_vjets_scale_muR, the systematic uncertainty associated with the renormalization scales for V+jets as shown in Figure 8.9. This is because the data has enough information to reduce the post-fit uncertainty of these nuisance parameters. A similar plot is also made for the γ parameters associated with the MC statistical uncertainties shown in Figure 8.10. All of the γ parameters are centered around one with smaller error bars and are consistent with the data.

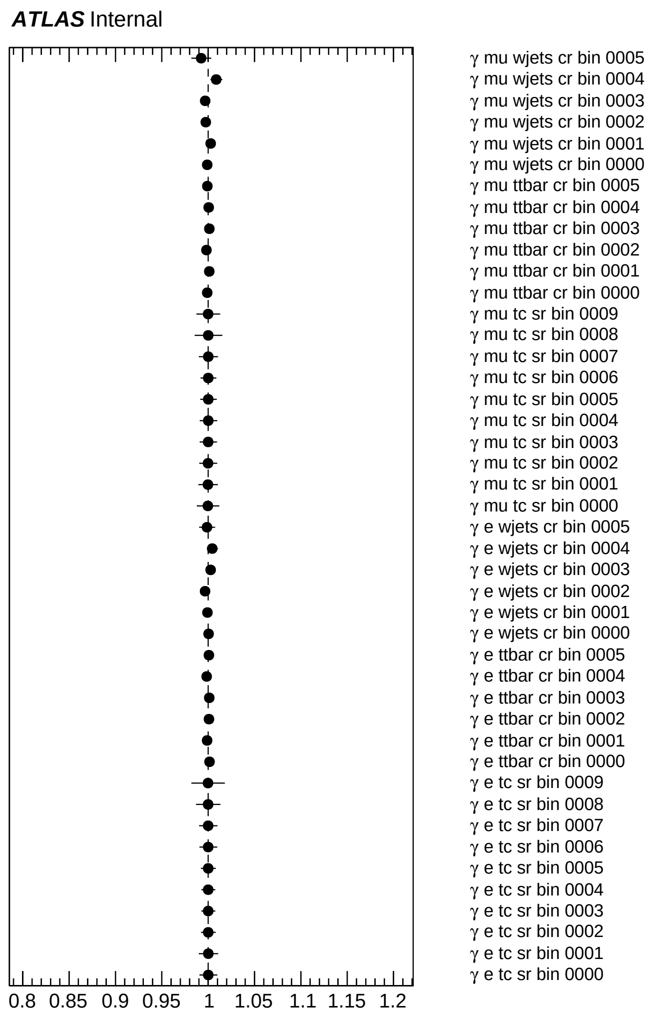


Figure 8.10: The post-fit γ factors.

The expected best fit value of signal strength for this analysis, μ_{tc}, is obtained using the binned BDT output distribution in the SR and E_T^{miss} distribution in the CRs. The post-fit normalization factors for the $t\bar{t}$ and $W^{\pm}$+jets, as well as the expected signal strength, are shown in Figure 8.11. The best fit value of μ_{tc} is 1, since we use the Asimov data in the signal region. The error or the uncertainty on μ_{tc} determines how well we can measure the single top + charm under the signal+background hypothesis. In this fit, μ can be measured with an expected uncertainty of 6%, indicating a high sensitivity to the signal process.

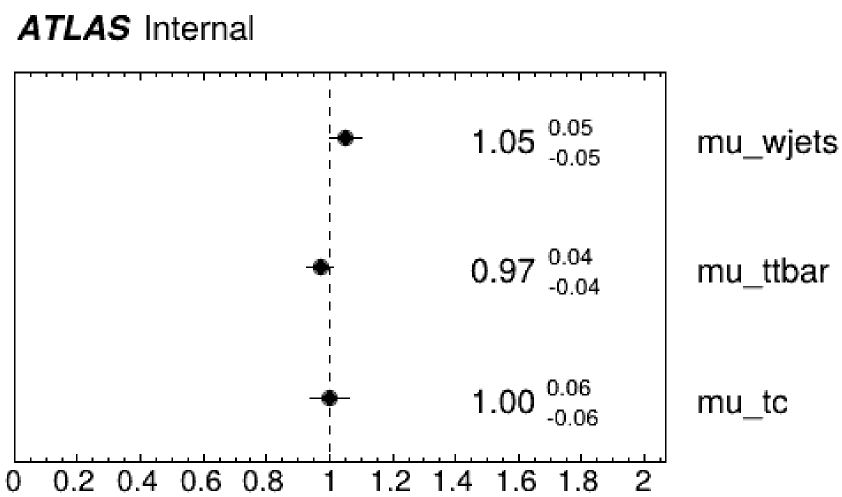
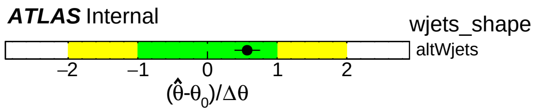
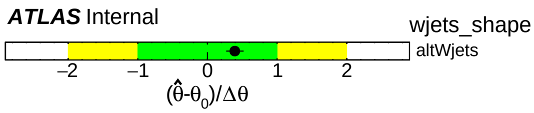


Figure 8.11: The post fit normalization factors for $t\bar{t}$ and $W^{\pm}$+jets and expected signal strength μ_{tc} with error.

The shape variation applied to the $W^{\pm}$+jets samples has been doubled because of possible mis-modeling in the $W^{\pm}$+jets samples as mentioned in Chapter 7. Figure 8.12 shows how the constraint changes when you triple the shape variation of $W^{\pm}$+jets. The pull decreased, and the constraint increased. It has also been observed that there is less than 0.01% difference between the μ_{tc} uncertainty because of these changes. Similarly, as discussed in Chapter 7, by varying the flat overall systematic uncertainty that can be applied to fakes, say from 40% to 0%, the percentage difference between uncertainty on μ_{tc} is only less than 0.5%.



(a)



(b)

Figure 8.12: Change in constraint in the $W^{\pm}$+jets shape variation pull plots (a) doubled (total error on μ_{tc}, 0.0616621) (b) tripled (total error on μ_{tc}, 0.616541)

The impact [96] of a given nuisance parameter on μ_{tc} is estimated using the covariance matrix. The impact I is defined as:

$$I = \frac{\sigma_a^{\text{total}} C_{a,n}}{\sigma_n}, \tag{8.14}$$

where σ_a^{total} is the total uncertainty on the POI a, in this analysis, μ_{tc}, $C_{a,n}$ is the correlation between μ_{tc} and the nuisance parameter and σ_n is the uncertainty on the nuisance parameter. Table 8.1 shows the impact of several uncertainty categories on the POI:

- TOT_ERROR is the total uncertainty on the POI.
- SYST_ERROR is the combined uncertainty of all the nuisance parameters.
- MCSTAT_ERROR represents the MC statistical uncertainty, the γ parameters.
- TOTSYST_ERROR is the total systematic uncertainty, which is SYST_ERROR and MCSTAT_ERROR added in quadrature.
- STAT_ERROR is the statistical uncertainty obtained by subtracting TOTSYST_ERROR from TOT_ERROR in quadrature.

- NF_ERROR is the uncertainty from all NormFactors.

- STAT_ERR_NO_NF is calculated by subtracting NF_ERROR from STAT_ERROR.

The systematic and MC statistical uncertainties appear to have equal impact on μ_{tc}, with negligible effect from the normalization factor uncertainty. The fit is not limited by statistical or systematic uncertainties since they contribute almost equally. The impacts of different sources of systematic uncertainties on the POI are also given in Table 8.2. The flavor tagging systematic uncertainties have the most impact on μ_{tc}. This is expected because SR and CRs are defined based on the number of flavor-tagged jets. The breakdown of the impact of systematic and statistical errors is shown in the ranking plot in Figure 8.13, which displays the top 15 parameters. The plot includes flavor-tagging uncertainties, jet uncertainties, and the theoretical uncertainty associated with FSR, which are the top 3 sources of systematic uncertainty affecting μ_{tc}.

Uncertainty Type	
TOT_ERROR	0.0616621
STAT_ERROR	0.0350251
SYST_ERROR	0.0431865
MCSTAT_ERROR	0.0266529
TOTSYST_ERROR	0.050749
STAT_ERR_NO_NF	0.0350139
NF_ERROR	0.000882806

Table 8.1: Impact of different uncertainties on the POI

Systematic Source	
DataDriven	0.000703107
EW_correction	0.00126708
Energy_scale_res	0.00120528
FSR	0.0199307
Flavor_tagging	0.021266
ISR_var	0.00220934
Jets	0.0164367
Leptons	0.00179372
Luminosity	0.00939129
E_T^{miss}	0.00906983
PDF_var	0.0134216
Pileup	0.0145924
Scale_var	0.0116952
ttbar_shape	0.00560131
wjets_shape	0.00144824

Table 8.2: Impact of several systematic uncertainties on the POI

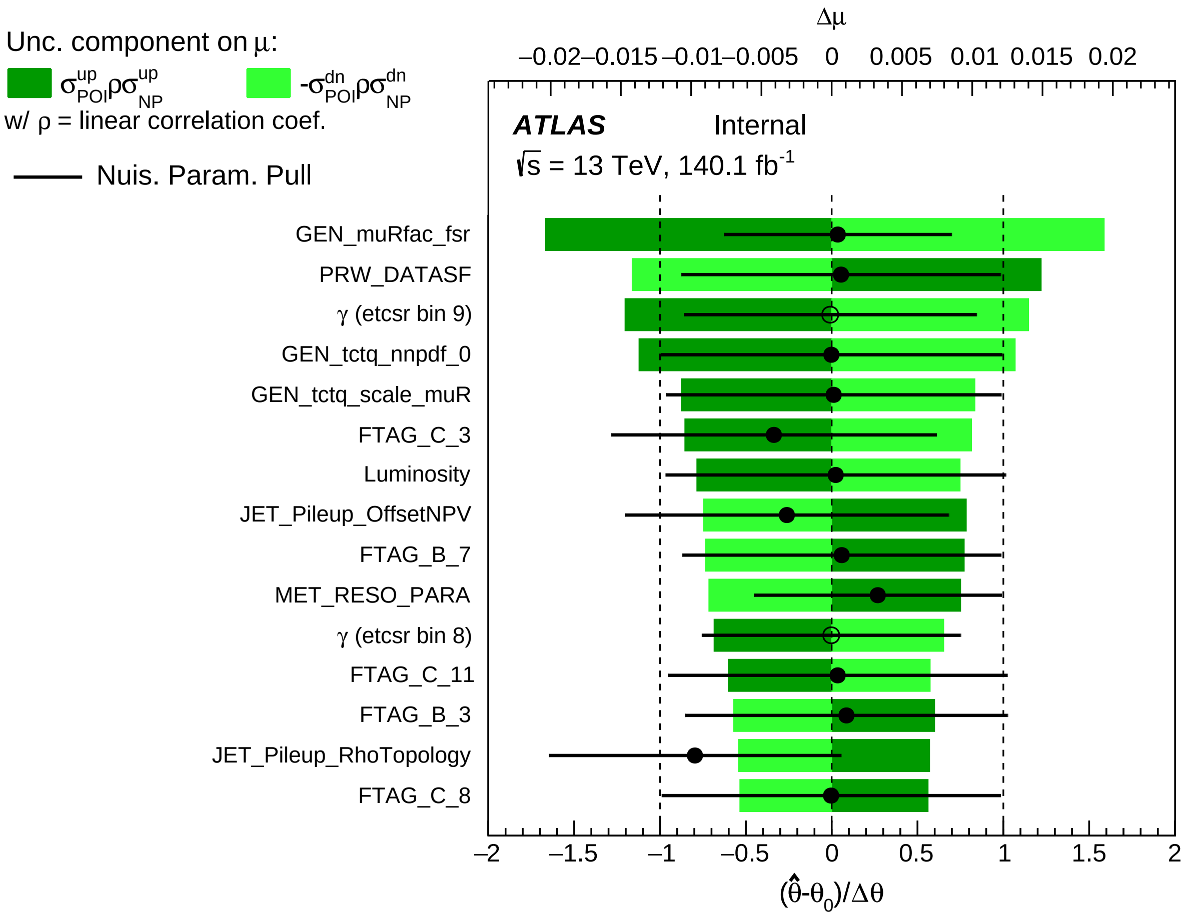
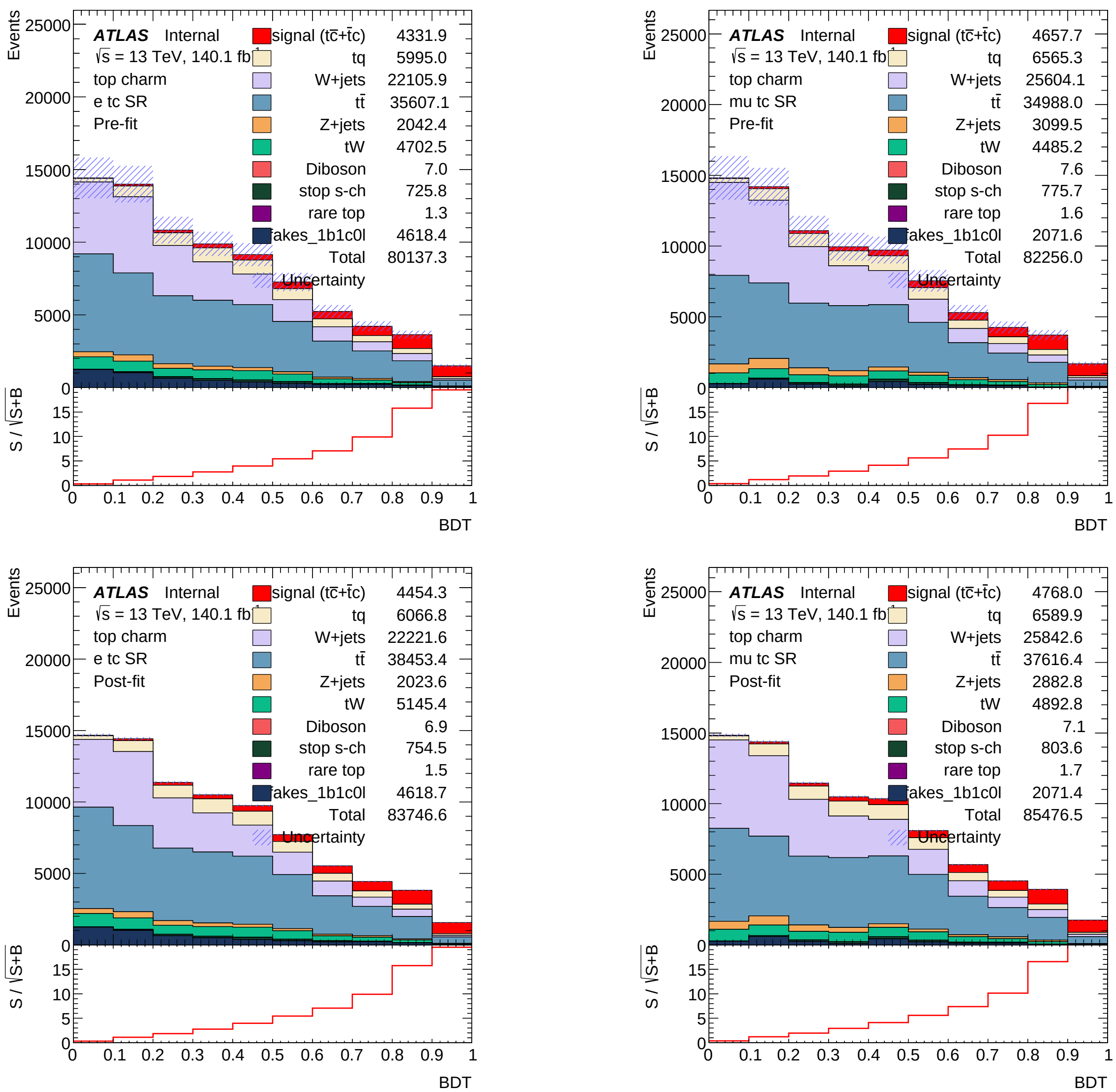


Figure 8.13: Impact of top 15 parameters on the POI

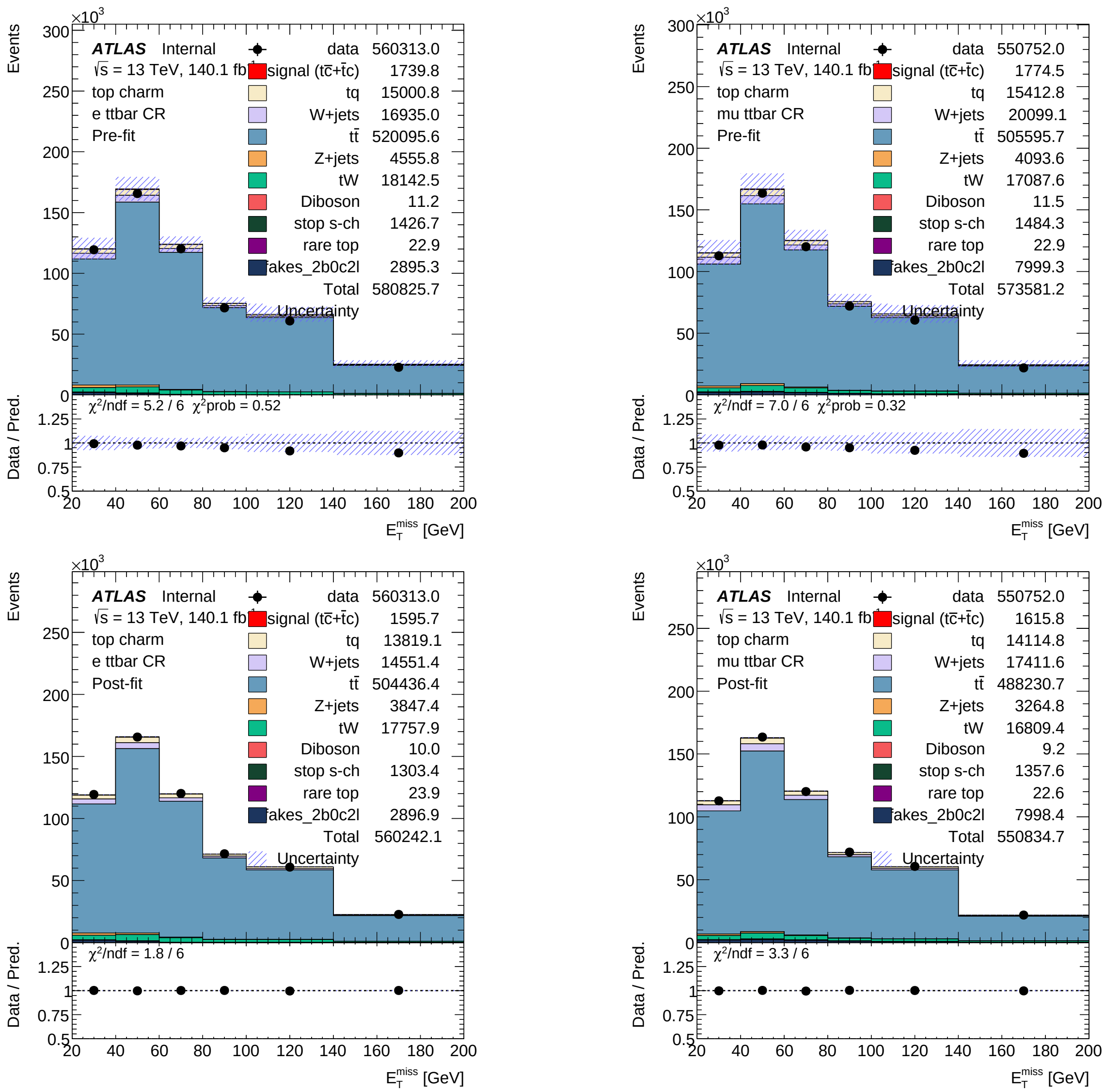
Figure 8.14, 8.15, 8.16 shows the pre-fit and post-fit plots for all the SR and CRs in both the electron and muon channels. The normalization factors of $t\bar{t}$ and $W^{\pm}$+jets from Figure 8.11 are very close to unity, which shows good agreement between data and MC predictions. The uncertainties on both the normalization factors are small, indicating that the control region constraints it well. Additionally low χ^2/ndf values in CR post-fit plots suggest good consistency between data and MC. The best fit parameter values are extrapolated to the $t\bar{t}$ and $W^{\pm}$+jets validation regions. The $t\bar{t}$ validation region contains events with exactly one lepton (e, μ) with $p_T > 27$ GeV, two b-tagged jets and one c-tagged jet with $p_T > 20$ GeV, zero light jets, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV and $MWT > 40$ GeV. The $W^{\pm}$+jets validation region contains events with exactly one lepton (e, μ) with $p_T > 27$ GeV, one c-tagged jet, and two light jet with $p_T > 20$ GeV, zero b-tagged jets, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, and $MWT > 40$ GeV. The validation plots are shown in Figure 8.17 and Figure 8.18. The first few bins of the $t\bar{t}$ validation region are blinded because they contain significant signal events. The χ^2 probability value of the $W^{\pm}$+jets validation region indicates good agreement between data and MC.

Thus, the analysis presented in this thesis provides the first expected observation of a single top with an associated charm, with an expected uncertainty of 6% on the signal strength parameter μ_{tc}. This indicates a high sensitivity to the measurement of the process once the SR is unblinded. The expected significance, calculated to quantify sensitivity, exceeds 17 and is well above the 5σ discovery threshold.



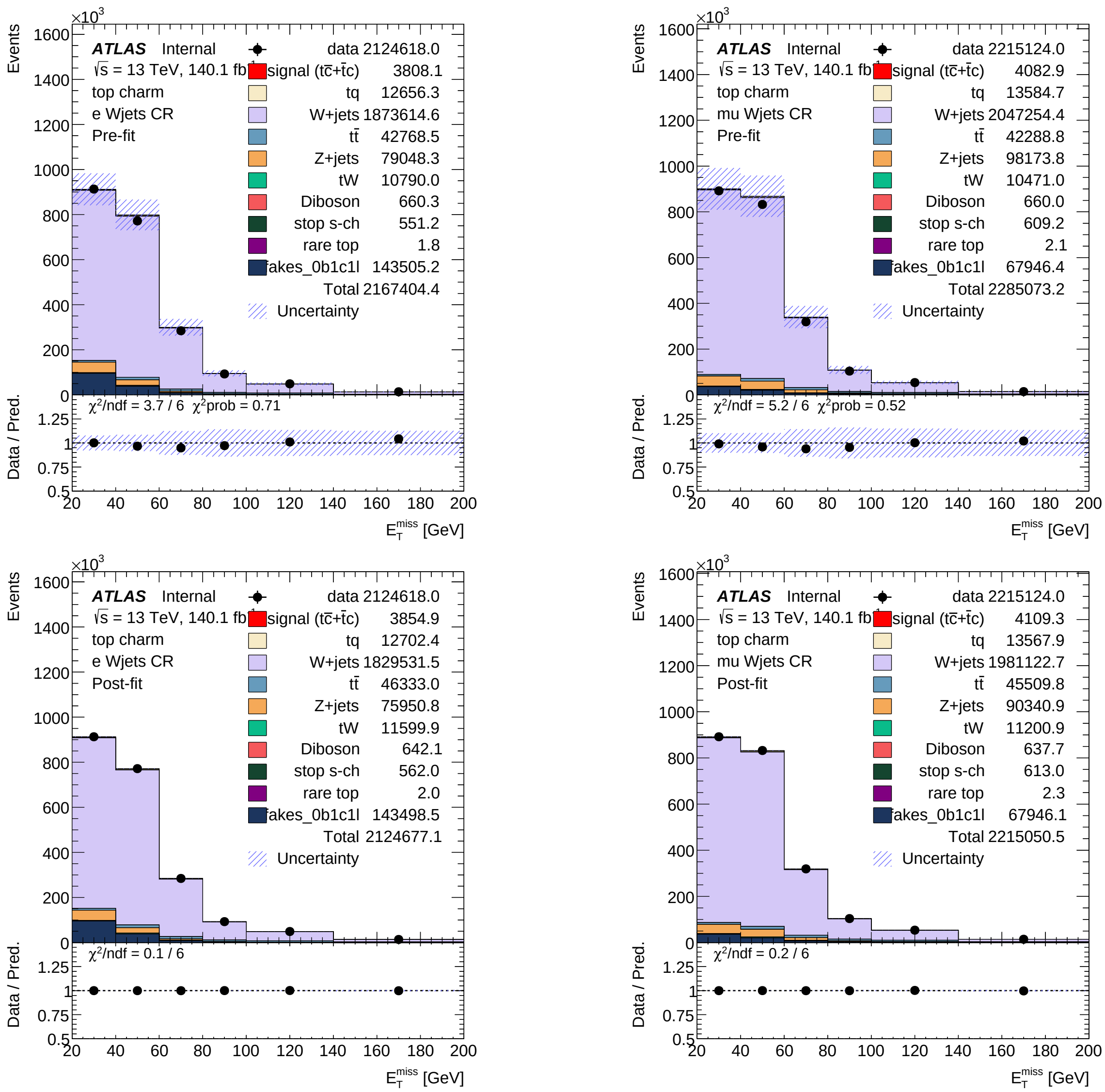
(a) (b)

Figure 8.14: (a) Pre-fit and (b) post-fit BDT distribution in signal region



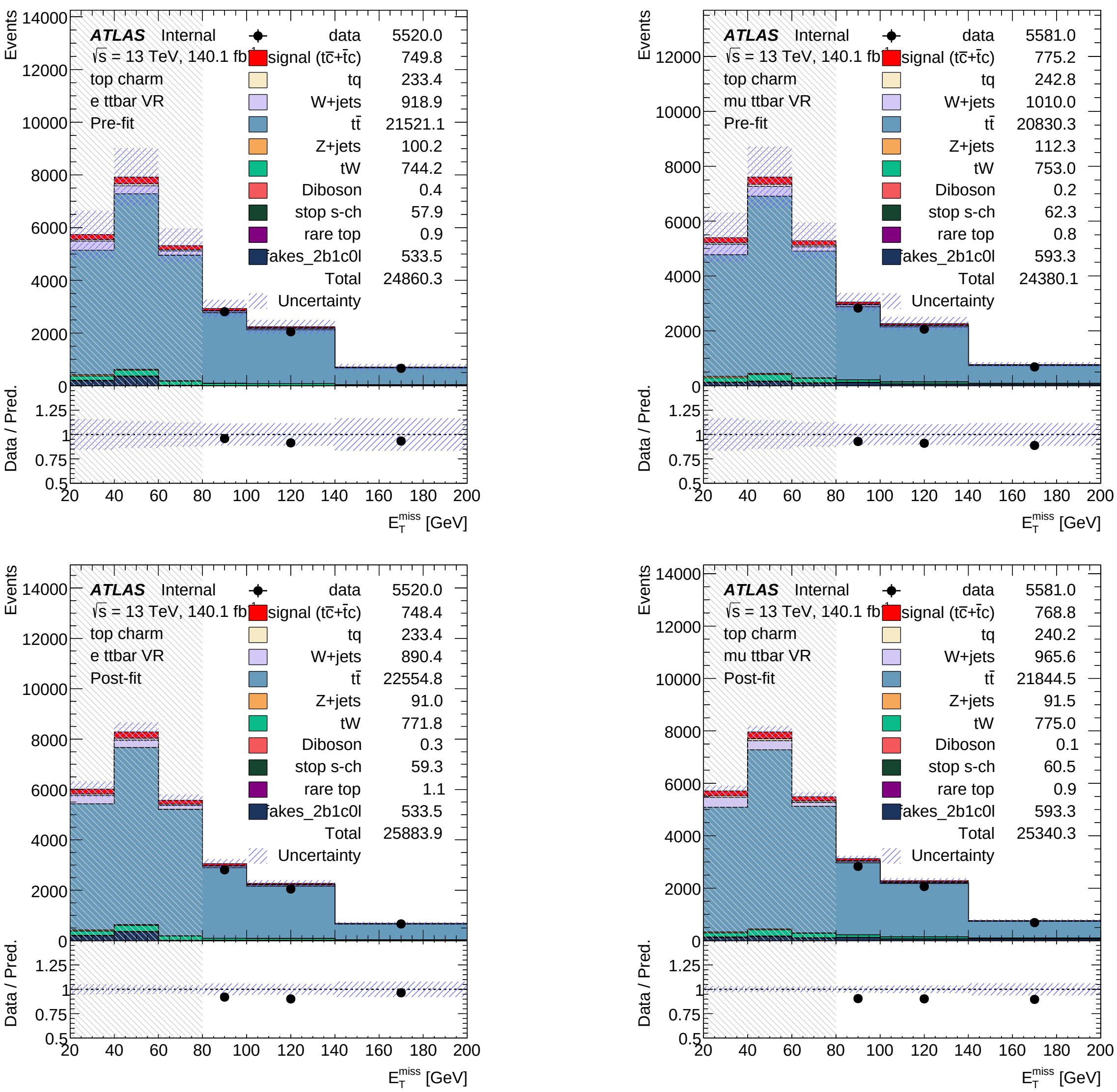
(a) (b)

Figure 8.15: (a) Pre-fit and (b) post-fit distribution of E_T^{miss} in $t\bar{t}$ control region



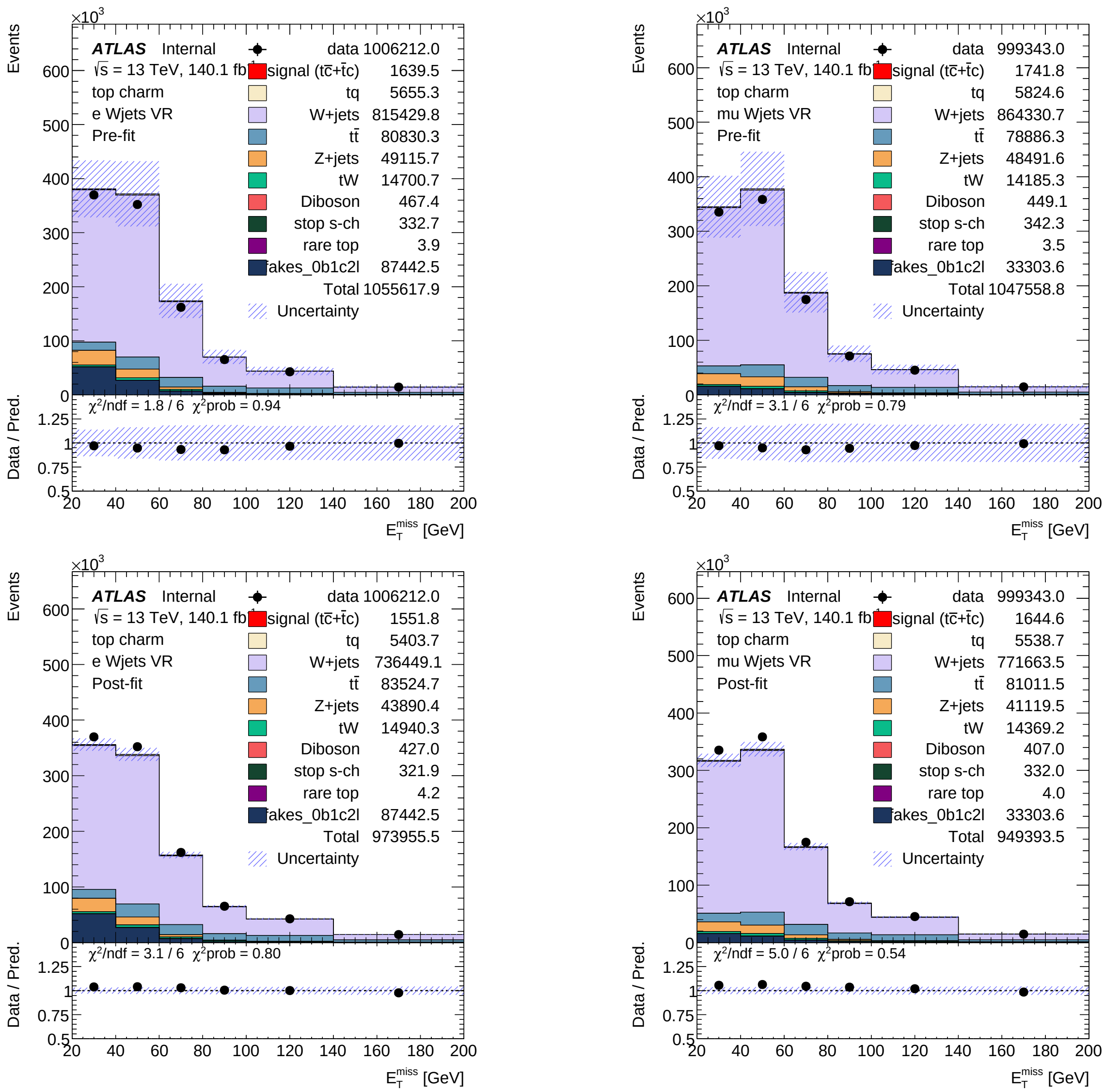
(a) (b)

Figure 8.16: (a) Pre-fit and (b) post-fit distribution of E_T^{miss} in $W^{\pm}$+jets control region



(a) (b)

Figure 8.17: (a) Pre-fit and (b) post-fit distribution of E_T^{miss} in $t\bar{t}$ validation region



(a) (b)

Figure 8.18: (a) Pre-fit and (b) post-fit distribution of E_T^{miss} in $W^{\pm}$+jets validation region

Chapter 9

Conclusion

This thesis presents an analysis that leads to the first observation of the single top quark with an associated charm quark. The data used for the analysis are recorded by the ATLAS detector at the LHC using proton-proton collisions at a center-of-mass energy of 13 TeV, with a full integrated luminosity of $140.1 fb^{-1}$ during the 2015-2018 data-taking period.

The analysis is performed by defining a SR with events consisting of exactly one lepton (e, μ) with $p_T > 27$ GeV, zero light jets, one b-tagged jet and one c-tagged jet with $p_T > 20$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ and $MWT > 40$ GeV and two CRs dominated by $t\bar{t}$ and $W^{\pm}$+jets samples.

A BDT was trained and used in the SR region to increase the sensitivity of the measurement. A binned likelihood fit is performed simultaneously on the SR and CRs using the BDT distribution in the SR and E_T^{miss} distribution in the CRs.

The expected uncertainty on the POI, μ_{tc} is measured to be 6%. The expected significance is more than 17, which is well above the 5σ discovery threshold. These results indicate a very high sensitivity to the measurement of the single top quark with an associated charm quark upon unblinding the SR, after approval from the ATLAS collaboration. This precision measurement can lead to beyond the standard model searches in the same final state, which are studied in a different thesis.

References

[1] David J. Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, Weinheim, Germany, 2nd, revised edition, 2008. Chapter 2: Elementary Particle Dynamics. 1.1

[2] 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*, 716:1–29, 2012. URL: https://doi.org/10.1016/j.physletb.2012.08.020. 1.1

[3] Schwartz, Matthew D. *Quantum Field Theory and the Standard Model*. Cambridge University Press, Cambridge, 1st edition, 2014. 1.1

[4] Peskin, Michael E. and Schroeder, Daniel V. *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, Massachusetts, USA, 1995. 1.1

[5] Glashow, Sheldon L. Partial-Symmetries of Weak Interactions. *Nuclear Physics*, 22(4):579–588, 1961. URL: https://doi.org/10.1016/0029-5582(61)90469-2. 1.1

[6] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Physical Review*, 127:965–970, 1962. URL: https://doi.org/10.1103/PhysRev.127.965. 1.1

[7] Antonio Pich. The standard model of electroweak interactions. *arXiv e-prints*, 2012. URL: https://doi.org/10.48550/arXiv.1201.0537. 1.1

[8] ATLAS Collaboration. Higgs potential diagram, 2013. 1.2

[9] Jens Erler and Andreas Freitas. Electroweak model and constraints on new physics, 2022. URL: https://pdg.lbl.gov/2022/reviews/rpp2022-rev-standard-model.pdf. 1.1

[10] Franz Gross and et al. 50 years of quantum chromodynamics. *European Physical Journal C*, 83(12):1125, 2023. URL: https://doi.org/10.1140/epjc/s10052-023-11949-2. 1.1

[11] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Physical Review Letters*, 30(26):1343–1346, 1973. URL: https://doi.org/10.1103/PhysRevLett.30.1343. 1.2

[12] Joël Feltesse. Introduction to parton distribution functions. *Scholarpedia*, 5(11):10160, 2010. URL: https://www.scholarpedia.org/article/Introduction_to_Parton_Distribution_Functions. 1.2

[13] CERN. Atlas. URL: https://greybook.cern.ch/experiment/detail?id=ATLAS. 2

[14] Christopher J. Rhodes. Large hadron collider (lhc). *Science Progress*, 96(Pt 1):95–109, March 2013. doi:10.3184/003685013X13623370524107. 2.1

[15] Lyndon Evans. The large hadron collider. In *eConf C06-07-17: Proceedings of the 40th SLAC Summer Institute on Particle Physics*, volume C060717 of *SLAC eConf*, page L003, Stanford, CA, USA, 2006. URL: https://www.slac.stanford.edu/econf/C060717/papers/L003.PDF. 2.1

[16] Esma Mobs. The cern accelerator complex, 2016. URL: https://cds.cern.ch/record/2197559. 2.1

[17] ATLAS Collaboration. Measurement of jet energy resolution with pile-up in ATLAS using the Run 2 dataset. Atl-phys-pub/ana-jetm-2023-08, European Organization for Nuclear Research (CERN), August 2023. URL: https://cds.cern.ch/record/2904445/files/ANA-JETM-2023-08-PAPER.pdf. 2.1

[18] ATLAS Collaboration. Luminosity public results run2, 2025. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. 2.2

[19] ATLAS Collaboration. *A 25-Year Insider Story of the LHC Experiment*. WORLD SCIENTIFIC, 2019. URL: https://doi.org/10.1142/11030. 2.2

[20] Joao Pequenao. Computer generated image of the whole ATLAS detector, 2008. URL: https://cds.cern.ch/record/1095924. 2.3

[21] The ATLAS Collaboration and Andreas Korn. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(8):S08003, December 2008. URL: https://doi.org/10.1088/1748-0221/3/08/S08003. 2.2.1

[22] Giles Strong. On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case. *Machine Learning: Science and Technology*, 1, 09 2020. URL: https://doi.org/10.48550/arXiv.2002.01427. 2.4

[23] ATLAS Collaboration. Performance of the reconstruction of large impact parameter tracks in the inner detector of atlas. *European Physical Journal C*, 83(11):1081, 2023. URL: https://doi.org/10.1140/epjc/s10052-023-12024-6. 2.2.2

[24] João Pequenao. Computer generated image of the ATLAS inner detector, 2008. URL: https://cds.cern.ch/record/1095926. 2.5

[25] ATLAS Collaboration. The ATLAS Inner Detector commissioning and calibration. *European Physical Journal C*, 70:787–821, 2010. URL: https://doi.org/10.1140/epjc/s10052-010-1366-7. 2.2.2

[26] ATLAS Collaboration. The upgraded pixel detector of the atlas experiment for run 2 at the large hadron collider. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 831:18–23, 2016. URL: https://doi.org/10.1016/j.nima.2016.05.018. 2.2.2

[27] ATLAS Collaboration. Production and integration of the atlas insertable b-layer. *Journal of Instrumentation*, 13:T05008, 2018. URL: https://doi.org/10.48550/arXiv.1803.00844. 2.2.2

[28] ATLAS Collaboration. Operation and performance of the atlas semiconductor tracker in lhc run2. *Journal of Instrumentation*, 17:P01013, 2022. URL: https://doi.org/10.48550/arXiv.2109.02591. 2.2.2

[29] ATLAS Collaboration. Atlas transition radiation tracker (trt): Straw tubes for tracking and particle identification at the large hadron collider. In *Proceedings of the 14th Vienna Conference on Instrumentation (VCI2016), Vienna, Austria, 15–19 Feb 2016*, volume 845 of *Nucl. Instrum. Methods Phys. Res. A*, pages 257–261, 2017. URL: https://doi.org/10.1016/j.nima.2016.04.026. 2.2.2

[30] ATLAS Collaboration. *ATLAS calorimeter performance: Technical Design Report*. Technical Design Report. ATLAS; 1. CERN, Geneva, Switzerland, 1996. URL: https://cds.cern.ch/record/331059. 2.2.3

[31] João Pequenao. Computer Generated Image of the ATLAS calorimeter, 2008. URL: https://cds.cern.ch/record/1095927. 2.6

[32] Christian W. Fabjan and Fabiola Gianotti. Calorimetry for particle physics. *Reviews of Modern Physics*, 75(4):1243–1286, October 2003. URL: https://doi.org/10.1103/RevModPhys.75.1243. 2.2.3

[33] ATLAS Collaboration. The ATLAS Muon Spectrometer. Sn-atlas-2003-030, CERN, July 2003. URL: https://doi.org/10.1140/epjcd/s2004-04-013-y. 2.2.4

[34] João Pequenao. Computer generated image of the ATLAS Muons subsystem, 2008. URL: https://cds.cern.ch/record/1095929. 2.7

[35] ATLAS Collaboration. Operation of the atlas trigger system in run 2. *Journal of Instrumentation*, 15(10):P10004, 2020. URL: https://doi.org/10.48550/arXiv.2007.12539. 2.2.5

[36] Nils Braun. *Combinatorial Kalman Filter and HighLevel Trigger Reconstruction for the Belle II Experiment*. PhD thesis, Karlsruhe Institute of Technology (KIT), Germany, 2019. URL: https://doi.org/10.1007/978-3-030-24997-7. 3.1

[37] ATLAS Collaboration. Tracking and vertexing with the atlas inner detector in the lhc run-2. In *Proceedings of the International Conference on Technology and Instru-*

mentation in Particle Physics (TIPP 2017), Beijing, China, 21–26 May 2017, volume 213 of *Springer Proceedings in Physics*, pages 400–403, 2018. URL: https://doi.org/10.1007/978-981-13-1316-5_75. 3.1

[38] Nora Pettersson. Brem recovery diagram (atlas tracking tutorial). Image in ATLAS Tracking Software Tutorial, ATLAS Tracking CP group, 2024. URL: https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/tracksf/. 3.1

[39] ATLAS Collaboration. Primary vertex reconstruction with the atlas detector. In *Proceedings of the 8th International Workshop on Semiconductor Pixel Detectors for Particles and Imaging (PIXEL2016), Sestri Levante, Italy, 5–9 Sep 2016*, volume 11 of *Journal of Instrumentation*, page C12060, 2016. URL: https://10.1088/1748-0221/11/12/C12060. 3.1

[40] Frühwirth, Rudolf and Strandlie, Are. *Secondary Vertex Reconstruction*, pages 159–165. Springer International Publishing, 2021. URL: https://doi.org/10.1007/978-3-030-65771-0_9. 3.1

[41] ATLAS Collaboration. Topological cell clustering in the atlas calorimeters and its performance in lhc run 1. *European Physical Journal C*, 77:490, 2017. URL: https://doi.org/10.1140/epjc/s10052-017-5004-5. 3.2

[42] ATLAS Collaboration. Improved electron reconstruction in atlas using the gaussian sum filter-based model for bremsstrahlung. Atlas-conf-2012-047, CERN, May 2012. URL: https://cds.cern.ch/record/1449796/files/ATLAS-CONF-2012-047.pdf. 3.3

[43] ATLAS Collaboration. Electron reconstruction and identification in the atlas experiment. *European Physical Journal C*, 79:639, 2019. URL: https://doi.org/10.1140/epjc/s10052-019-7140-6. 3.3

[44] Otilia Anamaria Ducu. Electron efficiency in lhc run-2 with the atlas experiment. Atl-phys-proc-2024-001, CERN, January 2024. URL: https://cds.cern.ch/record/2885845/files/ATL-PHYS-PROC-2024-001.pdf. 3.3

[45] ATLAS Collaboration. Muon reconstruction performance of the atlas detector in proton–proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$. *European Physical Journal C*, 76:292, 2016. URL: https://doi.org/10.1140/epjc/s10052-016-4120-y. 3.4

[46] ATLAS Collaboration. Muon reconstruction and identification efficiency in atlas using the full run 2 pp collision dataset at $\sqrt{s} = 13$ TeV. *European Physical Journal C*, 81:578, 2021. URL: https://doi.org/10.1140/epjc/s10052-021-09233-2. 3.4

[47] ATLAS Collaboration. Atlas jet and missing energy reconstruction, calibration and performance in lhc run-2. In *Proceedings of the International Conference on Instrumentation for Colliding Beam Physics (INSTR-2017), Novosibirsk, Russia, 27 Feb – 3 Mar 2017*, volume 12 of *Journal of Instrumentation*, page C06038, 2017. URL: https://doi.org/10.1088/1748-0221/12/06/C06038. 3.5

[48] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the atlas detector. *European Physical Journal C*, 77:466, 2017. URL: https://doi.org/10.1140/epjc/s10052-017-5031-2. 3.5

[49] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 04:063, 2008. doi:10.1088/1126-6708/2008/04/063. 3.5

[50] ATLAS Collaboration. Tagging and suppression of pileup jets with the atlas detector. Atlas-conf-2014-018, CERN, May 2014. URL: https://cds.cern.ch/record/1700870/files/ATLAS-CONF-2014-018.pdf. 3.5

[51] ATLAS Collaboration. Transforming jet flavour tagging at atlas. *Preprint submitted to Nature Communications*, 2025. URL: https://doi.org/10.48550/ARXIV.2505.19689. 3.5.1

[52] ATLAS Collaboration. The performance of missing transverse momentum reconstruction and its significance with the atlas detector using $140\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 13\,\mathrm{tev}$ pp collisions. *European Physical Journal C*, 85:606, 2025. URL: https://doi.org/10.1140/epjc/s10052-025-14062-8. 3.6

[53] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the atlas detector at the lhc. *European Physical Journal C*, 83:982, 2023. doi:10.1140/epjc/s10052-023-11747-w. 4.1

[54] ATLAS Collaboration. The atlas electron and photon trigger performance in run 2. *International Journal of Modern Physics A*, 35(34&35):2044007, 2020. URL: https://doi.org/10.1142/S0217751X20440078. 4.1

[55] ATLAS Collaboration. Performance of the atlas muon triggers in run 2. *Journal of Instrumentation*, 15(09):P09015, September 2020. URL: https://doi.org/10.1088/1748-0221/15/09/P09015. 4.1

[56] ATLAS Collaboration. Introduction to monte carlo simulations. ATLAS Open Data documentation, 2025. URL: https://opendata.atlas.cern/docs/documentation/monte_carlo/introduction_MC. 4.2

[57] Christian Bierlich and et al. *A Comprehensive Guide to the Physics and Usage of PYTHIA 8.3*. SciPost Phys. Codebases, 2022. URL: https://doi.org/10.21468/SciPostPhysCodeb.8. 4.2

[58] M. Bähr and et al. Herwig++ physics and manual. *arXiv e-print*, 2008. URL: https://doi.org/10.48550/arXiv.0803.0883. 4.2

[59] Enrico Bothmann and et al. Event generation with sherpa 2.2. *SciPost Physics*, 7(3):034, 2019. URL: https://doi.org/10.21468/SciPostPhys.7.3.034. 4.2

[60] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5 : Going beyond. *Journal of High Energy Physics*, 06:128, 2011. URL: https://doi.org/10.1007/JHEP06(2011)128. 4.2

[61] Carlo Oleari. The POWHEG box. *Nuclear Physics B – Proceedings Supplements*, 205–206:36–41, 2010. URL: https://doi.org/10.1016/j.nuclphysbps.2010.08.016. 4.2

[62] S. Agostinelli and et al. Geant4 -a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*

Equipment, 506(3):250–303, 2003. URL: https://doi.org/10.1016/S0168-9002(03)01368-8. 4.2

[63] ATLAS Collaboration. ATLFAST3: the next generation of fast simulation in ATLAS. Technical Report ATL-PHYS-PUB-2019-019, CERN, 2019. URL: https://cds.cern.ch/record/2684071. 4.2

[64] The NNPDF Collaboration. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 04:040, 2015. URL: https://doi.org/10.1007/JHEP04(2015)040. 4.2.1, 4.2.2

[65] ATLAS Collaboration. Measurement of t-channel production of single top quarks and antiquarks in pp collisions at 13 tev using the full atlas run 2 data sample. *Journal of High Energy Physics*, 05:305, 2024. URL: https://doi.org/10.1007/JHEP05(2024)305. 4.2.1

[66] J. Alwall and et al. A standard format for les houches event files. *Computer Physics Communications*, 176:300–304, 2007. URL: https://doi.org/10.1016/j.cpc.2006.11.010. 4.2.1

[67] ATLAS Collaboration. Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the atlas experiment. Ana-pmgr-2021-01, CERN, December 2021. URL: https://cds.cern.ch/record/2798348/files/ANA-PMGR-2021-01-PAPER.pdf. 4.2.2

[68] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *Journal of High Energy Physics*, 12:039, 2008. URL: https://doi.org/10.1088/1126-6708/2008/12/039. 4.2.2

[69] Fabrizio Buccioni, Stefano Pozzorini, and Max F. Zoller. Openloops 2: automated generation of tree and one-loop scattering amplitudes beyond *qcd*. *European Physical Journal C*, 79(10):866, 2019. URL: https://doi.org/10.1140/epjc/s10052-019-7306-2. 4.2.2

[70] Steffen Schumann and Frank Krauss. A parton shower algorithm based on catani–seymour dipole factorisation. *Journal of High Energy Physics*, 2008(03):038, 2008. URL: https://doi.org/10.1088/1126-6708/2008/03/038. 4.2.2

[71] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. Qcd matrix elements and truncated showers. *Journal of High Energy Physics*, 2009(05):053, 2009. URL: https://doi.org/10.1088/1126-6708/2009/05/053. 4.2.2

[72] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in monte carlo simulations. *Journal of High Energy Physics*, 03:015, 2013. URL: https://doi.org/10.1007/JHEP03(2013)015. 4.2.2

[73] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a w boson. *Journal of High Energy Physics*, 2008(07):029, 2008. URL: https://doi.org/10.1088/1126-6708/2008/07/029. 4.2.2

[74] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *arXiv:physics/0408124*, 2004. URL: https://doi.org/10.48550/arXiv.physics/0408124. 5.3

[75] Pushpalatha C. Bhat. Advanced analysis methods in particle physics. Fermilab-pub-10-054-e, Fermi National Accelerator Laboratory, 2010. URL: https://inspirehep.net/files/dfeaa32148f323372b79cc2a24a9a374. 5.3

[76] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *arXiv e-prints*, arXiv:1603.02754, 2016. URL: https://doi.org/10.48550/arXiv.1603.02754. 5.3.1

[77] Christoph Molnar. Permutation feature importance —interpretable machine learning, 2020. 5.3.1

[78] ATLAS Collaboration. Tools for estimating fake/non-prompt lepton backgrounds with the atlas detector at the lhc. *Journal of Instrumentation*, 18(11):T11004, 2023. URL: https://doi.org/10.1088/1748-0221/18/11/T11004. 6

[79] Arnulf Quadt and Baptiste Ravina and Daniel Werner and et al. TopCP-Toolkit: A framework for n-tuple production and analysis in top-quark physics within ATLAS, 2024. URL: https://www.baptisteravina.com/talk/topcptoolkit-analysis-software-for-run-3-and-beyond/. 6

[80] Arnulf Quadt, Baptiste Ravina, and et al. Fastframes: A root rdataframe-based framework for histogram and n-tuple production in atlas analyses, 2024. URL: https://gitlab.cern.ch/atlas-amglab/fastframes. 6

[81] ATLAS Collaboration. Truthclassification package in athena. URL: https://gitlab.cern.ch/atlas/athena/-/tree/7689c1b632a058bbc87f372590018008f4cafc78/PhysicsAnalysis/AnalysisCommon/TruthClassification. 6.1

[82] ATLAS Collaboration. TRExFitter: A statistical framework for the interpretation of ATLAS results in high energy physics. Technical Report ATL-PHYS-PUB-2018-017, CERN, 2018. URL: https://cds.cern.ch/record/2625971. 7

[83] S. van der Meer. Calibration of the effective beam height in the isr. *CERN Internal Report*, (ISR-PO-68-31), 1968. URL: https://cds.cern.ch/record/296752. 7.1.1

[84] G. Avoni et al. The new lucid-2 detector for luminosity measurement and monitoring in atlas. *Journal of Instrumentation*, 13(7):P07017, 2018. URL: https://cds.cern.ch/record/2633501/files/document.pdf, doi:10.1088/1748-0221/13/07/P07017. 7.1.1

[85] ATLAS Collaboration. Atl-soft-proc-2013-030: [title as in document]. Technical Report ATL-SOFT-PROC-2013-030, CERN, 2013. URL: https://cds.cern.ch/record/1616394/files/ATL-SOFT-PROC-2013-030.pdf. 7.1.2

[86] Timothée Thévenaux-Pelzer and on behalf of the ATLAS CDF CMS & D0 Collaborations Fernández Menéndez, Javier. Leptons: e, μ, systematic uncertainties. In *Proceedings of the DESY Workshop "Top Physics at Tevatron, LHC and Future*

Colliders", 2014 (DESY-PROC-2014-02), 2014. URL: https://doi.org/10.3204/DESY-PROC-2014-02/3. 7.1.3

[87] ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ tev with the atlas detector. *European Physical Journal C*, 81:689, 2021. URL: https://doi.org/10.1140/epjc/s10052-021-09402-3. 7.1.4

[88] Zhibo Wu and on behalf of the ATLAS Collaboration Delmastro, Marco. Small-r b-jet tagging calibration for gn2, 2025. URL: https://indico.in2p3.fr/event/35618/contributions/150869/subcontributions/9405/attachments/90885/138146/LAPPBtagCalib20250224.pdf. 7.1.5

[89] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions at $\sqrt{s} = 13$ tev. *European Physical Journal C*, 78(11):903, 2018. URL: https://doi.org/10.1140/epjc/s10052-018-6288-9. 7.1.6

[90] Sergey Alekhin, Alioli, and et al. The pdf4lhc working group interim report. *arXiv preprint*, 2011. URL: https://doi.org/10.48550/arXiv.1101.0536, arXiv:1101.0536. 7.2

[91] Johannes Bellm and et al. Herwig 7.2 release note. *European Physical Journal C*, 80:452, 2020. doi:10.1140/epjc/s10052-020-8011-x. 7.2

[92] ATLAS Collaboration. Systematicsmoothingtool: Algorithm description, 2018. URL: https://gitlab.cern.ch/atlas/statistics/SystematicSmoothingTool/-/blob/master/doc/algorithm_description.md. 7.4

[93] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, Oxford; New York, 1998. URL: https://www.sherrytowers.com/cowan_statistical_data_analysis.pdf. 8.1

[94] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. URL: https://arxiv.org/pdf/1007.1727, arXiv:1007.1727, doi:10.1140/epjc/s10052-011-1554-0. 8.1

[95] Luc Demortier and Louis Lyons. Testing hypotheses in particle physics: plots of p_0 versus p_1. *arXiv preprint*, 2014. URL: https://doi.org/10.48550/arXiv.1408.6123. 8.2

[96] Tomáš Dado, Alexander Held, Michele Pinamonti, Mohamed Aly, and Loïc Valéry. "breakdown of uncertainties via covariance matrix – total breakdown", 2025. URL: https://trexfitter-docs.web.cern.ch/trexfitter-docs/latest/inference/breakdown_covariance/#total-breakdown. 8.2