Measurement of the Single Top tW-Channel Inclusive Cross Section in the Single Lepton Final State at 13 TeV with ATLAS

Federico Guillermo Diaz Capriles

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

April 2018

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachter: Priv-Doz. Dr. Philip Bechtle

Acknowledgements

I would like to thank my family and friends for their continued support with the distinction of my mother, and best friend whose contributions are what made this possible. Prof. Dr. Ian Brock for the opportunity to take part in this field, the incredible experiences this year has had, and most importantly for the unending kindness. Finally, to the rest of the Brock group for their incredible patience, help, and words of encouragement. In the end I see you not just as coworkers or colleagues but as friends.

Contents

CHAPTER 1

Preamble

Mankind has always had a drive to understand the world it inhabits. This drive to search for the unknown has influenced humanity to build colossal machines to probe what eyes cannot see. One of such machines is the Large Hadron Collider (LHC) located in the Swiss-French border near Geneva and one of its eyes is the ATLAS detector. It is well known that energy and matter are interconnected in such a way that they are like two faces of the same coin. The LHC accelerates protons to record breaking energies for the purpose of colliding them. These highly energetic collisions create matter, some of which is not commonly found on earth, that is then measured by the ATLAS detector directly or indirectly. The creation and destruction of particles is not a truly random event, rather these follow a set of rules described, to an extent, by the Standard Model of particle physics. By measuring these particles in detectors one can gather information on the underlying processes that govern these interactions and test the limits of current knowledge.

The Standard Model (SM), in a nutshell, predicts several particles and describes their interactions with one another. It further predicts physical quantities such as how often certain processes happen. These are the types of quantities that can be measured at particle colliders. Although it is currently the most accurate tool to describe particle physics interactions, it is incomplete. Several phenomena have been observed that are not described by this theory and so these observations motivate the search for new physics beyond the Standard Model.

The heaviest of all particles is known as the top-quark. It has unique properties and characteristics that make it an interesting subject of study. Its production and decay can be studied as it may be sensitive to new physics and it can be used to measure certain SM parameters with great precision. Fortunately, the LHC is able to produce top-quarks in large numbers.

In this thesis, the production channel of a single top-quark in association with a W boson, commonly referred to as the tW channel, is studied by the single lepton decay mode. The goal in this analysis is to measure the cross-section of this single top-quark production. The biggest challenge in this measurement comes from the top-quark itself as it is more likely to be produced in pairs. Therefore, the differentiation of single top-quark to pairs becomes a large obstacle. Two strategies are explored in this document for the purpose of finding the most accurate and precise measurement of the tW cross-section.

CHAPTER 2

Introduction

2.1 The Standard Model

The Standard Model (SM) is the most successful theory at describing and classifying elementary particles and their interactions with great precision. Following the discoveries of the top-quark (1995, Tevatron), tau neutrino (2000, DONUT), and Higgs boson (2012, LHC), one can see how effective the SM is at modeling the universe. Further giving credence to the theory, several experiments seek to test the limits of precision measurements and compare them with the known model. In figure 2.1, one can see the results published by the ATLAS collaboration which compare their cross-section measurements to the theoretical expectations. It also shows that most cross-section measurements are in agreement with theory. However, the SM is not perfect. There are a few questions that remain unanswered, for example: dark matter and energy are not predicted, neutrino masses are not generated, and gravity is also not implemented as there is no renormalizable way to generate the theory.

In brief, the Standard Model classifies elementary particles into two types with differing properties: bosons and fermions. Elementary particles are those which have no substructure and are point-like. Fermions are the type that form matter and bosons are the mediators of force between them. In figure 2.2, all the elementary particles predicted by the SM are classified and labeled with some physical properties given.

Fermions

Fermions are divided into quarks and leptons. Each particle comes with a variety of characteristics, some of which are: electric, weak and color charge, flavor, mass, and spin. All fermions in the SM have a spin of one half and have an antiparticle version which has opposite charge and quantum numbers. There are total of six quarks and six leptons which can be divided into three families to a total of twelve fermions.

Leptons are further subdivided into charged and neutral. Of the charged leptons, the most well known is the electron, e, which was discovered by J. J. Thomson in 1897. It has an electric charge of 1.602×10^{-19} C and a mass of $0.511\,\mathrm{MeV}\,c^{-2}$. There are two more charged leptons which are more massive than the electron and share the same charge, the muon and tau, listed in increasing mass.

Each charged lepton has a neutral counterpart called a neutrino, ν, which shares a lepton family number with their charged partner. Neutrinos were postulated by Pauli after studying the energy spectrum of β decay. They were directly detected much later in nuclear reactors by a process called beta capture where a proton interacts with an electron and the resulting particles are a neutron and an electron neutrino. In the SM, neutrinos are massless although neutrino oscillation, which is only possible if neutrinos have

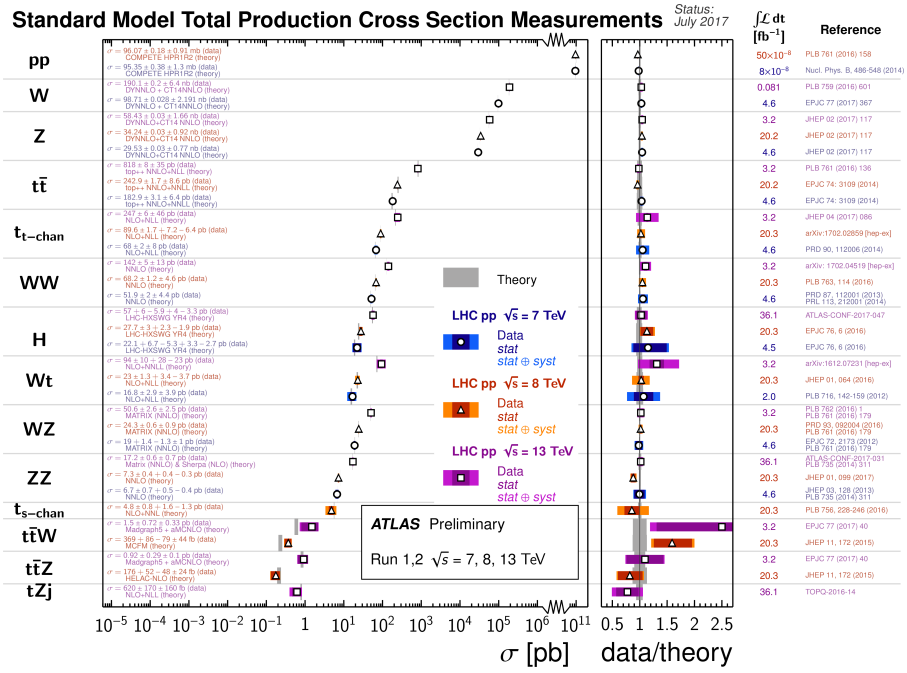


Figure 2.1: Public cross-section measurements for several processes and their theoretical value from the ATLAS collaboration. The left panel has the cross-section measurement and theory for different processes. The right side then shows a ratio of data to theory, the integrated luminosity and the reference to the measurement [1].

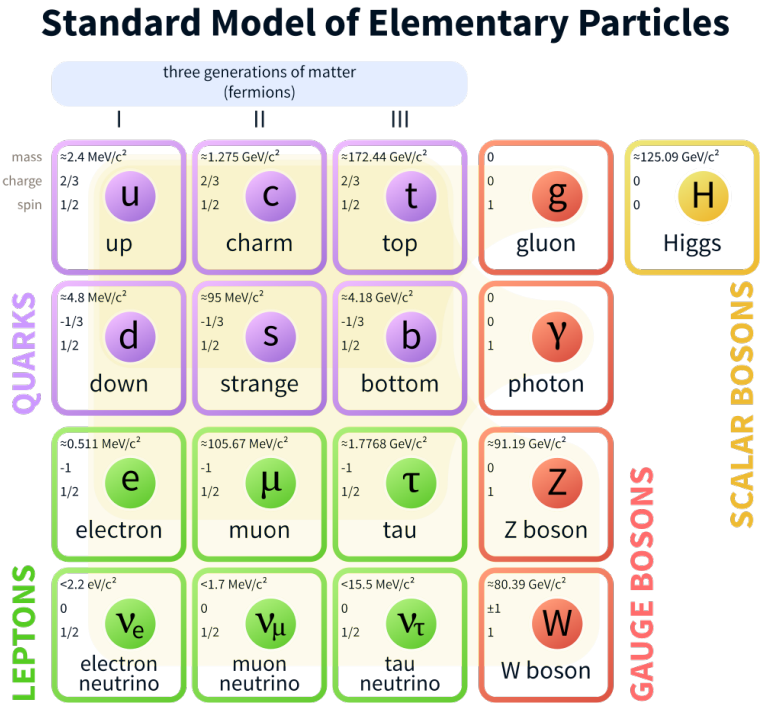


Figure 2.2: Schematic of the Standard Model of particle physics. It has color coded particle types. Each box contains the name, symbol, spin, mass, and charge of the particle denoted [2].

mass, has been observed. According to current experiments, no mass has been measured but instead upper limits have been imposed which are in the eV range.

Quarks, on the other hand, are all electrically charged but with a fraction of lepton charge. They are up, down, charm, strange, top, and bottom. These are also divided into up- and down-type quarks which have the same electric charge. Up-type quarks are (in increasing mass) up, charm, and top. They all have a positive fraction of two thirds the electron charge, unlike the down-type which have a negative fraction of one third the electron charge. An unique quality of quarks is their inability to exist freely, instead they can only exist in bound states with other quarks called hadrons (with the exception of the top-quark, more detail in section 2.2). This behavior is called confinement and it is a consequence of color charge. These charges are labeled red, green, and blue, while antiquarks have anticolor. The bound states must always be color neutral with the most common and known hadrons types being mesons ($q\bar{q}$) and baryons (qqq).

Bosons

The SM includes three out of the four forces of nature. The four forces are the strong, weak, electromagnetic, and gravitational. As mentioned earlier, there is no theory for gravity in the SM. Even though the gravitational force's range is infinite, the effect of it is incredibly small since all the masses of particles involved are very small. At least its effects can be neglected at the energy scale of this analysis. As for the other three forces, they all have one or more force carriers and have different strengths and act at distinct ranges.

All of the gauge bosons are a necessary consequence of physical laws having some form of symmetry at all points of space time. In the case of gauge symmetries, these are internal and unitary transformations that give some conservation law, via Noether's theorem, and can describe interactions with an exchange of a mediating particle. Specifically, the SM is a combination of three such transformations: $U(1) \times SU(2) \times SU(3)$, where the $U(1) \times SU(2)$ describes electroweak theory and $SU(3)$ represents the strong force.

The most well known force carrier is the photon, γ, which interacts with all things that are electrically charged. This is a massless boson with spin one and electrically neutral. The dynamics related to this boson are well described by quantum electrodynamics (QED).

The $W^{\pm}$ and Z bosons govern the weak interaction and are the only gauge bosons to have mass. Specifically, their masses are $(80.385 \pm 0.015)\,\mathrm{GeV}\,c^{-2}$ for the W bosons and $(91.1876 \pm 0.0021)\,\mathrm{GeV}\,c^{-2}$ for the Z [3]. They have spin one and, as their notation suggests, the W bosons are charged while the Z is neutral. The $W^{\pm}$ part of the weak force allows for flavor change through a mechanism known as mixing which is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix for quarks and Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix for neutrinos. The best description of all the weak and electromagnetic interactions are detailed in electroweak theory.

The strong force is mediated by the gluon, g, which also has the ability to self-interact. The gluon is massless, has spin one, and although the gluon is only noted once in figure 2.2, it can come in eight different color-anticolor combinations. This color combination is what allows them to interact among each other. This boson and all its interactions are described in quantum chromodynamics (QCD).

The last boson in the SM is called the Higgs, H, and is the latest addition to the model. It has a mass of about $125\,\mathrm{GeV}\,c^{-2}$, spin zero, and is electrically and color neutral. First theorized by Peter Higgs in 1964 and found by the Large Hadron Collider in 2012. The Higgs field plays a unique role in the SM as it explains the mechanism by which some particles have mass. In the SM lagrangian[1] , adding mass terms for some elementary particles breaks the symmetry needed for the interactions to work. The Higgs

[1] A lagrangian is a function that describes the state of a dynamic system and is equal to the difference between potential and kinetic energy.

mechanism explains why the photon and gluon are massless while at the same time explaining why the Z and $W^{\pm}$ are massive. It also generates the mass of quarks and charged leptons.

Structure of a Proton

As explained earlier, quarks are unable to exist isolated and therefore must form bound states with other quarks in order to produce hadrons. Arguably the most well known hadron is the proton, p, which is a baryon at the core of every atom. It has a positive charge and has a mass slightly less than the neutron, However its mass is much larger than the three valence quarks (two ups and one down) that compose it. The valence quarks masses added together make up only about 1 % of the total proton mass. The remaining mass comes from QCD binding energy, which is the kinetic energy of the quarks and energy of the gluon fields that bind them to each other.

One can consider a proton as composed by its three valence quarks. However, they will interact with each other by exchanging gluons which can interact with other gluons or decay into quark-antiquark pairs. These interactions are what make up the sea of partons inside a proton.

The interaction of quarks and gluons within the proton results in a distribution of momenta among these bodies within the proton. To describe these distributions, one can use a parton distribution function (PDF). One great quality of PDFs is that they are independent of the hard scattering process and can be extrapolated to higher energy scales. These distributions can be seen in figure 2.3. These were determined at HERA over several measurements where a $e^{\pm}$ beam would collide with a proton in order to probe its structure. The x value in the plot is called the *Bjorken x* and it described the longitudinal momentum fraction between the proton and the boson exchanged in the $e^{\pm}p$ collision. It can be seen that when x is high the valence quarks are more likely to be the interacting parton. At lower x values, the gluons dominate with some probability of finding sea quarks. Thus, at higher energies one can consider the proton as a uniform sea of partons. In order to probe in detail the structure of a proton, it is valuable to know the physics in particle colliders.

Physics at Hadron Colliders

Particle colliders, as their name implies, collide particles. This, in turn, allows one to measure the physical properties of elementary particles when they interact. One of the values that can be measured is the *cross-section*, or how often a process can happen. This is expressed as an effective area of the colliding particles and is typically measured in units of *barn*. For context, $1\,\mathrm{b} = 1 \times 10^{-28}\,\mathrm{m}^2$.

After collisions, particles will decay into other, more stable particles. One can measure the rate by which these particles decay and also divided into decay channels. The full rate is called *decay width*, Γ, and if one measures depending on the type (denoted by the i subscript) of decay it is then called a *partial decay width*, Γ_i. Furthermore, dividing the partial decay width by the full decay width one can obtain the branching fraction, or how likely any particular decay happens for a specific particle. Lastly, a physical property called the lifetime, τ, is directly related to the decay width as they are reciprocals of each other. As such, the decay width can be measured in units of energy, eV, or units of time by a conversion factor of $\hbar$.

In order to get any measurement about particles, one needs to reconstruct or rebuild an event. Before calculating event variables one needs to know the coordinate system that the detector will use. The Z axis points along the beam line, the Y axis points towards upward, and the X axis points horizontally and perpendicularly to Z and Y axises. Some variables that are valuable for describing an event are the transverse momentum, p_{T}, the azimuthal angle, ϕ, rapidity, y, the invariant mass, m, and the transverse mass between two objects (typically the lepton and its associated neutrino), m_{T}. Much of this information

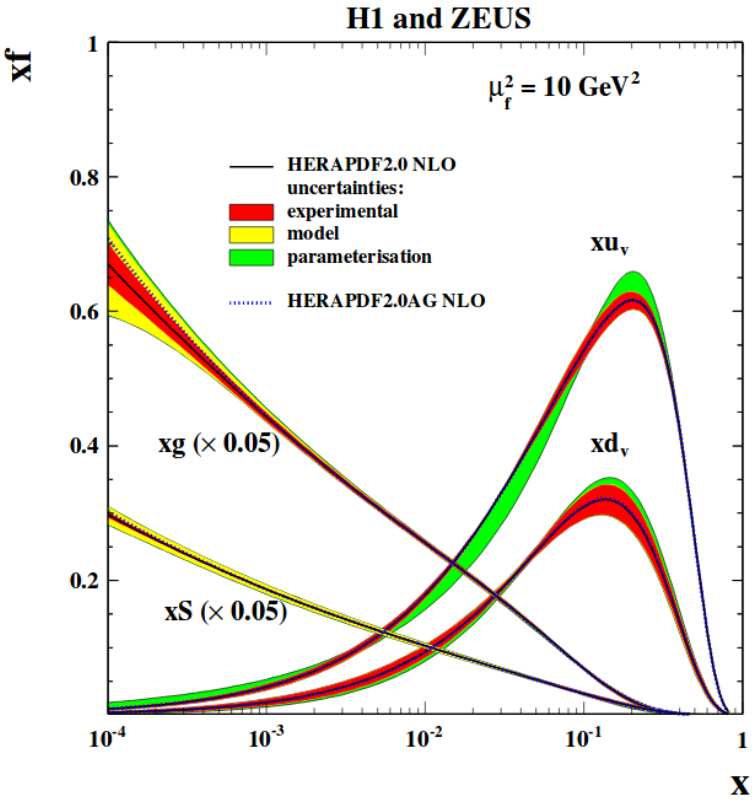


Figure 2.3: Parton distribution functions from a combination of measurements at HERA as a function of the longitudinal momentum fraction with uncertainties provided. The gluon and sea quarks PDFs have been scaled down by a factor of 20 [4].

can be gathered from the particle's four-momentum vector of (E, p_X, p_Y, p_Z) where E is the energy of the particle and p_i is the momentum pointing along the i-th axis:

$$p_{\mathrm{T}} = \sqrt{p_X^2 + p_Y^2}, \tag{2.1}$$

$$\phi = \begin{cases} \arctan(\frac{p_Y}{p_X}) & p_X \neq 0, \\ (\text{sign of } p_Y)\frac{\pi}{2} & p_X = 0. \end{cases} \tag{2.2}$$

$$y = \operatorname{arctanh}\frac{p_Z}{E}, \tag{2.3}$$

$$M = \sqrt{E^2 - p_X^2 - p_Y^2 - p_Z^2}, \tag{2.4}$$

$$\begin{aligned} m_{\mathrm{T}} &= \sqrt{M^2 + p_X^2 + p_Y^2} \\ &= \sqrt{2p_{\mathrm{T}}(\ell)p_{\mathrm{T}}(\nu_\ell)(1 - \cos(\phi(\ell) - \phi(\nu_\ell))}. \end{aligned} \tag{2.5}$$

At the energies that ATLAS measures currently, one can use the pseudorapidity, η, instead of the rapidity as it is a good approximation and can be calculated directly from the polar angle rather than the energy of the particle:

$$\eta = \text{arctanh}\frac{p_Z}{|\vec{p}|} = -\ln\tan\frac{\theta}{2}. \tag{2.6}$$

From η and ϕ one can then define the angular separation between any two objects as ΔR which is a function of the difference in pseudorapidities and azimuthal angles. The advantage of this definition is its invariance under boosts along the beam direction for massless objects and it is a good approximation for particles with small masses.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.7}$$

When two objects with substructure collide, typically debris is ejected. Similarly, proton-proton collisions dot not only eject one particle but a debris, or cloud, of other partons that can interact and radiate gluons. This radiation is typically done in the same general direction of the ejected partons. This cloud of partons that radiate gluons and hadronize is called a *parton shower*.

Typically, the parton shower happens with small ΔR around the particles with high energy from the collision. This spray of collimated particles is called a *jet*. Jets are identified by reconstruction algorithms which not only determine properties and identities of the original partons but also must be able to discern debris from what is of interest. The typical algorithm used for jets in ATLAS is the *anti-k_t* algorithm. More detail about the detector is in section 2.4

2.2 Top-Quark Physics

One of the most interesting particles is the top-quark. Being so heavy, it cannot hadronize as it decays almost immediately and practically always into a W and a b quark. The unique signature and inability to create a bound state with other quarks means that the top-quark can be described with great precision.

Being an up-type quark, the top-quark has all the properties described in Sec 2.1 for this subset of quarks. However, its unique quality of fast decay is measured in the order of 0.5×10^{-24} s which is much shorter than hadronization.

Production

The top-quark can be produced alone by weak interaction or as a pair of top-antitop-quark by strong interaction. In order to best illustrate the mechanisms that allow the top-quark to manifest, physicists have a tool that gives a graphical representation of the mathematical description of all particle physics interactions; this tool is called *Feynman diagrams*. The depiction of fermions is a solid line with an arrow that points with time for particle and against time for antiparticle. Bosons are described as a wavy or dashed line (depending on the convention) with the exception of the gluon which is a spring-like line. Herein, the electroweak bosons are described by wavy lines. To add on, when *tree level* or *leading order* is used to describe a process, it means that the interaction has no higher order, or loop, corrections applied. These loop corrections can be done iteratively in higher orders for better precision and are often called *next-to-leading order* and adding *next-to* iteratively to denote higher and higher orders.

The most common way to produce top-quarks is in pairs. Top pair production can be done via gluon fusion ($gg \to t\bar{t}$) or quark-antiquark annihilation ($q\bar{q} \to t\bar{t}$). Both processes are shown in figure 2.4. In proton-antiproton colliders like the Tevatron, the latter production is most common. However, since the LHC is a proton-proton collider, it is more common for gluons to fuse and generate top-quark pairs. Theory predicts that the top-quark pair cross-section at $\sqrt{s}$ =13 TeV at the LHC is $\sigma_{t\bar{t}} = (832\,^{+20}_{-29}\,(\text{scale})\,^{+35}_{-35}\,(\text{PDF}))$ pb which includes next-to-next-to-leading order (NNLO) in QCD and the

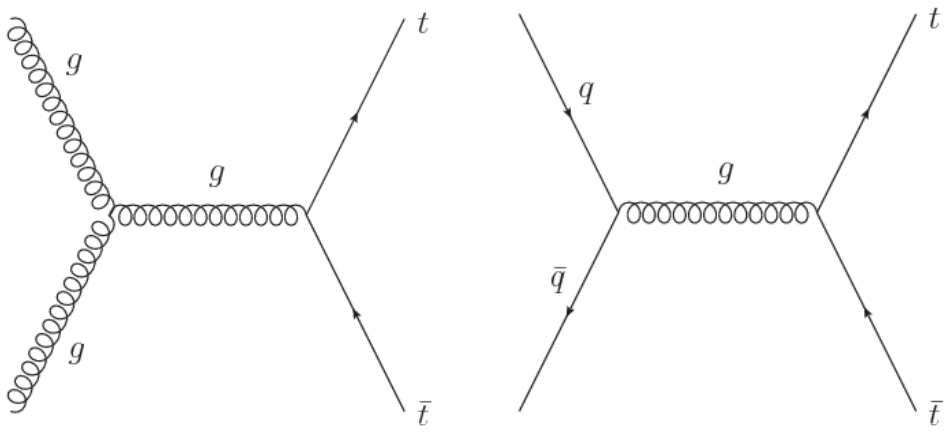


Figure 2.4: Feynman diagrams depicting a few of top-antitop-quark pair production mechanisms with (a) being gluon fusion and (b) depicting quark-antiquark annihilation

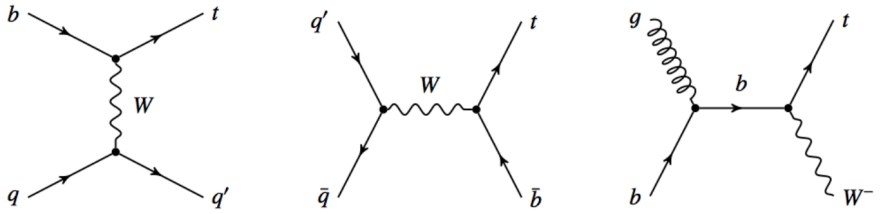


Figure 2.5: Feynman diagrams depicting the production of single top-quark via t-channel (a), s-channel (b), and t W channel (c)

resummation of next-to-next-to-leading logarithmic (NNLL) soft gluons and assumes the top-quark mass to be $m_t = 172.5\,\text{GeV}$.

The single top-quark production of interest for this analysis is called tW as it produces a top-quark in association with a W boson. However this is not the only way to produce a single top. More single top-quark production diagrams are shown in figure 2.5 where each will have its own signature, cross-section, and affinity towards some beyond the SM physics (which I will not cover in this thesis). Although, all of the single top-quark processes involve a Wbt vertex and can therefore be used to measure the V_{tb} element of the CKM matrix, of all these single top-quark productions, tW is the only one that will produce an on-shell[2] W. The cross-section for these processes at $\sqrt{s}$ =13 TeV are as follows:

$$\begin{aligned}
\sigma_{t-chan} &= (217\,^{+6.6}_{-4.6}\,(\text{scale})\,^{+6.2}_{-6.2}\,(\text{PDF}))\,\text{pb},\\
\sigma_{tW} &= (71.7\,^{+1.8}_{-1.8}\,(\text{scale})\,^{+3.4}_{-3.4}\,(\text{PDF}))\,\text{pb},\\
\sigma_{s-chan} &= (10.3\,^{+0.3}_{-0.2}\,(\text{scale})\,^{+0.3}_{-0.3}\,(\text{PDF}))\,\text{pb}.
\end{aligned}$$

[2] On-shell is when a particle field obeys the equation of motion and is a real particle.

These cross-sections include the top-quark and antitop-quark predictions added together. The fact that tW's cross-section is not even one tenth of the $t\bar{t}$ and one third of the t-channel cross-sections pose a difficulty in measuring tW. The addition of all single top-quark cross-sections are not even half of $t\bar{t}$ and this should point to $t\bar{t}$ becoming one of the main backgrounds for all single top-quark cross-section measurements.

At NLO, the tW channel can then have an intermediate top-quark that can become on-shell and therefore tW will have a similar signature to $t\bar{t}$. This interference is rather common at the LHC and there are two modeling methods to try and approach this interference. The two schemes are called Diagram Removal (DR) and Diagram Subtraction (DS) and are defined loosely in the following equations [5]:

$$\mathcal{A} = \mathcal{A}_{tW} + \mathcal{A}_{t\bar{t}}, \tag{2.8}$$

$$\begin{aligned} |\mathcal{A}|^2 &= |\mathcal{A}_{tW}|^2 + |\mathcal{A}_{tW}\mathcal{A}_{t\bar{t}}| + |\mathcal{A}_{t\bar{t}}|^2 \\ &= \mathcal{S} + \mathcal{I} + \mathcal{D}, \end{aligned} \tag{2.9}$$

$$DR : |\mathcal{A}|^2 = \mathcal{S}, \tag{2.10}$$

$$\begin{aligned} DS : |\mathcal{A}|^2 &= \mathcal{S} + \mathcal{I} + \mathcal{D} - \tilde{\mathcal{D}}, \\ &\simeq \mathcal{S} + \mathcal{I}. \end{aligned} \tag{2.11}$$

One can define the overall amplitude as the sum of the tW and $t\bar{t}$ amplitudes. However, the cross-section is calculated as the square of the amplitude and here is where the interference term will be introduced. The DR scheme calculates only the signal by *removing* the interference and $t\bar{t}$ terms from its calculation. In contrast, the DS calculates all of the amplitudes and then *subtracts* an estimate of the $t\bar{t}$ squared amplitude to cancel out the $t\bar{t}$ contribution. If both give similar results, then it can be seen that the contribution due to interference is small and the DR scheme has succeeded in modeling the tW decay channel correctly.

Top Decay

Earlier, it was mentioned that the top-quark decays almost always into a bottom quark and a W boson. This is due to the V_{tb} matrix element being very close to one, meaning that the top-quark decay will be a two-body decay. Given that the top-quark mass is more than twice the sum of W and bottom quark masses, it can produce both particles on shell. Moreover, the fast decay of the top-quark means that information based on spin can be drawn from its decay products since they will retain this information; therefore, top-quark polarization can also be studied with more precision than other quarks.

2.3 The Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton accelerator located near Geneva, on the French-Swiss border. It is one of the facilities used by the "Conseil Européen pour la Recherche Nucléaire" (CERN), or European Organization for Nuclear Research in English. The LHC lies in a tunnel that is nearly 27 km in circumference and is over 100 m underground. The current world record for center-of-mass energy achieved by any particle collider is held by the LHC which currently reaches 13 TeV by accelerating two proton beams in opposite direction at 6.5 TeV per beam. The tubes that carry these beams are kept at ultrahigh vacuum and are separated to prevent collisions outside of the detectors.

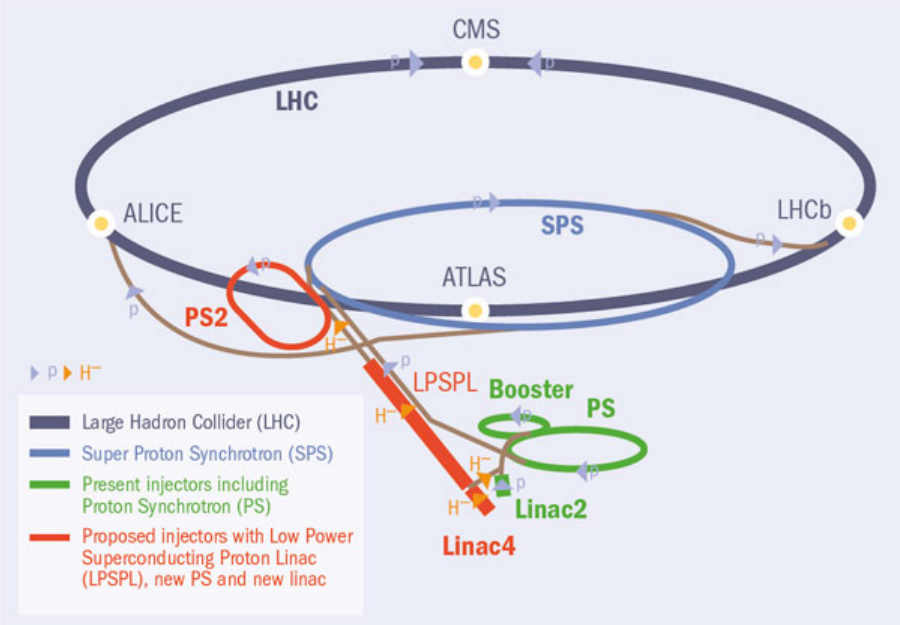


Figure 2.6: Schematic of the LHC and all its sequential accelerators [6]

Superconducting magnets are constructed to guide the beams around the accelerator ring and they are cooled to below 2 K.

The high energies provided by the LHC are designed to allow physicists to probe previously unseen physics, test BSM models like Super Symmetry (SUSY), and measure the properties of the Higgs boson. Currently, the LHC has provided over tens of petabytes per year of data; a major achievement in computing. As of 2017 over 170 computing centers in 42 countries participate in a grid-based network to provide the storage and analysis infrastructure for such a massive amount of data.

In figure 2.6, a schematic of different components in the LHC are illustrated. Before being loaded into the main ring, the protons need to be accelerated procedurally in smaller accelerators. Starting with the linear accelerator 2 (LINAC 2), which feeds 50 MeV protons to the Proton Synchrotron Booster (PSB). The PSB injects 1.4 GeV protons into the Proton Synchrotron (PS), until the beam reaches about 26 GeV. Finally, the Super Proton Synchrotron (SPS) accelerates them to 450 GeV and they are then injected to the main ring to finally reach the target energy of 6.5 TeV before collisions at the four designated interaction points where the experiments are housed.

The LHC houses many experiments which can be seen in figure 2.7. The major ones are: A Toroidal LHC Apparatus (ATLAS), Compact Muon Solenoid (CMS), LHC-beauty (LHCb), and A Large Ion Collider Experiment (ALICE). ATLAS and CMS are general purpose detectors. They are designed to study the Higgs boson, look for new physics, and perform precision SM tests. The LHCb seeks to understand matter-antimatter asymmetry. Lastly, the ALICE detector focuses on studying quark-gluon plasma in the interest of understanding the early universe.

This analysis uses datasets recorded by the ATLAS detector and as such it should be introduced in more detail.

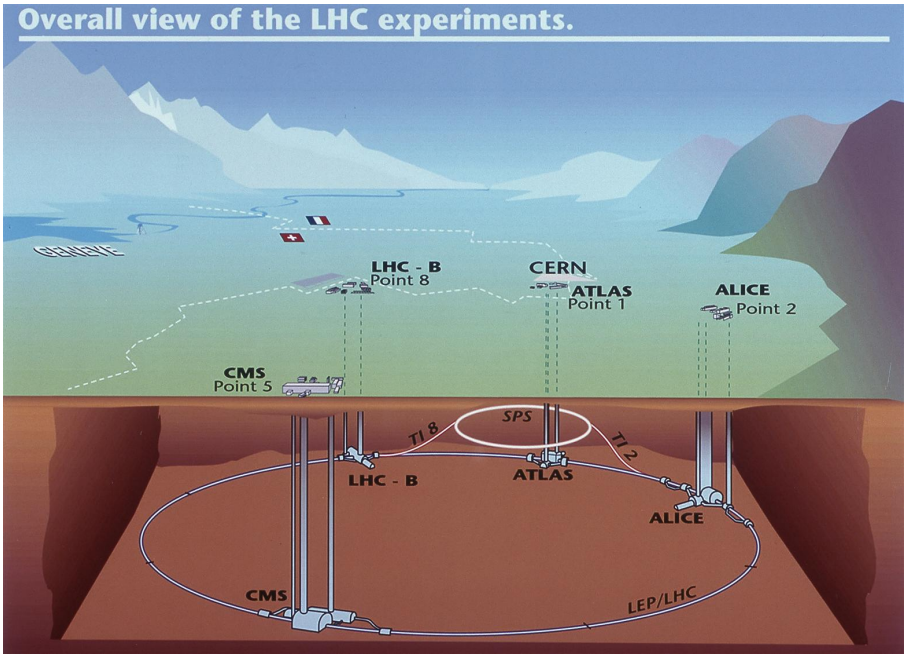


Figure 2.7: Drawing of the LHC and its four different detectors

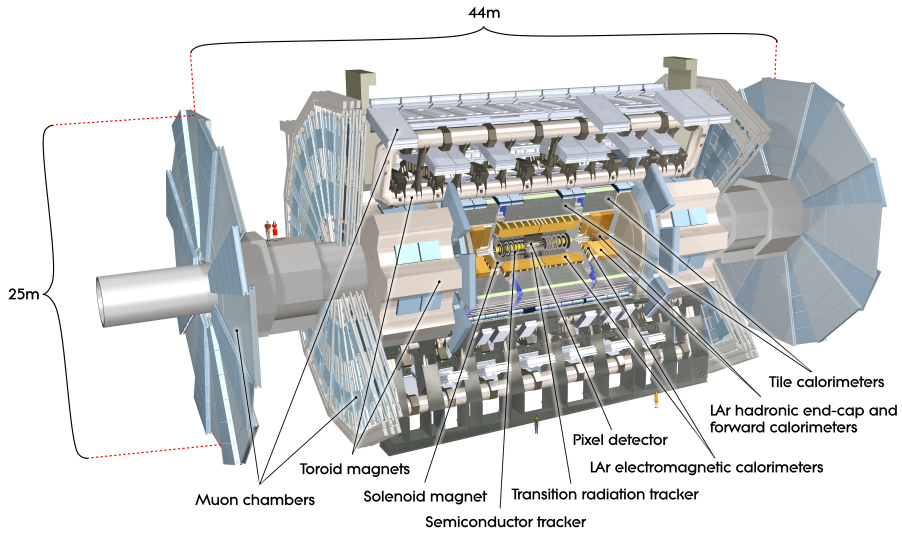


Figure 2.8: A detailed rendering of ATLAS and its components (labeled) [7].

2.4 ATLAS

The ATLAS detector is aptly named as its mythological counterpart was a colossal titan large enough to hold up the sky above all. ATLAS is the largest of all detectors, a true colossus by all definitions of the word. It weighs about 7 000 t, stands at 25 meters tall with similar width, and 44 meters long. The experiment collaboration has over 3 000 physicists as members from over 175 institutions in 38 countries.

In figure 2.8, the components that make up ATLAS are depicted and labeled. The ATLAS detector is

composed of smaller detectors that serve different purposes and are sensitive to distinct objects. These components are the inner detector (ID), electromagnetic and hadron calorimeters, muon detectors, and magnet system. All components that make up ATLAS encompass the beam pipe in a symmetric manner to provide as much coverage in all directions as possible. Each component will be detailed a bit more in the following subsections. More detail can be found in [8].

Inner Detector

The inner detector is the closest series of components to the interaction point, beginning only a few centimeters radially from the beam axis. The main function of the ID is to track charged particles as they pass through discrete points within it. The charged particles are then bent by the magnetic field that covers the inner detector and calorimeter system, which then gives information about the charge and momentum of the particle. Moreover, the information reconstructed in the ID can be used to associate particles with a vertex where they interact. As shown in figure 2.8, the ID is composed of a few other detectors that are the pixel detector (Pixel), semiconductor tracker (SCT), and transition radiation tracker (TRT). All subdetector systems cover up to a range of pseudorapidity $|\eta| < 2.5$.

The inner detector uses two types of particle detectors: silicon detectors (used in the Pixel and SCT systems) and straw trackers (in the TRT). A silicon detector uses an external electric field which creates an electric potential inside the silicon semiconductor. When a charged particle passes through[3], it knocks loose several electrons in the material creating electron-hole pairs that are attracted to opposite sides of the silicon due to the electric field. When the charges accumulates at the extremes of the silicon, they produce a current that flow into the electronics. Pixel detectors use this technology but the contacts at the extremes of the silicon are small, two-dimensional, and have separate circuits and electronics for the sake of precision. The SCT is very similar to the pixel detector but rather than individual pixels as contacts it uses a strip of material.

A straw tracker is a long tube that is filled with a gas which becomes ionized when a particle passes through it and a wire down the center of the cylinder. The wire and walls of the tube maintain a potential difference so the ionized gas particles will drift towards either wall or wire depending on the charge and generate a current. This current indicates a particle has passed within the straw and if enough straws are triggered, a track can be reconstructed.

The Pixel detector is the innermost part and contains three radial layers and three disk layers on each end-cap with the purpose of covering the forward regions. During the shutdown between Run 1 and Run 2, a new layer was added to the innermost part of the Pixel detector called the Insertable *B*-Layer (IBL) which improves track reconstruction [9]. Each other layer is composed of 1 744 modules, each module containing 16 readout chips and about 47 000 pixels. Each pixel has a size of $50\,\mu\text{m} \times 400\,\mu\text{m}$ and must be radiation hardened as they are so close to the interaction points. The intrinsic inaccuracies in the azimuthal ($R - \phi$) direction for all layers in both regions of the Pixel detector are 10 µm, 115 µm in the axial (z) direction for the barrel, and similarly for the disks along the radial R direction.

The SCT is the middle component and has a similar function to the pixel detector. However, instead of small pixels it has long, narrow strips measuring $80\,\mu\text{m} \times 12\,\text{cm}$. This means that it covers a larger area and is therefore a critical part of the inner detector. It has four double layers of silicon strips and measures over $60\,\text{m}^2$ in area. The intrinsic accuracy per module in the barrel and disks in the ($R - \phi$) direction are 17 µm, 580 µm (z) for the barrel, and in the disks 580 µm (R).

The TRT is the last component of the inner detector. The straw tracker are drift tubes that each measures 4 mm in diameter and can be as long as 144 cm. The TRT has about 298 000 straws in total.

[3] If the particle passes through the silicon without stopping then the detector is used as a tracker which is the intended purpose of the first two layers of the ID

The wire and walls of the tube maintain a potential difference of −1 500 V. In between the straws is the transition radiation component of the detector. Materials with varying indices of refraction cause ultra-relativistic charged particles to produce transition radiation and leave stronger signals in some straws. The amount of transition radiation depends on how relativistic a particle is with the greatest radiation coming from the lightest charged particles. By construction, the TRT only provides information in the $R - \phi$ plane and has an uncertainty of 130 μm per straw.

Calorimeters

The calorimeters can be found outside of the solenoid magnet that surround the inner detector. Unlike the inner detector, this section of ATLAS is not meant to be easily penetrated by particles. Instead, its intended function is to absorb particles that pass through the inner detector in order to measure their energy. The calorimeter is divided into two systems: the electromagnetic calorimeter (EM) followed by the hadron calorimeter on the outside. They both sample the shape of the resulting particle shower from absorbing an incoming particle. In contrast to the ID, these calorimeters cover a much greater range in pseudorapidity ($|\eta| < 4.9$)

The electromagnetic calorimeter measures the energy of charged particles and photons. It is divided into three parts that cover different η ranges: The central region is covered by the barrel in the range of $|\eta| < 1.475$ while the end-caps have two coaxial wheels each that cover $1.375 < |\eta| < 3.2$. The barrel's energy absorbing materials are lead and stainless steel, with argon as the sampling material. Using an accordion geometry for this detector, one can cover the complete polar range with no cracks in the azimuthal direction.

For the inner detector region ($|\eta| < 2.5$), the EM calorimeter is divided into three sections to improve precision. The strip section acts as a preshower detector and provides precision measurements in η and help with particle identification. All three regions give a high level of detail, with the highest coming from the strip $\Delta\eta \times \Delta\phi = 0.003 \times 0.1$. Predictably, the granularity in the end-caps is worse than in the inner region but it is still sufficient for jet reconstruction and missing transverse energy measurements.

The hadron calorimeter is intended to measure all particles that pass through the EM calorimeter and interact strongly. The detector is composed of steel to absorb energy and has scintillating tiles that sample in the barrel region $|\eta| < 1.7$ called the tile calorimeters. In the mid-outer region of $1.5 < |\eta| < 3.2$ there is a hadron end-cap that is composed of liquid argon like the EM calorimeter but uses copper as the absorbing material instead (LAr calorimeters). In the forward region of $3.1 < |\eta| < 4.9$, a high-density forward calorimeter is used. It is composed of copper in the first layer and outer layers are made of tungsten.

Muon Spectrometer

The outermost part of the ATLAS detector is the muon spectrometer. As muons typically penetrate most sub-detectors without depositing much energy in them, this component is meant to identify muons and measure their momenta. The detector is composed of three components: three toroidal magnets that provide a magnetic field, several chambers that measure the tracks of outgoing muons, and a set of triggering chambers with accurate time-resolution.

The magnetic force from the toroidal magnets will cause muon tracks to bend much like in the inner detector and with the same purpose of measuring momenta. This bend is tracked by drift tubes similar to the TRT but with resolution of 80 μm in the inner chamber and 60 μm in the forward region. The Thin Gap Chambers (TGC) trigger when a muon passes through in the forward regions. In the barrel

region, the Resistive Plate Chambers (RPC) are located for the same purpose as the TGC. Their combined coverage range is limited to $|\eta| < 2.4$.

Magnet System

The magnet system is composed of two components which are superconducting magnets: the central solenoid magnet and a toroidal system. The first one is located right outside the inner detector and provides a field strength of 2 T. The strength of the magnetic field is necessary as the bending is inversely proportional to the energy of the particle but directly proportional to the field strength. This forces high energy particles to bend and have their momentum measured.

The toroid magnet system is composed of three parts with one of them in the center region and the other two in the end-caps. This magnet has eight separate coils that generate maximum magnetic fields of 4 T. The toroidal magnets need to be placed in a cryostat as their working points are a maximum of 4.5 K [10]. Because of their large size, the barrel toroid has a separate cryostat per coil.

Triggers and Data Acquisition

As mentioned before in section 2.3, ATLAS produces an incredible amount of data and not all data is of interest. In order to keep only the most interesting events, the trigger and data acquisition system (TDAQ) manages to select about one event per every 200 000 that is interesting for physics. The TDAQ selects events that pass two triggers. In Run 1, the last trigger was two separate triggers that were merged to reduce complexity and allow for dynamic resource sharing. More detail in [11]

The first trigger is entirely hardware based. This part of the selection process defines interesting based on information from the components of detectors that have reduced granularity. Typically, these are high p_{T} objects, τ leptons decaying into hadrons, and large missing and total transverse energy. This trigger segment reaches a conclusion about an event in only 2 µs after the interaction takes place. Events that pass this trigger have their full detector information read from the electronics into readout drivers and then into readout buffers.

The second trigger filters events at 75 kHz based on reconstruction of the collision data. It then decides if the event is worth keeping purely on the information given by the first trigger and performs a full reconstruction. Events that do not pass this trigger are discarded and all who pass are then transfered to storage associated with the event filter. The latency of this component is between 1 to 10 ms. Lastly, it makes the final selection to keep the events for offline analysis. The information is reduced by a factor of 10 from the second trigger level. In the end of all three triggers, the data rate of recorded events reaches about $\sim$300 MB s^{-1}.

2.5 Reconstruction and Identification of Objects

As mentioned previously, ATLAS is composed of several layers that identify particles and are able to measure their kinematic properties. Figure 2.9 shows how some well known particles are detected and where they are measured. To reiterate, all charged particles will leave a curved track in the inner detector which will yield information regarding their momentum and charge. Neutral particles will leave no track here since they are unable to ionize gas in the TRT or interact with the Pixel/SCT detectors. Electrons and photons will deposit most, if not all, of their energy in the EM calorimeter while muons will deposit very little energy and carry on to the muon spectrometer and even past that. Charged hadrons will deposit some of their energy and continue onwards to be stopped in the hadron calorimeter. Neutral hadrons will not deposit any energy in the EM calorimeter and will instead be absorbed in the hadron calorimeter.

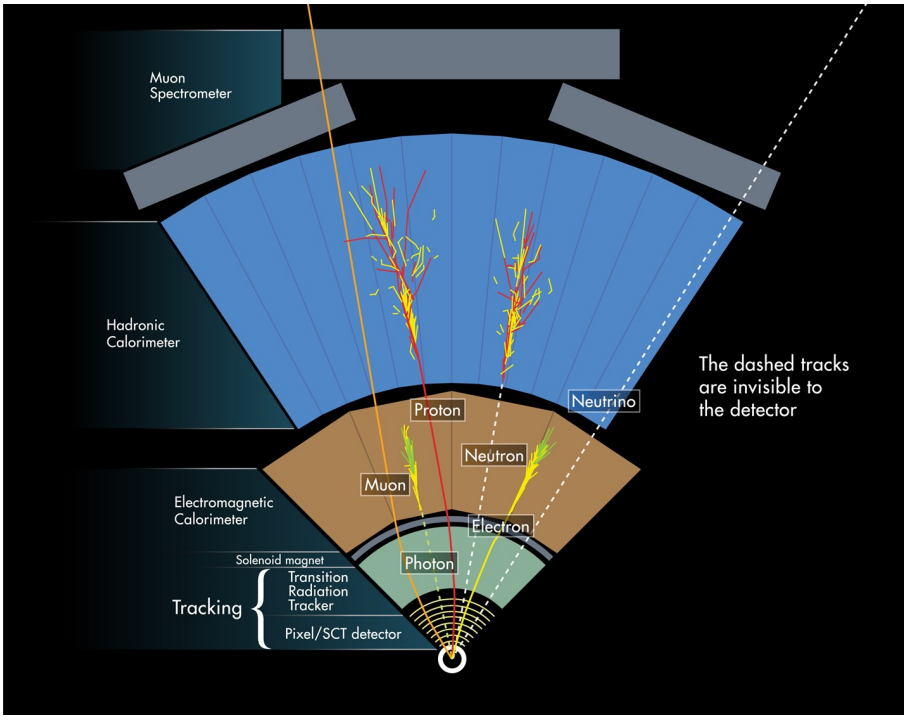


Figure 2.9: Cross sectional image of the ATLAS detector which shows where some well known particles are detected [12].

Neutrinos, unlike all the other particles mentioned, will not interact with any components and instead are indirectly measured as missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$.

The tW channel has decay objects that are mostly detectable directly by ATLAS. As such, they will be explained more in depth in the following sections. It should be noted that for this analysis, leptons refers to muons and electrons only since taus will not be taken into consideration.

Electrons

In order for an object to be considered an electron, it must have a EM calorimeter deposit associated with ID tracks, p_{T} of at least 20 GeV and be in the inner detector region ($|\eta| < 2.47$) excluding the pseudorapidity range $1.37 < |\eta| < 1.52$. Other properties of the event like EM shower shape and ratio between calorimeter energy to tracker momentum are taken into account to calculate a likelihood-based discriminant [13].

Other criteria for a candidate electron are based on the distance from the ID track to the reconstructed primary vertex. Primary vertex is defined as the vertex with largest summed p_{T}^2 of all associated tracks. The distance to the beamline can be described with two variables. The transverse impact parameter, d_0, must satisfy $|d_0|/\sigma_{d_0} < 5$, where σ_{d_0} is the uncertainty in the impact parameter. The longitudinal impact parameter, z_0, must satisfy $|\Delta z_0 \sin\theta| < 0.5$ mm, where Δz_0 is the longitudinal distance from the primary vertex and θ is the angle of the track with respect to the beam line.

Lastly, electron candidates should be sufficiently isolated. Track isolation is calculated from the sum of p_{T} of all objects in a ΔR cone that are not the candidate divided by the p_{T} of the electron candidate. Lower values for isolation therefore correspond to greater isolation. Calorimetric isolation is defined as the sum of transverse energies deposited in the calorimeters within a ΔR cone. More detail can be found

in [14].

Muons

Muon candidates are identified by matching tracks in the muon system with ID tracks. The matching depends on momentum measurement and track compatibility. Similarly to the electron candidates, these objects must have p_{T} of at least 20 GeV and be in the inner detector region ($|\eta| < 2.5$). Also, they must be submitted to similar but slightly looser criteria of distance to a primary vertex. The transverse impact parameter must satisfy $|d_0|/\sigma_{d_0} < 3$ and otherwise the criteria for z_0 is the same. Isolation is also required for these objects similar to electron candidates.

Jets

Jets are reconstructed using the anti-k_t algorithm [15] with a R parameter of 0.4. Jets are rejected if they are within $\Delta R < 0.2$ of an accepted electron, must be located within the inner detector ($|\eta| < 2.5$), and have $p_{\mathrm{T}} > 25$ GeV. Their energies are corrected to account for pile-up[4] and calibrated using corrections derived from data driven methods in Run 2.

There is a specific type of jet that will be used in this analysis and is a key component of the tW signal. These jets contain a b-hadron and are therefore called b-tagged jets. b-tagging is done by using the long lifetime of hadrons containing b-quarks that decay weakly and the large invariant mass of their decay products relative to other light hadrons to generate a multivariate discriminant [16, 17]. The b-tagging efficiency used for this analysis is 77 %. This means that out of all the b-jets, 77 % are correctly identified while the rest are not. Not only is b-jet "loss" possible but also mis-tag a light jet as a b-jet however this is much more unlikely. The reported rejection factors per quark is about 4.5 for c-jets, and 140 for light and gluon jets [18].

Missing Transverse Energy

Missing transverse energy is a quantity that cannot be directly measured, instead it must be calculated as the negative vectorial sum of the transverse momentum of all particles in the event. $E_{\mathrm{T}}^{\mathrm{miss}}$ is then calculated from the information in the calorimeters and muon system. This is the method by which one can measure the energy of neutrinos that escape detection.

[4] In the LHC, protons are not accelerated in bunches where multiple collisions can take place at once. This is referred to as *pile-up* and it is caused by additional proton-proton collisions within the same bunch crossings or in previous/subsequent bunch crossings.

CHAPTER 3

Event Selection

In this analysis, the single-lepton decay mode of the tW production channel is investigated. Section 3.1 covers the topology of signal and background processes, explores the differences between a dilepton and single-lepton analysis, and explores how the backgrounds can bleed into the signal region. More detail on datasets and Monte Carlo simulations can be found in sections 3.2 and 3.3 with technical details in appendix A. Multi-jet, or fake-lepton, background estimation is shown in section 3.4. Finally, cut optimization will be detailed in the last section (3.5) where cutting on some kinematic values is shown to improve the signal over background ratio.

3.1 Signal and Sources of Background

3.1.1 The tW Decay Channel

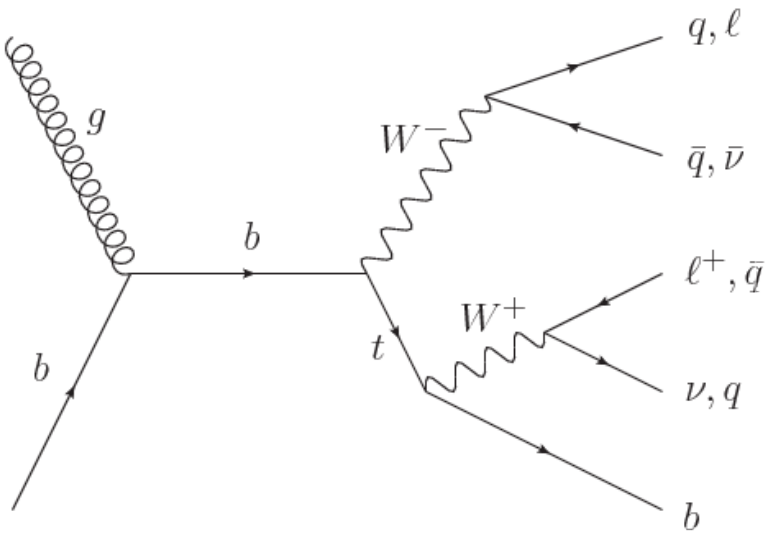


Figure 3.1: Feynman diagram of the tW channel single-lepton decay mode.

In section 2.2, it was shown that the tW channel includes an on-shell W boson and a top-quark which will also decay weakly into a W boson and b-quark. It is known that the W boson can decay hadronically ($q\overline{q}'$) or leptonically ($\ell\nu_\ell$) via the mechanisms described in bosons subsection of section 2.1. So the

tW decay will have an intermediate step of bWW that will then culminate in the final topologies of $bq\bar{q}'q''\bar{q}'''$ for a purely hadronic mode, $bq\bar{q}'\ell\nu_\ell$ for the single-lepton mode, and $b\ell\bar{\nu}_\ell\ell^{+\prime}\nu_\ell'$ for the dilepton mode. These modes are depicted via a tree level Feynman diagram in figure 3.1. The branching ratio for $t \to Wb$ is nearly 100 %. The W boson will decay hadronically most often as its branching ratio is about (67.60 ± 0.27) % where as the leptonic mode has about half the probability with (32.57 ± 0.28) % [3].

The branching ratio for tW to decay with only one lepton is ~44.0 % which is comparable to fully hadronic decay (45.7 %). The lowest branching ratio then goes to the dilepton channel which is 10.6 %. The fact that the single-lepton mode has a comparable branching ratio to the fully hadronic decay means that one gains the same amount of statistics from either mode but the single-lepton has the advantage that it has fewer jets in the final topology. Jets, in general, are dominated by experimental systematics and this is further explained in section 5.2. Furthermore, the dilepton channel has a branching ratio that is one fourth that of single-lepton or full hadronic decays. Thus, the dilepton channel will have less statistics.

The dilepton channel has a few advantages over the other two decay modes. Primarily, it does not deal with as many jets and has a much cleaner signal. The signal would consist of a high amount of $E_{\mathrm{T}}^{\mathrm{miss}}$, two well reconstructed leptons, and a b-tagged jet. Furthermore, instead of being plagued by multi-jet and boson + jets backgrounds, it only has $t\bar{t}$ as a main source of background which makes separation easier.

The single-lepton decay channel has one main advantage over the dilepton channel. Since the dilepton has two neutrinos in the final state, it cannot easily reconstruct the two W bosons and therefore the top-quark kinematics are harder to calculate. With a single-lepton decay, the reconstruction of the neutrino is much easier and therefore it is possible to attain the top-quark kinematics and perform a differential cross-section analysis with less struggle.

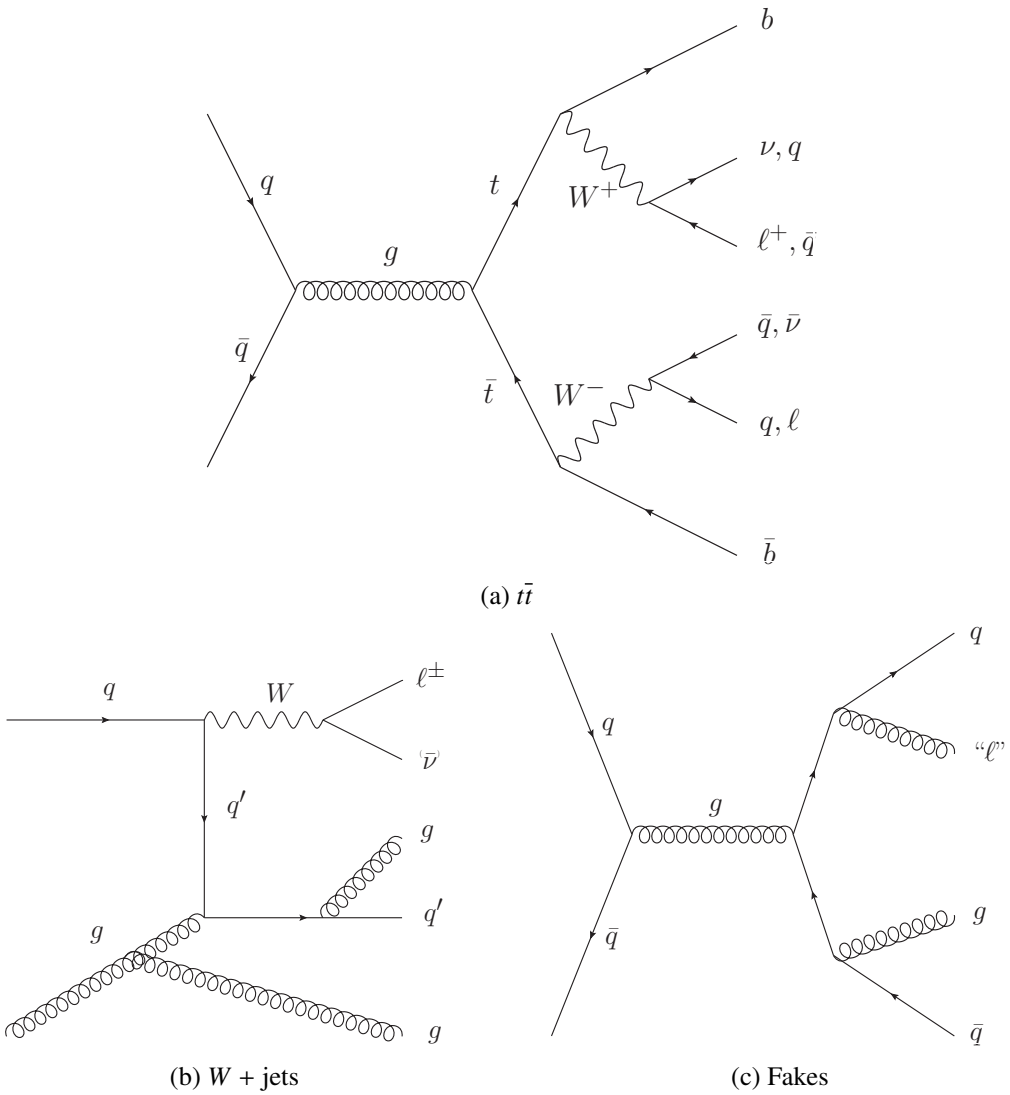
For the tW single-lepton analysis, the final state will be composed of three jets of which one will be b-tagged, one lepton, and some missing energy which is associated to the neutrino. It is worth noting that for this analysis the word “lepton” refers only to muons and electrons as mentioned previously in section 2.5. This is due to the fast decay time of the tau lepton. It is able to decay hadronically or leptonically via weak interaction and thus it would make the signal region harder to define, measure, and separate from background processes. This lowers the branching fraction for a single lepton tW channel decay from about 44 % to almost 30 %.

3.1.2 Backgrounds

Background processes can often appear signal-like. They can share topological features, have objects misidentified by reconstruction software, or have similar final states. The following sections will list the backgrounds by importance and detail them further.

Top-Quark Pair

Top-quark pair production is the most difficult background to separate from the signal. With a cross section of over ten times that of tW, it becomes the main source of events that dwarfs the tW by comparison. To make matters more difficult, the $t\bar{t}$ final state will look identical to tW if one of the two b-jets is not properly tagged. This is easy to see in figure 3.2(a) where the $t\bar{t}$ final state is shown. Furthermore, both $t\bar{t}$ and tW share several properties in their topology, e.g. they both contain on-shell W bosons that can be well reconstructed and the mass of a top-quark can be extracted from the event kinematics.



(a) $t\bar{t}$

(b) W + jets

(c) Fakes

Figure 3.2: Feynman diagram of the main backgrounds for the tW single-lepton channel with final states.

W Boson in Association with Jets

In figure 3.2(b), a Feynman diagram depicting a W + jets final state is shown. It can be seen that in order for the W + jets background to pass the selection, the W boson needs to decay leptonically and three jets must be radiated. From these three jets, one either must be a b-jet or be mis-tagged as such. The W boson must decay leptonically in order to produce a clean lepton and enough missing energy. Even with so many prerequisites to appear like the signal, its large cross-section makes it impossible to ignore. The W + jets process has the second largest cross-section of all backgrounds and is orders of magnitude larger than the tW cross-section. It is reported [19] that the W + jets cross section at $\sqrt{s} = 13\,\text{TeV}$ is $20.64 \pm 0.02(\text{stat.}) \pm 0.55(\text{sys.}) \pm 0.43(\text{lumi.})\,\text{nb}$.

Even with such a massive cross-section, this process rarely is associated with b-jets and typically has low jet multiplicity. It is also worth noting that the kinematics will help to separate signal from this background. For example, one of the main bodies that can be reconstructed in the tW single-lepton channel is a W boson that decays hadronically while W + jets will not have a well reconstructed hadronic

W boson.

Multi-jet (Fake Leptons)

Figure 3.2(c) shows one of the many ways that a multi-jet event can appear like the tW channel. It can be seen that the multi-jet background will not have any on-shell weak bosons or top-quarks, it will typically not have much missing transverse momentum (unless a jet is discarded or mis-reconstructed), and no real lepton. However, one of its jets may be misidentified as a lepton thus it is sometimes called the fake lepton background. It is incredibly unlikely that a purely QCD process would generate jets that would be mislabeled and so this background would normally be irrelevant. However, the cross-section for such events is incredibly large, even orders of magnitude above W + jets. The high frequency for these type of events to happen makes the multi-jet background very relevant to the analysis.

Another consequence of the high cross-section is the inability for this process to be modeled by MC simulation. A data driven method, named Matrix Method, will instead be used to estimate the fakes contribution. More on this method can be found in section 3.4.

Z Boson in Association with Jets

The Z + jets has a cross section similar to that of W + jets. However, considering that the Z boson can only decay into two opposite sign leptons, two neutrinos, or two quarks makes it easy for this background to be separated. It also has low jet multiplicity most of the time. This means that requiring three jets and one lepton removes a significant amount of this background. In order for a Z + jets process to appear like a single-lepton event, one of its leptons must be misidentified or outside the relevant part of the detector and some object must be badly reconstructed in order to appear as missing energy.

Single Top-Quark (t-channel and s-channel)

The t-channel has a cross section about three times that of tW and as a single top-quark event it has a similar final state. However, the main difference comes from the W boson. Being only used to change flavors and transfer energy, there is no W decay that does not come from the top-quark decay. Thus one jet would be missing and so it becomes one of the minor backgrounds in this analysis.

Similarly, the s-channel will only have one W to decay so it will be easily separated. This in combination with its very small cross section will yield very small amount of events in the signal region and thus it is a minor background.

Diboson (WW, ZZ, and WZ)

Diboson refers to an event where two W bosons, two Z bosons, or a combination of both are produced. Some diboson events may mimic the signal very well. In particular, the WW can have identical end states to tW with one W boson decaying hadronically and the other leptonically. The difference being a b-quark is absent. The lack of a b-quark in the final state mitigates most of this background. In ZZ events, the same arguments for Z + jets applies and so the contribution is minimal. Lastly, the mixed boson event can look similar to tW with a hadronic decaying Z boson and leptonic W boson. This particular diboson mode also is unable to produce a b-jet and therefore is easily separated. Altogether, the contribution from these processes are very small and so they are a minor background of this analysis.

3.2 Datasets

The data sets used in this analysis were collected by the ATLAS detector in the years of 2015 and 2016 with a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and total integrated luminosity[1] of $\mathcal{L} = 36.1\,\text{fb}^{-1}$. The average number of collisions per bunch crossing $\langle\mu\rangle$ for this period of data-taking was about 23 collisions. Data is only used for physics analysis when all detector systems are known to be operating normally.

3.3 Monte Carlo Simulation

Monte Carlo (MC) simulations are tools that model individual processes to be compared with data. Monte Carlo generators will take physical parameters, such as the masses of particles, and create predictions for physical processes. The parameters used in generators relevant to this analysis are masses and decay widths for the top-quark, W and Z bosons. These are $m_t = 172.5\,\text{GeV}$, $\Gamma_t = 1.32\,\text{GeV}$, $m_Z = 91.2\,\text{GeV}$, $\Gamma_Z = 2.09\,\text{GeV}$, $m_W = 80.4\,\text{GeV}$, and $\Gamma_W = 2.50\,\text{GeV}$.

One can also estimate uncertainties as the ones introduced by the limited resolution of the detector, and as a tool for validation. There are two type of generated MC sets: those who simulate the events as well as the detector in its full capacity (commonly referred to as "full sim") and fast simulations (fast sim, AtlFast2 or AFII) that use a simplified detector simulation. All samples used in this analysis are full sim. Full sim sets are used as the nominal samples that the analysis compares to data. Furthermore, they are used to estimate the experimental uncertainties introduced by the detector, such as jet energy scale (JES), jet energy resolution (JER), lepton identification uncertainties, calorimeter calibration, etc. Fast sim, on the other hand, are used to measure modeling systematics. Since they do not fully simulate the detector response (the most time consuming part of the process), several sets can be generated much faster and in greater numbers to achieve several tasks, e.g. neural network or boosted decision trees training as done in the single-lepton analysis at $\sqrt{s} = 8\,\text{TeV}$ [20]. Another source of error that can be estimated is the difference between generators. Generating two sets of the same process and either applying a full sim or fast sim to them will allow one to compare generators and establish an uncertainty to them.

There are four levels of simulation: full, library, fast, and parametric. The standard (full) simulator is called Geant4. It uses the full description of ATLAS and most precise simulation. The next tier of simulation uses what is called Frozen Showers. Simulated events in the forward region use up a significant amount of the CPU. To mitigate this, a library of pre-simulated showers is created so that when some particles fail to exceed an energy threshold, they are replaced with a pre-simulated shower from the library. This is mainly used in the forward calorimeter even though it is implemented for all parts of the LAr calorimeter. The last relevant tier is where AFII takes place. It uses a parametrized version of the calorimeters to duplicate energy profiles with a fine E/η grid [21]. It can simulate the energy response, average lateral shape, and energy fractions. It cannot simulate particle decays nor hadronic leakage into the muon spectrometer.

Physical processes like pile-up are introduced by overlaying other collisions from QCD generators. For all single top-quark and top-quark pair MC samples, Powheg [22] was used in conjunction with Pythia 6 [23] as the pile-up generator. Sherpa2.2.1 [24] is used to generate the W and Z + jets samples. Similarly to the top-quark samples, Powheg is used as a generator for diboson events with Pythia 8 [25]. Both Pythia generators as well as Sherpa are multi-purpose, leading-order (LO) generators. Pythia

[1] Instantaneous luminosity, often called just luminosity, is defined as collision rate divided by the cross-section of the interaction. Integrated luminosity is the total number of events divided by the cross-section of the interaction, or instantaneous luminosity integrated over time. It should be noted that the cross-section for calculating accelerator luminosity is not the cross-section of any process in particular but the probability for protons in the colliding bunches to interact.

generates QCD events, SHERPA is a multi-parton generator that includes hadronization, and POWHEG is a next-to-leading order (NLO) generator that creates NLO QCD events. After generation, events are reweighted so they match theoretical calculation of the total cross-sections.

Top-Quark Production

Nominal as well as experimental systematic simulations were created using the POWHEG-BOX v1 with the CT10 PDF set for the matrix-element calculations and PYTHIA 6 to generate the parton shower, hadronization, and pile-up. The tW sample had two production schemes that were referenced earlier at the end of section 2.2. The DR scheme is used to handle the interference between $t\bar{t}$ and tW while the DS scheme is made as a check to estimate the uncertainty in modeling the signal process. All top-quark samples used in this analysis are the nominal POWHEG with PYTHIA 6 full simulations. More information about the generation of top-quark samples can be found in [26]

Boson + Jets and Diboson Productions

Both W and Z + jets as well as diboson MC simulations were generated with SHERPA2.2.1 and put through the full GEANT4 simulation. These are divided by jet collection and leptonic decay as well. Each set is contains a boson decaying leptonically, to any of the three leptons, and is labeled to denote the flavor composition of jets associated with it. The jet flavor labels are light (for u-, d-, and s-quarks), charm-quark, and bottom-quark. Jet collections labeled with bottom- and charm-quarks will contain at least one jet with the noted flavor and may contain extra light jets.

Most diboson decay modes are considered in this analysis. The sets not included are with purely hadronic decays from both bosons. All sets are subjected to the same criteria regarding number of jets, leptons, and b-tagged jets so the dilepton and trilepton samples will be filtered out.

3.4 Modeling the Fake Lepton Background

Multi-jet backgrounds are incredibly difficult to model as the high cross-section means a gargantuan number of events would need to be generated costing too much storage space and CPU time. Instead of simulated events, the Matrix Method for estimating fake lepton contribution is employed. It is a data driven technique which was already employed for Run 1 in ATLAS [27].

As mentioned in section 2.5, lepton candidates undergo several criteria to be identified as a lepton in the event. A combination of conditions must be met in order for a lepton candidate to be considered *tight* and therefore part of the nominal sample.

Tight electrons must: pass the tight likelihood explained in detail here [13], have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $|\eta| < 2.5$, clusters must not be in crack[2] , candidates must pass gradient isolation defined in [14]. Electron candidates that only pass the medium likelihood [13] threshold are still recorded but are instead called *loose*.

Tight muons, on the other hand, need to pass the medium identification threshold [14], same transverse momentum and isolation requirement as electron candidates, and have $|\eta| < 2.4$. Loose muons need only pass the medium ID criteria.

There are objects that can pass these requirements without being prompt leptons[3]. For example, semileptonic decays of mesons in jets may cause the identification of the lepton rather than the jet.

[2] This is the area of the detector between the barrel and end-cap. In the pseudorapidity range of $1.425 < |\eta| < 1.5$.

[3] Prompt, also referred to as real, leptons are those which come from the decay of the W boson associated with the top-quark.

Photon conversion, cosmic rays, and jets that punch-through to the muon chamber can pass the tight lepton requirements and become fake leptons.

Theory

Consider a data sample that contains only single-lepton events. One can define the number of tight or loose as the sum of fake and real leptons within the selection:

$$N^l = N^l_r + N^l_f, \tag{3.1}$$

$$\begin{aligned} N^t &= N^t_r + N^t_f, \\ &= \epsilon_r N^l_r + \epsilon_f N^l_f, \end{aligned} \tag{3.2}$$

where the superscript N^t and N^l denote the total number of events that pass the tight and loose selection respectively, the subscript r and f denote real and fake leptons, respectively. Tight leptons are a fraction of loose events and can be expressed as some efficiency, ϵ, of loose events passing the tight selection.

After reorganizing the above equations, one can then estimate the number of fake leptons that pass the tight selection as a function of both efficiencies and the total number of loose and tight events.

$$N^t_f = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}(\epsilon_r N^l - N^t). \tag{3.3}$$

The efficiencies depend on event kinematics and characteristics. These can then be parametrized as a function of various properties of the event like number of b-tagged jets, pseudorapidity and p_{T} of the leading jet, lepton or b-tagged jet, etc. To account for the effects of kinematics, one can calculate an event weight from the efficiencies:

$$w_i = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}(\epsilon_r - \delta_i), \tag{3.4}$$

$$\delta_i = \begin{cases} 1 & \text{if the loose event passes the tight selection,} \\ 0 & \text{else.} \end{cases}, \tag{3.5}$$

the background estimate of the total fake contribution will then be the sum of all the weights.

The definitions for loose and tight must be sufficiently different or the $(\epsilon_r - \epsilon_f)$ factor in the denominator will cause a huge number of tight fake events. This would cause the Matrix Method to fail.

Efficiencies can be measured in control regions where the purity of fakes or real leptons is clearly defined. Table 3.1 gives a brief summary of the different regions and what efficiency can be measured.

Channel	n_{jet}/n_{b-jet}	Kinematic Cuts	Efficiency measured
ℓ + jets	≥ 1 jet, no tag	$m_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) < 20\,\mathrm{GeV}$, $E_{\mathrm{T}}^{\mathrm{miss}} + m_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) < 60\,\mathrm{GeV}$	ϵ_f
$\ell\ell$	≥ 1 jet, no tag	Opposite charge sign, $80\,\mathrm{GeV} < m(\ell,\ell) < 100\,\mathrm{GeV}$	ϵ_r

Table 3.1: Brief summary of control regions to measure the real and fake efficiencies used in Matrix Method [27].

In this analysis, the efficiencies were not measured. Instead, the Top Fakes group provides efficiencies to be used in top-quark analyses. This reference [28] details the process for generating the efficiencies for

$\sqrt{s} = 13$ TeV. The provided efficiency files used in this analysis can be found in the `TopFakes-00-00-19` package. The package provides standard efficiencies for the data release version used in this analysis.

Limitations

After the weights have been applied to the loose data set, one can see the fake estimation in 3.3. When lepton flavor is not accounted for, the agreement between data and Monte Carlo is reasonable. However, this does not hold true when flavors are assessed individually. In particular, the muon selection is underestimated as can be seen in 3.4(b). The systematic uncertainty shown stack plots such as figure 3.3 is further explained in section 5.2.

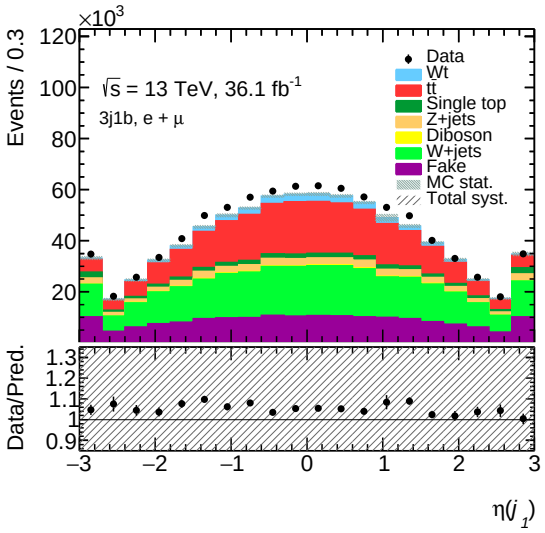
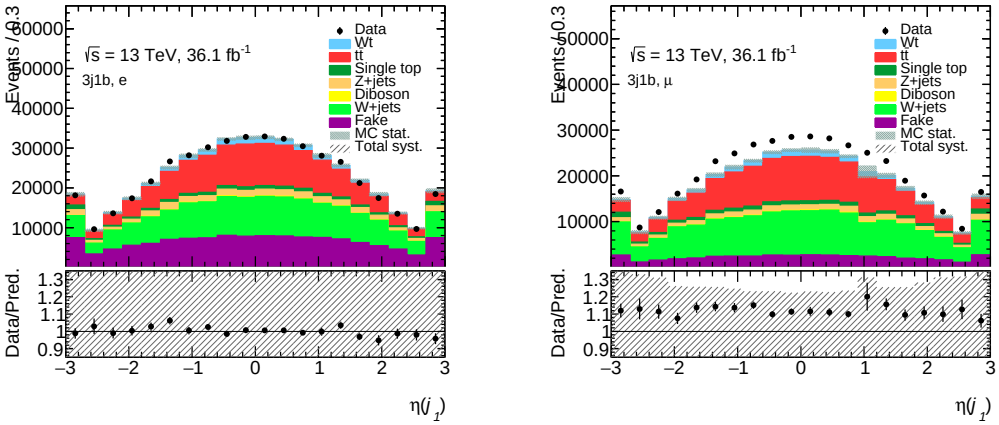


Figure 3.3: η of the leading jet. No cuts have been applied aside from lepton, jet, and b-jet number requirement. The systematic uncertainty shown in this plot is further explained in section 5.2

This behavior is believed to be limited to analyses that require b-tagging. The current reason for this is unknown and under investigation by the Top Fakes group. However, a few possible reasons have been proposed: there may be some correlation between parameters that the Matrix Method is unable to account for, or the efficiencies are incompatible with single top-quark research. Even though the muon channel is underestimated, the overall is still compatible with observed data and thus this problem should not have a massive impact in the overall analysis.

3.5 Cut Optimization

In order to improve the overall performance of the analysis, an attempt is made to maximize the number of signal events while minimizing background. Simply put, if an observable can show differences between the signal and some background processes, a cut can be made on the observable such that the events cut away are mostly background. One can perform scans for kinematic variables where values are cut away at different intervals. After each cut, the event yields are used to calculate the ratio of signal to the square root of total background events, sometimes referred to as *significance*. This can then be done sequentially, from variable to variable, until the background is minimized with only minimal losses in signal events.



(a) e only (b) μ only

Figure 3.4: η of the leading jet separated by lepton flavor. No cuts have been applied aside from lepton, jet, and b-jet number requirement.

This is illustrated in figure 3.5, where the leading jet is clearly softer, or has lower transverse momentum, in the W + jets collection. Therefore accepting events with a cut on transverse momentum of the leading jet will lead to a great reduction of W + jets background.

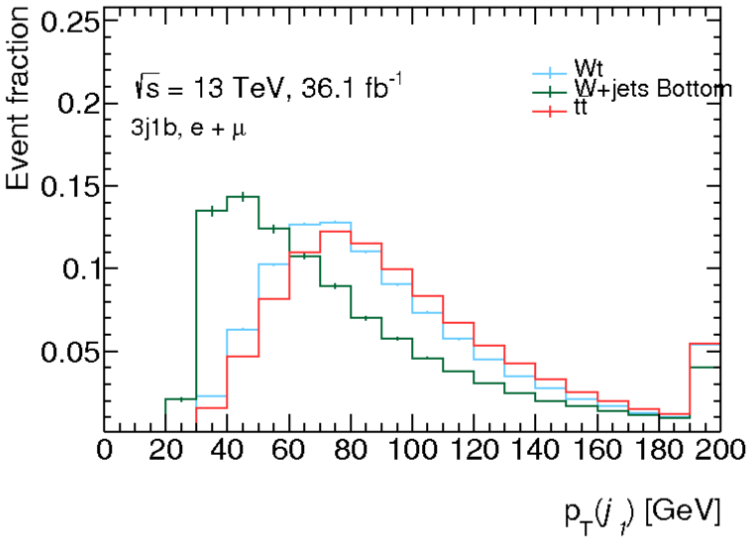


Figure 3.5: p_{T} of the leading jet. The b-jet associated set of W + jets, tW and $t\bar{t}$ are overlaid and normalized to show shape differences for the purpose of cutting.

In figure 3.6(f), one can see that the low m_{T} regime is where most of the fakes reside. Performing a high enough cut here will reduce their contribution. Since the tW channel will have one W boson that decays leptonically by definition, most of the events will accumulate at the W boson mass. It is beneficial to cut as many fakes as possible since they have the highest uncertainty and are hardest to estimate. Similarly, a fake jet will not have a neutrino in the final state and thus it will have little to no missing

transverse energy. Cutting on this variable will also reduce this background.

In general, the p_T and η of jets depends on the kinematics of whichever process they decay from. For example, the t-channel has a prominent signature as one of its leading jets is in the forward region of the detector. In contrast, all objects from the tW decay tend to be in the inner region of the detector. An example was already shown in section 3.3 where the tW is most noticeable in the $|\eta| < 2.5$ region. As mentioned earlier, jets coming from W in association with jets tend to be softer than ones from tW and top-quark pair processes. This is due to the jets coming from radiation or other interactions rather than the decay of a very heavy particle. Lastly, low p_T jets introduce a great amount of uncertainty and so cutting them as much as possible will improve the overall analysis.

The starting point for the cuts in this analysis will be the ones stated in Sebastian Mergelmeyer's thesis [20] which was a similar analysis but at $\sqrt{s} = 8\,\text{TeV}$. A scan is then performed to further optimize signal to background ratio. The cuts used in the 8 TeV analysis were: the lepton and all jets should have $p_T > 30\,\text{GeV}$, $E_T^{\text{miss}} > 30\,\text{GeV}$, and $m_T(\ell, \nu_\ell) > 50\,\text{GeV}$.

After optimizing cuts, the following cuts are implemented in this analysis: leptons and jets must be in the inner detector region $|\eta| < 2.5$. The transverse momentum of the lepton must be above the threshold for it to be considered a real lepton (i.e. $p_T(\ell) > 27\,\text{GeV}$), the event is vetoed if second real lepton is identified and has $p_T(\ell_2) > 25\,\text{GeV}$. The leading jet must have transverse momentum greater than 40 GeV, the sub-leading jet must also have $p_T(j_2) > 30\,\text{GeV}$, the softest jet must pass the threshold for reconstruction of $p_T(j_3) > 25\,\text{GeV}$, and one of the three jets must be b-tagged. Missing transverse energy must be greater than 30 GeV, and the transverse mass of the lepton-neutrino system must be greater than 40 GeV.

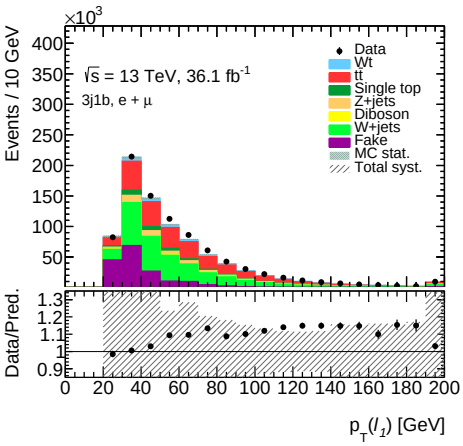
The plots shown in figure 3.6 have no cuts and show graphically the effect of uncertainty. The best cuts are depicted in figure 3.7, where it can be seen that MC agreement has remained within ~5 % on average yet the uncertainty has slightly improved. For comparison several kinematic variables are plotted in figures 3.7 and 3.6. The uncertainty difference is most obvious in the p_T of the leading jet but all six plots exhibit a small reduction in systematic uncertainty.

The event yields with uncertainties included are shown in table 3.2. A more detailed discussion regarding the systematics included in this analysis can be found in section 5.2.

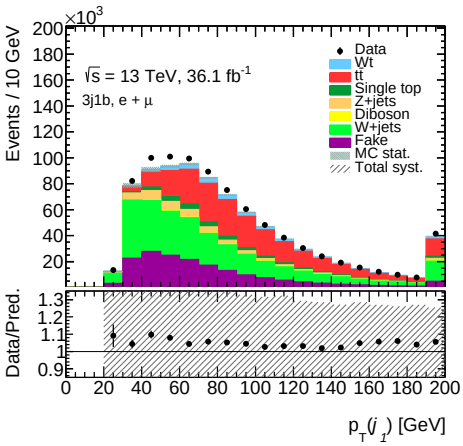
tW only consists about 5 % of the signal region while $t\bar{t}$ and W + jets are almost ten times in size individually. Each being about 38 % and 34 % of the total respectively. The following chapter will introduce the tools required to perform an analysis with the difficulties of low statistics and backgrounds that share characteristics with signal.

Process	Event Yields		
tW	23 300	±	1 000
$t\bar{t}$	179 000	±	14 000
W +jets	161 000	±	14 000
Fakes	60 200	±	15 000
Z +jets	19 100	±	1 100
Diboson	3 180	±	210
s-channel	738	±	49
t-channel	20 600	±	1 100
tZ	154	±	15
Total	468 000	±	25 000
Data	492596		
Data/MC	1.05		
$S/\sqrt{B}$	35.0		

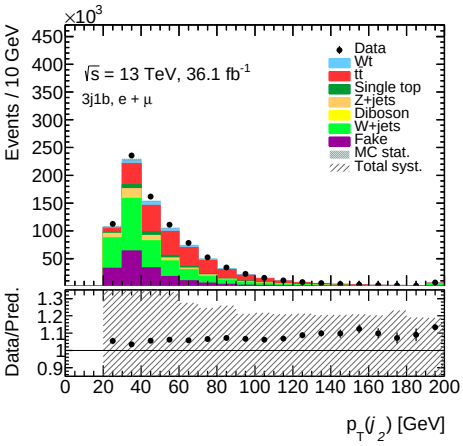
Table 3.2: Event yields with signal region requirements after cut optimization with error estimation from statistical uncertainty, experimental systematics (without JES) and DR/DS systematics.



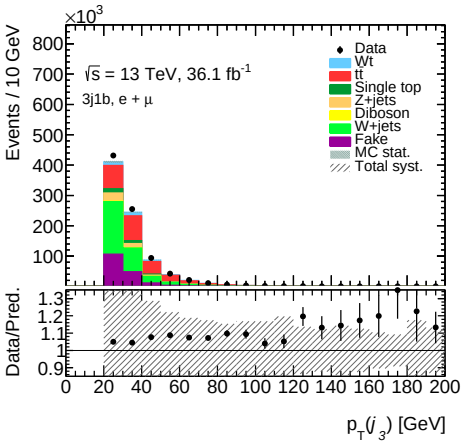
(a) p_{T} of the lepton



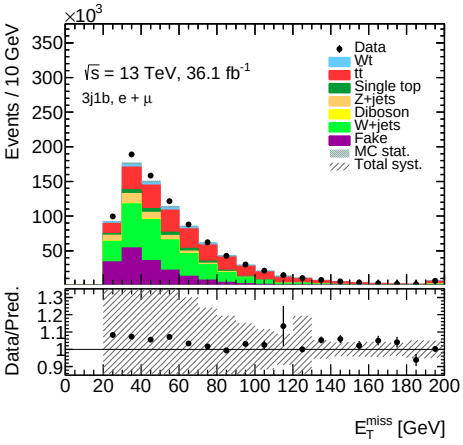
(b) p_{T} of the leading jet



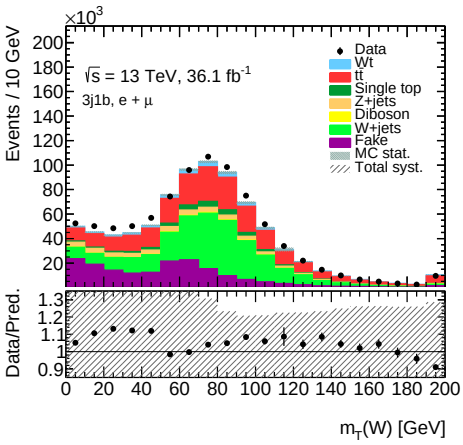
(c) p_{T} of the second leading jet



(d) p_{T} of the softest jet

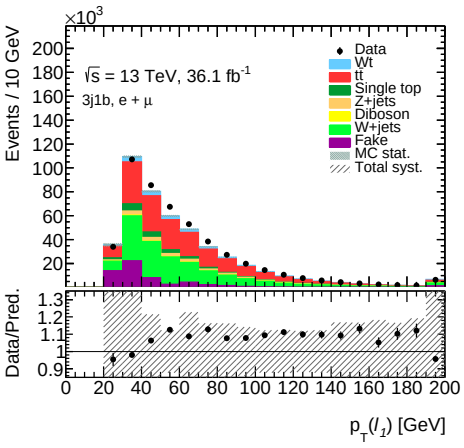


(e) $E_{\mathrm{T}}^{\mathrm{miss}}$

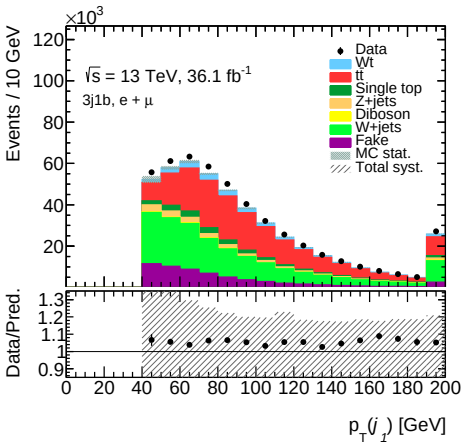


(f) m_{T} of the leptonic W boson

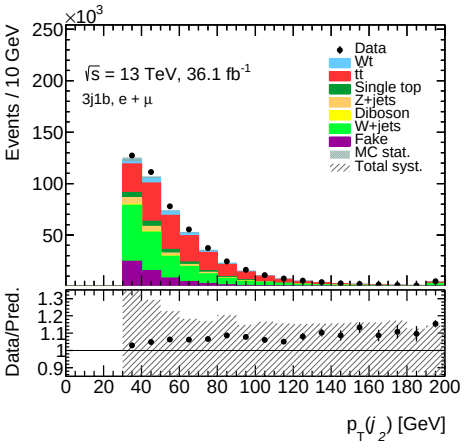
Figure 3.6: Several kinematic variables used for cutting away background. No cut on any kinematic variable has been implemented in these plots. Only uncertainties are experimental with the exception of JES.



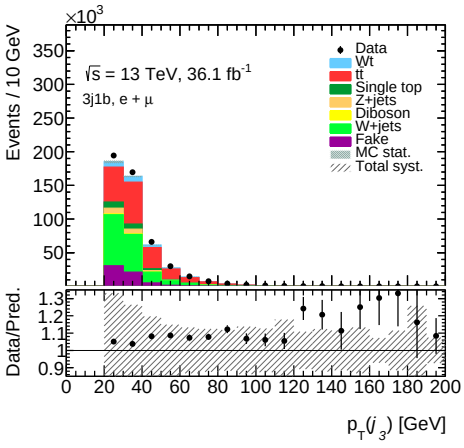
(a) p_{T} of the lepton



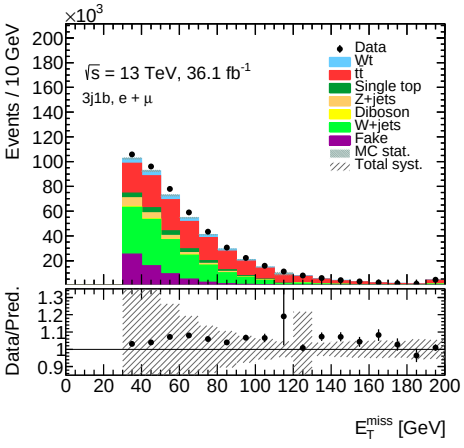
(b) p_{T} of the leading jet



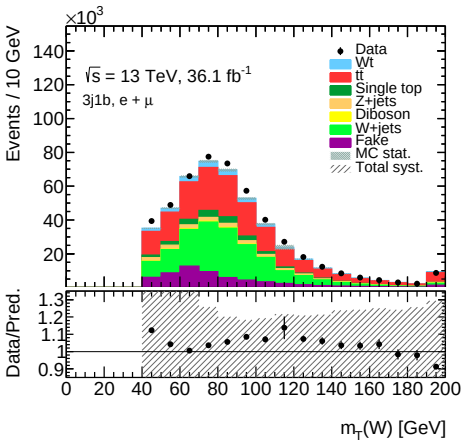
(c) p_{T} of the second leading jet



(d) p_{T} of the softest jet



(e) $E_{\mathrm{T}}^{\mathrm{miss}}$



(f) m_{T} of the leptonic W boson

Figure 3.7: Kinematic variables used for cut optimization but with optimized cuts applied. Only uncertainties are experimental with the exception of JES.

CHAPTER 4

Signal-Background Separation

Given the tW channel has low statistics by comparison to its backgrounds and similar features to processes like $t\bar{t}$, one needs to employ more sophisticated methods to perform a reasonable separation. The following sections detail the tools used for separating signal from background. Section 4.1 introduces Neural Networks (NN). These programs are shown the characteristics of signal and background events so that they learn to differentiate the two. Section 4.2 mentions which specific NN is used in this analysis. In section 4.4, the strategy used in the $\sqrt{s} = 8\,\text{TeV}$ analysis is outlined while section 4.5 details changes from the previous strategy and explain the reason for changing.

4.1 Introduction

A NN is an algorithm that is divided into two parts: a teacher and an expert. The teacher takes events labeled as signal and background as input in order to *teach* itself the differences between them. The expert then uses the training to classify data events as signal or background. There may be distinguishing features between samples and correlated variables that one, a priori, would now know. A NN can notice these features and calculate correlations without any previous knowledge of the processes.

In more detail, a NN is a collection of nodes (called artificial neurons) that are interconnected such that nodes can exchange information by artificial synapses. In this analysis, a *three-layered feed-forward* network is used. This type of NN has its nodes are separated into three layers: the input layer, the hidden layer, and the output layer. These layers are depicted in figure 4.1.

The input layer consists of preprocessed input nodes and a bias node. The input nodes take in the values from input variables. The bias node is used to aid in the success of learning. Specifically, it helps the network train in extreme cases like when all inputs are zero. The input layer is connected to the intermediate, hidden layer. Each connection has a weight associated with it; defined as a multiplicative factor which is applied to all values transmitted over the synapse. The hidden layer then transmits to the output layer by synapses with individual weights. The output layer then assigns a number between -1 to 1 to the event where negative means background-like and positive is signal-like.

4.2 NeuroBayes

The NeuroBayes NN is a sophisticated tool that performs multivariate analysis on data. It is a three-layered feed-forward NN, as is described in the previous section, combined with an automated preprocessing of the input variables [29]. For training, it can take in MC simulations where samples are designated as

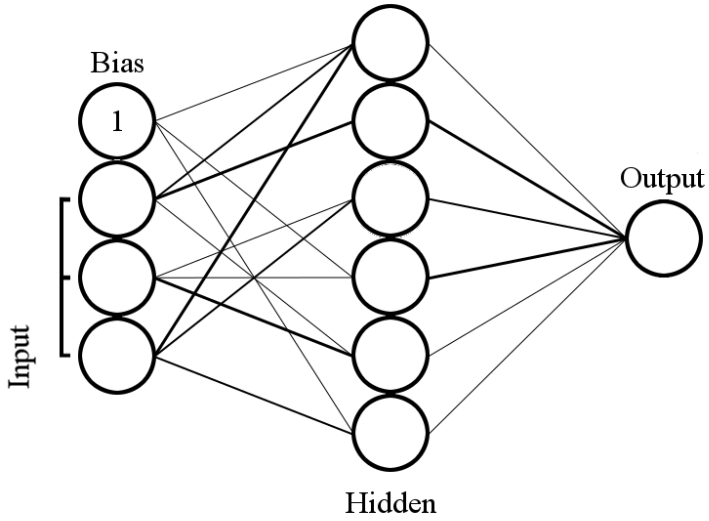


Figure 4.1: Illustration of a NN's node system. Each circle represents a node and each line is a connecting synapse. The input nodes receive the input values per event. The bias node is added to increase the flexibility of the learning and takes no input. Each node is connected by a synapse depicted by a line with the significance of the weight as the thickness. The output node gives the event a value that determines its likeness to signal or background.

backgrounds and signal by the user. It can perform preprocessing steps, which the user can change, and learn complex relationships between variables. Each node and weight significance is evaluated during training to ensure only significant parts of the network remain. NeuroBayes then outputs a series of plots for the user to evaluate the success of training.

Preprocessing

Preprocessing is the first step performed by NeuroBayes. All input values must first be prepared for the network to be used easily. First, all input variables are equalized. i.e the original distribution for all input variables are flattened by a nonlinear transformation. Next, it is transformed into a normal Gaussian distribution. Afterwards, a correlation matrix for each variable to the output is calculated. Once the correlation matrix is calculated, variables are removed one at a time and the correlation to the output is computed again. This is used to rank the variables by loss of correlation. The variable causing the least loss is then labeled the least significant variable and is discarded. This process is repeated until of all variables only one variable remains, i.e. the most significant variable. After ranking, the last preprocessing procedure is called *global preprocessing* which can de-correlate the input variables if set by the user [30].

NeuroBayes calculates four quantities after preprocessing for each variable that describes their importance:

Additional significance: Significance calculated by the iterative method described. Used for ranking and pruning of variables.

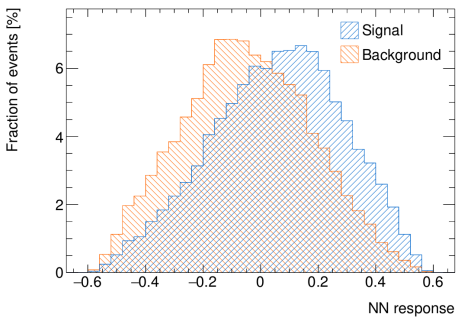
Significance of this variable only: Correlation of a variable to the output multiplied by the square root of the sample size, $\sqrt{n}$. This calculation does not account for correlation to other variables.

Significance loss when the variable is removed: The loss of correlation multiplied by $\sqrt{n}$ when only this variable is removed from the input set, i.e calculated at the first step of the iterative method but choosing this variable as the first variable.

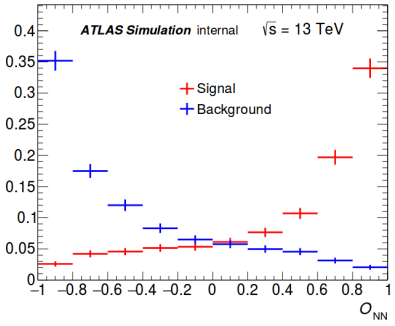
Global correlation to other variables: The correlation of a variable to all the others computed from the complete matrix.

Training Output

The NeuroBayes package outputs a multitude of useful plots in order to evaluate the success of the training. One of such plots show the separation for signal to background. Examples are seen in figure 4.2. These plots show how signal and background events are classified in the range described in section 4.1. Figure 4.2(b) exemplifies a reasonable separation that can be used as a discriminant for the analysis. In contrast, figure 4.2(a) shows very little separation and therefore it is weak as a discriminant alone. Thus, a more elaborate discriminant is constructed. This is further explained in sections 4.4 and 4.5.



(a) Plot with low separation from the tW analysis at $\sqrt{s} = 8$ TeV [20]



(b) Plot with good separation from the tZq analysis at $\sqrt{s} = 13$ TeV [31]

Figure 4.2: Example of the separation plot that the NeuroBayes package can output. Signal and background are labeled in each plot as their colors are dissimilar. The network output ranges from −1 to describe events as background-like and 1 denotes the likeness to signal for an event.

Another plot one can use to gauge separation is the signal efficiency vs. total efficiency plot. This plot is shown in figure 4.3. From this plot one can calculate a value called the *Gini index* (already given by the NeuroBayes package). It is a measure of the network's ability to discriminate signal from background and is calculated by dividing the area under the ROC curve[1] by the area under the no-discrimination line.[2] The upper limit to the Gini index is 50 % since the signal to background ratio is 50:50 for this analysis. The ratio is reached by weighting events.

Another output worth mentioning is the correlation matrix for variables. Each variable is also given a purity vs. efficiency curve.

4.3 First Approach

One can naïvely try to train the network to separate the signal from all backgrounds. One caveat about this analysis that should be noted is that one cannot train against fakes. Fake background is estimated

[1] The Receiver Operating Characteristic curve illustrates the diagnostic ability of a classifier system as its discrimination threshold is varied.

[2] When the ROC curve lies on this line, it shows that the network was unable to discriminate signal from background.

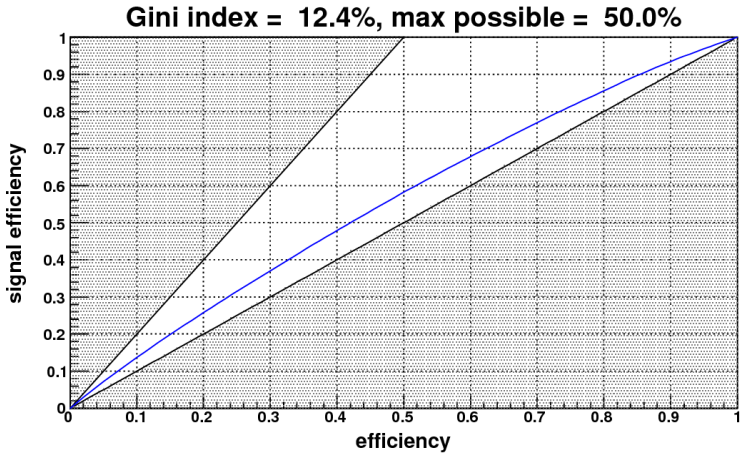
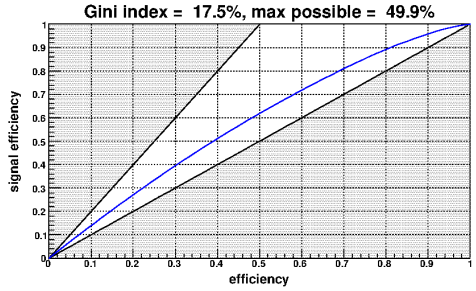


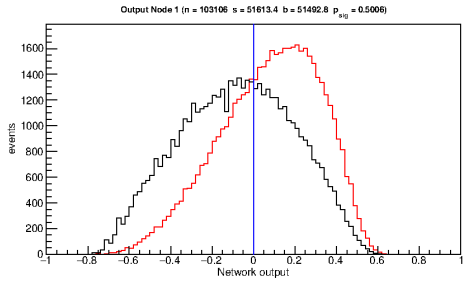
Figure 4.3: Signal efficiency vs. total efficiency plot from the output of the NeuroBayes package. The blue line is the ROC curve, the one-to-one diagonal line is the no-discrimination line, and the diagonal line above the other two denote the maximum shape the ROC curve can have. The Gini index is given above as well as the possible maximum.

from loose data events that may pass the tight selection. In this case, the fake events are not orthogonal to data and therefore one cannot train against them without introducing bias. In order to have an estimated fake lepton contribution that can be used for NN training, one would need to use a solely loose selection of data events. By construction, the Matrix Method requires the information of passing the tight selection to correctly derive the total number of events.

This training is considered naïve because it attempts to separate signal from all backgrounds with minimal consideration of the effects that systematic uncertainties have on variables picked. To elaborate, the training takes as input all kinematic (e.g. p_{T}, η, centrality) variables with at most two objects. The limit on number of objects is chosen to avoid having variables that essentially carry the same information, or are highly correlated, for the purpose of reducing training time.



(a) Signal efficiency vs. total efficiency



(b) Separation plot

Figure 4.4: NeuroBayes output plots that show the separation for training tW vs. all backgrounds except for fakes.

The first approach training gives the output shown in figure 4.4. Separation appears significant for this training. However, upon close inspection one can notice that the training discriminates very well for most backgrounds but fails at separating $t\bar{t}$ from tW (shown in figure 4.5). This is unacceptable as $t\bar{t}$ is the main background that one should separate. The network is not biased towards different processes and

gives significance to variables that separate most background events. Although top-quark pair production is the largest background, it yields less than half of the total events. In turn, the network picks variables that are well defined in tW which are not in most backgrounds. For example, the mass of the hadronically decaying W boson, $m(W_{\mathrm{H}})$, is exceptional at separating tW from all non-$t\bar{t}$ backgrounds as tW has an on-shell W boson that can be reconstructed from two jets. Unfortunately, $t\bar{t}$ shares this quality with tW and can therefore not be separated successfully. As a caveat, the W + jets collection does not contain a hadronically decaying W as it needs to decay leptonically in order to pass the event selection. In addition to being unable to discriminate $t\bar{t}$ from tW, variables like $m(W_{\mathrm{H}})$ are very sensitive to experimental systematic uncertainties. Thus it is an undesirable variable for training.

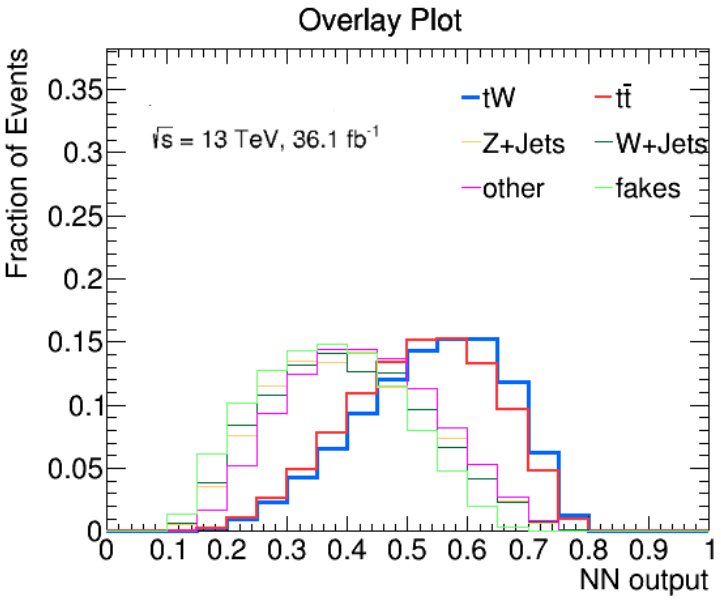


Figure 4.5: Overlay of the NN output after naïve training. Each MC process is normalized for comparison. To make the tW and $t\bar{t}$ stand out against other backgrounds, their lines are thicker.

Given these limitations and need for variables that are less sensitive to systematic uncertainties, one needs to consider a different strategy to ensure a successful training. Fortunately, an analysis has already done this. The following section will outline the strategy used for the tW analysis at $\sqrt{s} = 8$ TeV [20].

4.4 The $\sqrt{s} = 8$ TeV Strategy

As mentioned previously, the information from events with well reconstructed hadronically decaying W bosons is valuable as it helps separate many backgrounds from the signal. However, it is very susceptible to experimental uncertainties as it is reconstructed from two jets and it lacks separation power when it comes to tW and $t\bar{t}$. The goal of maintaining the information needed to separate other backgrounds while avoiding using variables with high sensitivity to uncertainties is achieved by placing constraints on the training. One such constraint is to require well reconstructed hadronically decaying W bosons by cutting around $m(W_{\mathrm{H}})$. Training events must have $m(W_{\mathrm{H}})$ between 65 and 92.5 GeV. This cut has the added advantage that the training is less affected by combinatoric background in $t\bar{t}$ events.

The performance of the $m(W_{\mathrm{H}})$ cut is affected by systematics as well. In order to further lessen the influence of systematics, the number of input variables is reduced to a total of four:

- The transverse momentum balance of the two reconstructed W bosons and the b-tagged jet:

$$\rho(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b) \stackrel{\mathrm{def}}{=} (p_{\mathrm{T}}\Sigma/\Sigma p_{\mathrm{T}})(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b) = \frac{p_{\mathrm{T}}(W_{\mathrm{L}} + W_{\mathrm{H}} + j_b)}{p_{\mathrm{T}}(W_{\mathrm{L}}) + p_{\mathrm{T}}(W_{\mathrm{H}}) + p_{\mathrm{T}}(j_b)},$$

- The invariant mass, $m(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b)$, of the tW system,
- The difference in pseudorapidities of the lepton and leading jet, $\Delta\eta(\ell, j_1)$,
- The pseudorapidity of the lepton, $\eta(\ell)$.

The transverse momentum balance is picked rather than the p_{T} of the system. It is mentioned in the 8 TeV analysis that using the p_{T} of the system led to the W + jets background to appear more signal-like. The invariant mass of the two W bosons and b-jet is shown to have a significant separation power as the tW would appear to be softer than $t\bar{t}$. The difference in pseudorapidity shows no separation power on its own unless it is combined with the invariant mass. Lastly, the pseudorapidity of the lepton is useful as it is less sensitive to systematic uncertainties. Furthermore, the training would only be tW vs. $t\bar{t}$ as the cut on $m(W_{\mathrm{H}})$ would ensure that the training would gain the information to separate the other backgrounds as well.

After training, the distribution (shown in figure 4.2(a)) is re-binned into eight irregularly sized bins in such a way that the number of events is approximately constant in each bin. The binning used is:

$$-1,\ -0.30407,\ -0.17779,\ -0.09170,\ -0.01704,\ 0.05733,\ 0.14100,\ 0.26829,\ 1.$$

After re-binning, a two-dimensional discriminant is constructed from the re-binned output and $m(W_{\mathrm{H}})$. The $m(W_{\mathrm{H}})$ is divided into several bins with varying sizes. Finer binning is done around the training cut while outside this range the binning becomes more coarse. The NN response is only meaningful for events with well reconstructed W bosons and so this NN output binning is only used within the $m(W_{\mathrm{H}})$ cut mentioned previously. A schematic of the binning can be seen in figure 4.6. The two-dimensional discriminant is then converted into a one-dimensional distribution for the sake of simplicity in the fitting stage. The bin numbers can be seen in the same figure which map the two-dimensional distribution to one dimension.

For completeness, the one-dimensional discriminant will have a total of 45 bins. The first, second, and third bins are divided solely in $m(W_{\mathrm{H}})$ such that the NN response is not taken into account. The next bins (4–35) will be divided by $m(W_{\mathrm{H}})$ and NN response as these will carry the information of well reconstructed hadronically decaying W bosons. These 32 bins are only in the range of $m(W_{\mathrm{H}}) \in [65, 92.5]$. The range is divided into four bins and the NN response is divided by eight. The numbering of bins is done from low to high NN response first and then from low to high $m(W_{\mathrm{H}})$. The last bins, similarly to the first three, are only divided in $m(W_{\mathrm{H}})$ and do not carry the NN response information.

4.5 Modification of the $\sqrt{s} = 8$ TeV Strategy

The first modification that this analysis makes to the strategy summarized in section 4.4 is the number of input variables. Initially, the four variables mentioned were used to attempt a separation between $t\bar{t}$ and tW. However, these variables were not enough and the training failed. In figure 4.7, the four variables mentioned earlier are plotted from the $\sqrt{s} = 13$ TeV samples. It can be seen in figures 4.7(a) and 4.7(b) that $t\bar{t}$ and tW (the processes used for training) are very similar. This similarity diminishes the separation performance as the NN is working with less useful variables. The two other variables shown in figures

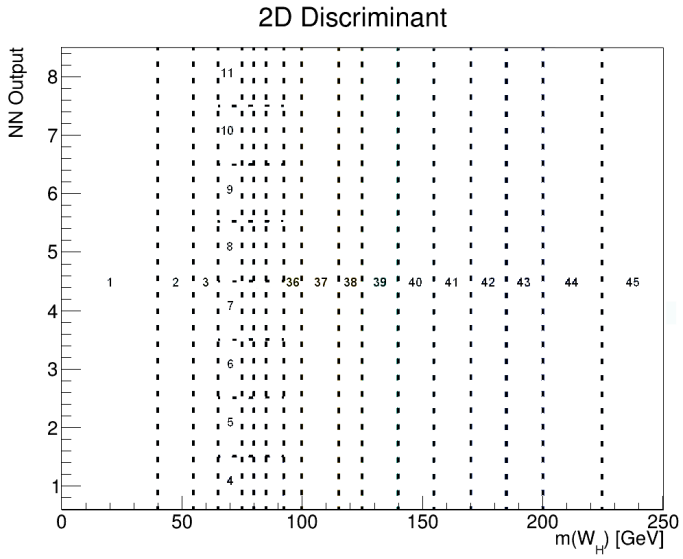


Figure 4.6: An illustration of the two dimensional distribution. Each numbered subdivision corresponds to the binning in one dimension.

4.7(c) and 4.7(d) show a more defined difference between $t\bar{t}$ and tW. With this in mind, this analysis expands the list of variables used for training in order to achieve better separation. However, one purpose of this training is to reduce the effect of systematics and thus avoid variables with dependence on jets. This is proven to be difficult and so the variables chosen for training will depend on at most one jet[3].

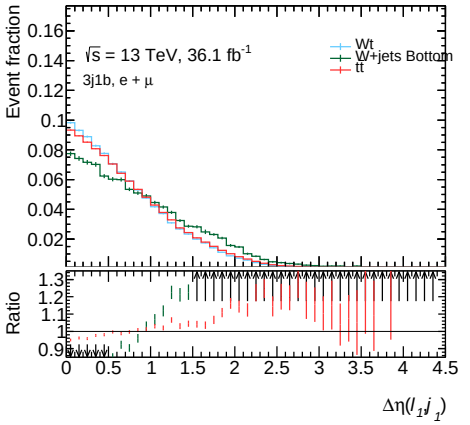
Forty variables, including the previous four, were used as input to be ranked by the NN (shown in table 4.1). Of the forty, only the variables that exceeded $5\,\sigma$ in significance were kept. As shown earlier, of the four variables only two had use for the NN.

The separation and efficiency plots can be seen in figure 4.8. When comparing the separation power of this training to the naïve training (figures 4.8(a) and 4.4(a)) one can see that the separation power has decreased. This is to be expected as this training is only attempting to separate tW and $t\bar{t}$ rather than backgrounds which are more easily separated. However, although the training was relatively successful at separating $t\bar{t}$ and tW, it fails at classifying important background processes as background. This can be seen in figure 4.9. Although $t\bar{t}$ events are generally background-like, the W + jets and fakes have particularly been separated with less success. Both W + jets and fakes are mostly in the middle with fakes appearing more signal-like.

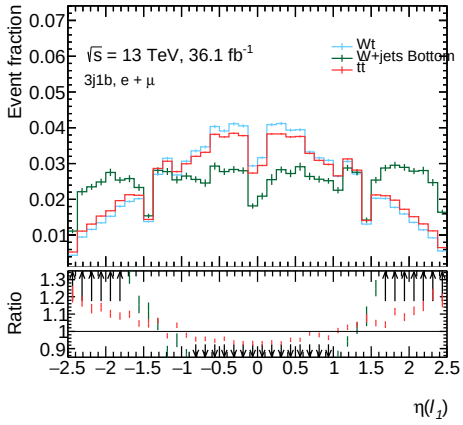
This misclassification can be mediated by adding a background that will not ruin the separation between tW and $t\bar{t}$ yet change the significance of variables to allow the separation between tW and other backgrounds. The best candidate to add is the W + jets collection as it is another major background and is the second largest contribution of events. However, the ratio of $t\bar{t}$ to W + jets will influence how the significance shifts in each variable to favor separation of either background. For this reason, only the collection Wb + jets is used for training. The events from this collection make up slightly less than half of the events contributing to the total W + jets background. This will make the ratio of $t\bar{t}$ to W + jets close to 75:25 in training.

The separation and efficiency plots are shown in figure 4.10. When comparing the Gini index between

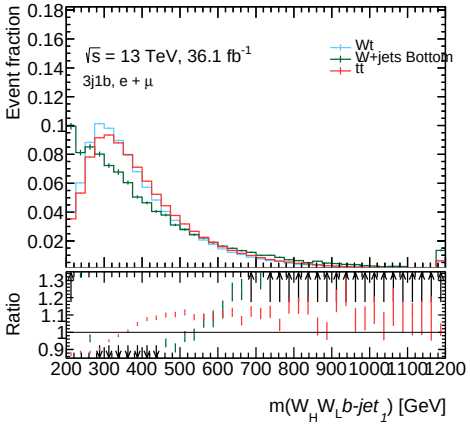
[3] Jets are the objects in this analysis that are most sensitive to experimental systematics (see section 5.2) and thus have the greatest uncertainty. Variables calculated from multiple jets are subject to greater uncertainty as both jets will have their errors propagated.



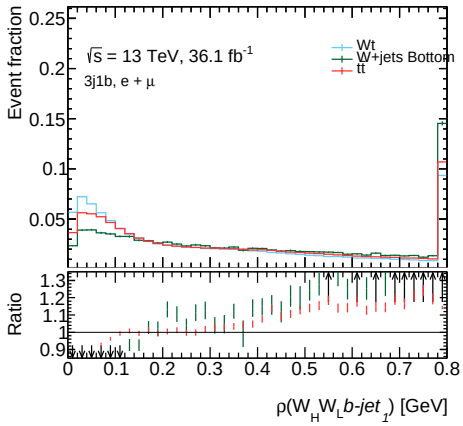
(a) $\Delta\eta(\ell, j_1)$



(b) $\eta(\ell)$



(c) $m(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b)$



(d) $\rho(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b)$

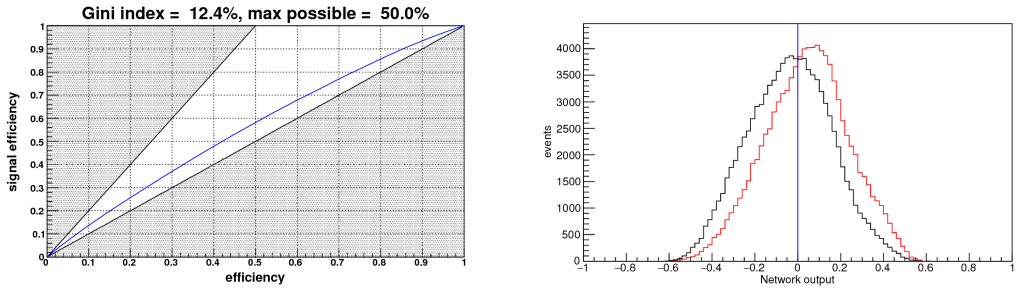
Figure 4.7: Overlay of the four variables used in the training at $\sqrt{s} = 8\,\mathrm{TeV}$. The ratio plot on the bottom shows the ratio of signal to background color coded by the legend and with statistical uncertainty. One W + jets collection is shown for comparison.

this training and the $t\bar{t}$ only training (figures 4.8(a) and 4.10(a)), a slight decrease of 0.4 % can be seen. This is due to the slight loss in separation between tW and $t\bar{t}$ as this background makes up most of the background training sample. However, this slight loss is acceptable as the overall separation is comparable to the previous training. Further comparison of figures 4.8(b) and 4.10(b) leads one to see that the signal peak from the $t\bar{t}$ only training has shifted left along with the background peak.

Figure 4.11 shows a more clear comparison between the two trainings. It can be seen that the signal peak has shifted to a more neutral output but gained a more linear tail towards the signal-like range of the plot. The $t\bar{t}$ only training has a very well defined peak in the signal region but with a sharper drop thereafter. The background histogram for the $t\bar{t}$ only training has a central peak with a long tail in the background-like region of the plot. In contrast, the new training offers a well defined background peak to the left of the zero line but with a sharp drop thereafter. In summary, the separation should be comparable as the features, such as peaks and tails, of signal and background distributions have been exchanged between trainings.

Variable	Significance(σ)
$\rho(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b)$*	45.74
$m_{\mathrm{T}}(j_2, E_{\mathrm{T}}^{\mathrm{miss}})$	33.09
$\Delta p_{\mathrm{T}}(\ell, j_1)$	26.69
$m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	22.6
Centrality(ℓ, j_2)	18.83
$p_{\mathrm{T}}(\ell)$	14.05
$m(W_{\mathrm{L}}, W_{\mathrm{H}}, j_b)$*	13.04
$\Delta\phi(\ell, j_1)$	13.42
$m_{\mathrm{T}}(j_3, E_{\mathrm{T}}^{\mathrm{miss}})$	12.6
$\eta(j_1)$	11.35
$m_{\mathrm{T}}(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$	8.04
$\Delta R(\ell, j_2)$	7.93
$p_{\mathrm{T}}(j_b)$	7.09
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, j_2)$	6.38
$\Delta\eta(\ell, j_3)$	5.7
$\Delta R(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	5.99
$\Delta p_{\mathrm{T}}(\ell, j_2)$	5.08

Table 4.1: Variables used as input to the NN ordered by their importance. Training was tW against $t\bar{t}$. Variables starred (*) are the ones used in the $\sqrt{s} = 8$ TeV analysis.



(a) Signal efficiency vs. total efficiency

(b) Separation plot

Figure 4.8: NeuroBayes output plots that show the separation for training tW vs. $t\bar{t}$.

In more detail, figure 4.12 shows where select processes are classified by the NN. It can be seen that although the signal is slightly less signal-like, $t\bar{t}$ is sufficiently separated with the addition of other processes being correctly separated.

After these modifications, the re-binning is done similarly as previously described in section 4.4. It should be noted that the bin widths proposed in the previous section work well at keeping similar event yields per bin for the $t\bar{t}$ only training. As for the other strategy, $t\bar{t}$ with Wb, binning in this manner caused a lack of events in the first bin (shown in figure 4.13(b)). Lack of events in a bin can cause the fit to fail, therefore the bins with significantly lower yields for all processes are merged with the following bin. That is to say that if bin one lacks events of all processes, then it is merged with bin two in the fit.

The trainings are further compared in figure 4.13. It is clear from comparing figures 4.13(a) and 4.13(b)

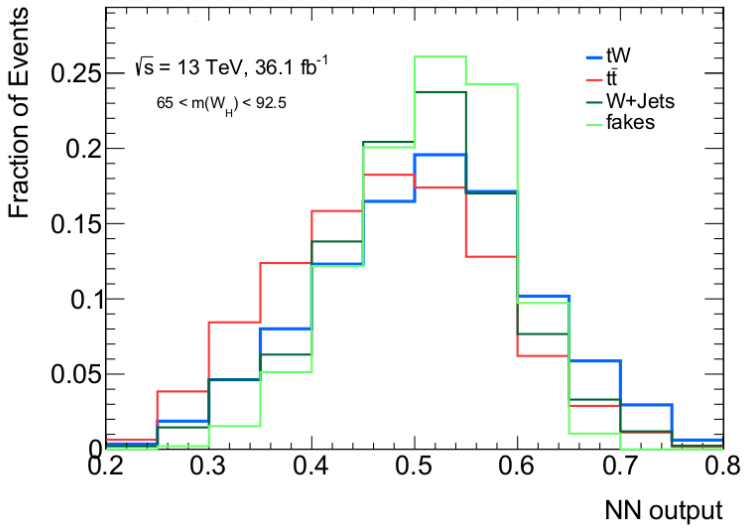
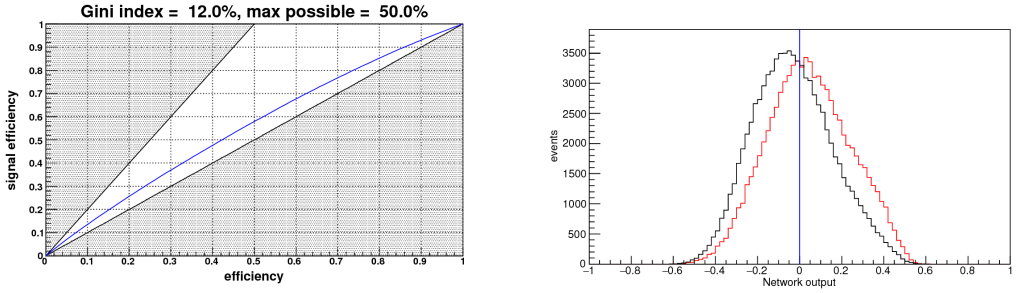


Figure 4.9: Normalized overlay plot of the NN output for different processes using the tW vs. $t\bar{t}$ training. The $m(W_{\mathrm{H}})$ cut is applied as this is the region of most interest. NN output has been shifted from the range of $[-1, 1]$ to $[0, 1]$.



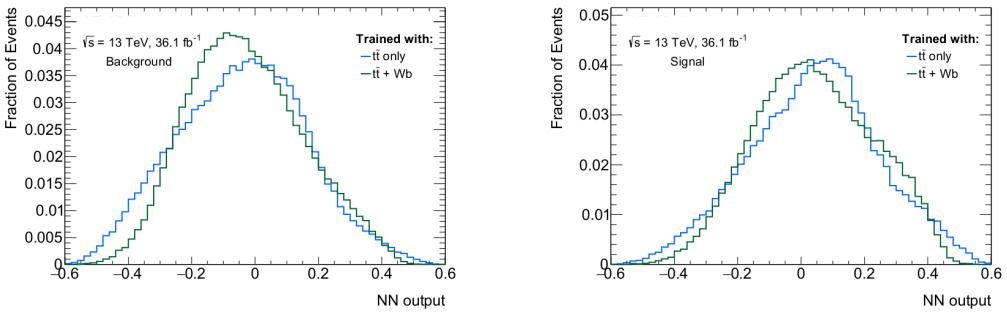
(a) Signal efficiency vs. total efficiency

(b) Separation plot

Figure 4.10: NeuroBayes output plots that show the separation for training tW vs. $t\bar{t}$ and Wb + jets.

that the $t\bar{t}$ only training separates $t\bar{t}$ better but other backgrounds appear signal-like. In particular, the W + jets is classified almost the same as tW. The $t\bar{t}$ and Wb training has much better overall separation of backgrounds while still separating tW and $t\bar{t}$ sufficiently well.

Once the NN response has been re-binned, the mapping of the two-dimensional discriminant onto one dimension is constructed. Figure 4.14 shows the one-dimensional distribution to be fit. All 45 bins have been described previously in section 4.4. Since the data obscures the signal contribution, a dashed, blue line shows the normalized to total background tW to make the shape clear and demonstrate the difference between it and the overall background distribution.



(a) Background (b) Signal

Figure 4.11: Normalized overlay of the NN output for both trainings.

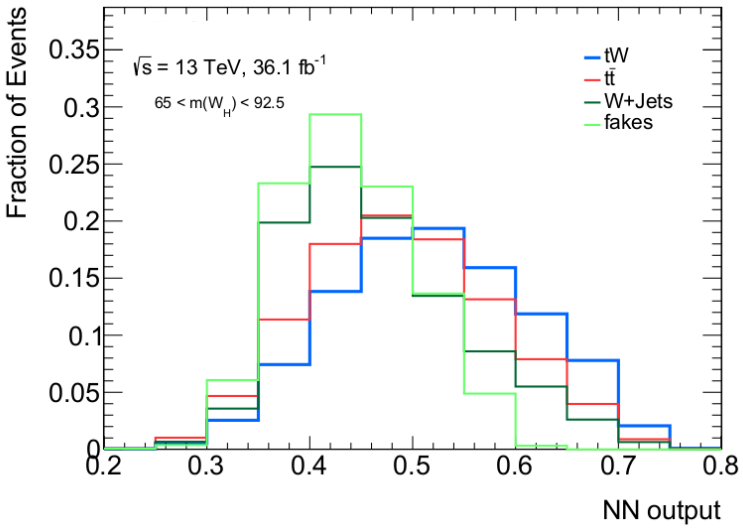
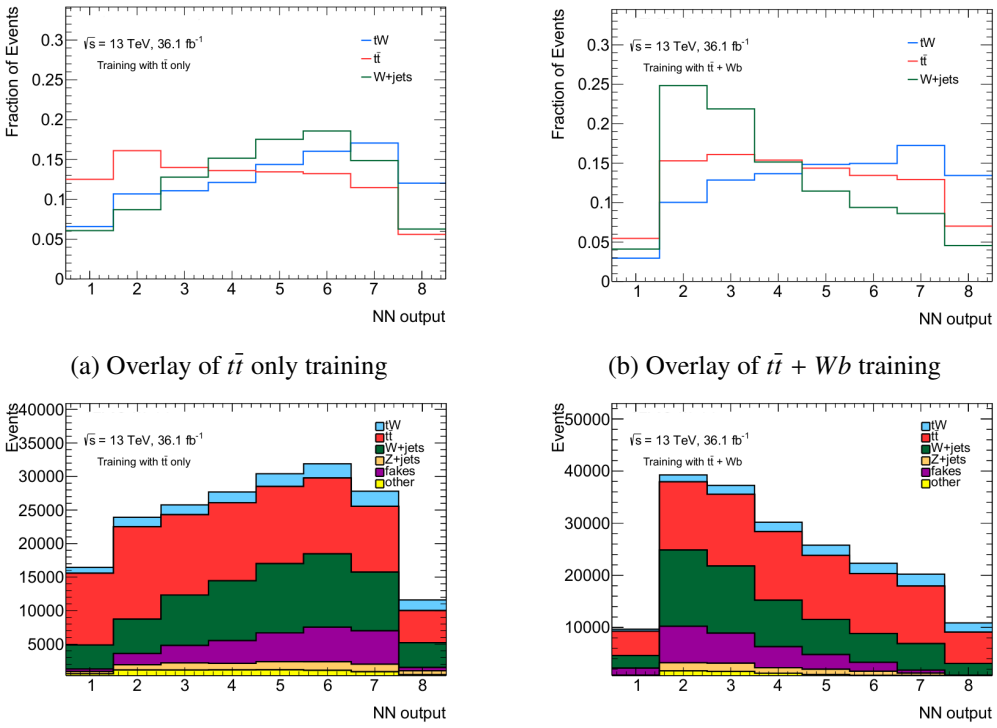


Figure 4.12: Normalized overlay plot of the NN output for different processes using the tW vs. $t\bar{t}$ and Wb + jets training. The $m(W_{\text{H}})$ cut is applied as this is the region of most interest. NN output has been shifted from the range of $[-1, 1]$ to $[0, 1]$.



(a) Overlay of $t\bar{t}$ only training

(b) Overlay of $t\bar{t}$ + Wb training

(c) Stack plot of $t\bar{t}$ only training

(d) Stack plot of $t\bar{t}$ + Wb training

Figure 4.13: Normalized overlay and stack plots comparing both trainings. The left two plots are $t\bar{t}$ only training and the right plots were generated from $t\bar{t}$ + Wb training samples. All plots contain only events where $m(W_{\mathrm{H}}) \in [65, 92.5]$.

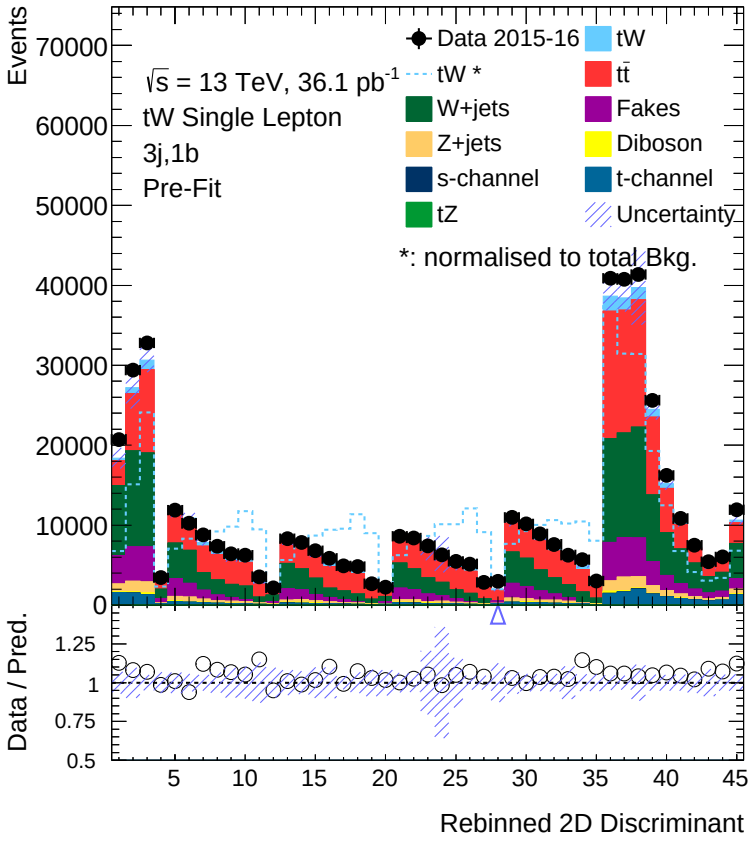


Figure 4.14: Mapping of the two-dimensional discriminant to one dimension to be used for fitting. Above is a stacked plot that shows the total contribution per process denoted in the legend. A normalized signal can be seen to compare against the background in the dashed, blue line. The bottom plot is a ratio plot which compares data to MC yields and the shaded area denotes the collective uncertainty per bin.

CHAPTER 5

Statistical Analysis

From the discriminant generated in the previous chapter, the number of tW events in the data can be determined. By knowing the number of events in the data that belong to tW, the cross-section can be extracted. In order to extract the tW cross-section, a binned profile likelihood fit is used. This fitting method can take systematic uncertainties as nuisance parameters to adapt to their effects on the discriminant. In addition, the fit can constrain the overestimated uncertainties while updating signal and backgrounds to the data.

The following section will describe the theory behind a likelihood fit (section 5.1). Afterwards, uncertainties will be defined and their influence on reconstructed objects will be described (section 5.2). And finally, the results of the fit are shown in section 5.3.

5.1 Likelihood Function

The likelihood function is a tool used to maximize the probability that the model describes the data. It takes the parametrized model with the best fit one provides and gives the representation of how likely different parameters for the distribution are. In this analysis, the cross-section is measured using the principle of *maximum likelihood*. This is the procedure of finding the parameters of a model in order to maximize a known likelihood function.

One can define the total events in the signal region, n, as a sum over signal and background processes:

$$n(\mu) = \mathcal{L}\epsilon_0\sigma_0\mu + \mathcal{L}\sum_j^{\text{bkg.}} \epsilon_j\sigma_j, \tag{5.1}$$

where $\mathcal{L}$ denotes the integrated luminosity, ϵ_j is the efficiency of reconstruction and selection of events of background process j, and σ_j is the cross-section for each background process. ϵ_0 and σ_0 are the efficiency and cross-section for the tW process, respectively. The theoretical cross-section is used multiplied with a scaling parameter, μ, in order to extract the tW cross-section.

The probability of observing N events (without accounting for uncertainties) in an interval follows a Poisson distribution:

$$P(N) = e^{-r}\frac{r^N}{N!}, \tag{5.2}$$

where r is the mean number of events per interval. With the total number of events previously defined, one can then write the likelihood function for one bin as:

$$P(N,n) = e^{-n(\mu)}\frac{n(\mu)^N}{N!}, \tag{5.3}$$

The one-dimensional mapping of the two-dimensional discriminant introduced in section 4.4 is composed with 45 bins that are statistically independent measurements. Thus the probabilities can be multiplied, defining the likelihood function for all bins:

$$L(\boldsymbol{N},\boldsymbol{n}) = \prod_i^{\text{bins}} P(N_i,n_i), \tag{5.4}$$

$$= \prod_i^{\text{bins}} e^{-n_i(\mu)}\frac{n_i(\mu)^{N_i}}{N_i!}, \tag{5.5}$$

where $\boldsymbol{N}$ and $\boldsymbol{n}$ are now vectors and the i subscript denotes the bin number.

Commonly, the natural logarithm of the likelihood function, known as *log-likelihood*, is used instead for convenience. This logarithmic function is strictly increasing, it achieves its maximum value at same points as the function itself, and in this case turns multiplication over bins into addition. The equation is as follows:

$$\Lambda(\boldsymbol{N},\boldsymbol{n}) = -2\ln(L(\boldsymbol{N},\boldsymbol{n})), \tag{5.6}$$

where the factor of -2 is introduced purely for convention.

The most likely value of the parameter of interest, $\hat{\mu}$, is the closest to μ and maximizes the likelihood function, L, or minimizes the log-likelihood function, Λ. The error for the nuisance parameter is calculated by shifting $\hat{\mu}$ until the log-likelihood function increases by one unit.

With one source of uncertainty included, the nuisance parameter, θ, is added to account for the influence that this systematic uncertainty has on the total number of events. θ has an influence on the total number of events, denoted by δ, that is estimated by varying the nuisance parameter by one standard deviation ($\theta = \pm 1$). This is shown in the following equations:

$$n_i(\mu,\theta) = n_i(\mu)(1+\delta\cdot\theta), \tag{5.7}$$

$$L(\boldsymbol{N},\boldsymbol{n}(\mu,\theta)) = \prod_i^{\text{bins}}[P(N_i n_i)]f_{\mathcal{N}}(\theta), \tag{5.8}$$

$$= L_{\text{nom.}}(\boldsymbol{N},\boldsymbol{n}(\mu))f_{\mathcal{N}}(\theta), \tag{5.9}$$

where the nom. subscript denotes a θ-less likelihood function, and $f_{\mathcal{N}}(\theta)$ is a normal Gaussian distribution describing the probability density distribution of θ. The calculation of δ is shown in section 5.2. More systematic uncertainties have an additive effect to the total number of events which translates to a multiplicative effect on the likelihood function. This can then be written as follows:

$$n_i(\mu, \theta) = n_i(\mu)(1 + \sum_k^{\text{unc.}} \delta_k \cdot \theta_k), \tag{5.10}$$

$$L(\boldsymbol{N}, \boldsymbol{n}(\mu, \theta)) = L_{\text{nom.}}(\boldsymbol{N}, \boldsymbol{n}(\mu)) \prod_k^{\text{unc.}} f_{\mathcal{N}}(\theta_k), \tag{5.11}$$

$$\Lambda(\boldsymbol{N}, \boldsymbol{n}) = \Lambda_{\text{nom.}}(\boldsymbol{N}, \boldsymbol{n}) + \sum_k^{\text{unc.}} \ln(f_{\mathcal{N}}(\theta_k)), \tag{5.12}$$

where k denotes the systematic associated with the nuisance parameter, θ, and error, δ. When there are multiple nuisance parameters, the log-likelihood function is minimized not just in μ but also in θ. Generally, the most likely nuisance parameter associated with an uncertainty, $\hat{\theta}$, can minimize the log-likelihood function for $\mu \neq \hat{\mu}$ thus broadening the parabolic shape. This broadened log-likelihood increases the uncertainty associated with $\hat{\mu}$.

Furthermore, the error intervals and contours can be approximated using a covariance matrix of the parameter estimates. The matrix is defined as follows:

$$\hat{V}_{ab}^{-1} = -\frac{\partial^2 L}{\partial\hat{\theta}_a \partial\hat{\theta}_b}, \tag{5.13}$$

where the subscripts a and b denote the uncertainty associated to the estimator $\hat{\theta}$. The μ parameter can be easily included by defining $\hat{\theta}_0 \stackrel{\text{def}}{=} \hat{\mu}$.

The *impact* of a systematic uncertainty on the measurement of the cross-section is to be calculated. This is defined as the shift of the maximum-likelihood estimate for μ when the nuisance parameter θ_k is shifted by $\pm\Delta\theta_k$. Similarly to the estimation of error intervals, this can be approximated using a covariance matrix:

$$\frac{\text{cov}[\hat{\mu}, \hat{\theta}_k]}{\sqrt{\text{cov}[\hat{\theta}_k, \hat{\theta}_k]}}. \tag{5.14}$$

Nuisance parameters are estimated from the MC simulation but are actually measured from, or constrained by, the data. This measurement is done by constraining processes as they have different event yields per bin. In more detail, bins that contain different signal to background ratios can be used to constrain the uncertainties in these processes. Furthermore, bins with no signal events can limit the uncertainties of background processes to the statistical uncertainty. Therefore, sufficient amount of data can reduce uncertainties to great effect.

The likelihood fit is performed by the TRExFitter program. It is capable of performing a *profile likelihood* fit, or a likelihood fit for models with more than one unknown parameter. The program builds a global likelihood function for all the bins and includes all parameters. One parameter in particular, denoted as the parameter of interest (POI), can be measured by performing a log-likelihood minimization on the global likelihood function.

5.2 Uncertainty

Throughout this thesis, there have been mentions of detector and modeling limitations. These, among other sources of uncertainty, contribute negatively to the performance of any analysis. In the following

sections, the uncertainties that affect this analysis are discussed.

Overview

Generally, systematic uncertainty error are calculated by the modification of parameters in the simulation or in the reconstruction of objects. As an example: given N total of events, a parameter associated with a given systematic, k, is varied such that N_k^+ and N_k^- are recorded where the plus and minus subscript denotes a deviation with higher and lower total events or events in a given bin, respectively. This can then be used to calculate the error, δ_k, as:

$$\delta_k = \frac{N_k^+ - N_k^-}{2N}. \tag{5.15}$$

For systematic uncertainties that are only one-sided, such as the Diagram Subtraction scheme, the error associated is taken as the relative deviation from the nominal rate. This example is illustrated in figure 5.1. In this plot, the difference between schemes in a bin by bin basis can be seen. These type of systematics deviations are taken into account in the fitting process which can be smoothed or taken as they are.

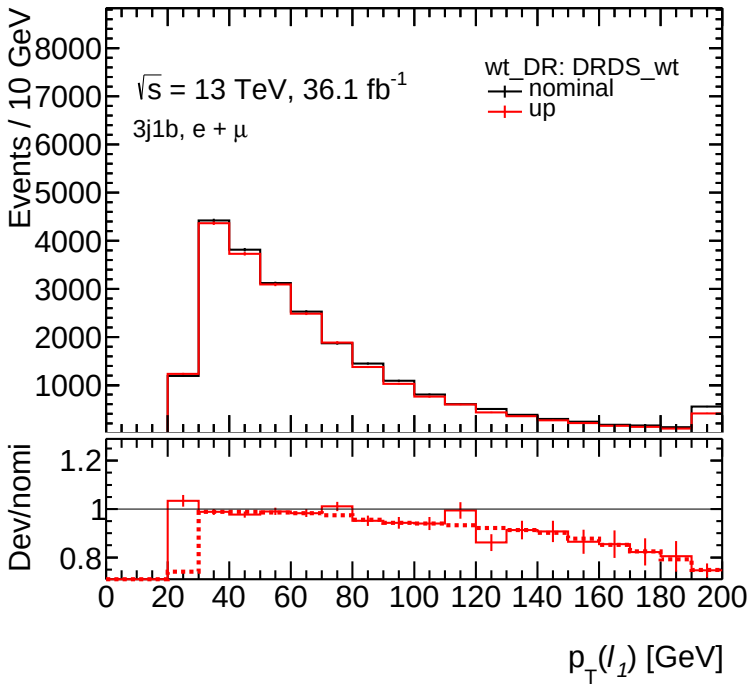


Figure 5.1: Overlay of Diagram Removal (DR) (nominal) tW and Diagram Subtraction (DS) tW samples. The top plot shows an overlay of the DR and DS distributions with respect to the p_{T} of the lepton (including e and μ). The bottom plot shows the deviation of DS from the nominal (DR) sample. The solid red line shows the the DS on the top plot and the ratio of DS to DR in the bottom plot. The dashed, red line is a smoothed version of the ratio between systematic and nominal.

The shape of systematics in the discriminant is the dominant effect for the performance of the fit. This is due to small systematic uncertainties that have similar shapes to the nominal distribution or two-sided

uncertainties that cross[1] as they can make the fit unstable or unreliable. For this reason, only important systematics should be included in the fitting process with keeping in mind that in some bins, these uncertainties will be dominated by statistics. The table containing the systematics used in this analysis can be found in appendix B as well their influence on select processes.

Details

Experimental uncertainties are related to the reconstruction, identification, and measurement of physics objects by the detector. In more detail, the uncertainty in detector resolution and calibration, luminosity, flavor composition and response, lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ scale, and pile-up effects adversely affect the performance of the measurement.

The Jet Energy Scale (JES) is the estimation of average detector bias for jet energy measured with relation to the true energy. This estimation has uncertainties related to the method used for its evaluation which affect the overall uncertainty of analyses that rely on jet reconstruction. The Jet Energy Resolution (JER) describes the resolution of the detector component, or the width of the detector response. Different regions will respond differently and this calibration will have uncertainties attached to it. Furthermore, jet flavor and pile-up also affect the scale. These are modeled and accounted for by the MC simulation. The uncertainty associated with JES and JER are derived using Run 2 data at $\sqrt{s} = 13\,\mathrm{TeV}$ with use of MC simulation and are further detailed in [32]. In this analysis, the JER is not taken into account as its effects on my results were unreasonably high. The reason for it could not be found within the time constraints.

Since neutrinos cannot be detected and measured, they are instead inferred from lepton and jets measurements. Low p_{T} pile-up jets and underlying activity can affect the estimation of missing transverse energy and thus an uncertainty is attached to this. Therefore, the uncertainty associated with lepton, jets, and pile-up effects are propagated into the uncertainty with respect to the $E_{\mathrm{T}}^{\mathrm{miss}}$.

The possibility for jets to be mis-tagged introduces a degree of uncertainty in the cross-section measurement. b-tagging efficiency is corrected by a scale factor that has uncertainty associated with it. This systematic is estimated by varying the efficiency correction in the MC simulation. As mentioned earlier, JES and JER are sensitive to flavor effects so there is an additional uncertainty associated to the JES related to b-jets.

Lepton momentum resolution, scale, trigger efficiency, and identification also have an uncertainty attached to them. In this thesis, the scale and resolution are taken into account for electrons while the muon has identification uncertainty estimated in addition to scale and resolution. Uncertainties related to lepton reconstruction and identification are estimated from dilepton Z boson decays in data. Typically, these uncertainties are small.

There are uncertainties that affect overall samples such as luminosity for all processes and cross-section measurement uncertainty for each individual process. Luminosity of the beam has an uncertainty of 2.2 % [33] attached to it that affects all processes. Cross-section measurement uncertainties are typically within (5–6)% for this analysis. Single top-quark processes [34] as well as Z + jets [35] cross-sections have a measured uncertainties of 5 % while diboson [36] has an uncertainty of 6 %.

In general, MC generators introduce a degree of uncertainty as they approximate reality rather than perfectly represent it. In this case, MC generators are compared for the same sample in order to estimate uncertainty related to using them. For this comparison, typically two samples of the same process are generated using different combinations of event and parton shower generators. The samples that pertained to estimating the systematic uncertainty introduced by MC generators were not available at the time of

[1] Crossing systematic uncertainty distributions refers to having bins where the up (responsible for the calculation of N_k^+) distribution is below the down or nominal distribution. Likewise, the down distribution being above either of the two is also referred to as crossing.

this analysis. For this reason, these uncertainties are unaccounted for but are expected to have a major impact as they were the dominant source of uncertainty in previous analyses.

Lastly, the estimation of fake lepton background is not perfect and therefore must have an uncertainty associated to it. An accurate method for estimating this uncertainty is a topic discussed often without a simple solution. For simplicity, the uncertainty is set to 50 % as an overestimate to be constrained by the fit.

5.3 Fit Results

For all fits, the μ scale parameters were not just set for the tW cross-section but instead this type of parameter was added to the main background processes. This is referred to as *floating* a process. The fit will allow the cross-section of $t\bar{t}$ and W + jets to change within the systematics and statistical uncertainties.

As an exercise, an *Asimov* fit was made in order to estimate the influence of uncertainties on the cross-section measurement and to compare both trainings. An Asimov fit is when the MC is fit to itself rather than data. By fitting to the MC yields as if they were data, the expected values for nuisance parameters are generated in a "perfect" fit. This is used to suppress the influence of statistical fluctuations. For easy comparison, table 5.1 shows the results of the Asimov fit from the $t\bar{t}$ only and the $t\bar{t}$ + Wb training strategies.

	$t\bar{t}$ Only	$t\bar{t}$ + Wb
μ_{tW}	1.00 ± 0.11	1.00 ± 0.18
$\mu_{t\bar{t}}$	1.00 ± 0.03	1.00 ± 0.04
$\mu_{W+\text{jets}}$	1.00 ± 0.07	1.00 ± 0.06

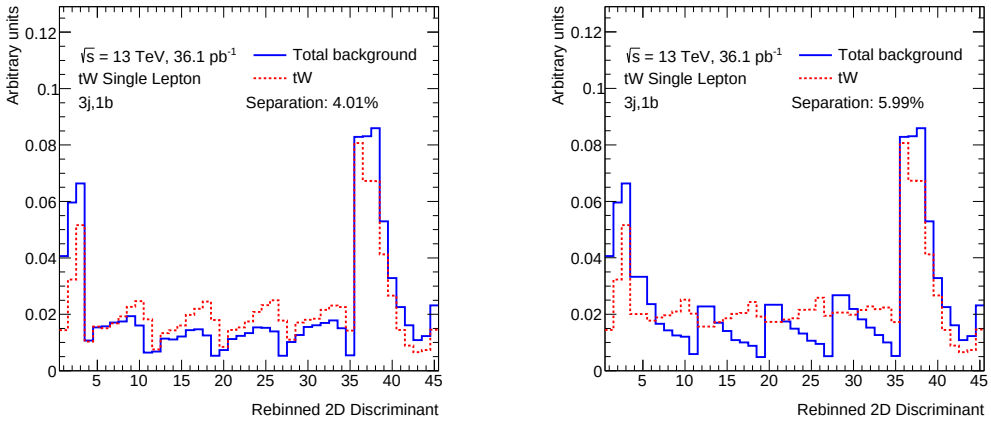
Table 5.1: Results of the Asimov fit comparing both training strategies. The cross-sections of the $t\bar{t}$ and W + jets were also given scaling parameters.

By comparing both columns, it can be seen that the trainings have different effects on the uncertainties. Overall, the compounded uncertainty has a tendency to increase when one uses the $t\bar{t}$ and Wb training. A 7 % difference in uncertainty for the signal scale factor is unwanted but the overall error is within 20 %. For comparison, the dilepton 13 TeV analysis and single lepton at 8 TeV reported errors around 30 %. Considering that modeling systematic uncertainties are likely to dominate this analysis, it is possible that the second training strategy may have similar uncertainty to other mentioned analyses. The uncertainty is further illustrated and compared in figure 5.3 while the overall separation can be seen in figure 5.2. It is clear that both trainings offer different advantages and as such, this document compares the strategies further by performing a fit to data on both discriminants.

The results from fitting against data are shown in two columns as it was done previously:

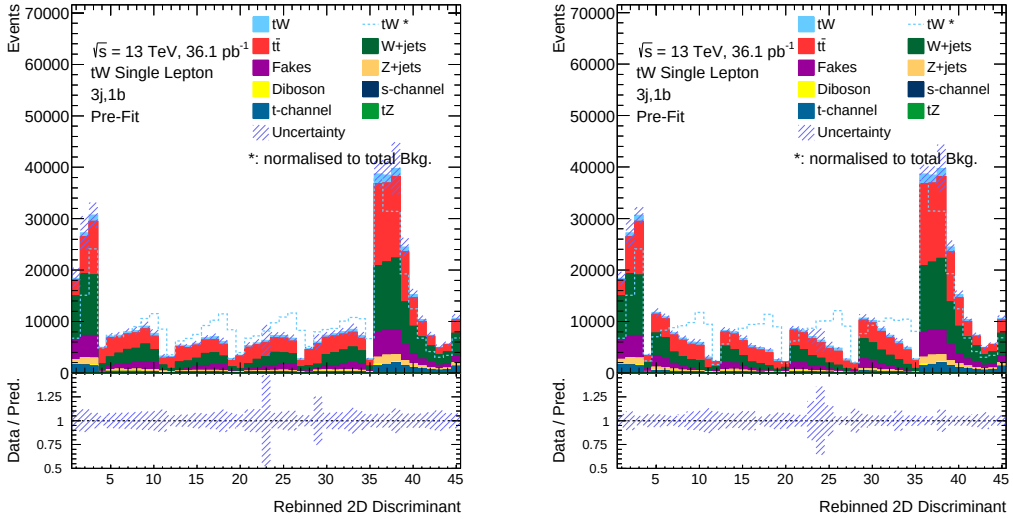
For reference, the predicted cross-section for the tW production channel is $\sigma_{tW}^{\text{Theo.}}(13\,\text{TeV}) = 71.7 \pm 3.8\,\text{pb}$. It is clear that both results are consistent with each other and errors are within 10 % of the theoretical prediction. As was mentioned earlier, not all uncertainties are taken into account and therefore the errors estimated in this measurement are underestimated.

Figures 5.4(a) and 5.4(b) compare one-dimensional discriminant before and after the fit, respectively, for the $t\bar{t}$ only training. Figures 5.5(a) and 5.5(b) show the same plots for the $t\bar{t}$ + Wb strategy. Figure compares the nuisance parameters after fitting for both trainings and figure 5.6 compares the ranking



(a) $t\bar{t}$ only training. (b) $t\bar{t}$ and Wb training.

Figure 5.2: Separation plot of the one-dimensional mapping generated from both training strategies with $t\bar{t}$ only on the left and $t\bar{t}$ with Wb on the right. The signal (red) and all backgrounds (black) are normalized and overlaid such that their shape can be compared.

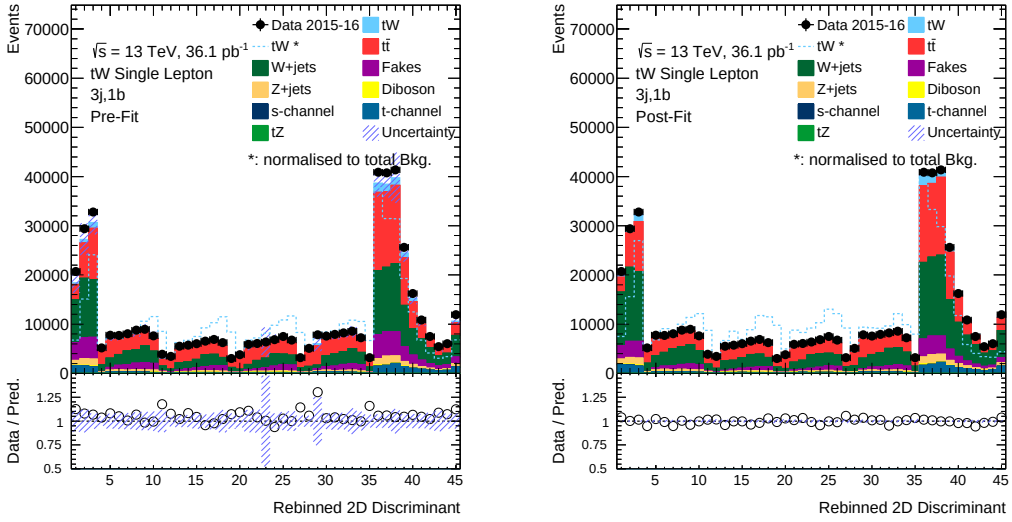


(a) $t\bar{t}$ only training. (b) $t\bar{t}$ and Wb training.

Figure 5.3: Asimov stack plot of the one-dimensional mapping generated from both training strategies with $t\bar{t}$ only on the left and $t\bar{t}$ with Wb on the right. The shaded area illustrates the combined uncertainty of systematics and statistical sources. The plot shows no data as it is an Asimov fit. The top plot is a stack plot with all processes labeled by the legend. The dashed, blue line shows the tW distribution normalized to the total background for comparison in the shapes. The bottom plot shows the overall effect of uncertainty per bin.

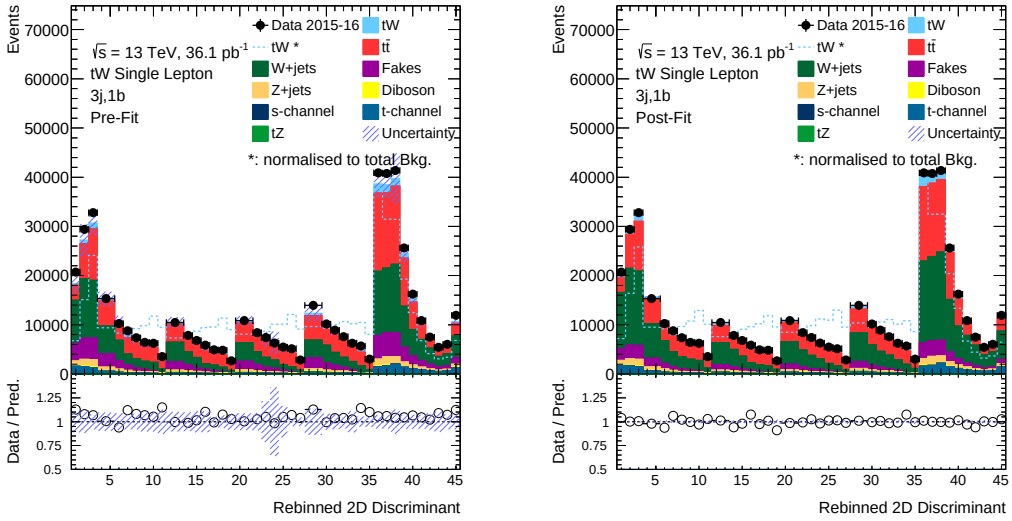
plots. Correlation matrices for both trainings are illustrated in figures 5.7 and 5.8. Supplementary fit plots, such as the comparison of nuisance parameters post-fit (figure C.1), can be found in section C.

It is clear that training on only $t\bar{t}$ leads to a lower uncertainty for the signal cross-section measurement. From comparing the correlation matrices, it can be seen that the scale parameter for the signal cross-section is not very correlated with any other parameter. This is not true for the second training as the



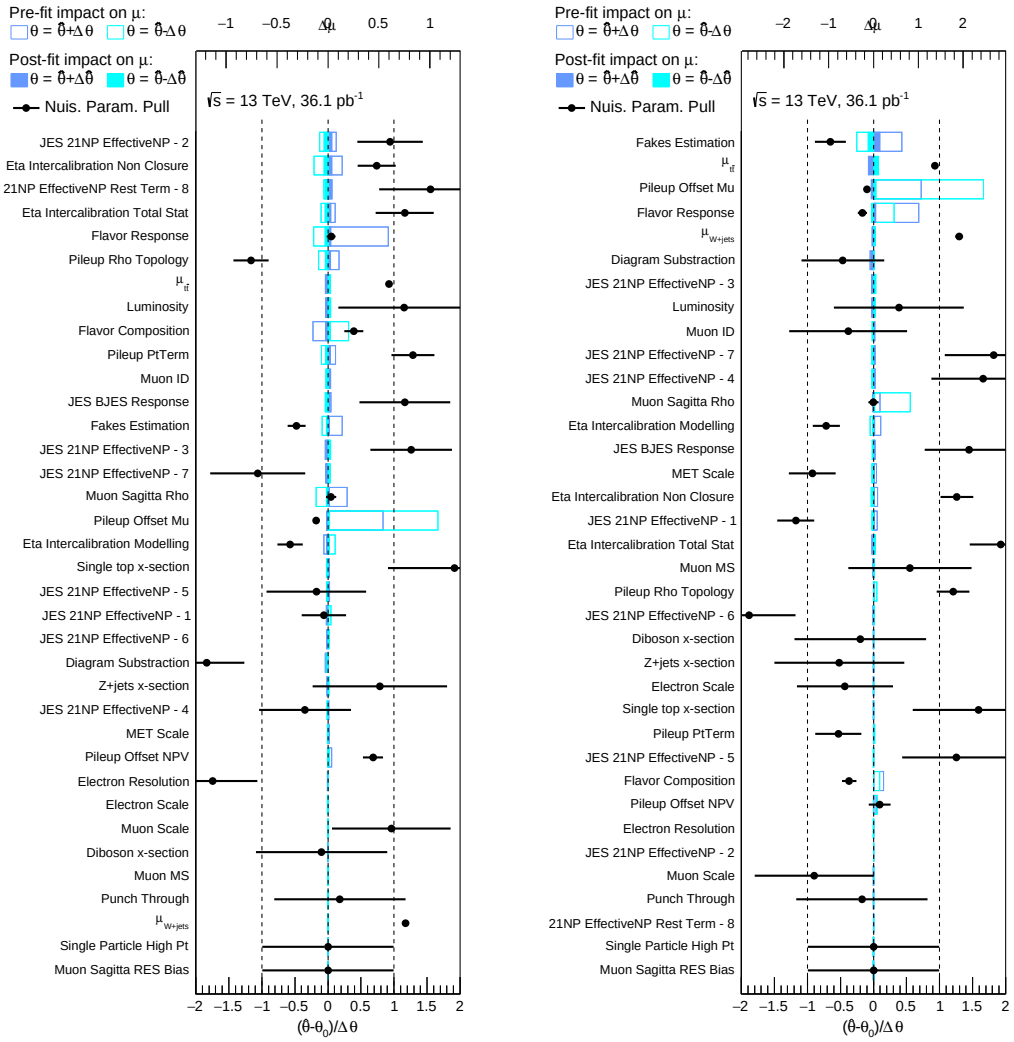
(a) Pre-fit. (b) Post-fit.

Figure 5.4: Comparison of the one-dimensional discriminant before and after the fit for the $t\bar{t}$ only training.



(a) Pre-fit. (b) Post-fit.

Figure 5.5: Comparison of the one-dimensional discriminant before and after the fit for the $t\bar{t}$ + Wb training.



(a) $t\bar{t}$ only. (b) $t\bar{t} + Wb$.

Figure 5.6: Comparison of the ranking plots between both trainings after fitting.

	$t\bar{t}$ Only	$t\bar{t}$ + Wb
μ_{tW}	1.17 ± 0.11	1.28 ± 0.18
$\mu_{t\bar{t}}$	0.92 ± 0.03	0.93 ± 0.04
$\mu_{W+\text{jets}}$	1.17 ± 0.07	1.29 ± 0.06
σ_{tW}(13 TeV)	84 ± 8	92 ± 13

Table 5.2: Results of the fit comparing both training strategies. The cross-sections of the $t\bar{t}$ and W + jets were also given scaling parameters. This table includes the translation of scale factor to measured cross-section.

signal scale parameter is heavily correlated to the fake estimation uncertainty and anti-correlated to the $t\bar{t}$ cross-section scale parameter. This is also shown in the ranking plots as the largest sources of uncertainty are these two parameters for this training strategy. The ranking plots also show almost all uncertainty sources are being pulled rather than being minimized near zero as it would be expected. This can be attributed to the crossing mentioned in section 5.2. Because the fit converges and improves the MC to data agreement (see figures 5.4(b) and 5.5(b)), it can be concluded that the fit works and it deals with the shape uncertainties by pulling the estimated parameters.

Table 5.3 shows the event yields after fitting for both training strategies with errors. It can be seen that the yields differ slightly but are not in disagreement.

	$t\bar{t}$ Only			$t\bar{t}$ + Wb		
tW	26 900	±	2 600	30 100	±	4 100
$t\bar{t}$	174 000	±	3 100	171 000	±	6 200
W+jets	198 000	±	6 200	208 000	±	8 200
Fakes	45 500	±	4 500	39 100	±	6 800
Z +jets	20 300	±	1 100	18 700	±	1 000
Diboson	3 240	±	210	3 170	±	210
s-channel	830	±	52	804	±	51
t-channel	23 200	±	1 200	22 400	±	1 200
tZ	174	±	16	168	±	16
Total	493 000	±	1 000	493 000	±	1 200
Data	492596					

Table 5.3: Yields of the analysis after the fit for both training strategies.

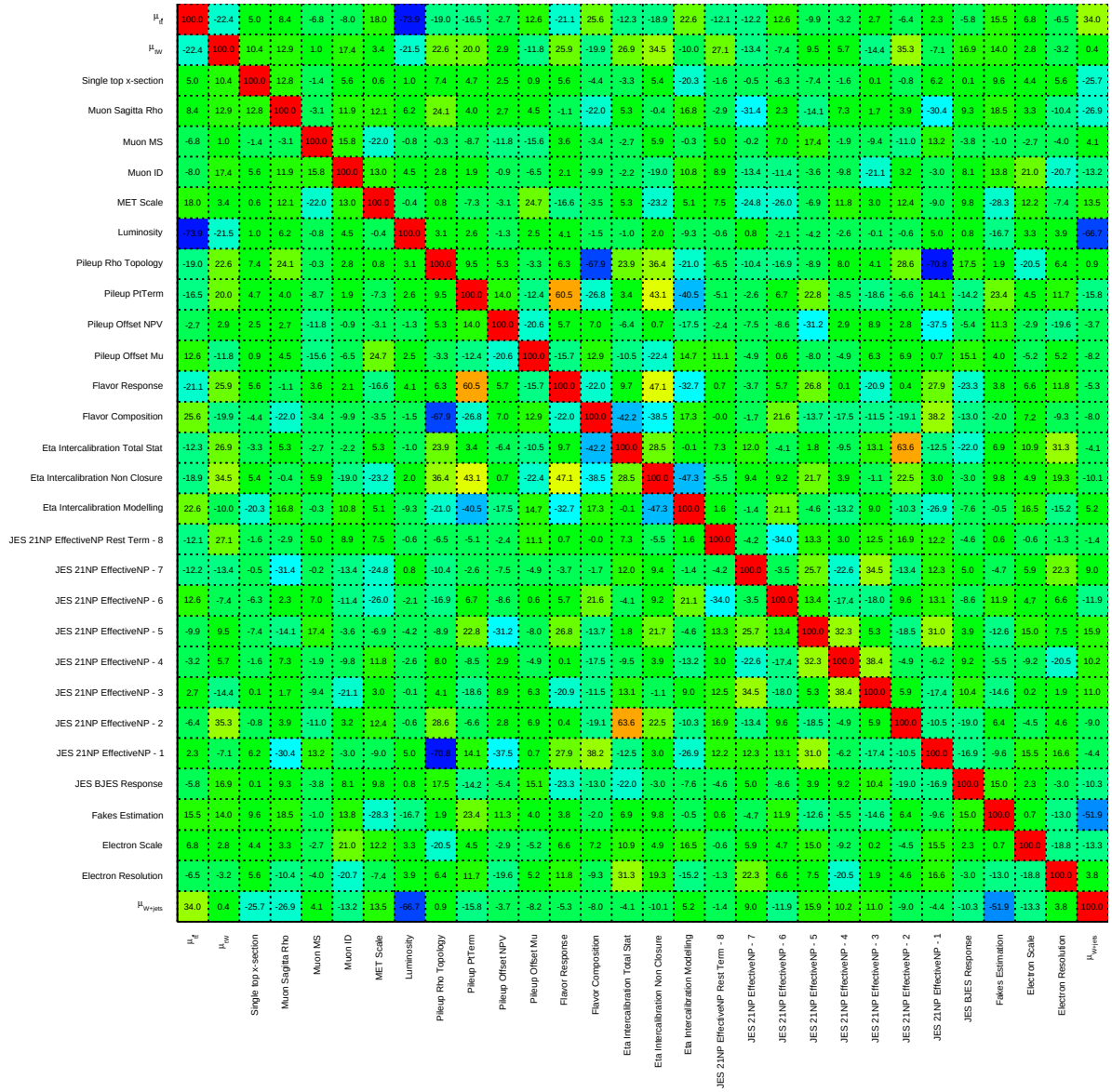


Figure 5.7: Correlation matrix for the $t\bar{t}$ only training.

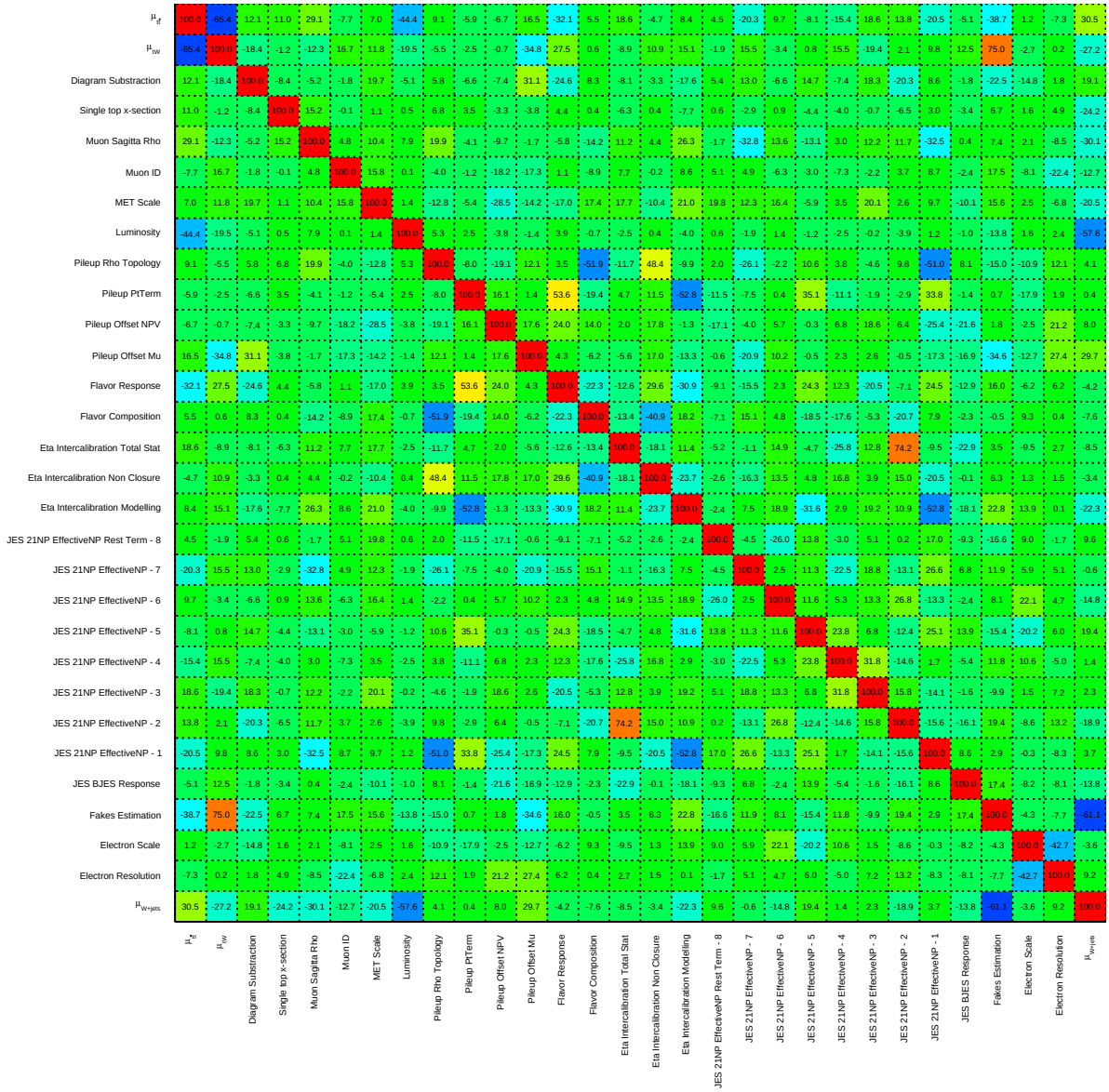


Figure 5.8: Correlation matrix for the $t\bar{t} + Wb$ training.

CHAPTER 6

Summary and Conclusion

To summarize, the tW decay channel is theorized to be sensitive to new physics that other single top-quark processes are not. Therefore, it is an interesting to study in conjunction to other single top-quark processes. In order to measure such productions, the signature of these processes needs to be found and reconstructed to its parent particles. In this analysis, the single lepton decay mode of the tW channel was measured whose final state is composed of one lepton with a neutrino (missing transverse energy), and three jets of which one must be a b-tagged jet. However, the measurement is not a simple task as the top-quark pair production is a dominant source of background and it has many similarities to the tW channel. This analysis puts heavy emphasis on separating signal from background by employing two training strategies for neural network. The basic strategy was to use topological and kinematic variables to differentiate between tW and $t\bar{t}$ by training on a constrained set of well reconstructed events. After the neural network applied its training to the datasets and the MC simulations, a two dimensional discriminant is built by using the information of the NN output and the mass of the hadronic decaying W boson. This discriminant is then mapped to a one-dimensional distribution for the purpose of performing a profile likelihood fit. The results for both trainings were compared as well as the influence systematic uncertainties had on them.

The measurement of the tW cross section was made using the data recorded in the years 2 015–2 016 with the ATLAS detector at the LHC from proton-proton collisions at a center-of-mass energy of 13 TeV. Moreover, this is the first measurement of the tW cross-section in the single lepton decay channel at 13 TeV. Two NN training methods were used to compare precision. The measured cross-section for both trainings are shown in table 6.1.

	$t\bar{t}$ Only	$t\bar{t} + Wb$
σ_{tW}(13 TeV)	84 ± 8	92 ± 13

Table 6.1: Measured cross-section results from both training strategies.

The measured cross-section, in both cases, fall slightly above the SM prediction. It should be taken into consideration that not all systematic uncertainties were taken into account when calculating the error. In particular, the modeling systematic uncertainties are not included in this analysis. Therefore, the resulting error is most likely underestimated as the missing sources of uncertainties were shown to have a major impact on past analyses. Furthermore, there are several improvements that could be made to refine the result. It is notable how the $t\bar{t}$ only training has a lower uncertainty to the $t\bar{t} + Wb$. It is possible that the second method requires a re-binning that spreads events more uniformly between the bins or

perhaps the ratio of Wb to $t\bar{t}$ events should be revisited; although training against $t\bar{t}$ only could be the strategy to follow after introducing more systematic uncertainties. The dominant uncertainty for the $t\bar{t}$ + Wb training was the fake estimation while the JES dominated the $t\bar{t}$ only training.

It is seen in other measurements that the dominating sources of uncertainty are experimental, in particular the JES. Given that these uncertainties are already mostly estimated in this analysis, the fact that the $t\bar{t}$ only training has about a 10 % uncertainty shows that this measurement is worth pursuing and completing as it has the potential to measure the tW channel with reasonable accuracy.

Currently, there are studies being made in machine learning where deep neural networks are being used to reduce the effect of systematic uncertainties. In an analysis like this where systematic uncertainties dominate, these techniques could be applied to improve the precision of this measurement.

Bibliography

[1] The ATLAS Collaboration, "Standard Model Total Production Cross Section Measurements", 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ (cit. on p. 4).

[2] Wikipedia, *The Standard Model — Wikipedia, The Free Encyclopedia*, [Online; accessed 08-01-2018], 2004, URL: https://en.wikipedia.org/wiki/Standard_Model (cit. on p. 4).

[3] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** (2016) 100001 (cit. on pp. 5, 20).

[4] H1 and ZEUS Collaborations, *Combination of Measurements of Inclusive Deep Inelastic $e^{\pm}p$ Scattering Cross Sections and QCD Analysis of HERA Data*, ArXiv e-prints (2015), arXiv: 1506.06042 [hep-ex] (cit. on p. 7).

[5] T. Loddenkötter, *Implementation of a kinematic fit of single top-quark production in association with a W boson and its application in a neural-network-based analysis in ATLAS*, BONN-IR-2012-06, PhD Thesis: University of Bonn, 2012, URL: http://hss.ulb.uni-bonn.de/diss_online (cit. on p. 10).

[6] L. Evans and L. Linssen, *The Super-LHC is on the starting blocks*, CERN Courier (2008) (cit. on p. 11).

[7] J. JPequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 12).

[8] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, ArXiv e-prints (2009), arXiv: 0901.0512 [hep-ex] (cit. on p. 13).

[9] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 13).

[10] *ATLAS magnet system: Technical Design Report, 1*, tech. rep., 1997, URL: https://cds.cern.ch/record/338080 (cit. on p. 15).

[11] ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, 2016, URL: https://cds.cern.ch/record/2133909 (cit. on p. 15).

[12] J. Pequenao and P. Schaffner, "An computer generated image representing how ATLAS detects particles", 2013, URL: https://cds.cern.ch/record/1505342 (cit. on p. 16).

[13] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: https://cds.cern.ch/record/2157687 (cit. on pp. 16, 24).

[14] ATLAS Collaboration, *ATLAS electron, photon and muon isolation in Run 2*, This note contains the Moriond 2017 recommendations. It will be updated when new recommendations become available., 2017, URL: https://cds.cern.ch/record/2256658 (cit. on pp. 17, 24).

[15] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, Journal of High Energy Physics **4**, 063 (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 17).

[16] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, Journal of Instrumentation **11** (2016) P04008, arXiv: 1512.01094 [hep-ex] (cit. on p. 17).

[17] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on p. 17).

[18] *Commissioning of the ATLAS b-tagging algorithms using $t\bar{t}$ events in early Run-2 data*, 2015, URL: https://cds.cern.ch/record/2047871 (cit. on p. 17).

[19] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, arXiv: 1603.09222 [hep-ex] (cit. on p. 21).

[20] S. Mergelmeyer, *Measurement of the Associated Production of a Single Top Quark and a W Boson in Single-Lepton Events with the ATLAS Detector*, BONN-IR-2016-03, Universität Bonn, 2016 (cit. on pp. 23, 28, 35, 37).

[21] R. Harrington, "ATLAS fast simulation and digitisation/reconstruction", Talk, 2014, URL: %7Bhttps://indico.cern.ch/event/279530/contributions/634994/attachments/511923/706532/LPCC_fastsim_robert.pdf%7D (cit. on p. 23).

[22] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, Journal of High Energy Physics **11**, 040 (2004) 040, eprint: hep-ph/0409146 (cit. on p. 23).

[23] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006** (2006) 026, URL: http://stacks.iop.org/1126-6708/2006/i=05/a=026 (cit. on p. 23).

[24] T. Gleisberg et al., *Event generation with SHERPA 1.1*, Journal of High Energy Physics **2**, 007 (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 23).

[25] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 23).

[26] ATLAS Collaboration, *Simulation of top-quark production for the ATLAS experiment at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2016-004, 2016, URL: https://cds.cern.ch/record/2120417 (cit. on p. 24).

[27] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS Detector*, ATLAS-CONF-2014-058, 2014, URL: https://cds.cern.ch/record/1951336 (cit. on pp. 24, 25).

[28] F. Derue, *Estimation of fake lepton background for top analyses using the Matrix Method with the 2015 dataset at $\sqrt{s}$=13 TeV with AnalysisTop-2.3.41*, internal report ATL-COM-PHYS-2016-198, CERN, 2016, URL: https://cds.cern.ch/record/2135116 (cit. on p. 25).

[29] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190 (cit. on p. 33).

[30] <phi-t> Physics Information Technologies, ed., *The NeuroBayes®User's Guide*, English, version October 11, 2013, 55 pp. (cit. on p. 34).

[31] ATLAS Collaboration, *Production of a top quark in association with a Z boson in the trilepton channel with ATLAS*, internal report ATL-COM-PHYS-2016-1758, CERN, 2016, URL: https://cds.cern.ch/record/2237442 (cit. on p. 35).

[32] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: https://cds.cern.ch/record/2037613 (cit. on p. 51).

[33] Luminosity Group, *Luminosity For Physics*, [Online; accessed 29-03-2018], 2018, URL: https://twiki.cern.ch/twiki/bin/view/Atlas/LuminosityForPhysics#2016_13_TeV_proton_proton_final (cit. on p. 51).

[34] Physics Modelling Group, *Top Focus Group*, [Online; accessed 29-03-2018], 2018, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopFocusGroup#Single_top_samples (cit. on p. 51).

[35] Physics Modelling Group, *Boson Jets Focus Group*, [Online; accessed 29-03-2018], 2018, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BosonJetsFocusGroup#Theory_uncertainties_on_the_gene (cit. on p. 51).

[36] Physics Modelling Group, *Multiboson Focus Group*, [Online; accessed 29-03-2018], 2018, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MultibosonFocusGroup#Systematic_uncertainties (cit. on p. 51).

APPENDIX A

Technical Details

The following section will give the technical details for the datasets and Monte Carlo simulations used. All of the sets given here are distributed by the SingleTop group at ATLAS.

Dataset Container
data15_13TeV.periodD.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data15_13TeV.periodE.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data15_13TeV.periodF.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data15_13TeV.periodG.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data15_13TeV.periodH.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data15_13TeV.periodJ.physics_Main.SGTOP1.grp15_v01_p2950.lj.v15
data16_13TeV.periodA.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodB.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodC.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodD.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodE.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodF.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodG.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodI.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodK.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15
data16_13TeV.periodL.physics_Main.SGTOP1.grp16_v01_p2950.lj.v15

Table A.1: List of the datasets containers used in this analysis.

Tables A.1 and A.2 detail of the n-tuples used in this analysis. The MC sets used are labeled as `v15` from Release 20.7 and only the sets containing one lepton in association with jets were used.

MC DSID	Tag	σ[pb]	k-Factor
Top-Quark Pair			
410000	e3698_s2608_s2183_r7725_r7676_p2952	377.9932	1.1949
Single Top-Quark, *t*-channel			
410011	e3824_s2608_s2183_r7725_r7676_p2952	43.739	1.0094
410012	e3824_s2608_s2183_r7725_r7676_p2952	25.778	1.0193

Single Top-Quark, tW Diagram Removal			
410013	e3753_s2608_s2183_r7725_r7676_p2952	34.009	1.054
410014	e3753_s2608_s2183_r7725_r7676_p2952	33.989	1.054
Single Top-Quark, tW Diagram Subtraction			
410062	e4132_s2608_s2183_r7725_r7676_p2952	32.384	1.054
410063	e4132_s2608_s2183_r7725_r7676_p2952	32.357	1.054
Single Top-Quark, s-channel			
410025	e3998_s2608_s2183_r7725_r7676_p2952	2.0517	1.005
410026	e3998_s2608_s2183_r7725_r7676_p2952	1.2615	1.022
Single Top-Quark, tZ			
410050	e4279_a766_a818_r7676_p2952	0.24013	1.0
410050	e4279_s2608_s2183_r7725_r7676_p2952	0.24013	1.0
W Boson in Association with Jets			
364156	e5340_s2726_r7772_r7676_p2952	15 770.0034	0.9702
364157	e5340_s2726_r7772_r7676_p2952	2 493.3784	0.9702
364158	e5340_s2726_r7772_r7676_p2952	844.19793	0.9702
364159	e5340_s2726_r7772_r7676_p2952	637.4241555	0.9702
364160	e5340_s2726_r7772_r7676_p2952	219.9656768	0.9702
364161	e5340_s2726_r7772_r7676_p2952	71.45937024	0.9702
364162	e5340_s2726_r7772_r7676_p2952	212.5554354	0.9702
364163	e5340_s2726_r7772_r7676_p2952	98.4371682	0.9702
364164	e5340_s2726_r7772_r7676_p2952	36.9147888	0.9702
364165	e5340_s2726_r7772_r7676_p2952	39.38245349	0.9702
364166	e5340_s2726_r7772_r7676_p2952	22.91781114	0.9702
364167	e5340_s2726_r7772_r7676_p2952	9.60864165	0.9702
364168	e5340_s2726_r7772_r7676_p2952	15.01	0.9702
364169	e5340_s2726_r7772_r7676_p2952	1.2344	0.9702
364170	e5340_s2726_r7772_r7676_p2952	15 769.63769	0.9702
364171	e5340_s2726_r7772_r7676_p2952	2 492.639	0.9702
364172	e5340_s2726_r7772_r7676_p2952	844.638035	0.9702
364173	e5340_s2726_r7772_r7676_p2952	630.3220976	0.9702
364174	e5340_s2726_r7772_r7676_p2952	215.4898229	0.9702
364175	e5340_s2726_r7772_r7676_p2952	97.7379615	0.9702
364176	e5340_s2726_r7772_r7676_p2952	202.8359871	0.9702
364177	e5340_s2726_r7772_r7676_p2952	98.4433455	0.9702
364178	e5340_s2726_r7772_r7676_p2952	36.9965304	0.9702
364179	e5340_s2726_r7772_r7676_p2952	39.24325044	0.9702
364180	e5340_s2726_r7772_r7676_p2952	22.846544	0.9702
364181	e5340_s2726_r7772_r7676_p2952	9.65665183	0.9702
364182	e5340_s2726_r7772_r7676_p2952	15.224	0.9702
364183	e5340_s2726_r7772_r7676_p2952	1.2334	0.9702
364184	e5340_s2726_r7772_r7676_p2952	15 799.4424	0.9702
364185	e5340_s2726_r7772_r7676_p2952	2 477.24902	0.9702
364186	e5340_s2726_r7772_r7676_p2952	854.554822	0.9702
364187	e5340_s2726_r7772_r7676_p2952	638.545523	0.9702

364188	e5340_s2726_r7772_r7676_p2952	210.3823406	0.9702
364189	e5340_s2726_r7772_r7676_p2952	98.018303	0.9702
364190	e5340_s2726_r7772_r7676_p2952	202.3332192	0.9702
364191	e5340_s2726_r7772_r7676_p2952	98.5776075	0.9702
364192	e5340_s2726_r7772_r7676_p2952	40.0623246	0.9702
364193	e5340_s2726_r7772_r7676_p2952	39.32514985	0.9702
364194	e5340_s2726_r7772_r7676_p2952	22.77896448	0.9702
364195	e5340_s2726_r7772_r7676_p2952	9.67021076	0.9702
364196	e5340_s2726_r7772_r7676_p2952	15.046	0.9702
364197	e5340_s2726_r7772_r7676_p2952	1.2339	0.9702
Z Boson in Association with Jets			
364100	e5271_s2726_r7772_r7676_p2952	1 630.2243	0.9751
364101	e5271_s2726_r7772_r7676_p2952	223.717472	0.9751
364102	e5271_s2726_r7772_r7676_p2952	127.1799342	0.9751
364103	e5271_s2726_r7772_r7676_p2952	75.0164716	0.9751
364104	e5271_s2726_r7772_r7676_p2952	20.3477432	0.9751
364105	e5271_s2726_r7772_r7676_p2952	12.3885125	0.9751
364106	e5271_s2726_r7772_r7676_p2952	24.28530322	0.9751
364107	e5271_s2726_r7772_r7676_p2952	9.2754186	0.9751
364108	e5271_s2726_r7772_r7676_p2952	6.01361075	0.9751
364109	e5271_s2726_r7772_r7676_p2952	4.77297475	0.9751
364110	e5271_s2726_r7772_r7676_p2952	2.265570784	0.9751
364111	e5271_s2726_r7772_r7676_p2952	1.491320988	0.9751
364112	e5271_s2726_r7772_r7676_p2952	1.7881	0.9751
364113	e5271_s2726_r7772_r7676_p2952	0.14769	0.9751
364114	e5299_s2726_r7772_r7676_p2952	1 627.176708	0.9751
364115	e5299_s2726_r7772_r7676_p2952	223.73136	0.9751
364116	e5299_s2726_r7772_r7676_p2952	126.4502953	0.9751
364117	e5299_s2726_r7772_r7676_p2952	76.292515	0.9751
364118	e5299_s2726_r7772_r7676_p2952	20.3360066	0.9751
364119	e5299_s2726_r7772_r7676_p2952	12.6227733	0.9751
364120	e5299_s2726_r7772_r7676_p2952	25.03001412	0.9751
364121	e5299_s2726_r7772_r7676_p2952	9.3719948	0.9751
364122	e5299_s2726_r7772_r7676_p2952	6.08263138	0.9751
364123	e5299_s2726_r7772_r7676_p2952	4.869231562	0.9751
364124	e5299_s2726_r7772_r7676_p2952	2.279979034	0.9751
364125	e5299_s2726_r7772_r7676_p2952	1.494370818	0.9751
364126	e5299_s2726_r7772_r7676_p2952	1.8081	0.9751
364127	e5299_s2726_r7772_r7676_p2952	0.14857	0.9751
364128	e5307_s2726_r7772_r7676_p2952	1 627.725872	0.9751
364129	e5307_s2726_r7772_r7676_p2952	223.881432	0.9751
364130	e5307_s2726_r7772_r7676_p2952	127.7329554	0.9751
364131	e5307_s2726_r7772_r7676_p2952	76.0261671	0.9751
364132	e5307_s2726_r7772_r7676_p2952	20.212279	0.9751
364133	e5307_s2726_r7772_r7676_p2952	12.29393	0.9751
364134	e5307_s2726_r7772_r7676_p2952	24.80341201	0.9751

364135	e5307_s2726_r7772_r7676_p2952	9.3282378	0.9751
364136	e5307_s2726_r7772_r7676_p2952	5.47909362	0.9751
364137	e5307_s2726_r7772_r7676_p2952	4.791190072	0.9751
364138	e5313_s2726_r7772_r7676_p2952	2.275625215	0.9751
364139	e5313_s2726_r7772_r7676_p2952	1.502837652	0.9751
364140	e5307_s2726_r7772_r7676_p2952	1.8096	0.9751
364141	e5307_s2726_r7772_r7676_p2952	0.14834	0.9751
Diboson			
361600	e4054_s2608_s2183_r7725_r7676_p2669	10.631	1.0
361601	e4054_s2608_s2183_r7725_r7676_p2952	4.4625	1.0
361602	e4054_s2608_s2183_r7725_r7676_p2669	2.7778	1.0
361603	e4054_s2608_s2183_r7725_r7676_p2952	1.2568	1.0
361604	e4054_s2608_s2183_r7725_r7676_p2669	0.92498	1.0
361606	e4054_s2608_s2183_r7725_r7676_p2669	44.176	1.0
361607	e4054_s2608_s2183_r7725_r7676_p2669	3.2849	1.0
361609	e4054_s2608_s2183_r7725_r7676_p2669	10.099	1.0
361610	e4054_s2608_s2183_r7725_r7676_p2669	2.2739	1.0

Table A.2: Detailed list of the MC sets used in this analysis.

APPENDIX B

Details on Systematic Uncertainties

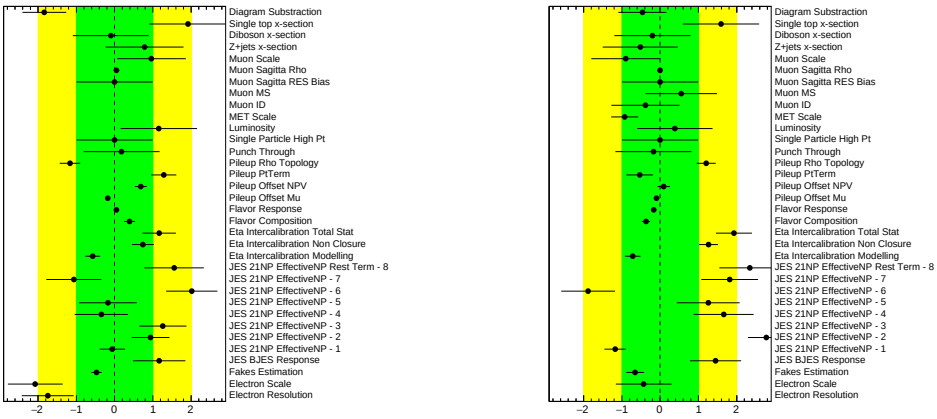
Table B.1 shows the impact that each systematic has on select processes. Other processes are excluded as their contributions are very small by comparison to the main backgrounds shown in the table. Values at zero denote a contribution that was less than 1×10^{-15}.

Systematic	Effect on Process [%]		
	tW	$t\bar{t}$	W + jets
Diagram Substraction	±3.07	0	0
Luminosity	±2.2	±2.2	±2.2
JES BJES Response	±0.0763	±0.228	±0.309
JES 21NP EffectiveNP - 1	±0.769	±2.9	±2.95
JES 21NP EffectiveNP - 2	±0.206	±0.737	±0.565
JES 21NP EffectiveNP - 3	±0.018	±0.136	±0.0719
JES 21NP EffectiveNP - 4	±0.0181	±0.091	±0.0381
JES 21NP EffectiveNP - 5	±0.032	±0.0113	±0.199
JES 21NP EffectiveNP - 6	±0.0168	±0.137	±0.252
JES 21NP EffectiveNP - 7	±0.0453	±0.207	±0.0472
JES 21NP EffectiveNP Rest Term - 8	±0.0283	±0.0318	±0.15
Eta Intercalibration Modelling	±0.73	±1.84	±2.03
Eta Intercalibration Non Closure	±0.288	±0.327	±0.118
Eta Intercalibration Total Stat	±0.184	±0.636	±0.79
Flavor Composition	±1.63	±5.19	±6.15
Flavor Response	±0.41	±1.52	$^{+2.02}_{-1.71}$
Pileup Offset Mu	±0.0654	±0.324	±0.498
Pileup Offset NPV	±0.279	±0.737	±0.73
Pileup PtTerm	±0.0807	±0.106	±0.583
Pileup Rho Topology	±1.01	±4.04	±3.85
Punch Through	$\pm 1.41 \times 10^{-3}$	$\pm 1.82 \times 10^{-3}$	$\pm 1.28 \times 10^{-3}$
Single Particle High Pt	0	0	$+3.6 \times 10^{-14}$
Electron Resolution	$\pm 6.23 \times 10^{-3}$	$\pm 4.41 \times 10^{-3}$	±0.0241
Electron Scale	±0.107	±0.0985	±0.226
MET Scale	±0.37	±0.204	±0.695
Muon ID	$\pm 6.53 \times 10^{-3}$	$\pm 5.89 \times 10^{-3}$	±0.0226
Muon MS	$\pm 4.19 \times 10^{-3}$	$\pm 2.96 \times 10^{-3}$	±0.0209
Muon Scale	±0.0258	±0.0351	±0.0805
Muon Sagitta RES Bias	0	0	0
Muon Sagitta Rho	0	0	0

Table B.1: Relative effect of each systematic on the yields.

APPENDIX C

Supplemental Fit Results



(a) $t\bar{t}$ only.

(b) $t\bar{t}$ + Wb.

Figure C.1: Comparison of the nuisance parameters after the fit between training strategies.

List of Figures

List of Tables